Model Selection - Regression and Time Series Applications

A thesis presented

by

Allan Ernest Clark

to

The Department of Statistical Sciences

in fulfillment of the requirements

for the degree of

Masters in Science

in the subject of

Statistics

University of Cape Town

August 2003

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Acknowledgments

I wish to express my sincere thanks to the following:

* *To my supervisor, Professor Casper Troskie, for always making time to answer my many questions, supplying me with many good ideas, all the data used in this thesis and encouraging me to pursue further studies*
* To *Professor Stewart, who initially encouraged me to pursue postgraduate studies*
* *To Professor Dunne, for proofreading a large portion of the initial work*
* *To the members of the stats department (UCT), Dr Christien Thiart, Freedom Gumedze and Professor Barr in particular*
* *To Dominique Katshunga who often helped me when I was stuck*
* *To my fellow Office members*
* *To my family and*
* *The National Research Foundation, DAAD and UCT for providing support during my studies.*

Synopsis

In any statistical analysis the researcher is often faced with the challenging task of gleaning relevant information from a sample data set in order to answer questions about the area under investigation. Often the exact data generating process that governs any data set is unknown, indicating that we have to estimate the data generating process by using statistical methods. Regression analysis and time series analysis are two statistical techniques that can be used to undertake such an analysis.

In practice researcher will propose one model or a group of competing models that attempts to explain the data being investigated. This process is known as model selection. Model selection techniques have been developed to aid researchers in finding a suitable approximation to the true data generating process. Methods have also been developed that attempt to distinguish between different competing models. Many of these techniques entail using an information criterion that estimates the "closeness" of a fitted model to the unknown data generating process. This study investigates the properties of Bozdogan's Information complexity measure (ICOMP) when undertaking time series and regression analysis.

Model selection techniques have been developed for both time series and regression analysis. The regression analysis techniques however often provide unsatisfactory results due to poor experimental designs. Poor experimental design could induce collinearities causing parameter estimates to become unstable with large standard errors. Time series analysis utilizes lagged autocorrelation- and lagged partial autocorrelation coefficients in order to specify the lag structure of the model. In certain data sets this process is not informative in determining the order of an ARIMA model. ICOMP guards against collinearity by considering the interaction between the parameters being estimated in a model.

This study investigates the properties of ICOMP when undertaking regression and time series analysis by means of a simulation study. The regression simulation study investigates the behaviour of ICOMP, AIC and BIC under various collinearity-, sample size- and residual variance- levels extending the model by Bozdogan and Haughton (1998). The time series simulation study investigates the behaviour of ICOMP and

other information criteria in a time series context. The study entails simulating stationary ARMA and GARCH models 1000 times and then fitting different time series models to the simulated series. Different series will be considered by changing the size of the residual variance. We consider all subset models that have at most three AR, MA, ARCH and GARCH terms. We then investigate the frequency of selecting the different models by each of the following information criteria: maximum log likelihood, AIC, BIC, ICOMP(IFIM), ICOMP(VanEmden) and COMP.

As an application of model selection we examine the stability of the correlation and the covariance structure of share returns of the largest market capitalization companies listed on the JSE. We use the model selection techniques developed in the study in order to model the conditional variance of share returns during the period January 1990- December 2000.

It was found that the performance of ICOMP in selecting the Kullback Leibler model is dependent on the degree of collinearity of the design matrix, the sample size, the amount of variation in the data set and the residual variance level.

When the variation in the design matrix is large it was found that as the collinearity levels in the design matrix increased, that the agreement percentages for all of the information criteria decreased monotonically and that the ICOMP criterion agreed with the Kullback Leibler model selected more often than both AIC and BIC. As the residual variance increases, the agreement percentages of all of the information criteria decreases, however as the sample size increased the agreement percentages of all information criteria increased.

When the variation in the design matrix is low it was found that when the collinearity is low, as the residual variance increases, the agreement percentages for all of the information criteria decreases monotonically such that the ICOMP criterion agreed more often with the model selected by the Kullback Leibler criterion than both AIC and BIC. When the collinearity in the design matrix is moderate, the agreement percentages for the ICOMP criterion and BIC decreases monotonically as the residual variance increased for all sample sizes considered. AIC however behaved differently. When the sample size is large, it behaves similar to the ICOMP criterion and BIC, however when the residual variance is increased from 2.5 to 5, the agreement percentage for AIC decreases. Note however that the agreement percentages are less than

the levels observed when the residual variance is low and that the ICOMP criterion selects the Kullback Leibler model more often than both AIC and BIC for all residual variance levels considered except when the residual variance is large and the sample size is small.

When applying ICOMP to time series data it was found that none of the information criteria considered selected the correct model under all of the different conditions. It is suggested that no single information criteria should be used independently of other information criteria. Any modelling decision should involve a careful examination of all of the different models proposed taking into consideration not only technical issues such as model fit, tests for significant parameters and correlated errors but we should also consider the economic plausibility of any proposed model.

As an application of model selection, this study investigated the stability of the covariance- and the correlation-matrix of thirty three companies as well as the JSE All share return index. It was found that both the covariance- and the correlation-matrix does appear to change over time. An attempt was made to model the conditional variance of these series by utilising GARCH models. It was found that the normality assumption for the error distribution was not appropriate and that in certain cases a standardised t distribution was more appropriate.

It is recommended that further simulation studies be undertaken in order to assess the performance of ICOMP in the presence of outliers and influential observations when the design matrix is highly collinear. Further investigation into the simultaneous use of time series analysis and regression analysis is warranted.

In order to assess the stability of the covariance- and correlation- matrix of South African share returns, it is recommended that the analysis found in this study be undertaken on monthly data. Future studies should also incorporate the use of multivariate GARCH models in order to model the covariances of the share returns simultaneously.

Contents

Chapter 1

Introduction

1.1 Background to the Problem

In any statistical analysis the researcher is often faced with the challenging task of gleaning relevant information from a sample data set in order to answer certain questions about the content of the data set being investigated. The process of investigation is known as statistical analysis. Often the exact data generating process that governs any data set is unknown, indicating that we have to estimate the data generating process by using statistical methods. Regression analysis and time series analysis (amongst others) are two statistical techniques that can be used to undertake such an analysis.

Often the researcher will propose one model or a group of competing models that attempts to explain the data being investigated. Model selection techniques have been developed to aid researchers in finding a suitable approximation to the true data generating process that attempts to distinguish between different competing models. Many of these techniques entails using a information criteria that estimates the *"closeness"* of a fitted model to the unknown data generating process. This idea was first proposed by Kullback and Leibler (1951) and later developed by Akaike (1973), Sugiura (1978), Sakamoto *et al.* (1986), Takeuchi (1976) and Rissanen (1989).

In a regression context, model building entails selecting those variables that are deemed important to the area under investigation. Draper and Smith (1966) adds that variable selection should be undertaken so as to provide a linear model that is *"useful for prediction purposes and includes as many variables as possible so as to provide adequate fitted values for a data set."* It is however stressed that researchers should consider the cost of acquiring information about the variables to be included

in the final model. In general variable selection entails making a compromise between the last two points since the monitoring of many variables may be too expensive. Miller (1990) notes the importance of finding a small subset of variables that provides adequate fit and precision. The following variable selection techniques are the most popular namely, (1) All Possible Regressions, (2) Stepwise Procedures and (3) Information criteria such as AIC and BIC. In a time series context, model building often involves investigating which lag structures of a data set might be appropriate. This involves the use of lagged autocorrelations, lagged partial autocorrelations and an examination of the residuals structure of the proposed model.

1.2 Statement of the Problem

When one utilises OLS multiple regression to model Y as a linear combination of p independent variables, $\hat{\beta} = (X'X)^{-1} X'Y$ is used as an estimate of the beta coefficients in the model $Y = X\beta + e$. This estimate is an unbiased estimate of β and it can be shown that $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$ if $e \sim N(0, \sigma^2 I)$ (Wetherill (1986)). However,when $X'X$ is singular or near-singular one cannot perform the inversion of $X'X$ and the normal equations $X'X\beta = X'Y$ do not have a unique solution. Parameter estimates are often unstable with large standard errors. This problem occurs since there exists at least one exact linear relationship or near linear relationship between some of the columns of the X matrix and it is said that the design matrix, X is ill-conditioned. Biased regression techniques such as ridge,- principal components,- latent roots- regression and L_p estimation have been proposed in order to attempt to solve the collinearity problem. (Draper and Smith (1981))

Occams Razor (Brewer and Room (1999)) suggests that a model selection criteria should ensure that **simpler** models are preferred to more **complex** ones. Simpler in this case suggests **parsimony** in terms of the number of variables included in a model, where as complex suggests a **high dimensional model**. Many information criteria select a model by maximising the log likelihood of a model and penalising it by some scalar value. Some penalises a model based on the number of variables in the model whereas others penalises the model based on the sample size and the number of variables in the model. These penalty functions aim to penalise more complex models. Notice however that they do not consider the issue of collinearity. Maximum

log likelihood based information criteria for example, will often select models that contain too many variables even if the design matrix is ill-conditioned.

Numerous authors have investigated the **complexity** of statistical models leading to the development of numerous definitions of the complexity of a statistical model. (for example, "Kolmogorov Complexity" (Cover, *et al.* (1989), "Shannon complexity" (Rissanen (1989)) and "Stochastic Complexity " (Rissanen (1986)). Bozdogan and Bearse (1999) defines the complexity of a system as " *a measure of the degree of interdependency between the whole system and a simple enumerative composition of its subsystems or parts.*" This definition suggests that the complexity of a model is not based solely on the number of variables included in a model but rather on how they are **related** with each other. Bozdogan and Haughton (1998) has used the above definition and derived a number of Information Complexity Criteria (ICOMP) used for model selection and states that ICOMP guards against collinearity.

1.3 Objectives of the Study

The main objective of the study is to investigate the properties of ICOMP when undertaking time series and regression analysis.

1.4 Limitations of the Study

This study investigates the properties of ICOMP when undertaking time series and regression analysis. This study does not consider the vast amount of literature available that discusses outlier analysis, the detection of influential observations, intervention analysis and thus does not investigate the performance of ICOMP in the presence of outliers and influential observations.

The performance of ICOMP when undertaking time series and regression analysis is investigated separately. The study does not investigate models in which a response vector's mean equation is modelled by both fixed explanatory variables and lagged response variables. (i.e. mixing regression and time series analysis.)

The regression and time series simulations are not exhaustive and thus any conclusions drawn should not be generalised to conditions that were not investigated in

this study.

The volatility application in the final chapter focuses specifically on the South African stock market during the period January 1990 to December 2000. The South African market is the 15th largest stock market in the world and the largest in Africa. It is classified as being an Emerging Market and thus behaves differently to many of the larger stock exchanges around the world. Thin trading is particularly severe in the South African stock market. This thesis does not explicitly account for this problem. Univariate GARCH models were used in order to model the time varying nature of the conditional variance of the share returns. Multivariate GARCH models were however not estimated. These models would model the joint evolution of the conditional covariances and conditional variances dynamically. Monthly share returns were not investigated indicating that the results found from the weekly share returns series should not be generalised to the monthly return case.

1.5 Plan of Development of the Thesis

This thesis is divided into five sections. Each section includes its own set of appendices and are included at the end of each chapter in each section. All references however are displayed in section five.

Part one introduces many of the concepts found in time series analysis. It discusses the different components of a time series by highlighting different techniques in order to gain a preliminary understanding of the variables of interest. It provides an examination of ARIMA models data series.

Part two provide a brief introduction to Volatility modelling.

Part three discusses model selection and introduces Bozdogan's Information Complexity criteria.

Part four investigates the properties of ICOMP when undertaking regression and time series analysis by means of a simulation study. The regression simulation study investigates the behaviour of ICOMP, AIC and BIC under various collinearity-, sample size- and residual variance- levels extending Bozdogan and Haughton (1998). The

time series simulation study investigates the behaviour of ICOMP and other information criteria in a **time series** context. The study entails simulating stationary ARMA and GARCH models 1000 times and then fitting different time series models to the simulated series. Different series will be considered by changing the size of the residual variance. We consider all subset models that have at most three AR, MA, ARCH and GARCH terms. We then investigate the frequency of selecting the different models by each of the following information criteria: maximum log likelihood, AIC, BIC, ICOMP(IFIM), ICOMP(VanEmden) and COMP.

Part five examines the stability of the correlation and the covariance structure of share returns of the largest market capitalization companies listed on the JSE. This section investigates the problem by utilising the model selection techniques developed in the study.

Part I

Time Series Analysis

Chapter 2

Fundamental Concepts

A *time series* is a sequence of observations ordered sequentially in time. The measurements could be taken continuously in time (e.g. an individual's brain activity measured from an EEG machine) or they could be taken at discrete time points (e.g. the closing price of a share each day). Often one observes a continuous variable or a stochastic process (e.g. the daily temperature at a place in South Africa, or the price of gold) and discretises the process by only recording the series at certain set times. Note however that the discretisation process could also occur if one aggregates a data set over a period in time. For example, *when modelling the monthly sales of a store, the daily sales level could be aggregated in order to generate a monthly total sales value. This monthly figure could then be used in order to make predictions about the monthly sales value for the next few months.*

The observations in a data set may be collected such that observations are equally spaced in time or they could be unequally spaced. Unequally spaced data could occur if there is some missing observations in a data set or if there is no underlying sampling interval (e.g. *the modelling of rare events*). The variable under investigation (Z_t) may be **continuous** (sea surface temperatures Z, at time t), or **discrete** (counts Z of living persons in a specified region, at time t), or **qualitative** (category Z at time t: non-student, student, graduate, employee, post-graduate student).

This thesis will only attempt to solve the former problem although models for unequally spaced data are available. (Krishnaiah and Rao (1988), Priestley (1993)) In the following three chapters we present methods for building, identifying, fitting, and checking models to time series data. Different time series models will be examined theoretically although useful practical examples that are relevant to the South African financial market place will be provided.

This chapter introduces many of the concepts found in time series analysis. It discusses the different components of a time series by highlighting different techniques in order to gain a preliminary understanding of the variables of interest. Chapter three examines the models by which to handle stationary data series and Chapter four will provide a brief introduction to Volatility modelling.

2.1 Objectives of Time Series Analysis

The main objectives of undertaking a time series analysis are:

1. **Description**: The salient features of a series are described by using summary statistics and/or graphical illustrations. A time plot of the series can be particularly valuable in order to identify the existence of any regular patterns (such as trends, seasonality or cycles). A time plot is a graph where we plot the data series y_t against time, t such that we connect adjacent points. Note that this does not imply that the series is necessarily continuous. Adjacent points are simply joined in order to emphasise the order of the series, (t, y_t). A time plot will allow one to pick up any observations that might not be consistent with patterns in the rest of a data set. Such points are termed outliers.

2 **Modelling**: Often we seek a statistical model that suitably describes the dynamics of a series or a group of series. Modelling of several variables may be undertaken in order to quantify the relationship between them. The process of building a suitable model is a stepwise procedure. It consists of initially choosing a model and thereafter scrutinising it by examining the residuals of the model and making a judgement on its adequacy. One could however also utilise various information criteria (such as AIC, Schwarz, LOGL and ICOMP) in order to choose between similar models.

   While building a model the analyst has to consider the possible existence of **delays** in any cause and effect process between different variables in a model. One also has to consider the fact that variables could influence one another as when changes give rise to subsequent **feedback effects**. For example, *as rainfall increases, crop yields increases thus increasing the abundance of food products. These increases may in turn increase the consumption of food (due to a fall in the price of food). This change may effect a decrease in the abundance of food stuff (due to increased consumption levels).*

3 **Forecasting**: Often one may wish to forecast the future value of a series or a group of series. Forecasting is particularly important in the investment community where analysts and fund managers are all searching for the best strategies by which to make the most money for themselves and their clients. Forecasting models are based on the premise that the statistical model that best represents

2.1 Objectives of Time Series Analysis

The main objectives of undertaking a time series analysis are:

1. **Description**: The salient features of a series are described by using summary statistics and/or graphical illustrations. A time plot of the series can be particularly valuable in order to identify the existence of any regular patterns (such as trends, seasonality or cycles). A time plot is a graph where we plot the data series y_t against time, t such that we connect adjacent points. Note that this does not imply that the series is necessarily continuous. Adjacent points are simply joined in order to emphasise the order of the series, (t, y_t). A time plot will allow one to pick up any observations that might not be consistent with patterns in the rest of a data set. Such points are termed outliers.

2 **Modelling**: Often we seek a statistical model that suitably describes the dynamics of a series or a group of series. Modelling of several variables may be undertaken in order to quantify the relationship between them. The process of building a suitable model is a stepwise procedure. It consists of initially choosing a model and thereafter scrutinising it by examining the residuals of the model and making a judgement on its adequacy. One could however also utilise various information criteria (such as AIC, Schwarz, LOGL and ICOMP) in order to choose between similar models.

   While building a model the analyst has to consider the possible existence of **delays** in any cause and effect process between different variables in a model. One also has to consider the fact that variables could influence one another as when changes give rise to subsequent **feedback effects**. For example, *as rainfall increases, crop yields increases thus increasing the abundance of food products. These increases may in turn increase the consumption of food (due to a fall in the price of food). This change may effect a decrease in the abundance of food stuff (due to increased consumption levels).*

3 **Forecasting**: Often one may wish to forecast the future value of a series or a group of series. Forecasting is particularly important in the investment community where analysts and fund managers are all searching for the best strategies by which to make the most money for themselves and their clients. Forecasting models are based on the premise that the statistical model that best represents

2.1 Objectives of Time Series Analysis

The main objectives of undertaking a time series analysis are:

1. **Description**: The salient features of a series are described by using summary statistics and/or graphical illustrations. A time plot of the series can be particularly valuable in order to identify the existence of any regular patterns (such as trends, seasonality or cycles). A time plot is a graph where we plot the data series y_t against time, t such that we connect adjacent points. Note that this does not imply that the series is necessarily continuous. Adjacent points are simply joined in order to emphasise the order of the series, (t, y_t). A time plot will allow one to pick up any observations that might not be consistent with patterns in the rest of a data set. Such points are termed outliers.

2 **Modelling**: Often we seek a statistical model that suitably describes the dynamics of a series or a group of series. Modelling of several variables may be undertaken in order to quantify the relationship between them. The process of building a suitable model is a stepwise procedure. It consists of initially choosing a model and thereafter scrutinising it by examining the residuals of the model and making a judgement on its adequacy. One could however also utilise various information criteria (such as AIC, Schwarz, LOGL and ICOMP) in order to choose between similar models.

   While building a model the analyst has to consider the possible existence of **delays** in any cause and effect process between different variables in a model. One also has to consider the fact that variables could influence one another as when changes give rise to subsequent **feedback effects**. For example, *as rainfall increases, crop yields increases thus increasing the abundance of food products. These increases may in turn increase the consumption of food (due to a fall in the price of food). This change may effect a decrease in the abundance of food stuff (due to increased consumption levels).*

3 **Forecasting**: Often one may wish to forecast the future value of a series or a group of series. Forecasting is particularly important in the investment community where analysts and fund managers are all searching for the best strategies by which to make the most money for themselves and their clients. Forecasting models are based on the premise that the statistical model that best represents

a data set will continue to be valid in the short term. Note however that numerous models could describe the dynamics of a series of data and that before one selects a final model with which to forecast, one should first ensure that the assumption of short term persistence is justified.

4 **Control**: Time series models may be developed in order to study the complex dynamics between different variables in order to control the relationship. For example, *Park officials from the Kruger national park could be interested in monitoring and controlling the number of elephants in the park by means of culling. Time series analysis could then be used in order to make forecasts about the future population sizes of elephants in the park in order to set up guidelines about how to implement the culling of the elephants.*

This thesis is primarily concerned with objectives number one to three.

2.2 Components of a Time Series

Traditional time series analysis decomposes the variation in a time series into the four components. They are as follows:

1. **Trend**: A series that exhibits a long-run growth or decline (at least over successive time periods) is said to be a trending series. Not all series have a trend component, however. The trend of a series may follow either a linear-, quadratic-, exponential- or even a polynomial- function. Other shapes are possible, but simple descriptive terms may be helpful. As an example we can use the All Share Index (ALSI). See Figure 1 below. The All Share Index is a series of the combined price movements of all of the shares listed on the Johannesburg Stock Exchange. *One could say that the ALSI has experienced dramatic growth during the 1980's to the year 2000 and thus the ALSI index is an upward trending series.*

2. **Cycles**: Cycles can be defined as recurring up and down movements around any trend line of a series. Cycles may vary in length and could be modelled as being part of the trend or being part of the residual process of a series. An example of a cycle might be the *business cycle of an economy which may move over a period of say five years depending on the country under examination.*

3 **Seasonal variation**: Seasonal variation is present in a data set when similar patterns in a data set are observed at similar times during the year or during the week (for example). An example of seasonal variation would be *the increase in spending that occurs during the Christmas period each year.* i.e. *Consumer spending generally starts to increase during November each year. It generally peaks during December and decreases after the holiday period.* Another example of a series that might display seasonal variation *is the level of jersey sales made by shops during the year. Jersey sales will generally increase at the start of winter and decline at the end of the winter months.*

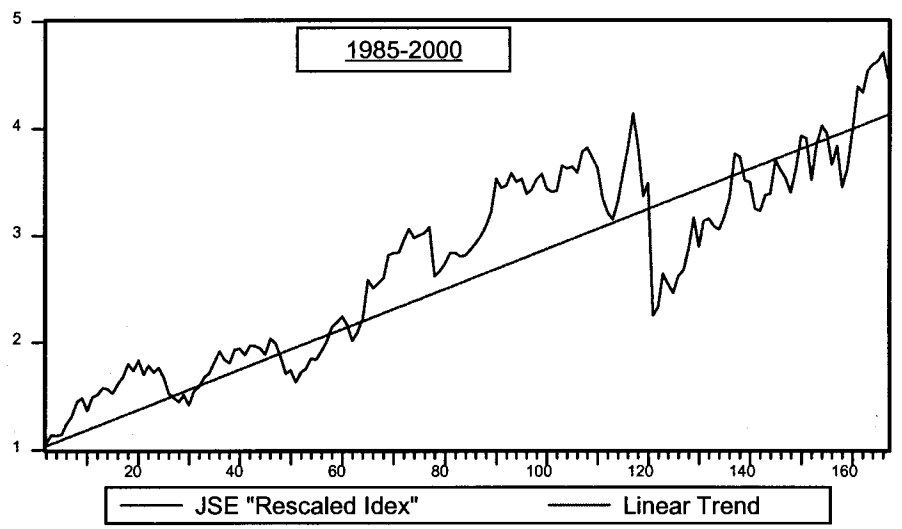


Figure 1: ALSI 1985-2000 (monthly data)

4 **Noise or Random fluctuation**: The noise of a series represents the variation that remains after one assigns the trend-, seasonal- and cyclical- component of a time series. The noise terms will generally be assumed to be uncorrelated however they may be correlated over a short period of time. When the noise is uncorrelated one is unable to forecast the future behaviour of this component from its past history. However, if there exists short term correlation amongst the noise terms one may be able to forecast the short term behaviour of the random component of the series, given knowledge of its current behaviour.

Classical methods of separating components of a time series often assume (at most) one trend, seasonal or cyclical components, over the entire time period of a series. Chatfield (2001) notes that "*classical methods work quite well when the variation is*

dominated by a regular linear trend and/or regular seasonality. However, they do not work very well when the trend and/or seasonal effects are changing through time or when successive values of the irregular fluctuations are correlated."

The next section of this chapter discusses different techniques that may be used in order to model the different components of a time series as a preliminary examination of the data.

2.3 Initial modelling of the components

2.3.1 Transformation of the data

Often patterns in a time series can be made simpler to describe if one transforms the series by either using **logs** or by **taking square roots**. Transformations may be used in order to stabilise the variance of the series (i.e. to make the variance homoskedastic). Transformations may be advised if a data set is skewed or asymmetric as time series literature may often depend on the assumption that the random variation component of a time series is normally distributed. One would also recommend the transformation of the data set if the seasonal effects of a model are multiplicative (see section below on seasonal variation.). A log transformation would make the seasonal effects additive (linear).

A general class of transformations is the Box-Cox family of transformations. The original series $x_1, x_2..., x_n$ is transformed to $y_1(\lambda), y_2(\lambda)..., y_n(\lambda)$ as follows:

$$y_t(\lambda) = \begin{cases} \frac{(x_t)^\lambda - 1}{\lambda} & if \quad \lambda \neq 0 \\ \log(x_t) & \iota f \quad \lambda = 0 \end{cases} \tag{2.1}$$

where λ is a parameter that has to be estimated. Trial and error can be used in order to estimate λ. The chosen value of λ should ensure that the transformed series has a stabilised variance and that it does not depend on time. Nelson and Granger (1979), and Chatfield (2001) both state that transformations are often used. However they suggest that one should only transform a series if there is a clear interpretation of the transformed series or if it stabilises the variance of the underlying series. For example, *one might transform a price series by using the log function (i.e. let the original series be* $x_1, x_2, ..., x_n$ and the transformed series be defined by

$y_t = \log(y_t) - \log(y_{t-1})$). *Then each of the log differences can be directly interpreted as being the return of a share during a particular period of time.*

2.3.2 Trend

We introduced the notion of trend by a simple example. Chatfield (2001) states that an exact definition of long-term trend is not necessarily obvious since it is dependent on what one interprets as a long term period. *For example, a long term trend might be apparent if one were to investigate a data set containing 5 years of history, but in a period of say thirty years, the apparent 5 year trend might simply be part of the a "long term oscillation "* of the time series. (Chatfield (1975))

When undertaking a time series analysis it is important to firstly identify whether or not a trend is present in a data set and thereafter remove its influence from the series by modelling the trend component by means of some mathematical model. The following section discusses several methods of modelling both a global (long term) trend and a local (short term) trend. Global and local trend models lead to the construction of distinct models as well as distinct forecasts. The nature of the series will determine which type of trend model is most appropriate.

Curve-Fitting

Data sets that have a long or global linear term trend could be modelled by using the following model:

$$\mu_t = \alpha + \beta t \tag{2.2}$$

where μ_t represents the local mean level at time t and α and β are fixed constants. This model is often termed the *deterministic* or *linear trend model.* It is very simple but is rarely used in practice since not many real-world data sets can be modelled by equation 2.2 alone. Researchers favour modelling the trend by means of a local trend component. (Newbold (1988), Chatfield (1996a))

Local linear trend models are of the form:

$$\mu_t = \alpha_t + \beta_t t \tag{2.3}$$

where α_t and β_t are allowed to change over time. They are more robust than the deterministic model since they are able to capture the short term variation present in data and are often also able to improve short term forecasting.

An alternative way of modelling the local trend is to utilise a state space model. The local trend could then be captured by modelling it as:

$$\mu_t = \alpha + \mu_{t-1} + \varepsilon_t \tag{2.4}$$

Notice that in this case the local trend is being modelled as a function of the previous mean level, μ_{t-1}. The local trend at time t thus depends on the local trend at time $t-1$. The noise term, ε_t, is generally modelled as a normal random variable with **zero mean** and **fixed variance,** σ^2. This ε_t process is known as a **white noise** process since it represents random movement.

