Forecasting Stock Returns: A Comparison of Five Models
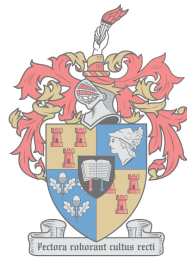
by

Vhahangwele Cedrick Ramuada

Thesis presented in partial fulfilment of the requirements
for the degree of Master of Science in Mathematics in the
Faculty of Science at Stellenbosch University

Supervisor: Prof. Jeff Sanders

Co-supervisor: Prof. Ronald Becker

December 2018

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: . . . . December 2018

Abstract

Forecasting Stock Returns: A Comparison of Five Models

Vhahangwele Cedrick Ramuada

Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MSc

October 2018

Forecasting the movement of stock returns prices has been of interest to researches for many decades. Due to the complex and chaotic nature of the stock market, it has been difficult for researches to find a model which can be used to accurately predict the movement of stock returns prices. Many statistical models have been proposed for forecasting the direction of movement of stock returns prices. The objective of this study was to use ARMA-type models and an Artificial Intelligence Neural Network model to predict the direction of movement of stock returns prices of four JSE listed companies, namely, Netcare Group Ltd, Santam Ltd, Sanlam Group Ltd, and Nedbank Group. The models were assessed in terms of their ability to predict whether the next day's returns price will go *down* or *up*.

Four ARMA-type models, namely, ARMA-Maximum Likelihood, ARMA-State Space, ARMA-Metropolis Hastings, AR(3)-AVGARCH(1,1)-Student-t model and an Artificial Neural Network (ANN) model were implemented to try to predict the direction of movement of stock returns prices. Historical (past) stock returns prices were used to make inference about future directional movement of stock returns prices. Empirical results show that the ARMA-Maximum Likelihood, ARMA-State Space, AR(3)-AVGARCH(1,1)-Student-t model, and Artificial Neural Network (ANN) models have a strong

ability to predict whether the next day's returns price will go *down* or *up* with acceptable accuracy. However, the ARMA-Metropolis Hastings model performed very poorly, its highest accuracy was a mere 68%. Overall, empirical results show that the Artificial Neural Network model was superior or outperformed all the ARMA-type models, the highest accuracy achieved by the model was 89%. The results of the Superior Ability Test also showed that the ANN model was indeed superior to the Box-Jenkins ARMA type models in at least 5 cases.

Uittreksel

Vooruitskatting van voorraadopbrengste :´ n Vergelyking van vfy modelle

("Forecasting Stock Returns: A Comparison of Five Models")

Vhahangwele Cedrick Ramuada

Departement Wiskuudige Wetenskappe,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MSc

Oktober 2018

Die voorspelling van die beweging van voorraad opbrengs pryse, is van groot belang vir navorsing vir dekades. As gevolg van die komplekse en chaotiese natuur van die aandele mark, dit mooilik vir navorsers om ´ n model te vind wat gebruik kan word om akkurate voorspelling van die beweging van die voorraad opbrengs pryse te maak. Verskeie statistiese modelle is voorgestel om rigting van beweging te voorspel van die aandele opbrengs prys. Die doel van hierdie studie was om die ARMA- tipe model en ´ n "kunsmatige intelligensie neurale netwerk"(Artificial Intelligence Neural Network) model te gebruik om die rigting van beweging van aandele obrengs prys van vier JSE genoteerde maatskappye te voorspel; naamlik, Netcare Group Ltd, Santam Ltd, Sanlam Group Ltd, and Nedbank Group. Die modelle is beoordeel in terme van hul vermoë om te voorspel of die volgende dag se pryse sal op of afwaarts gaan.

Vier ARMA-tipe modelle, naamlik ARMA-Maksimum Waarskynlikheid, ARMA-Staat Ruimte, ARMA- Metropolis Hastings, AR(3)-AVGARCH(1,1)-Student-t modelle en ´ n Kunsmatige Neurale Network (Artificial Neural Network :

ANN) model is geimplementeer om die bewegingsrigting van aandele op-
brengs pryse te voorspel. Historiese aandele pryse is gebruik om afleidings
te maak oor toekomstige rigtingbewegings van aandele pryse.

Gebaseer op ondervinding die resulte bewys dat die ARMA-Maksimum
Waarskynlikheid, ARMA-Staat Rruimte, AR(3)-AVGARCH(1,1)-Student-t
Modelle en Kunsmatige Neutral Netwerk (ANN) modelle ´n sterk vermöe
het, om die volgende dag se obrengs pryse af of hoër te voorspel met aan-
vaarbare akkuraatheid. Nietemin, die ARMA-Metropolis Hastings modelle
het baie swak gevaar , die hoogste akkuraatheid was ´n blote 68%. In die
algemeen, gebaseer op ondervinding die resultate wys dat die ANN model
beter was en die ARMA-tipe modelle geklop het, die hoogste akkuraatheid
behaal van die model was 89%. Die resultate van die Superior Ability Test
het aangetoon dat die ANN model beter was as die Box-Jenkins ARMA-
tipe modelle in ten minste 5 gevalle.

Acknowledgements

Firstly i would like to thank my supervisor Prof Ronnie Becker for his excellent guidance, insightful comments, patience, and encouragement throughout the duration of the study.

I would also like to thank the African Institute for Mathematical Sciences (AIMS) for providing funding for this study. To my colleagues, family and friends, thank you for your support.

Dedications

To my parents, Daniel and Itani Ramuada

Contents

CONTENTS   **x**

List of Figures

Nomenclature

ADF Augumented Dickey Fuller

ANN Artificial Neural Network

ARMA Autoregressive Moving Average

ARIMA Autoregressive Integrated Moving Average

ARCH Autoregressive Conditional Heteroscedasticity

DW Durbin Watson

EWMA Exponential Weighted Moving Average

JSE Johannesburg Stock Exchange

LM Lagrange Multiplier

MA Moving Average

MCMC Markov Chain Monte Carlo

MCSE Markov Chain Standard Error

MH Metropolis Hastings

MLP Multilayer Perceptron

PP Phillips Perron

SPA Superior Predictive Ability

Chapter 1

Introduction

Forecasting of stock prices is one of the most important topics in academic and financial studies. However, the stock market prices can be influenced by many factors such as political situations, economic events, and natural disasters etc. As a result, forecasting stock prices is a challenging task due to the complexity of the stock market. Financial time series forecasting involves trying to understand the data generation process using historic (past) observations and applying a chosen model to try to extrapolate the returns series into the future (Tang et al., 2003).

Many studies has been carried out to find the best models for forecasting stock prices. Two of the most popular categories for modelling time series are linear models e.g ARMA Models, and Exponential smoothing. The second category is nonlinear models (artificial intelligence based models) e.g Support Vector Machines, Genetic Algorithms, and Artificial Neural Networks. The most popular of the linear models is the ARMA model first developed by Box-Jenkins in 1976. The ARMA model is widely used in financial time series forecasting because its quite flexible as it is made up of a combination of different types of time series models, namely, the Autoregressive (AR) and the Moving Average (MA) model, hence the name Autoregressive Moving Average (ARMA).

However, many financial time series studies carried out have shown that there is little evidence that the stock market is perfectly linear. To deal with this problem, mixture models have been proposed. The most popular mixture model is the combination of ARMA models together with non-linear

models such as Autoregressive Conditional Heteroscedasticty (ARCH) and the Generalised Autoregressive Conditional Heteroscedasticty (GARCH). The ARMA-GARCH types models have been successfully used for forecasting of financial time series (Majumder and Hussian, 2007). The ARCH model was first introduced by Robert F. Engle and Clive W. Granger in 1982 to model time varying volatility. Engle and Granger were awarded the Nobel prize for Economics in 2003 for this work. In 1986, Bollerslev introduced the Generalised ARCH model. The Generalised ARCH model is simply an extension of the ARCH model.

The rise in computer power, the availability of large amount of data, and advancements made in neural network theory has lead to a wide use of Artificial Neural Networks (ANN) in financial literature. The main advantage of using ANN in predicting stock returns is their ability to find non-linear relationships in the data (Majumder and Hussian, 2007, p.3). ANN are used to overcome the drawbacks of linear models. Another advantage of ANN is that they can use non-linear activation functions such as the sigmoid, and tangent sigmoid which are able to detect non-linear patterns in a data set. As a result, accuracy of predictions can be improved by using these non-linear models (Majumder and Hussian, 2007),(Rather, 2014). Because of the chaos and complex nature of the system, traditional statistical models sometimes tend to be inadequate to understand the relationship between input data and output of the particular system. Artificial Neural Networks models are data driven, they do not require any prior assumption about the knowledge of the relationship that may exist between input and output of the system. Some researchers claim that the stock market behaves like a chaotic system. Neural networks have been widely used to model these chaotic systems because they have few assumptions about dynamic dependencies.

1.1　Aim of the Study

The aim of this project is to predict the financial log-returns of Johannesburg Stock Exchange (JSE) listed companies. We will use Autoregressive Moving Average model, State Space ARMA Model, Autoregressive Moving Average model with parameters estimated through the Metropolis Hastings

algorithm, Autoregressive GARCH type models, and an Artificial Neural Network to make predictions. We will then evaluate each of the model in terms of how good it can predict whether the next day's returns price will go *up* or *down*. The results will then be summarised using a confusion matrix. The Superior Predictive Ability (SPA) test will be used to determine the "best" model for predicting the direction of change of stock returns prices.

1.2   Research Objectives

Time series models and computational intelligence techniques have been widely used for forecasting of stock returns. However, few studies have been carried out to try to forecast the returns of emerging market such as the South African Market the JSE. Hence, the first objective of this project is to use ARMA-Metropolis Hastings algorithm, ARMA Maximum Likelihood, State Space ARMA, AR-GARCH-type, and Artificial Neural Networks models to forecast the log returns of JSE listed companies. The second objective is to assess whether the proposed models are able to accurately predict whether the next day's returns price will go *down* or *up*. The third objective would be to identify the best model for forecasting log returns of the JSE listed companies.

1.3   Significance of the Study

During periods of economic instability or financial market turmoil, investors can suffer huge losses. As a result, prediction of the direction of movement of stock returns has gained more attention in recent times, this is because if we can successfully predict direction of movement of stock returns, then investors can be able to make profits from trading and investing in the stock market. If a system can be developed to predict movements and trends of stock market, then the owner of such a system would be wealthy. As a result, findings emerging from this study will help investors to minimise risk to their portfolios, regulators in policy making, and hedge fund managers.

1.4 Structure of the Thesis

The structure of the research project is as follows: Chapter 2 reviews previous literature on forecasting stock returns. Chapter 3 gives a introduction to statistical models to be used in the study. In Chapter 4, we discuss the methods used for analysis, in Chapter 5 results are discussed. Finally in Chapter 6, we draw conclusion and also propose future work which may be carried out.

Chapter 2

Literature Review

In this chapter, we review some literature on stock returns forecasting. We will also review some of the models used in our study to see how well they performed when applied to different stock exchanges.

2.1   Studies on Forecasting Stock Returns Using Time Series and Computational Intelligence Models

A study was carried out by (Hansson, 2017) in an effort to forecast the log returns series using data from three stock markets the S&P 500, Bovespa (Brazil), and OMX 30 (Sweden). Two models used in the study for forecasting the log returns were the ARMA(1,1)-GJRGARCH(1,1) and the Long Short Term Memory (LSTM). Results emerging from the study showed that both the ARMA(1,1)-GJRGARCH(1,1) and the LSTM achieves similar results when regression approach is taken. But, the LSTM outperformed the ARMA(1,1)-GJRGARCH(1,1) in terms of prediction of direction of change for only the small Swedish market.

In their study, (Ferenstein and Gasowski, 2004b) used the AR(1)-GARCH types models to predict the stock returns prices of two companies from the Warsaw stock exchange. Results emerging from the study showed that the

5

AR(1)-GARCH(1,1) model outperformed the AR(1)-EGARCH model. The results showed that the AR(1)-GARCH(1,1) had the lowest Mean Squared Error (MSE) out of all models considered in the study.

The ARIMA-EGARCH model was used by (Hossain et al., 2015) to model stock volume data of the Dhaka Stock Exchange (DSH). Results from the study showed that the ARIMA-EGARCH model was able to forecast the stock volume series, with the model producing low Mean Absolute Error (MAE) and Mean Squared Error (MSE).

A study was carried out by (Majumder and Hussian, 2007), In an effort to try to predict the direction of the movement of the stock returns of the S&P 500, CNX 500, and NIFTY index using ANN based models. In their study, various features of the data and structures of the ANN model were used to predict direction of stock returns. Results emerging from the study showed that in terms of direction of price movements, the highest accuracy achieved by ANN was 89%, and an average of 69.72% was achieved for all different structures of ANN models.

Chapter 3

ARMA-Type Time Series and Artificial Intelligence Techniques

In this chapter, statistical properties of log returns will be discussed. We will also discuss the models to be used in the study, this include the assumptions and parameter estimation methods for these models.

3.1   Financial Log Returns

Instead of using prices of assets, most financial studies uses returns (Tsay, 2005, p.2). (Campbell et al., 1997) specified two major reasons for using the returns

- Asset returns have attractive statistical properties. As a result, they are easier to handle than actual prices series.

- For a typical average investor, returns are a scale free and complete summary of an investment opportunity (Tsay, 2005, p.2).

The financial log returns are given as

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right),\tag{3.1}$$

where P_t is the closing price at time t, and r_t are the log returns.

3.1.1   Statistical Properties of Financial Asset Returns

Financial log returns often exhibit the following statistical properties (Fryzlewicz, 2005, p.1-2).

- The expected value of the returns sample is close to *zero*.

- The returns have a distribution with heavy tails (higher kurtosis).

- They exhibits volatility clustering. That is, period of large changes in the price of a financial asset tend to be followed by period of large changes, and also period of small changes tend to be followed by period of small changes (Rachev et al., 2008, p.185).

- The Autocorrelations for the sample are relatively small for almost all the lags.

Weakly Stationary

Suppose r_t is a time series, then this time series is said to be weakly stationary if its long term expected value (mean) and variance converge to a constant as time goes to infinity ($t \to \infty$). Weakly stationary simply means that data points of the series fluctuate around a fixed level (Tsay, 2005, p.29-30). Under weakly stationary, we assume that the first moment (mean) and the second moment of r_t are finite. In statistical finance literature, asset returns series are usually assumed to be weakly stationary (Tsay, 2005, p.30).

White Noise

A time series r_t is said to be Gaussian White noise if it has a mean of *zero* and a constant variance, thus $E(r_t) = 0$ and $var(r_t) = \sigma^2$. This can be expressed mathematically as $\{r_t\} \sim WN(\mu = 0, \sigma^2)$ (Tsay, 2005, p.36).

3.2   AutoRegressive (AR) Model

A simple AR(1) model is given as (Tsay, 2005, p.37-38)

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t, \tag{3.2}$$

where r_{t-1} is the returns price at previous step $t - 1$, ϵ_t is a white noise series, and ϕ_0, ϕ_1 are the model parameters which will be estimated using the maximum likelihood approach (mle). The white noise series ϵ_t is independent and identically distributed with a mean of *zero* and constant variance.

3.2.1 Statistical Properties of the AR Model

Below we will discuss the statistical properties of AR(1) model, AR(2) model and then proceed to give a generalization of the results of the general AR(p) model. If we assume that the time series r_t is weakly stationary, then from the properties discussed earlier it follows that E(r_t)=μ and var(r_t)=ρ (Tsay, 2005, p.38)

Proof:

$$
\begin{aligned}
E(r_t) &= E(\phi_0 + \phi_1 r_{t-1} + \epsilon_t) \\
&= \phi_0 + \phi_1 E(r_{t-1}).
\end{aligned}
\tag{3.3}
$$

But, under the stationary assumption we have that E(r_t)=E(r_{t-1})=μ (Tsay, 2005, p.38-39). Therefore, according to (Tsay, 2005, p.38)

$$
\begin{aligned}
E(r_t) &= \phi_0 + \phi_1 E(r_{t-1}) \\
\mu &= \phi_0 + \phi_1 E(r_{t-1}) \\
\mu &= \phi_0 + \phi_1 \mu \\
E(r_t) = \mu &= \frac{\phi_0}{(1 - \phi_1)}.
\end{aligned}
\tag{3.4}
$$

The above results implies that the expected value of r_t exists if $\phi_1 \neq 1$, and $E(r_t) = 0$ if $\phi_0 = 0$.

Since $\mu = \frac{\phi_0}{1-\phi_1} \implies \phi_0 = \mu(1 - \phi_1)$, then we can rewrite the AR(1) model as (Tsay, 2005, p.38)

$$
\begin{aligned}
r_t &= \phi_0 + \phi_1 r_{t-1} + \epsilon_t \\
r_t &= \mu(1 - \phi_1) + \phi_1 r_{t-1} + \epsilon_t \\
r_t - \mu &= \phi_1(r_{t-1} - \mu) + \epsilon_t.
\end{aligned}
\tag{3.5}
$$

To find the variance, we take the square of the results above and apply the expected value (Tsay, 2005, p.38-39).

$$
r_t^2 - \mu^2 = \phi_1^2(r_{t-1} - \mu)^2 + \epsilon_t^2
\tag{3.6}
$$

$$
E(r_t^2) - \mu^2 = \phi_1^2 E(r_{t-1} - \mu)^2 + E(\epsilon_t^2).
\tag{3.7}
$$

$$
\tag{3.8}
$$

But, we know that var(r_t)=E(r_t^2)$-\mu^2$, var(r_{t-1})=E(r_{t-1}^2)$-\mu^2$, and E(ϵ_t^2)=σ_ϵ^2. Therefore,

$$
var(r_t) \;\; = \phi_1^2 var(r_{t-1}) + \sigma_\epsilon^2.
\tag{3.9}
$$

$$
\tag{3.10}
$$

But, under stationary assumption it is known that var(r_t)=var(r_{t-1}) (Tsay, 2005, p.39). Therefore,

$$var(r_t) = \phi_1^2 var(r_{t-1}) + \sigma_\epsilon^2$$
$$var(r_t) - \phi_1^2 var(r_t) = \sigma_\epsilon^2 \tag{3.11}$$
$$var(r_t) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2}, \quad \phi_1^2 < 1.$$

The condition that $\phi_1^2 < 1$ is to ensure that the variance is non-negative. Consequently, by definition of weakly stationary of AR(1) model we have that $-1 < \phi_1 < 1$. Therefore, the necessary condition required for Autoregressive model of order 1 to be weakly stationary is $\mid \phi_1 \mid < 1$ (Tsay, 2005, p. 39). For more details about the proof see (Tsay, 2005, p.37-39).

AR(2) Model

Lemma: The AR(2) model is weakly stationary if the roots x_1, x_2 of $x^2 - \phi_1 x - \phi_2 = 0$ satisfy $\mid x_1 \mid < 1$ (i=1,2) and $\mid \phi_1 \mid + \mid \phi_2 \mid < 1$. The long term mean

$$\mu = \lim_{t \to \infty} E(r_t)$$

and the long term variance

$$\rho = \lim_{t \to \infty} var(r_t)$$

are given by

$$E(r_t) = \mu = \frac{\phi_0}{(1 - (\phi_1 + \phi_2))} \tag{3.12}$$

and

$$var(r_t) = \rho = \frac{\sigma_\epsilon^2}{(1 - (\phi_1^2 + \phi_2^2))}. \tag{3.13}$$

The results above implies that the variance of return series r_t exists if $\phi_1^2 + \phi_2^2 \neq 1$.

Proof: Weakly stationary can be proved similarly to the AR(1) case by using the variation of constants formula for r_t. We omit this portion of proof.

AR(p) Model

The AR(p) model is a generalization of AR(1) model. It is given as (Tsay,
2005, p.46)

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + \epsilon_t, \tag{3.14}$$

where ϵ_t is the white noise process, ϕ_0 is the model constant, and $\phi_0, \phi_1, \cdots, \phi_p$
are the model parameters which are estimated using the maximum likeli-
hood method. Applying the same techniques as that of AR(1), and AR(2)
cases we have that (Tsay, 2005, p.46)

$$E(r_t) = \mu = \frac{\phi_0}{1 - (\phi_1 + \phi_2 +, \cdots, +\phi_p)} \tag{3.15}$$

and

$$var(r_t) = \frac{\sigma_\epsilon^2}{1 - (\phi_1 + \phi_2 +, \cdots, +\phi_p)}. \tag{3.16}$$

For a detailed explanation of Autoregressive models see (Tsay, 2005, p.36-
47).

3.3   Moving Average (MA) Model

The MA model is a finite linear combination of the residuals ϵ_t. In many
statistical finance literature, the MA model is simply known as the extension
of the white noise series (Tsay, 2005, p.57). The MA(1) model is given as
(Tsay, 2005, p.57-58)

$$r_t = c_0 + \epsilon_t - \theta_1 \epsilon_{t-1}, \tag{3.17}$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, this implies that ϵ_t are independent and identically
distributed, each with a gaussian distribution with mean of *zero* and a con-
stant variance, and c_0 is the model constant.

3.3.1   Statistical Properties of the MA Model

The first and second moment of r_t are time invariant, therefore Moving Av-
erage models are weakly stationary. To prove this, we will use the property
$E(\epsilon_i) = 0$ for all i.

Proof:

Taking the expected value of Equation (3.17)

$$
\begin{aligned}
E(r_t) &= E(c_0 + \epsilon_t - \theta_1 \epsilon_{t-1}) \\
E(r_t) &= E(c_0) + E(\epsilon_t) - \theta_1 E(\epsilon_{t-1}) = c_0.
\end{aligned}
\tag{3.18}
$$

Hence, $E(r_t)=c_0$ is time invariant as also shown on (Tsay, 2005, p.58). We can also show that the variance is time invariant.

$$
\begin{aligned}
var(r_t) &= E(r_t^2) - (E(r_t))^2 \\
&= E[(c_0 + \epsilon_t - \theta_1 \epsilon_{t-1})(c_0 + \epsilon_t - \theta_1 \epsilon_{t-1})] - c_0^2.
\end{aligned}
\tag{3.19}
$$

But for all i $E(\epsilon_i) = 0$. Therefore,

$$
\begin{aligned}
var(r_t) &= c_0^2 + E(\epsilon_t^2) + \theta_1^2 E(\epsilon_{t-1}^2) - c_0^2 \\
&= E(\epsilon_t^2) + \theta_1^2 E(\epsilon_{t-1}^2) \\
&= \sigma_\epsilon^2 + \theta_1^2 \sigma_\epsilon^2 \\
&= \sigma_\epsilon^2 (1 + \theta_1^2).
\end{aligned}
\tag{3.20}
$$

Therefore, we have show that $var(r_t)$ is time invariant. Similarly, we can also show that the properties of MA(1) model holds for MA(2) case. Consider MA(2) model given as

$$
r_t = c_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}.
\tag{3.21}
$$

Then, taking the expected value of the MA(2) model above we obtain

$$
\begin{aligned}
E(r_t) &= E(c_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}) \\
E(r_t) &= E(c_0) + E(\epsilon_t) - \theta_1 E(\epsilon_{t-1}) - \theta_2 E(\epsilon_{t-2}) = c_0.
\end{aligned}
\tag{3.22}
$$

Hence, the results are similar to those of the MA(1) case.

MA(q) Model

The MA(q) model is an extension of the simple MA(1) model. The MA(q) model is given as (Tsay, 2005, p.57-58)

$$
r_t = c_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q},
\tag{3.23}
$$

where q is the number of lags of the model, and c_0, ϵ_t were discussed earlier. We can also show that the expected value of the MA(q) model is time invariant. Taking the expectation of Equation (3.23), we obtain

$$
\begin{aligned}
E(r_t) &= E(c_0 + \epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q}) \\
E(r_t) &= E(c_0) + E(\epsilon_t) - \theta_1 E(\epsilon_{t-1}) - \cdots - \theta_q E(\epsilon_{t-q}) = c_0.
\end{aligned}
\tag{3.24}
$$

Hence, the expected value of MA(q) model is time invariant. Furthermore, the variance of the MA(q) model is also time invariant. To show this, we generalise the results obtained from the cases of MA(1) and MA(2) models. Therefore, the variance of MA(q) model is

$$
var(r_t) = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma_\epsilon^2,
\tag{3.25}
$$

see (Tsay, 2005, p.59) for more details.

3.4   Auto Regressive Moving Average (ARMA) Model

To sufficiently describe the structure of a data set, both the AR and MA models usually need to be of higher order. As a result, the models will have many parameters that needs to be estimated. The ARMA model was first introduced by Box-Jenkins in 1994 to deal with this problem. The ARMA stands for Autoregressive Moving Average. The ARMA model is a combination of the AR and the MA models. Suppose r_t are returns series, then the ARMA(1,1) model is given as (Tsay, 2005, p.64)

$$
r_t = \phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t,
\tag{3.26}
$$

where ϵ_t is the white noise series with mean *zero* and variance σ_ϵ^2, ϕ_0 is the model constant, and ϕ_1, θ are the Autoregressive and Moving average parameters respectively. Note that

$$
\epsilon_t = \sigma_t z_t, \text{where} \quad z_t \sim N(0,1).
\tag{3.27}
$$

Financial returns usually exhibit fat tails and higher peak, as a result other statistical distributions such as the Student-t and the Generalised Error distributions are used as the distribution of z_t. It is important to note that the ARMA model assumes that the variance is constant. Later we will introduce

an AR model which assumes that the model variance (σ_t^2) is non-constant. That is, an AR model which assumes that the variance follows a GARCH processs (Shumway and Stoffer, 2006, p.280). More details about the ARMA model can be found on (Tsay, 2005, p.64-66).

3.4.1   Properties of the ARMA(1,1) Model

Taking the expected value of ARMA(1,1) model in Equation (3.26) we obtain

$$E(r_t) \ = E(\phi_0) + \phi_1 E(r_{t-1}) - \theta_1 E(\epsilon_{t-1}) + E(\epsilon_t). \tag{3.28}$$

But, E(ϵ_i)=0 for all i. Therefore,

$$\begin{aligned} E(r_t) &= \phi_0 + \phi_1 E(r_{t-1}) \\ \mu &= \phi_0 + \phi_1 \mu \\ E(r_t) = \mu &= \frac{\phi_0}{1 - \phi_1}. \end{aligned} \tag{3.29}$$

The above results implies that the expected value of r_t exists if $\phi_1 \neq 1$, and E(r_t) = 0 if $\phi_0 = 0$. These results are similar to those of an AR(1) model described earlier. For simplicity reasons, suppose that we assume that $\phi_0 = 0$ and we multiply Equation (3.26) by ϵ_t, and taking the expectation we obtain (Tsay, 2005, p.64-65)

$$\begin{aligned} r_t \epsilon_t - \epsilon_t \phi_1 r_{t-1} &= \epsilon_t \phi_0 + \epsilon_t^2 - \theta_1 \epsilon_t \epsilon_{t-1} \\ E(r_t \epsilon_t) - \phi_1 E(\epsilon_t r_{t-1}) &= \phi_0 E(\epsilon_t) + E(\epsilon_t^2) - \theta_1 E(\epsilon_t \epsilon_{t-1}) \\ E(r_t \epsilon_t) &= E(\epsilon_t^2) - \theta_1 E(\epsilon_t \epsilon_{t-1}) \\ E(r_t \epsilon_t) &= E(\epsilon_t^2) = \sigma_\epsilon^2. \end{aligned} \tag{3.30}$$

The ARMA(1,1) model in Equation (3.26) can be written as (Tsay, 2005, p.64)

$$r_t - \phi_1 r_{t-1} = \phi_0 + \epsilon_t - \theta_1 \epsilon_{t-1}. \tag{3.31}$$

Recall that if the returns series r_t is weakly stationary, it then follows that $var(r_t) = var(r_{t-1})$. Taking the variance of Equation (3.31) we obtain (Tsay,

2005, p.65)

$$
\begin{aligned}
var(r_t) &= \phi_1^2 E(r_{t-1}^2) + E(\epsilon_t^2) - 2\phi_1\theta_1 E(r_{t-1}\epsilon_{t-1}) + \theta_1^2 E(\epsilon_{t-1}^2) \\
var(r_t) &= \phi_1^2 var(r_{t-1}) + \sigma_\epsilon^2 - 2\phi_1\theta_1\sigma_\epsilon^2 + \theta_1^2\sigma_\epsilon^2 \\
var(r_t) - \phi_1^2 var(r_{t-1}) &= \sigma_\epsilon^2(1 - 2\phi_1\theta_1 + \theta_1^2) \\
var(r_t) - \phi_1^2 var(r_t) &= \sigma_\epsilon^2(1 - 2\phi_1\theta_1 + \theta_1^2) \\
var(r_t) &= \frac{\sigma_\epsilon^2(1 - 2\phi_1\theta_1 + \theta_1^2)}{1 - \phi_1^2}.
\end{aligned}
$$

(3.32)

Since we know that the variance is always positive, we should have that $\mid \phi_1 \mid < 1$. This stationary condition is similar to that of the AR(1) and AR(2) models discussed earlier.

ARMA(p,q) Model

The ARMA(p,q) model is a natural extension of the ARMA(1,1) model with longer lags and it is given as (Tsay, 2005, p.66)

$$
r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t,
$$

(3.33)

where ϵ_t are the residuals of the series, p and q are non-negative integers representing the number of lags of the model. It is important to note that signs of the MA terms may be reversed from those in some statistical text-books. For more details about the ARMA model see (Shumway and Stoffer, 2006, p.93-97) and (Tsay, 2005, p.64-66).

3.5 Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA stands for Autoregressive Integrated Moving Average. Suppose that r_t is a time series, such that $\triangle r_t = r_t - r_{t-1}$ follows an invertible and stationary ARMA(p,q) model, then r_t is said to be an ARIMA(p,1,q) process (Tsay, 2005, p.76). Differencing is the most commonly used technique for converting a non-stationary time series to a stationary one. In finance literature, the asset price series are usually non-stationary. As a result, the

log returns described in Equation (3.1) are usually used for modelling stock prices, largely because they are stationary. In time series literature, transforming a time series from a non-stationary to a stationary one by using its changes series is known as *differencing*.

3.6   Parameter Estimation Using the Maximum Likelihood Method

After defining our models, we need to estimate the model parameters. The parameters for these models will be estimated using the maximum likelihood estimation method. If $r_1, r_2, \cdots, r_T$, for $i = 1, 2, 3, \cdots, T$ are iid data whose marginal probability density function is $f(r_t, \boldsymbol{\theta})$, then, the joint density function for that particular sample can be defined as the product of the marginal densities, that is,

$$f(\mathbf{r}; \boldsymbol{\theta}) = f(r_1, \cdots, r_T; \boldsymbol{\theta}) = \prod_{t=1}^{T} f(r_t; \boldsymbol{\theta}). \qquad (3.34)$$

Then, the likelihood function can be described as

$$L(\boldsymbol{\theta} \mid \boldsymbol{r}) = L(\boldsymbol{\theta} \mid r_1, \cdots, y_T) = \prod_{t=1}^{T} f(r_t; \boldsymbol{\theta}). \qquad (3.35)$$

Then, the log-likelihood function is given as

$$L(\boldsymbol{\theta} \mid \mathbf{r}) = \sum_{t=1}^{T} log f(r_t; \boldsymbol{\theta}). \qquad (3.36)$$

3.6.1   Estimating Parameters of the AR(1) Model

If we have samples that are independent and identically distributed, then the likelihood is defined as the product of marginal density of these individual samples. But, in time series analysis we cannot simply use the product of marginal density to evaluate the likelihood. In the study of time series, we use conditional density to evaluate the likelihood function. Consider a simple AR(1) model given as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t, \text{where} \quad \epsilon_t \sim N(\mu = 0, \sigma^2) \qquad (3.37)$$

for $t = 1, \cdots, T$. Then, from the model above we let $\boldsymbol{\theta} = (\phi_0, \phi_1, \sigma^2)$ to
be a vector which contains the parameters of AR(1) model that we want to
estimate. So, the goal of MLE is to find a set of parameters that are likely to
have generated the sample data $r_1, r_2, \cdots, r_T$. Earlier we showed that for a
stationary AR(1) process we have that

$$E(r_t) = \frac{\phi_0}{1 - \phi_1} \tag{3.38}$$

and

$$var(r_t) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2}. \tag{3.39}$$

Furthermore, since we assumed that $\{\epsilon_t\}_{t=-\infty}^{\infty}$ is a Gaussian process, it then
follows that r_1 is also Gaussian. Then, for the initial value r_1, the marginal
density is given as

$$r_1 \sim N\left(\frac{\phi_0}{1 - \phi}, \frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right) \tag{3.40}$$

$\Rightarrow$

$$
\begin{aligned}
f_{r_1}(r_1; \boldsymbol{\theta}) &= f_{r_1}(r_1; \phi_0, \phi_1, \sigma^2) \\
&= (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{ -\frac{1}{2\sigma^2}(r_1 - \mu)^2 \right\} \\
&= \left(2\pi\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right)^{-\frac{1}{2}} exp\left\{ -\frac{(1 - \phi_1^2)}{2\sigma_\epsilon^2}\left(r_1 - \frac{\phi_0}{1 - \phi_1}\right)^2 \right\}.
\end{aligned} \tag{3.41}
$$

The above is the marginal density for the initial value r_1. The log-likelihood
function of Equation (3.41) is

$$
\begin{aligned}
log f_{r_1}(r_1; \boldsymbol{\theta}) &= log\left(\left(2\pi\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right)^{-\frac{1}{2}} exp\left\{ -\frac{(1 - \phi_1^2)}{2\sigma_\epsilon^2}\left(r_1 - \frac{\phi_0}{1 - \phi_1}\right)^2 \right\}\right) \\
&= -\frac{1}{2}log\left(2\pi\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right) - \left(\frac{1 - \phi_1^2}{2\sigma_\epsilon^2}\right)\left(r_1 - \frac{\phi_0}{1 - \phi_1}\right)^2\right) \\
&= -\frac{1}{2}log(2\pi) - \frac{1}{2}log\left(\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right) - \frac{\left(r_1 - \frac{\phi_0}{1-\phi_1}\right)^2}{(2\sigma_\epsilon^2)/(1 - \phi_1^2)}.
\end{aligned} \tag{3.42}
$$

The equation above represents the log-likelihood function of the marginal
density for the initial value r_1. Generally, the value of $r_1, r_2, \cdots, r_{t-1}$ is re-
lated to r_t only through r_{t-1}. Furthermore, the density of observation at
time t conditional on the previous observations $t-1$ is given as

$$f_{r_t|r_{t-1},r_{t-2},\cdots,r_1}(r_t \mid r_{t-1},\cdots,r_1;\boldsymbol{\theta}) = f_{r_t|r_{t-1}}(r_t \mid r_{t-1};\boldsymbol{\theta})$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)$$

$$\text{(3.43)}$$

$$f_{r_t|r_{t-1}}(r_t \mid r_{t-1};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)$$

$$log(f_{r_t|r_{t-1}}(r_t \mid r_{t-1};\boldsymbol{\theta})) = \sum_{t=2}^{T} log\left((2\pi\sigma^2)^{-\frac{1}{2}}exp\left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)\right)$$

$$= \sum_{t=2}^{T}\left(-\frac{1}{2}log(2\pi) - \frac{1}{2}log(\sigma^2)\right.$$

$$+ \left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)\right)$$

$$= \left(-\frac{(T-1)}{2}log(2\pi) - \frac{(T-1)}{2}log(\sigma^2)\right.$$

$$+ \sum_{t=2}^{T}\left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)\right).$$

Then, the log likelihood function is given as

$$logL(\boldsymbol{\theta} \mid \mathbf{r}) = \sum_{t=2}^{T} logf_{r_t|r_{t-1}}(r_t \mid r_{t-1};\boldsymbol{\theta}) + logf_{r_1}(r_1;\boldsymbol{\theta}). \qquad \text{(3.44)}$$

The full log-likelihood function defined above is known as the exact log-likelihood. The first part on the RHS is known as the conditional log-likelihood and the second part is known as the marginal log-likelihood for the initial values. Using Equation (3.42) and (3.44), we can write the log likelihood function for a Gaussian AR(1) process with T sample size as

$$\log L(\boldsymbol{\theta} \mid r) = \left(-\frac{(T-1)}{2}log(2\pi) - \frac{(T-1)}{2}log(\sigma^2)\right.$$

$$+ \sum_{t=2}^{T}\left(-\frac{1}{2}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{\sigma^2}\right)\right)$$

$$- \frac{1}{2}log(2\pi) - \frac{1}{2}log\left(\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right) - \frac{\left(r_1 - \frac{\phi_0}{1-\phi_1}\right)^2}{(2\sigma_\epsilon^2)/(1 - \phi_1^2)}.$$

The equation above represents a Gaussian log-likelihood for an AR(1) process. The log-likelihood equation above can also be used for AR models of different lags.

3.6.2   Estimating Parameters of the MA Model

To estimate parameters of the Moving Average model, we will use the same approach used for estimation of Autoregressive model parameters. Consider a simple Gaussian Moving Average process of order 1 given as

$$r_t = c_0 + \epsilon_t - \theta_1 \epsilon_{t-1}, \text{where} \quad \epsilon_t \sim N(0, \sigma^2). \tag{3.45}$$

Let $\boldsymbol{\theta} = (c_0, \theta_1, \sigma^2)$ be a vector which contains parameters to be estimated. Using the results from Equation (3.19) and (3.20) we can construct the joint marginal densities for each observation. For $\left\{ r_t \right\}_{t=1}^{T}$, then

$$r_t \sim N\left(c_0, \sigma_\epsilon^2 (1 + \theta_1^2) \right).$$

The sequence $\{\epsilon_1, \epsilon_2, \cdots, \epsilon_T\}$ can be calculated using $r_1, r_2, \cdots, r_T$ through iterating on

$$\epsilon_t = r_t - c_0 - \theta_1 \epsilon_{t-1} \tag{3.46}$$

for $t = 1, 2, 3, \cdots, T$.

Then, the likelihood function for Moving Average process of order 1 is given as

$$L(\boldsymbol{\theta} \mid \mathbf{r}) = \prod_{t=1}^{T} f(r_t; \boldsymbol{\theta})$$
$$= -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma^2}. \tag{3.47}$$

3.6.3   Estimating Parameters of the ARMA Model

Consider an ARMA(1,1) process defined earlier in Equation (3.26) as

$$r_t = \phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t, \tag{3.48}$$

where $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. Let $\boldsymbol{\theta} = (\phi_0, \phi_1, \theta_1, \sigma^2)$ be a vector which contains parameters to be estimated. The sequence $\{\epsilon_1, \epsilon_2, \cdots, \epsilon_T\}$ can be calculated

using $r_1, r_2, \cdots, r_T$, through iterating on

$$\epsilon_t = r_t - (\phi_0 + \phi_1 r_{t-1}) - (\theta_1 \epsilon_{t-1}), \tag{3.49}$$

for $t = 1, 2, 3, \cdots, T$. Then, the log likelihood function is given as

$$
\begin{aligned}
logL(\boldsymbol{\theta}) &= \sum_{t=1}^{T} log\left((2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{ -\frac{1}{2\sigma^2}(r_t - \mu)^2 \right\} \right) \\
&= \sum_{t=1}^{T} log\left((2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{ -\frac{1}{2\sigma^2}(\epsilon_t)^2 \right\} \right) \\
&= -\frac{T}{2}log(2\pi) - \frac{T}{2}log(\sigma^2) - \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma^2}.
\end{aligned}
\tag{3.50}
$$

3.7   Estimation of Parameters Using Bayesian Methods

The Bayes theorem was first introduced by Thomas Bayes (1701-1761). In Bayesian methods, probabilities represents the degree of belief that we have about an unknown event (Rachev et al., 2008).

Bayes theorem: Suppose the probability distribution of θ is given as $p(\boldsymbol{\theta})$, then we can rewrite the posterior distribution of θ in terms of the Bayes theorem given as

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{\int_\theta p(y \mid \theta)p(\theta)d\theta}, \tag{3.51}$$

where $p(y \mid \theta)$ is the likelihood function, θ is the vector which contains model parameters, y is the data, and $p(\theta)$ is the prior distribution of θ. The prior distribution represent prior belief that we have before we see the data y. The posterior distribution given by Equation (3.51) represents the distribution of θ after we have seen the data y. In Equation (3.51), the denominator is computed by averaging all possible values of θ see (Rachev et al., 2008, p.19). As a result, the denominator expression is not dependent on θ, so we can rewrite Equation (3.51) as

$$p(\theta \mid y) \propto p(y \mid \theta)p(\theta). \tag{3.52}$$

Equation (3.52) is called the continuous version of the Bayes's theorem. In Bayesian methodology, all inference are made based on $p(\theta \mid y)$ (the posterior distribution).

Markov Chain Monte Carlo (MCMC)

The Markov Chain Monte Carlo is a simulation based algorithm which is used in Bayesian framework to sample from the posterior distribution $p(\theta \mid y)$ (Rachev et al., 2008, p.66). The MCMC algorithm generates samples, such that each sample is dependent only on the previous one, hence the name Markov Chains. Formally speaking, the Markov chain is a sequence of random variables where a given state of the process is dependent only on its previous state, this implies that the state does not depend on other states except its previous state (Rachev et al., 2008, p.67). The MCMC allows us to draw i.i.d samples from the posterior distribution which is of complicated form (Rachev et al., 2008, p.66). The Metropolis Hastings algorithm and the Gibbs sampler are the most commonly used MCMC methods for sampling from the posterior distribution.

Monte Carlo

Suppose y represents the observed data set, and θ is a vector which contains unknown parameters, then the posterior expectation (mean) of the function $g(\theta)$ is given by (Rachev et al., 2008, p.61-62).

$$Eg(\theta \mid y) = \int g(\theta)p(\theta \mid y)d\theta, \tag{3.53}$$

where $p(\theta \mid y)$ is the posterior distribution defined earlier in Equation (3.51). Suppose that $\theta^{(1)}, \cdots, \theta^{(N)}$ are N samples which were drawn from the posterior distribution $p(\theta \mid y)$. Then, by Law of large numbers the Monte Carlo approximation

$$\widehat{g_N(\theta)} = \frac{1}{N} \sum_{n=1}^{N} g(\theta^{(n)}) \tag{3.54}$$

converges to E $g(\theta \mid y)$ as N goes to infinity (Rachev et al., 2008, p.62). This implies that, if we can obtain a large samples from the posterior distribution, then we can be able to accurately approximate the mean of $g(\theta)$ (Rachev et al., 2008, p.62). This approximation is known as the Monte Carlo integration. See (Rachev et al., 2008, p.61-63) for more details about the

Monte Carlo integration.

The Monte Carlo Standard Error (MCSE)

The Monte Carlo Standard Error indicates the amount of error in the estimates as a result of using MCMC. The MSCE is given as

$$MCSE = \frac{SD}{\sqrt{n}}, \tag{3.55}$$

where SD is the standard deviation given as

$$SD = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (g(\theta^{(n)}) - \widehat{g_N(\theta)})^2}. \tag{3.56}$$

As the number of iterations increases the value of the MCSE approaches 0. From Equation (3.55) we can see that if the sample size n is large, then the MCSE will be small.

3.7.1 MCMC: Metropolis Hastings (M-H) Algorithm

The Metropolis Hastings is an MCMC algorithm which was first introduced by Nicholas Metropolis in 1953 and later modified by Wilfred Hastings in 1970, hence the name Metropolis Hastings (MH) algorithm. The algorithm is used to obtain or generate samples from the posterior distribution. The Markov Chain Monte Carlo algorithm works by returning a sequence of N samples $\theta^{(1)}, \cdots, \theta^{(N)}$ such that each sample is generated from the posterior distribution Rachev et al. (2008). That is,

$$\theta^{(i)} \sim p(\theta \mid y), \text{for} \quad i = 1, \cdots, N. \tag{3.57}$$

The MCMC algorithm produces a bag of samples $\left\{\theta^{(i)}\right\}_{i=1}^{N}$, such that this sample is a proxy for the posterior distribution $p(\theta \mid y)$.

Metropolis Hastings (MH) Algorithm (Rachev et al., 2008, p.67-69)

Let $\theta = (\theta_1, \theta_2, \cdots, \theta_n)$ be a vector which contains model parameters, and $q(\theta \mid \theta^{(t-1)})$ be the proposal (candidate) distribution. The M-H algorithm is made up of two major stages. In the first stage, a sample is drawn from

the proposal distribution. In the second stage, the drawn sample is either
accepted or rejected. The accepted sample are then used to form the Markov
chains, while the rejected ones are discarded.

1. Specify the initial value of the chain, that is, set a value for $\theta^{(0)}$.

2. At each time step t, draw θ^* from the proposal distribution $q(\theta \mid \theta^{(t-1)})$,
   where $\theta^{(t-1)}$ is the parameter value at the previous step $t-1$.

3. Compute the acceptance probability given as (Rachev et al., 2008, p.67-
   68) $\alpha(\theta^*, \theta^{(t-1)}) = \text{minimum}\left(1, \frac{p(\theta^*)q(\theta^{t-1}|\theta^*)}{p(\theta^{(t-1)})q(\theta^*|\theta^{(t-1)})}\right)$,

   where $p(\theta^{(t-1)})$ represents the probability that the next state of the
   chain is $\theta^{(t-1)}$ (the previous state), and $p(\theta^*)$ represents the probability
   that the next state of the chain is θ^* (the value drawn from the proposal
   distribution). To find the value of the acceptance probability we take
   the minimum of 1 and $\frac{p(\theta^*)q(\theta^{t-1}|\theta^*)}{p(\theta^{(t-1)})q(\theta^*|\theta^{(t-1)})}$

4. Draw a random variable u from the Uniform distribution whose sup-
   port is $(0,1)$, that is, $u \sim U(0,1)$.

   - if $u \leq \alpha(\theta^*, \theta^{(t-1)})$, then we set $\theta^{(t)}$ (current state of the chain) to
     be equal to θ^* (the value drawn from the proposal distribution),
     that is, $\theta^{(t)} = \theta^*$.

   - else, set $\theta^{(t)}$ (current value of chain) to be equal to $\theta^{(t-1)}$ (the value
     at the previous step), that is, $\theta^{(t)} = \theta^{(t-1)}$.

