A Novel Non-Linear GARCH framework for Modelling the Volatility of Heteroskedastic Time Series

Maneck Temti Bodhanwala

A thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy

Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, UK

November 2013

Abstract

Conventional methods of modelling financial data using Generalised AutoRegressive Conditional Heteroskedastic (GARCH) models make many assumptions involving the distribution of the data and the structures of the underlying mean and variance models. This research intends to develop a unified framework for estimating GARCH models using Non-Linear AutoRegressive Moving Average with eXogenous inputs (NARMAX) methodology without making many assumptions about the structures of the mean and variance models.

This thesis starts with a review of financial volatility and the different models that have been used to model financial volatility. These models are collectively termed as GARCH-class models. All GARCH-class models attempt to model financial volatility of a given return series by fitting a mean model, obtaining the residuals, and then fitting a variance model using the obtained residuals. Whilst a great deal of research has been done to develop different types of linear and non-linear variance model, researchers have ignored the possibility of the mean model being non-linear in nature, and most GARCH models use a very simple constant mean model to describe the means process.

In 2010, Zhao developed a NARMAX based Weighted Orthogonal Forward Regression method to identify non-linear mean models whilst assuming that the structure of the variance model is known. In this thesis, this method is extended to accommodate the case where the variance is unobservable. The working of this method is demonstrated with a simulated example. The method is also used to select and estimate the mean models of two real financial data sets and to demonstrate that a constant mean model is often inadequate and can result in inaccurate variance estimates.

A new Weighted Least Squares approach for the estimation of the variance model is also developed. Since the true variance is unobservable and unknown, a linear ARCH estimate of the variance is used as a proxy for the true variance. The results of simulations to demonstrate the working of the new method are also shown. Identification of a non-linear variance model is not possible using this method, since a linear estimate of the variance is used.

The thesis goes on to generate a non-linear estimate of the variance without making strong assumptions about the structure of the variance model making use of Radial Basis Function models. These have been used to create a generalised representation of linear and non-linear multivariate functions in other fields. In order to create a generalised non-linear variance estimate, a generalised RBF representation of a linear ARCH variance estimate is first created. The parameters of the obtained RBF model are optimised using Maximum Likelihood to generate a variance estimate that is much more accurate. The proposed method is tested and demonstrated in three simulations and succeeds in creating a non-linear variance estimate that is more accurate than a linear ARCH variance estimate in all the demonstrated simulations.

The methods introduced in this thesis build upon the newly developed application of NARMAX methodology to modelling financial volatility using GARCH models and provides a fresh new perspective to estimating financial volatility using GARCH-class models without making many assumptions about the structures of the underlying mean and variance processes.

Acknowledgements

I have been lucky to have three supervisors to guide me throughout this project and do not wish to belittle them by numbering them. They have all played an equally important role in supervising and guiding me through this unique project.

I would like to express my sincerest thanks and gratitude to Professor Stephen A. Billings for giving me the opportunity to work on this unique project. He has provided excellent supervision, immense support and guidance throughout the period of my work and I am thankful to him for being patient and understanding with me. I would like to sincerely thank Professor Daniel Coca for supervising my work and motivating me throughout the project. I have sincerely enjoyed the long brain-storming sessions and intense e-mail conversations with him and am grateful to him for guiding me and helping me overcome the obstacles that I faced throughout this project. I would like to express my heartfelt appreciation to Professor Kevin Dowd for helping me with the Economics aspect of this project and helping me through the lowest periods of my work when I most needed it. I have enjoyed all the talks we have had and would like to thank you for the unconditional support.

It has been an immense honour to share a meeting room with these three immensely knowledgeable and intelligent gentlemen and I have enjoyed all our brain-storming sessions. I would also like to thank Liang Zhao for getting me up to speed with all the work he had done, and helping me grasp all the concepts that were required initially.

I would like to thank the University of Sheffield and the European Research Council for the scholarship and funding. I would like to thank the complete staff of the Department of Automatic Control and Systems Engineering for all the provisions, especially Mrs. Linda Gray for the all the help and support she has provided me with over the years.

I would like to thank all my friends for their company, support and the memories, in particular, Mr Kushal D'Souza, Mr Uday Nair, Ms Pallavi Singh, Mr Sagnik Mukherjee, Mr Amit Tripathy, Mr Nishant Bhardwaj, Mr Siddharth Prem, Mr Kunal Gulati and Mr Karan Luthra.

Finally, I am forever indebted to my parents for all that they have been through to make me what I am, my sister, Diana, and my girlfriend, Shruti, for their unconditional love and support.

Contents

List of Tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 1

Introduction

1.1 Background and Motivation

Financial economists have long sought to forecast the prices and the financial asset returns. Forecasting such variables has proven to be immensely difficult, however, and over the last 30 years the focus has shifted to forecasting financial volatilities, that is to say, forecasting the variance of financial returns.

A key stylised fact of financial volatilities is that they vary over time, a property known as heteroskedasticity. The AutoRegressive Conditional Heteroskedastic (ARCH) model (Engle, 1982) was introduced to model this volatility. Robert Engle was awarded the Noble Prize for Economics in 2003 for this innovation which has laid the groundwork for a plethora of later variance models. One of the earliest and most popular such model is the Generalised AutoRegressive Conditional Heteroskedastic (GARCH) model introduced by Bollerslev (1986). Other models include, among others, the Integrated GARCH model (Bollerslev, 1986), the Exponential GARCH (Nelson, 1991), the Quadratic GARCH (Engle and Ng, 1993), the Non-Linear Asymmetric GARCH (Engle and Ng, 1993) and GJR-GARCH (Glosten et al., 1993). Each of these sought to model and forecast financial volatility more accurately than its predecessors.

The different types of GARCH models aim to capture some or all of the behaviour exhibited by financial time series and effectively model their volatility. One feature of most of these models is that they consist of 2 major equations to describe the behaviour of a financial return series. The first equation, termed as the mean model, describes the evolution of the mean of the returns. The second equation, termed as the variance model, describes the evolution of the variance of the returns. For the different models listed above, a simple constant mean model is usually used to describe the assumed mean process, while the variance models differ. Indeed, almost all the effort in this research field has gone into modelling the variance process, and very little effort has been made to improve the mean model.

The procedure used by financial economists to fit models to explain the mean and variance processes of a financial asset is different from the approach used by systems engineers and researchers to model physical systems. Economists perform initial tests on given financial data, fit models to the given data, validate the fitted model, and repeat the process until satisfactory results are achieved. After performing initial tests on the given data, systems engineers perform an exercise known as term selection whereby a large candidate model consisting of a variety of linear and non-linear terms is fitted to the given data. The terms that best describe the given data are selected to be included in the model, and the fitted model is then validated.

The leading model in the control engineering field is the Non-Linear AutoRegressive Moving Average with eXogenous inputs (NARMAX) model proposed by Leontaritis and Billings (1985a and 1985b). This model provides a unified framework for the identification and modelling of non-linear systems.

However, the simple Orthogonal Forward Regression (OFR) procedures of the NARMAX model can be ineffective in accurately selecting the terms of the underlying mean model of empirical return series due to their heteroskedastic nature. To address this problem, Zhao (2010) modified the simple OFR approach by using Weighted Least Squares (WLS) to deal with the underlying heteroskedasticity in the return series. He then used this Weighted OFR (or WOFR) approach to accurately select the terms and model the mean process of a given return series whilst assuming that the true variance process is known.

Zhao went on to examine the effects of under-fitting the mean model and demonstrated that under-fitting the mean model leads to the predictable elements of the mean process being discarded into the residuals. Since the residuals are used to estimate the variance, under-fitting the mean model leads to over-estimating the variance.

In the real world, however, the true variance of financial return series is unobservable, and hence, unknown. Zhao's WOFR method of accurately fitting a mean model therefore needs to be extended to the case where the structure of the variance model is unknown. Also, since it is common practice to fit a noise model in addition to fitting a process model to any time series, the WOFR method needs to be extended to handle such noise processes as well. There is, in addition, a need for more robust methods of model validation.

The standard approach to fitting a variance model to a given financial return series involves testing the returns for non-linearity, fitting a type of linear or non-linear variance model depending on the results of the previous test using maximum likelihood, performing model validation tests and selecting a different variance model if the validation tests fail (Engle, 2001). As in the case of the conventional approach to modelling the mean process, no framework for term selection of the variance model exists. Hence, a suitable method for term selection of the variance model when the variance is known and unknown (as is always the case with real financial data) needs to be developed as well.

Like the mean model, the variance model can also be non-linear in nature. There already exist a variety of non-linear variance models like the EGARCH (Nelson, 1991), QGARCH (Engle and Ng, 1993), NA-GARCH (Engle and Ng, 1993), GJR-GARCH (Glosten et al., 1993) models, to list a few, that can be used to describe a non-linear variance process. However, there exists no way of predetermining which non-linear variance model is required to be fitted to a given return series. There is, thus, a need for the development of a generalised approach to modelling a non-linear variance process that does not require strong assumptions about the true variance process.

A possible solution is to use Radial Basis Function (RBF) models. RBFs (Broomhead and Lowe, 1988) have been used to generalise and approximate linear and non-linear multivariate functions. RBF models are different from NARMAX models in the sense that a direct representation of the output using the inputs cannot be obtained by using RBF models. Instead, the output can be described as a combination of basis functions derived from a combination of the inputs. This generalised representation could be used to model the variance process of a given return series whilst using Maximum Likelihood to estimate the parameters of the RBF model.

1.2 Objectives

The main objectives of this thesis are:

(i) To review the most common GARCH-class volatility models in the literature and investigate the effects of incorrectly fitting the mean model on variance estimation.

The innovation of the ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) volatility models has laid down a solid framework to model the mean and variance of a financial return series. But the vanilla ARCH and GARCH models fail to capture a number of different behaviours exhibited by empirical financial volatility. To capture these behaviours, there exist at least 100 variations of GARCH models (Bollerslev, 2008), collectively termed as GARCH-class volatility models. The method of fitting these models remains the same. A mean model is first fitted to the returns. The one-step-ahead modelling error, known as the residuals, are obtained and then used to estimate the variance using a selected variance model. Since the mean model plays a pivotal role in variance estimation, the effects of fitting incorrect mean models on the accuracy of variance estimates are to be investigated.

(ii) To develop structure selection and parameter estimation methods for modelling the mean of the heteroskedastic processes when the structure of the variance model is unknown.

A vast amount of research has been done in this literature to improve the way the variance is modelled, but not much has been done to improve the way the means process is modelled. An exception is Zhao (2010) who introduced the NARMAX based WOFR method for the term selection and parameter estimation of the mean model in the case where the true structure of the variance is known. But for real financial data sets, the underlying variance is unobservable. Hence, the NARMAX based WOFR method needs to be extended to incorporate this real world scenario.

(iii) To develop structure selection and parameter estimation methods for modelling the variance of heteroskedastic processes.

The conventional method for fitting a variance model does not include term selection. Preliminary tests are carried out to determine whether a linear or a non-linear variance model is required to be fitted, after which a variance model of that type is fitted. The structure of the variance model is arbitrarily selected, and validated using model validation tests. There is a clear need for developing a technique for term selection for the variance model to simplify the modellers task and let the variance model be extracted from the given data.

(iv) To develop a general methodology for modelling both, the mean and the variance, of a heteroskedastic process without making any prior assumptions about the structure of these models.

Instead of testing the return series for non-linearity, fitting a variance model based upon the outcome of the test, validating the fitted model and re-fitting a different model if validation fails, it would make sense to have a general method for selecting the best variance model without making any prior assumptions about the structure and nature (linear or non-linear) of the variance.

(v) To demonstrate the applicability of the proposed methods through numerical simulation studies and by benchmarking these methods using real financial data sets.

The performances of the proposed methods need to be demonstrated through numerical simulation studies. Returns data will be simulated from known mean and variance models. The proposed methods will then be applied to the simulated returns data whilst assuming that nothing is known about the true mean and variance model and that the true variance is also unknown. The mean and variance models selected by the proposed methods will be compared to the true mean and variance models to demonstrate the accuracy of the proposed methods. Finally, the proposed methods will be applied to real financial data sets to investigate the nature (linear or non-linear) of the underlying mean and variance processes.

1.3 Layout of Thesis

This thesis is organised into 8 chapters as follows:

Chapter 2 provides an introduction to the fundamental basics of financial time series. The stylised facts about financial time series are explained and the various GARCH class of volatility models found commonly in the literature are reviewed.

Chapter 3 introduces the financial mean (return) model and reviews the different mean models found in the ARCH literature. The NARMAX based WOFR methodology for identifying and fitting linear and non-linear models (Zhao, 2010) is explained in detail. The WOFR method for identifying the mean model of a given return series when the structure of the variance model is known is also reviewed. The results of simulations demonstrating the impact of incorrectly fitting the mean model on variance estimation and the results of implementing the WOFR method on a simulated data set are shown.

Chapter 4 introduces structure selection and parameter estimation methods for modelling the mean of the returns process when the structure of the underlying variance process is unknown. The method is then modified to include a linear noise model in addition to the mean process model. New tests for model validation are introduced. The results of implementing the extended WOFR method on a simulated data set while assuming that the variance is unobservable are shown. The effects of under-fitting the mean model on mean and variance estimation are also investigated.

Chapter 5 demonstrates the applicability of the method proposed in Chapter 4 and showcases the results of the implementation of the extended WOFR method on two real financial data sets. The best mean models for the real data sets are obtained and the effects of under-fitting the mean model on mean and variance estimation for the given data sets are investigated.

Chapter 6 reviews the most commonly used method for the estimation of the variance model. A new NARMAX based approach using WLS to model the underlying variance of a given return series is introduced. The results of simulations demonstrating the effect of using a NARMAX based approach to model the variance without using WLS are shown. The results of simulations demonstrating the new WLS approach to model the variance in various scenarios are also shown.

Chapter 7 reviews RBF models, the different basis functions, and the method used to estimate (train) an RBF model. A method to generate a non-linear estimate of the variance using RBFs is then introduced. The results of three different simulations demonstrating the implementation of the new method to obtain a non-linear estimate of the variance from a known residual series are showcased. The accuracy of the generated non-linear variance estimates are compared to that of a linear variance estimate generated using the same residuals.

Chapter 8 lists the main contributions of this thesis, sets out some limitations of the work and suggests further areas that can be investigated.

Chapter 2

Modelling Financial Volatility

2.1 Introduction

The price series of an asset is a collection of the actual price of an asset over a period of time. Asset returns, on the other hand, denote the difference in the price of the asset over a period of time. The measure of the variation of the price of a financial asset over a period of time is termed as its volatility. Typically, the financial time-series data is a chronological collection of the returns of a certain option or derivative. The frequency of these returns can be hourly, daily, weekly, monthly or yearly, depending on the type of predictions required. The variance of the returns describes the magnitude of the volatility (commonly known as the risk) of the asset at that particular instant in time. This risk is not constant and varies with time. The financial return series is hence said to have time varying variance or in financial terms, it is known to be heteroskedastic. If the variance does not vary with time, it is termed as homoskedastic.

Volatility modelling of asset returns is an important aspect for many financial applications, especially option or derivative pricing and risk management. Market risk is an explosive new development in finance, which is driven by the statutory requirement for companies to monitor market exposure on a daily basis, and with the introduction of new derivatives on mortality and other financial products. A class of models known as GARCH (Generalised Auto Regressive Conditional Heteroskedasticity) models are usually used to model the volatility processes of financial time-series.

Volatility is dependent on several external factors. For example, bad news about a company can cause a decline in its stock prices or similarly, a rise in anticipation of its half-yearly financial results. This behaviour is known as volatility clustering. Financial systems also exhibit behaviour termed as the leverage effect. The volatility response to a large negative return is far greater than it is to a large positive return of the same magnitude.

The different types of GARCH models aim to capture some or all of this behaviour of financial time-series and effectively model their volatility. They consist of 2 major equations that help describe the return series data. The first equation describes the evolution of the mean of the returns (known as the mean model) and the second one describes the evolution of the variance of the returns (known as the variance model).

GARCH models have come a long way since their introduction (Bollerslev, 1986) and are used routinely by economists for risk analysis which is an important part of option pricing, asset pricing, risk management and portfolio optimisation (Engle, 2001). Despite their huge popularity, existing volatility models still have many shortcomings.

Various types of GARCH models have been developed in an attempt to capture the different types of behaviour of financial systems. As a result, a particular type of model is suitable to describe only certain trends/effects adequately. For example, a generic GARCH model is not capable of capturing and describing the asymmetric behaviour of the variance exhibited by the leverage effect. For this purpose, several different types of GARCH models have been developed. These models are said to belong to the GARCH class of models since they have been developed from GARCH models with similar implementation.

2.2 Financial Time-series

The two most commonly used and freely available financial time series are the price series and the return series. A majority of financial studies deal with data comprising of asset returns rather than asset prices. This is due to the simple fact that return series have several desirable properties over price series. For starters, return series appear to have a stable mean whilst this is not the case with price series.

Let $P(t)$ represent the price of an asset at any instant in time, t. The price series is represented as $\{P(t)\} = \{P(1), P(2), \dots, P(n)\}$ where n is the total number of samples of the price available. The returns series is represented as $\{y(t)\} = \{y(1), y(2), \dots, y(n - 1)\}$.

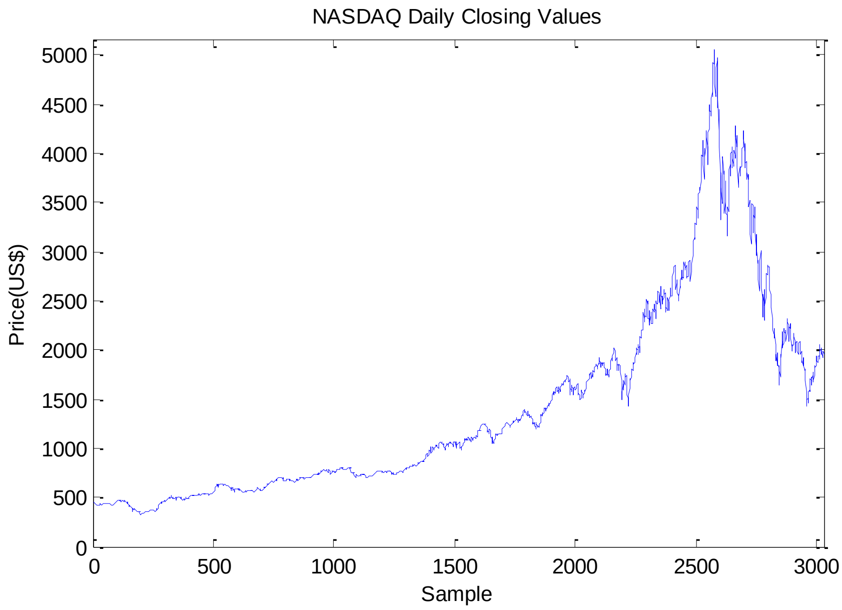


Figure 2.1 Price Series

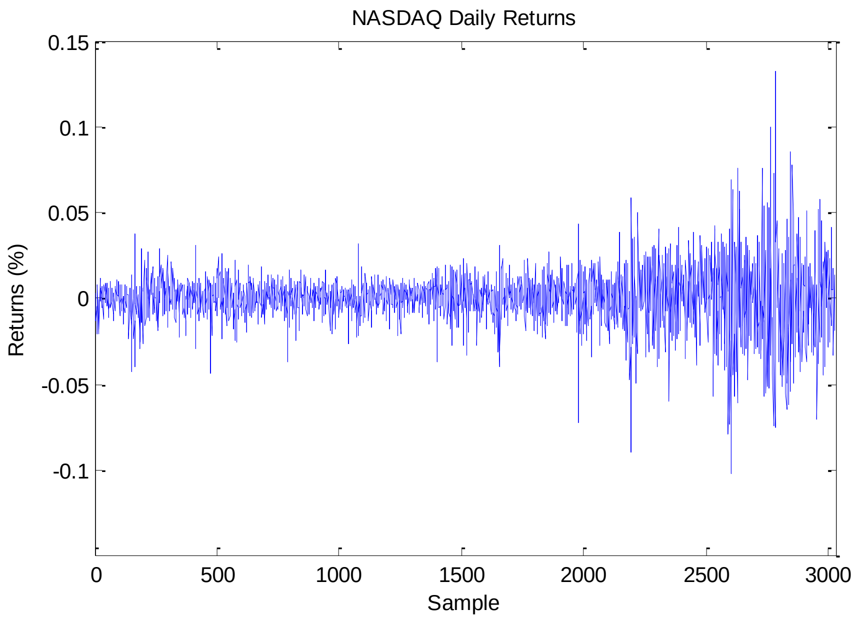


Figure 2.2 Return Series

A price series can be converted into a return series in two ways:

- Simple Periodic Compounding: $y(t) = \frac{P(t+1)-P(t)}{P(t)} = \frac{P(t+1)}{P(t)} - 1$ (2.1)

- Continuous Compounding: $y(t) = log\left(\frac{P(t+1)}{P(t)}\right) = log\, P(t+1) - log\, P(t)$ (2.2)

Continuous compounding is the preferable choice for most financial studies.

The return series of an asset is assumed to be driven by a white noise series. A white noise series is a collection of independent and identically distributed (i.i.d) random variables. For an ideal white noise series, the autocorrelation function is zero. In practice, if the autocorrelation function is within predefined confidence bounds (95% or 98%) the series is considered to be a white noise series.

The probability distribution of an asset return is assumed to be Gaussian and is defined by two factors: mean and variance. The variance of this probability distribution describes the volatility of the asset. The higher the variance of the return at that instant in time, the greater the risk. For any financial return series, the variance is time-varying i.e. the returns are heteroskedastic.

Asset returns tend to exhibit certain behaviour that is specific to financial time-series and exist mainly due to the way in which financial markets work.

- **Leverage Effect**: The volatility response to a large negative return is greater than it is to a large positive return of the same magnitude. In other words, it is observed that after a large negative return, the volatility of an asset increases immediately by a larger magnitude than that observed after a large positive return. This is caused partially due to the fact that after a negative return, the debt increases thereby increasing the volatility of the asset.

- **Volatility Clustering**: The volatility of an asset increases right before an anticipated announcement or right after an unanticipated announcement. It is often noticed that stock prices of a certain company rise before the announcement of their quarterly, half-yearly and annual financial results especially when good results are expected. In a similar way, it is often noticed that the stock prices of a company fall immediately after the release of news that negatively impacts the company or the sector it belongs to.

- **Mean Reversion**: On observation of any asset return series, it is noticed that the returns always tend to converge to a mean after a period of disturbance. This phenomenon is described as mean reversion. This phenomenon is not specific only to asset returns but is also noticed in asset prices and interest rates. The asset prices tend to revert to a historical average value. However, the likelihood of mean reversion of any of these factors is not 100%. It is less likely to occur if the outlook of the company changes due to a change in management or other unforeseen circumstances.

2.3 Modelling Financial Time-series

As mentioned earlier, dynamic modelling of asset returns and their volatilities is largely based on GARCH-type models. A lot of work has been done in this field and numerous types of models have been developed that attempt to model financial time-series accurately while also trying to accurately capture the various common effects described earlier.

The following section aims to summarise the progress in this field whilst providing a brief description about the different types of models developed, their use, advantages and disadvantages.

2.3.1 Historical Average or Moving Average (MA) Model

The Moving Average model is one of the simplest variance model used in the field of Econometrics. The model uses a simple concept to estimate the variance. The variance at time, t, is calculated by averaging the past n values of the squared returns. The value of n is user-defined.

$$\hat{h}(t) = \frac{1}{n}\sum_{i=1}^{n} y^2(t-i) \tag{2.3}$$

where $\hat{h}(t)$ is the estimate of the variance at time, t and $y(t)$ is the return at time, t.

This historical variance calculated using the past n values of the squared returns is often used as a measure of the risk of a portfolio. If n is selected to be fairly large, and unusually large shock in the returns would decrease the accuracy of the variance estimate.

In order to overcome this problem, n is selected to be reasonably small. Another advantage of selecting a reasonably small value of n is that the effect of volatility clustering can be effectively captured. The dynamics of the variance is still not effectively captured due to the fact that the squared returns are weighted equally irrespective of the order of their occurrence. In order to overcome this problem, the Exponentially Weighted Moving Average Model was introduced.

2.3.2 Exponentially Weighted Moving Average (EWMA) Model

Exponentially Weighted Moving Average (EWMA) models (Harrison, 1967) attempt to model the volatility of a return series by using past values of the volatility and the squared returns. The past values of volatility are weighted such that the most recent value of volatility has a higher weight than that of a term much further in the past. This implies that the most recent value of volatility has a higher influence on the current volatility. This ensures that shocks in the market impact the volatility to a larger extent i.e. the volatility of an asset will be high after a market shock.

The EWMA model comprises of 2 main parameters – the smoothing constant, λ, and time, t. Let $y(t)$ be the return at any instant in time, t. $\hat{h}(t)$ is the estimated variance of the return and λ is the smoothing constant. In general, the EWMA model is given by:

$$\hat{h}(t) = (1-\lambda)\sum_{i=1}^{n} \lambda^{i-1} y^2(t-i) \qquad (2.4)$$

In equation (2.4), λ is the most significant control parameter and an optimum value must be selected for the development of an accurate model. Another parameter to be considered is the number of effective observations to be used to generate the model (denoted by n).

Recursively, equation (2.4) can be written as

$$\hat{h}(t) = (1-\lambda)y^2(t-i) + \lambda\hat{h}(t-1) \qquad (2.5)$$

The first term, $(1-\lambda)y^2(t-i)$, is known as the reaction parameter. A low value of λ implies that the estimate of the variance at time, t, is largely dependent on the past values of squared returns and will react to large shocks to the returns.

The second term, $\lambda\hat{h}(t-1)$, is known as the persistence parameter. A high value of λ implies that the estimate of the variance at time, t, is largely dependent on the past values of the estimated variance. This implies that large shocks to the returns will not affect the current volatility much, thereby making the volatility persistent.

In order to ensure that recent values of squared returns are weighted more than the older values, λ is chosen to be between 0 and 1. Since both the above mentioned parameters are controlled by one parameter, λ, the volatility can be modelled either to make it persistent (high value of λ) or more reactive to shocks in the returns (low value of λ). This limitation is exploited to model the volatility in the foreign exchange market using the EWMA model (Alexander, 2001).

2.3.3 Auto Regressive Conditional Heteroskedastic (ARCH) Model

The Auto Regressive Conditional Heteroskedastic (ARCH) model is considered to be the foundation of volatility modelling and was developed by Engle in 1982 (Engle, 1982). It is based on the concept that the variance is dependent on the past values of the residuals of the returns. The returns are modelled as:

$$y(t) = g(y(t-1), y(t-2), \dots, y(t-m); b_1, b_2, \dots b_m) + e(t) \qquad (2.6)$$

Here, $g(y(t-1), y(t-2), \dots, y(t-m); b_1, b_2, \dots b_m)$ is a function of m past returns and $b_1, b_2, \dots b_m$ are the parameters of the past return terms. $e(t)$ is the residual and is considered to be serially uncorrelated and has zero mean. The residual is also written as $z(t)\sqrt{h(t)}$ where $\{z(t)\}$ is a sequence of independent and identically distributed (i.i.d) random variables with a zero mean and unit variance and $\sqrt{h(t)}$ represents the standard deviation of the return at time, t. The series, $\{z(t)\}$, is collectively known as the standardised residuals. The variance equation is thus written as follows:

$$h(t) = a_0 + a_1 e^2(t-1) + a_2 e^2(t-2) + \cdots + a_m e^2(t-m) \qquad (2.7)$$

Equations (2.6) and (2.7) bundled together represent the ARCH(m) model, where m is the number of lagged returns used. Large shocks in the return series imply a high value of $e(t-m)$ and hence imply a high value of $h(t)$. This means that large shocks to the returns cause an increase in the volatility but it still fails to capture the leverage effect discussed in Section 2.2. This is because squared values of the past residuals are used which fail to differentiate between negative and positive shocks.

2.3.4 Generalised Auto Regressive Conditional Heteroskedastic (GARCH) Model

The advantages of the simple ARCH model are outweighed by the disadvantages. For example, a financial time-series often requires an ARCH(m) model where the value of m is quite high which leads to the problem of estimating a high number of parameters thereby decreasing the overall accuracy of the model. It is for this very reason that Generalised Auto Regressive Conditional Heteroskedastic (GARCH) models were developed (Bollerslev, 1986).

GARCH models are able to capture most of the effects exhibited by financial time-series and are effective in modelling and forecasting the variances of asset returns. A financial return series can be described accurately using 2 parameters: mean and variance. GARCH models attempt to replicate a return series by providing models for the mean and the variance. Let $y(t)$ be the return at any instant in time, t. $h(t)$ is the variance of the return and $e(t)$ is the residual and is also known as innovation of the returns process. a_0, K, G_i and A_j are constants. The GARCH(p,q) model is written as:

Mean Model: $$y(t) = a_0 + e(t) \tag{2.8}$$

Variance Model: $$h(t) = K + \sum_{i=1}^{p} G_i h(t-i) + \sum_{j=1}^{q} A_j e^2(t-j) \tag{2.9}$$

$$p \geq 0, q > 0$$
$$K > 0, G_i \geq 0, i = 1, \dots, p$$
$$A_j \geq 0, j = 1, \dots, q$$
$$\sum_{i=1}^{\max(p,q)} G_i + A_i < 1$$

In the final condition, it is implied that $G_i = 0$ for $i > p$ and $A_j = 0$ for $j > q$. If $p = 0$, the model loses the lagged variance term to become an ARCH(q) model. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as

$$h(0) = \frac{K}{1 - \left(\sum_{i,j=1}^{i=p,j=q} G_i + A_j\right)} \tag{2.10}$$

The GARCH(1,1) model is the most basic and commonly used variation of the above given model and is also known as the vanilla GARCH model. The residual, $e(t)$, is represented as a product of the standard deviation of the return, $\sqrt{h(t)}$, and a random independent and identically distributed (i.i.d) term, $z(t)$, that has zero mean and a variance of 1. The series $\{z(t)\}$ is collectively known as the standardised residuals. Equation (2.14) just represents an alternative way of describing the mean model using $z(t)$.

$$y(t) = a_0 + e(t) \tag{2.11}$$

$$h(t) = K + Gh(t-1) + Ae^2(t-1) \tag{2.12}$$

$$e(t) = z(t)\sqrt{h(t)} \tag{2.13}$$

or

$$h(t) = K + Gh(t-1) + Ah(t-1)z^2(t-1) \qquad (2.14)$$

From equation (2.8) it can be noted that the mean model is a constant. This is a common occurrence in quite a lot of variations of the GARCH model. A few important ones are detailed in sections 2.3.5 to 2.3.11.

2.3.5 Integrated GARCH (IGARCH) Model

The GARCH model effectively models volatility that exhibits the phenomenon of mean reversion. But quite a few financial instruments like currencies and commodities do not tend to exhibit mean reversion. In such cases, the Integrated GARCH (IGARCH) model is helpful (Bollerslev, 1986). The IGARCH model looks exactly like the GARCH model apart from the fact that the coefficients of the lagged variance and error terms must add up to 1.

Generally, the mean model is an ARMA(*m*,*n*) model which is written as

$$y(t) = a_0 + \sum_{i=1}^{m} a_i y(t-i) + \sum_{j=0}^{n} b_j e(t-j) + e(t) \qquad (2.15)$$

where $y(t)$ represents the returns at time, t and $e(t)$ represents the residual or innovations at time, t. In most cases, the mean model is simply a constant mean model where *m* and *n* are both zero.

Let the variance of the return be represented as $h(t)$. An IGARCH(*p*,*q*) model can be written as follows:

Mean Model: $$y(t) = a_0 + e(t) \qquad (2.16)$$

Variance Model: $$h(t) = K + \sum_{i=1}^{p} G_i h(t-i) + \sum_{j=1}^{q} A_j e^2(t-j) \qquad (2.17)$$

and $$\sum_{i=1}^{p} G_i + \sum_{j=1}^{q} A_j = 1 \qquad (2.18)$$

In the IGARCH model, the variance equation has a unit root and a shock to the variance has a permanent effect on all the future predictions of the variance.

2.3.6 Exponential GARCH (EGARCH) Model

The vanilla GARCH model is the most commonly used model to model financial time-series. One of its major disadvantages is that it fails to capture the leverage effect (explained in Section 2.2). Since the past values of the square of the residuals are used in the model, a positive shock to the returns will have the same impact on the volatility as a negative shock of the same magnitude. This is not quite desirable and hence, in 1991, Nelson proposed the EGARCH model that effectively captures and describes the leverage effect.

Let $h(t)$ be the variance of the return, $e(t)$ be the residual of the returns process. $z(t)$ is a random independently and identically distributed (i.i.d) term that has zero mean and a variance of 1. The residuals can also be written as $e(t) = z(t)\sqrt{h(t)}$. The EGARCH(p,q) model is written as:

$$ln(h(t)) = a_0 + \sum_{i=1}^{p} \beta_i ln(h(t-i)) +$$

$$\left(1 + \textstyle\sum_{i=1}^{q} \alpha_i L^i\right)\left(1 + \textstyle\sum_{j=1}^{p} \beta_j L^j\right)^{-1} \left(\theta z(t-1) + \gamma \begin{bmatrix} |z(t-1)| - \\ E|z(t-1)| \end{bmatrix}\right) \quad (2.19)$$

Here, $a_0, \beta_i, \alpha_i, \gamma, \theta$ are all unknown parameters. L is the backshift operator (also known as the lag operator) such that $Lz(t) = z(t-1)$. Assume,

$$g(z(t)) = (\theta z(t) + \gamma[|z(t)| - E|z(t)|]) \quad (2.20)$$

This can also be rewritten as

$$g(z(t)) = \begin{cases} (\theta + \gamma)z(t) - \gamma E|z(t)| & if\ z(t) \geq 0, \\ (\theta - \gamma)z(t) - \gamma E|z(t)| & if\ z(t) < 0. \end{cases} \quad (2.21)$$

This clearly shows us that for a shock of positive value ($z(t) \geq 0$), the volatility response will be lesser compared to a negative shock of equal magnitude, thereby incorporating the leverage effect effectively in the model. Another advantage of the EGARCH model is that no inequality constraints are required for the parameters since the volatility will always be positive due to the presence of the logarithmic operator.

2.3.7 Quadratic GARCH (QGARCH) Model

The inability of the traditional GARCH model to capture the leverage effect gave rise to a number of models that had a similar framework. These models are collectively termed as asymmetric GARCH models. One such model is the Quadratic GARCH (QGARCH) model (Engle and Ng, 1993). The variance equation of the QGARCH model is as follows:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 (e(t-1) - \gamma)^2 \quad (2.22)$$

where $h(t)$ is the variance of the return, $e(t)$ is the residual of the returns process and a_0, a_1, β_1 and γ are the unknown parameters. γ is known as the leverage parameter. Equation (2.22) can be expanded and rewritten as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 e^2(t-1) + \gamma^2 \beta_1 - 2\gamma\beta_1 e(t-1) \quad (2.23)$$

Since a_0, a_1, γ and β_1 are all currently unknown, the coefficients can be congregated and written as one for each term. Equation (2.23) can be rewritten as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 e^2(t-1) - \varphi e(t-1) \quad (2.24)$$

Here, φ is a positive unknown constant. As we can see, a negative shock to the return series will make the final term in equation (2.24) positive thereby increasing the future volatility while a positive shock of the same magnitude will decrease the volatility thereby resulting in a GARCH model that effectively captures the leverage effect.

2.3.8 Non-Linear Asymmetric GARCH (NA-GARCH) Model

The NA-GARCH model (Engle and Ng, 1993) is one of the first attempts at capturing the non-linear effects of heteroskedastic volatility while also capturing the Leverage Effect. The variance equation of the NA-GARCH model is as follows:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 \left(e(t-1) - \gamma\sqrt{h(t-1)}\right)^2 \quad (2.25)$$

where $h(t)$ is the variance of the return, $e(t)$ is the residual of the returns process and a_0, a_1, β_1 and γ are the unknown parameters. γ is known as the leverage parameter. Equation (2.25) can be expanded and rewritten as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 e^2(t-1) + \gamma^2 \beta_1 h(t-1) - 2\gamma\beta_1 e(t-1)\sqrt{h(t-1)} \quad (2.26)$$

Since a_0, a_1, γ and β_1 are all currently unknown, the coefficients can be congregated and written as one for each term. Equation (2.26) can be rewritten as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 e^2(t-1) - 2\gamma\beta_1 e(t-1)\sqrt{h(t-1)} \quad (2.27)$$

This looks very similar to the QGARCH model but upon expansion of the model, it can be noted that the additional term includes a product of the past residual and the past standard deviation of the asset. A positive value of γ, the leverage operator, implies that the leverage effect is effectively captured as well.

2.3.9 Square-Root GARCH (SQR-GARCH) Model

The SQR-GARCH model (Heston and Nandi, 2000) has a structure similar to that of the QGARCH model. The variance equation of the SQR-GARCH model is written as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 \left(\frac{e(t-1)}{\sqrt{h(t-1)}} - \gamma \sqrt{h(t-1)} \right)^2 \quad (2.28)$$

where $h(t)$ is the variance of the return, $e(t)$ is the residual of the returns process and a_0, a_1, β_1 and γ are the unknown parameters. γ is known as the leverage parameter. Equation (2.28) can be expanded and rewritten as:

$$h(t) = a_0 + a_1 h(t-1) + \beta_1 \frac{e^2(t-1)}{h(t-1)} + \gamma^2 \beta_1 h(t-1) - 2\gamma\beta_1 e(t-1) \quad (2.29)$$

A positive value of γ, the leverage operator, implies that the leverage effect is effectively captured. A negative shock to the returns will make the last term of equation (2.29) positive and add to the variance, whilst a positive shock to the returns of the same magnitude will make that term negative and reduce the variance.

2.3.10 GJR-GARCH Model

Another famous asymmetric GARCH model variant that has the ability to capture the leverage effect is the GJR-GARCH model (Glosten et al., 1993). This is done by the addition of an identification term to the normal GARCH model. The variance equation of the GJR-GARCH(p, q) model is as follows:

$$h(t) = a_0 + \sum_{i=1}^{p} a_i h(t-i) + \sum_{j=1}^{q} \beta_j e^2(t-j) + \gamma e^2(t-1) I(t-1) \quad (2.30)$$

where $h(t)$ is the variance of the return, $e(t)$ is the residual of the returns process and $a_0, a_1, \dots, a_i, \beta_1, \dots, \beta_j$ and γ are the unknown parameters. γ is known as the leverage parameter. The identification term is the last term i.e. $\left(\gamma . e^2(t-1) . I(t-1)\right)$.

$$I(t) = \begin{cases} 1, & if\ e(t) < 0 \\ 0, & if\ e(t) \geq 0 \end{cases} \quad (2.31)$$

Hence, for a negative shock to the returns, the coefficient of the $e^2(t-1)$ term will be $(\beta_1 + \gamma)$ whilst for a positive shock of the same magnitude, the coefficient will be β_1 thereby resulting in a GARCH model that effectively captures the leverage effect.

2.3.11 GARCH in Mean (GARCH-M) Model

The return of an asset may depend on its volatility. It is for this very reason that the GARCH-M model (Engle et al., 1987) was developed. This model is different from all the models specified in sections 2.3.1 to 2.3.10 in the fact that it uses a simple GARCH variance model but a modified mean model. The GARCH-M model is as follows:

Mean Model: $$y(t) = a_0 + a_1 h(t) + e(t) \tag{2.32}$$

Variance Model: $$h(t) = K + \sum_{i=1}^{p} G_i h(t-i) + \sum_{j=1}^{q} A_j e^2(t-j) \tag{2.33}$$

where $y(t)$ represents the returns at time, t, $h(t)$ is the variance of the return, $e(t)$ is the residual of the returns process and $a_0, a_1, K, G_1 \ldots, G_i, A_1, \ldots, A_j$ are the unknown parameters.

In some cases, the standard deviation of the return $\left(\sigma(t) = \sqrt{h(t)}\right)$ is used in place of the variance ($h(t)$) in the mean model. A positive value of a_1 implies that the return of an asset is positively related to its volatility. The existence of the variance term in the mean model also implies that the returns are serially correlated.

2.4 Conclusions

Many variants of the GARCH model have been developed in the recent years. The variance model is the main feature that differentiates different types of models, as often, the mean model includes just a single constant term. Very little research has been carried out on the estimation of non-linear GARCH models when the structure of the mean and variance models is unknown. In particular, the importance of fitting an accurate mean model has been largely overlooked both by theoreticians and practitioners. In this respect, this thesis advocates the use of rigorous model selection, parameter estimation and model validation approaches to estimate linear or non-linear mean and variance models. It is demonstrated through numerical simulations that fitting an incorrect mean model results in the contamination of the residuals, $e(t)$, and leads to an incorrect estimate of the variance model.

Chapter 3

The Financial Mean Model

3.1 Introduction

The mean model is an important part of the GARCH class of models since the modelling residuals (prediction error) from the mean model drives the variance. Hence, a misspecified mean process has the potential to lead to a misspecified variance process even when the form of the variance process is assumed to be known.

Yet, it is striking how little attention the GARCH literature actually pays to the mean process. In most cases, the mean is assumed to be a constant. Only in a small number of cases (to be examined below) does the literature consider more general mean processes. Even then, these are typically linear.

Linear models may provide reasonable approximations in some instances, but most real life processes are non-linear in nature, and linear models often fail to provide good approximations of such processes (Willey, 1992). Different financial returns series have been tested for non-linearity by various researchers (Hinich and Patterson, 1985) (Terui and Kariya, 1997) (Lima, 1998) (Urrutia et al., 2002) and strong evidence of non-linearity and non-Gaussianity has been found. This evidence suggests the need for a suitable non-linear mean process, and yet there are very few of these in the GARCH literature. A notable exception is Meitz and Saikkonen (2008) who proposed a non-linear autoregressive model to model the mean of the returns and developed an asymptotic estimation theory for non-linear AR-GARCH models. There is, thus, a major gap in the literature to be filled.

NARMAX models (Leontaritis and Billings, 1985a and 1985b) provide a natural solution to this problem. These have commonly been used by systems engineers to model non-linear systems found in the real world. NARMAX models describe the system as a set of non-linear difference equations that relate the input(s) to the output(s) whilst taking into account the measurement noise, modelling errors and unmeasured noise, all combined into one variable.

The NARMAX methodology works on the assumption that the input and output data are homoskedastic. Since financial returns are heteroskedastic in nature, traditional methods of identifying and estimating financial models using the NARMAX methodology are likely to yield erroneous (and especially biased) results. A solution for this problem is to use a Weighted Orthogonal Forward Regression (WOFR) algorithm for the estimation of GARCH mean models (Zhao, 2010). Thus, the WOFR algorithm provides a way to select the terms in a non-linear mean model and then estimate the respective coefficients in a context in which the noise process is heteroskedastic.

The purpose of this chapter is to showcase the different types of linear and non-linear mean models used in the financial returns literature and to demonstrate the effectiveness of using NARMAX methods to model the mean of financial returns.

Section 3.2 gives a brief overview of the most commonly used mean models.

Section 3.3 explains NARMAX models and the methods used to identify and estimate them. It also explains the WOFR algorithm in a NARMAX model that can be used to accurately model the mean returns process.

Section 3.4 showcases the impact of incorrectly modelling a non-linear mean process as a linear model and also demonstrates how the WOFR algorithm can be used to correctly identify a non-linear mean model.

Section 3.5 concludes the chapter.

3.2 Financial Mean Models

3.2.1 Linear Mean Models

Over the past 30 years, ARCH and GARCH models have paved the way to new and improved methods to model financial volatility. Much work has been done in order to capture the various effects exhibited by financial returns series and to model the variance of these returns with new types of GARCH models but not much importance has been given to the mean model that is used to model the conditional mean of the returns. As noted earlier, the literature focuses on modelling the variance and the assumed mean processes are typically very simplistic and usually just a simple constant.

When Engle introduced the ARCH model (Engle, 1982), a linear mean model was used which was only introduced along with the simulations. The model used was

$$\Delta p(t) = \beta_1 + \beta_2 \Delta p(t-1) + \beta_3 \Delta p(t-4) + \beta_4 \Delta p(t-5) \\ + \beta_5 \big(p(t-1) - w(t-1)\big) \qquad (3.1)$$

where $\Delta p(t)$ is the first difference of the log of the quarterly consumer price index, w is the log of the quarterly index of manual wage rates and $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are the coefficients to be estimated. Since, $\Delta p(t) = p(t) - p(t-1)$, equation (3.1) can be rewritten as

$$p(t) - p(t-1) = \beta_2\big(p(t-1) - p(t-2)\big) + \beta_3\big(p(t-4) - p(t-5)\big) \\ + \beta_4(p(t-5) - p(t-6)) + \beta_5 p(t-1) - \beta_5 w(t-1) + \beta_1$$

$$p(t) = (\beta_2 + \beta_5 + 1)p(t-1) - \beta_2 p(t-2) + \beta_3 p(t-4) + (\beta_4 - \beta_3)p(t-5) \\ - \beta_4 p(t-6) - \beta_5 w(t-1) + \beta_1 \qquad (3.2)$$

Equation (3.2) depicts a linear ARX (AutoRegressive with eXogenous inputs) model of $p(t)$ with

- Five autoregressive (AR) terms, $p(t-1), p(t-2), p(t-4), p(t-5)$ and $p(t-6)$ and
- One lagged exogenous input term, $w(t-1)$.

In 1986, Bollerslev introduced a more generalised version of the ARCH model (Bollerslev, 1986), aptly termed as the GARCH model. The mean model used along with the new variance model was

$$\pi(t) = \beta_1 \pi(t-1) + \beta_2 \pi(t-2) + \beta_3 \pi(t-3) + \beta_4 \pi(t-4) + \beta_5 + e(t) \qquad (3.3)$$

where $\pi(t) = 100 log\left(\frac{GD(t)}{GD(t-1)}\right)$, $GD(t)$ is the implicit price deflator for Gross National Product, $e(t)$ is the modelling residual and $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are the coefficients to be estimated. Equation (3.3) depicts a linear AR model with 4 autoregressive terms, $\pi(t-1), \pi(t-2), \pi(t-3)$ and $\pi(t-4)$.

In 1991, Nelson introduced one of the first non-linear versions of the GARCH model, known as the Exponential GARCH (EGARCH) model (Nelson, 1991). The mean model used along with this new non-linear variance model was

$$y(t) = \beta_1 + \beta_2 y(t-1) + \beta_3 h(t) + e(t) \qquad (3.4)$$

where $y(t)$ is the excess return, $h(t)$ is the variance, $e(t)$ is the modelling residual and β_1, β_2 and β_3 are the coefficients to be estimated. Equation (3.4) depicts a linear AR variance in mean model with 1 autoregressive term, $y(t-1)$. Whilst Scholes and Williams (1977) suggested an MA(1) model to model the returns, Lo and MacKinlay (Lo and MacKinlay, 1988) suggested an AR(1) model for the same whilst stating that such simple models are not adequate enough to explain the short-term autocorrelation behaviour of market indices (Nelson, 1991).

In 1993, Glosten, Jagannathan and Runkle introduced the GJR-GARCH model specifically designed to capture the leverage effect exhibited by financial data (Glosten et al., 1993). The model was a modified version of a simple GARCH-M model in which the mean model used was

$$y(t) = \beta_1 + \beta_2 h(t-1) + e(t) \tag{3.5}$$

where $y(t)$ is the excess return, $h(t)$ is the variance, $e(t)$ is the modelling residual and β_1 and β_2 are the coefficients to be estimated. Equation (3.5) depicts a simple variance in mean model with only a constant term.

Whilst linear mean models such as the ones given above have been used in the literature, most economists and researchers use the most basic mean model

$$y(t) = c + e(t) \tag{3.6}$$

where c is a constant term. This model has been used to do a comparative study of the performance of the GARCH(1,1) model, the QGARCH model and the GJR-GARCH model to model weekly stock market volatility (Franses et al., 1996).

3.2.2 Non-Linear Mean Models

Not much work has been done in the implementation of non-linear mean models. Campbell, Lo, and MacKinlay mention the use of a simple non-linear MA model (Campbell et al., 1997)

$$y(t) = \beta_1 e^2(t-1) + e(t) \tag{3.7}$$

where $y(t)$ is the excess return and $e(t)$ is the modelling residual.

In 1992, LeBaron proposed an exponential AR model to model the returns (LeBaron, 1992) as

$$y(t) = \beta_0 + \left(\alpha_0 + \alpha_1 exp\left(-\frac{h(t)}{\alpha_2}\right)\right) y(t-1) + e(t) \tag{3.8}$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual and $\alpha_0, \alpha_1, \alpha_2$ and β_0 are the coefficients to be estimated.

A similar model was adopted to model the US stock market index (Bollerslev et al., 1994). The excess returns are modelled as

$$y(t) = \beta_0 + \left(\beta_1 + \beta_2 exp\left(-\frac{h(t)}{\beta_3}\right)\right) y(t-1) + \beta_4 h(t) + e(t) \tag{3.9}$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual, $h(t)$ is the variance and $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are the coefficients to be estimated. The model in equation (3.9) differs from equation (3.8) by an additional term, $\beta_4 h(t)$.

In 1992, Cao and Tsay used a Threshold AR model to model the returns (Cao and Tsay, 1992). The model is a piecewise linear model

$$y(t) = \begin{cases} \beta_0 + \beta_1 y(t-1) + \beta_2 y(t-2) + \beta_3 y(t-3) + e(t)\,, & y(t) < T \\ \alpha_0 + \alpha_1 y(t-1) + \alpha_2 y(t-2) + \alpha_3 y(t-3) + e(t)\,, & y(t) \geq T \end{cases} \tag{3.10}$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual, T is a threshold value and $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2$ and β_3 are the coefficients to be estimated. The model described in equation (3.10) can be combined together and represented as one using the logistic Smooth Transition Auto Regressive (STAR) function

$$L = -\frac{1}{1 + exp\big(-\lambda(y(t) - T)\big)} \tag{3.11}$$

The model can be rewritten as

$$\begin{aligned} y(t) = {} & \alpha_0 + \alpha_1 y(t-1) + \alpha_2 y(t-2) + \alpha_3 y(t-3) + \frac{-(\beta_0 - \alpha_0)}{1 + \exp\big(-\lambda(y(t) - T)\big)} \\ & + \frac{-(\beta_1 - \alpha_1)}{1 + \exp\big(-\lambda(y(t) - T)\big)} y(t-1) + \frac{-(\beta_2 - \alpha_2)}{1 + \exp\big(-\lambda(y(t) - T)\big)} y(t-2) \\ & + \frac{-(\beta_3 - \alpha_3)}{1 + \exp\big(-\lambda(y(t) - T)\big)} y(t-3) + e(t) \end{aligned} \tag{3.12}$$

Hence, the piecewise linear model in equation (3.10) is actually a model of non-linear nature represented in equation (3.12).

Another non-linear AR(p)-GARCH(1,1) model was introduced (Meitz and Saikkonen, 2008) where the mean model is a non-linear AR(p) model of the form

$$y(t) = \alpha_0 + \beta_0 F(y(t-d); \lambda_0) + \sum_{j=1}^{p} \left(\alpha_j + \beta_j F(y(t-d); \lambda_0)\right) y(t-j) + e(t) \quad (3.13)$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual, and $\alpha_0, \alpha_1, \dots, \alpha_p, \beta_0, \beta_1, \dots, \beta_p$ are the coefficients to be estimated. p is a user selected integer and d is fixed known integer between 1 and p. F is a non-linear function depending on $y(t-d)$ and takes values in [0, 1].

Meitz and Saikkonen emphasise the need of non-linear mean models and identified a major gap in the research, i.e. the absence of a good mean process.

This is where NARMAX models come in.

3.3 NARMAX Models

3.3.1 Non-Linear System Identification

System identification is the process of applying a set of known input signals to an unknown (real) system, recording the output signals and using this input-output data set to estimate a mathematical description of the underlying dynamic behaviour of the system.

The modelling process involves a number of stages. The first is model selection, the second is parameter estimation. Next, the estimated model is validated to ensure that system output can be predicted and the model can accurately generate a desired output when supplied with a known input. Finally, if the model is adequate, the behaviour of the system can be analysed and predicted.

System identification methods are routinely applied to develop financial models that are used for forecasting and process optimisation. As most systems in the real world are non-linear, linear system identification methods are complemented by powerful non-linear approaches such as the NARMAX method, which over the past few decades has been used as a *de facto* standard approach in non-linear system identification. The following section provides a brief overview of this methodology.

3.3.2 The NARMAX Methodology

NARMAX (Non-Linear AutoRegressive Moving Average with eXogenous inputs) models are one of the most well-known and effective methods for the identification and modelling of non-linear systems (Leontaritis and Billings, 1985a and 1985b). The NARMAX representation is a non-linear mapping which relates the output to past values of input(s), output(s) and noise.

Let

- $y(t)$ represent the output, $u(t)$ represent the value of the input and $e(t)$ represent the noise, all at an instant in time, t.
- n_y, n_u and n_e represent the maximum time lags of the output, input and errors respectively.
- d represents the time delay of the input.

The generalised representation of the NARMAX model is as follows:

$$y(t) = F\big(y(t-1), \dots, y(t-n_y), u(t-d), \dots, u(t-n_u), e(t-1), \dots, e(t-n_e)\big) + e(t) \tag{3.14}$$

where $F(\,)$ is an unknown non-linear mapping and is often implemented as a multivariable polynomial.

NARMAX models can be represented using non-linear mappings other than polynomial functions, like output-affine and rational models (Chen and Billings, 1989). However, this study considers only the polynomial representation.

The first step in NARMAX modelling is the concept of model structure selection, also known as term selection. This is important because a model that is more complex than required will over-fit the data rather than describe the underlying dynamics.

The next step is parameter estimation, which involves estimating the coefficients of the terms selected in the model in the previous step. Once the model has been estimated, the third step is to validate it effectively to ensure that it replicates the system and can accurately generate a desired output when supplied with a known input.

There are a number of simplified versions of the NARMAX model that are commonly used in practice.

- Non-Linear Autoregressive (NAR) – depends only on lagged values of output.

$$y(t) = F\left(y(t-1), \dots, y(t-n_y)\right) + e(t)$$

- Non-Linear Autoregressive Moving Average (NARMA) – depends on lagged values of output and prediction error.

$$y(t) = F\left(y(t-1), \dots, y(t-n_y), e(t-1), \dots, e(t-n_e)\right) + e(t)$$

- Non-Linear Autoregressive with exogenous inputs (NARX) – depends on lagged values of output and input.

$$y(t) = F\left(y(t-1), \dots, y(t-n_y), u(t-d), \dots, u(t-n_u)\right) + e(t)$$

Assuming that $F(\,)$ is a polynomial, a NAR model can be specified in terms of the polynomial order and the maximum output lag. For example, a NAR(2,3) model implies that $F(\,)$ is a second order polynomial and the maximum output lag in the model, n_y, is 3.

The conventional methods of term selection, parameter estimation and model validation have been developed under the assumption that the additive noise is homoskedastic but this is not the case when dealing with financial time-series data. However, recent studies have revealed that standard model selection algorithms fail to correctly identify the correct model structure when the noise is heteroskedastic (Zhao, 2010). Fortunately, we can remedy this problem using by replacing the OFR algorithm with the WOFR algorithm developed by Zhao (2010).

3.3.3 Structure Determination and Parameter Estimation using the Orthogonal Forward Regression Algorithm

The Weighted Orthogonal Forward Regression (WOFR) algorithm and the weighted ERR approach used later in the simulations in this thesis are based on the Orthogonal Forward Regression algorithm (Korenberg et al., 1988).

Polynomial NARMAX models are linear-in-the-parameter models that can be written as

$$y(t) = \sum_{i=1}^{M} \theta_i p_i\left(x(t)\right) + e(t), \quad t = 1, 2, \dots, N \tag{3.15}$$

where

- N is the data length,

- θ_i are the unknown parameters to be estimated,
- $p_i(.)$ are the selected polynomial model terms,
- $x(t)$ is a vector of lagged input, output and error variables, and
- $e(t)$ is the modelling error.

The above model can be written in matrix form as:

$$Y = P\Theta + \Xi \tag{3.16}$$

where

- $Y = [y(1), y(2), \dots, y(N)]^T$
- $P = [p_1, p_2, \dots, p_M]$
- M is the number of terms in the model.
- $p_i = [p_1(x(1)), p_2(x(2)), \dots, p_N(x(N))]^T$
- $\Theta = [\theta_1, \theta_2, \dots, \theta_M]$
- $\Xi = [e(1), e(2), \dots, e(N)]^T$

Since the correlation matrix P^TP is symmetric and positive definite, using the matrix decomposition theorem (Fox, 1964),

$$P^TP = A^TDA \tag{3.17}$$

where A is a M x M unit upper triangular matrix and D is a diagonal matrix all of whose elements are positive. Equation (3.16) can be rewritten as

$$Y = P(A^{-1}A)\Theta + \Xi = WG + \Xi \tag{3.18}$$

where

$$W = PA^{-1}, G = A\Theta \tag{3.19}$$

W is an N x M matrix with orthogonal columns $w_1, w_2, \dots, w_M$.

$$W^TW = (PA^{-1})^T(PA^{-1}) = (A^{-1})^T(P^TP)A^{-1} = (A^T)^{-1}A^TDAA^{-1} = D \tag{3.20}$$

$$W = \begin{bmatrix} w_0(1) & w_1(1) & \cdots & w_M(1) \\ w_0(2) & w_1(2) & \cdots & w_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ w_0(N) & w_1(N) & \cdots & w_M(N) \end{bmatrix} \tag{3.21}$$

Hence

$$D = \begin{bmatrix} \sum_{t=1}^{N} {w_0}^2(t) & 0 & 0 & 0 \\ 0 & \sum_{t=1}^{N} {w_1}^2(t) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sum_{t=1}^{N} {w_M}^2(t) \end{bmatrix} \tag{3.22}$$

and $$D^{-1} = diag \frac{1}{\sum_{t=1}^{N} {w_i}^2(t)}, \; i = 0,1,\dots,M$$

$$WA = P \tag{3.23}$$

Pre-multiplying by W^T gives

$$W^T W A = W^T P \tag{3.24}$$

$$A = (W^T W)^{-1} W^T P = D^{-1} W^T P \tag{3.25}$$

Using the matrix definitions of $D, W \; and \; P$ from equations (3.22), (3.21) and (3.16) respectively,

$$A = \begin{bmatrix} \frac{\sum_{t=1}^{N} w_0(t) p_0(t)}{\sum_{t=1}^{N} {w_0}^2(t)} & \frac{\sum_{t=1}^{N} w_0(t) p_1(t)}{\sum_{t=1}^{N} {w_0}^2(t)} & \cdots & \frac{\sum_{t=1}^{N} w_0(t) p_M(t)}{\sum_{t=1}^{N} {w_0}^2(t)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{t=1}^{N} w_M(t) p_0(t)}{\sum_{t=1}^{N} {w_M}^2(t)} & \frac{\sum_{t=1}^{N} w_M(t) p_1(t)}{\sum_{t=1}^{N} {w_M}^2(t)} & \cdots & \frac{\sum_{t=1}^{N} w_M(t) p_M(t)}{\sum_{t=1}^{N} {w_M}^2(t)} \end{bmatrix}$$

$$A = \begin{bmatrix} \alpha_{00} & \alpha_{01} & \cdots & \alpha_{0M} \\ \alpha_{10} & \alpha_{11} & \cdots & \alpha_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{M0} & \alpha_{M1} & \cdots & \alpha_{MM} \end{bmatrix} \tag{3.26}$$

A is unit upper triangular and its terms are defined as follows:

$$\alpha_{ij} = \begin{cases} 0, & \forall\, i > j \\ 1, & \forall\, i = j \\ \dfrac{\sum_{t=1}^{N} w_i(t)p_j(t)}{\sum_{t=1}^{N} w_i{}^2(t)} & \forall\, i < j \end{cases} \tag{3.27}$$

giving

$$A = \begin{bmatrix} 1 & \alpha_{01} & \alpha_{02} & \cdots & \alpha_{0M} \\ 0 & 1 & \alpha_{12} & \cdots & \alpha_{1M} \\ 0 & 0 & 1 & \cdots & \vdots \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.28}$$

The elements of W can be determined by

$$WA = P$$

$$W = P - WA + W = P - W(A - I) \tag{3.29}$$

$$\begin{bmatrix} w_0(1) & w_1(1) & \cdots & w_M(1) \\ w_0(2) & w_1(2) & \cdots & w_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ w_0(N) & w_1(N) & \cdots & w_M(N) \end{bmatrix} = \begin{bmatrix} p_0(1) & p_1(1) & \cdots & p_M(1) \\ p_0(2) & p_1(2) & \cdots & p_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(N) & p_1(N) & \cdots & p_M(N) \end{bmatrix}$$

$$- \begin{bmatrix} w_0(1) & w_1(1) & \cdots & w_M(1) \\ w_0(2) & w_1(2) & \cdots & w_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ w_0(N) & w_1(N) & \cdots & w_M(N) \end{bmatrix} \times \begin{bmatrix} 0 & \alpha_{01} & \alpha_{02} & \cdots & \alpha_{0M} \\ 0 & 0 & \alpha_{12} & \cdots & \alpha_{1M} \\ 0 & 0 & 0 & \cdots & \vdots \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.30}$$

$G = [g_1, g_2, \dots, g_M]^T$ is an auxiliary parameter vector and the estimate of G can be calculated using

$$\hat{G} = (W^T W)^{-1} W^T Y = D^{-1} W^T Y \tag{3.31}$$

or

$$\hat{g}_i = \frac{\langle Y, w_i \rangle}{\langle w_i, w_i \rangle} \quad , \qquad i \;=\; 1,2, \dots, M \tag{3.32}$$

where $\langle .\,,. \rangle$ denotes the inner product of 2 vectors and $\langle Y, w_i \rangle = \sum_{t=1}^{N} w_i(t) y(t)$

$$\hat{G} = A\hat{\Theta}$$

$$\hat{\Theta} = A^{-1}\hat{G} \tag{3.33}$$

The prediction errors are not known beforehand and are calculated using

$$\hat{e}(t) = y(t) - \sum_{i=0}^{N} \hat{g}_i \, w_i(t) \tag{3.34}$$

If $e(t)$ is uncorrelated with the past outputs, the output variance can be written as

$$\frac{1}{N} Y^T Y = \frac{1}{N} \sum_{i=1}^{M} g_i^2 \, w_i^T w_i + \frac{1}{N} \Xi^T \Xi \tag{3.35}$$

Here, $\frac{1}{N}\sum_{i=1}^{M} g_i^2 \, w_i^T w_i$ represents the desired output and $\frac{1}{N}\Xi^T \Xi$ represents the unexplained variance. Hence, $\frac{1}{N}\sum_{i=1}^{M} g_i^2 \, w_i^T w_i$is the increment to the explained desired output variance brought by w_i which leads to the definition of the ERR_i as

$$ERR_i = \frac{g_i^2 \langle w_i, w_i \rangle}{\langle Y, Y \rangle} \times 100\% = \frac{\langle Y, w_i \rangle^2}{\langle Y, Y \rangle \langle w_i, w_i \rangle} \times 100\%, \;\; i \; = \; 1,2, \ldots, M, \tag{3.36}$$

The Error Reduction Ratio (ERR) is often used in structure determination algorithms to determine which term should be included in the NARMAX model. The higher the ERR, the more likely it is that the term should be included in the model. There is often a cut-off value which specifies that if a term has an ERR value lower than that of the cut-off, the term should not be included in the model.

The OFR algorithm can be summarised in the following steps (Korenberg et al., 1988):

1. Select the values for n_y, n_u, n_e and d in (3.14). Set $e(t) = 0, t = 1, \ldots, N$. Select the threshold values for the ERR.

2. Estimate all the parameters of the terms which do not include any $e(.)$ terms by computing the elements of matrices W in equation (3.21), A in equation (3.26) and $\hat{G}$ in equation (3.31).

3. If $\hat{e}(t) = 0, t = 1, \ldots, N$ go to step 4 else use $\hat{e}(t)$ to estimate the parameters associated with the prediction error terms by computing W, A and $\hat{G}$.

4. Calculate ERR_i in equation (3.36), check against the thresholds and remove the insignificant terms.

5. Estimate the prediction errors using equation (3.34).

6. If any process model terms were deleted in step 4 then go to step 2 otherwise go to step 3 and repeat until convergence.

7. Estimate the NARMAX model coefficients by calculating $\hat{\Theta}$ from equation (3.33).

The procedure of orthogonal decomposition can be carried out using several algorithms such as Gram-Schmidt, modified Gram-Schmidt and the Householder transformation.

Although the ERR value indicates the terms to be included in the model, there can be instances when terms that are insignificant according to their ERR value will introduce bias if excluded from the final model. Simulations (Korenberg et al., 1988) show that this usually occurs only with the noise model or prediction error terms. A majority of the prediction error terms have a very low ERR value, which if deleted can cause the sequence $e(t)$ to become autocorrelated rather than white, thereby inducing possible bias in the model parameters.

3.3.4 Weighted Orthogonal Forward Regression (WOFR) Method

This section describes the algorithm of the WOFR procedure (Zhao, 2010) used in the simulations in this chapter.

Algorithm:

1. A non-linear candidate model is built which consists of all the terms that might possibly be a part of the original model. The OFR algorithm is then applied to the candidate model to calculate the parameter estimates of each term. Once this is done, one-step-ahead predictions of the returns are calculated which in turn are used to calculate the residuals by subtracting the one-step-ahead estimates from the actual return at that instant in time.

$$e(t) = y(t) - \hat{y}(t) \tag{3.37}$$

2. Once the residuals are calculated, an estimate of the variance is derived by inputting the residuals into a GARCH estimation algorithm. This algorithm estimates the GARCH parameters and the variance.

3. Each term in the candidate model is weighed by the square root of the estimated variance at each instant in time. The original model is

$$y(t) = \sum_{i=1}^{M} g_i w_i + z(t)\sqrt{h(t)} \quad ,\mathrm{t} \ = \ 1{,}2{,}\ldots,\mathrm{N} \tag{3.38}$$

After weighing each term, the model becomes:

$$\frac{y(t)}{\sqrt{h(t)}} = \sum_{i=1}^{M} \frac{g_i w_i}{\sqrt{h(t)}} + z(t) \quad ,\mathrm{t} \ = \ 1{,}2{,}\ldots,\mathrm{N} \tag{3.39}$$

4. The OFR algorithm is again applied to the weighted model and the ERR of each term is calculated. Note that this ERR will be the weighted ERR. A cut-off value is set and all the terms with an ERR above the cut-off value are selected. Thus ends the term selection phase.

5. The candidate model now consists only of the terms selected in the previous step. The OFR algorithm is applied to this model to obtain the correct parameter estimates.

6. Once the parameter estimates are obtained, the residuals are calculated and fed into the GARCH estimation algorithm which in turn calculates the new variance and new GARCH parameter estimates. The new variance is then used to weight the candidate model in step 4 and recalculate the parameter estimates of the mean model, the variance and eventually the parameter estimates of the variance model as well.

7. This procedure is then repeated until the parameter estimates of the mean and variance model converge to a fixed value. This occurs after about 10 iterations.

3.4 Simulations

A majority of the research papers which introduce and implement various GARCH-class models use a constant or a linear model to model the returns.

The following simulations demonstrate the impact of incorrectly fitting a linear mean model to a non-linear return series on variance estimation and also demonstrate the WOFR method described in the previous sections to correctly identify linear and non-linear mean models in the presence of heteroskedastic noise.

3.4.1 Impact of Incorrectly Fitting the Mean Model on Variance Estimation

Consider the following GARCH(1,1) model with a non-linear mean

$$y(t) = a_0 + a_1 y(t-1) + a_2 y^2(t-1) + e(t) \qquad (3.40)$$

$$h(t) = K + Gh(t-1) + Ae^2(t-1) \qquad (3.41)$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual and $h(t)$ is the variance, all at an instant in time, t. a_0, a_1 and a_2 are the parameters of the mean model and K, G and A are the parameters of the variance model. The values are listed in Table 3.1.

Table 3.1 Parameters of the Simulated GARCH(1,1) Model

Parameter of the Mean Model	Value	Parameter of the Variance Model	Value
a_0	5E-04	K	3E-07
a_1	0.1	G	0.924
a_2	-6	A	0.075

This model is assumed to be the 'true' model, i.e., that from which the data is generated. 5000 data points were generated and the first 1000 were discarded to avoid initial condition errors. The simulated returns and variance are shown in Figure 3.1. When simulating the returns and the variance from equations (3.40) and (3.41), the residuals $e(t)$ are modelled as $e(t) = z(t)\sqrt{h(t)}$, where $z(t)$ is a random independent and identically distributed (i.i.d) sequence that has zero mean and a variance of 1.

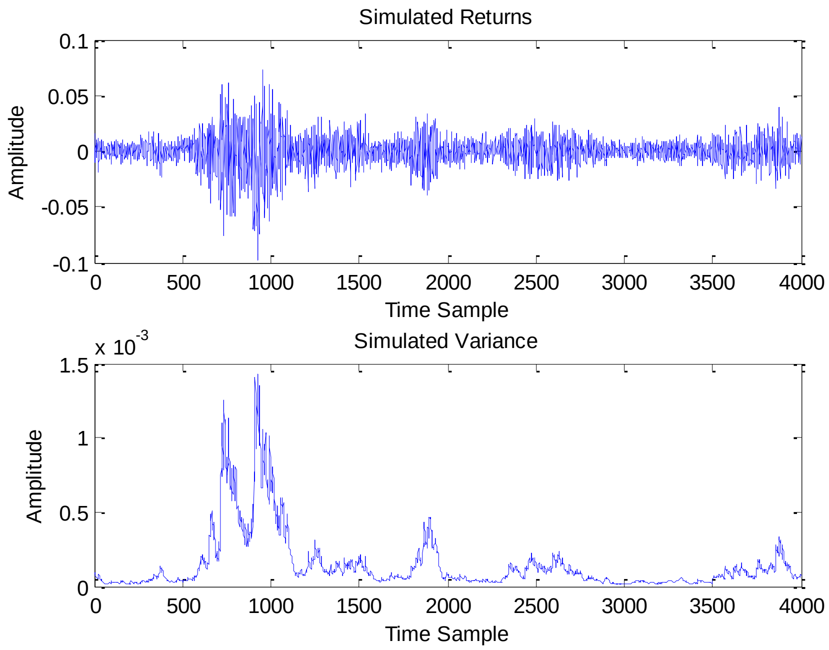


Figure 3.1 Simulated Returns and Variance of GARCH(1,1) Model

It is assumed that the modeller does not know that equation (3.40) is the true mean process, but instead incorrectly assumes that the mean process is the commonly used AR(1) model

$$y(t) = a_0 + a_1 y(t-1) + e(t) \tag{3.42}$$

It is also assumed that the modeller does however correctly believe that equation (3.41) is the variance process. The parameter estimates of the mean model are obtained and shown in Table 3.2. The one-step-ahead estimates of the returns are calculated and used to obtain the residuals, $e(t)$. Assuming that the structure of the variance model is known, a GARCH(1,1) model was fitted to these residuals to obtain an estimate of the variance (shown in Figure 3.2) and to obtain the parameter estimates of the GARCH variance model (shown in Table 3.2).