We may observe a non-linear global trend. Models for non-linear trend include logistic models and Gompertz models. The logistic and the Gompertz model are both known as S-curve models. They are most often used to model growth curves of living organisms (population studies) and the growth of new industries in an economy. The logistic model is represented by:

$$\mu_t = \frac{a}{1 + be^{-ct}} \tag{2.5}$$

where b and c are fixed constants obtained by fitting the curve to the data series and a is defined as the hypothesised upper limit of the growth of the μ_t. Figure 2 below displays a typical logistic curve. It can be seen that with t near the limits ∞ and $-\infty$ the logistic curve tends towards the upper asymptote, a and the lower asymptote 0 respectively. The inflection point occurs at $t = \frac{\ln(b)}{c}$, when $\mu_t = \frac{a}{2}$.

The Gompertz model is defined as:

$$\mu_t = ae^{-be^{ct}} \tag{2.6}$$

where b and c are constants obtained by fitting the curve to the data series and a is defined as the hypothesised upper limit of the growth of the μ_t. The Gompertz curve has similar properties to a logistic curve. As t tends towards the limits ∞ and $-\infty$ the Gompertz curve tends towards the upper asymptote, a and the lower asymptote respectively. The inflection point occurs at $t = \frac{\ln(b)}{c}$, when $\mu_t = \frac{a}{e}$.

After a trend curve has been fitted, the effect of the trend can be removed from the series by subtraction.

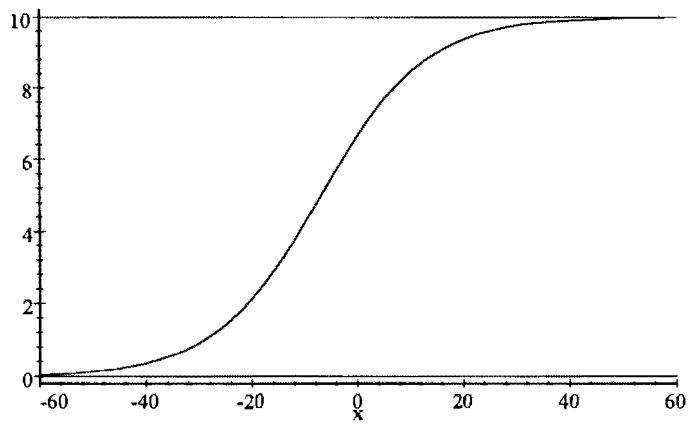


Figure 2: A logistic curve with parameters a=10, b=0.5, c=0.1

Filtering local trend

A second useful approach models the local trend of a series as the output of a *linear filter*. A linear filter transforms a time series $y_1, y_2 \ldots, y_n$ to $\widetilde{y}_1, \widetilde{y}_2 \ldots, \widetilde{y}_n$ by using the linear operation:

$$\widetilde{y}_t = \sum_{r=-q}^{s} w_r y_{t+r} \tag{2.7}$$

where $\{w_r\}$ is a set of weights that sum to one. Usually we set $s = q$ where q is some poistive integer.

This sum of weights property is useful because the filter preserves a constant mean. If $E(y_{t+r}) = \mu$ for all r and we set $w_i = w_{-i}$ for all $i = 1, 2, ..., q$ then:

$$E(\widetilde{y}_t) = E\left(\sum_{r=-s}^{s} w_r y_{t+r}\right) = \sum_{r=-s}^{s} w_r E(y_{t+r}) \tag{2.8.a}$$

$$= \left(\sum_{r=-s}^{s} w_r\right)\mu = \mu \quad \text{for all } t. \tag{2.8.b}$$

The filter also preserves linear trend. If $E(y_t) = \alpha + \beta t$ then:

$$E(\widetilde{y}_t) = E\left(\sum_{r=-s}^{s} w_r(\alpha + \beta(t+r))\right) \tag{2.9.a}$$

$$= \left(\sum_{r=-s}^{s} w_r\right)\alpha + \beta t + \left(\sum_{r=-s}^{s} w_r r\right)\beta = 1 * \alpha + \beta t + 0 * \beta = \alpha + \beta t \quad \text{for all } t.$$

A simple example of a linear filter is a **moving average (ma)**. For example, a MA(3) can be represented by the filter $\frac{1}{3}[1,1,1]$ similarly a MA(5) can be represented by $\frac{1}{5}[1,1,1,1,1]$ and a MA(9) by $\frac{1}{9}[1,1,1,1,1,1,1,1,1]$. Figure 3 below displays the monthly ALSI index over the periods 1980 to 2000 as well as two linear filters namely a MA(15) and a MA(41) respectively.

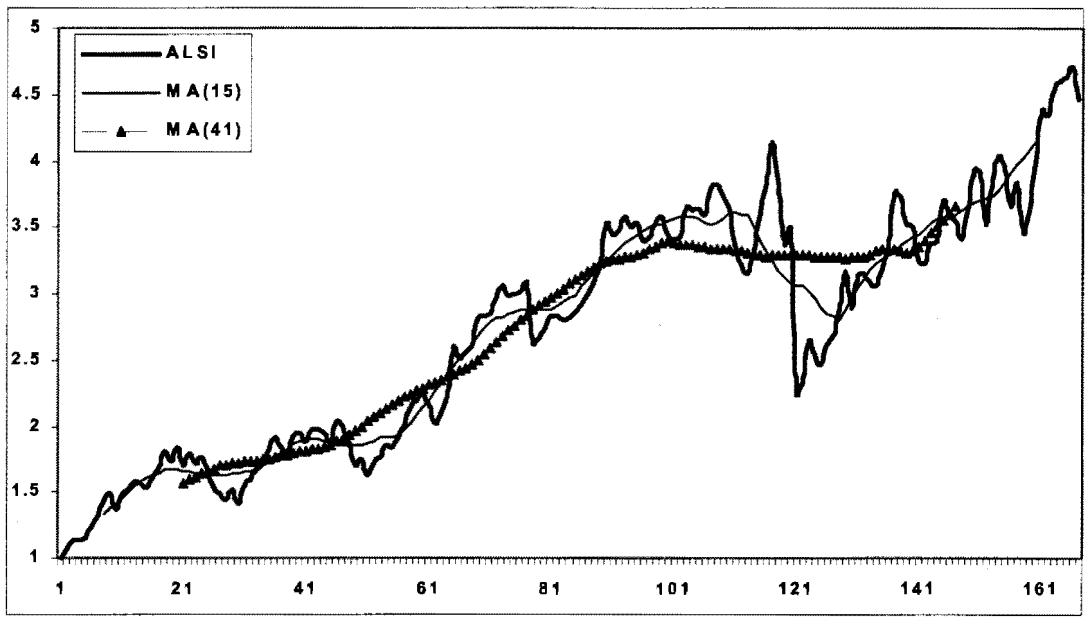


Figure 3: ALSI (Monthly Rescaled Index) 1980-2000 -and two linear filters

The MA(41) provides a smoother representation of the underlying trend of the ALSI series than the MA(15) since it groups 41 adjacent observations in order to generate one point in the smoothed series. The MA(41) is thus an example of a wide filter (width = 41) that removes most of the random variation of the series by averaging terms that are assumed to have zero mean and common variance. The MA(15) series is a narrower filter and thus tracks the ALSI much closer. It retains more detail about the ALSI series than the MA(41) series. Both of these moving averages are examples of *odd* moving averages. Odd moving averages are plotted against the mid-point of a time interval. When using a MA(15), both s and q is equal to 7. Fifteen observations of the ALSI series are used in order to plot one point on the smoothed series. This point will be situated at observation 8 (centred) for each rolling window of 15 observations. Similarly, the smoothed estimate of the local trend of a MA(41) will be situated at observation 21 for each rolling window of 41.

Special moving averages can be used to fit (estimate) the polynomial trend in a time series. Numerous examples are provided in table 1 and table 2. Note that all

of the filters are symmetric and that the elements of the filter sum to one. The first group of filters can be used to fit (estimate) at most a quadratic or cubic polynomial trend from a data series where as the second set of filters can be used in order to fit (estimate) at most a quartic or a quintic polynomial trend from a data series.

	Table 1: Quadratic and Cubic Moving Averages
[5]	$\frac{1}{35}[-3, 12, \underline{17}]$
[7]	$\frac{1}{21}[-2, 3, 6, \underline{7}]$
[9]	$\frac{1}{231}[-21, 14, 39, 54, \underline{59}]$
[11]	$\frac{1}{429}[-36, 9, 44, 69, 84, \underline{89}]$
[13]	$\frac{1}{143}[-11, 0, 9, 16, 21, 24, \underline{25}]$
[15]	$\frac{1}{1105}[-78, -13, 42, 87, 122, 147, 162, \underline{167}]$
[17]	$\frac{1}{323}[-21, -6, 7, 18, 27, 34, 39, 42, \underline{43}]$
[19]	$\frac{1}{2261}[-136, -51, 24, 89, 144, 189, 224, 249, 264, \underline{269}]$
[21]	$\frac{1}{3053}[-171, -79, 9, 84, 149, 204, 249, 284, 309, 324, \underline{329}]$

	Table 2: Quartic and Quintic Moving Averages
[7]	$\frac{1}{231}[5, -30, 75, \underline{131}]$
[9]	$\frac{1}{429}[15, -55, 30, 135, \underline{179}]$
[11]	$\frac{1}{429}[18, -45, -10, 60, 120, \underline{143}]$
[13]	$\frac{1}{2431}[110, -198, -135, 110, 390, 600, \underline{677}]$
[15]	$\frac{1}{46189}[2145, -2860, -2937, -165, 3755, 7500, 10125, \underline{11063}]$
[17]	$\frac{1}{4199}[195, -195, -260, -117, 135, 415, 660, 825, \underline{883}]$
[19]	$\frac{1}{7429}[340, -255, -420, -290, 18, 405, 790, 1110, 1320, \underline{1393}]$

Deriving a Polynomial trend Moving Average

Let $\{a_r\}$ be a set of symmetric filter weights that sum to one. Symmetric moving averages can then be derived by setting $s = q$ and thereafter solving for $\{a_r\}$ once a decision has been made on the type of function to fit for the trend. For example, suppose that one chose to fit a cubic as the trend, then:

$$y_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \varepsilon_t \tag{2.10}$$

where y_t is the time series under investigation, a_0, a_1, a_2, a_3 are the coefficients of the cubic trend that have to be estimated from the data and ε_t is a random error term representing the difference between the **true data** series and the smooth cubic trend.

The cubic polynomial could be fitted as a **global** trend using multiple regression to estimate a_0, a_1, a_2 and a_3 with explanatory variables $X_1 = t$, $X_2 = t^2$ and $X_3 = t^3$.

We now discuss how the cubic polynomial can be fitted as a local trend using an odd number $(2m + 1)$ of time points: $t - m$ to $t + m$. Our aim is to *minimise* (for all t):

$$\sum_{s=t-m}^{t+m} \left[y_s - \sum_{j=0}^{3} a_j s^j \right]^2 \tag{2.11}$$

by selecting values for a_0, a_1, a_2 and a_3 that depend upon the time series data at $(2m + 1)$ points at equal units.

Changing the origin from $s = t$ to $t = 0$ allows one to rewrite equation 2.9 as:

$$\sum_{t=-m}^{m} \left[y_t - \sum_{j=0}^{3} a_j t^j \right]^2 \tag{2.12}$$

Differentiating equation 2.12 with respect to a_i for $i = 0, 1, 2, 3$ and setting each derivative equal to zero leads to:

$$\sum_{t=-m}^{m} y_t t^i = \sum_{j=0}^{3} a_j \left[\sum_{t=-m}^{m} t^{j+i} \right] \qquad (0 \leq i \leq 3) \tag{2.13}$$

Equation 2.13 can be expanded into a set of four equations. If we set $m = 3$ we obtain $-3 \leq t \leq 3$ and:

$$\begin{array}{lcccccl} \sum y_t & = & 7a_0 & & +28a_2 & & (A) \\ \sum t y_t & = & & 28a_1 & & +196a_3 & (B) \\ \sum t^2 y_t & = & 28a_0 & & +196a_2 & & (C) \\ \sum t^3 y_t & = & & 196a_1 & & +1588a_3 & (D) \end{array} \tag{2.14}$$

In order to estimate the trend value, x_0, one has to solve for a_0 (i.e. $t = 0$).

$$7\,(A) - (C) \implies 21a_0 = 7\sum_{-3}^{3} y_t - \sum_{-3}^{3} t^2 y_t \tag{2.15.a}$$

$$\implies a_0 = \frac{1}{21}\left[-2y_{-3} + 3y_{-2} + 6y_{-1} + 7y_0 + 6y_1 + 3y_2 - 2y_3\right]$$

$$\implies a_0 = \frac{1}{21}\left[-2, 3, 6, \underline{7}\right] \tag{2.15.c}$$

The linear filter is $\frac{1}{21}[-2, 3, 6, \underline{7}]$. This filter fits a cubic polynomial to $2m+1 = 7$ points, but calculates the fitted value only for the mid point (fourth) amongst the seven. To obtain fitted values for the other six points would require solutions for a_1, a_2 and a_3 as well as a_0. However these filters are applied to 7 consecutive values in equally spaced time series data and the fitted values for time points 4 to $n-3$ are separately estimated by a_0 alone. The cubic polynomial is fitted by the successive moving values of the filter a_0.

Properties of a Moving Average

Assuming that a series does not have a seasonal component and that the series can be represented by:

$$y_t = \mu_t + \varepsilon_t \tag{2.16}$$

where μ_t is the mean of the series, ε_t is a identically distributed random variable with mean *zero* and variance σ^2 and $\hat{\mu}_t = \sum_{r=-s}^{s} a_r y_{t+r}$ is the smoothed trend, then the following properties hold:

1. $var\left(\hat{\mu}_t\right) = \left(\sum_{r=-s}^{s} a_r^2\right)\sigma^2$
2. $cov\left(\hat{\mu}_t, \hat{\mu}_{t+k}\right) = \left(\sum_{r=-s}^{r+s=k} a_r a_{r+k}\right)\sigma^2$
3. $corr\left(\hat{\mu}_t, \hat{\mu}_{t+k}\right) = \frac{\sum_{r=-s}^{r+s=k} a_r a_{r+k}}{\sum_{r=-s}^{s} a_r^2}$

Moving averages are very useful because they "*iron out*" noise in a data series while preserving any trend. This property can be seen as follows:

$$\begin{aligned}
var\left(\sum_{r=1}^{k} w_r y_t\right) &= \sum_{r=1}^{k} w_r^2 var\left(y_t\right) + \sum_{r\neq j}^{k}\sum_{r\neq j}^{k} w_r w_j cov\left(y_r, y_j\right) && \text{(2.17.a)}\\
&= \sigma^2\left[\sum_{r=1}^{k} w_r^2 + \sum_{r\neq j}^{k}\sum_{r\neq j}^{k} w_r w_j \rho_{rj}\right] && \text{(2.17.b)}\\
&< \sigma^2\left(\sum_{r=1}^{k} w_r\right)^2 && \text{(2.17.c)}\\
&< \sigma^2 \qquad \text{for } \sum w_t = 1 && \text{(2.17.d)}
\end{aligned}$$

Equation 2.17.*b* follows since $\left(\sum_{r=1}^{k} w_r\right)^2 = \sum_{r=1}^{k} w_r^2 + \sum_{r\neq j}^{k}\sum_{r\neq j}^{k} w_r w_j$ and since $|\rho_{rj}| \leq 1$ $\forall\,(r \neq j)$. Under suitable conditions moving averages can also be used to "*iron out*" seasonal effects while preserving the trend that might be present in a data series.

Exponential Smoothing

Exponential smoothing can also be used in order to fit (estimate) the trend from a data set. The filter that is applied to the data series is of the form:

$$a_j = \alpha^j (1-\alpha) \tag{2.18}$$

where $0 < \alpha < 1$ for $j = 0, 1, 2, ...$ The filter is an **assymmetric** infinite filter since the filter assigns exponentially decreasing weights to observations in the data series further distant in time. In contrast, other filters are **symmetric** and assign the smoothed value to the **mid-point** of a finite segment of time. When we construct an **EWMA** (exponentially weighted moving average), we say we use **single exponential smoothing (SES)**.

The smoothed series can be found by using the following equations:

$$s_t = (1-\alpha)\, y_t + \alpha s_{t-1} \tag{2.19}$$

where s_t is the smoothed estimate of observation t, y_t is the observed data value at time t and α is a constant such that $0 < \alpha < 1$. Assuming that one has an **infinite** number of observations in a data series then we can expand equation 2.19 by iteration as follows:

$$\begin{aligned}
s_t &= (1-\alpha)\, y_t + \alpha\left[(1-\alpha)\, y_{t-1} + \alpha s_{t-2}\right] && \text{(2.20.a)}\\
&= (1-\alpha)\, y_t + \alpha\,(1-\alpha)\, y_{t-1} + \alpha^2\left[(1-\alpha)\, y_{t-2} + \alpha s_{t-3}\right] && \text{(2.20.b)}\\
&= (1-\alpha)\, y_t + \alpha\,(1-\alpha)\, y_{t-1} + \alpha^2\,(1-\alpha)\, y_{t-2} + \alpha^3(1-\alpha) y_{t-3} + ... && \text{(2.20.c)}\\
&= \sum_{i=0}^{\infty} \alpha^i\,(1-\alpha)\, y_{t-i} && \text{(2.20.d)}\\
&= \sum_{i=0}^{\infty} w_i y_{t-i} \qquad \text{where } \sum_{i=0}^{\infty} w_i = 1 \text{ and } w_i = \alpha^i\,(1-\alpha) && \text{(2.20.e)}
\end{aligned}$$

Equation 2.20.*e* indicates that exponential smoothing utilises an infinite window of observations in order to smooth the data series.

In general one does not have a infinite series of observations. In order to handle the finite case we need an assumption about s_0, the smoothed estimate at time 0. Many authors suggest that s_0 should be set equal to the value of the first observation of the data series ($s_0 = y_0$) which allows one to use equation 2.19 in order to smooth the subsequent data series.

The degree to which the filter smoothes out a data series is dependent on the choice of α. A value of α close to 0 implies that the smoothed series will look very similar to the actual data series since very little weight is being assigned to previous observations found in the data series. A value of α close to 1 will cause the smoothed series to be highly dependent on many of the preceding observations, and perhaps on the initial assumption of s_0. One should take care in choosing an appropriate value for α. One criterion might be to choose the value of α such that it minimises the sums of squared residuals between the observed data series and the smoothed series.

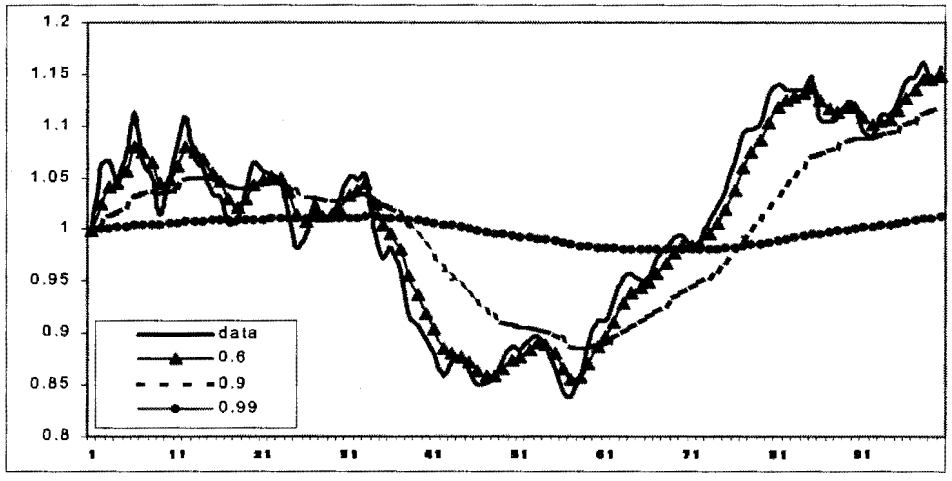


Figure 4: ALSI (Jan 1990/Dec 1991) and three exponential filters

Figure 4 above displays the Weekly ALSI index rescaled over the period 6 January 1990 to 14 December 1991 as well as three different exponential filters. All three filters use the assumption that $s_0 = 1$ (i.e. the first observation of the rescaled data series). It can be seen that the filter with $\alpha = 0.6$ tracks the actual index more closely than if one used $\alpha = 0.9$ or $\alpha = 0.99$ indicating that $\alpha = 0.6$ assigns more weight to recent observations than to earlier observations. The $\alpha = 0.9$ filter generates a smoother estimate of the long run trend than the $\alpha = 0.6$ filter. Figure 4 below displays the detrended series (data series - smoothed series) for two of the filters namely $\alpha = 0.6$ and $\alpha = 0.99$. From figure 5 it can be seen that the $\alpha = 0.99$ filter

does not adequately remove the trend in the data series since the detrended series is approximately proportional to the data series indicating that they both contain the same patterns.

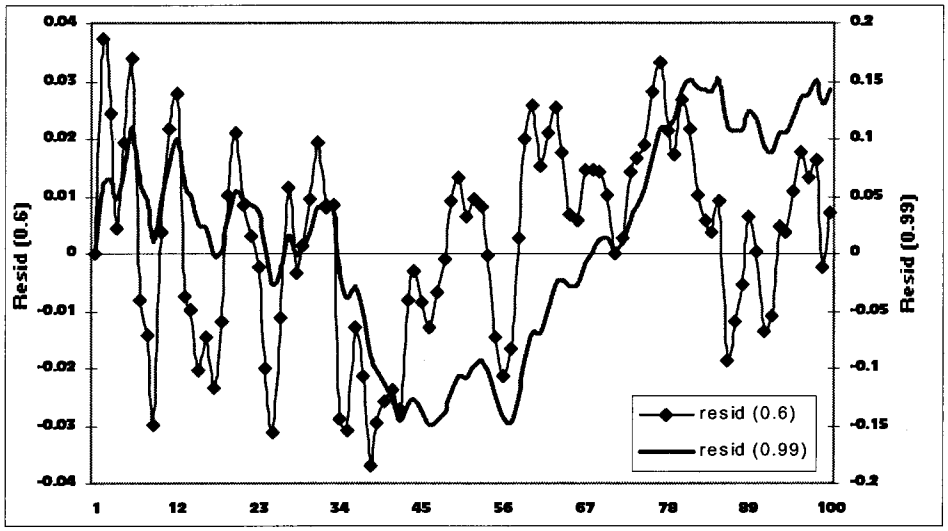


Figure 5: The detrended series $(y_t - \hat{\mu}_t)$

In this section we consider single exponential smoothing (**SES**) and double exponential smoothing (**DES**) with smoothing parameter α applied to y_t to obtain s_t, and then applied to s_t to obtain s_t^*. The following preliminary equations are used throughout the discussion.

Preliminary Equations: for $|\alpha| < 1$

$$\sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha} \tag{2.21.a}$$

$$\sum_{i=1}^{\infty} \alpha^i = \frac{\alpha}{1-\alpha} \tag{2.21.b}$$

$$\sum_{i=0}^{\infty} \alpha^{2i} = \frac{1}{1-\alpha^2} \tag{2.21.c}$$

$$\sum_{i=0}^{\infty} i\alpha^i = \sum_{i=1}^{\infty} i\alpha^i = \alpha \sum_{i=1}^{\infty} i\alpha^{i-1} = \alpha \sum_{i=1}^{i=\infty} \frac{\partial}{\partial \alpha}\left(\alpha^i\right) \tag{2.22.a}$$

$$= \alpha \frac{\partial}{\partial \alpha}\left(\sum_{i=1}^{\infty} \alpha^i\right) \tag{2.22.b}$$

$$= \alpha \frac{\partial}{\partial \alpha}\left(\frac{\alpha}{1-\alpha}\right) \tag{2.22.c}$$

$$= \frac{\alpha}{(1-\alpha)^2} \tag{2.22.d}$$

Set $\omega_0 = (1-\alpha)$ and $\omega_i = (1-\alpha)\alpha^i$ then the following equations hold:

$$\sum_{i=0}^{\infty} \omega_i = (1-\alpha)\sum_{i=0}^{\infty} \alpha^i = 1; \tag{2.23.a}$$

$$\sum_{i=0}^{\infty} \omega_i^2 = (1-\alpha)^2 \sum_{i=0}^{\infty} \alpha^{2i} = \frac{(1-\alpha)^2}{1-\alpha^2} = \frac{(1-\alpha)}{(1+\alpha)} \tag{2.23.b}$$

Let $y_t = \mu_t + \varepsilon_t = a + bt + \varepsilon_t$ for all t, with $\varepsilon_t \sim IID(0, \sigma^2)$. We assume we have the entire (infinite) history of the series. Now consider single exponential smoothing (**SES**) with smoothing parameter α applied to y_t to obtain s_t:

$$s_t = \sum_{i=0}^{\infty} \omega_i y_{t-i} = \sum_{i=0}^{\infty} \omega_i (a+bt) - b\sum_{i=0}^{\infty} \omega_i i + \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} \tag{2.24.a}$$

$$= (a+bt) - b\frac{\alpha}{(1-\alpha)} + e_t \tag{2.24.b}$$

$$= \mu_t - b\frac{\alpha}{(1-\alpha)} + e_t \tag{2.24.c}$$

Equation 2.24.*b* holds since $\sum_{i=0}^{i=\infty} \omega_i = 1$ and $\sum_{i=0}^{i=\infty} \omega_i i = \frac{\alpha}{(1-\alpha)}$ from 2.22.*d*. Note that we have defined $e_t = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i}$. Define s_t^* as the smoothed series after one undertakes double exponential smoothing to y_t.

$$s_t^* = \sum_{i=0}^{\infty} \omega_i s_{t-i} = \sum_{i=0}^{\infty} \omega_i (a+bt) - b\frac{\alpha}{(1-\alpha)} - b\sum_{i=0}^{\infty} \omega_i i + \sum_{i=0}^{\infty} \omega_i e_{t-i}$$

$$= (a+bt) - 2b\frac{\alpha}{(1-\alpha)} + e_t^* \tag{2.25.b}$$

$$= \mu_t - 2b\frac{\alpha}{(1-\alpha)} + e_t^* \tag{2.25.c}$$

The s_t result (Equation 2.24.*a* through 2.24.*c*) implies that when there is **no trend** in y_t, $b = 0$, we may use s_t as an **unbiased** estimator of the true constant mean $\mu_t = a$, since the coefficients sum to one, and:

$$E(s_t) = E(\sum \omega_i y_{t-i}) = \sum \omega_i E(y_{t-i}) = \sum \omega_i a = a \tag{2.26}$$

However, when there is **global linear trend** in y_t, the smoothed series s_t has the same slope as $\mu_t = E(Y_t) = a + bt$, namely b, but the smoothed series is biased by the term $-b\frac{\alpha}{(1-\alpha)}$. The **bias** is negative for slope $b > 0$, and is positive for slope $b < 0$. The fitted line is on average about $\left|b\frac{\alpha}{(1-\alpha)}\right|$ units below (or above) the μ_t on the true mean line $(a + bt)$ of the y_t. Another way of describing the bias is to say that the fitted values appear to be $\frac{\alpha}{(1-\alpha)}$ time units later than the true mean values. The fitted straight line is always to the right of the true mean line for $b > 0$.

The smoothed values s_t have variance $var(e_t) = \frac{(1-\alpha)}{(1+\alpha)}\sigma^2$, but these terms e_t are **correlated,** rather than independent, like ε_t, due to the **Slutsky-Yule** effect (Slutsky(1937)). Linear combinations of overlapping sets of independent variables will be correlated, precisely because the overlap induces non-zero covariances:

$$\begin{aligned} cov(s_t, s_{t+k}) &= cov(\sum \omega_i \varepsilon_{t-i}, \sum \omega_i \varepsilon_{t+k-i}) = \sum \omega_i \omega_{i+k} \sigma^2 = \alpha^k \frac{(1-\alpha)}{(1+\alpha)} \sigma^2 \\ corr(s_t, s_{t+k}) &= \alpha^k \end{aligned} \tag{2.27.b}$$

Slutsky (1937) observed that the use of moving averages in order to smooth data series, induces non-zero correlations between the smoothed series and the original series even if the original observations were independent. This correlation implies that apparent periodic behaviour in detrended series may be due to the use of moving averages.

