TIME SERIES MODELLING WITH APPLICATION TO SOUTH AFRICAN INFLATION DATA

By

AMOS CHINOMONA
(Hons. Statistics)

Submitted in fulfillment of a the academic requirements for the degree of

MASTER OF SCIENCE
in
Statistics

in the
School of Statistics and Actuarial Science
University of KwaZulu Natal
Pietermaritzburg

2009

Time series modelling with application to South African inflation data.

Amos Chinomona

March 19, 2010

Declaration

I declare that this research was carried out in accordance with the regulations of the School of Statistics and Actuarial Science, University of KwaZulu-Natal, Pietermaritzburg from January 2008 to December 2009 under the supervision of Prof. Henry Mwambi.

The research work represents original work by the author, and where use has been made of the work of others it is duly acknowledged by special reference in the text. Any views expressed in the dissertation are those of the author and in no way represent those of the University of KwaZulu-Natal. The work has not otherwise been submitted in any form for any degree or diploma to any University.

Signed:..... Date.....
Amos Chinomona (Student)

Signed:..... Date.....
Prof. Henry Mwambi (Supervisor)

Acknowledgement

This thesis would not have been possible if it were not for the tireless guidance and support I got from my supervisor Prof. Henry Mwambi. To you Professor I say thank you. To my sister Joanah and your family, I owe you my deepest gratitude for your support and assistance, you stood by me every step of the way. Many thanks go to the School of Statistics and Actuarial Science for allowing me the opportunity to work on this topic. I would like to thank the South African Reserve Bank for giving me the permission to use the South African inflation rate data for my thesis.

I am indebted to my colleagues, Dikokole, Phathisani, Mohammed, Morris and Makhala for your tremendous support. I would like to express my sincere gratitude to my fiancee Perpetua and my nephew Tariro for their unwavering support and encouragement.

To my parents, Agnes and Steven, thank you for your unconditional support. Lastly, it is my pleasure to thank those not mentioned by names but whose support made this thesis a success.

Abstract

The research is based on financial time series modelling with special application to modelling inflation data for South Africa. In particular the theory of time series is explored and applied to the inflation data spanning from January 1994 to December 2008. The data was obtained from the South African Reserve Bank.

Two families of time series models namely, the autoregressive integrated moving average (ARIMA) (with extension to seasonal ARIMA (SARIMA)) models and the autoregressive conditional heteroscedastic (ARCH) (with their extensions to generalized ARCH, (GARCH)) models were fitted to the data. Time series model building strategies as postulated by Box-Jenkins (1976) are explored in detail from a theoretical and practical stand point. The stages in the model building namely, identification, estimation and checking are explained explicitly and applied to the data. A best fitting model for each family of models was selected based on how well the model captures the variations in the data (goodness of fit). The goodness of fit is assessed via the Akaike information criteria (AIC), Bayesian information criteria (BIC) and the coefficient of determination, R^2.

A seasonal ARIMA model, SARIMA$(1,1,0)\times(0,1,1)_s$ was chosen to be the best fitting from the ARIMA family of models, while the GARCH(1,1) was chosen to be the best fitting from the ARCH-GARCH models. The selected models were used to compute two year forecasts for inflation using time series forecasting methods. The accuracy of the forecast were assessed via the forecast confidence width. Then comparisons of the

two selected models was carried out based on the goodness of fit and the forecasting power of the two models. It was established that the GARCH(1,1) model was superior to the SARIMA$(1,1,0)\times(0,1,1)$ model according to both criteria. This was especially so because the data was characterized by changing mean and variance, a characteristic that makes ARIMA modelling less appropriate than the ARCH-GARCH modelling.

Analysis and write-up were done using SAS version 9.1 and latex respectively. Presentation and explanations of results were aided by the use of graphs and tables.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Inflation modelling is one of the most important research area in monetary planning. To start with we state a quote from Kohn, (2005): "Nothing is more important to the conduct of monetary policy than understanding and predicting inflation. Achieving and maintaining price stability will be more efficient and effective the better we understand the causes of inflation and the dynamics of how it evolves". Monetary policy-makers worldwide are more interested in containing and reducing inflation through price stability. Moreover, so long as inflation expectations are well anchored, we can tolerate limited changes in inflation, but we need to know that a rise or fall is not the beginning of a more extended trend. Consequently, we focus closely on the reasons for any changes in inflation and their implications for the outlook.

Financial modelling defined as 'creation of representation of reality' is one way of trying to understand the underlying dynamics of inflation and how it evolves (www.statisticalforecasting.com). Models for finance and economics are heavily influenced by time, through both time reso-

lution and time horizon. The concept of resolution signifies how densely data are recorded varying from seconds to years and time horizon looks at the length of time the data spans. Financial analysis usually involves a study of price movement, usually given over time (financial time series), hence financial modelling focuses on building mathematical and statistical models to capture the price movements and the variations in the prices over time.

Inflation is usually defined as a sustained rise in a broadly based index of commodity prices over some period of time, (Bomberger and Makinen, 1979). The essential ingredients of the definition are that the rise in prices takes place in a broad group of goods and services, not in one or two commodities, and that the increase continues for a rather lengthy period rather than for one or two quarters. When the price level rises, each unit of currency buys fewer goods and services; consequently, inflation is also an erosion in the purchasing power of money, that is a loss of real value in the internal medium of exchange and unit of account in the economy (Stokes, 2009).

The most common way of measuring inflation is by reference to consumer price index (CPI), which is an annualized percentage change that measures the changes in prices of a basket of goods and services purchased by a representative set of households. In South Africa the CPI is calculated by the Statistics South Africa under the auspices of the Ministry of Finance. The CPI is a dynamic figure obtained by measuring the price movements of numerous goods and services consumed by a typical South African household on a monthly basis. Specific weights

are assigned to categories of expenditure based on consumer spending statistics. The different types of expenditure are grouped together in a basket and the high volume spending items carry the most weight and therefore have the most material impact on the calculated index.

The main shortfall of using CPI is that, it's a narrow measure of inflation and does not take into account other goods and services such as those used for intermediate consumption or the prices of capital products. An alternative measure is the gross domestic product (GDP) deflator, which is derived from the GDP measured in current prices and GDP measured in constant prices, that is prices in the base year. This is a weighted average of the prices of:

- goods and services consumed by households
- expenditure by government on goods, services and salaries
- fixed capital assets
- changes in inventories
- exports of goods and services
- imports of goods and services

Inflation dynamics and evolution can be studied using a stochastic modelling approach that captures the time dependent structure embedded in the time series of inflation as a stochastic process. ARIMA/Box Jenkins models are important in complementing financial modelling by describing the component structure of statistical time series especially to many financial time series that show seasonal behavior, random changes

and trends (non-stationary) time series. Stochastic volatility models, autoregressive conditional heteroscedastic (ARCH) and generalized ARCH (GARCH) models are also other approaches used to capture and model the volatility behaviour of financial time series.

Monetary policy in South Africa went through several regime and other changes in the past quarter century (approximately since 1980). The country's (through the South African Reserve Bank (SARB)) recent adoption of the explicit inflation targeting and protection of the internal and external value of the Rand increase the need for good forecasting models of inflation and models for understanding the monetary transmission mechanism (Aron and Muellbauer, 2004). Currently, monetary policy in South Africa is aimed at reducing inflation and improving international competitiveness. An understanding of the forces driving and perpetuating inflation in South Africa should shed some light on the Reserve Bank's control over the inflation process and the relative efficacy of alternative monetary policy frameworks and instruments, (Akinboade et al., 2001).

During the 1990s the SARB pursued an unofficial (implicit) inflation target regime. This did not strictly give a target range for inflation, but rather advocated for a floating or flexible way of controlling inflation. Although, compared to the 1990s, the explicit inflation targeting regime might have been marginally more successful in keeping inflation at lower levels, the inflation targeting system still displays limited success in keeping inflation within target range, (Burger and Marinkov, 2008). This is apparent when considering that since the implementation of inflation

targeting, the CPI exceeded its 6% upper bound in 29 of the 72 months between January 2002 (the first month of the first year that inflation had to be within its target range) and December 2007.

It is against this background that this research focuses on understanding the dynamics of the underlying inflation generating mechanism as a stochastic process. Mboweni, (2007) the former governor of the South African Reserve Bank applauded the ARIMA models in predicting inflation saying they are momentum-type models with no underlying economic theory. Meyler, Kenny and Quinn (1998) also used ARIMA models to forecast Irish inflation. The three placed emphasis on forecast performance based on minimizing 'out of sample forecast errors' rather than on maximizing 'in-sample goodness of fit'. Muhammed et al., (2005) found out that on the basis of both in-sample and out-of- sample forecast it can be concluded that their model has sufficient predictive powers to forecast inflation in Pakistan. Engle, (1982) used the ARCH processes to model inflation rates recognizing that the uncertainty inherent in inflation tends to change with time. Talke (2003) in his MSc research also applied ARCH models to modelling volatility in time series data. In his work Talke, (2003) applied the ARCH methods to changes in exchange rates of the South African rand against the US dollar, Swiss and the British pound. Rangan and Josine, (2009) used cosine-squared cepstrum to provide evidence that inflation in South Africa has become more volatile since the first quarter of 2000, when the country moved into an inflation targeting regime, than it would have been had the SARB continued with the more eclectic monetary policy approach pursued in the pre-target era. The current application will be based on the South African inflation data and

its associated variations and volatility over time.

1.1 Significance of the study

It is generally accepted that over the medium to long term, inflation is a monetary phenomenon, that is, it is entirely determined by the monetary policy. Over shorter horizons, however, various micro-economic shocks, including variations in economic activity have a bearing on the inflation. Therefore, a profound understanding of the inflation-generating process, in particular, the speed of inflation adjustment in response to such shocks is of crucial importance for a central bank whose policy is oriented towards price stability. The best way to study and understand the underlying phenomenon of the inflation process is to construct practical statistical models that account for the variation and volatility in the process. Policies to contain inflation are necessary for transition economies to grow and firms to restructure. Furthermore, a change in inflation is a response to intertwined change in macro-economic fundamentals such as prices, wages, money supply and exchange rates. In particular, by enhancing the transparency of the monetary policy, the inflation targeting regime is expected to provide a better anchor for inflation expectations and hence lead to favourable impacts on price-setting and wage-setting behaviour in the economy. Containment of inflation is useful to an open and small economy such as South Africa for a variety of reasons, among which are:

- to provide important information about the likely impact of shocks on the economy over time especially for the purpose of setting the monetary policy.

- to assist in the identification of the underlying pressures in the economy.

- to shed some light on the performance of the monetary policy under different policy regimes.

1.2 Objectives of study

The main objectives of our study are to:

- use time series analysis to study and understand the variations and volatility in inflation trend as a stochastic process in South Africa after 1994. Specifically, we shall use the time series idea that observations that are close together in time are more correlated or similar than observations that are far apart.

- develop time series models (ARIMA and ARCH models) that will be used to explain the time dependent structure embedded in the inflation process. This will be achieved by studying and understanding the underlying causes or mechanisms of the variation and volatility in the series.

- compare the models to find the best fit model using time series model selection techniques.

- demonstrate how to use the model selected to compute forecasts of future inflation rate.

The thesis is arranged in the following way: Chapters 2 and 3 focus on general theory of Box-Jenkins ARIMA models. Specifically, the theory

on identification of different models, model estimation, model checking and validation and forecasting are explained in detail in these chapters. Chapter 4 looks at the theory of the ARCH-GARCH models. The concept of heteroscedasticity and how the models are constructed is given in detail. Chapter 5 deals with the application of the two family of models, that is, the ARIMA and the ARCH-GARCH, to the inflation rate data. Steps in identifying the models, estimating the model parameters and validating the selected models are well followed. Model validation is done using residual analysis. Analysis is done using SAS version 9.1 procedures. Chapter 6 focuses on comparing the selected ARIMA and selected ARCH models using the best fit criteria and their forecasting power from both a theoretical and practical point of view. Conclusions, recommendations and areas of further research are also presented in this chapter.

Chapter 2

ARIMA modelling

2.1 Introduction

A time series is a series or sequence of data points measured typically at successive times. These data points are commonly equally spaced in time (Chatfield, 2004). It should be noted that time series analysis and longitudinal data analysis are fundamentally different although they share some common properties, (Diggle et al., 2002). In the latter, a sample of individuals (N) are recruited into a study e.g a clinical trial, then repeated measurements (n) are taken on the same subjects. Thus in this case, we have many short time series generated by each individual. In time series we have one long series of observations such as in the current study. Time series analysis comprises methods that attempt to understand such time series, often to understand the underlying generation process of the data points and construct a mathematical formula, (or a model) to represent the process. The constructed model is used to forecast future events based on known past events, that is, to forecast future data points before they are realized. Prediction on the other hand

is concerned with using a time series model to obtain data points within a realized time period.

A time series model will usually or should reflect the fact that observations close together in time domain are more correlated than observations further apart. When modelling variations in the level of a process realized over time, there are three main classes of models that are of importance. These are:

- Autoregressive (AR) structure which is defined as a stochastic process that assumes that the current data can be described by a weighted sum of its previous values and an error term.
- Moving average (MA) model defined as a type of finite impulse response filter process used to analyze a set of data points by creating a series of averages of different subsets of the full data set.
- Finally, an integrated time series is obtained by taking difference of successive observations.

As already mentioned earlier, time series modelling falls under a wider class of processes called stochastic processes. These are processes in which there is an indeterminacy in the future evolution of the process described by probability distributions. This means that even if the initial conditions (or starting point) is known, there are many possible paths that the process might take, but some paths are more probable and plausible than others. An observed time series is just one example of the infinite set of time series which can be treated as a simple case of a stochastic process and the infinite set is sometimes called an ensemble.

A useful way of describing a stochastic process is to give the moments of the process particularly the first and second moments, that is the mean and the autocovariance function (ACVF). Let $\{y_t\}$ be a random variable that is time dependent, and thus y_t will denote the observed time series. The mean function μ_t is defined for all t by $\mu_t = E[y_t]$. The variance function σ_t^2 is defined for all t by $\sigma_t^2 = E[(y_t - \mu_t)^2]$. Furthermore the autocovariance function (ACVF) denoted by $\gamma(t_1, t_2)$ is the covariance between y_{t_1} and y_{t_2} given by

$$\gamma(t_1, t_2) = E[(y_{t_1} - \mu(t_1))(y_{t_2} - \mu(t_2))].$$

The variance function is a special case of the ACVF when $t_1 = t_2$. When analyzing time series data an assumption of stationarity is usually made and this is defined and described in section (2.2) below.

2.2 Stationary process

A time series process is said to be strictly stationary if the joint distribution of $y_{t_1}, ..., y_{t_k}$ is the same as the joint distribution of $y_{t_1+\tau}, ..., y_{t_k+\tau}$ for all $t_1, ..., t_k, \tau$. That is, shifting the time origin by τ has no effect on the joint distribution which must therefore depend only on the intervals between the two set of points given by τ which is called the lag. A time series process is also called weak (or second order) stationary when the requirements of strict stationarity are only applied to two pairs of random variables from the time series. Formerly, a time series process is said to be second stationarity if the first and second order density function satisfies:

$$f_y(y_1 : t_1) = f_y(y_1 : t_1 + \tau) \tag{2.1}$$

and

$$f_y(y_1, y_2 : t_1 t_2) = f_y(y_1, y_2 : t_1 + \tau, t_2 + \tau) \tag{2.2}$$

for all t_1,t_2 and τ.

If $k = 1$, strict stationarity implies that the marginal distribution of y_t is the same for all t so that provided the first two moments are finite, we have $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$ which do not depend on the value of t. Furthermore if $k = 2$ the joint distribution of y_{t_1} and y_{t_2} depends only on the time difference $t_2 - t_1 = \tau$, implying strict stationarity. Thus the ACVF $\gamma(t_1, t_2)$ also depends only on $\gamma(t_2 - t_1)$ since $\gamma(t_1, t_2)$ is given by

$$\begin{aligned} \gamma(\tau) &= E[y_t - \mu][y_{(t+\tau)} - \mu] \\ &= Cov[y_t, y_{t+\tau}] \\ &= \gamma(\tau) \end{aligned} \tag{2.3}$$

which is called the autocovariance coefficient at lag τ. The autocorrelation function (ACF) is defined by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

which is a measure of the correlation between y_t and $y_{t+\tau}$ and it is an important function for identifying the order of an ARMA model as will be discussed in chapter 3. A process is said to satisfy second order stationarity if its mean is constant and its ACVF depends only on the lag, so that

$$E[y_t] = \mu$$

$$Cov[y_t, y_{t+\tau}] = \gamma(\tau)$$

If $\tau = 0$, then the mean and the variance are constant, implying strict stationarity. Application of Box-Jenkins ARIMA modelling assumes the series is stationary. In the event that the series exhibits nonstationarity, appropriate transformations, as given in detail in section (2.5), have to be applied to make the series stationary. Alternatively, other approaches such as smoothing, in the form of for example exponential smoothing may also be used to transform the data. We present the properties of the ACF of a stationary stochastic process.

2.2.1 The ACF and its properties

Suppose a stationary stochastic process $\{y_t\}$ has mean μ and variance σ^2, then the ACF at lag τ is defined as

$$\begin{aligned} \frac{Cov(y_t, y_{t+\tau})}{Var(y_t)} &= \frac{\gamma(\tau)}{\gamma(0)} \\ &= \rho(\tau) \end{aligned} \tag{2.4}$$

Note $\rho(\tau) = 1$ if $\tau = 0$. The plot of the ACF over time is used to identify the order and degree of nonstationarity of an ARIMA model.

Property 1

The ACF is an even function of its lag, that is,

$$\begin{aligned} \rho(\tau) &= \rho(-\tau) \\ &= \frac{Cov(y_{t-\tau}, y_t)}{Var(y_{t-\tau})} \end{aligned} \tag{2.5}$$

This means that correlation between y_t and $y_{t+\tau}$ is the same as that between y_t and $y_{t-\tau}$

Property 2

$|\rho(\tau)| \leq 1$, that is, the ACF lies between -1 and +1
Combining the two results gives $|\rho(\tau)| \leq 1$ as required.

Property 3

The ACF does not uniquely determine the underlying model. Although a given stochastic process has a unique covariance structure, the converse is not in general true. That is a unique covariance structure is not a guarantee for the existence of a stochastic process. It is generally possible to find many normal and non-normal processes with the same ACF. As a consequence, in model identification, invertibility (see section 2.3.3) must be ensured in order to uniquely identify a given model. For proofs of properties 1, 2 and 3 see Chatfield (2004) page 36.

2.3 Some basic useful models

This section describes a series of different types of stochastic processes, from simple to complex in a building block fashion that may be appropriate when setting up a model for a time series. The basis or criteria for identifying a model is the ACF and the PACF.

2.3.1 Purely random process

A time series is said to be purely random if it consist of a sequence of random variables, ε_t which are mutually independent and identically distributed with constant mean and variance σ_ε^2. From this definition it follows that the process has constant mean and variance, and the independent assumption means that

$$\gamma(k) = Cov(\varepsilon_t, \varepsilon_{t+k}) = \begin{cases} \sigma_\varepsilon^2 & \text{if } k = 0; \\ 0 & \text{if } k = \pm 1, \pm 2, ... \end{cases}$$

and therefore the autocorrelations are given by

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0; \\ 0 & \text{if } k = \pm 1, \pm 2, ... \end{cases}$$

Because the mean and the variance do not depend on time, the process is said to be second order stationarity. Further, the independence assumption implies that the process is strictly stationary.

2.3.2 Random Walk

Suppose we now define ε_t as an independent random process with mean μ and the variance σ_ε^2. Then a process $\{y_t\}$ is said to be a random walk if

$$y_t = y_{t-1} + \varepsilon_t. \tag{2.6}$$

Such a process is usually assumed to start at zero when $t = 0$ so that $y_1 = \varepsilon_1$ and in general

$$y_t = \sum_{i=1}^{t} \varepsilon_i$$

Thus we deduce that $E(y_t) = t\mu$ and $Var(y_t) = t\sigma_\varepsilon^2$. The process is non-stationary because the mean and variance depends on time t. However the first order difference of a random walk given by

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \tag{2.7}$$

forms a purely random process, which is stationary. The best known examples of time series which behave like random walks are share prices on successive days, that is, share price on day t is equal to price on day $t-1$ plus random error.