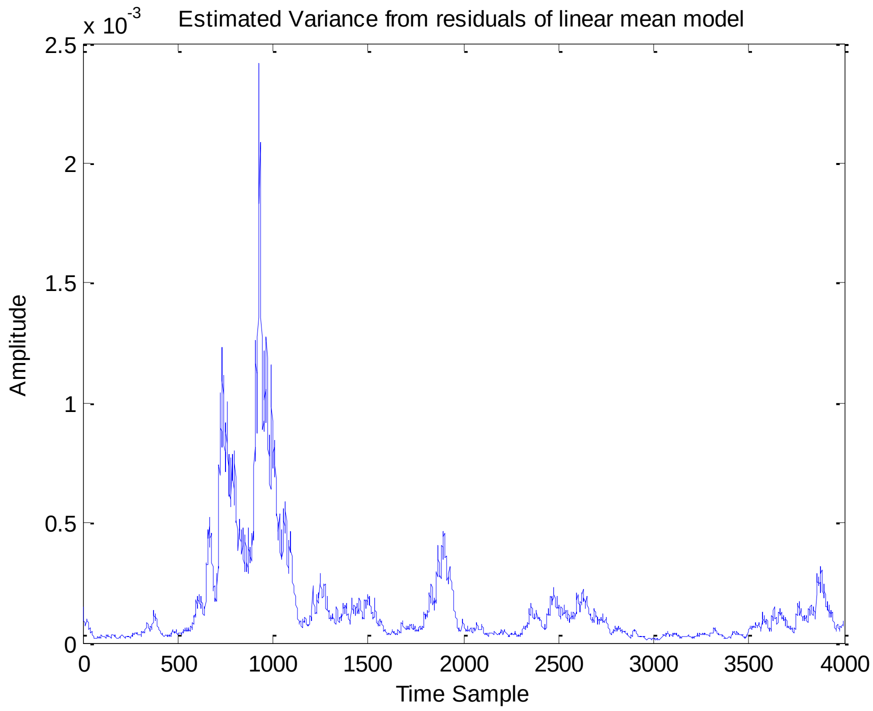


Figure 3.2 Estimated Variance after fitting Incorrect AR(1) Model to the Returns

Table 3.2 Parameter Estimates of the Estimated GARCH Model vs. True GARCH Model when the Incorrect Mean Model is fitted

Parameter of the Mean Model	Estimated Value	True Value
a_0	-3.069E-04	5E-04
a_1	0.1495	0.1
a_2	N/A	-6
Parameter of the Variance Model	**Estimated Value**	**True Value**
K	3.0948E-07	3E-07
G	0.9277	0.924
A	0.0711	0.075

The obtained parameter estimates in Table 3.2, a_0 and a_1, are clearly inaccurate.

In the second case, the mean model is correctly selected as

$$y(t) = a_0 + a_1 y(t-1) + a_2 y^2(t-1) + e(t) \tag{3.43}$$

The parameter estimates of the mean model are obtained and shown in Table 3.3. The one-step-ahead estimates of the returns are calculated and used to obtain the residuals, $e(t)$. Assuming that the structure of the variance model is known, a GARCH(1,1) model was fitted to these residuals to obtain an estimate of the variance (shown in Figure 3.3) and to obtain the parameter estimates of the GARCH variance model (shown in Table 3.3).

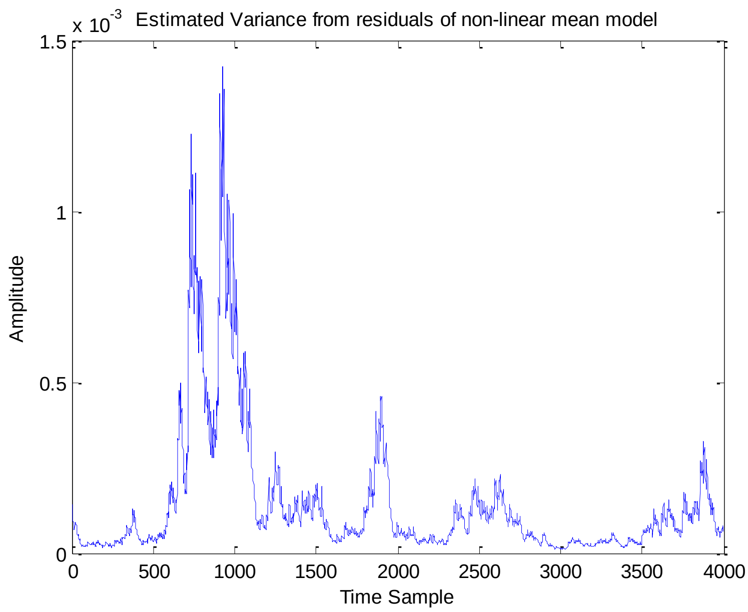


Figure 3.3 Estimated Variance after fitting the True Model to the Returns

Table 3.3 Parameter Estimates of the Estimated GARCH Model vs. True GARCH Model when the Correct Mean Model is fitted

Parameter of the Mean Model	Estimated Value	True Value
a_0	5.4157E-04	5E-04
a_1	0.0962	0.1
a_2	-5.7626	-6
Parameter of the Variance Model	**Estimated Value**	**True Value**
K	2.8604E-07	3E-07
G	0.9285	0.924
A	0.0705	0.075

The obtained parameter estimates in Table 3.3 are accurate and close to the true values.

The variance estimated from the residuals of the non-linear mean model (Figure 3.3) is noticeably closer to the simulated (true) variance than the variance estimated from the residuals of the linear mean model (Figure 3.2). To provide a better visualisation of this point, the absolute errors between the estimated GARCH(1,1) variance and the simulated (true) GARCH(1,1) variance for both the cases are shown in Figure 3.4.

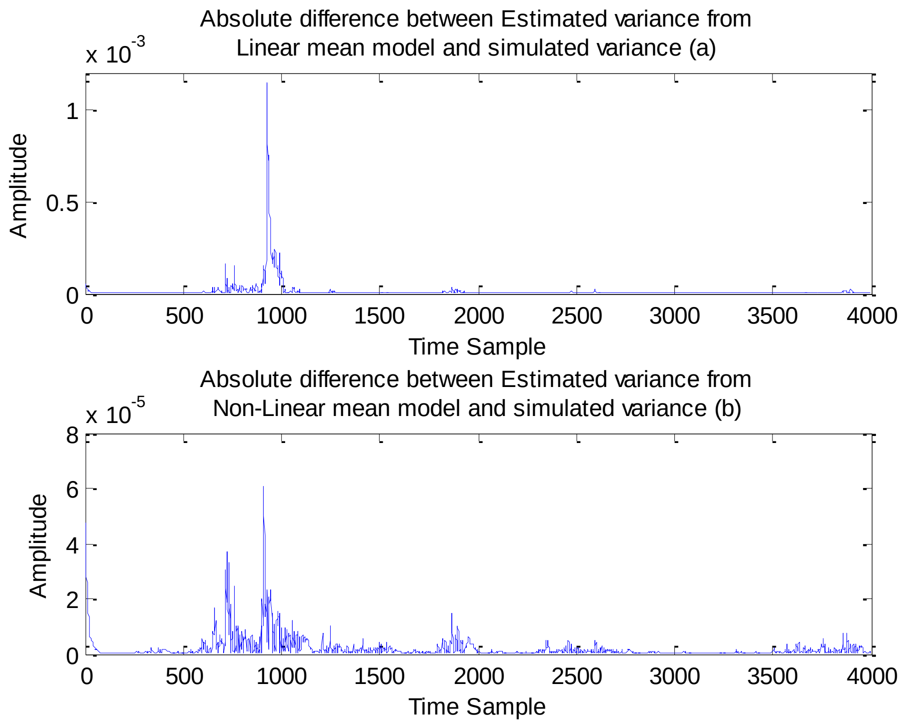


Figure 3.4 Absolute Differences between True GARCH(1,1) Variance and estimated GARCH(1,1) Variance from Residuals of (a) Linear and (b) Non-Linear Mean Model

The impact of selecting an incorrect mean model on the estimation of the variance model is evident from Figure 3.4 (a). Misspecification of the mean model can lead to inaccurate estimates of the variance, especially at the peaks (high volatility events) – as the spike in the upper sub-figure demonstrates. Figure 3.4 (b) shows that the magnitude of the absolute error is very small when the correct mean model is fitted.

In summary, if the modeller correctly assumes both the mean and the variance process, then the NARMAX OFR approach gives accurate results, but if the modeller is mistaken about the mean process, then the estimated variance process can be misspecified, even if the modeller is correct about the variance process itself.

This is where the WOFR method comes into play. Instead of making a guess about the mean process – which may or may not be correct – it is now assumed that the modeller proposes a very general candidate mean process and then applies the WOFR method.

3.4.2 Term Selection and Parameter Estimation of the Mean Model using WOFR

Consider the same GARCH(1,1) model with a non-linear mean as used in Section 3.4.1 (See equations (3.40) and (3.41), and Table 3.1).

The procedure followed is the same as that described in Section 3.3.4. First, a non-linear candidate model is chosen which consists of all the terms that could possibly be in the final mean model. A NAR(2,5) model is chosen as the candidate model.

$$
\begin{aligned}
y(t) = {} & a_1 y(t-1) + a_2 y(t-2) + a_3 y(t-3) + a_4 y(t-4) + a_5 y(t-5) \\
& + a_6 y^2(t-1) + a_7 y^2(t-2) + a_8 y^2(t-3) + a_9 y^2(t-4) \\
& + a_{10} y^2(t-5) + a_{11} y(t-1)y(t-2) + a_{12} y(t-1)y(t-3) \\
& + a_{13} y(t-1)y(t-4) + a_{14} y(t-1)y(t-5) + a_{15} y(t-2)y(t-3) \\
& + a_{16} y(t-2)y(t-4) + a_{17} y(t-2)y(t-5) + a_{18} y(t-3)y(t-4) \\
& + a_{19} y(t-3)y(t-5) + a_{20} y(t-4)y(t-5) + a_{21}
\end{aligned} \tag{3.44}
$$

Next, simple OFR is applied to the candidate model and the results are shown in the left-hand side of Table 3.4. This shows the terms selected in decreasing order of ERR. The one-step-ahead residuals of the mean model are calculated. The structure of the variance model was assumed to be known and a GARCH(1,1) model was fitted to these residuals to obtain an estimate of the GARCH variance.

The square root of the estimated GARCH variance is used to weight each term in the candidate mean model on the left and the right hand side of equation (3.44). Once again, simple OFR was applied to the weighted candidate mean model, the results of which are shown in the right-hand side of Table 3.4. This shows the terms selected in decreasing order of ERR.

Table 3.4 Ranking of Terms of the Candidate Mean Model when using OFR and WOFR

Rank	Standard OFR		Weighted OFR	
	Term	ERR (%)	Term	ERR (%)
1	$y^2(t-1)$	55.2096	$y^2(t-1)$	10.0976
2	$y(t-1)$	9.0819	$y(t-1)$	6.4782
3	$y(t-1)y(t-2)$	3.2117	1	4.3895
4	$y^2(t-3)$	1.4619	$y(t-2)$	0.8200
5	$y^2(t-2)$	1.8652	$y(t-1)y(t-2)$	0.4903
6	1	1.3808	$y^2(t-4)$	0.4280
7	$y^2(t-5)$	1.2395	$y(t-3)$	0.2619
8	$y(t-2)$	1.0791	$y(t-2)y(t-5)$	0.2273
9	$y(t-3)y(t-5)$	0.5381	$y(t-5)$	0.1376
10	$y(t-3)y(t-4)$	0.7534	$y(t-3)y(t-4)$	0.1296
11	$y(t-3)$	0.4844	$y(t-3)y(t-5)$	0.1543
12	$y(t-2)y(t-3)$	0.5886	$y^2(t-2)$	0.1069
13	$y(t-2)y(t-4)$	0.3771	$y^2(t-3)$	0.0784
14	$y(t-1)y(t-4)$	0.4395	$y(t-2)y(t-3)$	0.0561
15	$y^2(t-4)$	0.2590	$y(t-1)y(t-4)$	0.0411
16	$y(t-4)y(t-5)$	0.3221	$y(t-1)y(t-3)$	0.0324
17	$y(t-4)$	0.1537	$y(t-4)$	0.0156
18	$y(t-1)y(t-3)$	7.5016E-03	$y(t-4)y(t-5)$	0.0092
19	$y(t-2)y(t-5)$	3.4853E-03	$y(t-1)y(t-5)$	0.0016
20	$y(t-5)$	4.8341E-04	$y(t-2)y(t-4)$	1.2196E-04
21	$y(t-1)y(t-5)$	1.3478E-04	$y^2(t-5)$	3.9431E-05

The OFR algorithm ranks the constant term 6th, and ranks 3 terms ($y^2(t-2)$, $y^2(t-3)$ and $y(t-1)y(t-2)$) above the constant term that are not present in the original GARCH mean model (equation (3.40)). After weighting, the terms that are present in the original GARCH mean model are correctly selected. The ERR values of the top 3 ranked terms are significantly higher than the other terms in the candidate mean model. Hence these terms can be selected by setting the cut-off value to 2% (Wei and Billings, 2004). In short, OFR selects the wrong terms, and WOFR selects the right ones.

The parameters of the selected terms in the candidate mean model and the GARCH(1,1) variance model are re-estimated iteratively until convergence. The parameter estimates of the mean and the variance model converge after 10 iterations and the associated values are listed in Table 3.5.

Table 3.5 Parameter Estimates of Selected Mean Model and GARCH(1,1) Variance Model after 10 iterations of WOFR

Term of the Mean Model	Parameter Estimate	True Coefficient
Constant	5.2414E-04	5E-04
$y(t-1)$	0.0957	0.1
$y^2(t-1)$	-5.2413	-6
Parameter of Variance Model	**Parameter Estimate**	**True Coefficient**
K	2.7957E-07	3E-07
G	0.9282	0.924
A	0.0707	0.075

Thus, the WOFR algorithm correctly identifies and accurately estimates the true mean model even when the modeller was not assumed to know what the correct mean model actually was. To validate the performance of the WOFR algorithm, the estimated GARCH(1,1) variance obtained after 10 iterations is plotted against the simulated GARCH(1,1) variance along with the absolute difference between them. These are shown in Figure 3.5.

Figure 3.5 shows that the estimated GARCH(1,1) variance is extremely close to the simulated GARCH(1,1) variance and the absolute difference between the two is minimal.

If a GARCH model has been adequately fitted, the estimated standardised residuals, $\hat{z}(t) = \frac{\hat{e}(t)}{\sqrt{\hat{h}(t)}}$, are independent and identically distributed (i.i.d), where $\hat{e}(t)$ are the estimated residuals from the GARCH mean model and $\hat{h}(t)$ is the estimated GARCH variance. This implies that the autocorrelation of the squared estimated standardised residuals, $\hat{z}^2(t)$, should lie below the 95% significance boundary. The autocorrelation function for 50 lags is shown in Figure 3.6. It can be seen that the sample autocorrelations for all lags lie on or below the 95% significance level implying that the GARCH model has been adequately fitted. Thus, the estimated standardised residuals obtained from the WOFR algorithm pass validation tests for i.i.d behaviour.

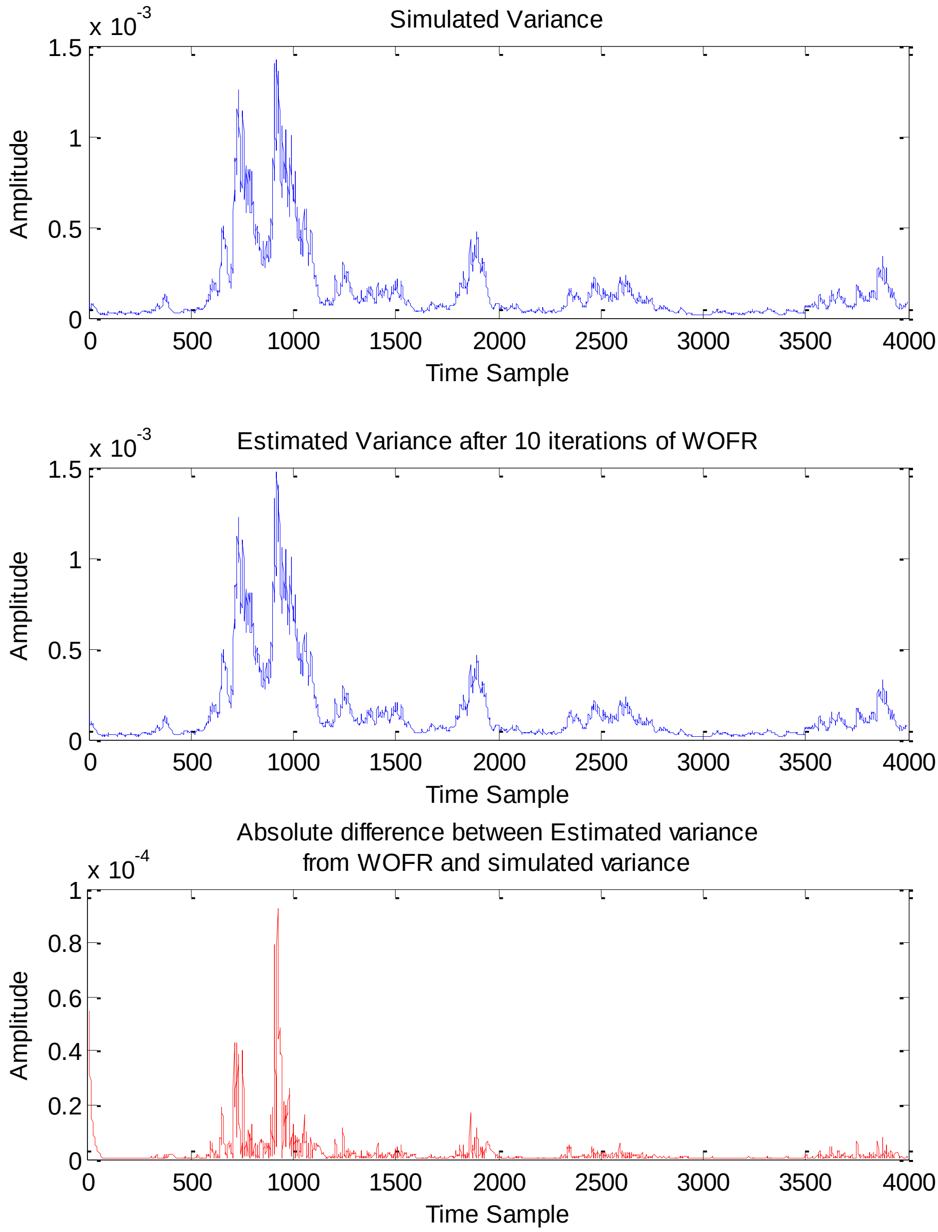


Figure 3.5 Simulated GARCH(1,1) Variance vs. Estimated GARCH(1,1) Variance from WOFR vs. Absolute Difference

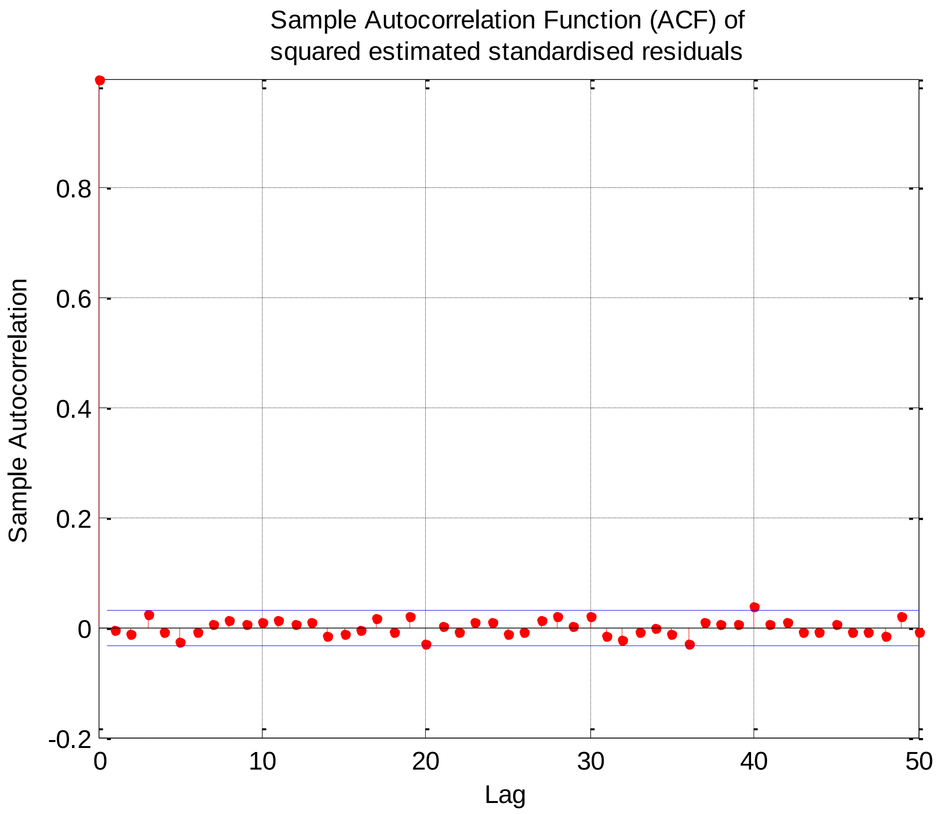


Figure 3.6 Sample Autocorrelation Function (ACF) of Squared Estimated Standardised Residuals for 50 Lags

3.5 Conclusions

This chapter introduces the financial mean model and the different types of linear and non-linear mean models used in the literature. It also introduces NARMAX models and the WOFR method of selecting the terms and estimating the parameters in a GARCH model. The simulations demonstrate the effect that fitting an incorrect linear model to returns that are actually described by a non-linear model has on variance estimation and also how the WOFR algorithm is used to correctly select the terms and estimate the parameters of a non-linear mean model. However, these results are predicated on the assumption that the variance model is known. How this latter assumption might be relaxed is examined next.

Chapter 4

Extended Weighted Orthogonal Forward Regression for the Estimation of the Mean Model

4.1 Introduction

In this chapter, a framework to model the mean of the returns is introduced, which does not require the modeller to know the terms present in the true mean model.

The framework aims to select the best type of mean model (constant, linear or non-linear) for a given return series without making any assumptions about the modeller knowing the true structure of the variance model. The framework will first be tested on a return series simulated from a known mean and variance model. Using the framework, the best mean model for the simulated return series will be selected, which will be compared to the true mean model of the series to demonstrate that the framework actually works.

Four different types of mean models are fitted to the returns: - a constant mean model, a linear mean model, a second order non-linear mean model and a third order non-linear mean model. The WOFR algorithm introduced in Chapter 3 is used to perform model term selection and parameter estimation. The ARCH Test is used in addition to standard higher order residual analysis tests to evaluate the model adequacy. The effects of inaccurate mean model estimation on variance estimation are also examined.

This chapter is organised as follows:

Section 4.2 reviews the correlation tests used in this chapter for model selection and model validation.

Section 4.3 reviews Engle's ARCH Test and explains how model validation and model selection is carried out using the ARCH Test.

Section 4.4 reviews the Akaike Information Criterion (AIC) and explains how the term selection of the mean model is carried out using the AIC.

Section 4.5 explains in detail the method to fit several mean models to a given return series, and to select the best model.

Section 4.6 showcases the results of applying the method introduced in Section 4.5 on a simulated data set. The true mean model of the simulated data is known, and hence the results of mean model selection using the new method are verified by checking whether the true mean model is selected.

Section 4.7 concludes the chapter.

4.2 Correlation Tests

Linear autocorrelation plots of the standardised residuals, $z(t)$, the squared standardised residuals, $z^2(t)$, and squared residuals, $e^2(t)$, obtained after fitting a mean model have long been used to validate the fitted mean model. The sample autocorrelation of both, $z(t)$ and $z^2(t)$, should ideally lie within the 95% confidence bands, indicating that the respective $z(t)$ series is white. The sample autocorrelation of $e^2(t)$ should ideally lie outside the 95% confidence bands indicating the need to fit a variance model to the given return series. In financial literature, the autocorrelation of the residuals, $e(t)$, is ignored, but this is a key test used to check the adequacy of the fitted model in systems engineering. Hence, to examine the adequacy of the fitted mean model, the autocorrelation of the residuals will be plotted as well.

In addition to linear correlation tests, the correlation based non-linear model validation tests introduced by Billings and Zhu (Billings and Zhu, 1994) can also be used to validate a candidate mean model by calculating higher order autocorrelations for $e^2(t)$ and $z^2(t)$. Also, the non-linear correlations from the tests introduced by Billings and Zhu (Billings and Zhu, 1994) and quantified by Friederich (Friederich, 2011) can be used to compare the higher order autocorrelations of $e^2(t)$ and $z^2(t)$ obtained by fitting different mean models.

To understand these non-linear correlation tests, consider a system with outputs, $y(t)$, and residuals, $e(t)$. The non-linear correlation test for lag, τ, can be represented as (Billings and Zhu, 1994)

$$\Phi_{(Y\varepsilon)'(\varepsilon^2)'}(\tau) = \kappa\delta(\tau) \qquad \forall\tau \tag{4.1}$$

where

- $(.)'$ describes the element-wise products of the mean removed series within the brackets.
- $(Y\varepsilon)'$ represents the element-wise product of mean removed $y(t)$ and mean removed $e(t)$, for $t = 1, \dots, N$.
- $(\varepsilon^2)'$ represents the element-wise product of mean removed $e(t)$ and mean removed $e(t)$, for $t = 1, \dots, N$.
- κ is a constant such that $0 < \kappa < 1$, and $\delta(.)$ is the Kronecker Delta (Weisstein, 2010).

From equation (3.15), a generalised representation of the mean of the returns can be given as:

$$y(t) = \sum_{i=1}^{M} \theta_i p_i\big(x(t)\big) + e(t) \tag{4.2}$$

$$e(t) = z(t)\sqrt{h(t)}$$

where

- $y(t)$ is the return,
- $e(t)$ is the residual of the mean of the returns,
- $z(t)$ is a random, independent and identically distributed (i.i.d) sequence that has zero mean and a variance of 1, also denoted as the standardised residuals,
- $h(t)$ is the variance,
- θ_i are the unknown parameters to be estimated,
- $p_i(.)$ are the selected polynomial model terms, and
- $x(t)$ is a vector of lagged output and error variables.

While performing WOFR, all the terms in the mean model are divided by $\sqrt{h(t)}$. Equation (4.2) becomes

$$\frac{y(t)}{\sqrt{h(t)}} = \frac{1}{\sqrt{h(t)}} \sum_{i=1}^{M} \theta_i p_i\big(x(t)\big) + z(t) \tag{4.3}$$

where $z(t)$ is the standardised residual of the mean of the returns.

Let:

- the collection of $y(t)$ for $t = 1, \dots, N$ be denoted as the vector Y,
- the collection of $e(t)$ for $t = 1, \dots, N$ be denoted as the vector ε, and
- the collection of $z(t)$ for $t = 1, \dots, N$ be denoted as the vector Z.

To validate and test the adequacy of a selected mean model, the higher order correlation functions, $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}(\tau), \Phi_{(\varepsilon^2)'(\varepsilon^2)'}(\tau)$ and $\Phi_{(Z^2)'(Z^2)'}(\tau)$, are calculated and plotted. The various normalised cross-correlation functions for τ lags are calculated as (Billings and Voon, 1986)

$$\Phi_{(Y\varepsilon)'(\varepsilon^2)'}(\tau) = \frac{\sum_{k=1}^{N-\tau}\big(ye(k) - \overline{ye}\big)\big(e^2(k+\tau) - \overline{e^2}\big)}{\sqrt{\left(\sum_{k=1}^{N-\tau}(ye(k) - \overline{ye})^2\right)\left(\sum_{k=1}^{N-\tau}\big(e^2(k) - \overline{e^2}\big)^2\right)}} \tag{4.4}$$

$$\Phi_{(\varepsilon^2)'(\varepsilon^2)'}(\tau) = \frac{\sum_{k=1}^{N-\tau}\big(e^2(k) - \overline{e^2}\big)\big(e^2(k+\tau) - \overline{e^2}\big)}{\sqrt{\left(\sum_{k=1}^{N-\tau}\big(e^2(k) - \overline{e^2}\big)^2\right)\left(\sum_{k=1}^{N-\tau}\big(e^2(k) - \overline{e^2}\big)^2\right)}} \tag{4.5}$$

$$\Phi_{(Z^2)'(Z^2)'}(\tau) = \frac{\sum_{k=1}^{N-\tau}\big(z^2(k) - \overline{z^2}\big)\big(z^2(k+\tau) - \overline{z^2}\big)}{\sqrt{\left(\sum_{k=1}^{N-\tau}\big(z^2(k) - \overline{z^2}\big)^2\right)\left(\sum_{k=1}^{N-\tau}\big(z^2(k) - \overline{z^2}\big)^2\right)}} \tag{4.6}$$

If the outputs, $y(t)$, have been correctly modelled, the values of $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}(\tau)$ and $\Phi_{(Z^2)'(Z^2)'}(\tau)$ lie within a 95% confidence band that is calculated as $\pm\frac{1.96}{\sqrt{N}}$ (Billings and Zhu, 1994). Small violations of the confidence band for larger values of τ are acceptable, but significant violations at small lags indicate an inadequate model.

The higher order correlation violation statistics, $V_{(Y\varepsilon)'(\varepsilon^2)'}, V_{(\varepsilon^2)'(\varepsilon^2)'}$, and $V_{(Z^2)'(Z^2)'}$, are also calculated. The following objective functions quantify the magnitude of violation of the correlation functions (Friederich, 2011)

$$V_{(Y\varepsilon)'(\varepsilon^2)'} = \begin{cases} \sum_{\tau} \left[exp\left(\alpha \left(\left| \Phi_{(Y\varepsilon)'(\varepsilon^2)'} \right| - \frac{1.96}{\sqrt{N}} \right) \right) - 1 \right] & \forall\tau/_{\tau = 0} > \frac{1.96}{\sqrt{N}} \\ 0 & else \end{cases} \tag{4.7}$$

$$V_{(\varepsilon^2)'(\varepsilon^2)'} = \begin{cases} \sum_{\tau} \left[exp\left(\alpha \left(\left| \Phi_{(\varepsilon^2)'(\varepsilon^2)'} \right| - \frac{1.96}{\sqrt{N}} \right) \right) - 1 \right] & \forall\tau/_{\tau = 0} > \frac{1.96}{\sqrt{N}} \\ 0 & else \end{cases} \tag{4.8}$$

$$V_{(Z^2)'(Z^2)'} = \begin{cases} \sum_{\tau} \left[exp\left(\alpha \left(\left| \Phi_{(Z^2)'(Z^2)'} \right| - \frac{1.96}{\sqrt{N}} \right) \right) - 1 \right] & \forall\tau/_{\tau = 0} > \frac{1.96}{\sqrt{N}} \\ 0 & else \end{cases} \tag{4.9}$$

where $\alpha > 1$ is a user-defined constant that punishes larger confidence violations exponentially (Friederich, 2011). After much testing, the value of $\alpha = 3$ is found to work the best in most cases of interest. For an adequately fit model, $V_{(Y\varepsilon)'(\varepsilon^2)'} = 0$ and $V_{(Z^2)'(Z^2)'} = 0$, implying that there exist no violations of the confidence band.

4.3 Engle's ARCH Test

Engle's ARCH Test (Engle, 1982) is a Lagrange Multiplier test used to detect the presence of ARCH effects in a given time series.

Consider the ARCH model with $h(t) = h(z_t\alpha)$ where h is a linear differentiable function, $z_t = \left(1, e^2(t-1), \dots, e^2(t-m)\right)$, $e(t)$ is the residual of the mean of the returns and $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m)$ are constants.

Under the null hypothesis, $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$ and $h(t)$ is a constant denoted as h^0. Writing $\frac{\partial h(t)}{\partial \alpha} = h' z_t'$ where h' is the scalar derivative of h, the score and information can be written as

$$\left.\frac{\partial l}{\partial \alpha}\right|_0 = \frac{h'}{2h^0} \sum_t z_t' \left(\frac{e^2(t)}{h^0} - 1 \right) = \frac{h'}{2h^0} z' f^0,$$

$$\mathcal{I}^0_{\alpha\alpha} = \frac{1}{2} \left(\frac{h^{0\prime}}{h^0} \right)^2 E z' z,$$

and, hence, the LM test statistic can be estimated by

$$\xi^* = \frac{1}{2} f^{0\prime} z (z'z)^{-1} z' f^0 \tag{4.10}$$

where $z' = (z'_1, \dots, z'_T)$ and f^0 is the column vector of $\left(\frac{e^2(t)}{h^0} - 1\right)$.

The ARCH test yields a test statistic and a critical value for a given dimension, $m \geq 1$. If the test statistic is greater than the critical value, the presence of ARCH effects in the given series is confirmed, and an ARCH(p) model, where $p \geq m$, is required to describe the variance of the given series (Engle, 1982).

In this chapter, the ARCH Test is used for model comparison, and is performed on the estimated residuals, $\hat{e}(t)$, obtained after fitting different candidate mean models to the given returns series. The more adequately the mean of the returns has been modelled, the lower the amount of heteroskedasticity present in the estimated residuals, and hence, the lower the ARCH Test statistic of $\hat{e}(t)$. Hence, from this perspective, the best fit is the one with the lowest ARCH Test statistic of $\hat{e}(t)$.

4.4 The Akaike Information Criteria (AIC)

The Akaike Information Criterion (AIC) is often used for model validation, and enumerates the quality of a model fitted to a given data set (Akaike, 1974). In this chapter, AIC is used to aid in term selection of the mean model.

The terms of the candidate mean model are reordered in decreasing order of the Error Reduction Ratio (ERR) values after performing WOFR. Starting with the term with the highest ERR value, the mean model is selectively updated, adding one term at a time. For each mean model, the one-step-ahead (OSA) residuals are calculated.

Since the true structure of the variance is unknown, a linear estimate of the variance is to be generated from the obtained residuals. ARCH(p) models provide a basic linear estimate of the variance using just the lagged squared residuals. Since a good approximation of the variance is required, the lag, p, needs to be large. Various simulations with p ranging from 10 to 100 were performed, and p is chosen to be a sufficient value of 25, since it provides a good linear variance estimate whilst keeping computational time to generate the variance estimate low. p is not required to be exactly 25. Any value close to or above 25 may also be used.

Hence, an ARCH(25) variance model is then fitted to the OSA residuals, and the AIC of the fitted ARCH(25) model is calculated as

$$AIC = (-2 \times LLF) + (2 \times p) \tag{4.11}$$

where LLF is the Log-Likelihood Function and can be calculated using equations (6.5) and (6.6), and p is the total number of parameters of the fitted variance model (25, in this case). When used for model comparison, the model that yields the lowest AIC value is selected.

4.5 Overview of the Method

The method used is as follows:

1. Consider a total of N samples of returns, denoted as y.

$$Returns = y = [y(1), y(2), \dots, y(N)] \tag{4.12}$$

2. The complete data set is split into three sets – Estimation, Validation and Testing Set. n_{est} denotes the number of samples used in the Estimation set, n_{val} denotes the number of samples used in the Validation set, and n_{test} denotes the number of samples in the Testing set.

$$N = n_{est} + n_{val} + n_{test}$$

   Of the last $(n_{val} + n_{test})$ samples of y, the first n_{val} samples are used for term selection and model validation, and the last n_{test} samples are used to demonstrate out-of-sample performance of the fitted model.

3. Consider a candidate mean model consisting of all possible non-linear and/or linear combinations of lagged returns. n_{est} samples of returns are used to estimate the mean model. The returns are split into three data sets: an Estimation subset of n_{est} samples, a Validation set of n_{val} samples, and a Testing set of n_{test} samples.

$$y_{model} = [y(1), y(2), \dots, y(n_{est})] \tag{4.13}$$

$$y_{val} = [y(n_{est}+1), y(n_{est}+2), \dots, y(N - n_{test})] \tag{4.14}$$

$$y_{test} = [y(N - n_{test}+1), y(N - n_{test}+2), \dots, y(N)] \tag{4.15}$$

4. The coefficients of the terms in the candidate model are estimated, and the one-step-ahead (OSA) estimates of the returns over the estimation, validation, and testing data sets (denoted as $\hat{y}_{model}, \hat{y}_{val}$ and $\hat{y}_{test}$) are calculated. The initial residuals, denoted as $\hat{e}_{model}, \hat{e}_{val}$ and $\hat{e}_{test}$, are also calculated:

$$\hat{e}_{model} = y_{model} - \hat{y}_{model}$$
$$\hat{e}_{val} = y_{val} - \hat{y}_{val}$$
$$\hat{e}_{test} = y_{test} - \hat{y}_{test} \tag{4.16}$$

5. Since the true structure of the variance model is assumed to be unknown to the modeller, an ARCH(25) model is fitted to the residuals, $\hat{e}_{model}$. The coefficients of the ARCH(25) variance model (denoted as $K, A_1, A_2, \dots, A_{25}$) are estimated, and the OSA estimates of variance over the estimation, validation and testing sets (denoted as $\hat{h}_{model}, \hat{h}_{val}$ and $\hat{h}_{test}$) are obtained.

$$\hat{h}_{model}(t) = K + \sum_{q=1}^{25} A_q \hat{e}_{model}{}^2(t-q)$$

$$\hat{h}_{val}(t) = K + \sum_{q=1}^{25} A_q \hat{e}_{val}{}^2(t-q)$$

$$\hat{h}_{test}(t) = K + \sum_{q=1}^{25} A_q \hat{e}_{test}{}^2(t-q) \quad (4.17)$$

where $K, A_1, A_2, \dots, A_{25}$ are estimated via Maximum Likelihood.

6. WOFR is performed on the candidate mean model. The terms are re-ordered in decreasing order of ERR values due to weighting by $\sqrt{\hat{h}_{model}}$.

7. If there are 100+ terms in the candidate mean model, a reasonably small ERR cut-off value like 0.01% is selected to reduce computational time. Else, the ERR cut-off value is selected to be 0%. All the terms in the candidate mean model that have an ERR less than the ERR cut-off are discarded. Starting with only the term with the highest ERR value, the terms are iteratively added to constitute the mean model being tested.
In each iteration, the term with the next highest ERR in the candidate mean model is added to the mean model to be tested. A noise model is also fitted to the terms in the mean model being tested.

8. The one-step-ahead (OSA) estimates of the returns, and hence the OSA residuals over the Estimation, Validation and Testing Sets are calculated.
An ARCH(25) variance model is fitted to the OSA residuals in the Estimation Set, and the AIC is calculated using equation (4.11). The AIC decreases as terms are iteratively added to the mean model. The percentage change in AIC in that iteration relative to the previous iteration is calculated as

$$AIC_{\%}(i) = 100 \times \left(\frac{AIC(i)}{AIC(i-1)} - 1 \right) \tag{4.18}$$

where $AIC(i)$ represents the AIC of the fitted ARCH(25) model, and $AIC_{\%}(i)$ represents the percentage change in AIC, both in iteration, i.

9. Once the correct terms in the mean model have been selected, $AIC_{\%}$ becomes minimal. Hence, only those terms are selected in the mean model that yield a significant value of $AIC_{\%}$. If there exists an element of doubt, ERR is used for verification. If a term has a significant value of ERR (that is to say, an ERR greater than 0.1%), it is selected.

10. Steps 3 to 8 are carried out for the different candidate mean models that are fitted to the returns. For the purpose of this chapter, four different types of candidate mean models are fit to the returns – a constant mean model, a linear mean model, a second order polynomial non-linear mean model, and finally, a third order polynomial non-linear mean model. The ARCH Test statistics of the different obtained $\hat{e}(t)$ series for the Validation Set, and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the complete data set are calculated and compared for all the different types of selected mean models in order to determine the best mean model for the given return series.

4.6 Simulated Data

The method is first demonstrated using a numerical simulation example. An assumed known heteroskedastic volatility model is used to generate synthetic data where the mean process is described by a non-linear autoregressive model and the variance process by a GARCH(1,1) model. We consider the situation faced by a modeller who does not know the true mean and variance models and has only the data themselves to work with. The proposed estimation method is used to identify the mean model from the simulated returns.

4.6.1 The Model

A return series is simulated using the second order non-linear mean model and a simple GARCH(1,1) variance model given in equations (4.19) and (4.20).

Mean Model: $y(t) = a_0 + a_1 y(t-1) + a_2 y(t-3) + a_3 y(t-1) y(t-2)$

$$+a_4 y(t-5)y(t-5) + a_5 y(t-3)y(t-4) + e(t) \quad (4.19)$$

Variance Model: $$h(t) = K + Gh(t-1) + Ae^2(t-1) \quad (4.20)$$

where

- $y(t)$ is the return,
- $e(t)$ is the modelling residual and
- $h(t)$ is the variance, all at an instant in time, t.
- a_0, a_1, a_2, a_3, a_4 and a_5 are the parameters of the mean model and K, G and A are the parameters of the variance model.

When simulating the returns and the variance, a random i.i.d sequence, $z(t)$, is generated. $e(t)$ is simulated as $e(t) = z(t)\sqrt{h(t)}$. The true values of the parameters are listed in Table 4.1. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10).

Table 4.1 Model Parameters of the Simulated Data

Parameter of the Mean Model	Value	Parameter of the Variance Model	Value
a_0	0.001	K	3E-07
a_1	0.2	G	0.924
a_2	0.15	A	0.075
a_3	-10		
a_4	8		
a_5	-5		

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. The simulated returns and simulated variance are shown in Figure 4.1.

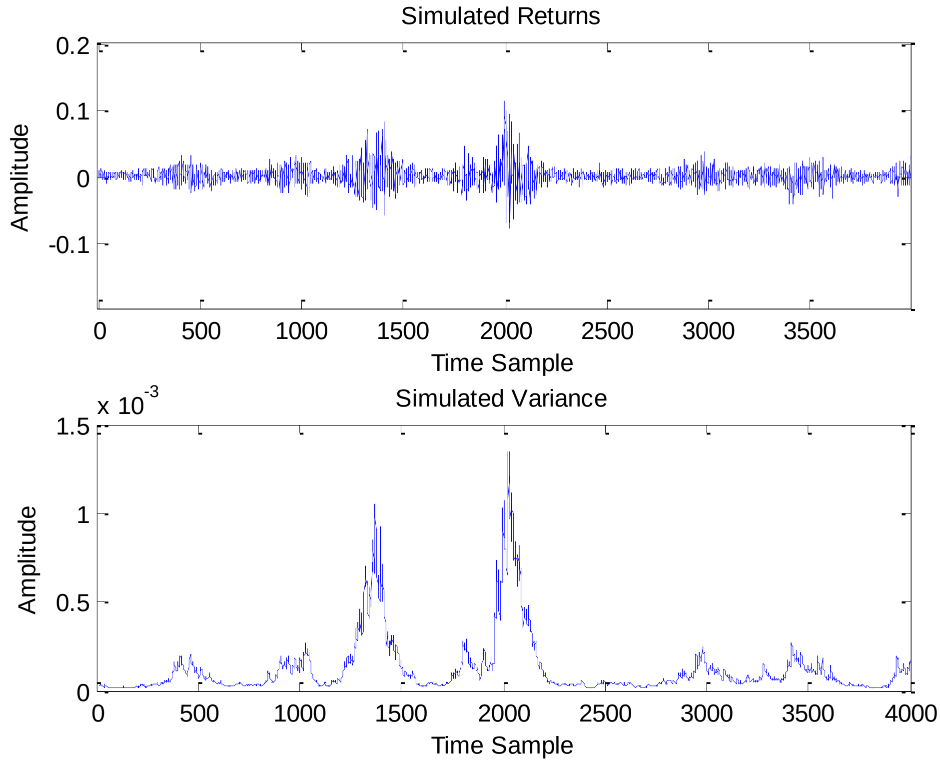


Figure 4.1 Simulated Returns and Simulated Variance

All the simulated data series are split into three sets – Estimation, Validation and Testing Sets. The number of samples in each of these sets are listed in Table 4.2.

Table 4.2 Number of Samples in Estimation, Validation and Testing Sets

No. of Samples in Estimation Set (n_{est})	3300
No. of Samples in Validation Set (n_{val})	500
No. of Samples in Testing Set (n_{test})	200

To capture the non-linearity present in the simulated data set, a large Estimation Set is required to enable accurate term selection and parameter estimation. Hence, a large Estimation Set comprising of 3300 samples is selected. For model validation results to be accurate and consistent, the Validation Set is chosen to have 500 samples. Out of sample model performance is tested with a Testing Set consisting of 200 samples.

4.6.2 Autocorrelation Plots of Returns and Squared Returns

4.6.2.1 Linear Autocorrelation

The linear sample autocorrelation of the returns and the squared returns for 20 lags are shown in Figure 4.2.

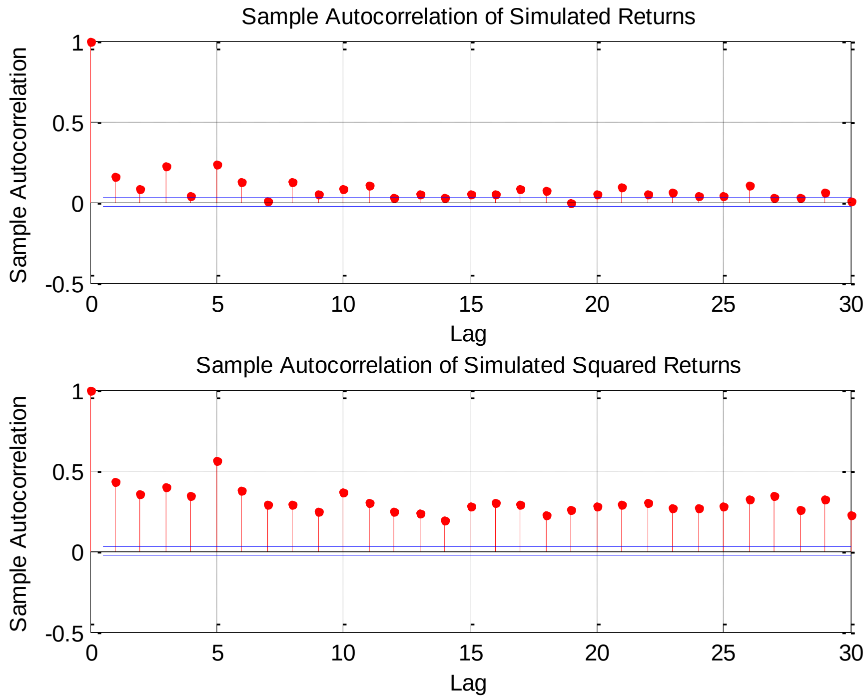


Figure 4.2 Sample Autocorrelation Plots of Returns and Squared Returns

The sample autocorrelation of the returns for lags 1, 3 and 5 lie significantly outside the 95% confidence bands, indicating the possibility of the presence of the terms, $y(t-1)$, $y(t-3)$ and $y(t-5)$, in the mean model. Note that the presence of non-linear terms cannot be indicated by linear autocorrelation plots.

The sample autocorrelation of the squared returns are significantly outside the 95% confidence bands for all the lags, indicating the need to fit a variance model to the returns, in addition to a mean model.

4.6.2.2 Higher Order Autocorrelation

The higher order autocorrelation of the returns and the squared returns for the Estimation, Validation, Testing and Complete Data Set are also calculated and plotted in Figure 4.3 and Figure 4.4 respectively.

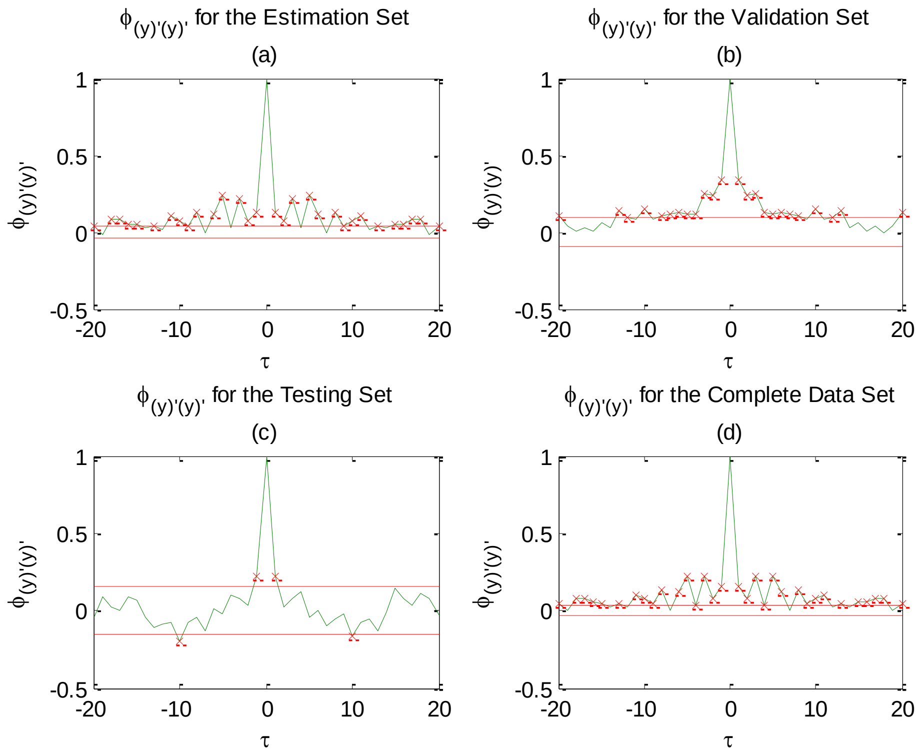


Figure 4.3 $\Phi_{(y)'(y)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Data Set

Figure 4.3 (d) shows that $\Phi_{(y)'(y)'}$ for lags 1, 3 and 5 is well outside the 95% confidence bands, suggesting the presence of linear and non-linear terms with these lags.

The corresponding higher order correlation violation statistic, $V_{(y)'(y)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set are listed in Table 4.3.

Table 4.3 $V_{(y)'(y)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y)'(y)'}$ (Estimation Set)	$V_{(y)'(y)'}$ (Validation Set)	$V_{(y)'(y)'}$ (Testing Set)	$V_{(y)'(y)'}$ (Complete Data Set)
6.0135E-06	1.6851E-05	6.3830E-07	5.8302E-06

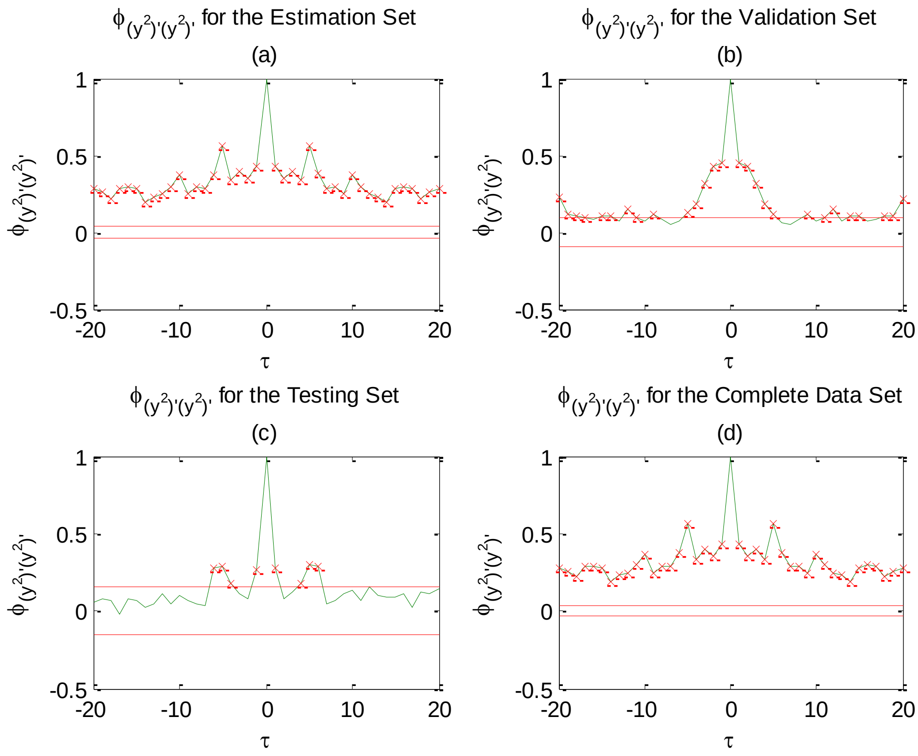


Figure 4.4 $\Phi_{(y^2)'(y^2)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Set

Figure 4.4 (d) shows that $\Phi_{(y^2)'(y^2)'}$ is significantly outside the 95% confidence bands for all the lags. This indicates the need to fit a variance model to the returns, in addition to fitting a mean model.

The corresponding non-linear correlation violation statistic, $V_{(y^2)'(y^2)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set are listed in Table 4.4.

Table 4.4 $V_{(y^2)'(y^2)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y^2)'(y^2)'}$ (Estimation Set)	$V_{(y^2)'(y^2)'}$ (Validation Set)	$V_{(y^2)'(y^2)'}$ (Testing Set)	$V_{(y^2)'(y^2)'}$ (Complete Data Set)
2.8329E-04	7.7942E-05	9.2299E-06	2.8023E-04

The presence of non-linear auto-regressive nature of the returns has been indicated by the correlation plots. The presence of heteroskedasticity in the returns has also been confirmed. The next step is to fit a variety of candidate mean models to the returns.

4.6.3 Candidate Mean Models

Four different candidate mean models are considered.

4.6.3.1 Constant Mean Model

The first is a constant mean model, $y(t) = a_0$, where a_0 is to be estimated.

4.6.3.2 Linear Candidate Mean Model

The second is a linear candidate mean model. The number of lagged linear terms to be included in the linear candidate mean model is denoted as n_l. As explained earlier, from Figure 4.2, the sample autocorrelation of the returns for lags 1,3 and 5 lie significantly outside the 95% confidence bands, indicating the presence of the terms $y(t-1), y(t-3)$ and $y(t-5)$ in the mean model. Hence, n_l is selected to be 5. The linear candidate model is an AR(5) model with a linear noise model consisting of 10 lagged noise terms written as

$$y(t) = a_0 + \sum_{i=1}^{5} a_i y(t-i) + \sum_{j=1}^{10} b_j e(t-j) \tag{4.22}$$

where a_0, a_i, and b_j are coefficients to be estimated.

Since the true mean model (equation (4.19)) includes an additive noise term, $e(t)$, to improve parameter estimation, a linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the linear mean model is carefully selected to be $n_e = 10$. It must be noted that a linear mean model with a large enough noise model can give residuals that are white.

4.6.3.3 Second Order Non-Linear Candidate Mean Model

The third candidate mean model is a second-order non-linear mean model. For the reasons explained in Section 4.6.4.2, n_l is selected to be 5. The number of terms in the candidate mean model increases exponentially with the maximum lag of the non-linear terms to be included in the candidate mean model. Hence, to keep the size of the candidate mean model and computational time reasonable, the maximum lag of the second order non-linear terms to be included in the non-linear candidate mean model is selected to be 5 as well. The second order non-linear mean model is written as

$$y(t) = a_0 + \sum_{i=1}^{5} a_i y(t-i) + \sum_{j=1,k=1,i=6}^{j=5,k=5,i=16} a_i y(t-j) y(t-k) + \sum_{l=1}^{10} b_l e(t-l) \tag{4.23}$$

where a_0, a_i, and b_l are coefficients to be estimated.

A linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 10$.

4.6.3.4 Third Order Non-Linear Candidate Mean Model

The final candidate model is a third-order non-linear mean model. For the reasons explained in Section 4.6.3.2, n_l is selected to be 5. The maximum lag of the second and third order non-linear terms to be included in the non-linear candidate mean model is also selected to be 5. A linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 10$. The third order non-linear mean model is written as

$$y(t) = a_0 + \sum_{i=1}^{5} a_i y(t-i) + \sum_{j=1,k=1,i=6}^{j=5,k=5,i=16} a_i y(t-j) y(t-k) + \sum_{j=1,k=1,l=1,i=17}^{j=5,k=5,l=5,i=35} a_i y(t-j) y(t-k) y(t-l) + \sum_{l=1}^{10} b_l e(t-l) \tag{4.24}$$

where a_0, a_i, and b_l are coefficients to be estimated.

4.6.4 Estimation of Constant Mean Model

First, a constant mean model with no noise model is fitted to the simulated returns. The values of the hyper-parameters used when fitting a constant mean model are given in Table 4.5.

Table 4.5 Hyper-parameters for fitting Constant Mean Model

Hyper-parameter	Value
No. of Samples in Estimation Set (n_{est})	3300
No. of Samples in Estimation Set (n_{val})	500
No. of Samples in Estimation Set (n_{test})	200
No. of Lagged Linear Noise Terms in Mean Model (n_e)	0
No. of Lags for non-linear Correlation Tests (n_{corr})	20

The constant term fitted is 0.0021 and has an ERR of 7.8113.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the constant mean model are plotted in Figure 4.5.

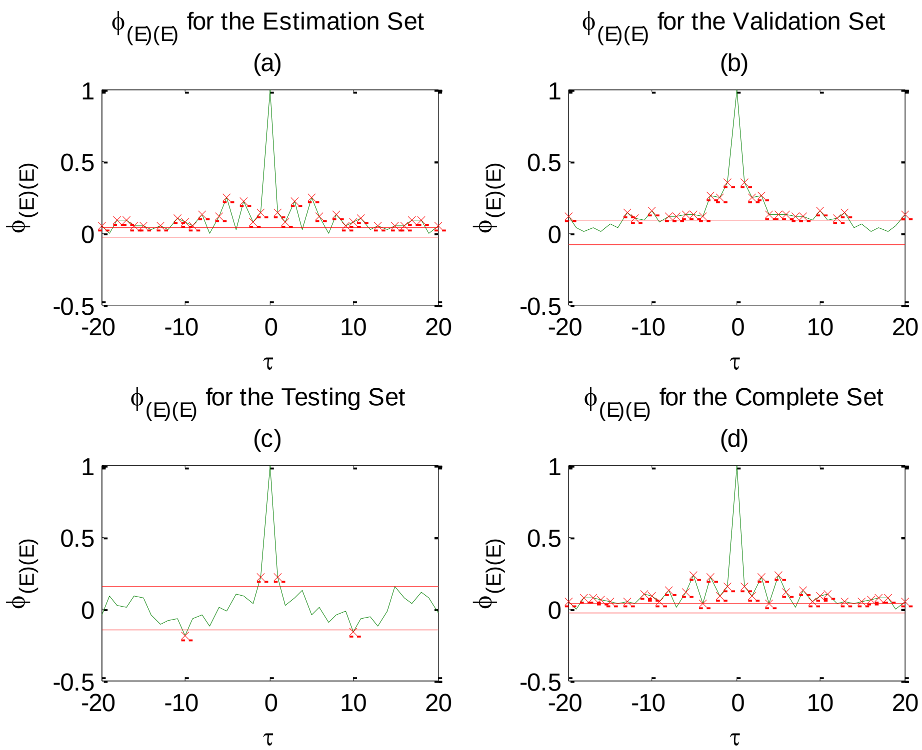


Figure 4.5 Autocorrelation Plots of Residuals for the Selected Constant Mean Model

In Figure 4.5 (d), note that there exists significant autocorrelation of the residuals at lags 1,2,3,5 and 6 implying that the fitted constant mean model is inadequate.

To validate the fitted mean and variance models, the higher order correlation of the squared residuals and squared standardised residuals obtained after fitting the selected constant mean model are plotted in Figure 4.6.

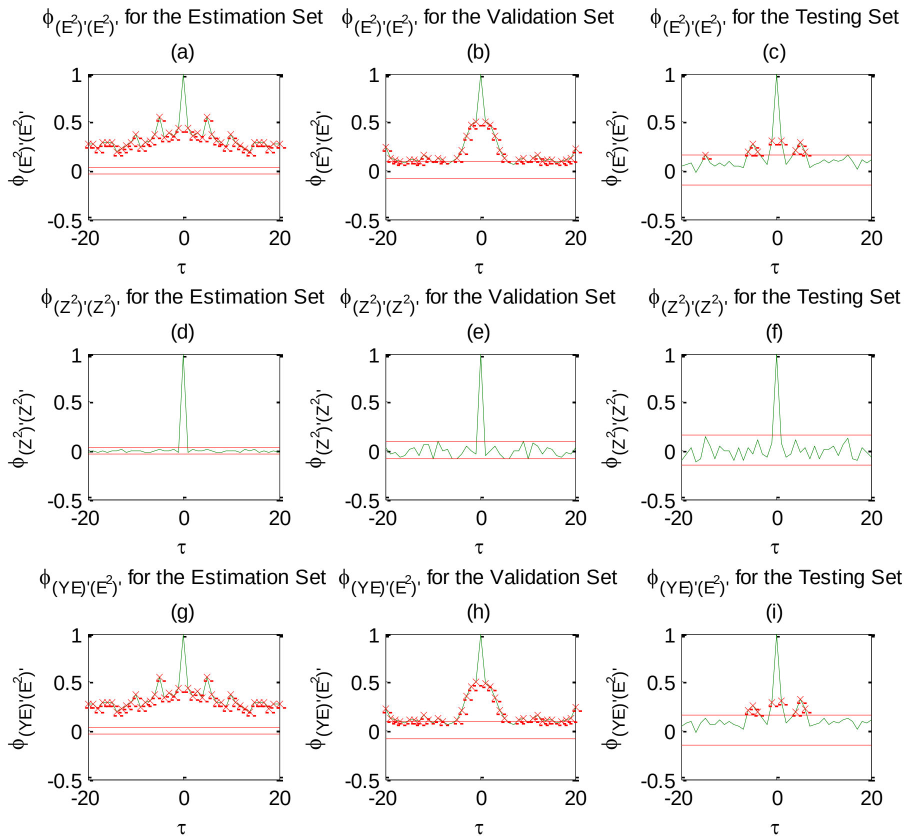


Figure 4.6 Higher Order Correlation Plots for the Selected Constant Mean Model

From Figure 4.6 (d), (e) and (f), the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained from fitting a constant mean model, do not indicate any higher order autocorrelation; these results suggest that the mean and variance of the returns are adequately fit.

4.6.4.1 $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(z^2)'(z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation, Validation and Testing Sets

The higher order correlation plots in Figure 4.6 are quantified by calculating the confidence violation statistics using equations (4.4), (4.5) and (4.6).

Table 4.6 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(z^2)'(z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation, Validation and Testing Sets for the constant mean model.

Table 4.6 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation, Validation and Testing Sets in Constant Mean Model

	Estimation Set	Validation Set	Testing Set
$V_{(\varepsilon^2)'(\varepsilon^2)'}$	2.7641E-04	1.2152E-04	8.0410E-06
$V_{(Z^2)'(Z^2)'}$	0	0	0
$V_{(Y\varepsilon)'(\varepsilon^2)'}$	2.7867E-04	1.1056E-04	8.2072E-06

From Figure 4.6 (a), (b), (c), (g), (h) and (i), the squared estimated residuals, $\hat{e}^2(t)$, obtained from fitting a constant mean model, indicate higher order correlation and autocorrelation. The corresponding confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$, for the Estimation and Validation Sets in Table 4.6 are of the order of 1E-04 which is considerably high in this context. These results suggest that a variance model is required to be fitted to the given return series.

For an adequately fit mean model, the value of $V_{(Z^2)'(Z^2)'}$ is required to be zero, thereby confirming the absence of any higher order correlation in the estimated standardised residuals. The magnitude of the value of $V_{(Z^2)'(Z^2)'}$ for the Validation Set is zero, which is acceptable. These results suggest that the fitted mean and variance models are adequate.

4.6.5 Estimation of Linear Mean Model using WOFR

The next type of mean model to be fitted to the simulated returns is the linear candidate mean model listed in Section 4.6.3.2. WOFR is performed on the linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select the linear mean model that best describes the mean of the simulated returns.

The values of the hyper-parameters used when fitting a linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 4.5. The only difference is that a linear noise model is to be fitted as well, hence n_e is selected to be 10.

The results of the WOFR analysis are given in Table 4.7. The terms shaded in blue represent the linear terms that are present in the true non-linear mean model (equation (4.19)) that was used to generate the simulated return series. Note that lags 2,4 and 5 appear as non-linear terms in equation (4.19).

Table 4.7 Terms in Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	1	0.0013	7.6504
2	$y(t-3)$	0.1583	2.3217
3	$y(t-1)$	0.1636	1.9363
4	$y(t-5)$	0.0803	0.4191
5	$y(t-4)$	-0.0454	0.1649
6	$y(t-2)$	-0.0053	0.0022

Using the hyper-parameters in Table 4.5, and $n_e = 10$, steps 7 to 10 of the method described in Section 4.5 are now carried out on the re-ordered terms of the linear candidate mean model.

4.6.5.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 4.7 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18).

Figure 4.7 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the linear mean model.

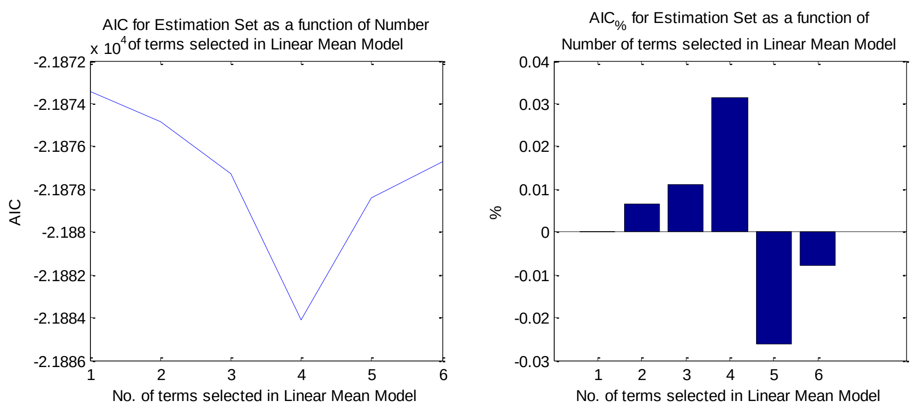


Figure 4.7 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Linear Mean Model

From Figure 4.7, AIC is the least (-2.1884E+04) when 4 terms from the top of Table 4.7 are included in the linear mean model. Also, $AIC_{\%}$ increases progressively until 4 terms from the top of Table 4.7 are included. Upon addition of the fifth term, the AIC increases, and $AIC_{\%}$ decreases. Hence, 4 terms from the top of Table 4.7 are selected to be included in the linear mean model.

Note that for all the linear terms present in the actual mean model to be selected, the first 3 terms from the top of Table 4.7 need to be selected. Since the actual mean model is non-linear in nature, it is acceptable for 1 or 2 extra linear terms get selected as well. Hence, 4 terms being selected in the linear mean model is acceptable, as long as all the linear terms present in the actual mean model are selected.

4.6.5.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the linear mean models using equations (4.4), (4.5) and (4.6). Figure 4.8 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 4.9 shows $V_{(Z^2)'(Z^2)'}$, and Figure 4.10 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

For the linear mean model comprising of the 4 terms from the top of Table 4.7, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 1.1609E-05. This confirms the term selection result arrived upon in Section 4.6.6.1 using AIC.

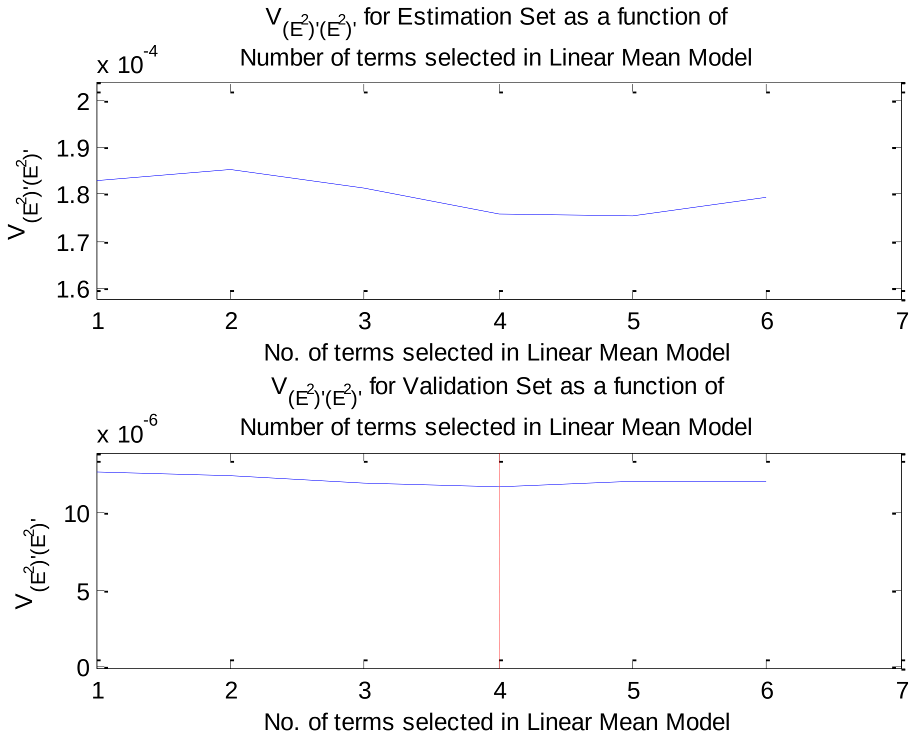


Figure 4.8 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

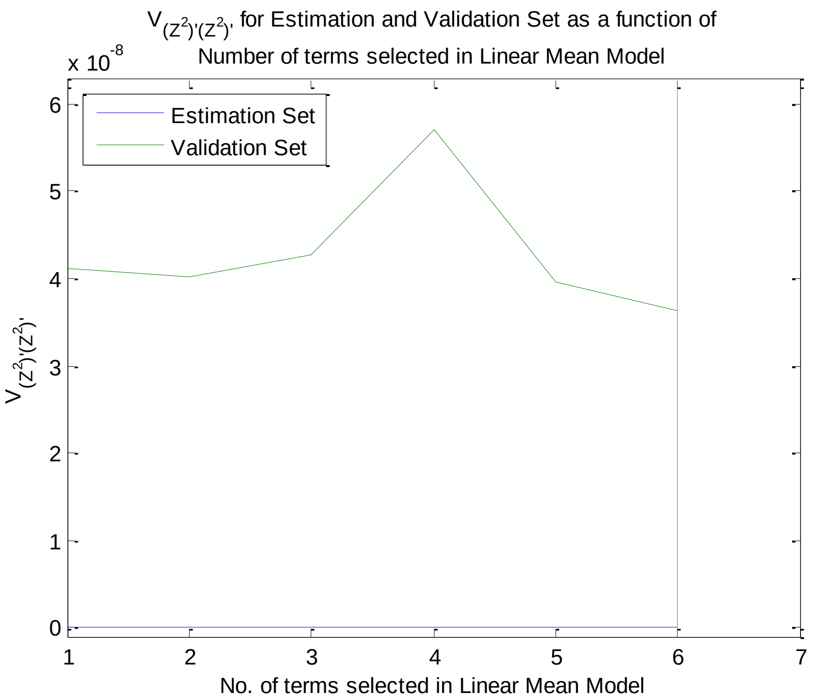


Figure 4.9 $V_{(z^2)'(z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of all the 6 terms from Table 4.7, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 3.6255E-08. In this case, 4 terms from the top of Table 4.7 are selected to be included in the linear mean model which yields a value of 5.7049E-08 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude and very close to the minimum value of 3.6255E-08. These results suggest that the mean and the variance of the returns have been adequately modelled.