5. Repeat the algorithm from step 2 to step 4 n times. This will produce
   a sequence of Markov chains $\{\theta^{(i)}\}_{i=1}^{N}$. In order to ensure that the
   samples drawn are indeed from the posterior distribution, we run the
   simulations for a long period of time.

For more information on the Metropolis Hastings algorithm see (Rachev
et al., 2008, p.63-69).

Diagnostics: Convergence of Markov Chains and the Burn-in Period

The samples obtained using MCMC are used to make inferences about the
posterior distribution. But, we have to first ensure that these chains has con-
verged to the stationary distribution. As stated earlier, we run the MCMC

simulations for a large number of time in order to ensure that these chains
converged. The starting point of the Markov Chain is influential. In order
to minimise the effects of the initial state of the chain, we discard a fraction
of these simulations. This fraction is known as the *burn-in fraction*. Only
subsequent draws would be used for posterior inference after we remove
the burn-in period (Rachev et al., 2008, p.75).

Estimating Parameters of a Model Using MCMC

- Start by specify the model and our prior knowledge about each parameter.

- Specify the initial value for each model parameter.

- Run the MCMC algorithm and obtain the chains.

- Discard the initial Burn-in period and use only subsequent draws to make inference about the posterior distribution.

3.8   Volatility Models

Volatility can simply be described as the variability of a financial time series
(Rachev et al., 2008, p.185). In many statistical finance literature, it is defined as the magnitude and speed of fluctuations of a financial time series
(Rachev et al., 2008, p.185). Volatility is expressed mathematically as the
standard deviation of a financial returns asset.

Characteristics of Volatility

- The volatility of financial asset returns is non-constant through time (Rachev et al., 2008, p.185). As a result, we say that asset returns are *heteroskedastic*.

- Volatility reacts differently to both negative and positive returns shocks. That is, volatility tends to be high when the markets falls than when they rise. (Rachev et al., 2008, p.186).

Volatility models were first introduced to try to explain these stylised facts about financial asset returns (Rachev et al., 2008, p.186).

3.8.1   ARCH Model

The Autoregressive Conditional Heteroscedasticity (ARCH) Model was first introduced by Engle in 1982 to capture time varying volatility in returns series. The ARCH(1) model is the simplest of the ARCH model. If we let p_t to be stock price at time t, then as shown earlier, the log returns denoted by r_t are given as

$$r_t = log\left(\frac{p_t}{p_{t-1}}\right). \tag{3.58}$$

The expected value of the log returns is very close to *zero* and the variance is generally dependent on time, and they also exhibits serial correlation. In this study, we model log returns by

$$r_t = \mu_t + \sigma_t \epsilon_t, \tag{3.59}$$

where σ_t, μ_t will be discussed later, ϵ_t is the white noise process which is independent and identically distributed with expected value of 0 and variance of 1. That is, $\epsilon_t \sim N(0,1)$. The assumption that ϵ_t is independent implies that the mean of ϵ_t is independent of the past, i.e

$$\epsilon_t \mid F_{t-1} \sim \epsilon_t, \tag{3.60}$$

where F_{t-1} represents a set of information available at previous step $t-1$. We will write

$$y_t = r_t - \mu_t. \tag{3.61}$$

Furthermore,

$$y_t = \sigma_t \epsilon_t. \tag{3.62}$$

The ARCH model is the simplest of all volatility models. It can be expressed as a deterministic function of past squared returns. In this case, volatility at time t is determined at time $t-1$ (Rachev et al., 2008, p.185). We assume that the equation for ARCH(1) model is given as (Shumway and Stoffer, 2006, p.281)

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2. \tag{3.63}$$

From Equation (3.61) and (3.62), it follows that when given y_{t-1}, the variance σ_t is a deterministic constant. The conditional mean of y_t given y_{t-1} is given as

$$E(y_t \mid y_{t-1}) = E(\sigma_t \epsilon_t \mid y_{t-1}) = \sigma_t E(\epsilon_t \mid y_{t-1}) = \sigma_t E(\epsilon_t) = 0. \qquad (3.64)$$

The conditional variance of y_t given y_{t-1} is given as

$$var(y_t \mid y_{t-1}) = E(y_t^2 \mid y_{t-1}) = E(\sigma_t^2 \epsilon_t^2 \mid y_{t-1}) = \sigma_t^2 E(\epsilon_t^2 \mid y_{t-1}) = \sigma_t^2 E(\epsilon_t^2) = \sigma_t^2 \qquad (3.65)$$

Using Equation (3.64) and (3.65), we have that

$$y_t \mid y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2). \qquad (3.66)$$

Using Equation (3.64), we have that

$$E(\mu_t \mid y_{t-1}) = E(r_t \mid y_{t-1}) \qquad (3.67)$$

and

$$E(\epsilon_t^2 \mid y_{t-1}) = E(\epsilon_t^2) = 1. \qquad (3.68)$$

Now

$$\sigma_t^2 = \sigma_t^2 E(\epsilon_t^2 \mid y_{t-1}) = E(\sigma_t^2 \epsilon_t^2 \mid y_{t-1}) = E(y_t^2 \mid y_{t-1}). \qquad (3.69)$$

Using Equation (3.61) and (3.67), we now have that

$$\sigma_t^2 = E((r_t - \mu_t)^2 \mid y_{t-1})) \qquad (3.70)$$
$$= var(r_t \mid y_{t-1}). \qquad (3.71)$$

From the results above μ_t, σ_t^2 are the conditional expectation and variance of r_t. Applying the tower law for conditional mean and using Equation (3.66) we obtain

$$E(y_t) = E(E(y_t \mid y_{t-1})) = E(0) = 0. \qquad (3.72)$$

From the above it is clear that the expected value of y_t is equal to *zero*. By subtracting Equation (3.62) and (3.81), we obtain (Shumway and Stoffer, 2006, p.281)

$$y_t^2 - (\alpha_0 + \alpha_1 y_{t-1}^2) = \sigma_t^2 \epsilon_t - \sigma_t^2 \qquad (3.73)$$

Now, the equation can be rewritten as

$$
\begin{aligned}
y_t &= (\alpha_0 + \alpha_1 y_{t-1}^2) + \sigma_t^2 \epsilon_t - \sigma_t^2 \\
y_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \sigma_t^2(\epsilon_t - 1) \\
y_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + v_t,
\end{aligned}
\tag{3.74}
$$

where $v_t = \sigma_t^2(\epsilon_t - 1)$. But, since ϵ_t^2 is a square of the Normal distribution with mean *zero* and variance *one* (i.e N(0,1)), then ϵ_{t-1}^2 is a χ_1^2 random variable whose mean is *zero* (Shumway and Stoffer, 2006, p.281-282). For an integer $p > 0$ and by using Equation (3.72) we have that

$$
\begin{aligned}
cov(y_{t+p}, y_t) &= E(y_t y_{t+p}) = EE(y_t y_{t+p} \mid y_{t+p-1}) \tag{3.75} \\
&= E(y_t E(y_{t+p} \mid y_{t+p-1})) = 0. \tag{3.76}
\end{aligned}
$$

Hence, we have shown that y_t are uncorrelated. For more details see (Shumway and Stoffer, 2006, p.281-282).

Assumption: Suppose that the variance of v_t is a a constant and $0 \leq \alpha_1 < 1$. From the assumption above, it follows that y_t is weakly stationary, and as a result the unconditional variance of y_t is constant.

Proof:

$$
var(y_t) = E(y_t^2) = E(E(y_t^2 \mid y_{t-1})) = E(\alpha_0 + \alpha_1 y_{t-1}^2) = \alpha_0 + \alpha_1 E(y_{t-1}^2)
$$
$$
= \alpha_0 + \alpha_1 E(y_t^2).
\tag{3.77}
$$

But, $E(y_t^2) = var(y_t)$, therefore

$$
\begin{aligned}
var(y_t) &= \alpha_0 + \alpha_1 var(y_t) \\
var(y_t) &= \frac{\alpha_0}{1 - \alpha_1}.
\end{aligned}
\tag{3.78}
$$

According to (Shumway and Stoffer, 2006, p.282), the results above implies that $E(y_t^2)$ and $var(y_t^2)$ are constant with respect to time.

$$
E(y_t^4) = \frac{3\alpha_0^2}{(1 - \alpha_1)^2} \frac{(1 - \alpha_1^2)}{1 - 3\alpha_1^2}, \quad 3\alpha_1^2 < 1.
\tag{3.79}
$$

Then, the kurtosis is (Shumway and Stoffer, 2006, p.282), (Tsay, 2005, p.118)

$$
\begin{aligned}
k &= \frac{E(y_t^4)}{[E(y_t^2)]^2} \\
&= \frac{\frac{3\alpha_0^2}{(1-\alpha_1)^2}\frac{(1-\alpha_1^2)}{1-3\alpha_1^2}}{\left(\frac{\alpha_0}{1-\alpha_1}\right)^2} \\
&= 3\frac{1-\alpha_1^2}{1-3\alpha_1^2}.
\end{aligned}
\tag{3.80}
$$

Notice that the kurtosis will always be greater than 3 (unless if $\alpha_1 = 0$). But, we know that the Normal distribution has a kurtosis of 3. The results above implies that the distribution of y_t has fat "tails", or it is leptokurtic. For more details about the proof see (Shumway and Stoffer, 2006, p.282) and (Tsay, 2005, p.118).

ARCH(p) Model

The ARCH(p) is an extension of the basic ARCH(1) model, it is given as

$$
\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2,
\tag{3.81}
$$

where $\alpha_j \geq 0$, and p represents the number of lags of the model. The parameters α_j, ω are the error and constant parameters respectively. To ensure that the conditional variance σ_t^2 is positive we restrict α_1 to take positive values only, thus $\alpha_1 > 0$ (Shumway and Stoffer, 2006). From Equation (3.81), we can see that the conditional variance is dependent upon its own previous residuals. The residuals are innovations of a time series.

Weakness of the ARCH Model

The ARCH model assumes that negative and positive shocks have the same impact on volatility. In reality, stock prices reacts differently to positive and negative shocks. For more details about the characteristics, properties, and some applications of ARCH model see (Shumway and Stoffer, 2006, p.280-289). Also, (Tsay, 2005, p.109-131) gives a detailed presentation of the ARCH model and some excellent examples of the model being applied to solve financial problems.

3.8.2 Generalised Autoregressive Conditional Heteroscedasticity (GARCH) Model

The restrictions placed on the parameters of ARCH(p) model defined earlier can be violated. As a result, Bollerslev in 1986 introduced an extension of the ARCH model called the *GARCH model* to deal with these restrictions. GARCH stands for Generalised ARCH. To show some properties of the GARCH model, we will use properties of the ARCH model which were discussed earlier. The symmetric GARCH model assumes that negative and positive shocks have the same magnitude or effect on volatility (Alexander, 2009, p.132). Consider a mean equation given as

$$y_t = c + u_t, \tag{3.82}$$

where c is the constant parameter and $E(u_t) = 0$. The Equation (3.82) represents the mean of y_t. We write

$$var(u_t) = \sigma_t^2 = E(u_t^2), \tag{3.83}$$

where $E(u_t^2)$ represents the conditional variance of y_t given some available information F_{t-1}. Note that σ_t represents the volatility of process y_t. The volatility process σ_t is non-constant, as a result y_t is said to be heteroskedastic (a process where the variance is not constant).

The GARCH(1,1) model is given as (Shumway and Stoffer, 2006, p.286)

$$\sigma_t^2 = \gamma \bar{\sigma}^2 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3.84}$$

where $0 < \alpha + \beta < 1$ and $\omega = \gamma \bar{\sigma}^2$. We will discuss the notation $\bar{\sigma}^2$ later.

Suppose we write $m_t = E(\sigma_t^2) = E(u_t^2)$. Therefore, if we take the expectation of Equation (3.84) we have that

$$
\begin{aligned}
E(\sigma_t^2) &= E(\gamma \bar{\sigma}^2 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2) \\
m_t &= \gamma \bar{\sigma}^2 (\alpha + \beta) m_{t-1}.
\end{aligned}
\tag{3.85}
$$

Suppose that we let $\gamma = 0$, then a homogeneous equation is obtained. That is,

$$m_t = (\alpha + \beta) m_{t-1}. \tag{3.86}$$

It can be show that

$$m_t = m_0(\alpha + \beta)^t. \tag{3.87}$$

Given the value of m_0, it then follows that the solution of the difference equation is unique. From the chain of equalities below, it can be easily seen that

$$m_t = (\alpha + \beta)m_{t-1} = (\alpha + \beta)^2 m_{t-2} = \cdots = (\alpha + \beta)^2 m_0. \tag{3.88}$$

If we also include $\gamma\bar{\sigma}^2$, then, it follows that we have a general solution for homogeneous equation. We can be able to shown that $m_t = \bar{\sigma}^2$ satisfies the GARCH model equation described earlier. Then, it follows that the general solution is given as

$$m_t = m_0(\alpha + \beta)^n + \bar{\sigma}^2.$$

But, we have restrictions placed on parameters α and β such that $0 < \alpha + \beta < 1$. Therefore,

$$\lim_{t \to \infty} m_t = \bar{\sigma}^2. \tag{3.89}$$

The above results implies that the process m_t would converge to a long term equilibrium $\bar{\sigma}^2$ (the equilibrium is the long term variance of the GARCH model). The process described in Equation (3.89) is known as the *mean-reverting process*. Mean reversion implies that the volatility of asset returns would eventually settle down to a mean state, where the mean is that of the GARCH model.

Persistence in Volatility

The GARCH model parameters α and β are together known as the *persistence parameters*. According to (Rachev et al., 2008, p.192) these parameters determines the speed of mean reversion of the volatility to its long term average value. From Equation (3.84), if $\alpha + \beta = 1$, then $m_t = \bar{\sigma}^2$ would no longer be a solution to Equation (3.84). As a result, we say that the GARCH model has a unit root (non-stationary).

Long Term Volatility in the Market

In the absent of market shocks, the variance of the GARCH model would eventually settle down to a constant or state value (Alexander, 2008). That

is, $\sigma_t^2 = \bar{\sigma}^2$, where $\bar{\sigma}^2$ is called the *unconditional variance* of GARCH model defined in Equation (3.84). In many financial econometrics literature, the unconditional variance $\bar{\sigma}^2$ is called *long term variance*.

Proof:

Properties of the ARCH model which where discussed earlier will be used to derive the long term volatility equation. The unconditional variance is constant through time (it does not change), as a result we have that

$$var(u_t) = var(u_{t-1}) = var(u_{t-2}) \cdots = var(u_{t-n}). \tag{3.90}$$

Using results from Equation (3.83), it then follows that

$$E(u_t^2) = E(u_{t-1}^2) = \cdots = E(u_{t-n}^2). \tag{3.91}$$

Now

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2, \text{apply the expectation} \\
E(\sigma_t^2) &= E(\omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2) \\
E(\sigma_t^2) &= E(\omega) + \alpha E(u_{t-1}^2) + \beta E(\sigma_{t-1}^2) \\
\lim_{t \to \infty} E(\sigma_t^2) &= \omega + \alpha \lim_{t \to \infty} E(u_{t-1}^2) + \beta \lim_{t \to \infty} E(\sigma_{t-1}^2).
\end{aligned}
\tag{3.92}
$$

By using Equation (3.83) and Equation (3.89), we have that (Rachev et al., 2008, p.187)

$$
\begin{aligned}
\bar{\sigma}^2 &= \omega + \alpha \bar{\sigma}^2 + \beta \bar{\sigma}^2 \\
\omega &= \bar{\sigma}^2 - \alpha \bar{\sigma}^2 - \beta \bar{\sigma}^2 \\
\omega &= \bar{\sigma}^2 (1 - \alpha - \beta).
\end{aligned}
\tag{3.93}
$$

Finally, the long term variance is given by

$$\bar{\sigma}^2 = \frac{\omega}{1 - (\alpha + \beta)}. \tag{3.94}$$

To find the long term unconditional volatility, we take the square root of the unconditional variance ($\bar{\sigma}^2$). The condition that $0 < \alpha + \beta < 1$ ensures that the variance is positive. The volatility would be relatively high in the market if the value of the unconditional volatility is large (Alexander, 2008).

The GARCH(p,q) is an extension of the simple GARCH(1,1) model with lags *p* and *q*. It is given as (Shumway and Stoffer, 2006, p.286)

$$\sigma_t^2 = \omega + \alpha \sum_{i=1}^{p} u_{t-j}^2 + \beta \sum_{j=1}^{q} \sigma_{t-j}^2.$$

From the equation above, we can see that large shocks would result in large variance. The GARCH model assume symmetric, that is, negative and positive shocks have the same effect.

Properties of the GARCH(1,1) Process

The GARCH(1,1) process has the following four important properties (Rachev et al., 2008, p.190-192).

- Stationarity is the most important feature of the GARCH(1,1) process. A stochastic process is said to be stationary if it has finite moments, that is, its variance, mean, and covariance are constant through time (Rachev et al., 2008, p.190). If we assume that the returns **r_t** are normally distributed, then, the GARCH(1,1) process is stationary if its persistence parameters $\alpha + \beta$ are less than *one* (Rachev et al., 2008, p.190).

- If the returns are assumed to be Student-t distributed, then the stationary inequality of the GARCH(1,1) model is given as (Rachev et al., 2008, p.191)

$$\alpha \frac{\nu}{\nu - 2} + \beta < 1, \tag{3.95}$$

  where ν is the degree of freedom parameter. On the other hand, if we assume that the returns follows a Normal distribution, then the necessary condition for GARCH(1,1) process to be considered stationary is (Rachev et al., 2008, p.191)

$$\alpha + \beta < 1 \tag{3.96}$$

- The long-term unconditional variance is given as (Rachev et al., 2008, p.191).

$$\sigma^2 = \frac{\omega}{1 - (\alpha + \beta)}. \tag{3.97}$$

- Since the autocovariance between returns at time t and $t - h$ is equal to *zero*, that is, $cov(r_{t-h}, r_t) = 0$, then the autocorrelations between the returns is equal to *zero* (Rachev et al., 2008, p.191).

3.8.3 Estimating Parameters of the GARCH Model Using the Maximum Likelihood Method

The Maximum Likelihood method is used to estimate parameters of the GARCH model. Suppose that $r_1, r_2, \cdots, r_T$, for $i = 1, 2, 3, \cdots, T$ is some available data and $\boldsymbol{\theta}$ is a vector of parameters. Then, the Likelihood function of parameter $\boldsymbol{\theta}$ given the data $r_1, r_2, \cdots, r_T$ is given as

$$L(\boldsymbol{\theta} \mid r_1, r_2, \cdots, r_n) = \prod_{t=1}^{T} f(r_t \mid \boldsymbol{\theta}) = f(r_1 \mid \theta) f(r_2 \mid \theta) \cdots f(r_2 \mid \theta), \quad (3.98)$$

where $f(r_1 \mid \theta), f(r_2 \mid \theta), \cdots, f(r_n \mid \theta)$ are known as the conditional density of the series $\mathbf{r_t}$ given parameter vector θ. Note that parameter $\boldsymbol{\theta}$ is a vector which contains the model parameters to be estimated.

Lemma:

According to (Rachev et al., 2008, p.192), if we assume that returns r_t are Normally distributed with expected value $\bar{r}$ and conditional variance σ_t^2, then we can describe the log-likelihood of the GARCH(1,1) model as

$$
\begin{aligned}
\log L(\boldsymbol{\theta} \mid \mathbf{r}) &= \sum_{t=1}^{T} \log f(r_t \mid \boldsymbol{\theta}) \\
&= const - \frac{1}{2} \sum_{t=1}^{T} \left(log(\sigma_t^2) + (\frac{y_t}{\sigma_t})^2 \right),
\end{aligned}
\quad (3.99)
$$

where $\mathbf{r} = r_1, r_2, \cdots, r_T$, $\boldsymbol{\theta} = (\omega, \beta, \alpha)$ is a vector which contains the GARCH model parameters that we want to estimate, and σ_t^2 is the variance given by Equation (3.84).

Proof:

Since we assumed that the returns r_t are Normally distributed, then we can define their distribution as

$$f(\mathbf{r_t} \mid \bar{r}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mathbf{r}_t - \bar{r}}{2\sigma^2} \right)^2. \quad (3.100)$$

Therefore, the likelihood function is given as

$$L(\boldsymbol{\theta} \mid \mathbf{r}) = \prod_{t=1}^{T} f(\mathbf{r_t} \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \quad (3.101)$$

Taking the log of the likelihood in Equation (3.101) we obtain

$$
\begin{aligned}
\log L(\boldsymbol{\theta} \mid \mathbf{r}) &= \sum_{t=1}^{T} \log\left(\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \right) \\
&= \sum_{t=1}^{T} \log\left((2\pi)^{-\frac{1}{2}} (\sigma_t^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \right) \\
&= \sum_{t=1}^{T} \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \\
&= \sum_{t=1}^{T} \left(-\frac{1}{2}\log(2\pi) - \sum_{t=1}^{T}\frac{1}{2}\log(\sigma_t^2) - \sum_{t=1}^{T}\frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \\
&= -\frac{T}{2}\log(2\pi) - \sum_{t=1}^{T} \left(\frac{1}{2}\log(\sigma_t^2) + \frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \\
&= const - \sum_{t=1}^{T} \left(\frac{1}{2}\log(\sigma_t^2) + \frac{1}{2}\left(\frac{r_t - \bar{r}}{\sigma_t}\right)^2 \right) \\
&= const - \frac{1}{2}\sum_{t=1}^{T} \left(\log(\sigma_t^2) + \left(\frac{y_t}{\sigma_t}\right)^2 \right),
\end{aligned}
$$

$$(3.102)$$

Equation (3.102) represents the log-Likelihood of the GARCH model with
normal distributed returns.

The GARCH(1,1) Model With Student-t Returns

Financial returns usually have a distribution with fat tails and higher peak.
As a result, a Student-t distribution has been widely used as an alternative
distribution for returns. A Student-t distribution with v degree of freedom
is given as

$$f(y_t; v) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)\sigma_t^2}} \left(1 + \frac{y_t^2}{(v-2)\sigma_t^2} \right)^{-\frac{v+1}{2}}, \qquad v > 2 \quad (3.103)$$

where σ_t^2 is conditional variance represented by Equation (3.84) and $\boldsymbol{\theta}$ is a vector which contains model parameters, thus $\boldsymbol{\theta} = (\omega, \beta, \alpha, v)$, $\Gamma(v) = \int_0^\infty e^{-x} x^{v-1} dx$ is the Gamma function, and parameter v measures the thickness of the tail of the distribution. According to (Alberg et al., 2008, p.1203), if $\mathbf{r_t}$ are financial asset returns which are Student-t distributed, then we can express the log-Likelihood of these returns as $\boldsymbol{\theta}$

$$\log L(\boldsymbol{\theta} \mid r_t) = T\left(log\Gamma\left(\frac{v+1}{2}\right) - log(\frac{v}{2}) - \frac{1}{2}log(\pi(v-2)) \right)$$
$$- \frac{1}{2} \sum_{t=1}^{T} \left(log(\sigma_t^2) + (1+v)log\left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right) \right)$$

Proof:

$$L(r_t \mid \boldsymbol{\theta}) = \prod_{t=1}^{T} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)\sigma_t^2}} \left(1 + \frac{y_t^2}{(v-2)\sigma_t^2}\right)^{-\frac{v+1}{2}}. \tag{3.104}$$

Taking the log-likelihood of Equation (3.104), then it follows that

$$\log L(r_t \mid \boldsymbol{\theta}) = \sum_{t=1}^{T} \log\left(\Gamma\left(\frac{v+1}{2}\right)\Gamma\left(\frac{v}{2}\right)^{-1} \left(\pi(v-2)\right)^{-\frac{1}{2}} (\sigma_t^2)^{-\frac{1}{2}} \right.$$
$$\left. \left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right)^{-\frac{v+1}{2}} \right)$$

$$= T\log\left(\Gamma\left(\frac{v+1}{2}\right)\Gamma\left(\frac{v}{2}\right)^{-1} \left(\pi(v-2)\right)^{-\frac{1}{2}} \right) - \frac{1}{2}\sum_{t=1}^{T}\log(\sigma_t^2) -$$
$$\sum_{t=1}^{T}\left(\frac{v+1}{2}\right)\log\left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right)$$

$$= T\log\left(\Gamma\left(\frac{v+1}{2}\right)\Gamma\left(\frac{v}{2}\right)^{-1} \left(\pi(v-2)\right)^{-\frac{1}{2}} \right) - \frac{1}{2}\left(\sum_{t=1}^{T}\log(\sigma_t^2) + \sum_{t=1}^{T}\left(\frac{v+1}{2}\right)\right.$$
$$\left.\log\left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right)\right)$$

$$= T\left(log\Gamma\left(\frac{v+1}{2}\right) - log(\frac{v}{2}) - \frac{1}{2}log(\pi(v-2)) \right)$$
$$- \frac{1}{2} \sum_{t=1}^{T} \left(log(\sigma_t^2) + (1+v)\log\left(1 + \frac{y_t^2}{\sigma_t^2(v-2)}\right) \right).$$

The equation above represents the log-likelihood function of a GARCH model with Student-t returns.

3.9   GARCH-type Models for Volatility Estimation

Some of the drawbacks of the GARCH model are:

- The GARCH model tends to react slowly to large isolated return shocks. As a result, the model usually over predict volatility (Shumway and Stoffer, 2006, p.287).

- Since constraints are placed on the model parameters, it results in the model being too restrictive (Shumway and Stoffer, 2006, p.287).

- The GARCH model assumes that negative and positive returns have the same effect, as a result of the model being dependent upon past squared returns (Shumway and Stoffer, 2006, p.287).

Some of the weakness of the GARCH model are similar to those of the ARCH model described earlier. To overcome some of the shortcomings of the GARCH model mentioned above, various extensions of the GARCH model have been proposed. In this study, we look at the Threshold Autoregressive Conditional Heteroscedasticity (TARCH), Exponential Weighted Moving Average (EWMA), Absolute Value Generalised Autoregressive Conditional Heteroscedasticity (AVGARCH), and the Exponential Generalised Autoregressive Conditional Heteroscedasticity (EGARCH) models.

3.9.1   Exponential GARCH (EGARCH) Model

The GARCH model is not adequate to capture asymmetric volatility. In 1991, Nelson proposed the Exponential GARCH model to deal with the issue of asymmetric volatility. Asymmetric volatility can simply be described as a condition in which the impact of negative stock price moves are bigger on future volatility than that of positive price moves (Sjöholm, 2015). In many financial literatures, asymmetric volatility is described as a phenomenon in which there are higher market volatility levels in declining markets than in rising markets. The proposed EGARCH model assumes

that negative and positive shocks do not have the same effect on volatility.
As a result, the EGARCH model is able to capture asymmetric volatility.
The EGARCH(p,q) model is given as (Tsay, 2005, p.144)

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \frac{\mid y_{t-i} \mid + \gamma_i y_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{p} \ln(\sigma_{t-j}^2), \tag{3.105}$$

where γ is just a constant. Since we model the $\log(\sigma_t^2)$, then even if the
model parameters are negative, σ_t^2 will always be positive. More impor-
tantly, the model is able to account for leverage effects, that is, if the relation-
ship between volatility and returns is negative, then γ will be negative. In
financial markets, negative shocks tends to have a bigger impact on volatil-
ity than positive shocks, this condition is known as *leverage effects*.

3.9.2 Exponential Weighted Smoothing Moving Average (EWMA) Model

The EWMA model is given as

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) y_{t-1}^2, \tag{3.106}$$

where λ is a constant whose value lies between 0 and 1. The parameter λ
is known as the *smoothing parameter*. Large values of λ are an indication of
slow decay in the returns series. The EWMA model is a simplified version
of the GARCH model. The advantage of the EWMA model is that more
weight is given to the most recent returns, this implies that current returns
would have more weight than past returns.

3.9.3 Absolute Value Generalised Autoregressive Conditional Heteroscedastic (AVGARCH) Model

The Absolute Value GARCH (AVGARCH) is also commonly known as the
Taylor-Schwart GARCH (TS-GARCH). The model is given as (Basu and
Saha, 2013, p.263)

$$\sigma_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \mid y_{t-i} \mid + \sum_{j=1}^{q} \beta_j \sigma_{t-j}. \tag{3.107}$$

The AVGARCH model reduces the influence of large observations by using the absolute value (Basu and Saha, 2013, p.263). The model assumes that negative and positive shocks have the same effect on volatility.

3.9.4   Threshold ARCH (TARCH) Model

In the TARCH model, volatility is modelled using absolute values. The TARCH model is closely related to the GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). The variance process in the Threshold ARCH (1,1) model is given as (Bollerslev, 2008, p.29)

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \gamma I_{t-1}[y_{t-1} < 0]y_{t-1}^2 + \beta \sigma_{t-1}^2, \qquad (3.108)$$

where I_{t-1} is known as a *dummy variable*, if y_{t-1} is negative it takes a value of 1 and takes a value of 0 if $y_{t-1} \geq 0$ (Franses, 1998, p.172). When $\gamma > 0$, then negative shocks will have a greater impact on volatility than positive shocks (Franses, 1998, p.172). Parameter γ is known as the *Assymetric term*. If $\gamma = 0$, then the model becomes a standard GARCH model. The main advantage of the TARCH is that it is able to capture asymmetric effects. Three excellent books on GARCH type models are (Rachev et al., 2008),(Shumway and Stoffer, 2006), and (Tsay, 2005).

3.10   AutoRegressive-GARCH type Models

In many financial literature, returns are modelled using an Autoregressive model whose random disturbances follows a GARCH process. Earlier we described the AR(p) and GARCH type models. We will use the properties described earlier to form an AR(p) model with GARCH noise. Suppose that **r$_t$** are returns series, then an Autoregressive model of order p with GARCH noise of order p, q is given as (Ferenstein and Gasowski, 2004a, 56).

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + \epsilon_t,$$
$$\epsilon_t = \sigma_t z_t, \qquad (3.109)$$

where ϵ_t is the white noise process with mean of *zero* and variance of *one*, and σ_t is the volatility. The volatility process σ_t satisfies the recurrence equa-

tion

$$\sigma_t^2 = \omega + \alpha \sum_{i=1}^{p} u_{t-j}^2 + \beta \sum_{j=1}^{q} \sigma_{t-j}^2,$$

which is the GARCH(p,q) model described earlier. The volatility process σ_t can also be described by any of the models in Equation (3.105),(3.106),(3.107), and (3.108). In financial literature, the most popular distributions for z_t are the Normal, Student-t, Generalised error and Hyperbolic distributions. The latter *three* distributions have heavier tails (leptokurtic property) than the Gaussian distribution (Ferenstein and Gasowski, 2004a, p.56). The AR(p)-GARCH type model described above is able to better capture the features of financial returns series such as leptokurtic property and volatility clustering.

3.11   ARMA Model in State-Space form

The state space approach is also known as the unobserved components time series models, because they try to incorporate unobserved components into a statistical time series model (Commandeur and Koopman, 2007, p.133). The state space approach is an alternative to the Box-Jenkins approach used for time series analysis. For the state space approach, we assume that if we have a time series $(\mathbf{r_t})_{t=1}^{n}$, then this time series depends on some un-observed states $(\mathbf{x_t})_{t=1}^{n}$. So, a model which uses the state space approach tries to specify a relationship between time series $\mathbf{r_t}$ and unobserved states $\mathbf{x_t}$. The unobserved state $\mathbf{x_t}$ encodes the unknown state of the system which we are studying, and $\mathbf{r_t}$ represents the observable quantities that dependent upon the state.

In the Box$-$Jenkins (ARIMA) framework, seasonal and trend effects are removed from the time series before statistical analysis can be performed. The state space methods do not require stationarity of a time series (Commandeur and Koopman, 2007, p.133). To shows properties of the state space method, we will consider several examples. Consider an AR(2) model which is expressed as

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \epsilon_t, \text{where} \quad \epsilon_t \sim \text{NID}(0, \sigma^2). \tag{3.110}$$

Then, the AR(2) model in Equation (3.110) can be written in state space form.

Let the state $\mathbf{X_t}$ be defined by $\mathbf{X_t} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \mathbf{X_{t-1}} + \mathbf{W_t}$, where $\mathbf{W}_t = \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$,

$\mathbf{r_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{X}_t$, and $\mathbf{X_{t-1}} = \begin{bmatrix} \mathbf{r_{t-1}} \\ \mathbf{r_{t-2}} \end{bmatrix}$. Using the properties above, we have that

$$\mathbf{r_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{X_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \mathbf{X_{t-1}} + \mathbf{W_t} \tag{3.111}$$

$$\mathbf{r_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r_{t-1}} \\ \mathbf{r_{t-2}} \end{bmatrix} + \mathbf{W_t}$$

$$= \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \mathbf{r_{t-1}} \\ \mathbf{r_{t-2}} \end{bmatrix} + \mathbf{W_t}$$

$$= \phi_1 r_{t-1} + \phi_2 r_{t-2} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

$$= \phi_1 r_{t-1} + \phi_2 r_{t-2} + \epsilon_t.$$

Hence, we have shown that the AR(2) model can be written in state space form.

3.11.1 ARMA(1,1) Model in State Space form

Consider an ARMA(1,1) model given as

$$r_t = \phi r_{t-1} + \theta \epsilon_{t-1} + \epsilon_t, \text{where} \quad \epsilon_t \sim \text{NID}(0, \sigma^2), \tag{3.112}$$

where θ and ϕ are model parameters.

Then, the state space representation of ARMA(1,1) model is

$$\mathbf{X_t} = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \mathbf{X_{t-1}} + \mathbf{W_t}, \tag{3.113}$$

where $\mathbf{W_t} = \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$ and $\mathbf{X_{t-1}} = \begin{bmatrix} r_{t-1} \\ \theta \epsilon_{t-1} \end{bmatrix}$.

Proof:

$$\mathbf{r_t} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{X_t}$$

$$r_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \mathbf{X_{t-1}} + \mathbf{W_t}$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \theta\epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix} \tag{3.114}$$

$$= \begin{bmatrix} \phi & 1 \end{bmatrix} \begin{bmatrix} r_{t-1} \\ \theta\epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

$$= \phi r_{t-1} + \theta\epsilon_{t-1} + \epsilon_t.$$

Hence we have shown that the ARMA(1,1) model can be written in state space form.

3.12   Artificial Neural Network (ANN)

The first concept of a neuron based model was introduced by Walter Pitts and Warren McCullock in 1943. The model was designed to "mimic" the human brain. Neurons are nerve cells in the brain that are responsible for processing and transmitting electrical and chemical signals, as illustrated in Figure below.
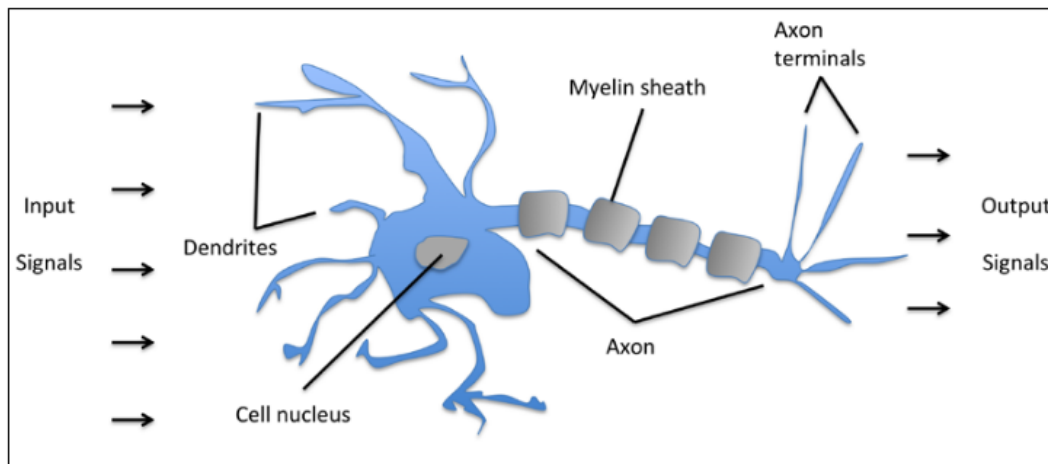
Figure 3.1: A biological neuron (Raschka, 2015, p.18)

Inputs signals arrives at the dendrites, they are then incorporated into the cell body. Then, the signals are summed together, if the sum exceeds a specified threshold then the neuron fires (output signal is generated). In 1957, Frank Rosenblatt introduced the concept of perceptron learning rule based on the neuron model proposed by Walter Pitts and Warren McCullock. The perceptron learning rule algorithm proposed by Rosenblatt automatically learns optimal weights, these weights are then multiplied by the input features (Raschka, 2015, p.17-18). Then, a decision will be made whether the neuron fires or not (Raschka, 2015, p.17-18). In supervised learning and classification problems, such an algorithm will then be used to predict whether a given sample falls within one class or another. Summary of the concept of perceptron learning rule is illustrated in the figure below.
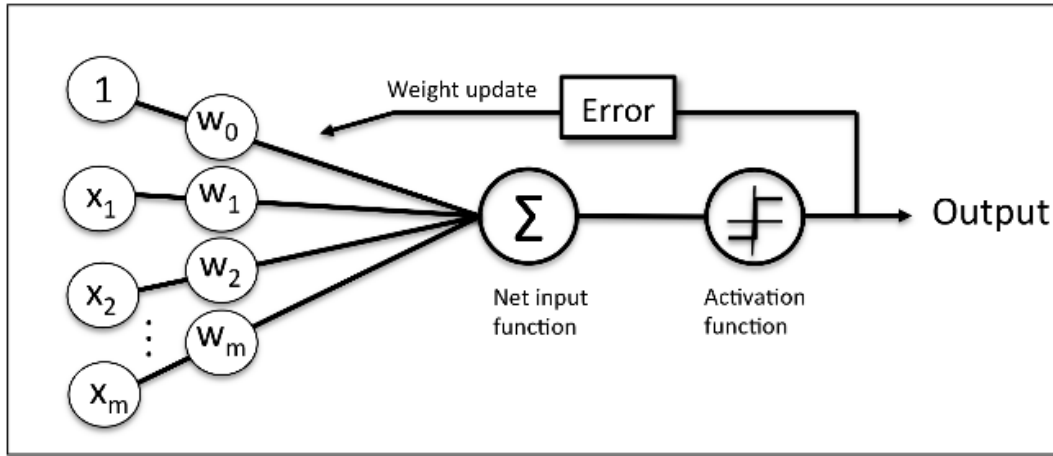
Figure 3.2: Concept of perceptron (Raschka, 2015, p.24)

The preceding figure shows how the perceptron works. The inputs **X** are combined with weights **W** to obtain the net input. This net input is then passed through the activation function. The activation function would then generate an binary output or continuous output depending on whether we are dealing with a classification or a regression problem. During the learning phase, the output obtained is used to compute the error of the predictions and to update the synoptic weights (Raschka, 2015, p.24). These weights shows the strength of the connections. The larger the value of the weights the stronger the connections. The types of activation function that can be used would be discussed later. More formally, the above can be summarised mathematically. Suppose z is the net input given as $z = w_1 x_1 + \cdots + w_n x_n$, then the activation function is given as $\phi(z)$. This activation function is a linear combination of input features **X** and their corresponding weights vector **W**, such that

$$
\mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{3.115}
$$

Minimizing the Cost Function Using the Gradient Descent Technique

The most important aspect of supervised machine learning algorithm is the need to define a cost function J(w) that would be optimized during the

learning process. One of the most commonly used cost function is the sum of squared error (SSE) which is given as (Raschka, 2015, p.34)

$$J(W) = \frac{1}{2} \sum_i \left(y^{(i)} - \phi(z^{(i)}) \right)^2, \tag{3.116}$$

where $y^{(i)}$ is the target value of sample i, $\phi(z^{(i)})$ is the activation function output for sample i, and the term $\frac{1}{2}$ is added for convenience reasons (Raschka, 2015, p.34). Suppose we have a linear activation function

$$\phi(z) = \mathbf{w}^T\mathbf{x}, \tag{3.117}$$

then this activation function is just an identity function of the net input defined earlier. The main advantage of using a continuous linear activation function is that the cost function in Equation (3.116) becomes differentiable. Another advantage of the SSE cost function is that it is convex, as result, we can use a powerful optimization algorithm known as the *gradient descent*. This algorithm is used to find the weights that will minimise the cost function $J(W)$. The figure below illustrate the principle behind gradient descent algorithm. Gradient descent can simply be described as climbing down the
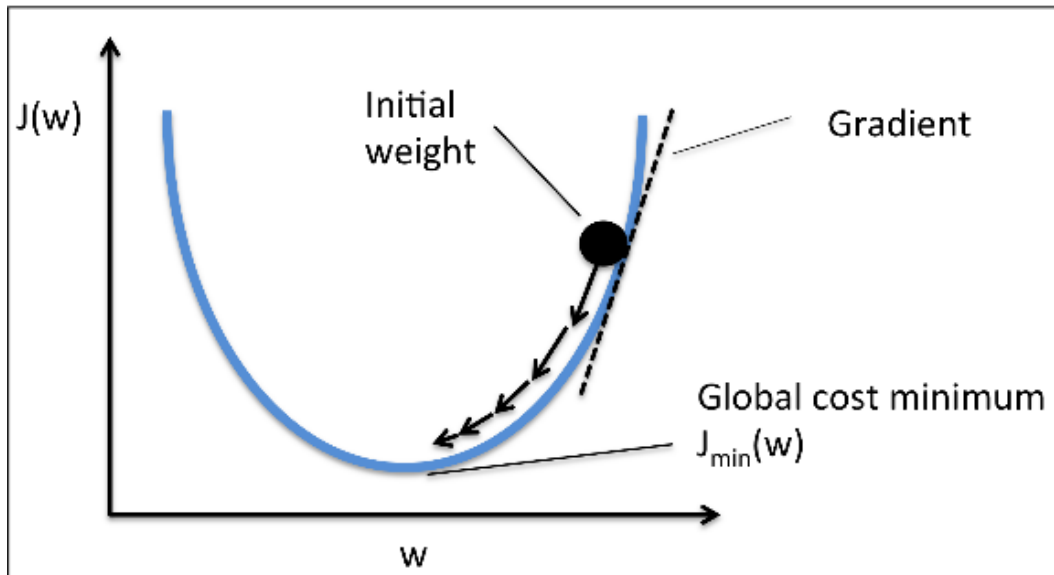
Figure 3.3: Gradient descent (*climbing down a hill*) (Raschka, 2015, p.35)

hill until we reach global or local cost minimum (Raschka, 2015, p.34). At each iteration, we take a step in the opposite direction of gradient descent,

and the size of the step is determined by the value of the learning rate λ (Raschka, 2015, p.34). We update the weights by taking a step away or in the opposite direction of gradient $\triangledown J(W)$ (gradient of the cost function).

$$W := W + \triangle W. \tag{3.118}$$

The weight change $\triangle W$ is defined by

$$\triangle W = -\lambda \triangle J(W), \tag{3.119}$$

where $-\lambda$ is the learning rate. In order to compute the gradient of the cost function $J(W)$, we have to compute the partial derivative of $J(W)$ with respect to individual weights W_j. Therefore, using Equation (3.116) and $\phi(z) = \mathbf{w}^T \mathbf{x}$, then we have that

$$\begin{aligned}
\frac{\partial J}{\partial W_j} &= \frac{\partial}{\partial W_j} \frac{1}{2} \sum_i \left(y^{(i)} - \phi(z^{(i)}) \right)^2 \\
&= \frac{1}{2} \sum_i 2 \left(y^{(i)} - \phi(z^{(i)}) \right) \frac{\partial}{\partial W_j} \left(y^{(i)} - \phi(z^{(i)}) \right) \\
&= \sum_i \left(y^{(i)} - \phi(z^{(i)}) \right) \frac{\partial}{\partial W_j} \left(y^{(i)} - \sum_i W_j^{(i)} x_j^{(i)} \right) \\
&= \sum_i \left(y^{(i)} - \phi(z^{(i)}) \right) \left(-x_j^{(i)} \right).
\end{aligned} \tag{3.120}$$

The results above implies that

$$\triangle W = -\lambda \triangle J(W) = -\lambda \sum_i \left(y^{(i)} - \phi(z^{(i)}) \right) \left(-x_j^{(i)} \right). \tag{3.121}$$

The weight update is computed using all samples in the training set, hence this is why the approach is also known as "batch" gradient descent (Raschka, 2015, p.36).