2.3.3 Moving Average (MA) process

Suppose that ε_t is a purely random process with mean zero and variance σ_ε^2, then a process $\{y_t\}$ is said to be a moving average process of order q, (MA(q)) if

$$y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q} \tag{2.8}$$

where the $\beta_i's$ are constants. It immediately follows that

$$E[y_t] = 0$$

$$Var(y_t) = \sigma_\varepsilon^2 \sum_{i=0}^{q} \beta_i^2$$

since ε's are independent. The εs are scaled so that β_0=1. We also have

$$\begin{aligned} \gamma(k) &= Cov(y_t, y_{t+k}) \\ &= Cov(\beta_0\varepsilon_t + ... + \beta_q\varepsilon_{t-q}, \beta_0\varepsilon_{t+k} + ... + \beta_q\varepsilon_{t+k-q}) \end{aligned} \tag{2.9}$$

$$\gamma(k) = \begin{cases} 0 & \text{if } |k| > q; \\ \sigma_\varepsilon^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & \text{if } k = 0, 1, 2, ..., q; \\ \gamma(-k) & \text{if } -q < k < 0 \end{cases}$$

As $\gamma(k)$ does not depend on t and the mean is a constant, the process satisfies second order stationarity for all values of the β_i. The ACF of the MA(q) process is given by

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0; \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2} & \text{if } k = 0, 1, 2, ..., q; \\ \rho(-k) & \text{if } -q < k < 0; \\ 0 & \text{if } k > q. \end{cases}$$

Note that the ACF cuts off at lag q which is a special feature of MA processes, and in particular, the MA(1) process with $q = 1$ has an ACF given by

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0; \\ \frac{\beta_1}{1+\beta_1^2} & \text{if } k = \pm 1; \\ 0 & \textit{otherwise}. \end{cases}$$

A restriction is always imposed on the β_i to ensure that the process satisfies a condition called invertibility. This condition may be explained by use of an example. Consider the following first order MA processes

$$\begin{aligned} A : y_t &= \varepsilon_t + \theta \varepsilon_{t-1} \\ B : y_t &= \varepsilon_t + \frac{1}{\theta} \varepsilon_{t-1} \end{aligned} \tag{2.10}$$

It can be shown that the two processes are different but have the same ACF, that is $\frac{\theta}{1+\theta^2}$. Thus we cannot identify uniquely an MA process from a given ACF. If we invert models A and B in equation (2.10) by expressing ε_t in terms of $y_t, y_{t-1}, ...$ we get

$$\begin{aligned} A : \varepsilon_t &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - ... \\ B : \varepsilon_t &= y_t - \frac{1}{\theta} y_{t-1} + \frac{1}{\theta^2} y_{t-2} - ... \end{aligned} \tag{2.11}$$

If $|\theta| < 1$ the series of the coefficients of y_{t-j} for model A in equation (2.11) converges whereas that of B does not. Thus generally a process $\{y_t\}$ is said to be invertible if the random disturbance at time t, sometimes called the innovation can be expressed as a convergent sum of past and present values of y_t in the form

$$\sum_{j=0}^{\infty} \Pi_j y_{t-j}$$

where $\sum_{j=0}^{\infty} |\Pi_j| < \infty$, that is the process can be written in the form of an autoregressive (AR) process, possibly of infinity order but whose coefficient form a convergent sum. The AR process is described in section (2.3.4) below. The imposition of the invertible condition ensures there is a unique MA process for a given ACF.

The invertibility condition for an MA process of any order is best expressed by using a backward shift operator denoted by **B** which is defined by
$\mathbf{B}^j y_t = y_{t-j}$ for all j. For example an MA process of order q, that is an MA(q), given by

$$y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + ... + \beta_q\varepsilon_{t-q} \tag{2.12}$$

can be written using the shift operator as

$$\begin{aligned} y_t &= (\beta_0 + \beta_1\mathbf{B} + \beta_2\mathbf{B}^2 + ... + \beta_q\mathbf{B}^q)\varepsilon_t \\ &= \Theta\mathbf{B}\varepsilon_t \end{aligned} \tag{2.13}$$

where $\Theta\mathbf{B}$ is a polynomial of order q in $\mathbf{B}$. It can be shown that an MA(q) process is invertible if the roots of the equation

$$\Theta\mathbf{B} = \beta_0 + \beta_1\mathbf{B} + \beta_2\mathbf{B}^2 + ... + \beta_q\mathbf{B}^q = 0 \tag{2.14}$$

all lie outside the unit circle, where we regard $\mathbf{B}$ as a complex variable and not as an operator.

2.3.4 Autoregressive (AR) process

Suppose that ε_t is a purely random process with mean zero and variance σ_ε^2, then a process $\{y_t\}$ is said to be an AR process of order p, that is, AR(p) if

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \varepsilon_t \tag{2.15}$$

If we let

$$H_t = \{y_{t-1}, ..., y_{t-p}\},$$

then we can express the AR(p) model as a conditional mean given by

$$E(y_t|H_t) = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p}, \tag{2.16}$$

where H_t is the history at time t. It is like a multiple regression model, where y_t is regressed on past values of itself which explains the prefix 'auto'.

First order AR process

The process $\{y_t\}$ in equation (2.15) is said to be an AR process of order 1 if it is given by

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{2.17}$$

which is sometimes called the Markov first order property of the AR. Here the history at time t is given by $H_t = \{y_{t-1}\}$. By successive substitution we may write

$$\begin{aligned} y_t &= \alpha(\alpha y_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\ &= \alpha^2(\alpha y_{t-3} + \varepsilon_{t-2}) + \alpha\varepsilon_{t-1} + \varepsilon_t \end{aligned} \tag{2.18}$$

and hence eventually we find that y_t may be expressed as an infinite-order MA process in the form

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + ... \tag{2.19}$$

where $\beta_i = \alpha^i$ provided $-1 < \alpha < 1$ to ensure convergence. Thus there is a link between an MA and AR processes. To show the link easily we can use the backward shift operator $\mathbf{B}$, on equation (2.17) and obtain

$$(1 - \alpha\mathbf{B})y_t = \varepsilon_t \tag{2.20}$$

implying

$$\begin{aligned} y_t &= \frac{\varepsilon_t}{1 - \alpha\mathbf{B}} \\ &= (1 + \alpha\mathbf{B} + \alpha^2\mathbf{B}^2 + ...)\varepsilon_t \\ &= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + ... \end{aligned} \tag{2.21}$$

When it is expressed in this format, it is easy to see that

$$E(y_t) = 0$$

and the variance is given by

$$\begin{aligned} Var(y_t) &= \sigma_\varepsilon^2(1 + \alpha^2 + \alpha^4 + ...) \\ &= \frac{\sigma_\varepsilon^2}{1 - \alpha^2} \\ &= \sigma_y^2. \end{aligned} \tag{2.22}$$

Thus the variance is finite provided that $|\alpha| < 1$ which is the convergence condition of the infinite MA series given in equation (2.19). Thus since $E(y_t) = 0$ the ACVF is given by

$$\begin{aligned} \gamma_k &= E[y_t y_{t+k}] \\ &= E[(\sum \alpha^i \varepsilon_{t-i})(\sum \alpha^j \varepsilon_{t+k-j})] \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \\ &= \frac{\alpha^k \sigma_\varepsilon^2}{1 - \alpha^2} \\ &= \alpha^k \sigma_y^2. \end{aligned}$$

It can easily be shown that $\gamma_k = \gamma_{-k}$ since γ_k does not depend on t. Thus an AR process of order 1 is second-order stationary provided that $|\alpha| < 1$, and the ACF is then given by

$$\rho(k) = \alpha^k, \text{ for } k = 0, 1, 2, ...$$

General AR process

Now consider the general AR(p) process stated in equation (2.15). Then an AR process of any finite order can be written as an MA process of infinite order by successive shift operator. Recall that an AR(p) process is given by

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \varepsilon_t \tag{2.23}$$

Using the backward shift operator approach we have

$$(1 - \alpha_1 \mathbf{B} - ... - \alpha_p \mathbf{B}^p) y_t = \varepsilon_t \tag{2.24}$$

or that

$$\begin{aligned} y_t &= \frac{\varepsilon_t}{1 - \alpha_1 \mathbf{B} - ... - \alpha_p \mathbf{B}^p} \\ &= (1 + \beta_1 \mathbf{B} + \beta_2 \mathbf{B}^2 + ...)\varepsilon_t. \end{aligned} \tag{2.25}$$

Thus the finite AR(p) process can be expressed as an MA process of infinite order and it implies that $E(y_t) = 0$. The variance is finite provided that $\sum \beta_i^2$ converges. The ACVF is given by

$$\gamma(k) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \beta_i \beta_{i+k}$$

where $\beta_0 = 1$. This is a necessary but not sufficient condition for stationarity. We can in principle find the ACVF of the general order AR process using the above procedure, but β_i may be algebraically hard to find. The alternative simpler way is to assume the process stated in equation (2.23) is stationary. Next multiply through by y_{t-k} and take expectations and

divide by σ_y^2 assuming the variance of y_t is finite. Then using the fact that $\rho_k = \rho_{-k}$ for all k we find

$$y_t y_{t-k} = y_{t-k}[\alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \varepsilon_t] \tag{2.26}$$

.

Thus

$$E[y_t y_{t-k}] = \alpha_1 E[y_{t-1} y_{t-k}] + ... + \alpha_p E[y_{t-p} y_{t-k}] \tag{2.27}$$

hence

$$\gamma(k) = \alpha_1 \gamma(k-1) + ... + \alpha_p \gamma(k-p) \tag{2.28}$$

.

Thus

$$\frac{\gamma(k)}{\sigma_y^2} = \alpha_1 \frac{\gamma(k-1)}{\sigma_y^2} + ... + \alpha_p \frac{\gamma(k-p)}{\sigma_y^2} \tag{2.29}$$

,

which implies

$$\rho(k) = \alpha_1 \rho(k-1) + ... + \alpha_p \rho(k-p) \tag{2.30}$$

.

Equations (2.28) and (2.30) are called the Yule Walker Equations. These are a set of difference equations and have the general solution (given for ρ_k)

$$\rho(k) = A_1 \pi_1^{|k|} + ... + A_p \pi_p^{|k|},$$

where π_i are the roots of the so called auxiliary equation

$$y^p - \alpha_1 y^{p-1} - ... - \alpha_p = 0.$$

The constants A_i are chosen to satisfy the initial conditions depending on $\rho(0) = 1$ which means that $\sum A_i = 1$. The Yulle-Walker equations are fully explained in Kendall and Ord, (1990).

2.3.5 Mixed AR and MA (ARMA) models

ARIMA models are meant to borrow from the strength of both the AR and MA process. Thus these models are formed by combining both the AR and MA structures to model the stochastic process $\{y_t\}$. A model containing p AR terms and q MA terms is said to be an ARMA process of order (p, q). This hybrid model is therefore given by

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_q \varepsilon_{t-q} \tag{2.31}$$

Using the backward shift operator $\mathbf{B}$ equation (2.31) can be written as

$$\phi(\mathbf{B}) y_t = \theta(\mathbf{B}) \varepsilon_t \tag{2.32}$$

where

$$\phi(\mathbf{B}) y_t = y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} - ... - \alpha_p y_{t-p} \tag{2.33}$$

and

$$\theta(\mathbf{B}) \varepsilon_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_q \varepsilon_{t-q} \tag{2.34}$$

thus $\phi(\mathbf{B})$ and $\theta(\mathbf{B})$ are polynomials of order p and q respectively, such that

$$\phi(\mathbf{B}) = 1 - \alpha_1 \mathbf{B} - ... - \alpha_p \mathbf{B}^p$$

and

$$\theta(\mathbf{B}) = 1 + \beta_1 \mathbf{B} + ... + \beta_q \mathbf{B}^q.$$

The constraints required on the model parameters to render the processes stationary and invertible are the same as for a pure AR or pure MA processes. That is the values of α_i, which guarantee stationarity, are such

that the roots of $\phi(\mathbf{B}) = 0$ lie outside the unit circle, while the values of β_i, which guarantee invertibility, are such that the roots of $\theta(\mathbf{B}) = 0$ lie outside the unit circle. The importance of the ARMA formulation lies in the fact that a stationary time series may often be adequately modelled by an ARMA model requiring much fewer parameters than a pure MA or AR process by itself. Thus such a formulation can potentially lead to a more parsimonious model than the respective separate formulations. The principle of parsimony prefers a model with as few parameters as possible, but which gives an adequate representation of the data being studied.

2.3.6 Integrated ARMA (or ARIMA) models

In practice most time series are non-stationary. In order to fit and exploit the nice properties of a stationary model, it is necessary to remove non-stationary sources of variation. If the observed time series is non-stationary in the mean, then we can difference the series. Thus for this reason let Δ denotes the 1-step difference operator. If y_t is replaced by $\Delta^d y_t$ where $d = 1, 2, ...$ then we have a model capable of describing certain types of non-stationary series. In this case d is the order of differencing. The resultant model is called an 'integrated' model because the stationary model that is fitted to the differenced data has to be summed or 'integrated' to provide a model for the original non-stationary data. Let

$$W_t = \Delta^d y_t = (1 - \mathbf{B})^d y_t.$$

Then the general form of the ARIMA model process is given by

$$W_t = \alpha_1 W_{t-1} + ... + \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + ... + \beta_q \varepsilon_{t-q} \qquad (2.35)$$

.

Equation (2.35) may be written in the form

$$\phi(\mathbf{B})W_t = \theta(\mathbf{B})\varepsilon_t$$

or

$$\phi(\mathbf{B})(1-\mathbf{B})^d y_t = \theta(\mathbf{B})\varepsilon_t$$

where the shift polynomials $\phi(\mathbf{B})$ and $\theta(\mathbf{B})$ are as defined in equations (2.33) and (2.34) respectively. The model for W_t, which describes the d^{th} order difference of Y_t is said to be an ARIMA process of order (p, d, q). In practice the first differencing is often found to be adequate to make the series stationary, and so the value of d is often taken to be one. ARIMA models can be generalized to include seasonal terms as will be described in detail in section 2.4.

2.4 Seasonal ARIMA (SARIMA models)

Many practical time series contain a seasonal periodic component which repeats every, say s observations, in which case s denotes the seasonal period. For example with monthly observations, with $s = 12$, we may expect time series y_t to depend on values at annual lags such as y_{t-12}, y_{t-24} and on more recent non-seasonal values such as y_{t-1} and y_{t-2}. A seasonal ARIMA model can be defined as

$$\alpha_p(\mathbf{B})\phi_P(\mathbf{B}^s)X_t = \beta_q(\mathbf{B})\theta_Q(\mathbf{B}^s)\varepsilon_t, \tag{2.36}$$

where $\mathbf{B}$ denotes the backward shift operator, α_p, ϕ_P, β_q θ_Q are polynomials of order p, P, q, Q respectively, ε_t is a purely random process and

$$X_t = \Delta^d \Delta_s^D y_t$$

denotes the differenced series. Model (2.36) is called SARIMA model of order $(p, d, q) \times (P, D, Q)_s$. D denotes the order of seasonal differencing, P denotes the order of seasonal AR process, Q denotes the order of seasonal MA process and s the seasonality. The differenced series $\{X_t\}$ is formed from the original series $\{y_t\}$ by appropriate differencing to remove non-stationarity terms. d is used to remove trend by simple differencing while seasonal differencing Δ_s is used to remove seasonality. For example if $d = D = 1$ and $s = 12$, then

$$\begin{aligned} X_t &= \Delta\Delta_{12} y_t = \Delta_{12} y_t - \Delta_{12} y_{t-1} \\ &= y_t - y_{t-12} - (y_{t-1} - y_{t-13}). \end{aligned} \tag{2.37}$$

Now consider the seasonal AR term $\phi_P(\mathbf{B}^s)$ and suppose $P = 1$, then $\phi_1(\mathbf{B}^s)$ will be of the form $(1 - c \times \mathbf{B}^s)$ where c denotes a constant, which simply means that X_t will depend on X_{t-s}, since $\mathbf{B}^s X_t = X_{t-s}$. Similarly, seasonal MA term of order one means that X_t will depend on ε_{t-s}. As an example, consider a SARIMA model of order $(1, 0, 0) \times (0, 1, 1)_{12}$ where $s = 12$. Here we have a non-seasonal AR term, one seasonal MA term and one seasonal differencing. Thus we can write this as

$$(1 - \alpha\mathbf{B}) X_t = (1 + \beta\mathbf{B}^{12}) \varepsilon_t \tag{2.38}$$

,

where $X_t = \Delta_{12} y_t$ and α and β are constant parameters. Equation (2.38) implies that

$$(1 - \alpha\mathbf{B}) \Delta_{12} y_t = (1 + \beta\mathbf{B}^{12}) \varepsilon_t$$

or

$$\Delta_{12} y_t - \alpha \Delta_{12} y_{t-1} = \varepsilon_t + \beta \varepsilon_{t-12}$$

or

$$y_t - y_{t-12} - \alpha(y_{t-1} - y_{t-13}) = \varepsilon_t + \beta \varepsilon_{t-12}$$

and finally we have

$$y_t = y_{t-12} + \alpha(y_{t-1} - y_{t-13}) + \varepsilon_t + \beta \varepsilon_{t-12} \tag{2.39}$$

so that $\{y_t\}$ depends on y_{t-1}, y_{t-12} and y_{t-13} as well as the innovations at time $(t-12)$. When fitting a seasonal model to data, the first task is to determine values of d and D which reduce the series to stationarity and the seasonality. Suitable values of p, P, q and Q are determined by looking at the ACF and partial ACF of the differenced series which aids in choosing an appropriate SARIMA model.

2.5 Transformations

As stated earlier most applied time series are nonstationary, that is, they have nonconstant mean or varying second moments such as the variance. In order to achieve stationarity, the time series can undergo some differencing as described above or it can also be transformed through transformations such as logarithm and square-root transformations. Although some time series are nonstationary due to some equilibrium forces, different parts of these series behave very much alike except for their difference in their mean levels. These are called homogeneous nonstationary series, (Box et al). These models can be made stationary by taking suitable

differencing of the general series. In other words the series Y_t is nonstationary but its d^{th} order differenced series $(1-\mathbf{B})^d Y_t$ for some integer $d \geq 1$, is stationary. If d^{th} order differenced series follows a white noise phenomenon, we have, $Y_t = Y_{t-1} + \varepsilon_t$ and the conditional mean function is

$$\mu_t = Y_{t-1} \tag{2.40}$$

since $E(\varepsilon_t) = 0$ which is subject to the stochastic disturbance at time $(t-1)$. In other words, the mean level of the process Y_t in $(1-\mathbf{B})^d Y_t$ for $d \geq 1$ changes through time stochastically and we characterize the process as having a stochastic trend. A process that is nonstationary in the mean is more likely to be nonstationary in the variance and autocovariance also. If the variance and covariance are nonconstant, that is, say

$$Var(Y_t) = cf(\mu_t)$$

for some positive constants c and function f, the question is, how do we find a function T so that the transformed series $T(Y_t)$ has a constant variance? To illustrate the method we approximate the desired function by a first order Taylor series expansion about the point μ_t as follows. Let

$$T(Y_t) \simeq T(\mu_t) + T'(\mu_t)(Y_t - \mu_t), \tag{2.41}$$

where $T'(\mu_t)$ is the first derivative of $T(Y_t)$ evaluated at μ_t.
Now taking the variance on both sides leads to

$$\begin{aligned} Var[T(Y_t)] &= [T'(\mu_t)]^2 Var(Y_t) \\ &= c[T'(\mu_t)]^2 f(\mu_t). \end{aligned} \tag{2.42}$$

The derivation given in equation (2.42) is infact the delta method approximation to $var[T(Y_t)]$. Thus in order that the variance of $T(Y_t)$ be constant we conclude that the variance stabilizing transformation $T(Y_t)$ must be chosen so that

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}} \tag{2.43}$$

since then $var[T(Y_t)] = c$ in equation (2.42) and which also implies that

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t \tag{2.44}$$

.