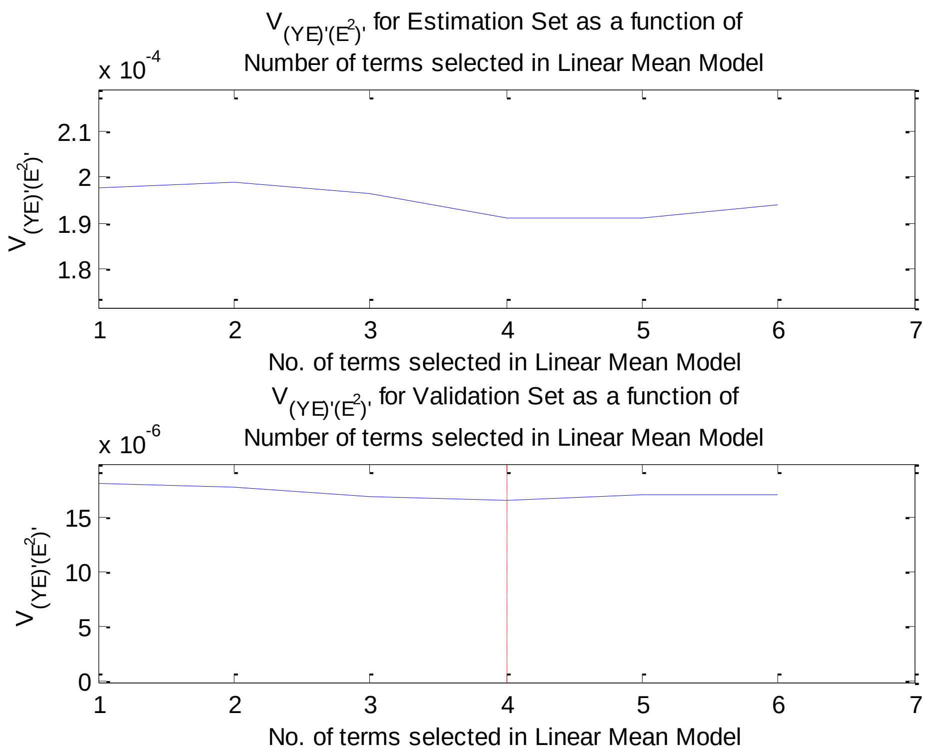


Figure 4.10 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of the 4 terms from the top of Table 4.7, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 1.6469E-05. This confirms the term selection result arrived upon in Section 4.6.6.1 using AIC.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected linear mean model are plotted in Figure 4.11.

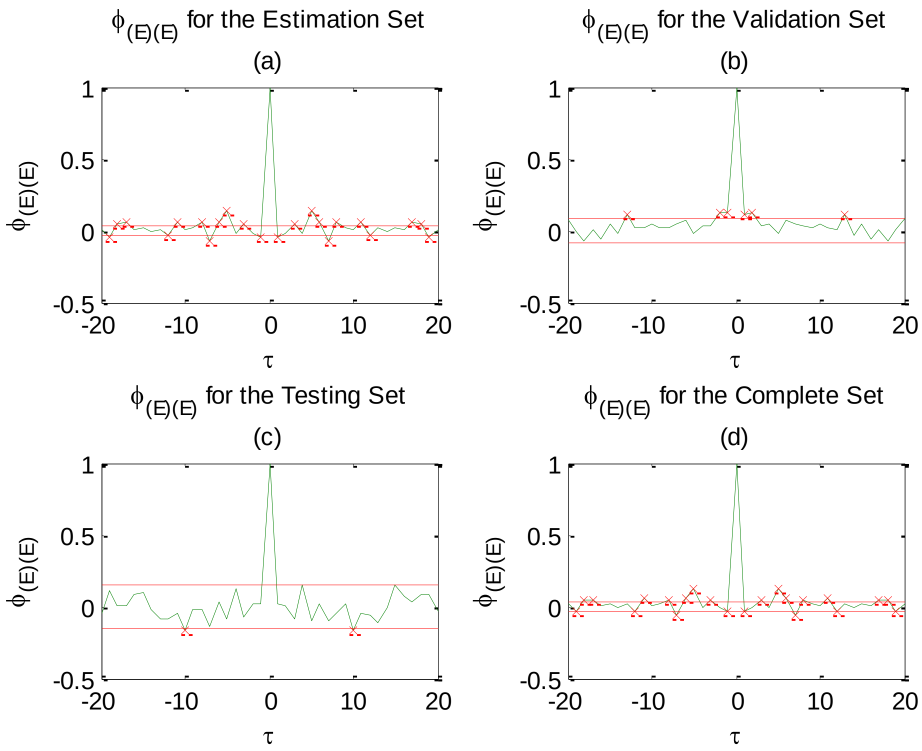


Figure 4.11 Autocorrelation Plots of Residuals for the Selected Linear Mean Model

In Figure 4.11 (d), note that there exists significant autocorrelation of the residuals at lags 4 and 5 implying that the fitted linear mean model is inadequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 4.5 (d)) which indicates that the selected linear mean model captures the predictable elements of the mean of the returns much better than the selected constant mean model.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected linear mean model (with 4 process terms) are shown in Figure 4.12.

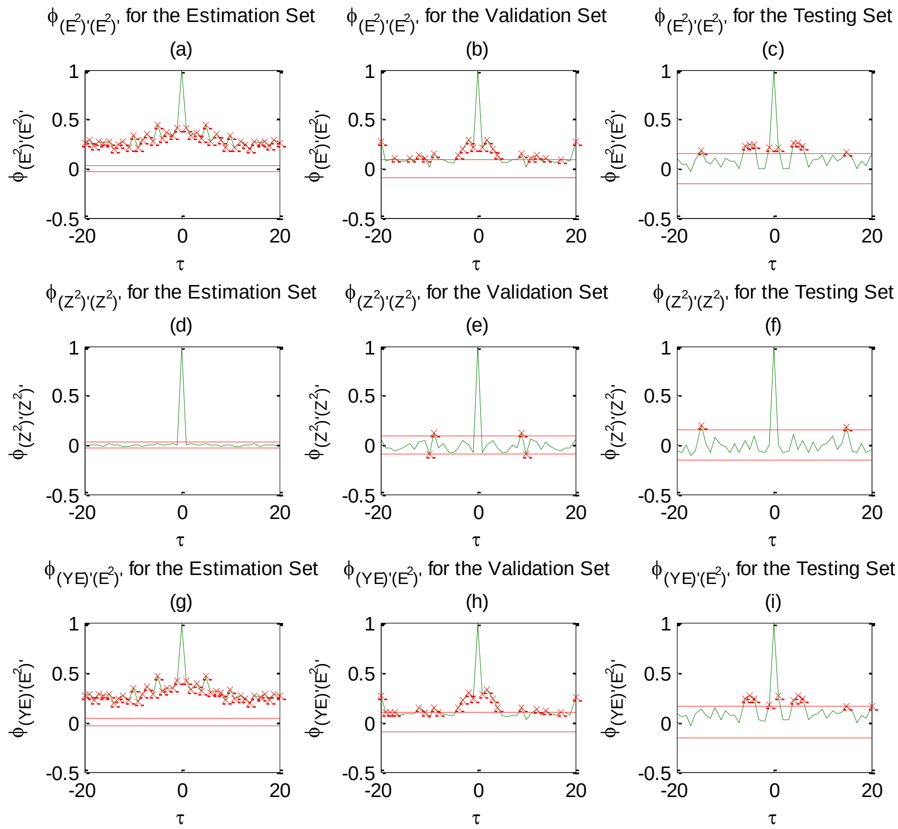


Figure 4.12 Higher Order Correlation Plots for the Selected Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(Z^2)'(Z^2)'}$, should exist. In Figure 4.12 (d), there exist no violations of the 95% confidence bands. In Figure 4.12 (e), note that the violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(Z^2)'(Z^2)'}$, are negligible, once again indicating that the mean of the returns has been adequately modelled.

The terms selected in the linear mean model along with the coefficient estimates and ERR values are listed in Table 4.8. The coefficient estimates of the 10 noise terms are also included.

Table 4.8 Linear Mean Model

No.	Term	Parameter Estimate	True Coefficient	ERR
1	1	0.0008	0.001	7.6499
2	$y(t-3)$	-0.4279	0.15	0.0759
3	$y(t-1)$	0.7326	0.2	2.3046
4	$y(t-5)$	0.3206	N/A	0.4337
5	$e(t-1)$	-0.5603	N/A	0.0342
6	$e(t-2)$	-0.1050	N/A	0.0062
7	$e(t-3)$	0.5725	N/A	1.9520
8	$e(t-4)$	-0.0351	N/A	0.0287
9	$e(t-5)$	-0.2511	N/A	0.0214
10	$e(t-6)$	-0.0141	N/A	0.0024
11	$e(t-7)$	-0.0089	N/A	0.0048
12	$e(t-8)$	0.0123	N/A	0.0721
13	$e(t-9)$	-0.0390	N/A	0.0421
14	$e(t-10)$	0.0396	N/A	0.1474

So far, the constant and the linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in both cases.

4.6.6 Estimation of Second Order Non-Linear Mean Model using WOFR

The next type of candidate mean model to be fitted to the simulated returns is the second order non-linear candidate mean model listed in Section 4.6.3.3. As in the case of the linear candidate mean model, WOFR is performed on the second order non-linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select a second order non-linear mean model that best describes the mean of the simulated returns.

The values of the hyper-parameters used when fitting a second order non-linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 4.5. The only difference is that a linear noise model is to be fitted as well, hence n_e is selected to be 10.

The WOFR analysis results are shown in Table 4.9. The terms shaded in blue represent the terms that are present in the true non-linear mean model used to generate the simulated return series.

Table 4.9 Terms in Second Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	1	0.0012	7.4602
2	$y(t-5)y(t-5)$	7.8869	3.3629
3	$y(t-3)$	0.1521	2.1289
4	$y(t-1)$	0.1931	1.7372
5	$y(t-1)y(t-2)$	-9.8374	2.0654
6	$y(t-3)y(t-4)$	-3.7392	0.5453
7	$y(t-2)y(t-3)$	-2.5696	0.0904
8	$y(t-4)y(t-4)$	-1.2752	0.1070
9	$y(t-4)$	-0.0352	0.0569
10	$y(t-3)y(t-3)$	1.4093	0.0412
11	$y(t-2)y(t-4)$	1.6808	0.0405
12	$y(t-5)$	0.0185	0.0259
13	$y(t-2)y(t-2)$	-0.7980	0.0155
14	$y(t-1)y(t-3)$	-0.8652	0.0151
15	$y(t-4)y(t-5)$	-0.8417	0.0115
16	$y(t-2)$	0.0113	0.0093
17	$y(t-1)y(t-1)$	-0.3226	0.0047
18	$y(t-3)y(t-5)$	-0.3221	0.0011
19	$y(t-1)y(t-5)$	0.2424	0.0016
20	$y(t-2)y(t-5)$	0.1721	0.0005
21	$y(t-1)y(t-4)$	-0.1687	0.0005

Note that the terms that are present in the actual mean model are all at the top of Table 4.9. Using the hyper-parameters in Table 4.5, and $n_e = 10$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the second order non-linear candidate mean model.

4.6.6.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 4.9 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18).

Figure 4.13 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the second order non-linear mean model.

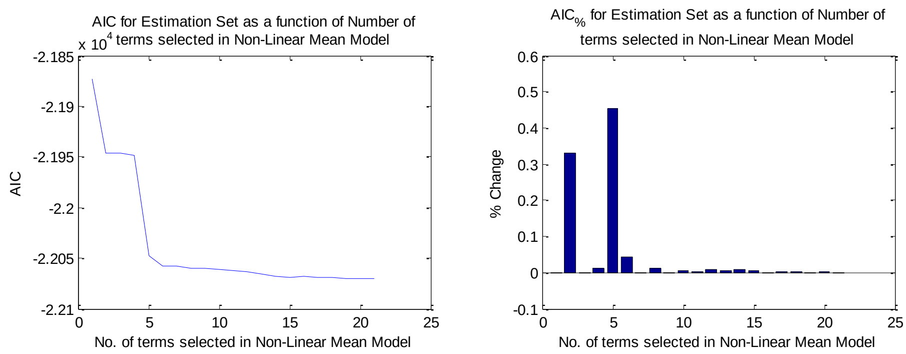


Figure 4.13 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Second Order Non-Linear Mean Model

From Figure 4.13, the AIC decreases drastically to a value of -2.2058E+04 till 6 terms from the top of Table 4.9 are selected to be included in the second order non-linear mean model. Also, $AIC_{\%}$ for each iteration is significant until 6 terms from the top of Table 4.9 are included in the mean model. The addition of further terms to the mean model does not decrease the value of AIC drastically.

Term selection could be stopped at 5 terms, since $AIC_{\%}$ in the 6th iteration is not very significant, but from Table 4.9, the ERR of the 6th term is 0.5453%, which is greater than 0.1% and hence suggestive that the 6th term should be included as well (Wei and Billings, 2004).

Note that for all the terms present in the actual mean model to be selected, the first 6 terms from the top of Table 4.9 need to be selected, which is the case.

4.6.6.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the linear mean models using equations (4.7), (4.8) and (4.9). Figure 4.14 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 4.15 shows $V_{(Z^2)'(Z^2)'}$, and Figure 4.16 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected in the second order non-linear mean model. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

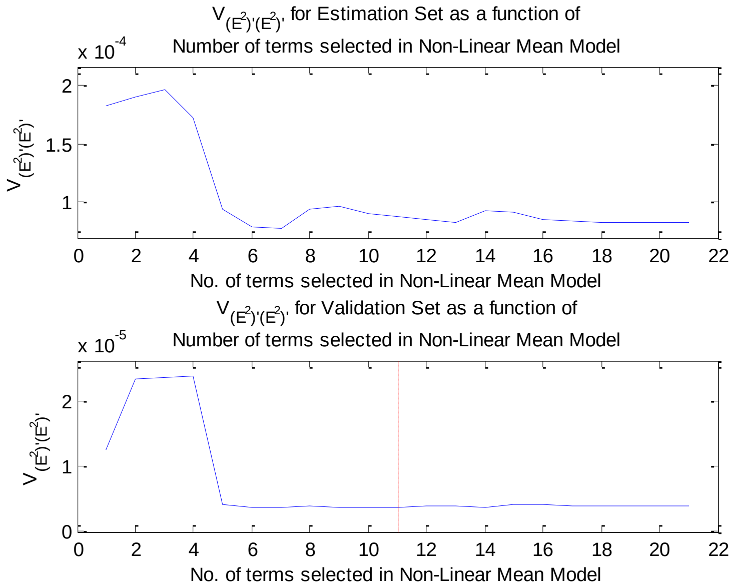


Figure 4.14 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ **for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model**

$V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least (3.7365E-06) when 11 terms from the top of Table 4.9 are selected to be included in the mean model. This does not coincide with term selection using AIC (6 terms). But note that $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set drops drastically when 5 terms from the top of Table 4.9 are selected to be included in the mean model, and remains at this minimum level as more terms are included in the mean model.

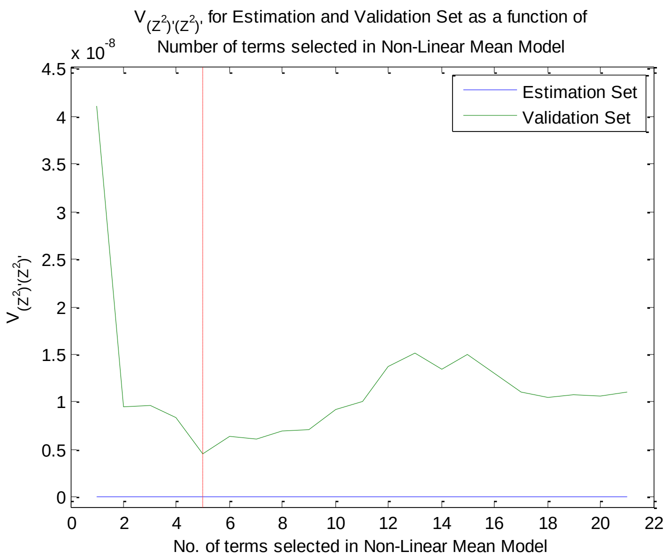


Figure 4.15 $V_{(z^2)'(z^2)'}$ **for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model**

$V_{(z^2)'(z^2)'}$ for the Validation Set is the least (4.5356E-09) when 5 terms from the top of Table 4.9 are selected to be included in the mean model. In this case, 6 terms from the top of Table 4.9 are selected to be included in the second order non-linear mean model which yields a value of 6.3454E-09 for $V_{(z^2)'(z^2)'}$ for the Validation Set, which is of considerably less magnitude and very close to the minimum value of 4.5356E-09, implying that the mean and the variance of the returns have been adequately modelled.

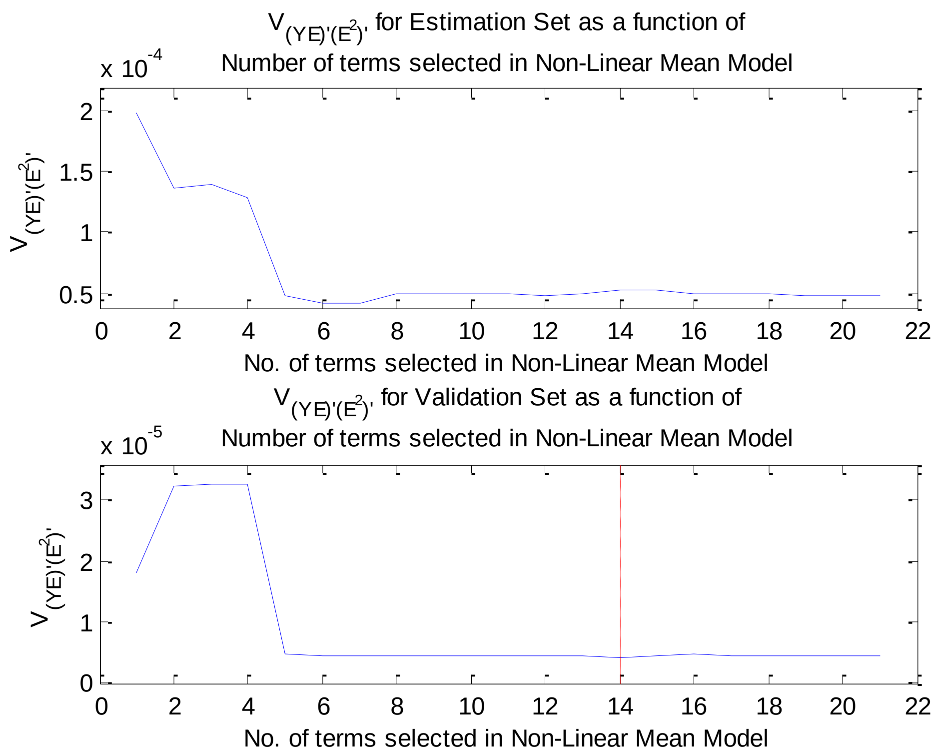


Figure 4.16 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

$V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least (4.2156E-06) when 14 terms from the top of Table 4.9 are included in the mean model. This does not coincide with term selection using AIC (6 terms). But note that like $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set also drops drastically when 5 terms from the top of Table 4.9 are included in the mean model, and remains at this minimum level as more terms are included in the mean model.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 4.17.

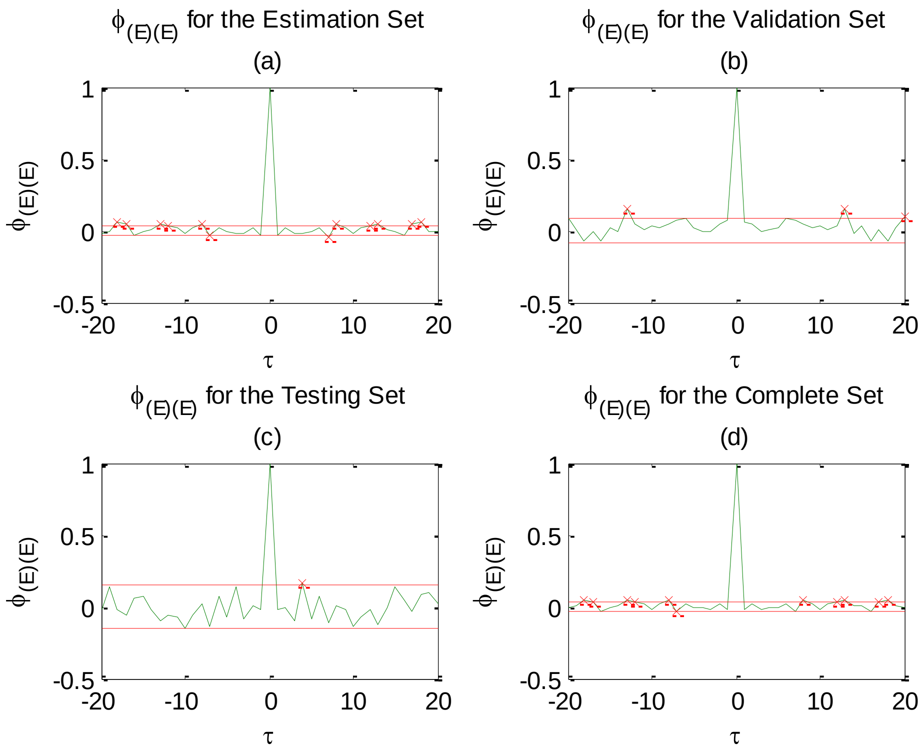


Figure 4.17 Autocorrelation Plots of Residuals for the Selected Second Order Non-Linear Mean Model (6 Process Terms)

In Figure 4.17 (d), note that there exists no significant autocorrelation of the residuals implying that the fitted second order non-linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 4.5 (d)) and the residuals obtained after fitting the selected linear mean model (Figure 4.11 (d)) which indicates that the selected second order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant and linear mean models.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected second order non-linear mean model (with 6 process terms) are shown in Figure 4.18.

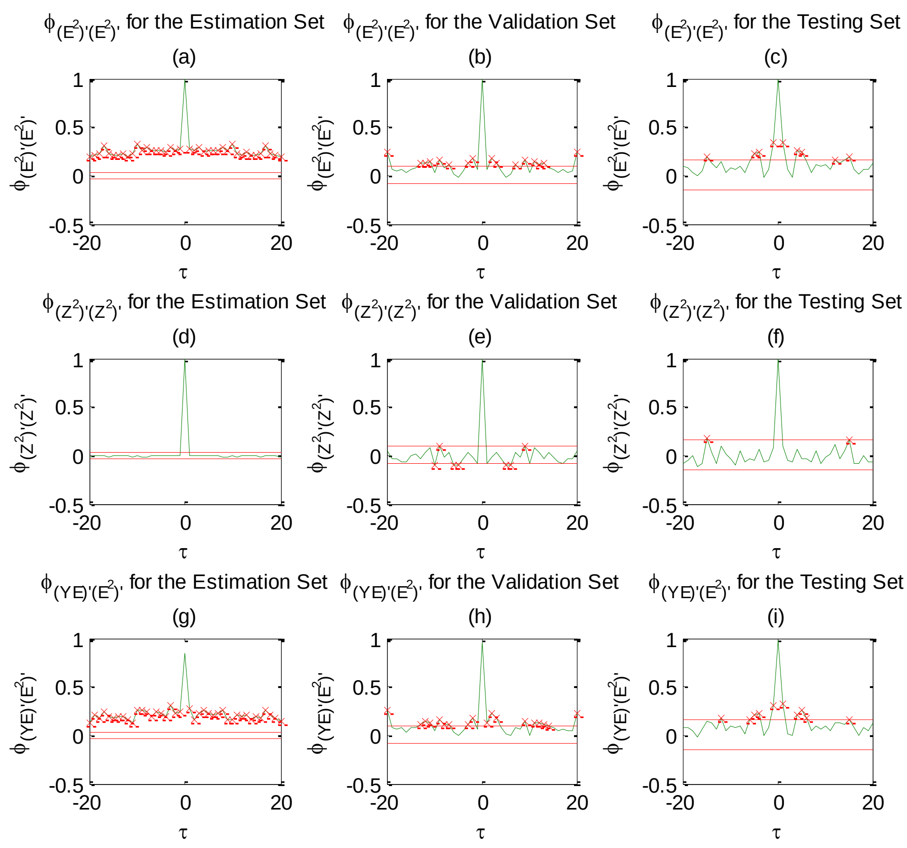


Figure 4.18 Higher Order Correlation Plots for the Selected Second Order Non-Linear Mean Model (6 Process Terms)

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. In Figure 4.18 (e), note that the violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(Z^2)'(Z^2)'}$, are negligible, indicating that the mean and the variance of the returns have been adequately modelled.

The selected terms along with the parameter estimates and ERR values are listed in Table 4.10. The coefficient estimates of the 10 noise terms are also included. Note that the terms selected in the second order non-linear mean model are all present in the true mean model.

Table 4.10 Second Order Non-Linear Mean Model

No.	Term	Parameter Estimate	True Coefficient	ERR
1	1	0.0012	0.001	7.5171
2	$y(t-5)y(t-5)$	7.2544	8	3.1913
3	$y(t-3)$	-0.0104	0.15	0.0006
4	$y(t-1)$	0.3375	0.2	1.9056
5	$y(t-1)y(t-2)$	-10.6898	-10	2.2140
6	$y(t-3)y(t-4)$	-3.4399	-5	0.3714
7	$e(t-1)$	-0.1492	N/A	0.0565
8	$e(t-2)$	-0.0081	N/A	0.0034
9	$e(t-3)$	0.1653	N/A	2.1272
10	$e(t-4)$	-0.0325	N/A	0.0461
11	$e(t-5)$	0.0124	N/A	0.0106
12	$e(t-6)$	-0.0112	N/A	0.0103
13	$e(t-7)$	0.0117	N/A	0.0115
14	$e(t-8)$	0.0306	N/A	0.0712
15	$e(t-9)$	-0.0254	N/A	0.0469
16	$e(t-10)$	0.0458	N/A	0.1329

So far, the constant, the linear and the second order non-linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 3 cases.

4.6.7 Estimation of Third Order Non-Linear Mean Model using WOFR

The last type of candidate mean model to be fitted to the simulated returns is the third order non-linear candidate mean model listed in Section 4.6.3.4. As in the case of the previous candidate mean models, WOFR is performed on the third order non-linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select a third order non-linear mean model that best describes the mean of the simulated returns.

The values of the hyper-parameters used when fitting a third order non-linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 4.5. The only difference is that a linear noise model is to be fitted as well, hence n_e is selected to be 10.

The WOFR analysis results are shown in Table 4.11. The terms shaded in blue represent the terms that are present in the true non-linear mean model used to generate the simulated return series.

Table 4.11 Terms in Third Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-5)y(t-5)$	9.0225	7.6816
2	1	0.0013	4.3588
3	$y(t-3)$	0.1606	2.1920
4	$y(t-1)$	0.1780	1.4907
5	$y(t-1)y(t-2)$	-11.9361	2.4490
6	$y(t-3)y(t-4)$	-2.6153	0.8945
7	$y(t-1)y(t-1)y(t-3)$	-73.2853	0.2585
8	$y(t-2)y(t-3)$	-3.8379	0.0995
9	$y(t-2)y(t-3)y(t-3)$	50.2891	0.1171
10	$y(t-4)$	-0.0351	0.1391
11	$y(t-1)y(t-2)y(t-3)$	109.5623	0.0892
12	$y(t-4)y(t-4)$	-3.9871	0.0660
13	$y(t-2)y(t-3)y(t-4)$	-65.2268	0.1184
14	$y(t-1)y(t-1)$	1.8861	0.0791
15	$y(t-4)y(t-4)y(t-4)$	51.1969	0.1002
16	$y(t-2)y(t-4)$	3.8161	0.0480
17	$y(t-1)y(t-2)y(t-4)$	-0.6808	0.0673
18	$y(t-3)y(t-3)y(t-3)$	25.8487	0.0618
19	$y(t-3)y(t-5)y(t-5)$	-55.4806	0.0618
20	$y(t-4)y(t-5)$	-1.0299	0.0453
21	$y(t-3)y(t-3)y(t-4)$	-48.5025	0.0386
22	$y(t-5)$	0.0219	0.0365
23	$y(t-4)y(t-4)y(t-5)$	-58.7921	0.0274
24	$y(t-1)y(t-1)y(t-4)$	-56.0390	0.0426
25	$y(t-2)y(t-4)y(t-5)$	115.6252	0.0253
26	$y(t-3)y(t-3)y(t-5)$	47.5560	0.0474
27	$y(t-2)y(t-5)$	-1.9805	0.0262
28	$y(t-3)y(t-4)y(t-4)$	41.9477	0.0262
29	$y(t-2)y(t-2)y(t-2)$	31.2245	0.0230
30	$y(t-2)y(t-2)$	-1.2913	0.0303
31	$y(t-1)y(t-1)y(t-2)$	-40.8939	0.0243
32	$y(t-1)y(t-1)y(t-1)$	16.7382	0.0144
33	$y(t-3)y(t-4)y(t-5)$	-40.8867	0.0137
34	$y(t-1)y(t-1)y(t-5)$	34.5182	0.0178

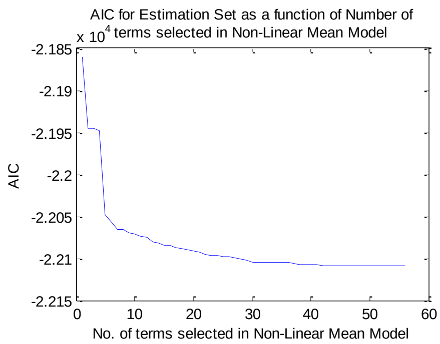
35	$y(t-3)y(t-5)$	0.9191	0.0115
36	$y(t-2)y(t-4)y(t-4)$	-41.8216	0.0111
37	$y(t-1)y(t-4)y(t-5)$	-39.2105	0.0158
38	$y(t-1)y(t-4)$	1.5876	0.0120
39	$y(t-5)y(t-5)y(t-5)$	-12.6208	0.0115
40	$y(t-2)$	0.0138	0.0105
41	$y(t-1)y(t-2)y(t-5)$	-32.8960	0.0084
42	$y(t-2)y(t-2)y(t-5)$	22.2874	0.0087
43	$y(t-2)y(t-3)y(t-5)$	-29.0874	0.0114
44	$y(t-1)y(t-3)y(t-4)$	-26.7274	0.0083
45	$y(t-2)y(t-5)y(t-5)$	-7.3906	0.0043
46	$y(t-2)y(t-2)y(t-4)$	-10.4421	0.0026
47	$y(t-1)y(t-4)y(t-4)$	-11.4072	0.0017
48	$y(t-1)y(t-3)$	-0.4703	9.7492E-04
49	$y(t-2)y(t-2)y(t-3)$	-4.0594	7.8821E-04
50	$y(t-1)y(t-5)$	0.2454	7.6123E-04
51	$y(t-1)y(t-2)y(t-2)$	4.2755	5.9736E-04
52	$y(t-4)y(t-5)y(t-5)$	-3.7607	2.8501E-04
53	$y(t-3)y(t-3)$	0.0960	1.3231E-04
54	$y(t-1)y(t-3)y(t-3)$	2.4271	1.0838E-04
55	$y(t-1)y(t-3)y(t-5)$	-1.3464	1.8784E-05
56	$y(t-1)y(t-5)y(t-5)$	0.4518	3.7099E-06

Note that all the terms present in the actual mean model are at the top of Table 4.11. Using the hyper-parameters in Table 4.5, and $n_e = 10$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the candidate model.

4.6.7.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 4.11 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18).

Figure 4.19 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the second order non-linear mean model.



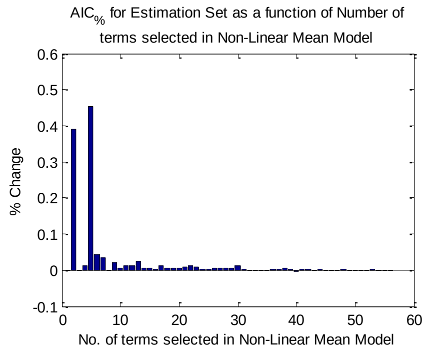


Figure 4.19 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Third Order Non-Linear Mean Model

From Figure 4.19, the AIC decreases drastically to a value of -2.2065E+04 till 7 terms from the top of Table 4.9 are selected to be included in the second order non-linear mean model. Also, $AIC_{\%}$ for each iteration is significant until 7 terms from the top of Table 4.11 are included in the mean model. The addition of further terms to the mean model does not decrease the value of AIC drastically.

Knowing the true structure of the mean model, there does not exists any third order polynomial term in the actual mean model. The 7th term that is selected is not actually present in the actual mean model. But, from Table 4.11, the ERR of the 7th term is 0.2585%, which is greater than 0.1% and hence suggestive that the 7th term should be selected (Wei and Billings, 2004).

Note that for all the terms present in the actual mean model to be selected, the first 6 terms from the top of Table 4.11 need to be selected, which is the case.

4.6.7.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the linear mean models using equations (4.7), (4.8) and (4.9). Figure 4.20 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 4.21 shows $V_{(Z^2)'(Z^2)'}$, and Figure 4.22 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected in the third order non-linear mean model. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

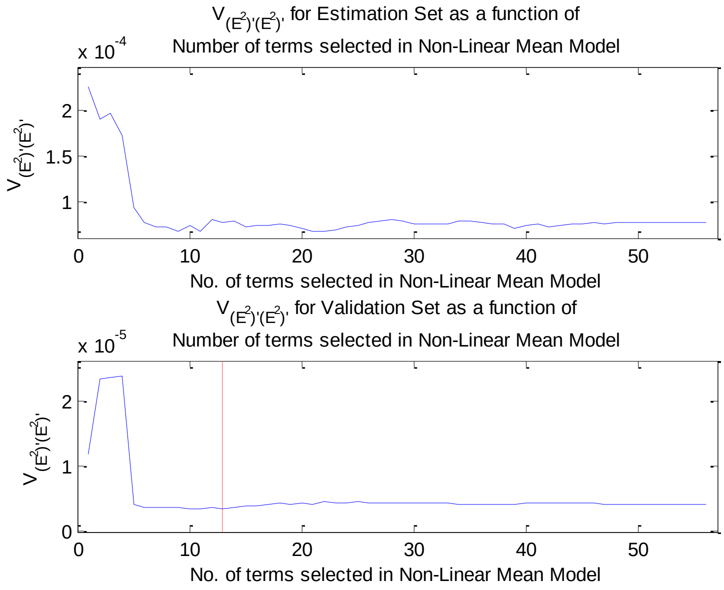


Figure 4.20 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

$V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least (3.5650E-06) when 13 terms from the top of Table 4.11 are selected to be included in the mean model. Note that $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set drops drastically when 5 terms from the top of Table 4.11 are selected to be included in the mean model, and remains at this minimum level as more terms are included in the mean model.

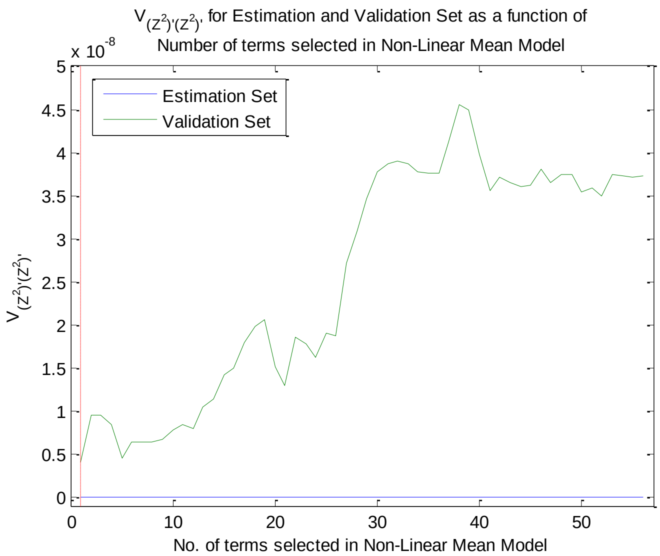


Figure 4.21 $V_{(z^2)'(z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

$V_{(z^2)'(z^2)'}$ for the Validation Set is the least (4.0902E-09) when only the first term from Table 4.11 is selected to be included in the mean model.

In this case, 7 terms from the top of Table 4.11 are selected to be included in the second order non-linear mean model which yields a value of 6.3352E-09 for $V_{(z^2)'(z^2)'}$ for the Validation Set, which is of significantly less magnitude and very close to the minimum value of 4.0902E-09. These results indicate that the mean and the variance of the returns have been adequately modelled.

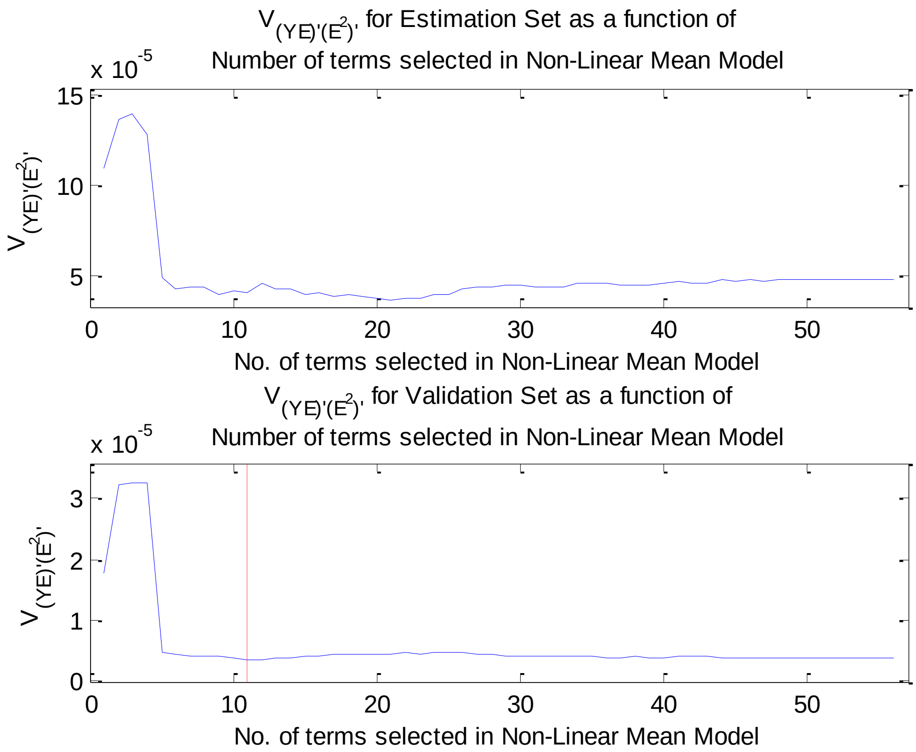


Figure 4.22 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

$V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least (3.6153E-06) when 11 terms from the top of Table 4.11 are included in the mean model. Note that $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set drops drastically when 5 terms from the top of Table 4.11 are selected to be included in the mean model, and remains at this minimum level as more terms are included in the mean model.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 4.23.

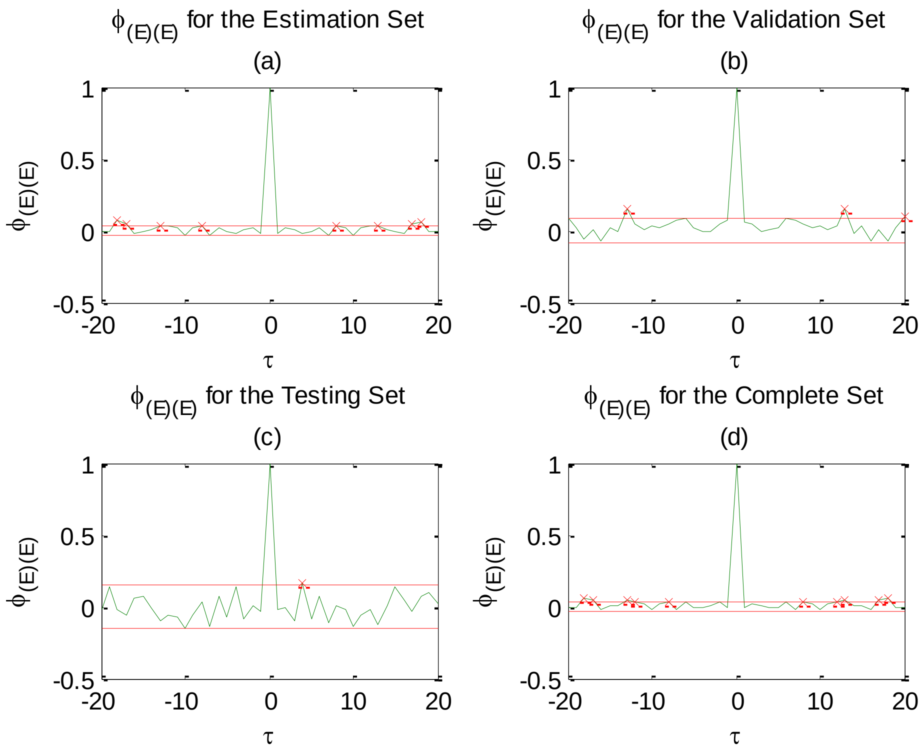


Figure 4.23 Autocorrelation Plots of Residuals for the Selected Third Order Non-Linear Mean Model

In Figure 4.23 (d), note that there exists no significant autocorrelation of the residuals implying that the fitted second order non-linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 4.5 (d)) and the residuals obtained after fitting the selected linear mean model (Figure 4.11 (d)) which indicates that the selected second order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant and linear mean models.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected second order non-linear mean model (with 7 process terms) are shown in Figure 4.24.

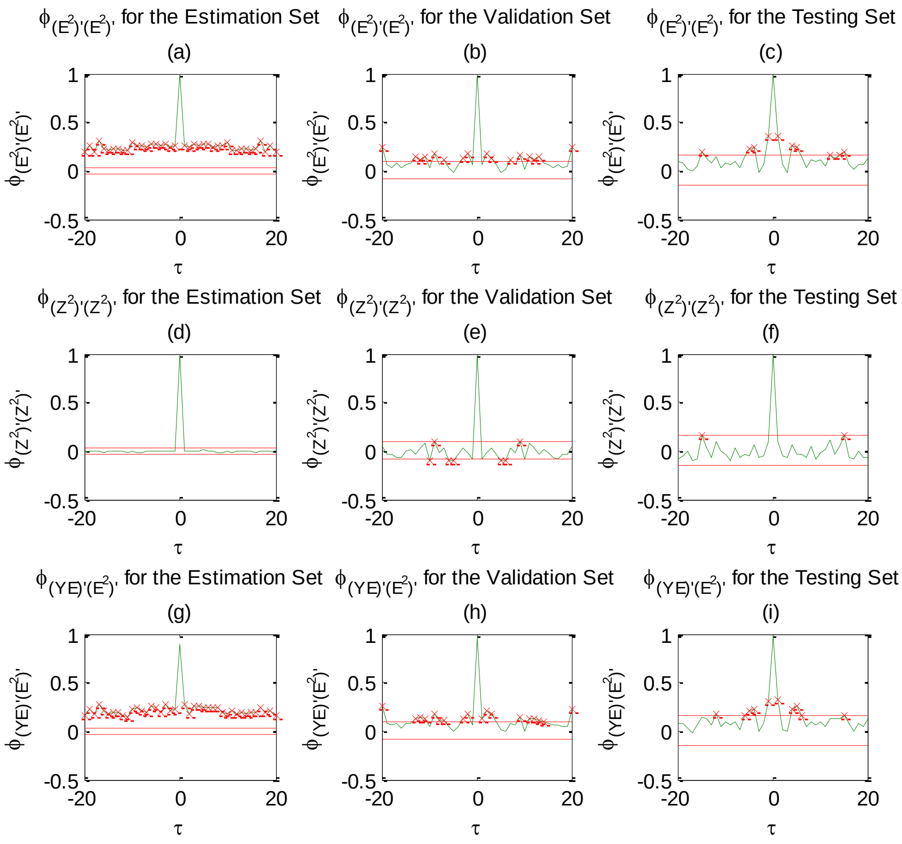


Figure 4.24 Higher Order Correlation Plots for the Selected Third Order Non-Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(Z^2)'(Z^2)'}$, should exist. In Figure 4.24 (d), there exist no violations of the 95% confidence bands. In Figure 4.24 (e), note that the violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(Z^2)'(Z^2)'}$, are negligible, indicating that the mean and the variance of the returns have been adequately modelled.

The selected terms along with the coefficient estimates and ERR values are listed in Table 4.12. The coefficient estimates of the 10 noise terms are also included.

Table 4.12 Third Order Non-Linear Mean Model

No.	Term	Parameter Estimate	True Coefficient	ERR
1	$y(t-5)y(t-5)$	7.4958	8	3.3562
2	1	0.0011	0.001	7.5139
3	$y(t-3)$	0.0211	0.15	0.0025
4	$y(t-1)$	0.3607	0.2	1.8986
5	$y(t-1)y(t-2)$	-11.4529	-10	2.2406
6	$y(t-3)y(t-4)$	-3.0877	-5	0.4075
7	$y(t-1)y(t-1)y(t-3)$	-49.4778	N/A	0.0745
8	$e(t-1)$	-0.1679	N/A	0.0853
9	$e(t-2)$	-0.0101	N/A	0.0063
10	$e(t-3)$	0.1416	N/A	2.1519
11	$e(t-4)$	-0.0417	N/A	0.0647
12	$e(t-5)$	0.0109	N/A	0.0086
13	$e(t-6)$	-0.0140	N/A	0.0071
14	$e(t-7)$	0.0114	N/A	0.0114
15	$e(t-8)$	0.0287	N/A	0.0698
16	$e(t-9)$	-0.0255	N/A	0.0456
17	$e(t-10)$	0.0459	N/A	0.1412

So far, all the mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 4 cases.

4.6.8 Comparison of ARCH Test Statistics and Non-Linear Correlation Statistics of All Mean Models

Note that higher order correlation tests and the ARCH Test statistics of $\hat{z}(t)$ do not work for the purpose of comparison of the performance of the different mean models. This is because the mean model can be underfitted, and the variance model can be overfitted, thereby yielding a $\hat{z}(t)$ series that passes all the mentioned model validation tests. Hence, to compare the fitted mean models, the ARCH Test statistics and the higher order correlation statistics of $\hat{e}(t)$ need to be used.

The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the selected mean models are listed in Table 4.13. The minimum values are shaded in blue.

Table 4.13 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Mean Models

Type of Mean Model	**ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)**	**$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)**
Constant	129.1306	2.7469E-04
Linear	23.3885	1.8930E-04
Second Order Non-Linear	2.0435	4.1168E-05
Third Order Non-Linear	2.1966	4.2576E-05

The second order non-linear mean model (terms and parameter estimates listed in Table 4.10) has the lowest ARCH Test Statistic of $\hat{e}(t)$ for the Validation Set. This implies that the mean of the returns have been modelled better by the second order non-linear mean model than the third order non-linear, linear or constant mean model, and hence is the appropriate choice. Thus, the proposed procedure correctly identifies the true mean model.

Note that:

- The constant mean model has the highest value of ARCH test statistic of $\hat{e}(t)$ for the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set amongst all the mean models.
- The linear mean model has the second highest value of ARCH test statistic of $\hat{e}(t)$ for the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set amongst all the mean models.
- Lastly, the selected second order non-linear mean model, which is in fact the true mean model of the simulated data, has the lowest value of ARCH test statistic of $\hat{e}(t)$ for the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set amongst all the mean models.

This decreasing trend in the values of the ARCH test statistic of $\hat{e}(t)$ for the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set reinforces the idea that these statistics are a good indicator of how adequately the mean of the returns has been modelled.

Model selection using the ARCH test statistic of $\hat{e}(t)$ for the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set seems to detect over-fitting as well.

4.6.9 Comparison of Variance Estimates of All Mean Models

In the case of the simulated example, the true (also termed as simulated) variance is known. The different ARCH(25) variance estimates obtained after fitting the different kinds of mean models to the simulated returns are compared to the simulated variance. The ARCH(25) variance estimates obtained using the constant mean model, the linear mean model, the second order non-linear mean model, and the third order non-linear mean model are plotted in Figure 4.25, Figure 4.26 and Figure 4.27. The Normalised Root Mean Squared Error (NRMSE) between the simulated variance and the ARCH(25) variance estimates are calculated for all the models and listed in Table 4.14. The minimum value is shaded in blue.

Table 4.14 NRMSE of ARCH(25) Variance Estimates of all Mean Models

Type of Mean Model	NRMSE (%)
Constant	117.96
Linear	91.77
Second Order Non-Linear	17.36
Third Order Non-Linear	19.41

The ARCH(25) variance estimate from the second order non-linear mean model has the lowest NRMSE (17.36%). The variance estimates from the constant and linear mean models perform badly compared to second order and third order non-linear mean model variance estimates.

The ARCH(25) variance estimate from the constant mean model has the highest NRMSE of 117.96%. This, along with the results of the ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the constant mean model can be safely used to conclude that a constant mean model is not adequate to estimate the mean of the given returns series, even if the higher order correlation model validation tests suggest that the mean and the variance of the returns have been adequately modelled. At the very least, a linear mean model with a noise model should be used to model the mean of the returns in order to allow for more accurate estimation of the variance.

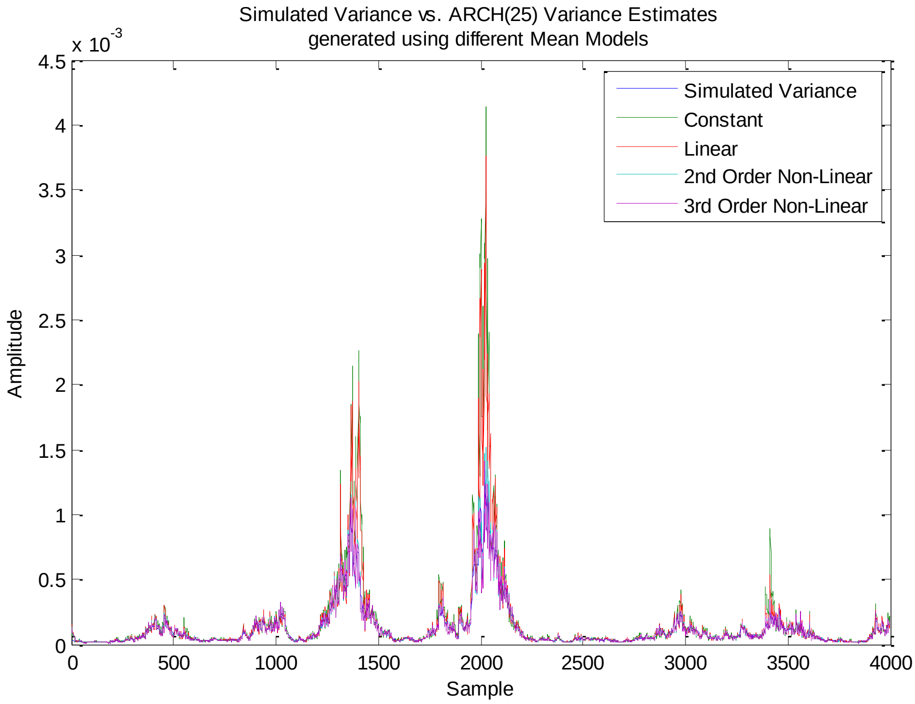


Figure 4.25 Simulated Variance vs. ARCH(25) Variance Estimates generated using different Mean Models

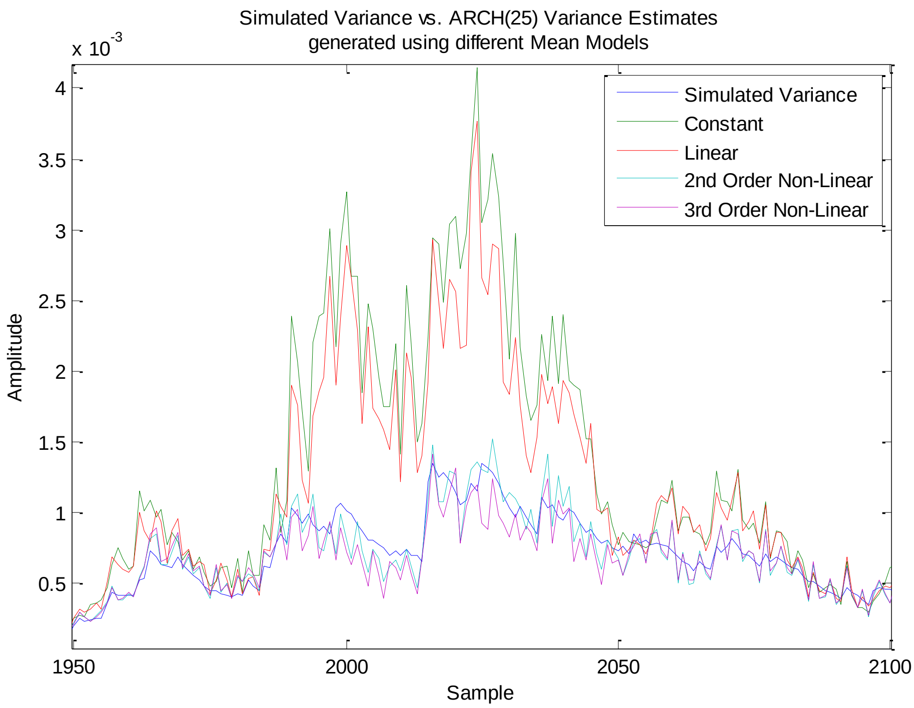


Figure 4.26 Simulated Variance vs. ARCH(25) Variance Estimates generated using different Mean Models (Samples 1950 to 2100)

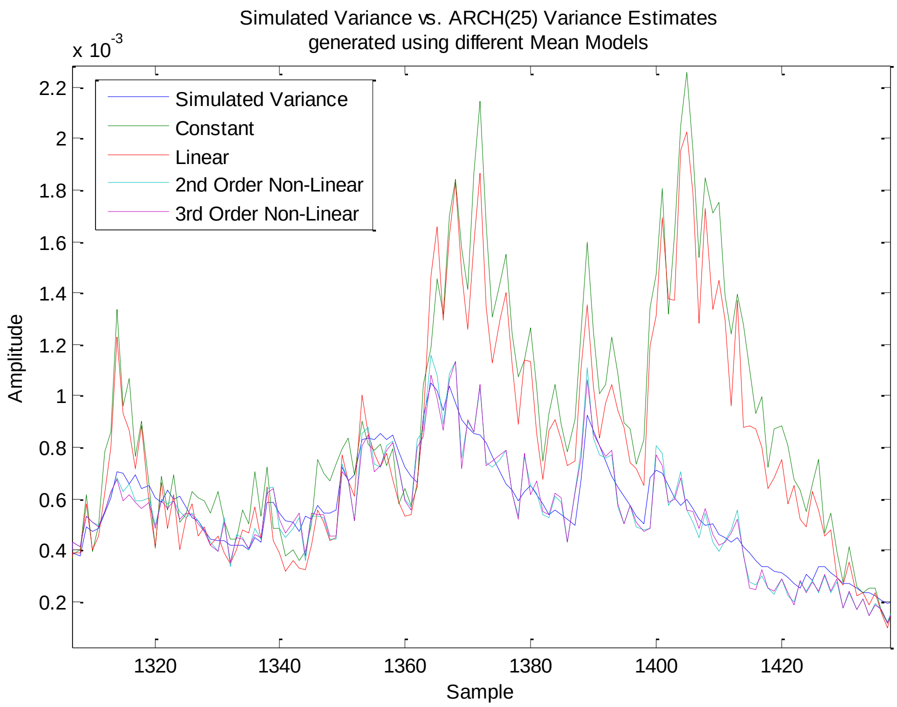


Figure 4.27 Simulated Variance vs. ARCH(25) Variance Estimates generated using different Mean Models (Samples 1300 to 1440)

Figure 4.26 and Figure 4.27 show that fitting a Constant or a Linear mean model to the simulated non-linear return series leads to noticeably large overestimation of the variance of the returns at the peaks. The variance estimates are much closer to the true variance if a second or third order non-linear mean model is fitted. It can be argued that this may be due to the usage of an ARCH(25) variance estimate due to the structure of the variance model being unknown. Hence, this was studied next.

The term selection and parameter estimation of all the models was carried out again, but this time it was assumed that the structure of the variance model was known (GARCH(1,1)). The Normalised Root Mean Squared Error (NRMSE) between the simulated variance and the GARCH(1,1) variance estimates are calculated for all the models and listed in Table 4.15. The minimum value is shaded in blue.

Table 4.15 NRMSE of GARCH(1,1) Variance Estimates of all Mean Models

Type of Mean Model	NRMSE (%)
Constant	120.09
Linear	92.56
Second Order Non-Linear	7.64
Third Order Non-Linear	12.07

The NRMSE of the GARCH(1,1) variance estimate from the second order non-linear mean model is the least (7.64%). From Table 4.14 and Table 4.15, note that the magnitude of NRMSE values of all the GARCH(1,1) variance estimates for the second and third order non-linear mean models are 7% to 10% lesser than that of the ARCH(25) variance estimates for the respective mean models. This demonstrates the importance of knowing the structure of the variance model.

But knowing the true structure of the mean model is more important. If the mean model was wrongly fitted whilst the variance model was correctly fitted, the NRMSE of the GARCH(1,1) variance estimate compared to the true variance would still be between 90% to 120%.

The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the selected mean models select and fitted using a GARCH(1,1) variance model are listed in Table 4.16. The minimum values are shaded in blue.

Table 4.16 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Mean Models when GARCH(1,1) Variance Model is used

Type of Mean Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Constant	129.4422	2.7453E-04
Linear	24.0756	1.8907E-04
Second Order Non-Linear	2.1353	4.1644E-05
Third Order Non-Linear	2.2874	4.3569E-05

Note that the second order non-linear mean model has the least ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data compared to the other mean models. Also, the constant mean model, which almost all modellers currently employ, has the worst performance statistics amongst all the mean models. Even when the structure of the variance model is known, the constant mean model fails the model selection test and yields a variance estimate that performs badly compared to the variance estimates obtained using other types of mean models.

Knowing the true structure of the variance model is important, but knowing the true structure of the mean model is much more important. Under-fitting the mean of the returns leads to a lot of the predictable elements being left in the residuals, $\hat{e}(t)$, thereby increasing their magnitude. When a variance model is fitted to these 'larger' residuals, the estimated variance is of bigger magnitude compared to the true variance. Model validation tests may suggest that the mean and the variance of the given return series have been adequately modelled, but that does not necessarily imply that the variance can be accurately estimated. There is a high possibility of the mean being under-fitted and the variance being over-fitted.

4.6.10 Comparison of Selected Mean Model to Vanilla GARCH Model

A vanilla GARCH model (constant mean model with a GARCH(1,1) variance model) is fitted to the simulated return series. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the fitted vanilla GARCH model have been listed and compared to all the fitted mean models in Table 4.17. The minimum values are shaded in blue.

Table 4.17 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Vanilla GARCH and Selected Mean Model fitted to Simulated Return Series

Type of Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Vanilla GARCH	129.4422	2.7453E-04
Selected Second Order Non-Linear Mean Model with ARCH(25) Variance Model	2.0435	4.1168E-05

The selected second order non-linear mean model with an ARCH(25) variance model performs much better than a vanilla GARCH model. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the selected model are lower than those of the vanilla GARCH model.

Figure 4.28 and Figure 4.29 show the one-step-ahead (OSA) return estimates generated using the vanilla GARCH model and the selected second order non-linear mean model.

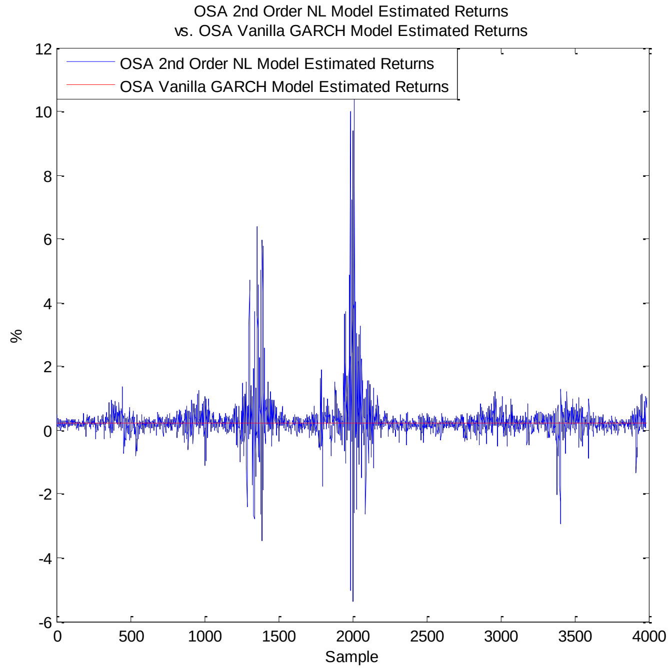


Figure 4.28 OSA Return Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for the Simulated Return Series

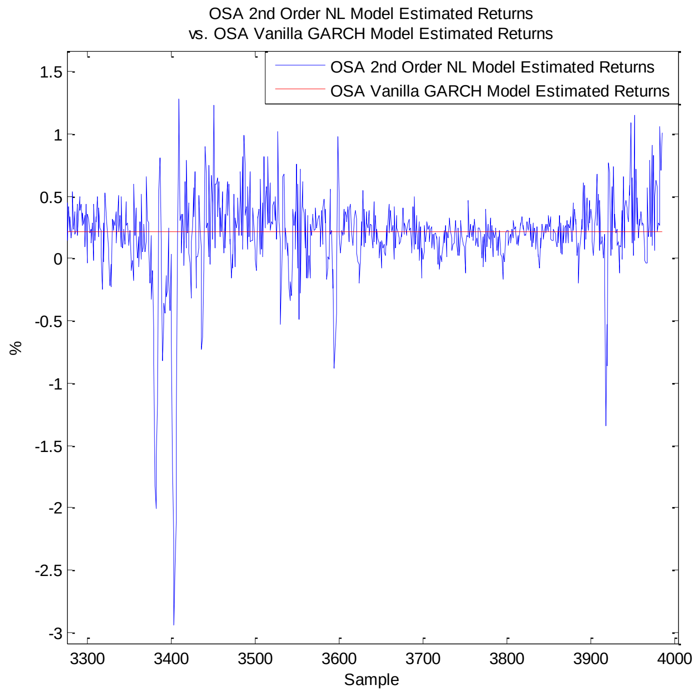


Figure 4.29 OSA Return Estimates generated using Vanilla GARCH Model and Second Order Non-Linear Mean Model for the Simulated Return Series (Samples 3300 to 4000)

From Figure 4.28 and Figure 4.29, it can be noted that using the second order non-linear mean model certainly captures the predictable elements of the mean of the returns, rather than just using a constant mean model and passing off the predictable elements to be included in the residuals. The standard deviation of the various return series are listed in Table 4.18.

Table 4.18 Standard Deviation of various return series for Simulated Data

Series	Standard Deviation
OSA Return Estimate of Constant Mean Model	0
OSA Return Estimate of Second Order NL Mean Model	0.0069
True Return Series	0.0129

Comparing the magnitudes of the Validation and Testing Sets of the OSA return estimates generated using the second order non-linear mean model to those of the true return series, 443 samples of 700 samples have the same magnitude. Hence, the magnitude of the returns is predicted right 63.2857% of the time.

4.7 Conclusions

1. The WOFR algorithm (Zhao, 2010) to model the mean of the returns was a major step forward towards fitting a non-linear mean model to the returns. One drawback was that no provision was made for the variance model of the returns being unknown, which is usually the case with real financial data. Also, there lacked a method to select the appropriate ERR cut-off value for term selection. Until now, there was no method to compare and determine which type of mean model (constant, linear or non-linear) is to be fitted to a given return series whose true mean and variance model is unknown.

2. In this Chapter, a new framework to model the mean of the returns based upon WOFR (Zhao, 2010), when the structure of the true variance model is unknown, is introduced. A method for term selection based upon the Akaike Information Criterion is introduced. This framework also helps select the best mean model from a selection of models.

3. The best mean model for a given return series is one that yields the least ARCH Test statistic of $\hat{e}(t)$ for the Validation Set and the least $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Complete data set.

4. Non-Linear correlation plots and test statistics (Friederich, 2011) and the ARCH Test statistic of $\hat{z}(t)$ for the Validation Set are used for model validation. The autocorrelation of the estimated residuals after fitting a mean model help validate the fitted mean model accurately.

5. The framework correctly identifies the mean model of a simulated return series with a non-linear mean model. All the fitted mean models pass standard financial model validation tests, but only one mean model is correct.

6. The effects of fitting different types of mean models on variance estimation are also examined. Fitting an inadequate mean model leads to inaccurate variance estimation, especially at the peaks. This is due to a lot of the predictable elements being left in the residuals, thereby increasing their magnitude. When a variance model is fitted to these 'larger' residuals, the estimated variance is of bigger magnitude compared to the true variance. Model validation tests may suggest that the mean and the variance of the given return series have been adequately modelled, but that does not necessarily imply that the variance has been accurately estimated.

7. The effects of not knowing the true structure of the variance model, and using an ARCH model to estimate the variance are also examined. Misspecifying the mean model leads to a more inaccurate variance estimate than that obtained by misspecifying the variance model.

The framework introduced in this chapter is applied to 2 real financial data sets in the next chapter.

Chapter 5

Application of the Extended WOFR approach to Financial Data

5.1 Introduction

The aim of this chapter is to demonstrate the applicability and benefits of the model selection and estimation approach introduced in Chapter 4, for estimating non-linear dynamic conditional mean models of a heteroskedastic time series, using FTSE100 and NASDAQ data sets.

The advantage of the proposed approach is demonstrated by comparing the resulting variance estimates obtained with the variance estimates generated by a 'vanilla' GARCH model (constant mean model with a GARCH(1,1) variance model).

This chapter is organised as follows. Section 5.2 presents the modelling and validation results for the FTSE100 index whilst Section 5.3 describes the modelling and validation results for the NASDAQ index. The conclusions are summarized in Section 5.4.

5.2 Application to the FTSE100

This section presents an analysis of a FTSE100 share index time series, which aims to identify the best conditional, linear or non-linear, mean model for this data set.

5.2.1 The Data

4001 samples of the price of the FTSE100 index, dating from 23rd September 1997 to 29th July 2013 are considered. The data is obtained from Yahoo! Finance (2013a). The price series is converted to a return series comprising of 4000 samples via continuous compounding (see equation (2.2)). The true structure of the mean and variance model of the obtained return series are of course unknown. The price and returns of the FTSE100 index are plotted in Figure 5.1.

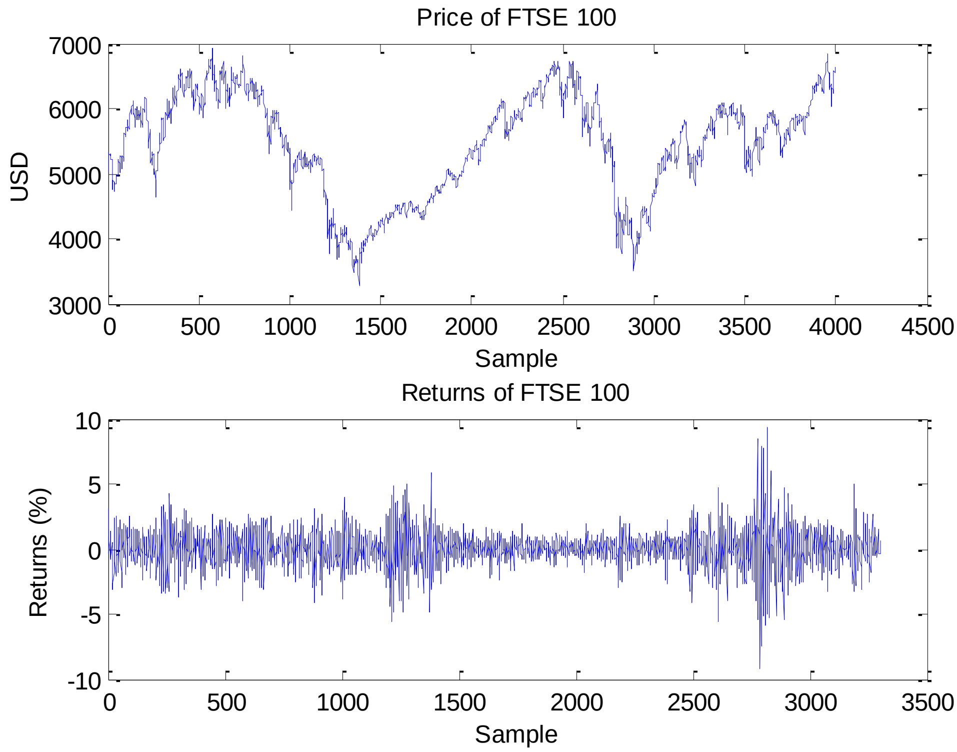


Figure 5.1 Price and Returns of the FTSE100 index

As in the case of the simulated example, the return series is split into three sets – Estimation, Validation and Testing Sets. The number of samples in each of these sets is the same as in the case of the simulated data example, and are listed in Table 4.2.

5.2.2 Autocorrelation Plots of Returns and Squared Returns

5.2.2.1 Linear Autocorrelation

The linear sample autocorrelation of the returns and the squared returns for 20 lags are shown in Figure 5.2.

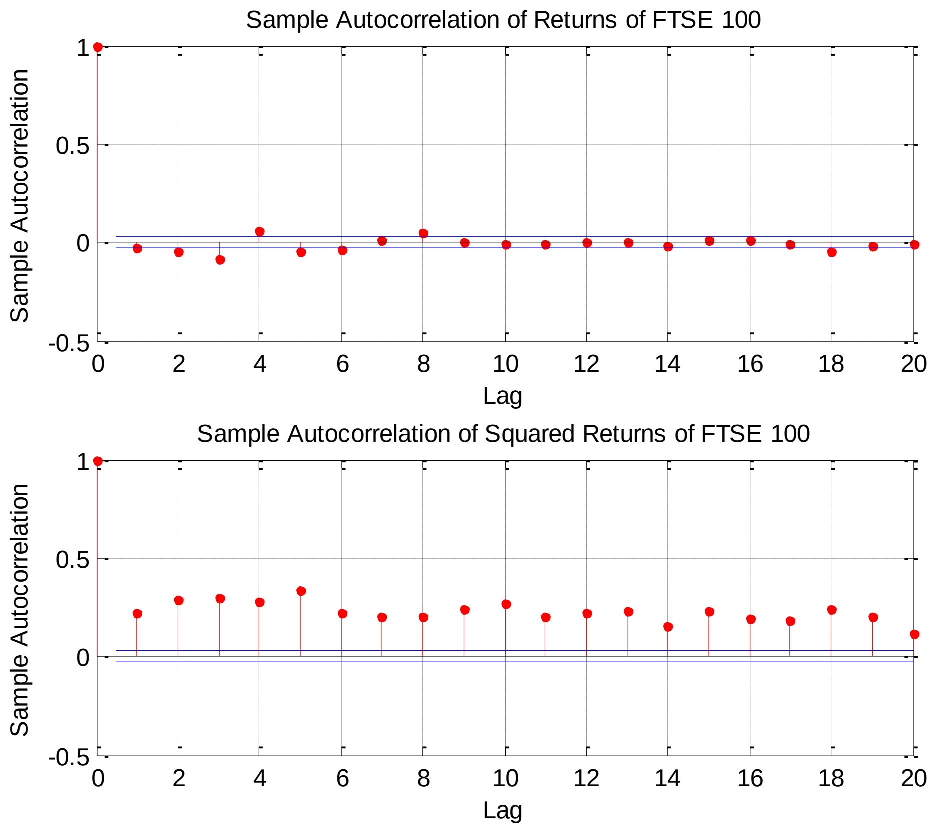


Figure 5.2 Sample Autocorrelation Plots of Returns and Squared Returns of the FTSE100

The sample autocorrelation of the returns for lags 3 and 4 lie outside the 95% confidence bands, indicating the possibility of the presence of the terms, $y(t-3)$ and $y(t-4)$, in the mean model.