Activation Function

One of the most important aspect to consider when constructing an ANN is choosing appropriate activation functions for the hidden and output layers. The activation function should be differentiable, this is because calculating the backpropagated error signal requires that we should first compute the

gradient of the activation function. The most commonly used activation
functions are the linear activation function, the hyperbolic tangent, and the
logistic sigmoid function. The hyperbolic tangent is given as

$$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{\sinh(z)}{\cosh(z)}. \tag{3.122}$$

The hyperbolic tangent is sigmoidal ("s-shaped") and it outputs values in
the range (-1,1). Its derivative is given as

$$
\begin{aligned}
\phi(z)' &= \frac{\partial}{\partial z} \frac{\sinh(z)}{\cosh(z)} \\
&= \frac{\frac{\partial}{\partial z}\sinh(z) * \cosh(z) - \frac{\partial}{\partial z}\cosh(z) * \sinh(z)}{\cosh^2(z)} \\
&= \frac{\cosh^2(z) - \sinh^2(z)}{\cosh^2(z)} \\
&= 1 - \frac{\sinh^2(z)}{\cosh^2(z)} \\
&= 1 - \tanh^2(h).
\end{aligned}
\tag{3.123}
$$

The logistic sigmoid is given as

$$\phi(z) = \frac{1}{1 + e^{-z}}. \tag{3.124}$$

It outputs values in the range (0,1), and its derivation is given as

$$
\begin{aligned}
\phi(z)' &= \frac{\partial}{\partial z}\left(\frac{1}{1 + e^{-z}}\right), \quad \text{using the chain rule we obtain} \\
&= \frac{e^{-z}}{(1 + e^{-z})^2} \\
&= \frac{1 + e^{-z} - 1}{(1 + e^{-z})^2} \\
&= \frac{1 + e^{-z}}{(1 + e^{-z})^2} - \frac{1}{(1 + e^{-z})^2}, \quad \text{but} \quad \phi(z) = \frac{1}{1 + e^{-z}}, \text{so that} \\
&= \phi(z)(1 - \phi(z))
\end{aligned}
\tag{3.125}
$$

Introduction to Multilayer Neural Network (MLP) Architecture

The figure below illustrates how we can connect multiple neurons to form a multilayer feedforward neural network, this type of fully connected neural network is known as the *Multilayer Perceptron (MLP)*
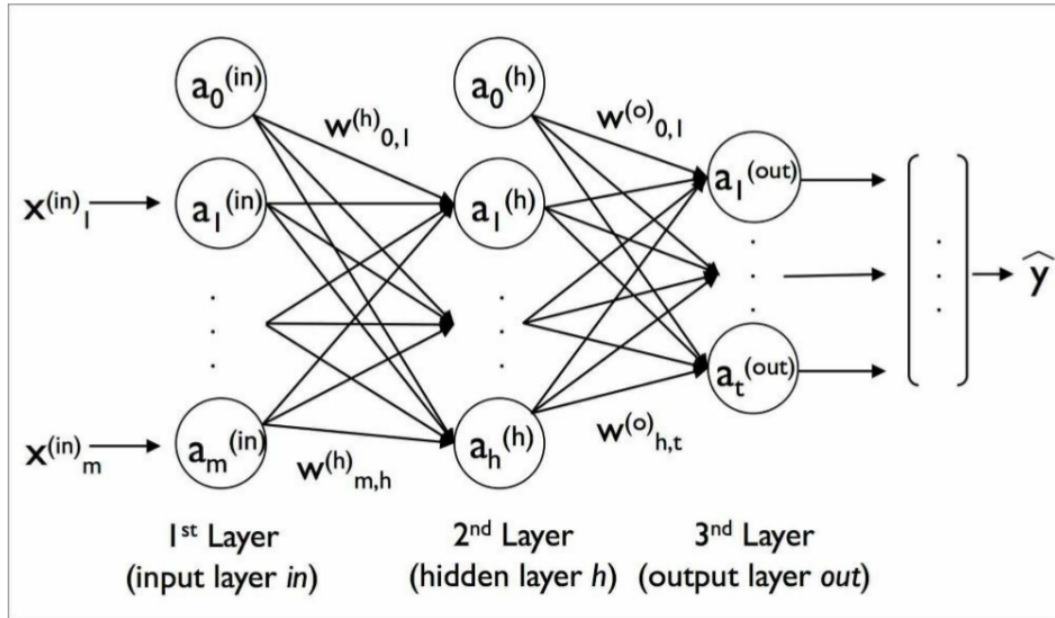
Figure 3.4: Concept of Multilayer Perceptron (MLP) (Raschka and Mirjalili, 2017, p.548)

The Multilayer Perceptron (MLP) depicted in Figure 3.4 is made up of one input layer, one hidden layer, and one output layer. The input layer is fully connected to the hidden layer, and all units in the hidden layer are fully connected to the output layer of the network. A neural network with more than one hidden layer is often referred to as a *Deep Artificial Neural Network*. From the preceding figure, $a_i^{(l)}$ represents the ith activation unit in the lth layer of the network. Therefore, $a_i^{(in)}$ is the ith unit in the input layer, $a_i^{(h)}$ represents the ith unit in the neural network hidden layer, and $a_i^{(out)}$ represents the ith unit in the final (output) layer (Raschka and Mirjalili, 2017, p.549). The weight coefficient $W_{k,j}^{(l)}$ represents the connection weight between kth unit in layer l and jth unit in the next layer $l+1$. Therefore, $W^{(h)}$ is the weight matrix that connects the input layer to the hidden layer of the

network, and $W^{(out)}$ represents the weight matrix that connects the hidden
layer to the output layer.

Forward Propagation

Forward propagation process is used to compute the output of a Multilayer
Perceptron model (MLP). To better understand how forward propagation
works, we summarize the MLP learning process as follows (Raschka and
Mirjalili, 2017, p.553).

- Beginning from the input layer, we forward propagate our training
  data through the neural network so that we can generate the output.

- Using the network's generated output, we compute the error that will
  minimise the cost function in Equation (3.116).

- Then, in the final stage we backpropagated the error, and then com-
  pute the derivative of the error with respect to individual weights in
  the neural network, and then update our model.

The forward propagation process described above can be expressed mathe-
matically as

$$z^{(h)} = a^{(in)} W^{(h)}, \tag{3.126}$$

$$a^{(h)} = \phi(z^{(h)}), \tag{3.127}$$

$$z^{(out)} = a^{(h)} W^{(out)}, \tag{3.128}$$

where $z^{(h)}$ is the net input of the hidden layer, $z^{(out)}$ is the net input of the
output layer, $a^{(h)}$ and $W^{(out)}$ were defined earlier. For more details about the
forward propagation process see (Raschka and Mirjalili, 2017, p.553-556).

Training a Neural Network

The backpropagation algorithm is the most popular algorithm used to train-
ing an artificial neural network. In simple terms, we can define the back-
propagation algorithm as machine learning algorithm used in multilayer
neural networks to compute the partial derivatives of a cost function (Raschka
and Mirjalili, 2017, p.584) . Then, we use those derivatives to learn the
weights for parameterizing that particular multilayer neural network (Raschka

Compute the gradient:

$$\frac{\partial}{\partial w_{i,j}^{(out)}} J(\boldsymbol{W}) = a_j^{(h)} \delta_i^{(out)}$$

Error term of the output layer:

$$\boldsymbol{\delta}^{(out)} = \boldsymbol{a}^{(out)} - \boldsymbol{y}$$

Input **x**

Output **ŷ** ← Target **y**

Error term of the hidden layer:

$$\boldsymbol{\delta}^{(h)} = \boldsymbol{\delta}^{(out)} \left(\boldsymbol{W}^{(out)}\right)^T \odot \frac{\partial \phi\left(z^{(h)}\right)}{\partial z^{(h)}}$$

Compute the gradient:

$$\frac{\partial}{\partial w_{i,j}^{(h)}} J(\boldsymbol{W}) = a_j^{(in)} \delta_i^{(h)}$$
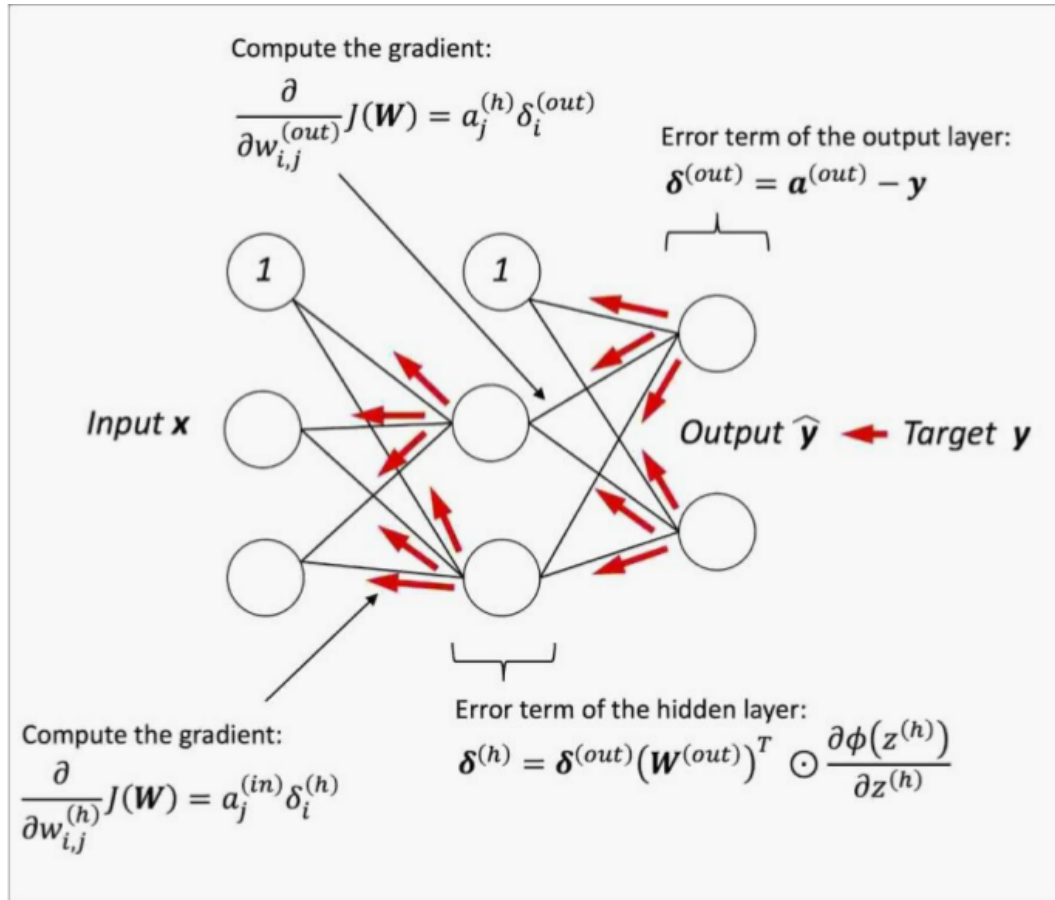
Figure 3.5: Backpropagation (Raschka and Mirjalili, 2017, p.586)

and Mirjalili, 2017, p.584). The figure above illustrates how the backpropagation algorithm works.

Before we apply the backpropagation algorithm, we first need to apply the forward propagation process discussed earlier. The first step in backpropagation algorithm is to compute the error of the output layer using :

$$\delta^{(out)} = a^{(out)} - y, \tag{3.129}$$

where y is the vector of actual values, and $a^{(out)}$ is the activation of the network output layer. Then, the error of the network's hidden layer is com-

puted by (Raschka and Mirjalili, 2017, p.587-588)

$$\delta^{(h)} = \delta^{(out)} \left(W^{(out)}\right)^{(T)} * \frac{\partial \phi(z^{(h)})}{\partial z^{(h)}}, \tag{3.130}$$

where $\frac{\partial \phi(z^{(h)})}{\partial z^{(h)}}$ is the derivative of the activation function, this activation function could be one of those described in Equation (3.122),(3.117), and (3.124). Suppose we choose the logistic sigmoid as the activation function, then taking the derivative of this activation function, the error of the hidden layer becomes (Raschka and Mirjalili, 2017, p.588)

$$\delta^{(h)} = \delta^{(out)} \left(W^{(out)}\right)^{(T)} * \phi(z)(1 - \phi(z)). \tag{3.131}$$

The name backpropagation comes from the fact that we start to compute the delta (error) from the output layer and then go backward to compute the delta terms for the preceding layers. After we obtain the δ (errors), we can now compute the derivation of the cost function in Equation (3.116) with respect to the weights.

$$\frac{\partial}{\partial W_{i,j}^{(out)}} J(w) = a_j^{(h)} \delta_i^{(out)}, \tag{3.132}$$

$$\frac{\partial}{\partial W_{i,j}^{(h)}} J(w) = a_j^{(in)} \delta_i^{(h)}. \tag{3.133}$$

The next step is to a accumulate all the partial derivatives in every layer and the error of the node in the next layer of the neural network. Expressed mathematically as
$\Delta^{(out)} = \Delta^{(out)} + (a^{(h)})^T \delta^{(out)}$ and $\Delta^{(h)} = \Delta^{(h)} + (a^{(in)})^T \delta^{(h)}$. The last step is to add the regularization term to the accumulated partial derivatives.

$$\Delta^{(l)} = \Delta^{(l)} + \rho^{(l)}, \tag{3.134}$$

where $\rho^{(l)}$ is the regularization term. Regularization is a technique used in machine learning to prevent overfitting. The regularization term could be $L1$ or $L2$ given respectively as

$$L1 = \eta \sum_{i=1}^{m} \mid w_i \mid, \tag{3.135}$$

$$L2 = \eta \sum_{i=1}^{m} w_i^2. \tag{3.136}$$

The *L*2 regularization is simply the sum of squares of the weights, while *L*1 is the sum of the weights. After we compute the gradients, we can now be able to update the weights by taking a step in the opposite direction of gradient for each layer *l* in the neural network. That is,

$$W^{(l)} = W^{(l)} - \lambda \Delta^{(l)}. \tag{3.137}$$

For more details on how to construct a multilayer perceptron model see (Raschka, 2015, p.17-47),(Raschka and Mirjalili, 2017, p.565-592).

Chapter 4

Methodology

In this Chapter, we discuss the data to be used in the study. We also look at the techniques and methods which will be used for data analysis. Lastly, we will discuss the model diagnostics.

Data

The time series data used in this study was from the Johannesburg Stock Exchange (JSE). The JSE is the largest and oldest stock exchange in Africa, with a market capitalisation of around R10 trillion rands as of 2013. It is located in Sandton, Johannesburg, and was founded on 18 November 1887. The time series data contains the historical opening, closing, high, and low prices. The prices are in South African Rands. The log returns are computed using Equation (3.1). The data used in this study was obtained from Google finance website. Data to be used is from 02 January 2013 to 04 January 2017. In this study, we will model financial log-returns using closing prices from Four JSE listed companies namely

- Netcare Health Limited, Sanlam Limited group, Nedbank Group Limited, Santam Limited.

4.1   Techniques For Data Analysis

As stated earlier in Chapter 1, statistical time series models such as the Autoregressive Moving Average model, State Space ARMA Model, Autoregressive Moving Average model with parameters estimated through the

Metropolis Hastings algorithm, Autoregressive GARCH type models, and an Artificial Neural Network will be used for forecasting stock prices of the Four JSE listed companies. We will split the data into two training and testing sets, as follows

- Train 500 Test 200

- Train 700 Test 200

Then, our models will be evaluated on how best they can predict the direction of movement of prices, we will use a Confusion matrix to summarize the results. The SPA test will be used to choose the "best" model out of all competing models.

4.1.1   Decay rate and Amplitudes

The stationarity of the ARMA and GARCH models implies that there could be an initially large component which decays quickly to 0. The rate of decay is in fact dependent on the initial value and the coefficients, and unfortunately in these models, it affects the amplitude and the variance as well, but as a kind of positive amplifier. However, the data available does not reflect the decay but only the steady state, so there is not sufficient data to ensure that the decay rate and the amplitudes are correct. We do not deal with this problem in this thesis, since it needs much further investigation, but for our up/down problem it is not necessary to know the amplitudes, so we defer the investigation to another context.

4.2   Model Diagnostics and Statistical Tests

(Tsay, 2005, p.36-50), (Box and Jenkins, 1976) suggested that some diagnostics and statistical tests should be carried out to determine if the fitted statistical time series model is adequate.

Testing for Presence of ARCH or GARCH Effects

Before applying the ARCH or GARCH model, it is important to first test for the presence of ARCH \ GARCH effects. Suppose we let $\epsilon_t = r_t - \mu_t$ to

be the residuals of the mean equation described earlier. Then, we can use ϵ_t^2 to detect the presence of ARCH\GARCH effects. Two tests normally used for testing for the presence of the effects are Ljung-Box test statistic, and the Lagrange Multiplier test. The Lagrange Multiplier test is equivalent to the *F* statistic test for testing that all parameters of a linear regression are equal to zero (Tsay, 2005, p.114). According to (Andersen et al., 2009, p.g121-124), the test is applied to the residuals from the fitted conditional mean equation or to the squared and absolute financial returns. Now

$$\epsilon_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \cdots + \phi_m \epsilon_{t-m}^2 + e_t, \tag{4.1}$$

for $t = m + 1, \cdots, T$, where e_t is the error term, T denotes the sample size, and m represent a positive integer. The null hypothesis is that there is no ARCH effects, that is,

$$H_0 = \phi_1 = \phi_2 = \cdots = \phi_m = 0. \tag{4.2}$$

We then have that,

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - 2m - 1)}, \tag{4.3}$$

where $SSR_0 = \sum_{t=m+1}^{T}(\epsilon_t^2 - \bar{\omega})^2$, $\bar{\omega} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$ (mean of squared residuals), and $SSR_1 = \sum_{t=m+1}^{T}\hat{e}_t^2$ (where $\hat{e}_t^2$ represents the least square residuals of the linear regression we defined earlier). It is important to know that the *F* statistic is asymptotically distributed as chi-square distribution with *m* degree of freedom (Tsay, 2005, p.114). The decision rule is that we reject the null hypothesis if $F > \chi_m^2(\alpha)$. For more details about the Lagrange Multiplier test and the importance of testing for ARCH effects see (Andersen et al., 2009, p.121-123), and (Tsay, 2005, p114-115).

Testing for Stationarity

The **Augumented Dickey Fuller (ADF)** test is used to test for stationarity of the returns series. Nonstationarity in stock market occurs because stock prices have no fixed price, they fluctuate depending on many factors, e.g political events, natural disasters (Tsay, 2005, p.72). The ADF test is based on the model

$$\Delta y_t = \alpha + \gamma y_{t-1} + \delta_t + \Delta y_{t-1}\zeta_1 + \zeta_2 \Delta y_{t-2} + \cdots \zeta_k \Delta y_{k-t} + e_t, \tag{4.4}$$

where α is the constant term of the model, ζ_i are the model parameters, δ_t is the time trend, k is the number of lags, and Δ represents the differencing operator. The ADF tests the null hypothesis of presence of unit root (nonstationary) against alternative hypothesis of no unit root (stationarity). From Equation (4.4), the null hypothesis of unit root is equivalent to letting $\gamma = 0$ and the alternative hypothesis of no unit root is equivalent to setting $\gamma > 0$. The test statistic for the ADF test is given as

$$\tau = \frac{\hat{\gamma}}{\sigma_{\hat{\gamma}}}, \tag{4.5}$$

where $\sigma_{\hat{\gamma}}$ represents the standard error of the coefficient estimate, and $\hat{\gamma}$ represents the estimated coefficient. The decision rule is that we reject the null hypothesis if the value of the test statistic is less than the critical values for the 0.05 significance level. The smaller the p-value the stronger the evidence that the returns series is stationary.

The **Phillips Perron (PP)** test is another statistical test which is used to test for the presence of a unit root in a series. It is named after Peter Phillips and Pierre Perron. It tests the null hypothesis that the series contains a unit root against the alternative hypothesis that the series was generated by a stationary process. The PP test has become popular in the analysis of financial time series data. For a univariate time series y, the PP test tests the null hypothesis of the presence of a unit root . The test is based on the model

$$y_t = c + \delta_t + a y_{t-1} + e_t, \tag{4.6}$$

where δ and c are the deterministic trend and drift coefficients respectively, e_t represents the innovations of the process. The null hypothesis restricts a to be equal to 1. For a detailed explanation of the Phillips Perron test see the original work of (Phillips and Perron, 1988).

Testing for Independence

The Ljung-Box test statistic was named after Greta Ljung and George P. Box. It is a statistical test used to test whether autocorrelations of a particular time series are different from zero. It tests the null hypothesis that the data are independently distributed against an alternative hypothesis that the data

are not independently distributed. Suppose r_t and r_{t-l} are returns at time t and $t - l$ respectively, then the correlation coefficient is given as (Tsay, 2005, p.31)

$$\rho_l = \frac{\text{Cov}(r_t, r_{t-l})}{\sqrt{\text{Var}(r_t)\text{Var}(r_{t-l})}}, \tag{4.7}$$

where $\text{Cov}(r_t, r_{t-l})$ is the covariance between r_t and r_{t-l}, $\text{Var}(r_t)$ is the variance of r_t, and $\text{Var}(r_{t-l})$ is the variance of r_{t-l}. For a weakly stationary series, we have that $\text{Var}(r_t) = \text{Var}(r_{t-l})$ (Tsay, 2005, p.31). For more details about the correlation coefficient see (Tsay, 2005, p.30-32). The Ljung Box test statistic is given as (Tsay, 2005, p.32)

$$Q(n) = K(K+2) \sum_{l=1}^{n} \frac{\hat{\rho}_l^2}{K-l}, \tag{4.8}$$

where K is the size of the sample. It tests the null hypothesis that $H_0 : \rho_1 = \rho_2 = \cdots = \rho_n = 0$ against alternative hypothesis $H_a : \rho_i \neq 0$. The decision rule is to reject the null hypothesis if $Q(n) > \chi_\alpha^2$, where χ_α^2 represents a $100(1-\alpha)th$ percentile of the well known Chi-square distribution which has n degrees of freedom. Also note that if the null hypothesis is true, $Q(n)$ has a Chi-square distribution with n degrees of freedom. If the p-value is less than the significance level (denoted by α) of 0.05, then we reject the null hypothesis of independence of residuals and conclude that the model does not meet the independence assumption. Once the null hypothesis is rejected, the proposed model may not fit the data and we should be cautious when we interpret the results. A small p-value is an indication of strong evidence against the null hypothesis. For more details about the Ljung Box test see (Tsay, 2005, p.30-33).

Testing for Normality

One of the most important task in statistical analysis is the need to characterize the variability and location of a data set. This includes using techniques such as *Skewness* and *Kurtosis*.

Kurtosis is usually described as the measure of thickness of the tails of a distribution relative to that of a Gaussian distribution. A data set with a high kurtosis value tend to have "heavier" tails or many outliers. Similarly,

a data set with small kurtosis value tend to have few outliers or "light" tails. Suppose that y is a random variable, then its kurtosis (K) is given as

$$\text{Kurtosis}(K) = \frac{E(y - \mu)^4}{\sigma^4}, \tag{4.9}$$

where σ and μ are the standard deviation and the expected value of y. For a Gaussian distribution, kurtosis is equal to *three*. If kurtosis $(K) > 3$, then the distribution of that particular data set has heavy tails and it is highly peaked around the mean. If kurtosis $(K) < 3$, then the sample data has a flatter distribution than that of a Gaussian distribution.

Skewness can be described as the measure of symmetry or lack of symmetry of a distribution. A distribution is said to be symmetric if it "looks" the same on both the right and left hand side of the main centre point of the distribution. The Skewness (S) for a random variable y is given as

$$\text{Skewness}(S) = \frac{E(y - \mu)^3}{\sigma^3}, \tag{4.10}$$

where σ and μ are the standard deviation and expected value of y. If the skewness value is equal to *zero*, it is an indication that the distribution would be symmetric around the mean. While a negative skewness value indicates that the distribution will have a long left tail. On the other hand, a positive skewness value indicates that the distribution will have a long right tail.

Shapiro Wilk Test

The Shapiro Wilk test was first introduced by Samuel Sanford Shapiro and Martin Wilk in 1965, hence the name Shapiro Wilk. The Shapiro Wilk test tests the null hypothesis that the sample $y_1, y_2, \cdots, y_n$ comes from a normally distributed population against an alternative hypothesis that the sample does not come from a population which is normally distributed. The Shapiro Wilk test statistics is given as

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_{(i)} \right)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{4.11}$$

where a_i are constants which are generated from the variances, covariances, and means of the order statistics of a normal distribution with a sample of

size n, and $y_{(i)}$ are the ordered sample values (they are arranged from the smallest $y_{(1)}$ to the largest $y_{(n)}$). If the *p-value* is less than the 0.05 significance level, then we reject the null hypothesis of normality. If the data fails the normality test, then we can state with 95% confidence that the data is not Gaussian distributed. On the other hand, if the normality assumption is accepted, then we can conclude that there is no significant departure from normality. For more details about the Shapiro Wilk test see (Shapiro and Wilk, 1965).

Testing for Autocorrelations

The **Durbin Watson (DW)** test statistic is used to detect the presence of autocorrelations in the residuals. It tests the null hypothesis that there is no significant autocorrelations in the residuals against the alternative hypothesis of the presence of autocorrelations in the residuals. The DW test statistic is given as

$$DW = \frac{\sum_{t=1}^{T} \left(e_t - e_{t-1}\right)^2}{\sum_{t=1}^{T} e_t^2}, \tag{4.12}$$

where e_t is the error at time step t. The DW test statistic is approximately equal to $2(1-\rho)$, where ρ represents the sample autocorrelation of the model residuals. If $\rho = 0$, then this is an indication that there is no serial correlation in the residuals, and DW test statistic would be equal to 2. The value of the DW test statistic always lies between 0 and 4. A DW value near 2 indicate that there is no autocorrelation in the sample. On the other hand, values approaching 0 are a clear indication of the presence of positive autocorrelations, while values approaching 4 are an indication of presence of negative autocorrelation. To get conclusive results from the test, the DW test statistic is compared to the upper and lower critical bounds from the Durbin Watson significance table. The upper critical value is denoted by d_u and the lower critical value by d_l. The significance level for the DW test is 0.05. For more details about the Durbin Watson test statistic see the original paper of (Durbin and Watson, 1951).

Assessing the Adequacy of the Fitted ARMA Model

After fitting any ARMA type model, It is important to examine the model

residuals to check for model inadequacy (Tsay, 2005, p.50). The residuals of the fitted model should behave as a white noise process in order for the fitted model to be considered adequate (Tsay, 2005, p.50). Suppose the estimate of ϕ_i is denoted by $\hat{\phi}_i$ and of θ_i is denoted by $\hat{\theta}_i$, then the fitted ARMA(1,1) model is given as

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} - \hat{\theta}_1 \epsilon_{t-1} + \epsilon_t, \qquad (4.13)$$

then the associated residual is given as (Tsay, 2005, p.49-50)

$$\hat{\epsilon}_t = r_t - \hat{r}_t, \qquad (4.14)$$

where r_t is given by Equation (3.26). The series $\{\hat{\epsilon}_t\}$ is known as the *residual series*. The residuals in Equation (4.13) are assumed to be white noise, this implies that they are independent and identically distributed from a Gaussian distribution with a mean of *zero* and constant variance. The Ljung-Box test and the ACF are used to check how close the residuals are to the white noise process (Tsay, 2005, p.50). For more details on checking for model adequacy see (Tsay, 2005, p.49-51).

Significance of Coefficients of ARMA Model

For each estimated ARMA model coefficient, the Z score is calculated as

$$\text{Z-score} = \frac{\text{Estimated coefficient}}{\text{Standard error of coefficeint}}. \qquad (4.15)$$

If $\mid$ Z-score $\mid > 1.96$, then the estimated ARMA model coefficient is significantly different from zero. It is well known that 95% of the area under the normal curve lies within 1.96 standard deviations of the expected value (mean). So, by the central limit theorem we construct the confidence interval using 1.96. Large Z-scores negative or positive are an indication that the estimated coefficients are significantly different from zero.

Evaluating the Performance of the Model

A confusion matrix is a statistical technique used for summarizing the performance of a classification model. It gives us a general idea of how the

model is performing. Classification can simply be described as the ratio of correct model predictions to total number of all predictions made. Suppose that the predictions made by a classification algorithm is given on the table below:

Predicted	Down	Up
Down	γ	β
Up	ρ	θ

Table 4.1: A Confusion Matrix

Where $\gamma, \beta, \rho, \theta \in \mathbb{N}$. Then, from Table 4.1, we can deduce the following properties of a Confusion Matrix:

- **Accuracy**: It represents the percentage of all correct predictions made by the classification algorithm, and it is given as:

$$\text{Accuracy} = \frac{\gamma + \theta}{\gamma + \theta + \beta + \rho}. \tag{4.16}$$

- **Misclassification**: It represents the percentage of all incorrect predictions made by the classifier, and it is given as

$$\text{Misclassification} = 1 - \frac{\gamma + \theta}{\gamma + \theta + \beta + \rho}. \tag{4.17}$$

The misclassification rate is also known as the *Error rate*. There are other statistics such as Sensitivity, Specificity, and Prevalence which can be obtained from the output of a Confusion matrix, but in this study, we will only focus on the accuracy and misclassification rate. In financial stock returns forecasting, classification is done by predicting whether the next day's stock returns price will be lower or higher than the stock returns price at the previous day.

Superior Predictive Ability (SPA) Test

The Superior Predictive Ability test also known as the *Reality Check* is a statistical test used to determine whether any model in a given set of models outperforms the benchmark model. It was first introduced by Peter Hansen in 2001. The procedure takes losses as inputs. As a result, smaller values are favoured over large values. To implement the procedure we need $n-$element array of benchmark model losses and n by m element array of model losses. The SPA tests the null hypothesis that no model outperforms or is better than the benchmark model, that is,

$$H_0 = \max_{i} \quad E[L_i] \geq E[L_{bm}], \tag{4.18}$$

where L_{bm} are losses from the benchmark model, and L_i are losses from model i. The p-values from the SPA test represents the probability that the null hypothesis is true. The test has three p-values: the lower, consistent, and upper p-value. Each of the p-value represents different re-centerings of the model losses. we should have that

$$\text{lower p-value} \geq \text{consistent p-value} \leq \text{upper p-value} \tag{4.19}$$

We reject the null hypothesis if the p-values are less than the 0.05 significance level. The p-value is a probability that measures the "amount" of evidence we have against the null hypothesis of a statistical test. For more details about the SPA test see the original papers (Hansen et al., 2003), (Hansen et al., 2001).

Chapter 5

Data Analysis and Results

In this chapter, results and analysis of the fitted models will be discussed. The statistical software packages used for data analysis are Python 2 and Python 3. Analysis of the JSE stock returns using ARMA-type time series models is performed in six steps, that is, preliminary tests on returns series, model fitting, model diagnostics, assessing adequacy of the fitted model, evaluating the performance of the model, and selecting the overall best model using Superior Predictive Ability test. Preliminary tests are carried out on the returns series, this includes tests such as ADF test, PP test, Shapiro Wilk test for normality, and LM test for testing for ARCH effects. The models would then be fitted to the returns series after the appropriate statistical preliminary tests have been carried out. Diagnostics tests are carried out on the residuals of the fitted model. The ACF is then used to assess the adequacy of the fitted time series model. Accuracy of the models is evaluated using a Confusion Matrix, and the SPA test is then used to choose the overall "best" model.

Plots of the Actual Closing Price Series
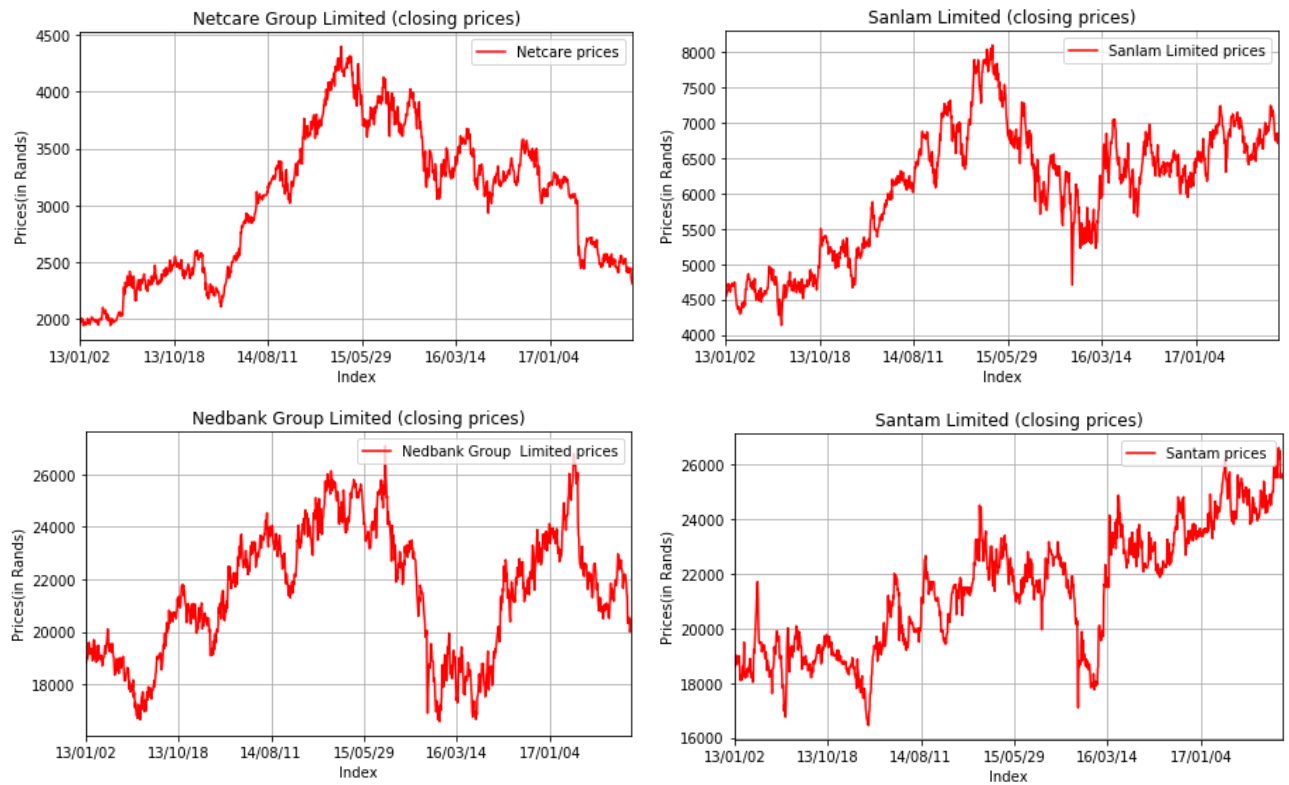
Figure 5.1: **Daily closing prices series of JSE listed companies**

From Figure 5.1 above, we can see that the mean and the variance of the
closing prices are not asymptotically constant, this implies that the series
are not stationary. So, we have to perform some transformations in order to
make the time series stationary.

Plots of the Log Returns Series
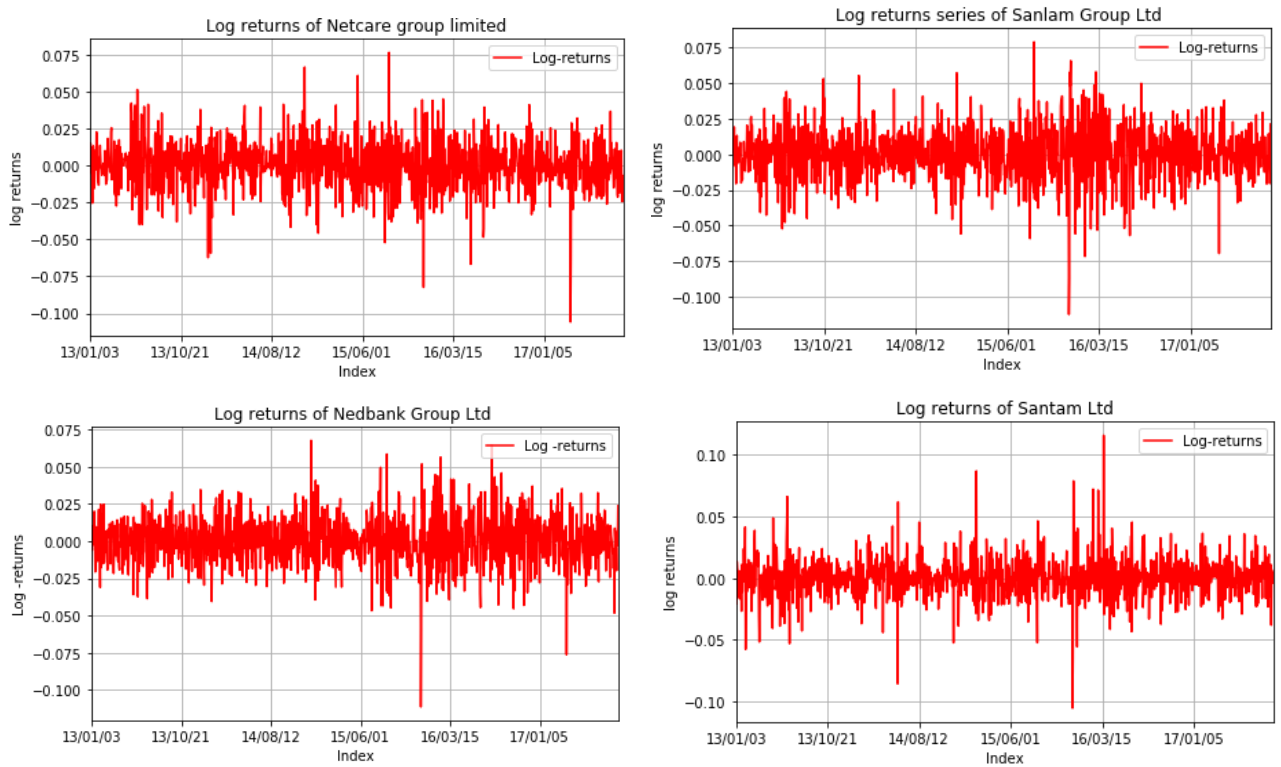
Figure 5.2: **Log returns series of JSE listed companies**

It can be seen from the time series plots in Figure 5.2 that periods of low volatility tend to be followed by period of low volatility and periods of high volatility tend to be also followed by period of high volatility. This phenomenon is common in financial markets and is known as *Volatility clustering*.

Table 5.1: **Summary statistics for the entire returns series**

	Mean	Skewness	Kurtosis
Netcare	0.000132	-0.307577	2.937779
Sanlam	0.000349	-0.339099	2.238172
Nedbank	6.78049469903e-05	-0.221069	2.642788
Santam	0.000254	0.000254	6.127645

As stated earlier in Chapter 2, the kurtosis of the Normal distribution is equal to 3. For Santam group limited returns series, the kurtosis is greater than that of a Normal distribution, as a result, we say that the return series exhibits excess kurtosis. In simple terms, this implies that the Santam returns series data has heavy tails. We conclude that the distribution of Santam Group Limited returns series is leptorkutic since its kurtosis is larger than 3. Leptorkutosis is a common phenomenon in financial time series data. For Netcare group limited, Sanlam limited, and Nedbank group limited returns series the kurtosis is close to that of a Normal distribution, so in this case the distribution of these returns series does not have heavy tails. The Normal distribution is symmetric around the mean, as a result its skewness parameter is *zero*. For Nedbank, Sanlam, and Netcare returns series the skewness parameter is negative, this implies that their distribution has longer negative tails. The skewness parameter for Santam group returns series is positive and almost close to zero, so the skewness value is close to that of a Normal distribution.

5.1   ARMA(1,1) Model Results

Train:500 Test:200

Preliminary tests for the ARMA(1,1) model

Table 5.2: **Testing for Stationarity of returns series using Augmented Dickey Fuller (ADF)**

Stock	Test statistic	P-value
Netcare	-9.050	0.000
Sanlam	-9.272	0.000
Nedbank	-8.738	0.000
Santam	-9.272	0.000

Critical Values: $-3.44(1\%)$, $-2.86(5\%)$, $-2.57(10\%)$

For all the returns series data, the ADF test statistic is less than the critical value at 1%, 5%, and 10% significance levels. This implies that we reject the null hypothesis of a unit root and conclude that the returns series data for all Four JSE listed companies are stationary. The stationary assumption simply means that the time series fluctuates around a common mean. The p-values are also extremely small, suggesting that we have strong evidence against the null hypothesis of presence of a unit root.

Table 5.3: **Testing for Stationarity of returns series using Phillips Perron test**

Stock	Test statistic	P-value
Netcare	-36.361	0.000
Sanlam	-35.677	0.000
Nedbank	-36.165	0.000
Santam	-39.027	0.000

Critical Values: $-3.44(1\%)$, $-2.86(5\%)$, $-2.57(10\%)$

The Phillips Perron test is another statistical test which can be used to test for stationarity of a time series. From Table 5.3, we can see that the test statistic for all the returns series is less than the critical values at 1%, 5%, and

10% significance levels. As a result, stationary of the series has be achieved. These results confirms the findings of the ADF test shown earlier in Table 5.2. We are confident that all the returns series are stationary.

```
                       ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  699
Model:                     ARMA(1, 1)  Log Likelihood                1893.003
Method:                         mle   S.D. of innovations              0.016
Date:               Wed, 10 Jan 2018  AIC                          -3780.007
Time:                       09:35:36  BIC                          -3766.358
Sample:                            0  HQIC                         -3774.730

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.3842      0.176      2.188      0.029       0.040       0.728
ma.L1.y       -0.5170      0.162     -3.195      0.001      -0.834      -0.200
                                     Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            2.6028           +0.0000j            2.6028            0.0000
MA.1            1.9342           +0.0000j            1.9342            0.0000
------------------------------------------------------------------------------
```

Figure 5.3: **Summary results of ARMA(1,1) fit on Netcare returns series**

From the output of the ARMA(1,1) model in Figure 5.3 above, we can see that the estimated coefficients lies within the 95% confidence interval, so we are 95% confident that the true parameters are within the computed intervals. The standard errors are a measure of how precise our estimates are. From the results above, we can see that the computed standard errors are relatively small. Small standard errors are an indication that the estimated model's coefficients are close to the true values. For the ARMA(1,1) model to be considered stationary, all the roots of the model's characteristic equation should be greater than *one* in unity. From the output above, we can see that the roots of the characteristic equation are greater than 1 in modulus. Hence, the ARMA(1,1) model fit on Netcare returns series is stationary (does not have a unit root). The estimated coefficients are significantly different from zero since all the Z scores are greater than 1.96. The estimated ARMA(1,1) model fitted on Netcare returns series is given by

$$r_t = 0.3842 r_{t-1} - (-0.5170)\epsilon_{t-1} + \epsilon_t. \tag{5.1}$$

```
                             ARMA Model Results
==============================================================================
Dep. Variable:                     y   No. Observations:                  699
Model:                    ARMA(1, 1)   Log Likelihood                1860.581
Method:                          mle   S.D. of innovations              0.017
Date:               Wed, 10 Jan 2018   AIC                          -3715.161
Time:                       09:48:17   BIC                          -3701.512
Sample:                            0   HQIC                         -3709.885

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.0399      0.478      0.083      0.934      -0.898       0.978
ma.L1.y       -0.1279      0.475     -0.269      0.788      -1.059       0.803
                                    Roots
==============================================================================
                  Real           Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1           25.0686           +0.0000j           25.0686            0.0000
MA.1            7.8212           +0.0000j            7.8212            0.0000
------------------------------------------------------------------------------
```

Figure 5.4: **Summary results of ARMA(1,1) fit on Sanlam returns series**

From the output of the ARMA(1,1) model above, both the Auto Regressive and the Moving Average coefficients lies within 95% confidence interval. The values of the standard errors are close to *zero*, this is an indication that the estimated AR and MA coefficients are close to the true value. The roots of the characteristic equation are greater than 1 in modulus, so this shows that the ARMA(1,1) model is stationary. The Z scores for both the Autoregressive and Moving Average parameters are less than 1.96, so there is room for further improvement of the model since the estimated coefficients are not significantly different from zero. The estimated ARMA(1,1) model for Sanlam return series is given as

$$r_t = 0.0399r_{t-1} - (-0.1279)\epsilon_{t-1} + \epsilon_t. \tag{5.2}$$

```
                           ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  699
Model:                   ARMA(1, 1)   Log Likelihood                1916.673
Method:                         mle   S.D. of innovations              0.016
Date:               Wed, 10 Jan 2018   AIC                          -3827.346
Time:                      09:40:15   BIC                          -3813.697
Sample:                           0   HQIC                         -3822.069

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.9308      0.022     42.562      0.000       0.888       0.974
ma.L1.y       -0.9789      0.011    -85.806      0.000      -1.001      -0.957
                                  Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            1.0743           +0.0000j            1.0743            0.0000
MA.1            1.0215           +0.0000j            1.0215            0.0000
------------------------------------------------------------------------------
```

Figure 5.5: **Summary results of ARMA(1,1) fit on Santam return series**

As seen in Figure 5.5, the estimated ARMA(1,1) model's coefficients lies within the 95% confidence interval. So, we are 95% confident that the true parameters are within the computed intervals. The estimated standard errors are close to 0, so, this shows that the estimated model parameters are close to the true values. Since the roots of the characteristic equation are greater than 1 in modulus, then the ARMA(1,1) model is stationary. The Z-score for the Autoregressive (AR) parameter is greater than 1.96, that is, $\mid Z = 42.562 \mid> 1.96$. So, the estimated AR parameter is significantly different from zero. For the Moving Average (MA) parameter, the Z score is also greater than 1.96, so the estimated MA parameter is significantly different from zero. The estimated ARMA(1,1) model fitted on Santam returns series is given as

$$r_t = 0.9308 r_{t-1} - (0.9789)\epsilon_{t-1} + \epsilon_t. \tag{5.3}$$

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  699
Model:                   ARMA(1, 1)   Log Likelihood                1969.662
Method:                         mle   S.D. of innovations              0.014
Date:                Wed, 10 Jan 2018   AIC                          -3933.324
Time:                      09:51:51   BIC                          -3919.675
Sample:                           0   HQIC                         -3928.047