Note that if $\alpha \to 1$ then the variance of s_t is small but the bias is large. If $\alpha \to 0$, there is virtually no smoothing, the variance is virtually unchanged, and the bias $-b\frac{\alpha}{(1-\alpha)} \to 0$.

Use of $\alpha < 0$ would result in sign changes in s_t, and also an increase in variance. Thus **SES** with $0 < \alpha < 1$ can reduce variance (and hence better reflect underlying pattern in the mean), but may introduce bias (if linear global trend is present). The graphs of y_t and s_t will be parallel, but at a vertical distance equal to the bias.

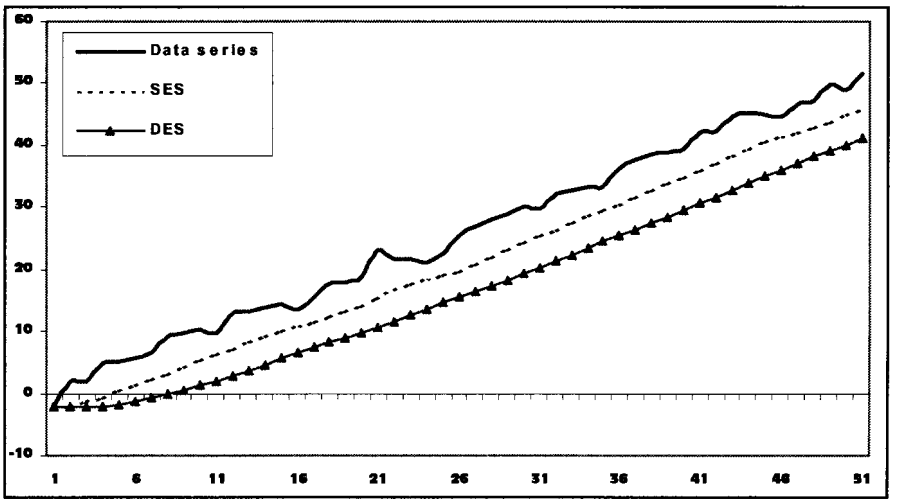


Figure 6: A time series with the SES and DES series

Figure 6 above displays the series $y_t = 0.8 + 1 * t + e_t$ (where e_t is a standard normal random variable and $\alpha = 0.8$), the single exponentially smoothed series and

the double exponentially smoothed series. It can be seen that the single exponentially smoothed series is to the right of y_t. Similarly the double exponentially smoothed series is to the right of the single exponentially smoothed series.

Can we remove the bias? The s_t^* (Equation 2.25.a through 2.25.c) result implies we can use **DES** to remove the bias in **SES** when there is a linear trend. Define $\widehat{\mu}_t$ as the unbiased estimate of the trend present in the data series, y_t.

$$\widehat{\mu}_t = 2s_t - s_t^* \Rightarrow E(\widehat{\mu}_t) = 2E(s_t) - E(s_t^*) = a + bt = \mu_t \tag{2.28}$$

These unbiasedness results extend to quadratic, cubic and higher order **polynomial global trend.** Quadratic trend may be recovered and bias removed by using **triple exponential smoothing,** and so on for higher order polynomials. The only difficulty that occurs when one tries to apply double and triple exponential smoothing (as well as higher exponential smoothing) is the fact that the calculation of the distribution characteristics of the filter becomes increasingly complicated. The calculation of the variance of s_t^* is equal to $\sigma^2 \frac{2(1-\alpha)}{(1+\alpha)^3}$ $(var(s_t^*))$ but the calculation of $var(\widehat{\mu}_t)$ is very challenging. An intuitive derivation of $var(e_t^*)$ is supplied below:

$$\begin{aligned}
e_t^* &= \sum_{i=0}^{\infty} \omega_i e_{t-i} \\
&= \sum_{i=0}^{\infty} \omega_i \sum_{j=0}^{\infty} \omega_j \varepsilon_{t-i-j} \\
&= (1-\alpha)^2 \sum_{i=0}^{\infty} \alpha^i \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-i-j} \\
&= (1-\alpha)^2 \left[\sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \alpha^{j+1} \varepsilon_{t-1-j} + \sum_{j=0}^{\infty} \alpha^{j+2} \varepsilon_{t-2-j} + \ldots \right]
\end{aligned}$$

Note that the square brackets can be rewritten as the sum of the elements in the following lines:

$$\begin{array}{cccccccccc}
\alpha^0\varepsilon_t & \alpha^1\varepsilon_{t-1} & \alpha^2\varepsilon_{t-2} & \ldots & \ldots & \alpha^{\infty}\varepsilon_{t-\infty} & & & & \\
 & \alpha^1\varepsilon_{t-1} & \alpha^2\varepsilon_{t-2} & \ldots & \ldots & \ldots & \alpha^{\infty+1}\varepsilon_{t-1-\infty} & & & \\
 & & \alpha^2\varepsilon_{t-2} & \ldots & \ldots & \ldots & \ldots & \alpha^{\infty+2}\varepsilon_{t-2-\infty} & & \\
 & & & \alpha^3\varepsilon_{t-3} & \ldots & \ldots & \ldots & \ldots & \ldots & \\
 & & & & \ldots & \ldots & \ldots & \ldots & \ldots & \\
 & & & & & \ldots & \ldots & \ldots & \ldots &
\end{array}$$

such that:

$$e_t^* = \sum_{m=0}^{\infty}\sum_{j=0}^{m}(1-\alpha)^2\alpha^m\varepsilon_{t-m} = \sum_{m=0}^{\infty}(1-\alpha)^2(m+1)\alpha^m\varepsilon_{t-m} \tag{2.29}$$

$$\begin{aligned}
var(e_t^*) &= \sigma^2(1-\alpha)^4\sum_{m=0}^{\infty}(m+1)^2\alpha^{2m} && (2.30.a)\\
&= \sigma^2(1-\alpha)^4\sum_{m=0}^{\infty}\left[(m+1)m+(m+1)\right]\alpha^{2m}\\
&= \sigma^2(1-\alpha)^4\left\{\sum_{m=1}^{\infty}(m+1)m\alpha^{2(m-1)}\alpha^2+\sum_{m=0}^{\infty}(m+1)\alpha^{2m}\right\}\\
&= \sigma^2(1-\alpha)^4\left\{\alpha^2\sum_{m=1}^{\infty}\left[\frac{\delta^2}{\delta\beta^2}\beta^{m+1}\right]_{\beta=\alpha^2}+\sum_{m=0}^{\infty}\left[\frac{\delta}{\delta\beta}\beta^{m+1}\right]_{\beta=\alpha^2}\right\}\\
&= \sigma^2(1-\alpha)^4\left\{\left[\beta\frac{\delta^2}{\delta\beta^2}\sum_{m=1}^{\infty}\beta^{m+1}\right]_{\beta=\alpha^2}+\left[\frac{\delta}{\delta\beta}\sum_{m=0}^{\infty}\beta^{m+1}\right]_{\beta=\alpha^2}\right\}\\
&= \sigma^2(1-\alpha)^4\left\{\left[\beta\frac{\delta^2}{\delta\beta^2}\frac{\beta^2}{(1-\beta)}+\frac{\delta}{\delta\beta}\frac{\beta}{(1-\beta)}\right]_{\beta=\alpha^2}\right\}\\
&= \sigma^2(1-\alpha)^4\left\{\left[\frac{2\beta}{(1-\beta)^3}+\frac{1}{(1-\beta)^2}\right]_{\beta=\alpha^2}\right\}\\
&= \sigma^2(1-\alpha)^4\frac{(1+\alpha^2)}{(1-\alpha^2)^3}\\
&\quad \sigma^2\frac{(1-\alpha)(1+\alpha^2)}{(1+\alpha)^3} && (2.30.i)
\end{aligned}$$

Differencing

Instead of fitting the trend by means of smoothing (and then removing it by means of subtraction), we now attempt to remove trend directly by means of differencing. The difference operator ∇ is by:

$$\nabla X_t = X_t - X_{t-1} = (1-B)\,X_t \tag{2.31}$$

where B is the backshift operator defined by:

$$BX_t = X_{t-1} \tag{2.32}$$

Differencing once removes a constant trend, and makes a linear trend appear constant.

$$\nabla Y_t = (a-a) = 0 \quad \text{for any } a,t \tag{2.33.a}$$

$$\nabla Y_t = (a+bt) - [a+b(t-1)] = b \quad \text{for any } a,b,t \tag{2.33.b}$$

Powers of the operator B and ∇ are defined as follows:

$$B^k(X_t) = X_{t-k} \tag{2.34.a}$$

$$\nabla^k(X_t) = \left(\nabla^{k-1}(X_t)\right) \quad \text{for } k \geq 1 \text{ with } \nabla^0(X_t) = X_t \tag{2.34.b}$$

As an example we investigate $\nabla^2(X_t)$.

$$\nabla^2(X_t) = \nabla[\nabla(X_t)] = \nabla(X_t - X_{t-1}) \tag{2.35.a}$$

$$= \nabla(X_t) - \nabla(X_{t-1}) = X_t - 2X_{t-1} + X_{t-2} \tag{2.35.b}$$

The same result is obtained from $\nabla^2(X_t) = (1-B)^2 X_t = (1-2B+B^2)X_t$

2.3.3 Seasonal Variation

Seasonal variation is by definition a regular pattern repeated over each set of a number of equally spaced data. For example, *total rainfall varies over the four quarters of a year with some type of regularity, traffic volumes vary over the seven days of the week. Car sales may vary over the twelve months of the year.*

These patterns imply that the mean of all first quarter values will be distinct from the mean of at least one of the other quarters. The Sunday traffic volumes will differ from at least one of the other days, as will the mean of the Mondays. The mean car sales for the twelve months of the year will not all be equal. We say that there is an effect associated with each month, which makes some month means differ from the average of all the months together.

Seasonal effects can be removed from a data series by either choosing a suitable moving average or by using a suitable differencing operator. Seasonal effects can be **removed** from monthly data by using $\frac{1}{12}\left[\frac{1}{2},1,1,1,1,1,1,\underline{1}\right]$, which smooth the data by the **uncentred** and **centred** 12 month moving average filters, by constructing successive averages of overlapping but complete one year periods. If differencing is preferred, then one could use the operator ∇_{12} where $\nabla_{12}(X_t) = X_t - X_{t-12}$. The seasonal effect is removed because the new series $\nabla_{12}(X_t)$ contains only the annual

change for a specific month. For quarterly data, the seasonal effects can removed by using $\frac{1}{4}\left[\frac{1}{2}, 1, \underline{1}\right]$.

If the seasonal effect remains more or less the same regardless of the mean level of the series, then the seasonal effect is said to be additive. However if the seasonal effect increases in direct proportion to the level of the mean of the series, then the seasonal effect is said to be multiplicative. The seasonal effect can be **estimated** by means of subtraction or by means of division depending on whether or not the seasonal effect is additive or multiplicative.

An alternative seasonal adjustment procedure is to utilise the X-11 method (Shiskin *et al.* (1967)).

2.3.4 Short term Persistence

In addition to trend and seasonal fluctuations most series exhibit short term persistence implying that there exists a positive correlation between neighbouring values of the equal-interval time series. The sample autocorrelation function, correlagram (a plot of the autocorrelation at different lags against the size of the lag) and the partial autocorrelation function are useful tools in quantifying this type of pattern. The autocorrelation function measures the correlation between observations in a series that are at specified distances apart. For example, *the autocorrelation of lag 1, denoted* ρ_1 *is the correlation between* y_t *and* y_{t-1}, *the autocorrelation of lag 2, denoted* ρ_2 *is the correlation between* y_t *and* y_{t-2}. *Similarly the autocorrelation of lag* j *is* ρ_j, *the correlation between* y_t *and* y_{t-j}. The estimates are $\hat{\rho}_1$, $\hat{\rho}_2$, ..., $\hat{\rho}_j$ are sometimes written as r_1, r_2, ..., r_j.

Let Y_t be the observed data value at time t then the $j'th$ sample autocorrelation is defined as:

$$\hat{\rho}_j = \frac{cov(Y_t, Y_{t-j})}{\sqrt{var(Y_t)}\sqrt{var(Y_{t-j})}} = \frac{\sum_{t=1+j}^{T}(Y_t - \bar{Y}_t)(Y_{t-j} - \bar{Y}_t)}{\sum_{t=1}^{T}(Y_t - \bar{Y}_t)^2} \tag{2.36}$$

where $\bar{Y}_t$ is the sample mean , $\frac{1}{T}\sum_{t=1}^{T} y_t$, $\hat{\sigma}_{t,t-j}$ is the $j'th$ sample covariance such that $\widehat{cov}(Y_t, Y_{t-j}) = \frac{1}{T}\sum_{t=1+j}^{T}(Y_t - \bar{Y}_t)(Y_{t-j} - \bar{Y}_t)$, the sample variance is equal to $\hat{\sigma}^2 = \widehat{var}(Y_t)$

change for a specific month. For quarterly data, the seasonal effects can removed by using $\frac{1}{4}\left[\frac{1}{2}, 1, \underline{1}\right]$.

If the seasonal effect remains more or less the same regardless of the mean level of the series, then the seasonal effect is said to be additive. However if the seasonal effect increases in direct proportion to the level of the mean of the series, then the seasonal effect is said to be multiplicative. The seasonal effect can be **estimated** by means of subtraction or by means of division depending on whether or not the seasonal effect is additive or multiplicative.

An alternative seasonal adjustment procedure is to utilise the X-11 method (Shiskin *et al.* (1967)).

2.3.4 Short term Persistence

In addition to trend and seasonal fluctuations most series exhibit short term persistence implying that there exists a positive correlation between neighbouring values of the equal-interval time series. The sample autocorrelation function, correlagram (a plot of the autocorrelation at different lags against the size of the lag) and the partial autocorrelation function are useful tools in quantifying this type of pattern. The autocorrelation function measures the correlation between observations in a series that are at specified distances apart. For example, *the autocorrelation of lag 1, denoted* ρ_1 *is the correlation between* y_t *and* y_{t-1}, *the autocorrelation of lag 2, denoted* ρ_2 *is the correlation between* y_t *and* y_{t-2}. *Similarly the autocorrelation of lag* j *is* ρ_j, *the correlation between* y_t *and* y_{t-j}. The estimates are $\hat{\rho}_1$, $\hat{\rho}_2$, ..., $\hat{\rho}_j$ are sometimes written as r_1, r_2, ..., r_j.

Let Y_t be the observed data value at time t then the $j'th$ sample autocorrelation is defined as:

$$\hat{\rho}_j = \frac{cov(Y_t, Y_{t-j})}{\sqrt{var(Y_t)}\sqrt{var(Y_{t-j})}} = \frac{\sum_{t=1+j}^{T}(Y_t - \bar{Y}_t)(Y_{t-j} - \bar{Y}_t)}{\sum_{t=1}^{T}(Y_t - \bar{Y}_t)^2} \tag{2.36}$$

where $\bar{Y}_t$ is the sample mean , $\frac{1}{T}\sum_{t=1}^{T} y_t$, $\hat{\sigma}_{t,t-j}$ is the $j'th$ sample covariance such that $\widehat{cov}(Y_t, Y_{t-j}) = \frac{1}{T}\sum_{t=1+j}^{T}(Y_t - \bar{Y}_t)(Y_{t-j} - \bar{Y}_t)$, the sample variance is equal to $\hat{\sigma}^2 = \widehat{var}(Y_t)$

$= \frac{1}{T}\sum_{t=1}^{T}(Y_t - \bar{Y}_t)^2$ and T is the sample size.

Stuart (1994) has shown that $\hat{\rho}_j$ is a biased estimate of ρ_j since $E(\hat{\rho}_j) = \rho_j + 0(\frac{1}{T})$. This bias can be substantial when the sample size T is small. Quenouille (1949) proposed estimating the $j'th$ sample autocorrelation coefficient as $\hat{\rho}_{j^*} = 2\hat{\rho}_j - 0.5(\hat{\rho}_{j1} + \hat{\rho}_{j2})$ where $\hat{\rho}_{j1}$ and $\hat{\rho}_{j2}$ are calculated by dividing the sample into two sections and then using equation 2.36 to estimate the autocorrelations over each of the two parts. Quenouille's estimate is also a biased estimate, however its bias is of order $\frac{1}{T^2}$ and thus the estimate converges to the true autocorrelation much faster than equation 2.36 as T increases.

The correlogram is a plot of the autocorrelation at different lags against the size of the lag. It is often used as a visual aid by which to identify a type of process that best models a data series.

Tests on the autocorrelation function

Bartlett (1946) has shown that the sample autocorrelation coefficients are asymptotically normally distributed with mean 0 and variance $\frac{1}{T}$ if only if $\{Y_t\}$ is *iid* and $E(Y_t^2) < \infty$. This result is often used in practice to test the null hypothesis: $H_0 : \rho_j = 0$ versus $H_1 : \rho_j \neq 0$ for any j. The test statistic $\left(\sqrt{T}\hat{\rho}_j\right)$ is normally distributed with mean 0 and variance 1 and thus the null hypothesis will be rejected if $\left|\sqrt{T}\hat{\rho}_j\right| > z_{1-\frac{\alpha}{2}}$ where $z_{1-\frac{\alpha}{2}}$ is the limit at $100\frac{\alpha}{2}\%$ significance level.

Portmanteau Tests

Box and Pierce (1970) proposed the Portmanteau statistic $Q(m) = T\sum_{j=1}^{m}\hat{\rho}_j^2$ as a test statistic for the null hypothesis $H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ against the alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, 2, ..., m\}$. Assuming that $\{Y_t\}$ is *iid* and certain moment conditions hold, $Q(m) \sim \chi_m^2$. (Note that if $\{Y_t\}$ represents the residuals from *ARIMA* *(still to be discussed)* estimation, the appropriate degrees of freedom should be adjusted to represent the number of autocorrelations less the number of AR and MA terms previously estimated.)

Ljung and Box (1979) modified the above statistic to increase the power of the test to $Q^*(m) = T(T-2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j}$. The $Q^*(m)$-statistic is often used to test whether a data

series is a white noise series (random series). There remains the practical problem of choosing the order of lag to use for the test. If you choose too small a lag, the test may not detect serial correlation at high-order lags. However, if you choose too large a lag, the test may have low power since one or more significant correlation at one lag may be diluted by insignificant correlations at other lags. For further discussion, see Ljung and Box (1979) and Harvey (1990,1993).

Simulation studies suggests that $m \approx \ln(T)$ provides better power performance. See (Dezhbaksh, Hashem (1990)).

Example: Weekly Returns on the JSE (1985-2000)

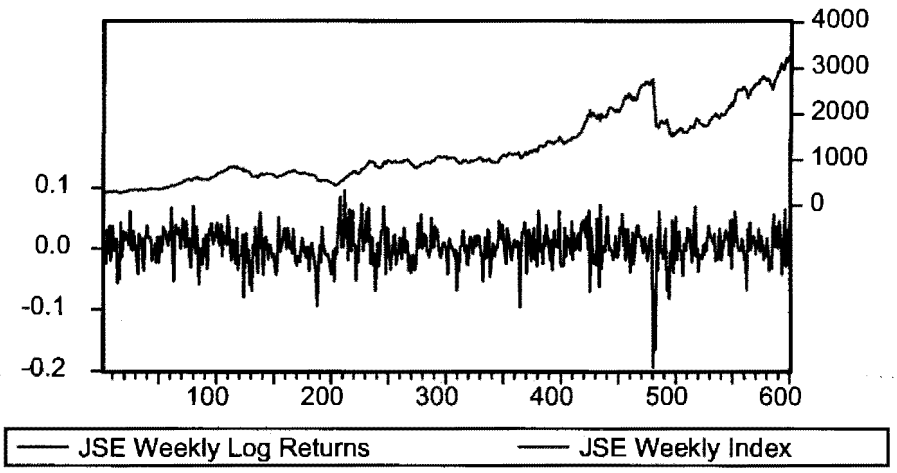


Figure 7: JSE Weekly Returns and Index

Table 3 below displays the sample autocorrelations of the Weekly log returns of the JSE and the associated $Q^*(m)$ statistic and p-values (as output from Eviews 3.1) for the first ten lags. Figure 7 plots the weekly log returns as well as the closing index of the JSE over the period 1985 to 2000.

If one were to test for significant lag structures $(H_0 : \rho_j = 0)$, the first and the seventh lags would be significant at the 5% level since their associated test statistics (8.412 and 3.252) both exceed the cut off value of 1.96 since $T = 598$. None of the other lags are significant at the 5% level.

LAG	ACF	PAC	Q*(m)	Prob
1	**0.344**	0.344	71.418	0.000
2	0.080	-0.044	75.298	0.000
3	0.017	0.004	75.480	0.000
4	0.043	0.044	76.579	0.000
5	0.042	0.015	77.639	0.000
6	-0.022	-0.049	77.934	0.000
7	**-0.133**	-0.124	88.648	0.000
8	-0.048	0.046	90.058	0.000
9	-0.014	-0.010	90.177	0.000
10	0.005	0.010	90.192	0.000

Table 3: Sample Autocorrelations and Portmanteau statistics

2.3.5 Stochastic Processes

Definitions (Dunne and Clark (2003))

A **stochastic process** $\{Z_t : t \in T\}$ is a family of random variables, all involving a response variable Z, indexed by time of observation t, in a time-frame T. The word **stochastic** implies Z_t affected by chance or random variation, and is a random variable with a distribution, for every t. The word **process** conveys development over time. For $s \neq t$, the random variable Z_s may have a different distribution from that of Z_t.

The variable Z_t may be **continuous** (sea surface temperatures Z, at time t), or **discrete** (counts Z of living persons in a specified region, at time t), or **qualitative** (category Z at time t: non-student, student, graduate, employee, post-graduate student).

A time **origin** $t = 0$ may be meaningful or arbitrary, and may or may not be a **time of observation** (*i.e.* there need not be a Z_t observable at $t = 0$). We will allow infinite **continuous** time of observation, and in both positive and negative directions; *e.g.* $(0 \leq t < \infty)$ or $(T = \mathcal{R}^+)$, and $(-\infty < t < \infty)$ or $(T = \mathcal{R})$. In this thesis we will assume the time-frame T is **discrete**. Discrete time processes may involve a **finite** or a **countably infinite** number of variables in $\{Z_t : t \in T\}$.

The term **time series** is often used to describe a stochastic process which is observed at **discrete time** points, rather than observed continuously in time. Thus we might use notation $\{Z_{t_0}, Z_{t_1}, Z_{t_2}, \ldots Z_{t_n}\}$ to suggest observations in discrete time explicitly, but we allow each Z_{t_i} to assume one of the values in a continuous range.

For convenience of notation we will have $t_0 < t_1 < t_2 < \ldots < t_n$. We may imagine the array of **univariate marginal** *pdf's* of the Z_{t_i} to be a sequence of curves $f_{t_i}(z)$, located on a horizontal T-axis at the time points $t_0, t_1, t_2, \ldots\ldots, t_n$. Each curve $f_{t_i}(z)$ is defined over a vertical axis Z. These univariate marginal *pdf's* need not all be identical, but it will be convenient for analysis if they are identical.

It may also be useful to think of the **observed time series data** as being bivariate observations (t_i, z_{t_i}) or (t_i, z_i), where $i = 0, 1, 2, \ldots n$. It is common practice to link pairs of successive points (t_i, z_{t_i}) and $(t_{i+1}, z_{t_{i+1}})$ with line segments, to indicate that the data points are not simply bivariate data, but in addition arise from a continuous process in time.

The random variable Z_t is **observable** at time t, and its observed value $Z_t = z_t$ is a realisation from the univariate (marginal) distribution $f_t(z)$ of Z_t. We know how to model a single random variable Z_t by a probability distribution $f_t(z)$. We now attempt to conceptualise all the different **simultaneous** models for the entire family $\{Z_t\}$.

For any two time points s and t, the joint behaviour of Z_t and Z_s is determined by their joint bivariate distribution. For $s < t$, we will say time s **leads** time t by $m = t - s$ time units and t **lags** time s by $m = t - s$ units. Hence we also say Z_s leads Z_t by m time units, and Z_t lags Z_s by m time units.

Specifically we allow in the pair (Z_s, Z_t), for any t and s, that Z_t and Z_s are stochastically **dependent**. It will be meaningful to consider both conditional distributions $Z_t | Z_s = z_s$ and $Z_s | Z_t = z_t$, regardless of whether or not $s < t$. In other words we will allow conditional inferences **forward** or **backward** in time, but governed by the available conditional information. For given values (s, t), the **covariance** $cov(Z_s, Z_t) = \gamma_z(t, s)$ and $|\rho_Z(t, s)| \leq 1$. The **correlation** $corr(Z_s, Z_t) = \rho_z(t, s) = \gamma_z(t, s) \div \sqrt{(\gamma_z(s, s) \times \gamma_z(t, t))}$ are partial measures of the strength of the dependence between Z_s and Z_t and by definition $cov(Z_s, Z_t) = cov(Z_t, Z_s)$. We often expect stronger relationships between observations that are closer in time.

Similarly, we presume that internal **dependencies** will often exist in the entire sequence or stream of observations on $\{Z_t\}$, and these dependencies will be carried

into the marginal distributions of m-tuples or other subsets within $\{Z_t\}$, and not just for $m = 2$, or pairs (s, t).

We will not necessarily study the whole set of all marginal distributions of a stochastic process. In many cases, much (but not necessarily all) of the relevant structural information about dependencies is carried in the **mean function** $\{\mu_t = \mu_z(t)\}$, the **variance function** $\{\sigma_t^2 = \sigma_z^2(t) = cov(Z_t, Z_t)\}$, and also more generally in the **covariance function** $\{\gamma_{t,s} = \gamma_z(t, s) = cov(Z_t, Z_s)\}$. The mean and the variance may be viewed as **continuous** functions over time. The covariance is a **continuous** function over the two-dimensional space of vectors (t, s).

For stochastic processes in discrete time these continuous functions reduce to sequences of means and variances and covariances. Thus we obtain a **mean vector**, a **variance vector** and a **covariance matrix** structure $V = [\gamma_{t,s}]$.

Using the definition of the **correlation coefficient** $\rho_{t,s}$ of Z_t and Z_s, we write $\rho_{t,s} = \gamma_{t,s} \div (\sigma_t \cdot \sigma_s)$, we may find the **correlation matrix** structure $P = [\rho_{t,s}]$.

In many contexts it will be possible and practical to observe data $Z_t = z_t$ at **equally spaced time-points**, say $t = \delta \pm k.u$ where $0 \leq \delta < u$, and $k = 0, 1, 2, \ldots$, where u is the **constant interval** between points of observation, and δ is the arbitrary **location shift**. Often we will set $\delta = 0$ and $u = 1$. We may under these conditions write data (t_i, Z_{t_i}) as (i, Z_i), where $i = 0, 1, 2, \ldots n$, without losing any information. Thus the notation $\{Z_0, Z_1, \ldots, Z_n\}$ is used to convey a first observable Z_0 at any time origin, followed by a subsequent observable at each of n equally spaced time points, one unit of time apart. The unit on the T-axis will vary from one context to another, as seconds, minutes, hours, days, weeks, months, years, and so on, depending upon the common interval between observations.