The sample autocorrelation of the squared returns are outside the 95% confidence bands for all the lags, indicating the need to fit a variance model to the returns, in addition to a mean model.

5.2.2.2 Higher Order Autocorrelation

The higher order autocorrelation of the returns and the squared returns for the Estimation, Validation, Testing and Complete Data Set are also calculated and plotted in Figure 5.3 and Figure 5.4 respectively.

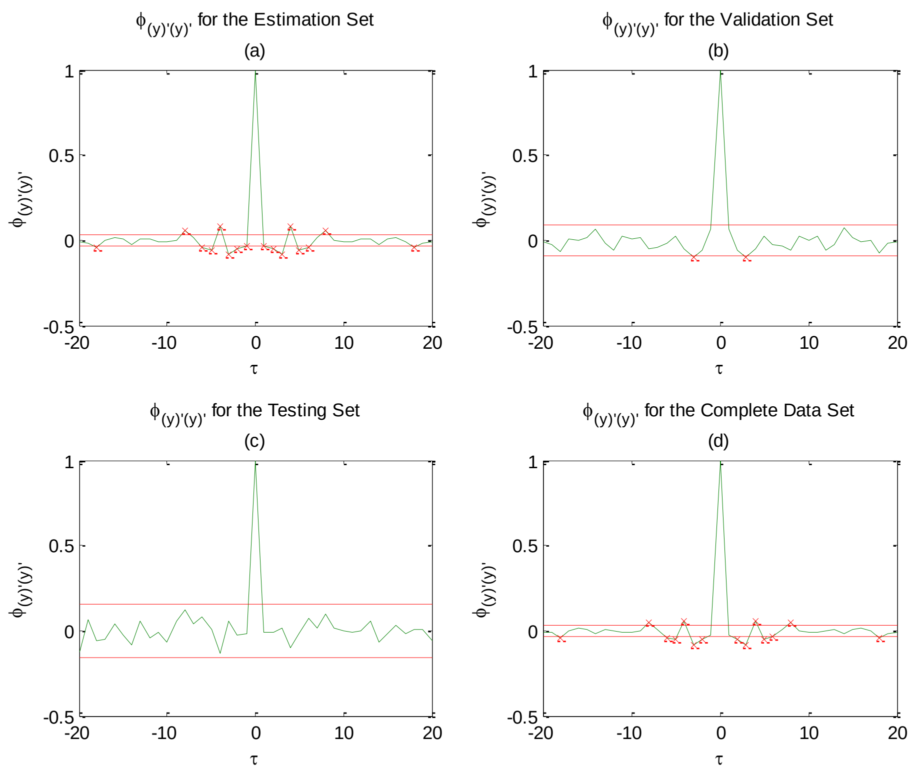


Figure 5.3 $\Phi_{(y)'(y)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Data Set

Figure 5.3 (d) shows that $\Phi_{(y)'(y)'}$ for lags 2, 3, 4, 5, 6 and 8 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags.

The higher order correlation violation statistic, $V_{(y)'(y)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set that enumerate the plots shown in Figure 5.3 are calculated and listed in Table 5.1.

Table 5.1 $V_{(y)'(y)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y)'(y)'}$ **(Estimation Set)**	$V_{(y)'(y)'}$ **(Validation Set)**	$V_{(y)'(y)'}$ **(Testing Set)**	$V_{(y)'(y)'}$ **(Complete Data Set)**
6.9041E-08	2.2930E-09	0	4.8384E-08

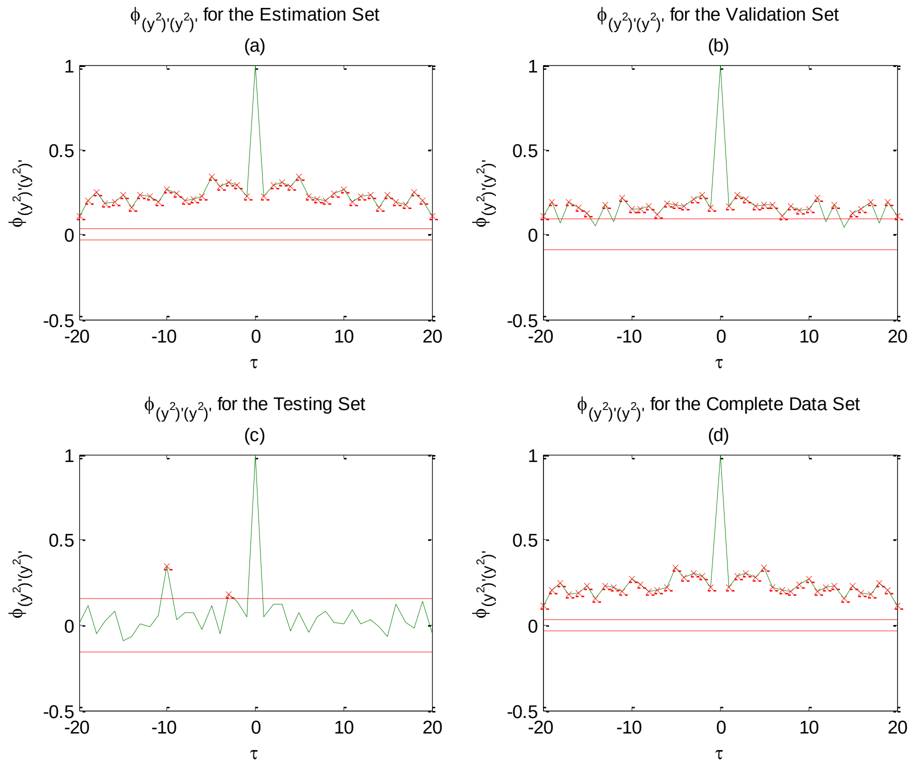


Figure 5.4 $\Phi_{(y^2)'(y^2)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Set

Figure 5.4 (d) shows that $\Phi_{(y^2)'(y^2)'}$ is significantly outside the 95% confidence bands for all the lags. This indicates the need to fit a variance model to the returns, in addition to fitting a mean model. The non-linear correlation violation statistic, $V_{(y^2)'(y^2)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set that enumerate the plots shown in Figure 5.4 are calculated and listed in Table 5.2.

Table 5.2 $V_{(y^2)'(y^2)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y^2)'(y^2)'}$ **(Estimation Set)**	$V_{(y^2)'(y^2)'}$ **(Validation Set)**	$V_{(y^2)'(y^2)'}$ **(Testing Set)**	$V_{(y^2)'(y^2)'}$ **(Complete Data Set)**
6.7169E-05	9.8607E-06	4.1998E-06	6.5828E-05

5.2.3 Candidate Mean Models

Four different candidate mean models are fitted to the returns of the FTSE100 index.

5.2.3.1 Constant Mean Model

The first is a constant mean model, $y(t) = a_0$, where a_0 is to be estimated.

5.2.3.2 Linear Candidate Mean Model

The second is a linear candidate mean model. This is an AR(10) model with a linear noise model consisting of 5 lagged noise terms

$$y(t) = a_0 + \sum_{i=1}^{10} a_i y(t-i) + \sum_{j=1}^{5} b_j e(t-j) \tag{5.1}$$

where a_0, a_i, and b_j are coefficients to be estimated.

The number of lagged linear terms to be included in the linear candidate mean model is denoted as n_l. As explained in Section 5.2.2.2, from Figure 5.3 (d), $\Phi_{(y)'(y)'}$ for lags 2, 3, 4, 5, 6 and 8 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags. Hence, n_l is selected to be a round figure of 10.

In order to improve parameter estimation, a linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the linear mean model is selected to be $n_e = 5$. It must be noted that a linear mean model with a very large noise model can make the model unstable and give unstable one-step-ahead estimates or the returns.

5.2.3.3 Second Order Non-Linear Candidate Mean Model

The third candidate mean model is a second-order non-linear mean model. This is

$$\begin{aligned} y(t) &= a_0 + \sum_{i=1}^{10} a_i y(t-i) \\ &+ \sum_{j=1,k=1,i=11}^{j=5,k=5,i=21} a_i y(t-j) y(t-k) + \sum_{l=1}^{5} b_l e(t-l) \end{aligned} \tag{5.2}$$

where a_0, a_i, and b_l are coefficients to be estimated.

As explained in Section 5.2.2.2, from Figure 5.3 (d), $\Phi_{(y)'(y)'}$ for lags 2, 3, 4, 5, 6 and 8 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags. Hence, n_l is selected to be 10.

The number of terms in the candidate mean model increases exponentially as the maximum lag of the non-linear terms to be included in the candidate mean model is increased. Hence, to keep the size of the candidate mean model reasonable, so as not to have extremely high computational time, the maximum lag of the second order non-linear terms to be included in the non-linear candidate mean model is selected to be 5.

A linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 5$.

5.2.3.4 Third Order Non-Linear Candidate Mean Model

The final candidate model is a third-order non-linear mean model. This is

$$y(t) = a_0 + \sum_{i=1}^{10} a_i y(t-i) + \sum_{j=1,k=1,i=11}^{j=5,k=5,i=21} a_i y(t-j)y(t-k)$$
$$+ \sum_{j=1,k=1,l=1,i=22}^{j=5,k=5,l=5,i=40} a_i y(t-j)y(t-k)y(t-l) + \sum_{l=1}^{5} b_l e(t-l) \qquad (5.3)$$

where a_0, a_i, and b_l are coefficients to be estimated. n_l is selected to be 10 for the same reasons specified for the second order non-linear candidate mean model in Section 5.2.3.3. The maximum lag of the second order non-linear terms to be included in the non-linear candidate mean model is selected to be 5. A linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 5$.

5.2.4 Estimation of Constant Mean Model

The values of the hyper-parameters used when fitting a constant mean model are given in Table 5.3.

Table 5.3 Hyper-parameters for fitting Constant Mean Model

Hyper-parameter	Value
No. of Samples in Estimation Set (n_{est})	3300
No. of Samples in Estimation Set (n_{val})	500
No. of Samples in Estimation Set (n_{test})	200
No. of Lagged Linear Noise Terms in Mean Model (n_e)	0
No. of Lags for non-linear Correlation Tests (n_{corr})	20

The constant term fitted is 1.1900E-04, and has an ERR of 0.0167.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the constant mean model are plotted in Figure 5.5.

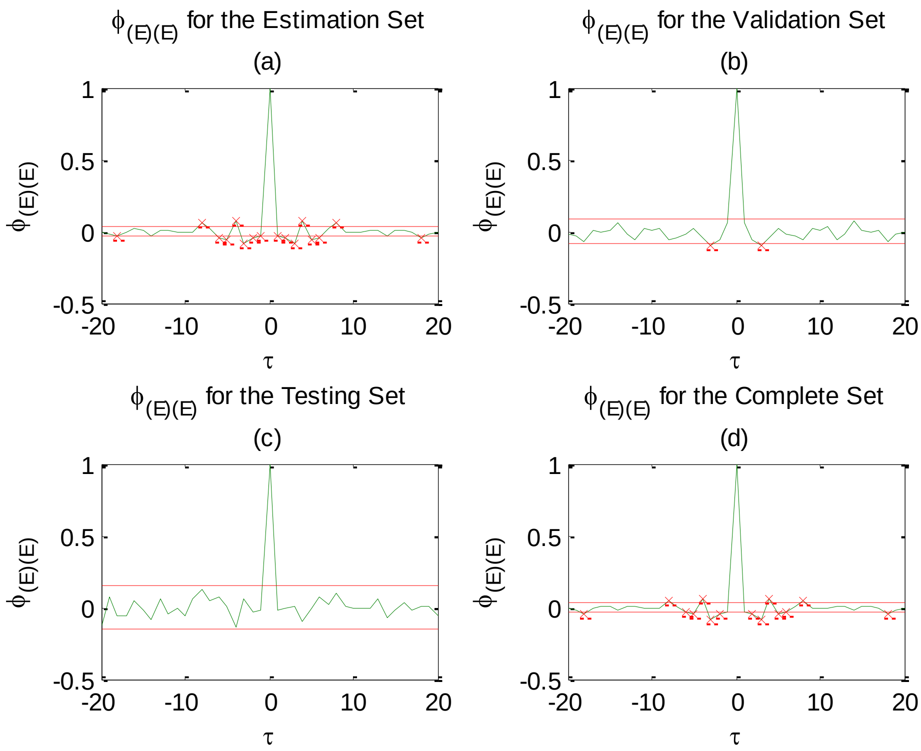


Figure 5.5 Autocorrelation Plots of Residuals for the Selected Constant Mean Model

In Figure 5.5 (d), note that there exists significant autocorrelation of the residuals at lags 2, 3, 4 and 5 implying that the fitted constant mean model is inadequate.

To validate the fitted mean and variance models, the higher order correlation of the squared residuals and squared standardised residuals obtained after fitting the selected constant mean model are plotted in Figure 5.6.

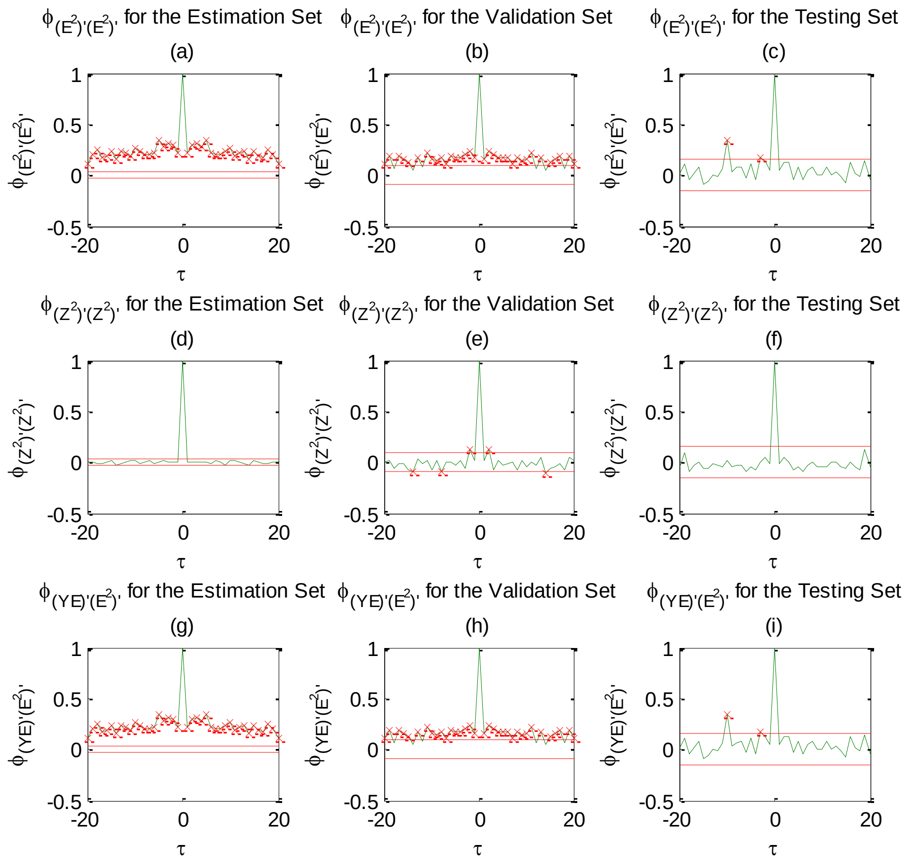


Figure 5.6 Higher Order Correlation Plots for the Selected Constant Mean Model

From Figure 5.6 (d) and (f), the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained from fitting a constant mean model, do not indicate any higher order autocorrelation. In Figure 5.6 (e), $\Phi_{(z^2)'(z^2)'}$ lies on or just outside the 95% confidence band for a few lags. The confidence violations are not considerably large, and the ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 0.0457, which is lesser than the critical value of 3.8415. Hence, the mean and the variance of the returns can be considered to be adequately fitted.

5.2.4.1 $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(z^2)'(z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation, Validation and Testing Sets

The higher order correlation plots in Figure 5.6 are quantified by calculating the confidence violation statistics using equations (4.7), (4.8) and (4.9). Table 5.4 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(z^2)'(z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation, Validation and Testing Sets for the constant mean model.

Table 5.4 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation, Validation and Testing Sets in Constant Mean Model

	Estimation Set	Validation Set	Testing Set
$V_{(\varepsilon^2)'(\varepsilon^2)'}$	6.7520E-05	9.8071E-06	4.1489E-06
$V_{(Z^2)'(Z^2)'}$	0	8.2247E-08	0
$V_{(Y\varepsilon)'(\varepsilon^2)'}$	6.7434E-05	9.8074E-06	4.0716E-06
ARCH Test Statistic of $\hat{z}(t)$	0.2156	0.0457	0.0168

From Figure 5.6 (a), (b), (c), (g), (h) and (i), the squared estimated residuals, $\hat{e}^2(t)$, obtained from fitting a constant mean model, indicate higher order correlation and autocorrelation. The corresponding confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$, for the Estimation and Validation Sets in Table 5.4 are of the order of 1E-04 which is considerably high in this context. These results suggest that a variance model is required to be fitted to the given return series.

For an adequately fit mean and variance model, the value of $V_{(Z^2)'(Z^2)'}$ is required to be zero and the value of the ARCH Test Statistic of $\hat{z}(t)$ is required to be lesser than the critical value of 3.8415, thereby confirming the absence of any higher order correlation in the estimated standardised residuals. The magnitude of the value of $V_{(Z^2)'(Z^2)'}$ for the Validation Set is extremely small, and close to zero, which is acceptable. Also, the values of ARCH Test Statistic of $\hat{z}(t)$ for all the data sets are less than 3.8415. These results imply that no heteroskedastic effects are present in $\hat{z}(t)$ suggesting that the mean and the variance of the returns have been adequately modelled.

5.2.5 Estimation of Linear Mean Model using WOFR

The next type of mean model to be fitted to the returns of the FTSE100 index is the linear candidate mean model listed in Section 5.2.3.2. WOFR is performed on the linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select the best linear mean model that best describes the mean of the returns of FTSE100.

The values of the hyper-parameters used when fitting a linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 5.3. The only difference is that a linear noise model is to be fitted as well, hence n_e is selected to be 5.

The results of the WOFR analysis are given in Table 5.5.

Table 5.5 Terms in Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-5)$	-0.0367	0.1102
2	$y(t-3)$	-0.0376	0.1056
3	$y(t-2)$	-0.0261	0.0578
4	$y(t-1)$	-0.0215	0.0413
5	$y(t-10)$	0.0195	0.0347
6	1	1.2241E-04	0.0215
7	$y(t-8)$	0.0142	0.0175
8	$y(t-4)$	0.0113	0.0106
9	$y(t-7)$	0.0052	0.0026
10	$y(t-6)$	-0.0042	0.0015
11	$y(t-9)$	4.9791E-04	2.4340E-05

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are now carried out on the re-ordered terms of the linear candidate mean model.

5.2.5.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 5.5 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18). Figure 5.7 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the linear mean model.

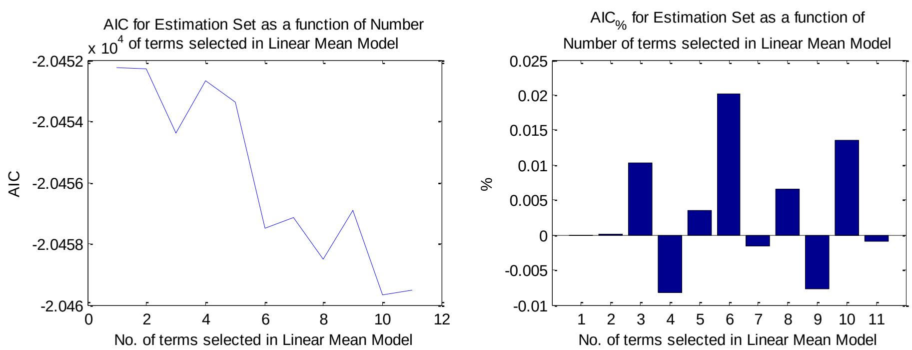


Figure 5.7 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Linear Mean Model

From Figure 5.7, $AIC_{\%}$ is the highest (0.0201%) when 6 terms from the top of Table 5.5 are included. Also, the AIC is -2.0458E+04 which is very close to the minimum value of 2.0460E+04. Hence, 6 terms from the top of Table 5.5 are selected to be included in the linear mean model.

5.2.5.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the linear mean models using equations (4.7), (4.8) and (4.9). Figure 5.8 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.9 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.10 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

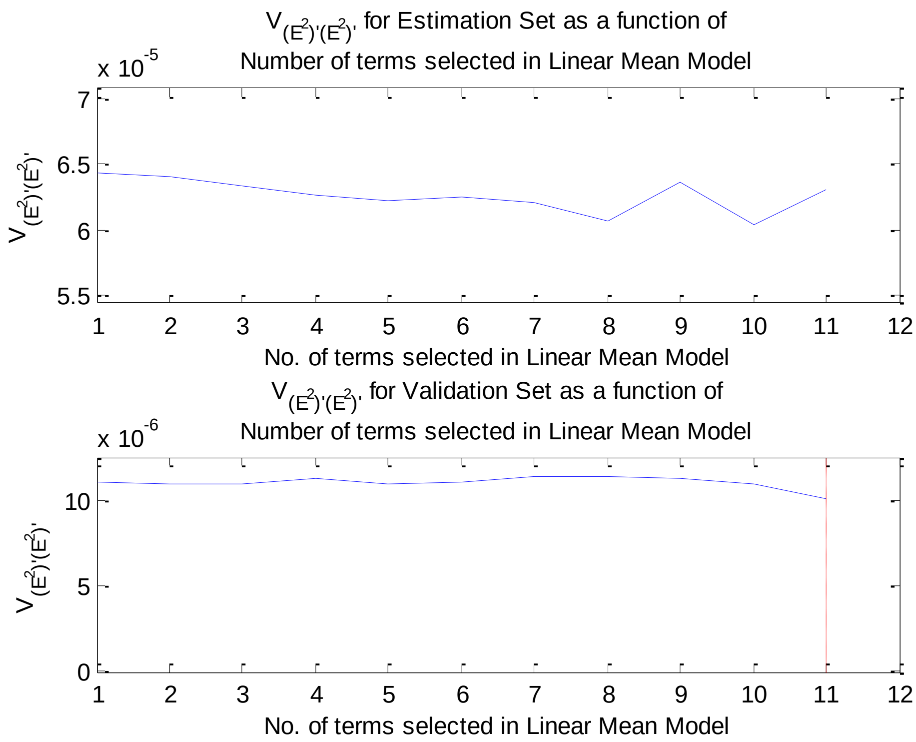


Figure 5.8 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of all the terms from the top of Table 5.5, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 1.0091E-05. For the linear mean model comprising of 6 terms from the top of Table 5.5, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 1.1086E-05 which is very close to the minimum value.

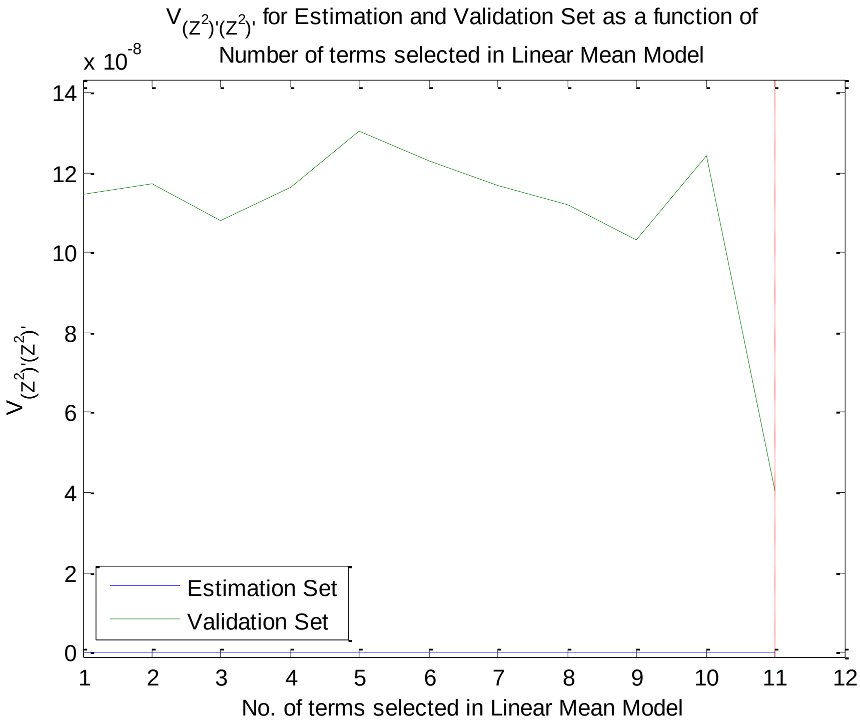


Figure 5.9 $V_{(Z^2)'(Z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of all the 11 terms from Table 5.5, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 4.0589E-08. In this case, 6 terms from the top of Table 5.5 are selected to be included in the linear mean model which yields a value of 1.2288E-07 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude and very close to the minimum value of 4.0589E-08. These results suggest that the mean and the variance of the returns have been adequately modelled.

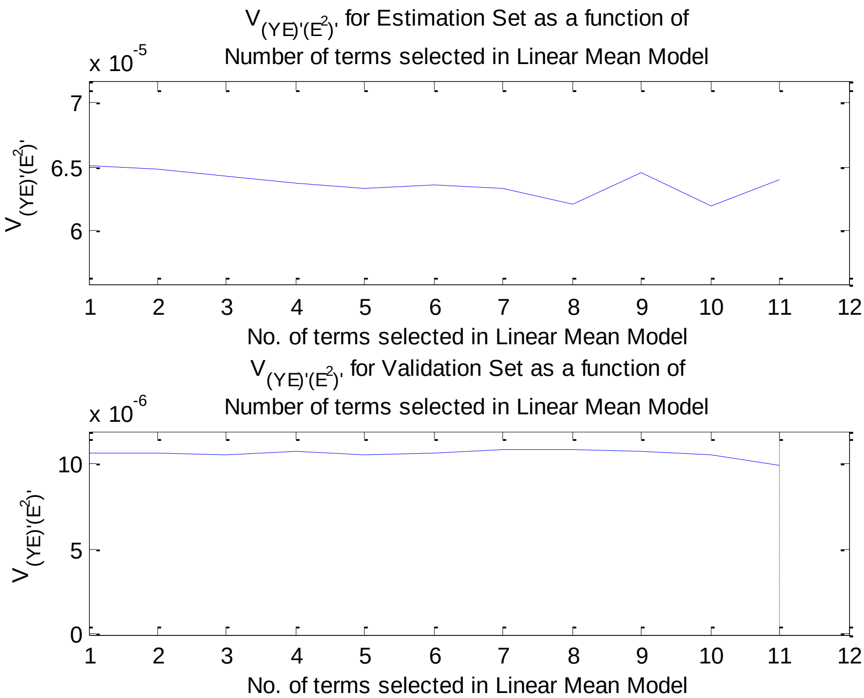


Figure 5.10 $\boldsymbol{V_{(Y\varepsilon)'(\varepsilon^2)'}}$ **for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model**

For the linear mean model comprising of all the terms from the top of Table 5.5, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 9.9211E-06. For the linear mean model comprising of 6 terms from the top of Table 5.5, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is 1.0637E-05 which is very close to the minimum value.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected linear mean model are plotted in Figure 5.11.

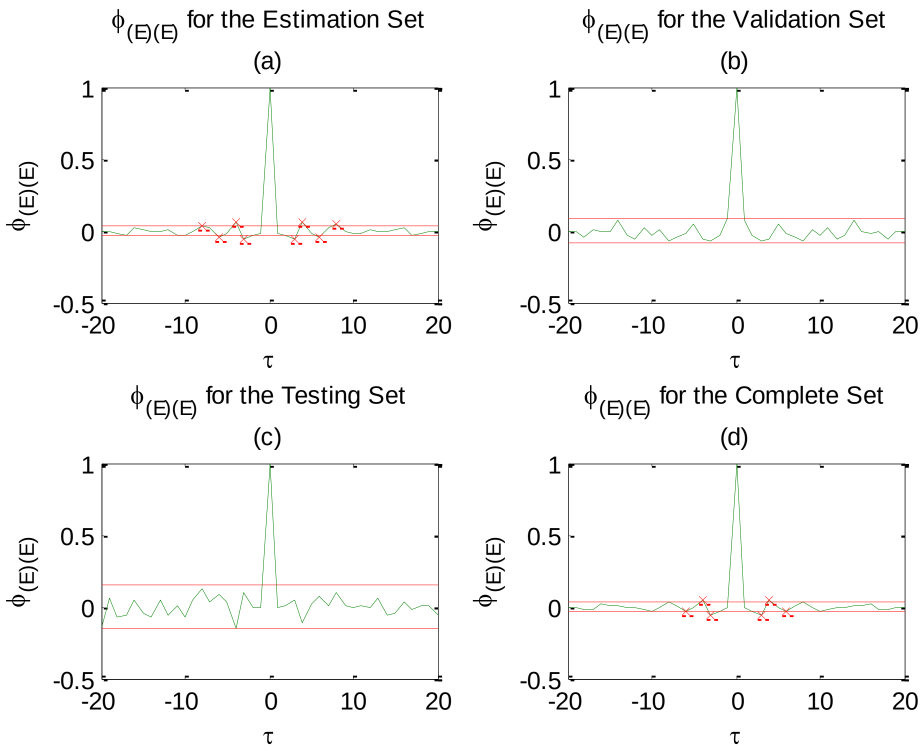


Figure 5.11 Autocorrelation Plots of Residuals for the Selected Linear Mean Model

In Figure 5.11 (d), note that there exists significant autocorrelation of the residuals at lags 2 and 3 implying that the fitted linear mean model is inadequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.5 (d)) which indicates that the selected linear mean model captures the predictable elements of the mean of the returns much better than the selected constant mean model.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected linear mean model (with 4 process terms) are shown in Figure 5.12.

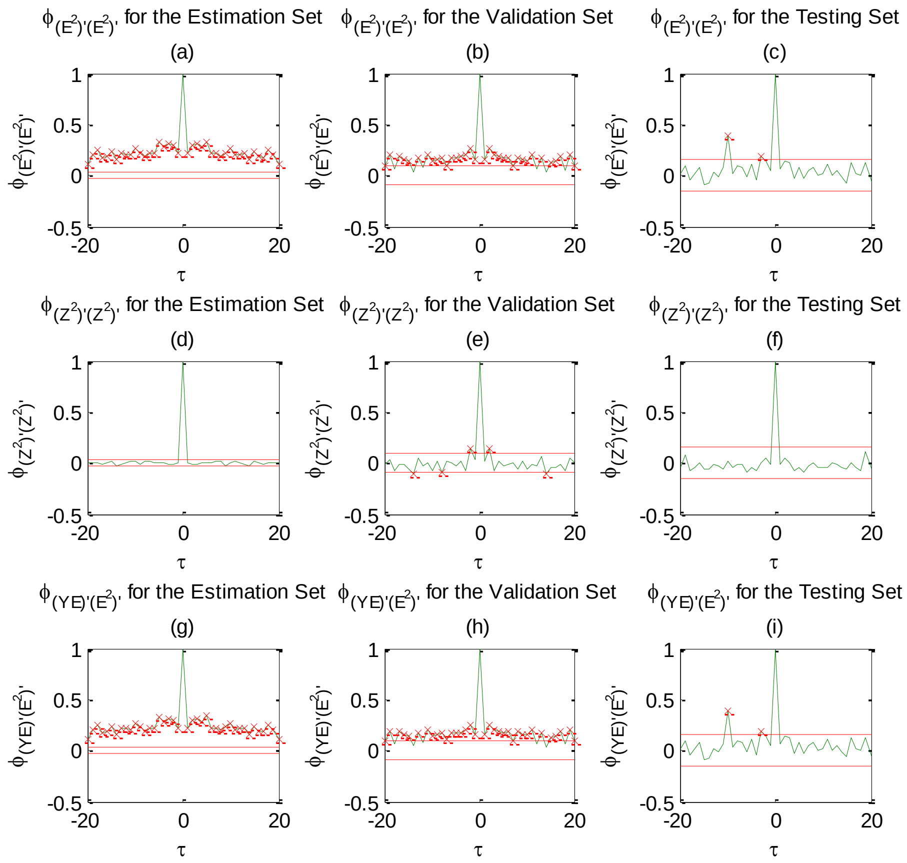


Figure 5.12 Higher Order Correlation Plots for the Selected Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.12 (d) and (f), there exist no violations of the 95% confidence bands. In Figure 5.12 (e), $\Phi_{(z^2)'(z^2)'}$ lies on or just outside the 95% confidence band for a few lags. The confidence violations are not considerably large, and the ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 0.0628, which is smaller than the critical value of 3.8415. This suggests that the mean and the variance of the returns have been modelled adequately.

The terms selected in the linear mean model along with the coefficient estimates and ERR values are listed in Table 5.6. The coefficient estimates of the 5 noise terms are also included.

Table 5.6 Linear Mean Model fitted to FTSE100

No.	Term	Parameter Estimate	ERR
1	$y(t-5)$	-0.2096	0.1048
2	$y(t-3)$	-0.1994	0.1029
3	$y(t-2)$	0.0035	2.2423E-06
4	$y(t-1)$	0.5263	0.0273
5	$y(t-10)$	0.0107	0.0390
6	1	9.7124E-05	0.0185
7	$e(t-1)$	-0.5484	0.0438
8	$e(t-2)$	-0.0172	0.0563
9	$e(t-3)$	0.1771	0.0057
10	$e(t-4)$	0.0273	0.0125
11	$e(t-5)$	0.1637	0.0186

So far, the constant and the linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in both cases.

5.2.6 Estimation of Second Order Non-Linear Mean Model using WOFR

The next type of mean model to be fitted to the returns of the FTSE100 index is the second order non-linear candidate mean model listed in Section 5.2.3.3. As in the case of the linear candidate mean model, WOFR is performed on the second order non-linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select a second order non-linear mean model that best describes the mean of the returns of FTSE100.

The values of the hyper-parameters used when fitting a second order non-linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 5.3. A linear noise model is also fitted. The maximum noise lag, n_e, is selected to be 5.

The results of the WOFR analysis are given in Table 5.7.

Table 5.7 Terms in Second Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-3)$	2.5447	0.1246
2	$y(t-5)$	-0.0382	0.1167
3	$y(t-3)$	-0.0380	0.1127
4	$y(t-2)$	-0.0242	0.0622
5	$y(t-10)$	0.0188	0.0415
6	$y(t-3)y(t-5)$	1.7582	0.0318
7	$y(t-2)y(t-3)$	1.8644	0.0350
8	$y(t-1)y(t-2)$	1.4161	0.0307
9	1	2.0147E-04	0.0281
10	$y(t-5)y(t-5)$	-1.0810	0.0256
11	$y(t-4)y(t-5)$	-1.5215	0.0305
12	$y(t-1)$	-0.0146	0.0225
13	$y(t-1)y(t-5)$	1.1186	0.0196
14	$y(t-8)$	0.0140	0.0162
15	$y(t-3)y(t-4)$	-0.9456	0.0115
16	$y(t-2)y(t-4)$	0.7326	0.0085
17	$y(t-4)$	0.0104	0.0079
18	$y(t-3)y(t-3)$	0.5437	0.0087
19	$y(t-4)y(t-4)$	-0.4512	0.0056
20	$y(t-7)$	0.0049	0.0024
21	$y(t-6)$	-0.0049	0.0021
22	$y(t-9)$	-0.0036	0.0013
23	$y(t-1)y(t-1)$	0.1424	9.7772E-04
24	$y(t-1)y(t-4)$	-0.1316	3.1291E-04
25	$y(t-2)y(t-2)$	0.1026	2.8300E-04
26	$y(t-2)y(t-5)$	0.0645	6.3475E-05

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the second order non-linear candidate mean model.

5.2.6.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 5.7 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18).

Figure 5.13 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the second order non-linear mean model.

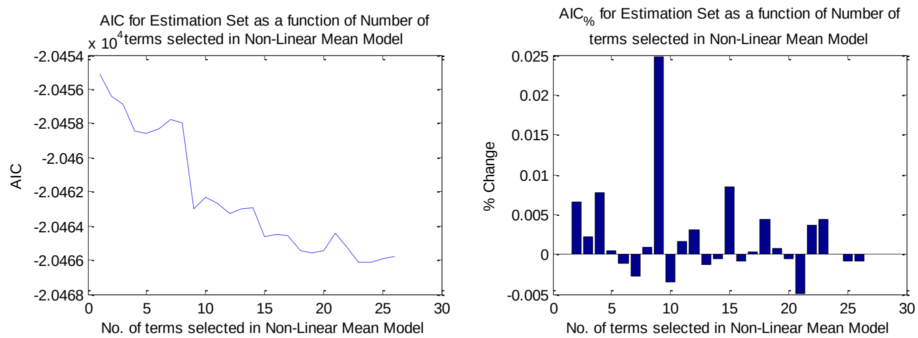


Figure 5.13 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Second Order Non-Linear Mean Model

From Figure 5.13, the AIC drastically decreases to a value of -2.0463E+04 till 9 terms from the top of Table 5.7 are selected to be included in the second order non-linear mean model. Also, $AIC_{\%}$ for the 9th iteration is the highest (0.0247%). The addition of further terms to the mean model decreases the value of AIC gradually to a minimum value of -2.0466E+04.

5.2.6.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the non-linear mean models using equations (4.7), (4.8) and (4.9). Figure 5.14 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.15 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.16 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

For the non-linear mean model comprising of only the first term from Table 5.7, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 1.0678E-05. For the non-linear mean model comprising of 9 terms from the top of Table 5.7, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 1.2065E-05 which is very close to the minimum value.

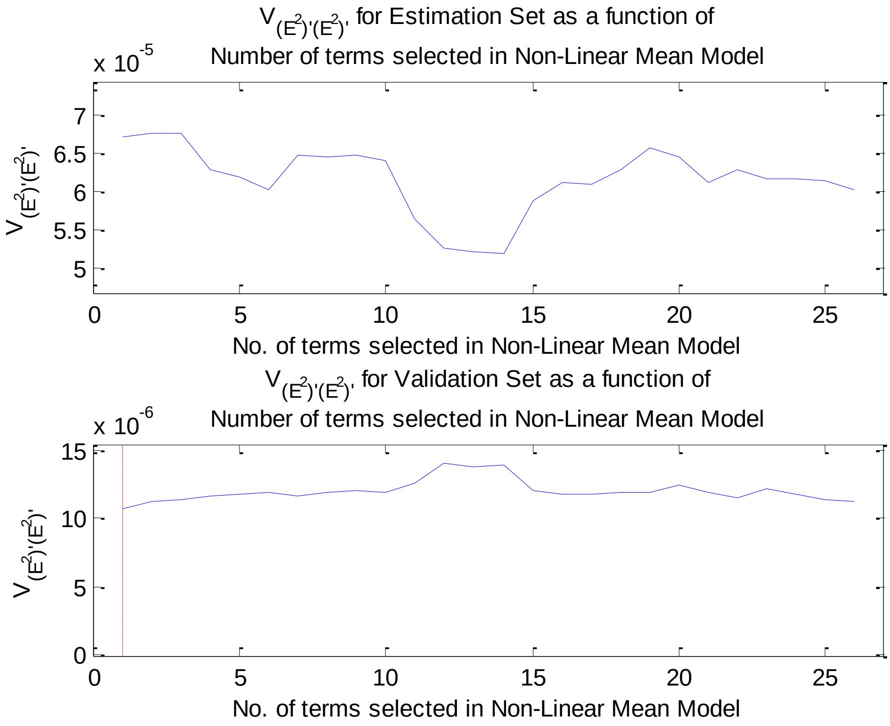


Figure 5.14 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

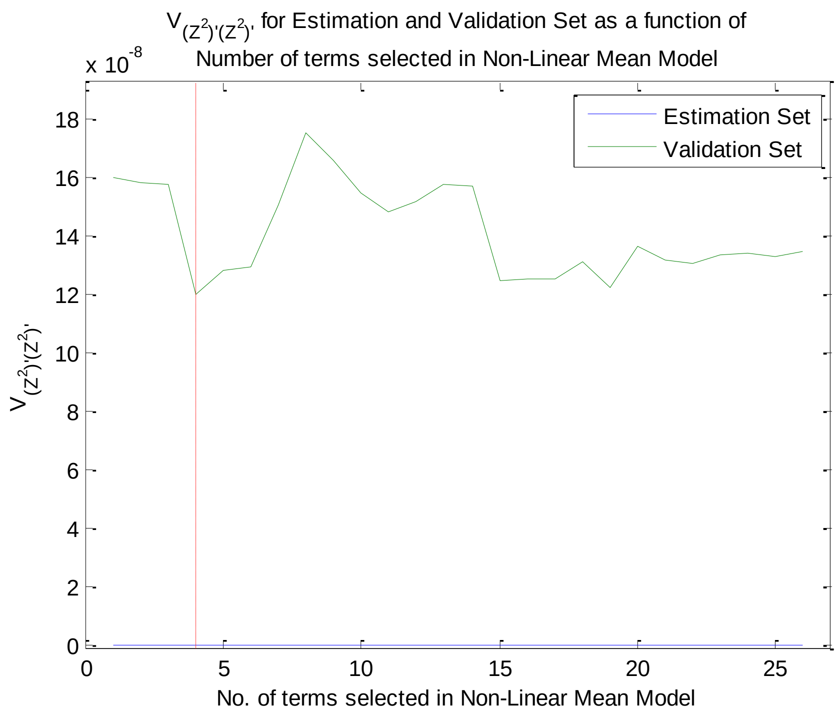


Figure 5.15 $V_{(z^2)'(z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

For the non-linear mean model comprising of 4 terms from Table 5.7, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 1.2006E-07. In this case, 9 terms from the top of Table 5.7 are selected to be included in the non-linear mean model which yields a value of 1.6548E-07 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude and very close to the minimum value. These results suggest that the mean and the variance of the returns have been adequately modelled.

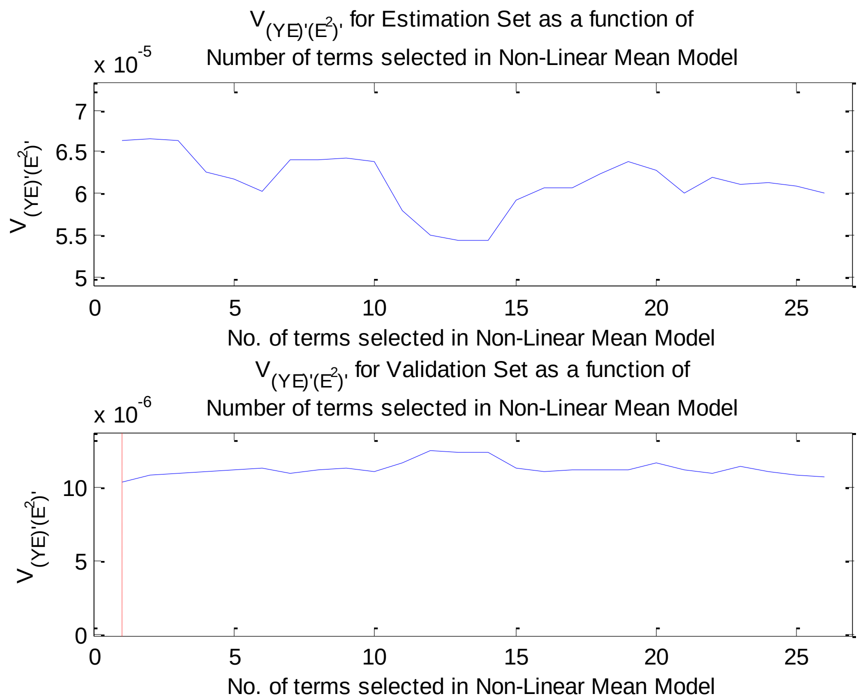


Figure 5.16 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

For the non-linear mean model comprising of the first term from Table 5.7, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 1.0353E-05. For the non-linear mean model comprising of 9 terms from the top of Table 5.7, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is 1.1221E-05 which is very close to the minimum value.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 5.17.

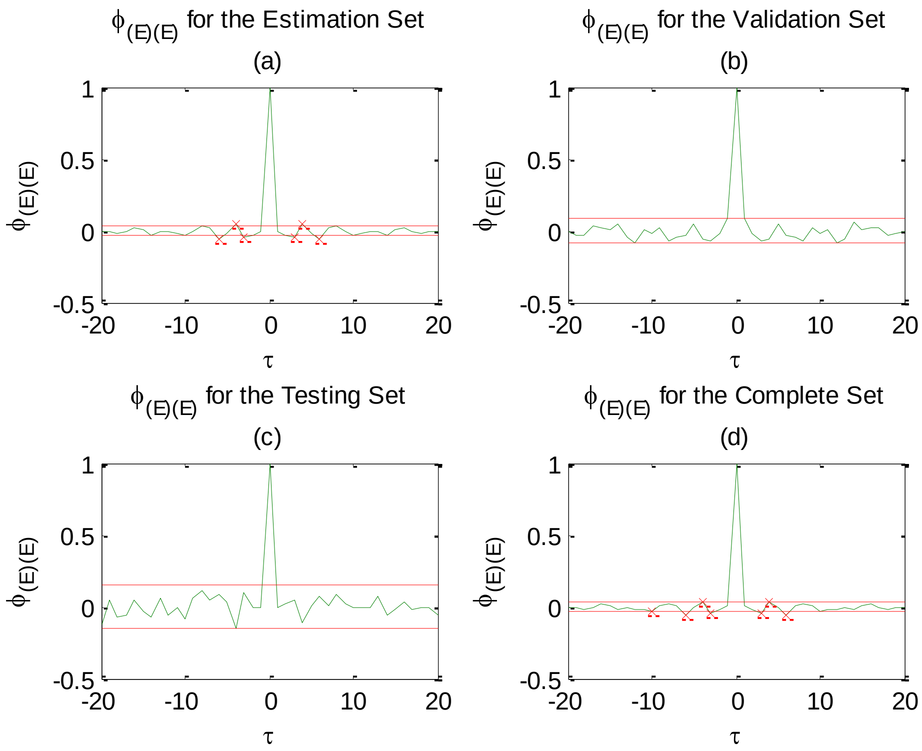


Figure 5.17 Autocorrelation Plots of Residuals for the Selected Second Order Non-Linear Mean Model

In Figure 5.17 (d), note that there exists slight autocorrelation of the residuals at lags 2, 3 and 4 implying that the fitted second order non-linear mean model may be inadequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.5 (d)) and the residuals obtained after fitting the selected linear mean model (Figure 5.11 (d)) which indicates that the selected second order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant and linear mean models.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected second order non-linear mean model are shown in Figure 5.18.

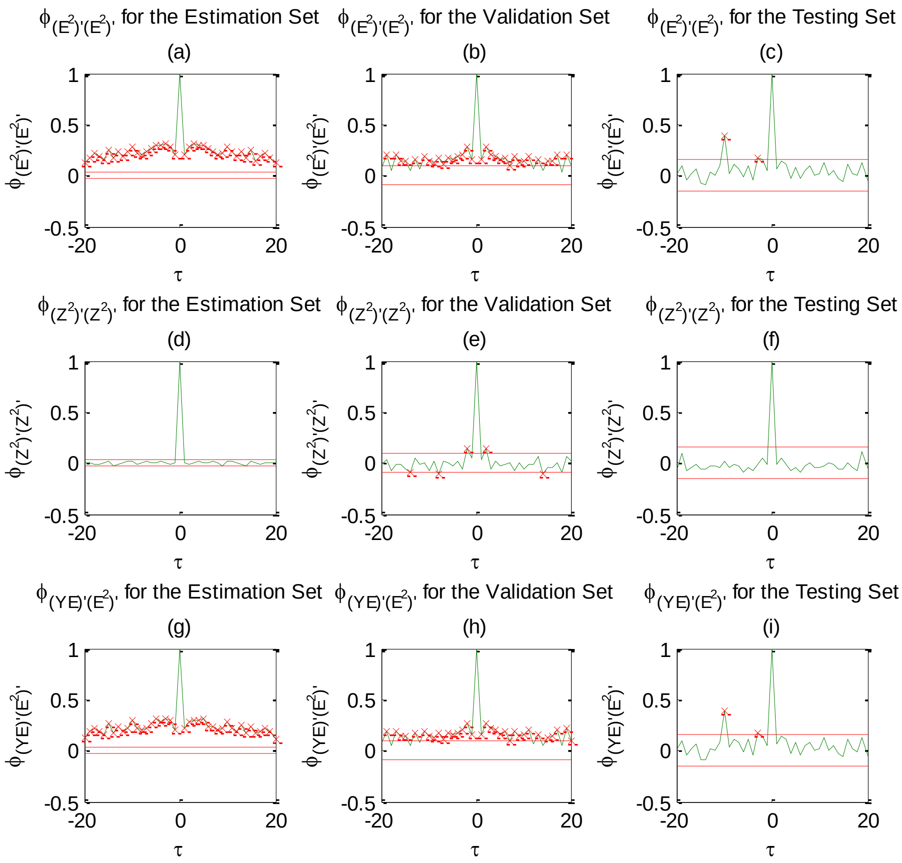


Figure 5.18 Higher Order Correlation Plots for the Selected Second Order Non-Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.18 (d) and (f), there exist no violations of the 95% confidence bands. In Figure 5.18 (e), $\Phi_{(z^2)'(z^2)'}$ lies on or just outside the 95% confidence band for a few lags. The confidence violations are not considerably large, and the ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 0.4220, which is smaller than the critical value of 3.8415 suggesting that the mean and the variance of the returns have been fitted adequately.

The selected terms along with the parameter estimates and ERR values are listed in Table 5.8. The coefficient estimates of the 5 noise terms are also included.

Table 5.8 Second Order Non-Linear Mean Model fitted to FTSE100

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-3)$	2.9097	0.1506
2	$y(t-5)$	-0.1873	0.1091
3	$y(t-3)$	-0.0097	0.1115
4	$y(t-2)$	0.1064	0.0027
5	$y(t-10)$	0.0135	0.0411
6	$y(t-3)y(t-5)$	1.4767	0.0312
7	$y(t-2)y(t-3)$	1.7885	0.0304
8	$y(t-1)y(t-2)$	1.3905	0.0268
9	1	1.4516E-04	0.0261
10	$e(t-1)$	-0.0176	0.0318
11	$e(t-2)$	-0.1330	0.0668
12	$e(t-3)$	-0.0245	2.1716E-04
13	$e(t-4)$	0.0146	0.0104
14	$e(t-5)$	0.1545	0.0091

So far, the constant, the linear and the second order non-linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 3 cases.

5.2.7 Estimation of Third Order Non-Linear Mean Model using WOFR

The last mean model to be fitted to the returns of the FTSE100 index is the third order non-linear candidate mean model listed in Section 5.2.3.4. As in the case of the previous candidate mean models, WOFR is performed on the third order non-linear candidate mean model, and the terms are re-ordered in decreasing order of their ERR values. Term selection is then carried out to select a third order non-linear mean model that best describes the mean of the returns of FTSE100.

The values of the hyper-parameters used when fitting a third order non-linear mean model are the same as in the case of fitting a constant mean model, and are given in Table 5.3. The only difference is that a linear noise model is to be fitted as well, hence n_e is selected to be 5.

The results of the WOFR analysis are shown in Table 5.9.

Table 5.9 Terms in Third Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-4)y(t-4)y(t-4)$	33.5657	0.4170
2	$y(t-1)y(t-3)y(t-5)$	-66.2826	0.3457
3	$y(t-1)y(t-5)y(t-5)$	-61.1157	0.3030
4	$y(t-1)y(t-1)y(t-2)$	-100.8959	0.1674
5	$y(t-5)$	-0.0315	0.0986
6	$y(t-3)$	-0.0298	0.0950
7	$y(t-2)y(t-3)y(t-3)$	-65.7263	0.0834
8	$y(t-3)y(t-5)$	2.6237	0.0715
9	$y(t-10)$	2.0103E-02	0.0631
10	$y(t-2)y(t-2)y(t-4)$	118.6700	0.0532
11	$y(t-1)y(t-1)y(t-5)$	73.7220	0.0489
12	$y(t-2)y(t-2)y(t-2)$	37.1373	0.0483
13	$y(t-1)y(t-2)y(t-2)$	113.6097	0.0487
14	$y(t-3)y(t-4)y(t-5)$	81.5845	0.0455
15	$y(t-1)y(t-2)$	2.3277	0.0422
16	$y(t-3)y(t-5)y(t-5)$	-65.5028	0.0348
17	$y(t-5)y(t-5)y(t-5)$	-22.6070	0.0427
18	$y(t-1)y(t-3)y(t-3)$	-80.8411	0.0549
19	$y(t-1)y(t-3)$	0.9579	0.0381
20	$y(t-4)$	-0.0143	0.0300
21	$y(t-2)y(t-4)$	2.0816	0.0330
22	$y(t-2)y(t-5)y(t-5)$	-56.9164	0.0318
23	1	0.0003	0.0269
24	$y(t-5)y(t-5)$	-1.1067	0.0284
25	$y(t-4)y(t-4)y(t-5)$	-41.0855	0.0294
26	$y(t-1)y(t-2)y(t-4)$	-59.0133	0.0299
27	$y(t-3)y(t-3)$	0.8629	0.0225
28	$y(t-2)y(t-3)$	1.7678	0.0224
29	$y(t-1)y(t-3)y(t-4)$	-75.3693	0.0170
30	$y(t-3)y(t-3)y(t-4)$	-41.6680	0.0261
31	$y(t-4)y(t-4)$	-0.6487	0.0144
32	$y(t-1)y(t-1)y(t-4)$	-24.7767	0.0158
33	$y(t-7)$	0.0112	0.0127
34	$y(t-2)y(t-3)y(t-4)$	58.2569	0.0111
35	$y(t-2)y(t-4)y(t-5)$	57.1653	0.0181
36	$y(t-1)$	-0.0169	0.0120
37	$y(t-1)y(t-2)y(t-5)$	55.6623	0.0110
38	$y(t-9)$	-0.0098	0.0087

39	$y(t-8)$	0.0083	0.0063
40	$y(t-2)y(t-2)y(t-5)$	-30.5734	0.0063
41	$y(t-4)y(t-5)$	-0.8589	0.0069
42	$y(t-3)y(t-3)y(t-3)$	13.7276	0.0059
43	$y(t-2)$	-0.0088	0.0043
44	$y(t-3)y(t-4)$	-0.5514	0.0035
45	$y(t-2)y(t-5)$	0.4365	0.0033
46	$y(t-1)y(t-5)$	0.5647	0.0031
47	$y(t-4)y(t-5)y(t-5)$	-14.9617	0.0022
48	$y(t-1)y(t-4)y(t-4)$	17.1400	0.0019
49	$y(t-3)y(t-4)y(t-4)$	19.3294	0.0023
50	$y(t-3)y(t-3)y(t-5)$	19.3956	0.0019
51	$y(t-1)y(t-1)y(t-3)$	-18.0446	0.0027
52	$y(t-2)y(t-2)$	-0.2825	0.0018
53	$y(t-2)y(t-3)y(t-5)$	12.5042	0.0016
54	$y(t-1)y(t-4)$	0.2763	0.0011
55	$y(t-2)y(t-4)y(t-4)$	9.8964	0.0012
56	$y(t-6)$	-0.0038	0.0011
57	$y(t-1)y(t-2)y(t-3)$	-13.5193	8.5656E-04
58	$y(t-1)y(t-1)$	-0.1047	3.9998E-04
59	$y(t-2)y(t-2)y(t-3)$	-7.5895	3.4938E-04
60	$y(t-1)y(t-1)y(t-1)$	1.7408	9.9794E-05
61	$y(t-1)y(t-4)y(t-5)$	-0.4417	8.5290E-07

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the third order non-linear candidate mean model.

5.2.7.1 Progression of AIC for the Estimation Set

As explained in Section 4.5, each term from Table 5.9 is added to the mean model iteratively, starting with the first term. A linear noise model is fitted in addition to the terms in the mean model, and the OSA residuals, $\hat{e}(t)$, are calculated. An ARCH(25) variance model is fitted to $\hat{e}(t)$, and the AIC of the fitted variance model is calculated using equation (4.11). Starting from the second iteration, $AIC_{\%}$ is also calculated using equation (4.18).

Figure 5.19 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the third order non-linear mean model.

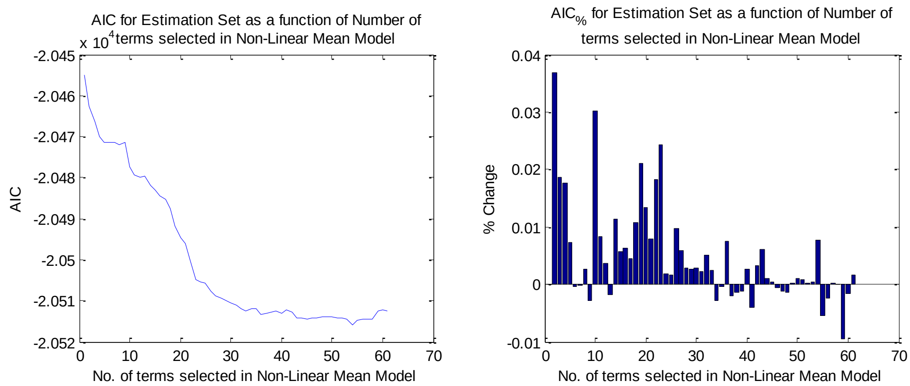


Figure 5.19 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Third Order Non-Linear Mean Model

From Figure 5.19, the AIC drastically decreases to a value of -2.0505E+04 till 23 terms from the top of Table 5.9 are selected to be included in the third order non-linear mean model. Also, $AIC_{\%}$ for the 23rd iteration is the third highest (0.0243%). The addition of further terms to the mean model decreases the value of AIC gradually to a minimum value of -2.0516E+04.

5.2.7.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

The higher order correlation confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ and $V_{(Z^2)'(Z^2)'}$, are calculated for all the non-linear mean models using equations (4.7), (4.8) and (4.9). Figure 5.20 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.21 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.22 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

For the non-linear mean model comprising of all the 61 terms from Table 5.9, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 8.1624E-06. For the non-linear mean model comprising of 23 terms from the top of Table 5.9, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 1.0088E-05 which is very close to the minimum value.

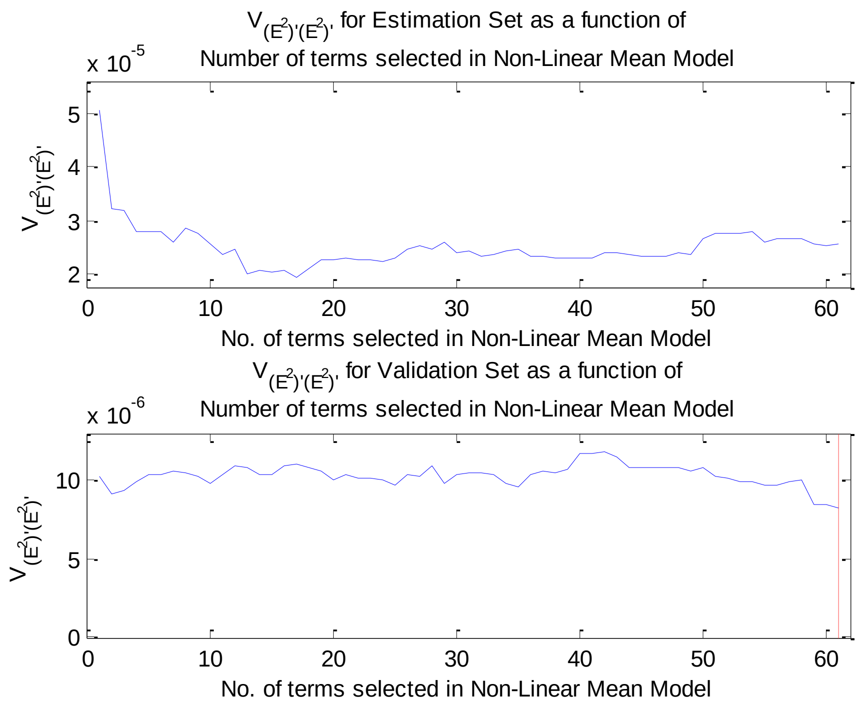


Figure 5.20 $\boldsymbol{V_{(\varepsilon^2)'(\varepsilon^2)'}}$ **for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model**

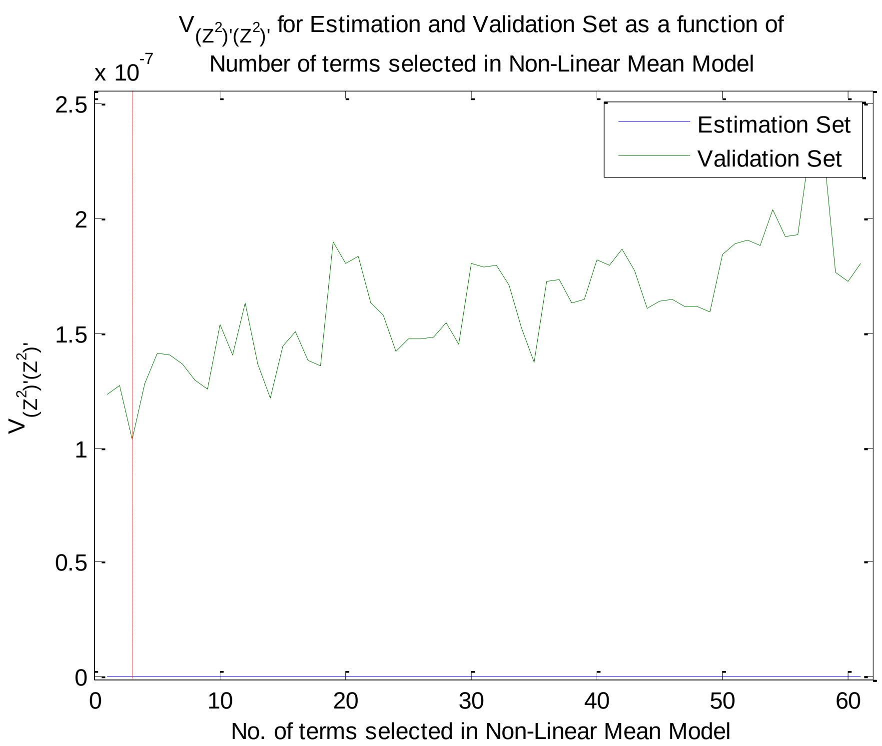


Figure 5.21 $\boldsymbol{V_{(z^2)'(z^2)'}}$ **for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model**

For the non-linear mean model comprising of 3 terms from Table 5.9, $V_{(z^2)'(z^2)'}$ for the Validation Set is the least with a value of 1.0329E-07. In this case, based on the AIC criterion, 23 terms from the top of Table 5.9 are selected to be included in the non-linear mean model. The model yields a value of 1.5785E-07 for $V_{(z^2)'(z^2)'}$ for the Validation Set, which is close to the minimum value. These results suggest that the mean and the variance of the returns have been adequately modelled.

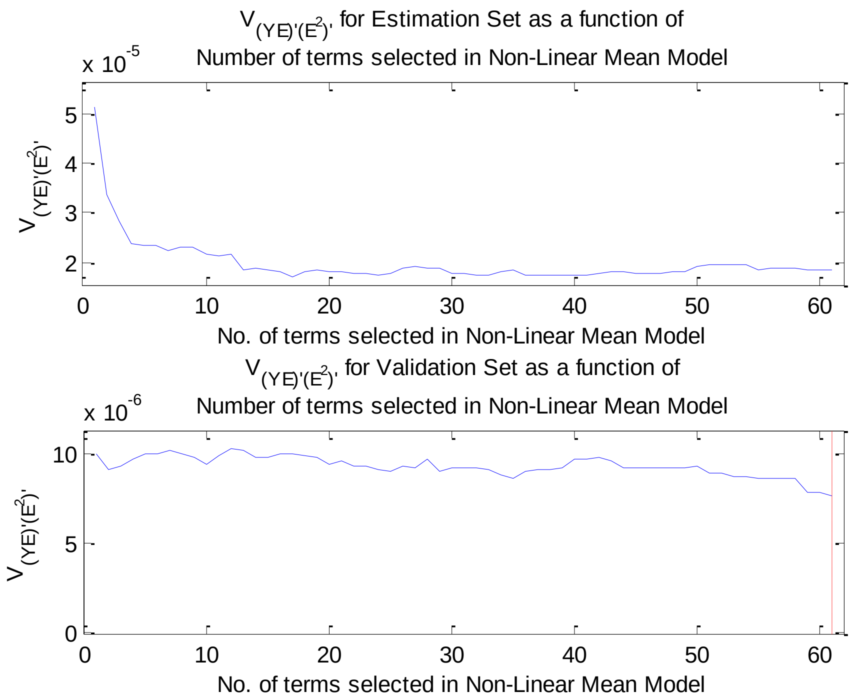


Figure 5.22 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

For the non-linear mean model comprising of all the 61 terms from Table 5.9, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 7.6625E-06. For the non-linear mean model comprising of 23 terms from the top of Table 5.9, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is 9.3438E-06 which is very close to the minimum value.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 5.23.

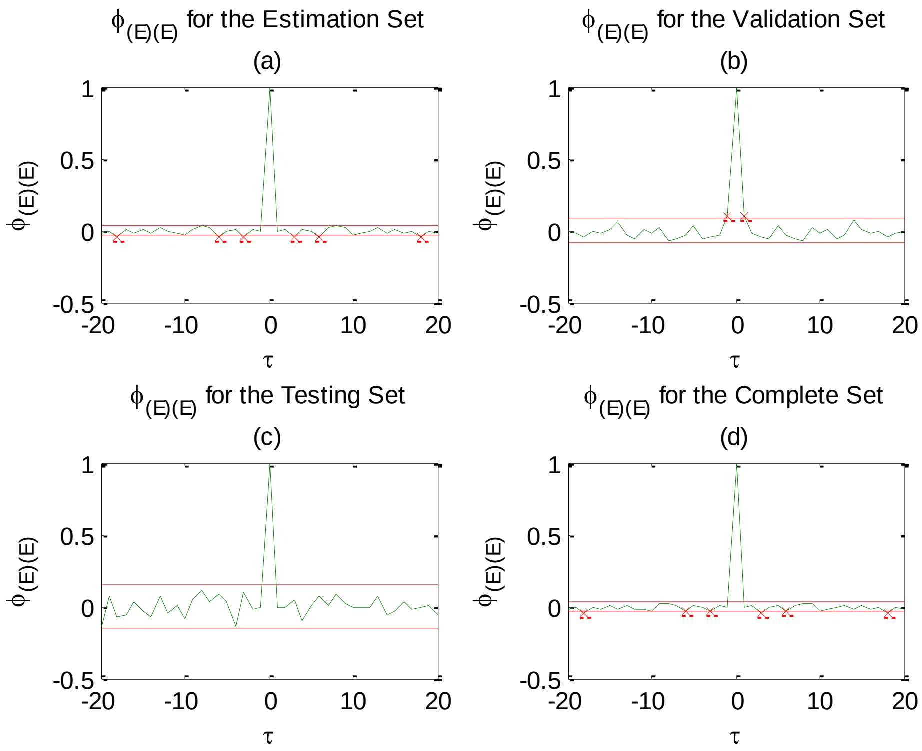


Figure 5.23 Autocorrelation Plots of Residuals for the Selected Third Order Non-Linear Mean Model

In Figure 5.23 (d), note that there exists very small autocorrelation of the residuals at lag 3 implying that the fitted third order non-linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.5 (d)), the residuals obtained after fitting the selected linear mean model (Figure 5.11 (d)) and the residuals obtained after fitting the selected second order non-linear mean model (Figure 5.17 (d)) which indicates that the selected third order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant, linear and second order non-linear mean models.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected third order non-linear mean model are shown in Figure 5.24.

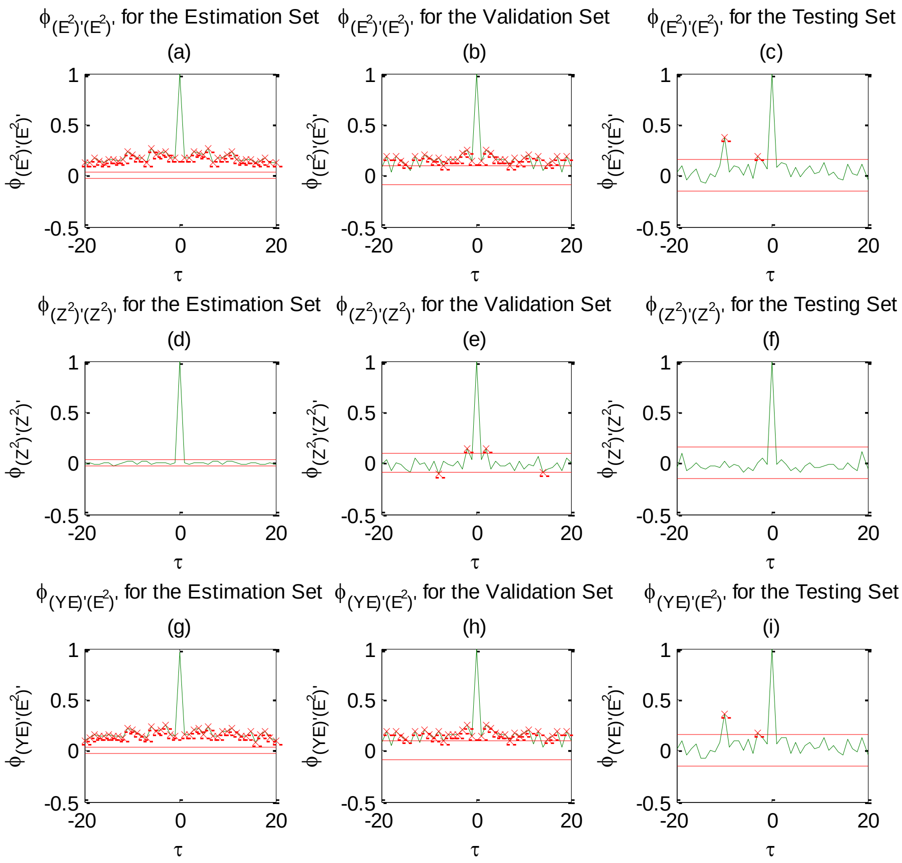


Figure 5.24 Higher Order Correlation Plots for the Selected Third Order Non-Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.24 (d) and (f), there exist no violations of the 95% confidence bands. In Figure 5.24 (e), $\Phi_{(z^2)'(z^2)'}$ lies on or just outside the 95% confidence band for a few lags. The confidence violations are not large, and the ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 0.3188, which is smaller than the critical value of 3.8415 suggesting that the mean and the variance of the returns have been fitted adequately.

The selected terms along with the coefficient estimates and ERR values are listed in Table 5.10. The coefficient estimates of the 5 noise terms are also included.