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.4345      0.160      2.710      0.007       0.120       0.749
ma.L1.y       -0.5993      0.143     -4.195      0.000      -0.879      -0.319
                                     Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            2.3014           +0.0000j            2.3014            0.0000
MA.1            1.6686           +0.0000j            1.6686            0.0000
------------------------------------------------------------------------------
```

Figure 5.6: **Summary results of ARMA(1,1) fit on Nedbank returns series**

Summary results of the ARMA(1,1) model fit on Nedbank returns series in Figure 5.6 shows that the coefficients of the estimated model lies within 95% confidence interval. Since the estimated standard errors are close to 0, we can conclude that the estimated model parameters are close to the true parameter values. The ARMA(1,1) model fitted on Nedbank returns series is stationary since all the roots of the characteristic equations are greater than 1 in Modulus. Since all the Z scores are greater than 1.96, then this implies that both the estimated Autoregressive and the Moving average coefficients are significantly different from zero. The estimated ARMA(1,1) model fitted on the Nedbank returns series is given by

$$r_t = 0.4345r_{t-1} - (-0.5993)\epsilon_{t-1} + \epsilon_t. \tag{5.4}$$

Assessing the Adequacy of the Fitted ARMA(1,1) Models

The Autocorrelation Function (ACF) plots of the residuals from the fitted ARMA(1,1) models in Figure 5.7,5.8,5.9 shows that the spikes do not exceeds the significance bounds from the first lag to the tenth lag. The ACF of the residuals also do not show any pattern. This implies that the residuals are independent. Hence, the ACF of the residuals confirms that we did not misspecified the model. The ACF plot in Figure 5.10 shows one spike

around the 9^{nth} lag which slightly exceeds the significance bound, but overall we can see that the residuals appears to be independent.
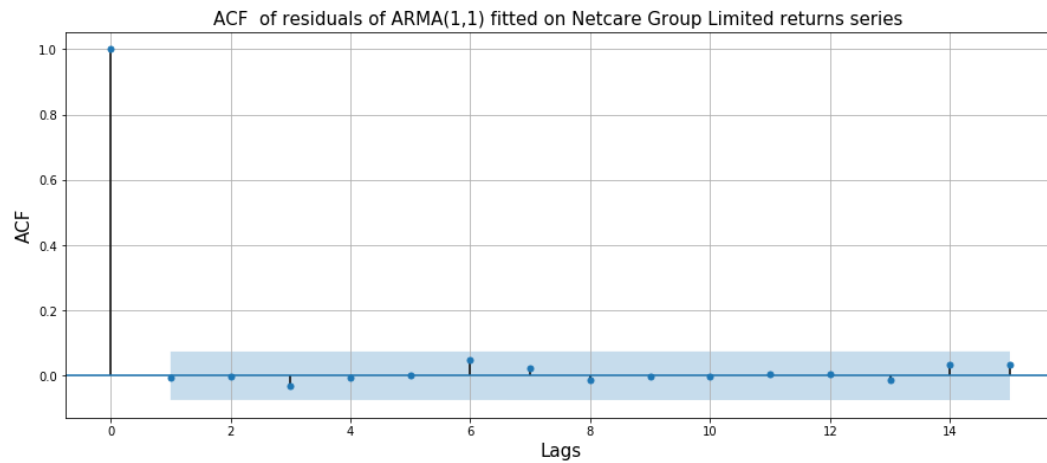
Figure 5.7: **ACF plot of residuals of ARMA(1,1) model fitted on Netcare returns series**
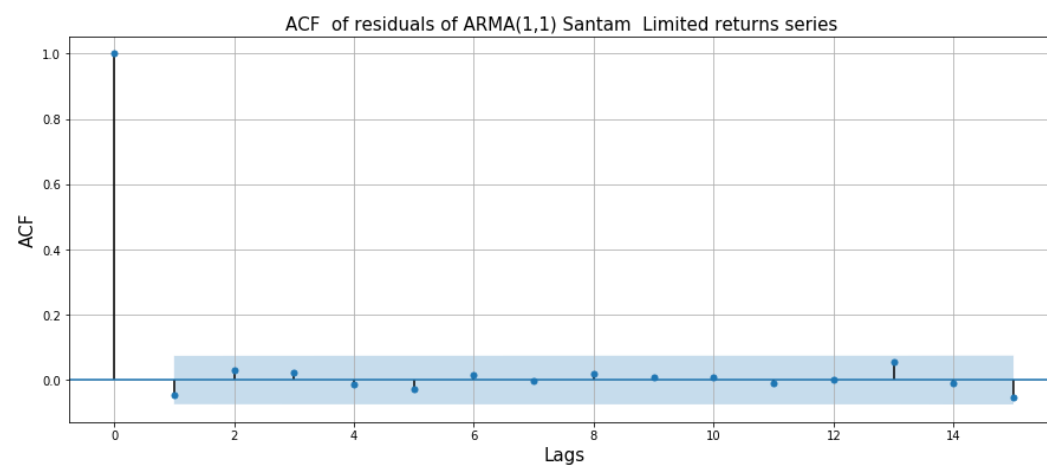
Figure 5.8: **ACF plot of residuals of ARMA(1,1) model fitted on Santam returns series**
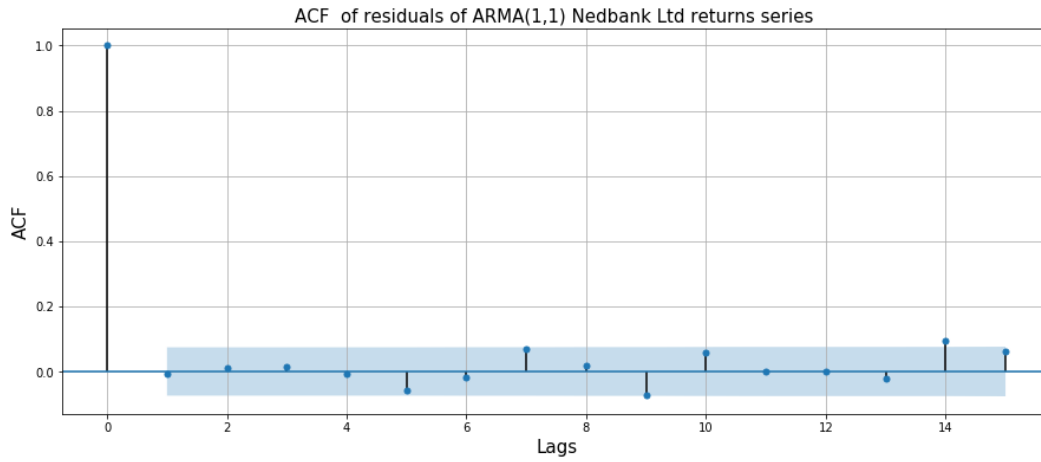
Figure 5.9: **ACF plot of residuals of ARMA(1,1) model fitted on Nedbank returns series**
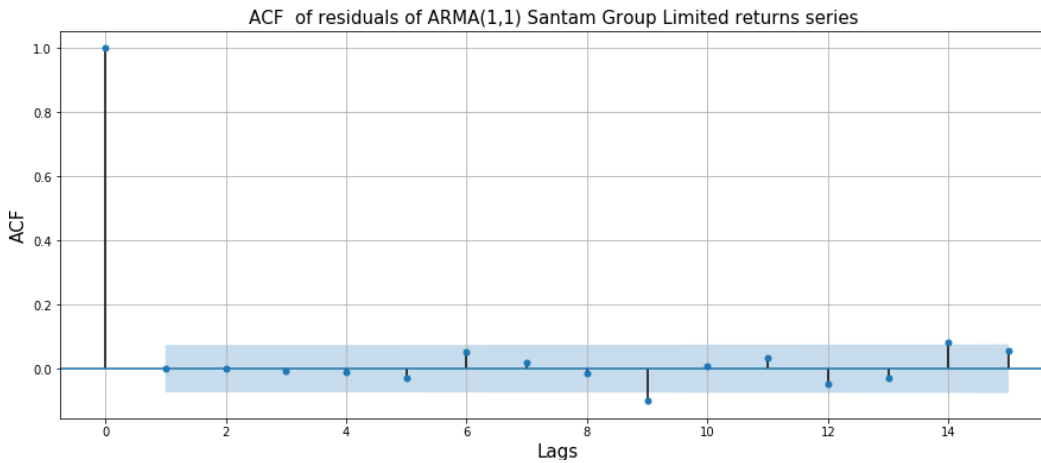
Figure 5.10: **ACF plot of residuals of ARMA(1,1) model fitted on Sanlam returns series**

Table 5.4: **Independence test of residuals of ARMA(1,1) model fitted on Netcare series**

Lag	Ljung statistic	P-value
0	0.0179	0.8935
5	2.4368	0.8755
10	2.9631	0.9912
15	5.4338	0.9931
20	6.0147	0.9994

From Table 5.4, the results of the Ljung Box test shows that the residuals from the fitted model are independent at 5% significance level since all p-values are greater than the 0.05 level of significance. Hence, we conclude that the residuals of the ARMA(1,1) model meets the independence assumption.

Table 5.5: **Independence test of residuals of ARMA(1,1) model fitted on Sanlam returns series**

Lag	Ljung statistic	P-value
0	0.0011	0.9733
5	2.6204	0.8548
10	11.0985	0.4350
15	20.6624	0.1919
20	26.8007	0.1775

All the p-values are greater than the 0.05 significance level, so we fail to reject the null hypothesis of independence. Hence, the residuals from the ARMA(1,1) model fitted on Sanlam return series are independent.

Table 5.6: **Independence test of residuals of ARMA(1,1) model fitted on Santam returns series**

Lag	Ljung statistic	P-value
0	0.15208	0.2175
5	3.4569	0.7497
10	3.9195	0.9722
15	8.4999	0.9326
20	11.6163	0.9494

From Table 5.6, it is clear that the residuals of the model are independent since all the p-values are greater than the 0.05 significance level. Large p-values are an indication of strong evidence against the alternative hypothesis.

Table 5.7: **Independence test of residuals of ARMA(1,1) model fitted on Nedbank returns series**

Lag	Ljung statistic	P-value
0	0.0267	0.8703
5	2.7699	0.8371
10	12.5540	0.3235
15	25.6425	0.0592
20	38.3579	0.0117

The p-values for the Ljung Box test statistic for lags 1 through 19 are well above the 0.05 significance level. This is an indication of independence of the model's residuals. However, the p-value for Ljung Box test at lag 20 is less than the 0.05 significance level, indicating some degree of non-independence of residuals of the fitted ARMA(1,1) model.

Table 5.8: **Testing for autocorrelation of residuals from the fitted ARMA(1,1) model**

Stock	Durbin Watson test statistic
Netcare	1.9982
Santam	2.0882
Sanlam	1.9988
Nedbank	2.0105

As seen in Table 5.8, all the DW test statistic values are close to 2. This is an indication that the the residuals from the fitted ARMA(1,1) models do not have any significant autocorrelation.

Histogram of residuals of ARMA(1,1) fit on Netcare return series
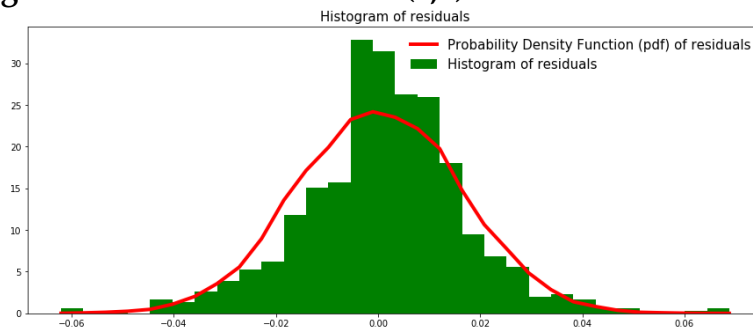
Histogram of residuals of ARMA(1,1) fit on Santam return series
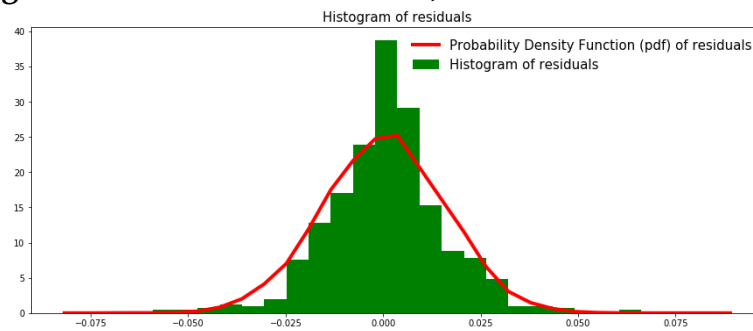
Histogram of residuals of ARMA(1,1) fit on Sanlam return series
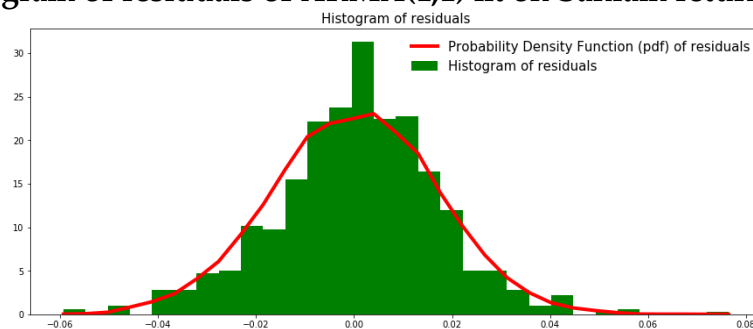
Histogram of residuals of ARMA(1,1) fit on Nedbank return series
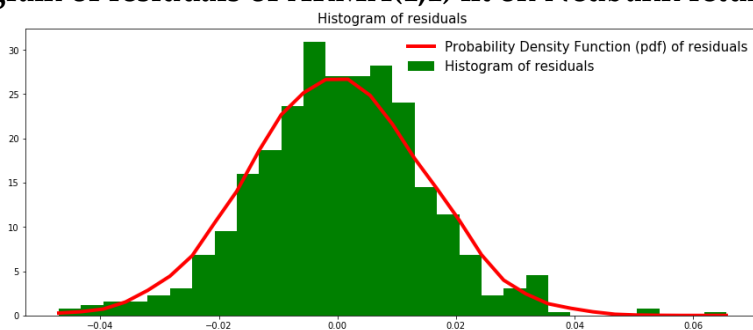
Figure 5.11: Histograms of residuals from the ARMA(1,1) models

Figure 5.11 shows the histogram of residuals (green) from the fitted ARMA(1,1) models and probability density function (pdf) whose mean is *zero* and a standard deviation of residuals. As seen in Figure 5.11, the residuals of ARMA(1,1) model fit on Netcare returns series has a few outliers in both the left and right side of the centre of the distribution. But, the distribution of the residuals seems to be symmetric. The residuals of ARMA fit on Santam series are centered around the mean suggesting some bell shaped characteristic. For ARMA(1,1) fit on Nedbank returns series, residuals are clustered around the mean and are slightly skewed to the right of the distribution due to the presence of some outliers.

Results for ARMA(1,1) Model's Predictions

The graphs in Figure 5.12 and 5.13 shows the predicted returns series using ARMA(1,1) model. The Confusion matrix in Table 5.9,5.10,5.11, and 5.12 shows the accuracy of the fitted ARMA(1,1) model in predicting whether the next day's stock return price will go down or up.
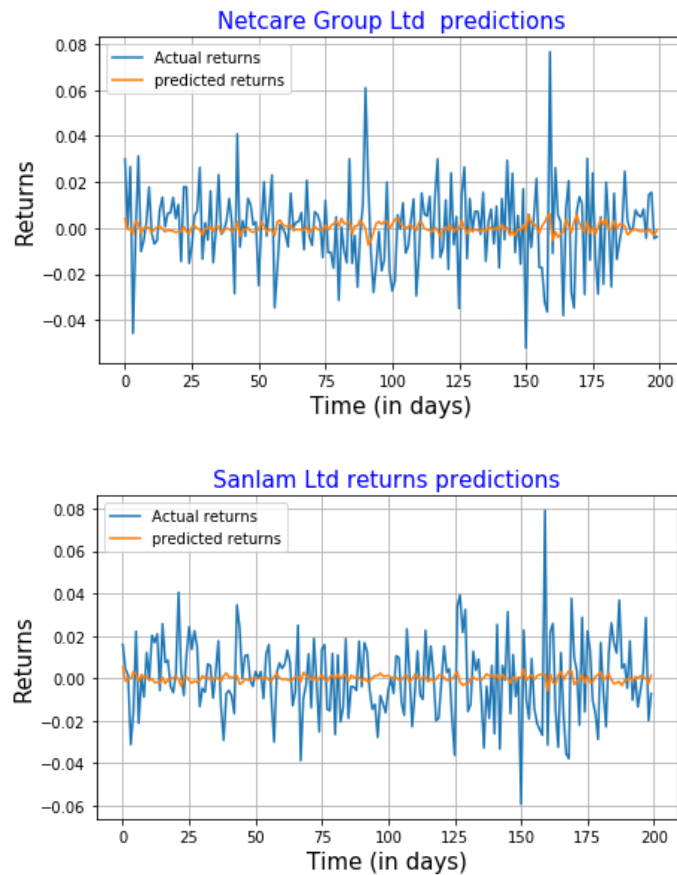
Figure 5.12: **Daily returns series predictions using ARMA(1,1) model**

From Figure 5.12 above, blue indicates the actual returns and yellow indicates the predicted returns. In Figure 5.12, the top panel shows that the ARMA(1,1) model was able to capture the volatility clusters well as compared to the bottom panel.

	Predicted Down	**Predicted Up**
True Down	75	19
True Up	27	79

Accuracy= 77%

Table 5.9: **Confusion Matrix for ARMA(1,1) predictions of Netcare returns series**

	Predicted Down	Predicted Up
True Down	68	39
True Up	25	68

Accuracy= 68%

Table 5.10: **Confusion Matrix for ARMA(1,1) predictions of Sanlam Ltd returns series**

The confusion matrix in Table 5.9 shows that the ARMA(1,1) model achieved a relatively "good" accuracy of 77% in terms of predicting the direction of movement of Netcare Ltd returns series. On the other-hand, the model achieved an accuracy of 68% for the Sanlam Ltd returns series as seen in Table 5.10.
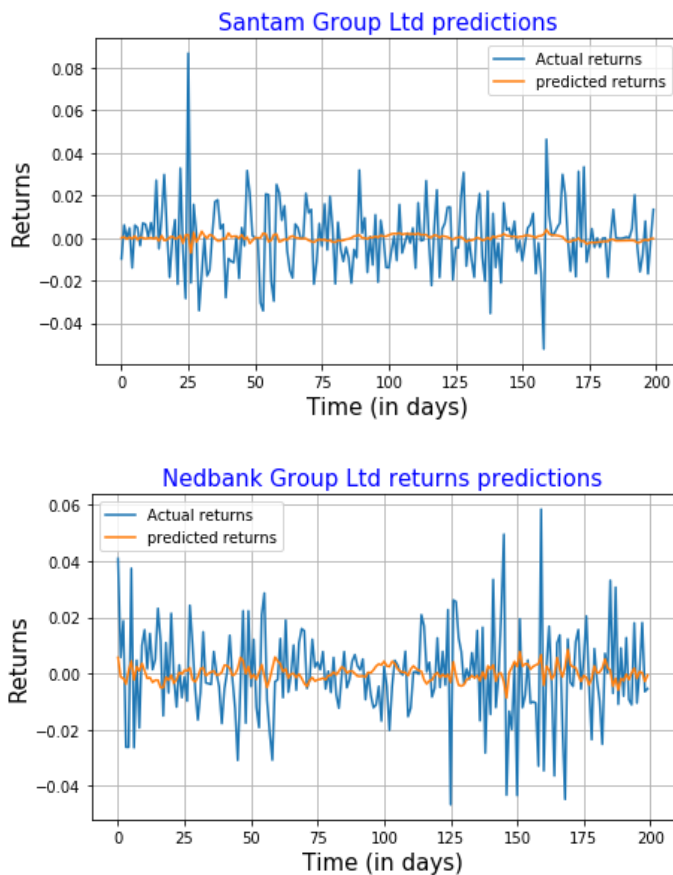
Figure 5.13: **Daily returns series predictions using ARMA(1,1) model**

In Figure 5.13, the top panel shows that the ARMA(1,1) model did not adequately capture the volatility clusters, but according to (Hansson, 2017, p.23), better capturing variance process of the series does not necessarily imply that we will have better forecasts. The model seems to be operating around the mean of the series. By visual investigation, there seems to be high variance in the predictions of Nedbank Group series as seen in the bottom panel in Figure 5.13. The ARMA(1,1) model has successfully captured the variance at the end of the series .

	Predicted Down	**Predicted Up**
True Down	60	41
True Up	29	70

Accuracy= 65%

Table 5.11: **Confusion Matrix for ARMA(1,1) predictions of Santam Ltd returns series**

	Predicted Down	**Predicted Up**
True Down	73	28
True Up	25	74

Accuracy=73.5%

Table 5.12: **Confusion Matrix for ARMA(1,1) predictions of Nedbank Ltd returns series**

From the confusion matrix in Table 5.11 and Table 5.12, the ARMA(1,1) model has achieved an accuracy of 65% and 73.5% for the Santam Ltd and the Nedbank Group Ltd returns series respectively.

Table 5.13: **Summary results of predictions of the ARMA(1,1) model**

JSE listed company	Accuracy
Netcare Group Ltd	77%
Sanlam Group Ltd	68%
Santam Ltd	65%
Nedbank Group Ltd	73.5%

From Table 5.13, we can see that the highest accuracy achieved by the ARMA(1,1) model is 77% for the Netcare Group Ltd returns series and the lowest accuracy is 65% for the Santam Ltd returns series.

Discussion

Using steps suggested by (Box and Jenkins, 1976), we constructed the ARMA(1,1) model and then used the estimated model to make predictions. The confusion matrix shows that the ARMA(1,1) model has a good ability to predict whether the next day's return price will up or down.

5.2   AR(3)-AVGARCH(1,1)-Student-t Model Results

Train:500 Test:200

Preliminary Statistical Tests on the Returns Series

Table 5.14: **Test for normality of returns using Shapiro Wilk test**

Stock	Shapiro Wilk test statistic	Kurtosis	Skew-ness
Netcare	0.9806 (0.000003)	1.5979	-0.07229
Santam	0.9506 (0.00000)	3.7361	-0.3081
Sanlam	0.9833 (0.00002)	1.1667	-0.1727
Nedbank	0.9922 (0.01055)	0.7639	0.1072

Notes: p-values of the test are in parentheses

Table 5.14 shows values of the Shapiro Wilk test statistic computed from returns of JSE listed companies, with p-values in parentheses. There is clear evidence of non-normality of all the returns series since the p-values are all less than the 0.05 level of significance. The kurtosis of a Gaussian distribution is equal to 3, however from the results above it is clear that all the returns series have a kurtosis value which is either less than or greater than that of a Gaussian distribution. A Gaussian distribution has a skewness value of 0, for the Netcare, Santam, and Sanlam returns series the skewness is negative, while for Nedbank returns series the skewness is positive suggesting that there is more weight in the right tail of the constructed distribution. In all cases the null hypothesis of normality of returns series is rejected.

Table 5.15: **Testing for ARCH effects in the returns using Lagrange Multiplier test**

Stock	Lagrange Multiplier test statistic	P-value
Netcare	36.0644	0.0069
Santam	50.1404	0.00007
Sanlam	38.3782	0.0034
Nedbank	17.6544	0.4786

Table 5.15 shows values of the Lagrange Multiplier (LM) test for ARCH effects in the returns series. The null hypothesis of no ARCH effects is rejected

for Netcare, Santam, and Sanlam return series since all the p-values are less than 0.05 level of significance. However, the p-value for LM test on Nedbank return series is greater than 0.05 level of significance, so we fail to reject the null hypothesis of no ARCH effects.

Netcare Limited Group: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                         AR - AVGARCH Model Results
===============================================================================
Dep. Variable:                       Close   R-squared:
Mean Model:                             AR   Adj. R-squared:
Vol Model:                         AVGARCH   Log-Likelihood:
Distribution:     Standardized Student's t   AIC:                            -
Method:                Maximum Likelihood    BIC:                            -
                                             No. Observations:
Date:                    Fri, Jan 12 2018    Df Residuals:
Time:                            14:13:52    Df Model:
                                 Mean Model
===============================================================================
               coef     std err          t      P>|t|        95.0% Conf. Int.
-------------------------------------------------------------------------------
Const      1.6853e-03  5.657e-04      2.979  2.892e-03 [5.765e-04,2.794e-03]
Close[1]      -0.1109  4.697e-02     -2.361  1.824e-02 [ -0.203,-1.882e-02]
Close[2]      -0.0532  4.860e-02     -1.094      0.274 [ -0.148,4.206e-02]
Close[3]      -0.0465  4.635e-02     -1.002      0.316 [ -0.137,4.439e-02]
                              Volatility Model
===============================================================================
               coef     std err          t      P>|t|        95.0% Conf. Int.
-------------------------------------------------------------------------------
omega      1.3133e-03  6.991e-04      1.879  6.029e-02 [-5.684e-05,2.683e-03]
alpha[1]      0.1761  5.334e-02      3.302  9.593e-04    [7.159e-02,  0.281]
beta[1]       0.7890  7.029e-02     11.224  3.093e-29    [  0.651,  0.927]
                                Distribution
===============================================================================
               coef     std err          t      P>|t|   95.0% Conf. Int.
-------------------------------------------------------------------------------
nu            5.1141      1.202      4.256  2.082e-05 [  2.759,  7.469]
===============================================================================
```

Figure 5.14: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fit on Netcare returns series**

Figure 5.14 shows summary output of AR(3)-AVGARCH(1,1)-Student-t model. The standard errors of the mean and volatility models are extremely small, this indicates that the estimated parameters of the AR(3)-AVGARCH(1,1)-Student-t model are close to the true parameters. The degree of freedom parameter $\nu(nu) = 5.11$, hence confirming non-normality of returns. The volatility model's parameter beta (β) is equal to 0.7890, this value is relatively large indicating that the impact of last period's variance to the cur-

rent conditional variance is large. The last two parameters of the mean model are statistically insignificant since their t-statistics are less than two. So, they do not provide significant information to the model. The AR(3)-AVGARCH(1,1)-Student-t model estimated for the Netcare Group Ltd returns series is of the form

$$r_t = 0.0016853 + (-0.1109)r_{t-1} + (-0.0532)r_{t-2} + (-0.0465)r_{t-3} + \epsilon_t,$$
$$(5.5)$$

where the error term $\epsilon_t = \sigma_t z_t$, such that z_t is a Student-t process with ($v \approx 5$) degree of freedom and the conditional variance

$$\sigma_t^2 = 0.0013133 + 0.1761 \mid \epsilon_{t-1}^2 \mid +0.7890\sigma_{t-1}^2, \qquad (5.6)$$

where ϵ_t represents the error term in the regression model.

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.16: **Diagnostics on standardised residuals of AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
20.2833 (0.3171)	1.9335	18.1114 (0.6419)

Notes: p-values of the tests are in parentheses

The Durbin Watson test statistic value is close to 2, this is an indication of non-autocorrelation of the standardised residuals. On the other hand, the p-values for the ARCH effects and Ljung box independence tests are all greater than 0.05 level of significance, suggesting that the standardised residuals have no ARCH effects and they are independent. These results suggests that the AR(3)-AVGARCH(1,1)-Student-t model is adequate for forecasting Netcare returns series.
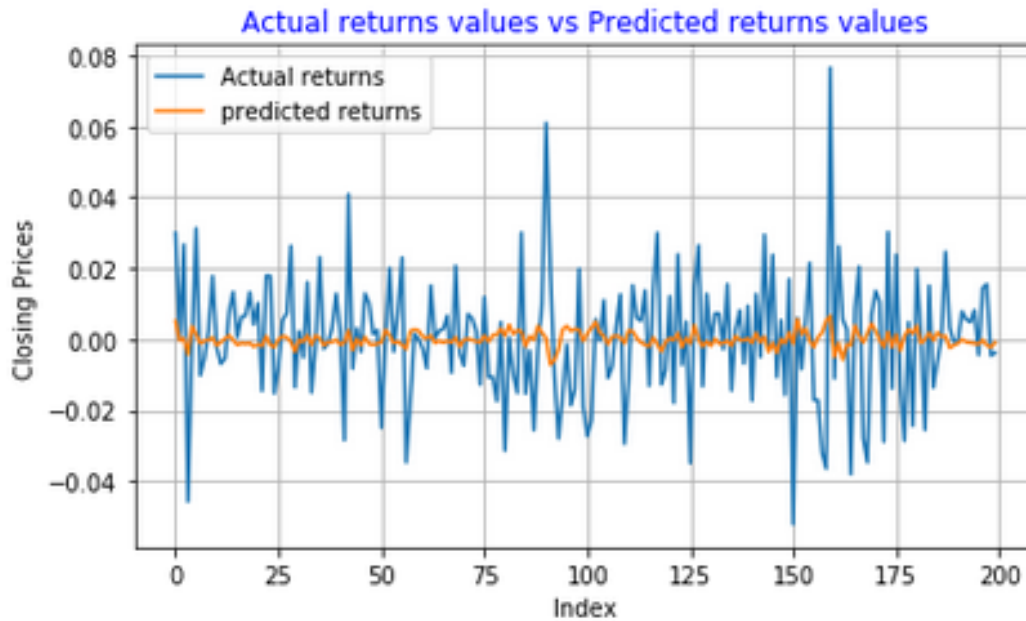
Figure 5.15: **Netcare Group Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	Predicted Up
True Down	76	18
True Up	26	80

Accuracy= 78%

Table 5.17: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Netcare Group Ltd returns series**

Figure 5.15 shows the actual returns (blue lines) and predicted returns (yellow lines) obtained using a AR(3)-AVGARCH(1,1)-Student-t model. The confusion matrix presented in Table 5.17 shows that the AR(3)-AVGARCH(1,1)-Student-t model has a good ability to predict the direction of stock returns since the model achieved a relatively high accuracy of 78%.

Santam Limited: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                       AR - AVGARCH Model Results
==============================================================================
Dep. Variable:                   Close   R-squared:                     0.001
Mean Model:                         AR   Adj. R-squared:               -0.005
Vol Model:                     AVGARCH   Log-Likelihood:              1417.15
Distribution:    Standardized Student's t   AIC:                     -2818.29
Method:             Maximum Likelihood   BIC:                        -2784.63
                                         No. Observations:                497
Date:                  Fri, Jan 12 2018  Df Residuals:                    489
Time:                          14:25:09  Df Model:                          8
                                Mean Model
==============================================================================
                coef    std err          t      P>|t|      95.0% Conf. Int.
------------------------------------------------------------------------------
Const       1.2019e-04  1.596e-04      0.753      0.452  [-1.927e-04,4.331e-04]
Close[1]       -0.0853  5.730e-03    -14.891  3.750e-50  [-9.655e-02,-7.409e-02]
Close[2]    -4.1582e-03 3.981e-03     -1.045      0.296  [-1.196e-02,3.644e-03]
Close[3]        0.0189  5.363e-04     35.240 4.855e-272   [1.785e-02,1.995e-02]
                             Volatility Model
==============================================================================
                coef    std err          t      P>|t|      95.0% Conf. Int.
------------------------------------------------------------------------------
omega       2.0311e-03  1.654e-03      1.228      0.219 [-1.211e-03,5.273e-03]
alpha[1]        0.1707  8.296e-02      2.058  3.960e-02   [8.121e-03,  0.333]
beta[1]         0.7639      0.144      5.298  1.171e-07   [  0.481,  1.047]
                              Distribution
==============================================================================
                coef    std err          t      P>|t|   95.0% Conf. Int.
------------------------------------------------------------------------------
nu              3.3534      0.544      6.162  7.172e-10 [  2.287,  4.420]
```

Figure 5.16: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fit on Santam group returns series**

Figure 5.16 shows the estimated AR(3)-AVGARCH(1,1)-Student-t model's parameters together with their corresponding standard errors. The standard errors are a measure of how precise our estimated parameters are. The standard errors for the parameters of the mean equation (AR(3) model) and the volatility (AVGARCH(1,1)-Student-t) model are extremely small, this suggests that the estimated parameters are close to the true values. The degree of freedom parameter is relatively small (about 3), this confirms non-normality of the returns series. The constant parameter and the first parameter of the mean equation (AR model) are statistically insignificant since their t-values are less than 2, suggesting that they do not provide significant information to the model.

The estimated AR(3)-AVGARCH(1,1)-Student-t model for Santam Group

Ltd returns series is of the form

$$r_t = 0.00012019 + (-0.0853)r_{t-1} + (-0.0041582)r_{t-2} + 0.0189r_{t-3} + \epsilon_t,$$
$$(5.7)$$

where $\epsilon_t = \sigma_t z_t$, and z_t is a Student-t distributed process with $v \approx 3$ (degrees of freedom). The conditional variance process is given as

$$\sigma_t^2 = 0.0020311 + 0.1707 \mid \epsilon_{t-1}^2 \mid +0.7639\sigma_{t-1}^2.$$
$$(5.8)$$

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.18:   **Diagnostics on standardised residuals of AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
20.2662 (0.3180)	2.0372	13.0417 (0.9072)

Notes: p-values of the tests are in parentheses

Based on the p-value of the Ljung Box test statistic at lag 20, one can not reject the null hypothesis of independence of the residuals. Furthermore, the p-value for Lagrange multiplier test is greater than 0.05 level of significance, this suggests that there are no remaining ARCH effects in the residuals series. If the residuals are uncorrelated, the Durbin Watson test statistic value should be close to 2, in our case, it is 2.0372, therefore we have strong evidence that the residuals from AR(3)-AVGARCH(1,1)-Student-t fit on Santam return series are uncorrelated. Since none of the statistical tests is significant, this suggests that the AR(3)-AVGARCH(1,1)-Student-t fits the Santam returns series "well".
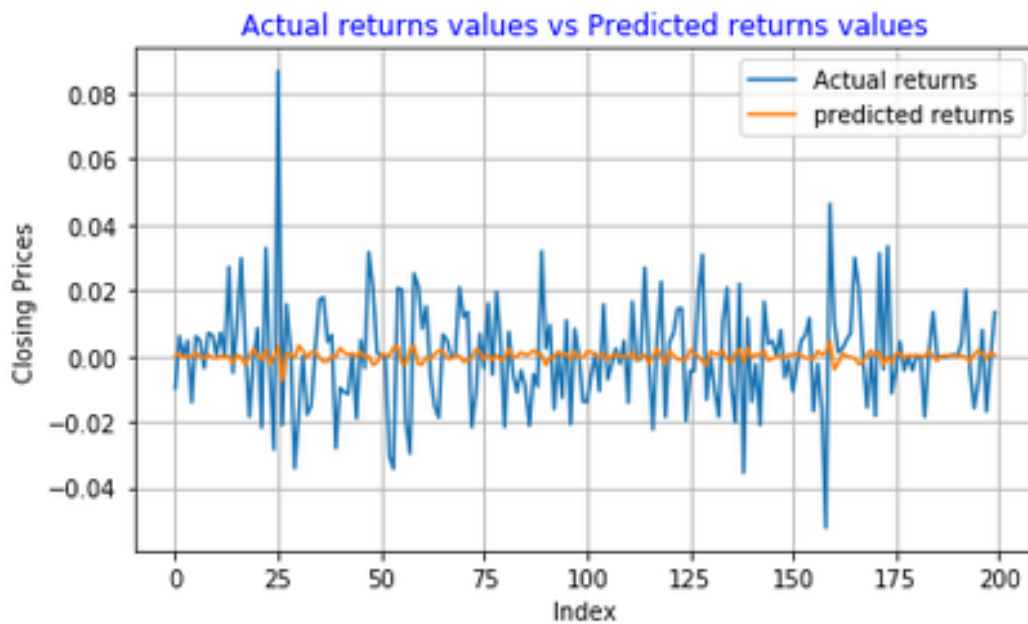
Figure 5.17: **Santam Group returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	Predicted Up
True Down	74	27
True Up	30	69

Accuracy= 71.5%

Table 5.19: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Santam Group returns series**

Predictions obtained using AR(3)-AVGARCH(1,1)-Student-t model are shown in Figure 5.17. Although the model did not capture periods of high variance well, this does not necessarily lead to poor forecasts (Hansson, 2017, p.g 23). As seen in the confusion matrix presented in Table 5.19, the model achieved a good accuracy (about 71.5%).

Sanlam Group Limited: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                         AR - AVGARCH Model Results
===============================================================================
Dep. Variable:                        Close   R-squared:
Mean Model:                              AR    Adj. R-squared:
Vol Model:                           AVGARCH   Log-Likelihood:
Distribution:      Standardized Student's t    AIC:                          -
Method:                  Maximum Likelihood     BIC:                          -
                                                No. Observations:
Date:                      Fri, Jan 12 2018    Df Residuals:
Time:                              14:30:13    Df Model:
                                 Mean Model
===============================================================================
                 coef     std err          t      P>|t|      95.0% Conf. Int.
-------------------------------------------------------------------------------
Const       1.0480e-03   6.726e-04      1.558      0.119 [-2.704e-04,2.366e-03]
Close[1]       -0.1088   4.278e-02     -2.542  1.102e-02  [ -0.193,-2.491e-02]
Close[2]       -0.0210   5.322e-02     -0.395      0.693  [ -0.125,8.327e-02]
Close[3]       -0.0186   5.356e-02     -0.348      0.728  [ -0.124,8.635e-02]
                              Volatility Model
===============================================================================
                 coef     std err          t      P>|t|      95.0% Conf. Int.
-------------------------------------------------------------------------------
omega       8.5291e-04   6.971e-04      1.224      0.221 [-5.133e-04,2.219e-03]
alpha[1]        0.0639   3.612e-02      1.770  7.676e-02  [-6.870e-03,  0.135]
beta[1]         0.9011   6.357e-02     14.175  1.299e-45  [  0.777,  1.026]
                                Distribution
==============================================================================
                 coef     std err          t      P>|t|  95.0% Conf. Int.
-------------------------------------------------------------------------------
nu              6.3255       1.966      3.218  1.291e-03 [  2.473, 10.178]
==============================================================================
```

Figure 5.18: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fit on Sanlam group returns series**

Based on the small standard errors for the estimated parameters, both the mean and the volatility model's estimated parameters are close to the true values. The t-statistics for the second and third parameter of the mean model (AR(3)) are statistically insignificant. The degree of freedom parameter (ν) is relatively small (about 6), indicating non-normality of the returns. From the output in Figure 5.18, the estimated AR(3)-AVGARCH(1,1)-Student-t model for the Sanlam Group returns series is

$$r_t = 0.001048 + (-0.1088)r_{t-1} + (-0.0210)r_{t-2} + (-0.0186)r_{t-3} + \epsilon_t, \quad (5.9)$$

where the error term is expressed as $\epsilon_t = \sigma_t z_t$, and z_t represents a Student-t distributed process with ($\nu \approx 6$) degrees of freedom. The conditional vari-

ance is given as

$$\sigma_t^2 = 0.00085291 + 0.0639 \mid \epsilon_{t-1}^2 \mid +0.9011\sigma_{t-1}^2. \qquad (5.10)$$

Table 5.20: **Diagnostics on standardised residuals of AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
22.8606 (0.1960)	1.9574	26.5428 (0.1865)

Notes: p-values of the tests are in parentheses.

The Ljung Box test for independence of the standardised residuals at lag 20 has a p-value greater than 0.05 level of significance, suggesting that we can not reject the null hypothesis of independence of the standardised residuals. Since the p-value for the Lagrange Multiplier test is greater than the 5% level of significance, this suggests that there are no ARCH effects remaining in the standardised residuals. On the other hand, the DW test suggests that the residuals are uncorrelated since the test statistic is very close to 2. Consequently, the fitted AR(3)-AVGARCH(1,1)-Student-t model is adequate for forecasting stock returns of Sanlam Group Limited.
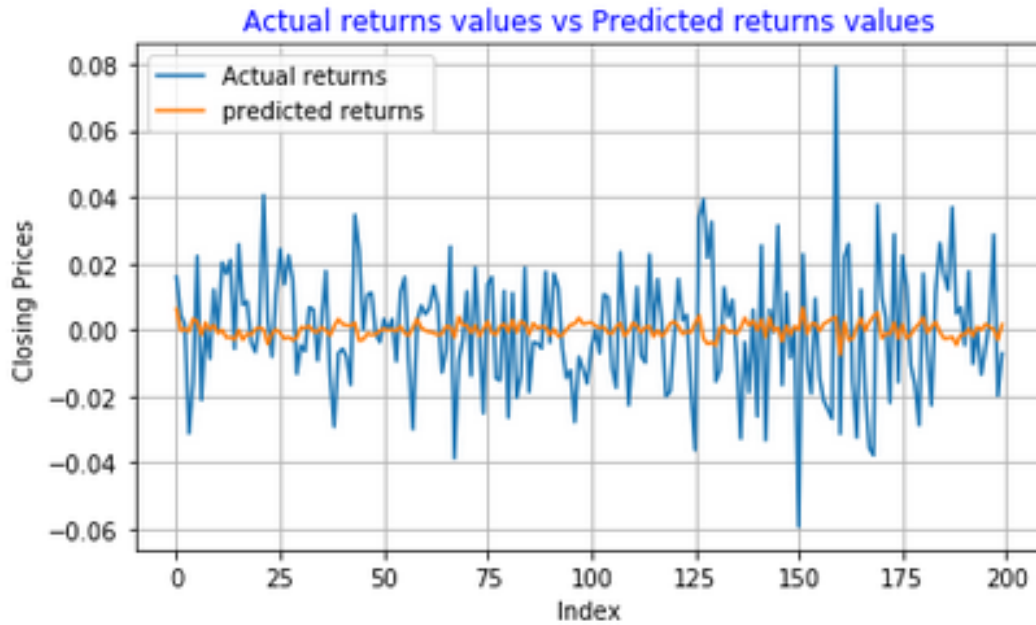
Figure 5.19: **Sanlam Group Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	Predicted Up
True Down	69	38
True Up	24	69

Accuracy= 69%

Table 5.21:   **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Sanlam Group Ltd returns series**

As seen in Figure 5.19, the variance of Sanlam Group Ltd predictions seems to be similar to that in Figure 5.15. Table 5.21 shows that the AR(3)-AVGARCH(1,1)-Student-t model has a fairly adequate ability to accurately predict whether the next day's return price will go *down* or *up*.