For example, if the standard deviation of a series is proportional to the mean so that

$$var(Y_t) = c^2\mu_t^2 \tag{2.45}$$

then

$$T(\mu_t) = \int \frac{1}{\mu_t} d\mu_t = ln(\mu_t). \tag{2.46}$$

Hence a logarithmic transformation of the series $ln(\mu_t)$ will give a constant variance. If the variance of the series is proportional to the mean so that

$$var(Y_t) = c\mu_t \tag{2.47}$$

then

$$\begin{aligned} T(\mu_t) &= \int \frac{1}{\sqrt{\mu_t}} d\mu_t \\ &= 2\sqrt{\mu_t}. \end{aligned} \tag{2.48}$$

Thus a square root transformation of the series $\sqrt{\mu_t}$ will give a constant variance. If the standard deviation of the series is proportional to the square of the level so that

$$var(Y_t) = c^2\mu_t^4 \tag{2.49}$$

then

$$\begin{aligned} T(\mu_t) &= \int \frac{1}{\sqrt{\mu_t^4}} d\mu_t \\ &= -\frac{1}{\mu_t}. \end{aligned} \tag{2.50}$$

a desired transformation that gives a constant variance will be the reciprocal $\frac{1}{\mu_t}$. More generally, to stabilize the variance, we can use the power transformation which is more flexible given by

$$\begin{aligned} T(Y_t) &= Y_t^{(\lambda)} \\ &= f(Y_t, \lambda) \\ &= \frac{Y_t^{\lambda} - 1}{\lambda} \end{aligned} \tag{2.51}$$

introduced by Box and Cox (1964) and λ is called the transformation parameter. In table (2.1) we give some commonly used values of λ and their corresponding transformations. The beauty of the power link model is that, a range of values of λ can be explored in order to find the most optimal one. The main reason to consider the Box-Cox transformation family is to analyze the data in a more normal scale. Such a scale typically leads to models that are simpler and easier to interpret, with possibly well

Table 2.1: Different values of λ and their corresponding transformations

value-of λ	Transformation
-1.0	$\frac{1}{\varepsilon_t}$
-0.5	$\frac{1}{\sqrt{\varepsilon_t}}$
0.0	$ln\varepsilon_t$
0.5	$\sqrt{\varepsilon_t}$
1.0	ε_t

behaved residuals and clearer inference. It should however be noted that if the time series data is known to come from non-Gaussian distribution then a time series model accounting for non-Gaussian data can be used, Zeger and Qaqish, (1988); Zeger and Liang ,(1991); Peng and Dominic, (2008)

Chapter 3

Model identification, estimation, checking and forecasting

3.1 Introduction

Time series analysis comes with a number of ways in which to build a suitable model for a given time series, however for this thesis the three stage approach proposed by Box, et al) is adopted. The two postulated the following three step process namely:

- model identification (or model specification)
- model estimation (or model fitting)
- model checking (or model verification/diagnostic checking).

This chapter focuses on explaining these steps in detail.

3.2 Model selection

This stage involves identifying the appropriate ARIMA model by determining what the appropriate orders of the AR, MA and Integrated parts are. Model identification is also aided by examining the time plot of the series, which include the construction of the correlogram (autocorrelation function), the partial and inverse autocorrelation functions. In the presence of several competing models, use of model selection criteria becomes necessary. Such criteria include the Akaike's Information Criteria (AIC) (with its modification to AIC_c, Bayesian Information Criteria (BIC) and Schwartz's Bayesian Criteria (SBC) are used to choose the most appropriate model. The formal expressions for the above criteria in terms of the log-likelihood are:

- AIC = -2ln(likelihood) + 2r
- AIC_c = -2ln(likelihood) +$\frac{2rN}{N-r-1}$
- BIC = -2ln(likelihood) + $(r + rlnN)$
- SBC = -2ln(likelihood) + $rlnN$

where r denotes the number of parameters and N denotes the number of observations. The AIC_c is a modified form of AIC to correct for small sample data and the BIC is a modified SBC with an extra penalty of r in addition to $rlnN$. In considering the statistics given above, there is need to strike a balance between the need for a parsimonious model, which uses as few parameters as possible but maintains its usefulness and a model that is too simple and overlooks the important effects. The model selection criteria are revisited and described fully in section (3.6) below.

3.3 Model estimation

Model estimation involves estimating the parameters for the selected model using the computational methods suggested by Box and Jenkins. These are the maximum likelihood method (ML), the unconditional least squares, and the conditional least squares. In this thesis the least squares method is discussed fully and is applied. An ARMA(p, q) model contains three different kinds of parameters: the p-AR parameters, the q-MA parameters and the variance of the error term ε_t to be estimated. This gives a total of $p + q + 1$ parameters to be estimated from the stationary time series. Sometimes it is necessary to include a constant term in which case there will be a total of $p+q+2$ parameters to be estimated. We will consider how to estimate these separately.

3.3.1 Estimating parameters of AR processes

Suppose we define the AR(p) process where each term is corrected to the mean μ as follows

$$y_t - \mu = \alpha_1(y_{t-1} - \mu) + ... + \alpha_p(y_{t-p} - \mu) + \varepsilon_t. \tag{3.1}$$

Given N observations $y_1, ..., y_N$, the parameters $\mu, \alpha_1, ..., \alpha_p$ may be estimated using least squares method by minimizing the sums of squares

$$S = \sum_{t=p+1}^{N} \{(y_t - \mu - \alpha_1(y_{t-1} - \mu) - ... - \alpha_p(y_{t-p} - \mu)\}^2 \tag{3.2}$$

with respect to $\mu, \alpha_1, ..., \alpha_p$. If the ε_t process is normal, then the least square estimates are also maximum likelihood estimates. In the first order case with $p = 1$, and according to Chatfield, (2004), it can be established that

$$\hat{\mu} = \frac{\bar{y}_2 - \hat{\alpha}_1 \bar{y}_1}{1 - \hat{\alpha}_1} \tag{3.3}$$

and

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N}(y_t - \hat{\mu})(y_{t-1} - \hat{\mu})}{\sum_{t=1}^{N}(y_t - \hat{\mu})^2} \tag{3.4}$$

where $\bar{y}_1$ and $\bar{y}_2$ are the means of the first and last $N-1$ observations in the series. By taking $\bar{y}_1 \simeq \bar{y}_2 \simeq \bar{y}$ so we can approximate $\hat{\mu} = \bar{y}$. Substituting $\hat{\mu} = \bar{y}$ into equation (3.4) we obtain

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{N}(y_t - \bar{y})^2}. \tag{3.5}$$

This is exactly the estimator that would arise if we treat the AR equation

$$Y_t - \bar{y} = \alpha_1(y_{t-1} - \bar{y}) + \varepsilon_t \tag{3.6}$$

as an ordinary regression with $y_{t-1} - \bar{y}$ as the independent variable. As in Chatfield, (2004), it can be shown that equation (3.5) can be regarded as an approximation of α_1

$$\hat{\alpha}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \rho_1.$$

This approximate estimate for α_1 is also appealing since $\hat{\rho}_1$ is an estimator of ρ_1 and $\rho_1 = \alpha_1$ for a first order AR process as shown in section (2.3.4). For a second order AR process where $p = 2$ similar approximations to those given above can be given as, (Chatfield, 2004):

$$\hat{\mu} = \bar{y}$$

$$\hat{\alpha}_1 = \frac{\hat{\gamma}_1(1-\hat{\gamma}_2)}{(1-\hat{\gamma}_1)}$$

$$\hat{\alpha}_2 = \frac{\hat{\gamma}_2 - \hat{\gamma}_1^2}{1-\hat{\gamma}_1^2}.$$

The coefficient $\hat{\rho}_2$ is called the sample autocorrelation coefficient of order two as it measures the excess correlations two steps apart not accounted for by the autocorrelation at lag 1 namely ρ_1. Higher order AR processes may also be fitted in a similar way by taking $\hat{\mu} = \bar{y}$ and substituting the sample autocorrelation coefficients into the first p Yulle Walker equations, as given in section (2.3.4), and solving for $\{\hat{\alpha}_1, ..., \hat{\alpha}_p\}$. In matrix form these equations are given as:

$$\mathbf{R}\hat{\alpha} = \mathbf{r}$$

where $\hat{\alpha}^T = (\hat{\alpha}_1, ..., \hat{\alpha}_p)$, $\mathbf{r}^T = (\hat{\gamma}_1, ..., \hat{\gamma}_p)$ and

$$\mathbf{R} = \begin{pmatrix} 1 & \hat{\gamma}_1 & \hat{\gamma}_2 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & 1 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-2} \\ \hat{\gamma}_2 & \hat{\gamma}_1 & 1 & \cdots & \hat{\gamma}_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \hat{\gamma}_{p-3} & \cdots & 1 \end{pmatrix}$$

is a $p \times p$ matrix. For large N the estimated values given are very close to the exact least squares estimates.

3.3.2 Estimating parameters of an MA process

Estimation of parameters of MA processes is more challenging than the AR processes because, as stated in Chatfield, (2004), efficient explicit estimators cannot be easily found. Some form of numerical iterations must be employed. First consider the first-order MA process

$$y_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} \tag{3.7}$$

where μ and β_1 are constant parameters and ε_t denotes a purely random process. Our aim would be to write the residual sum of squares $\sum \varepsilon_t^2$ solely in terms of the observed y_t and parameters μ and β_1 and differentiate with respect to μ and β_1 and hence find the least squares estimates. Unfortunately as in Chatfield, (2004), this cannot be done. Therefore explicit least square estimates cannot be found, nor is it wise to simply equate sample and theoretical first order autocovariance

$$\hat{\gamma}_1 = \frac{\hat{\beta}_1}{1 + \hat{\beta}_1^2} \tag{3.8}$$

and choose the solution $\hat{\beta}_1$ such that $|\hat{\beta}_1| > 1$, since it can be shown that this gives rise to an inefficient estimator. Alternatively starting values for μ and β_1 such as $\mu = \bar{y}$ and the value of β_1 given by the solution of equation (3.8) can be obtained. Then we can calculate corresponding residual sum of squares using equation (3.7) recursively in the form

$$\varepsilon_t = y_t - \mu - \beta_1 \varepsilon_{t-1}.$$

Taking $\varepsilon_0 = 0$ we calculate $\varepsilon_1 = y_1 - \mu$ and $\varepsilon_2 = y_2 - \mu - \beta_1 y_1$ and so on until $\varepsilon_N = y_N - \mu - \beta_1 \varepsilon_{N-1}$. Then the residual sum of squares $\sum_{t=1}^{N} \varepsilon_t^2$ is calculated conditional on the given values of the parameters and $\varepsilon_0 = 0$. This procedure is repeated for other neighbouring values of μ and β_1 so that the residual sum of squares $\sum \varepsilon_t^2$ is computed on a grid of points in the (μ, β_1) plane. Values of μ and β_1 that minimize $\sum \varepsilon_t^2$ can be determined by inspection. This procedure gives the least squares estimates which are also maximum likelihood estimates conditional on $\varepsilon_0 = 0$ provided that ε_t is normally distributed. For higher order processes a similar type of an iterative procedure to that described above could be used. For instance with a second-order MA process one

would guess starting values of μ, β_1 and β_2, then compute the residuals recursively using

$$\varepsilon_t = y_t - \mu - \beta_1\varepsilon_{t-1} - \beta_2\varepsilon_{t-2}$$

and compute $\sum \varepsilon_t^2$. Other values of μ, β_1 and β_2 can be tried perhaps on a grid of point until a minimum value of $\sum \varepsilon_t^2$ is found.

3.3.3 Estimating the parameters of an ARMA process

Suppose a model with both AR and MA components, that is, an ARMA process is thought to explain the data. The estimation procedure is similar to those for an MA process in that an iterative procedure has to be used as explained above. The residual sum of squares can be calculated at every point on a suitable grid of the parameter values, and the values, which give the minimum sum of squares, may then be assessed. Consider, as an example, the ARMA(1,1) process given by

$$y_t - \mu = \alpha_1(y_{t-1} - \mu) + \varepsilon_t + \beta_1\varepsilon_{t-1} \tag{3.9}$$

Given observations up to time t, $\{y_1, ..., y_t\}$ we can guess starting values for μ, α_1 and β_1, set $\varepsilon_0 = 0$ and $y_0 = \mu$ and then calculate the residuals recursively by

$$\varepsilon_1 = y_1 - \mu$$

$$\varepsilon_2 = y_2 - \mu - \alpha_1(y_1 - \mu) - \beta_1\varepsilon_1$$

$$\vdots$$

$$\varepsilon_t = y_t - \mu - \alpha_1(y_{t-1} - \mu) - \beta_1\varepsilon_{t-1}.$$

The residuals sum of squares $\sum_{i=1}^{t} \varepsilon_i^2$ for $i = 1, 2, ..., t$ may then be calculated. Then other values of μ, α_1 and β_1, may be tried until the minimum residual sum of squares is found. The conditional least squares estimates, as pointed out by Ansley and Newbold, (1986) and also Chatfield, (2004), provide very accurate estimates if the series is 'long'. The procedure for estimating an ARMA process also applies to ARIMA model parameter estimation. As earlier stated in an ARIMA process, stationarity is achieved by differencing, hence the reason for referring to the series as integrated. As was pointed out in section (2.3.6) ARIMA models can also be generalized to accommodate seasonal terms. This generalization leads to a class of models called seasonal ARIMA (SARIMA) which are discussed in section (2.4) in chapter 2.

3.4 Model checking

When a model has been fitted to a time series, it is advisable to check that the model really does provide an adequate description of the data. As with most statistical models, this is usually done by looking at the residuals, which are defined by

residual=observation-fitted value

For a univariate time series model, the fitted value is the one-step-ahead forecast, as defined in section (3.7) below, so that the residual is the one-step-ahead forecast error. For example, for the AR(1) model, with $y_t = \alpha y_{t-1} + \varepsilon_t$ where α is estimated by least squares, the fitted value at time t is $\hat{y}_{t-1}$ so that the residual corresponding to the observed value y_t is

$$\varepsilon_t = y_t - \alpha y_{t-1}.$$

Note that $E(y_t|y_{t-1}) = \alpha y_{t-1}$ since $E(\varepsilon_t) = 0$. If a model fits the data well, then we expect the residuals to be random and close to zero. Thus model validation usually consists of plotting residuals in various ways to see whether this is the case. The following steps are usually performed to assess the extent of model adequacy.

- Plot residuals against time and look for unusual values, or for any increasing (decreasing) dispersion which may suggest the need to transform the data.
- Examine the appropriate t-ratio parameter estimates to see whether any terms should be dropped from the model.
- Examine the correlograms derived from the residuals to determine whether additional terms are required.
- Check whether the selected model (after differencing) is stationary and invertible.
- Check the overall fit of the model (although this is less often done than in regression analysis)

Residual analysis can also involve formal tests in the form of the Box-Pierce-Ljung test and information criteria as given in sections (3.5) and (3.6) respectively below.

3.5 The Box-Pierce-Ljung test

It is necessary to check the model for adequacy before it is used for forecasting. The Box-Pierce-Ljung test is one of the widely used lack-of-

fit test, that is, the appropriateness of the fitted model. It was developed by Box and Pierce (1970) and modified by Ljung and Box (1978). They suggested a 'portmanteau' test to test the null hypothesis

$$H_0 : \rho_1 = \rho_2 = \rho_3 = ... = \rho_k = 0$$

against the alternative hypothesis

$$H_1 : \text{not all } \rho_j = 0$$

where ρ_j is the ACF at lag $j = 1, 2, ..., k$. Based on the residual correlogram, they suggested the statistic

$$Q = Q(k) = n(n+2) \sum_{j=1}^{k} \frac{r_j^2}{n-j}. \tag{3.10}$$

where n denotes the length of the series after any differencing and r_j denote the residuals. Box and Pierce (1970) showed that under H_0, Q is asymptotically distributed as chi-squared with $(k - p - q)$ degrees of freedom, where p and q are the order of the AR and MA processes in the hybrid ARIMA model given in equation (2.31) in chapter 2.

3.6 Information criteria

This is a tool used to compare two or more competing, but not necessarily nested, models (with different number of parameters) using their likelihood functions (LF). For instance if model 1, with k_1 parameters with $LF = L_1$ is correct and is compared with model 2, which has an additional $(k_2 - k_1)$ parameters and its $LF = L_2$, it is well known fact that minus twice the log-likelihood ratio asymptotically follows a chi-squared distribution with $(k_2 - k_1)$ degrees of freedom. Thus for large samples

$$E[-2ln(L_1/L_2)] = k_2 - k_1. \tag{3.11}$$

This led to Akaike (1974) to propose an information criteria (AIC), to deal with models not necessarily nested, of the form

$$AIC = -2ln(L) + k,$$

where k denotes the number of parameters in the model. The model with the smallest AIC is deemed best in the sense of minimizing the forecast mean square error (FMSE). However, it was pointed out by Schwartz (1978) that AIC is not a consistent criterion in that it does not select the true model with probability approaching 1 as $n \to \infty$. To overcome this problem, Schwartz proposed a Bayesian information criterion (BIC) given by:

$$BIC = -2ln(L) + kln(n),$$

where n is the length of the time series or the sample size. After an estimated model has been validated, forecasting future values of the series can be considered. This is the subject of discussion in section 3.7 below.

3.7 Forecasting

One of the main objectives of this study and for time series analysis in general is to use the constructed model to compute forecasts. In time series, forecasting is a mathematical way of estimating future values of a series using present and historical values of the series. Suppose we have an observed time series $y_1, y_2, ..., y_t$, then the basic problem is to estimate future values such as y_{t+l}, for $l = 1, 2, ...$, where the integer l is called the lead time or forecasting horizon. The forecast of y_{t+l} made at time t for l

steps ahead is typically denoted by $\hat{y}_t(l)$. The method used in forecasting depends on the purpose to which the forecast are intended for:

- we may require a forecast which is to be the basis of our own action for instance sales, that is, to what our sales would be, say in light of a marketing strategy.
- forecasts for general purpose, say, population forecast, growth rates, disease rates, inflation rates, etc.

Furthermore, the time frame for forecasting may be divided into three horizons, short term, medium term and long term. There are several forecasting techniques applicable for time series. Some key forecasting techniques are discussed in the sections below.

3.7.1 Simple exponential smoothing

This is a method that relies on simple updating equations to calculate forecasts assuming the data is non-seasonal and showing no systematic trend. Given a non-seasonal time series $y_1, y_2, ..., y_t$ with no systematic trend, a one-step-ahead forecast y_{t+1} can be forecasted by means of a weighted sum of past and present observations given by

$$\hat{y}_t(1) = c_0 y_t + c_1 y_{t-1} + c_2 y_{t-2} + ... \tag{3.12}$$

where $\{c_i\}$ are the weights. Clearly it is more sensible to give more weight to recent observations. A good example are the geometric weights which decrease by a constant ratio for every unit increase in the lag. The weights are required to sum to one, thus one can take

$$c_i = \alpha(1-\alpha)^i, \text{ for } i = 0, 1, 2, ...$$

where α is a constant such that $0 < \alpha < 1$, (Chatfield, 2004). Equation (3.12) becomes

$$\hat{y}_t(1) = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + ... + \alpha(1-\alpha)^i y_{t-i} + ... \quad (3.13)$$

Equation (3.13) implies infinite number of past observations, however in practice there will only be a finite number. Equation (3.13) can be rewritten in a recurrence form as

$$\begin{aligned} \hat{y}_t(1) &= \alpha y_t + (1-\alpha)[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + ...] \\ &= \alpha y_t + (1-\alpha)\hat{y}_{t-1}(1). \end{aligned} \quad (3.14)$$

If we set $\hat{y}_1(1) = y_1$, then equation (3.14) can be used recursively to compute subsequent forecasts. Equation (3.14) implies that forecasts can be updated using only the latest observation and the previous forecast. This process is called simple exponential smoothing (SES), (exponential in the sense that the geometric weights lie on an exponential curve) (Chatfield, 2004). SES is optimal if the underlying model for the time series is given by

$$y_t = \mu + \alpha \sum_{i<t} \varepsilon_i + \varepsilon_t. \quad (3.15)$$

where $\{\varepsilon\}$ denotes a purely random process. Values of the smoothing constant α between 0.1 and 0.3 are commonly used and produce forecasts that depend on a large number of past observations. Values of α close to one give forecasts that depend much on recent observations. If we write equation (3.14) in its error correction form we get

$$\begin{aligned}\hat{y}_t(1) &= \alpha[y_t - \hat{y}_{t-1}(1)] + \hat{y}_{t-1}(1) \\ &= \alpha e_t + \hat{y}_{t-1}(1) \qquad (3.16)\end{aligned}$$

where $e_t = y_t - \hat{y}_{t-1}(1)$ is the prediction error at time t hence using different values of α, we can obtain different sums of squares of one-step ahead forecasts as $\sum_{i=2}^{t} e_i^2$ where if

$$\hat{y}_1(1) = y_1$$

then

$$e_2 = y_2 - \hat{y}_1(1)$$

and

$$\hat{y}_2(1) = \alpha e_2 + \hat{y}_1(1)$$

then

$$e_3 = y_3 - \hat{y}_2(1)$$

and so on until

$$e_t = y_t - \hat{y}_{t-1}(1).$$

A value of α that minimize $\sum_{i=2}^{t} e_i^2$ is selected either by inspection or by an algorithmic numerical procedure, and that will be the value deemed to give the best forecasts.