Table 5.10 Third Order Non-Linear Mean Model fitted to FTSE100

No.	Term	Parameter Estimate	ERR
1	$y(t-4)y(t-4)y(t-4)$	32.6956	0.2849
2	$y(t-1)y(t-3)y(t-5)$	-103.9749	0.2883
3	$y(t-1)y(t-5)y(t-5)$	-54.4378	0.2239
4	$y(t-1)y(t-1)y(t-2)$	-78.8744	0.1364
5	$y(t-5)$	-0.3832	0.0878
6	$y(t-3)$	0.2420	0.0984
7	$y(t-2)y(t-3)y(t-3)$	-62.4940	0.0361
8	$y(t-3)y(t-5)$	1.9704	0.0377
9	$y(t-10)$	0.0054	0.0025
10	$y(t-2)y(t-2)y(t-4)$	130.2594	0.1343
11	$y(t-1)y(t-1)y(t-5)$	79.3634	0.0489
12	$y(t-2)y(t-2)y(t-2)$	53.3504	0.0489
13	$y(t-1)y(t-2)y(t-2)$	109.7360	0.0475
14	$y(t-3)y(t-4)y(t-5)$	97.2760	0.0465
15	$y(t-1)y(t-2)$	2.2638	0.0427
16	$y(t-3)y(t-5)y(t-5)$	-80.1125	0.0571
17	$y(t-5)y(t-5)y(t-5)$	-15.9892	0.0134
18	$y(t-1)y(t-3)y(t-3)$	-45.5746	0.0407
19	$y(t-1)y(t-3)$	2.1186	0.0795
20	$y(t-4)$	-0.1560	0.0188
21	$y(t-2)y(t-4)$	1.2857	0.0228
22	$y(t-2)y(t-5)y(t-5)$	-78.0188	0.0566
23	1	1.6779E-04	0.0260
24	$e(t-1)$	-0.0204	0.0256
25	$e(t-2)$	-0.0132	0.0116
26	$e(t-3)$	-0.2651	0.1141
27	$e(t-4)$	0.1379	0.0332
28	$e(t-5)$	0.3501	0.1553

So far, all the mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 4 cases.

5.2.8 Comparison of ARCH Test Statistics and Non-Linear Correlation Statistics of All Mean Models

Note that higher order correlation tests and the ARCH Test statistics of $\hat{z}(t)$ do not work for the purpose of comparison of the performance of the different mean models. This is because the mean model can be underfitted, and the variance model can be overfitted, thereby yielding a $\hat{z}(t)$ series that passes all the mentioned model validation tests. Hence, to compare the fitted mean models, the ARCH Test statistics and the higher order correlation statistics of $\hat{e}(t)$ need to be used.

The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the selected mean models are listed in Table 5.11. The minimum values are shaded in blue.

Table 5.11 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Mean Models

Type of Mean Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Constant	14.0218	6.6068E-05
Linear	13.5820	6.2440E-05
Second Order Non-Linear	13.2098	6.2705E-05
Third Order Non-Linear	11.5644	1.8281E-05

The third order non-linear mean model (terms and parameter estimates listed in Table 5.10) has the lowest ARCH Test Statistic of $\hat{e}(t)$ for the Validation Set. This implies that the mean of the returns have been modelled better by the third order non-linear mean model than the second order non-linear, linear or constant mean model, and hence is the appropriate choice.

5.2.9 Comparison of Selected Mean Model to Vanilla GARCH Model

A vanilla GARCH model (constant mean model with a GARCH(1,1) variance model) is fitted to the given FTSE100 return series. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the fitted vanilla GARCH model are listed and compared to the selected third order non-linear mean model in Table 5.12. The minimum values are shaded in blue.

Table 5.12 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Vanilla GARCH and Selected Mean Model fitted to FTSE100

Type of Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Vanilla GARCH	13.9934	6.5966E-05
Selected Third Order Non-Linear Mean Model with ARCH(25) Variance Model	11.5644	1.8281E-05

The selected third order non-linear mean model with an ARCH(25) variance model performs much better than a vanilla GARCH model. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the selected model are lower than those of the vanilla GARCH model.

Figure 5.25 and Figure 5.26 show the one-step-ahead (OSA) return estimates generated using the vanilla GARCH model and the selected third order non-linear mean model.

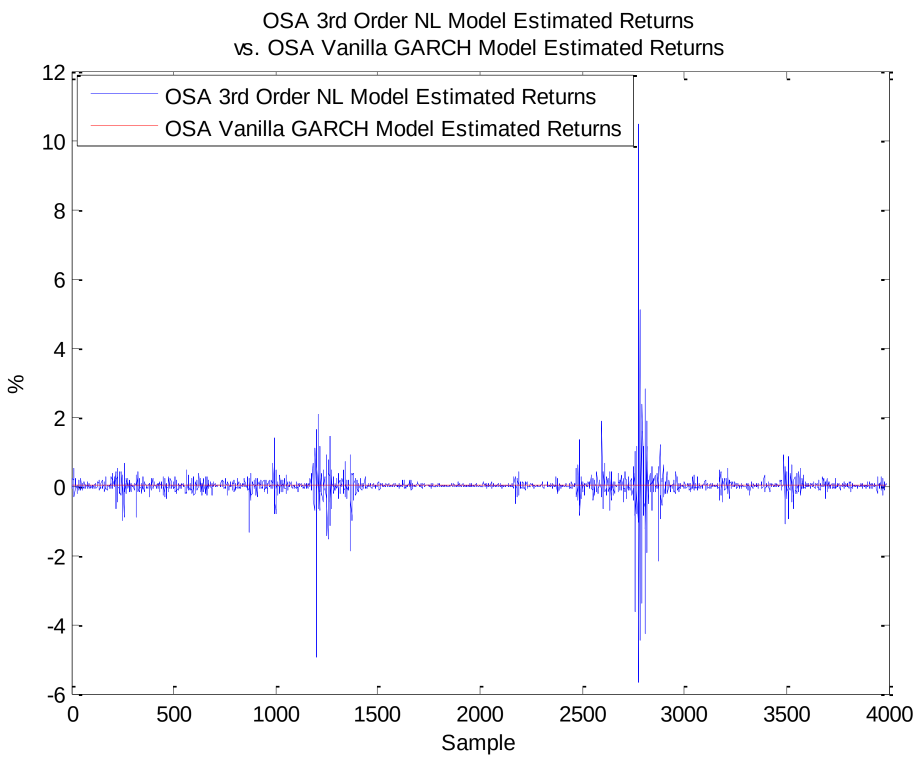


Figure 5.25 OSA Return Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for the FTSE100

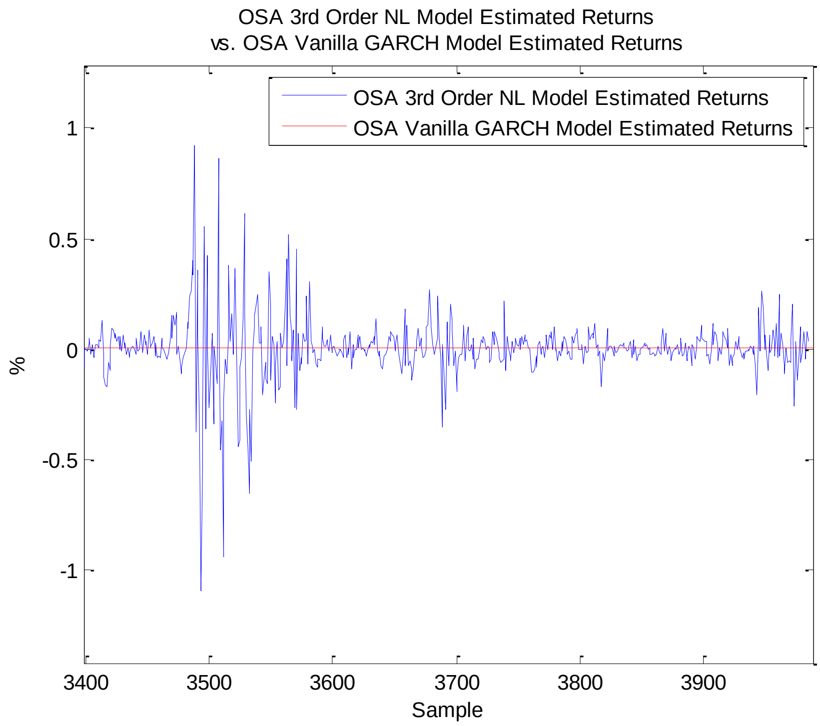


Figure 5.26 OSA Return Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for the FTSE100 (Samples 3400 to 4000)

From Figure 5.25 and Figure 5.26, it can be noted that using the third order non-linear mean model certainly captures the predictable elements of the mean of the returns, rather than just using a constant mean model and passing off the predictable elements to be included in the residuals. The standard deviation of the various return series are listed in Table 5.13.

Table 5.13 Standard Deviation of various return series for FTSE100

Series	Standard Deviation
OSA Return Estimate of Constant Mean Model	0
OSA Return Estimate of Third Order NL Mean Model	0.0037
True Return Series	0.0127

Comparing the magnitudes of the Validation and Testing Sets of the OSA return estimates generated using the third order non-linear mean model to those of the true return series, 359 samples of 700 samples have the same magnitude. Hence, the magnitude of the returns is predicted right 51.2857% of the time.

Figure 5.27 and Figure 5.28 show the variance estimates generated using the vanilla GARCH model and the selected third order non-linear mean model.

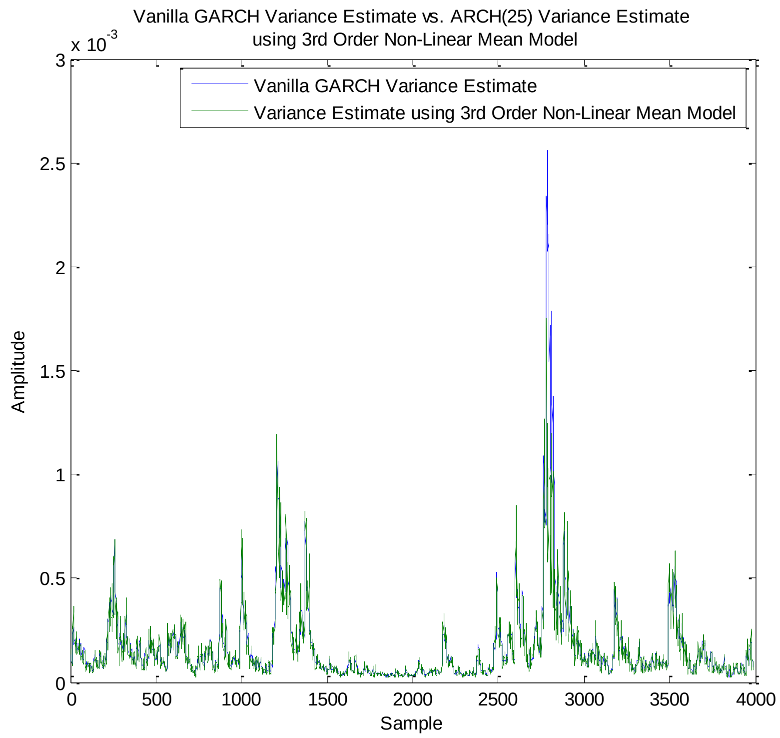


Figure 5.27 Variance Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for the FTSE100

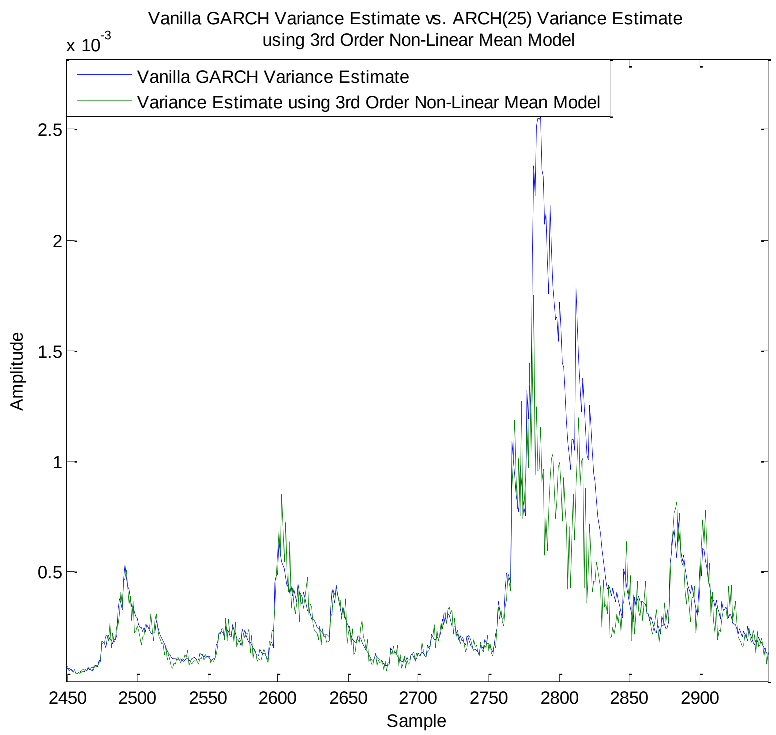


Figure 5.28 Variance Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for the FTSE100 (Samples 2450 to 2950)

Note that the variance estimates are similar during periods of low volatility, but different during periods of high volatility. This suggests that a constant mean model does not capture all the predictable elements in the mean of the returns and adds them to the residuals, thereby increasing the magnitude of the variance estimate (obtained using a constant mean model) at the peaks.

To summarise, constant, linear, second order non-linear and third order non-linear candidate mean models were fitted to the given returns of FTSE100. After carrying out term selection, the ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the models were compared to evaluate how well each model described the mean of the given return series. As in the case of the simulated example, all the fitted mean models passed standard model validation tests, but the third order non-linear mean model was found to have the lowest values for the ARCH Test statistic of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set, implying that the mean of the returns have been modelled best by the third order non-linear mean model.

5.3 Application to NASDAQ

The framework introduced in this chapter is used on the returns of the NASDAQ index. The procedure remains the same as in the case of the returns of the FTSE100 index, and a summary of the results, rather than a detailed description of the approach, will be given.

5.3.1 The Data

4001 samples of the price of the NASDAQ index, dating from 23rd September 1997 to 29th July 2013 are considered. The data is obtained from Yahoo! Finance (2013b). The price series is converted to a return series comprising of 4000 samples via continuous compounding (see equation (2.2)). The price and returns of the NASDAQ index are plotted in Figure 5.29.

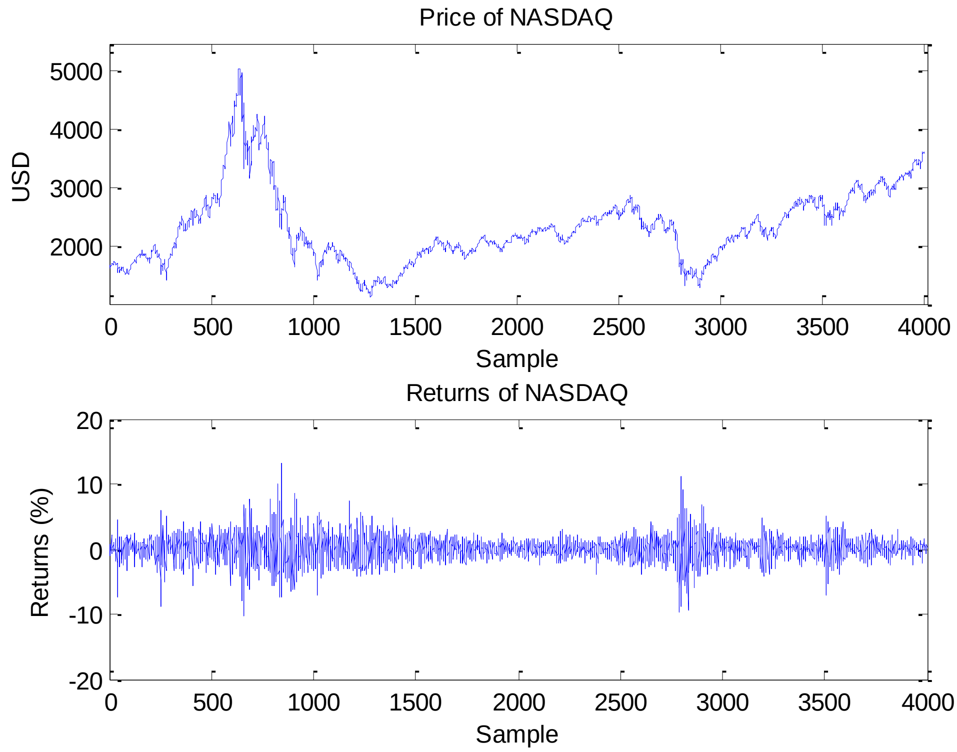


Figure 5.29 Price and Returns of NASDAQ

The number of samples in the Estimation, Validation and Testing Sets is the same as in the case of the simulated data example, and are listed in Table 4.2.

5.3.2 Autocorrelation Plots of Returns and Squared Returns

5.3.2.1 Linear Autocorrelation

The linear sample autocorrelation of the returns and the squared returns for 20 lags are shown in Figure 5.30.

The sample autocorrelation of the returns for lags 2, 12 and 13 lie outside the 95% confidence bands, indicating the possibility of the presence of the terms, $y(t-2), y(t-12)$ and $y(t-13)$, in the mean model.

The sample autocorrelation of the squared returns are outside the 95% confidence bands for all the lags, indicating the need to fit a variance model to the returns, in addition to a mean model.

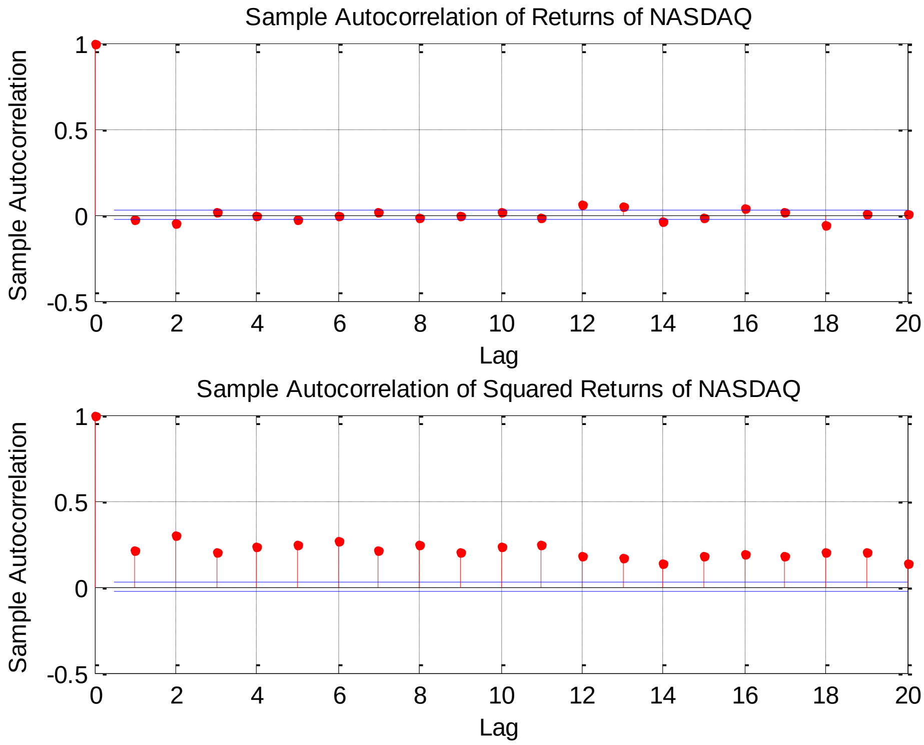


Figure 5.30 Sample Autocorrelation Plots of Returns and Squared Returns of NASDAQ

5.3.2.2 Higher Order Autocorrelation

The higher order autocorrelation of the returns and the squared returns for the Estimation, Validation, Testing and Complete Data Set are also calculated and plotted in Figure 5.31 and Figure 5.32 respectively.

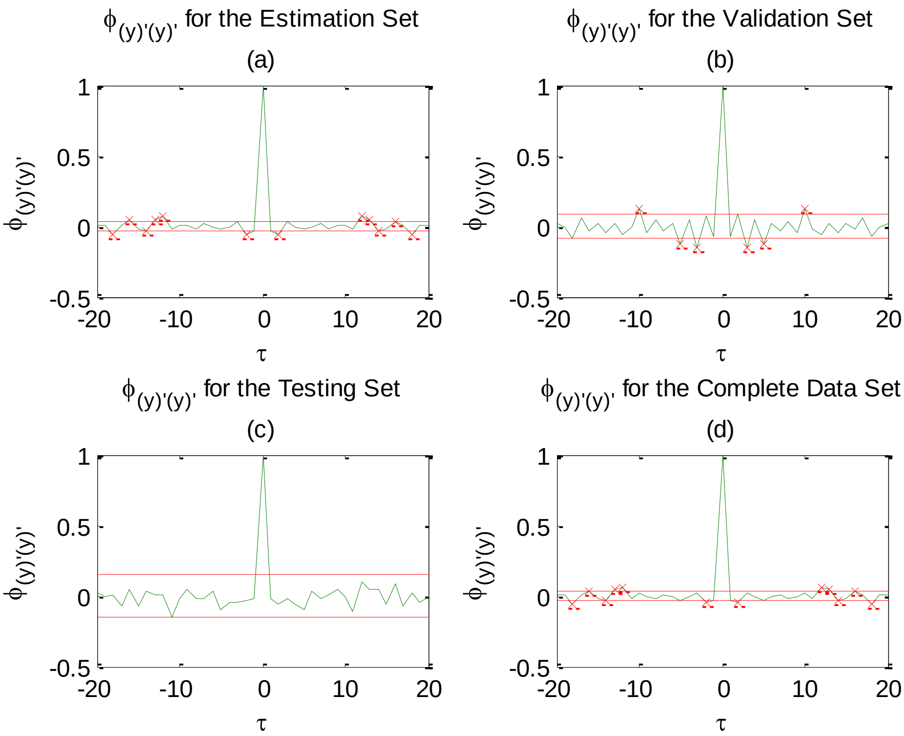


Figure 5.31 $\Phi_{(y)'(y)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Data Set

Figure 5.31 (d) shows that $\Phi_{(y)'(y)'}$ for lags 2, 12 and 13 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags.

The higher order correlation violation statistic, $V_{(y)'(y)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set that enumerate the plots shown in Figure 5.31 are calculated and listed in Table 5.14.

Table 5.14 $V_{(y)'(y)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y)'(y)'}$ (Estimation Set)	$V_{(y)'(y)'}$ (Validation Set)	$V_{(y)'(y)'}$ (Testing Set)	$V_{(y)'(y)'}$ (Complete Data Set)
2.2521E-08	3.0694E-07	0	1.5368E-08

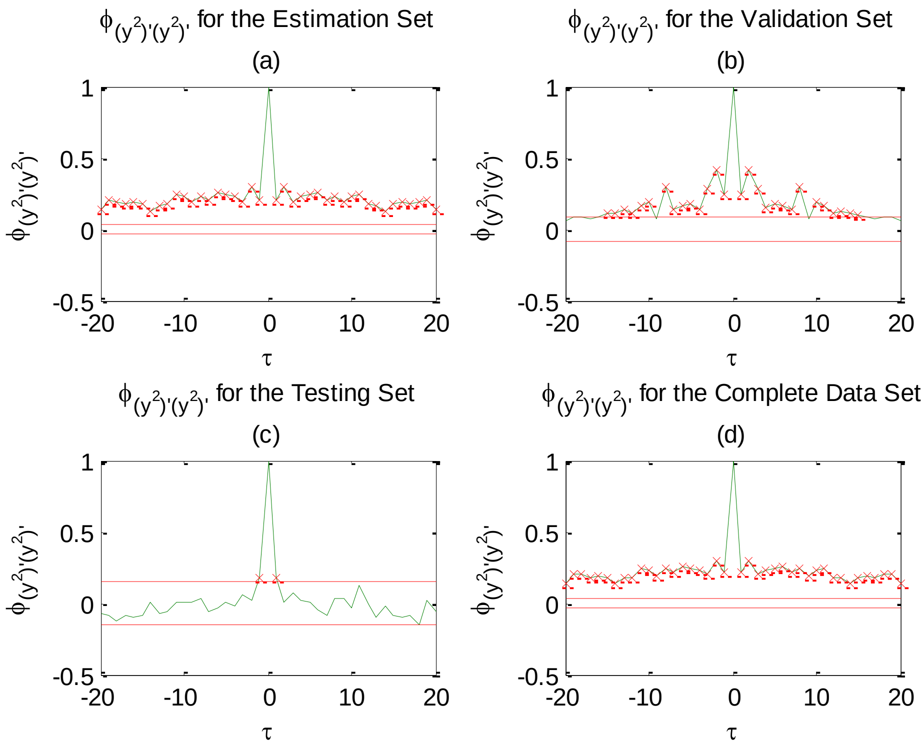


Figure 5.32 $\Phi_{(y^2)'(y^2)'}$ for (a) Estimation, (b) Validation, (c) Testing, and (d) Complete Set

Figure 5.32 (d) shows that $\Phi_{(y^2)'(y^2)'}$ is significantly outside the 95% confidence bands for all the lags. This indicates the need to fit a variance model to the returns, in addition to fitting a mean model. The non-linear correlation violation statistic, $V_{(y^2)'(y^2)'}$, for the Estimation Set, Validation Set, Testing Set, and the Complete Set that enumerate the plots shown in Figure 5.32 are calculated and listed in Table 5.15.

Table 5.15 $V_{(y^2)'(y^2)'}$ for Estimation, Validation, Testing, and Complete Data Sets

$V_{(y^2)'(y^2)'}$ (Estimation Set)	$V_{(y^2)'(y^2)'}$ (Validation Set)	$V_{(y^2)'(y^2)'}$ (Testing Set)	$V_{(y^2)'(y^2)'}$ (Complete Data Set)
3.8219E-05	4.3321E-05	5.0229E-08	4.3612E-05

5.3.3 Candidate Mean Models

Four different candidate mean models are fitted to the returns of the NASDAQ index.

5.3.3.1 Constant Mean Model

The first is a constant mean model, $y(t) = a_0$, where a_0 is to be estimated.

5.3.3.2 Linear Candidate Mean Model

The second is a linear candidate mean model. This is an AR(13) model with a linear noise model consisting of 5 lagged noise terms

$$y(t) = a_0 + \sum_{i=1}^{13} a_i y(t-i) + \sum_{j=1}^{5} b_j e(t-j) \tag{5.4}$$

where a_0, a_i, and b_j are coefficients to be estimated.

From Figure 5.31 (d), $\Phi_{(y)'(y)'}$ for lags 2, 12 and 13 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags. Hence, n_l is selected to be 13.

The maximum lag of the error terms to be included in the linear mean model is selected to be $n_e = 5$.

5.3.3.3 Second Order Non-Linear Candidate Mean Model

The third candidate mean model is a second-order non-linear mean model. This is

$$\begin{aligned} y(t) &= a_0 + \sum_{i=1}^{13} a_i y(t-i) \\ &\quad + \sum_{j=1,k=1,i=11}^{j=5,k=5,i=21} a_i y(t-j) y(t-k) + \sum_{l=1}^{5} b_l e(t-l) \end{aligned} \tag{5.5}$$

where a_0, a_i, and b_l are coefficients to be estimated.

From Figure 5.31 (d), $\Phi_{(y)'(y)'}$ for lags 2, 12 and 13 are just outside the 95% confidence bands, indicating the possibility of the presence of linear and non-linear terms with these lags. Hence, n_l is selected to be 13.

The maximum lag of the second order non-linear terms to be included in the non-linear candidate mean model is selected to be 5. A linear noise model is also fitted to the candidate mean model once term selection has been carried out. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 5$.

5.3.3.4 Third Order Non-Linear Candidate Mean Model

The final candidate model is a third-order non-linear mean model. This is

$$y(t) = a_0 + \sum_{i=1}^{13} a_i y(t-i) + \sum_{j=1,k=1,i=11}^{j=5,k=5,i=21} a_i y(t-j)y(t-k)$$

$$+ \sum_{j=1,k=1,l=1,i=22}^{j=5,k=5,l=5,i=40} a_i y(t-j)y(t-k)y(t-l) + \sum_{l=1}^{5} b_l e(t-l) \qquad (5.6)$$

where a_0, a_i, and b_l are coefficients to be estimated. n_l is selected to be 13. The maximum lag of the second order non-linear terms to be included in the non-linear candidate mean model is selected to be 5. The maximum lag of the error terms to be included in the non-linear mean model is selected to be $n_e = 5$.

5.3.4 Estimation of Constant Mean Model

The values of the hyper-parameters used when fitting a constant mean model are the same as in the case of the FTSE100 and are given in Table 5.3.

The constant term fitted is 3.1437E-04, and has an ERR of 0.0602. To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the constant mean model are plotted in Figure 5.34.

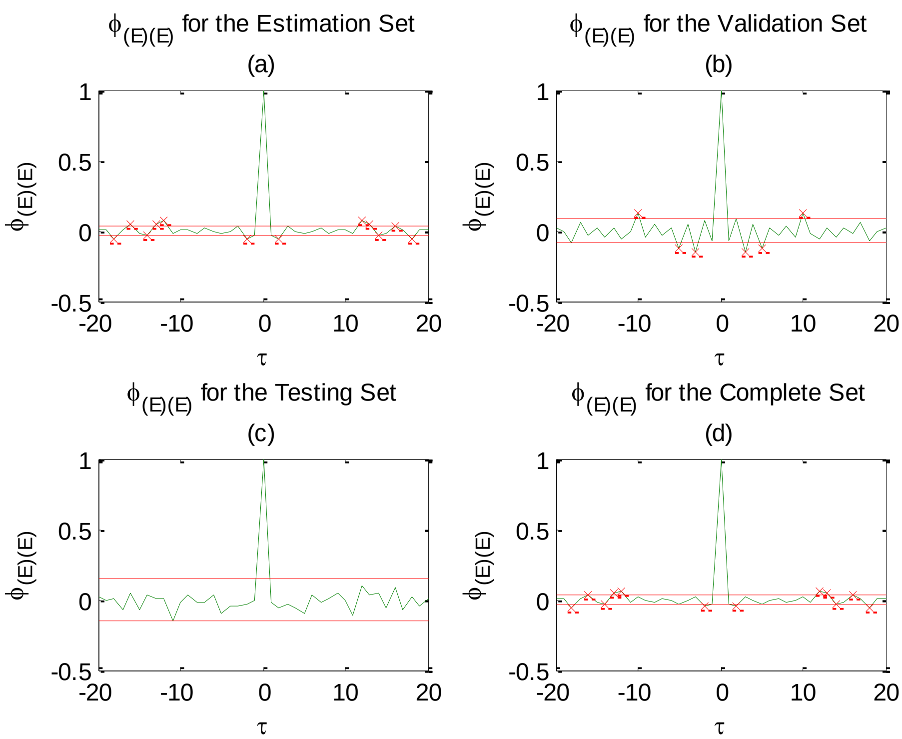


Figure 5.33 Autocorrelation Plots of Residuals for the Selected Constant Mean Model

In Figure 5.34 (d), note that there exists slight autocorrelation of the residuals at lag 2 implying that the fitted constant mean model may be inadequate.

To validate the fitted mean and variance models, the higher order correlation of the squared residuals and squared standardised residuals obtained after fitting the selected constant mean model are plotted in Figure 5.34.

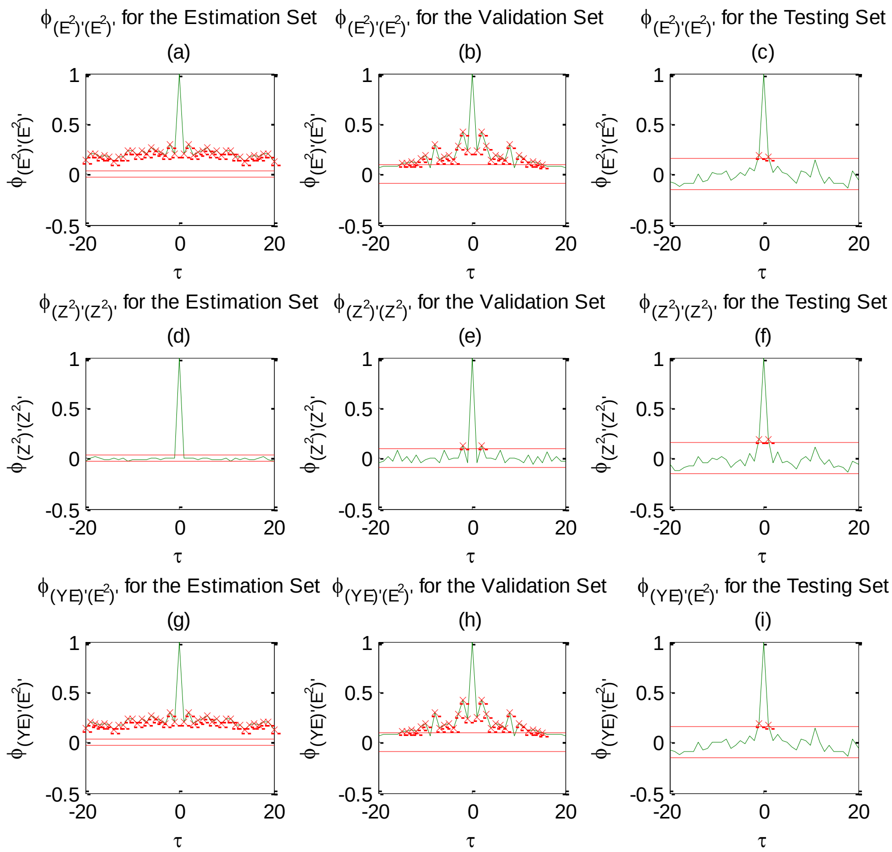


Figure 5.34 Higher Order Correlation Plots for the Selected Constant Mean Model

From Figure 5.34 (d) and (e), the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained from fitting a constant mean model, do not indicate any higher order autocorrelation. In Figure 5.34 (f), $\Phi_{(z^2)'(z^2)'}$ lies just outside the 95% confidence band for lag = 1. The ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 0.9775, which is lesser than the critical value of 3.8415. Hence, the mean and the variance of the returns can be considered to be adequately fitted.

5.3.4.1 $\boldsymbol{V}_{(\varepsilon^2)'(\varepsilon^2)'}, \boldsymbol{V}_{(Z^2)'(Z^2)'}$ and $\boldsymbol{V}_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation, Validation and Testing Sets

The higher order correlation plots in Figure 5.34 are quantified by calculating the confidence violation statistics using equations (4.7), (4.8) and (4.9). Table 5.16 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}, V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation, Validation and Testing Sets for the constant mean model.

Table 5.16 $\boldsymbol{V}_{(\varepsilon^2)'(\varepsilon^2)'}, \boldsymbol{V}_{(Z^2)'(Z^2)'}$ and $\boldsymbol{V}_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation, Validation and Testing Sets in Constant Mean Model

	Estimation Set	Validation Set	Testing Set
$\boldsymbol{V}_{(\varepsilon^2)'(\varepsilon^2)'}$	3.8999E-05	4.3892E-05	8.4453E-08
$\boldsymbol{V}_{(Z^2)'(Z^2)'}$	0	5.3441E-08	1.3511E-07
$\boldsymbol{V}_{(Y\varepsilon)'(\varepsilon^2)'}$	3.8807E-05	4.3740E-05	8.5185E-08
ARCH Test Statistic of $\hat{\boldsymbol{z}}(\boldsymbol{t})$	0.2415	0.9775	5.0025

From Figure 5.34 (a), (b), (c), (g), (h) and (i), the squared estimated residuals, $\hat{e}^2(t)$, obtained from fitting a constant mean model, indicate higher order correlation and autocorrelation. The corresponding confidence violation statistics, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$, for the Estimation and Validation Sets in Table 5.16 are of the order of 1E-05 which is considerably high in this context. These results suggest that a variance model is required to be fitted to the given return series.

The magnitude of the value of $V_{(Z^2)'(Z^2)'}$ for the Validation Set is extremely small, and close to zero, which is acceptable. The ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is lesser than the critical value of 3.8415. Hence, the mean and the variance of the returns can be considered to be adequately fitted.

5.3.5 Estimation of Linear Mean Model using WOFR

The next type of mean model to be fitted to the returns of the NASDAQ index is the linear candidate mean model listed in Section 5.3.3.2. The results of the WOFR analysis are given in Table 5.17.

Table 5.17 Terms in Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-12)$	0.0502	0.2427
2	$y(t-13)$	0.0307	0.0939
3	$y(t-10)$	0.0226	0.0537
4	$y(t-2)$	-0.0240	0.0427
5	$y(t-6)$	-0.0194	0.0281
6	1	0.0002	0.0289
7	$y(t-1)$	-0.0152	0.0226
8	$y(t-9)$	-0.0123	0.0148
9	$y(t-3)$	0.0100	0.0093
10	$y(t-8)$	-0.0098	0.0086
11	$y(t-11)$	0.0070	0.0045
12	$y(t-5)$	-0.0062	0.0035
13	$y(t-4)$	-0.0043	0.0015
14	$y(t-7)$	0.0039	0.0013

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are now carried out on the re-ordered terms of the linear candidate mean model.

5.3.5.1 Progression of AIC for the Estimation Set

Figure 5.35 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the linear mean model.

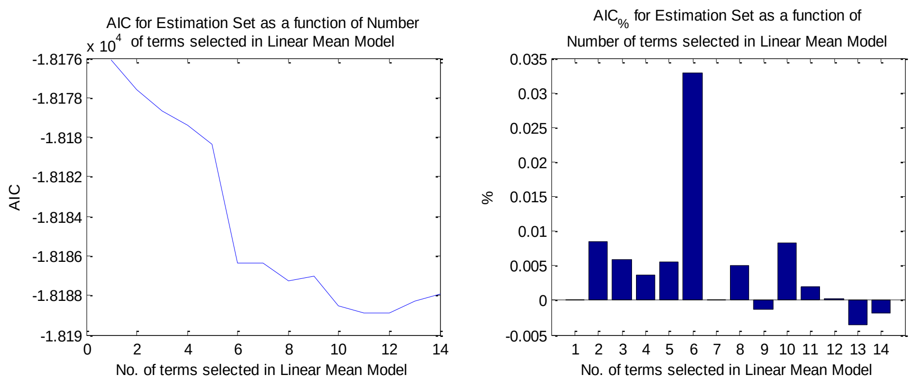


Figure 5.35 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Linear Mean Model

From Figure 5.35, $AIC_{\%}$ is the highest (0.0329%) when 6 terms from the top of Table 5.17 are included. Also, the AIC decreases rapidly to -1.8186E+04 when the first 6 terms are added, and then gradually decreases to the minimum value of 1.8189E+04 as more terms are added. Hence, 6 terms from the top of Table 5.17 are selected to be included in the linear mean model.

5.3.5.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

Figure 5.36 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.37 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.38 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

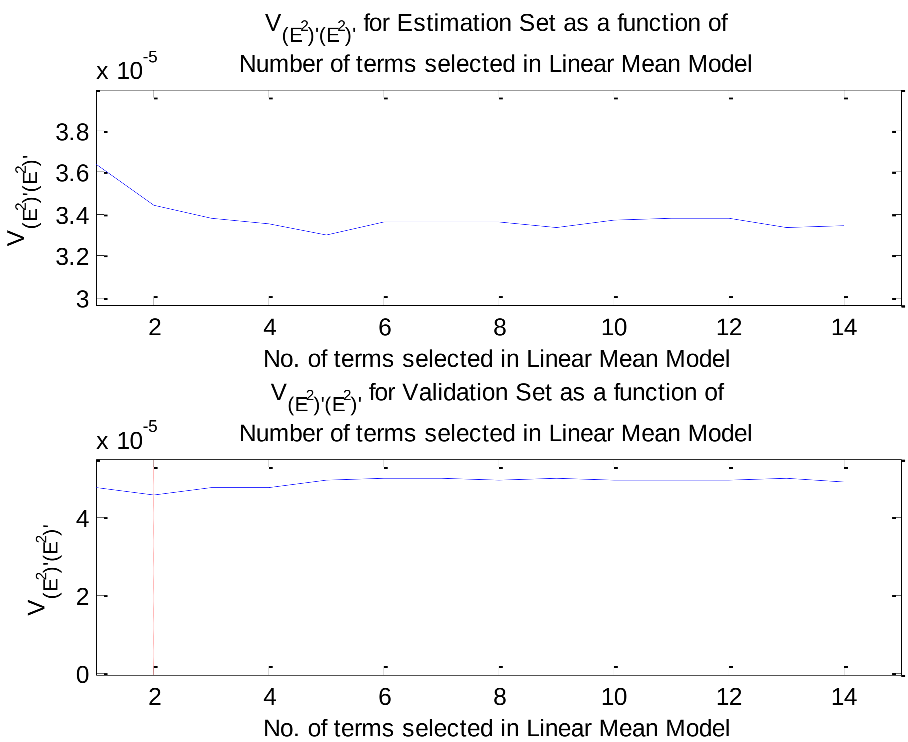


Figure 5.36 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of 2 terms from the top of Table 5.17, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 4.6048E-05. For the linear mean model comprising of 6 terms from the top of Table 5.17, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 5.0195E-05 which is very close to the minimum value.

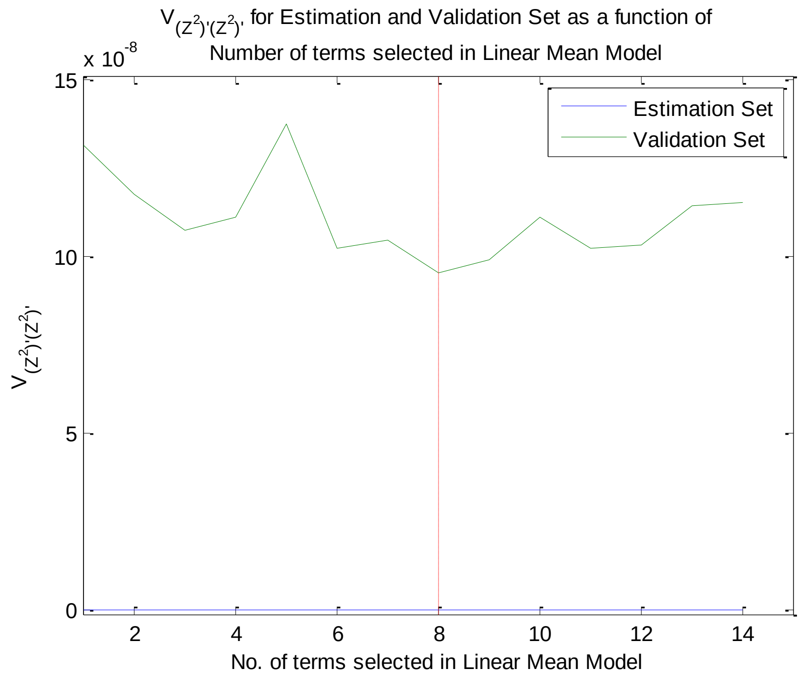


Figure 5.37 $V_{(Z^2)'(Z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

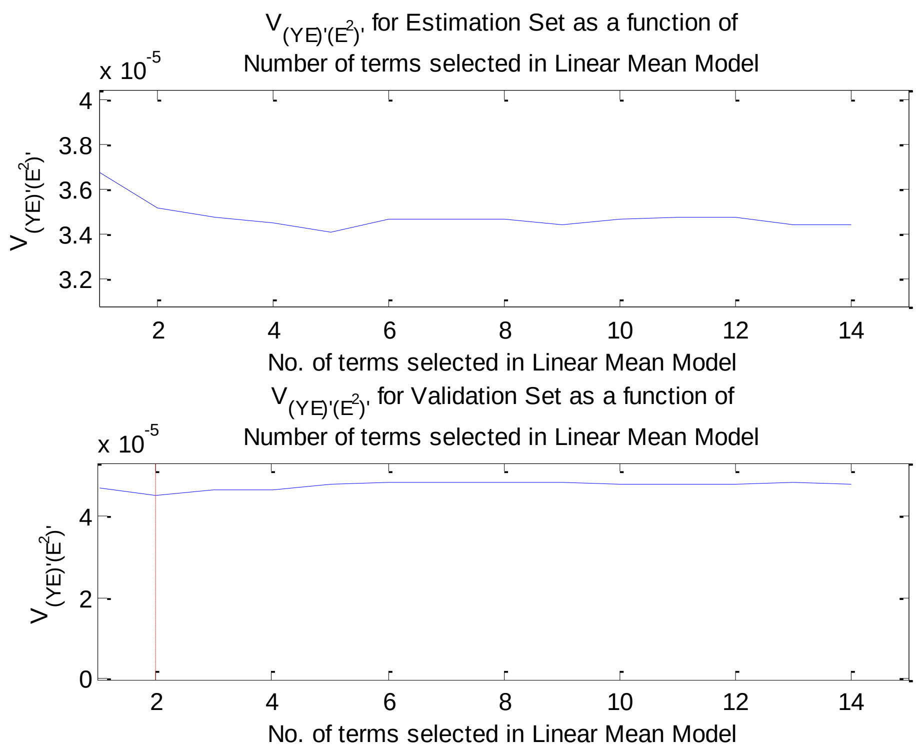


Figure 5.38 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Linear Mean Model

For the linear mean model comprising of 8 terms from Table 5.17, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 9.5345E-08. In this case, 6 terms from the top of Table 5.17 are selected to be included in the linear mean model which yields a value of 1.0244E-07 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude and very close to the minimum value of 9.5345E-08. These results suggest that the mean and the variance of the returns have been adequately modelled.

For the linear mean model comprising of 2 terms from the top of Table 5.17, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 4.5356E-05. For the linear mean model comprising of 6 terms from the top of Table 5.17, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is 4.8358E-05 which is very close to the minimum value.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected linear mean model are plotted in Figure 5.39.

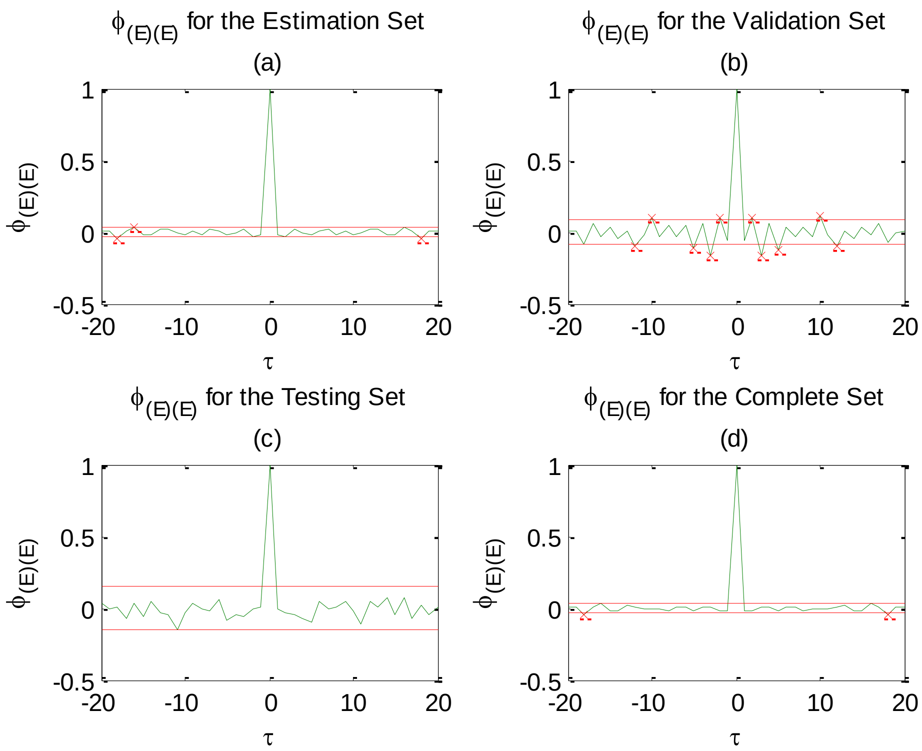


Figure 5.39 Autocorrelation Plots of Residuals for the Selected Linear Mean Model

In Figure 5.39 (d), note that there exists no autocorrelation of the residuals implying that the fitted linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.33 (d)) which indicates that the selected linear mean model captures the predictable elements of the mean of the returns much better than the selected constant mean model.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected linear mean model are shown in Figure 5.40.

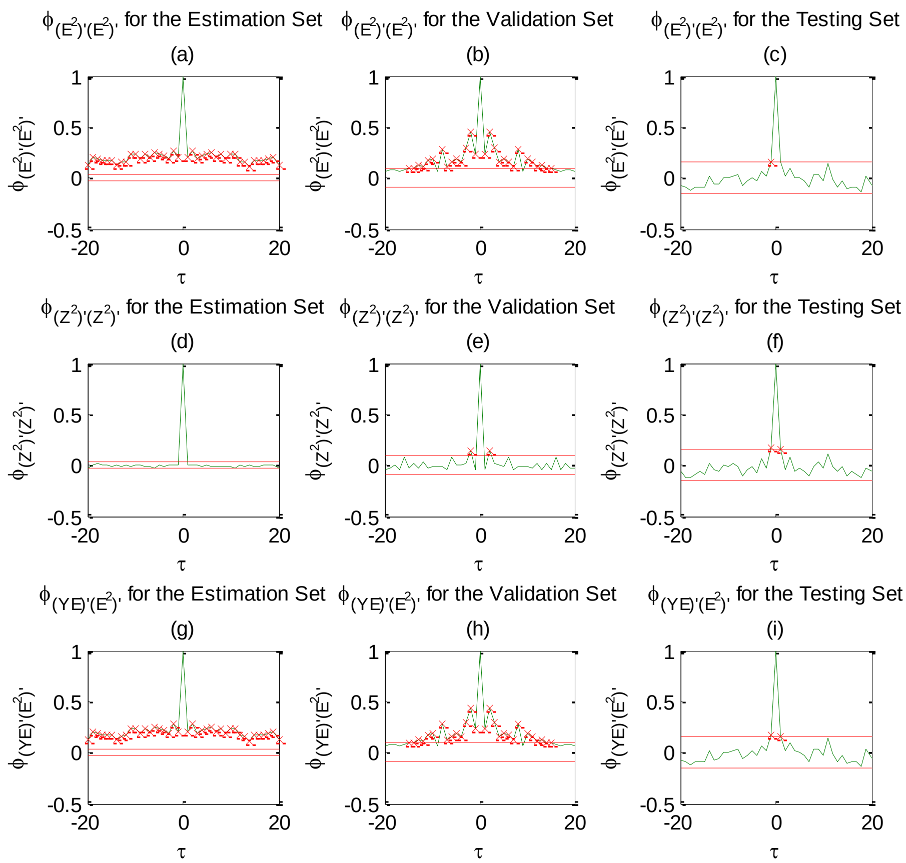


Figure 5.40 Higher Order Correlation Plots for the Selected Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.40 (d), there exist no violations of the 95% confidence bands. In Figure 5.40 (e), $\Phi_{(z^2)'(z^2)'}$ lies just outside the 95% confidence band for lag = 2. In Figure 5.40 (f), $\Phi_{(z^2)'(z^2)'}$ lies just outside the 95% confidence band for lag = 1. The ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 1.0300, which is smaller than the critical value of 3.8415 suggesting that the mean and the variance of the returns have been fitted adequately.

The terms selected in the linear mean model along with the coefficient estimates and ERR values are listed in Table 5.18. The coefficient estimates of the 5 noise terms are also included.

Table 5.18 Linear Mean Model fitted to NASDAQ

No.	Term	Parameter Estimate	ERR
1	$y(t-12)$	0.0539	0.2467
2	$y(t-13)$	0.0298	0.0953
3	$y(t-10)$	0.0243	0.0563
4	$y(t-2)$	-0.1941	0.0399
5	$y(t-6)$	-0.0204	0.0293
6	1	0.0003	0.0322
7	$e(t-1)$	-0.0153	0.0223
8	$e(t-2)$	0.1720	0.0114
9	$e(t-3)$	0.0080	0.0052
10	$e(t-4)$	-0.0097	0.0084
11	$e(t-5)$	-0.0039	0.0013

So far, the constant and the linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in both cases.

5.3.6 Estimation of Second Order Non-Linear Mean Model using WOFR

The next type of mean model to be fitted to the returns of the NASDAQ index is the second order non-linear candidate mean model listed in Section 5.3.3.3. The results of the WOFR analysis are given in Table 5.19.

Table 5.19 Terms in Second Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-3)$	3.0414	0.2707
2	$y(t-12)$	0.0509	0.2503
3	$y(t-3)y(t-4)$	-2.0058	0.1390
4	$y(t-2)y(t-4)$	2.0189	0.1431
5	$y(t-13)$	0.0357	0.1055
6	$y(t-1)y(t-1)$	0.8178	0.1059

7	$y(t-2)y(t-2)$	-1.2547	0.0603
8	$y(t-3)y(t-3)$	0.8572	0.0634
9	$y(t-10)$	0.0228	0.0494
10	$y(t-1)y(t-5)$	1.1415	0.0436
11	$y(t-6)$	-0.0179	0.0277
12	$y(t-2)$	-0.0176	0.0223
13	$y(t-5)y(t-5)$	0.6013	0.0241
14	$y(t-1)y(t-2)$	0.8240	0.0169
15	$y(t-9)$	-0.0137	0.0155
16	$y(t-2)y(t-5)$	0.7902	0.0118
17	$y(t-4)y(t-5)$	0.7158	0.0130
18	$y(t-3)y(t-5)$	0.5906	0.0103
19	$y(t-11)$	0.0089	0.0082
20	$y(t-3)$	0.0078	0.0062
21	$y(t-2)y(t-3)$	0.4789	0.0053
22	$y(t-8)$	-0.0080	0.0049
23	$y(t-5)$	-0.0063	0.0027
24	$y(t-4)y(t-4)$	-0.2193	0.0020
25	1	0.0001	0.0025
26	$y(t-7)$	0.0037	0.0012

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the second order non-linear candidate mean model.

5.3.6.1 Progression of AIC for the Estimation Set

Figure 5.41 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the second order non-linear mean model.

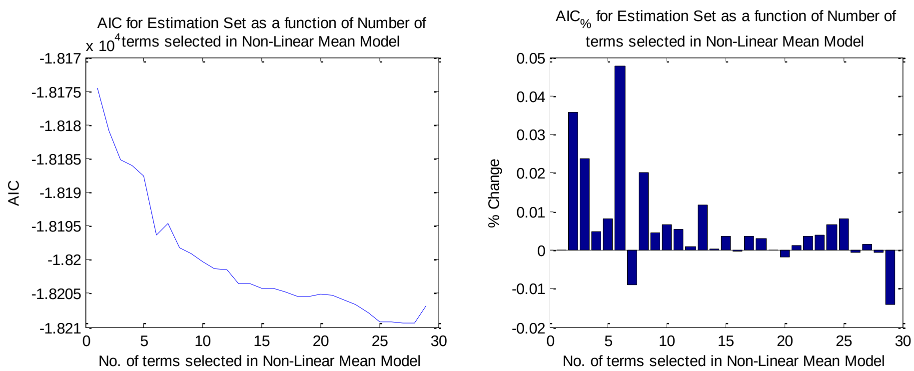


Figure 5.41 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Second Order Non-Linear Mean Model

From Figure 5.41, the AIC drastically decreases to a value of -1.8196E+04 till 6 terms from the top of Table 5.19 are selected to be included in the second order non-linear mean model. Also, $AIC_{\%}$ for the 6th iteration is the highest (0.0477%). The addition of further terms to the mean model decreases the value of AIC gradually to a minimum value of -1.8209E+04.

5.3.6.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

Figure 5.42 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.43 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.44 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

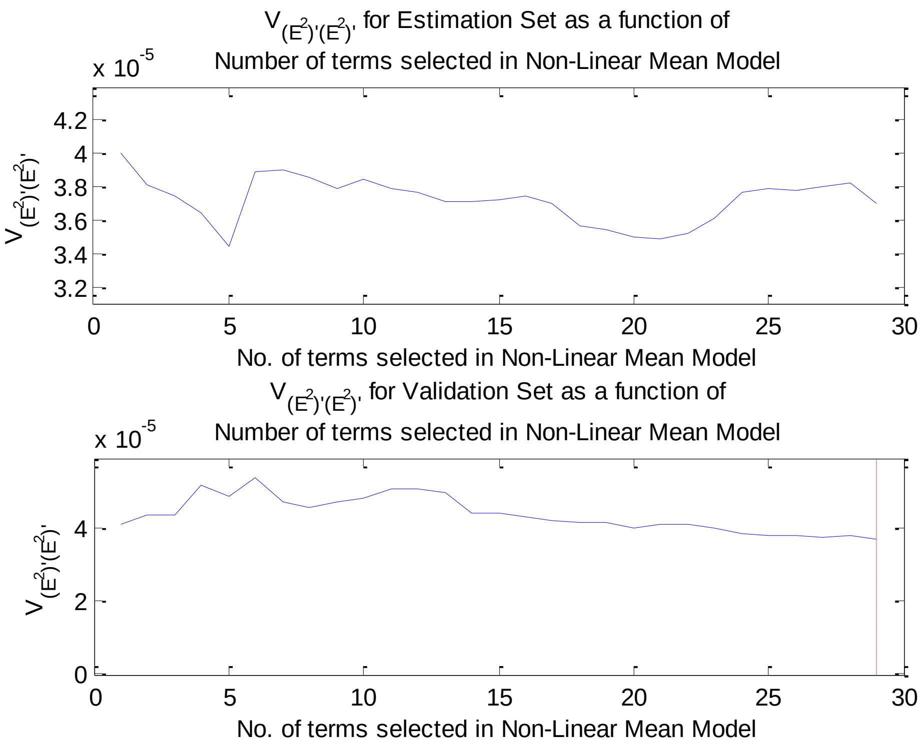


Figure 5.42 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

For the non-linear mean model comprising of all the 29 terms from Table 5.19, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 3.6769E-05. For the non-linear mean model comprising of 6 terms from the top of Table 5.19, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 5.3708E-05 and is the highest.

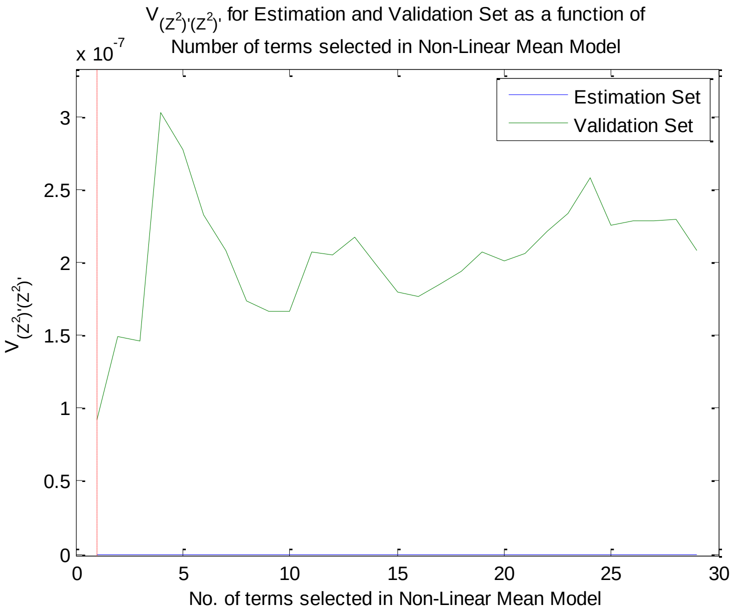


Figure 5.43 $V_{(z^2)'(z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

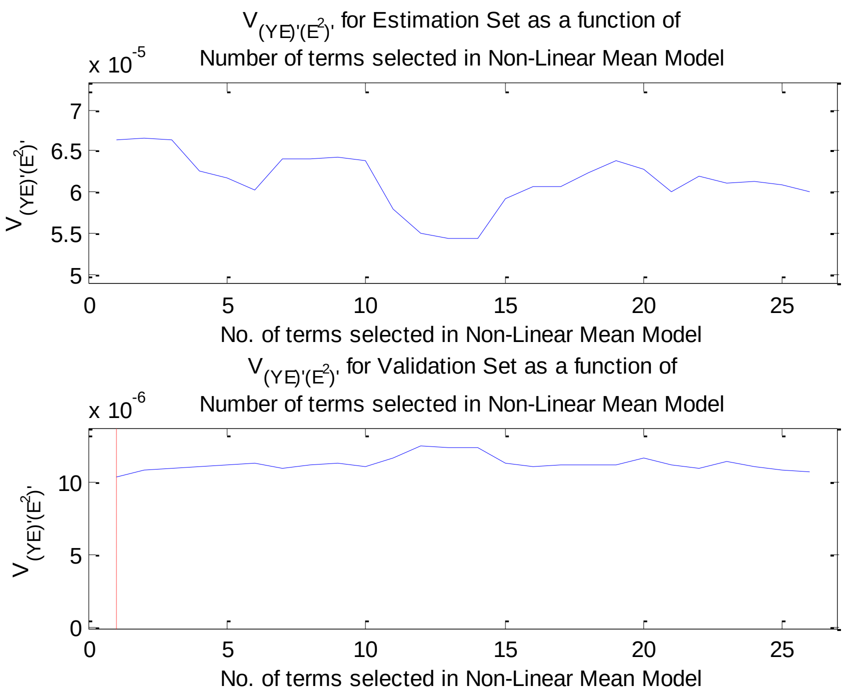


Figure 5.44 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Second Order Non-Linear Mean Model

For the non-linear mean model comprising of 1 term from Table 5.19, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 9.2401E-08. In this case, according to the AIC criterion, 6 terms from the top of Table 5.19 are selected to be included in the non-linear mean model which yields a value of 2.3247E-07 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude.

For the non-linear mean model comprising of all the 29 terms from Table 5.19, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 3.6395E-05. For the non-linear mean model comprising of 6 terms from the top of Table 5.19, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is 4.6856E-05 which is close to the minimum value.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 5.46.

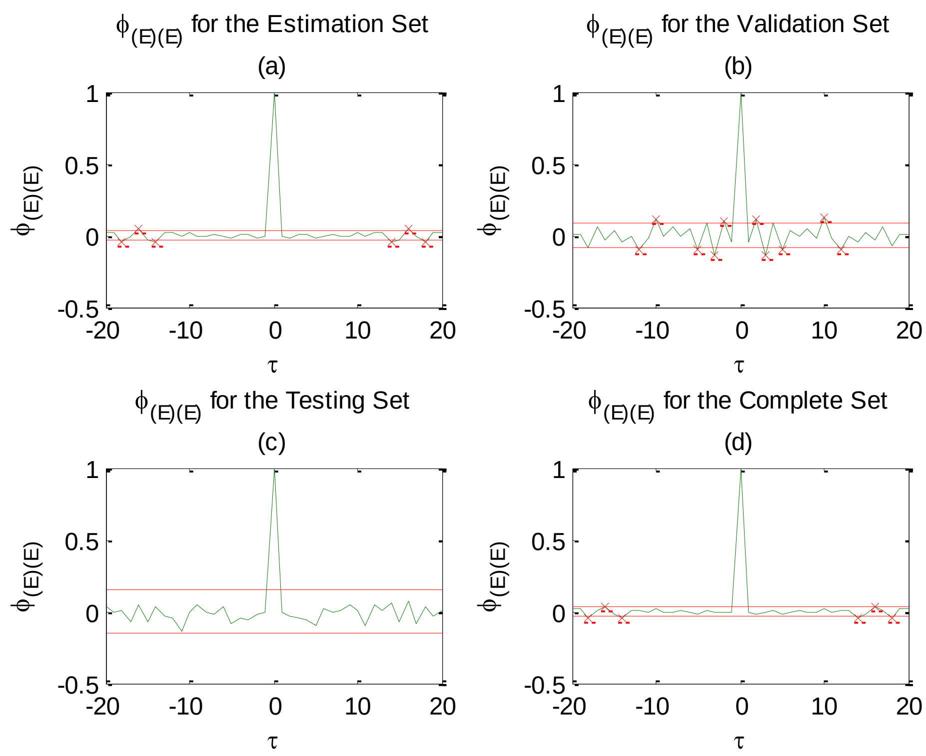


Figure 5.45 Autocorrelation Plots of Residuals for the Selected Second Order Non-Linear Mean Model

In Figure 5.46 (d), note that there exists no autocorrelation of the residuals implying that the fitted second order non-linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.33 (d)) which indicates that the selected second order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant mean model.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected non-linear mean model are shown in Figure 5.46.

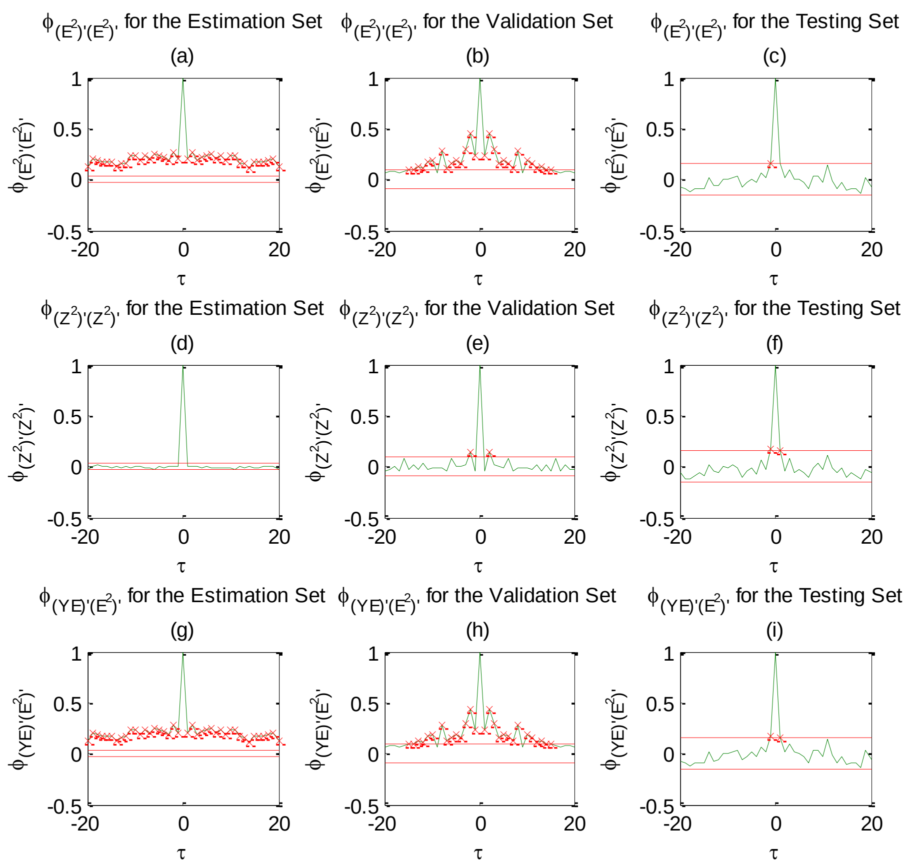


Figure 5.46 Higher Order Correlation Plots for the Selected Second Order Non-Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.46 (d), there exist no violations of the 95% confidence bands. In Figure 5.46 (e), $\Phi_{(z^2)'(z^2)'}$ lies just outside the 95% confidence band for lag = 2. In Figure 5.46 (f), $\Phi_{(z^2)'(z^2)'}$ lies just outside the 95% confidence band for lag = 1. The ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 1.0475, which is lesser than the critical value of 3.8415. Hence, the mean and the variance of the returns can be considered to be adequately fitted.

The selected terms along with the parameter estimates and ERR values are listed in Table 5.20. The coefficient estimates of the 5 noise terms are also included.

Table 5.20 Second Order Non-Linear Mean Model fitted to NASDAQ

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-3)$	3.1289	0.2965
2	$y(t-12)$	0.0518	0.2630
3	$y(t-3)y(t-4)$	-2.3644	0.1446
4	$y(t-2)y(t-4)$	2.1340	0.1438
5	$y(t-13)$	0.0339	0.1121
6	$y(t-1)y(t-1)$	1.0174	0.1122
7	$e(t-1)$	-0.0038	0.0014
8	$e(t-2)$	-0.0132	0.0134
9	$e(t-3)$	0.0114	0.0111
10	$e(t-4)$	-0.0020	0.0003
11	$e(t-5)$	-0.0046	0.0018

So far, the constant, the linear and the second order non-linear mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 3 cases.

5.3.7 Estimation of Third Order Non-Linear Mean Model using WOFR

The last mean model to be fitted to the returns of the NASDAQ index is the third order non-linear candidate mean model listed in Section 5.3.3.4. The results of the WOFR analysis are shown in Table 5.21.

Table 5.21 Terms in Third Order Non-Linear Candidate Mean Model reordered after weighting

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-1)y(t-1)$	-31.5330	0.3196
2	$y(t-1)y(t-1)y(t-2)$	-64.9987	0.3305
3	$y(t-12)$	0.0516	0.2390

4	$y(t-1)y(t-3)$	1.5830	0.1247
5	$y(t-3)y(t-4)$	-1.6931	0.1323
6	$y(t-13)$	0.0363	0.1236
7	$y(t-2)y(t-4)$	1.7179	0.1125
8	$y(t-1)y(t-4)y(t-4)$	22.4857	0.1036
9	$y(t-3)y(t-3)y(t-3)$	29.0255	0.0577
10	$y(t-1)y(t-2)y(t-3)$	-59.8278	0.0895
11	$y(t-3)y(t-5)y(t-5)$	-31.4166	0.0694
12	$y(t-1)y(t-3)y(t-5)$	-34.4527	0.0654
13	$y(t-2)y(t-2)y(t-5)$	-52.7194	0.0719
14	$y(t-2)y(t-2)$	-1.9478	0.0592
15	$y(t-3)y(t-3)$	1.0544	0.0973
16	$y(t-10)$	0.0200	0.0497
17	$y(t-2)y(t-2)y(t-4)$	-43.1814	0.0344
18	$y(t-1)y(t-1)y(t-4)$	47.9860	0.0651
19	$y(t-2)y(t-4)y(t-4)$	-36.5570	0.0589
20	$y(t-1)y(t-1)y(t-3)$	-8.9310	0.0408
21	$y(t-6)$	-0.0198	0.0294
22	1	0.0002	0.0315
23	$y(t-2)y(t-2)y(t-2)$	23.3448	0.0209
24	$y(t-1)y(t-2)y(t-2)$	36.5415	0.0312
25	$y(t-2)y(t-2)y(t-3)$	19.3295	0.0196
26	$y(t-3)y(t-5)$	0.9461	0.0181
27	$y(t-9)$	-0.0132	0.0179
28	$y(t-1)y(t-2)y(t-5)$	27.6813	0.0182
29	$y(t-1)y(t-2)y(t-4)$	-40.0591	0.0190
30	$y(t-3)y(t-4)y(t-5)$	-42.8410	0.0145
31	$y(t-2)y(t-3)y(t-5)$	-40.2467	0.0150
32	$y(t-3)y(t-4)y(t-4)$	-28.4840	0.0132
33	$y(t-5)y(t-5)y(t-5)$	13.4714	0.0126
34	$y(t-4)y(t-4)y(t-5)$	-32.1786	0.0178
35	$y(t-4)y(t-4)y(t-4)$	-12.1889	0.0174
36	$y(t-4)y(t-5)y(t-5)$	27.2226	0.0161
37	$y(t-4)y(t-5)$	0.8930	0.0167
38	$y(t-1)y(t-1)$	0.4863	0.0109
39	$y(t-1)y(t-5)$	0.6713	0.0118
40	$y(t-2)y(t-5)$	0.4744	0.0090
41	$y(t-1)y(t-3)y(t-4)$	-26.8202	0.0085
42	$y(t-1)y(t-4)y(t-5)$	-30.1561	0.0107
43	$y(t-4)y(t-4)$	-0.3924	0.0100
44	$y(t-2)y(t-3)y(t-3)$	-16.1998	0.0079
45	$y(t-2)y(t-3)$	-0.6487	0.0078
46	$y(t-1)y(t-3)y(t-3)$	11.6571	0.0082

47	$y(t-11)$	0.0083	0.0060
48	$y(t-8)$	-0.0079	0.0058
49	$y(t-2)y(t-4)y(t-5)$	-22.4254	0.0046
50	$y(t-1)y(t-1)y(t-5)$	11.1024	0.0058
51	$y(t-2)y(t-3)y(t-4)$	-16.3538	0.0026
52	$y(t-3)y(t-3)y(t-4)$	-9.6988	0.0028
53	$y(t-1)y(t-4)$	-0.3272	0.0031
54	$y(t-1)$	0.0048	0.0020
55	$y(t-3)y(t-3)y(t-5)$	-5.7301	0.0015
56	$y(t-1)y(t-2)$	0.2656	0.0014
57	$y(t-2)y(t-5)y(t-5)$	-4.8777	0.0010
58	$y(t-4)$	-0.0037	6.0922E-04
59	$y(t-2)$	-0.0036	6.5641E-04
60	$y(t-1)y(t-5)y(t-5)$	3.2473	5.3158E-04
61	$y(t-3)$	-0.0026	3.6045E-04
62	$y(t-7)$	0.0019	2.6246E-04
63	$y(t-5)y(t-5)$	0.0705	2.7812E-04
64	$y(t-5)$	0.0008	2.8878E-05

Using the hyper-parameters in Table 5.3, and $n_e = 5$, steps 7 to 10 of the method described in Section 4.5 are carried out on the re-ordered terms of the third order non-linear candidate mean model.