Nedbank Group Limited: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                          AR - AVGARCH Model Results
==================================================================================
Dep. Variable:                       Close    R-squared:
Mean Model:                             AR    Adj. R-squared:
Vol Model:                          AVGARCH    Log-Likelihood:
Distribution:      Standardized Student's t    AIC:                            -
Method:                  Maximum Likelihood    BIC:                            -
                                              No. Observations:
Date:                     Fri, Jan 12 2018    Df Residuals:
Time:                             14:34:53    Df Model:
                                Mean Model
==================================================================================
              coef     std err         t      P>|t|        95.0% Conf. Int.
----------------------------------------------------------------------------------
Const     5.4670e-04  6.262e-04      0.873      0.383  [-6.805e-04,1.774e-03]
Close[1]     -0.1281  4.595e-02     -2.788  5.309e-03  [ -0.218,-3.804e-02]
Close[2]     -0.0842  4.355e-02     -1.933  5.323e-02  [ -0.170,1.171e-03]
Close[3]      0.0119  4.603e-02      0.258      0.796  [-7.834e-02,  0.102]
                             Volatility Model
==================================================================================
              coef     std err         t      P>|t|        95.0% Conf. Int.
----------------------------------------------------------------------------------
omega     2.3470e-03  1.541e-03      1.523      0.128  [-6.739e-04,5.368e-03]
alpha[1]  3.7599e-14  3.463e-02  1.086e-12      1.000  [-6.787e-02,6.787e-02]
beta[1]      0.8315       0.103      8.044  8.686e-16  [  0.629,  1.034]
                                Distribution
==================================================================================
              coef     std err         t      P>|t|  95.0% Conf. Int.
----------------------------------------------------------------------------------
nu           15.6120      10.273      1.520      0.129  [ -4.523, 35.747]
==================================================================================
```

Figure 5.20: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t fit on Nedbank group returns series**

Figure 5.20 shows the estimated parameters of the AR(3)-AVGARCH(1,1)-Student-t model and their corresponding standard errors. All the estimated coefficients lies within the 95% confidence interval, this suggests that we are 95% confident that the true parameters lies within the computed intervals. Small standard errors are an indication that our estimated coefficients are close to the true values. Also notice that the constant parameter and the last parameter of the mean model are statistically insignificant. This implies that they can be dropped from the model as they do not provide useful information in the model. The degree of freedom parameter ν is quite high (about 15), this suggests that the distribution of the returns is closer to the normal distribution. For the Nedbank returns series, we obtain the follow-

ing AR(3)-AVGARCH(1,1)-Student-t model

$$r_t = 0.00054678 + (-0.1281)r_{t-1} + (-0.0842)r_{t-2} + 0.0119r_{t-3} + \epsilon_t. \quad (5.11)$$

The model errors $\epsilon_t = \sigma_t z_t$, such that z_t is a Student-t distributed process with $v \approx 15$ degrees of freedom. Furthermore, the conditional variance model is of the form

$$\sigma_t = 0.002347 + (3.7599e - 14) \mid \epsilon_{t-1}^2 \mid +0.8315\sigma_{t-1}^2. \quad (5.12)$$

Table 5.22: **Diagnostics on standardised residuals of AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
13.3214 (0.4346)	1.9616	25.0738 (0.2440)

Notes: p-values of the tests are in parentheses.

The p-values for the Lagrange Multiplier (at lag 20) and the Ljung Box test are 0.4346 and 0.2440 respectively, they are relatively large and are greater than the 0.05 level of significance. Hence, this suggests that the standardised residuals are independent and have no ARCH effects. The Durbin Watson test statistic is close to 2, this indicates that there are no correlations in the residuals. Consequently, the fitted AR(3)-AVGARCH(1,1)-Student-t is adequate for forecasting stock returns of Nedbank Group Limited.
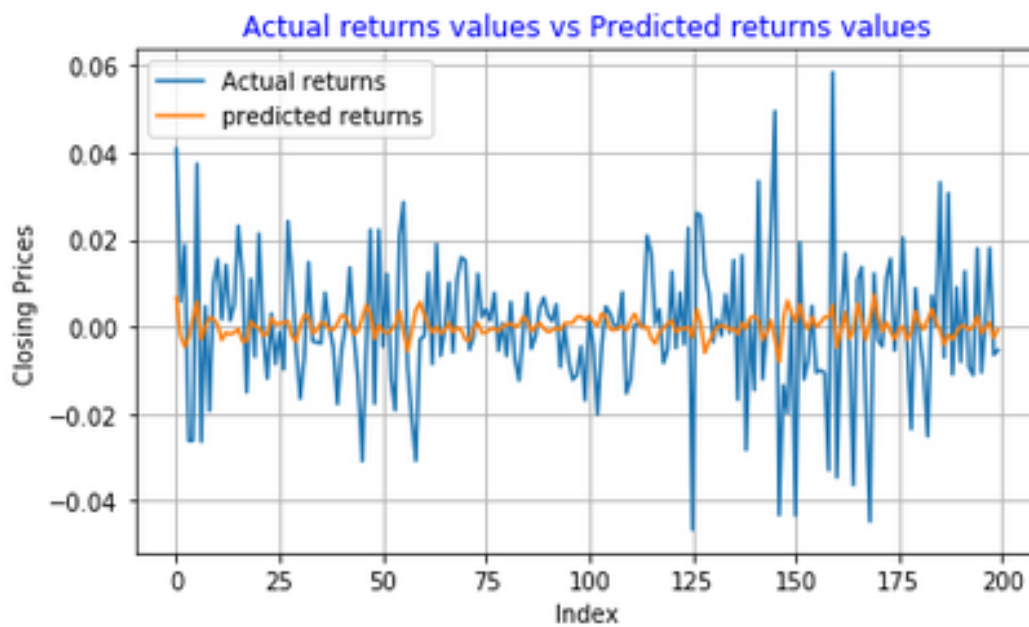
Figure 5.21: **Nedbank Group Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	**Predicted Up**
True Down	66	35
True Up	26	73

Accuracy= 69.5%

Table 5.23: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Nedbank Group Ltd returns series**

As seen in Figure 5.21, when predicting Nedbank Group returns series, the AR(3)-AVGARCH(1,1)-Student-t model seems to have captured the volatility clustering well especially at the end of the series. However, the ability to capture the variance of the series does not necessarily results in better forecast (Hansson, 2017, p.g 23). The results of the AR(3)-AVGARCH(1,1)-Student-t model's ability to predict the direction of movement of returns prices is presented in Table 5.23. The model achieved an "acceptable" accuracy of 69.5%.

Table 5.24: **Summary accuracy results of AR(3)-GARCH type models**

Prediction accuracy AR(3)-GARCH type models			
Stock	**Model**	**Distribution**	**Accuracy (%)**
Netcare	AR(3)-GARCH(1,1)	Normal	76.5%
	AR(3)-EGARCH(1,1)	Student-t	78.5%
	AR(3)-EWMA	Normal	77.5%
	AR(3)-EWMA	Student-t	77.5%
	AR(3)-TARCH(1,1,1)	Normal	74.5%
	AR(3)-TARCH(1,1,1)	Student-t	78%
	AR(3)-AVGARCH(1,1)	Normal	74.5%
	AR(3)-AVGARCH(1,1)	Student-t	78%
Santam	AR(3)-GARCH(1,1)	Normal	66%
	AR(3)-EGARCH(1,1)	Student-t	70.5%
	AR(3)-EWMA	Normal	68%
	AR(3)-EWMA	Student-t	67.5%
	AR(3)-TARCH(1,1,1)	Normal	65.5%
	AR(3)-TARCH(1,1,1)	Student-t	72%
	AR(3)-AVGARCH(1,1)	Normal	65%
	AR(3)-AVGARCH(1,1)	Student-t	71.5%
Sanlam	AR(3)-GARCH(1,1)	Normal	69%
	AR(3)-EGARCH(1,1)	Student-t	65.5%
	AR(3)-EWMA	Normal	67%
	AR(3)-EWMA	Student-t	68.5%
	AR(3)-TARCH(1,1,1)	Normal	70%
	AR(3)-TARCH(1,1,1)	Student-t	70%
	AR(3)-AVGARCH(1,1)	Normal	68.5%
	AR(3)-AVGARCH(1,1)	Student-t	69%
Nedbank	AR(3)-GARCH(1,1)	Normal	68.5%
	AR(3)-EGARCH(1,1)	Student-t	76%
	AR(3)-EWMA	Normal	69%
	AR(3)-EWMA	Student-t	70%
	AR(3)-TARCH(1,1,1)	Normal	68.5%
	AR(3)-TARCH(1,1,1)	Student-t	69%
	AR(3)-AVGARCH(1,1)	Normal	68.5%
	AR(3)-AVGARCH(1,1)	Student-t	69.5%

Table 5.24 shows summary accuracy results of AR(3)-GARCH type models in predicting whether the next day's returns price will go *up* or *down*. For the Netcare Group Ltd returns series, the AR(3)-EGARCH(1,1)-Student-t achieved the highest accuracy (about 78.5%). The AR(3)-TARCH(1,1,1)-

Student-t model achieved the highest accuracy (72%) for Santam Ltd return series. Again for Sanlam Group Ltd returns series, the AR(3)-TARCH(1,1,1) model under the Student-t and Normal distribution achieved the highest accuracy (about 70%). The AR(3)-EGARCH(1,1)-Student-t achieved the highest accuracy (about 76%) for Nedbank Group Ltd returns series.

Discussion:

AR(3)-GARCH type models were fitted on various JSE listed companies returns series, these models performed well in terms of forecasting the direction of movement of returns prices. Summary results presented in Table 5.24 shows that the models have a good ability to predict whether the next day's returns price will go *down* or *up*. The highest accuracy achieved by one of these models was about 78.5%.

5.3   ARMA(1,1)-State Space Model Results

Train:500 Test:200

Netcare Group Limited: Estimated State Space ARMA(1,1) Model

```
                    Statespace Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  500
Model:                       ARMA11   Log Likelihood                1370.640
Date:               Mon, 15 Jan 2018  AIC                          -2735.281
Time:                      16:41:41   BIC                          -2722.637
Sample:                           0   HQIC                         -2730.319
                              - 500
Covariance Type:                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.4305      0.339      1.270      0.204      -0.234       1.095
theta         -0.5184      0.321     -1.616      0.106      -1.147       0.110
sigma2         0.0002   1.15e-05     21.239      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                       38.81   Jarque-Bera (JB):            61.77
Prob(Q):                              0.52    Prob(JB):                     0.00
Heteroskedasticity (H):               0.95    Skew:                        -0.06
Prob(H) (two-sided):                  0.76    Kurtosis:                     4.72
==============================================================================
```

Figure 5.22: **Summary results of State Space ARMA(1,1) model fit on Netcare Limited Group returns series**

Figure 5.22 shows the estimated parameters of the State Space ARMA(1,1) model fit on Netcare returns series. The estimated parameters lies within the confidence interval, so we are confident that the true parameters lies within the computed intervals. Small standard errors are an indication that the estimated parameters are close to the true values. For an ARMA model, the residuals are assumed to have a constant variance. The p-value for the heteroskedasticity test (0.76) is greater than the 0.05 level of significance, so we fail to reject the null hypothesis that the residuals of the fitted State Space ARMA(1,1) model have a constant variance. The p-value for the Jarque Bera (JB) test is extremely small (about 0.00), suggesting that the residuals are not normally distributed. Furthermore, the kurtosis of residuals is around 4, this confirms the results of the JB test that the residuals are not from a normal distribution. A large p-value for the Ljung Box test (about 0.52) is an indication that the residuals of the State Space ARMA(1,1) model fit on Netcare Limited Group series are independent. Consequently, other times series models which assume non normality of residuals should be explored to better capture the statistical properties of the Netcare Limited Group returns series. The estimated model is given as

$$r_t = 0.4305 r_{t-1} + (-0.5184)\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 = 0.0002) \qquad (5.13)$$
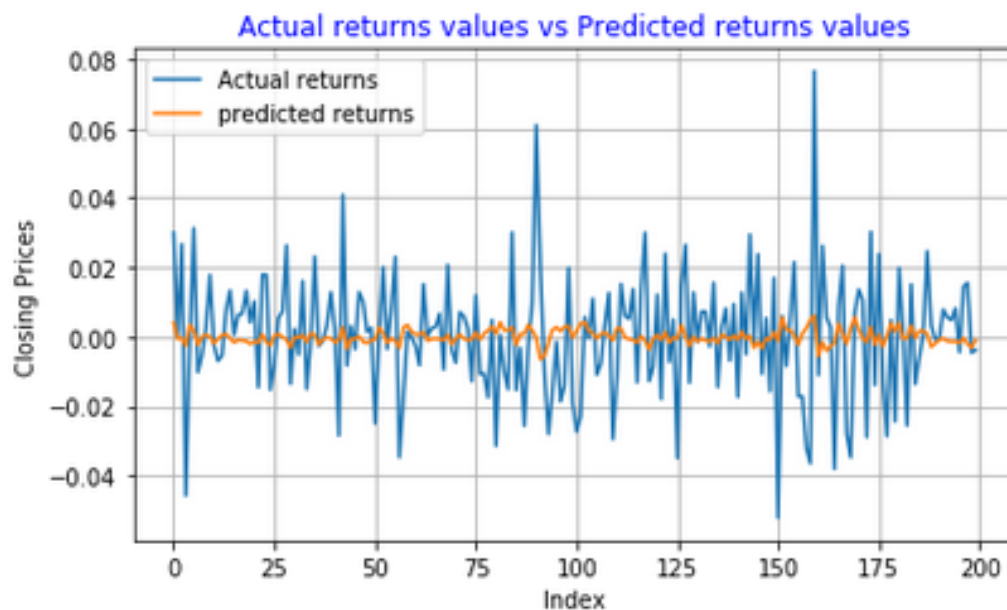
Figure 5.23: **Netcare Group Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	Predicted Up
True Down	75	19
True Up	27	79

Accuracy= 77%

Table 5.25: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Netcare returns series**

The confusion matrix in Table 5.25 shows the accuracy of the fitted State Space ARMA(1,1) model in predicting whether the next day's return price will go down or up.

Santam Limited (Ltd): Estimated State Space ARMA(1,1) Model

```
                        Statespace Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  500
Model:                         ARMA11   Log Likelihood                1374.559
Date:                Mon, 15 Jan 2018   AIC                          -2743.118
Time:                        16:43:49   BIC                          -2730.475
Sample:                             0   HQIC                         -2738.157
                              - 500
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.5524      0.644      0.857      0.391      -0.711       1.815
theta         -0.5825      0.636     -0.916      0.360      -1.829       0.664
sigma2         0.0002   9.12e-06     26.257      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                       34.80   Jarque-Bera (JB):          294.18
Prob(Q):                              0.70   Prob(JB):                    0.00
Heteroskedasticity (H):               0.71   Skew:                       -0.30
Prob(H) (two-sided):                  0.03   Kurtosis:                    6.71
==============================================================================
```

Figure 5.24: **Summary results of State Space ARMA(1,1) fit on Santam Limited returns series**

Summary results of the State Space ARMA(1,1) model fit on Santam Ltd series are shown in Figure 5.24. The estimated parameters are within their computed confidence intervals. The estimated variance parameter (sigma2) has a very small standard error, this suggests that the parameter is close to the true value. The null hypothesis that the residuals have the same variance (homoscedastic) is rejected since the Heteroskedasticity test statistic value is extremely small (around 0.71) and the p-value is less than the 0.05

level of significance. Since the p-value for the Jarque Bera test is less than
0.05 level of significance, we reject the null hypothesis that the residuals
are normally distributed. The skewness value is -0.30, this suggests that
the distribution of residuals is fairly skewed to the left. The distribution of
the residuals has heavier tails than the normal distribution since its kurto-
sis value (around 6) is greater than 3 (that of a normal distribution). Lastly,
the Ljung Box test does not reject the null hypothesis of independence of
residuals since the p-value is greater than 0.05 (level of significance). The
estimated model for Santam Limited returns series is given as

$$r_t = 0.5524 r_{t-1} + (-0.5825)\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 = 0.0002) \quad (5.14)$$
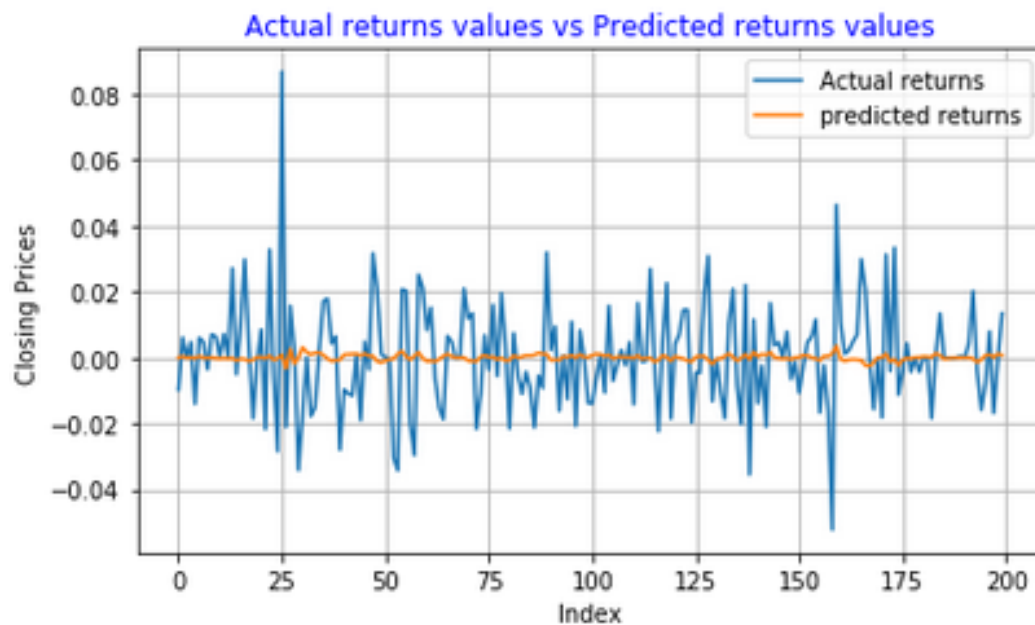
Figure 5.25: **Santam Group returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	Predicted Up
True Down	73	28
True Up	26	73
	Accuracy= 73%	

Table 5.26: **Confusion Matrix for the State Space ARMA(1,1) model's predictions of Santam Group returns series**

Using a confusion matrix, we evaluated the ability of the State Space ARMA(1,1) model to predict whether the next day's Santam Group Ltd returns price will go down or up (as seen in Table 5.26). The model achieved a satisfactory accuracy (about 73%).

Sanlam Group Ltd: Estimated State Space ARMA(1,1) Model

```
                      Statespace Model Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                  500
Model:                         ARMA11   Log Likelihood                1352.542
Date:                Mon, 15 Jan 2018   AIC                          -2699.084
Time:                        16:45:07   BIC                          -2686.441
Sample:                      01-13-2003   HQIC                        -2694.123
                           - 01-15-2005
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi           -0.0704      0.393     -0.179      0.858      -0.841       0.700
theta         -0.0240      0.393     -0.061      0.951      -0.794       0.746
sigma2         0.0003   1.32e-05     19.856      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                    48.31   Jarque-Bera (JB):            32.32
Prob(Q):                           0.17   Prob(JB):                     0.00
Heteroskedasticity (H):            0.77   Skew:                        -0.16
Prob(H) (two-sided):               0.10   Kurtosis:                     4.20
==============================================================================
```

Figure 5.26: **Summary results of the State Space ARMA(1,1) model fit on Sanlam Group Limited returns series**

Figure 5.26 shows the summary of the estimated State Space ARMA(1,1) model fitted on Sanlam Group Limited returns series. The estimated parameters are within their confidence intervals, suggesting that the true parameters lie within the computed intervals. Since the p-value for the heteroscedasticity test is fairly large (around 0.10), this indicates that we have weak evidence against the null hypothesis. Hence, the residuals of the fitted State Space ARMA(1,1) model have a constant variance. For the residuals to be considered normally distributed, the Jarque Bera test statistic value should be close to 0. In our case, we reject the null hypothesis of normality since the JB test statistic is very large (about 32). The kurtosis value is around 4, suggesting that the distribution of residuals has slightly "heavier tails" than the normal distribution. We accept the null hypothesis of independence of the residuals since the p-value for the Ljung Box test is greater than the 0.05 significance level. The estimated model for the Sanlam Group

Ltd returns series is given as

$$r_t = (-0.0704)r_{t-1} + (-0.0240)\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 = 0.0003) \quad (5.15)$$
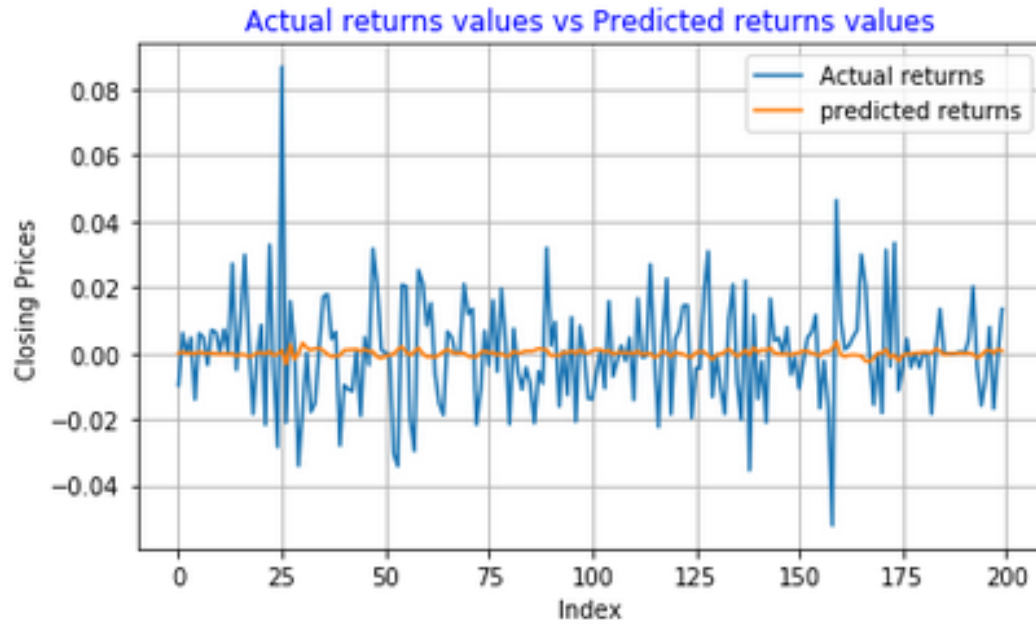
Figure 5.27: **Sanlam Group Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	70	37
True Up	24	69

Accuracy= 69.5%

Table 5.27: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Sanlam Group Ltd returns series**

Results of the confusion matrix presented in Table 5.27 shows that the State Space ARMA(1,1) achieved an accuracy of 69.5% and a misclassification rate (percentage of incorrect predictions) of about 30.5%.

Nedbank Group Ltd: Estimated State Space ARMA(1,1) model

```
                     Statespace Model Results
==============================================================================
Dep. Variable:                 Close   No. Observations:                  500
Model:                        ARMA11   Log Likelihood                1430.199
Date:                Mon, 15 Jan 2018  AIC                          -2854.398
Time:                       16:50:00   BIC                          -2841.754
Sample:                    01-13-2003  HQIC                         -2849.436
                         - 01-15-2005
Covariance Type:                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.5955      0.157      3.789      0.000       0.288       0.904
theta         -0.7221      0.132     -5.489      0.000      -0.980      -0.464
sigma2         0.0002   1.04e-05     18.458      0.000       0.000       0.000
===================================================================================
Ljung-Box (Q):                     37.40   Jarque-Bera (JB):               13.29
Prob(Q):                            0.59   Prob(JB):                        0.00
Heteroskedasticity (H):             1.20   Skew:                            0.11
Prob(H) (two-sided):                0.23   Kurtosis:                        3.77
===================================================================================
```

Figure 5.28: **Summary results of State Space ARMA(1,1) fit on Nedbank Group Ltd returns series**

Summary results of the estimated State Space ARMA(1,1) model fit on Nedbank Group Ltd returns series are presented in Figure 5.28. The estimated Autoregressive parameter phi (ϕ) and Moving Average parameter theta (θ) are all within the confidence intervals, suggesting that the true parameters lie within the computed intervals. The heteroscedasticity test statistic value is 1.20 (this value is extremely small), suggesting that we can not reject the null hypothesis that the residuals have the same variance. The Jarque Bera test statistic value is large (about 13.29), this suggests that the null hypothesis of normality of residuals can not be accepted. The kurtosis for a normal distribution is equal to 3, in our case the kurtosis of the residuals is around $3.77 \approx 4$, this implies that the distribution of the residuals has "slightly" heavier tails than the normal distribution. The distribution of the residuals is "slightly" right skewed since the skewness value is positive (about 0.11). A more appropriate time series model which assumes non normality of returns should be considered for modelling stock returns of Nedbank Group Ltd returns series. The estimated State Space ARMA(1,1) model for Nedbank Group Ltd returns series is given as

$$r_t = 0.5955 r_{t-1} + (-0.7221)\epsilon_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2 = 0.0002) \qquad (5.16)$$
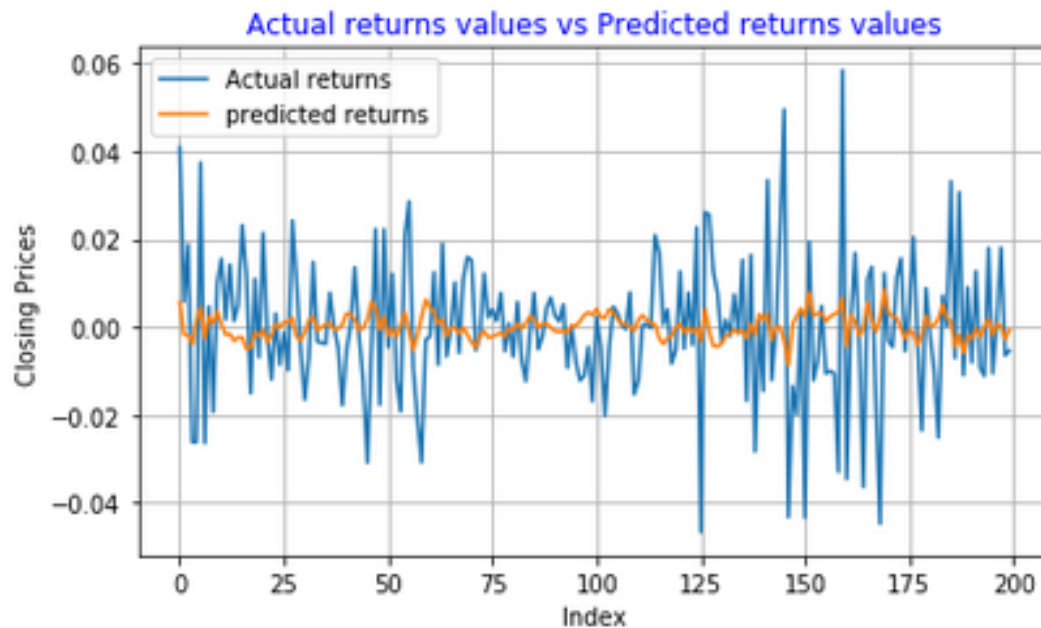
Figure 5.29: **Nedbank Group Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	74	27
True Up	25	74

Accuracy= 74%

Table 5.28: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Nedbank Group Ltd returns series**

For the Nedbank Group Ltd returns series, the State Space ARMA(1,1) model achieved a very good accuracy (74%), the high accuracy show that the model has a strong ability to predict the direction of movement of the returns series.

Discussion:

As seen in the confusion matrix presented in Table 5.25, 5.26, 5.27, and 5.28 , accuracy achieved by the State Space ARMA(1,1) model for all four time series was satisfactory. However, some of the statistical assumptions such as normality of residuals were not entirely satisfied by the fitted model, this

suggests that other advanced time series model such as the ARMA-GARCH type models can be appropriate for modelling of JSE listed companies returns series as they can assume other statistical distributions such as the Student-t, and Skewed Student-t distribution.

5.4   ARMA(1,1)-Metropolis Hastings (MH) Algorithm Results

Train:500 Test:200

We implemented the ARMA(1,1) model (with parameters estimated via Metropolis Hastings algorithm) using a powerful Python 2 package called **Pymc**. Recall that the Metropolis-Hastings ARMA(1,1) model has the following three parameters: The autoregressive parameter phi (ϕ), the moving average parameter theta (θ), and the variance parameter σ^2 (precision). Following (Fulton, 2017), we specify Uniform prior distributions for ϕ and θ, with support (-1,1). For the precision parameter $1/\sigma^2$ we assume a Gamma prior distribution, with support $\Gamma(2, 4)$.

Netcare Group Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
precision:

        Mean            SD              MC Error        95% HPD interval
        --------------------------------------------------------------
        61.869          3.819           0.092           [ 55.079  69.671]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        54.615          59.153         61.936          64.457          69.437


theta:

        Mean            SD              MC Error        95% HPD interval
        --------------------------------------------------------------
        -0.033          0.479           0.018           [-0.878  0.826]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.867          -0.417         -0.05           0.339           0.855


phi:

        Mean            SD              MC Error        95% HPD interval
        --------------------------------------------------------------
        -0.026          0.508           0.019           [-0.923  0.837]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.909          -0.44          -0.025          0.39            0.879
```

Figure 5.30: **Summary results of ARMA(1,1) Metropolis Hastings model fit on Netcare Group Limited returns series**

In Bayesian methods, the expected value (mean) is used as a best estimator of a parameter. The Monte Carlo Standard Error (MC error) indicates the amount of error in the estimates due to the use of MCMC method. As seen in Figure 5.30, for all parameters the MC error are close to 0, this is a clear indication that the estimated mean parameters are close to the true value. In Bayesian statistics, standard deviation (SD) is used to describe uncertainty

in the estimated mean parameter. Since the standard deviations (SD) for the autoregressive (phi) and moving average parameters (theta) are small, this suggests that the degree of uncertainty in the estimated mean parameters is relatively small. However, there is some degree of uncertainty in the variance parameter (precision (σ^2)) since its SD is relatively high (about 3.819). From Figure 5.30, the estimated ARMA(1,1)-Metropolis Hastings model for the Netcare Group Ltd returns series is of the form

$$r_t = (-0.026)r_{t-1} + (-0.033)\epsilon_{t-1} + \epsilon_t, \tag{5.17}$$

where $\epsilon_t \sim N(0, \sigma^2)$, the precision parameter $1/\sigma^2 = \frac{1}{61.869}$.
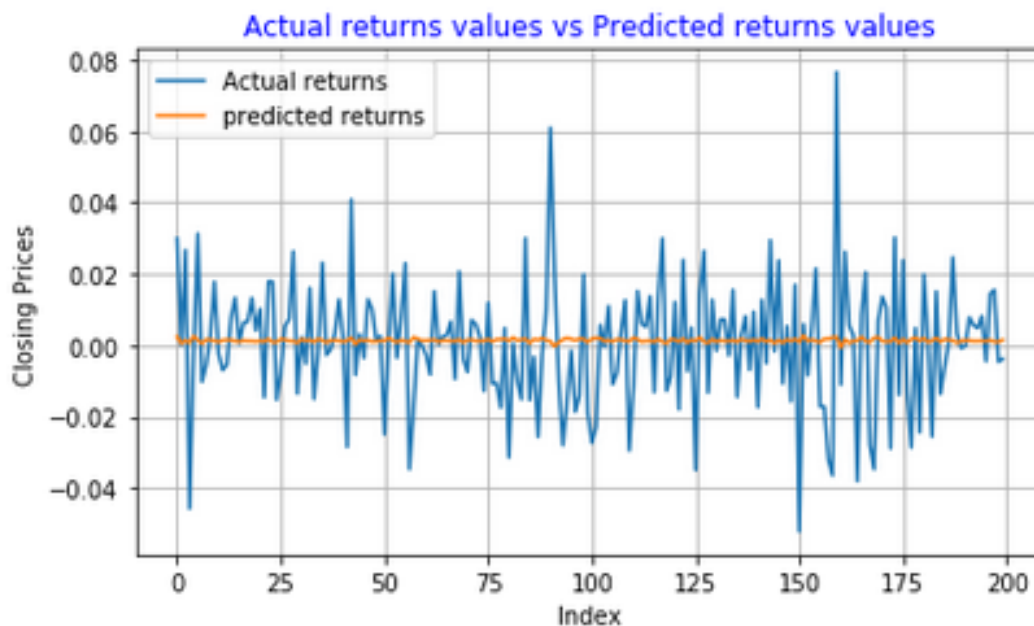
Figure 5.31: **Netcare Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	68	26
True Up	37	69

Accuracy= 68.5%

Table 5.29: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Netcare Group Ltd returns series**

Figure 5.31 shows the actual (blue line) and predicted (yellow line) returns series obtained using ARMA(1,1)-Metropolis Hastings model. By visual investigating the plot in Figure 5.31, we can clearly see that there is low or small variance in the predictions made by the ARMA(1,1)-MH model. The model seems to be operating around the expected value (mean). Confusion matrix for ARMA(1,1)-MH model's prediction results are presented in Table 5.29.

Santam Limited (Ltd): Estimated ARMA(1,1)-Metropolis Hastings model

```
phi:

        Mean            SD              MC Error         95% HPD interval
        --------------------------------------------------------------------
        -0.039          0.509           0.021            [-0.946  0.833]


        Posterior quantiles:

        2.5             25              50               75              97.5
        |---------------|===============|===============|---------------|
        -0.909          -0.457          -0.042           0.361           0.885

theta:

        Mean            SD              MC Error         95% HPD interval
        --------------------------------------------------------------------
        -0.012          0.464           0.019            [-0.896  0.785]


        Posterior quantiles:

        2.5             25              50               75              97.5
        |---------------|===============|===============|---------------|
        -0.862          -0.366          -0.008           0.339           0.832

precision:

        Mean            SD              MC Error         95% HPD interval
        --------------------------------------------------------------------
        61.953          3.713           0.092            [ 54.941  69.175]


        Posterior quantiles:

        2.5             25              50               75              97.5
        |---------------|===============|===============|---------------|
        55.092          59.351          61.801           64.473          69.491
```

Figure 5.32: **Summary results of ARMA(1,1) Metropolis Hastings model fit on Santam Ltd returns series**

The degree of uncertainty in the precision parameter is quite high since its standard deviation (SD) value is large (3.713). The Monte Carlo Standard Error (MC error) for all parameters are extremely small, this suggests that the estimated mean parameters of the model are very close to the true values. Furthermore, all the estimated mean parameters lies within the 95% Highest Posterior Density (HPD) interval, suggesting that the true estimates are

within the computed interval. The estimated ARMA(1,1)-Metropolis Hastings model fitted on Santam Ltd returns series is of the form

$$r_t = (-0.039)r_{t-1} + (-0.012)\epsilon_{t-1} + \epsilon_t, \tag{5.18}$$

where $\epsilon_t \sim N(0, \sigma^2)$. The precision parameter $1/\sigma^2 = \frac{1}{61.953}$, this value is close to that in Figure 5.30.
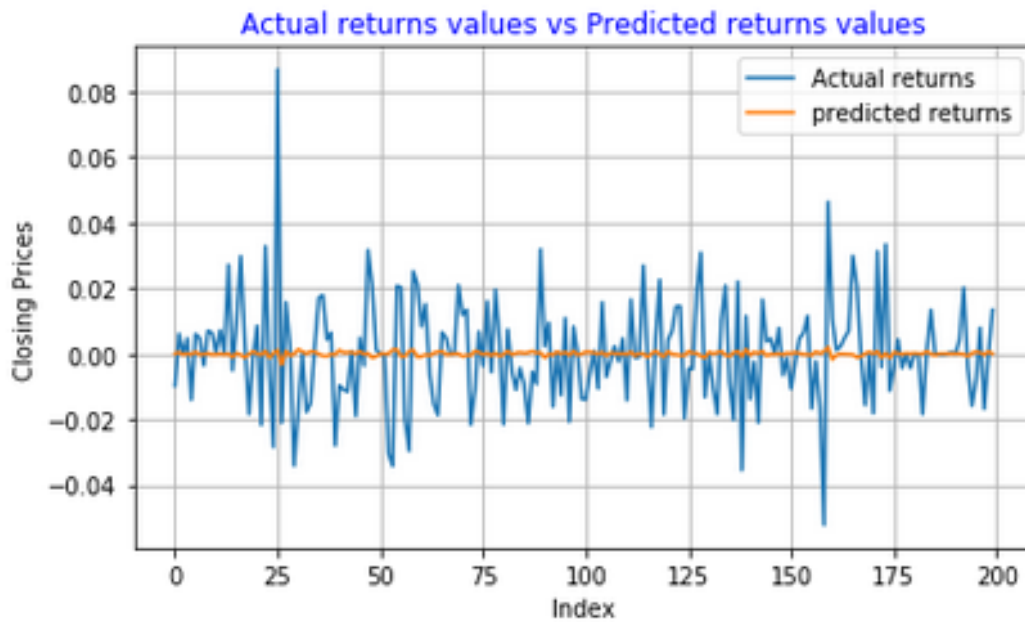
Figure 5.33: **Santam Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	Predicted Up
True Down	69	32
True Up	30	69

Accuracy= 69%

Table 5.30: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Santam Ltd returns series**

Santam Ltd returns series predictions made by the ARMA(1,1)-MH model are presented in Figure 5.33. The structure of these predictions seems to be similar to those in Figure 5.31.. The confusion matrix presented in Table 5.30 shows that the ARMA(1,1)-Metropolis Hastings model achieved an overall accuracy of 69%.

Sanlam Group Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
phi:
        Mean            SD            MC Error        95% HPD interval
        --------------------------------------------------------------
        -0.039          0.509         0.021           [-0.865  0.859]


        Posterior quantiles:

        2.5             25            50              75              97.5
        |---------------|==============|==============|---------------|
        -0.885          -0.472        -0.067          0.369           0.855


precision:
        Mean            SD            MC Error        95% HPD interval
        --------------------------------------------------------------
        61.791          3.754         0.1             [ 54.595  68.932]


        Posterior quantiles:

        2.5             25            50              75              97.5
        |---------------|==============|==============|---------------|
        54.697          59.129        61.654          64.343          69.244


theta:
        Mean            SD            MC Error        95% HPD interval
        --------------------------------------------------------------
        -0.029          0.476         0.02            [-0.866  0.789]


        Posterior quantiles:

        2.5             25            50              75              97.5
        |---------------|==============|==============|---------------|
        -0.848          -0.411        -0.057          0.35            0.825
```

Figure 5.34: **Summary results of ARMA(1,1) Metropolis Hastings model fit on Sanlam Group Ltd returns series**

The standard deviations for the autoregressive (phi) and moving average (theta) parameters are 0.509 and 0.476 respectively. The degree of uncertainty in these estimated mean parameters is insignificant since their standard deviations are very small. But for the variance parameter (precision) the uncertainty is quite high since its standard deviation has a very large value (about 3.754). Small values for the Markov Chain standard (MC) er-

rors for all the estimated mean parameters are an indication that the error in the estimates is quite low, suggesting that the estimated parameters are close to the true value. All the estimates lies within the 95% HPD intervals, this indicates that the true parameters values lies within the computed intervals. The estimated ARMA(1,1)-Metropolis Hastings model for the Sanlam Group Ltd returns series is of the form

$$r_t = (-0.039)r_{t-1} + (-0.029)\epsilon_{t-1} + \epsilon_t, \tag{5.19}$$

where $\epsilon_t \sim N(0, \sigma^2)$, and the precision parameter value is $1/\sigma^2 = \frac{1}{61.791}$
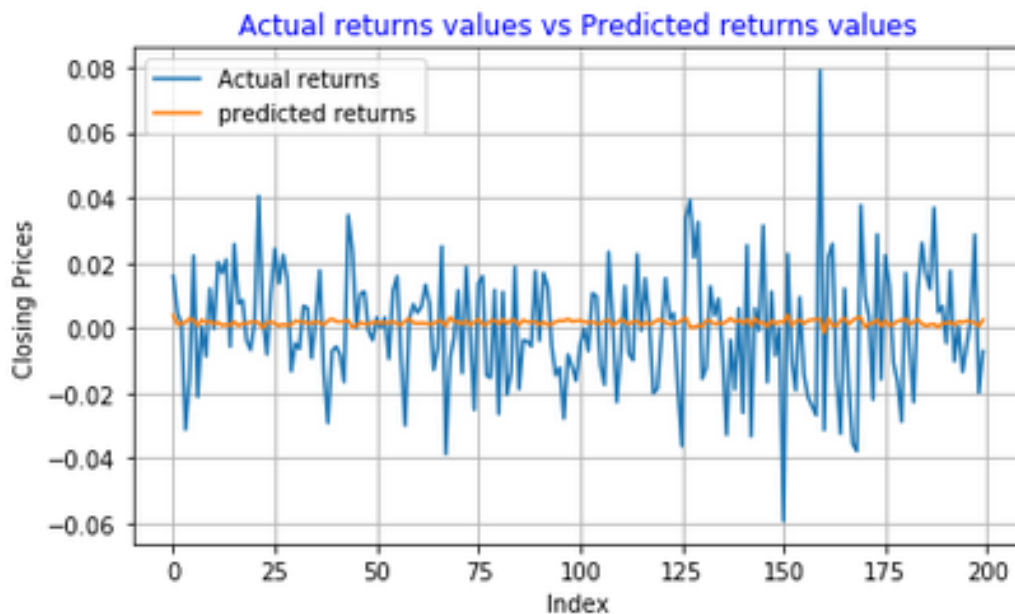
Figure 5.35: **Sanlam Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	68	39
True Up	24	69
Accuracy= 68.5%		

Table 5.31: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Sanlam Group Ltd returns series**

As seen in the predictions of ARMA(1,1)-MH model in Figure 5.35, the model seems to be operating around the mean and the structure of the predictions is similar to that in Figure 5.31, and 5.33. For the Sanlam Group Ltd returns series, the ARMA(1,1)-Metropolis Hastings model achieved a fairly adequate accuracy of 68.5%.

Nedbank Group Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
theta:

        Mean            SD              MC Error        95% HPD interval
        ---------------------------------------------------------------
        -0.015          0.493           0.017           [-0.841  0.886]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |---------------|===============|===============|---------------|
        -0.867          -0.429          -0.014          0.39            0.869


precision:

        Mean            SD              MC Error        95% HPD interval
        ---------------------------------------------------------------
        62.195          3.883           0.1             [ 54.91  69.965]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |---------------|===============|===============|---------------|
        54.802          59.497          62.173          64.796          69.892


phi:

        Mean            SD              MC Error        95% HPD interval
        ---------------------------------------------------------------
        -0.057          0.5             0.018           [-0.924  0.827]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |---------------|===============|===============|---------------|
        -0.902          -0.462          -0.057          0.334           0.865
```

Figure 5.36: **Summary results of ARMA(1,1) Metropolis Hastings model fit on Nedbank Group Ltd returns series**

The standard deviation for the variance parameter (precision) is 3.883, this value is similar to the values obtain in Figure 5.30, 5.32, and 5.34. Such a large value indicates the existence of some degree of uncertainty in the estimated mean parameters. All the Monte Carlo (MC) errors are small, this implies that there are relatively small errors in the estimated mean parameters. Since all the estimated mean parameters lies within the 95% HPD interval, this suggests that true parameters will be within the computed intervals. The estimated ARMA(1,1)-Metropolis Hastings model fitted on Nedbank Group Ltd returns series is given by

$$r_t = (-0.057)r_{t-1} + (-0.015)\epsilon_{t-1} + \epsilon_t, \tag{5.20}$$

where $\epsilon_t \sim N(0, \sigma^2)$. The precision parameter is $1/\sigma^2 = \frac{1}{62.195}$.
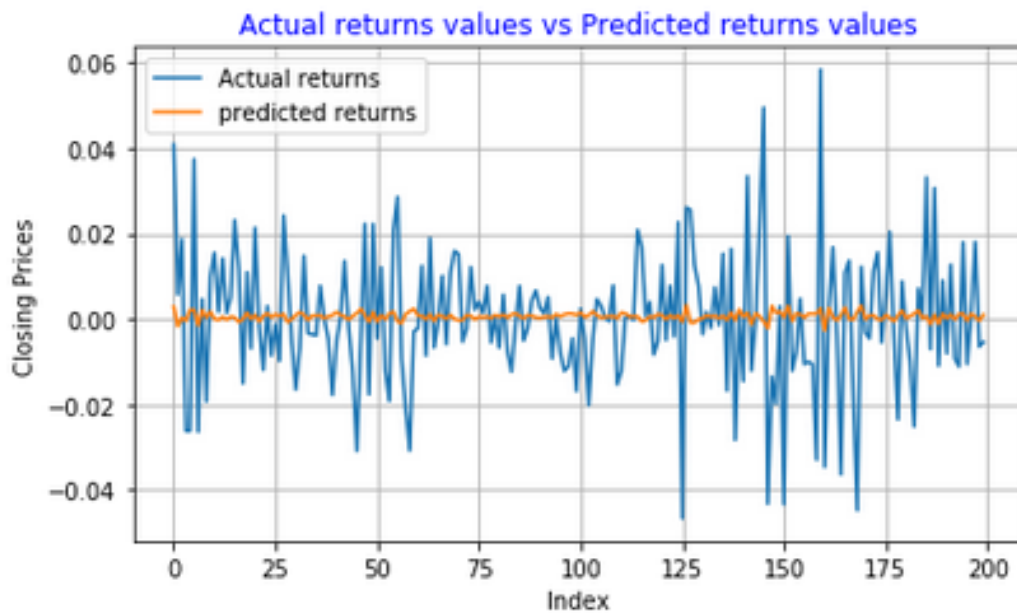
Figure 5.37: **Nedbank Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	Predicted Up
True Down	69	32
True Up	29	70
Accuracy= 69.5%		

Table 5.32: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Nedbank Group Ltd returns series**

As seen in Figure 5.37, when predicting the Nedbank Group Ltd returns series the ARMA(1,1)-Metropolis Hastings model seems to have poorly captured the volatility clusters in the series. However, (Hansson, 2017) argue that accurately capturing the variance does not necessarily imply that we will obtain better forecasts of the series. The structure of predictions in Figure 5.37 is similar to those obtain in Figure 5.31,5.33, and 5.35. The predictions made by the ARMA(1,1)-MH model are extremely close the mean.

Discussion:

We constructed an ARMA(1,1)-Metropolis Hastings model using **pymc** package in Python 2. The model was fitted on various time series and predictions were made. Results shows that the ARMA(1,1)-MH model has a fairly adequate ability to forecast the direction of returns prices. However, the model seems to be operating around the mean. This phenomenon is similar to that observed in a study carried out by (Hansson, 2017), were they used ARMA(1,1)-GJRGARCH(1,1) and LSTM models to predict S&P and OMX returns series. In their study, both the models made predictions which are close to the mean.

5.5   Artificial Neural Network (ANN) results

Netcare Group Ltd: ANN model results

For the Netcare Group Ltd returns series, we constructed an ANN model with the following architecture: 2 hidden layers, a learning rate of 0.0006, 4800 epochs, and a l2 regularization term of 0.0006. A tangent activation function is used on the input and output layers of the neural network.
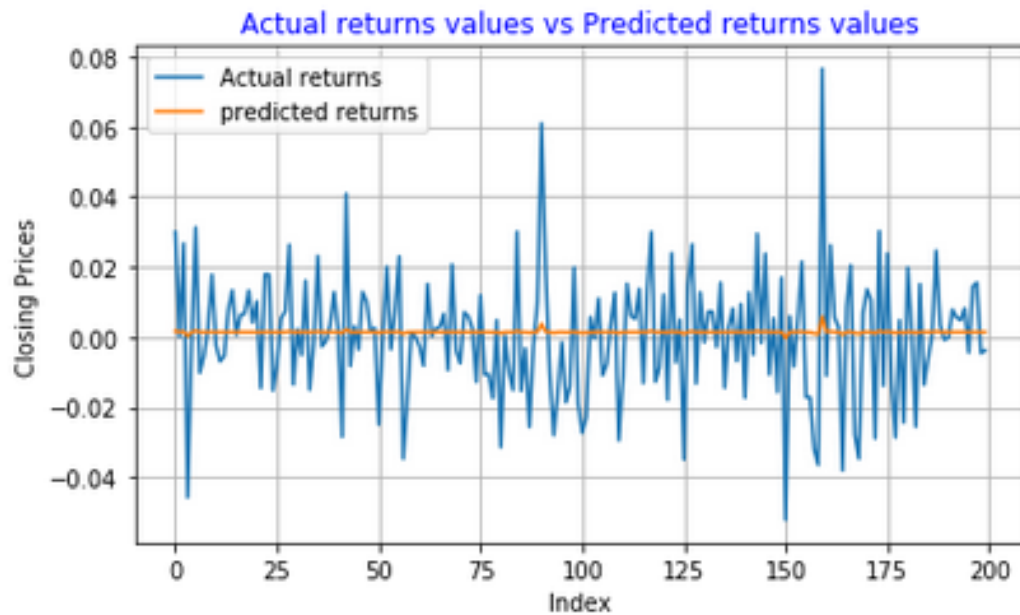
Figure 5.38: **Netcare Group Ltd returns series predictions using ANN model**

	Predicted Down	**Predicted Up**
True Down	75	19
True Up	29	77
	Accuracy= 76%	

Table 5.33: **Confusion Matrix for ANN model's predictions of Netcare Group Ltd returns series**

Figure 5.38 shows the out of sample predictions of Netcare Group Ltd returns series using ANN model. By visual inspecting the plot in Figure 5.38, there seems to be low variance in the predictions. The ANN model seems to be operating around the mean. This phenomenon was also observed in a study carried out by (Hansson, 2017, p.g 22-26). Confusion matrix for ANN model's prediction results is presented in Table 5.33. The results shows that the model has a good ability to predict the direction of Netcare returns series. An "acceptable" accuracy of 76% was achieved by the model.