3.7.2 The Holt and Holt-Winters forecasting procedures

The simple exponential smoothing procedure discussed above works well when dealing with non-seasonal time series with no trend, however in practice time series data is usually seasonal and contains trend variation. Holt's two parameter exponential smoothing is used to obtain forecasts in non-seasonal data with trend. When data is seasonal and has trend, a version called Holt-Winters three parameter procedure can be used. The idea in these procedures is to generalize the equations for the SES by introducing trend and seasonal terms, which are updated by exponential smoothing. The one-step-ahead forecast from the SES can be thought of as an estimate of the local mean level of the series, and SES can be regarded as a way of updating the local level of the series, say L_t. For this, equation (3.14) can be written as

$$L_t = \alpha y_t + (1 - \alpha) L_{t-1}. \tag{3.17}$$

Suppose we now wish to include a trend term, say T_t, which is the expected increase or decrease per unit time period in the current level. A reasonable pair of equations for updating the values of L_t and T_t in recurrence form are the following

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma) T_{t-1}$$

where α and γ are smoothing parameters generally chosen to lie in the range (0,1). Then the l-step-ahead forecast for non-seasonal data with trend at time t can now be stated as

$$\hat{y}_t(l) = (L_t + lT_t) \text{ for } l = 1, 2, ...$$

To cope with seasonality the above procedure can be generalized by including an additional term namely I_t which denotes a seasonal index. The interpretation of I_t depends on whether seasonality is multiplicative or additive. If additive, a deseasonalized value is $y_t - I_t$ and if multiplicative, it is $\frac{y_t}{I_t}$. Since all three quantities L_t, T_t and I_t need to be estimated, we need at least three equations with three smoothing parameters say α, γ and δ all chosen in the range (0,1). Suppose the observations are monthly, and that the seasonal variation is multiplicative. Then the (recurrence form) equations for updating L_t, T_t and I_t when a new observation y_t becomes available are

$$\begin{aligned} L_t &= \alpha\Big(\frac{y_t}{I_{t-12}}\Big) + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \\ I_t &= \delta\Big(\frac{y_t}{L_t}\Big) + (1-\delta)I_{t-12} \end{aligned} \tag{3.18}$$

and the l-step-ahead forecast from time t are then

$$y_t(l) = (L_t + lT_t)I_{t-12+l} \text{ for } l = 1, 2, ..., 12,$$

where l spans over 12 time intervals if the series is made up of monthly observations. There are similar formulae for additive seasonal cases and where the seasonality is of the length s, rather than 12 as for monthly observations. After forecasts have been computed, it is of paramount importance to assess the accuracy of the forecasts. Section 3.8 below explains the methods of assessing the forecast accuracy.

3.8 Forecast accuracy

Suppose we have a known model such as

$$y_t = \mu_t + \varepsilon_t \tag{3.19}$$

where μ_t is a known function of past y values, and ε_t is independent of μ_t. It is apparent that a suitable forecast for y_t is μ_t and that the forecast error after y_t is observed is $\varepsilon_t = y_t - \mu_t$. The performance of a forecasting model at a single point in time does not tell us much about the overall quality of the method rather it is the long-run average properties of the ε_t that will be informative. Precisely, we may record the errors at time periods $t = n+1, n+2, ..., n+m$ and examine the corresponding observed forecast errors $\varepsilon_{n+1}, \varepsilon_{n+2}, ..., \varepsilon_{n+m}$. Assuming the ε_t are independent and identically distributed, with mean zero and standard deviation σ we consider the forecast mean square error (FMSE)

$$FMSE = FMSE(\mu_t) = E[(y_t - \mu_t)^2] = E(\varepsilon_t^2)$$

which may be estimated by

$$\hat{FMSE} = \sum_{i=1}^{m} \frac{e_{n+i}^2}{m}$$

Given the assumption that the $\{e_t\}$ are iid it follows that $FMSE = \sigma^2$, but under more general conditions

$$\begin{aligned} FMSE(\hat{\mu}_t) &= E[(y_t - \hat{\mu}_t)^2] \\ &= E[(y_t - \mu_t)^2] + E[(\mu_t - \hat{\mu}_t)^2] \end{aligned} \tag{3.20}$$

where $\hat{\mu}_t$ is the estimator for μ_t based upon the estimated parameter values. The above decomposition follows from the independence of the error terms. Other measures of forecast performance include

- Forecast Mean Absolute Error (FMAE) given by

$$FMAE = E[|y_t - \mu_t|] = E(|\varepsilon_t|)$$

- Forecast Mean Absolute Percentage Error (FMAPE) given by

$$FMAPE = E[|y_t - \mu_t|/y_t]100\%$$

  provided y_t takes on positive values.

The discussion so far presented relies only on sampling errors. However in practice other sources of errors can also be considered, prompting the use or consideration of the best linear predictors as an alternative.

3.8.1 Best Linear Predictors

Now assume that the process has a known ARIMA form and the question of interest is about what is the best forecast achievable given the model can be expressed in terms of random shocks as

$$y_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t+k-j} \tag{3.21}$$

where Ψ_j is the j^{th} parameter. The information we have available is equivalent to finding $[\varepsilon_t, \varepsilon_{t-1}, ...]$ so that any linear predictor for k steps ahead at time t is

$$m_t(k) = \sum_{j=0}^{\infty} \eta_{j+k} \varepsilon_{t-j}. \tag{3.22}$$

We now determine the best linear predictor using $FMSE$ as the criterion. The $FMSE$ is

$$\begin{aligned} FMSE[m_t(k)] &= E[y_{t+k} - m_t(k)]^2 \\ &= E\Big[\sum_{j=0}^{k-1} \Psi_j \varepsilon_{t+k-j} + \sum_{j=k}^{\infty} (\Psi_j - \eta_j)\varepsilon_{t+k-j}\Big]^2 . \end{aligned} \quad (3.23)$$

Since the ε_t are independent with zero mean and variance σ^2, this reduces to

$$FMSE[m_t(k)] = \sigma^2\Big[\sum_{j=0}^{k-1} \Psi^2 + \sum_{j=k} (\Psi_j - \eta_j)^2\Big]. \quad (3.24)$$

This shows that $FMSE$ is minimized when $\Psi_j = \eta_j, j \geq k$, that is, the best linear predictor is

$$y_t(k) = \textstyle\sum_{j=k}^{\infty} \Psi_j \varepsilon_{t+k-j}$$

with

$$FMSE(k) = \sigma^2 \textstyle\sum_{j=0}^{k-1} \Psi_j^2.$$

A $100(1-\alpha)\%$ prediction interval for y_{t+k} can be set up assuming the errors to be normally distributed as

$$y_t(k) \pm z_{\alpha/2}(FMSE)^{\frac{1}{2}},$$

where z denotes the z value in the normal probability table. An alternative determination of the best linear predictor is to consider the minimization of

$$E[(y_{t+k} - \mu_{t+k})^2 | D_t]$$

where $D_t = [y_t, y_{t-1}, ..., \varepsilon_t, \varepsilon_{t-1}, ...]$ denotes the available information at time t implying that

$$\begin{aligned} \mu_{t+k} &= E(y_{t+k}|D_t) \\ &= y_t(k). \end{aligned} \tag{3.25}$$

Since the ARIMA models are linear in y and ε, it follows that for any ARIMA(p, d, q) scheme

$$y_t(k) = \sum_{j=1}^{p} \phi_j E(y_{t+k-j}|D_t) - \sum_{j=0}^{q} \theta_j E(e_{t+k-j}|D_t).$$

For any past values with $s < 0$

$$E(y_{t+s}|D_t) = y_{t+s}$$

and

$$E(\varepsilon_{t+s}|D_t) = \varepsilon_{t+s},$$

whereas for $s > 0$

$$E(y_{t+s}|D_t) = y_t(s)$$

and

$$E(\varepsilon_{t+s}|D_t) = 0.$$

Chapter 4

Conditional Heteroscedasticity: ARCH-GARCH models

4.1 Introduction

So far the modelling of the time series considered has assumed time-invariant or constant variance. However in real life financial data, variance may change with time (a phenomenon defined as heteroscedasticity), hence there is a need for models that accommodate this possible variation in variance. The autoregressive conditional heteroscedastic (ARCH) models, with its extension to generalized ARCH, (GARCH) models as introduced by Engle (1982) and Bollerslev (1986) respectively, accommodates the dynamics of conditional heteroscedasticity. This is done by relating the error variance to the previous errors in the case of ARCH. In the case of GARCH, the previous conditional variances are included. Het-

eroscedasticity affects the accuracy of forecast confidence limits and thus has to be handled properly by constructing appropriate non-constant variance models. To test for heteroscedasticity, the portmanteau test statistic and the Engle's lagrange multiplier tests have been proposed and frequently in use. The ARCH-GARCH modelling considers the conditional error variance as a function of the past realization of the series. Campbell et al., (1997) argued that it is both logically inconsistent and statistically inefficient to use and model volatility measures that are based on the assumption of constant variance over some period when the resulting series moves or progress through time. In the case of financial data for example, large and small errors occur in clusters, which implies that large returns are followed by more large returns and small returns by further small returns. In the context of the current study this is equivalent to saying periods of high inflation are usually followed by further periods of high inflation, while low inflation is likely to be followed by much low inflation. The theory of the ARCH models and the GARCH models are explained in detail in sections 4.2 and 4.3 below respectively.

4.2 ARCH(q) model

Let $\{y_t\}$ be the mean-corrected return or rate of inflation, ε_t be the Gaussian white noise with zero mean and unit variance and H_t be the information set at time t given by $H_t = \{y_1, y_2, ..., y_{t-1}\}$, then the process $\{y_t\}$ is ARCH(q) (Engle, 1982) if

$$y_t = \sigma_t \varepsilon_t \tag{4.1}$$

where

$$E(y_t|H_t) = 0 \tag{4.2}$$

$$V(y_t|H_t) = \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 \tag{4.3}$$

and the error term ε_t is such that

$$E(\varepsilon_t|H_t) = 0 \tag{4.4}$$

$$V(\varepsilon_t|H_t) = 1. \tag{4.5}$$

Equations (4.4) and (4.5) show that the error term ε_t is a conditionally standardized martingale difference defined as follows: A stochastic series $\{y_t\}$ is a martingale difference if its expectation with respect to past values of another stochastic series X_i is zero. That is

$$E(y_{t+i}|X_i, X_{i-1}, ...) = 0$$

for $i = 1, 2, ...$ In this type of the model the impact of the past on the present volatility is assumed to be a quadratic function of lagged innovations. The coefficients $(\alpha_0, \alpha_1, ..., \alpha_q)$ can consistently be estimated by regressing y_t^2 on $y_{t-1}^2, y_{t-2}^2, ..., y_{t-q}^2$. To ensure non-negative volatility we require $\alpha_0 \geq 0, \alpha_i \geq 0$ for all $i = 1, ..., q$.

4.2.1 ARCH(1) model

The ARCH(1) model is a particular case of the general ARCH(q) model and is defined as follows: Let $\{y_t\}$ be the mean-corrected return and ε_t be a Gaussian white noise with mean zero and unit variance. If H_t is the information set available at time t then the process $\{y_t\}$ is ARCH(1) if $q = 1$ given by

$$y_t = \sigma_t \varepsilon_t \tag{4.6}$$

and the conditional variance σ_t^2 is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{4.7}$$

where α_0 and α_1 are unknown parameters. Thus under the normality assumption of ε_t the process can be stated conditionally in terms of H_t similar to the variance σ_t^2 as given above. To ensure non-negativity condition for the conditional variance the constraints $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$ must be satisfied. Equations (4.6) and (4.7) suggest that a large past squared mean-corrected return implies a large conditional variance σ_t^2 resulting in y_t being large in absolute value. The ARCH(1) is a special case of ARCH(q), therefore what applies for ARCH(q) also applies for ARCH(1).

4.2.2 Estimation of the ARCH(1) and the ARCH(q) models

Based on the assumption of normality made on the ε_t the method of maximum likelihood estimation is adopted. Let $y_1, y_2, ..., y_t$ be a realization from an ARCH(1) process, then the likelihood of the data can be written as a product of the conditionals as

$$f(y_1, y_2, ..., y_t)|\theta) = f(y_t|y_{t-1})f(y_{t-1}|y_{t-2})...f(y_2|y_1)f(y_1|\theta) \tag{4.8}$$

where θ=$(\alpha_0, \alpha_1)'$. It is more practical to condition on y_1 since the form $f(y_1|\theta)$ is difficult to obtain. Usually y_1 is assumed known and equal to its observed value. This allows us to use the conditional likelihood given by

$$f(y_1, y_2, ..., y_t|\theta; y_1) = f(y_t|y_{t-1})f(y_{t-1}|y_{t-2})...f(y_2|y_1)f(y_1|\theta; y_1) \tag{4.9}$$

Since $y_t|H_t \sim N(0, \sigma_t^2)$ it follows that

$$f(y_t|H_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} exp\left[\frac{-y_t^2}{2\sigma_t^2}\right] \tag{4.10}$$

where $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2$. The conditional log-likelihood is expressed as

$$\begin{aligned} l &= l_c(\alpha_0, \alpha_1|y_1) \\ &= lnf(y_2, ...y_t|y_1; \theta) \\ &= -\frac{1}{2}\sum_{i=2}^{t} ln(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=2}^{t}\frac{y_i^2}{\sigma_i^2} \end{aligned} \tag{4.11}$$

The maximum likelihood estimates are obtained by maximizing this function with respect to α_0 and α_1, (Tsay, 2002). Note that the function is nonlinear in these parameters and thus its maximization must be done using appropriate nonlinear optimization routine. Let a process $[y_t]_{t=1}^T$ be a series generated by an ARCH(1) process, where T is the sample size. Conditioning on the initial observation, the joint density function can be written as

$$f(y) = \prod_{t=2}^{T} f(y_t|H_t). \tag{4.12}$$

To find the conditional maximum likelihood estimates of α_0 and α_1, first one needs the derivatives of the the conditional log-likelihood with respect to α_0 and α_1 given by

$$\begin{aligned} \frac{\partial l_t}{\partial \alpha_0} &= \frac{1}{2(\sigma_t^2)}\Big(\frac{y_t^2}{\sigma_t^2} - 1\Big)\frac{\partial l_t}{\partial \alpha_0} \\ &= \frac{1}{2\sigma_t^2}\Big(\frac{y_t^2}{\sigma_t^2} - 1\Big)\frac{\partial l_t}{\partial \sigma_t^2} \times \frac{\partial \sigma_t^2}{\partial \alpha_0} \end{aligned} \tag{4.13}$$

and

$$\begin{aligned}\frac{\partial l_t}{\partial \alpha_1} &= \frac{1}{2(\sigma_t^2)}\Big(\frac{y_t^2}{\sigma_t^2}-1\Big)\frac{\partial l_t}{\partial \alpha_1} \\ &= \frac{1}{2\sigma_t^2}\Big(\frac{y_t^2}{\sigma_t^2}-1\Big)\frac{\partial l_t}{\partial \sigma_t^2}\times\frac{\partial \sigma_t^2}{\partial \alpha_1}. \end{aligned} \tag{4.14}$$

More generally the partial derivative of l is

$$\begin{aligned}\frac{\partial l}{\partial \theta} &= \sum_{t=2}^{T}\frac{\partial l_t}{\partial \sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta} \\ &= -\frac{1}{2}\sum_{t=2}^{T}\left(\frac{1}{\sigma_t^2}-\frac{y_t^2}{\sigma_t^4}\right)\begin{pmatrix}1\\ y_{t-1}^2\end{pmatrix} \\ &= \frac{1}{2}\sum_{t=2}^{T}\left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2}\begin{pmatrix}1\\ y_{t-1}^2\end{pmatrix} \end{aligned} \tag{4.15}$$

recalling that $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2$. Since $\frac{\partial^2 \sigma_t^2}{\partial\theta\partial\theta'} = 0$, the Hessian is given by

$$\begin{aligned}\frac{\partial^2 l}{\partial\theta\partial\theta'} &= \sum_{t=2}^{T}\frac{\partial^2 l_t}{\partial\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} \\ &= -\frac{1}{2}\sum_{t=2}^{T}\left(\frac{y_t^2}{(\sigma_t^2)^3}+\left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^4}\right)\begin{pmatrix}1 & y_t^2\\ y_t^2 & y_t^4.\end{pmatrix}\end{aligned}$$

The Fisher information matrix denoted by g is defined to be the negative of the expected value of the Hessian, that is,

$$g = -E[\tfrac{\partial^2 l}{\partial\theta\partial\theta'}]$$

since

$$E_{y_t|H_t}\left((\frac{y_t^2}{\sigma_t^2}-1)\frac{1}{\sigma_t^4}\begin{pmatrix}1 & y_{t-1}^2\\ y_{t-1}^2 & y_{t-1}^4\end{pmatrix}\right)=0$$

and

$$E_{y_t|H_t}\{\frac{y_t^2}{(\sigma_t^2)^3}\}=\frac{E_{y_t|H_t}(y_t^2)}{(\sigma_t^2)^3}.$$

It then follows that

$$g=\frac{1}{2}\sum_{t=2}^{T}\frac{1}{\sigma_t^4}\begin{pmatrix}1 & y_{t-1}^2\\ y_{t-1}^2 & y_{t-1}^4\end{pmatrix}$$

as in Engle, (1982). Non-linear optimization routines are iterative, thus if θ^i denotes the parameter estimates after the i^{th} iterations, then θ^{i+1} has the form

$$\theta^{i+1}=\theta^i+\lambda M^{-1}\{\frac{\partial l}{\partial\theta}\}$$

where λ is a step-length chosen to maximize the likelihood function in the direction $\frac{\partial l}{\partial\theta}$. For Newton Raphson based routines $\lambda=1$ and $M=\frac{\partial^2 l}{\partial\theta\partial\theta'}$ and for the Fisher scoring method $\lambda=1$ and $M=g$ (Mills,1994 and Engle, 1982).

4.2.3 Forecasting with the ARCH model

As stated in section 3.7 on ARIMA modelling, forecasting is one of the main aims of developing a time series model. The ARCH models also provide good estimates of the series before it is realized. We now provide the theory of forecasting with the ARCH models in detail. Let $y_1, y_2, ..., y_t$ be an observed time series, then the l-step ahead forecast, for $l=1,2,...,$ at the origin t, denoted as $y_t(l)$, is taken to be the minimum mean squared error predictor, that is, $y_t(l)$ minimizes

$$E(y_{t+l} - f(y))^2$$

where $f(y)$ is a function of the observations, then

$$y_t(l) = E[y_{t+l}|y_1, ..., y_t]$$

(Tsay, 2002). However for the ARCH(1) model

$$y_t(l) = E[y_{t+l}|y_1, ..., y_t] = 0$$

as seen in section 4.1. The forecasts for the y_t series provide no much helpful information. It is therefore important to consider the squared returns y_t^2 given as

$$y_t^2 = E[y_{t+l}^2|y_1^2, ..., y_t^2]$$

(Shephard, 1996). Hence the 1-step ahead forecast for y_t^2 is given by

$$y_t^2(1) = \hat{\alpha_0} + \hat{\alpha_1} y_t^2$$

which is equivalent to

$$\begin{aligned} \sigma_t^2(1) &= E(\sigma_{t+1}^2|y_t) \\ &= \hat{\alpha_0} + \hat{\alpha_1} y_t^2 \end{aligned}$$

where $\hat{\alpha_0}$ and $\hat{\alpha_1}$ are the conditional maximum likelihood estimates of α_0 and α_1. Similarly a 2-step ahead forecast for y_t^2 is given by

$$\begin{aligned} y_t^2(2) &= E[y_{t+2}^2|y_t] \\ &= E[\sigma_{t+2}^2|y_t] \end{aligned}$$

$$
\begin{aligned}
&= \hat{\alpha}_0 + \hat{\alpha}_1 E(y_{t+1}^2|y_t) \\
&= \hat{\alpha}_0 + \hat{\alpha}_1(\hat{\alpha}_0 + \hat{\alpha}_1 y_t^2) \\
&= \hat{\alpha}_0(1 + \hat{\alpha}_1) + \hat{\alpha}_1^2 y_t^2 \\
&= \sigma_t^2(2).
\end{aligned}
$$

In general the l-step ahead forecast for the y_t^2 is given by

$$
\begin{aligned}
y_t^2(l) &= E(y_{t+l}^2|y_t) \\
&= \hat{\alpha_0}(1 + \hat{\alpha_1} + \hat{\alpha}_1^2 + ... + \hat{\alpha}_1^{l-1}) + \hat{\alpha}_1^l y_t^2 \\
&= \sigma_t^2(l).
\end{aligned}
$$

The obvious possible problem in using the ARCH formulation is that the approach can lead to a highly parametric model if the lag q is large. This necessitated the introduction of the GARCH model as an extension to the ARCH model. Section 4.3 below gives an account of the theory of the GARCH modelling.

4.3 The GARCH model

Generalized ARCH (GARCH), as developed by Bollerslev (1986), is an extension of the ARCH model similar to the extension of an AR to ARMA process. When modelling using ARCH, there might be need for a large value for the lag q, hence a large number of parameters. This may result in a model with a large number of parameters, violating the principle of parsimony and this can present difficulties when using the model to adequately describe the data. An ARMA model may have less parameters

than an AR model, similarly a GARCH model may contain fewer parameters as compared to an ARCH model, thus a GARCH model may be preferred to an ARCH model. There are a variety of extensions of the ARCH family of models that include the exponential GARCH (EGARCH), the integrated GARCH (IGARCH). These are not discussed in this study. For an interested reader the thesis by Talke (2003) is a good source of such information. We now consider the GARCH(1,1) model a particular case of the GARCH(p, q) modelling.