To study a countably infinite stochastic process $\{\ldots, Z_{-2}, Z_{-1}, Z_0, Z_1, \ldots\}$ we should technically be able to describe the set of joint **marginal distributions** of every possible finite and infinite subsequence of the random variables in the process.

For a **finite** stochastic process with equally spaced observations $\{Z_0, Z_1, \ldots, Z_n\}$, we would in general wish to study the joint multivariate distribution of all $n + 1$ variables, from which we could find all joint marginal and conditional distributions

of all possible subsets of any size $m \leq n$: say $\{Z_{t_1}, Z_{t_2}, \ldots Z_{t_m}\}$. There are $(2^{n+1} - 1)$ such subsets, so that such a study would be very complex, and would require many assumptions before the problem became tractable.

For a finite process $\{Z_0, Z_1, \ldots, Z_n\}$, we may construct a diagonal **scaling matrix** D with diagonal entries determined by the variance function $\gamma_z(i,i) = cov(Z_i, Z_i) = \sigma_i^2$. Using the definition of the **correlation coefficient** $\rho_{i,j}$ of Z_i and Z_j, we may find the **correlation matrix** structure $P = [\rho_{i,j}]$, and we may write $P = D^{-\frac{1}{2}} V D^{-\frac{1}{2}}$ and $V = D^{\frac{1}{2}} P D^{\frac{1}{2}}$. By construction, P has ones on the diagonal and all off-diagonal entries have absolute value at most one. These structures P and V are **symmetric non-negative definite** matrices, but are otherwise arbitrary.

When we consider two or more stochastic processes, say $\{Z_t\}$ and $\{W_t\}$, we will be interested in contrasting them on the basis of their mean functions $\{\mu_z(t)\}$ and $\{\mu_w(t)\}$ and covariance structures, $\gamma_z(t,s)$ and $\gamma_w(t,s)$.

If we combine two or more stochastic processes to make a new process, say $\{X_t\}$ where $X_t = Z_t + W_t$, we will want to know how the means $\{\mu_x(t)\}$ and covariances $\gamma_x(t,s)$ relate to the corresponding structures of $\{Z_t\}$ and $\{W_t\}$.

Strict Stationarity

It is useful to begin the study of stochastic processes with a simpler task, examining a small subset of stochastic processes that are easier to handle. Initially we consider only those processes that have a property called strict stationarity.

The infinite process $\{Z_t\}$ is **strictly stationary** if and only if, for every count n, all time points $\{t_1, t_2, \ldots t_n\}$, and for all **time-shifts** $m > 0$, the joint marginal distribution of $\{Z_{t_1}, Z_{t_2}, \ldots Z_{t_n}\}$ is identical to the joint marginal distribution of $\{Z_{t_1-m}, Z_{t_2-m}, \ldots Z_{t_n-m}\}$. Note the effect of varying n and m through all possible values.

This strict stationarity condition is very strong. For just the count $n = 1$, it requires that for any t, Z_t and Z_{t-m} are identically distributed for every time-shift m. Thus all Z_t have the same univariate marginal distribution (and hence **common mean** and **common variance**).

Setting $n = 2$, the condition requires that for any (t, s), (Z_t, Z_s) and (Z_{t-m}, Z_{s-m}) are identically distributed for every time-shift m. Thus all pairs (Z_{t-m}, Z_{s-m}) have the same bivariate distribution, and while this common distribution is a function of (t, s), it depends on (t, s) only through the common **time difference** $|t - s|$. We say the distribution of (Z_t, Z_s) is a function of the **lag** $k = |t - s|$. Consequently, for any (t, s), we may derive a **common covariance**, since:

$$\gamma_{t,s} = \begin{cases} \gamma_{t-m,s-m} & for \quad \text{all } m \\ = \gamma_{0,s-t} = \gamma_{t-s,0} = \gamma_{|t-s|} & for \quad m = t \text{ and } m = s \end{cases} \tag{2.37}$$

Under these conditions we define the **autocovariance function (ACV)** as a function of the lags $k = |t - s|$ alone:

$$\gamma_k = \gamma(k) = \gamma_{|t-s|} = \gamma(-k) = \gamma_{-k} \qquad \text{for all } t, s \tag{2.38}$$

Note that $\gamma_0 = \sigma_t^2$ for all t. Hence we define the **autocorrelation function (ACF)** using the equations:

$$\rho_k = \rho(k) = \rho_{|t-s|} = (\gamma_k \div \gamma_0) \tag{2.39}$$

For our purposes the major consequence of strict stationarity is the enormous simplification of the mean and covariances structures. These simplifications allow us to define the two important functions **ACV** and **ACF.** The facts, firstly that the mean is a constant, and secondly that the covariances at any lag k are constant and functions of k, are **necessary** conditions arising from the definition of strict stationarity. However these two conditions alone are **not sufficient** to guarantee that a process is strictly stationary, because further necessary conditions arise when we consider m-tuples when $m = 3, 4, ...$

If we also presume the variables are jointly **multivariate Gaussian**, the sequence of means $\{\mu_Z(t) = \mu_t\}$ and the covariance matrix $V = [\gamma_{t,s}]$ fully specify the joint distribution, and all the marginals and conditionals, whether or not stationarity applies. However $\{\mu_Z(t) = \mu_t = \mu\}$ and $V = [\gamma_{t,s}] = [\gamma_{|t-s|}]$ for the stationary Gaussian.

An important strictly stationary process is $\{A_t\}$ with all A_t identically and independently distributed (**iid**) and **zero mean.** This process is called **(pure) white noise.** It is easy to verify that the **iid** variables satisfy the criteria for strict stationarity. Furthermore we obtain the constant zero mean and a diagonal covariance

Recall that the covariances will allow us to establish all the correlation coefficients, but that the correlation matrix structure P will determine V only partly, as it will not give us the value of γ_0. If we are given P and γ_0 we may find $V = \gamma_0 \cdot P$.

$$P = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \cdots & \rho_n \\ \rho_1 & 1 & \rho_1 & \ddots & \ddots & \vdots \\ \rho_2 & \rho_1 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \rho_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \rho_1 \\ \rho_n & \cdots & \cdots & \rho_2 & \rho_1 & 1 \end{bmatrix}. \tag{2.42}$$

We will be concerned to find ways of distinguishing between stationary processes that present us with similar or identical mean and covariance structures. Two infinite discrete-time processes $\{Z_t\}$ and $\{Y_t\}$, obtained from pure white noise $\{A_t\}$ by the equations $Z_t = 2A_t - A_{t-1}$ and $Y_t = A_t - 2A_{t-1}$, will give us identical covariance matrices of the form.

$$V = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & +5 & -2 & +0 & \ddots \\ \ddots & -2 & +5 & -2 & \ddots \\ \ddots & +0 & -2 & +5 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{2.43}$$

From those identical structures we will not be able distinguish which of the two processes, $\{Z_t\}$ or $\{Y_t\}$, we are observing. Additional information or methods will be required to make that distinction.

Another important facet of the relationships between the values observable at different points in the stationary time series, arises from the **partial autocorrelation function** or **PACF**. This function focusses upon the strength of the relationship across a **lag** of k intervals, between say (Z_t, Z_{t-k}), but specifically after all the $(k-1)$ intervening values $\{Z_{t-1}, \ldots, Z_{t-k+1}\}$ have been taken into account. The function is

defined by:

$$\phi(k) = corr(Z_t, Z_{t-k} \mid Z_{t-1}, \cdots, Z_{t-k+1}) \tag{2.44.a}$$

$$= \frac{cov(Z_t, Z_{t-k} \mid Z_{t-1}, \cdots, Z_{t-k+1})}{\sqrt{var(Z_t \mid Z_{t-1}, \cdots, Z_{t-k+1}) \times var(Z_{t-k} \mid Z_{t-1}, \cdots, Z_{t-k+1})}} \tag{2.44.b}$$

Notice that the value of $\phi(k)$ is constant over all t in stationary time series. A zero value for a partial correlation coefficient at lag k suggests that after fitting all $(k-1)$ intervening variables, neither Z_t nor Z_{t-k} have any further explanatory dependence upon or relationship to one another, even if they were as a pair originally strongly correlated.

Non-stationarity

There will be other processes that are **non-stationary**, such as those with either a mean function $\{\mu_t\}$ that changes over time, or those with covariances whose values are functions of (t, s) other than a multiple of $|t - s|$. We will have either to adapt any methods for stationary processes before they can also be applied to non-stationary processes, or to transform non-stationary processes to render them stationary, and then apply the methods.

2.4 Model Building Processes

A brief overview of the model building process is represented graphically below (as adapted from Box *et al.* (1994)). It can be seen that the process is iterative since one is seldom able to obtain a deterministic model for a time series. A model is said to be **deterministic** if the time series could be modelled exactly by some mathematical formula. For example, *the path travelled by a body in space could theoretically be known since there exists some mathematical formulae that model precisely the movement of bodies in space.* Models that are not deterministic are called **stochastic processes.** These processes by definition include a component that is random, operating at each time observation. A stochastic process is a collection of random variables indexed by time. The process is not governed by a deterministic model since it is affected by chance or random variation. For example, *the process of monthly returns of a company cannot be predicted with a high degree of accuracy since it may be influenced*

by a large number of variables. We may model the returns using variables we assume to be most important, and regard the remaining as one aggregate component.

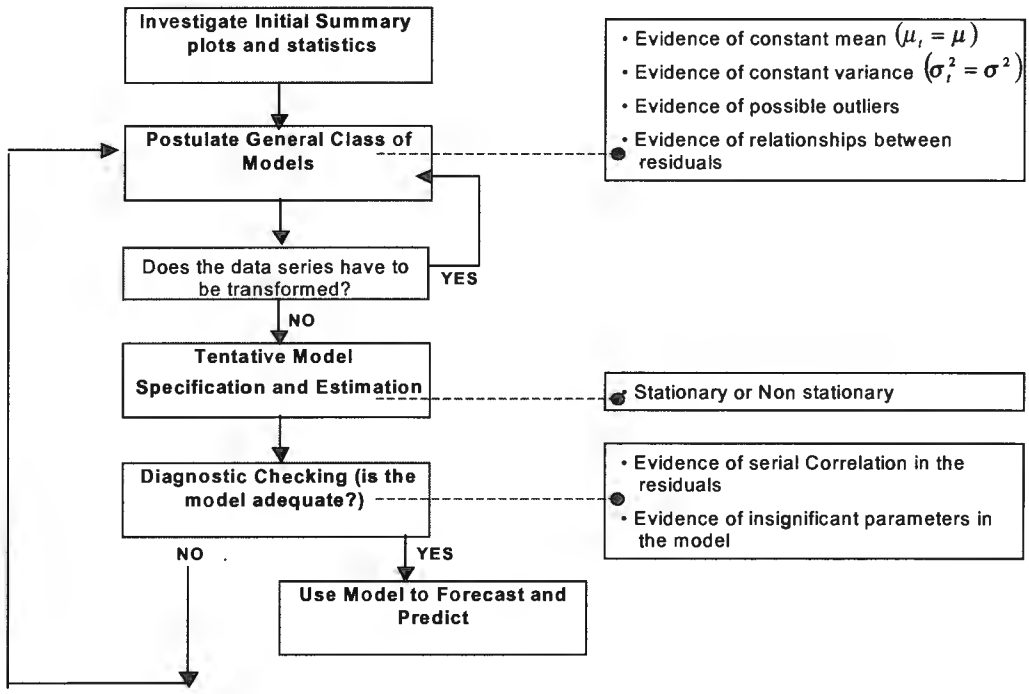


Figure 8: The Model Building Process

1. The first step in any time series analysis is to plot each of the variables that one is interested in against time. This step should provide an initial summary of the salient features of the series. At this stage one should observe whether or not any abnormal data points are present in the series. These points could either be due to incorrect data collection or they may be valid but potential outliers. Outliers are observations that are very different to the bulk of a data set. Note however that in a financial context, abnormal values may be classified as being outliers even though they may have arisen due to some important economic activity. For example, *values associated with a market crash may appear as outliers in a plot covering a long period of time.* From a time series point of view, outliers could be handled by using the intervention analysis of Box and Tiao (1975). One should also determine whether or not the series has to be transformed, as transformations can give rise to series of values that appear less widely spread.

2. From the examination of the initial time plots as well as other graphical aids (e.g. correllograms and histograms) one should be able to tentatively identify a class of model that might fit a series adequately to, i.e. one should be able to

indicate whether or not the series is stationary or not.

3. Once a tentative model class has been chosen, the analyst should then impose a particular model to the data set and undertake parameter estimation.

4. Diagnostic checking is the process of scrutinising a model in order to determine its adequacy. This scrutiny would entail checking the residuals for anything untoward. Parameter estimates are also checked in order to discard any parameters for which there is insufficient statistical evidence to warrant keeping in the model.

Chapter 3

Linear Time Series Modelling

3.1 Simple Introductory Models

This section describes several types of stochastic processes which could be useful models for time series.

A White Noise Process

A white noise process, also called a purely random process is a sequence of random variables, $\{e_t\}$ such that $e_t \sim iid(0, \sigma^2)$. This means that each of the observations from the sequence $\{e_t\}$ is a (random) realization from a $iid(0, \sigma^2)$ distribution. By definition this series is random and one cannot predict future values of such a series. The best one step ahead estimate is $E(\varepsilon_{t+1}|\varepsilon_t) = 0$. Similarly $E(\varepsilon_{t+j}|\varepsilon_t) = 0$ for any arbitrary integer j. The random component of a time series is often modelled as a white noise process within stationary and non-stationary models.

Random Walk Process

A discrete process $\{Y_t\}$ follows a random walk process if the random variables $\{Y_t\}$ is defined as:

$$Y_t = Y_{t-1} + e_t \text{ where } e_t \sim iid(0, \sigma^2) \text{ for all } t \text{ and } y_1 = e_1 \tag{3.1}$$

We note that $Y_t = \sum_{i=1}^{t} e_i$ since the Y_t can be recursively expanded using equation 3.1. The autocovariance and the autocorrelation structure are defined as follows:

$$var\,(Y_t) = var\left(\sum_{i=1}^{t} e_i\right) = t\sigma^2 \tag{3.2.a}$$

$$cov\,(Y_t, Y_s) = cov\,(Y_s, Y_t) = \gamma\,(t,s) = E\left(\sum_{i=1}^{t} e_i \sum_{i=1}^{s} e_i\right) \tag{3.2.b}$$

$$= s\sigma^2 \quad \text{if } s \leq t \tag{3.2.c}$$

$$corr\,(Y_t, Y_s) = \sqrt{\frac{s}{t}} \quad \text{if } 1 \leq s \leq t \tag{3.2.d}$$

The random walk process is non-stationary despite the constant mean ($\mu = 0$) because the covariance structure depends on time (as seen by equations 3.2.a through to 3.2.d). Figure 1 below displays three different random walk processes each having been generated by assuming that the residual series, $\{e_t\}$ is a standard normal random variable and that $e_1 = 0.5$.

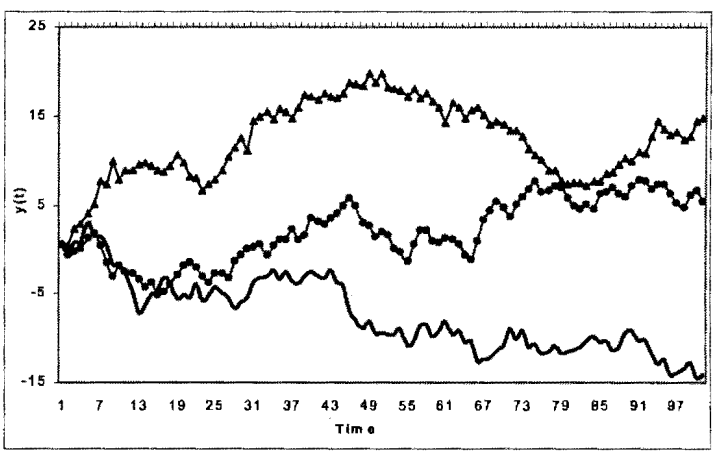


Figure 1: Three different Random Walk Processes

Random Walk Process with constant drift

A discrete process $\{Y_t\}$ follows a random walk process with constant drift μ, if the random variables $\{Y_t\}$ is defined as:

$$Y_t = \mu + Y_{t-1} + e_t \text{ where } e_t \sim iid(0, \sigma^2) \text{ for } t = 1, 2, 3, \ldots \text{ and } y_0 = 0 \tag{3.3}$$

We note that $Y_t = t\mu + \sum_{i=1}^{t} e_i$ because the Y_t can be recursively expanded using equation 3.3. Thus $E(Y_t) = t\mu$. The autocovariance and the autocorrelation structure are defined as follows:

$$var\,(Y_t) = E\left([y_t - t\mu]^2\right) \tag{3.4.a}$$

$$= E\left(\left[\sum_{i=1}^{t} e_i\right]^2\right) = t\sigma^2 \tag{3.4.b}$$

$$
\begin{aligned}
cov\left(Y_t, Y_s\right) &= cov\left(Y_s, Y_t\right) = \gamma\left(t, s\right) && \text{(3.4.c)} \\
&= E\left(\sum_{i=1}^{t} e_i \sum_{i=1}^{s} e_i\right) && \text{(3.4.d)} \\
&= s\sigma^2 \qquad \text{if } s \leq t && \text{(3.4.e)} \\
corr\left(Y_t, Y_s\right) &= \sqrt{\frac{s}{t}} \qquad \text{if } 1 \leq s \leq t && \text{(3.4.f)}
\end{aligned}
$$

The random walk process with constant drift is non-stationary because both the covariance structure and the mean equation depends on time t (as seen by equations 3.4.a through to 3.4.f). Differencing the series $\{Y_t\}$ will generate a weakly stationary series $\{Z_t\}$ such that:

$$
\begin{aligned}
Z_t &= \nabla\left(Y_t\right) = \nabla\left(t\mu + \sum_{i=1}^{t} e_i\right) && \text{(3.5.a)} \\
&= t\mu + \sum_{i=1}^{t} e_i - \left\{\left(t-1\right)\mu + \sum_{i=1}^{t-1} e_i\right\} && \text{(3.5.b)} \\
&= t\mu + \sum_{i=1}^{t} e_i - t\mu + \mu - \sum_{i=1}^{t-1} e_i && \text{(3.5.c)} \\
&= \mu + e_t && \text{(3.5.d)}
\end{aligned}
$$

$\{Z_t\}$ is weakly stationary because the covariance structure and the mean equation does not depend on time: $E(Z_t) = \mu$, $cov\left(Z_t, Z_s\right) = 0$ for all $t \neq s$ and $var(Z_t) = \sigma^2$.

3.2 Models for Stationary Time Series

3.2.1 General Linear Processes

Let $\{\varepsilon_t\}$ be a white noise process with mean 0 and variance σ^2. A general linear process $\{Z_t\}$ is a series that can be defined as a weighted linear combination of the white noise process $\{\varepsilon_t\}$. $\{Z_t\}$ is defined as:

$$Z_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \text{ with } \left|\sum_{i=0}^{\infty} \psi_i^2\right| < \infty \text{ and } \sum_{i=0}^{\infty} \psi_i^2 < \infty \text{ and } \psi_0 = 1 \tag{3.6}$$

We investigate the properties of $\{Z_t\}$ if $\psi_i = \phi^i$ for $\phi \in (0,1)$. $Z_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$. $E(Z_t) = 0$ for all t. The autocovariance and autocorrelation structure is defined as:

$$var(Z_t) = E\left(\left[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right]^2\right) = E\left(\sigma^2 \sum_{i=0}^{\infty} \phi^{2i}\right) = \frac{\sigma^2}{1-\phi^2} \tag{3.7}$$

$$cov(Z_t, Z_{t+k}) = E\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \sum_{i=0}^{\infty} \phi^i \varepsilon_{t+k-i}\right) \quad \text{for } k = 0, 1, 2, ... \tag{3.8.a}$$

Now expanding the second summation we get

$$E\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \left[\varepsilon_{t+k} + \phi\varepsilon_{t+k-1} + \phi^2\varepsilon_{t+k-2} + ... + \phi^k\varepsilon_t + \phi^{k+1}\varepsilon_{t-1} + \phi^{k+2}\varepsilon_{t-2} + ...\right]\right)$$

$$= E\left(\left[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right]\left[\phi^k \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right]\right) \tag{3.8.c}$$

$$= \phi^k E\left(\left[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}\right]^2\right) \tag{3.8.d}$$

$$= \frac{\phi^k \sigma^2}{1-\phi^2} \tag{3.8.e}$$

$$corr(Z_t, Z_{t+k}) = \phi^k \quad \text{for } k = 0, 1, 2, ... \tag{3.9}$$

Equation 3.8.c follows due to the independence of the error terms. i.e. $E(e_i e_j) = 0$ for all $i \neq j$. Note that general linear processes are weakly stationary of order two since both the mean and the covariance structure does not depend on time t. For a

general linear process $Z_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ the following properties hold:

$$E(Z_t) = 0 \quad \text{for all } t \tag{3.10.a}$$

$$cov(Z_t, Z_{t+k}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \quad \text{for all } t \text{ and } k \tag{3.10.b}$$

$$corr(Z_t, Z_{t+k}) = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2} \quad \text{for all } t \text{ and } k \tag{3.10.c}$$

3.2.2 Moving Average Processes (MA Processes)

The moving average process of order q is defined as:

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots\ldots - \theta_q \varepsilon_{t-q} \tag{3.11}$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are constants and $\{\varepsilon_t\}$ is white noise process. The term moving average comes from the fact that $\{Z_t\}$ is constructed from a weighted sum, akin to an average of the q most recent values of ε_t. The MA models are always weakly stationary because they are finite linear combinations of a white noise sequence for which the first two moments are time invariant. In the following section we will examine the properties of different moving average models.

First-Order Moving Average Processes

A moving average of order one, MA(1), is defined as follows:

$$Z_t = \varepsilon_t - \theta \varepsilon_{t-1} \tag{3.12.a}$$

$$Z_t = \theta(B) \varepsilon_t \tag{3.12.b}$$

where $\{\varepsilon_t\}$ is white noise process , $\theta(B) = 1 - \theta B$ and B^k is the backshift operator defined as $B^k(Z_t) = Z_{t-k}$. The expectation of Z_t is $E(Z_t) = 0$. The variance of Z_t is equal to:

$$Var(Z_t) = E(Z_t^2) = E(\varepsilon_t - \theta \varepsilon_{t-1})^2 = (1 + \theta^2)\sigma^2 \tag{3.13}$$

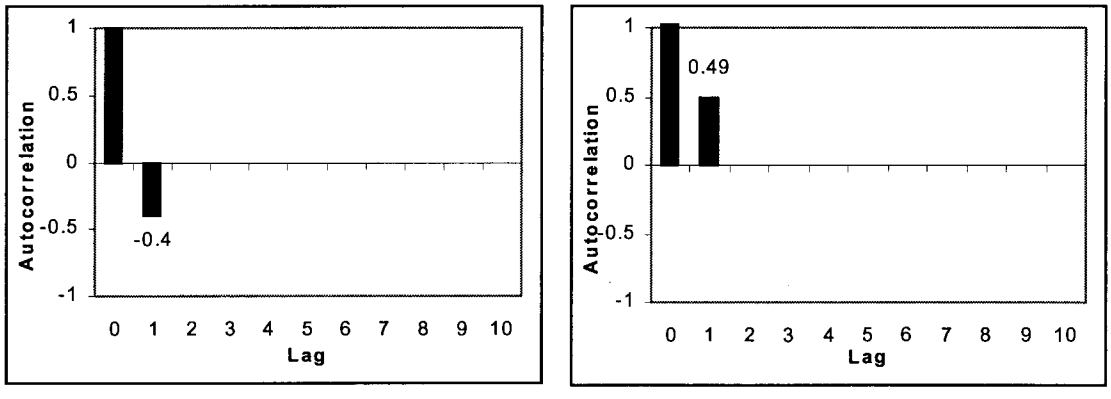
since: $E(\varepsilon_t \varepsilon_{t-1}) = 0$ for all t and $E(\varepsilon_t \varepsilon_t) = \sigma^2$ for all t. The autocovariance and autocorrelation structure is defined as follows:

$$cov(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}) \quad \text{for } k \geqslant 1 \tag{3.14.a}$$

$$= E(\{\varepsilon_t - \theta \varepsilon_{t-1}\}\{\varepsilon_{t+k} - \theta \varepsilon_{t+k-1}\}) \tag{3.14.b}$$

$$cov(Z_t, Z_{t+k}) = \begin{cases} -\theta\sigma^2 & \text{for} \quad k = 1 \\ 0 & \text{for} \quad |k| \geq 2 \end{cases} \tag{3.14.c}$$

$$corr(Z_t, Z_{t+k}) = \begin{cases} \frac{-\theta}{1+\theta^2} & \text{for} \quad k = 1 \\ 0 & \text{for} \quad |k| \geq 2 \end{cases} \tag{3.14.d}$$



$\theta = 0.5$ $\theta = -0.8$

Figure 2: Two different autocorrelation functions for an MA(1)

Since both the first and the second moments of the series are invariant with respect to time, the MA(1) process is weakly stationary of order two. Figure 2 above displays the autocorrelation function when $\theta = 0.5$ and $\theta = -0.8$ respectively. It can be seen that by definition $\rho_0 = 1$ and that the MA(1) autocorrelation function cuts off after lag one (is equal to zero). This cutting off property of an $MA(1)$ process will be used in order to identify a process to model sample data by examining the sample autocorrelation function of time series data.

Second-Order Moving Average Processes

A moving average of order two, MA(2), is defined as follows:

$$Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \tag{3.15.a}$$

$$Z_t = \theta(B)\varepsilon_t \tag{3.15.b}$$

where $\{\varepsilon_t\}$ is white noise process , $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$. The expectation of Z_t is $E(Z_t) = 0$. The variance of Z_t is equal to:

$$Var(Z_t) = E(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})^2 = (1 + \theta_1^2 + \theta_2^2)\sigma^2 \tag{3.16}$$

The autocovariance and autocorrelation function (ACF) is defined as follows:

$$cov(Z_t, Z_{t+k}) = E(Z_t Z_{t+k}) \quad \text{for } k \geqslant 1 \tag{3.17.a}$$

$$= E(\{\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}\}\{\varepsilon_{t+k} - \theta_1\varepsilon_{t+k-1} - \theta_2\varepsilon_{t+k-2}\}) \tag{3.17.b}$$

$$= \begin{cases} (-\theta_1 + \theta_1\theta_2)\,\sigma^2 & \text{for } k = 1 \\ -\theta_2\sigma^2 & \text{for } k = 2 \\ 0 & \text{for } |k| \geq 3 \end{cases} \tag{3.17.c}$$

$$corr(Z_t, Z_{t+k}) = \begin{cases} \frac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{for } k = 1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & \text{for } k = 2 \\ 0 & \text{for } |k| \geq 3 \end{cases} \tag{3.17.d}$$

Notice that the ACF for the MA(2) cuts off at lag two where as the ACF for the MA(1) cut off at lag one. This property holds for other MA models as well. It will be seen that for an MA(q) model, the lagged autocorrelation functions up to and including q are non zero, but for $k > q$ the ACF values is equal to zero.