Santam Ltd: ANN model results

A simple ANN model with 2 hidden layers, a learning rate of 0.0006, 4800 passes over the training data set, and a L2 regularization term of 0.0006 is constructed for Santam Ltd series. For the hidden layers and the output layer a tangent activation function is used.
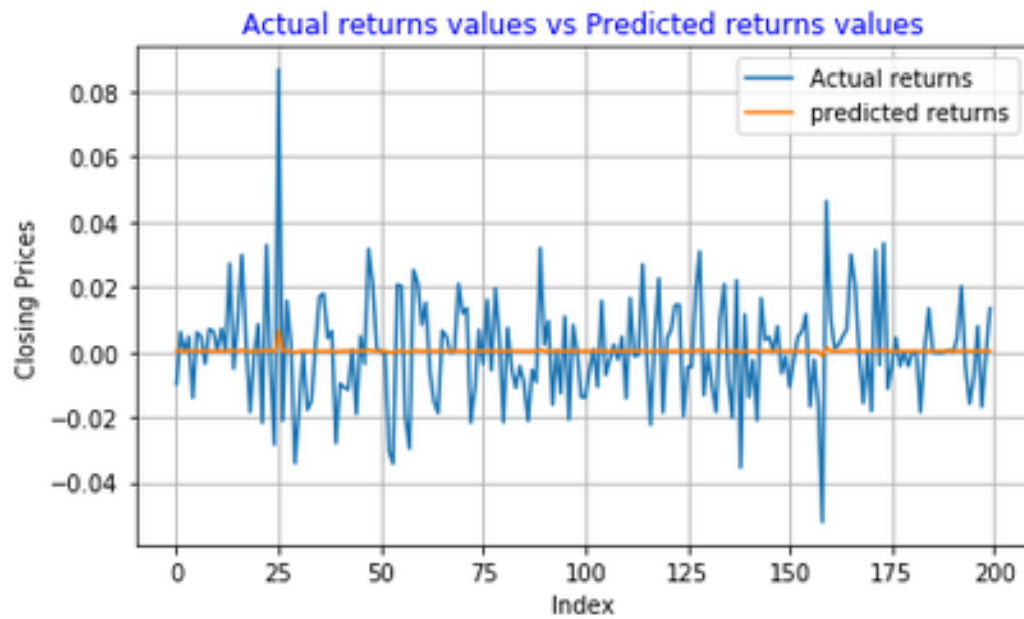
Figure 5.39: **Santam Ltd returns series predictions using ANN model**

	Predicted Down	Predicted Up
True Down	75	26
True Up	32	67

Accuracy= 71%

Table 5.34: **Confusion Matrix for ANN model's predictions of Santam Ltd returns series**

Figure 5.39 shows the Santam Ltd forecasts made by ANN model, the structure of the predictions is similar to that in Figure 5.38. Again, the model is operating around the mean. A seen in the confusion matrix presented in Table 5.34, the model achieved an accuracy of 71%. The reason for the adequate performance of the model could be due to the model being unable to capture the dynamics of the returns series data.

Sanlam Group Ltd: ANN model results

We constructed an ANN neural network with 2 hidden layers, a learning rate of 0.0007, an L2 regularisation term of 0.00054, and 3930 (epochs) passes over the training data set. A tangent activation function is used for the input and output layers.
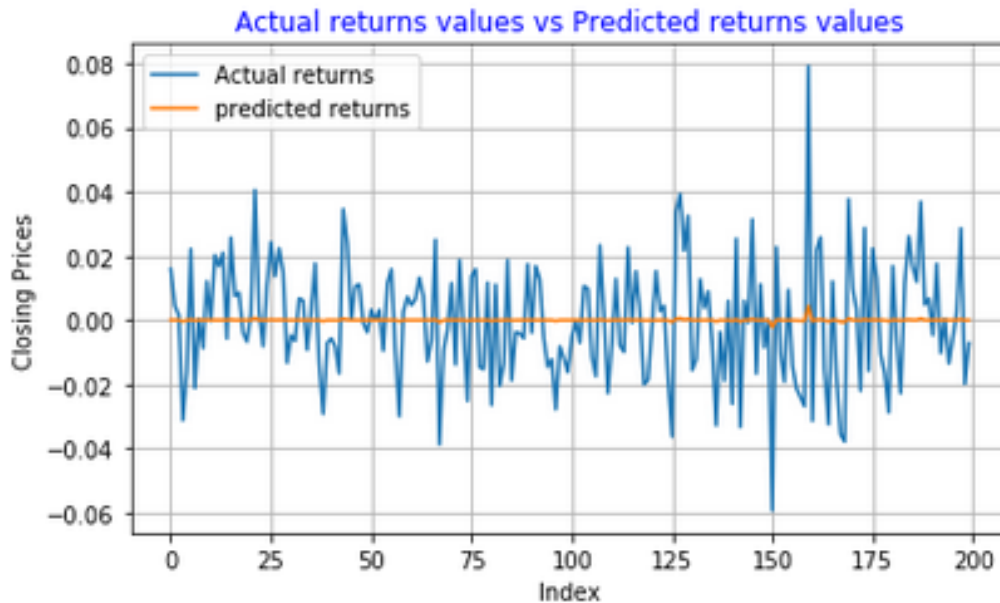
Figure 5.40: **Sanlam Group Ltd returns series predictions using ANN model**

	Predicted Down	Predicted Up
True Down	90	17
True Up	16	77

Accuracy= 83.5%

Table 5.35: **Confusion Matrix for ANN model's predictions of Sanlam Group Ltd returns series**

As seen in Figure 5.40, the ANN model predictions (yellow line) are extremely close to the mean. The variance of these predictions is also very small. When predicting log returns of S&P and OMX stock indices using LSTM and ARMA(1,1)-GJRGARCH(1,1) model, (Hansson, 2017, p.g 22-26)

also observed the same phenomenon. However, the confusion matrix presented in Table 5.35 shows that the ANN model has a good ability to predict the direction of Sanlam Group Ltd returns series.

Nedbank Group Ltd: ANN model results

Architecture of the ANN model fitted on Nedbank Group Ltd returns series is as follows: the model has 2 hidden layers, a learning rate and L2 regularization term of 0.0006, 5400 epochs (number of iterations over training set), and a tangent activation function for the hidden layers and the output layer.
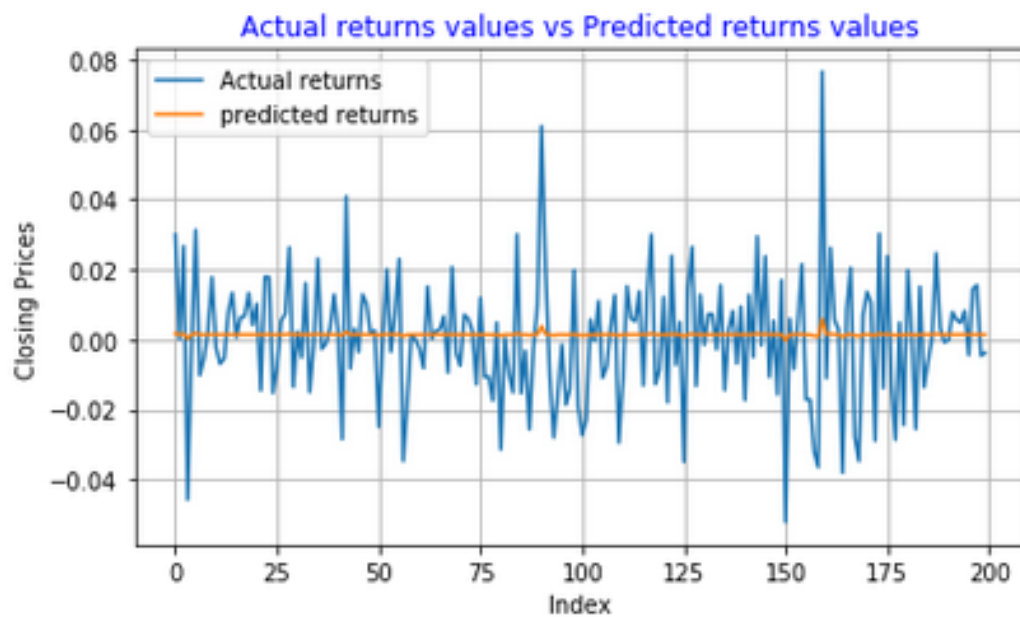
Figure 5.41: **Nedbank Group Ltd returns series predictions using ANN model**

	Predicted Down	Predicted Up
True Down	92	9
True Up	12	87

Accuracy= 89%

Table 5.36: **Confusion Matrix for ANN model's predictions of Nedbank Group Ltd returns series**

ANN model's Nedbank Group Ltd returns series predictions are presented in Figure 5.41. The structure of the predictions is similar to that in Figure 5.38,5.39, and 5.40. The predictions are extremely close to the mean. As seen in the results of the confusion matrix presented in Table 5.36, the ANN model has a good ability to correctly predict the direction of Nedbank Group Ltd returns series. The model achieved an accuracy of 89%.

Discussion

We fitted an Artificial Neural Network model on four JSE listed companies returns series. The model was evaluated in terms of its ability to correctly predict the direction of returns prices. The model has a strong ability to predict whether the next day's return price will go up or down, the lowest accuracy achieved by the model was 71% and the highest was 89%. The predictions from the fitted ANN models were operating around the mean. This phenomenon was also observed in studies carried out by (Hansson, 2017) and (Björklund and Uhlin, 2017).

5.6 ARMA(1,1) Model Results, Train:700 Test:200

Preliminary tests for ARMA model

Table 5.37: **Testing for Stationarity of the returns series using Augmented Dickey Fuller (ADF)**

Stock	Test statistic	P-value
Netcare	-6.335	0.000
Sanlam	-6.221	0.000
Nedbank	-6.205	0.000
Santam	-7.618	0.000

Critical Values: $-3.44(1\%)$, $-2.87(5\%)$, $-2.57(10\%)$

To apply a time series data to an ARMA model, it is important to first test for stationarity of the time series. We reject the null hypothesis of non-stationarity (unit root) since the ADF test statistic is less than the critical values at 1%,5%, and 10% significance level. Hence, all the time series are stationary. Small p-values are also a clear indication that there is strong evidence against the null hypothesis of existence of a unit root in the series.

Table 5.38: **Testing for Stationarity of the returns series using Phillips Perron test**

Stock	Test statistic	P-value
Netcare	-30.460	0.000
Sanlam	-28.946	0.000
Nedbank	-31.916	0.000
Santam	-28.934	0.000

Critical Values: $-3.44(1\%)$, $-2.87(5\%)$, $-2.57(10\%)$

For all four time series, the Phillips Perron test statistic is less than the critical values at 1%,5%, and 10% level of significance. So, the null hypothesis of non-stationarity is rejected in favour of the alternative hypothesis of no unit root. The p-values for the test statistic are also extremely small, suggesting strong evidence against the null hypothesis.

Netcare Group Ltd: Estimated ARMA(1,1) model

```
                            ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  899
Model:                    ARMA(1, 1)  Log Likelihood                2395.619
Method:                         mle   S.D. of innovations              0.017
Date:                Sun, 21 Jan 2018 AIC                          -4785.238
Time:                       17:52:03  BIC                          -4770.834
Sample:                            0  HQIC                         -4779.735
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.4960      0.140      3.539      0.000       0.221       0.771
ma.L1.y       -0.6027      0.128     -4.727      0.000      -0.853      -0.353
                                    Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            2.0160           +0.0000j            2.0160            0.0000
MA.1            1.6592           +0.0000j            1.6592            0.0000
------------------------------------------------------------------------------
```

Figure 5.42: **Summary results of ARMA(1,1) model fit on Netcare returns series**

Figure 5.42 shows summary results of the estimated ARMA(1,1) model fitted on Netcare Group Ltd returns series. The estimated coefficients for the Autoregressive (AR) and the Moving Average (MA) parameters lies within the confidence interval, this suggests that the true parameters lie within the computed intervals. When using the 95% confidence level, the critical Z scores are ± 1.96 standard deviations. The p-value associated with the 95% confidence level is 0.05. Since the $\mid Z \mid$- statistic values are greater than the critical value (1.96), this suggests that the estimated AR and MA parameters are significantly different from zero. Small standard errors are a clear indication that the estimated parameters are close to the true values. The roots of the AR and MA parameters are greater than 1 in modulus, suggesting that the fitted ARMA(1,1) model is stationary. The estimated ARMA(1,1) model for Netcare Group Ltd returns series is

$$r_t = 0.4960 r_{t-1} + (-0.6827)\epsilon_{t-1} + \epsilon_t, \tag{5.21}$$

where error term $\epsilon_t \sim N(\mu = 0, \sigma^2)$.

Assessing the Adequacy of the ARMA(1,1) Model Fitted on Netcare Returns Series
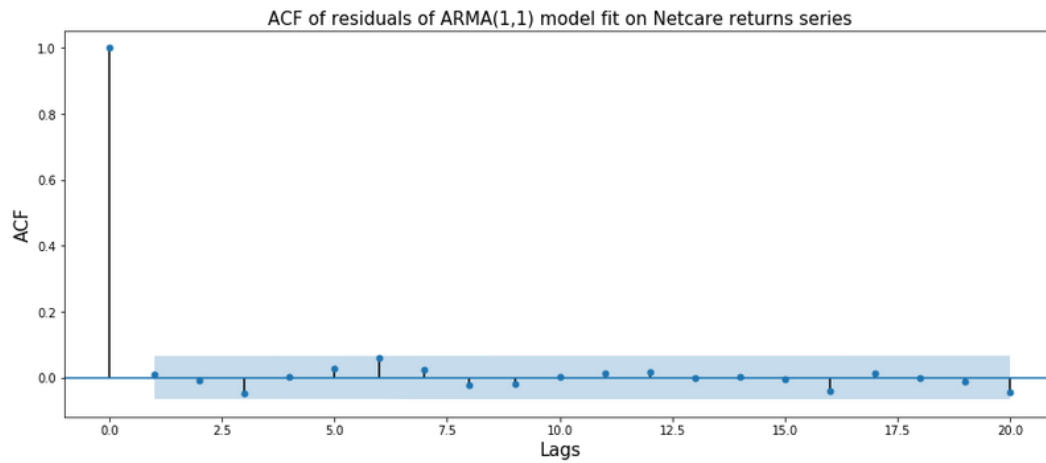
Figure 5.43: **ACF of residuals of ARMA(1,1) model fitted on Netcare returns series**

Table 5.39: **Independence test of residuals of ARMA(1,1) fitted on Netcare series**

Lag	Ljung statistic	P-value
0	0.0651	0.7956
5	6.1609	0.4054
10	7.6300	0.7460
15	9.4744	0.8926
20	11.5930	0.9499

The ACF plot in Figure 5.43 shows that the residuals seems to be uncorrelated and they appear to be patternless. Hence, this indicates that the is no autocorrelations in the residuals. Summary results of Ljung Box test statistic presented in Table 5.39 shows that the residuals are independent since all the p-values are greater than the 0.05 level of significance.
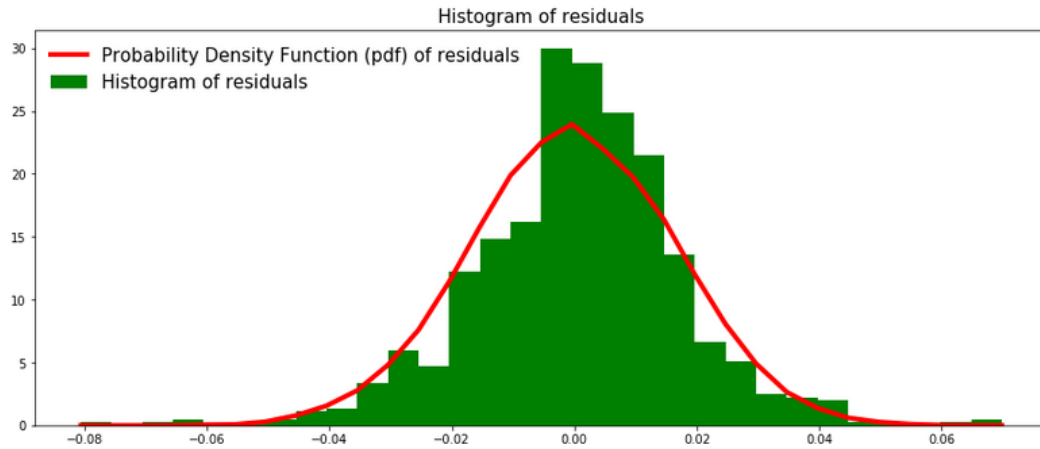
Figure 5.44: **Histogram of residuals of ARMA(1,1) model fitted on Netcare Group Ltd returns series**

Apart from summary statistics, histogram of residuals and density plots are used to better understand the distribution of residuals (errors). Figure 5.44 shows the histogram of residuals (green) and probability density function (pdf) of a Gaussian distributed process with mean of *zero* and standard deviation of residuals. The distribution seems to be symmetric and the residuals are clustered around the mean. The Durbin Watson (DW) test statistic value for residuals from the fitted ARMA(1,1) model is 1.9797, this value is very close to 2 suggesting that there is no autocorrelations in the residuals.
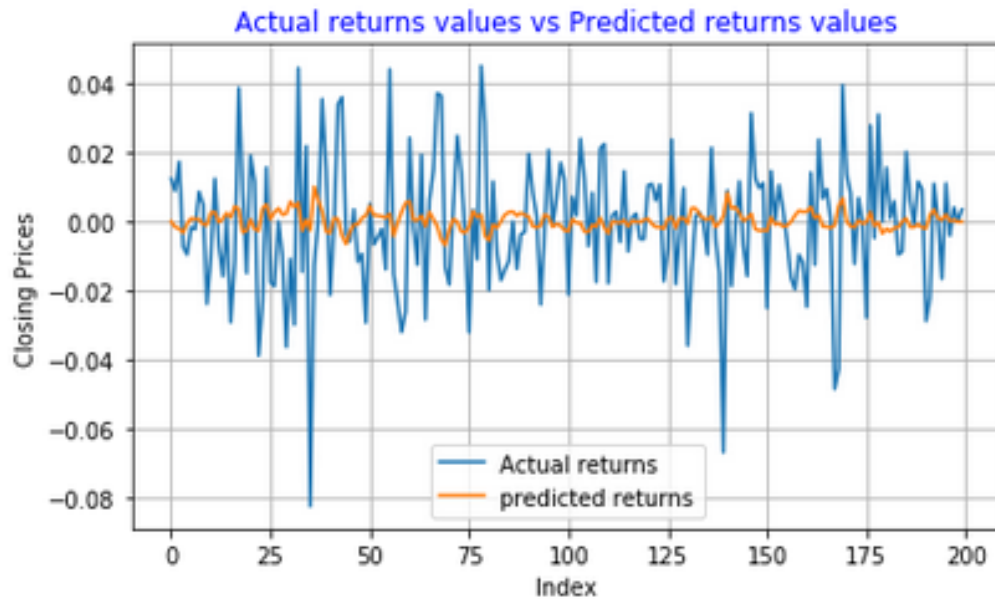
Figure 5.45: **Netcare Group Ltd returns series predictions using ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	79	25
True Up	25	71
Accuracy= 75%		

Table 5.40: **Confusion Matrix for ARMA(1,1) model's predictions of Netcare Group Ltd returns series**

Figure 5.45 shows the predicted Netcare returns series using ARMA(1,1) model. The model seems to have adequately captured the volatility clusters in the series. As seen in the summary results of the confusion matrix presented in Table 5.40, the model has a good ability to accurately predict the direction of movement of Netcare returns prices. The model achieved an accuracy of 75%.

Santam Ltd: Estimated ARMA(1,1) model

```
                        ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  899
Model:                   ARMA(1, 1)   Log Likelihood                2380.279
Method:                         mle   S.D. of innovations              0.017
Date:                Sun, 21 Jan 2018   AIC                          -4754.559
Time:                      17:55:29   BIC                          -4740.155
Sample:                           0   HQIC                         -4749.056

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.5427      0.282      1.923      0.055      -0.010       1.096
ma.L1.y       -0.6330      0.261     -2.423      0.016      -1.145      -0.121
                                Roots
==============================================================================
                 Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1           1.8427           +0.0000j           1.8427            0.0000
MA.1           1.5798           +0.0000j           1.5798            0.0000
------------------------------------------------------------------------------
```

Figure 5.46: **Summary results of ARMA(1,1) model fit on Santam Ltd returns series**

The summary results of the estimated ARMA(1,1) model fit on Santam Ltd returns series are presented in Figure 5.46. All the estimated ARMA model's parameters lies within the 95% confidence interval, indicating that the true parameters lie within the computed intervals. The standard errors for the AR and MA parameters are 0.282 and 0.261 respectively, this suggests that the estimated parameters are close to the true values. The Z score for the AR parameter is slightly less than the 1.96 critical value, suggesting that the parameter may be insignificantly different from zero. The estimated ARMA(1,1) model is

$$r_t = 0.5427 r_{t-1} + (-0.6338)\epsilon_{t-1} + \epsilon_t, \tag{5.22}$$

where $\epsilon_t \sim N(0, \sigma^2)$.

Assessing the adequacy of the ARMA(1,1) model fitted on Santam returns series
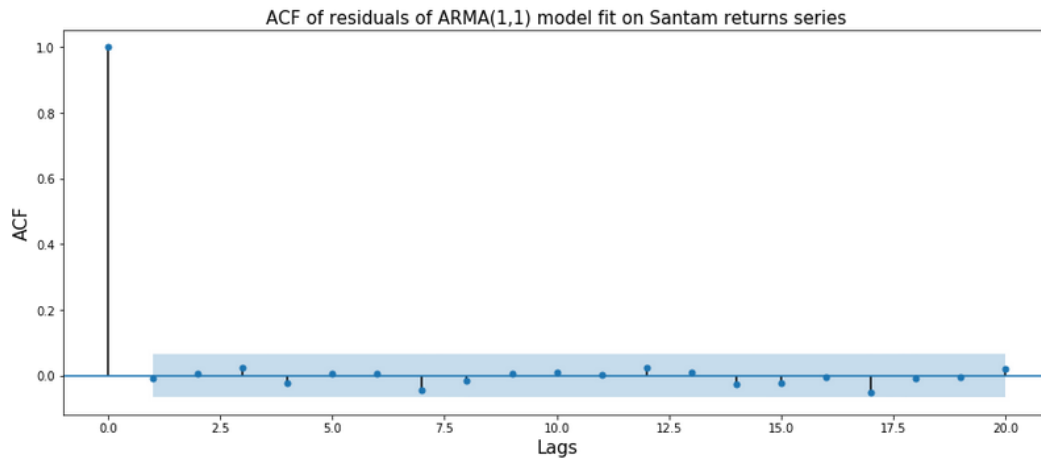
Figure 5.47: **ACF of residuals of ARMA(1,1) model fitted on Santam returns series**

Table 5.41: **Independence test of residuals from ARMA(1,1) model fitted on Santam series**

Lag	Ljung statistic	P-value
0	0.0459	0.8311
5	1.1485	0.9794
10	3.4202	0.9839
15	5.1466	0.9950
20	8.7889	0.9900

By visual inspecting the ACF plot in Figure 5.47, we can see that there are no "spikes" which exceeds the significance bounds, this indicates that the residuals from the ARMA(1,1) model fitted on Santam returns series are uncorrelated. Since all the p-values for the Ljung Box test statistic are greater than 0.05 level of significance, this suggests that we have weak evidence against the null hypothesis of independence of the residuals of the fitted model.
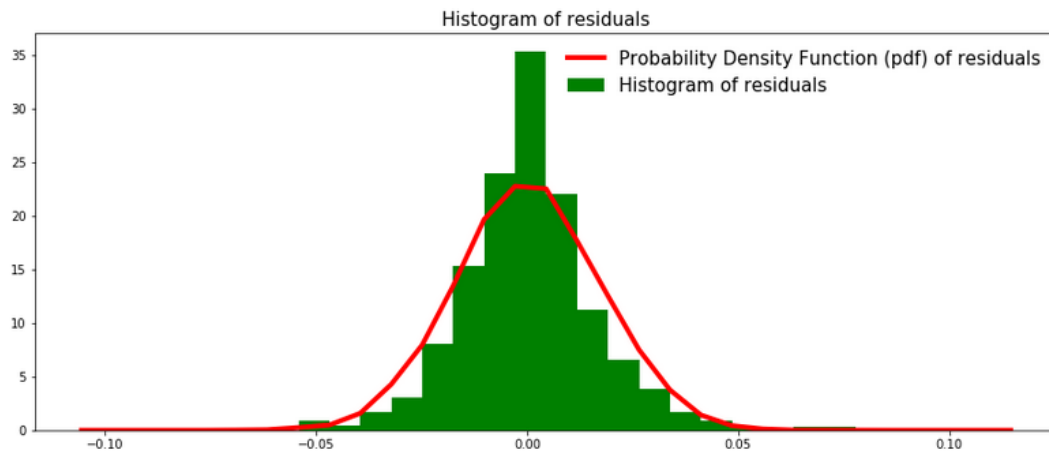
Figure 5.48: **Histogram of residuals of the ARMA(1,1) model fitted on Santam Ltd returns series**

Figure 5.48 shows the histogram of residuals (green) from the fitted ARMA(1,1) model and the pdf of a Gaussian process with mean of *zero* and standard deviation of residuals. By visual inspecting the histogram in Figure 5.48, the residuals seems to be centered around the mean, the distribution has "thin" tails and is symmetric. The DW test statistic for independence of residuals test is 2.0137, so we fail to reject the null hypothesis that the residuals are uncorrelated since the test statistic value is close to 2.
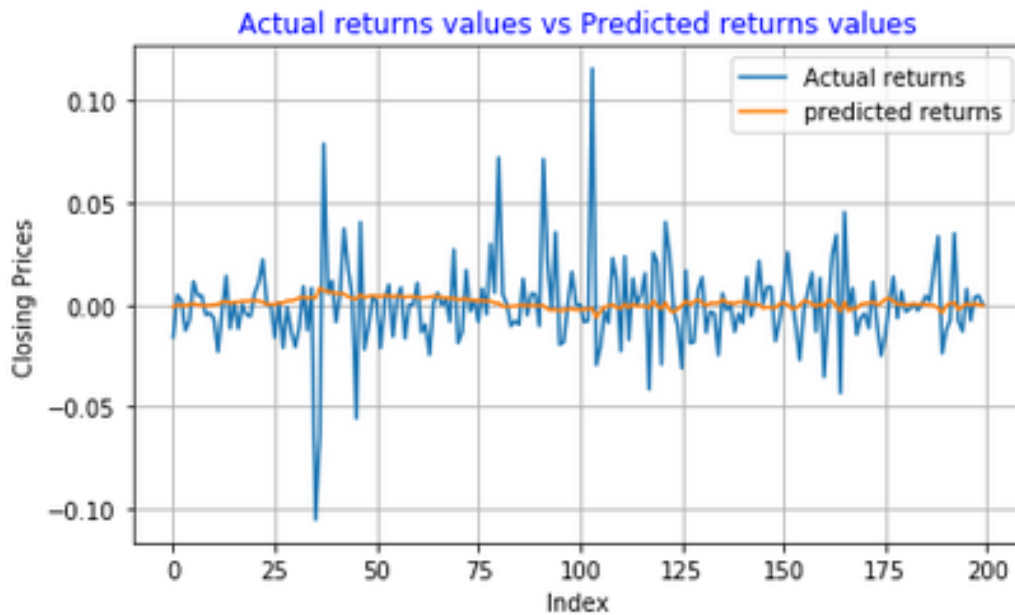
Figure 5.49: **Santam Ltd returns series predictions using ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	77	19
True Up	27	77

Accuracy= 77%

Table 5.42: **Confusion Matrix for ARMA(1,1) model's predictions of Santam Ltd returns series**

ARMA(1,1) model's Santam Ltd returns series predictions are presented in Figure 5.49. The model seems to be operating extremely close to the mean. Results from the confusion matrix presented in Table 5.42 shows that the model has a "very strong" ability to predict whether the next day's returns price will go up or down. It achieved an overall accuracy of 77%.

Sanlam Group Ltd: Estimated ARMA(1,1) model

```
                            ARMA Model Results
==============================================================================
Dep. Variable:                    y   No. Observations:                  899
Model:                   ARMA(1, 1)   Log Likelihood                2263.081
Method:                         mle   S.D. of innovations              0.020
Date:                Sun, 21 Jan 2018   AIC                          -4520.161
Time:                      17:57:49   BIC                           -4505.757
Sample:                           0   HQIC                          -4514.658
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.9181      0.038     24.036      0.000       0.843       0.993
ma.L1.y       -0.9529      0.029    -33.301      0.000      -1.009      -0.897
                                   Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            1.0892           +0.0000j            1.0892            0.0000
MA.1            1.0495           +0.0000j            1.0495            0.0000
------------------------------------------------------------------------------
```

Figure 5.50: **Summary results of the ARMA(1,1) model fit on Sanlam Group Ltd returns series**

Summary results of the ARMA(1,1) model fitted on Sanlam Group Ltd returns series are presented in Figure 5.50. Both the estimated AR and MA parameters lies within the 95% confidence level, this is a clear indication that the true parameters lie within the computed intervals. The standard errors are extremely close to zero, this indicates that the estimated AR and MA parameters are close to the true values. The estimated AR and MA parameters are significantly different from zero since they have $|Z|$ statistics values which are greater than the 1.96 critical value. The fitted ARMA(1,1) model is stationary since all its roots are greater than 1 in modulus. The estimated ARMA(1,1) model for Sanlam Group Ltd series is

$$r_t = 0.9181 r_{t-1} + (-0.9529)\epsilon_{t-1} + \epsilon_t, \tag{5.23}$$

where $\epsilon_t \sim N(0, \sigma^2)$.

Assessing the Adequacy of the ARMA(1,1) Model Fitted on Sanlam Returns Series
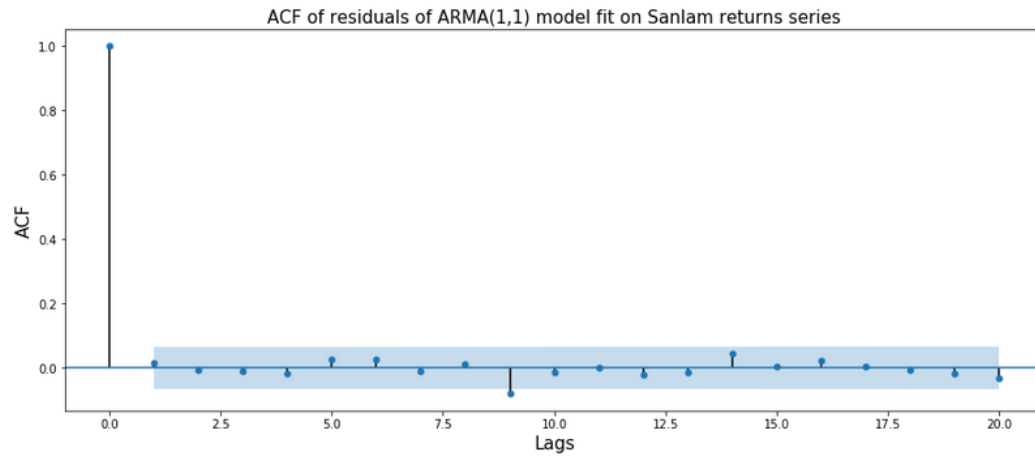
Figure 5.51: **ACF of residuals from the ARMA(1,1) model fitted on Sanlam returns series**

Table 5.43: **Independence test of residuals from the ARMA(1,1) model fitted on Sanlam returns series**

Lag	Ljung statistic	P-value
0	0.1799	0.6714
5	1.7578	0.9406
10	7.7706	0.7337
15	10.6101	0.8329
20	12.1246	0.9362

From the ACF plot of residuals in Figure 5.51, there is one significant "spike" at the 9th lag, but overall the the residuals appears to be patternless. Summary results of the Ljung Box test statistic in Table 5.43 shows that we have strong evidence to accept the null hypothesis of independence of the residuals since all p-values are greater than 0.05 (level of significance).
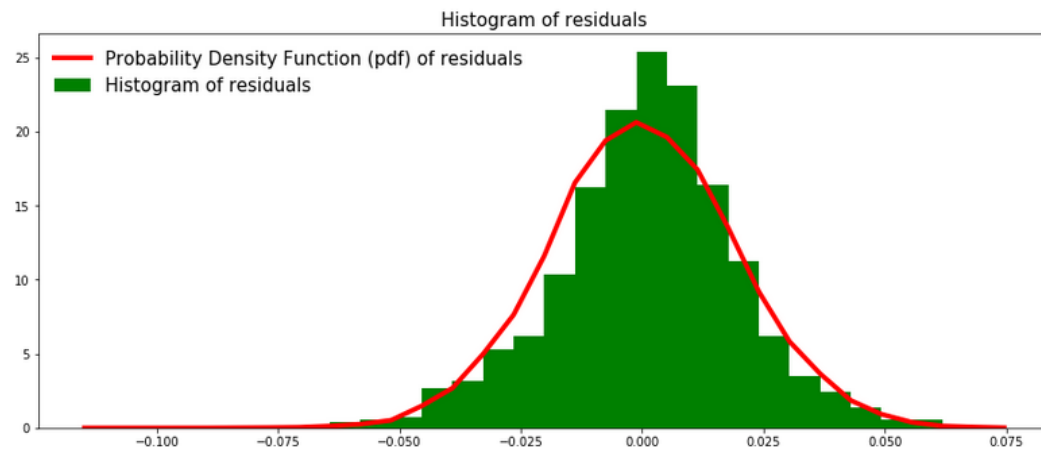
Figure 5.52: **Histogram of residuals of ARMA(1,1) model fitted on Sanlam Group Ltd returns series**

If the distribution of the residuals is distinctly non-normal (Gaussian), it implies that the assumption made by the ARMA modelling process are incorrect and a different modelling technique should be considered. The histogram of residuals in Figure 5.52 looks Gaussian, and roughly symmetric. The DW test statistic is 1.9690 , so we fail to reject the null hypothesis that the residuals from ARMA(1,1) model fitted on Sanlam Group Ltd returns series are uncorrelated.
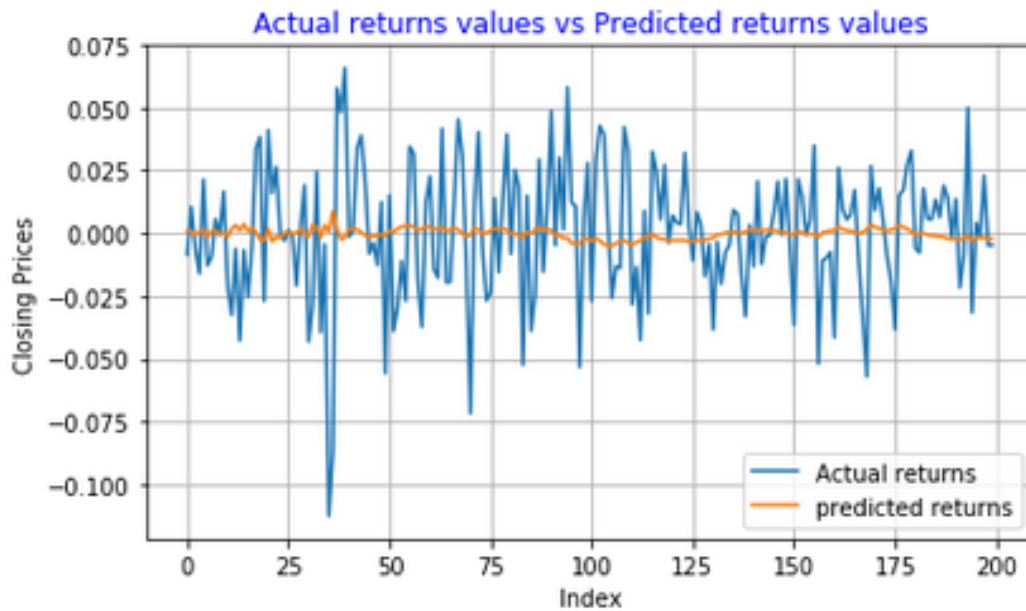
Figure 5.53: **Sanlam Group Ltd returns series predictions using ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	73	28
True Up	28	71

Accuracy= 72%

Table 5.44: **Confusion Matrix for ARMA(1,1) model's predictions of Sanlam Group Ltd returns series**

Figure 5.53 looks similar to Figure 5.53. The model did not adequately capture volatility clustering in the series. But, better capturing the variance of the series does not necessarily mean that the model would produce better forecasts (Hansson, 2017, pg. 24). The model achieved an accuracy of 72%.

Nedbank Group Ltd: Estimated ARMA(1,1) model

```
                          ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  899
Model:                     ARMA(1, 1)   Log Likelihood                2421.639
Method:                           mle   S.D. of innovations              0.016
Date:                Sun, 21 Jan 2018   AIC                          -4837.277
Time:                        17:59:32   BIC                          -4822.873
Sample:                             0   HQIC                         -4831.775

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1.y        0.7116      0.108      6.575      0.000       0.500       0.924
ma.L1.y       -0.7859      0.095     -8.317      0.000      -0.971      -0.601
                                    Roots
==============================================================================
                  Real          Imaginary           Modulus         Frequency
------------------------------------------------------------------------------
AR.1            1.4053           +0.0000j            1.4053            0.0000
MA.1            1.2724           +0.0000j            1.2724            0.0000
------------------------------------------------------------------------------
```

Figure 5.54: **Summary results of ARMA(1,1) model fit on Nedbank Group Ltd returns series**

Figure 5.54 shows summary results of the estimated ARMA(1,1) model fitted on Nedbank Group Ltd returns series. Since the estimated AR and MA parameters lies within 95% confidence interval, this suggests that we are 95% confident that the true parameters lies within the computed intervals. Small standard errors are an indication that the estimated ARMA model's parameters are close to the true values. The roots of the AR and MA parameters are 1.4053 and 1.2724 respectively, since these values are greater than 1 this suggests that the fitted ARMA(1,1) model is stationary. Relatively large Z statistic values indicates that both the AR and MA parameters are significantly different from zero. The estimated ARMA(1,1) model for Nedbank Group Ltd returns series is

$$r_t = 0.7116 r_{t-1} + (-0.7859)\epsilon_{t-1} + \epsilon_t, \tag{5.24}$$

where ϵ_t is a normally distributed process with mean of *zero* and constant variance.

**Assessing the Adequacy of the ARMA(1,1) Model Fitted on Nedbank Returns Series**
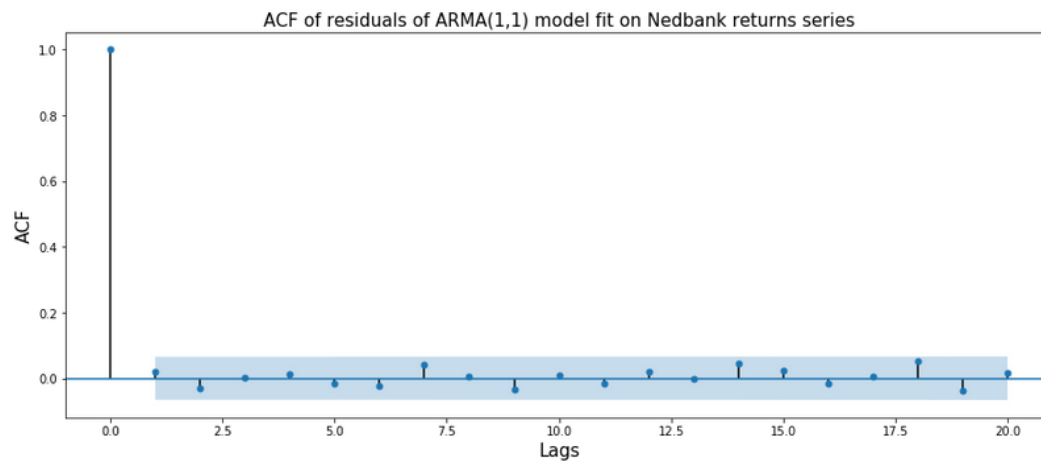
Figure 5.55: **ACF of residuals from the ARMA(1,1) fitted on Nedbank returns series**

Table 5.45: **Independence test of residuals of ARMA(1,1) model fitted on Nedbank series**

Lag	Ljung statistic	P-value
0	0.3495	0.5544
5	2.0293	0.9170
10	5.0414	0.9291
15	8.0586	0.9471
20	12.4164	0.9277

The residuals from the ARMA(1,1) model fitted on Nedbank Group returns series seems to be patternless and uncorrelated as seen in the ACF plot in Figure 5.55. No spikes exceeds the significance bounds, this suggest that there is no significant autocorrelations in the residuals. We cannot reject the null hypothesis of independence of residuals since all the p-values for the Ljung Box test statistic are greater than the 0.05 level of significance.
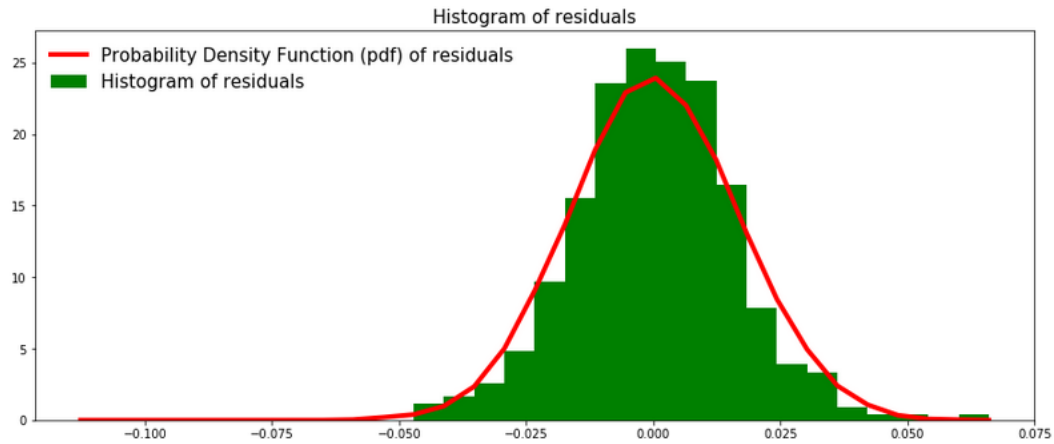
Figure 5.56: **Histogram of residuals of ARMA(1,1) model fitted on Nedbank Group Ltd returns series**

As seen in Figure 5.56, the residuals from the ARMA(1,1) model are clustered around the mean (center of the distribution) and they look Gaussian. Hence, this suggests that the assumptions of the modelling process are correct. The DW test statistic value for residuals from the ARMA(1,1) model fitted on Nedbank Group Ltd returns series is 1.9587, this value is close to 2, suggesting that we can not reject the null hypothesis that the residuals are uncorrelated.
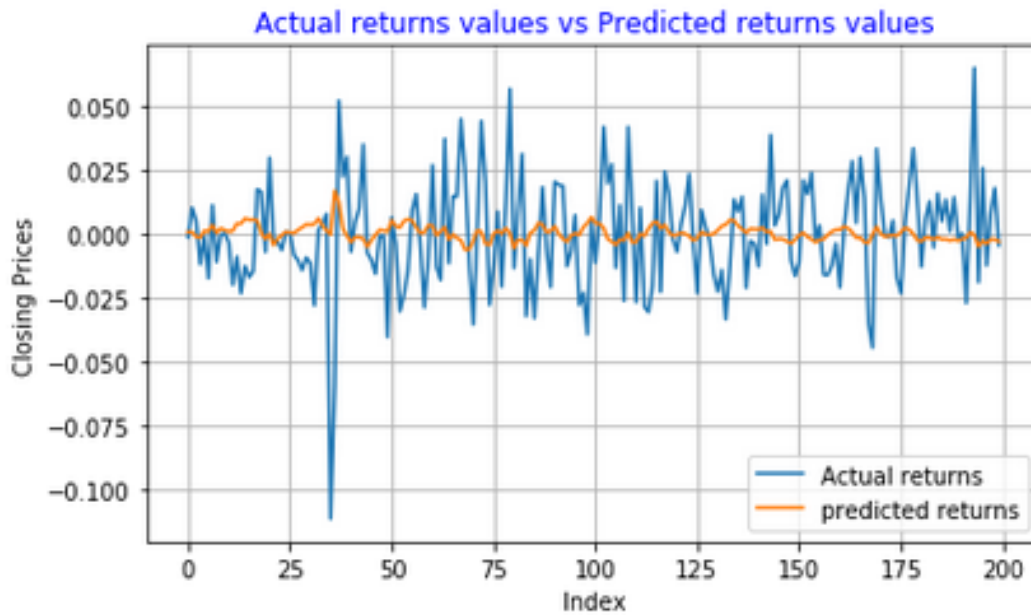
Figure 5.57: **Nedbank Group Ltd returns series predictions using ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	72	28
True Up	34	66

Accuracy= 69%

Table 5.46: **Confusion Matrix for ARMA(1,1) model's predictions of Nedbank Group Ltd returns series**

When predicting the Nedbank Group returns series, the ARMA(1,1) model seems to have adequately captured the volatility clusters in the series. But, (Hansson, 2017) argue that better capturing the variance of the series does not necessarily result in better forecast of the returns series. The model achieved a "fairly adequate" accuracy of 69%.

Discussion

An ARMA(1,1) model was fitted on the Netcare, Sanlam, Santam, and Nedbank returns series using steps recommended by (Box and Jenkins, 1976). Summary results of the confusion matrix shows that the model has a "good"

ability to correctly predict the direction of movement of these time series. The highest accuracy achieved by the model was 77% (for Santam Ltd returns series) and the lowest was 69% (for Nedbank Group Ltd returns series). All the assumptions of the ARMA modelling process were satisfied, this suggests that the model is an appropriate modelling technique for these time series.

5.7 AR(3)-AVGARCH(1,1) -Student-t results, Train:700 Test:200

Preliminary statistical tests on the returns series

Table 5.47: **Test for normality of returns using Shapiro Wilk test**

Stock	Shapiro Wilk test statistic	Kurtosis	Skew-ness
Netcare	0.9822 (0.0000)	1.6783	0.0323
Santam	0.9580 (0.0000)	3.6849	0.0087
Sanlam	0.9898 (0.0000)	1.1448	-0.0271
Nedbank	0.9917 (0.0006)	1.0232	0.0881

Notes: p-values of the test are in parentheses

Table 5.47 shows summary results of the Shapiro Wilk test statistic, kurtosis, and Skewness values. Since all the p-values for the Shapiro Wilk test statistics are less than 0.05 level of significance, we have strong evidence to reject the null hypothesis of normality of the returns series. Hence, all the returns series are non-normal. A normal distribution has a kurtosis of 3 and a skewness of 0. A kurtosis value of less than 3 for the Netcare, Sanlam , and Nedbank returns series suggests that their distribution has "lighter" tails than the normal distribution. The skewness value for all the returns series is extremely close to 0, this suggests that the distributions of these returns series are almost symmetric.

Table 5.48: **Testing for ARCH effects in the returns using Lagrange Multiplier (LM) test**

Stock	Lagrange Multiplier test statistic	P-value
Netcare	48.0507	0.0004
Santam	48.3021	0.0004
Sanlam	67.3852	0.0000
Nedbank	62.1455	0.0000

In order to model a time series using GARCH-type models, it is important to first test for presence of ARCH behaviour (effects) in the series. Table 5.48 shows summary results of the Lagrange Multiplier test for presence of ARCH effects. Since all the p-values for LM test are less than 0.05 level of significance, we have strong evidence against the null hypothesis of no ARCH effects. Hence, we reject the null hypothesis and conclude that there are ARCH effects in all four time series. Hence, we can model these time series using GARCH-type models.