4.3.1 GARCH(1,1) model

Under GARCH(1,1), the conditional variance depends not only on the previous returns, but also on the previous conditional variance. The GARCH(1,1) can formally be defined as follows: Let $\{y_t\}$ be the mean corrected return, ε_t be a Gaussian white noise with mean zero and unit variance. Let also H_t be the information set or history at time t given by $H_t = \{y_1, ..., y_{t-1}\}$ as in the ARCH model, then the process $\{y_t\}$ is GARCH(1,1) if

$$y_t = \sigma_t \varepsilon_t \tag{4.16}$$

and

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4.17}$$

.

The restrictions $\alpha_0 > 0, \alpha_1 \geq 0$, and $\beta_1 \geq 0$ are imposed in order for the variance σ_t^2 to be positive. Clearly equations (4.16) and (4.17) show that large past mean-corrected squared returns y_{t-1}^2 or past conditional variances σ_{t-1}^2 give rise to large values of σ_t^2 (Tsay, 2002). The conditional

mean $E(y_t|H_t) = 0$ implying that $\{y_t\}$ is a martingale difference, thus $E(y_t) = 0$ and $\{y_t\}$ is an uncorrelated series. At this point it is important to point out that the information set is now strictly given by

$$[y_1; \sigma_1^2, ..., y_{t-1}; \sigma_{t-1}^2].$$

Further ARCH(1) can also be defined as GARCH(0,1). Taking $v_t = y_t^2 - \sigma_t^2$, we can express

$$\begin{aligned} y_t^2 &= \sigma_t^2 + v_t \\ &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1(y_{t-1}^2 - v_{t-1}) + v_t \\ &= \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + v_t - \beta_1 v_{t-1}. \end{aligned} \tag{4.18}$$

As in ARCH(1) model $E(v_t|H_t) = 0$ and v_t is another martingale difference, meaning, $E(v_t) = 0$ and $Cov(v_t, v_{t-k}) = 0$ for $k \geq 1$ and v_t is serially uncorrelated. From equation (4.16)

$$\begin{aligned} E(y_t^2) &= \alpha_0 + (\alpha_1 + \beta_1) E(y_{t-1}^2) \\ &= \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \end{aligned}$$

provided $|\alpha_1 + \beta_1| < 1$.

4.3.2 Estimation of the GARCH(1,1) model

Estimation of the parameters of the GARCH(1,1) model is performed in the same way as in the ARCH(1) model. However, since the conditional variance of the GARCH(1,1) model depends also on the past conditional

variance, an initial value of the past conditional variance σ_1^2 is needed. The unconditional variance of y_t can be taken as an initial value for this variance, that is, σ_1^2 can be taken to be $\frac{\alpha_0}{1-\alpha_1-\beta_1}$. Sometimes the sample variance of the return series can be taken to be the initial value. As in the ARCH modelling the maximum likelihood estimates are obtained by maximizing the conditional log-likelihood given by

$$lnf(y_2, ..., y_t, \sigma_2, ..., \sigma_t | y_1, \sigma_1^2, \theta) = -\frac{1}{2}\sum_{t=2}^{t} ln(2\pi\sigma^2) - \frac{1}{2}\sum_{t=2}^{t}\frac{y_t^2}{\sigma_t^2} \quad (4.19)$$

where $\theta = (\alpha_0, \alpha_1, \beta_1)'$. The gradient, the Hessian and the optimization procedure are the same as the ARCH(1) modelling except that σ_t^2 has a different formulation.

4.3.3 GARCH(p, q)

The GARCH(p, q) is an generalization of GARCH(1,1) with p as the autoregressive lag and q is the moving average lag. Formally a process $\{y_t\}$ is GARCH(p, q) if

$$y_t = \sigma_t \varepsilon_t$$

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \\ &= \alpha_0 + \alpha(\mathbf{B}) y_t^2 + \beta(\mathbf{B}) \sigma_t^2 \end{aligned} \quad (4.20)$$

where ε_t is Gaussian white noise while $\alpha(\mathbf{B})$ and $\beta(\mathbf{B})$ are polynomials in the backshift operator given by

$$\alpha(\mathbf{B}) = \alpha_1 \mathbf{B} + ... + \alpha_q \mathbf{B}^q$$

and

$$\beta(\mathbf{B}) = \beta_1\mathbf{B} + ... + \beta_p\mathbf{B}^p$$

with the restrictions $\alpha_0 > 0, \alpha_i \geq 0$ and $\beta_j \geq 0$ for $i = 1, 2, ..., q$ and $j = 1, 2, ..., p$ being imposed in order to have the conditional variance remaining positive. Equation (4.20) can be expressed as

$$(1 - \beta(\mathbf{B}))\sigma_t^2 = \alpha_0 + \alpha(\mathbf{B})y_t^2.$$

Note that GARCH(0,q) model is the same as an ARCH(q) model and that for $p = q = 0$ we have a GARCH(0,0) model, which is simple, white noise.

Similar to the ARCH(q) model, the conditional mean of $\{y_t\}$ is zero, that is $E(y_t|H_t) = 0$ and hence the series $\{y_t\}$ is a martingale difference and observing $E(y_t) = 0$ implies the series $\{y_t\}$ is uncorrelated Gourieroux, et.al (1997). Assuming the GARCH(p, q) process is second order stationary, that is,

$$\begin{aligned} Var(y_t) &= E(y_t^2) \\ &= E(\sigma_t^2\varepsilon_t^2) \\ &= E(\sigma_t^2 E(\varepsilon_t^2|y_{t-1})) \\ &= E(\alpha_0 + \alpha(\mathbf{B})y_t^2 + \beta(\mathbf{B})\sigma_t^2) \\ &= \alpha_0 + \alpha(\mathbf{B})E(y_t^2) + \beta(\mathbf{B})E(\sigma_t^2) \\ &= \frac{\alpha_0}{1 - \sum_{i=1}^{q}\alpha_i - \sum_{j=1}^{P}\beta_j} \end{aligned} \tag{4.21}$$

The autocovariance of a GARCH(p, q) model for $k \geq 1$ where k is the lag is

$$E(y_t y_{t-k}) = 0$$

since y_t is a martingale difference Gourieroux, et.al (1997). Thus the GARCH(p,q) model does not show autocorrelation in the return series $\{y_t\}$. However the squared returns show autocorrelation even though the returns are not correlated. Considering writing y_t^2 in terms of $v_t = y_t^2 - \sigma_t^2$ yields

$$\begin{aligned} y_t^2 &= \sigma_t^2 + v_t \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j (y_{t-j}^2 - v_{t-j}) + v_t \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j y_{t-j}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t. \end{aligned} \tag{4.22}$$

Now let $m = max(p,q)$, then

$$y_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) y_{t-i}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t \tag{4.23}$$

where $\alpha_i = 0$ for $i > q$ and $\beta_j = 0$ for $i > p$. Thus the equation of y_t^2 has an ARMA(m,p) representation. In order to find the GARCH(p,q) process, we consider solving for α_0 in equation (4.21) and let the variance of y_t be σ_ε^2 which yields

$$\alpha_0 = \sigma_\varepsilon^2 (1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j) \tag{4.24}$$

and substituting the value of α_0 in equation (4.24) into equation (4.23) gives

$$
\begin{aligned}
y_t^2 &= \sigma_\varepsilon^2[1-\sum_{i=1}^{m}(\alpha_i+\beta_j)]+\sum_{i=1}^{m}(\alpha_i+\beta_i)y_{t-i}^2-\sum_{j=1}^{p}\beta_j v_{t-j}+v_t \\
&= \sigma_\varepsilon^2+\sum_{i=1}^{m}(\alpha_i+\beta_i)(y_{t-i}^2-\sigma_\varepsilon^2)-\sum_{j=1}^{p}\beta_j v_{t-j}+v_t. \qquad (4.25)
\end{aligned}
$$

Therefore

$$
y_t^2-\sigma_\varepsilon^2=\sum_{i=1}^{m}(\alpha_i+\beta_i)(y_{t-i}^2-\sigma_\varepsilon^2)-\sum_{j=1}^{p}\beta_j v_{t-j}+v_t \qquad (4.26)
$$

.

Multiplying both sides by $(y_{t-k}^2-\sigma_\varepsilon^2)$ yields

$$
\begin{aligned}
(y_{t-k}^2-\sigma_\varepsilon^2)(y_t^2-\sigma_\varepsilon^2) &= \sum_{i=1}^{m}(\alpha_i+\beta_i)(y_{t-i}^2-\sigma_\varepsilon^2)(y_{t-k}^2-\sigma_\varepsilon^2) \\
&- \sum_{j=1}^{p}\beta_j v_{t-j}(y_{t-k}^2-\sigma_\varepsilon^2)+v_t(y_{t-k}^2-\sigma_\varepsilon^2)
\end{aligned}
$$

and taking expectations

$$
\begin{aligned}
E[(y_{t-k}^2-\sigma_\varepsilon^2)(y_t^2-\sigma_\varepsilon^2] &= E[\sum_{i=1}^{m}(\alpha_i+\beta_i)(y_{t-i}^2-\sigma_\varepsilon^2)(y_{t-k}^2-\sigma_\varepsilon^2)] \\
&- E[\sum_{j=1}^{p}\beta_j v_{t-j}(y_{t-k}^2-\sigma_\varepsilon^2)]+E[v_t(y_{t-k}^2-\sigma_\varepsilon^2)].
\end{aligned}
$$

But

$$
E[v_t(y_{t-k}^2-\sigma_\varepsilon^2)]=E[(y_{t-k}^2-\sigma_\varepsilon^2)E(v_t|y_t)]=0
$$

since v_t is a martingale difference and also

$$
E[\beta_j v_{t-j}(y_{t-k}^2-\sigma_\varepsilon^2)]=E[(y_{t-k}^2-\sigma_\varepsilon^2)E(v_{t-j}|y_{t-k})]=0 \qquad (4.27)
$$

for $k < j$. Thus the autocovariance of the squared returns for the GARCH(p, q) model is given by

$$\begin{aligned} Cov(y_t^2, y_{t-k}^2) &= E[\sum_{i=1}^{m}(\alpha_i + \beta_i)(y_{t-k}^2 - \sigma_\varepsilon^2)(y_{t-k}^2 - \sigma_\varepsilon^2)] \\ &= \sum_{i=1}^{m}(\alpha_i + \beta_i)Cov(y_t^2, y_{t-k+i}^2). \end{aligned} \tag{4.28}$$

Dividing both sides by $var(y_t^2)$ gives the autocorrelation function at lag k as

$$\rho_k = \sum_{i=1}^{m}(\alpha_i + \beta_i)\rho_{k-i} \tag{4.29}$$

for $k \geq p + 1$. This result is analogous to the Yule-Walker equations for an AR process. Thus the autocorrelation function (ACF) and the partial ACF (PACF) of the squared returns in a GARCH(p, q) process has the same pattern as those of an ARMA(m, p) process. As in ARMA modelling the ACF and the PACF are useful in identifying the orders p and q of the GARCH(p, q) process. The ACF are also important for checking model adequacy, in which case, the ACF's of residuals should be indicative of a white noise process if the model is adequate. Thus the first p autocorrelations depend on the parameters $\alpha_1, \alpha_2, ..., \alpha_p; \beta_1, \beta_2, ..., \beta_q$, but given $\rho_p, ..., \rho_{p+1-m}$, the expression in (4.29) determines uniquely the autocorrelations at higher lags, (Bollerslev, 1986). Thus letting ϕ_{mm} denotes the m^{th} partial autocorrelation for y_t^2, then

$$\rho_k = \textstyle\sum_{i=1}^{m} \phi_{mi}\rho_{k-i}, k = 1, ..., m$$

By equation (4.29) ϕ_{mm} cuts off after lag q for an ARCH(q) process such that $\phi_{mm} \neq 0$ for $k \leq q$ and $\phi_{mm} = 0$ for $k > q$. This is identical to the

PACF for an AR(q) process and decays exponentially (Bollerslev, 1986). After identifying the orders p and q, we now can estimate the parameters $(\alpha_1, \alpha_2, ..., \alpha_p; \beta_1, \beta_2, ..., \beta_q)$ of the GARCH(p, q) model as explained in section (4.3.4) below.

4.3.4 Estimation of GARCH(p, q) model

The maximum likelihood estimation is also applicable in estimating the parameters of the model. As in the GARCH(1,1) model estimation, initial values of both the squared returns and past conditional variances are needed in estimating the parameters of the model. As suggested by Bollerslev, (1986) and Tsay, (2002), the unconditional variance given in equation (4.21) or the past sample variance of the returns for the past variances may be used as initial values. Therefore assuming $y_1, ..., y_q$ and $\sigma_1^2, ..., \sigma_p^2$ are known, the conditional maximum likelihood estimates can be obtained by maximizing the conditional log-likelihood given by

$$\begin{aligned} l &= lnf(y_{q+1}, ..., y_t, \sigma_{p+1}^2, ..., \sigma_t^2 | \theta, y_1, ..., y_q, \sigma_1^2, ..., \sigma_p^2) \\ &= -\frac{1}{2} \sum_{t=m+1}^{T} ln(2\pi\sigma_t^2) - \frac{1}{2} \sum_{t=m+1}^{T} \frac{y_t^2}{\sigma_t^2} \end{aligned}$$

with $\theta = (\alpha_0, ..., \alpha_q, \beta_1, ..., \beta_p)$ and $m = max(p, q)$

4.4 Model checking

Goodness of fit of the ARCH-GARCH model are based on residuals and more specifically on the standardized residuals (Talke, 2003). The residuals are assumed to be independently and identically distributed

following either a normal or a standardized t-distribution (Tsay, 2002) and (Gourieroux, 1997). Plots of the residuals such as the histogram, the normal probability plot and the time plot of residuals can be used. If the model fits the data well the histogram of residuals should be approximately symmetric. The normal probability plot should be a straight line while the time plot should exhibit random variation ($http://en.wikipedia.org/wiki/Q-Q-plot$). The ACF and the PACF of the standardized residuals are used for checking the adequacy of the conditional variance model. The lagrange multiplier and the Ljung Box Q-test (given in section 3.5) are used to check the validity of the ARCH effects in the data. Having established that our model fits the data well, we can now use the fitted model to compute forecasts just the same way as in the ARIMA modelling.

4.4.1 Forecasting with GARCH(p, q) models

As we have outlined, the GARCH and the ARMA models are similar, such that forecasting using the GARCH model is the same as using the ARMA model. Thus the conditional variance of $\{y_t\}$ is obtained simply by taking the conditional expectation of the squared mean corrected returns. Assuming a forecasting origin of T, then the 1-step ahead volatility forecast is given by

$$\begin{aligned} y_t^2(1) &= E[y_{t+1}^2|y_t] \\ &= \alpha_0 + \sum_{i=1}^{m}(\alpha_i + \beta_i)E(y_{t+1-i}^2|y_t) - \sum_{i=1}^{p}\beta_i(v_{t+1-i}|y_t) \end{aligned}$$

where $y_t^2, ..., y_{t+1-m}^2$, $\sigma_t^2, ..., \sigma_{t+1-p}^2$ are assumed known at time t and the true parameter values α_i and β_i for $i = 1, ..., m$ are replaced by their estimates. Furthermore, the l-step ahead forecast of the conditional variance in a GARCH(p, q) model is given by

$$\begin{aligned}\sigma_t^2(l) &= E(y_{t+l}^2|y_t) \\ &= \alpha_0 + \sum_{i=1}^{m}(\alpha_i + \beta_i)E(y_{t+l-i}^2|y_t) - \sum_{i=1}^{p}\beta_i(v_{t+l-i}|y_t)\end{aligned}$$

where $E(y_{t+l-i}^2|y_t)$ for $i < l$ can be given recursively as

$$E(y_{t+l-i}^2|y_t) = y_{t+l-i}^2 \text{ for } i \geq l$$

$$E(v_{t+l-i}|y_t) = 0 \text{ for } i < l$$

$$E(v_{t+l-i}|y_t) = v_{t+l-i} \text{ for } i \geq l.$$

We now consider the techniques that are used for selecting the best fitting model in light of two or more competing models based on the likelihood ratios.

4.5 Model selection criteria

There are many model selection procedures or criteria for deciding amongst the competing ARCH-GARCH models. However the most common are the AIC defined (as in section 3.2) by

AIC=-2(loglikelihood)+2(number of parameter)

or provided the sample is large the SBC given as

SBC=-2(log-likelihood)+(number of parameter)[log(number of observations)]

. A desirable model is one that minimizes the AIC or the SBC. The other criteria is the R^2 associated with the model which is the proportion of variability in a data set that is accounted for by the statistical model. However, as Harvey, (1991) indicated that R^2 has a limitation in that a model that can pick out the trend reasonably well will have an R^2 close to unit. In this study the R^2 is used alongside the AIC and the SBC to be on the safer side. In general a model selected by two different criteria mentioned above may differ and thus it should be emphasized that the selection of an ARCH-GARCH model depends on the selection criteria used (Talke, 2003).

4.6 Forecasting Performance

As pointed out in section 3.7, forecasting is one of the most important objectives of time series modelling, thus one of the criteria for selecting the best model can be based or centered on the best forecasting model. Forecast in this context means the estimates of the conditional variances obtained from the models. There are several measures for assessing the predictive accuracy of an ARCH-GARCH model. Among these are the mean square error (MSE). The MSE is defined as the average of the squared difference between the actual variance and volatility forecast σ_t^2. In the absence of the observed true variance the squared time series observations y_t^2 is used. The MSE is thus given by

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t^2 - \hat{\sigma}_t^2)^2$$

where $\hat{\sigma}_t^2$ for $t = 1, ..., T$ is the estimated conditional variance obtained from fitting ARCH-GARCH model. The MSE is criticized, as in Tsay, (2002), in that, although the y_t^2 is a consistent estimator of σ_t^2, it is nevertheless noisy, therefore unstable. Among the alternative measures are the mean absolute error (MAE) by Lopez, (1999) defined by

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t^2 - \hat{\sigma}_t^2|$$

and the MSE of the log of the squared error defined by

$$MSLE = \frac{1}{T} \sum_{t=1}^{T} (ln(y_t^2) - ln(\hat{\sigma}_t^2))^2 \tag{4.30}$$

The advantage of the MSE of the log of the squared error is that it penalizes inaccurate variance forecasts more heavily when the squared innovations y_t^2 is low.

Chapter 5

Application of ARIMA and ARCH-GARCH models

5.1 Introduction

Having explored the theory of ARIMA and ARCH-GARCH models in the preceding chapters, this chapter is dedicated to fitting the two families of models to the South African inflation rate data which we obtained curtesy of the South African Central Bank. A description of the data is given in section (5.2) below and application of ARIMA and ARCH-GARCH are given in section 5.3 and in section (5.4) respectively. Analysis was done using the following SAS version 9.1 procedures: ARIMA, AUTOREG, REG and FORECASTING procedures.

5.2 Data description

As mentioned in section (5.1) above, the data was obtained from the South African Reserve Bank. These are monthly inflation rates spanning

from January 1994 to December 2008, which covers the post-apartheid period. We defined the rate of inflation as the dynamic percentage rate of change of prices of an economic (representative) basket of goods and services of a typical South African household over time. That is, the dynamic rate of decrease in the purchasing power of money. In South Africa, the inflation rate is calculated as an index number termed the Consumer Price Index (CPI). The CPI is a yardstick of the increase in the general level of prices in the economy. The total SA CPI basket consists of about 1500 different goods and services which are classified into more than 40 groups and subgroups for which separate indices are calculated. Statistics South Africa construct the CPI according to the following five steps:

- the selection of goods and services to be included in the basket
- the assignment of weights to each good or service to indicate its relative importance in the basket,
- the choice of a base year for calculating the CPI,
- the choice of a formula for calculating the CPI,
- the collection of prices each month to calculate the value of the CPI for that month.

The inflation rate is obtained by calculating the monthly percentage change of average prices from one period to the next. The formula for calculating the inflation can be formally defined as follows. Let P_t be the current average price level of an economic basket of goods and services and P_{t-1} be the average price level of the basket a year ago, then the inflation rate I_t at time t is calculated as:

Table 5.1: Covariances and Correlations at each lag

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
Covariance	22.4	21.6	21.3	20.7	20.0	19.3	18.6	18.0	17.4	16.6	16.0	15.4	14.7
correlation	1	.96	.95	.92	.89	.86	.83	.80	.77	.74	.71	.68	.65

$$I_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100\%.$$

The rate is usually given in annualized terms. That is, a 10% inflation rate means prices are increasing at a rate of 10% per year at time t. This precisely is the reason why we used the term 'dynamic' in our definition to emphasize the time dependence in the inflation rate calculation.