The q^{th} order Moving Average Process

Generalising, we define a MA(q) process. It is defined as follows:

$$Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ..... - \theta_q\varepsilon_{t-q} \tag{3.18.a}$$

$$Z_t = \theta(B)\varepsilon_t \tag{3.18.b}$$

where: $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ..... - \theta_q B^q$. The mean of the process is *zero* for all t and the variance (γ_0) of the process is:

$$var\,(Z_t) = E(Z_t^2) \tag{3.19.a}$$

$$= E(\{\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ..... - \theta_q\varepsilon_{t-q}\}^2) \tag{3.19.b}$$

$$= (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma^2 \tag{3.19.c}$$

since $E(\varepsilon_t\varepsilon_{t-q}) = 0$ for all $q > 0$ and for all t. Note that equation 3.18.a suggests that the process $\{Z_t\}$ is dependent upon only the recent history of error terms. (q periods in all.) The earlier history is thus irrelevant and hence any term Z_{t-k} for $k > q$ will have no common error term.

The autocovariance and autocorrelation structure is non-zero for $1 \leq k \leq q$ such

that:

$$cov(Z_t, Z_{t+k}) = \begin{cases} (-\theta_k+\theta_{k+1}\theta_1+\theta_{k+2}\theta_2+... + \theta_q\theta_{q-k})\sigma^2 & \text{(3.20.a)} \\ \frac{-\theta_k+\theta_{k+1}\theta_1+\theta_{k+2}\theta_2+...+\theta_q\theta_{q-k}}{1+\theta_1^2+\theta_2^2+\theta_3^2+.....+\theta_q^2} & \text{(3.20.b)} \end{cases}$$

The infinite Moving Average Process

The process can be defined as follows:

$$Z_t = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} \tag{3.21}$$

with $\omega_0 = 1$ and $\omega_i = -\theta_i$ for $i = \{1, 2, ..., \infty\}$

We allow T to play the role of n, the data set size, by using the symbol T for the number of terms in a series. The mean of the process is 0 and the variance is:

$$\gamma_0 = \lim_{T \to \infty} (\omega_0^2 + \omega_1^2 + \omega_2^2 + ... + \omega_T^2)\sigma^2$$

The covariance and the autocovariance structure is defined as:

$$\begin{aligned} cov\,(Z_t, Z_{t+k}) &= E\,(Z_t Z_{t+k}) \\ &= \lim_{T \to \infty} E\left(\sum_{i=0}^{T} \omega_i \varepsilon_{t-i} \sum_{j=0}^{T} \omega_j \varepsilon_{t+k-j}\right) && \text{(3.22.b)} \\ &= \lim_{T \to \infty} \sigma^2 \sum_{i=0}^{T} \omega_i \omega_{k+i} && \text{(3.22.c)} \end{aligned}$$

$$corr\,(Z_t, Z_{t+k}) = \lim_{T \to \infty} \frac{\sum_{i=0}^{T} \omega_i \omega_{k+i}}{\sum_{i=0}^{T} \omega_i^2} \tag{3.22.d}$$

Equation 3.22.c holds because $E\left(\sum_{i=0}^{T} \omega_i \varepsilon_{t-i} \sum_{j=0}^{T} \omega_j \varepsilon_{t+k-j}\right)$ will only have non-zero terms when $j = k + i$.

Identifying the order of a MA model

The ACF function can be used in order to identify the order of an MA model. Recall that for a MA(q) model, the ACF terms at lags greater than q, is zero. This property is used to identify the order of the process. This criterion can be stated emperically as follows: for a time series Z_t with ACF ρ_k, if $|\rho_k| > \frac{2}{\sqrt{T}}$ for $1 \leq k \leq q$, but $\rho_k = 0$ for $k > q$, then Z_t follows a MA(q) model.

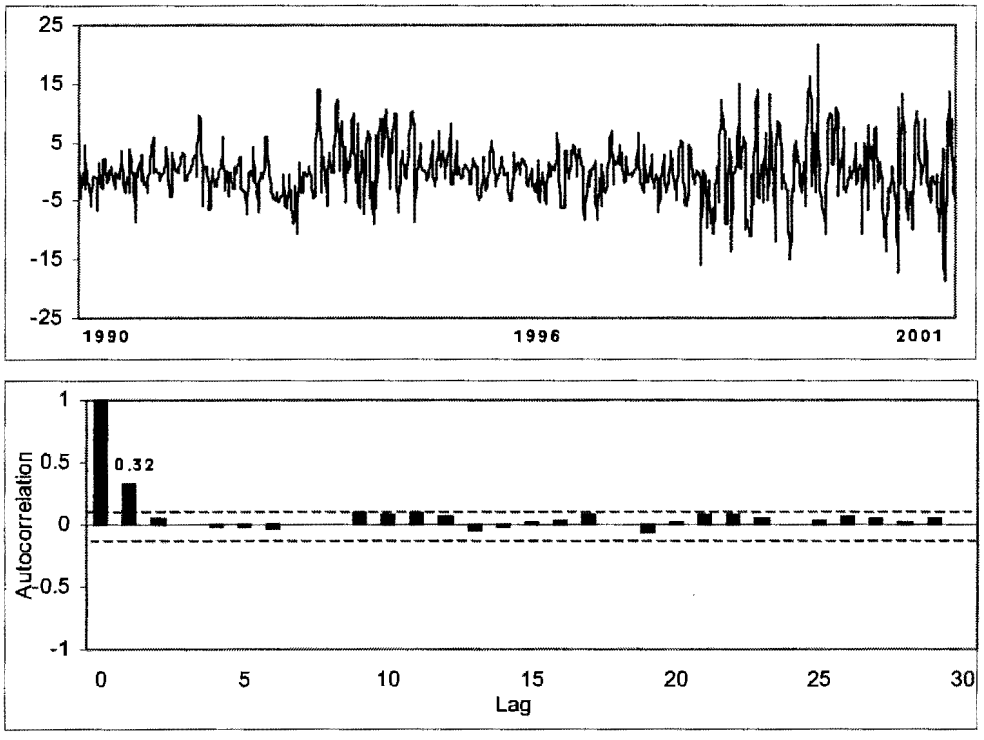


Figure 3: Time plot and sample ACF of ISCOR weekly returns (1990-2000)

Figure 3 above displays the time plot of the weekly log returns of ISCOR over the period 6 January 1990 to 30 December 2000 and the sample autocorrelation function of the series. The two dashed lines indicate the two standard deviation cut off limits $\left(\pm\frac{2}{\sqrt{T}} = \pm 0.084\right)$. It can be seen that the series has a significant autocorrelation (0.323) at lag one. Based on the sample autocorrelation function the MA(1) model, $\varepsilon_t - \theta\varepsilon_{t-1}$ is identified for the series. We have the autocorrelation $\hat{\rho}_1 = 0.323$ and $\hat{\rho}_1 = \frac{-\theta}{1+\theta^2}$ suggests $\hat{\theta} = -2.729\,6$,or $\hat{\theta} = -0.36645$. (Notice that $-2.729\,6 = \frac{1}{-0.366\,45}$.)

Estimation of the MA(1)Process.

The principle of Maximum Likelihood (ML) is usually used in order to estimate the parameters in MA models. Two approaches are commonly used to evaluate the likelihood function. The first assumes that the initial shocks $e_t = 0$, for $t \leq 0$ and is a conditional likelihood method. The second is an exact likelihood method and is the preferred method. The next section contrasts the two ML approaches by estimating the parameters of a MA(1) model. The following methods can be generalised to higher order MA models but will not be discussed at this stage.

The Conditional Likelihood Function Approach (Hamilton (1994a))

Consider the Gaussian $MA(1)$ process:

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{3.23}$$

with $\varepsilon_t \backsim i.i.d\ N(0, \sigma^2)$. Let $\theta' = \begin{pmatrix} \theta_1 & \sigma^2 \end{pmatrix}'$ denote the population parameters to be estimated. If the value of ε_{t-1} were known with certainty, then $Z_t|\varepsilon_{t-1} \sim N(-\theta_1 \varepsilon_{t-1}, \sigma^2)$ and the conditional density of $Z_t|(\varepsilon_{t-1}, \boldsymbol{\theta})$ is equal to:

$$f(Z_t|\varepsilon_{t-1}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Z_t + \theta_1 \varepsilon_{t-1})^2}{2\sigma^2}\right) \tag{3.24}$$

Assume we knew for certain that $\varepsilon_0 = 0$. Then $Z_1|(\varepsilon_0 = 0, \boldsymbol{\theta}) \backsim N(0, \sigma^2)$. Given observation Z_1 the value of ε_1 is then known with certainty as $\varepsilon_1 = Z_1$ so that:

$$f(Z_2|Z_1, \varepsilon_0 = 0; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Z_2 + \theta_1 \varepsilon_1)^2}{2\sigma^2}\right) \tag{3.25}$$

Since ε_1 is known with certainty, ε_2 can be calculated from $\varepsilon_2 = Z_2 + \theta_1 \varepsilon_1$.

Thus, proceeding in this way, given $\varepsilon_0 = 0$, the full sequence $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T\}$ can be calculated from $\{Z_1, \ldots, Z_T\}$ by iterating on $\varepsilon_t = Z_t + \theta_1 \varepsilon_{t-1}$ for $t = 1, 2, \ldots, T$. The t^{th} conditional density can then be calculated from equation 3.25 and is equal to:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Z_t + \theta \varepsilon_{t-1})^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon_t^2\right) \tag{3.26}$$

The sample likelihood would then be the product of these individual densities:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{T} f_{Z_t, Z_{t-1}, \ldots Z_1|\varepsilon_0=0}(Z_t|Z_{t-1}, \ldots, Z_1, \varepsilon_0 = 0; \boldsymbol{\theta}) \tag{3.27.a}$$

$$= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}\varepsilon_t^2) \tag{3.27.b}$$

The conditional log likelihood is equal to:

$$l(\boldsymbol{\theta}) = \frac{-T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \tag{3.28}$$

The function $l(\boldsymbol{\theta})$ is a complicated non-linear function of σ^2 and θ_1 and the maximum likelihood estimates can only be found by using numerical optimization.

The Exact Likelihood Function Approach (Hamilton (1994))

Where as the conditional method focuses upon the conditional distribution given $\varepsilon_0 = 0$, the exact method focusses upon the marginal distribution of $Z_1, \ldots\ldots, Z_T$ within the complete stochastic process $\{Z_t\}$. Let $Z' = \begin{pmatrix} Z_1 & \ldots & Z_T \end{pmatrix}'$ have zero mean and the $(T \times T)$ covariance matrix be $E(\mathbf{Z}\mathbf{Z}') = \Omega$ with:

$$\Omega = \sigma^2 \begin{bmatrix} (1+\theta_1^2) & -\theta_1 & 0 & \dots & 0 \\ -\theta_1 & (1+\theta_1^2) & -\theta_1 & \dots & 0 \\ 0 & -\theta_1 & (1+\theta_1^2) & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & -\theta_1 \\ 0 & 0 & 0 & -\theta_1 & (1+\theta_1^2) \end{bmatrix} \tag{3.29}$$

from equation 3.13 and 3.14.*c*. The likelihood function is then equal to:

$$f_{\mathbf{Z}}(\mathbf{Z};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{T/2}\,|\Omega|^{1/2}} \exp[-\frac{1}{2}(\mathbf{Z}'\Omega^{-1}\mathbf{Z})]. \tag{3.30}$$

Using the triangular factorization:

$$\Omega = \mathbf{ADA}' \tag{3.31}$$

where A is a lower triangular matrix and D is a diagonal matrix. Substituting equation 3.31 into equation 3.30 we obtain:

$$f_{\mathbf{Z}}(\mathbf{Z};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{T/2}\,|\mathbf{ADA}'|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{Z}'[\mathbf{A}']^{-1}\mathbf{D}^{-1}\mathbf{A}^{-1}\mathbf{Z}\right) \tag{3.32}$$

From matrix theory the matrix A is given by:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ \frac{-\theta_1}{1+\theta_1^2} & 1 & 0 & 0 & & 0 \\ 0 & \frac{-\theta_1(1+\theta_1^2)}{1+\theta_1^2+\theta_1^4} & 1 & 0 & & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{-\theta_1[1+\theta_1^2+\theta_1^4+\dots+\theta_1^{2(T-2)}]}{1+\theta_1^2+\theta_1^4+\dots+\theta_1^{2(T-1)}} & 1 \end{bmatrix} \tag{3.33}$$

and the diagonal matrix D is given by:

$$\mathbf{D} = \sigma^2 \begin{bmatrix} 1+\theta_1^2 & 0 & 0 & \dots & \dots & 0 \\ 0 & \frac{1+\theta_1^2+\theta_1^4}{1+\theta_1^2} & & \dots & \dots & 0 \\ 0 & 0 & \frac{1+\theta_1^2+\theta_1^4+\theta_1^6}{1+\theta_1^2+\theta_1^4} & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & \frac{1+\theta_1^2+\theta_1^4+\theta_1^6+\dots+\theta_1^{2T}}{1+\theta_1^2+\theta_1^4+\dots+\theta_1^{2(T-1)}} \end{bmatrix} \tag{3.34}$$

The matrix A is a lower triangular matrix with $1's$ along the diagonals, hence

$|\mathbf{A}| = 1$ and:

$$|\mathbf{ADA}'| = |\mathbf{A}|\,|\mathbf{D}|\,|\mathbf{A}'| = |\mathbf{D}|\,. \tag{3.35}$$

Let $\tilde{Z} = A^{-1}Z$. Then the likelihood can be written as:

$$f_{\tilde{\mathbf{Z}}}(\tilde{\mathbf{Z}};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{T/2}\,|\mathbf{D}|^{1/2}} \exp\left(-\frac{1}{2}\tilde{\mathbf{Z}}'\mathbf{D}^{-1}\tilde{\mathbf{Z}}\right) \tag{3.36}$$

Since $A\tilde{Z} = Z$ the first equation of this system can be written as $\tilde{Z}_1 = Z_1$ while the t^{th} row implies that:

$$\tilde{Z}_t = Z_t + \frac{\theta_1\left[1+\theta_1^2+\theta_1^4+\cdots+\theta_1^{2(t-2)}\right]}{1+\theta_1^2+\theta_1^4+\cdots+\theta_1^{2(t-1)}}\tilde{Z}_{t-1}. \tag{3.37}$$

The vector $\tilde{Z}$ can thus be computed by iterating on equation 3.37 for $t = 2, 3, \ldots, T$ starting with $\tilde{Z}_1 = Z_1$. Note that:

$$|\mathbf{D}| = \prod_{t=1}^{T} d_{tt} \quad \text{with } d_{tt} = \sigma^2\frac{1+\theta_1^2+\theta_1^4+\cdots+\theta_1^{2t}}{1+\theta_1^2+\theta_1^4+\cdots+\theta_1^{2(t-1)}} \tag{3.38}$$

since D is a diagonal matrix and

$$\tilde{\mathbf{Z}}'\mathbf{D}^{-1}\tilde{\mathbf{Z}} = \sum_{t=1}^{T}\frac{\tilde{Z}_t^2}{d_{tt}}. \tag{3.39}$$

The likelihood is then equal to:

$$f_{\tilde{\mathbf{Z}}}(\tilde{\mathbf{Z}};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{T/2}\left(\prod_{t=1}^{T} d_{tt}\right)^{1/2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\frac{\tilde{Z}_t^2}{d_{tt}}\right) \tag{3.40}$$

with log likelihood equal to:

$$l(\boldsymbol{\theta}) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log d_{tt} - \frac{1}{2}\sum_{t=1}^{T}\frac{\tilde{Z}_t^2}{d_{tt}}. \tag{3.41}$$

Given reasonable numerical values for $\theta_0' = \begin{pmatrix} \theta_{1,0} & \sigma_0^2 \end{pmatrix}'$ the sequence $\tilde{Z}_t$ is calculated by iterating on equation 3.37 starting with $\tilde{Z}_1 = Z_1$ with d_{tt} given by equation 3.38.

Figure 4 below displays the estimation results for the fitting of a MA(1) to the ISCOR example by ML. The conditional maximum likelihood method produces the fitted model $Z_t = \varepsilon_t + 0.337\varepsilon_{t-1}$ with $\hat{\sigma}^2 = 4.675$. The standard error of the estimate is 0.039. The t statistic of 8.567 indicates that the parameter should be included in the model. The Ljung-Box statistics (see section 2.3.4) of the residuals give $Q^*(2) = 1.0491$ with p value 0.306. and $Q^*(10) = 9.1362$ with p value 0.425. These results indicate that the fitted MA(1) model is adequate since there does not appear to be any

serial correlation between the estimated residuals of the model. The Durbin Watson statistic of 1.976 is close to 2 confirms that the MA(1) residuals have no first-order autocorrelation. The exact maximum likelihood method produces the fitted model $Z_t = \varepsilon_t + 0.3369\varepsilon_{t-1}$ with $\hat{\sigma}^2 = 4.669$. Thus indicating that the two optimization techniques lead to similar results.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	**0.337**	0.039	8.567	0.000
R-squared	0.108		Mean dependent var	-0.118
Adjusted R-squared	0.108		S.D. dependent var	4.949
S.E. of regression	4.675		Akaike info criterion	5.924
Sum squared resid	12,500.730		Schwarz criterion	5.932
Log likelihood	-1,696.233		Durbin-Watson stat	1.976
Inverted MA Roots	-0.340			

Figure 4: The MA(1) estimation results for ISCOR weekly returns (1990-2000)

Forecasting using MA models

For the one step ahead forecast of an MA(1) process, the model states that $Z_t = \varepsilon_t - \theta\varepsilon_{t-1}$. Taking conditional expectations, we have that $\tilde{Z}_{t+1} = E(Z_{t+1}|Z_t, Z_{t-1}, ..., Z_1) = -\theta\varepsilon_t$ since $\varepsilon_{t+1}|Z_t \sim iid(0, \sigma^2)$. The one step ahead forecast error is thus $\tilde{e}(1) = Z_{t+1} - \tilde{Z}_{t+1} = \varepsilon_{t+1}$ and the variance of $\tilde{e}(1)$, $(var(\tilde{e}(1)))$ is equal to σ^2.

The two step ahead forecast can be found similarly and is equal to *zero* which implies that $\tilde{e}(2) = \varepsilon_{t+2} - \theta\varepsilon_{t+1}$ such that $var(\tilde{e}(2)) = (1 + \theta^2)\sigma^2$.

Similarly for a MA(q) model, we have: $Z_{t+l} = \varepsilon_{t+l} - \theta_1\varepsilon_{t-1+l} - \theta_2\varepsilon_{t-2+l} - ..... - \theta_q\varepsilon_{t-q+l}$ from which we can obtain the l step ahead forecasts:

$$\begin{aligned}
\tilde{Z}_{t+1} &= -\theta_1\varepsilon_t - \theta_2\varepsilon_{t-1} - ..... - \theta_q\varepsilon_{t-q+1} \\
\tilde{Z}_{t+2} &= -\theta_2\varepsilon_t - \theta_3\varepsilon_{t-1} - ..... - \theta_q\varepsilon_{t-q+2} \\
\tilde{Z}_{t+l} &= -\sum_{j=l}^{q}\theta_j\varepsilon_{t-j+l} \qquad \text{for } 1 \leq l \leq q
\end{aligned} \tag{3.42}$$

and thus the variance of the forecast error is:

$$var\left(\tilde{e}\left(l\right)\right) = \begin{cases} \sigma^2 & for \quad l = 1 \\ \sigma^2 \left(1 + \sum_{j=1}^{l-1} \theta_j^2\right) & for \quad 2 \leq l \leq q \\ \sigma^2 \left(1 + \sum_{j=1}^{q} \theta_j^2\right) & for \quad l > q \end{cases}$$

From equation 3.42 it can be seen that the l step ahead forecasts of an MA(q) is equal to the mean of the series for forecasts greater than q steps. Similarly equation 3.43 indicates that the forecasts' error variance increases as l increases so that it is equal to the variance of the series for forecasts greater than q steps. In short we are forecasting so far ahead (1 step) that the remote history of the MA(q) no longer has any effect.

3.2.3 Autoregressive Processes (AR Processes)

Before we investigate the properties of autoregressive processes we first examine the correlogram (a plot of the sample autocorrelation function against time) of the weekly log returns of METCASH over the period 6 January 1990 to 30 December 2000. Figure 7 (page 64) below displays this correlogram. It can be seen that some of the sample autocorrelations are significant at the 5% level (since the two bands indicate the two standard deviation cut off values.) We have seen that one may identify moving average processes by examining the sample autocorrelations of a series. In this section we attempt to model a series based on the significant lagged partial autocorrelations.

First-Order Autoregressive Processes

A time series $\{Z_t\}$ is said to be a first order autoregressive model (also called the Markov Model) if it is a weighted linear sum of the previous value of Z_t and a random error term, ε_t. Thus $\{Z_t\}$ can be defined as:

$$Z_t = \phi Z_{t-1} + \varepsilon_t \tag{3.44}$$

Notice that the process could also be represented as $\varepsilon_t = Z_t - \phi Z_{t-1} = (1 - \phi B) Z_t = \phi(B) Z_t$ where B is the backshift operator. By recursive substitution Z_t can be written as follows:

$$\begin{aligned}
Z_t &= \phi(\phi Z_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2 Z_{t-2} \\
&= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2(\phi Z_{t-3} + \varepsilon_{t-2}) \\
&= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3 Z_{t-3} \\
&= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \phi^4\varepsilon_{t-4} + \ldots\ldots \\
&= \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}
\end{aligned} \tag{3.45}$$

Equation 3.45 is an infinite moving average process provided that $|\phi| < 1$. The mean of the process is *zero* and its variance is equal to $\sigma^2 \sum_{i=0}^{\infty} \phi^{2i} = \frac{\sigma^2}{1-\phi^2}$. The autocovariance structure of Z_t can be calculated as follows:

$$cov(Z_t, Z_{t+k}) = E\left(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}, \sum_{j=0}^{\infty} \phi^j \varepsilon_{t+k-j}\right) \qquad \text{for } k \geq 0$$

$$\begin{aligned} cov\left(Z_t, Z_{t+k}\right) &= E\left(\sum_{i=0}^{\infty}\phi^i\varepsilon_{t-i}\left[\phi^k\sum_{j=0}^{\infty}\phi^j\varepsilon_{t-j}\right]\right) \\ &= \phi^k\sigma^2\sum_{i=0}^{\infty}\phi^{2i} \\ &= \frac{\phi^k\sigma^2}{1-\phi^2} \end{aligned} \tag{3.46}$$

Note that in the derivation of the autocovariance structure that the second line follows because the product of the two summations will only have non-zero covariance terms when $j = k$. Equation 3.45 and equation 3.46 imply that $corr\left(Z_t, Z_{t+k}\right) = \phi^k$ for $k \geq 0$. The autocorrelation function is thus an exponentially decreasing function of k. If $0 < \phi < 1$, the autocorrelation function is non-negative for all k. However, if $-1 < \phi < 0$, then the autocorrelation function oscillates from being negative to positive for increasing $k > 0$.

If one assumes that Z_t is weakly stationary of order two (Constant mean, constant autocovariance matrix over time.) it can be shown that:

$$\gamma_k = \phi\gamma_{k-1} \quad \text{for } k = 1, 2, 3, ... \text{ where } \gamma_0 = \frac{\sigma^2}{1-\phi^2} \tag{3.47.a}$$

$$\rho_k = \phi\rho_{k-1} \quad \text{for } k = 1, 2, 3, ... \text{ where } \rho_0 = 1 \tag{3.47.b}$$

Equation 3.47.*b* is a form of the Yule-Walker equations. It is used extensively when undertaking parameter estimation.

Second-Order Autoregressive Processes

In this section we investigate the properties of a autoregressive process of order two. Such a process is defined as $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \varepsilon_t$ where we assume that ε_t is independent of $Z_{t-1}, Z_{t-2}, ...$ Once again the model can be expressed by using the backshift operator: $(1 - \phi_1 B - \phi_2 B^2)Z_t = \phi\left(B\right) Z_t = \varepsilon_t$. Assuming that Z_t is weakly stationary of order two, the expected value of the process is equal to *zero* for all t.

The autocovariance and autocorrelation structure can also be derived if one assumes that Z_t is weakly stationarity of order two.

$$\gamma_k = E\left(Z_t Z_{t-k}\right) = \phi_1 E\left(Z_{t-1}Z_{t-k}\right) + \phi_2 E\left(Z_{t-2}Z_{t-k}\right) \quad \text{for } k \geq 1 \tag{3.48.a}$$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} \tag{3.48.b}$$

The variance of Z_t can be represented in terms of ϕ_1, ϕ_2 and σ^2 as $\gamma_0 = \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + \sigma^2 + 2\phi_1\phi_2\gamma_1$. Now from equation 3.48.*b*, $\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$ which implies that:

$$\gamma_0 = \left(\frac{1-\phi_2}{1+\phi_2}\right)\frac{\sigma^2}{(1-\phi_2)^2 - \phi_1^2} \tag{3.49}$$

Dividing both sides of equation 3.48.*b* by γ_0 implies that:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \quad \text{for } k \geq 1 \tag{3.50}$$

Equations 3.48.*b* and 3.50 are called the Yule-Walker equations. They are used to derive the autocorrelation function. Setting $k = 1$, $\rho_0 = 1$ and $\rho_{-k} = \rho_k$ for all k we obtain $\rho_1 = \phi_1 + \phi_2\rho_1 \Rightarrow \rho_1 = \frac{\phi_1}{1-\phi_2}$. Similarly $\rho_2 = \phi_1\rho_1 + \phi_2 = \frac{\phi_2(1-\phi_2+\phi_1^2)}{1-\phi_2}$.

Equation 3.50 states that the autocorrelation function of a weakly stationary series of order two satisfies the second order difference equation $(1 - \phi_1 B - \phi_2 B^2)\rho_k = 0$. Thus $(1 - \phi_1 B - \phi_2 B^2) = 0$. Now dividing both sides by B^2 and setting $x = \frac{1}{B}$ we have $x^2 - \phi_1 x - \phi_2 = 0$. This equation is known as the characteristic equation. It's roots are called the characteristic roots and they are defined as $x_i = \frac{\phi_1 \pm \sqrt{\phi_1^2+4\phi_2}}{2}$. Theorem one (see later) states that in general an AR(p) process is weakly stationary of order two if and only if the characteristic roots lie inside the unit circle on the complex plane. This property can be used to show that an AR(2) model is weakly stationary of order two if and only if $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ and $|\phi_2| < 1$.

pth-Order Autoregressive Process

The AR(p) model is $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t$ indicating that the process at time t is a linear combination of its own p previous time points, and some random error. $\{Z_t\}$ can also be represented as $\phi(B)Z_t = \varepsilon_t$ where $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$. Z_t is weakly stationary of order two provided that the roots of $x^p - \phi_1 x^{p-1} - \phi_2 x^{p-2} - ...\phi_p = 0$ lies inside the unit circle. The constant mean of the process $E(Z_t)$ is equal to *zero*. Notice however that if one were to include a non-zero constant term as the intercept in the formulation of Z_t, i.e. $Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t$ then $E(Z_t) = \frac{c}{1-\phi_1-\phi_2-.....-\phi_p}$ provided that $\sum \phi_i \neq 1$ and that Z_t is weakly stationary of order two. The autocovariance function can then be defined as:

$$\gamma_k = \begin{cases} \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + ... + \phi_p\gamma_{k-p} & for \quad k > 0 \\ \phi_1\gamma_1 + \phi_2\gamma_2 + ... + \phi_p\gamma_p + \sigma^2 & for \quad k = 0 \end{cases} \tag{3.51}$$

and the Yule-Walker equations are:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + ... + \phi_p\rho_{k-p} \quad for \quad k > 0 \tag{3.52}$$

Identifying the order of a AR model

Our aim is to utilise the data in order to choose a suitable value for p. This choice is made after investigating the sample partial autocorrelation function of a series. Multiple regression can be used in order to explain what the partial autocorrelation function (PACF) is intended to measure. Consider the following set of AR models, each estimated under a least squares criterion and in the time order provided below.

$$\begin{aligned} Z_t &= \phi_{1,0} + \underline{\phi_{1,1}} Z_{t-1} + \varepsilon_{1,t} \\ Z_t &= \phi_{2,0} + \phi_{2,1} Z_{t-1} + \underline{\phi_{2,2}} Z_{t-2} + \varepsilon_{2,t} \\ Z_t &= \phi_{3,0} + \phi_{3,1} Z_{t-1} + \phi_{3,2} Z_{t-2} + \underline{\phi_{3,3}} Z_{t-3} + \varepsilon_{3,t} \end{aligned}$$

where $\phi_{k,0}$, $\phi_{k,j}$ and $\varepsilon_{k,t}$ are the constant term, the coefficient of Z_{t-j}, and the error of the regression equation . $\phi_{1,1}$ of the first equation is called the lag one PACF of Z_t. Similarly $\phi_{j,j}$ is called the lag j PACF of Z_t. $\phi_{2,2}$ measures the lag two correlation after removing the correlation of the first lag. Similarly the partial autocorrelation at lag k is the regression coefficient of Z_{t-k} when Z_t is regressed on a constant, $Z_{t-1}, ..., Z_{t-k}$. This is a partial correlation since it measures the correlation of Z_t values that are k periods apart after removing the correlation from the intervening lags.