5.3.7.1 Progression of AIC for the Estimation Set

Figure 5.47 shows AIC and $AIC_{\%}$ as a function of the number of terms selected in the third order non-linear mean model.

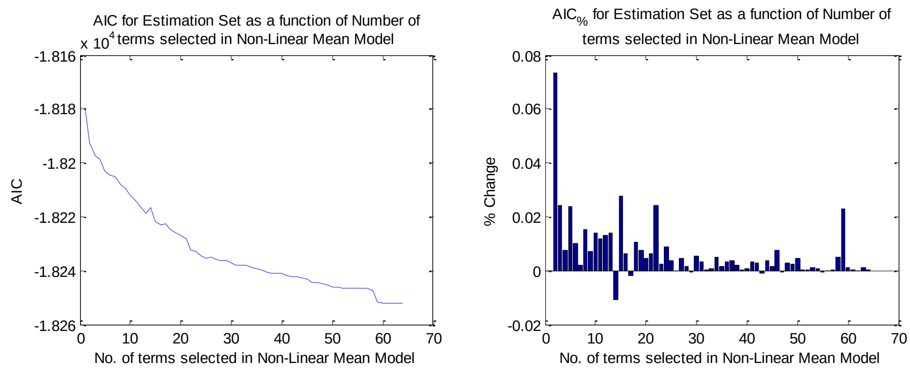


Figure 5.47 AIC and $AIC_{\%}$ as a function of Number of terms selected in the Third Order Non-Linear Mean Model

From Figure 5.47, the AIC drastically decreases to a value of -1.8232E+04 till 22 terms from the top of Table 5.21 are selected to be included in the third order non-linear mean model. Also, $AIC_{\%}$ for the 22nd iteration is the third highest (0.0244%). The addition of further terms to the mean model decreases the value of AIC gradually to a minimum value of -1.8252E+04. Hence, 22 terms from the top of Table 5.21 are selected to be included in the third order non-linear mean model.

5.3.7.2 $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Estimation and Validation Sets

Figure 5.48 shows $V_{(\varepsilon^2)'(\varepsilon^2)'}$, Figure 5.49 shows $V_{(Z^2)'(Z^2)'}$, and Figure 5.50 shows $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for 20 lags for the Estimation and Validation Sets as a function of the number of terms selected. The number of terms that yield the least $V_{(\varepsilon^2)'(\varepsilon^2)'}$, $V_{(Z^2)'(Z^2)'}$ and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set are denoted by a red dashed line in each figure.

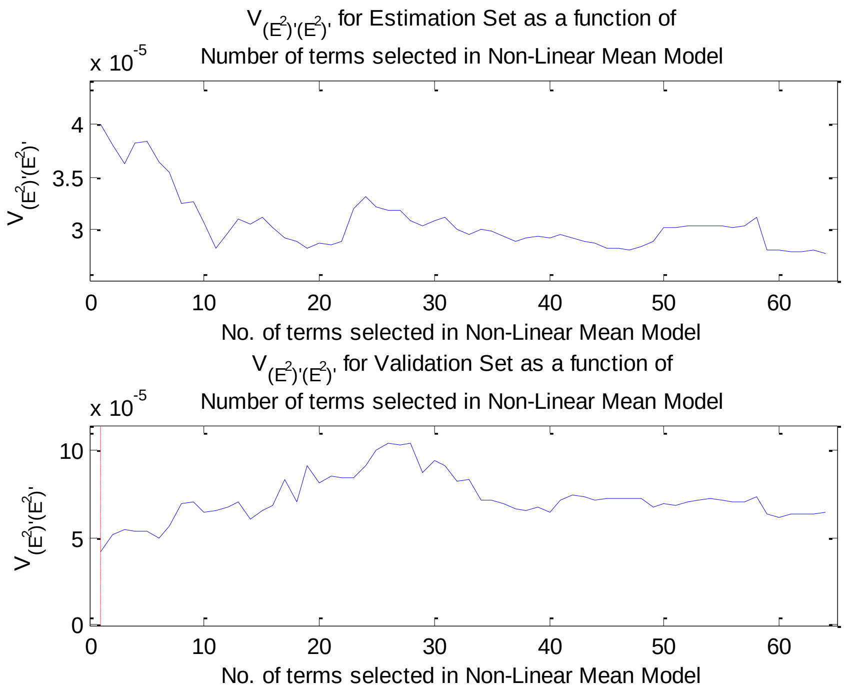


Figure 5.48 $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

For the non-linear mean model comprising of the first term from Table 5.21, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 4.1532E-05. For the non-linear mean model comprising of 22 terms from the top of Table 5.21, $V_{(\varepsilon^2)'(\varepsilon^2)'}$ for the Validation Set is 8.4167E-05.

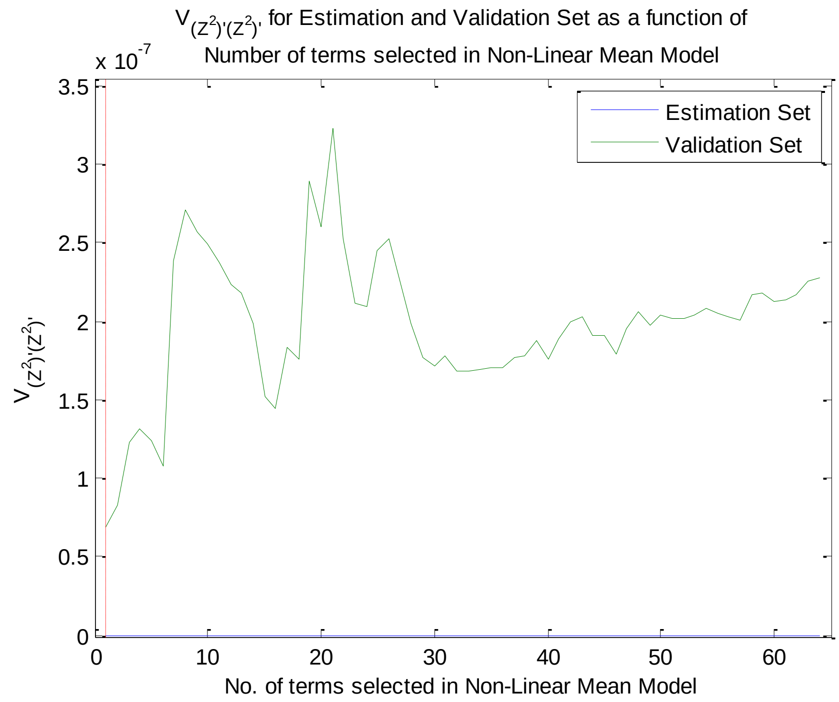


Figure 5.49 $V_{(z^2)'(z^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

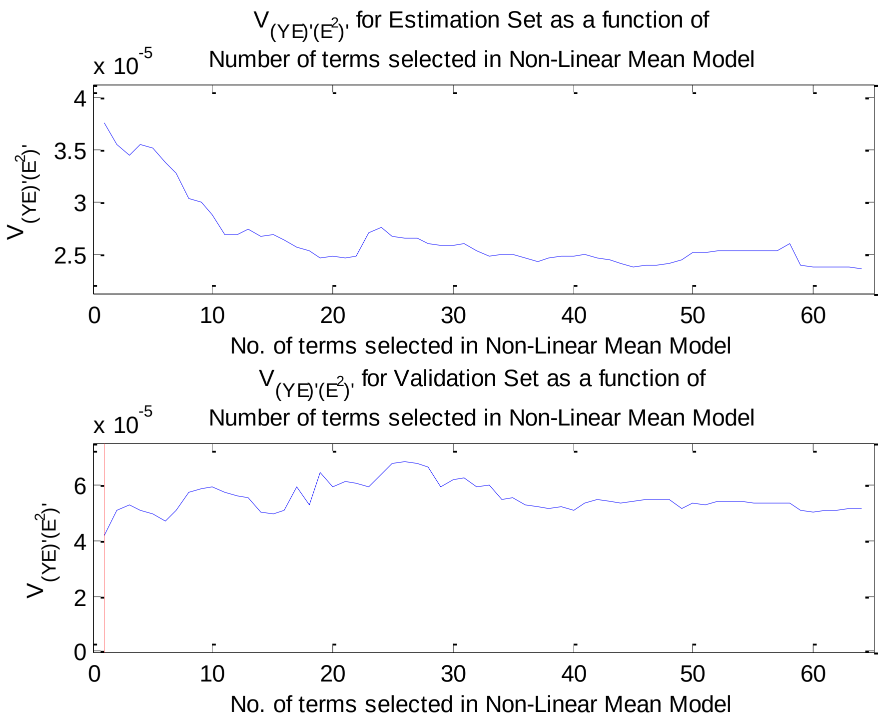


Figure 5.50 $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for Estimation and Validation Sets as a function of Number of terms selected in Third Order Non-Linear Mean Model

For the non-linear mean model comprising of the first from Table 5.21, $V_{(Z^2)'(Z^2)'}$ for the Validation Set is the least with a value of 6.8680E-08. In this case, according to the AIC criterion, 22 terms from the top of Table 5.21 are selected to be included in the non-linear mean model which yields a value of 2.5262E-07 for $V_{(Z^2)'(Z^2)'}$ for the Validation Set, which is of significantly less magnitude.

For the non-linear mean model comprising of the first term from Table 5.21, $V_{(Y\varepsilon)'(\varepsilon^2)'}$ for the Validation Set is the least with a value of 4.2056E-05.

To validate the fitted mean model, the linear autocorrelation of the residuals obtained after fitting the selected second order non-linear mean model are plotted in Figure 5.51.

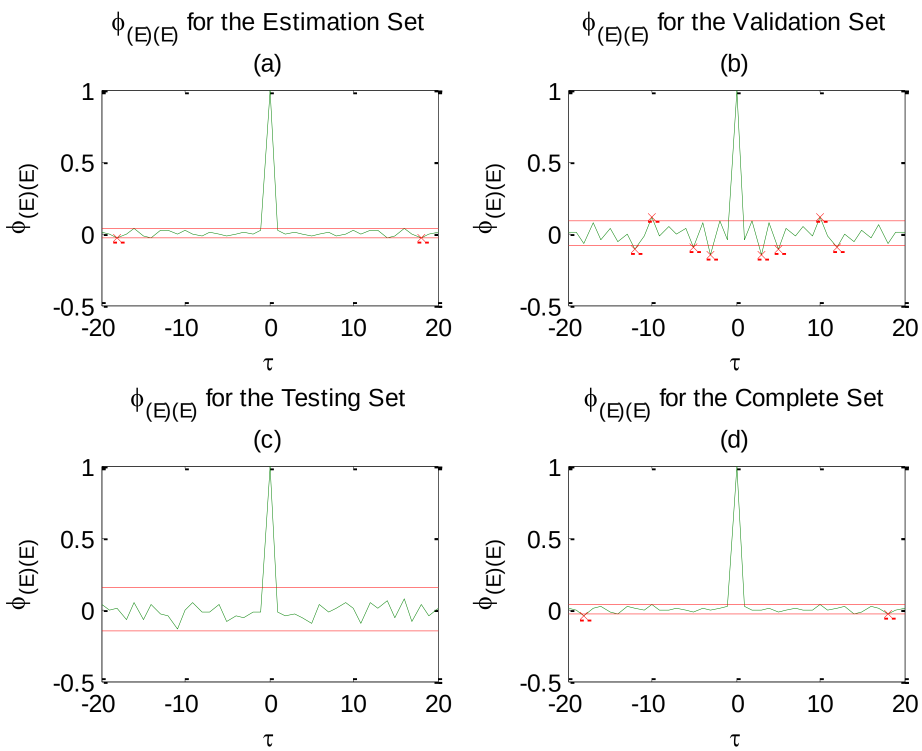


Figure 5.51 Autocorrelation Plots of Residuals for the Selected Third Order Non-Linear Mean Model

In Figure 5.51 (d), note that there exists no autocorrelation of the residuals implying that the fitted third order non-linear mean model is adequate. The magnitude of autocorrelation is lesser than that in the residuals obtained after fitting the constant mean model (Figure 5.33 (d)) which indicates that the selected third order non-linear mean model captures the predictable elements of the mean of the returns much better than the selected constant mean model.

To validate the fitted mean and variance models, the higher order correlation plots of the squared estimated residuals, $\hat{e}^2(t)$, and the squared estimated standardised residuals, $\hat{z}^2(t)$, obtained after fitting the selected non-linear mean model are shown in Figure 5.52.

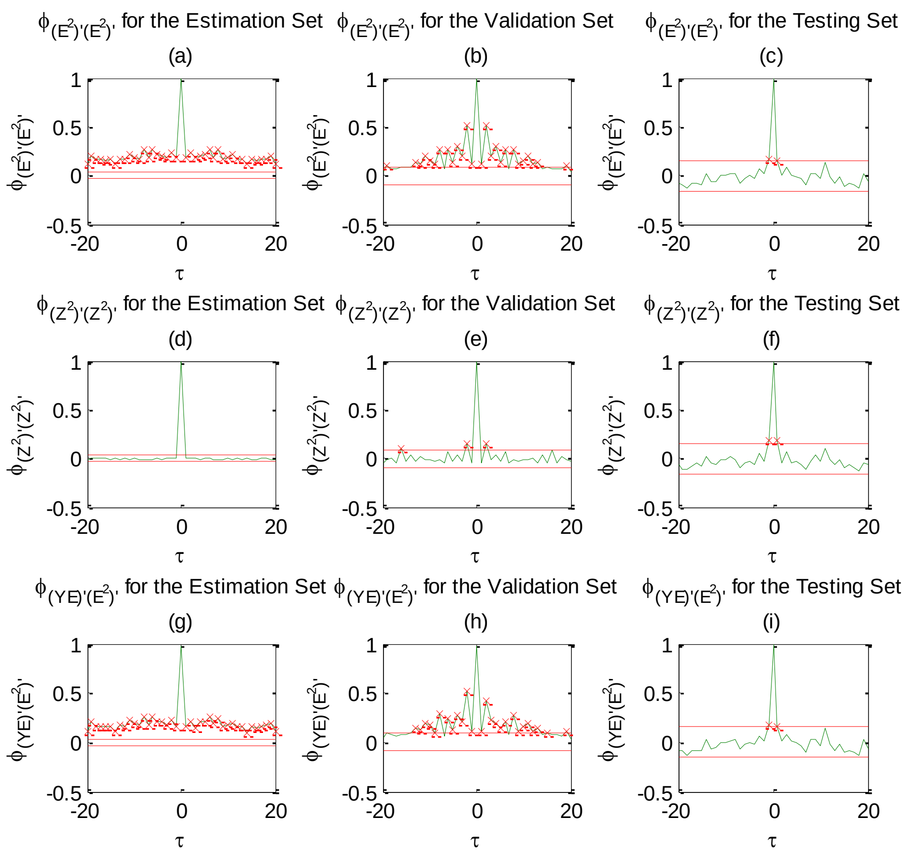


Figure 5.52 Higher Order Correlation Plots for the Selected Third Order Non-Linear Mean Model

The violation of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(\varepsilon^2)'(\varepsilon^2)'}$ and $\Phi_{(Y\varepsilon)'(\varepsilon^2)'}$, indicate the need to fit a variance model to the mean of the returns. For the fitted variance model to be adequate, no violations of the 95% confidence bands in the higher order autocorrelation, $\Phi_{(z^2)'(z^2)'}$, should exist. In Figure 5.52 (d), there exists no violations of the 95% confidence bands. In Figure 5.52 (e), $\Phi_{(z^2)'(z^2)'}$ lies on or just outside the 95% confidence band for lag = 2. The confidence violations are not large, and the ARCH Test Statistic of $\hat{z}(t)$ for the Validation Set is calculated to be 1.1221, which is lesser than the critical value of 3.8415. Hence, the mean and the variance of the returns can be considered to be adequately fitted.

The selected terms along with the coefficient estimates and ERR values are listed in Table 5.22. The coefficient estimates of the 5 noise terms are also included.

Table 5.22 Third Order Non-Linear Mean Model fitted to NASDAQ

No.	Term	Parameter Estimate	ERR
1	$y(t-1)y(t-1)y(t-1)$	-32.3482	0.3636
2	$y(t-1)y(t-1)y(t-2)$	-65.6457	0.3008
3	$y(t-12)$	0.0521	0.2416
4	$y(t-1)y(t-3)$	1.5498	0.1493
5	$y(t-3)y(t-4)$	-1.5961	0.1466
6	$y(t-13)$	0.0341	0.1161
7	$y(t-2)y(t-4)$	1.5201	0.1067
8	$y(t-1)y(t-4)y(t-4)$	41.3240	0.1311
9	$y(t-3)y(t-3)y(t-3)$	29.8598	0.0645
10	$y(t-1)y(t-2)y(t-3)$	-58.7972	0.0849
11	$y(t-3)y(t-5)y(t-5)$	-24.7672	0.0499
12	$y(t-1)y(t-3)y(t-5)$	-62.1985	0.0598
13	$y(t-2)y(t-2)y(t-5)$	-42.5019	0.0723
14	$y(t-2)y(t-2)$	-2.0335	0.0854
15	$y(t-3)y(t-3)$	0.9609	0.0721
16	$y(t-10)$	0.0215	0.0505
17	$y(t-2)y(t-2)y(t-4)$	-44.8129	0.0334
18	$y(t-1)y(t-1)y(t-4)$	45.7599	0.0584
19	$y(t-2)y(t-4)y(t-4)$	-32.1252	0.0483
20	$y(t-1)y(t-1)y(t-3)$	-24.7960	0.0406
21	$y(t-6)$	-0.0219	0.0333
22	1	0.0003	0.0354
23	$e(t-1)$	0.0115	0.0089
24	$e(t-2)$	0.0111	0.0077
25	$e(t-3)$	-0.0038	0.0009
26	$e(t-4)$	-0.0120	0.0091
27	$e(t-5)$	0.0062	0.0026

So far, all the mean models seem to pass standard financial model validation tests and suggest that the mean and the variance of the returns have been adequately modelled in all the 4 cases.

5.3.8 Comparison of ARCH Test Statistics and Non-Linear Correlation Statistics of All Mean Models

The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the selected mean models are listed in Table 5.23. The minimum values are shaded in blue.

Table 5.23 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Mean Models

Type of Mean Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Constant	30.2857	4.4260E-05
Linear	28.1956	3.9677E-05
Second Order Non-Linear	12.3207	4.2073E-05
Third Order Non-Linear	7.3515	2.8984E-05

The third order non-linear mean model (terms and parameter estimates listed in Table 5.22) has the lowest ARCH Test Statistic of $\hat{e}(t)$ for the Validation Set. This implies that the mean of the returns have been modelled better by the third order non-linear mean model than the second order non-linear, linear or constant mean model, and hence is the appropriate choice.

5.3.9 Comparison of Selected Mean Model to Vanilla GARCH Model

A vanilla GARCH model (constant mean model with a GARCH(1,1) variance model) is fitted to the given NASDAQ return series. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the fitted vanilla GARCH model are listed and compared to the selected third order non-linear mean model in Table 5.24. The minimum values are shaded in blue.

Table 5.24 ARCH Test statistic of $\hat{e}(t)$ of Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of Complete data set for all Vanilla GARCH and Selected Mean Model fitted to NASDAQ

Type of Model	ARCH Test Statistic of $\hat{e}(t)$ (Validation Set)	$V_{(Y\varepsilon)'(\varepsilon^2)'}$ (Complete Set)
Vanilla GARCH	30.3239	4.3598E-05
Selected Third Order Non-Linear Mean Model with ARCH(25) Variance Model	7.3515	2.8984E-05

The selected third order non-linear mean model with an ARCH(25) variance model performs much better than a vanilla GARCH model. The ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of the selected model are lower than those of the vanilla GARCH model.

Figure 5.53 and Figure 5.54 show the one-step-ahead (OSA) return estimates generated using the vanilla GARCH model and the selected third order non-linear mean model.

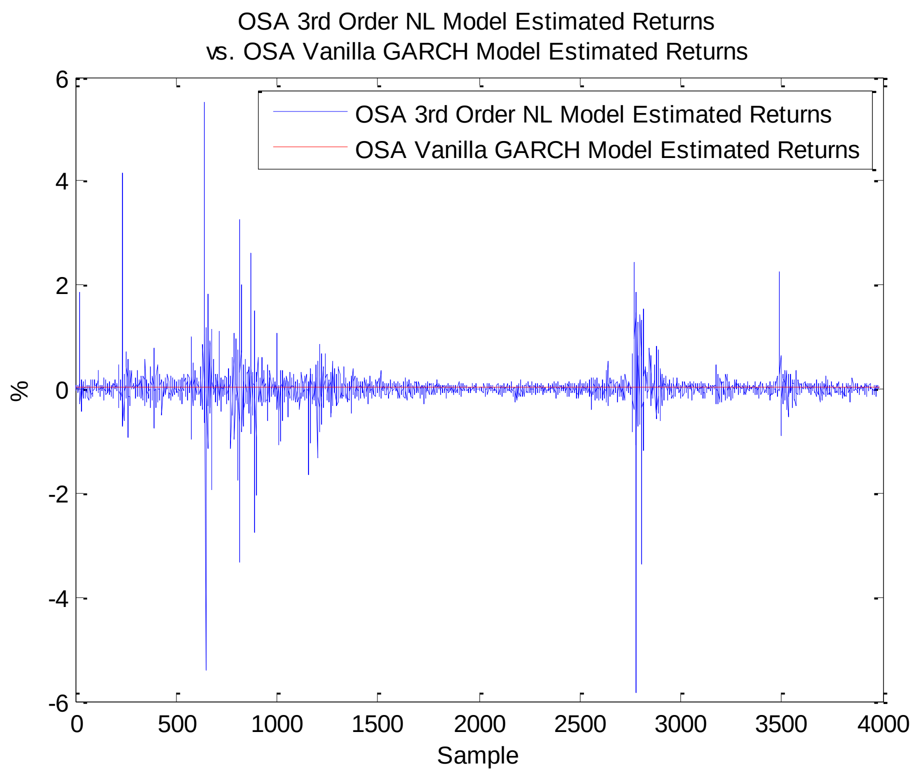


Figure 5.53 OSA Return Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for NASDAQ

From Figure 5.53 and Figure 5.54, it can be noted that using the third order non-linear mean model certainly captures the predictable elements of the mean of the returns, rather than just using a constant mean model and passing off the predictable elements to be included in the residuals. The standard deviation of the various return series are listed in Table 5.25.

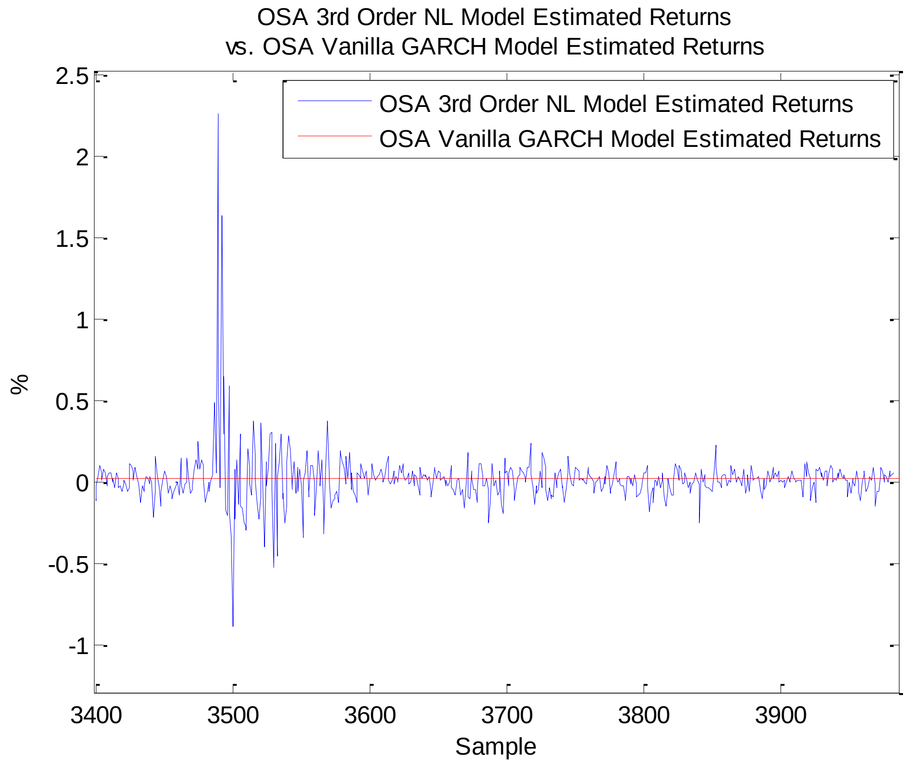


Figure 5.54 OSA Return Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for NASDAQ (Samples 3400 to 3985)

Table 5.25 Standard Deviation of various return series for NASDAQ

Series	Standard Deviation
OSA Return Estimate of Constant Mean Model	0
OSA Return Estimate of Third Order NL Mean Model	0.0031
True Return Series	0.0176

Comparing the magnitudes of the Validation and Testing Sets of the OSA return estimates generated using the third order non-linear mean model to those of the true return series, 354 samples of 700 samples have the same magnitude. Hence, the magnitude of the returns is predicted right 50.5714% of the time.

Figure 5.55 and Figure 5.56 show the variance estimates generated using the vanilla GARCH model and the selected third order non-linear mean model.

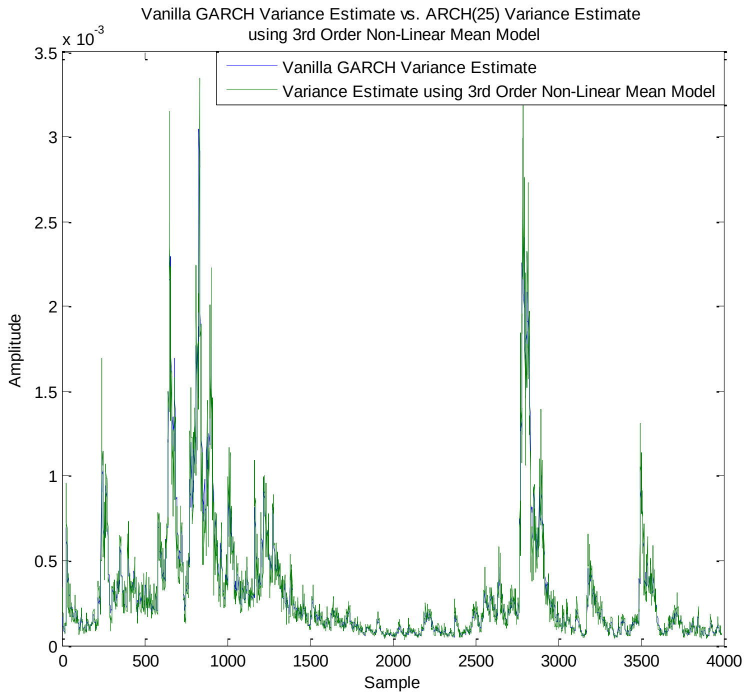


Figure 5.55 Variance Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for NASDAQ

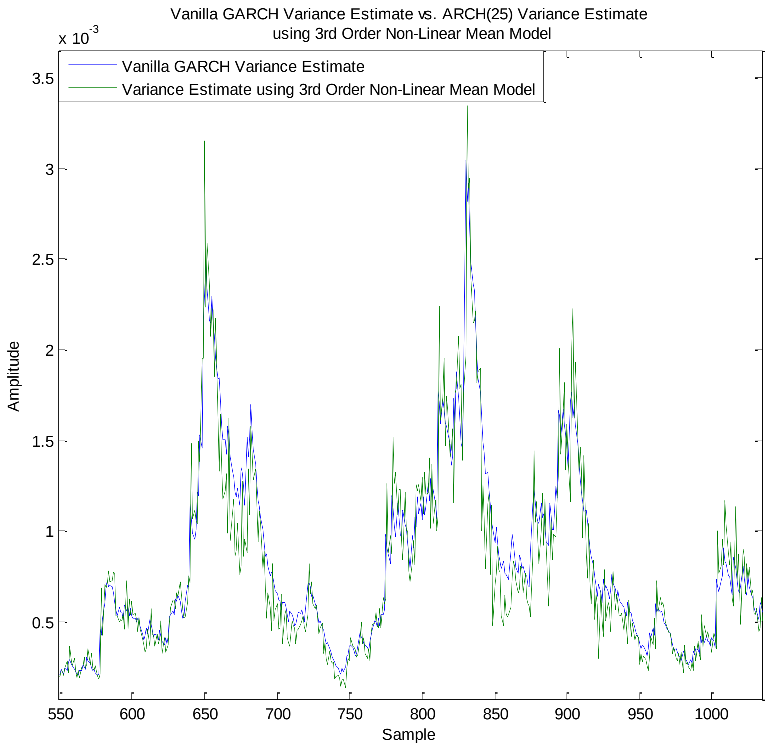


Figure 5.56 Variance Estimates generated using Vanilla GARCH Model and Third Order Non-Linear Mean Model for NASDAQ (Samples 550 to 1050)

Note that the variance estimates are similar during periods of low volatility, but a little different during periods of high volatility. The difference is not as pronounced as in the case of the FTSE100, but it still exists.

To summarise, constant, linear, second order non-linear and third order non-linear candidate mean models were fitted to the given returns of NASDAQ. After carrying out term selection, the ARCH Test statistics of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set of all the models were compared to evaluate how well each model described the mean of the given return series. As in the case of the simulated example, all the fitted mean models passed standard model validation tests, but the third order non-linear mean model was found to have the lowest values for the ARCH Test statistic of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set, implying that the mean of the returns have been modelled best by the third order non-linear mean model.

5.4 Conclusions

The framework is used to identify the best mean model for 2 real financial return series. For each real data set, the best mean model amongst a plethora of mean models is selected. For both the data sets, a constant mean model performs the worst in terms of the ARCH Test statistic of $\hat{e}(t)$ of the Validation Set and $V_{(Y\varepsilon)'(\varepsilon^2)'}$ of the Complete data set. A third order non-linear mean model is preferred over a constant or a linear mean model, even though all the fitted mean models pass standard model validation tests.

Fitting vanilla GARCH models to the real financial data sets yield worse performance statistics (ARCH Test Statistics and Non-Linear Correlation Statistics) than those obtained from fitting linear and non-linear mean models to these data sets.

For the FTSE100 data set, the difference between the variance estimates obtained using the vanilla GARCH model and the selected third order non-linear mean model (Figure 5.27 and Figure 5.28) is much more pronounced at the peaks, than in the case of the NASDAQ data set(Figure 5.55 and Figure 5.56). For the FTSE100 data set, this suggests that a constant mean model does not capture all the predictable elements in the mean of the returns and adds them to the residuals, thereby increasing the magnitude of the variance estimate at the peaks. The fitted third order non-linear mean model yields a variance estimate that is much lower in magnitude during periods of high volatility.

For the case of the NASDAQ data set, this indicates that fitting a third order non-linear mean model does not impact variance estimation much as compared to fitting a constant mean model. Hence, correctly selecting and fitting a non-linear mean model to the returns is recommended in order to obtain more accurate variance estimates and to obtain standardised residuals that are more ‘white’ in nature.

Chapter 6

Weighted Least Squares Estimation of the Variance Model

6.1 Introduction

Of the two parts of a GARCH model, the variance model has been given much more importance by researchers and a truly vast number of GARCH variance models have been developed over the years.

The standard approach to fitting a variance model to a given financial return series involves testing the returns for non-linearity, choosing a type of linear or non-linear variance model to fit depending on the results of the previous test, estimating the model using some Maximum Likelihood (ML) method, performing model validation tests to check whether the selected variance model can adequately describe the given data set and selecting a different variance model if the validation tests fail (Engle, 2001).

As noted, the standard paradigm involves estimating the model by ML. There are two problems with this, however. First, there is no term selection. The structure of the variance model cannot be determined and has to be selected – that is to say, assumed - beforehand. There exists no simple method of predetermining the exact type of variance model to fit the data accurately. The second, related, problem is that there is no definite method to select the best model. It is therefore possible for two different models to give plausible fits without being able to determine which is best.

The NARMAX methodology for system identification avoids both these problems. Term selection and parameter estimation can be carried out easily using Orthogonal Forward Regression (OFR) (Billings et al., 1988, 1989; Korenberg et al., 1988, Chen et al., 1989, Billings and Zhu, 1994) or WOFR (Zhao, 2010). As in the case of the GARCH mean model, fitting the variance model using OFR does not give accurate results due to the presence of heteroskedasticity in the financial return series data (Bjorck, 1996), but this problem can be overcome using WOFR, as it was with the mean model considered in the previous chapter.

System identification techniques require the inputs and outputs of the system to be observable and measurable (Billings and Coca, 2001). This is a major drawback of using NARMAX methods to model the GARCH variance model since the variance of the returns at any instant in time is not observable. To work around this limitation, an estimate of the GARCH variance needs to be used. The quality of the estimated variance then determines the accuracy of term selection and parameter estimation. To anticipate the conclusions of this chapter, if the underlying true GARCH variance is linear, this estimate works well, but if the underlying true GARCH variance is non-linear, then the estimate does not work well and a better estimate of the GARCH variance is required to capture the underlying non-linearity.

The purposes of this chapter are to introduce the financial variance model, the current methods used for the estimation of a GARCH variance model and to suggest a new method to successfully select the terms and estimate the coefficients of a GARCH variance model using NARMAX methodology.

This chapter is laid out as follows. Section 4.2 gives a brief overview of the financial variance model. Section 4.3 showcases the maximum likelihood estimation method used to estimate the GARCH variance model. Section 4.4 describes the NARMAX methodology and introduces a new method to select and estimate the terms in a GARCH variance model using NARMAX methods. Simulations are included to demonstrate the performance of the method introduced in this section. Section 4.5 concludes the chapter.

6.2 The GARCH Financial Variance Model

In a GARCH(p, q) model, the variance is modelled as

$$h(t) = K + \sum_{i=1}^{p} G_i . h(t-i) + \sum_{j=1}^{q} A_j . e^2(t-j) \tag{6.1}$$

$$\begin{gathered} p \geq 0, q > 0 \\ K > 0, G_i \geq 0, i = 1, \dots, p \\ A_j \geq 0, j = 1, \dots, q \\ \sum_{i=1}^{max(p,q)} G_i + A_i < 1 \end{gathered}$$

where $h(t)$ is the variance of the return and $e(t)$ is the residual and is also known as innovation of the returns process. K, G_i and A_j are constants. In the final condition, it is implied that $G_i = 0$ for $i > p$ and $A_j = 0$ for $j > q$.

It can be seen that the GARCH(p, q) variance model is a linear AutoRegressive Moving Average model with eXogenous inputs (ARMAX). Here, the squared residual, $e^2(t)$, is the exogenous input.

There exist several non-linear variance models as well such as the EGARCH model (Section 2.3.6), the QGARCH model (Section 2.3.7), the NA-GARCH model (Section 2.3.8), the SQR-GARCH model (Section 2.3.9) and the GJR-GARCH model (Section 2.3.10). Most of the variance models, linear and non-linear, are driven by lagged variance $(h(t-i))$, lagged squared residual $(e^2(t-j))$, and in some cases, lagged residual $(e(t-i))$ terms and possibly a combination of these.

6.3 Maximum Likelihood Estimation of the GARCH Variance Model

6.3.1 Probability Density Function

Let the vector $x = [x_1, x_2, \dots, x_n]$ represent the data to be modelled where n represents the total number of observations of the sample data. From a statistical point of view, x is a random sample from an unknown population and is generated by a model. Every population has a unique probability distribution (also known as probability density function or PDF) that is generated by a predefined set of model parameters. A change in the model parameters changes the probability distribution.

Let $F(x|\theta)$ denote the probability distribution function of x depending on the model parameters $\theta = [\theta_1, \theta_2, \dots, \theta_m]$ in the parameter space Θ, where m represents the total number of parameters. The PDF of the data series, x, can be expressed as a product of the PDFs of the individual variance observations

$$F(x|\theta) = F_1(x_1|\theta)F_2(x_2|\theta) \dots F_n(x_n|\theta) \tag{6.2}$$

6.3.2 Likelihood Function

Given that the data, x, and the probability density function of the data, $F(x|\theta)$, are assumed known, the set of model parameters, θ, that correspond to the known PDF need to be found. This is done by defining a likelihood function

$$L(\theta|x) = F(x|\theta) \tag{6.3}$$

The values of θ within the parameter space Θ that maximise the likelihood function $L(\theta|x)$ are then the ideal or likelihood-maximising set of parameters.

6.3.3 Maximum Likelihood Estimation Method

For the GARCH class of variance models, the variance, $h(t)$, the random i.i.d sequence, $z(t)$, and the residuals, $e(t) = z(t)\sqrt{h(t)}$ are assumed to have a Gaussian probability distribution function.

$$F(x|\theta) = \frac{1}{\sqrt{2\pi h(t)}} exp\left(-\frac{e^2(t)}{2h(t)}\right) \tag{6.4}$$

Let the average log-likelihood be denoted by l and the log-likelihood of the t^{th} observation be denoted by $l(t)$. For N samples,

$$l = \frac{1}{N}\sum_{t=1}^{N} l(t) \tag{6.5}$$

$$\begin{aligned} l(t) &= ln\left(\frac{1}{\sqrt{2\pi h(t)}} exp\left(-\frac{e^2(t)}{2h(t)}\right)\right) \\ &= -\frac{1}{2} ln\big(2\pi h(t)\big) + \left(-\frac{e^2(t)}{2h(t)}\right) \\ &= -\frac{1}{2} ln(2\pi) - \frac{1}{2} ln(h(t)) - \frac{e^2(t)}{2h(t)} \end{aligned}$$

Since, $-\frac{1}{2} log(2\pi)$ is a constant, it can be ignored while performing maximisation. The modified log-likelihood function to use for maximisation for the t^{th} observation becomes

$$l_m(t) = -\frac{1}{2} ln\big(h(t)\big) - \frac{e^2(t)}{2h(t)} \tag{6.6}$$

For a general GARCH(p, q) model, the mean is usually modelled as, $y(t) = a_0 + e(t)$, and the variance is modelled as shown in equation (6.1). The parameter constraints of the variance model are also as shown in equation (6.1).

Within this constrained parameter space, the values of $K, G_1, \ldots, G_p, A_1, \ldots, A_q$ are found such that the average log-likelihood function l_m is maximised.

Several minimisation routines can be implemented via MATLAB. Minimising the negative of the average log-likelihood function is the same as maximising the average log-likelihood function. Hence, the cost function supplied to the minimisation routine in MATLAB is the negative of the average log-likelihood function. Throughout this chapter, all negative log-likelihood minimisation routines were performed using 'fmincon'.

6.4 The NARMAX Approach for the Identification and Estimation of the GARCH Variance Model

6.4.1 Introduction

The NARMAX methodology can be used to identify and represent any system whose inputs and outputs are measurable. All that is required for the methods to be applied are the data series of the input(s) and output(s) of the system to be modelled. For a GARCH(p,q) variance model, the variance, $h(t)$ is the output and the lagged variance, $h(t-i)$, and lagged squared residuals, $e^2(t-j)$, are the inputs.

For any real financial asset, only the price and hence, the returns are observable. The variance of the returns of an asset is not observable. Hence, the time series data of the true variance of the returns of an asset is never available. This highlights a major setback to using the NARMAX approach for modelling the GARCH variance. To be able to use the NARMAX approach to model the GARCH variance, an estimate of the variance must be generated from the residuals of the GARCH mean model. The closer the estimate is to the true GARCH variance, the better the performance of the NARMAX approach.

Another problem with using the NARMAX approach to identify GARCH variance models is the underlying heteroskedastic nature of the GARCH variance. The standard NARMAX methodology works accurately only if the system being modelled is homoskedastic in nature. Modelling a heteroskedastic system using NARMAX methods usually produces biased results (Bjorck, 1996).

This is not a problem when the true GARCH variance is used to model the GARCH variance model using the standard NARMAX approach. The problem arises when an estimate of the GARCH variance is used. The difference between the true GARCH variance and the estimated GARCH variance can be considered as additional noise in the true output (here, the true variance). This noise is heteroskedastic in nature and hence produces biased results whilst using the simple NARMAX approach (Bjorck, 1996; Zhao, 2010).

A WOFR algorithm was developed to counteract this problem of the presence of heteroskedasticity in the GARCH mean model (Zhao, 2010). Weighted Least Squares has been used to counteract the problem of biased parameter estimation of the GARCH variance model using simple Least Squares due to the presence of heteroskedasticity in the GARCH variance model (Tofallis, 2009). A similar approach can be used to implement the NARMAX methodology to accurately select the terms and estimate the respective coefficients in a GARCH variance model when the true GARCH variance is unavailable and an estimate of the GARCH variance is generated from the residuals of the GARCH mean model.

6.4.2 Weighted Orthogonal Forward Regression for the GARCH Variance Model

The Orthogonal Forward Regression (OFR) procedure forms the basis of this algorithm and has been explained in Section 3.3.3. The algorithm for the accurate identification and estimation of the GARCH variance model when an estimate of the GARCH variance is used is as follows:

1. The GARCH mean model is estimated and the one-step-ahead estimates of the returns, $\hat{y}(t)$, are calculated. The modelling residuals of the GARCH mean model, $e(t)$, are then calculated by subtracting the estimated returns, $\hat{y}(t)$, from the true returns, $y(t)$.
$$e(t) = y(t) - \hat{y}(t) \tag{6.7}$$
2. The squared residuals, $e^2(t)$, are then used to fit an ARCH(25) model to the GARCH variance estimate which yields an estimate of the true GARCH variance, $\hat{h}(t)$.
3. The ARCH(25) estimate of the variance, $\hat{h}(t)$, is defined as the output. The lagged squared residuals, $e^2(t-i)$, and the lagged values of the estimated variance, $\hat{h}(t-j)$, are set as the inputs. A GARCH(5,5) model is selected as the candidate model.
$$\hat{h}(t) = K + \sum_{i=1}^{5} G_i \hat{h}(t-i) + \sum_{j=1}^{5} A_j e^2(t-j) \tag{6.8}$$
If N represents the total number of data samples, the set of equations describing the candidate model can be written in matrix form as

$$\begin{bmatrix} \hat{h}(6) \\ \hat{h}(7) \\ \vdots \\ \hat{h}(N) \end{bmatrix} = \begin{bmatrix} 1 & \hat{h}(1) & \cdots & \hat{h}(5) & e^2(1) & \cdots & e^2(5) \\ 1 & \hat{h}(2) & \cdots & \hat{h}(6) & e^2(2) & \cdots & e^2(6) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{h}(N-5) & \cdots & \hat{h}(N-1) & e^2(N-5) & \cdots & e^2(N-1) \end{bmatrix} \begin{bmatrix} K \\ G_1 \\ \vdots \\ G_5 \\ K_1 \\ \vdots \\ K_5 \end{bmatrix} \quad (6.9)$$

4. In an attempt to eliminate the bias caused due to the presence of heteroskedasticity, both sides of equation (6.8) are divided by the term $\hat{h}(t)$ (Tofallis, 2009). Equation (6.8) now becomes

$$1 = \frac{K}{\hat{h}(t)} + \frac{1}{\hat{h}(t)} \sum_{i=1}^{5} G_i \hat{h}(t-i) + \frac{1}{\hat{h}(t)} \sum_{j=1}^{5} A_j e^2(t-j) \quad (6.10)$$

Equation (6.9) becomes

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\hat{h}(6)} & \frac{\hat{h}(1)}{\hat{h}(6)} & \cdots & \frac{\hat{h}(5)}{\hat{h}(6)} & \frac{e^2(1)}{\hat{h}(6)} & \cdots & \frac{e^2(5)}{\hat{h}(6)} \\ \frac{1}{\hat{h}(7)} & \frac{\hat{h}(2)}{\hat{h}(7)} & \cdots & \frac{\hat{h}(6)}{\hat{h}(7)} & \frac{e^2(2)}{\hat{h}(7)} & \cdots & \frac{e^2(6)}{\hat{h}(7)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\hat{h}(N)} & \frac{\hat{h}(N-5)}{\hat{h}(N)} & \cdots & \frac{\hat{h}(N-1)}{\hat{h}(N)} & \frac{e^2(N-5)}{\hat{h}(N)} & \cdots & \frac{e^2(N-1)}{\hat{h}(N)} \end{bmatrix} \begin{bmatrix} K \\ G_1 \\ \vdots \\ G_5 \\ K_1 \\ \vdots \\ K_5 \end{bmatrix} \quad (6.11)$$

5. OFR is now applied to the weighted candidate model. The terms are listed in decreasing order of ERR.
6. A suitable cut-off value is selected for the ERR and the terms that have a greater ERR value are selected. The unwanted terms are removed from the candidate model and steps 3 to 5 are repeated in order to obtain the parameter estimates of the selected terms.

6.4.3 Simulations

The following simulations introduce a new approach to modelling the variance model of any GARCH-class model. The approach is based on NARMAX methods and offers the ability to select the terms that are actually present in the variance model from a broader candidate model.

6.4.3.1 Using OFR for the Identification of the Variance Model when the True Variance and Residuals are Both Known

Consider the following GARCH(3,2) variance model

$$h(t) = K + \sum_{p=1}^{3} G_p h(t-p) + \sum_{q=1}^{2} A_q e^2(t-q) \quad (6.12)$$

where $h(t)$ is the variance and $e^2(t)$ is the squared residual, both at an instant in time, t. K, G_1, G_2, G_3, A_1 and A_2 are the parameters of the variance model. The values of the parameters of the simulated variance model are listed in Table 6.1.

Table 6.1 Parameters of the Simulated GARCH(3,2) Variance Model

Parameter of the Variance Model	Value
K	1.4E-05
G_1	0
G_2	0.3648
G_3	0.3520
A_1	0.0543
A_2	0.1870

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. The simulated variance and squared residuals are shown in Figure 6.1. When simulating the variance from equation (6.12), the residuals $e(t)$ are modelled as $e(t) = z(t)\sqrt{h(t)}$, where $z(t)$ is a random independent and identically distributed (i.i.d) term that has zero mean and a variance of 1. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10).

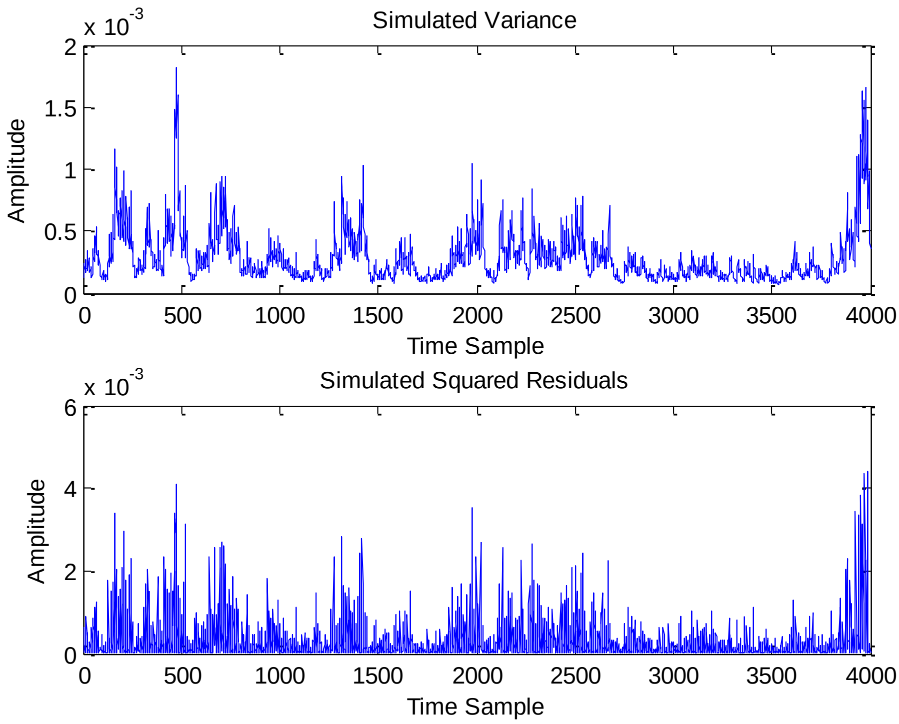


Figure 6.1 Simulated Variance and Squared Residuals for GARCH(3,2) Model

A GARCH(5,5) model is selected as the candidate variance model. The simulated variance, $h(t)$, is defined as the output. The lagged square of the residuals, $e^2(t-i)$, and the lagged values of the simulated variance, $h(t-j)$, are set as the inputs. OFR is applied to the candidate model. Table 6.2 shows the terms selected in decreasing order of ERR.

Table 6.2 Ranking of Terms of Candidate Variance Model selected by OFR

Rank	**Standard OFR**	
	Term	**ERR (%)**
1	$h(t-2)$	92.9966
2	$e^2(t-2)$	5.3308
3	$h(t-3)$	1.1595
4	$e^2(t-1)$	0.4572
5	1	0.0558
6	$h(t-4)$	1.3504E-26
7	$e^2(t-5)$	4.1104E-27
8	$h(t-1)$	3.3389E-27
9	$e^2(t-3)$	0.8256
10	$e^2(t-4)$	0.4965
11	$h(t-5)$	0.2886

The highlighted terms represent the terms present in the original model. It can be seen that all the terms present in the original model have been selected. The ERR cut-off value is set as 0.05% and the parameter estimates of the selected terms are obtained and listed in Table 6.3. The modelling residuals obtained after parameter estimation are shown in Figure 6.2.

Table 6.3 Parameter Estimates of Selected GARCH(3,2) Variance Model using OFR

Parameter of the variance model	**Parameter Estimate**	**True Coefficient**
K	1.4E-05	1.4E-05
G_1	0	0
G_2	0.3648	0.3648
G_3	0.3520	0.3520
A_1	0.0543	0.0543
A_2	0.1870	0.1870

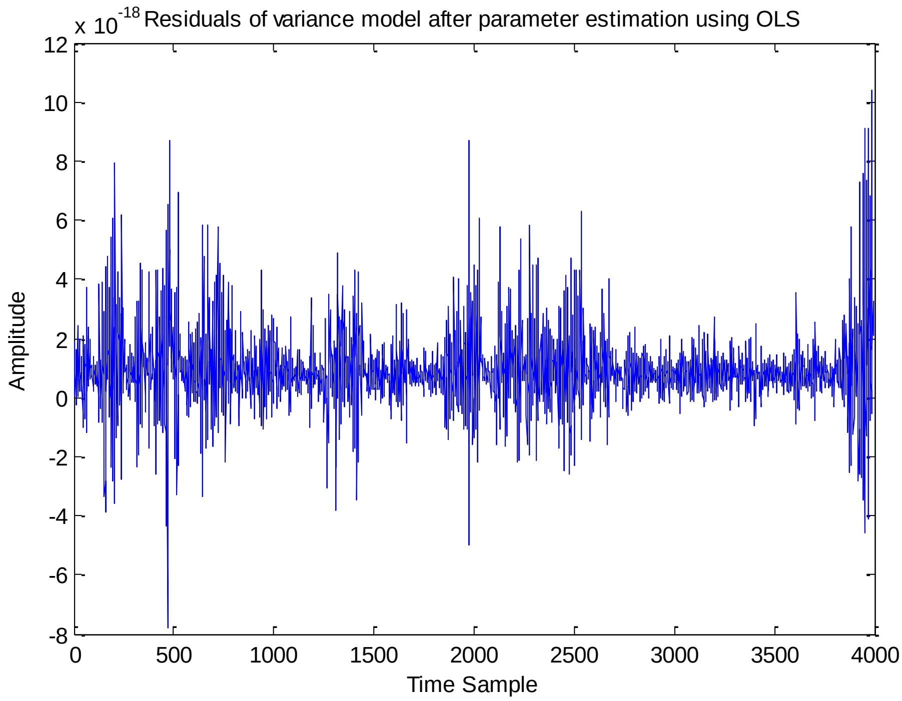


Figure 6.2 Residuals of GARCH(3,2) Variance Model after Parameter Estimation using OFR

From Table 6.3, it can be seen that the parameter estimates of the selected GARCH(3,2) variance model are identical to the true values of the coefficients of the simulated GARCH(3,2) variance model. If the true variance and the true squared residuals are known, OFR works very well to accurately select the terms in a GARCH variance model and estimate their coefficients. Furthermore, Figure 6.2 reveals that the modelling residuals of the GARCH variance model are of very small magnitude (order of E-18) – although they also appear to be heteroskedastic in nature.

In short, the approach proposed appears to work well when the candidate variance model encompasses the true variance model even if the true model is unknown to the modeller, provided the true variance and residuals are known.

Cases where the true variance and/or residuals are unknown are now examined.

6.4.3.2 Using Weighted OFR for the Identification of the GARCH Variance Model when the True Variance is Unknown and the True Residuals are Known

The data used for the following simulation is generated from the model used in the previous simulation (Section 6.4.3.1, see equation (6.12)). It is assumed that the residuals, $e(t)$, are known and that the true GARCH(3,2) variance is unknown.

Since the true variance is now unknown, an estimate of the variance is required. An estimate is obtained by fitting an ARCH(25) model to the residuals using ML. This estimated variance, $\hat{h}(t)$, will be used in place of the unknown true GARCH variance.

$$\hat{h}(t) = K' + \sum_{n=1}^{25} A'_n e^2(t-n) \tag{6.13}$$

where K' and A'_n for $n = 1,2,\ldots,25$ are coefficients of the ARCH(25) model.

A GARCH(5,5) model is selected as the candidate variance model. The estimated ARCH(25) variance, $\hat{h}(t)$, is defined as the output. The lagged square of the residuals, $e^2(t-i)$, and the lagged values of the estimated variance, $\hat{h}(t-j)$, are set as the inputs. OFR is applied to the candidate model. The left-hand side of Table 6.4 shows the terms selected in decreasing order of ERR.

The highlighted terms represent the terms present in the original model. It can be seen that most of the terms present in the original model except the constant term have been selected. The modelling residuals obtained are shown in Figure 6.2. The modelling residuals appear to be heteroskedastic in nature – they are spikey over a period of time in bursts. As explained earlier in Section 6.2.2, the inverse of the estimated variance, $\frac{1}{\hat{h}(t)}$, is selected as the weight and weighted OFR is applied to the candidate model. The right-hand side of Table 6.4 shows the terms selected in decreasing order of ERR.

Table 6.4 Ranking of Terms of Candidate Variance Model selected by OFR and Weighted OFR

Rank	Standard OFR		Weighted OFR	
	Term	ERR (%)	Term	ERR (%)
1	$\hat{h}(t-2)$	90.2449	$\hat{h}(t-2)$	91.9727
2	$e^2(t-2)$	6.3338	$e^2(t-2)$	4.0756
3	$\hat{h}(t-3)$	1.6278	$\hat{h}(t-3)$	1.7373
4	$e^2(t-1)$	0.7204	$e^2(t-1)$	0.7167
5	$e^2(t-5)$	0.0875	1	0.1996
6	1	0.0367	$\hat{h}(t-4)$	0.0324
7	$e^2(t-3)$	0.0258	$e^2(t-4)$	0.0353
8	$\hat{h}(t-4)$	0.0166	$\hat{h}(t-5)$	0.0575
9	$e^2(t-4)$	0.0107	$e^2(t-5)$	0.0073
10	$\hat{h}(t-5)$	0.0127	$e^2(t-5)$	0.0011
11	$\hat{h}(t-1)$	0.0091	$\hat{h}(t-1)$	0.0046

Compared to the term selection results of OFR, the results of WOFR are better since the priority of the constant term has increased.

Moving ahead with WOFR, the ERR cut-off value is set as 0.15% and parameter estimation is carried out only with the selected terms included in the candidate variance model. The parameter estimates of the selected terms are obtained and listed in Table 6.5.

The modelling residuals obtained after parameter estimation using WOFR are shown in Figure 6.3. The effect weighting had on the modelling residuals can be seen – the heteroskedasticity has reduced a lot.

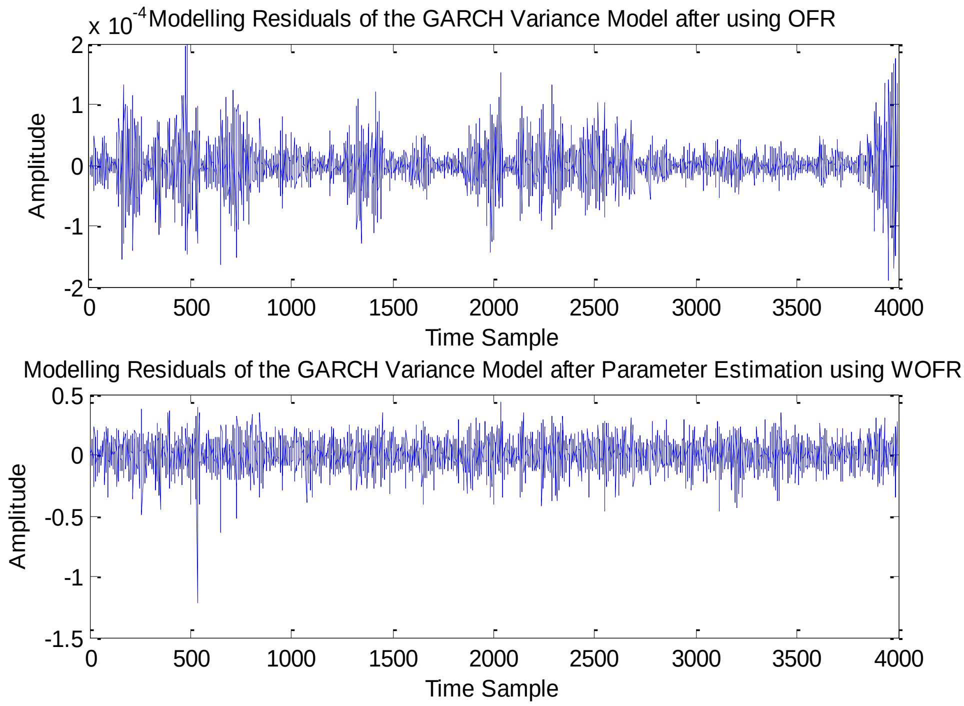


Figure 6.3 Modelling Residuals of the GARCH Variance Model after Parameter Estimation using Standard and Weighted OFR

In an attempt to improve the accuracy of the parameter estimates of the selected GARCH variance model, a noise model with fifteen lagged noise terms is fitted in addition to the terms selected in the variance model. These lagged noise terms can be regarded as proxies for the errors in our variance estimator. WOFR is applied, the noise terms are then recalculated and updated in the candidate model. WOFR is applied again to the updated model, and so forth. This procedure is repeated 50 times (or as long as it takes for the parameter estimates to converge). The parameter estimates of just the selected terms (noise terms not included) are listed in Table 6.5.

Table 6.5 Parameter estimates of selected terms in the variance model using WOFR before and after fitting a noise model

Parameter of the Variance Model	Parameter Estimate without fitting Noise Model	Parameter Estimate after fitting Noise Model	True Coefficient
K	1.4705E-05	1.4593E-05	1.4E-05
G_2	0.3218	0.3400	0.3648
G_3	0.3281	0.3190	0.3520
A_1	0.0686	0.0680	0.0543
A_2	0.2059	0.2045	0.1870
NRMSE	11.4676%	10.7321%	

The Normalised Root Mean Squared Error (NRMSE) between the one-step-ahead variance estimate obtained via WOFR after fitting a noise model to the selected GARCH variance model and the true GARCH(3,2) variance is calculated to be 10.7321%. This is a small improvement over the one-step-ahead variance estimate obtained via WOFR without fitting a noise model to the selected GARCH variance model (NRMSE = 11.4676%).

6.4.3.3 Using Weighted OFR for the Identification of the GARCH Mean and Variance Model when the True Variance and Residuals are Both Unknown – GARCH(3,2) Variance Model

This simulation is the first of three which examine the effectiveness of using WOFR for the identification of the variance model when the true variance and the true residuals of a GARCH model are both unknown.

In this first case, the GARCH mean model is estimated from the given returns using the WOFR method described in Section 3.3.4. Once the GARCH mean model is estimated, the modelling residuals of the GARCH mean model (simply known as the residuals of the GARCH model) are used to generate an ARCH(25) variance estimate. The estimated residuals and the ARCH(25) variance estimate are in turn used to estimate the GARCH variance model using Weighted OFR.

Consider the following GARCH(3,2) model with a non-linear mean

$$y(t) = a_0 + a_1 y(t-1) + a_2 y(t-1) y(t-2) + a_3 y(t-2) + e(t) \qquad (6.14)$$

$$h(t) = K + \sum_{p=1}^{3} G_p h(t-p) + \sum_{q=1}^{2} A_q e^2(t-q) \tag{6.15}$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual and $h(t)$ is the variance, all at an instant in time, t. a_0, a_1, a_2 and a_3 are the parameters of the mean model and K, G_1, G_2, G_3, A_1 and A_2 are the parameters of the variance model. The true values of the parameters are listed in Table 6.6.

Table 6.6 Parameters of the Simulated GARCH(3,2) Model

Parameter of the Mean Model	Value	Parameter of the Variance Model	Value
a_0	0.001	K	1.4E-05
a_1	0.2	G_1	0
a_2	-4	G_2	0.3648
a_3	0.1	G_3	0.3520
		A_1	0.0543
		A_2	0.1870

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. The simulated returns and GARCH(3,2) variance are shown in Figure 6.4. When simulating the returns and the variance from equations (6.14) and (6.15), the residuals, $e(t)$, are modelled as $e(t) = z(t)\sqrt{h(t)}$, where $z(t)$ is a random independent and identically distributed (i.i.d) term that has zero mean and a variance of 1. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10).

A NAR(2,5) model is chosen as the candidate mean model.

$$\begin{aligned}
y(t) = {} & a_1 y(t-1) + a_2 y(t-2) + a_3 y(t-3) + a_4 y(t-4) + a_5 y(t-5) \\
& + a_6 y^2(t-1) + a_7 y^2(t-2) + a_8 y^2(t-3) + a_9 y^2(t-4) \\
& + a_{10} y^2(t-5) + a_{11} y(t-1)y(t-2) + a_{12} y(t-1)y(t-3) \\
& + a_{13} y(t-1)y(t-4) + a_{14} y(t-1)y(t-5) + a_{15} y(t-2)y(t-3) \\
& + a_{16} y(t-2)y(t-4) + a_{17} y(t-2)y(t-5) + a_{18} y(t-3)y(t-4) \\
& + a_{19} y(t-3)y(t-5) + a_{20} y(t-4)y(t-5) + a_{21}
\end{aligned} \tag{6.16}$$

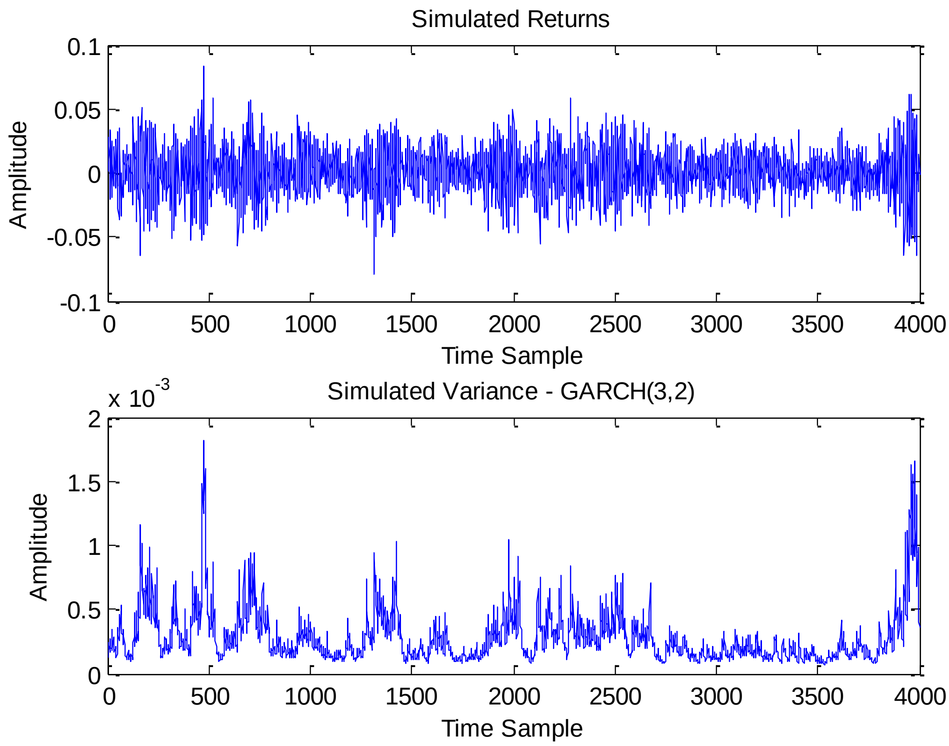


Figure 6.4 Simulated Returns and Variance for GARCH(3,2) Model

WOFR is carried out to select the terms in the GARCH mean model and to calculate the estimates of the respective coefficients. The terms selected, coefficient estimates and true values of the coefficients are listed in Table 6.7.

Table 6.7 Parameter Estimates of Selected Mean Model after 10 Iterations of WOFR

Term of the Mean Model	Coefficient Estimate	True Coefficient
1	9.3712E-04	0.001
$y(t-1)$	0.1981	0.2
$y(t-1)y(t-2)$	-4.2947	-4
$y(t-2)$	0.0788	0.1

The modelling residuals of the estimated GARCH mean model, $\hat{e}(t)$, are calculated. Using maximum likelihood, an ARCH(25) model is fit to the estimated residuals and an estimate of the GARCH variance, $\hat{h}(t)$, is obtained.

A GARCH(5,5) model is selected as the candidate variance model. The ARCH(25) variance estimate, $\hat{h}(t)$, is defined as the output. The lagged estimated squares residuals, $\hat{e}^2(t-i)$, and the lagged values of the estimated variance, $\hat{h}(t-j)$, are set as the inputs.

OFR is applied to the candidate variance model. The left-hand side of Table 6.8 shows the terms selected in decreasing order of ERR. The highlighted terms represent the terms present in the original GARCH(3,2) variance model. It can be seen that most of the terms present in the original model except the constant term have been selected.

Next, the inverse of the ARCH(25) variance estimate, $\frac{1}{\hat{h}(t)}$, is selected as the weight and weighted OFR is applied to the candidate variance model. The right-hand side of Table 6.8 shows the terms of the candidate variance model selected in decreasing order of ERR.

Table 6.8 Ranking of Terms of Candidate Variance Model selected by Standard and Weighted OFR

Rank	Standard OFR		Weighted OFR	
	Term	ERR (%)	Term	ERR (%)
1	$\hat{h}(t-2)$	90.2256	$\hat{h}(t-2)$	91.9489
2	$e^2(t-2)$	6.3679	$e^2(t-2)$	4.1074
3	$\hat{h}(t-3)$	1.6216	$\hat{h}(t-3)$	1.7460
4	$e^2(t-1)$	0.7334	$e^2(t-1)$	0.7288
5	$e^2(t-5)$	0.0949	1	0.2007
6	1	0.0356	$e^2(t-5)$	0.0363
7	$e^2(t-3)$	0.0267	$e^2(t-3)$	0.0214
8	$\hat{h}(t-4)$	0.0102	$\hat{h}(t-4)$	0.0127
9	$e^2(t-4)$	0.0085	$e^2(t-4)$	0.0292
10	$\hat{h}(t-5)$	0.0130	$\hat{h}(t-5)$	0.0292
11	$\hat{h}(t-1)$	0.0046	$\hat{h}(t-1)$	0.0022

The highlighted terms represent the terms present in the original GARCH(3,2) model. Once again, WOFR has improved term selection. The ERR cut-off value is set as 0.15% and parameter estimation is carried out only with the selected terms included in the candidate variance model. The parameter estimates of the selected terms are obtained and listed in Table 6.9.

In an attempt to improve the accuracy of the parameter estimates of the variance model, a noise model with fifteen lagged noise terms is then fitted in addition to the terms selected from the candidate GARCH variance model. The parameter estimates of just the selected terms are listed in Table 6.9.

Table 6.9 Parameter Estimates of Selected Terms in the Variance Model using WOFR before and after Fitting a Noise Model – GARCH(3,2) Model

Parameter of the Variance Model	Parameter Estimate before Fitting a Noise Model	Parameter Estimate after Fitting a Noise Model	True Coefficient
K	1.2791E-05	1.2456E-05	1.4E-05
G_2	0.3210	0.3405	0.3648
G_3	0.3276	0.3173	0.3520
A_1	0.0694	0.0687	0.0543
A_2	0.2073	0.2059	0.1870
NRMSE	11.6934%	10.9902%	

The Normalised Root Mean Squared Error (NRMSE) between the one-step-ahead variance estimate obtained via WOFR after fitting a noise model to the selected GARCH variance model and the true GARCH(3,2) variance is calculated to be 10.9902%. This is an improvement over the one-step-ahead variance estimate obtained via WOFR without fitting a noise model to the selected GARCH variance model (NRMSE = 11.6934%).

The proposed method works well in identifying and estimating all the correct terms in the linear GARCH(3,2) variance model and fitting a noise model improves the accuracy of the variance estimate.

6.4.3.4 Using Weighted OFR for the Identification of the GARCH Mean and Variance Model when the Variance and Residuals are Both Unknown – GARCH(1,5) Variance Model

Now consider the following GARCH(1,5) model with a non-linear mean

$$y(t) = a_0 + a_1 y(t-1) + a_2 y(t-1)y(t-2) + a_3 y(t-2) + e(t) \qquad (6.17)$$

$$h(t) = K + G_1 h(t-1) + \sum_{q=1}^{5} A_q e^2(t-q) \qquad (6.18)$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual and $h(t)$ is the variance, all at an instant in time, t. a_0, a_1, a_2 and a_3 are the parameters of the mean model and $K, G_1, A_1, A_2, A_3, A_4$ and A_5 are the parameters of the variance model. The true values of the parameters are listed in Table 6.10.

Table 6.10 Parameters of the Simulated GARCH(1,5) Model

Parameter of the Mean Model	Value	Parameter of the Variance Model	Value
a_0	0.001	K	4.25E-06
a_1	0.2	G_1	0.6388
a_2	-4	A_1	0.0431
a_3	0.1	A_2	0.0401
		A_3	0.1025
		A_4	0.0955
		A_5	0.0631

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. The simulated returns and GARCH(1,5) variance are shown in Figure 6.5. For a GARCH(*p*,*q*) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10).