5.3 Application of the ARIMA modelling

Figure (5.1) shows the time plot of the inflation rate in South Africa from January 1994 to December 2008. The plot shows that the series is characterized by a nonconstant mean and unstable variance over time. It also shows a slight decreasing trend in the years between 1994 and 2004 followed by a steady rise from 2004 to 2008. The sinusoidal pattern suggests the presence of seasonality which is one of the components in a time series and we wish to account for it in modelling. The autocorrelations given in Table (5.1) show large correlations at large lags. The trend, seasonality effects, the nonconstant mean and unstable variance together with the large autocorrelations are indicative of nonstationarity as was defined in section (2.2). The immediate challenge here is that the Box-Jenkins technique which we wish to apply in this section assumes

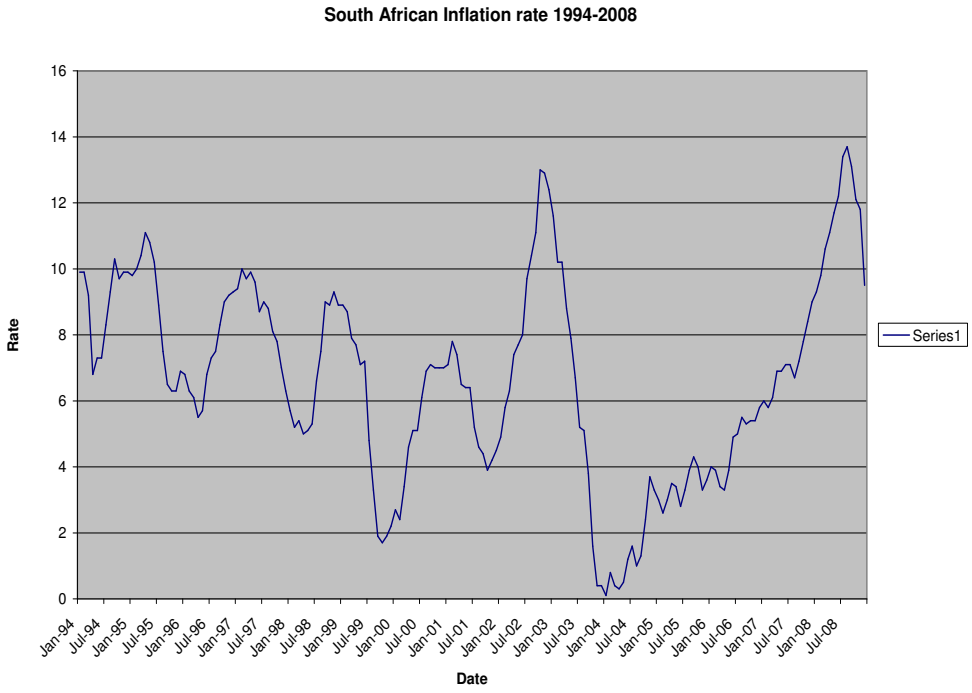

Figure 5.1: Time Plot of Inflation in South Africa

stationarity, hence to apply this technique, there is need to transform the data to make it stationary by for example taking transformations as explained in section (2.5) in order to stabilize the variance and achieve constant mean.

The first transformation we considered is the ordinary differencing and seasonal differencing as explained in sections (2.5) and (2.4). Figure (5.2)

Table 5.2: Covariances and Correlations of first differencing of inflation at each lag

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
Covariance	1.6	-.5	.3	0	.1	0	0	-.1	.1	0	-.1	.1	-1.0
correlation	1	-.3	.2	0	0	0	0	-.1	.1	0	-.05	.1	-.4

Table 5.3: Optimal order of differencing

order-of-differencing	Non	First	Second	Third
standard-deviation	4.734	1.271	1.459	1.816

gives the plot of the differenced (both ordinary and seasonal) series, and the plot shows near constant mean and stable variance. This shows differencing is necessary to make the series stationary. Table (5.3) gives the order of differencing together with the associated standard deviations. Box-Jenkins rule of thumb states that the optimum order of differencing is the one with the smallest standard deviation, hence in our case the optimum order of differencing is one with standard deviation of 1.271 which corresponds to order one. The autocorrelations and covariance for the first ordinary and seasonal differencing of inflation is given in Table (5.2). The autocorrelations tails-off fast with oscillating positive and negative values which is overwhelming evidence of stationarity.

The next transformation considered is the log transformation, as explained again in section (2.5). The time plot of the logarithm transformations is given in figure (5.3). The plot shows the mean and the variance still changing markedly with time, which is indicative of nonstationarity.

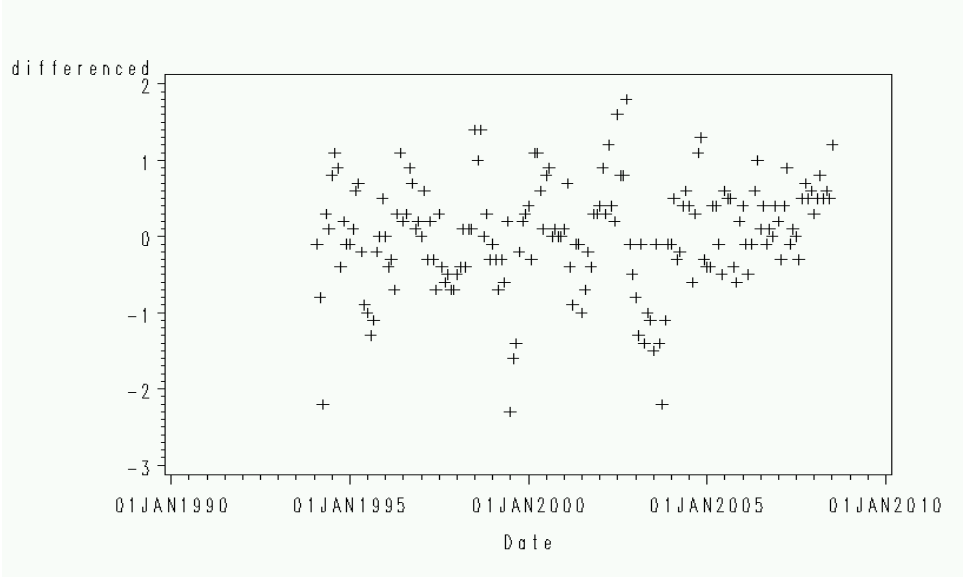

Figure 5.2: Plot of ordinary and seasonally difference of the inflation

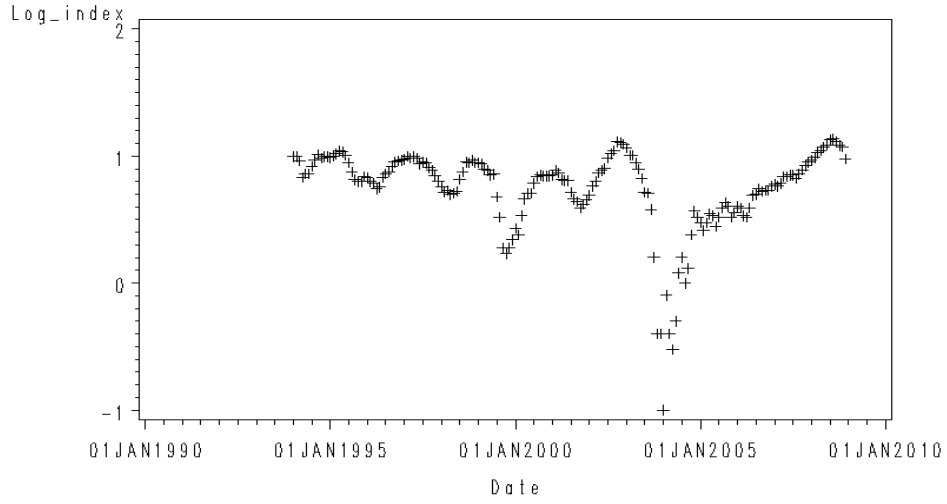

Figure 5.3: Logarithms of Inflation

Hence log transformation seems inappropriate for stabilizing the variance of the series. An alternative transformation is the square root transfor-

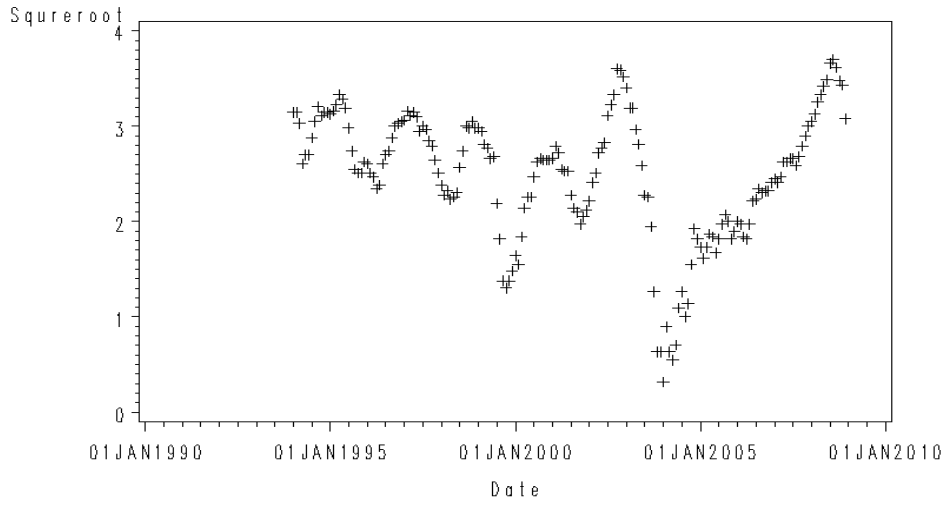

Figure 5.4: Square root of Inflation

mation. A time plot of the square root transformation of the inflation is given in figure (5.4). The plot also exhibit marked change in mean and variance with time, not very different from the untransformed series and the log-transformed series. Thus square root transformation also proves inadequate in transforming the data to stationarity.

Therefore looking at the transformations taken, the differencing seems to give the most stable variance and near constant mean. Thus in our subsequent analyses the differenced series is going to be considered or used. Nonetheless the plots of the log and the square root are provided for comparison purposes only. Now that we have adopted differencing as a means to achieve stationarity the next step is to employ the Box

Table 5.4: Model identification criteria

Model	ACF	PACF
$AR(1)$	decays exponentially	single spike
$MA(1)$	single spike	decays exponentially
$AR(p)$	decays exponentially with damped oscillations	p spikes
$MA(q)$	q spikes	decays exponentially with damped oscillations
$ARMA(p,q)$	decay exponentially and damped oscillations	both decay exponentially

Jenkins modelling techniques to identify an appropriate model that fits the inflation series. Thus section (5.3.1) below is dedicated to model identification.

5.3.1 Model Identification

Autocorrelations functions and partial autocorrelations functions discussed in section (2.1) are the sample statistics utilized in identifying an appropriate ARMA model. Model identification procedures allow us to build up a model systematically taking into account the main features exhibited in the sample ACF and sample PACF. Table (5.4) gives the criteria for model identification given by Kendall and Ord, (1990.) A close look at the ACF plot in figure (5.5) obtained from the series up to lag 12, it can be seen that there is evidence of exponential decay and damped oscillations. This may suggest the presence of both AR and MA parameters, hence an ARMA process with both ordinary and seasonal terms can be considered. There is a large spike at lag 12 which suggest a

seasonal parameters. Thus in our analyses, one of the model to be fitted is the ARMA model to X_t where $X_t = \Delta\Delta_{12}y_t$. The identification process was done using the ARIMA procedure in SAS version 9.1. We fit a number of models as suggested by the ACF and PACF selection criteria and choose the best model using the criteria explained in section 5.1

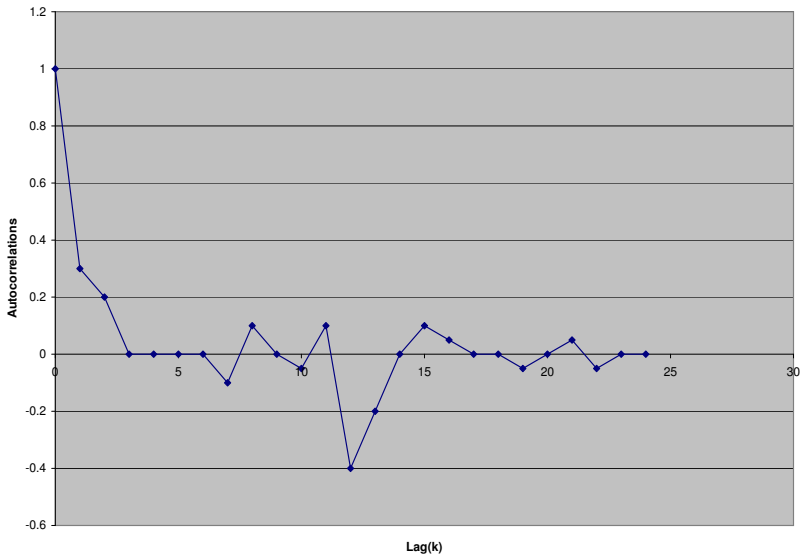

Figure 5.5: ACF of the ordinary and seasonally differenced series

5.3.2 Selection of the best fit ARIMA model

In light of a number of competing models, the criteria used to select the best fit model to the data are the AIC, SBC, and the significant tests of the individual parameters as explained in section 3.2. A parsimonious model (one with fewer parameters) and a lower AIC and SBC is the best

fit model. Our aim is to obtain the values of p, q, d, P, Q and D of the SARIMA model, where P and Q represents the seasonal AR and MA orders respectively and D represents the order of seasonal differencing. The analysis was also done in SAS using proc ARIMA. As a rule of thumb, any lag with an ACF or PACF coefficient whose magnitude is not within $\frac{2}{\sqrt{N}}$ where N is the size of the series is taken to be significantly different from zero, (Chatfield, 2004). With reference to Table (5.2), the significant spikes were found to be at lag 0, 1 and 12 and also 24 (not provided in the table). A large spike at lag 12 and a small spike at lag 24 may indicate a seasonal MA term. After several trial and error we established that the intercept is insignificant thus the models suggested will not have the constant term. The parameter estimates, standard errors and the significant test probabilities given in Table (5.5) are obtained from SAS ARIMA procedure. The values that gave a model of best fit are, $p = 1$, $q = 0$, $P = 0$, $Q = 1$, $d = 1$ and $D = 1$ with $s = 12$ since we have monthly data. Therefore we have seasonal ARIMA$(1, 1, 0) \times (0, 1, 1)_{12}$ model. Taking $X_t = \Delta\Delta_{12}Y_t$ the general form of the model as suggested by the results in Table (5.5) is:

$$(1 - \alpha_1\mathbf{B})Y_t = (1 + \beta_1\mathbf{B}^{12})\varepsilon_t$$

where $\alpha_1 = 0.528$, $\beta_1 = 0.998$, $(\mathbf{B})Y_t = Y_{t-1}$ and $(\mathbf{B}^{12})\varepsilon_t = \varepsilon_{t-12}$ implying that

$$\begin{aligned}
(1 - 0.528\mathbf{B})Y_t &= (1 + 0.998\mathbf{B}^{12})\varepsilon_t \\
\Delta\Delta_{12}Y_t &= 0.528\Delta\Delta_{12}Y_t + \varepsilon_t + 0.998\varepsilon_{t-12} \\
\Delta_{12}Y_t - \Delta_{12}Y_{t-1} &= 0.528(\Delta_{12}Y_t - \Delta_{12}Y_{t-1}) + \varepsilon_t +
\end{aligned}$$

Table 5.5: Parameter estimates for an SARIMA$(1,1,0) \times (0,1,1)_{12}$

parameter	AR(p=1)	MA(Q=1)
Estimate	0.52763	0.99814
Standard error	0.06948	0.01169
$P(>\|t\|)$	< 0.0001	< 0.0001

$$
\begin{aligned}
& 0.998\varepsilon_{t-12} \\
y_t - Y_{t-12} - Y_{t-1} + Y_{t-13} &= 0.528[Y_t - Y_{t-12} - Y_{t-1} + Y_{t-13}] \\
+\varepsilon_t + 0.998\varepsilon_{t-12} & \\
0.472(Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}) &= \varepsilon_t + 0.998\varepsilon_{t-12} \\
\hat{Y}_t &= Y_{t-1} + Y_{t-12} - Y_{t-13} + 2.12\varepsilon_t \\
+2.11\varepsilon_{t-12} & \qquad (5.1)
\end{aligned}
$$

The standard errors are used to assess the accuracy of the estimates, the smaller the better. The model fit statistics used to asses how well the model fit the data are the AIC and SBC. The corresponding values are $AIC = 411.3$ and $SBC = 423.7$. Table (5.6) gives the correlations of the parameter estimates. The values of the correlations between model parameters are low showing that they do not explain the same variation in the model. The parameters are estimated by using the least squares method in proc ARIMA statement in SAS version 9.1.

The second model to be considered was the ARIMA(0,1,1), also without the intercept since it was found to be insignificant. Here $p = P = Q = 0$ and $d = q = 1$. The model is given by:

Table 5.6: Correlations of parameter estimates

parameter	AR(p=1)	MA(Q=1)
AR(p=1)	1	0.215
MA(Q=1)	0.215	1

Table 5.7: Parameter estimates for an ARIMA(0,1,1)

Parameter	estimate	standard error	t-value	$P(>\|t\|)$
$q = 1$	-0.37972	0.07192	-5.28	< 0.0001

$$\begin{aligned} \Delta\Delta_{12}Y_t &= \beta_1\varepsilon_{t-1} + \varepsilon_t \\ Y_t - Y_{t-1} - Y_{t-12} - Y_{t-13} &= \beta_1\varepsilon_{t-1} \\ \hat{Y}_t &= Y_{t-1} + Y_{t-12} + Y_{t-13} + \beta_1\varepsilon_{t-1} \end{aligned} \tag{5.2}$$

The parameter estimates from SAS ARIMA procedure are given in Table (5.7) giving rise to the model

$$\hat{y}_t = y_{t-1} - 0.37972\varepsilon_{t-1} \tag{5.3}$$

$$\Delta\Delta_{12}y_t = -0.37972\varepsilon_{t-1} + \varepsilon_t$$

$$y_{t-1}(1) = y_{t-1} + y_{t-12} + y_{t-13} - 0.37972\varepsilon_{t-1}$$

The corresponding fit statistics are: $AIC = 423.7$ and $SBC = 454.9$. Thus the ARIMA(0,1,1) has bigger AIC and SBC values than the seasonal ARIMA$(1,1,0) \times (0,1,1)_{12}$ indicating that the seasonal model is

Table 5.8: Parameter estimates for an ARIMA(2,1,1)

Parameter	estimate	standard error	t-value	$P(>\|t\|)$
α_1	0.63095	0.66828	0.30	0.3464
α_2	0.01553	0.34496	0.94	0.9641
β_1	0.19586	0.66338	0.05	0.7682

superior to the MA model (Kendall and Ord, 1990). A corresponding seasonal ARIMA$(0,1,1)\times(0,1,0)_{12}$ was also seen not to fit the data well. Another candidate model tested was the ARIMA(2,1,1). Here $p=2$, $d=q=1$ and $P=Q=D=1$ implying no seasonal differencing. The proc ARIMA analysis show a nonsignificant intercept. Thus the model we finally fit is given by

$$\begin{aligned} \Delta\Delta_{12}y_t &= \alpha_1 y_{t-1}+\alpha_2 y_{t-2}+\beta_1\varepsilon_{t-1} \\ y_t - y_{t-1} &= \alpha_1 y_{t-1}+\alpha_2 y_{t-2}+\beta_1\varepsilon_{t-1} \\ \hat{y}_t &= (\alpha_1+1)y_{t-1}+\alpha_2 y_{t-2}+\alpha_1\varepsilon_{t-1} \end{aligned} \tag{5.4}$$

The Parameter estimates as given by the SAS ARIMA procedure are given in Table (5.8). The suggested model is:

$$\hat{y}_t = 1.6309y_{t-1}+0.01553y_{t-2}+0.19586\varepsilon_{t-1}.$$

Significant tests via p-values show that all the parameters are nonsignificant. The fit statistics are: $AIC=504.6$ and $SBC=581.6$ which are also greater than the seasonal ARIMA$(1,1,0)\times(0,1,0)_{12}$ model suggesting that the SARIMA$(2,1,1)\times(0,1,0)$ is less appropriate than the

Table 5.9: Comparison of selected models

model	estimate	$P(>\|t\|)$	AIC	SBC
ARIMA(1,1,0)	$\beta_1 = 0.4904$	< 0.0001	421.3	443.4
$\mathbf{SARIMA(1,1,0) \times (0,1,1)_{12}}$	$\boldsymbol{\alpha_1 = 0.528}$	$\mathbf{< 0.0001}$	**411.3**	**423.7**
$\mathbf{SARIMA(1,1,0) \times (0,1,1)_{12}}$	$\boldsymbol{\beta_{12} = 0.998}$	$\mathbf{< 0.0001}$	**411.3**	**423.7**
SARIMA $(0,1,1) \times (0,1,0)_{12}$	$\alpha_1 = -0.37972$	< 0.0001	423.7	454.9
SARIMA $(2,1,1) \times (0,1,0)_{12}$	$\beta_1 = 0.63095$	0.3464	504.6	581.6
SARIMA $(2,1,1) \times (0,1,0)_{12}$	$\beta_2 = 0.01553$	0.9641	504.6	581.6
SARIMA $(2,1,1) \times (0,1,0)_{12}$	$\alpha_1 = 0.19586$	0.7682	504.6	581.6
SARIMA $(1,1,1) \times (0,1,0)_{12}$	$\beta_1 = 0.6602$	0.1315	430.2	463.8
SARIMA $(1,1,1) \times (0,1,0)_{12}$	$\alpha_1 = 0.2244$	< 0.0001	430.2	463.8
SARIMA $(0,1,12) \times (0,1,0)_{12}$	$\alpha_1 = -0.19282$	0.0017	499.1	509.8
SARIMA($0,1,12) \times (0,1,0)_{12}$	$\alpha_{12} = 0.61162$	< 0.0001	499.1	508.9

seasonal ARIMA$(1,1,0) \times (0,1,0)_{12}$ model. Note also that the parameters α_1, α_2 and β_1 under SARIMA$(2,1,1) \times (0,1,0)_{12}$ are nonsignificant. At this point in time, it is established that, amongst all the identified models, the SARIMA$(1,1,0) \times (0,1,1)_{12}$ proves the best fit model. Section 5.3.3 below provides a comprehensive comparison of the identified models.