By definition $\phi_{kk} = corr\,(Z_t, Z_{t-k}|Z_{t-1}, Z_{t-2}, ...Z_{t-k+1})$ for normally distributed $\{Z_t\}$. If $\{Z_t\}$ is non-normal then $\phi_{kk} = corr\left(Z_t - \Sigma_{j=1}^{k-1}\beta_j Z_{t-j}, Z_{t-k} - \Sigma_{j=1}^{k-1}\beta_j Z_{t-k+j}\right)$ where β_j is chosen so as to minimise the mean square error of prediction. The mean square error of prediction is defined as $\Sigma_{t=T-m+1}^{T}\frac{Z_t-\tilde{Z}_t}{m}$ where $\tilde{Z}_t$ is the one step ahead forecast (or best linear predictor of Z_t) and m indicates the number of forecasts undertaken by the analyst.

For an AR(p) model with $k > p$, $\tilde{Z}_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p}$ and thus the covariance between $Z_t - \tilde{Z}_t$ and $Z_{t-k} - \tilde{Z}_{t-k}$ is equal to zero since $cov\left(\varepsilon_t, Z_{t-k} - \tilde{Z}_{t-k}\right) = 0$. Hence the partial autocorrelation function (PACF) of a pure autoregressive process of order p cuts off at lag p such that $\phi_{kk} = 0$ for $k > p$.

The autocorrelation function of an AR(p) process can be represented by the Yule-

Walker equations as follows:

$$\rho_j = \phi_{p1}\rho_{j-1} + \phi_{p2}\rho_{j-2} + \ldots + \phi_{pp}\rho_{j-p} \quad \text{for} \quad j = 1, 2, \ldots, p \tag{3.53}$$

where ϕ_{pj}, is the j^{th} coefficient in an autoregressive process of order p, such that ϕ_{pp} is the last coefficient in equation 3.53. Equation 3.53 can be expressed by the following system of equations:

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{p1} \\ \phi_{p2} \\ \vdots \\ \phi_{pp} \end{pmatrix} \tag{3.54}$$

or in matrix notation as $\underline{\mathbf{P}}_p\underline{\boldsymbol{\phi}}_p = \underline{\boldsymbol{\rho}}_p$. These equations can easily be solved for $p = 1, 2, 3\ldots$.

As an example, consider $p = 2$. From equation 3.53, $\rho_1 = \phi_{21} + \phi_{22}\rho_1 \Rightarrow \phi_{21} = (1 - \phi_{22})\rho_1$. Similarly $\rho_2 = \phi_{21}\rho_1 + \phi_{22} = (1 - \phi_{22})\rho_1^2 + \phi_{22} \Rightarrow \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$. Notice that ϕ_{22} is the partial autocorrelation of lag 2. For an autoregressive process of order p, the partial autocorrelation coefficient will be non-negative for $k \leq p$ where k indicates the order of the partial autocorrelation coefficient. In general, the k^{th} partial autocorrelation coefficient can be recursively calculated by using:

$$\phi_{kk} = \begin{cases} \rho_1 & for \quad k = 1 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j} & for \quad k > 1 \end{cases} \tag{3.55}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$ for $j = 1, 2, ..., p-1$ and ρ_k is the lag k autocorrelation. It can be shown that the asymptotic variance of ϕ_{kk} is $\frac{1}{T}$ for $k > p$. If the absolute value of the partial autocorrelation coefficient is greater than $\frac{2}{\sqrt{T}}$, then the partial autocorrelation is said to be significant at the 5% level.

Parameter Estimation

Once an experimenter has identified that a particular sample $(z_1, \ldots . z_T)$ was generated by an $AR(p)$ process, regression analysis could be used in order to estimate the parameters $(c, \phi_1, \ldots, \phi_p)$. Notice that in this formulation a constant term c, is being entertained. Note also that the explanatory variables $(Z_{t-1,\ldots}, Z_{t-p})$ are stochastic and are not fixed as assumed in the normal ordinary least squares (OLS) model. The

explanatory variables will be independent of the error structure if ε_t is a white noise- or Gaussian- process thus allowing one to use OLS estimation.

Define $\hat{\beta}_p' = \begin{pmatrix} \hat{c} & \hat{\phi}_1 & \hat{\phi}_2 & \dots & \dots & \hat{\phi}_p \end{pmatrix}'$ as the matrix containing the estimated parameters. Notice that in a sample of $(z_1, \dots, z_T)$ only $T-p$ of the observations are used as response variable values in estimating $\hat{\beta}_p$, because the first p observations enter the estimation process only as explanatory variable values. The system of equations to be solved is:

$$\begin{aligned} z_{p+1} &= c + \phi_1 z_p + \dots + \phi_p z_1 + \varepsilon_{p+1} \\ z_{p+2} &= c + \phi_1 z_{p+1} + \dots + \phi_p z_2 + \varepsilon_{p+2} \\ &\vdots \\ z_T &= c + \phi_1 z_{T-1} + \dots + \phi_p z_{T-p} + \varepsilon_T \end{aligned}$$

which can be rewritten into matrix notation such that $Z_p = X_p \Phi_p + e$ where:

$$Z_p = \begin{pmatrix} z_{p+1} \\ z_{p+2} \\ \dots \\ z_T \end{pmatrix}, \quad e = \begin{pmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \dots \\ \varepsilon_T \end{pmatrix}, \quad \text{and} \quad X_p = \begin{pmatrix} 1 & z_p & z_{p-1} & \cdots & z_1 \\ 1 & z_{p+1} & z_p & \cdots & z_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{T-1} & z_{T-2} & \cdots & z_{t-p} \end{pmatrix}$$

Note that the column of $1's$ in X_p is required in order to estimate the constant term. Since Z_t is an AR process it can be expressed by iterative substitution as an MA(∞) which is a linear combination of the error structure. Assuming that ε_t is normally distributed, Z_t itself would be normally distributed. Then OLS may be used as a criterion to estimate $\hat{\beta}_p$. The estimate would be $\hat{\beta}_p = (X_p' X_p)^{-1} X_p' Z_p$. If $\hat{\beta}_p$ is a Maximum Likelihood estimate $\hat{\beta}_p$ tends in distribution to a $N(\beta_p, \sigma^2 (X_p' X_p)^{-1})$. The variance-covariance matrix can be consistently estimated by $\hat{\Sigma} = \hat{\sigma}^2 (X_p' X_p)^{-1}$ where $\hat{\sigma}^2 = \frac{(Z_p - X_p \hat{\beta}_p)'(Z_p - X_p \hat{\beta}_p)}{T-(2p+1)}$ is the OLS estimate of the variance σ^2 of the white noise process ε_t.

The degrees of freedom used is $(T-p) - (p+1)$ rather than $T - (p+1)$ since there are only $T-p$ complete observations (response plus explanatory variables) with which to estimate the $(p+1)$ parameters.

Example

A simulation generated 1000 observations from an AR(2) which had the following formulation $Z_t = 0.2 + 0.8Z_{t-1} + 0.1Z_{t-2} + \varepsilon_t$ with starting values $\sigma_\varepsilon^2 = 1$, $Z_1 = 5$ and $Z_2 = 6$. The estimation procedure was applied. The results are presented below. Figure 5 is a graphical display of the sample data.

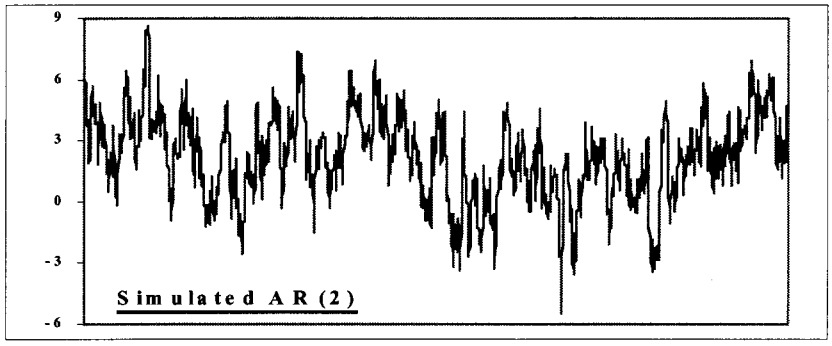


Figure 5: AR(2) Simulated Series

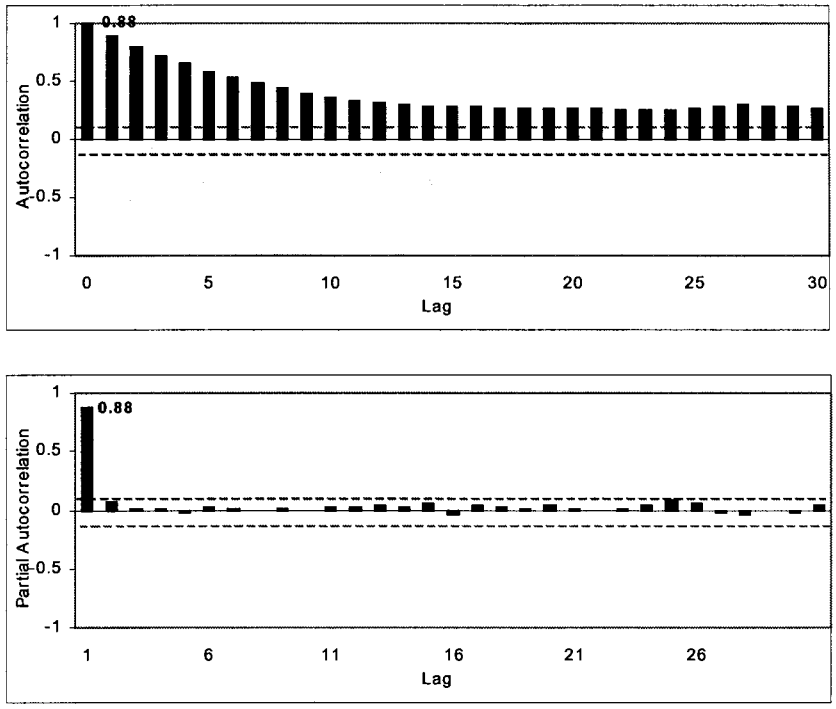


Figure 6: The sample ACF and the sample PACF of the simulated series

Figure 6 above displays the sample autocorrelation function as well as the sample partial correlation coefficient function. It can be seen that all of the lagged sample autocorrelation coefficients are significant at the 5% level and only the first and the second sample partial autocorrelation is significant at the 5% level. The sample partial autocorrelation function indicates that one should tentatively use a AR(2) in order to

model the simulated series. The estimation results for the AR(2) model is displayed below.

Table 1: AR(2) Regression Results	
Correlation Coefficient	0.8850
R Squared Statistic	0.7833
Adjusted R Squared Statistic	0.7828
Standard Error of Regression	0.9931
Number of Observations	998

Table 2: AR(2) Coefficients and Standard Errors

Coefficients	Estimates	Standard Error	t Statistic
c	0.2345	0.0458	5.1207
Y_{t-1}	0.8157	0.0316	25.8195
Y_{t-2}	0.0768	0.0316	2.4329

Table 3: AR(2) Statistica Results

Coefficients	Estimates	Standard Error	t (997)	p-value	Lower 95% Conf Lev	Upper 95% Conf Lev
c	2.4361	0.2938	8.2912	0.000	1.8595	3.0127
Y_{t-1}	0.8203	0.0316	25.9304	0.000	0.7582	0.8823
Y_{t-2}	0.0742	0.0317	2.3413	0.0194	0.0120	0.01364
Mean Square Error of Residuals = 0.99384						

From table 1 it can be seen that the Adjusted Coefficient of determination is 0.7828. The estimated coefficients are listed in table 2. The estimated coefficients are close to the actual parameters. All three parameters have t values greater than the 95% cut-off values thus indicating that the coefficients are not non-zero. The estimate of the constant is equivalent to $\hat{c} = \hat{\mu}(1 - \hat{\theta}_1 - \hat{\theta}_2)$ where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t$. Note however that the standard error each coefficient are relatively large. The 95% confidence intervals for the coefficients are as follows:

$$\hat{c} \in [0.1446 , 0.3243]$$

$$\hat{\theta}_1 \in [0.7537 , 0.8777]$$

$$\hat{\theta}_2 \in [0.0149 , 0.1388]$$

at lag 2 and lag 6. The large Q-statistic indicates that the residual series is not a white noise process in that they still seem to be serially correlated.

If we now change the model and we fit $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + \phi_5 Z_{t-5} + \phi_6 Z_{t-6} + \varepsilon_t$, then we have $Q^*(6) = 2.964$ with a p-value of 0.085, $Q^*(10) = 6.779$ with a p-value of 0.238. The Q-statistic indicates that the residual series is a possible white noise process, and hence the model is adequate. The final model is $Z_t = 0.291 Z_{t-1} - 0.168 Z_{t-2} + 0.205 Z_{t-3} + 0.225 Z_{t-5} - 0.140 Z_{t-6} + \hat{\varepsilon}_t$ where $\hat{\sigma}^2 = 5.023$. The standard errors of the estimated parameters are 0.041, 0.042, 0.042, 0.0044 and 0.045 respectively, indicating that all of the parameters are significant. Z_{t-4} is not included in the final model.

If a pure AR(6) was fitted to the METCASH series, all of the coefficients excluding the Z_{t-4} term would be significant at the 5% level. The Z_{t-4} term has a p-value of 0.077 and would thus be significant at the 10% significance level.

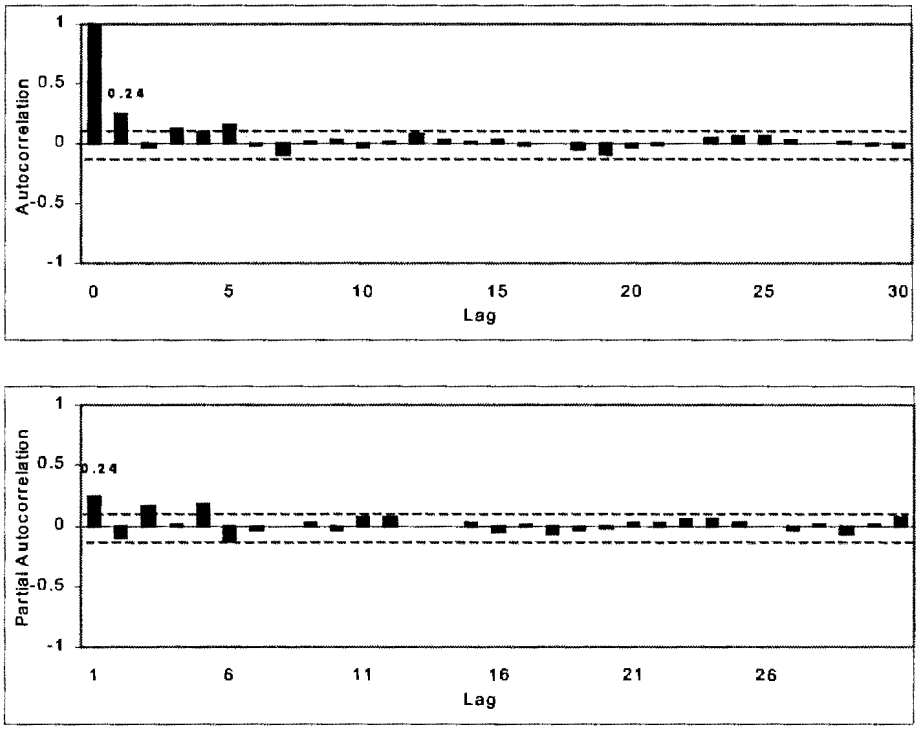


Figure 7: The sample ACF and the sample PACF of METCASH

Forecasting

Forecasts are undertaken so as to minimise the squared error loss function i.e. $\tilde{Z}_{t+l}$ is chosen so as to minimise $E\left(Z_{t+l} - \tilde{Z}_{t+l}\right)^2$ where $l > 0$. This procedure is equivalent to taking the conditional expectation of Z_{t+l} such that $\tilde{Z}_{t+1} = E\left(Z_{t+1} | Z_t, Z_{t-1}, ..., Z_1\right) =$

$\sum_{i=0}^{p-1} \phi_i Z_{t-i}$. The associated error of the forecast is equal to $\tilde{e}(1) = \varepsilon_{t+1}$ which has a forecast variance of σ^2.

Iterative substitution can be used in order to calculate the results associated with a 2 step ahead forecast.

$$\begin{aligned} Z_{t+2} &= \phi_1 Z_{t+1} + \phi_2 Z_t + ... + \phi_p Z_{t+2-p} + \varepsilon_{t+2} \text{ which implies that} \\ \tilde{Z}_{t+2} &= E(Z_{t+2}|Z_t, Z_{t-1}, Z_{t-2}, ..., Z_1) \\ &= \phi_1 \tilde{Z}_{t+1} + \phi_2 Z_t + ... + \phi_p Z_{t+2-p} \end{aligned}$$

Thus $\tilde{e}(2) = \phi_1 \left(Z_{t+1} - \tilde{Z}_{t+1}\right) + \varepsilon_{t+2} = \phi_1 \tilde{e}(1) + \varepsilon_{t+2} = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}$ such that the 2 step ahead forecast variance is equal to $\sigma^2 \left(1 + \phi_1^2\right)$.

Extending the methodology to a AR(p) process implies that $\tilde{Z}_{t+l} = \sum_{i=0}^{p} \phi_i Z_{t+l-i}$. The l-step ahead forecast error is $\tilde{e}(l) = Z_{t+l} - \tilde{Z}_{t+l}$ which can recursively estimated by using the formula for $\tilde{Z}_{t+l}$. Note that it can be shown that for a stationary AR(p) process that $\tilde{Z}_{t+l} \rightarrow E(Z_t)$ as $l \rightarrow \infty$. This property is known as mean reversion and means that for a very large time series that the conditional means tends towards the unconditional mean. Similarly the variance of the forecast error tends towards the unconditional variance of the series.

Stationarity and Invertibility

Theorem 1 *If $x_1, x_2, ..., x_p$ denotes the roots of the p^{th} order equation $\phi(\frac{1}{x}) = 0$, then the AR(p) model $\phi(B) X_t = \varepsilon_t$ can be written as a convergent series in $\varepsilon_t, \varepsilon_{t-1,...}$ such that:*

$$X_t = (\phi(B))^{-1} \varepsilon_t = \Psi(B)\varepsilon_t = \sum_{r=0}^{\infty} \psi_r \varepsilon_{t-r}$$

only if $|x_i| < 1$ for all roots $x_i, i = 1, 2, ..., p$ where $\Psi(B) = 1 + \psi_1 B + ... + \psi_c B^c$ for a constant c tending to ∞.

We now have $(\phi(B))^{-1} = \Psi(B) = \sum_{r=0}^{\infty} \psi_r B^r$ so that:

$$X_t = \Psi(B) \varepsilon_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2+.....} \tag{3.57}$$

If equation 3.57 holds it is said that X_t is invertible and the AR(q) process will only be stationary if the roots of the characteristic equation, $\phi(\frac{1}{x}) = 0$ lies within the unit circle.

3.2.4 Autoregressive Moving Average Processes

The ARMA(p,q) model is defined as $\Phi(B) Z_t = \phi_0 + \Theta(B)\varepsilon_t$ where $\Phi(B) = 1 - \phi_1 B - ... - \phi_p B$, $\Theta(B) = 1 - \theta_1 B - ... - \theta_q B$ and ϕ_0 is a intercept term. This model can be rewritten as:

$$Z_t - \phi_1 Z_{t-1} - ... - \phi_p Z_{t-p} = \phi_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q} \tag{3.58}$$

Remarks

1. The model will be AR-invertible and stationary if the roots of $\Phi\left(\frac{1}{x}\right) = 0$ all lie in the unit circle. A necessary condition for the stationarity of the process is that $\phi_1 + \phi_2 + ... + \phi_p < 1$.
2. The model will be MA-invertible if the roots of $\Theta\left(\frac{1}{x}\right) = 0$ all lie in the unit circle.
3. In general the autocovariance and autocorrelation structure of a ARMA(p,q) process is not easily solved by inspection (Pindyck and Rubinfeld (1997)) however it can be shown that for $k \geq q+1$ that:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + ... + \phi_p \gamma_{k-p} \tag{3.59.a}$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + ... + \phi_p \rho_{k-p} \tag{3.59.b}$$

In the following section we briefly examine the properties of an ARMA(1,1) process.

Properties of ARMA(1,1) Models

A time series Z_t follows an ARMA(1,1) model if it satisfies $Z_t - \phi Z_{t-1} = \phi_0 + \varepsilon_t - \theta \varepsilon_{t-1}$ where $\{\varepsilon_t\}$ is a white noise series. The expected value of Z_t is equal to $\frac{\phi_0}{1-\phi}$ if Z_t is weakly stationary of order two (i.e. if $|\phi| < 1$), $\phi_0 \neq 0$ and $\phi \neq 1$.

For simplicity lets assume that $\phi_0 = 0$ and that the process is weakly stationary of order two. The variance of Z_t can then be derived as follows:

$$\begin{aligned} Var(Z_t) &= E\left(Z_t^2\right) \\ &= E\left(\left[\phi Z_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}\right]^2\right) \\ &= \phi^2 var\left(Z_t\right) + \sigma^2 + \theta^2 \sigma^2 - 2\phi\theta E\left(Z_{t-1}\varepsilon_{t-1}\right) + 2\phi E\left(Z_{t-1}\varepsilon_t\right) - 2\theta E\left(\varepsilon_t \varepsilon_{t-1}\right) \end{aligned}$$

$$\begin{aligned} Var(Z_t) &= \phi^2 var(Z_t) + \sigma^2 + \theta^2\sigma^2 - 2\phi\theta E(Z_{t-1}\varepsilon_{t-1}) \\ &= \phi^2 var(Z_t) + \sigma^2 + \theta^2\sigma^2 - 2\phi\theta\sigma^2 \end{aligned}$$

since $E(\varepsilon_t\varepsilon_{t-1}) = 0$, $E(Z_{t-1}\varepsilon_t) = 0$ and $E(Z_t\varepsilon_t) = \sigma^2$ for all t. The variance of a ARMA(1,1) process is thus equal to:

$$\gamma_0 = \frac{(1+\theta^2-2\phi\theta)}{1-\phi^2}\sigma^2 \text{ such that } |\phi| < 1 \tag{3.60}$$

Note that $E(Z_{t-1}\varepsilon_t) = 0$ for all t. We can now determine the covariances $\gamma_1, \gamma_2, ...$ recursively such that:

$$\begin{aligned} \gamma_1 &= E(Z_{t-1}[\phi Z_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}]) = \phi\gamma_0 - \theta\sigma^2 \\ &= \frac{\phi(1+\theta^2-2\phi\theta) - \theta(1-\phi^2)}{1-\phi^2}\sigma^2 \\ &= \frac{(1-\phi\theta)(\phi-\theta)}{1-\phi^2}\sigma^2 \qquad (3.61.c) \\ \gamma_2 &= E(Z_{t-2}[\phi Z_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}]) = \phi\gamma_1 \qquad (3.61.d) \end{aligned}$$

Similarly $\gamma_k = \phi\gamma_{k-1}$ for $k \geq 2$. The autocorrelation function is then equal to:

$$\begin{aligned} \rho_1 &= \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta} \qquad (3.62.a) \\ \rho_k &= \phi\rho_{k-1} \quad for\ k \geq 2 \qquad (3.62.b) \end{aligned}$$

Notice that the autocorrelation function of the ARMA(1,1) is very similar to that of a AR(1) model. As k increases the ACF tends towards zero. In general the ACF will exhibit exponential decay and would oscillate between positive and negative values if ϕ is negative.

Identifying ARMA processes

In general it is difficult to identify the order of p and q in one simple step. The ACF and PAC does not provide much insight into the order of p and q, however they could be used in order to tentatively assign AR and MA components of the ARMA process. Model building is an iterative process. One should tentatively assign AR and MA components. The residuals of the initial model should then be examined (i.e. investigate the sample ACF, sample PACF, Bartlett test statistics and Portmanteau test statistics of the residual series in order to test for the existence of serial autocorrelation.) in order to identify additional AR and MA components that could be included in the model. If additional components are required, refit the model and repeat the above process until the residual series of your model does not exhibit any

serial autocorrelation.

Forecasting using ARMA Models

In this section we investigate the properties of ARMA forecasts. The one step ahead forecast of Z_{t+1} is equal to:

$$\tilde{Z}_{t+1} = E\left(Z_{t+1}|Z_t, Z_{t-1}, Z_{t-2}, ...\right) = \phi_0 + \sum_{i=1}^{p} \phi_i Z_{t+1-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t+1-i} \tag{3.63}$$

The associated error of the forecast is equal to $\tilde{e}\left(1\right) = Z_{t+1} - \tilde{Z}_{t+1} = \varepsilon_{t+1}$ which has a forecast variance of σ^2.

The above procedure can be used in order to calculate the results associated with a 2 step ahead forecast.

$$\begin{aligned} Z_{t+2} &= \phi_0 + \sum_{i=1}^{p} \phi_i Z_{t+2-i} + \varepsilon_{t+2} - \sum_{i=1}^{q} \theta_i \varepsilon_{t+2-i} \text{ which implies that} \\ \tilde{Z}_{t+2} &= E\left(Z_{t+2}|Z_t, Z_{t-1}, Z_{t-2}, ...\right) \\ \tilde{Z}_{t+2} &= \phi_0 + \phi_1 \tilde{Z}_{t+1} + \sum_{i=2}^{p} \phi_i Z_{t+2-i} - \sum_{i=2}^{q} \theta_i \varepsilon_{t+2-i} \end{aligned} \tag{3.64.c}$$

Thus $\tilde{e}\left(2\right) = \phi_1 \left(Z_{t+1} - \tilde{Z}_{t+1}\right) + \varepsilon_{t+2} - \phi_1 \varepsilon_{t+1} = \phi_1 \tilde{e}\left(1\right) + \varepsilon_{t+2} - \theta_1 \varepsilon_{t+1} = \left(\phi_1 - \theta_1\right) \varepsilon_{t+1} + \varepsilon_{t+2}$ such that the forecast variance is equal to $\sigma^2 \left(1 + \left[\phi_1 - \theta_1\right]^2\right)$.