Netcare Group Ltd: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                        AR - AVGARCH Model Results
==================================================================================
Dep. Variable:                        Close   R-squared:
Mean Model:                              AR   Adj. R-squared:
Vol Model:                          AVGARCH   Log-Likelihood:
Distribution:       Standardized Student's t  AIC:
Method:                Maximum Likelihood      BIC:
                                              No. Observations:
Date:                     Wed, Jan 24 2018    Df Residuals:
Time:                            17:37:14     Df Model:
                                Mean Model
==================================================================================
               coef     std err         t     P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------------
Const       4.7091e-03  1.066e-03     4.416   1.004e-05  [2.619e-03,6.799e-03]
Close[1]       -0.8063     0.201      -4.006   6.181e-05  [ -1.201, -0.412]
Close[2]       -0.5580     0.150      -3.728   1.933e-04  [ -0.851, -0.265]
Close[3]       -0.1874   7.203e-02    -2.602   9.256e-03  [ -0.329,-4.628e-02]
                             Volatility Model
==================================================================================
               coef     std err         t     P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------------
omega          0.0449   2.884e-03     15.559   1.388e-54  [3.922e-02,5.053e-02]
alpha[1]    1.4166e-05     0.580    2.444e-05    1.000    [ -1.136,  1.136]
beta[1]        0.0693     0.116       0.600      0.548    [ -0.157,  0.296]
                               Distribution
==================================================================================
               coef     std err         t     P>|t|    95.0% Conf. Int.
----------------------------------------------------------------------------------
nu            57.1710     4.395       13.009   1.089e-38  [ 48.557, 65.785]
==================================================================================
```

Figure 5.58: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Netcare returns series**

Figure 5.58 shows summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Netcare Group Ltd returns series. Since the standard errors are close to 0, this indicates that the estimated parameters of the AR(3)-AVGARCH(1,1)-Student-t model are close to the "true" parameters. As the degree of freedom parameter approaches infinity (i.e $\nu \to \infty$), the Student-t distribution reduces to a Normal distribution. In our case, the degree of freedom parameter ν is equal to 57.17, this value is large, this suggests that the distribution of the residual series is closer to the normal distribution. All the parameters of the Mean model (AR(3)) are significantly different from 0 since their t-statistics are greater than 2. The estimated AR(3)-AVGARCH(1,1)-Student-t model for Netcare Group Ltd returns se-

ries is given as

$$r_t = 0.0047091 + (-0.8063)r_{t-1} + (-0.5580)r_{t-2} + (-0.1874)r_{t-3} + \epsilon_t, \tag{5.25}$$

where the error term of the model $\epsilon_t = \sigma_t z_t$, and z_t is a Student-t distributed process with $\nu = 57.17$ degrees of freedom. The conditional variance equation is

$$\sigma_t^2 = 0.0449 + 0.000014166 \mid \epsilon_{t-1}^2 \mid +0.0693\sigma_{t-1}^2. \tag{5.26}$$

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.49: **Diagnostics on standardised residuals of the AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
24.3141 (0.4512)	1.9865	20.1856 (0.7651)

Notes: p-values of the tests are in parentheses

Results of the diagnostics tests carried out on the standardised residuals from the fitted model are presented in Table 5.49. There are no ARCH effects remaining in the residuals since the p-value for the LM test is greater than the 0.05 level of significance. We accept at 5% significance level the null hypothesis that the standardised residuals are uncorrelated since the DW test statistic is extremely close to 2. The p-value of the Ljung Box test is greater than 0.05 (level of significance), hence the null hypothesis that the residuals are independent is accepted.
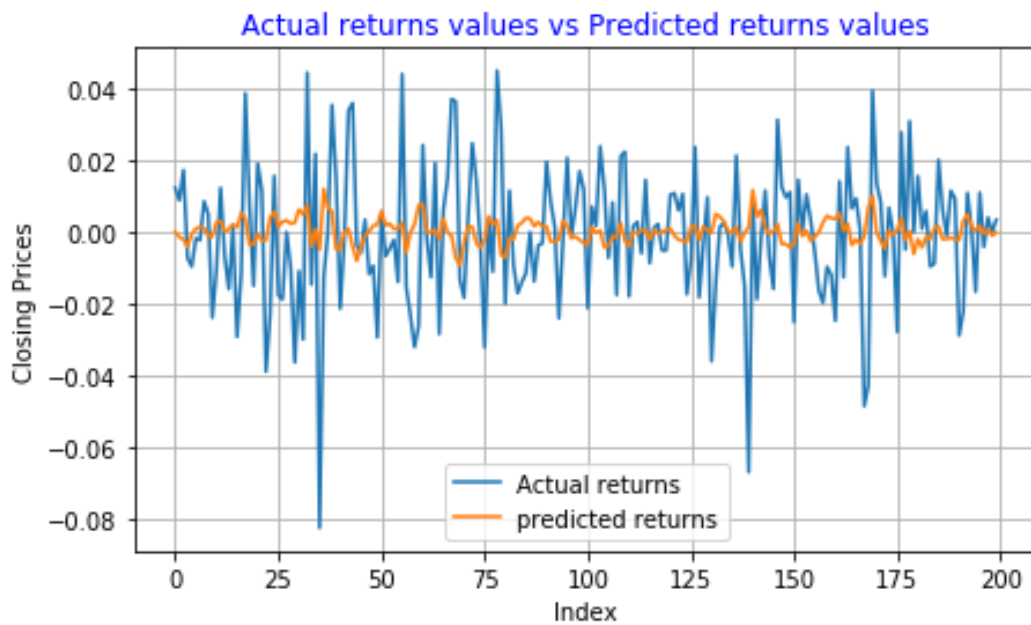
Figure 5.59: **Netcare Group Ltd returns series predictions using the AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	**Predicted Up**
True Down	71	33
True Up	25	71

Accuracy= 71%

Table 5.50: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Netcare Group Ltd returns series**

In Figure 5.59, the blue line indicates the true returns and the yellow line indicates the predictions. The variance of Netcare predictions is very high, and the model seems to have captured the volatility clusters well. Summary results of the confusion matrix for AR(3)-AVGARCH(1,1)-Student-t model prediction results is presented in Table 5.50. The model has a "very good" ability to predict the direction of movement of Netcare Group Ltd returns prices.

Santam Ltd: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                        AR - AVGARCH Model Results
==================================================================================
Dep. Variable:                        Close   R-squared:
Mean Model:                              AR    Adj. R-squared:
Vol Model:                          AVGARCH    Log-Likelihood:                 1
Distribution:       Standardized Student's t   AIC:                           -3
Method:                   Maximum Likelihood    BIC:                           -3
                                               No. Observations:
Date:                      Wed, Jan 24 2018    Df Residuals:
Time:                              17:45:50    Df Model:
                               Mean Model
==================================================================================
               coef      std err         t        P>|t|        95.0% Conf. Int.
----------------------------------------------------------------------------------
Const      3.4273e-05   4.331e-04   7.914e-02      0.937  [-8.145e-04,8.830e-04]
Close[1]      -0.0884   4.234e-02      -2.087  3.691e-02    [ -0.171,-5.368e-03]
Close[2]      -0.0204   3.570e-02      -0.570      0.569  [-9.031e-02,4.961e-02]
Close[3]   -3.5617e-03  3.437e-02      -0.104      0.917  [-7.092e-02,6.380e-02]
                            Volatility Model
==================================================================================
               coef      std err         t        P>|t|        95.0% Conf. Int.
----------------------------------------------------------------------------------
omega      1.8598e-03   8.814e-04       2.110  3.485e-02  [1.323e-04,3.587e-03]
alpha[1]      0.1540   4.961e-02        3.105  1.903e-03    [5.680e-02,   0.251]
beta[1]       0.7791   7.512e-02       10.371  3.360e-25    [  0.632,   0.926]
                              Distribution
==================================================================================
               coef      std err         t        P>|t|   95.0% Conf. Int.
----------------------------------------------------------------------------------
nu            3.8586       0.643       6.000  1.972e-09  [  2.598,   5.119]
==================================================================================
```

Figure 5.60: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Santam returns series**

As seen in Figure 5.60, the estimated coefficients of the Mean model have extremely small standard errors, this suggests that these estimated parameters are close to the true parameters. The constant, second and third parameters of the Mean model are statistically insignificant since their t-statistics values are less than 2, this suggest that these parameters do not provide much information in the model. On the other hand, the t-statistics for the estimated parameters of the volatility model are all greater than 2, this indicates that they are significantly different from 0 (they provide useful information in the model). The estimated AR(3)-AVGARCH(1,1)-Student-t model for the Santam returns series is

$$r_t = 0.000034273 + (-0.0884)r_{t-1} + (-0.0204)r_{t-2} + (-0.0035617)r_{t-3} + \epsilon_t,$$
$$(5.27)$$

where ϵ_t is a Student-t distributed process with 3.86 degrees of freedom. The conditional variance equation is

$$\sigma_t^2 = 0.0018598 + 0.1540 \mid \epsilon_{t-1}^2 \mid +0.7791\sigma_{t-1}^2. \qquad (5.28)$$

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.51:   **Diagnostics on standardised residuals of the AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
18.2346 (0.5720)	2.0316	13.6354 (0.8847)

Notes: p-values of the tests are in parentheses

Table 5.51 shows summary results of diagnostics tests carried out on standardised residuals from the AR(3)-AVGARCH(1,1)-Student-t model fitted on Santam Ltd returns series. Since the p-value for the Lagrange Multiplier test is greater than the 0.05 level of significance, we cannot reject the null hypothesis that there are no ARCH effects in the residuals. The null hypothesis that the standardised residuals are uncorrelated at 5% significance level is accepted since the DW test statistic value (2.0316) is close to 2. Large p-value for the Ljung Box test statistic indicates that we cannot reject the null hypothesis that the standardised residuals are independent.
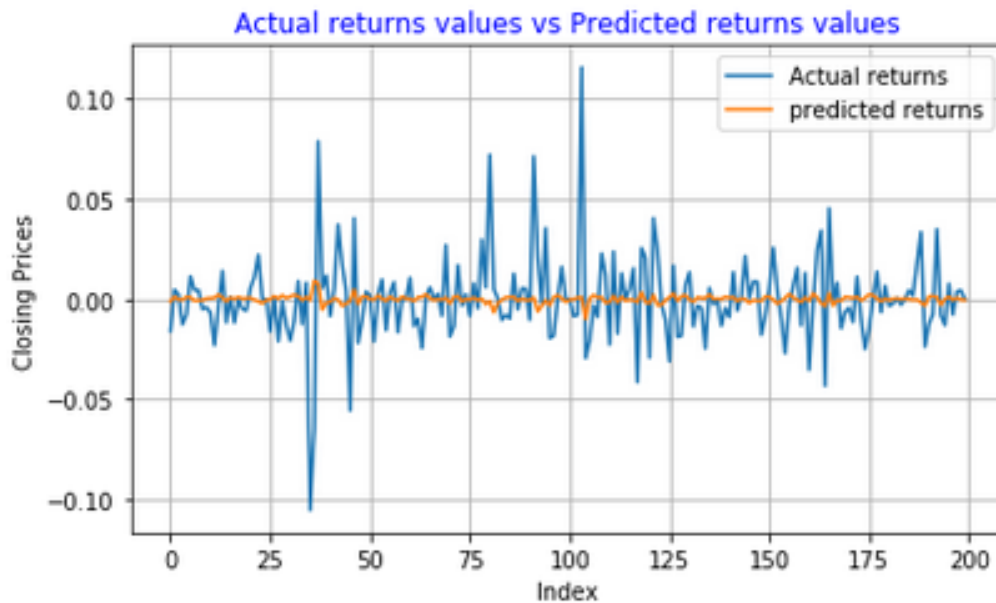
Figure 5.61: **Santam Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	**Predicted Up**
True Down	70	26
True Up	30	74

Accuracy= 72%

Table 5.52: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Santam Ltd returns series**

Figure 5.61 shows Santam Ltd returns series predictions obtained using AR(3)-AVGARCH(1,1)-Student-t model. The predictions from the model have low variance and the model seems to be operating near the mean. The confusion matrix for AR(3)-AVGARCH(1,1)-Student-t model predictions results is presented in Figure 5.54. The model achieved an accuracy of 72%, this shows that the model has a "fairly" strong ability to correctly predict the direction of the returns prices.

Sanlam Group Ltd: Estimated AR(3)-AVGARCH(1,1.)-Student-t model

```
                       AR - AVGARCH Model Results
==================================================================
Dep. Variable:                        Close   R-squared:
Mean Model:                              AR    Adj. R-squared:
Vol Model:                          AVGARCH    Log-Likelihood:
Distribution:     Standardized Student's t     AIC:                    -
Method:                 Maximum Likelihood      BIC:                    -
                                                No. Observations:
Date:                     Wed, Jan 24 2018      Df Residuals:
Time:                            17:42:44       Df Model:
                             Mean Model
==================================================================
                 coef    std err        t      P>|t|      95.0% Conf. Int.
------------------------------------------------------------------
Const       6.8512e-04  5.905e-04    1.160      0.246 [-4.723e-04,1.842e-03]
Close[1]       -0.0850  3.684e-02   -2.307  2.104e-02  [ -0.157,-1.279e-02]
Close[2]       -0.0181  4.214e-02   -0.430      0.667  [ -0.101,6.449e-02]
Close[3]    8.7625e-03  4.015e-02    0.218      0.827 [-6.993e-02,8.746e-02]
                          Volatility Model
==================================================================
                 coef    std err        t      P>|t|      95.0% Conf. Int.
------------------------------------------------------------------
omega       6.3434e-04  3.646e-04    1.740  8.191e-02 [-8.032e-05,1.349e-03]
alpha[1]       0.0590   2.281e-02    2.587  9.685e-03   [1.430e-02,  0.104]
beta[1]        0.9175   3.520e-02   26.067 8.619e-150   [  0.849,  0.987]
                            Distribution
==================================================================
                 coef    std err        t      P>|t|  95.0% Conf. Int.
------------------------------------------------------------------
nu             9.1677      2.800     3.274  1.061e-03 [  3.679, 14.656]
==================================================================
```

Figure 5.62: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Sanlam returns series**

Figure 5.62 shows summary results of the AR(3)-AVGARCH(1,1)-Student-t model fitted on Sanlam Group Ltd returns series. The standard errors for both the Mean (AR(3)) and the Volatility (AVGARCH(1,1)-Student-t) models are extremely small, this suggests that the estimated parameters for both of these models are close to the true parameters. The constant, the second, and the third parameters of the Mean model are statistical insignificant (i.e they do not provide significant information in the model). The degrees of freedom parameter ν is 9.17, this suggests that the Student-t distribution is an appropriate distribution for the returns series. The estimated AR(3)-AVGARCH(1,1)-Student-t model for Sanlam Group Ltd returns series is

$$r_t = 0.00068512 + (-0.0850)r_{t-1} + (-0.0181)r_{t-2} + (0.0087625)r_{t-3} + \epsilon_t,$$
$$(5.29)$$

where ϵ_t is a Student-t distributed process with $\nu = 9.17$ degrees of freedom. The conditional variance is given as

$$\sigma_t^2 = 0.00063434 + 0.0590 \mid \epsilon_{t-1}^2 \mid + 0.9175\sigma_{t-1}^2. \qquad (5.30)$$

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.53: **Diagnostics on the standardised residuals of the AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
34.5603 (0.0226)	1.9948	23.1133 (0.3371)

Notes: p-values of the tests are in parentheses

We reject the null hypothesis of no ARCH effects at 5% significance level since the p-value for the Lagrange Multiplier test is 0.0226. A DW test statistic value close to 2 is an indication that there are no autocorrelation in the standardised residuals from the fitted AR(3)-AVGARCH(1,1)-Student-t model. We fail to reject the null hypothesis that the standardised residuals are independent since the p-value (about 0.3371) for the Ljung box test statistic is is greater than the 0.05 significance level.
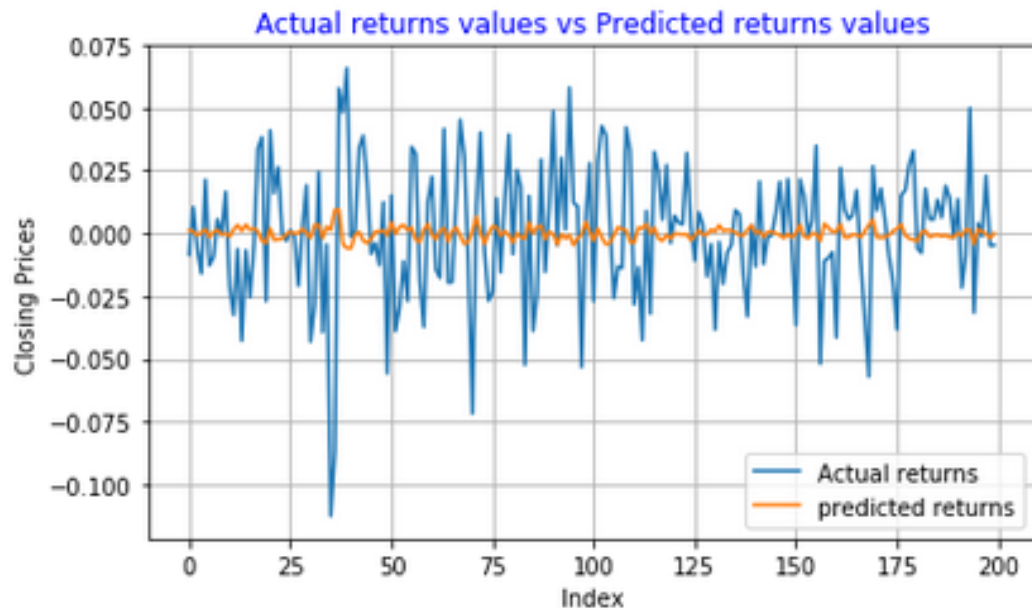
Figure 5.63: **Sanlam Group Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	Predicted Up
True Down	71	30
True Up	30	69

Accuracy= 70%

Table 5.54: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Sanlam Group Ltd returns series**

Figure 5.63 shows AR(3)-AVGARCH(1,1)-Student-t model's Sanlam Group Ltd returns predictions. These predictions looks similar to those in Figure 5.61. The model achieved an accuracy of 70%, this shows that the model has an "adequate" ability to predict the direction of Sanlam Group Ltd returns series.

Nedbank Group Ltd: Estimated AR(3)-AVGARCH(1,1)-Student-t model

```
                          AR - AVGARCH Model Results
==================================================================================
Dep. Variable:                        Close   R-squared:
Mean Model:                              AR    Adj. R-squared:
Vol Model:                          AVGARCH    Log-Likelihood:
Distribution:       Standardized Student's t   AIC:                             -
Method:                  Maximum Likelihood     BIC:                             -
                                                No. Observations:
Date:                     Wed, Jan 24 2018     Df Residuals:
Time:                             17:50:14     Df Model:
                                Mean Model
==================================================================================
               coef     std err          t      P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------------
Const      5.0487e-04  5.078e-04      0.994      0.320 [-4.905e-04,1.500e-03]
Close[1]      -0.1426  4.177e-02     -3.413  6.416e-04 [ -0.224,-6.071e-02]
Close[2]      -0.0793  3.739e-02     -2.122  3.387e-02 [ -0.153,-6.044e-03]
Close[3]      -0.0354  3.985e-02     -0.888      0.374 [ -0.114,4.271e-02]
                             Volatility Model
==================================================================================
               coef     std err          t      P>|t|       95.0% Conf. Int.
----------------------------------------------------------------------------------
omega      1.0823e-03  5.135e-04      2.107  3.507e-02 [7.576e-05,2.089e-03]
alpha[1]      0.0586  2.138e-02      2.740  6.148e-03  [1.667e-02,  0.100]
beta[1]       0.8794  3.946e-02     22.289 4.677e-110  [  0.802,  0.957]
                                Distribution
==================================================================================
               coef     std err          t      P>|t|  95.0% Conf. Int.
----------------------------------------------------------------------------------
nu           11.6752      4.791      2.437  1.482e-02 [  2.285, 21.066]
==================================================================================
```

Figure 5.64: **Summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Nedbank returns series**

Figure 5.64 shows summary results of the estimated AR(3)-AVGARCH(1,1)-Student-t model fitted on Nedbank Group Ltd returns series. Since the standard errors for both the Mean and Volatility models are small, this suggests that the estimated coefficients for these models are close to the true parameters. All the estimated parameters lies within the 95% confidence interval, so we are 95% confident that the true parameters are within the computed intervals. The constant and third parameter of the Mean model are statistically insignificant (they do not provide significant information in the model) since their t-statistics are less than 2 (critical value). The estimated AR(3)-AVGARCH(1,1)-Student-t model is

$$r_t = 0.00050487 + (-0.1426)r_{t-1} + (-0.0793)r_{t-2} + (-0.0354)r_{t-3} + \epsilon_t,$$

(5.31)

where the residuals (error) term ϵ_t is a Student-t distributed process with $\nu = 11.68$ (degrees of freedom). The conditional variance equation is given as

$$\sigma_t^2 = 0.0010823 + 0.0586 \mid \epsilon_{t-1}^2 \mid +0.8794\sigma_{t-1}^2. \tag{5.32}$$

Post-estimation diagnostics for the fitted AR(3)-AVGARCH(1,1)-Student-t model

Table 5.55: **Diagnostics on the standardised residuals of the AR(3)-AVGARCH(1,1)-Student-t model**

ARCH effects	Durbin Watson	Ljung Box
23.5574 (0.2623)	2.0074	32.6364 (0.0504)

Notes: p-values of the tests are in parentheses

There are no ARCH effects remaining in the residuals since the p-value for the Lagrange Multiplier test statistic is greater than the 0.05 significance level. The p-value for the Ljung Box test statistic is 0.0504, this value is "slightly" greater than the 0.05 level of significance suggesting that the null hypothesis of independence of residuals is "barely" accepted. The DW test statistic is 2.0074, this value is extremely close to 2, implying that the standardised residuals are independent.
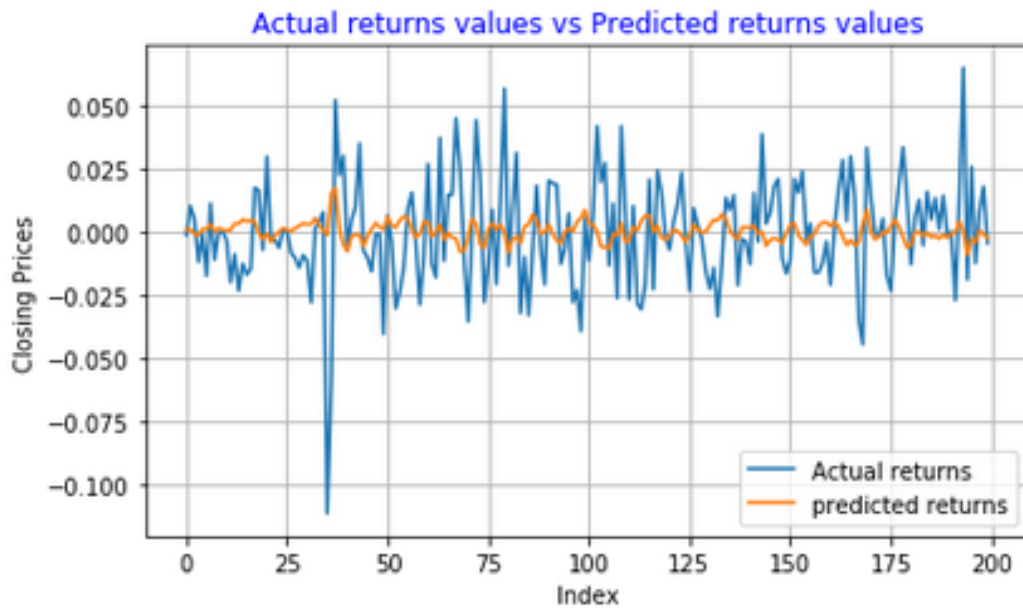
Figure 5.65: **Nedbank Group Ltd returns series predictions using AR(3)-AVGARCH(1,1)-Student-t model**

	Predicted Down	Predicted Up
True Down	70	30
True Up	29	71
Accuracy= 70.5%		

Table 5.56: **Confusion Matrix for AR(3)-AVGARCH(1,1)-Student-t model's predictions of Nedbank Group Ltd returns series**

The variance of Nedbank Group Ltd returns series predictions is slightly higher than that in Figure 5.61, and 5.63. The model has adequately captured the volatility clusters in the Nedbank returns series. Confusion matrix in Figure 5.56 shows that in terms of predicting the direction of movement of Nedbank Group Ltd returns series, the model achieved an accuracy of 70.5%.

Table 5.57: **Summary accuracy results of AR(3)-GARCH type models**

Prediction accuracy AR(3)-GARCH type models			
Stock	**Model**	**Distribution**	**Accuracy (%)**
Netcare	AR(3)-GARCH(1,1)	Normal	73%
	AR(3)-EGARCH(1,1)	Student-t	71.5%
	AR(3)-EWMA	Normal	72.5%
	AR(3)-EWMA	Student-t	72.5%
	AR(3)-TARCH(1,1,1)	Normal	70.5%
	AR(3)-TARCH(1,1,1)	Student-t	72.5%
	AR(3)-AVGARCH(1,1)	Normal	70.5%
	AR(3)-AVGARCH(1,1)	Student-t	75%
Santam	AR(3)-GARCH(1,1)	Normal	68.5%
	AR(3)-EGARCH(1,1)	Student-t	71.5%
	AR(3)-EWMA	Normal	71%
	AR(3)-EWMA	Student-t	68.5%
	AR(3)-TARCH(1,1,1)	Normal	72%
	AR(3)-TARCH(1,1,1)	Student-t	71.5%
	AR(3)-AVGARCH(1,1)	Normal	72%
	AR(3)-AVGARCH(1,1)	Student-t	72%
Sanlam	AR(3)-GARCH(1,1)	Normal	68.5%
	AR(3)-EGARCH(1,1)	Student-t	66%
	AR(3)-EWMA	Normal	68.5%
	AR(3)-EWMA	Student-t	68%
	AR(3)-TARCH(1,1,1)	Normal	69.5%
	AR(3)-TARCH(1,1,1)	Student-t	69%
	AR(3)-AVGARCH(1,1)	Normal	70%
	AR(3)-AVGARCH(1,1)	Student-t	70%
Nedbank	AR(3)-GARCH(1,1)	Normal	71%
	AR(3)-EGARCH(1,1)	Student-t	71%
	AR(3)-EWMA	Normal	71%
	AR(3)-EWMA	Student-t	70%
	AR(3)-TARCH(1,1,1)	Normal	71.5%
	AR(3)-TARCH(1,1,1)	Student-t	71%
	AR(3)-AVGARCH(1,1)	Normal	71.5%
	AR(3)-AVGARCH(1,1)	Student-t	70.5%

Table 5.57 shows summary results of predictions accuracy of various AR(3)-GARCH type models. For the Netcare Group Ltd returns series, the AR(3)-AVGARCH(1,1)-Student-t model achieved the highest accuracy (about 75%), this shows that the model has a "good ability" to predict the direction of

movement of the returns series. Both the AR(3)-AVGARCH(1,1)-Student-t model and the AR(3)-AVGARCH(1,1)-Normal model achieved the highest accuracy (about 72%) for Santam returns series. For the Sanlam Group Ltd returns series, the highest accuracy (about 70%) was achieved by the AR(3)-AVGARCH(1,1)-Student-t, and AR(3)-AVGARCH(1,1)-Normal models. In terms of predicting the direction of movement of Nedbank Group Ltd returns series, the highest accuracy achieved was 71.5% by the AR(3)-AVGARCH(1,1)-Normal and AR(3)-TARCH(1,1)-Normal models.

Discussion:

AR(3)-GARCH type models were fitted on various JSE returns series using steps suggested by (Shumway and Stoffer, 2006), and (Tsay, 2005). Summary accuracy results presented in Table 5.57 shows that the AR(3)-GARCH type models have a "strong" ability to correctly predict the direction of movement of the returns series. However, the models can be further improved by including the moving average component in the mean model.

5.8 State Space ARMA(1,1) Model results, Train:700 Test:200

Netcare Group Limited (Ltd): Estimated State Space ARMA(1,1) model

```
                         Statespace Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  700
Model:                         ARMA11   Log Likelihood                1896.196
Date:                Fri, 26 Jan 2018   AIC                          -3786.392
Time:                        16:16:58   BIC                          -3772.739
Sample:                             0   HQIC                         -3781.115
                                - 700
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.3854      0.205      1.879      0.060      -0.017       0.787
theta         -0.5182      0.189     -2.743      0.006      -0.889      -0.148
sigma2         0.0003   1.04e-05     24.910      0.000       0.000       0.000
===================================================================================
Ljung-Box (Q):                       37.32   Jarque-Bera (JB):            79.03
Prob(Q):                              0.59   Prob(JB):                     0.00
Heteroskedasticity (H):               1.35   Skew:                        -0.01
Prob(H) (two-sided):                  0.02   Kurtosis:                     4.65
===================================================================================
```

Figure 5.66: **Summary results of the State Space ARMA(1,1) fitted on Netcare Group Ltd returns series**

Figure 5.66 shows summary results of the estimated State Space ARMA(1,1) model fitted on Netcare Group Ltd returns series. Both the estimated Autoregressive (phi) and Moving Average (theta) parameters lies within the computed intervals, this suggests that the true parameters also lie within this intervals. The standard errors are relatively small, this is a clear indication that the estimated parameters are close to the true parameters. The Moving Average (theta) and the Variance (sigma2) parameters are statistically significant since their Z-statistics are greater than 1.96 (critical value). On the other hand, the Autoregressive parameter is statistically insignificant since its Z-statistic is "slightly" less than the 1.96 critical value. The p-value for the Ljung Box test statistic is 0.59, this value is extremely large, suggesting that we have weak evidence against the null hypothesis of independence of the residuals. The residuals from the fitted State Space ARMA(1,1) model are not normally distributed since the p-value for the Jarque Bera

test is less than 0.05 significance level. The p-value for the Heteroscedasticity test is 0.02, this value is less than the 0.05 level of significance, suggesting that we reject the null hypothesis that the residuals have a constant variance. The distribution of the residuals has "heavier" tails and is "slightly" skewed to the left since it has a kurtosis value of 4.65 and a negative skewness value. Since the normality and constant variance assumptions are violated, this suggests that an ARMA model which include a conditional variance process such as GARCH-Student-t process would be more appropriate for modelling the Netcare returns series. The estimated State Space ARMA(1,1) model is

$$r_t = 0.3854r_{t-1} + (-0.5182)\epsilon_{t-1} + \epsilon_t, \tag{5.33}$$

where $\epsilon_t \sim N(\mu, \sigma^2 = 0.0003)$.
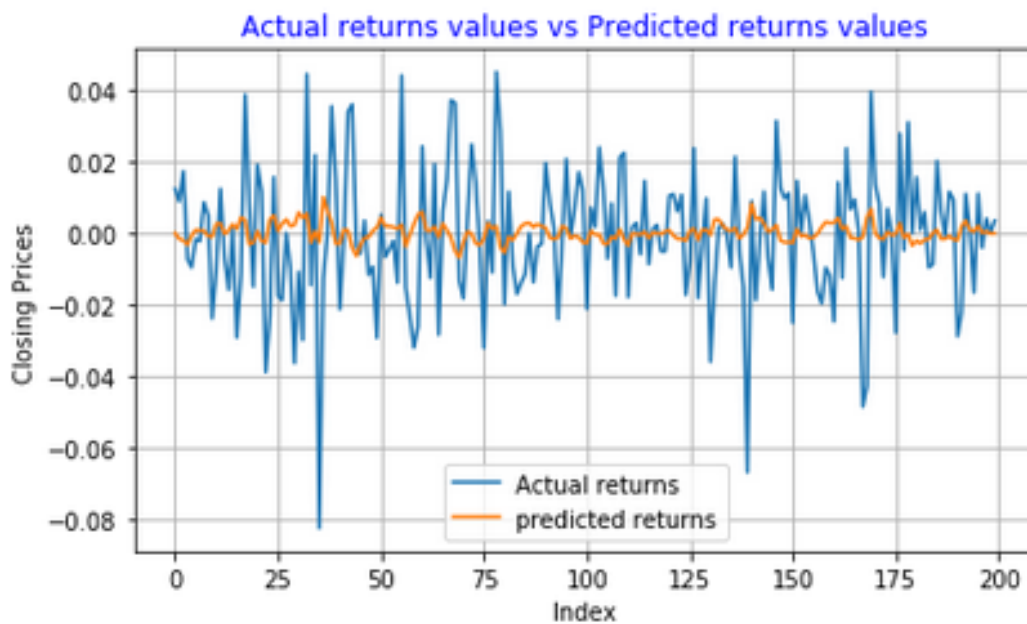
Figure 5.67: **Netcare Group Ltd returns series predictions using the State Space ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	78	26
True Up	25	71

Accuracy= 74.5%

Table 5.58: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Netcare returns series**

Figure 5.67 shows the actual (true) and predicted returns obtained using State Space ARMA(1,1) model. Confusion matrix in Table 5.58 shows prediction accuracy of the State Space ARMA(1,1) model . The model achieved an accuracy of 74.5%, so it has a "strong" ability to predict the direction of movement of Netcare Group returns prices.

Santam Ltd: Estimated State Space ARMA(1,1) model

```
                         Statespace Model Results
==============================================================================
Dep. Variable:                        y   No. Observations:                  700
Model:                           ARMA11   Log Likelihood                1915.653
Date:                  Fri, 26 Jan 2018   AIC                          -3825.306
Time:                          16:23:37   BIC                          -3811.653
Sample:                               0   HQIC                         -3820.028
                                  - 700
Covariance Type:                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.5368      0.258      2.082      0.037       0.031       1.042
theta         -0.6020      0.250     -2.408      0.016      -1.092      -0.112
sigma2         0.0002    7.9e-06     31.098      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                       24.92   Jarque-Bera (JB):           378.43
Prob(Q):                              0.97   Prob(JB):                     0.00
Heteroskedasticity (H):               0.98   Skew:                         0.00
Prob(H) (two-sided):                  0.85   Kurtosis:                     6.60
==============================================================================
```

Figure 5.68: **Summary results of the State space ARMA(1,1) model fitted on Santam Ltd returns series**

The standard errors for the estimated parameters are fairly small, this suggests that these parameters are close to the true values. The estimated parameters are all significantly different from zero since their Z-statistics are greater than the 1.96 critical value. The null hypothesis that residuals are independent at 5% significance level is accepted since the p-value for the

Ljung Box test statistic (about 0.97) is greater than 0.05. The Heteroscedasticity test is inconclusive since the p-value of the test is equal to the significance level (0.05). The Jarque Bera test statistic value is extremely large (about 378), this indicates that we cannot accept the null hypothesis that the residuals are normally distributed. A skewness value of 0.00 suggests that the distribution of residuals is symmetric. The distribution of the residuals has "heavier" tails than that of a normal distribution since its kurtosis value is larger than 3. Due to violations of some of the assumptions of the ARMA modelling process, an ARMA model which incorporate a conditional GARCH process would be more appropriate for modelling Santam returns series as the model would be able to capture the statistical dynamics of the series. The estimated State Space ARMA(1,1) model for Santam Ltd returns series is

$$r_t = 0.5368r_{t-1} + (-0.6020)\epsilon_{t-1} + \epsilon_t, \tag{5.34}$$

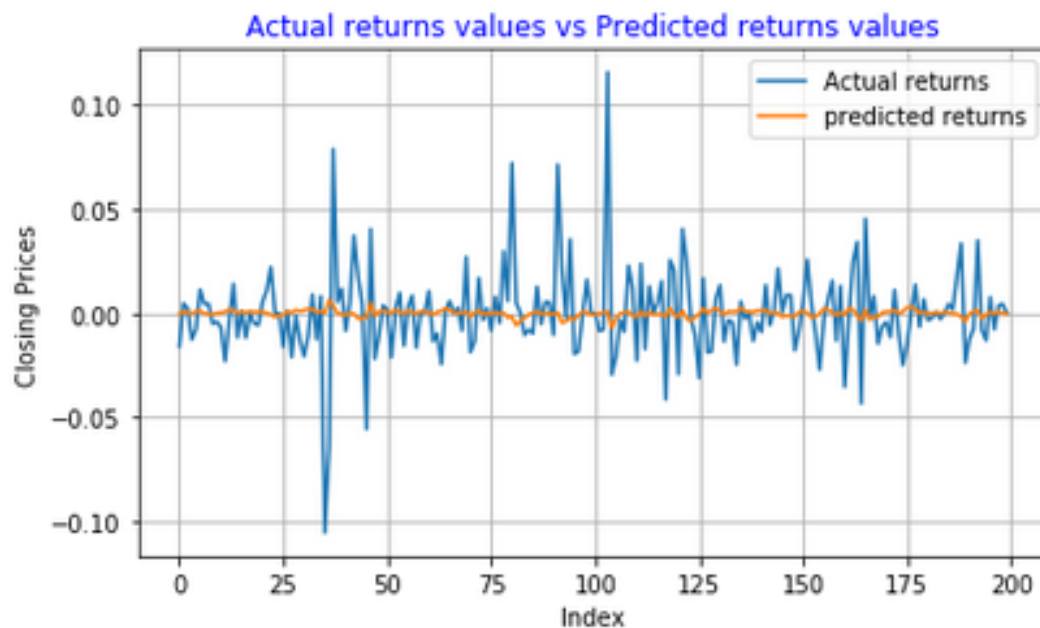where the error term $\epsilon_t \sim N(\mu, \sigma^2 = 0.0002)$

Figure 5.69: **Santam Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	72	24
True Up	28	76

Accuracy= 74%

Table 5.59: **Confusion Matrix for the State Space ARMA(1,1) model's predictions of Santam returns series**

As seen in Figure 5.69, the variance of the predictions made by the State Space ARMA(1,1) model is very small. (Hansson, 2017) also observed that when predicting the log returns series, Box and Jenkins ARMA-type models operates around the mean. In terms of predicting the direction of movement of Santam returns prices, the model achieved an accuracy of 74%.

Sanlam Group Ltd: Estimated State Space ARMA(1,1) model

```
                        Statespace Model Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                  700
Model:                         ARMA11   Log Likelihood                1863.130
Date:                Fri, 26 Jan 2018   AIC                          -3720.260
Time:                        16:30:19   BIC                          -3706.606
Sample:                     01-13-2003   HQIC                         -3714.982
                          - 10-15-2021
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.5164      0.317      1.631      0.103      -0.104       1.137
theta         -0.5832      0.297     -1.961      0.050      -1.166      -0.000
sigma2         0.0003   1.24e-05     22.966      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                       47.19   Jarque-Bera (JB):            34.09
Prob(Q):                              0.20   Prob(JB):                     0.00
Heteroskedasticity (H):               1.26   Skew:                        -0.05
Prob(H) (two-sided):                  0.08   Kurtosis:                     4.08
==============================================================================
```

Figure 5.70: **Summary results of State Space ARMA(1,1) model fit on Sanlam Group Ltd returns series**

Figure 5.70 shows summary results of the estimated State Space ARMA(1,1) model fitted on Sanlam Group Ltd returns series. Small standard error are an indication that the estimated parameters are close to the true parameters. The Z-statistics for the Moving Average (theta) and Variance (sigma2) parameters are greater than the 1.96 critical value, this suggests that these

parameters are significantly different from zero. On the other hand, the Autoregressive parameter (phi) is statistically insignificant since its Z-statistic value is less than the 1.96 (critical value). The null hypothesis that the residuals are independent at 5% significance level is accepted since the p-value for the Ljung Box test statistic is greater than 0.05 (level of significance). The residuals of the fitted State Space ARMA(1,1) have a constant variance since the p-value for the Heteroscedasticity test is greater than 0.05 significance level. Results of the Jarque-Bera test suggests that the residuals are not normally distributed since the p-value of the test statistic is less than 0.05 (level of significance). The estimated State Space ARMA(1,1) model for Sanlam Group Ltd returns series is

$$r_t = 0.5164r_{t-1} + (-0.5832)\epsilon_{t-1} + \epsilon_t, \tag{5.35}$$

where ϵ_t is a normal distributed process with mean μ and variance σ^2, that is, $\epsilon_t \sim N(\mu, \sigma^2)$.
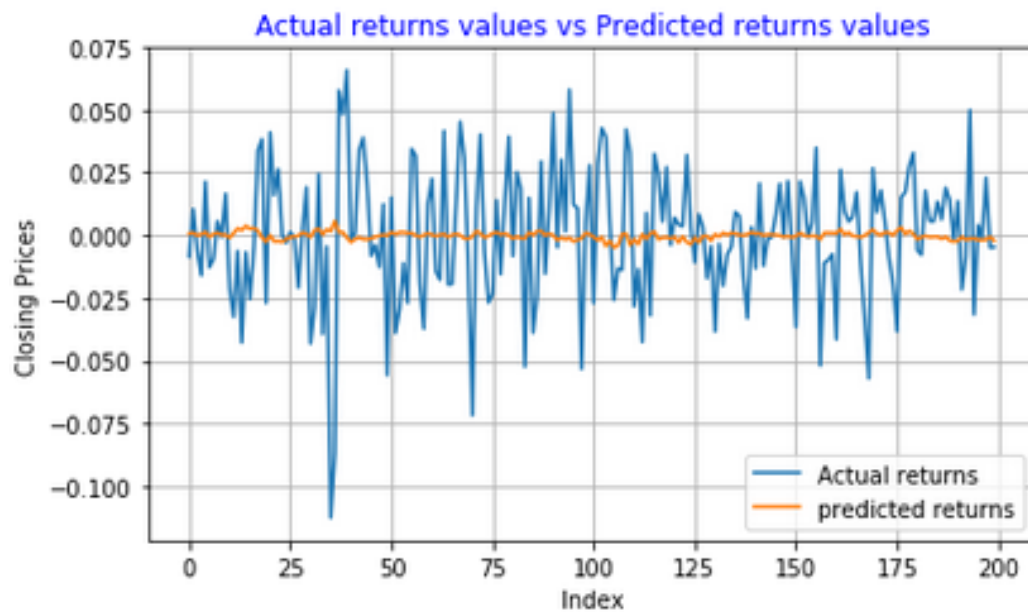

Figure 5.71: **Sanlam Group Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	Predicted Up
True Down	62	39
True Up	36	63

Accuracy= 62.5%

Table 5.60: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Sanlam returns series**

Plot in Figure 5.71 shows that the predictions from the State Space ARMA(1,1) model are extremely close to the mean and they have low or small variance. The predictions looks similar to those in Figure 5.69. The accuracy achieved by the model is very poor (about 62.5%). The model is as good as a random walk.

Nedbank Group Ltd: Estimated State Space ARMA(1,1) model

```
                        Statespace Model Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                  700
Model:                         ARMA11   Log Likelihood                1972.910
Date:                Fri, 26 Jan 2018   AIC                          -3939.819
Time:                        16:37:40   BIC                          -3926.166
Sample:                     01-13-2003   HQIC                         -3934.541
                           - 10-15-2021
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
phi            0.4628      0.138      3.349      0.001       0.192       0.734
theta         -0.6247      0.123     -5.077      0.000      -0.866      -0.384
sigma2         0.0002   9.24e-06     22.572      0.000       0.000       0.000
==============================================================================
Ljung-Box (Q):                        56.22   Jarque-Bera (JB):          26.80
Prob(Q):                               0.05   Prob(JB):                   0.00
Heteroskedasticity (H):                1.60   Skew:                       0.05
Prob(H) (two-sided):                   0.00   Kurtosis:                   3.95
==============================================================================
```

Figure 5.72: **Summary results of State Space ARMA(1,1) model fitted on Nedbank Group Ltd returns series**

The estimated coefficients for the Autoregressive, Moving Average, and the Variance parameters are 0.4628, -0.6247, and 0.0002 respectively. All these estimated coefficients lies within the confidence intervals, suggesting that the true parameters are also within these intervals. All the estimated State Space ARMA(1,1) model's parameters are statistically significant since their

Z-statistics are greater than the 1.96 critical value. The Ljung Box test for independence of residuals is inconclusive since the p-value of the test is equal to the 0.05 level of significance. The Jarque-Bera test rejects the null hypothesis that the residuals are normally distributed at 5% significance level. Residuals from the fitted State Space ARMA(1,1) model do not have a constant variance since the p-value for the Heteroscedasticity test is less than the 0.05 level of significance. Consequently, since the normality and the heteroscedasticity test are violated, this suggests that an ARMA model with a conditional variance process (such as the GARCH or EGARCH) should be used, so that the dynamics of the returns series can be better captured. The estimated State Space ARMA(1,1) model for the Nedbank Group Ltd returns series is

$$r_t = 0.4628r_{t-1} + (-0.6247)\epsilon_{t-1} + \epsilon_t, \tag{5.36}$$
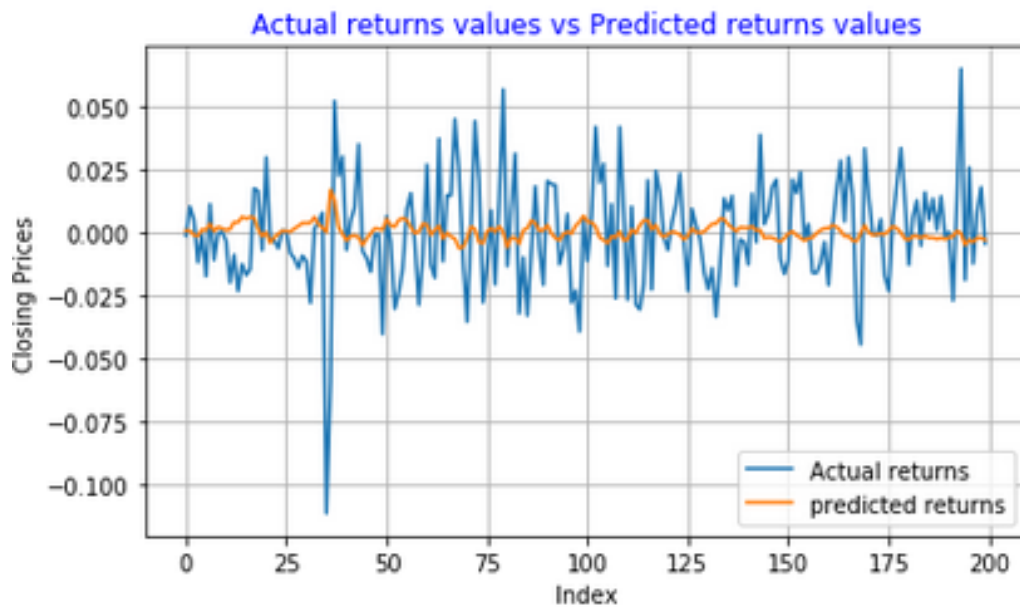
where $\epsilon_t \sim N(\mu, \sigma^2 = 0.0002)$.