5.3.3 A comparison of the fitted ARIMA models

Table (5.9) shows the competing ARIMA models together with the corresponding estimates, standard errors, significant probabilities, AIC and

SBC for each parameter. The SARIMA$(1,1,0)\times(0,1,1)$ has the smallest AIC=411.3 and SBC=423.7 and both of the parameters are significant. Therefore this model is selected to be the best fit of all the models fitted. All the other models have greater AIC and SBC, therefore they were provided only for comparison purposes. The interpretation of the selected model is that, holding all other factors constant, this month's inflation rate is a linear function of the previous monthly inflation rate and that of the twelve and thirteenth months earlier and some innovation terms. The term with lag 12 looks plausible since it was found that there is a significant seasonal variation in concordance with monthly collected data. Having constructed our model, the next task is to assess how well the model fits the data. This is done through residual analysis or model checking as explained in section (3.4).

5.3.4 Diagnostic checking of the seasonal ARIMA$(1,1,0)\times(0,1,1)_{12}$ model

Residuals from a model that fits the data well should be white noise, that is, they should have mean zero, uncorrelated and portray unsystematic uniformly random variability over time, (Chatfield, 1989). Table (5.10) gives the results of the autocorrelation check for residuals at the respective lags. The table shows that none of the chi-square-statistics are statistically significant implying that there is no important information that is left in the residuals. The ARIMA procedure in SAS version 9.1 also gives the Q statistic for the Box-Pierce-Ljung test to check for the lack of fit as explained in section (3.5). The Q statistics at all lags were found to be non-significant indicating absence of autocorrelation in the

Table 5.10: Autocorrelation check for residuals for the SARIMA model

lag	χ^2 statistics	df	Pr> χ^2
6	1.80	2	0.4070
12	13.70	8	0.0898
18	17.36	14	0.2374
24	27.50	20	0.1217
30	32.38	26	0.1809

Table 5.11: Time plot of residuals from the SARIMA model

residuals, further supporting evidence that the model selected fits the data well. Figure (5.11) gives the time plot of the residuals. From the plot, the residuals look randomly distributed around zero and exhibit no clear pattern. The pattern of residuals exhibited in figure (5.11) together with the Q-statistics give overwhelming evidence that the residuals are independent implying the model fits the data well. The residuals can also be checked for normality by looking at the histogram of the residuals. If the residuals are near normally distributed, the histogram should be symmetric and bell-shaped. Figure (5.6) gives the histogram of the residuals, which is almost symmetric. Now that we have identified and estimated a model that fits the data, the next step is to use the model to forecast future values of the series, which ideally is the principle goal of time series modelling and analysis.

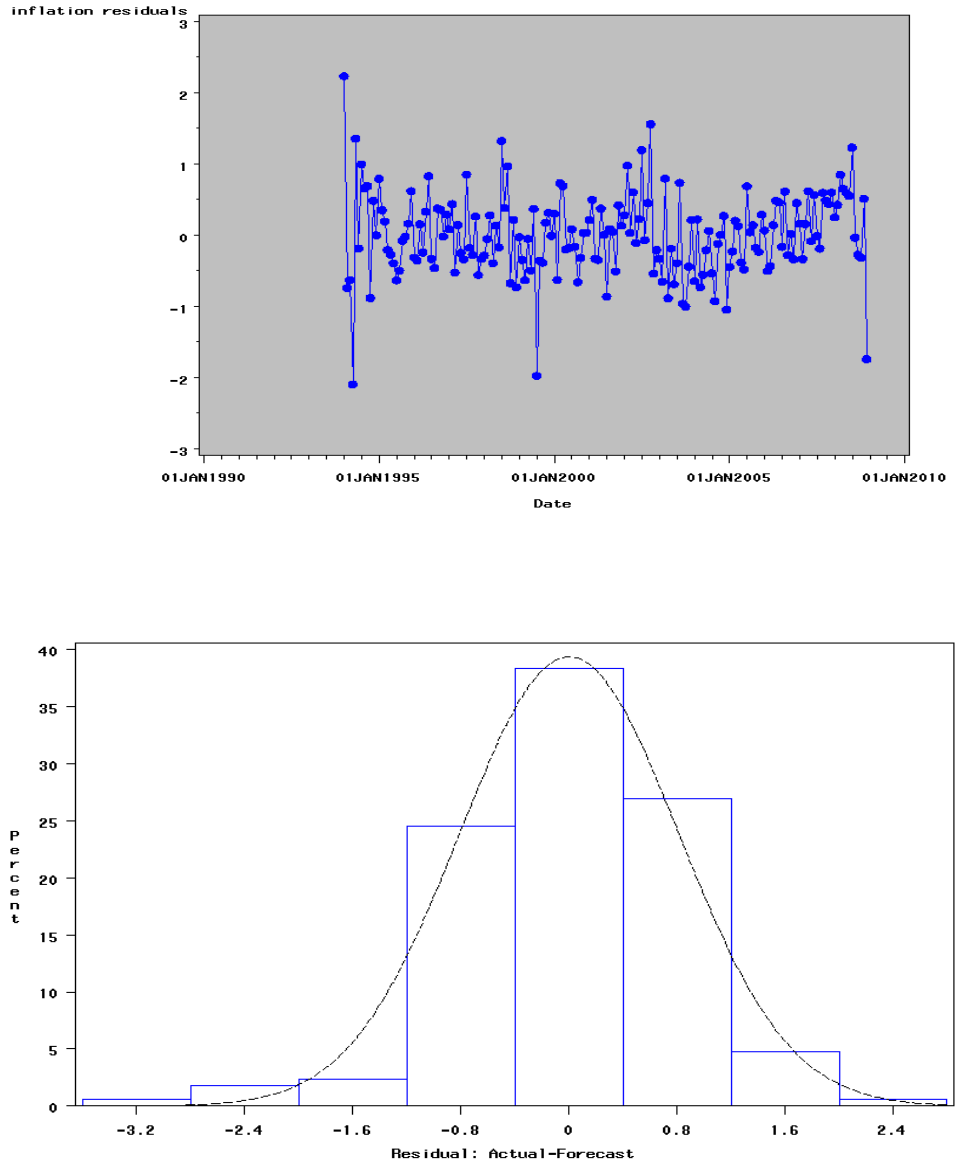

Figure 5.6: Histogram of residuals from the SARIMA model

5.3.5 Forecasting with the ARIMA model

One of the objectives of this thesis is to use the constructed model to attempt forecasting future inflation rate before they are realized, which is

thus the main focus of this section. The forecasting techniques discussed in section (3.7) are employed as procedures for obtaining the forecasts. Analysis is done using SAS proc ARIMA and proc FORECAST. The Holt-Winters forecasting procedure that was discussed in section (3.7.2) is used since it has the capacity to cope with both trend and seasonality. The seasonal ARIMA$(1,1,0) \times (0,1,1)_{12}$ model given above was used to generate the forecast for inflation given in table (5.12) for the years 2009 and 2010. The lower and upper 95% confidence limits are used to graphically assess how good the forecasts are. This implies that the forecasts are expected to lie within the confidence limits with 95% confidence. As expected, the further into the future a forecast is the less precise it is, hence the wider the confidence limits. The forecasts are also given on a plot together with the confidence limits as shown in figure (5.7). The confidence limits look wide indicating that the model has low forecasting power although it fits the data well. The models which account for possible heteroscedasticity are now evaluated.

5.4 Application of the ARCH-GARCH modelling

We now cosider applying the ARCH-GARCH modelling to the inflation rate data. First consider the time plot of inflation rate given in figure (5.1). As discussed in section (5.3) the plot shows changing mean and changing variance with time, implying nonstationarity. The changing variance nature of the data is also known as heteroscedasticity. As was explained in detail in section (4.1), ARCH models describe the variance

Table 5.12: Two year forecasts of inflation obtained from the SARIMA model

Date	Forecast	Std. error	Lower 95% CL	Upper 95% CL	Interval width
Jan -2009	8.7	0.8	7.1	10.3	3.2
Feb -2009	8.1	1.4	5.4	10.8	5.4
Mar-2009	8.2	1.9	4.5	11.9	7.4
Apr-2009	8.7	2.3	4.2	13.3	9.1
May-2009	8.8	2.7	3.4	14.2	10.7
Jun-2009	9.0	3.1	3.0	15.1	12.1
Jul-2009	9.3	3.4	2.6	16.1	13.4
Aug-2009	9.1	3.7	1.8	16.4	14.6
Sep-2009	9.3	4.0	1.4	17.1	15.7
Oct-2009	9.4	4.3	1.1	17.8	16.7
Nov-2009	9.8	4.5	1.0	18.7	17.7
Dec-2009	9.6	4.7	0.3	18.9	18.6
Jan -2010	9.6	5.0	-0.2	19.5	19.8
Feb -2010	9.9	5.3	-0.6	20.4	21.0
Mar-2010	10.5	5.6	-0.6	21.6	22.2
Apr-2010	11.0	5.9	-0.6	22.6	23.3
May-2010	11.5	6.2	-0.7	23.6	24.4
Jun-2010	11.9	6.5	-0.8	24.6	25.5
Jul-2010	12.8	6.7	-0.4	26.1	26.5
Aug-2010	13.0	7.0	-0.7	26.7	27.5
Sep-2010	12.6	7.2	-1.6	26.8	28.4
Oct-2010	11.9	7.5	-2.7	26.6	29.3
Nov-2010	11.8	7.7	-3.3	26.9	30.2
Dec-2010	10.1	7.9	-5.4	25.7	31.1

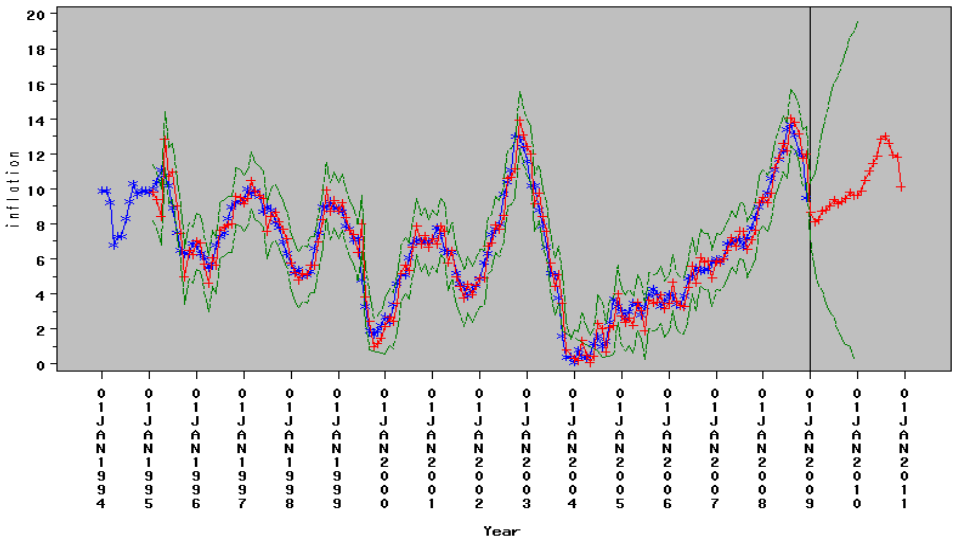

Figure 5.7: Plot of inflation together with forecasts from SARIMA model and 95% confidence interval

of the current error term as a function of the variances of the previous time periods' error terms. ARCH-GARCH modelling accommodates heteroscedasticity, hence there is no need for taking transformations as a pre-requisite for stationarity. Before applying the ARCH-GARCH modelling a formal test for heteroscedasticity was carried out in order to establish the presence of ARCH effect in the data. This was done using the Lagrange multiplier (LM) test and the Q-test as explained in section (4.1). SAS AUTOREG procedure in SAS version 9.1 was used to perform the tests. The null hypothesis of homoscedasticity, the opposite of heteroscedasticity was tested and Table (5.13) gives the results of the SAS AUTOREG procedure for the Q-test and Lagrange Multiplier (LM) test for heteroscedasticity. All the p-values show that the Q-values and the LM-values at all lags are significant giving overwhelming evidence of pres-

Table 5.13: Test for heteroscedasticity

Lag	Q-value	$(P > Q)$ CL	LM	$P >$LM
1	158.6532	< 0.0001	156.2002	< 0.0001
2	276.9920	< 0.0001	162.0527	< 0.0001
3	354.5199	< 0.0001	162.2445	< 0.0001
4	398.8997	< 0.0001	162.3780	< 0.0001
5	420.4065	< 0.0001	162.4967	< 0.0001
6	428.4414	< 0.0001	162.5365	< 0.0001
7	430.6922	< 0.0001	163.0859	< 0.0001
8	431.1828	< 0.0001	163.1714	< 0.0001
9	431.3162	< 0.0001	163.1794	< 0.0001
10	431.4116	< 0.0001	163.1812	< 0.0001
11	431.5616	< 0.0001	163.1917	< 0.0001
12	431.9856	< 0.0001	163.3124	< 0.0001

ence of heteroscedasticity in the data, hence ARCH-GARCH modelling was deemed appropriate. According to Engle, (1982) any autocorrelations in the series have to be removed before an ARCH-GARCH model is constructed. This was done by regressing the squares of the series y_t on its past squared values $y_t^2, y_{t-1}^2, ...$ with the number of lags determined by the form of the ACF and the PACF. The ACF suggested an AR(2) process, thus an AR(2) model was used resulting in all autocorrelations being removed. Hence we consider fitting the ARCH-GARCH models to the data.

5.4.1 Selection of the best fitting ARCH-GARCH model

The strategy used in selecting the appropriate model from competing models is based on the AIC, SBC, R^2, MSE and on the significance tests for each parameter as discussed in section (4.5). SAS proc AUTOREG was used to perform trial and error to determine the best fitting model. The GARCH model with AR errors is given by

$$y_t = \sigma_t \varepsilon_t \tag{5.5}$$

where

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{5.6}$$

The order of the parameters are determined by studying the ACF and the PACF in the same way as was done in the ARIMA modelling. The method used to estimate the parameters is the maximum likelihood method. Table (5.14) gives the suggested models with their respective fit statistics. As was seen in section (3.2), the idea is to have a parsimonious model that captures as much variation in the data as is possible. Therefore in table (5.14), the smaller the AIC, SBC and the MSE and the larger the R^2 the better. The model given in bold was judged to be the most appropriate according to the criteria above. Therefore we adopted the model

$$y_t = \sigma_t \varepsilon_t \tag{5.7}$$

for

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{5.8}$$

Table 5.14: Comparison of suggested ARCH-GARCH models with intercepts

Model	AIC	SBC	MSE	R^2
$GARCH(0,1)$	403.498	419.463	0.540	0.8926
GARCH$(1,1)$	**357.555**	**376.713**	**0.404**	**0.957**
$GARCH(0,2)$	358.355	377.706	0.393	0.928
$GARCH(1,2)$	359.532	381.912	0.540	0.943
$GARCH(2,0)$	357.555	376.713	0.404	0.957
$GARCH(2,1)$	369.781	392.144	0.393	0.951
$GARCH(2,2)$	373.912	402.705	0.393	0.951

Before the parameters are estimated, the same set of models are considered but without the intercept. Table (5.15) gives the models together with selection criteria. Although there is an improvement in the R^2, the large AIC's, BIC's and the MSE's makes these models unfavorable. Most of the parameter estimates for these models were also found to be nonsignificant. Therefore the GARCH(1,1) in table (5.14) is chosen to be the most appropriate, hence we proceed to estimate the associated model parameters.

5.4.2 Estimating parameters of the GARCH(1,1)

The first step is to estimate α_0, α_1 and β_1 as explained in detail in section (4.3.4). SAS proc AUTOREG was used to obtain these estimates specifying the maximum likelihood estimation as the method in the model statement. Table (5.16) gives the parameter estimates, the standard er-

Table 5.15: Comparison of suggested ARCH-GARCH models without intercepts

Model	AIC	SBC	MSE	R^2
$GARCH(0,1)$	487.98	503.945	1.10	0.9795
$GARCH(1,1)$	486.1	505.26	1.09	0.9797
$GARCH(0,2)$	485.02	504.18	1.11	0.9794
$GARCH(1,2)$	513.58	532.7	1.06	0.9803
$GARCH(2,0)$	587.98	503.9	1.11	0.9795
$GARCH(2,1)$	494.6	516.9	1.11	0.9793
$GARCH(2,2)$	515.58	541.1	1.15	0.9785

rors and p-values. It can be seen from the table that all the parameter estimates are statistically significant, that is, different from zero. The standard errors are quiet small suggesting quite precise estimates. The AR terms are provided since an AR(2) model was fitted in order to remove autocorrelations present in the original series. Table (5.16) suggests that equations (5.7) and (5.8) can now be written as

$$y_t = \sigma_t \varepsilon_t$$

where

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 y_t^2 + \beta_1 \sigma_{t-1}^2 \\ &= 0.3605 + 0.1141 y_t^2 + 0.6247 y_t^2. \end{aligned} \tag{5.9}$$

Having estimated our parameters, the next step is to check how well the model fits the data. As with ARIMA modelling, this was done through

Table 5.16: Parameter estimates for GARCH(1,1)

Parameter	Estimate	Standard-error	t-values	$Pr > \|t\|$
intercept	9.7780	1.8768	5.21	< 0.0001
$AR(1)$	-1.5065	0.0726	-20.75	< 0.0001
$AR(2)$	0.5492	0.0703	7.81	< 0.0001
α_0	0.3605	0.0436	8.27	< 0.0001
α_1	0.1141	0.1013	7.13	0.0259
β_1	0.6247	0.1389	4.50	< 0.0001

residual analysis.