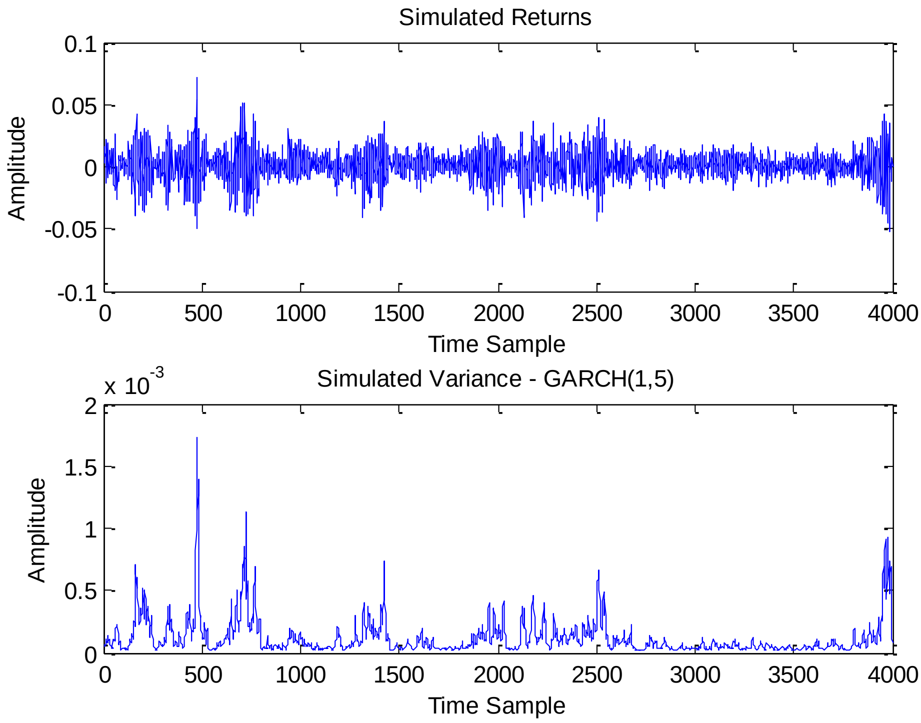


Figure 6.5 Simulated Returns and Variance for GARCH(1,5) Model

A NAR(2,5) model is chosen as the candidate model (See equation (6.16)).

WOFR is carried out to select the terms in the GARCH mean model and to calculate the estimates of the respective coefficients. The terms selected, coefficient estimates and true values of the coefficients are listed in Table 6.11.

Table 6.11 Parameter Estimates of Selected Mean Model after 10 Iterations of WOFR

Term of the Mean Model	Coefficient Estimate	True Coefficient
1	9.8352E-04	1E-03
$y(t-1)$	0.2001	0.2
$y(t-1)y(t-2)$	-4.3674	-4
$y(t-2)$	0.0810	0.1

The modelling residuals of the estimated GARCH mean model, $\hat{e}(t)$, are calculated. Using maximum likelihood, an ARCH(25) model is fit to the estimated residuals and an estimate of the GARCH variance, $\hat{h}(t)$, is obtained. A GARCH(5,5) model is selected as the candidate variance model. The ARCH(25) variance estimate, $\hat{h}(t)$, is defined as the output. The lagged estimated squared residuals, $\hat{e}^2(t-i)$, and the lagged values of the estimated variance, $\hat{h}(t-j)$, are set as the inputs.

OFR is applied to the candidate variance model. The left-hand side of Table 6.12 shows the terms selected in decreasing order of ERR. The highlighted terms represent the terms present in the original GARCH(1,5) variance model. It can be seen that most of the terms present in the original variance model except the constant term have been selected. Next, $\frac{1}{\hat{h}(t)}$, is selected as the weight and WOFR is applied to the candidate variance model. The right-hand side of Table 6.12 shows the terms selected in decreasing order of ERR.

Table 6.12 Ranking of terms of candidate variance model selected by Standard and Weighted OFR

Rank	Standard OFR		Weighted OFR	
	Term	ERR (%)	Term	ERR (%)
1	$\hat{h}(t-1)$	97.6930	$\hat{h}(t-1)$	97.3557
2	$\hat{e}^2(t-4)$	0.5916	$\hat{e}^2(t-4)$	0.4207
3	$\hat{e}^2(t-3)$	0.5144	$\hat{e}^2(t-3)$	0.4600
4	$\hat{e}^2(t-1)$	0.3478	$\hat{e}^2(t-2)$	0.3520
5	$\hat{e}^2(t-2)$	0.2177	$\hat{e}^2(t-1)$	0.2725
6	$\hat{e}^2(t-5)$	0.0303	1	0.1686
7	$\hat{h}(t-2)$	0.0377	$\hat{e}^2(t-5)$	0.0867
8	$\hat{h}(t-4)$	0.0459	$\hat{h}(t-2)$	0.0874
9	$\hat{h}(t-5)$	0.0585	$\hat{h}(t-4)$	0.0508
10	1	0.0198	$\hat{h}(t-5)$	0.0318
11	$\hat{h}(t-3)$	0.0030	$\hat{h}(t-3)$	0.0208

The highlighted terms represent the terms present in the original GARCH(1,5) variance model. Once again, WOFR has improved term selection. The ERR cut-off value is set as 0.08% and parameter estimation is carried out only with the selected terms included in the candidate variance model. The parameter estimates of the selected terms are obtained and listed in Table 6.13.

In an attempt to improve the accuracy of the parameter estimates of the variance model, a noise model with fifteen lagged noise terms is fit in addition to the terms selected from the candidate GARCH variance model. The parameter estimates of the just the selected terms are listed in Table 6.13.

Table 6.13 Parameter Estimates of Selected Terms in the Variance Model using WOFR before and after Fitting a Noise Model – GARCH(1,5) Model

Parameter of the Variance Model	Parameter Estimate before Fitting a Noise Model	Parameter Estimate after Fitting a Noise Model	True Coefficient
K	3.3969E-06	3.1735E-06	4.25E-06
G_1	0.6399	0.6643	0.6388
A_1	0.0492	0.0495	0.0431
A_2	0.0545	0.0517	0.0401
A_3	0.0830	0.0785	0.1025
A_4	0.1010	0.0958	0.0955
A_5	0.0363	0.0292	0.0631
NRMSE	8.9028%	8.6772%	

The Normalised Root Mean Squared Error (NRMSE) between the one-step-ahead variance estimate obtained via WOFR after fitting a noise model to the selected GARCH variance model and the true GARCH(1,5) variance is calculated to be 8.6772%. This is an improvement over the one-step-ahead variance estimate obtained via WOFR without fitting a noise model to the selected GARCH variance model (NRMSE = 8.9028%).

The proposed method works well in identifying and estimating all the correct terms in the linear GARCH(1,5) variance model and fitting a noise model improves the accuracy of the variance estimate.

6.4.3.5 Using Weighted OFR for the Identification of the Variance Model when the Variance and Residuals are Both Unknown – Non-Linear GARCH Variance Model

Finally, consider the following non-linear GARCH model with a non-linear mean

$$y(t) = a_0 + a_1 y(t-1) + a_2 y(t-1)y(t-2) + a_3 y(t-2) + e(t) \qquad (6.19)$$

$$h(t) = K + G_1 h(t-1) + A_1 e^2(t-1) + B_1 h(t-1)e^2(t-1) + B_2 h^2(t-1) + B_3 e^4(t-1) \qquad (6.20)$$

where $y(t)$ is the excess return, $e(t)$ is the modelling residual and $h(t)$ is the variance, all at an instant in time, t. a_0, a_1, a_2 and a_3 are the parameters of the mean model and K, G_1, A_1, B_1, B_2 and B_3 are the parameters of the variance model. The true values of the parameters are listed in Table 6.14.

Table 6.14 Parameters of the Simulated Non-Linear GARCH Model

Parameter of the Mean Model	Value	Parameter of the Variance Model	Value
a_0	0.001	K	1E-05
a_1	0.2	G_1	0.92
a_2	-4	A_1	0.05
a_3	0.1	B_1	2
		B_2	2
		B_3	5

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10). The simulated returns and Non-Linear GARCH variance are shown in Figure 6.6.

A NAR(2,5) model is chosen as the candidate model (See equation (6.16)).

WOFR is carried out to select the terms in the GARCH mean model and to calculate the estimates of the respective coefficients. The terms selected, coefficient estimates and true values of the coefficients are listed in Table 6.15.

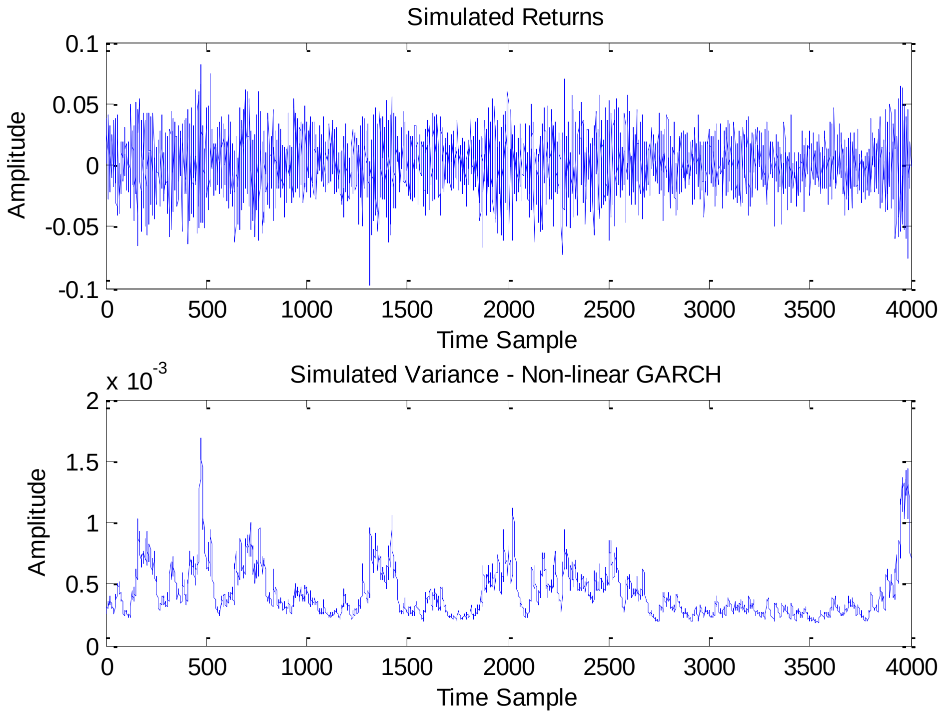


Figure 6.6 Simulated Returns and Variance for Non-Linear GARCH Model

Table 6.15 Parameter Estimates of Selected Mean Model after 10 Iterations of WOFR

Term of the Mean Model	Coefficient Estimate	True Coefficient
1	Not Selected	0.001
$y(t-1)$	0.2004	0.2
$y(t-1)y(t-2)$	-3.8266	-4
$y(t-2)$	0.0838	0.1

The modelling residuals of the estimated GARCH mean model, $\hat{e}(t)$, are calculated. Using maximum likelihood, an ARCH(25) model is fit to the estimated residuals and an estimate of the GARCH variance, $\hat{h}(t)$, is obtained.

A GARCH(5,5) model with the non-linear terms present in the original model is selected as the candidate variance model. The ARCH(25) variance estimate, $\hat{h}(t)$, is defined as the output. The lagged estimated squares residuals, $\hat{e}^2(t-i)$, and the lagged values of the estimated variance, $\hat{h}(t-j)$, are set as the inputs.

OFR is applied to the candidate variance model. Table 6.16 shows the terms selected in decreasing order of ERR. The highlighted terms represent the terms present in the original Non-Linear GARCH variance model. It can be seen that most of the terms present in the original variance model haven't been selected.

Next, inverse of the estimated variance, $\frac{1}{\hat{h}(t)}$, is selected as the weight and weighted OFR is applied to the candidate model. Table 6.16 shows the terms selected in decreasing order of ERR.

Table 6.16 Ranking of terms of Candidate Variance Model Selected by Standard and Weighted OFR

Rank	Standard OFR		Weighted OFR	
	Term	ERR (%)	Term	ERR (%)
1	$\hat{h}(t-1)$	96.3476	$\hat{h}(t-1)$	96.6015
2	$\hat{e}^2(t-1)$	0.8773	$\hat{e}^2(t-1)$	0.7576
3	$\hat{h}(t-2)$	0.7072	$\hat{h}(t-2)$	0.6056
4	$\hat{e}^2(t-2)$	0.2371	$\hat{e}^2(t-2)$	0.2154
5	$\hat{h}(t-3)$	0.3274	$\hat{h}(t-3)$	0.2534
6	$\hat{e}^2(t-5)$	0.0768	1	0.0784
7	1	0.0212	$\hat{e}^2(t-5)$	0.0395
8	$\hat{e}^2(t-3)$	0.0265	$\hat{e}^2(t-3)$	0.0315
9	$\hat{h}(t-4)$	0.0411	$\hat{h}(t-4)$	0.0367
10	$\hat{e}^2(t-4)$	0.0336	$\hat{e}^2(t-4)$	0.0480
11	$\hat{h}(t-5)$	0.0108	$\hat{h}(t-5)$	0.0053
12	$\hat{h}(t-1)\hat{e}^2(t-1)$	5.9741E-04	$\hat{h}(t-1)\hat{e}^2(t-1)$	1.1438E-04
13	$\hat{h}^2(t-1)$	1.5531E-04	$\hat{e}^4(t-1)$	8.2449E-06
14	$\hat{e}^4(t-1)$	7.3592E-06	$\hat{h}^2(t-1)$	1.1066E-07

The highlighted terms represent the terms present in the original Non-Linear GARCH variance model. WOFR does not seem to have improved term selection. This can be attributed to the fact that the ARCH(25) estimate of the variance is linear and not accurate enough to capture the non-linearity present in the true GARCH variance. This, in turn, suggests that the suggested approach does not work nearly so well when the true GARCH variance process is unknown.

To fully demonstrate the accuracy of the GARCH variance estimate obtained if the algorithm introduced in this chapter is applied to the given non-linear GARCH variance model, the ERR cut-off value for the selection of the terms from the candidate GARCH variance model is set as 0.05% and parameter estimation is carried out only with the selected terms included in the candidate GARCH variance model.

In an attempt to improve the accuracy of the parameter estimates of the variance model, a noise model with fifteen lagged noise terms is fit in addition to the terms selected from the candidate GARCH variance model. The Normalised Root Mean Squared Error (NRMSE) between the one-step-ahead variance estimate obtained via WOFR after fitting a noise model to the selected GARCH variance model and the true Non-Linear GARCH variance is calculated to be 20.1294%. This is an improvement over the one-step-ahead variance estimate obtained via WOFR without fitting a noise model to the selected GARCH variance model (NRMSE = 22.0539%) but is still not accurate enough.

6.5 Conclusions

This chapter introduces the financial variance model and the current methods of the estimation of the GARCH variance model, namely, maximum likelihood. The problems involving maximum likelihood are highlighted and the NARMAX methodology for system identification is introduced as a viable solution. Term selection and parameter estimation of a GARCH variance model is easily carried out with the use of Weighted OFR. Since the true variance of the returns of an asset is unobservable, an estimate of the true GARCH variance is derived from the squared residuals. This estimate of the GARCH variance is then used with WOFR to select the terms and estimate the parameters of the GARCH variance model.

The simulations demonstrate how the presence of heteroskedasticity affects term selection and parameter estimation whilst estimating the variance model using OFR and how WOFR is used to achieve accurate results when some or all of the properties of the true GARCH model are known and when some or all of the properties of the true GARCH model are unknown (estimates are used). The simulations also suggest the following important conclusion: *the estimate of the true GARCH variance is linear and works well if the true GARCH variance is linear, but is unsatisfactory when the true GARCH variance is non-linear*. This in turn indicates the need to derive a more accurate estimate of the true GARCH variance that contains some or all of the non-linearity present in the true GARCH variance. This is explored in the next chapter.

Chapter 7

Maximum Likelihood Estimation of Non-Linear Variance in GARCH Models using Radial Basis Functions

7.1 Introduction

Using the Weighted Orthogonal Forward Regression (WOFR) method (as explained in Section 6.4.2) of system identification to model the variance of a GARCH model requires the input(s) and output(s) of the GARCH model to be observable. This is a major setback of the method, which is why maximum likelihood has always been used to model the variance in a GARCH model (Bollerslev, 1986). In order to work around this problem and to use WOLS for the identification of the variance in a GARCH model, a linear estimate of the variance is used in place of the true variance of the GARCH model, and term selection and parameter estimation can be successfully carried out, as shown in Chapter 6.

A linear estimate of the variance of a GARCH model is insufficient for accurate term selection and parameter estimation when the underlying true variance of the GARCH model is non-linear in nature (as shown in Chapter 6, Section 6.4.3.5). This creates the need for an accurate non-linear variance estimate that can capture the non-linearity present in the true variance of the GARCH model, and should result in accurate term selection and parameter estimation when used to predict the GARCH variance model using WOLS.

Since their introduction, Radial Basis Functions (Broomhead and Lowe, 1988), have been used to approximate linear and non-linear multivariate functions. The GARCH system will be modelled as a non-linear mapping of the inputs to the outputs with the help of non-linear basis functions.

Fitting an RBF model requires the inputs and outputs of the system to be modelled to be observable and measurable (Orr, 1996). Since the output of a GARCH variance model (i.e. the true variance) is unobservable (Bollerslev, 1986), an RBF model cannot be used to estimate a GARCH variance model. Instead, an RBF model is used as a means to obtain a non-linear estimate of the variance of a GARCH model that is more accurate than a linear estimate of the variance of the GARCH model. This non-linear variance estimate may then be used to identify the GARCH variance model using WOLS.

The idea is to fit a Radial Basis Function (RBF) model to a linear estimate of the variance of a GARCH model that needs to be identified. The coefficients of the obtained RBF model are then further optimised to maximise the log-likelihood of the one-step-ahead estimate of the RBF model in an attempt to obtain a non-linear estimate of the GARCH variance that is more accurate than a linear ARCH estimate of the GARCH variance.

Broadly classified, two methods of maximisation of the likelihood are tested – constrained and unconstrained maximisation. Unconstrained maximisation implies that the coefficients of the RBF model (fitted to the linear ARCH estimate of the GARCH variance) are optimised around their initial starting values in an attempt to find the maximum of the log-likelihood of the variance estimate. This log-likelihood value is more likely a local maximum than a global maximum.

Constrained maximisation implies that the coefficients of the RBF model (fitted to the linear ARCH estimate of the GARCH variance) are optimised within user-specified constraints in an attempt to find the global maximum of the log-likelihood of the variance estimate. Several starting points are uniformly selected within the given constraints and the coefficients of the RBF model are optimised around the starting points whilst staying within the specified parameter constraints. A shortcoming of this method is that the parameter constraints within which the coefficients of the obtained RBF model need to be optimised are not known. Since the terms of the model are non-linear in nature and are complex non-linear functions of the inputs, the parameter bounds cannot be calculated as well. Hence, for the purpose of this chapter, very tight constraints are set within which an attempt to find the global maximum of the log-likelihood may or may not be fruitful.

This method of creating a non-linear estimate of the variance is theoretically advantageous compared to fitting one of the non-linear variance models, like an EGARCH or an NGARCH model (explained in Chapter 2), since no assumption about the structure of the variance model needs to be made. The proposed algorithm is tested on three different non-linear GARCH models, and for each example, four different non-linear estimates of the GARCH variance are generated. The accuracy of an estimate of the

GARCH variance (linear or non-linear) is determined by calculating the Normalised Root Mean Square Error (NRMSE) between the true/simulated GARCH variance and the estimate of the GARCH variance. The accuracy of all the non-linear estimates of the GARCH variance is compared to each other and to a linear ARCH estimate of the same GARCH variance to assess the performance of the RBF approach.

The purpose of this chapter is to introduce Radial Basis Functions and to introduce and test the new algorithm that uses RBF models to create a non-linear estimate of the GARCH variance from the squared residuals and a linear ARCH estimate of the GARCH variance.

Section 7.2 introduces Radial Basis Functions and describes how RBFs are used to approximate and estimate non-linear multivariate functions. Section 7.3 explains in detail the type of basis function used in this Chapter. Section 7.4 explains in detail the algorithm used to train an RBF model. Section 7.5 introduces and explains in detail an algorithm to create a non-linear estimate of a GARCH variance by using maximum likelihood to optimise the coefficients of an RBF model fit to a linear ARCH estimate of the GARCH variance. Section 7.6 puts the proposed algorithm to test on three different simulated non-linear GARCH variance models and compares the results of the simulations. Section 7.7 concludes the chapter.

7.2 Introduction to Radial Basis Functions

Radial basis functions (RBFs) were first developed to interpolate a set of data points in a multidimensional space (Powell, 1987). Since then, several advancements have been made and numerous methods have been developed that enable the use of radial basis functions to approximate non-linear multivariate functions using only the observed input(s) and output(s) of the system (Orr, 1996).

Consider a system with inputs,

$$X = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ u_{k1} & u_{k2} & \cdots & u_{kn} \end{bmatrix} \tag{7.1}$$

where each row represents a different input, k is the total number of inputs and n is the total number of samples.

The output is represented as $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$. For any integers i and j,

$$x_i = \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{ki} \end{bmatrix} \; and \; c_j = \begin{bmatrix} u_{1j} \\ u_{2j} \\ \vdots \\ u_{kj} \end{bmatrix} \text{ where } 1 \le i \le n, 1 \le j \le n \qquad (7.2)$$

A function $f(X)$ must be found that maps each input to the output such that

$$y_i = f(x_i) \qquad \text{where } i = 1,2,\dots,n \qquad (7.3)$$

According to Powell (1987), an exact mapping can be achieved by using n basis functions of the form $\varphi(\|x_i - c_j\|)$, where $\varphi(.)$ is some non-linear function termed as the basis function, x_i is the input data coordinate, c_j is the coordinate representing the centre of the basis function, and || || denotes the Euclidean distance between x_i and c_j.

The Euclidean distance, between x_i and c_j is calculated as

$$\|x_i - c_j\| = \sqrt{(u_{1i} - u_{1j})^2 + (u_{2i} - u_{2j})^2 + \cdots + (u_{ki} - u_{kj})^2} \qquad (7.4)$$

Hence, in equation (7.2), $f(x_i)$ can be represented as a linear combination of these basis functions as

$$y_i = f(x_i) = w_0 + \sum_{j=1}^{n} w_j \, \varphi_{ij}(\|x_i - c_j\|) \qquad (7.5)$$

where w_j denotes the weights of the respective basis functions and w_0 is a constant term. In matrix form, equation (7.5) can be written as

$$Y = \Phi W \qquad (7.6)$$

$$where \; Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ 1 & \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nn} \end{bmatrix}, \quad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

For a number of different types of functions, the matrix Φ is non-singular if the data samples are distinct. This implies that a least squares estimate of W in equation (7.6) can be calculated.

$$\widehat{W} = (\Phi^T \Phi)^{-1} \Phi^T Y \tag{7.7}$$

Figure 7.1 shows a diagrammatic representation of the Radial Basis Function network with N input samples, m basis functions, one input layer, one hidden layer and one output layer.

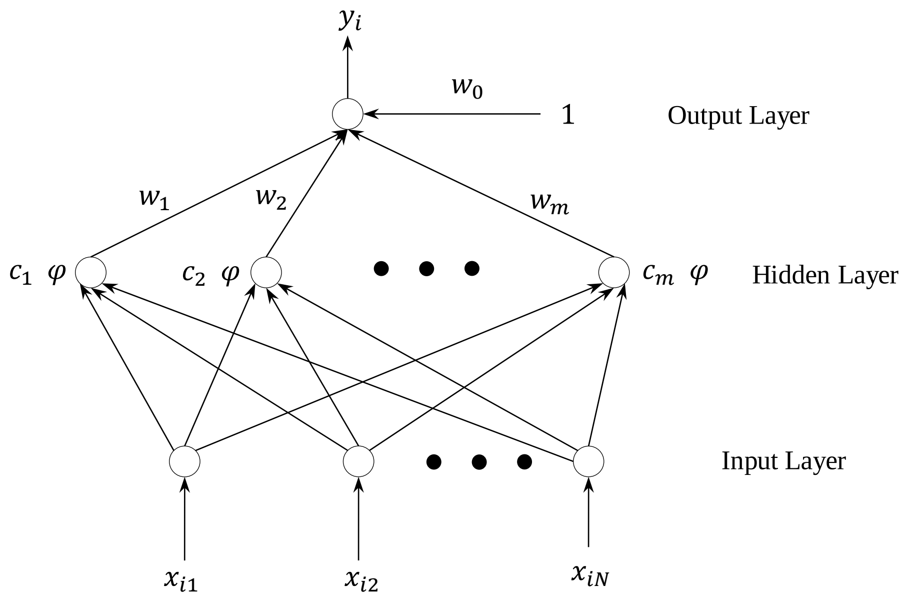


Figure 7.1 Architecture of a Radial Basis Function Network

In reality, training an RBF model to the whole data set is not ideal since this implies that the model would predict poorly due to any underlying noise in the data being modelled as well. To avoid this, several modifications are made to the modelling/training procedure.

1. The data set to be modelled is split into 2 parts – the training set and the validation set. The training set is used to generate the RBF model and the validation set is used to test the predictive ability of the generated model. The Normalised Root Mean Square Error (NRMSE) of the model predicted output vs. the true output is calculated for both data sets. The model that yields the least NRMSE for the validation set is selected.

2. The number of basis functions, m, is lesser than the number of data points, n.

3. A bias term is included in the linear sum in equation (7.7) whilst calculating the weights (coefficients) of the RBF model. This is done to penalise large weights.

$$W = (\Phi^T \Phi + \lambda \mathrm{I})^{-1} \Phi Y \tag{7.8}$$

where λ is the regularisation parameter used to control the trade-off between fitting the data and penalising large weights.

7.3 Gaussian Radial Basis Function

There exist many different types of basis functions such as Gaussian, Thin Plate Spline, Multiquadric, Inverse Quadratic and Inverse Multiquadric. For the purpose of this chapter, the choice of the basis function was not that important because upon comparison, they all worked well. In this chapter, the Gaussian basis function has been used to generate RBF models and is explained in detail.

The Gaussian basis function is one of the most commonly used basis functions.

$$\varphi_{ij}(\|x_i - c_j\|) = exp\left(\frac{-\|x_i - c_j\|^2}{2\sigma^2}\right) \tag{7.9}$$

where

- x_i is the input data coordinate, c_j is the coordinate representing the centre of the basis function,
- σ is the width of the basis function that controls the smoothness properties of the model, and

For improved training performance, multiple widths can be used, one for each input.

$$\varphi_{ij}(\|x_i - c_j\|) = exp\left(\frac{-(u_{1i} - u_{1j})^2}{2\sigma_{u1}{}^2} - \cdots - \frac{(u_{k\mathrm{i}} - u_{k\mathrm{j}})^2}{2\sigma_{uk}{}^2}\right) \tag{7.10}$$

where the inputs, u_1, u_2, …, u_k, are as defined in equations (7.1) and (7.3), and σ_{u1}, σ_{u2}, … , σ_{uk} are the widths of the gaussian basis functions for the respective inputs.

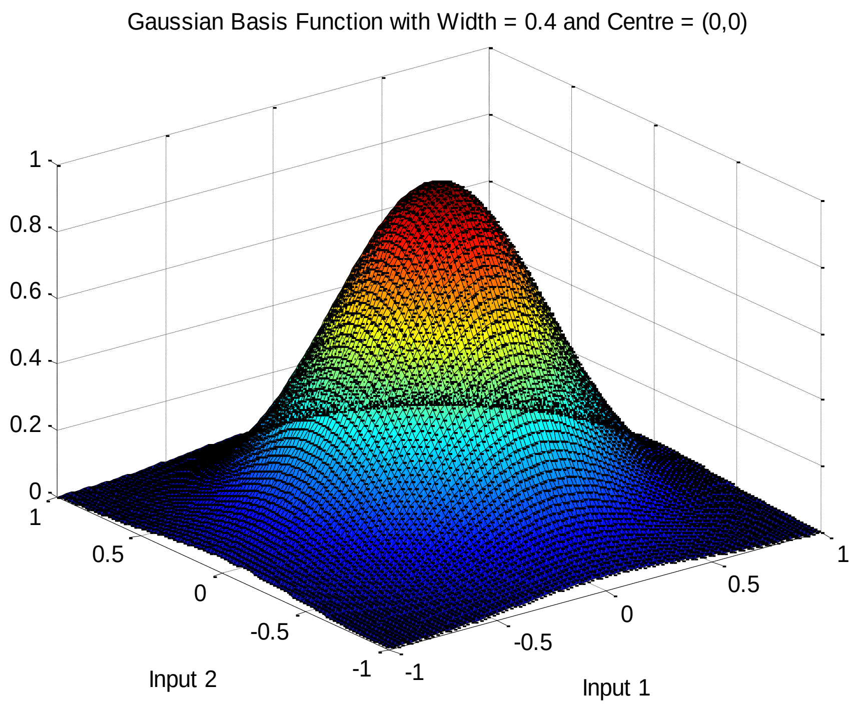


Figure 7.2 Gaussian Basis Function with Centre at (0,0) for 2 Inputs

7.4 Training a Radial Basis Function Network/Model

For the purpose of this chapter, a simple, single layer Radial Basis Function Network is used to map two inputs to one output. While training the network, three sets of parameters need to be optimised if using the Gaussian basis function, namely, the centres of the basis functions, the widths of the basis functions and the regularisation parameter.

7.4.1 Initialisation

Consider the original data set to have N samples. The data set (input(s) and output(s)) is split into two halves – training set and validation set. The training set consists of n samples and the validation set consists of $(N - n)$ samples.

X represents the input matrix, consisting of 2 different input series, u_1 and u_2, and Y represents the output matrix, consisting of 1 output.

$$X_{train} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \end{bmatrix}, \quad X_{val} = \begin{bmatrix} u_{1(n+1)} & u_{1(n+2)} & \cdots & u_{1N} \\ u_{2(n+1)} & u_{2(n+2)} & \cdots & u_{2N} \end{bmatrix},$$

$$Y_{train} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad Y_{val} = \begin{bmatrix} y_{n+1} \\ y_{n+2} \\ \vdots \\ y_N \end{bmatrix}$$

The input matrix can also be written as

$$X = [x_1 \quad \cdots \quad x_n] \quad where \quad x_i = \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix}, \quad i = 1,2,\dots,N$$

Initialise the regularisation parameter, λ, to a reasonable value, say, 1E-08. This value is carefully selected after testing a range of different values using various simulations.

If using Gaussian basis functions, initialise the widths for each set of inputs, σ_{u1} and σ_{u2} to

$$\sigma_{u1} = \frac{u_{1max} - u_{1min}}{\sqrt{2 \times n}}, \qquad \sigma_{u2} = \frac{u_{2max} - u_{2min}}{\sqrt{2 \times n}} \tag{7.11}$$

where u_{1max} and u_{1min} are the maximum and minimum values, respectively, of the input, u_1, amongst the given n samples in the training set. u_{2max} and u_{2min} are the maximum and minimum values, respectively, of the input, u_2, amongst the given n samples in the training set.

7.4.2 Selection of Centres

Step 1. The RBF model is trained using the training set, Y_{train}. The number of basis functions is initialised to be the same as the number of data samples, n. All the data samples of the input, X, are set as the centres, c_j, of the basis functions.

$$c_j = \begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix}, \quad j = 1,2,\dots,N \tag{7.12}$$

$$Y_{train} = \Phi_{train} W \tag{7.13}$$

$$where \quad \Phi_{train} = \begin{bmatrix} 1 & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ 1 & \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nn} \end{bmatrix}, \qquad W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix},$$

x_i is the input data coordinate, c_j is the coordinate representing the centre of the basis function, and W is the weight sequence. φ_{ij} is calculated as shown in equation (7.10).

Step 2. Orthogonal Forward Regression with ERR (Chen *et al.*, 1989) (explained in Section 3.3.3) is used on equation (7.13) to determine the key centres of the basis functions since each column in the matrix, Φ_{train}, represents the value of a basis function at a particular centre.

Step 3. The ERR cut-off value is set to the upper limit of a user-specified range, and the number of centres, and hence, number of basis functions is reduced to M. Equation (7.13) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1M} \\ 1 & \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2M} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nM} \end{bmatrix} \begin{bmatrix} w_0{}' \\ w_1{}' \\ w_2{}' \\ \vdots \\ w_M{}' \end{bmatrix} \tag{7.14}$$

$$Y_{train} = \Phi_{train}{}' W\,' \tag{7.15}$$

Step 4. The optimal weight matrix, $W\,'$ is calculated using

$$W\,' = [(\Phi_{train}{}')^T \Phi_{train}{}' + \lambda \mathrm{I}]^{-1} \Phi_{train}{}' Y \tag{7.16}$$

where λ is the regularisation parameter.

Step 5. Next, the performance of the above obtained RBF model is tested on the validation set. With the centres as selected in step 3, the one-step ahead estimates of the output for the validation set are calculated as follows:

$$Y_{val}{}' = \Phi_{val} W\,' \tag{7.17}$$

$$\begin{bmatrix} y_{n+1}{}' \\ y_{n+2}{}' \\ \vdots \\ y_N{}' \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{(n+1)1} & \varphi_{(n+1)2} & \cdots & \varphi_{(n+1)M} \\ 1 & \varphi_{(n+2)1} & \varphi_{(n+2)2} & \cdots & \varphi_{(n+2)M} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{NM} \end{bmatrix} \begin{bmatrix} w_0{}' \\ w_1{}' \\ w_2{}' \\ \vdots \\ w_M{}' \end{bmatrix} \tag{7.18}$$

Step 6. The Normalised Root Mean Square Error (NRMSE) between the actual output, Y_{val}, and the predicted output, $Y_{val}{}'$, is calculated.

$$NRMSE(Y_{val}, Y_{val}{}') = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N-n} (y_{n+i} - y_{n+i}{}')}}{SSD(Y_{val})} \tag{7.19}$$

$$where \quad SSD(Y_{val}) = Sample\ Standard\ Deviation\ of\ Y_{val}$$

$$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-n} (y_{n+i} - \overline{y_{val}})^2}$$

$$and \qquad \overline{y_{val}} = mean\ of\ Y_{val} = \frac{1}{N} \sum_{i=1}^{N-n} y_{n+i}$$

Step 7. The ERR cut-off value is gradually lowered to the lower limit of the user-specified range, in increments. As the ERR cut-off value is reduced, the number of centres being included in the model, M, increases. Steps 3 to 6 are carried out for each iteration. The NRMSE of the validation set is calculated for each iteration and the number of centres, m, yielding the least NRMSE is selected.

7.4.3 Selection of Width(s) of the Gaussian Basis Functions

Step 8. The widths of the Gaussian basis functions, σ_{u1} and σ_{u2}, are to be optimised to values that yield the least NRMSE for the validation set. The widths are optimised within a user-specified range, one at a time. First, initialise σ_{u1} to the lower limit of the user-specified range. The value of σ_{u2} kept constant (as calculated in equation (7.11)).

Step 9. Once the centres of the basis functions have been selected, the number of basis functions is m. The RBF model can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1m} \\ 1 & \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nm} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{7.20}$$

$$Y_{train} = \Phi_{train}{}' W\,' \tag{7.21}$$

φ_{ij} is calculated as shown in equation (7.10).

Step 10. The optimal weight matrix, $W\,'$ is calculated using

$$W\,' = [(\Phi_{train}{}')^T \Phi_{train}{}' + \lambda \mathrm{I}]^{-1} \Phi_{train}{}' Y \tag{7.22}$$

where λ is the regularisation parameter.

Step 11. Next, the performance of the above obtained RBF model is tested on the validation set. The one-step ahead estimates of the output for the validation set are calculated as follows:

$$Y_{val}{}' = \Phi_{val} W\,' \tag{7.23}$$

$$\begin{bmatrix} y_{n+1}{}' \\ y_{n+2}{}' \\ \vdots \\ y_N{}' \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{(n+1)1} & \varphi_{(n+1)2} & \cdots & \varphi_{(n+1)m} \\ 1 & \varphi_{(n+2)1} & \varphi_{(n+2)2} & \cdots & \varphi_{(n+2)m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{N1} & \varphi_{N2} & \cdots & \varphi_{Nm} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{7.24}$$

Step 12. The Normalised Root Mean Square Error (NRMSE) between the one-step-ahead estimate of the output, Y_{val}', and the actual output, Y_{val}, is calculated using equation (7.19).

Step 13. The value of σ_{u1} is gradually increased to the upper limit of the user-specified range, in increments. Steps 9 to 12 are carried out for each iteration. The NRMSE of the validation set is calculated for each iteration and the width, σ_{u1}', yielding the least NRMSE is selected.

Step 14. Keeping σ_{u1} constant as the calculated optimal value, σ_{u1}', steps 9 to 12 are carried out for σ_{u2} for a user-specified range. The NRMSE of the output vs. the one-step-ahead estimate of the RBF model for the validation set is calculated for each iteration and the width, σ_{u2}', yielding the least NRMSE is selected.

7.4.4 Selection of Regularisation Parameter

Step 15. The centres have been selected and the widths have been optimised to σ_{u1}' and σ_{u2}'. Using these optimised values, steps 9 to 12 are carried out for λ for a user-specified range of regularisation parameters. The NRMSE of the output vs. the one-step-ahead estimate of the RBF model for the validation set is calculated for each iteration and the regularisation parameter, λ', yielding the least NRMSE is selected.

7.5 Generating a Non-Linear Estimate of the Variance in a GARCH Model Using RBFs

An ARCH estimate of the variance of a GARCH model, when the true variance is unknown, works well as a linear estimate of the variance of a GARCH model. But if the true variance is non-linear, an ARCH estimate is unable to capture the underlying non-linearity of the true variance of the GARCH model (Franses and Van Dijk, 1996). Since the true variance and the true residuals of a GARCH model are unobservable, an RBF model is used as a means to obtain a non-linear estimate of the variance of a GARCH model that is more accurate than a linear estimate of the variance of the GARCH model.

The residuals in a GARCH model can be estimated using Weighted Orthogonal Forward Regression (Zhao, 2010) described in Section 3.3.4. These residuals can then be used to generate an initial linear ARCH estimate of the variance. An RBF model is then fitted to the ARCH variance estimate using the past squared residuals and the past values of the ARCH variance estimate. Finally, the coefficients of the obtained RBF model fitted to the ARCH variance estimate are further optimised using maximum likelihood.

7.5.1 Algorithm

Consider a simulated non-linear GARCH variance model where $h(t)$ is the simulated (true) variance (unknown and unavailable in the real world) and $e^2(t)$ is the squared residual, both at an instant in time, t. It is assumed that the true residuals, $e(t)$, and the true squared residuals, $e^2(t)$, are available.

$(N + 1000)$ data points are generated and the first 1000 are discarded to avoid initial condition errors. When simulating the variance, the residuals $e(t)$ are modelled as $e(t) = z(t)\sqrt{h(t)}$ (Engle, 2001), where $z(t)$ is a random independent and identically distributed (i.i.d) term that has zero mean and a variance of 1.

Step 1. An ARCH(q) estimate of the variance is obtained from the simulated residuals, $e(t)$, using maximum likelihood.

$$\hat{h}(t) = a_0 + a_1 e^2(t-1) + a_2 e^2(t-2) + \cdots + a_q e^2(t-q) \qquad (7.25)$$

where $\hat{h}(t)$ is the ARCH estimate of the variance, $e^2(t-i)$ is the lagged squared residual, q is the total number of coefficients to be optimised and $a_1, a_2, \ldots, a_q$ are the coefficients to be optimised using maximum likelihood (See Section 6.3).

Step 2. Divide all the available data series into 2 sets - training set and validation set. The training set consists of n samples and the validation set consists of $(N - n)$ samples.

Step 3. $\hat{h}_{train}(t)$ is set as the output series for the training set. $\hat{h}_{val}(t)$ is set as the output series for the validation set. The inputs can be set as one or more of the following: $\hat{h}_{train}(t-i), {e_{train}}^2(t-j)$ and $e_{train}(t-k)$, where i, j and k are integers ≥ 1. The respective inputs for the validation set are formulated using $\hat{h}_{val}(t-i), {e_{val}}^2(t-j)$ and $e_{val}(t-k)$. Consider $i = j = k = 1$. The input and output matrices are constructed as

$$X_{train} = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \ldots & \hat{h}_{n-1} \\ {e^2}_1 & {e^2}_2 & \ldots & {e^2}_{n-1} \\ e_1 & e_1 & \ldots & e_{n-1} \end{bmatrix}, \quad X_{val} = \begin{bmatrix} \hat{h}_n & \hat{h}_{n+1} & \ldots & \hat{h}_{N-1} \\ {e^2}_n & {e^2}_{n+1} & \ldots & {e^2}_{N-1} \\ e_n & e_{n+1} & \ldots & e_{N-1} \end{bmatrix},$$

$$\hat{h}_{train} = \begin{bmatrix} \hat{h}_2 \\ \hat{h}_3 \\ \vdots \\ \hat{h}_n \end{bmatrix}, \quad \hat{h}_{val} = \begin{bmatrix} \hat{h}_{n+1} \\ \hat{h}_{n+2} \\ \vdots \\ \hat{h}_N \end{bmatrix}$$

$$X = [x_1 \quad \cdots \quad x_N], \quad where \quad x_f = \begin{bmatrix} \hat{h}_f \\ {e^2}_f \\ e_f \end{bmatrix}, \quad f = 1,2,\ldots,N$$

Using the inputs, X_{train} and X_{val}, and outputs, $\hat{h}_{train}$ and $\hat{h}_{val}$, in the training and validation set, a suitable RBF model is obtained using the training algorithm described in Section 7.4. This RBF model is referred to as Model 1.

Step 4. The number of centres (or basis functions) in the obtained RBF model (Model 1) is represented by m. The centres are represented as

$$c_g = \begin{bmatrix} \hat{h}_g \\ {e^2}_g \\ e_g \end{bmatrix}, \qquad g = 1,2,\ldots,m \tag{7.26}$$

The obtained RBF model is represented as follows

For the training set, $\quad \hat{h}_{train} = \Phi_{train} W \qquad (7.27)$

$$\begin{bmatrix} \hat{h}_2 \\ \hat{h}_3 \\ \vdots \\ \hat{h}_n \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1m} \\ 1 & \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{(n-1)1} & \varphi_{(n-1)2} & \cdots & \varphi_{(n-1)m} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{7.28}$$

For the validation set, $\quad \hat{h}_{val} = \Phi_{val} W \qquad (7.29)$

$$\begin{bmatrix} \hat{h}_{n+1} \\ \hat{h}_{n+2} \\ \vdots \\ \hat{h}_N \end{bmatrix} = \begin{bmatrix} 1 & \varphi_{n1} & \varphi_{n2} & \cdots & \varphi_{nm} \\ 1 & \varphi_{(n+1)1} & \varphi_{(n+1)2} & \cdots & \varphi_{(n+1)m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_{(N-1)1} & \varphi_{(N-1)2} & \cdots & \varphi_{(N-1)m} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{7.30}$$

For Gaussian basis functions,

$$\varphi_{fg} = exp\left(\frac{-\left(\hat{h}_f - \hat{h}_g\right)^2}{2{\sigma_{\hat{h}}}^2} - \frac{\left({e^2}_f - {e^2}_g\right)^2}{2{\sigma_{e^2}}^2} - \frac{\left(e_f - e_g\right)^2}{2{\sigma_e}^2} \right) \tag{7.31}$$

x_f is the input data coordinate, c_g is the coordinate representing the centre of the basis function, and W is the weight sequence. $\sigma_{\hat{h}}, \sigma_{e^2}$ and σ_e are the widths of the basis functions corresponding to the three respective inputs.

Step 5. The optimal weight matrix, W is calculated using

$$W = [(\Phi_{train})^T \Phi_{train} + \lambda \mathrm{I}]^{-1} \Phi_{train} \hat{h}_{train} \tag{7.32}$$

where λ is the regularisation parameter.

Step 6. As explained in section 6.3.3, for the GARCH class of variance models, the variance, $\hat{h}(t)$, the random i.i.d sequence, $z(t)$ and the residuals, $e(t) = z(t)\sqrt{\hat{h}(t)}$ are assumed to have a Gaussian probability distribution function as shown in equation (6.4). The average log-likelihood is calculated as shown in equation (6.5). Due to the use of the regularisation parameter, λ, in the calculation of W, the average log-likelihood is now calculated as

$$l = \sum_{t=1}^{N}\left(-\frac{1}{2}ln\left(\hat{h}(t)\right) - \frac{e^2(t)}{2\hat{h}(t)}\right) - \lambda|W|^2 \tag{7.33}$$

where N denotes the number of samples, and $|W|$ denotes the l^2-norm of the weight matrix, W, and is calculated as

$$|W| = \sqrt{\sum_{k=1}^{m} {w_k}^2} \tag{7.34}$$

Step 7. The next step is to optimise the weights, W, of Model 1 such that the log-likelihood function, l, is maximised. Four approaches of optimisation are tested and the final results are compared.

Approach 1. *Unconstrained Maximisation of the Log-Likelihood of the Variance for the Validation Set:* Unconstrained optimisation of W is carried out while maximising the log-likelihood of the one-step-ahead variance estimate of Model 1 for the validation set. The 'fminunc' routine in MATLAB is used to perform unconstrained minimisation of the negative log-likelihood function.

Approach 2. *Unconstrained Maximisation of the Log-Likelihood of the Variance for the Training and Validation Set:* Unconstrained optimisation of W is carried out while maximising the sum of the log-likelihood values of the one-step-ahead variance estimate of Model 1 for both, the training and validation sets. The 'fminunc' routine in MATLAB is used to perform unconstrained minimisation of the negative log-likelihood function.

Approach 3. *Constrained Maximisation of the Log-Likelihood of the Variance for the Validation Set:* Constrained optimisation of W is carried out while maximising the log-likelihood values of the one-step-ahead variance estimate of Model 1 for the validation set. Of all the local maximums obtained, the parameters that yield the maximum sum of log-likelihood of the one-step-ahead variance estimate for the training and validation sets is selected. The 'MultiStart' routine in MATLAB is used to perform constrained minimisation of the negative log-likelihood function.

Approach 4. *Constrained Maximisation of the Log-Likelihood of the Variance for the Training and Validation Set:* Constrained optimisation of W is carried out while maximising the sum of the log-likelihood of the one-step-ahead variance estimate of Model 1 for the training and validation sets. The 'MultiStart' routine in MATLAB is used to perform constrained minimisation of the negative log-likelihood function.

Constrained maximisation of the likelihood poses a problem – the parameter bounds for the optimisation of the weights, W, of Model 1, are unknown and cannot be determined. For this reason, to determine suitable target weights to be achieved via optimisation of W, Model 1 is fitted to the true/simulated variance, $h(t)$. The weights that estimate the true variance using the structure of Model 1 are obtained.

Step 8. Consider the model obtained in Step 4. The true variance, $h(t)$, is split into 2 sets – the training set and the validation set. The training set consists of n samples and the validation set consists of $(N-n)$ samples.

$$h_{train} = \begin{bmatrix} h_2 \\ h_3 \\ \vdots \\ h_n \end{bmatrix}, \quad h_{val} = \begin{bmatrix} h_{n+1} \\ h_{n+2} \\ \vdots \\ h_N \end{bmatrix}$$

Using Φ_{train} from equation (7.28), the weights, W_{true}, that best estimate the true variance are calculated.

$$W_{true} = [(\Phi_{train})^T \Phi_{train} + \lambda \mathrm{I}]^{-1} \Phi_{train} h_{train} \tag{7.35}$$

Ideally, optimisation of the weights, W, using maximum likelihood (described in Steps 6 and 7), should result in a final weight matrix that is equivalent to W_{true}.

Step 9. The parameter bounds required in Steps 6 and 7 for constrained optimisation of the weights using maximum likelihood are calculated using the starting weights, W, and the target weights, W_{true}, obtained in Step 8. Let w_i be the i^{th} element in W, and wt_i be the i^{th} element in W_{true}. The upper bounds and lower bounds for every possible combination of w_i and wt_i are listed in Table 7.1.

Table 7.1 Calculation of Upper and Lower Bounds for Constrained Maximum Likelihood

Condition		Lower Bound	Upper Bound
$w_i > 0\ and$	$wt_i < 0 < w_i$	$wt_i \times 1.05$	$w_i \times 1.05$
	$0 < wt_i < w_i$	$wt_i \times 0.95$	$w_i \times 1.05$
	$w_i < wt_i$	$w_i \times 0.95$	$wt_i \times 1.05$
	$w_i = wt_i$	$w_i \times 0.95$	$w_i \times 1.05$
$w_i < 0\ and$	$wt_i < 0 < w_i$	$wt_i \times 1.05$	$w_i \times 0.95$
	$0 < wt_i < w_i$	$w_i \times 1.05$	$wt_i \times 0.95$
	$w_i < wt_i$	$w_i \times 1.05$	$wt_i \times 1.05$
	$w_i = wt_i$	$w_i \times 1.05$	$w_i \times 0.95$

Step 10. Let W' represent the final optimised weights obtained after optimisation. The final one-step-ahead variance estimates for the training and validation set are calculated as

$$\hat{h}_{train}{}' = \Phi_{train} W' \tag{7.36}$$

$$\hat{h}_{val}{}' = \Phi_{val} W' \tag{7.37}$$

where Φ_{train} and Φ_{val} are as calculated in step 4.

Step 11. The accuracy of the estimates is obtained by calculating NRMSE($h_{train}, \hat{h}_{train}{}'$) and NRMSE($h_{val}, \hat{h}_{val}{}'$) as shown in equation (7.19). If these values are lesser than NRMSE($h_{train}, \hat{h}_{train}$) and NRMSE($h_{val}, \hat{h}_{val}$), where $\hat{h}_{train}$ and $\hat{h}_{val}$ are ARCH estimates of the true GARCH variance, the obtained non-linear variance estimates, $\hat{h}_{train}{}'$ and $\hat{h}_{val}{}'$, are better than the linear ARCH estimates, $\hat{h}_{train}$ and $\hat{h}_{val}$.

7.6 Simulations

7.6.1 Simulation 1: SQR-GARCH Model

Consider the following SQR-GARCH variance model

$$h(t) = K + A_1 h(t-1) + A_2 z^2(t-1) + A_3 e(t-1) \qquad (7.38)$$

where $h(t)$ is the simulated/true variance, $e(t)$ is the residual, and $e^2(t)$ is the squared residual, all at an instant in time, t. K, A_1, A_2 and A_3 are the parameters of the variance model. When simulating the variance from equation (7.38), the residuals $e(t)$ are modelled as $e(t) = z(t)\sqrt{h(t)}$, where $z(t)$ is a random independent and identically distributed (i.i.d) term that has zero mean and a variance of 1. The values of the parameters of the SQR-GARCH model are listed in Table 7.2.

Table 7.2 Parameters of the simulated SQR-GARCH variance model

Parameter of the Variance Model	Value
K	1.0E-05
A_1	0.8
A_2	1E-04
A_3	1E-03

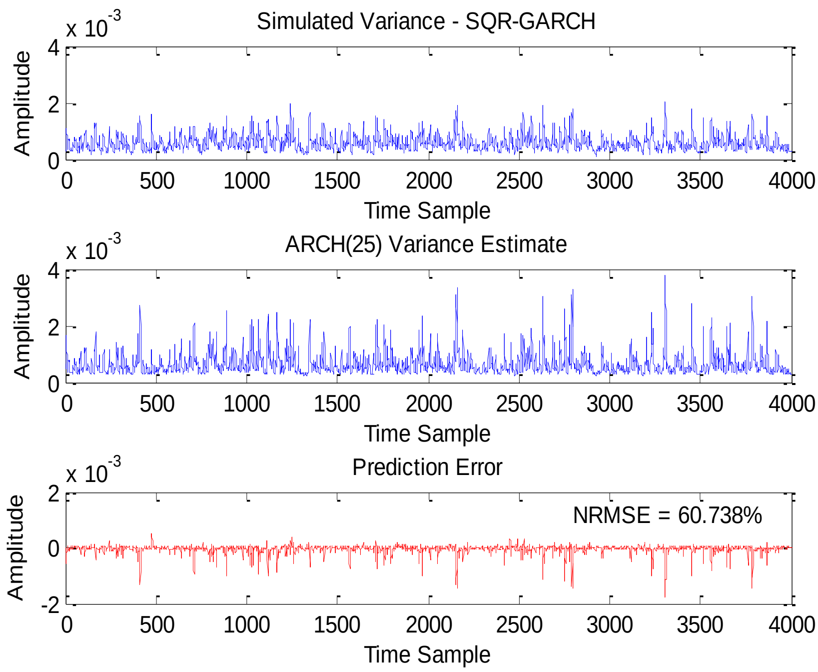


Figure 7.3 Simulated SQR-GARCH Variance, Estimated ARCH(25) Variance and Prediction Error

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. Hence, the total number of samples, N, is 4000. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10). For the same reasons given in Section 4.4, an ARCH(25) variance estimate, $\hat{h}_A(t)$, is generated using the true squared residuals, $e^2(t)$. The simulated variance, the ARCH(25) variance estimate and the prediction error between the two are shown in Figure 7.3. The NRMSE between the simulated variance and the ARCH(25) variance estimate is calculated to be 60.7380%.

All the available data series are split into 2 sets – training set and validation set. The training set consists of the first 2000 samples and the validation set consists of the remaining 2000 samples ($n = 2000$).

An RBF model is to be fitted to the ARCH(25) variance estimate, $\hat{h}_A(t)$. $\hat{h}_A(t-1)$, $e(t-1)$ and $e^2(t-1)$ are used as inputs. The hyper-parameters of the Gaussian basis functions are initialised as explained in Section 7.4.1.

Selection of Centres: The ERR cut-off values are reduced from 1% to 0.1% in increments of 0.1%, and 0.09% to 0.05% in increments of 0.01%. The significant centres are selected via OFR and the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are calculated for each iteration. The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets while selecting the centres is shown in Figure 7.4.

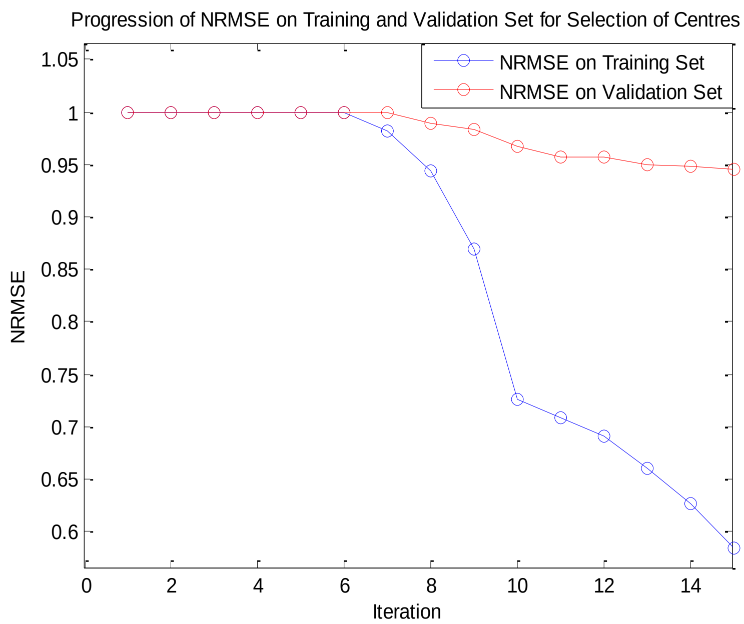


Figure 7.4 Progression of NRMSE for Training and Validation Sets for Selection of Centres of Gaussian Basis Functions for Example 1

The NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are least in the final iteration. The number of centres selected, m, is 123. Figure 7.5 shows the selected centres in red circles for all the 3 inputs.

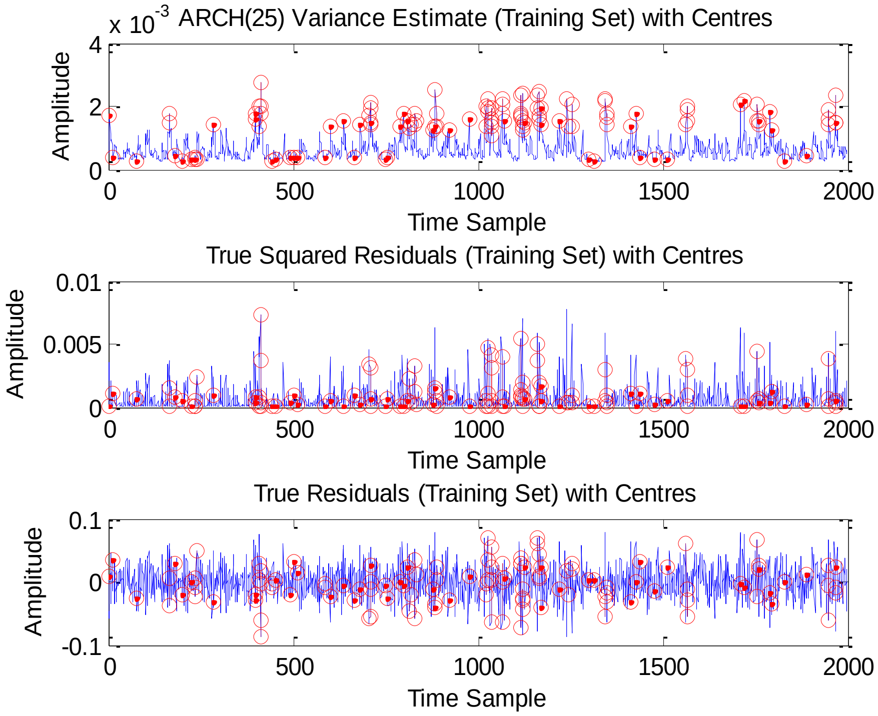


Figure 7.5 Selected Centres (Red Circles) for all Inputs for Example 1

Selection of Hyper-parameters of Basis Functions: The widths of the basis functions, $\sigma_{\hat{h}}, \sigma_{e^2}$ and σ_e, and the regularisation parameter, λ, are all optimised. $\sigma_{\hat{h}}, \sigma_{e^2}$ and σ_e are the widths of the basis functions corresponding to the three respective inputs. The initial values, range of optimisation, and optimised values of the hyper-parameters are listed in Table 7.3.

Table 7.3 Optimised Values of Hyper-parameters of the Gaussian Basis Functions for Example 1

Hyper-parameters of Gaussian Basis Function	Initial Value	Range of Optimisation	Optimised Value
$\sigma_{\hat{h}}$	1E-05	1E-05 to 1E-01	0.0016
σ_{e^2}	1E-05	1E-05 to 1E-01	0.0056
σ_e	0.03	2.9E-02 to 5	0.031
λ	1E-08	4E-09 to 1E-01	6.3E-04

The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets for the selection of the regularisation parameter is shown in Figure 7.6, and the same for the selection of the widths is shown in Figure 7.7.

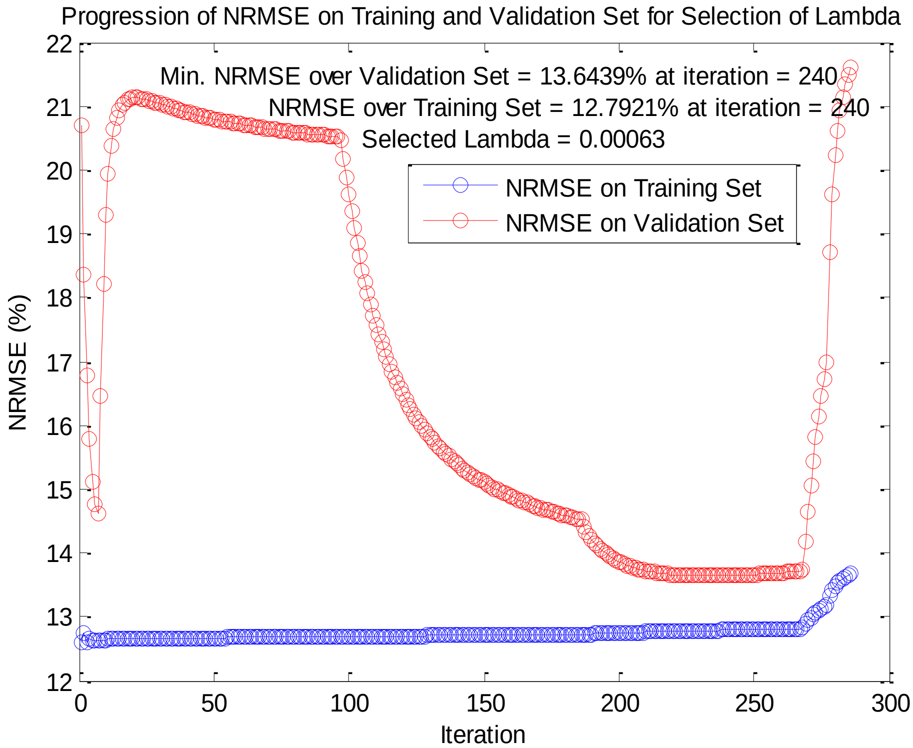


Figure 7.6 Progression of NRMSE for Training and Validation Sets for Selection of Regularisation Parameter for Example 1

This obtained RBF Model with optimised hyper-parameters, fitted to the ARCH(25) variance estimate is referred to as Model 1. Figure 7.8 shows the plots of the ARCH(25) variance estimate, one-step-ahead estimate of Model 1 ($\hat{h}_{m1}(t)$), and the prediction error between them, all for the training and validation set.

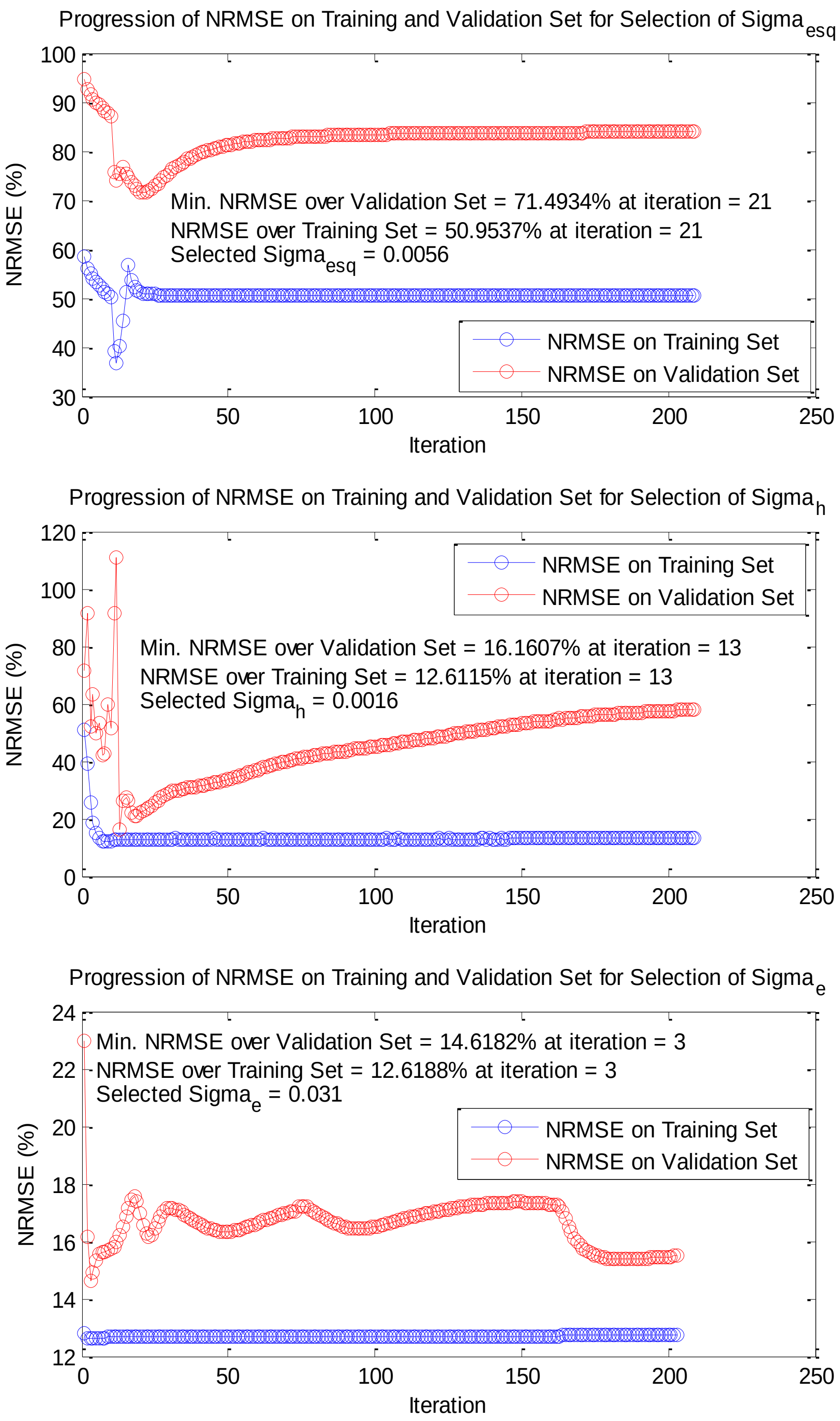


Figure 7.7 Progression of NRMSE for Training and Validation Sets for Selection of Widths of Gaussian Basis Functions for Example 1

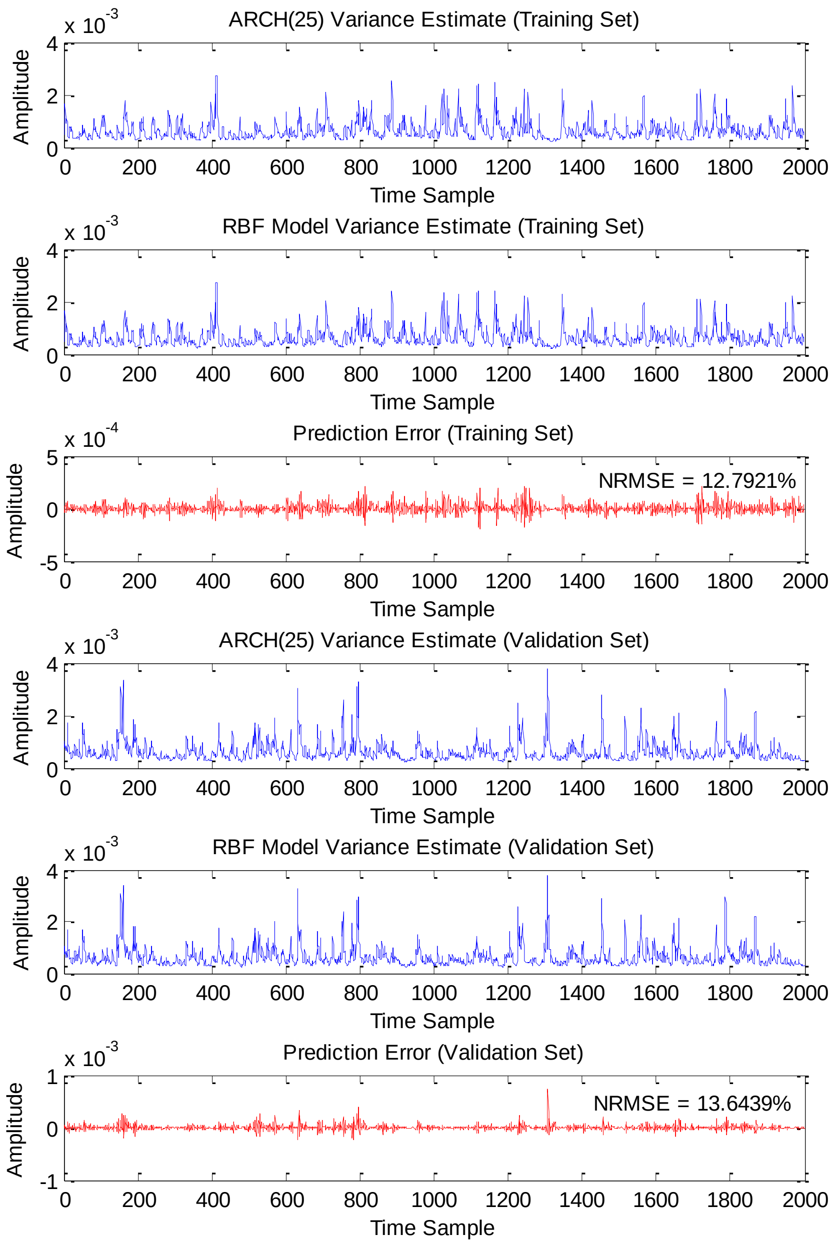


Figure 7.8 Outputs (ARCH(25) Variance Estimate) vs. RBF Model Estimate vs. Prediction Error for Training and Validation Set for Example 1

Obtaining RBF Model Parameter Estimates for the Simulated/True Variance: The parameter estimates of Model 1 when fitted to the true variance are obtained, as explained in Step 9 of Section 7.5.1. This model is referred to as Model 2. The one-step-ahead estimate of Model 2, $\hat{h}_{m2}(t)$, is generated.

Table 7.4 shows the values of the NRMSE between the true variance, $h(t)$, and all the model estimates, $\hat{h}_A(t)$, $\hat{h}_{m1}(t)$ and $\hat{h}_{m2}(t)$, for the training set and the validation set.

Table 7.4 NRMSEs between True Variance and All Model Estimates for Example 1

	$\boldsymbol{NRMSE}\left(\boldsymbol{h(t)},\boldsymbol{\hat{h}_A(t)}\right)$	$\boldsymbol{NRMSE}\left(\boldsymbol{h(t)},\boldsymbol{\hat{h}_{m1}(t)}\right)$	$\boldsymbol{NRMSE}\left(\boldsymbol{h(t)},\boldsymbol{\hat{h}_{m2}(t)}\right)$
Training Set	57.3505%	55.6473%	20.9679%
Validation Set	63.7313%	60.5615%	23.5443%
Average	60.5409%	58.1044%	22.2561%

$NRMSE\left(h(t),\hat{h}_A(t)\right)$ is the NRMSE between the true variance and the ARCH(25) variance estimate. $NRMSE\left(h(t),\hat{h}_{m1}(t)\right)$ is the NRMSE between the true variance and the one-step-ahead estimate of Model 1. It represents the initial NRMSE before the parameters of Model 1 are optimised via maximum likelihood. $NRMSE\left(h(t),\hat{h}_{m2}(t)\right)$ is the NRMSE between the true variance and the one-step-ahead estimate of Model 2. It represents the desired NRMSE after optimisation of Model 1 parameters via maximum likelihood.

The final step is the optimisation of the parameters of Model 1 using maximum likelihood. As explained in Step 7 of Section 7.5.1, four approaches of optimisation are tested and the final results are compared. $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$ are the one-step-ahead variance estimates generated after optimisation via Approach 1, 2, 3 and 4 respectively. Table 7.5 shows the log-likelihood values of all the model estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.6 shows the values of the NRMSE between the true variance, $h(t)$, and all the model estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.5 Log-Likelihood Values of All ML-Optimised RBF Model One-Step-Ahead Estimates for Example 1

	$\boldsymbol{L\left(\hat{h}_{NL1}(t)\right)}$	$\boldsymbol{L\left(\hat{h}_{NL2}(t)\right)}$	$\boldsymbol{L\left(\hat{h}_{NL3}(t)\right)}$	$\boldsymbol{L\left(\hat{h}_{NL4}(t)\right)}$
Training Set	4626.20	4631.20	4629.50	4636.80
Validation Set	4688.80	4688.80	4695.20	4692.30
Average	4657.50	4660.00	4662.35	4664.55

Table 7.6 NRMSEs between True Variance and All ML-Optimised RBF Model One-Step-Ahead Estimates for Example 1

	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL1}(t)}\right)$	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL2}(t)}\right)$	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL3}(t)}\right)$	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL4}(t)}\right)$
Training Set	41.4969%	41.9827%	46.4295%	38.0683%
Validation Set	39.8547%	43.1538%	60.2026%	47.3739%
Average	40.6578%	42.5682%	53.3161%	42.7211%

Of all the four approaches, Approach 4 yields a one-step-ahead variance estimate that has the highest log-likelihood and Approach 1 yields a one-step-ahead variance estimate that has the least NRMSE with the true variance. A variance estimate that has much better accuracy than a linear ARCH(25) estimate is obtained.