Note that it can be shown that for a stationary ARMA(p,q) process that $\tilde{Z}_{t+l} \rightarrow E\left(Z_t\right)$ as $l \rightarrow \infty$ thus $\tilde{Z}_{t+l}$ moves towards the unconditional mean as the forecast length is increased. Similarly the variance of the forecast error will tend towards the unconditional variance of the series. (as for the AR(p) process.)

3.3 Autoregressive Integrated Moving Average Processes (ARIMA Processes)

In practice, many time series are non-stationary implying that one requires special models in order to handle such time series. In this section we attempt to model non-stationary time series by using a suitable difference operator and a ARMA(p,q) specification.

Z_t is said to be a homogenous non-stationary series of order d if $W_t = \nabla^d Z_t$ is a stationary series. This means that if Z_t is a non-stationary series, differencing Z_t, d times results in a stationary series, W_t. If W_t is an ARMA(p,q) process, then the model is denoted as ARIMA(p,d,q). The process can thus be represented as:

$$\Phi(B) W_t = \Phi(B) \nabla^d Z_t = \phi_0 + \Theta(B) \varepsilon_t \tag{3.63}$$

If W_t only has an AR component then the process is known as an integrated autoregressive process of order (p, d) and is denoted as ARI(p,d,0). Similarly if W_t does not have a MA component, then the process is known as an integrated moving average process of order (d, q) and is denoted as IMA(0,d,q).

Identification of ARIMA Models

Before one can specify the order of p and q in an ARIMA(p,d,q) model, an appropriate value of d has to be chosen. The value of d is chosen as follows:

1. First examine the autocorrelation function of the original series Z_t and identify whether or not it is stationary. Note that the autocorrelation function of a stationary series approaches zero as the lag k increases. This property as well as the Bartlett and Portmanteau test statistics could be used in order to ascertain whether or not a series is stationary.

2. Assuming that the original series is non-stationary, difference the series and examine the autocorrelation function of ∇Z_t.

3. Check to see whether or not ∇Z_t is stationary. If ∇Z_t is stationary, attempt to model ∇Z_t by means of an ARMA(p,q) process. If not, difference ∇Z_t until it is stationary and then attempt to model $\nabla^d Z_t$ by means of an ARMA(p,q) process.

Note that the choice of d is not that simple to ascertain in many real life examples since overdifferncing can introduce unnecessary correlations into the model which could complicate the model. For example, suppose our observed series (Z_t) is a random walk so that $W_t = Z_t - Z_{t-1} = \varepsilon_t$. If we difference once more we obtain $\nabla W_t = \varepsilon_t - \varepsilon_{t-1}$ which is an MA(1) with $\theta = 1$. With an observed series we would unnecessarily have to estimate θ.

3.4 The Box-Jenkins Model Building Procedure

Figure 8 below provides a brief summary of the Box Jenkins model building procedure.

Box-Jenkins Model Building Procedure for ARIMA Models

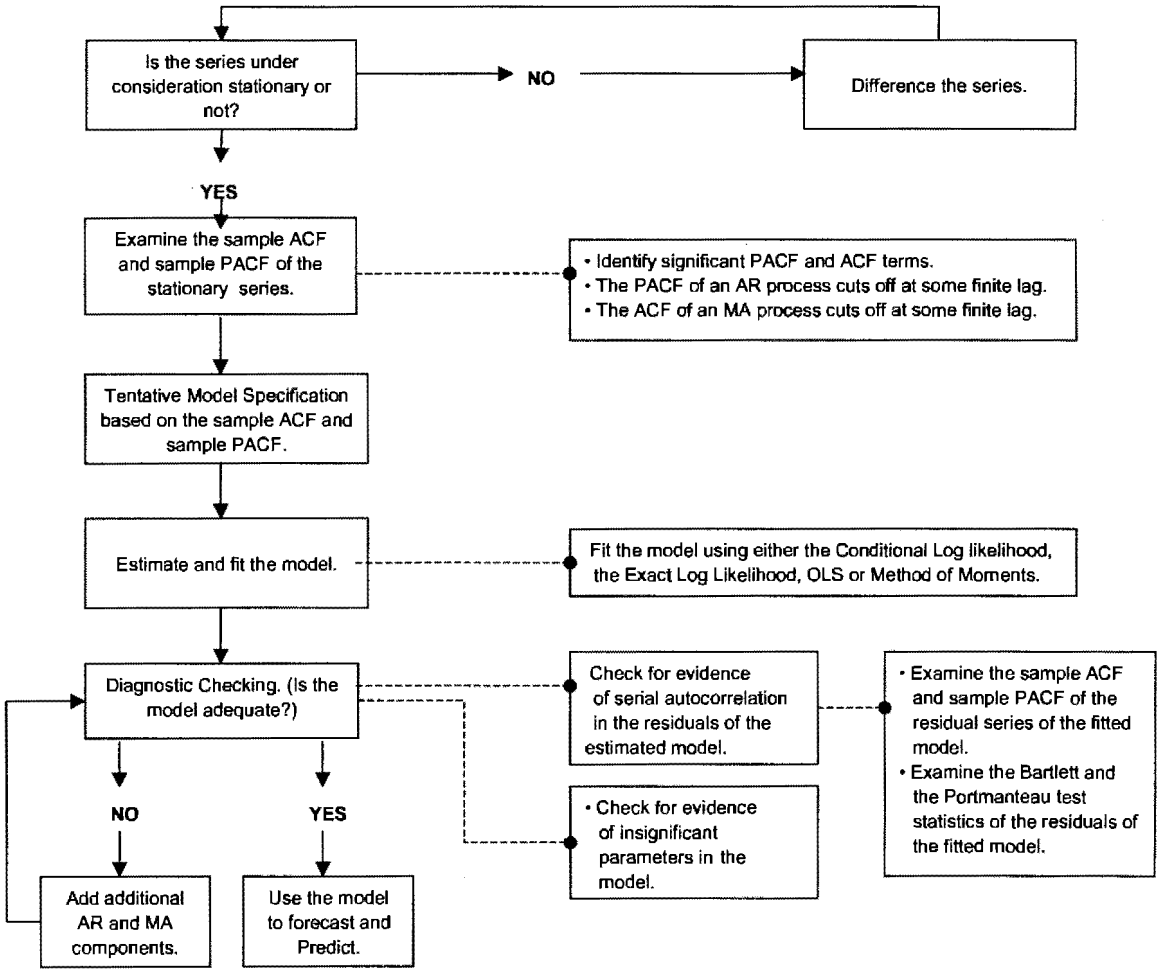


Figure 8: The Box Jenkins Model Building Procedure

Part II

Volatility Modelling

Chapter 4

Conditional Heteroskedastic Models

4.1 Introduction

Risk as measured by variance plays a major role in many financial applications. The Capital Asset Pricing Model (CAPM) or Market Model (Sharpe (1964), Litner (1965), Mossin(1966) and Merton(1973)) for example posits a positive relationship between the expected returns of a share and the market portfolio. This hypothesis relies on the assumption of a constant market variance. Portfolio managers as well as traders thus require an estimate of the level of future market uncertainty (variance) if they believe in the validity of the Capital Asset Pricing Model (CAPM). Volatility Models, first developed by Engle (1982), model the conditional volatility (conditional variance) of time series.

Volatility models have become increasingly important since asset return variances as well as the covariances of asset returns were found not to be constant over time (Bollerslev (1986)) but rather that they evolve over time. Many researchers have used volatility models in order to model share returns, exchange rate movements as well as interest rate movements. Both univariate time series models as well as their multivariate extensions are important to portfolio managers. For example, since covariances are time varying one could use a multivariate volatility model in order to solve the asset allocation problem of Markowitz (1952). Volatility models have also been used in order to test certain economic theories. Bollerslev *et al.* (1988) used a trivariate CAPM in order to test the CAPM and the assumption of constant covariances and variances amongst the three assets used (treasury bills, bonds and

stocks). They found that the conditional covariances between the three assets were varied over time. It was also shown that this movement could be forecasted.

Currently volatility models are most often used by derivatives analysts since they rely on estimates of the variance (vols) of certain financial instruments in order to use these estimates as inputs into their asset pricing models. Banks and large corporates however also use volatility models in order to estimate their daily value at risk (VAR). From a statistical point of view volatility models are important since many commonly used tools are no longer valid in the presence of non-constant variances. For example, standard regression type models assume that the residuals from the regression model is homoskedastic (constant over time). If this assumption is violated the analyst should adjust any results from such a model in order to compensate for heteroskedastic errors. ARCH in itself does not invalidate standard OLS inference. However, ignoring ARCH effects may result in the loss of efficiency of the estimated beta coefficients.

From the above discussion it can be seen that volatility models are important to both the financial practitioner as well as to the practicing statistician. In what follows is a discussion on both the theory and the use of volatility models in the financial markets. Most of the references pertain to the American stock market however some South African examples will also be supplied.

4.1.1 The Characteristics of Volatility

As mentioned previously variance has an important role to play in financial applications. Note however, that the variance of an assets returns it is not directly observable and thus one has to use statistical models in order to estimate daily, monthly as well as annual volatility. There are however a few characteristics associated with volatility that is commonly seen in asset returns. They are as follows:

1. There exists volatility clusters. i.e. volatility may be high for certain periods and low during other periods. Mandelbrot (1963) was the first to document this and stated that "... *large changes tend to be followed by large changes-of either sign- and small changes by small changes...*".

2. Volatility evolves over time in a continuous manner (Bollerslev (1986)). This can be seen by investigating the implied volatility of option pricing models. The Black Scholes (Black and Scholes (1973)) model can be used in order to

determine the price of options and futures. The model requires a number of estimates of which annual standard deviation is one of them. Implied volatility is the value of the annual standard deviation that solves the Black Scholes formula when the standard deviation is the only unknown variable in the model. It can be used as an estimate of what the "*market*" believes the volatility of a share is.

3. Volatility behaves asymmetrically to different types of news. Black (1976) states that volatility tends to rise in response to bad news and similarly fall due to good news.

4.1.2 Sources of ARCH and variables used to model Volatility

Volatility models were established in order to model the time varying nature of the conditional variance of time series. A number of researchers have tried to explain why conditional variances should be serially correlated. Diebold and Nerlove (1989) believe that share returns are heteroskedastic due to the existence of a serially correlated news process. Gallant *et al.* (1989) offer evidence in favour of the above hypothesis however Engle *et al.* (1990a,b) were unable to provide any satisfactory explanation for the dependence in the underlying news arrival process.

Lamoureux and Lastrapes (1990b) argue that heteroskedasticity found in share returns are due to the clustering in trading volumes. When they included trading volumes into their variance equation they found that the lagged residuals were not significant thus substantiating their claim. Karpoff (1987) however show that trading volumes and the price of a share are highly correlated. This could be the reason why Lamoureux and Lastrapes (1990b) found a significant loading on the trading volume in their variance equation.

In the quest to model the conditional variance of a shares return series, researchers have unearthed numerous economic variables that are related to the conditional variance of share returns. Gallant *et al.et al.* (1990) found that lagged volume were positively related to the conditional volatility on the NYSE where as Campbell (1987) and Glosten *et al.* (1991) found that nominal interest rates and volatility were related. Attanasia (1991) and Attanasia and Wadhwani (1989) show that dividend yields were

a significant determinant of volatility where as Engel and Rodrigues (1989) found a positive relationship between M1 money supply, the oil price and conditional volatility.

4.2 The ARCH model

4.2.1 Model Specification

The ARCH model was proposed by Engle (1982) in order to model the variance of inflation in the United Kingdom. ARCH(1) models the **conditional variance** at time period t (denoted h_t) of the return of a share at time period t (denoted r_t) by using the square of the first lagged return. Define r_t as the return of a share during time period t and R_t as the **mean adjusted return** of a share during time period t. The ARCH(1) model can then be represented as:

$$R_t = \sqrt{h_t}\varepsilon_t \quad \text{where} \quad h_t = Var(R_t|I_{t-1}) = \alpha_0 + \alpha_1 R_{t-1}^2 \tag{4.1}$$

where ε_t is assumed to be a **white noise process with variance 1** and I_{t-1} is the information set available at time $t-1$. Often ε_t is modelled as a **standard normal** random variate or as a **standardised Students t distribution**. In order to ensure that $h_t \geq 0$ it is assumed that $\alpha_0 >$ and $\alpha_1 \geq 0$. The above process can be generalised to specify an ARCH(p) model as follows:

$$R_t = \sqrt{h_t}\varepsilon_t \quad \text{where} \quad h_t = \alpha_0 + \alpha_1 R_{t-1}^2 + \alpha_2 R_{t-2}^2 + ... + \alpha_p R_{t-p}^2 \tag{4.2}$$

This particular specification of heteroskedasticity was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals (Bollerslev (1986)). From equation 4.1 and equation 4.2 it can be seen that the model implies that large prior returns will generate large conditional variances at time t since the conditional variance at time t is dependent on the previous p lagged returns of a share. Notice also that the ARCH methodology treats both positive and negative returns the same way since the square of lagged returns are used in order to model the conditional variance structure of the returns of a share.

4.2.2 Properties of an ARCH(1) Model

The **conditional** mean of R_t is $E(R_t|I_{t-1}) = 0$ for all t since ε_t and h_t are to be independent. Assuming that h_t and ε_t are independent and noting that $var(\varepsilon_t) = 1$,

the conditional variance is

$$var(R_t|I_{t-1}) = E\left(\left(\sqrt{h_t}\varepsilon_t\right)^2|I_{t-1}\right) = E\left(h_t\varepsilon_t^2|I_{t-1}\right) \quad (4.3.a)$$

$$= E(h_t|I_{t-1}) = E(\alpha_0 + \alpha_1 R_{t-1}^2) \quad (4.3.b)$$

for all t. Assuming that both the **unconditional** mean and the variance are time invariant, $\{R_t\}$ can be viewed as being **weakly stationary**. Thus $var(R_t) = var(R_{t-1})$ $= E(R_{t-1}^2)$ since $E(R_{t-1}) = 0$ implying that the **unconditional** variance is equal to $var(R_t) = \frac{\alpha_0}{1-\alpha_1}$. From this point onwards all expectations and variances are **conditional** expectations and **conditional** variances unless stated otherwise. Conditioning on the prior information will be dropped in all equations in order to improve notation.

Assuming that ε_t is normally distributed it can be shown that the **unconditional kurtosis** is greater than 3, indicating that the ARCH(1) model has heavier tails than the normal distribution. One of the limitations of ARCH type models is the restrictions one has on the parameters α_i for all i. This can be seen by examining $E(R_t^4)$. Assuming that the process $\{R_t\}$ is fourth order stationary $E(R_t^4) = m_4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$. Since $var(R_t) \geq 0$ and $\alpha_0 > 0 \Rightarrow 0 \leq \alpha_1 \leq 1$. Since $m_4 \geq 0 \Rightarrow 0 \leq \alpha_1 < \sqrt{\frac{1}{3}}$

4.2.3 Testing for ARCH Errors

ARCH LM Test

A Lagrange multiplier (LM) test for the presence of ARCH effects in the residuals of a time series is available (Engle 1982). To test the null hypothesis that there is no ARCH up to order p in the residuals of a time series model, run the regression:

$$e_t^2 = \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 e_{t-2}^2 + ... + \beta_{t-p} e_{t-p}^2 \quad (4.4)$$

where $\{e_t\}$ is the residual series from a time series model (e.g. a ARMA model or a regression and time series model). This is a regression of the squared residuals on a constant and the lagged squared residuals up to order p. The Lagrange Multiplier test statistic, nR^2 is asymptotically distributed χ_p^2 where n is the number of observations in the time series and R^2 is the coefficient of determination of the fitted time series model. The null hypothesis of no ARCH effects up to order p is rejected if $nR^2 > \chi_p^{2\,(\alpha)}$. A complete proof is provided in Gourieroux (1997).

4.2.4 Order Determination

Define $\varepsilon_t^* = R_t^2 - \sigma_t^2$ where σ_t^2 is the true conditional variance at time t. ε_t^* is uncorrelated with mean 0. Now substituting $E(h_t)$ as a estimate of σ_t^2 into ε_t^* we have $R_t^2 = \alpha_0 + \alpha_1 R_{t-1}^2 + \alpha_2 R_{t-2}^2 + ... + \alpha_p R_{t-p}^2 + \varepsilon_t^*$. From this it can be seen that the ARCH(p) specification can be regarded as an $AR(p)$ process for R_t^2. Note however that $\{\varepsilon_t^*\}$ is not a *iid* sequence. Box Jenkins methods as well as the Extended Sample Autocorrelation Function (ESACF) methodology (Tsay and Tiao (1984)) can be applied in order to tentatively determine the order of the ARCH process.

4.2.5 Estimation

We can estimate the parameters of an ARCH(p) model by using the conditional log likelihood function. In general we assume that ε_t is either a standard normal variate or that it has a standardised Students t distribution. If ε_t is normally distributed then the conditional log likelihood is:

$$\log\left\{\prod_{t=p+1}^{T}\frac{1}{\sqrt{2\pi h_t}}e^{\left(-\frac{R_t^2}{2h_t}\right)}\right\} = \sum_{t=p+1}^{T}\left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(h_t) - \frac{1}{2}\frac{R_t^2}{h_t}\right\} \tag{4.5}$$

where $h_t = \alpha_0 + \alpha_1 R_{t-1}^2 + \alpha_2 R_{t-2}^2 + ... + \alpha_p R_{t-p}^2$. Equation 4.5 should be evaluated iteratively for each observation in order to maximise the conditional log likelihood function. It can be seen that $R_t \sim N(0, h_t)$.

Fat tailed error distributions can be modelled by using a standardised t distribution. This can be done by transforming the error term to $\varepsilon_t = \frac{X}{\sqrt{\frac{v}{v-2}}}$ for $v > 2$ where X has a students t distribution with v degrees of freedom. It can be shown that the pdf of ε_t is:

$$f_{\varepsilon_t}(\varepsilon_t|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi}}\left(1 + \frac{\varepsilon_t^2}{v-2}\right)^{-\frac{(v+1)}{2}} \qquad -\infty < \varepsilon_t < \infty \tag{4.6}$$

and thus the pdf of $R_t = \sqrt{h_t}\varepsilon_t$ is:

$$f_{R_t}(R_t|v) = \frac{\Gamma\left(\frac{(v+1)}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi h_t}}\left(1 + \frac{R_t^2}{(v-2)h_t}\right)^{-\frac{(v+1)}{2}} \qquad -\infty < R_t < \infty \tag{4.7}$$

The conditional log likelihood function used when undertaking parameter estimation is thus equal to $\sum_{t=p+1}^{T}\left\{\log\left(\frac{\Gamma\left(\frac{(v+1)}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi}}\right) - \frac{v+1}{2}\log\left(1 + \frac{R_t^2}{(v-2)h_t}\right) - \frac{1}{2}\log(h_t)\right\}$.

4.2.6 Model Checking

The ARCH model assumes that the residual term, ε_t, is a either normally distributed or follows a standardised Student t distribution. This assumption should be tested once a possible model has been chosen. If a particular ARCH model is appropriate to model the conditional variance of a time series then $\hat{\varepsilon}_t = \frac{R_t}{\sqrt{\hat{h}_t}}$ should be either a standard normal random variable or follow a Standardised Student t distribution depending on the initial assumption of the specification of the distribution of ε_t. Portmanteau test statistics can be used in order to test the adequacy of the mean equation of the model and the sample PACF can be used to check the adequacy of the variance equation. Plotting the distribution of the estimated residuals by means of a histogram would also be useful in checking the validity of the distribution assumption. χ^2 goodness of fit statistics can be calculated in order to formally test whether or not the estimated residual series follows assumed distributional form. QQ plots could also used in order to test the error distribution assumption.

4.2.7 Forecasting Volatility

Assuming that an ARCH(p) model accurately captures the conditional volatility of a time series then it can be used to forecast the future conditional volatility of a time series. A one step ahead forecast of time $t+1$ would be:

$$\tilde{\sigma}^2_{t+1} = \alpha_0 + \alpha_1 R^2_t + \alpha_2 R^2_{t-1} + ... + \alpha_p R^2_{t-p+1} \tag{4.8}$$

Similarly a two step ahead forecast is:

$$\begin{aligned} \tilde{\sigma}^2_{t+2} &= \alpha_0 + \alpha_1 \tilde{R}^2_{t+1} + \alpha_2 R^2_t + ... + \alpha_p R^2_{t+2-p} \\ &= \alpha_0 + \alpha_1 \tilde{\sigma}^2_{t+1} + \alpha_2 R^2_t + ... + \alpha_p R^2_{t+2-p} \end{aligned} \tag{4.9}$$

The above process can be continued such that the l step ahead forecast is $\tilde{\sigma}^2_{t+l} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \tilde{\sigma}^2_{t+l-i}$ where $\tilde{\sigma}^2_{t+l-i} = R^2_{t+l-i}$ for $i \geq l$. Recall that the ARCH(p) model can be viewed as a AR(p) model for the square of the return (i.e. R^2_t) at time t and thus as $l \to \infty$, the forecast volatility tends towards the unconditional volatility of the time series, $\frac{\alpha_0}{1-\alpha_1-\alpha_2-.....-\alpha_p}$.

Example: The weekly log returns of the All Share Index***

Figure 1 below displays the weekly log returns of the JSE (***calculated as the

log difference of the average of the highest and the lowest traded price of the JSE for the week) over the period 4 January 1990 to 30 December 2000. It can be seen that there is significant volatility clustering as well as a non-constant variance in the time series. The returns during the period 1990 to early 1997 fluctuates between 5 and -5% per week. The Asian contagion (Started in September 1998) brought about the collapse of Asian markets and the devaluation of the Russian Roubel. World markets returns became increasingly more volatile. Returns on the JSE followed suit since international investors classified South Africa as an Emerging market and started investing in larger markets such as the USA and London and Germany.

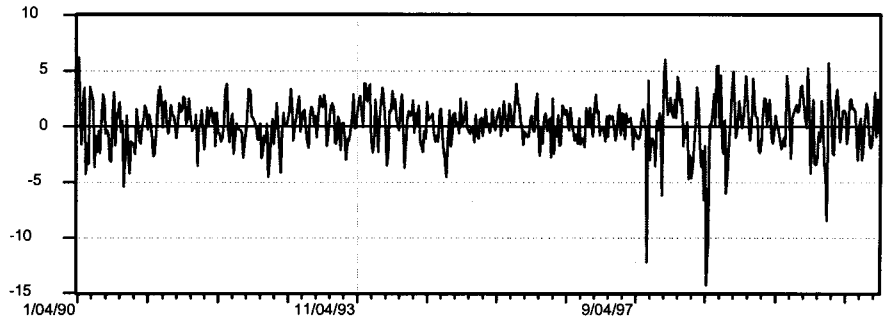


*Figure 1: The weekly log returns of the JSE**

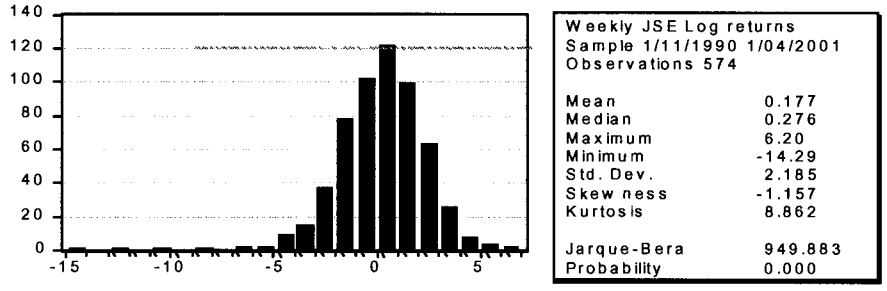


Figure 2: Histogram and Summary Statistics

Figure 2 above displays the histogram and the summary statistics of the weekly returns of the JSE*. It can be seen that the distribution is slightly **skewed** to the left with a mean 0.177% per week (This would provide a annual return of 9.267%.). Note however that the kurtosis is 8.862 which is about three times the level of the kurtosis of a normal distribution. The Jarque Bera statistic of 949.883 exceeds the cut off value of $\chi_2^{2(5\%)}$ and thus the hypothesis of normally distributed weekly returns is rejected at the 5% level. Figure 2 above indicates that the returns distribution has heavier tails than that of a normal distribution. The normal distribution does not

adequately capture the distribution of the weekly JSE return series. Although for this example lets assume that the assumption is adequate.

Figure 3 below displays the sample ACF and PACF for the log return series. The first and the seventh lagged sample autocorrelation coefficient is significant at the 5% level since $0.34 > \left|\frac{2}{\sqrt{574}}\right|$ and $0.118 > \left|\frac{2}{\sqrt{574}}\right|$ respectively (using Bartlett's tests). Similarly the, first and the seventh lagged partial autocorrelation coefficient is significant at the 5% level since $0.34 > \left|\frac{2}{\sqrt{574}}\right|$ and $0.119 > \left|\frac{2}{\sqrt{574}}\right|$ respectively. Note also that $Q^*(1) = 21.393 > \chi_1^{2(5\%)} = 3.841$.

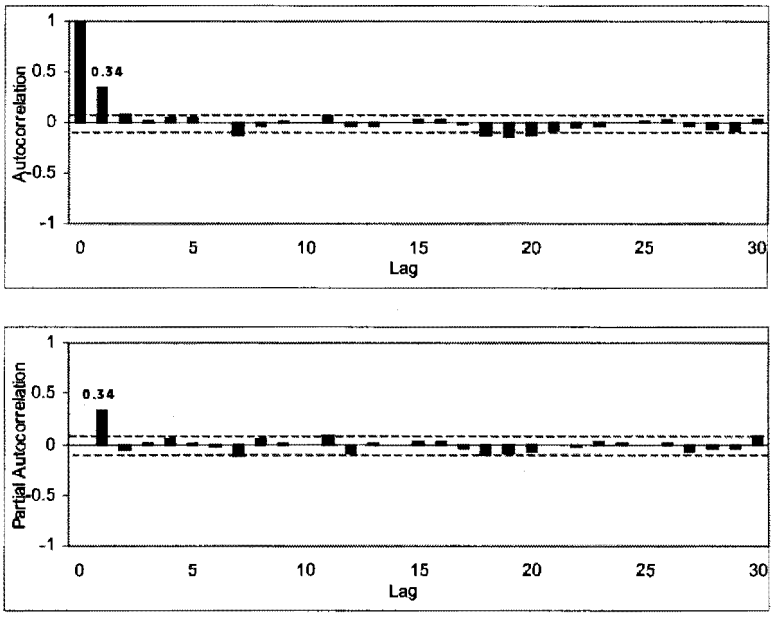


Figure 3: The sample PACF and ACF of JSEW

The above results suggests that we could potentially fit five different models as the mean equation of the time series. They are as follows: a pure AR(1), a pure MA(1), a ARMA(1,1) a moving average with only MA(1) and MA(7) terms and an autoregressive model with only AR(1) and AR(7) terms. Table 1 below displays the Akaike information criterion value (AIC) and the $Q^*(m)$ statistics for each of the models. It can be seen that the AR(1) AR(7) minimises the AIC, suggesting that it is the most appropriate model for the mean equation of the log return series. Notice however that all of the models do not show any significant series autocorrelation when tested over the first fifteen lags of the residual series.