5.4.3 Diagnostic checking of the GARCH(1,1) model

As was explained in detail in section (4.4) model checking is done through analyzing the residuals from the fitted model. If the model fits the data well, the residuals are expected to be random, independent and identically distributed following the normal distribution. The time plot of the residuals given in figure (5.8) is used to check whether the residuals are random. The normality check is also done by analyzing the histogram of residuals and the normal probability plot. Figure (5.9) gives the histogram of the residuals from the GARCH(1,1) model. The histogram shows almost a symmetric bell-shaped distribution which is indicative of the residuals following a normal distribution. The slight negative skewness is expected since the residuals may also come from student's t distribution. The negative skewness tendency is also supported by the negative large residuals in figure (5.8). The normality check of the residuals can

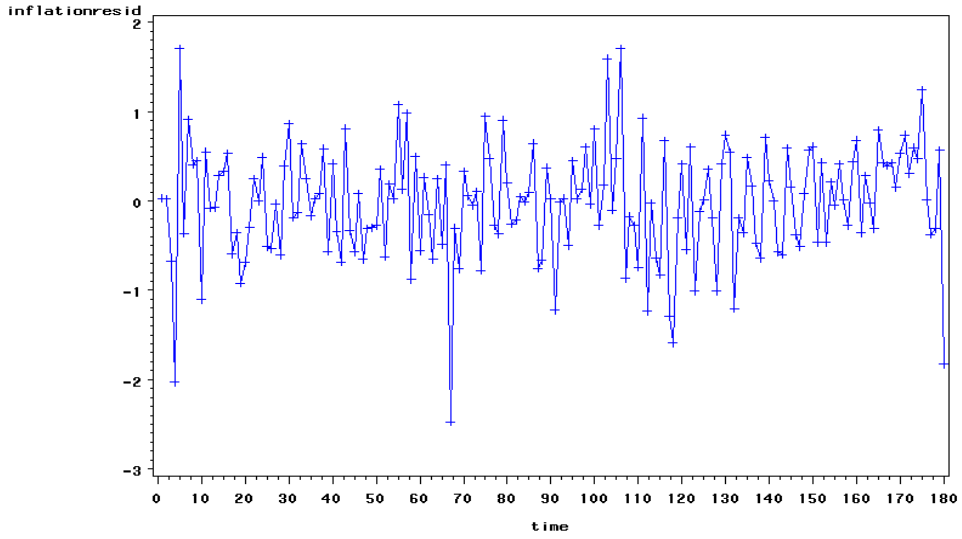

Figure 5.8: Time plot of residuals from GARCH(1,1)

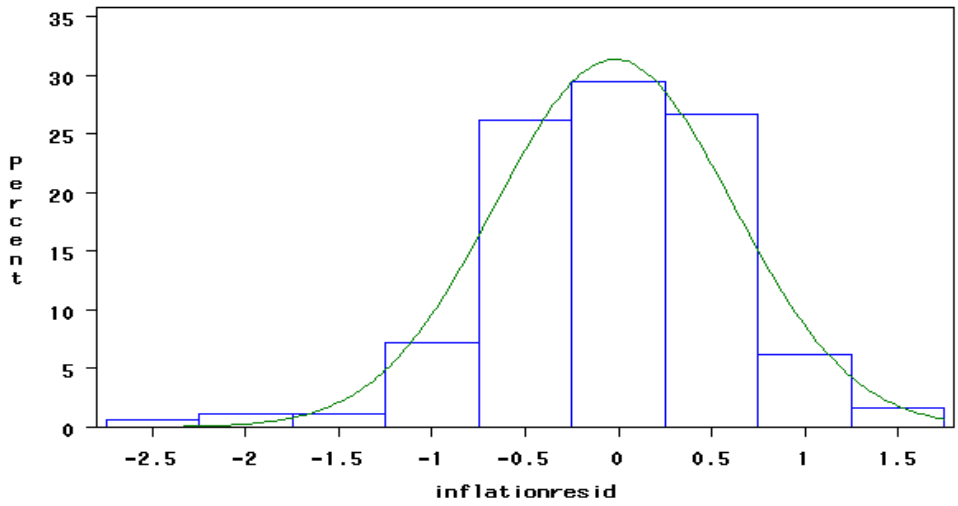

Figure 5.9: Histogram of residuals from GARCH(1,1)

also be done by the normal probability plot. As was seen in section (4.4), if the residuals come from the normal distribution the plot should be a straight line. Figure (5.10) gives the probability plot of the residuals. The plot shows a near straight line suggesting that the residuals follow an approximately normal distribution but a student t distribution cannot be ruled out. Therefore we proceed to use the model to forecast future values of the inflation series.

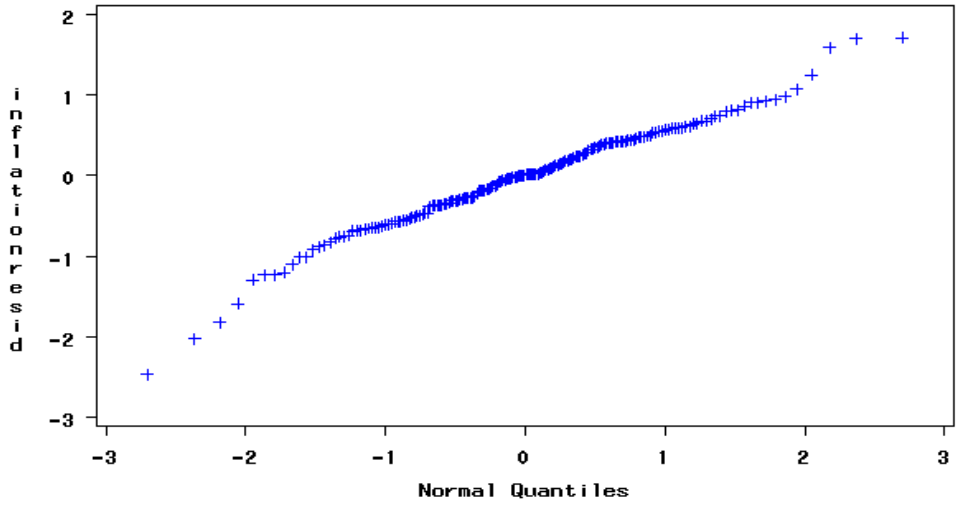

Figure 5.10: Normal probability plot of residuals from the GARCH(1,1)

5.4.4 Forecasting with the GARCH(1,1) model

As explained in section (5.3.5), forecasting is one of our main objectives of constructing a time series model and also of this research. A model that fits the data well is bound to give good forecasts. Figure (5.11) gives the time plot of inflation together with the 95% confidence interval and

the two year forecasts, that is, from January 2009 up to and including December 2010. The 95% confidence intervals provide a measure of accuracy in the forecasts. The wider a confidence interval, the less accurate the forecasts are. The confidence widths are much narrower than those based on the SARIMA model indicating that the ARCH-GARCH model has a reasonably strong forecasting power. The forecasts can also be

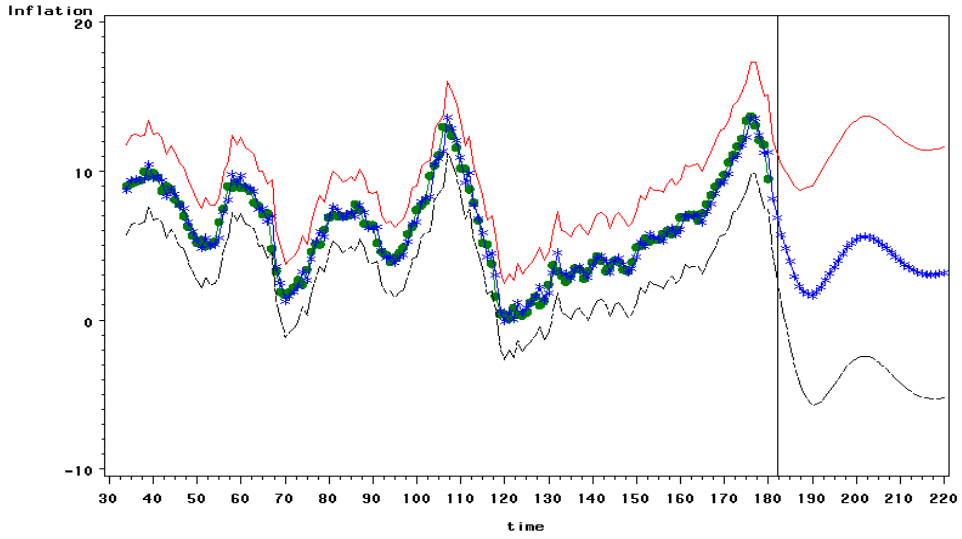

Figure 5.11: Time plot of inflation, 95% confidence intervals and two year forecasts from GARCH(1,1)

represented in a table as in the ARIMA case and hence table (5.17) gives the forecasts. From the table it can easily be seen that the widths of the confidence interval increase the further a forecast is into the future. Having identified, estimated and validated a model from each family of models, the subject of discussion of chapter 6 below is to give comparisons of the models based on their goodness of fit and forecasting power.

Table 5.17: Two year forecasts of inflation obtained from the GARCH(1,1) model

Date	Forecast	Lower 95% CL	Upper 95% CL	Interval width
Jan-2009	8.2	4.3	12.0	7.7
Feb-2009	8.8	2.6	11.1	8.5
Mar-2009	5.7	1.0	10.4	9.4
Apr-2009	4.8	-0.3	9.9	10.3
May-2009	4.0	-1.6	9.6	11.2
Jun-2009	2.9	-3.1	9.0	12.1
Jul-2009	2.3	-4.2	8.7	12.9
Aug-2009	1.9	-5.0	8.8	13.7
Sep-2009	1.7	-5.4	8.9	14.3
Oct-2009	1.6	-5.7	9.0	14.8
Nov-2009	1.9	-5.7	9.5	15.2
Dec-2009	2.2	-5.5	10.0	15.5
Jan-2010	2.7	-5.2	10.5	15.6
Feb-2010	3.1	-4.7	11.0	15.7
Mar-2010	3.5	-4.3	11.5	15.8
Apr-2010	4.0	-3.9	11.9	15.8
May-2010	4.4	-3.4	12.4	15.9
Jun-2010	4.8	-3.0	12.8	15.9
Jul-2010	5.2	-2.8	13.2	15.9
Aug-2010	5.4	-2.6	13.4	16.0
Sep-2010	5.6	-2.4	13.6	16.0
Oct-2010	5.6	-2.4	13.7	16.1
Nov-2010	5.6	-2.5	13.7	16.1
Dec-2010	5.4	-2.6	13.6	16.2

Chapter 6

Comparisons of the ARIMA and ARCH-GARCH models and conclusions

6.1 Introduction

The ARIMA and ARCH-GARCH modelling are both time series modelling procedures. However, the two have various differences. The aim of this chapter is to give a comparison of the two families of models, both from a theoretical and practical point of view. The comparisons are presented in the form of an account of the steps in the development of the thesis in brief since the theories and characteristics of each of the family of models were given in detail in the preceding chapters. As was explained in chapter 4, the main differences come about due to the assumptions and characteristics of the models, particularly the assumption of constant variance. We now present the comparisons in the theory and

the results as brought about by the findings in the research.

6.2 Comparisons of the ARIMA and ARCH-GARCH models

The main objective of the research was to understand the theory of time series analysis (in particular the ARIMA and ARCH-GARCH modelling) and apply the models to modelling rate of inflation in South Africa, and compute forecasts using the models. This involved discussing the stages in the model formulations of the models, exploring the Box-Jenkins three stage model formulation (that is identification, estimation and checking) technique applied to both the ARIMA and the ARCH-GARCH families of models. The theory behind each stage was explained in depth. As we explained in chapter 3, the tools used in the identification stage (for identifying the parameters of the models) are the plots and tables of the autocorrelations. In particular the plots and tables of the ACF (correlogram) were used. Where two or more models are identified, criteria based on likelihood functions such as the AIC and SBC were used to select the best fitting model.

The theory of the methods used for estimating the parameters of the models are given in chapter 3 and are applied in chapter 5. The methods are the least squares method and the maximum likelihood method under the normality assumption. Plots of residuals from the estimated models and significance test via the p-values are used to validate the goodness of fit of the estimated models. As explained in the chapters 3 and 5 resid-

uals from a model that fit the data well should not contain significantly important information that could be used to explain the variation in the data. An account of the theory of forecasting and forecasting methods was outlined for both family of models. Particular emphasis was given, as in section (3.7), on methods that accommodate the different components of the time series. These methods are the Holt and Holt-Winters forecasting methods which capture trend and seasonality. The methods of evaluating the accuracy of the forecasts from the estimated models were given in section (3.8).

The two family of models were applied to the South African inflation rate data. Every stage of the developments (that is the identification, estimation and checking) of the models was explained in detail in chapter 5. Plots and tables of the results analyzed using SAS version 9.1 procedures were given to enhance the explanations, however some plots were produced in Excel. An analysis of the inflation rate data showed that the data is characterized by changing mean level and unstable variance, with a dominant trend and seasonal variation. This prompted us to fit ARIMA models with both trend and seasonal terms in order to capture these variations. ARCH-GARCH models were also established to be plausible as they accommodate the time-varying variance nature of the data, hence a GARCH(1,1) model was fit to the data. The best fitting ARIMA and ARCH-GARCH models are SARIMA$(1,1,0)\times(0,1,1)$ and GARCH(1,1) respectively.

Analysis of the ARIMA and ARCH-GARCH family of models to the inflation rate series has shown that the ARCH-GARCH modelling is su-

Table 6.1: AIC and BIC for SARIMA$(1,1,0) \times (0,1,1)_{12}$ and GARCH(1,1)

Statistic	SARIMA(1,1,0)×(0,1,1)	GARCH(1,1)
AIC	411.03	351.5
BIC	423.7	376.7

perior than the ARIMA modelling although the two capture almost the same variation in the series. The Box-Jenkins ARIMA model is simple to construct but it has a shortfall of losing some observations through ordinary and seasonal differencing. The ARCH-GARCH models have a demerit of having very little theory available hence they are difficult to construct. With reference to sections (5.3.2) and (5.4.1), table (6.1) can be constructed to summarize the difference in how well the two models fit the data using the AIC and the BIC criteria. The GARCH model has smaller AIC and BIC which is indicative it explains the variation in the data better than the seasonal ARIMA model.

The models were both used to compute two year forecasts for the inflation series. The forecasts from the seasonal ARIMA and GARCH models are given in tables (5.12) and (5.17) respectively together with their respective 95% confidence intervals (CI) for each forecast value. As was explained in sections (5.3.5) and (5.4.4), the confidence intervals give the forecasting accuracy of the models. The narrower the confidence interval the better the forecasts, (Granger and Newbold,1986 and Granger, 1989). The CI's from the ARIMA model are narrower than the CI's from the GARCH model in the early months of the forecasting period

and became wider the further into the future a forecast is. This probably indicate that the ARIMA model is better for short term forecasting than the GARCH model. However figures 5.7 and 5.11 clearly indicate that projecting into the future is better under the GARCH than ARIMA, the later having relatively wider confidence limits than the former.

6.3 Conclusions

The study has presented us with an opportunity to have an extensive understanding of the theory of time series analysis and its application to a real life situation. Stages in the model building strategy have been fully explored and utilized. It is clear from the study that the South African inflation rate series is characterized by both trend and seasonal variations. Hence a model that captures these variations is appropriate.

Although the ARIMA model captured these variations, the need to transform the data to stationarity makes the model rely on rigid assumptions resulting in the GARCH model being superior. The GARCH model fits the data well, however further studies can be considered as extensions and improvements to the GARCH model. These are the integrated GARCH (IGARCH), the exponential GARCH (EGARCH) and the stochastic volatility models. Although the application in this research is based on inflation (financial) data, other areas of application include environmental and pollution data (Peng and Dominici, 2008), health research in the context of longitudinal data (Diggle et al., 2002; Molenberghs and Verbeke, 2005), geostatistics and agriculture just to mention a few. Future work will include extending the methodology to

include data which is not necessarily Gaussian to cater for time series in the form of counts and other types.

Bibliography

[1] Akaike, H (1974) A new look at the statistical model identification. *IEEE transactions on automatic control* volume 19 number 6 page 719-723.

[2] Akinboade, O.A, Niedermeier, E.W and Siebrits, F.K (2001) Inflation dynamics: implications for policy. *Discussion paper*, Department of Economics, University of South Africa.

[3] Ansley, C.F (1979) An algorithm for the exact likelihood of a mixed autoregressive moving average processes. *Biometrica volume 66 page 59-65.*

[4] Ansley, C.F and Newbold, P (1986) Stochastic processes and their applications, *Science direct* volume 11 no. 2 page 201-206

[5] Aron, J and Muellbaur, J (2004) Construction of CPIX data for forecasting and modelling in South Africa. *Economic journal of South Africa, vol 72 no. 5, page 884-912.*

[6] Bomberger, W.A and Makinen, G.E (1979) Inflation, unemployment and the expectations in Latin America: some simple test. *Southern economic journal, vol 45, no. 4, page 1299-1302.*

[7] Bollerslev, T (1986) Generalized autoregressive conditional heterescedasticity. *Journal of econometrics* volume 51 pages 307-327.

[8] Bollerslev, T, Engle, R.F and Nelson, D.B (1993) ARCH models. *Discussion paper* University of California San Diego.

[9] Box, G.E and Cox, D (1964) An analysis of transformations. *Journal of the Royal Statisticial Society, Series B* Number 26 pg 211-252.

[10] Box, G, Jenkins G.M and Gregory C.R (1976), *Time series analysis: Forecasting and control* 4th Edition. Wiley series, probability and statistics, Prentice Hall Inc., London.

[11] Box, G.E.P and Pierce, D.A (1970) Distribution of the autocorrelation in ARIMA time series models. *Journal of American statistical association* volume 65 page 1509-1526.

[12] Brown, R.G (1962) *Smoothing, forecasting and prediction of discrete time series*. Prentice Hall Inc., London.

[13] Burger, P and Marinkov, M (2008) Inflation targeting and inflation performance in South Africa. *Discussion paper* Trade and industrial policy strategy.

[14] Chatfield, C (1989) *The analysis of time series: an introduction*, 2nd edition. Chapman and Hall text in statistical science, London.

[15] Chatfield, C (2004) *The analysis of time series: an introduction* 3rd edition. Chapman and Hall text in statistical science, London.

[16] Campbell, J.Y, Lo, A.W and Mackinlay, A.C (1997) Modelling daily value-at-risk using realized volatility and ARCH type models, *Journal of empirical finance* volume 11, no. 3 pages 379-398

[17] Diggle, P, Heargerty, P and Liang, K.Y (2002) *Analysis of longitudinal data*, Oxford University press.

[18] Engle, R.F (1982) Autoregressive Conditional Heteroscedasticity with estimates of variance of the United Kingdom inflation. *Econometrica* vol 4 number 50 page 987-1006.

[19] Engle R (1982) The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspective* vol 15 number 4 pg 157-168.

[20] Gourieroux, C (2001) *Financial Econometrics, Problems, models and methods*. Princeton series in finance, New Jersey, USA .

[21] Granger, C.W.J (1989) *Forecasting in Business and Economics*, 2^{nd} edition. Academic press Inc USA.

[22] Granger, C.W.J and Newbold, P (1986) *Forecasting Economic time series* 2^{nd} edition. Economic theory, Econometrics and Mathematical Economics, Academic Press, USA

[23] Gourieroux, C, Monfort, A and Philips, P.C (1997) Time series and dynamic models, Cambridge university press, United Kingdom.

[24] Harvey, A.C (1981) *Time series models*. Philip Allan publishers limited, Oxford.

[25] Harvey, A.C (1991) Forecasting, structural time series models and kalman filter. Cambridge university press.

[26] Kendall, M and Ord, K.J (1990) *Time series*. 3rd Edition, Hodder Arnold, Lonon.

[27] Kohn D.L (2005), Modeling inflation: A policymaker's perspective, *International research forum on monetary policy coference*, Frankfurt.

[28] Ljung, G.M and Box, G.E.P (1978) On a measure of lack of fit in time series models, *Biometrika* volume 65 page 297-303.

[29] Lopez, J.A (1999) Evaluating the predictive accuracy of volatilty models, *Economic research, Department of federal reserve bank of San Francisco*

[30] Mann H.B and Wald, A (1943) On the statistical treatment of linear difference equations. *Econometrica, Journal of econometric society* Volume 11 number 3 page 173-22o.

[31] Martin, D.R and Yohai, V.J (1986) Influence functions for time series, *Annals of statistics, volume 14 Number 3 pages 838-840.*

[32] Mboweni, T (2007) The role of forecasting in monetary policy formulation, *Reuters Economist of the Year Award* Johannesburg.

[33] Meyler,A, Kenny, G and Quinn, T (1998) Forecasting Irish iflation using ARIMA models. University library of Munich, German, MPRA paper.

[34] Mills, T.C (1994) *Time series techniques for economists*. Cambridge University Press.

[35] Molenberghs, G and Verbeke, G (2005) *Models for discrete longitudinal data*, Springer Series in Statistics, New York.

[36] Muhammed, A.S, Shazia, S and Feridun, M (2005) Modelling and forecasting Pakistnan's inflation, *working paper, State bank of Pakistan*.

[37] Peng R.D and Dominici F (2008) *Statistical methods for Environmental Epidemiology with R* Springer Series in Statistics, New York.

[38] Rangan, G and Josine, U (2009) Comparing South African inflation volatility across monetary policy regimes: an application of saphe cracking. *Department of economics working paper*, University of Pretoria.

[39] SAS version 9.1.

[40] Schwart, G.E (1978) Estimating the dimension of a model, *Annals of statistics* volume 6 number 2 page 461-464.

[41] Shephard, N (1996) Statistical aspects of ARCH and stochastic volatility time series in Econometrics, finance and other fields, Chapman and hall, London.

[42] Stokes, G (2009), FAnews on line *South Africas premier financial and advisory news and information portal*.

[43] Talke, I. S (2003) Modelling volatility in time series data, MSc thesis, University of Kwa-Zulu Natal.

[44] Test of normality. $http://en.wikipedia.org/wiki/Q-Q-plot$.

[45] Tsay R.S (2002), *Analysis of financial time series* 2nd Edition. New York, John Wiley and Sons.

[46] www.statisticalforecasting.com, Statistical forecasting: The fine tuned art of estimation.

[47] Zeger, S.L and Liang, K.Y, (1991) Feedback models for discrete and continuous time series. *Statistica Sinica* Vol 1 page 51-64.

[48] Zeger, S.L and Qaqish, B (1988) Markov regression models for time series: a quasi-likelihood approach, *Biometrics vol 44 number 4*.