7.6.2 Simulation 2: NL-GARCH Model

Consider the following NL-GARCH variance model

$$\begin{aligned} h(t) = K + G_1 h(t-1) + G_2 h(t-2) + A e^2(t-1) + B_1 h(t-1) e^2(t-1) \\ + B_2 h^2(t-1) + B_3 e^4(t-1) + B_4 h(t-2) e^2(t-2) \end{aligned} \quad (7.39)$$

where all the terms have the same definition as in the previous example (Section 7.6.1, equation (7.38)). The values of the parameters of the NL-GARCH model are listed in Table 7.7.

Table 7.7 Parameters of the simulated NL-GARCH variance model

Parameter of the Variance Model	**Value**
K	1E-05
G_1	0.92
G_2	-0.06
A	0.05
B_1	8
B_2	8
B_3	30
B_4	0.1

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. Hence, the total number of samples, N, is 4000. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10). An ARCH(25) variance estimate, $\hat{h}_A(t)$, is generated using the true squared residuals, $e^2(t)$. The simulated variance, the ARCH(25) variance estimate and the prediction error between the two are shown in Figure 7.9. The NRMSE between the simulated variance and the ARCH(25) variance estimate is calculated to be 35.7650%.

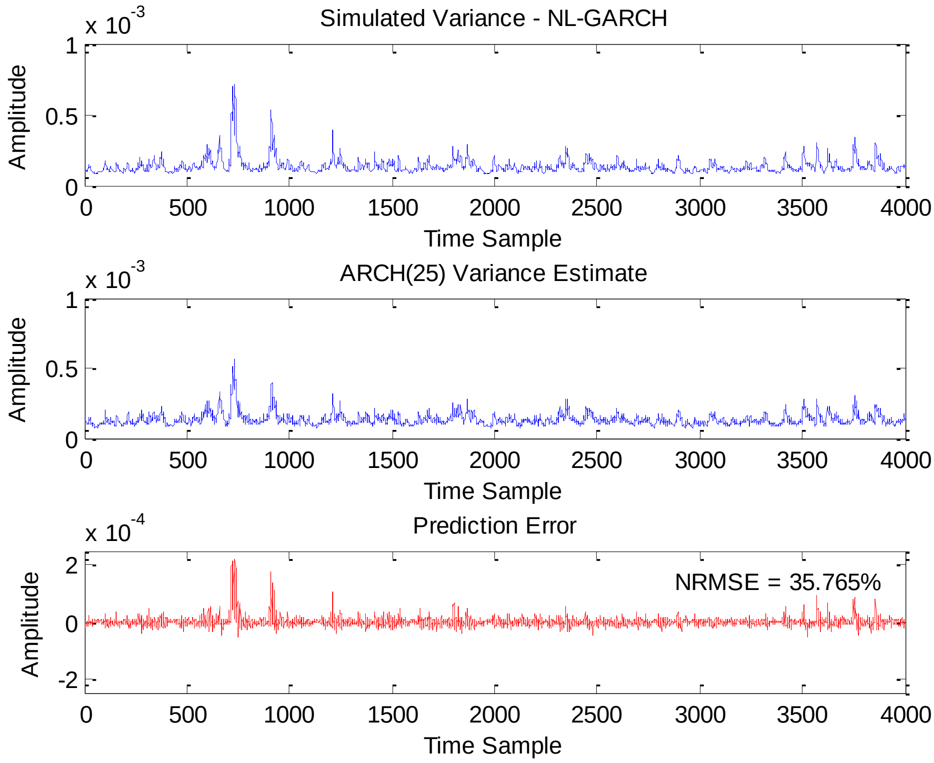


Figure 7.9 Simulated NL-GARCH Variance, Estimated ARCH(25) Variance and Prediction Error

All the available data series are split into 2 sets – training set and validation set. The training set consists of the first 2000 samples and the validation set consists of the remaining 2000 samples ($n = 2000$).

An RBF model is to be fitted to the ARCH(25) variance estimate, $\hat{h}_A(t)$. $\hat{h}_A(t-1), e(t-1)$ and $e^2(t-1)$ are used as inputs. The hyper-parameters of the Gaussian basis functions are initialised as explained in Section 7.4.1.

Selection of Centres: The ERR cut-off values are reduced from 1% to 0.1% in increments of 0.1%, and 0.09% to 0.07% in increments of 0.01%. The significant centres are selected via OFR and the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are calculated for each iteration. The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets while selecting the centres is shown in Figure 7.10.

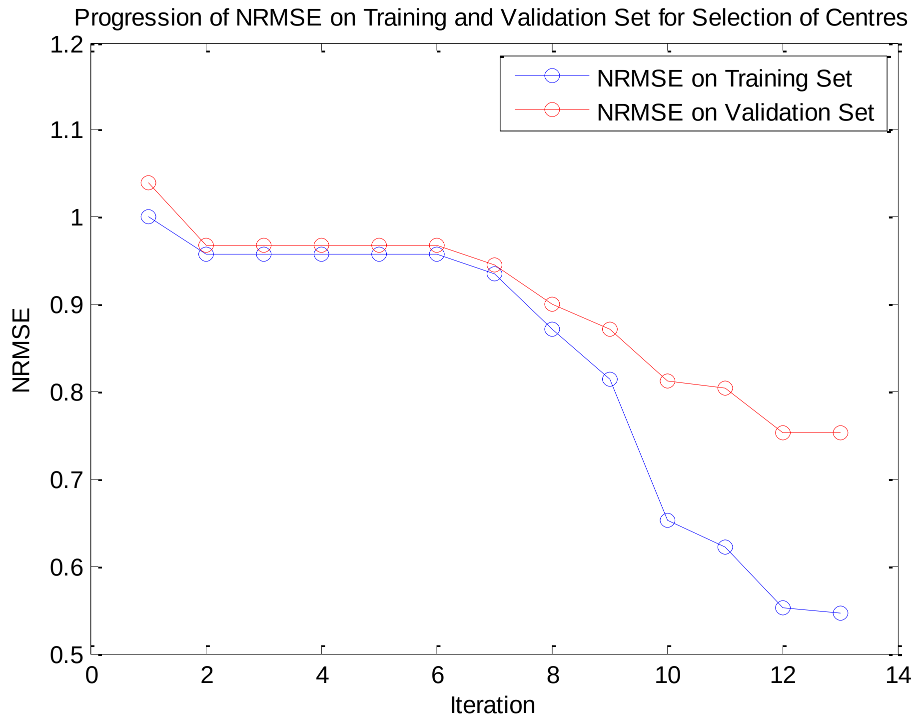


Figure 7.10 Progression of NRMSE for Training and Validation Sets for Selection of Centres of Gaussian Basis Functions for Example 2

The NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are least in the final iteration. The number of centres selected, m, is 48. Figure 7.11 shows the selected centres in red circles for all the 3 inputs.

Selection of Hyper-parameters of Basis Functions: The widths of the basis functions, $\sigma_{\hat{h}}, \sigma_{e^2}$ and σ_e, and the regularisation parameter, λ, are all optimised. The initial values, range of optimisation, and optimised values of the hyper-parameters are listed in Table 7.8.

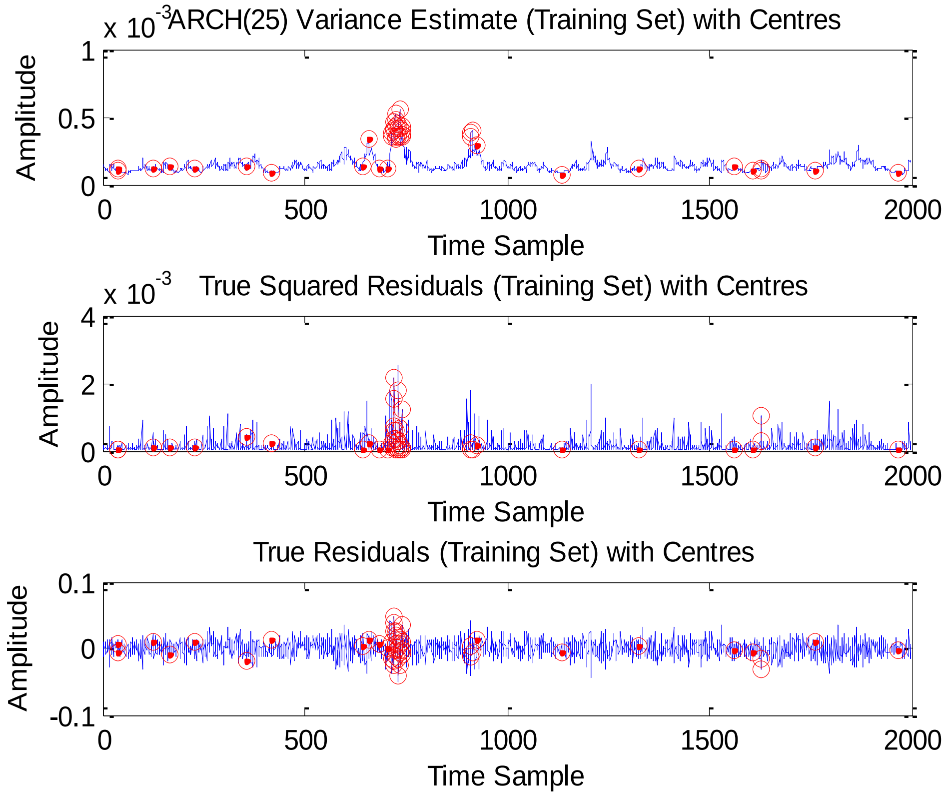


Figure 7.11 Selected Centres (Red Circles) for all Inputs for Example 2

Table 7.8 Optimised Values of Hyper-parameters of the Gaussian Basis Functions for Example 2

Hyper-parameters of Gaussian Basis Function	Initial Value	Range of Optimisation	Optimised Value
$\sigma_{\hat{h}}$	1E-05	1E-06 to 1E-01	0.0146
σ_{e^2}	1E-05	1E-06 to 1E-01	0.0006
σ_e	0.03	1.6E-02 to 1	0.0280
λ	1E-08	5E-11 to 1E-01	1E-08

The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets for the selection of the regularisation parameter is shown in Figure 7.12, and the same for the selection of the widths is shown in Figure 7.13.

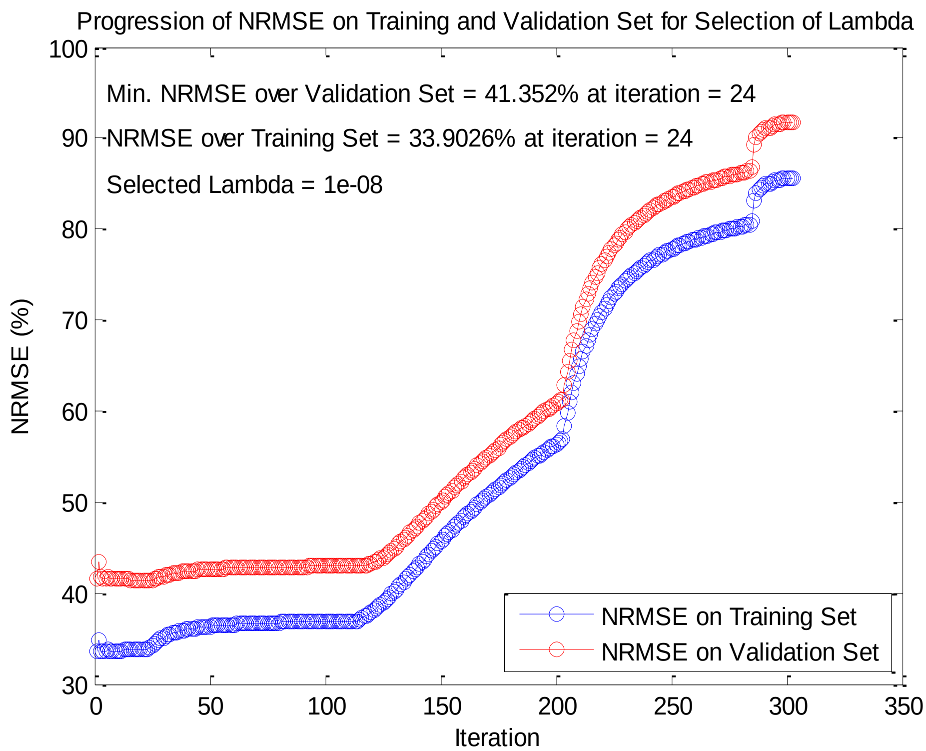


Figure 7.12 Progression of NRMSE for Training and Validation Sets for Selection of Regularisation Parameter for Example 2

This obtained RBF Model with optimised hyper-parameters, fitted to the ARCH(25) variance estimate is referred to as Model 1.

Obtaining RBF Model Parameter Estimates for the Simulated/True Variance: The parameter estimates of Model 1 when fitted to the true variance are obtained, as explained in Step 9 of Section 7.5.1. This model is referred to as Model 2. The one-step-ahead estimate of Model 2, $\hat{h}_{m2}(t)$, is generated.

Table 7.9 NRMSEs between True Variance and All Model Estimates for Example 2

	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_A(t)}\right)$	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{m1}(t)}\right)$	$\boldsymbol{NRMSE}$ $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{m2}(t)}\right)$
Training Set	36.0663%	36.9880%	27.4546%
Validation Set	35.8235%	33.3562%	36.1114%
Average	35.9449%	35.1721%	31.7830%

Table 7.9 shows the values of the NRMSE between the true variance, $h(t)$, and all the model estimates, $\hat{h}_A(t)$, $\hat{h}_{m1}(t)$ and $\hat{h}_{m2}(t)$, for the training set and the validation set.

$NRMSE\left(h(t), \hat{h}_A(t)\right)$, $NRMSE\left(h(t), \hat{h}_{m1}(t)\right)$ and $NRMSE\left(h(t), \hat{h}_{m2}(t)\right)$ have the same definitions as in the previous simulation (Section 7.6.1).

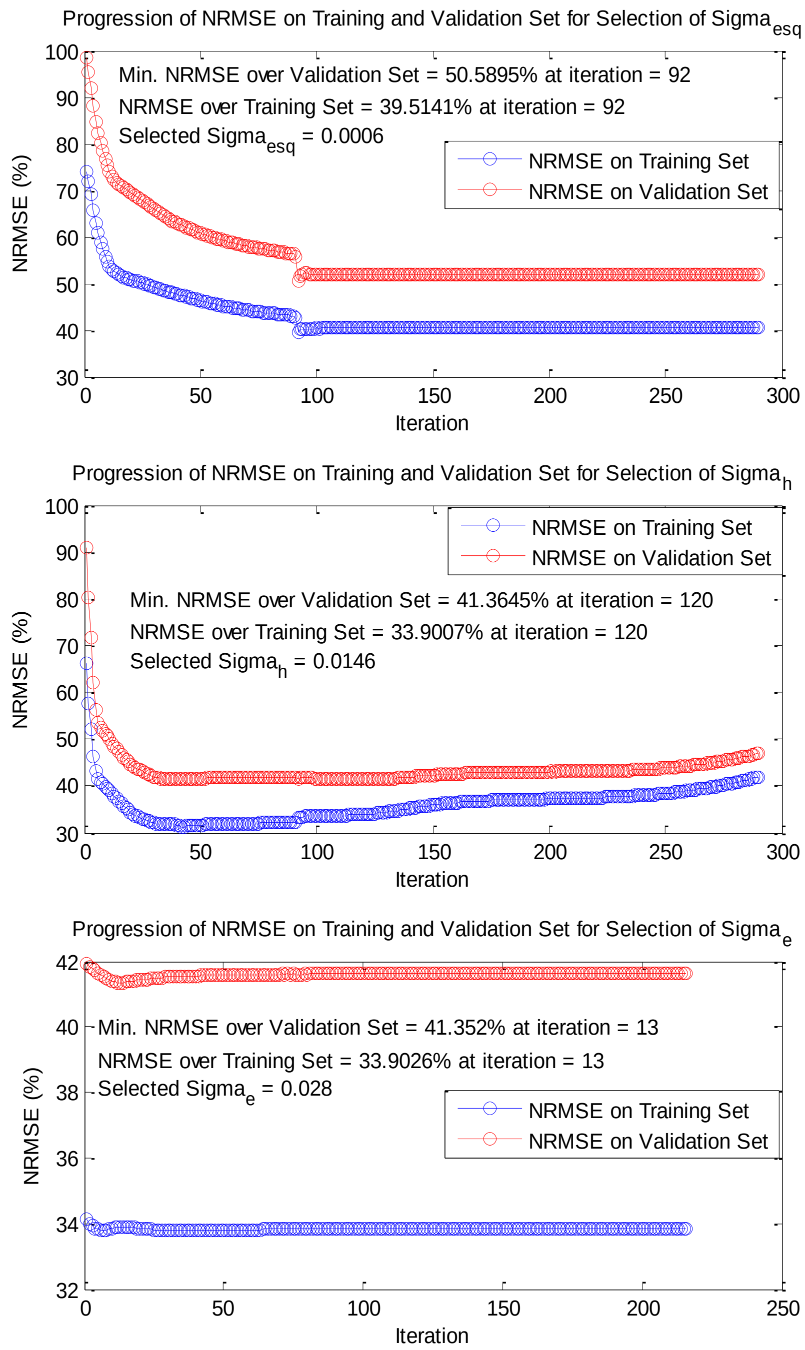


Figure 7.13 Progression of NRMSE for Training and Validation Sets for Selection of Widths of Gaussian Basis Functions for Example 2

The final step is the optimisation of the parameters of Model 1 using maximum likelihood. As explained in Step 7 of Section 7.5.1, four approaches of optimisation are tested and the final results are compared. $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$ have the same definitions as in the previous simulation (Section 7.6.1).

Table 7.10 shows the log-likelihood values of all the one-step-ahead model estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.11 shows the values of the NRMSE between the true variance, $h(t)$, and all the one-step-ahead estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.10 Log-Likelihood Values of All ML-Optimised RBF Model One-Step-Ahead Estimates for Example 2

	$\boldsymbol{L(\hat{h}_{NL1}(t))}$	$\boldsymbol{L(\hat{h}_{NL2}(t))}$	$\boldsymbol{L(\hat{h}_{NL3}(t))}$	$\boldsymbol{L(\hat{h}_{NL4}(t))}$
Training Set	6065.20	6067.00	6065.30	6067.60
Validation Set	6167.20	6166.60	6167.90	6166.00
Average	6116.20	6116.80	6116.60	6116.80

Table 7.11 NRMSEs between True Variance and All ML-Optimised RBF Model One-Step-Ahead Estimates for Example 2

	$\boldsymbol{NRMSE}$ $\boldsymbol{(h(t), \hat{h}_{NL1}(t))}$	$\boldsymbol{NRMSE}$ $\boldsymbol{(h(t), \hat{h}_{NL2}(t))}$	$\boldsymbol{NRMSE}$ $\boldsymbol{(h(t), \hat{h}_{NL3}(t))}$	$\boldsymbol{NRMSE}$ $\boldsymbol{(h(t), \hat{h}_{NL4}(t))}$
Training Set	35.7383%	34.9004%	42.9073%	39.9970%
Validation Set	36.3289%	32.6695%	40.2260%	36.9785%
Average	36.0336%	33.7849%	41.5667%	38.4877%

Of all the four approaches, Approaches 2 and 4 yield one-step-ahead variance estimates that have the highest log-likelihood and Approach 2 yields a one-step-ahead variance estimate that has the least NRMSE with the true variance. A variance estimate that has slightly better accuracy than a linear ARCH(25) estimate is obtained.

7.6.3 Simulation 3: NA-GARCH Model

Consider the following NL-GARCH variance model

$$h(t) = K + Gh(t-1) + A\left(e(t-1) + B\sqrt{h(t-1)}\right)^2 \qquad (7.40)$$

where all the terms have the same definition as in the first example (Section 7.6.1, equation (7.38)). The values of the parameters of the NL-GARCH model are listed in Table 7.12.

Table 7.12 Parameters of the simulated NA-GARCH variance model

Parameter of the Variance Model	Value
K	1E-05
G	0.8322
A	0.1074
B	-0.3072

5000 data points are generated and the first 1000 are discarded to avoid initial condition errors. Hence, the total number of samples, N, is 4000. For a GARCH(p,q) model, the initial condition, $h(0)$ is calculated as shown in equation (2.10). An ARCH(25) variance estimate, $\hat{h}_A(t)$, is generated using the true squared residuals, $e^2(t)$. The simulated variance, the ARCH(25) variance estimate and the prediction error between the two are shown in Figure 7.14. The NRMSE between the simulated variance and the ARCH(25) variance estimate is calculated to be 32.3715%.

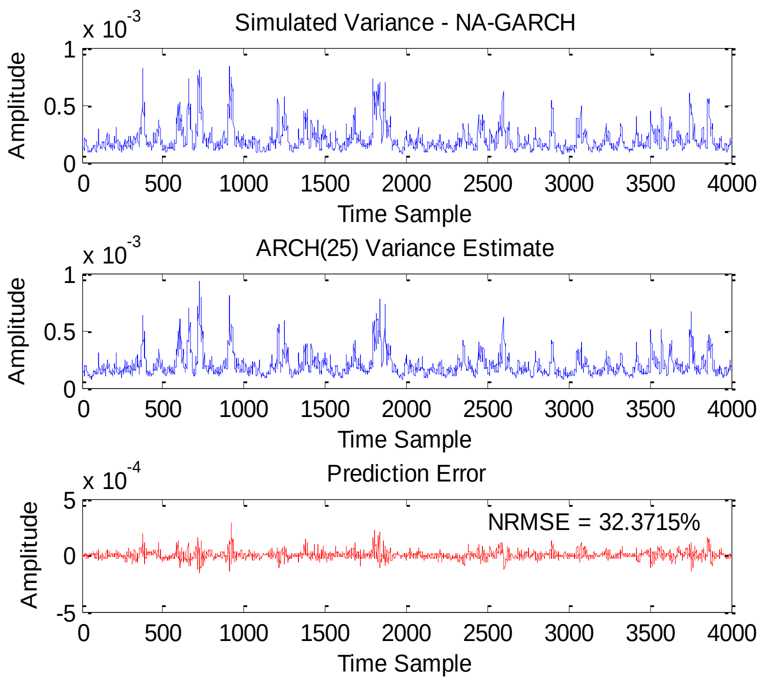


Figure 7.14 Simulated NA-GARCH Variance, Estimated ARCH(25) Variance and Prediction Error

All the available data series are split into 2 sets – training set and validation set. The training set consists of the first 2000 samples and the validation set consists of the remaining 2000 samples ($n = 2000$).

An RBF model is to be fitted to the ARCH(25) variance estimate, $\hat{h}_A(t)$. $\hat{h}_A(t-1), e(t-1)$ and $e^2(t-1)$ are used as inputs. The hyper-parameters of the Gaussian basis functions are initialised as explained in Section 7.4.1.

Selection of Centres: The ERR cut-off values are reduced from 1% to 0.1% in increments of 0.1% and 0.09% to 0.07% in increments of 0.01%. The significant centres are selected via OFR and the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are calculated for each iteration. The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets while selecting the centres is shown in Figure 7.15.

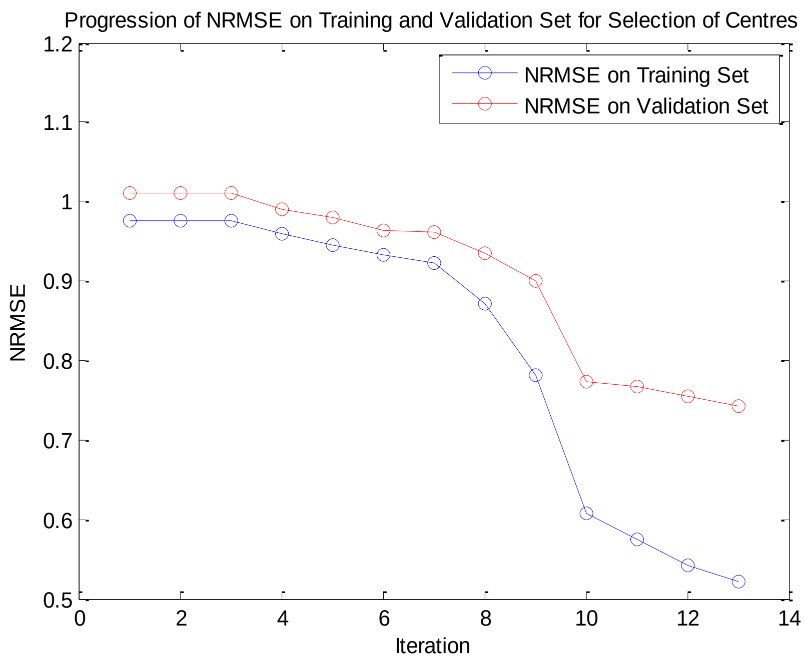


Figure 7.15 Progression of NRMSE for Training and Validation Sets for Selection of Centres of Gaussian Basis Functions for Example 3

The NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets are least in the final iteration. The number of centres selected, m, is 92. Figure 7.16 shows the selected centres in red circles for all the 3 inputs.

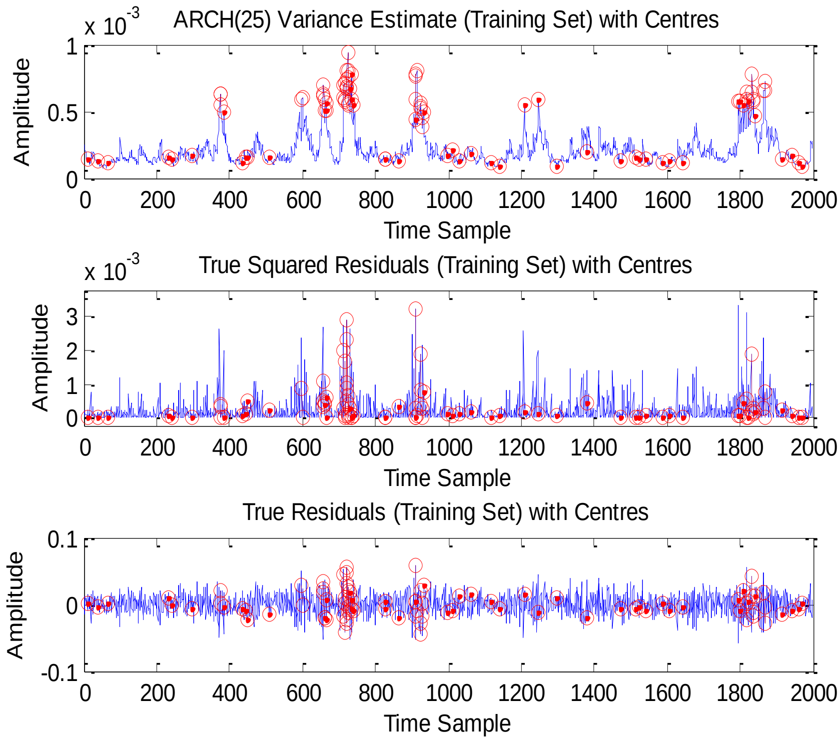


Figure 7.16 Selected Centres (Red Circles) for all Inputs for Example 3

Selection of Hyper-parameters of Basis Functions: The widths of the basis functions, $\sigma_{\hat{h}}, \sigma_{e^2}$ and σ_e, and the regularisation parameter, λ, are all optimised. The initial values, range of optimisation, and optimised values of the hyper-parameters are listed in Table 7.13.

Table 7.13 Optimised Values of Hyper-parameters of the Gaussian Basis Functions for Example 3

Hyper-parameters of Gaussian Basis Function	Initial Value	Range of Optimisation	Optimised Value
$\sigma_{\hat{h}}$	1E-05	1E-06 to 1E-01	0.0061
σ_{e^2}	1E-05	1E-06 to 1E-01	0.0031
σ_e	0.03	1.3E-02 to 1	0.21
λ	1E-03	2E-10 to 1E-01	1.0E-03

The progression of the NRMSE between the RBF model one-step-ahead estimate and the ARCH(25) variance estimate for the training and validation sets for the selection of the regularisation parameter is shown in Figure 7.17, and the same for the selection of the widths is shown in Figure 7.18.

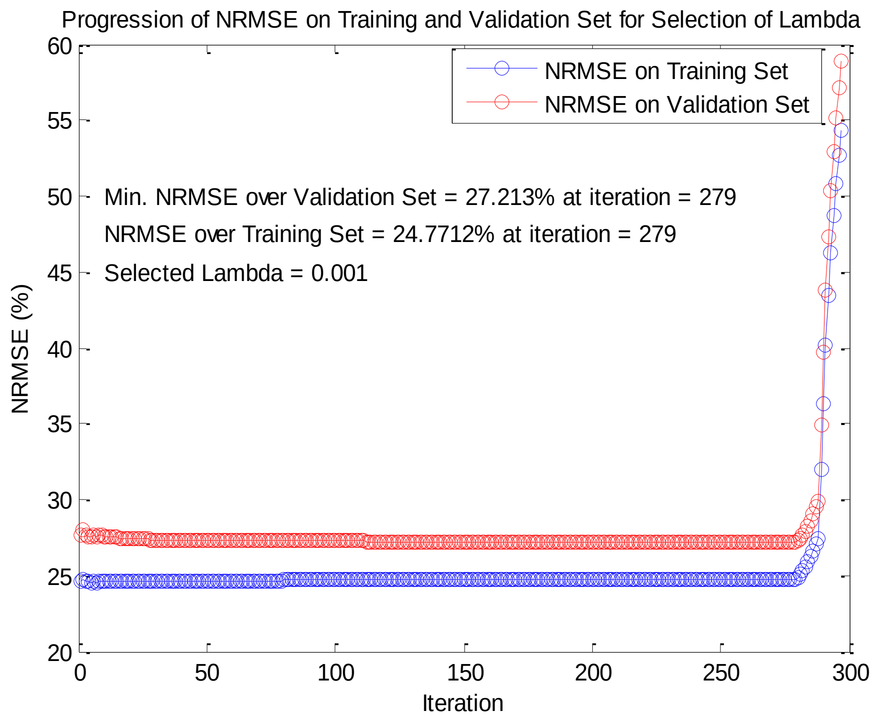


Figure 7.17 Progression of NRMSE for Training and Validation Sets for Selection of Regularisation Parameter for Example 3

This obtained RBF Model with optimised hyper-parameters, fitted to the ARCH(25) variance estimate is referred to as Model 1.

Obtaining RBF Model Parameter Estimates for the Simulated/True Variance: The parameter estimates of Model 1 when fitted to the true variance are obtained, as explained in Step 9 of Section 7.5.1. This model is referred to as Model 2. The one-step-ahead estimate of Model 2, $\hat{h}_{m2}(t)$, is generated. Table 7.14 shows the values of the NRMSE between the true variance, $h(t)$, and all the model estimates, $\hat{h}_A(t)$, $\hat{h}_{m1}(t)$ and $\hat{h}_{m2}(t)$, for the training set and the validation set. $NRMSE\left(h(t), \hat{h}_A(t)\right)$, $NRMSE\left(h(t), \hat{h}_{m1}(t)\right)$ and $NRMSE\left(h(t), \hat{h}_{m2}(t)\right)$ have the same definitions as in the first simulation (Section 7.6.1).

Table 7.14 NRMSEs between True Variance and All Model Estimates for Example 3

	$\boldsymbol{NRMSE\left(h(t), \hat{h}_A(t)\right)}$	$\boldsymbol{NRMSE\left(h(t), \hat{h}_{m1}(t)\right)}$	$\boldsymbol{NRMSE\left(h(t), \hat{h}_{m2}(t)\right)}$
Training Set	30.1543%	28.8163%	27.9729%
Validation Set	37.6921%	36.3301%	36.1646%
Average	33.9232%	32.5732%	32.0687%

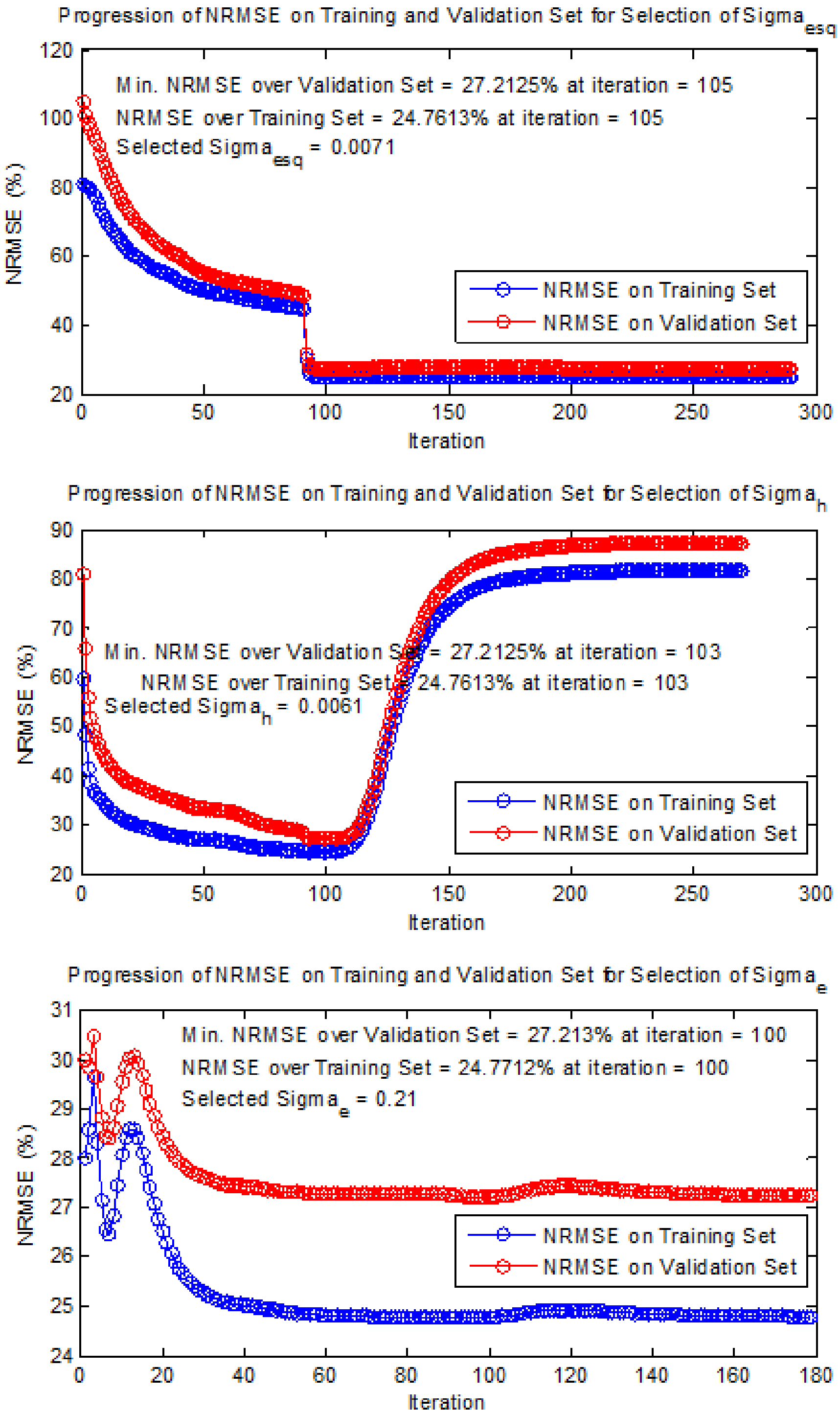


Figure 7.18 Progression of NRMSE for Training and Validation Sets for Selection of Widths of Gaussian Basis Functions for Example 3

The final step is the optimisation of the parameters of Model 1 using Maximum Likelihood. As explained in Step 7 of Section 7.5.1, four approaches of optimisation are tested and the final results are compared. $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$ have the same definitions as in the first simulation (Section 7.6.1).

Table 7.15 shows the log-likelihood values of all the one-step-ahead model estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.16 shows the values of the NRMSE between the true variance, $h(t)$, and all the one-step-ahead estimates, $\hat{h}_{NL1}(t)$, $\hat{h}_{NL2}(t)$, $\hat{h}_{NL3}(t)$ and $\hat{h}_{NL4}(t)$, for the training set and the validation set.

Table 7.15 Log-Likelihood Values of All ML-Optimised RBF Model One-Step-Ahead Estimates

	$L(\hat{h}_{NL1}(t))$	$L(\hat{h}_{NL2}(t))$	$L(\hat{h}_{NL3}(t))$	$L(\hat{h}_{NL4}(t))$
Training Set	5714.70	5715.00	5715.10	5715.00
Validation Set	5856.80	5856.70	5856.60	5856.70
Average	5785.75	5785.85	5785.85	5785.85

Table 7.16 NRMSEs between True Variance and All ML-Optimised RBF Model One-Step-Ahead Estimates for Example 3

	$NRMSE\left(h(t), \hat{h}_{NL1}(t)\right)$	$NRMSE\left(h(t), \hat{h}_{NL2}(t)\right)$	$NRMSE\left(h(t), \hat{h}_{NL3}(t)\right)$	$NRMSE\left(h(t), \hat{h}_{NL4}(t)\right)$
Training Set	29.0727%	28.8463%	28.8163%	28.9237%
Validation Set	36.3739%	36.3133%	36.3301%	36.2886%
Average	32.7233%	32.5798%	32.5732%	32.6061%

Of all the four approaches, Approaches 2, 3 and 4 yield one-step-ahead variance estimates that have the highest log-likelihood and Approach 3 yields a one-step-ahead variance estimate that has the least NRMSE with the true variance. A variance estimate that has slightly better accuracy than a linear ARCH(25) estimate is obtained.

7.6.4 Comparison of Simulation Results

A major disadvantage of constrained optimisation of the parameters of Model 1 (RBF model fit to an ARCH estimate of the true GARCH variance) is that the actual constraints within which the parameters are to be optimised are unknown and cannot be determined. Hence, for the purpose of this chapter, very tight constraints are set within which an attempt to find the global maximum of the log-likelihood of the one-step-ahead variance estimate of Model 1 is made and it is unknown whether or not the global maximum of the log-likelihood is found.

The performance of all the four approaches of optimisation of the parameters of Model 1 is compared for the three examples. Table 7.17 lists the values of the average NRMSE between the true variance and all the one-step-ahead model estimates of all the approaches, and the average NRMSE between the true variance and the ARCH(25) estimate, for all the three examples.

Table 7.17 Average NRMSEs between True Variance and All ML-Optimised RBF Model One-Step-Ahead Estimates for All Examples

Example	Avg. NRMSE $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_A(t)}\right)$	Avg. NRMSE $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL1}(t)}\right)$	Avg. NRMSE $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL2}(t)}\right)$	Avg. NRMSE $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL3}(t)}\right)$	Avg. NRMSE $\left(\boldsymbol{h(t)}, \boldsymbol{\hat{h}_{NL4}(t)}\right)$
1	60.5409%	40.6578%	42.5682%	53.3161%	42.7211%
2	35.9449%	36.0336%	33.7849%	41.5667%	38.4877%
3	33.9232%	32.7233%	32.5798%	32.5732%	32.6061%

For each variance model, the approach that yields a non-linear variance estimate with the least NRMSE (calculated vs. the true variance) highlighted. Looking at Table 7.17, the following observations are made:

- Estimates that are more accurate than the linear ARCH(25) estimates are achieved for all the examples.
- For all the examples, Approach 2 is the only method of optimisation of the parameters of Model 1 that yields a non-linear variance estimate that has a lesser NRMSE (calculated vs. the true variance) than a linear ARCH(25) variance estimate.
- For highly non-linear GARCH variance models (like Example 1) where a linear ARCH(25) is highly inaccurate, the algorithm seems to provide an estimate of the GARCH variance with much better accuracy.

There does not seem to be a definite link between an approach that yields an estimate with the highest log-likelihood and an approach that yields an estimate with the least NRMSE (calculated vs. the true variance).

7.7 Conclusions

This chapter explains in detail the concept of Radial Basis Functions and how RBF models are used to approximate linear and non-linear multivariate functions. The procedure used to train an RBF model that maps the given input(s) of the system to be modelled to the output(s) of the system is also explained. The training procedure involves the optimisation of the position of the centres and the hyper-parameters of the RBF model, namely, the width(s) of the basis functions and the regularisation parameter.

An algorithm to create a non-linear variance estimate using four different approaches of maximum likelihood to optimise the coefficients of an RBF model fit to a linear ARCH estimate of the variance of a non-linear GARCH model is introduced. The accuracies of the four non-linear variance estimates generated as a result of the four different maximum likelihood approaches are compared to each other and also to a linear ARCH estimate of the true GARCH variance.

To test whether the algorithm can yield non-linear estimates of a GARCH variance with better accuracy than a linear ARCH estimate of the same variance, the algorithm is implemented on three different simulated non-linear variance models. Four variations of the maximisation of the log-likelihood of the one-step-ahead variance estimate (of an RBF model fit to an ARCH estimate of the true GARCH variance) are tested of which unconstrained maximisation of the log-likelihood of the one-step-ahead variance estimate for the validation set (Approach 1), unconstrained maximisation of the log-likelihood of the one-step-ahead variance estimate for the training and validation set (Approach 2) and constrained maximisation of the log-likelihood of the one-step-ahead variance estimate for the validation set (Approach 3) seem to yield estimates that are more accurate than a linear ARCH estimate. Of all the four approaches, only Approach 2 yields a non-linear variance estimate that has a lesser NRMSE (calculated vs. the true GARCH variance) than a linear ARCH(25) variance estimate for all the examples.

In reality, since the true variance is unavailable, the NRMSE between the true GARCH variance and the variance estimates generated by the four approaches of optimisation cannot be calculated. Hence, a method to select the most accurate non-linear variance estimate out of the four generated, that does not use the NRMSE (calculated vs. the true variance), must be devised. Or, since Approach 2 seems to provide more accurate non-linear variance estimates, it can be further tested on different non-linear GARCH models. Since there is no definite link between an approach that yields an estimate with the highest log-likelihood and an approach that yields an estimate with the least NRMSE (calculated vs. the true variance), the idea of selecting the approach that yields a non-linear variance estimate with the highest log-likelihood value is dismissed.

Chapter 8

Conclusions and Future Work

Modelling financial volatility is an important topic which has given rise to an enormous literature in applied financial economics. Many GARCH-class volatility/variance models have been used to model the volatility of financial return series across many markets. But compared to the vast effort that has gone into modelling the variance process, the means process is modelled in a very primitive way. The mean model is often ignored or a basic constant model is fitted to the mean. At the very best, a linear mean model is fitted. The possibility of the mean model being non-linear has only recently been explored in detail by Zhao (2010).

A powerful NARMAX based framework to model the means process accurately was developed (Zhao, 2010). Accurate term selection and parameter estimation was achieved when the underlying means process was non-linear in nature and the structure of the variance model was assumed to be known. This thesis extended Zhao's WOFR framework to incorporate the fact that the empirical variance is unknown, and suggested new methods for model validation. The impact of under-fitting the mean model on the accuracy of the variance estimate was also studied in detail.

The extended WOFR framework was used to analyse the means process of 2 real financial data sets. The results concluded that non-linear models were preferred over linear and constant models to describe the means process. The effects of under-fitting the mean model on the 2 data sets were also studied.

The conventional method to fit a variance model also did not involve term selection. The structure of the model is arbitrarily chosen based on preliminary tests on the data set, and the fitted variance model is then validated. If the model fails validation, a different model is selected. The need for a NARMAX based term selection framework for fitting a variance model was evident. A NARMAX based WLS approach was developed for fitting a variance model as well.

The WLS approach for fitting a variance model identifies the true variance model correctly only when the underlying variance is linear in nature. The method fails to accurately select the non-linear terms present in the true variance model. This is due to the fact that the true variance is unobservable, and to model the variance using the WLS approach when the true variance is unavailable, a linear estimate of the variance is used. Hence, the need for generating a non-linear estimate of the variance without making any assumptions about the structure of the variance model is identified. An RBF based method to generate a non-linear variance estimate is then introduced to address this problem.

8.1 Main Contributions

The main contributions of this thesis can be summarised as follows:

(i) In this thesis, the various GARCH-class volatility models used in the GARCH literature and the methods of estimation used have been summarised. The NARMAX based WOFR method for the term selection and parameter estimation of the mean model when the true variance is known is summarised. The effects of incorrectly fitting a linear mean model when the underlying means process is non-linear in nature are also studied. The variance estimates are found to be largely inaccurate, especially during periods of high volatility, when the mean model is under-fitted.

(ii) Zhao's WOFR method of fitting a mean model is extended to include fitting a linear noise model and to work when the structure of the variance model is unknown as well. Several tests including the ARCH Test and higher order non-linear correlation tests are also used to make term selection and model validation more robust. The working of the extended framework is first demonstrated on a simulated data set, the mean and variance model of which are known. The effects of under-fitting the mean model on the accuracy of variance estimates when the true structure of the variance model is unknown are also examined. The variance estimates are found to be largely inaccurate, especially during periods of high volatility, when the mean model is under-fitted.

(iii) The extended framework is then used to analyse 2 real financial data sets – the FTSE100 and the NASDAQ. The models that best describe the underlying mean process of both data sets are derived. The effects of fitting a simple constant mean model to these data sets on mean and variance estimation are also examined. The magnitude of the variance estimate generated after fitting a constant mean model is found to be larger than the variance estimate generated after fitting the mean model selected using the extended WOFR approach. This difference is especially prominent during high periods of volatility. This supports the notion that under-fitting the mean model causes the predictable elements of the means process to be added on to the residuals, which when used to generate a variance estimate, yields a largely inaccurate estimate of the variance, especially during periods of high volatility.

(iv) A WLS approach similar to the one used to accurately select the terms in the mean model (Zhao, 2010) is introduced to facilitate term selection for the variance model. The performance of the approach in various cases is analysed, and the results listed. The approach is found to work well in correctly selecting and estimating the terms of a linear variance model, when both, the true residuals and the true variance are unknown. However, the approach failed to correctly select the non-linear terms when the underlying variance was non-linear in nature. This failure is chalked up to the fact that a linear estimate of the variance is used in place of the true variance due to the unobservable nature of the true variance.

(v) An RBF based method for the generation of a non-linear estimate of the variance without making any assumptions about the true structure of the variance model was developed. The method involved fitting an RBF model to a linear estimate of the variance. The parameters of the RBF model were then re-estimated using maximum likelihood in order to yield a more accurate estimate of the variance. The method was tested on three different data sets. For each data set, the test involved comparing the accuracy of a linear variance estimate and the accuracy of a non-linear variance estimate. Typically, a residual series was first used to generate a non-linear variance series (referred to as the true variance). The same residual series was then used to generate a linear estimate of the variance. The newly developed RBF method was then used to generate a non-linear estimate of the variance from the residuals. The accuracy of the linear variance estimate was compared to that of the non-linear variance estimate. For all the three cases that this test was run on, the RBF based method yielded more accurate variance estimates.

8.2 Limitations

Some limitations of the work in this thesis are:

(i) The non-availability of the true variance process.

(ii) In Chapter 5, a non-linear model is proven to be better than a constant model to describe the mean process of a return series, but the constant mean model passes standard model validation tests, deeming it to be acceptable, even when the true mean process is non-linear in nature.

(iii) In Chapter 6, a linear estimate of the variance is used in place of the true variance. However, non-linear variance models cannot be identified using the WLS approach for the selection of the variance model introduced in Chapter 6 due to the fact that a linear estimate of the variance is used as a proxy for the true variance.

(iv) In Chapter 7, four approaches to obtain a non-linear RBF estimate of the variance are investigated. The NRMSE of the variance estimates generated using the four methods are calculated against the true variance to select the estimate that is the closest to the true variance. As mentioned before, however, because the true variance is unavailable, the log-likelihood function values of the variance estimates cannot be used to select the best variance estimate, since results show that the best variance estimate does not necessarily have the highest log-likelihood function value.

8.3 Suggestions for Future Research

The suggestions for future research are:

(i) Alternative proxies for the true variance, like implied volatility and realised volatility, can be looked into. However, there are a few obstacles to overcome. The implied volatility series is model dependent and many different implied volatility series can be derived for a given data set. To extract a legitimate and usable implied volatility series, the given options pricing data needs to be cleaned a lot and can only be obtained from a data set that has been obtained during a well traded market (high trade volumes). This is very time consuming. In this thesis, the frequency of price and returns data used is daily. A realised volatility series is obtained from high frequency data (1 minute, 5 minute or 15 minute intervals). The basic realised volatility of a day is calculated by summing up the high frequency returns for the entire day. More advanced formulae exist too.

(ii) The framework developed in Chapter 4 is implemented on daily returns of stocks in Chapter 5, but can be extended to high frequency (5 minute, 15 minute etc.) returns. High frequency data exhibits completely different behaviour than that described in this thesis and the returns data are highly heteroskedastic, and hence, the models used to model the returns and the variance will be different.

(iii) The method developed and showcased in Chapters 4 and 5 for one period ahead forecasting can be extended to multi-period forecasting and density forecasting. This can be achieved by using data of lower frequency (weekly or monthly data). The magnitude of underlying variance in lower frequency returns data is much less than that observed in daily returns data. Hence, the returns data is less heteroskedastic and the mean and variance modelling techniques introduced in this thesis will work better.

(iv) This latter extension in turn suggests the need to look into back-testing (or forecast evaluation) methods appropriate to multi-period models. Forecast performance will need to be evaluated in order to determine the accuracy of models fitted on different data sets, which in turn will indicate how well the model selection and estimation method works.

(v) A non-linear variance estimate generated using the RBF method introduced in Chapter 7 can be combined with the WLS approach to fit a variance model introduced in Chapter 6. Accuracy of term selection using this non-linear variance estimate when the underlying true variance is non-linear in nature can then be investigated. Ideally, the non-linear terms in the variance model need to be selected when a non-linear variance estimate using the method described in Chapter 7 is used as a proxy for the true variance.

References

Abhyankar, A., Copeland, L. S. and Wong, W. (1997) 'Uncovering non-linear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100', *Journal of Business and Economic Statistics,* 15(1), 1-14.

Aguirre, L. A., Rodrigues, G. G. and Jacome, C. R. F. (1998) 'Identifying non-linear systems using NARMAX polynomial models - a review of novel results', *Identificacao de sistemas nao-lineares utilizando modelos NARMAX polinomiais - uma revisao e novos resultados,* 9(2), 90-106.

Akaike, H. (1974) 'A new look at the statistical model identification', *Automatic Control, IEEE Transactions on,* 19(6), 716-723.

Alberg, D., Shalit, H. and Yosef, R. (2008) 'Estimating stock market volatility using asymmetric GARCH models', *Applied Financial Economics,* 18(15), 1201-1208.

Alexander, C. (2001) *Market models : A guide to financial data analysis,* Chichester: Wiley.

Al-Loughani, N. and Chappell, D. (2001) 'Modelling the day-of-the-week effect in the Kuwait Stock Exchange: a non-linear GARCH representation', *Applied Financial Economics,* 11(4), 353 - 359.

Andersen, T. G. and Bollerslev, T. (1998) 'Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts', *International Economic Review,* 39(4), 885-905.

Angelidis, T. and Degiannakis, S. (2008) 'Volatility forecasting: Intra-day versus inter-day models', *Journal of International Financial Markets, Institutions and Money,* 18(5), 449-465.

Bai, J. S. and Chen, Z. H. (2008) 'Testing multivariate distributions in GARCH models', *Journal of Econometrics,* 143(1), 19-36.

Billings, S., Chen, S. and Korenberg, M. (1989) 'Identification of MIMO non-linear systems using a forward-regression orthogonal estimator', *International Journal of Control*, 49(6), 2157-2189.

Billings, S., Korenberg, M. and Chen, S. (1988) 'Identification of non-linear output-affine systems using an orthogonal least-squares algorithm', *International Journal of Systems Science,* 19(8), 1559-1568.

Billings, S. A. and Coca, D. (2001) *Identification of NARMAX and related models*, Sheffield: University of Sheffield, Department of Automatic Control and Systems Engineering.

Billings, S. A., Guo, L. Z. and Zhu, D. Q. (2008) 'An extended orthogonal forward regression algorithm for system identification using entropy', *International Journal of Control,* 81(4), 690-699.

Billings, S. A., Tsang, K. M. and Tomlinson, G. R. (1987) *Application of the NARMAX method to non-linear frequency response estimation*, Sheffield: University of Sheffield, Dept. of Automatic Control and Systems Engineering.

Billings, S. A. and Voon, W. S. F. (1986) 'Correlation based model validity tests for non-linear models', *International Journal of Control,* 44(1), 235-244.

Billings, S. A. and Wei, H. L. (2005a) 'A new class of wavelet networks for non-linear system identification', *IEEE Transactions on Neural Networks,* 16(4), 862-874.

Billings, S. A. and Wei, H. L. (2005b) 'The wavelet-NARMAX representation: A hybrid model structure combining polynomial models with multiresolution wavelet decompositions', *International Journal of Systems Science*, 36(3), 137-152.

Billings, S. A. and Wei, H. L. (2008) 'An adaptive orthogonal search algorithm for model subset selection and non-linear system identification', *International Journal of Control,* 81(5), 714-724.

Billings, S. A. and Zhu, Q. M. (1994) 'Non-Linear model validation using correlation tests', *International Journal of Control,* 60(6), 1107-1120.

Bjorck, A. (1996) *Numerical methods for least squares problems,* Society for Industrial Mathematics.

Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics,* 31(3), 307-327.

Bollerslev, T. (2008) *Glossary to ARCH (GARCH),* School of Economics and Management, University of Aarhus.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994) 'Chapter 49 Arch models' in Robert, F. E. and Daniel, L. M., eds., *Handbook of Econometrics,* Elsevier, 2959-3038.

Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B. (2001) 'A test for independence based on the correlation dimension', *Growth theory, non-linear dynamics, and economic modelling: scientific essays of William Allen Brock,* 15(3), 324.

Broomhead, D. S. and Lowe, D. (1988) *Radial basis functions, multi-variable functional interpolation and adaptive networks,* DTIC Document.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets,* Princeton, NJ: Princeton University Press.

Cao, C. Q. and Tsay, R. S. (1992) 'Nonlinear time-series analysis of stock volatilities', *Journal of Applied Econometrics,* 7(S1), S165-S185

Chang, S. L. and Ogunfunmi, T. (1997) 'Recursive orthogonal least squares method and its application in non-linear adaptive filtering', in Anon, ed. *IEEE Military Communications Conference MILCOM,* Monterey, CA, USA, 3-5 Nov 1997, New Jersey: IEEE, 1392-1396.

Chelley-Steeley, P. L. and Steeley, J. M. (2005) 'The leverage effect in the UK stock market', *Applied Financial Economics,* 15(6), 409-423.

Chen, C. W. S. and Yu, T. H. K. (2005) 'Long-term dependence with asymmetric conditional heteroscedasticity in stock returns', *Physica A: Statistical Mechanics and its Applications,* 353(1-4), 413-424.

Chen, S. and Billings, S. A. (1989) 'Representations of non-linear systems - The NARMAX Model', *International Journal of Control,* 49(3), 1013-1032.

Chen, S., Billings, S. A., Cowan, C. F. N. and Grant, P. M. (1990) 'Practical identification of NARMAX models using radial basis functions', *International Journal of Control,* 52(6), 1327-1350.

Chen, S., Billings, S. A. and Luo, W. (1989) 'Orthogonal least squares methods and their application to non-linear system identification', *International Journal of Control,* 50(5), 1873-1896.

Chong, C. W., Ahmad, M. I. and Abdullah, M. Y. (1999) 'Performance of GARCH models in forecasting stock market volatility', *Journal of Forecasting,* 18(5), 333-343.

Choudhry, T. (2001) 'Month of the year effect and January effect in pre-WWI stock returns: Evidence from a non-linear GARCH model', *International Journal of Finance & Economics,* 6(1), 1-11.

Coca, D. (1996) *Class of wavelet multiresolution decompositions for non-linear system identification and signal processing*, Unpublished thesis (PhD), University of Sheffield.

Coca, D. and Billings, S. A. (2001a) 'Non-Linear system identification using wavelet multiresolution models', *International Journal of Control,* 74(18), 1718-1736.

Coca, D. and Billings, S. A. (2001b) *System identification using wavelets,* Sheffield: University of Sheffield, Department of Automatic Control and Systems Engineering.

Engle, R. F. (1982) 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica,* 50(4), 987-1007.

Engle, R. F. (2001) 'GARCH 101: The use of ARCH/GARCH models in applied econometrics', *Journal of Economic Perspectives,* 15(4), 157-168.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987) 'Estimating time varying risk premia in the term structure: the ARCH-M model', *Econometrica: Journal of the Econometric Society,* 391-407.

Engle, R. F. and Ng, V. K. (1993) 'Measuring and testing the impact of news on volatility', *Journal of Finance,* 48(5), 1749-1778.

Fox, L. (1964) *An introduction to numerical linear algebra,* Clarendon Press Oxford.

Franses, P. H. and Van Dijk, D. (1996) 'Forecasting stock market volatility using (non-linear) Garch models', *Journal of Forecasting,* 15(3), 229-235.

Friederich, U. (2011) *Nonlinear System Identification and Analysis of Adaptive Neural Information Processing in the Fly Early Visual System*, Unpublished thesis (PhD), The University of Sheffield.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993) 'On the relation between the expected value and the volatility of nominal excess return on stocks', *Journal of Finance,* 48(5), 1779-1801.

Gokcan, S. (2000) 'Forecasting volatility of emerging stock markets: Linear versus non-linear GARCH models', *Journal of Forecasting,* 19(6), 499-504.

Graps, A. (1995) 'An Introduction to Wavelets', [online], available: http://www.amara.com/IEEEwave/IEEEwavelet.html [accessed 11th February 2010]

Guo, L. Z. and Billings, S. A. (2007) 'A modified orthogonal forward regression least-squares algorithm for system modelling from noisy regressors', *International*

Journal of Control, 80(3), 340-348.

Guo, L. Z. and Billings, S. A. (2008) 'Identification and analysis of spatio-temporal dynamical systems using wavelets', *International Journal of Systems Science,* 39(3), 315-334.

Harrison, P. (1967) 'Exponential smoothing and short-term sales forecasting', *Management Science,* 13(11), 821-842.

Heston, S. L. and Nandi, S. (2000) 'A closed-form GARCH option valuation model', *Review of financial studies,* 13(3), 585-625.

Hinich, M. J. and Patterson, D. M. (1985) 'Evidence of Non-Linearity in Daily Stock Returns', *Journal of Business & Economic Statistics,* 3(1), 69-77.

Holt, M. T. and Aradhyula, S. V. (1990) 'Price risk in supply equations: An application of GARCH time-series models to the U.S. broiler market', *Western Journal of Agricultural Economics,* 57(1), 230-242.

Hung, J. C. (2009a) 'A Fuzzy Asymmetric GARCH model applied to stock markets', *Information Sciences,* 179(22), 3930-3943.

Hung, J. C. (2009b) 'A fuzzy GARCH model applied to stock market scenario using a genetic algorithm', *Expert Systems with Applications*, 36(9), 11710-11717.

Jondeau, E. and Rockinger, M. (2003) 'Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements', *Journal of Economic Dynamics and Control,* 27(10), 1699-1737.

Kim, M. B. (1993) 'Forecasting the volatility of financial markets - ARCH GARCH models and the AUTOREG Procedure', in *Sugi 18: Proceedings of the Eighteenth Annual Sas Users Group International Conference,* New York, USA, 9-12 May 1993, SAS Publishing, 304-312.

Korenberg, M., Billings, S. A., Liu, Y. P. and Mcilroy, P. J. (1988) 'Orthogonal parameter-estimation algorithm for non-linear stochastic-systems', *International Journal of Control*, 48(1), 193-210.

Lanne, M. and Saikkonen, P. (2005) 'Non-Linear GARCH models for highly persistent volatility', *Journal of Econometrics*, 8(2), 251-276.

LeBaron, B. (1992) 'Some relations between volatility and serial correlations in stock market returns', *Journal of Business*, 65(2), 199-219.

Lee, K. L. and Billings, S. A. (2002) 'Effects of noise reduction on the prediction accuracy of time series', *International Journal of Systems Science*, 33(15), 1207-1216.

Lee, K. L. and Billings, S. A. (2003) 'A new direct approach of computing multi-step ahead predictions for non-linear models', *International Journal of Control*, 76(8), 810-822.

Leontaritis, I. J. and Billings, S. A. (1985a) 'Input-output parametric models for non-linear systems Part I: deterministic non-linear systems', *International Journal of Control*, 41(2), 303-328.

Leontaritis, I. J. and Billings, S. A. (1985b) 'Input-output parametric models for non-linear systems Part II: stochastic non-linear systems', *International Journal of Control*, 41(2), 329-344.

Li, L. M. (2000) *Continuous time non-linear system identification*, Unpublished thesis (PhD), University of Sheffield.

Li, L. M. and Billings, S. A. (2001) 'Continuous time non-linear system identification in the frequency domain', *International Journal of Control*, 74(11), 1052-1061.

Lima, P. J. F. d. (1998) 'Non-Linearities and Nonstationarities in Stock Returns', *Journal of Business & Economic Statistics*, 16(2), 227-236.

Lin, K. (1997) 'The ABC's of BDS', *Journal of Computational Intelligence in Finance*, 5(4), 23-26.

Ling, S. Q. (2007) 'Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models', *Journal of Econometrics*, 140(2), 849-873.

Liu, G. P., Billings, S. A. and Kadirkamanathan, V. (2000) 'Non-Linear system identification using wavelet networks', *International Journal of Systems Science*, 31(12), 1531-1541.

Liu, W. and Morley, B. (2009) 'Volatility forecasting in the hang seng index using the GARCH approach', *Asia-Pacific Financial Markets*, 16(1), 51-63.

Liu, Y. P. (1988) *Identification of non-linear systems : the NARMAX polynomial approach*, Unpublished thesis (PhD), University of Sheffield.

Lo, A. W. and MacKinlay, A. C. (1988) 'Stock market prices do not follow random walks: Evidence from a simple specification test', *Review of financial studies*, 1(1), 41-66.

Mao, K. Z. (1998) *Non-Linear systems identification using the NARMAX method*, Unpublished thesis (PhD), University of Sheffield.

Mao, K. Z. and Billings, S. A. (1997) 'Algorithms for minimal model structure detection in non-linear dynamic system identification', *International Journal of Control*, 68(2), 311-330.

Matias, J. M., Febrero-Bande, M., Gonzalez-Manteiga, W. and Reboredo, J. C. (2010) 'Boosting GARCH and neural networks for the prediction of heteroskedastic time series', *Mathematical and Computer Modelling*, 51(3-4), 256-271.

McMillan, D. G. and Garcia, R. Q. (2009) 'Intra-day volatility forecasts', *Applied Financial Economics*, 19(8), 611-623.

McMillan, D. G. and Speight, A. E. H. (2004) 'Daily volatility forecasts: Reassessing the performance of GARCH models', *Journal of Forecasting*, 23(6), 449-460.

McMillan, D. G., Speight, A. E. H. and Apgwilym, O. (2000) 'Forecasting UK stock market volatility', *Applied Financial Economics,* 10(4), 435-448.

Meitz, M. and Saikkonen, P. (2008) 'Stability of non-linear AR-GARCH models', *Journal of Time Series Analysis,* 29(3), 453-475.

Mendes, E. M. A. M. and Billings, S. A. (2001) 'An alternative solution to the model structure selection problem', *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.,* 31(6), 597-608.

Mohamed Vall, O. M. and M'Hiri, R. (2008) 'An approach to polynomial NARX/NARMAX systems identification in a closed-loop with variable structure control', *International Journal of Automation and Computing,* 5(3), 313-318.

Morimoto, T. and Kawasaki, Y. (2005) 'An empirical comparison of GARCH models based on intraday Value at Risk' in Simos, T. and Maroulis, G., eds., *Advances in Computational Methods in Sciences and Engineering 2005, Vols 4 A & 4 B*, Greece: Brill Academic Publications, 1299-1302.

Nakatsuma, T. and Tsurumi, H. (1996) 'ARMA-GARCH models: Bayes estimation versus MLE, and Bayes non-stationarity test' in *American Statistical Association - 1996 Proceedings of the Section on Bayesian Statistical Science*, 343-348.

Nason, G. P. (1999) 'A little introduction to wavelets', *IEE Colloquium (Digest),* (63), 1-6.

Nelson, D. B. (1991) 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica,* 59(2), 347-370.

Nelson, D. B. (1992) 'Filtering and forecasting with misspecified ARCH models I : Getting the right variance with the wrong model', *Journal of Econometrics,* 52(1-2), 61-90.

Orr, M. J. L. (1996) 'Introduction to radial basis function networks', *University of*

Edinburgh.

Pintelon, R. and Schoukens, J. (2001) *System Identification : a frequency domain approach,* New York: IEEE Press.

Poon, S. H. and Granger, C. W. J. (2003) 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature,* 41(2), 478-539.

Popov, A. A. and Bykhanov, K. V. (2005) 'Modeling volatility of time series using Fuzzy GARCH models', in *9th Russian-Korean International Symposium on Science and Technology, KORUS-2005*, Novosibirsk, 26 June - 2 July 2005, 687-692.

Powell, M. J. D. (1987) 'Radial basis functions for multivariable interpolation: a review' in *Algorithms for approximation,* Clarendon Press, 143-167.

Rahrooh, A. and Shepard, S. (2009) 'Identification of non-linear systems using NARMAX model', *Non-Linear Analysis, Theory, Methods and Applications,* 71(12), e1198-e1202.

Scholes, M. and Williams, J. (1977) 'Estimating betas from nonsynchronous data', *Journal of financial economics,* 5(3), 309-327.

Seferoglu, A. M. (1996) *Forecasting volatility using ARCH and GARCH modelling,* Unpublished thesis (PhD), University of Sheffield.

Shi, H., Cai, Y. and Qiu, Z. (2005) 'Improved system identification approach using wavelet networks', *Journal of Shanghai University (English Edition),* 9(2), 159-163.

Storti, G. (2008) 'Modelling asymmetric volatility dynamics by multivariate BL-GARCH models', *Statistical Methods and Applications,* 17(2), 251-274.

Terui, N. and Kariya, T. (1997) 'Gaussianity and Nonlinearity of Foreign Exchange Rates', *Lecture Notes-Monograph Series,* 337-350.

Thomaidis, N. S. and Dounias, G. (2008) 'A hybrid intelligent system for financial time-series forecasting', *Engineering Intelligent Systems for Electrical Engineering and Communications,* 16(4), 193-213.

Tofallis, C. (2009) 'Least squares percentage regression', *Journal of Modern Applied Statistical Methods.*

Tsay, R. S. (2002) *Analysis of financial time series : financial econometrics, Wiley series in probability and statistics,* New York: Wiley.

Tseng, C. H., Cheng, S. T. and Wang, Y. H. (2009) 'New hybrid methodology for stock volatility prediction', *Expert Systems with Applications,* 36(2 Part 1), 1833-1839.

Urrutia, J. L., Vu, J., Gronewoller, P. and Hoque, M. (2002) 'Non-Linearity and Low Deterministic Chaotic Behavior in Insurance Portfolio Stock Returns', *The Journal of Risk and Insurance,* 69(4), 537-554.

Walczak, B. and Massart, D. (2000) 'Local modelling with radial basis function networks', *Chemometrics and Intelligent Laboratory Systems,* 50(2), 179-198.

Wang, J. (2009) 'Stock market volatility and the forecasting performance of stock index futures', *Journal of Forecasting,* 28(4), 277-292.

Wang, Y. H. and Lin, C. T. (2008) 'Forecasting volatility for the stock market: A new hybrid model', *International Journal of Computer Mathematics,* 85(11), 1697-1707.

Wei, H. L. (2004) *A wavelet-based approach for non-linear system identification and non-stationary signal processing,* Unpublished thesis (PhD), University of Sheffield.

Wei, H. L., Balikhin, M. A. and Billings, S. A. (2003) *Non-Linear time-varying system identification using the NARMAX model and multiresolution wavelet expansions*, Sheffield: University of Sheffield, Department of Automatic Control and Systems Engineering.

Wei, H. L. and Billings, S. A. (2002) 'Identification of time-varying systems using multiresolution wavelet models', *International Journal of Systems Science,* 33(15), 1217-1228.

Wei, H. L. and Billings, S. A. (2006) 'An efficient non-linear cardinal B-spline model for high tide forecasts at the Venice Lagoon', *Non-Linear Processes in Geophysics,* 13(5), 577-584.

Wei, H. L. and Billings, S. A. (2006) 'Long term prediction of non-linear time series using multiresolution wavelet models', *International Journal of Control,* 79(6), 569-580.

Wei, H. L., Billings, S. A. and Balikhin, M. A. (2006) 'Wavelet based non-parametric NARX models for non-linear input-output system identification', *International Journal of Systems Science,* 37(15), 1089-1096.

Wei, H. L., Billings, S. A. and Liu, J. (2004) 'Term and variable selection for non-linear system identification', *International Journal of Control,* 77(1), 86-110.

Wei, H. L., Lang, Z. Q. and Billings, S. A. (2008) 'Constructing an overall dynamical model for a system with changing design parameter properties', *International Journal of Modelling, Identification and Control,* 5(2), 93-104.

Weisstein, E. W. (2010) '"Kronecker Delta." From MathWorld--A Wolfram Web Resource', [online], available: http://mathworld.wolfram.com/KroneckerDelta.html [accessed 8th September 2012]

Wenbing, X., Qi, F. and Weimin, C. (2008) 'Forecasting Chinese stock markets volatility based on neural network combining', in *4th International Conference on Natural Computation, ICNC 2008,* Jinan, 18-20 Oct 2008, 23-27.

Willey, T. (1992) 'Testing for non-linear dependence in daily stock indices', *Journal of Economics and Business,* 44(1), 63-76.

Yim, J. (2002) 'A comparison of neural networks with time series models for forecasting returns on a stock market index' in Hendtlass, T., ed. *Developments in Applied Artificial Intelligence, Proceedings*, 25-35.

Yahoo! Finance (2013a) '^FTSE Historical Prices | FTSE 100 Stock - Yahoo! Finance', [online], available: http://finance.yahoo.com/q/hp?s=%5EFTSE+Historical+Prices [accessed 30th July 2013]

Yahoo! Finance (2013b) '^IXIC Historical Prices | NASDAQ Composite Stock - Yahoo! Finance', [online], available: http://finance.yahoo.com/q/hp?s=%5EIXIC+Historical+Prices [accessed 30th July 2013]

Zhao, L. (2010) *System Identification for Complex Financial System*, Unpublished thesis (PhD), The University of Sheffield.

Zheng, G. L. and Billings, S. A. (1996) 'Radial basis function network configuration using mutual information and the orthogonal least squares algorithm', *Neural Networks*, 9(9), 1619-1637.

Zhu, Q. M. and Billings, S. A. (1992) *A comparison of polynomial and rational NARMAX models for non-linear system identification*, Sheffield: University of Sheffield, Dept. of Automatic Control and Systems Engineering.