Figure 5.73: **Nedbank Group Ltd returns series predictions using State Space ARMA(1,1) model**

	Predicted Down	**Predicted Up**
True Down	72	28
True Up	33	67

Accuracy= 69.5%

Table 5.61: **Confusion Matrix for State Space ARMA(1,1) model's predictions of Nedbank returns series**

The predictions made by the State Space ARMA(1,1) model have higher variance than those in Figure 5.69, and 5.71. The model achieved an accuracy of 69.5% and a misclassification rate (percentage of incorrect predictions) of 30.5%.

Discussion

In some cases, the assumptions of the ARMA modelling process were not satisfied, this includes the Normality and Heteroscedasticty assumptions. Consequently, an ARMA model with a conditional variance process could be used to accurately capture the dynamics of the returns series. The advantage of including a conditional variance process in the ARMA process is that we can assume other statistical distribution such as Skewed Student-t, Student-t, and Generalised Error Distribution (GED) for the returns series. The highest accuracy achieved by the State Space ARMA(1,1) model is 74.5% for the Netcare returns series and the lowest is 62.5% for the Nedbank returns series (suggesting that this model is as good as a random walk for this particular series).

5.9 ARMA(1,1)-Metropolis Hastings (MH) algorithm Model Results, Train:700 Test:200

The Metropolis-Hastings ARMA(1,1) model was implemented using the following assumptions for the prior distributions: the Autoregressive (phi) and Moving Average (theta) parameters are assumed to have a Uniformly distributed prior whose support is (-1,1), that is , $U \sim (-1,1)$. The precision parameter is assumed to have a Gamma prior, that is, $1/\sigma^2 \sim Gamma(2,4)$.

Netcare Group Ltd: Estimated Metropolis-Hastings -ARMA(1,1) model

```
phi:

        Mean              SD              MC Error          95% HPD interval
        ------------------------------------------------------------------
        -0.032            0.495           0.025             [-0.863  0.845]


        Posterior quantiles:

        2.5              25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.867          -0.425         -0.033          0.36           0.844


precision:

        Mean              SD              MC Error          95% HPD interval
        ------------------------------------------------------------------
        85.843            4.591           0.125             [ 77.367  94.976]


        Posterior quantiles:

        2.5              25              50              75              97.5
         |--------------|==============|==============|--------------|
        76.633          82.795         85.94           89.018         94.703


theta:

        Mean              SD              MC Error          95% HPD interval
        ------------------------------------------------------------------
        -0.069            0.47            0.024             [-0.892  0.765]


        Posterior quantiles:

        2.5              25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.855          -0.449         -0.08           0.288          0.831
```

Figure 5.74: **Summary results of ARMA(1,1)- Metropolis Hastings model fitted on Netcare Group Limited returns series**

Figure 5.74 shows summary results of the estimated Metropolis-Hastings ARMA(1,1) model fitted on Netcare Group Ltd returns series. The estimated Autoregressive, Moving Average, and precision parameters are $-0.032, -0.069$, and 85.843 respectively. The standard deviation (SD) for the precision parameter is relatively large (about 4.591), this suggests that there is some degree of uncertainty in the estimated precision parameter. Since the Monte

Carlo (MC) errors for all the estimated parameters are small, this is a clear indication that there is a small amount of error in the estimates.
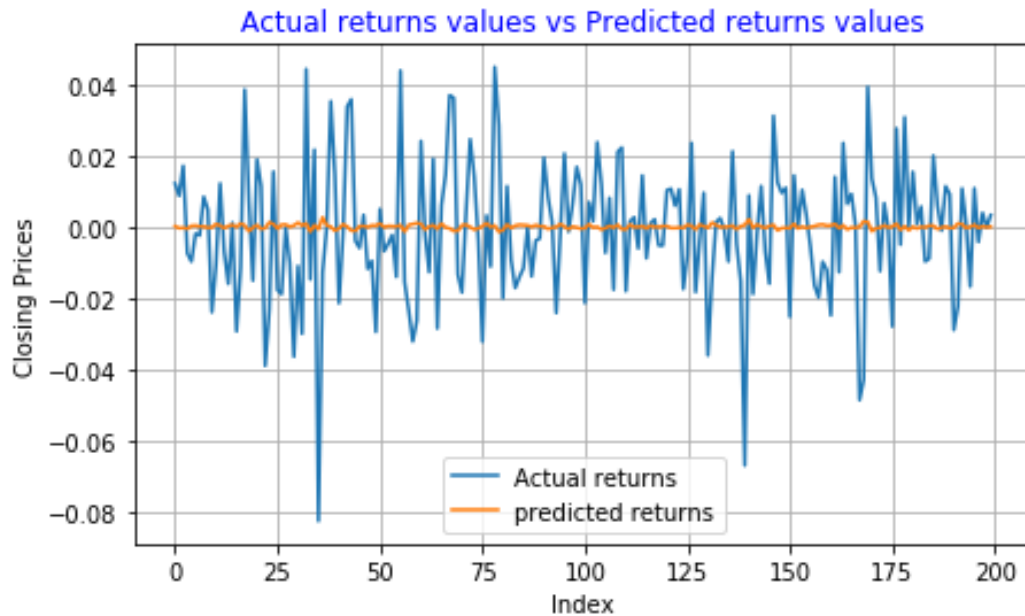

Figure 5.75: **Netcare Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	65	39
True Up	30	66

Accuracy= 65.5%

Table 5.62: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Netcare Group Ltd returns series**

Figure 5.75 presents predictions of Netcare Group Ltd returns series obtained using the ARMA(1,1)-Metropolis-Hastings model. Yellow line indicates the predicted returns and blue line indicates the true returns. The structure of the predictions is similar to that in Figure 5.31 (they have small variance). The model did not adequately capture the variance in the Netcare series (although this does not necessarily imply that we will have poor forecasts). Accuracy results of the ARMA(1,1)-MH model's predictions are

presented in Table 5.62. The model achieved an accuracy of 65%, hence the model is almost as good as a random walk model.

Santam Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
precision:

        Mean              SD            MC Error        95% HPD interval
        ------------------------------------------------------------------
        85.69             4.831         0.193           [ 76.189  95.444]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |--------------|==============|==============|--------------|
        75.946          82.536          85.766          88.789          95.3

theta:

        Mean              SD            MC Error        95% HPD interval
        ------------------------------------------------------------------
        -0.055            0.496         0.024           [-0.912  0.827]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |--------------|==============|==============|--------------|
        -0.911          -0.46           -0.063          0.333           0.831

phi:

        Mean              SD            MC Error        95% HPD interval
        ------------------------------------------------------------------
        -0.004            0.517         0.025           [-0.882  0.896]


        Posterior quantiles:

        2.5             25              50              75              97.5
        |--------------|==============|==============|--------------|
        -0.896          -0.428          -0.013          0.42            0.891
```

Figure 5.76: **Summary results of the ARMA(1,1) Metropolis Hastings model fitted on Santam Ltd returns series**

The estimated precision parameter and its standard deviation are close to that in Figure 5.74. In Bayesian statistics, standard deviations represents uncertainty in the estimated parameters. The degree of uncertainty in the Autoregressive and Moving Average parameters is very low since their stan-

dard deviations are close to zero. All the estimated parameters lies within their 95% HPD interval.
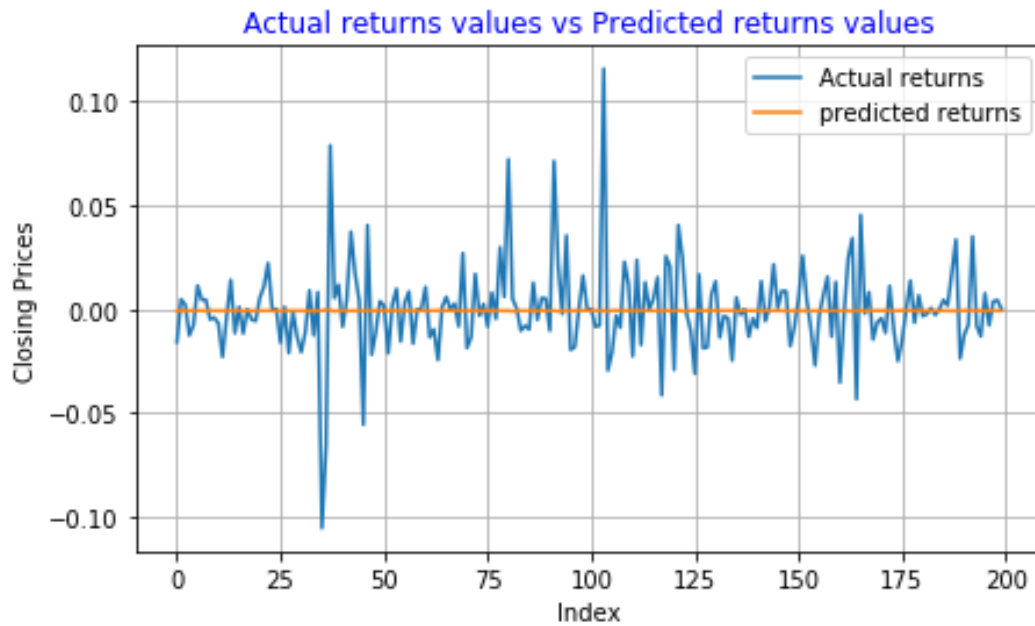

Figure 5.77: **Santam Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	69	27
True Up	37	67

Accuracy= 68%

Table 5.63: **Confusion Matrix for ARMA(1,1)-MH predictions of Santam Ltd returns series**

Figure 5.77 shows predictions from ARMA(1,1)-Metropolis Hastings model. In terms of the ability to predict the direction of Santam returns series, the ARMA(1,1)-MH model achieved an accuracy of 68%.

Sanlam Group Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
precision:

        Mean            SD              MC Error         95% HPD interval
        ------------------------------------------------------------------
        85.81           4.443           0.114            [ 77.673  94.962]


        Posterior quantiles:

        2.5             25              50               75            97.5
         |---------------|===============|===============|---------------|
        77.176          82.679          85.791          88.782        94.599

theta:

        Mean            SD              MC Error         95% HPD interval
        ------------------------------------------------------------------
        -0.035          0.485           0.025            [-0.918  0.797]


        Posterior quantiles:

        2.5             25              50               75            97.5
         |---------------|===============|===============|---------------|
        -0.887          -0.43           -0.036          0.346          0.84

phi:

        Mean            SD              MC Error         95% HPD interval
        ------------------------------------------------------------------
        -0.034          0.512           0.027            [-0.921  0.847]


        Posterior quantiles:

        2.5             25              50               75            97.5
         |---------------|===============|===============|---------------|
        -0.894          -0.448          -0.038          0.388          0.889
```

Figure 5.78: **Summary results of ARMA(1,1) Metropolis Hastings model fitted on Santam Group Ltd returns series**

The estimated precision parameter and its standard deviations are close to those in Figure 5.74, and 5.76. Since the MC errors for all the estimated model's parameters are close to zero, this suggests that the amount of error in these estimates is small. Since all the estimated parameters lies within the 95% HPD interval, then we are sure that the true parameters also lies within this interval.
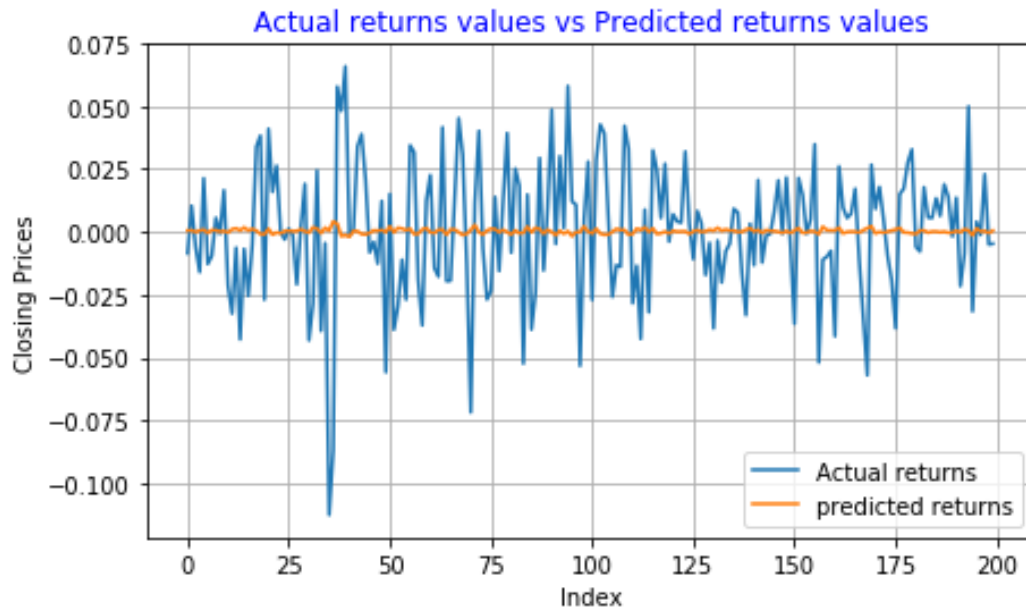
Figure 5.79: **Sanlam Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	66	35
True Up	32	67

<div align="center">**Accuracy**= 66.5%</div>

Table 5.64: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Sanlam Group Ltd returns series**

As seen in Figure 5.79, the variance of Sanlam returns series predictions seems to be less volatile and similar to that in Figure 5.75. The model achieved an accuracy of 66.5% and a misclassification rate of about 33.5%. This shows that the model has a "poor" ability to correctly predict the direction of movement of Sanlam returns series.

Nedbank Group Ltd: Estimated ARMA(1,1)-Metropolis Hastings model

```
precision:

        Mean            SD              MC Error        95% HPD interval
        ------------------------------------------------------------------
        86.146          4.639           0.106           [ 77.466  95.744]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        77.104          82.99          86.12           89.352          95.527

theta:

        Mean            SD              MC Error        95% HPD interval
        ------------------------------------------------------------------
        -0.049          0.48            0.025           [-0.891  0.795]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.847          -0.428         -0.081          0.338           0.844

phi:

        Mean            SD              MC Error        95% HPD interval
        ------------------------------------------------------------------
        -0.07           0.492           0.025           [-0.989  0.734]


        Posterior quantiles:

        2.5             25              50              75              97.5
         |--------------|==============|==============|--------------|
        -0.932          -0.458         -0.077          0.347           0.818
```

Figure 5.80: **Summary results of ARMA(1,1) Metropolis Hastings model fitted on Nedbank Group Ltd returns series**

A large standard deviation (SD) for the precision parameter (about 4.639) indicates that there is some degree of uncertainty in this parameter. On the other hand, the Monte Carlo error for this parameter is small, suggesting that the amount of error in this estimate is small. The Monte Carlo (MC) errors for the Autoregressive and Moving Average parameters are equal

(about 0.025), this value is extremely close to zero, this suggests that the amount of error in these estimates is very small.
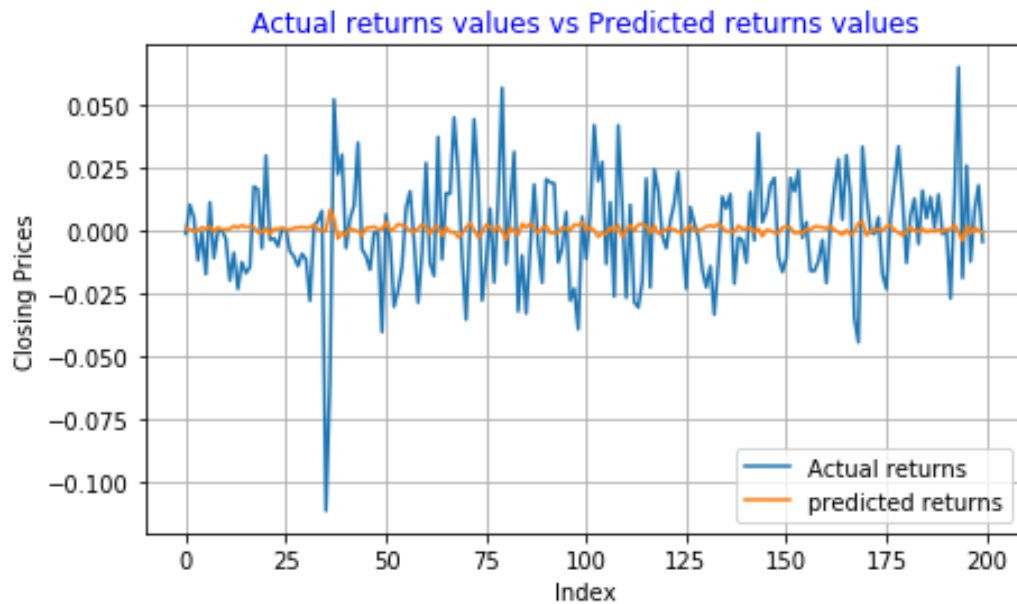

Figure 5.81: **Nedbank Group Ltd returns series predictions using ARMA(1,1)-MH model**

	Predicted Down	**Predicted Up**
True Down	65	35
True Up	35	65

Accuracy= 65%

Table 5.65: **Confusion Matrix for ARMA(1,1)-MH model's predictions of Nedbank Group Ltd returns series**

Figure 5.81 shows forecasts for Nedbank returns series obtained using ARMA(1,1)-MH model. Blue indicates the true returns and yellow indicates the predictions. The model has a "poor" ability to predict the direction of Nedbank returns series. It is as good as a random walk since its accuracy is only 65%.

Discussion:

For all four time series data sets considered, the ARMA(1,1)-Metropolis

Hastings model failed to achieve an accuracy of more than 68%. The poor performance could be due to the inability of the model to capture the dynamics of the returns series. Stock returns are non-linear in nature, as a result a linear model such as the ARMA(1,1)-Metropolis Hastings model cannot accurately capture the complex structure of these returns series.

5.10   Artificial Neural Network (ANN) results, Train:700 Test:200

Netcare Group Ltd: ANN model results

Architecture for ANN model constructed for the Netcare Group Ltd returns series: 2 hidden layers, 4600 epochs (passes over the training set), a learning rate of 0.0006, and a l2 regularization term of 0.0006. A tangent activation function is used on the input and output layers of the neural network, it output values between -1 and 1.
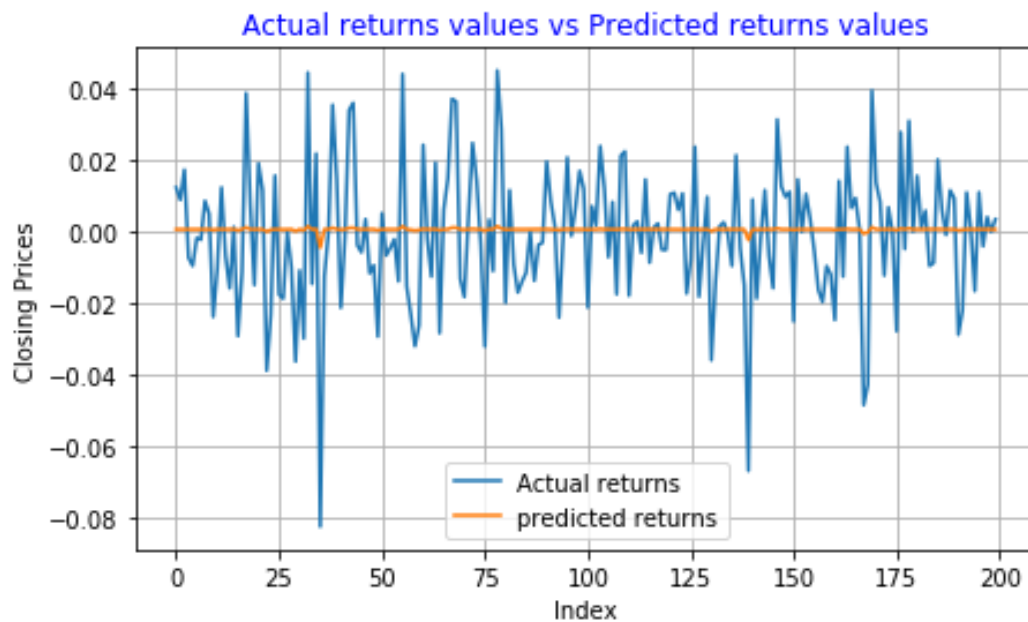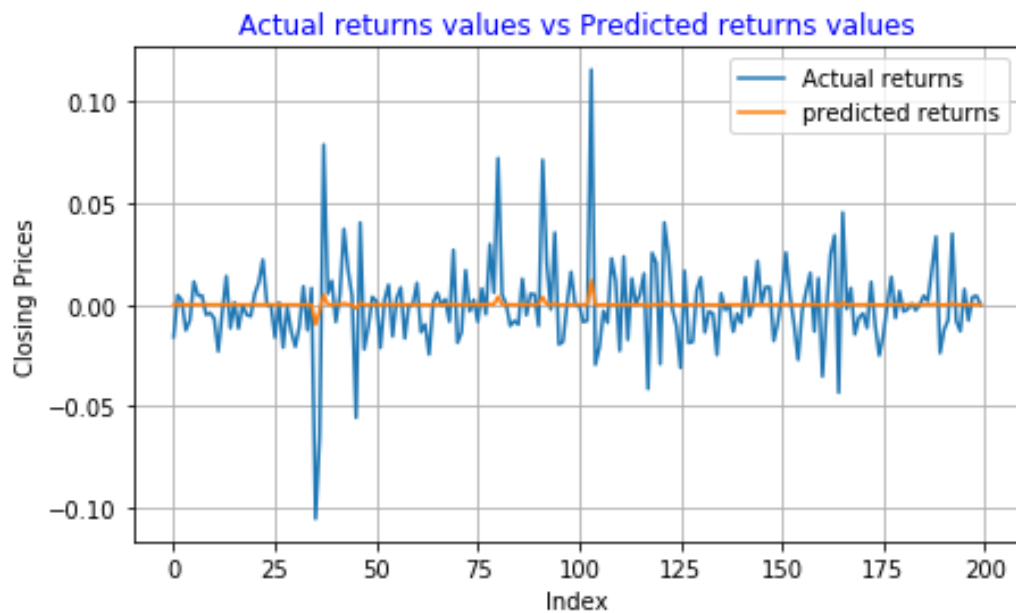

Figure 5.82: **Netcare Group Ltd returns series predictions using ANN model**

	Predicted Down	Predicted Up
True Down	74	30
True Up	24	72

<div align="center">**Accuracy**= 73%</div>

Table 5.66: **Confusion Matrix for ANN model's predictions of Netcare Group Ltd returns series**

Graph in Figure 5.82 shows the comparison between the true returns and the predicted returns. The structure of the predictions is similar to those in a study carried out by (Hansson, 2017). As seen in Table 5.66, the model achieved an accuracy of 73%.

Santam Ltd: ANN model results

Architecture for ANN model constructed for the Santam Ltd returns series: 2 hidden layers, 9800 epochs (passes over the training set), a learning rate of 0.0003, and a l2 regularization term of 0.0006. A tangent activation function is used in the input and output layers of the neural network.

Figure 5.83: **Santam Ltd returns series predictions using ANN model**

	Predicted Down	**Predicted Up**
True Down	74	22
True Up	23	81

Accuracy= 77.5%

Table 5.67: **Confusion Matrix for ANN model's predictions of Santam Ltd returns series**

Figure 5.83 shows the comparison between the actual returns and the predicted returns (from ANN model). The confusion matrix in Table 5.67 shows that the ANN model has a strong ability to learn the complex (non-linear) relationship that exists between the output and the input. The model has a misclassification rate of only 22.5%. These results indicates that the ANN model is a suitable machine learning algorithm for prediction of direction of movement of Santam Ltd returns series.

Sanlam Group Ltd: ANN model results

Architecture for ANN model constructed for the Sanlam Group Ltd returns series: 2 hidden layers, 4600 epochs, a learning rate of 0.00056, and a l2 regularization term of 0.00054. For the input and output layers of the neural network, a tangent activation function is used (it output values between -1 and 1).
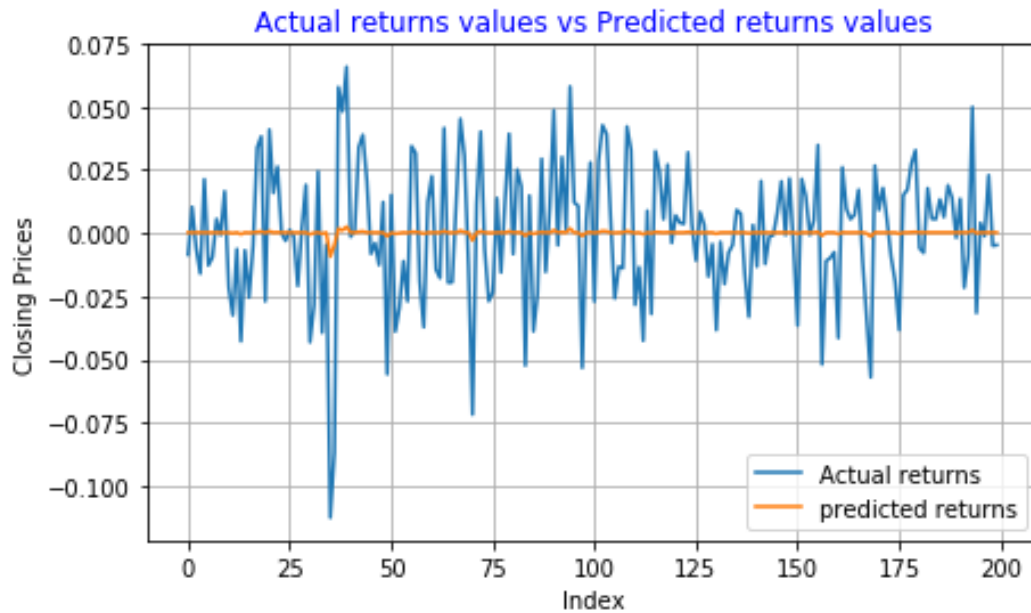
Figure 5.84: **Sanlam Group Ltd returns series predictions using ANN model**

	Predicted Down	Predicted Up
True Down	75	26
True Up	25	74

Accuracy= 74.5%

Table 5.68: **Confusion Matrix for ANN model's predictions of Sanlam Group Ltd returns series**

Figure 5.84 shows the predicted returns (yellow) and the actual returns (blue). Better capturing the volatility of the series does not necessarily lead to better forecasting accuracy as illustrated in the study by (Hansson, 2017). In terms of accurately predicting whether the next day's Sanlam Group Ltd returns price will go *up* or *down*, the ANN model achieved an accuracy of 74.5%.

Nedbank Group Ltd: ANN model results

Architecture for ANN model constructed for the Nedbank Group Ltd returns series: 2 hidden layers, 4900 passes over the training data set (epochs),

a learning rate of 0.0006, a l2 regularization term of 0.0006, and a tangent activation function for both the input and output layers.
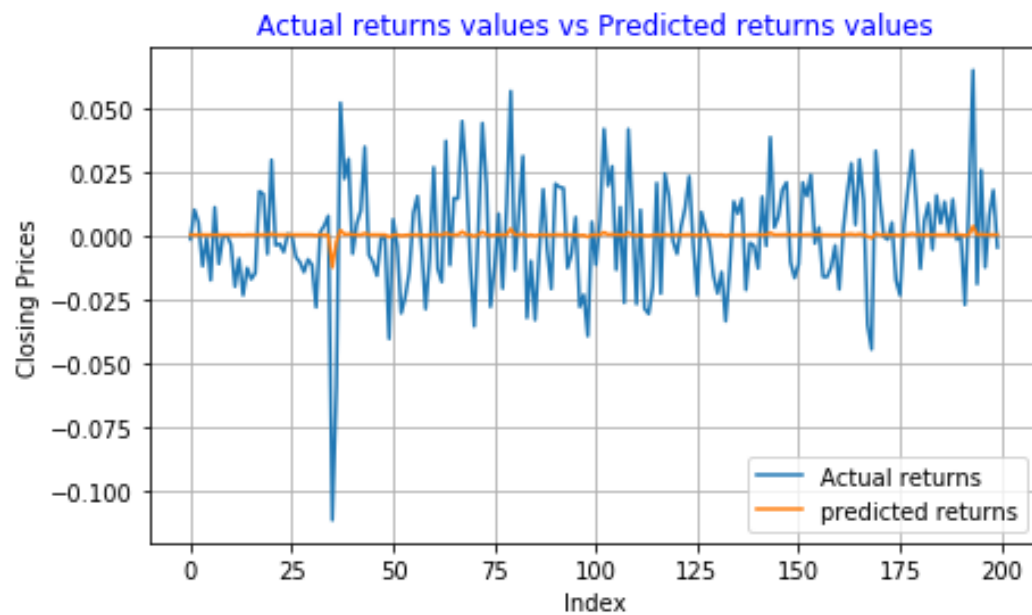

Figure 5.85: **Nedbank Group Ltd returns series predictions using ANN model**

	Predicted Down	**Predicted Up**
True Down	83	17
True Up	26	74

Accuracy= 78.5%

Table 5.69: **Confusion Matrix for ANN model's predictions of Nedbank Group Ltd returns series**

The graph in Figure 5.85 shows the comparison between the true returns (blue) and the predicted returns (yellow). The confusion matrix in Figure 5.69 illustrates how accurate the ANN model is in terms of its ability to predict the direction of movement of returns prices. The model achieved a very high accuracy of 78.5% and a misclassification rate (percentage of incorrect predictions) of only 21.5%.

Discussion:

One notable observation from the predictions of the ANN models is that they are extremely close to the mean. This phenomenon was also observed in studies carried out by (Hansson, 2017), and (Björklund and Uhlin, 2017). One of the advantage of Artificial Intelligence models is that they are able to capture the dynamics of "complex" data sets. In our case, the ANN model was able to correctly predict (with high accuracy) whether the next day's returns price for the various series will go *up* or *down*. The highest accuracy achieved by the model was 78.5% for the Nedbank returns series, and the lowest was 73% for the Netcare Group Ltd returns series.

5.11   Discussion of the Results

Accuracy Results

Table 5.70: **A comparison of forecasting accuracy of ARMA-type models and an Artificial Neural Network model, Train: 500 Test: 200**

Technique (Model)	Netcare	Santam	Sanlam	Nedbank
ARMA(1,1)-Maximum Likelihood	77%	65%	68%	73.5%
ARMA(1,1)-State Space	77%	73%	69.5%	74%
ARMA(1,1)-Metropolis Hastings	68.5%	69%	68.5%	69.5%
AR(3)-AVGARCH(1,1)-std	78%	71.5%	69%	69.5%
ANN Model	76%	71%	83.5%	89%

Table 5.71: **A comparison of forecasting accuracy of ARMA-type models and an Artificial Neural Network model, Train: 700 Test: 200**

Technique (Model)	Netcare	Santam	Sanlam	Nedbank
ARMA(1,1)-Maximum Likelihood	75%	77%	72%	69%
ARMA(1,1)-State Space	74.5%	74%	62.5%	69.5%
ARMA(1,1)-Metropolis Hastings	65.5%	68%	66.5%	65%
AR(3)-AVGARCH(1,1)-std	75%	72%	70%	70.5%
ANN Model	73%	77.5%	74.5%	78.5%

Table 5.70 and 5.71 shows forecasting accuracy results of ARMA-type models and an Artificial Intelligence model. As seen in Table 5.70, the ANN model outperforms the ARMA-type models in two cases (for the Sanlam and Nedbank returns series). Furthermore, as seen in Table 5.71 the ANN model again outperforms ARMA-type models in terms of predicting whether the next day's returns price will go up or down in three cases (for Santam, Sanlam, and Nedbank returns series). The results indicate that the ANN model is a suitable machine learning algorithm for prediction of direction of movement of stock returns of the four JSE listed companies. The ANN technique outperforms the ARMA-type models in most cases this may be because stock returns data have a "complex" structure, are nonlinear, and they behave in a "chaotic" manner. As a result, linear models such as the Box Jenkins ARMA-type models cannot adequately capture these behaviours.

Test for Superior Model

The Superior Predictive Ability (SPA) test is used to find the model with a superior forecasting ability out of all competing models, that is, we use the procedure to compare two or more competing statistical models. We test if there is any model which produce better forecasts than the benchmark model. We evaluate the forecasts using a pre-specified loss function (Hansen et al., 2003, p.g 6). In our study, we used the Mean Square Error (MSE) loss function. The model with the smallest expected (average) loss is chosen as the "best" forecasting model (Hansen et al., 2003, p.g 6). In general, we test if there is any model out of all competing models which produces better forecasting results than the benchmark model (Costa, 2017, p.g 35).

We test the null hypothesis that no model is better than the benchmark model against the alternative hypothesis that there is a competing model which is better than the benchmark (Costa, 2017, p.g 35). The SPA test has three p-values: lower p-value, consistent p-value, and the upper p-value. However, only the *consistent* p-value is used to evaluate the forecasting abil-

ity of the statistical models (Costa, 2017, p.g 35). If the *Consistent* p-value is small, one should therefore reject the null hypothesis and conclude that there is another model which is better the benchmark model.

Table 5.72: **Superior Predictive Ability (SPA) test, Train: 500 Test: 200**

Netcare Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.230	SPA_{Lower}=0.650
ARMA(1,1)-State Space	0.230	
ARMA(1,1)-Metropolis Hastings	0.315	$SPA_{Consistent}$=0.855
AR(3)-AVGARCH(1,1)-std *	**0.220**	
ANN Model	0.240	SPA_{Upper}=0.908
Santam Group		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.350	SPA_{Lower}=0.453
ARMA(1,1)-State Space *	**0.270**	
ARMA(1,1)-Metropolis Hastings	0.310	$SPA_{Consistent}$=0.685
AR(3)-AVGARCH(1,1)-std	0.285	
ANN Model	0.290	SPA_{Upper}=0.774
Sanlam Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.320	SPA_{Lower}=0.455
ARMA(1,1)-State Space	0.305	
ARMA(1,1)-Metropolis Hastings	0.315	$SPA_{Consistent}$=0.545
AR(3)-AVGARCH(1,1)-std	0.310	
ANN Model *	**0.165**	SPA_{Upper}=0.998
Nedbank Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.265	SPA_{Lower}=0.405
ARMA(1,1)-State Space	0.265	
ARMA(1,1)-Metropolis Hastings	0.305	$SPA_{Consistent}$=0.506
AR(3)-AVGARCH(1,1)-std	0.315	
ANN Model *	**0.105**	SPA_{Upper}=1.00

Note: This table shows the computed expected (average) loss for each model and the p-values for the Superior Predictive Ability (SPA) test. The table is divided into four parts, each part corresponding to a time series (by order : Netcare Group Ltd, Santam Group, Sanlam Group Ltd, Nedbank Group). For each time series data, the model with the smallest expected (average) loss is used as the benchmark model and is indicated by an asterisk.

In Table 5.72, the *consistent* p-values are greater than the 5% level of significance this suggests that for all four time series considered, the are no models which produces better forecasting results than the respective benchmark models. Large consistent p-values provides strong evidence against the alternative hypothesis. Hence, for the Netcare Group Ltd returns series no model outperforms the benchmark ARMA(1,1)-MLE model, for the Santam Group series the benchmark ARMA(1,1)-State Space model is better than all the competing models since it has the lowest average loss and its SPA test consistent p-value (about 0.685) is greater than 0.05 level of significance. For the Sanlam Group Ltd and the Nedbank Group Ltd returns series the benchmark ANN model outperformed all the competing models.

Table 5.73: **Superior Predictive Ability (SPA) test, Train: 700 Test: 200**

Netcare Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood *	**0.250**	$SPA_{Lower}=0.653$
ARMA(1,1)-State Space	0.255	
ARMA(1,1)-Metropolis Hastings	0.345	$SPA_{Consistent}=0.855$
AR(3)-AVGARCH(1,1)-std	0.290	
ANN Model	0.270	$SPA_{Upper}=0.975$
Santam Group		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.230	$SPA_{Lower}=0.459$
ARMA(1,1)-State Space	0.260	
ARMA(1,1)-Metropolis Hastings	0.320	$SPA_{Consistent}=0.631$
AR(3)-AVGARCH(1,1)-std	0.280	
ANN Model *	**0.225**	$SPA_{Upper}=0.648$
Sanlam Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.280	$SPA_{Lower}=0.486$
ARMA(1,1)-State Space	0.375	
ARMA(1,1)-Metropolis Hastings	0.335	$SPA_{Consistent}=0.748$
AR(3)-AVGARCH(1,1)-std	0.300	
ANN Model*	**0.255**	$SPA_{Upper}=0.841$
Nedbank Group Ltd		
Technique	**Average Loss**	**SPA test p-values**
ARMA(1,1)-Maximum Likelihood	0.310	$SPA_{Lower}=0.466$
ARMA(1,1)-State Space	0.305	
ARMA(1,1)-Metropolis Hastings	0.350	$SPA_{Consistent}=0.959$
AR(3)-AVGARCH(1,1)-std	0.295	
ANN Model *	**0.215**	$SPA_{Upper}=0.974$

Note: This table shows the computed average (expected) loss of each
model and the p-values for the SPA test. The model with the smallest
average (expected) loss is used as a benchmark model and is indicated by
an asterisk.

As seen in Table 5.73, for the Netcare Group Ltd returns series, the ARMA(1,1)-

MLE model had the lowest expected loss and was used as the benchmark. The SPA test shows that the model produces "better" forecasting results than all other models since the *consistent* p-value is relatively large (about 0.855). For the Santam Group, Sanlam Group Ltd, and the Nedbank Group Ltd returns series the ANN achieved the lowest average loss and the SPA test suggests that the model produces better forecasting results than all other competing models.

Chapter 6

Conclusion

In this chapter, conclusion that is aligned to the research problem and objectives (outlined in Chapter 1) is presented. In this Chapter we discuss the findings and explore possible limitations of the study. We further propose future work which can be carried out.

6.1   Research Findings

The models were evaluated in terms of their ability to predict whether the next day's returns price will go *down* or *up* i.e. the study only concentrated on trying to predict the direction of movement of returns series of JSE listed companies. The SPA test was used to find the "best" model out of all the competing models.

The returns series from the four JSE listed companies exhibited the following characteristics: volatility clustering, leverage effects, and non-normality of returns. These characteristics are common in financial markets (especially in emerging markets). One notable observation from fitting ARMA model on the returns series was that the constant variance and the normality assumption were usually violated. These violations suggest that an ARMA model which includes a conditional variance process such as the GARCH model should be used. One of the advantage of including a conditional variance process in the ARMA model is that we can assume that

the returns are from a specific statistical distribution such as the Student-t, Skewed Student-t, Hyperbolic or the Generalised Error Distribution (GED). The four distributions have leptorkurtic properties (heavier tails than the Gaussian/Normal distribution) (Ferenstein and Gasowski, 2004a, p.g.56).

The predictions from all the fitted models were operating around the mean, that is, they were very close to the mean. One notable characteristic of these predictions was that their variance was very "small". This phenomenon was also observed in studies carried out by (Hansson, 2017) and (Björklund and Uhlin, 2017). According to (Hansson, 2017, pg.23), better capturing the variance does not necessarily lead to better forecasts. The rate of decay of the transient is part of the formula for the variance, but since the data does not contain evidence of the transient component, it means that the estimation procedure does not estimate the rate correctly and therefore it affects the size of the variance.

Results emerging from the study showed that the ARMA-Maximum Likelihood, ARMA-State Space, AR(3)-AVGARCH(1,1)-Student-t, and ANN models have a strong ability to predict the direction of future returns prices (whether the price will go *down* or *up*). The highest accuracy achieved was 89% by the Artificial Intelligence Neural Network model. However, the ARMA-Metropolis Hastings model performed poorly, overall it failed to achieve an accuracy greater than 68%. Overall, the Artificial Intelligence Neural Network model outperformed all the time series ARMA-type models. The SPA test also confirmed the superiority of the ANN model over other models in at least 5 cases. In those cases, the model produced the smallest expected (average) loss. These results are similar to those of a study by (Marwala, 2010) which showed that when predicting the direction of movement of the JSE stock index artificial intelligence techniques such as ANN, Support Vector Machines, and Fuzzy models outperforms the ARIMA model and the Random Walk. A study carried out by (Adebiyi et al., 2014) also found that in terms of forecasting accuracy, Artificial Neural Network models were superior to Box Jenkins ARIMA models. Our study has shown that past stock returns can be used to predict the direction of movement of future returns prices.

6.2   Limitations and Recommendations

There are many factors which affect the movement of returns prices e.g political events, interest rates, natural disasters e.tc. This study was only confined in using past stock returns to make inference about future direction of movement of returns prices. As a result, our study is only limited to using one input (past stock returns) to try to understand the future movement of stock returns. The results emerging from the study sends a good message to fund managers, stock traders, and investors that statistical time series models and artificial neural networks can be used for predicting the direction of movement of future returns prices (whether the price will go *down* or *up*). Individual discretion has to be exercised if the investor wants to use these models to make a profit.

6.3   Areas of Further Research (Furture Work)

Future work should include combining the unique strength of linear ARMA-type models with Artificial Neural Network model to form a hybrid model, this would enable us to better capture both the linear and non-linear chaotic behaviour of stock returns. Another area of interest would be to use the Bayesian GARCH-type model as the conditional variance process of the ARMA model. Future research should also look at using macroeconomic variables such as GDP, inflation, interest rate etc. as inputs of the models to see if they contribute in improving the forecasting/ predictive power of the models.

Appendices

Appendix A

Python Codes

Statistical software packages **Python 2** and **Python 3** were used for analysing the data. The codes can be found on github: `https://github.com/vhahangwele/Thesis-Python-Codes`

List of References

A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo. Comparison of arima and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 2014:1–8, 2014.

D. Alberg, H. Shalit, and R. Yosef. Estimating stock market volatility using asymmetric garch models. *Applied Financial Economics*, 18(15):1201–1208, 2008.

C. Alexander. Market risk analysis: Practical financial econometrics (vol. ii), 2008.

C. Alexander. *Market Risk Analysis, Value at Risk Models*, volume 4. John Wiley & Sons, 2009.

T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. V. Mikosch. *Handbook of financial time series*. Springer Science & Business Media, 2009.

A. K. Basu and M. Saha. *Studies in Accounting and Finance: Contemporary Issues and Debates*. Pearson Education India, 2013.

S. Björklund and T. Uhlin. *Artificial neural networks for financial time series prediction and portfolio optimization*. PhD thesis, MA thesis. Linköping University, 2017.

T. Bollerslev. Glossary to arch (garch). 2008.

G. E. Box and G. M. Jenkins. *Time series analysis: forecasting and control, revised ed*. Holden-Day, 1976.

J. Y. Campbell, A. W.-C. Lo, and A. C. MacKinlay. *The econometrics of financial markets*. princeton University press, 1997.

J. J. Commandeur and S. J. Koopman. *An introduction to state space time series analysis*. Oxford University Press, 2007.

F. J. M. Costa. *Forecasting Volatility Using GARCH Models*. PhD thesis, 2017.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159–177, 1951.

E. Ferenstein and M. Gasowski. Modelling stock returns with ar-garch processes. *SORT-Statistics and Operations Research Transactions*, 28(1):55–68, 2004a.

E. Ferenstein and M. Gasowski. Modelling stock returns with ar-garch processes. *SORT-Statistics and Operations Research Transactions*, 28(1):55–68, 2004b.

P. H. Franses. *Time series models for business and economic forecasting*. Cambridge university press, 1998.

P. Fryzlewicz. Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics*, 32(5):503–528, 2005.

C. Fulton. Estimating time series models by state space methods in python: Statsmodels. 2017.

P. R. Hansen, J. Kim, and A. Lunde. Testing for superior predictive ability using ox, a manual for spa for ox. *Testing for superior predictive ability using Ox, a manual for SPA for Ox*, 2003.

P. R. Hansen et al. An unbiased and powerful test for superior predictive ability. Technical report, 2001.

M. Hansson. On stock return prediction with lstm networks. 2017.

A. Hossain, M. Kamruzzaman, and M. A. Ali. Arima with garch family modeling and projection on share volume of dse. *Economics*, 3(7-8):171–184, 2015.

M. Majumder and M. A. Hussian. Forecasting of indian stock market index using artificial neural network. *Information Science*, pages 98–105, 2007.

L. R. Marwala. *Forecasting the Stock Market Index Using Artificial Intelligence Techniques*. PhD thesis, 2010.

P. C. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

S. T. Rachev, J. S. Hsu, B. S. Bagasheva, and F. J. Fabozzi. *Bayesian methods in finance*, volume 153. John Wiley & Sons, 2008.

S. Raschka. *Python machine learning*. Packt Publishing Ltd, 2015.

S. Raschka and V. Mirjalili. *Python machine learning*. Packt Publishing Ltd, 2017.

A. M. Rather. A hybrid intelligent method of predicting stock returns. *Advances in Artificial Neural Systems*, 2014:4, 2014.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2006.

S. Sjöholm. Heteroscedasticity models and their forecasting performance, 2015.

H. Tang, K.-C. Chiu, and L. Xu. Finite mixture of arma-garch model for stock price prediction. In *Proceedings of the Third International Workshop on Computational Intelligence in Economics and Finance (CIEF'2003), North Carolina, USA*, pages 1112–1119, 2003.

R. S. Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.