Table 1: AIC and $Q^*(m)$ Statistics for the initial model				
Model	AIC	$Q^*(5)$	$Q^*(10)$	$Q^*(15)$
AR(1) AR(7)	**4.245**	3.133	5.466	15.252
MA(1) MA(7)	4.274	7.465	10.033	17.840
MA(1)	4.284	6.190	16.545	23.838
ARMA(1,1)	4.269	2.678	12.889	22.108
AR(1)	4.268	4.772	15.307	24.260
$Q^*(m)$ 5% cut off value		**11.1**	**18.3**	**25.0**

The mean equation fitted to the weekly log returns is $r_t = 0.3560r_{t-1} - 0.1118r_{t-7} + e_t$. e_t is assumed to be normally distributed. Notice that the intercept term was found to be insignificant in all of the models. The standard errors of parameters are 0.0390 and 0.0386 respectively suggesting that both parameters in the mean equation are significant at the 5% level. The *adjusted* R^2 *statistic* is equal to 0.133 indicating that the ARIMA model is only able to explain a very small proportion of the variation of the return series. This is to be expected since ARIMA models in general do not fit asset returns very well. One should thus use them in conjunction with an econometric model in order to improve the fit of the model.

Table 2:ACF, PACF and $Q^*(m)$ statistics for the AR(1) AR(7) squared residuals										
	Lags									
	1	2	3	4	5	6	7	8	9	10
ACF	0.254	0.140	-0.004	0.021	0.017	-0.008	0.021	-0.008	0.018	0.097
PACF	0.254	0.081	-0.061	0.027	0.016	-0.023	0.028	-0.016	0.017	0.102
Q*(m)	36.683	47.859	47.867	48.131	48.291	48.330	48.587	48.624	48.804	54.285
Q*(m) cutoff	3.84	5.99	7.81	9.49	11.1	12.6	14.1	15.5	16.9	18.3
Bartletts cutoff 0.0835										
5% Cut off values used										

Table 1 above displays the $Q^*(m)$ statistics and their cut off values at lag 5, 10, and 15. The AR(1) AR(7) is a adequate representation of the mean equation of the log return series since the residual series does not show any significant serial autocorrelation when tested over the first fifteen lags. Table 2 above displays the

sample ACF, PACF and $Q^*(m)$ statistics for the AR(1) AR(7) squared residuals. It can be seen that the sample PACF indicates that a ARCH(2) model might be appropriate to capture the conditional variance of the log return series eventhough the lag 2 PACF is slightly insignificant. The ARCH LM test statistic is equal to 36.426 $\left(> \chi_2^{2\,(0.05)} = 5.99\right)$. This further emphasises that the ARCH(2) model might be appropriate when testing at the 5% level. The **joint** estimation of both the conditional mean and conditional variance equations results in the model

$$\begin{aligned} r_t &= 0.3844 r_{t-1} - 0.1109 r_{t-7} + \sqrt{h_t}\varepsilon_t \\ h_t &= 2.8244 + 0.1235 R_{t-1}^2 + 0.1525 R_{t-2}^2 \end{aligned}$$

where the standard error of the parameters in the mean equation is equal to 0.0476 and 0.0336 respectively and the standard error of the parameters in the variance equation is equal to 0.1104, 0.0460 and 0.0234 respectively. All of the parameters are significant at the 5% level. The residual series of the joint model indicates that the model does adequately capture the conditional mean and variance equation since $Q^*(10) = 8.859$ with a p-value of 0.354 for the residual series. $Q^*(10) = 2.048$ with a p-value of 0.980 for the square of the residual series. The Q^*-statistics of the residual series and the square of the residual series indicates that the residuals of joint model is a white noise process.

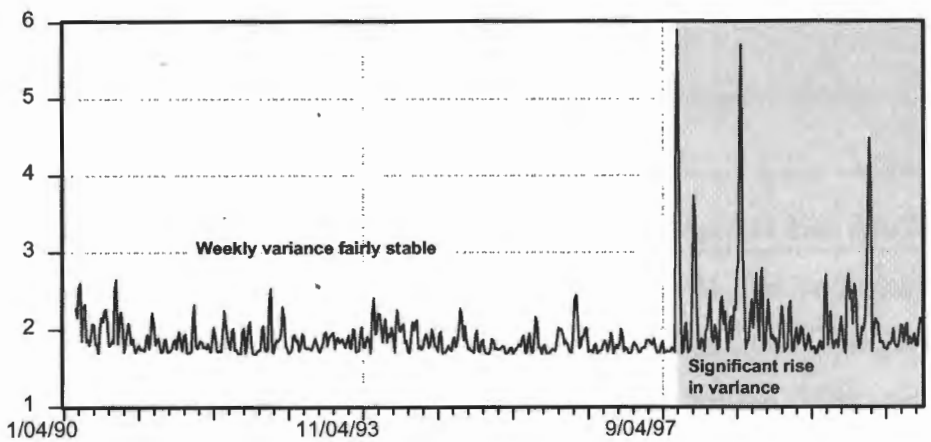


Figure 4: The fitted conditional variance for JSEW

Figure 4 above displays the fitted weekly variance over the period under investigation. It can be seen that the ARCH specification does capture the significant increase in the variability of returns post September 1998. Note however that the distribution of the fitted residuals is not normally distributed since the standardized

residuals have a kurtosis of 8.5073, a skewness of -1.0122 and a Jarque-Bera statistic of 813.389. The following section discusses two different specifications of the error distribution.

4.2.8 Different Specifications of the error distribution

A number of distributions have been suggested in order to model the error process $\{\varepsilon_t\}$ of a volatility model. Bollerslev (1987) suggests using a standardized Student t distribution with v degrees of freedom, where v has to be estimated from the data. Other parametric densities have also been considered. Jorion (1988) suggested using a normal-Poisson mixture distribution, Baillie and Bollerslev (1989a) proposed using the power exponential distribution, Hsieh (1989a) used the normal-lognormal mixture distribution, Nelson (1990a) used a generalised exponential distribution and McCulloch (1985) used an infinite variance leptokurtic stable Paretion distribution in order to model the distribution of the error process.

Semiparametric density estimation techniques have been proposed by Gallan and Nychka (1987), Gallant and Tauchen (1989), Gallant *et al.* (1989), Gallant *et al.* (1990) and Engle and Gonzalez-Rivera (1991). In general the semiparametric approach leads to a loss in asymptotic efficiency when one compares the estimation approach to maximum likelihood estimation with a correctly specified density function for the error process of the volatility model.

ARCH Models with a t distribution

Bollerslev (1987) suggests modelling the error process $\{\varepsilon_t\}$ of the ARCH model as a standardized Student t distribution. It was found that this specification for the innovations better captures the observed kurtosis in returns data. (Ballie and Bollerslev (1989), Hsieh (1989a), and Palm and Vlaar (1997)). Milhoj (1985) and Bollerslev (1986) however state that the specification remains inadequate for many financial time series since the observed time series have fatter tails than the Students t distribution.

Example: The weekly log returns of the All Share Index***

In the following example the weekly log returns of the JSE*** (over the period 4 January 1990 to 30 December 2000) will be modelled by using a ARMA ARCH

model. The first step in the modelling process is to fit a suitable ARMA model to the weekly log returns of the JSE series. This has already been done in the previous example. The second step is to model the residual series by means of a suitable ARCH process. It is assumed that the error process of the ARCH model follows a **standardised Students t distribution** where the degrees of freedom v has to be estimated from the data.

If a ARCH(1) model is entertained then the maximum likelihood estimation results leads to the model:

$$R_t = \sqrt{h_t}\varepsilon \qquad h_t = 3.0333 + 0.1930R_{t-1}^2 \qquad \text{with} \qquad v = 5.7872$$

where $\{R_t\}$ is the residuals series after fitting an ARMA model as the mean equation of r_t. The standard errors of the parameters are 0.2944, 0.0761 and 0.9736 respectively. All of the estimates are significant at the 5% level. The unconditional variance of R_t is 3.7587. The Ljung-Box statistic of the standardised residuals of this model is $Q^*(10) = 7.1335$ with a p-value of 0.713, indicating that the mean equation is adequate. The Ljung-Box statistic of the squared standardised residuals of this model is $Q^*(10) = 9.2833$ with a p-value of 0.505, suggesting that the variance equation is also adequate. Note however that $Q^*(2) = 5.630$ (for the square of the standardised residuals) which is marginally significant at the 5% level. The p-value is equal to 0.044.

The estimation results for a ARCH(2) is displayed in table 3 below. It can be seen that all of the parameters are significant at the 5% level except for the R_{t-2}^2 coefficient. If one compare these two models with the ARCH model under the normality assumptions it can be seen that the use of a heavy tailed distribution reduces the ARCH effect.

Table 3: Estimation Results of the ARCH(2)

	Coefficient	Std Error	t Statistic	p-value
α_0	2.7766	0.3133	8.8608	0.0000
α_1	0.1751	0.0739	2.3707	0.0178
α_2	**0.0796**	**0.0464**	**1.7170**	**0.0860**
v	6.2233	1.2013	5.1806	0.0000

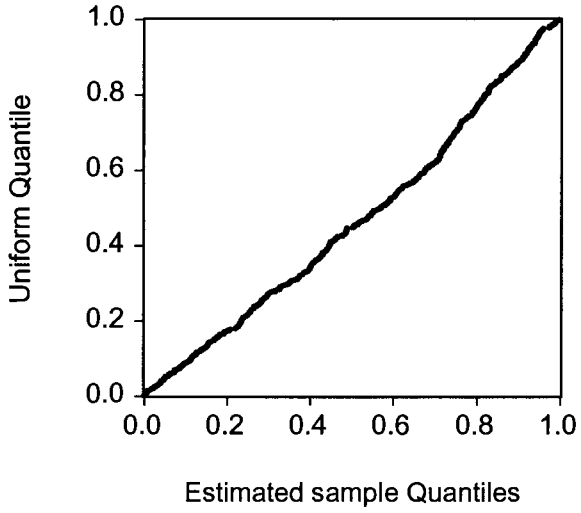


Figure 4: QQ plot

The above results suggests that the ARCH(1) model with standardised Student t distribution residuals adequately describes the conditional variance of the weekly returns of the JSE*. An examination of the probability plot/QQ plot in figure 4 indicates that the standardised Students t distribution assumption is an appropriate one.

ARCH Models with a Skewed Standardised t distribution

Share returns are often skewed. Symmetric distributions such as the normal and Students t distribution do not adequately capture this property of share returns. Fernández and Steel (1998) proposed using a skewed Students t distribution to model the distribution of share returns. Lambert and Laurent (2001) used this idea in order to model the error distribution of a volatility model as a standardised skewed Students t distribution.

A Skewed density function of a symmetric unimodal density function $g(u)$ (where U is continuous) can be constructed by transforming U to $\epsilon = X\xi|U| - (1-X)\frac{1}{\xi}|U|$ where X is a Bernoulli random variable, with probability of success $\frac{\xi^2}{1+\xi^2}$. We can show that the unconditional density $f(\epsilon|\xi)$ of ϵ is equal to:

$$\begin{aligned} f(\epsilon|\xi) &= \Pr(X=0)g(\epsilon|\xi, x=0) + \Pr(X=1)g(\epsilon|\xi, x=1) \\ &= \frac{2}{\xi+\frac{1}{\xi}}\left[g(\epsilon\xi)I_{(-\infty,0)}(\epsilon) + g\left(\frac{\epsilon}{\xi}\right)I_{[0,\infty)}(\epsilon)\right] \end{aligned} \tag{4.10}$$

where $I(\epsilon)$ is an indicator variable such that $I_{(a,b)}(\epsilon) = 1$ if $a < \epsilon < b$. Fernández and

Steel (1998) show that if the r^{th} moment of $g(.)$ exists then the moments of $f(\epsilon|\xi)$ are defined and is equal to:

$$E(\epsilon^r|\xi) = M_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}} \quad \text{where} \quad M_r = \int_0^\infty 2s^r g(s)\, ds \tag{4.11}$$

ϵ does not have a zero mean or a variance of 1. Lambert and Laurent (2001) proposed standardising ϵ so that it could be used as the error distribution of a of a volatility model (Lambert and Laurent (2001) used a GARCH(1,1) to model the conditional volatility. GARCH models will be covered in the following section.) If z_t is the standardised random variable then the density function of z_t is:

$$\left(\frac{2s}{\xi + \frac{1}{\xi}}\right)\left[g\left[\xi(sz+m)\right] I_{(-\infty,0)}\left(z + \frac{m}{s}\right) + g\left[\frac{(sz+m)}{\xi}\right] I_{[0,\infty)}\left(z + \frac{m}{s}\right)\right] \tag{4.12}$$

where $z_t = \frac{\epsilon_t - m}{s}$, $m = E(\epsilon_t|\xi)$ and $s^2 = Var(\epsilon_t|\xi)$. If U is assumed to be a standardised Students t distribution then the contribution of the t^{th} observation to the loglikelihood function is equal to:

$$\ln\left(\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(v-2)\pi}}\right) + \ln\left(\frac{2\xi s}{1+\xi^2}\right) - \frac{1}{2}\left\{\ln\left(\sigma_t^2\right) + (1+v)\left(1 + \frac{(sz_t+m)^2}{v-2}\xi^{-2A_t}\right)\right\} \tag{4.13}$$

where $z_t = \frac{\epsilon_t - m}{s}$, $m = \sqrt{\frac{v-2}{\pi}}\frac{\Gamma\left(\frac{v-1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right)$, $s^2 = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2$ and

$$A_t = \begin{cases} 1 & if \quad z_t \geq \frac{m}{s} \\ -1 & if \quad z_t < \frac{m}{s} \end{cases}$$

Lambert and Laurent (2001) undertook a Monte Carlo simulation in order to test the efficiency of the MLE estimation procedure for the GARCH(1,1) model. They found that the parameter estimates showed small biases when one wrongly assumes that the error distribution is a skewed t distribution. It was however noted that the specification of the skewed t distribution as the error distribution was superior to Quasi Maximum Likelihood estimation (i.e. the normality assumption) since it allows for both skewness and kurtosis to be incorporated into the model. The benefits of the skewed Standardised t distribution was demonstrated by examining twelve years of **daily** returns data on the NASDAQ from January 1985 to December 1996. Pearson goodness of fit tests were used to test assumptions about the error distribution of the GARCH process. It was found that both the normal and the Students t distribution could not adequately model the returns series. The skewed Standardised t distribution

however was found to be adequate.

4.2.9 Linear Regression with ARCH error terms

Suppose that one wishes to estimate the parameters of a regression model with *ARCH* errors. Let the regression equation be:

$$Y_t = X_t'\beta + e_t = X_t'\beta + \sqrt{h_t}\varepsilon_t \tag{4.14}$$

where $e_t = \sqrt{h_t}\varepsilon_t$, X_t contains k exogenous variables and $h_t = \alpha_0 + \alpha_1(Y_{t-1} - X_{t-1}'\beta)^2 + \alpha_2(Y_{t-2} - X_{t-2}'\beta)^2 + ... + \alpha_p(Y_{t-p} - X_{t-p}'\beta)^2$ and $\{\varepsilon_t\}$ is a iid random variable with mean 0 and variance 1.

If $\varepsilon_t \sim N(0,1)$ independent of X_t then the conditional distribution $Y_t|X_t \sim N(X_t'\beta, h_t)$ and thus the conditional density is:

$$f_{Y_t|X_t}(y_t|x_t) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-\left(y_t - x_t'\beta\right)^2}{2h_t}\right) \tag{4.15}$$

The conditional log likelihood is:

$$l_c(y_{p+1}, y_{p+2}, ..., y_T|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{t=p+1}^{T} \left\{ -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(h_t) - \frac{1}{2}\frac{\left(y_t - x_t'\beta\right)^2}{h_t} \right\} \tag{4.16}$$

This conditional log likelihood can evaluated by using the *Berndt, Hall,Hall* and *Hausman* (1974) algorithm. The optimisation algorithm uses numerical differentiation in order to solve for the unknown parameters.

Weiss (1984,1986), Bollerslev and Wooldridge (1992) and Glosten *et al.* (1991) show that maximising a normal log likelihood will provide **consistent** estimates even if the residuals of the fittted model is not normally distributed (provided that the residuals has a zero mean and unit variance.) They stress however that the standard errors of the estimates will have to be adjusted. This result can be stated more formally in the following theorem.

Definition 2 $s_t(\theta) = \frac{d\log f(y_t)}{d\theta}$ *(i.e. the matrix containing all of the first derivatives of the log likelihood with respect to the unknown parameters)*

Definition 3 $S = \lim t \to \infty \frac{1}{T}\sum_{i=1}^{T}\left[s_t(\theta)\right]\left[s_t(\theta)\right]'$ *which can be consistently estimated by* $\hat{S} = \frac{1}{T}\sum_{i=1}^{T}\left[s_t\left(\hat{\theta}\right)\right]\left[s_t\left(\hat{\theta}\right)\right]'$

Definition 4 $D = \lim t \to \infty \frac{1}{T} \sum_{i=1}^{T} E\left(-\frac{d(s_t(\theta))}{d\theta'}\right)$ *which can be consistently estimated by* $\hat{D} = \frac{1}{T} \sum_{i=1}^{T} \left(-\frac{d\left(s_t(\hat{\theta})\right)}{d\theta'}\right)$

Theorem 5 *If* $\hat{\theta}$ *is a matrix containing the estimates of* θ *found by maximising a normal log likelihood. Under certain regularity conditions,* $\sqrt{T}\left(\hat{\theta} - \theta\right) \to N\left(0, D^{-1}SD^{-1}\right)$ *even if the residual series of a ARCH model is not normally distributed.*

The above estimation procedure is known as Quasi Maximum likelihood estimation. Engle and Gloria Gonzalez Rivera (1991) investigated the relative efficiency of quasi maximum likelihood estimates compared to maximum likelihood estimation for the GARCH(1,1) model by considering different gamma and t distribution representations of the residual series. They found that QML ensures the consistency of the parameter estimates however they stress that the loss of efficiency due to the normality assumption could be significant.

4.2.10 Other ARCH specifications

The ARCH in Mean Model

Much of financial economic theory relies on the assumption of the existence of a positive relationship between expected return and risk (measured by means of the variance of a shares returns). For example, the Capital Asset Pricing Model (CAPM) of Sharpe (1964),Litner (1965), Mossin (1966), Merton(1973) or the Consumption Based CAPM of Breedon (1979), Lucas (1978), the Arbitrage Pricing Theory (APT) of Ross (1976), Chamberlain and Rothschild (1983). Due to this relationship researchers developed the ARCH-M or ARCH in mean model in order to model the conditional expectation of a share as a function of the conditional volatility. Formally the model can be expressed as:

$$r_t = \mu_t + \delta h_t + \varepsilon_t \tag{4.17}$$

where δ is a risk aversion factor and all of the common assumptions hold as in the above model specification.

Intuitively δ should be non-negative however as shown by Ballie and DeGennaro (1990), the sign of δ is dependent on the assumption of the error distribution of the volatility model. They found that the δ coefficient changes from being significant

at the 5% level to being insignificant by simply changing the error distribution assumption from being normal to being Students t. Bollerslev and Wooldridge (1991) found similar results by using robust standard errors. Further evidence against the ARCH-M specification was provided by Glosten *et al.* (1991). They found that the sign of δ is sensitive to the specification of the mean and variance equation. As a final criticism, it is important to note the findings of Pagan and Ullah (1998). They state that the parameters in the conditional mean equation (of the ARCH-M) are not asymtotically independent of the variance equation and thus any misspecification in the variance equation leads to inconsistent and biased estimates of the parameters in the mean equation. In many cases one should be careful when using the ARCH-M model due to the many problems associated with consistency and the sign of the δ coefficient.

4.2.11 The weaknesses of the ARCH Model

The following are the weaknesses of the ARCH methodology:

1. The model is sign independent with respect to returns at time t and thus does not allow one to model assymetric information flows. (Black (1976))
2. Due to the stationarity assumptions many of the parameters are restrictively constrained. This could cause convergence problems when estimating the parameter values by means of MLE.
3. ARCH models often over-predict volatility, since they respond slowly to outlier returns.
4. ARCH models are mechanical in nature and does not provide any new financial insight to portfolio and risk management.

4.3 The GARCH Model

Bollerslev (1986) developed the Generalized ARCH models since it was found that often ARCH models required relatively long lag structures in the conditional variance equation. The GARCH model is an extension of the univariate ARCH similar to the way in which ARMA models are extensions of AR and MA models. The model can be defined as follows:

Let R_t be the mean adjusted return of a share at time t and r_t is the return of a share at time t. $R_t = r_t - \mu_t$. It is then said that R_t follows a GARCH(p,q) process if and only if:

$$R_t = \sqrt{h_t}\varepsilon_t \quad \text{with} \quad h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i R_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \tag{4.18}$$

where ε_t is a white noise random variable such that $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$. Notice that in this formulation p represents the number of ARCH terms where as q represents the number of GARCH terms.

If R_t is a GARCH(p,q) process it can be shown that R_t^2 is an ARMA(max(p,q),q) process. Box Jenkins methods and the Extended Sample Autocorrelation Function ($ESACF$) methodology could be used in order to identify the order of the GARCH process.

This can be seen by rearranging equation 4.18 as follows:

$$h_t = \alpha_0 + (\alpha_1 R_{t-1}^2 + ... + \alpha_p R_{t-p}^2) + (\beta_1 h_{t-1} + ... + \beta_q h_{t-q}) \tag{4.19}$$

Now adding R_t^2 to both sides and rearranging one gets:

$$\begin{aligned} h_t + R_t^2 &= \alpha_0 + (\alpha_1 R_{t-1}^2 + ... + \alpha_p R_{t-p}^2) + (\beta_1 h_{t-1} + ... + \beta_q h_{t-q}) + R_t^2 \qquad (4.20) \\ &= \alpha_0 - \beta_1(R_{t-1}^2 - h_{t-1}) - \beta_2(R_{t-2}^2 - h_{t-2}) - ... - \beta_q(R_{t-q}^2 - h_{t-q}) + \\ &\quad \beta_1 R_{t-1}^2 + \beta_2 R_{t-2}^2 + ... + \beta_q R_{t-q}^2 + (\alpha R_{t-1}^2 + \alpha_2 R_{t-2}^2 + ... + \alpha_p R_{t-p}^2) + R_t^2 \end{aligned}$$

Now rearranging:

$$\begin{aligned} R_t^2 &= \alpha_0 + (R_t^2 - h_t) - \beta_1(R_{t-1}^2 - h_{t-1}) - \beta_2(R_{t-2}^2 - h_{t-2}) - ... - \beta_q(R_{t-q}^2 - h_{t-q}) + \\ &\quad + \beta_1 R_{t-1}^2 + \beta_2 R_{t-2}^2 + ... + \beta_q R_{t-q}^2 + \alpha_1 R_{t-1}^2 + \alpha_2 R_{t-2}^2 + ... + \alpha_p R_{t-p}^2 \end{aligned}$$

Such that:

$$R_t^2 = \alpha_0 + \left((\alpha_1 + \beta_1)R_{t-1}^2 + ... + (\alpha_p + \beta_p)R_{t-p}^2\right) + \left(w_t - \beta_1 w_{t-1} - ... - \beta_q w_{t-q}\right) \tag{4.21}$$

where $w_t = R_t^2 - h_t$ so that

$$R_t^2 = \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) R_{t-i}^2 + w_t - \sum_{j=1}^{q} \beta_j w_{t-j} \tag{4.22}$$

From this one can see that if R_t is a GARCH(p,q) process then R_t^2 is an ARMA(a, b) process where the AR coefficients are $\alpha_i + \beta_i$, the MA coefficients are β_i, a=max(p,q) and b=q. The GARCH process is thus only stationary if R_t^2 is stationary implying that $|\alpha_i + \beta_i| < 1$ and $|\beta_i| < 1$. Since $h_t \geq 0$, the stationarity conditions are: $0 < \alpha_i + \beta_i < 1$ and $0 < \beta_i < 1$ and thus $0 < \sum \alpha_i + \sum \beta_i < 1$.

4.3.1 The GARCH(1,1) Model

The model is specified as follows:

$$R_t = \sqrt{h_t}\varepsilon_t \quad \text{where} \quad h_t = \alpha_0 + \alpha_1 R_{t-1}^2 + \beta_1 h_{t-1} \tag{4.23}$$

or

$$R_t^2 = \alpha_0 + (\alpha_1 + \beta_1) R_{t-1}^2 + w_t - \beta_1 w_{t-1} \tag{4.24}$$

Assuming that the error distribution is a N(0,1) random variable then the log likelihood function is identical to equation 4.5 although the summation starts at 2 and not $p+1$. Note however that since h_t depends on the first lagged conditional variance one requires an estimate of h_1. The unconditional sample estimate is often used as a simple approximation (as recommended by Bollerslev (1986)). h_1 could also be set equal to zero.

4.3.2 Forecasting the conditional volatility using a GARCH(1,1) Model

Forecasting using this model is quite simple. The one step ahead forecast is:

$$\sigma_{t+1|t}^2 = \alpha_0 + \alpha_1 R_{t|t}^2 + \beta_1 \sigma_{t|t}^2 \tag{4.25}$$

It can be shown that a l step forecast is:

$$\sigma^2_{t+l|t} = \alpha_0 + (\alpha_1 + \beta_1)\,\sigma^2_{t+l-|t} \tag{4.26}$$

Since R^2_t is an ARMA(1,1) model with parameters α_0 and $(\alpha_1 + \beta_1)$. The l step forecast will converge to the unconditional variance , $\left(\lim_{l\to\infty} \sigma^2_{t+l|t} = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}\right)$ as l tends to ∞.

4.3.3 Limitations of the GARCH Model

GARCH model are generalizations of ARCH models and thus have many of the same limitations as the ARCH model. They are however able to model the conditional variance of a time series by using relatively few parameters. A GARCH(1,1) is most often used since in general it is difficult to specify the values of p and q. Notice also that the GARCH model treats both positive and negative returns in the same way since the square of lagged returns are used in order to model the conditional variance equation of shares returns.

Assymetric volatility models were developed in order to incorporate the sign of lagged returns into the specification of the variance equation. Threshold ARCH (TARCH) (Zakoian (1990), and (Glosten *et al.* (1993)) and Exponential GARCH (EGARCH) are two types of assymetric volatility models. The following section briefly discusses the EGARCH methodology.

4.4 The EGARCH Model

Black (1976) and Christie (1982) both found that there exists a negative correlation between current returns and future volatility of share returns. i.e. volatility tends to rise in response to bad news and fall with good news (where good news indicates the situation in which returns are larger than the concensus view). The standard GARCH model cannot capture this observation since it models the conditional volatility as the sum of squared lagged returns and lagged conditional variances. The size of the returns at each time period is thus more important than the sign in a GARCH context. Nelson (1990c)developed the exponential GARCH (EGARCH) in order to

capture the above observation. The conditional variance is modelled as:

$$\log(h_t) = \alpha_t + \sum_{j=1}^{\infty} \beta_j g(\varepsilon_{t-j}) \tag{4.27}$$

where $\{\alpha_t\}_{t=-\infty,\infty}$ and $\{\beta_j\}_{j=1,\infty}$ are real, non-stochastic scalars, while $g(.)$ is a function of both the size and sign of the return during time period t. *Nelson* (1990c) proposed using:

$$g(\varepsilon_t) = \theta\varepsilon_t + \gamma\left[|\varepsilon_t| - E(|\varepsilon_t|)\right] \tag{4.28}$$

and thus:

$$g(\varepsilon_t) = \begin{cases} (\theta+\gamma)\varepsilon_t - E(|\varepsilon_t|) & for \quad \varepsilon_t > 0 \\ (\theta-\gamma)\varepsilon_t - E(|\varepsilon_t|) & for \quad \varepsilon_t \leq 0 \end{cases} \tag{4.29}$$

It can be seen that $g(.)$ models the conditional variance asymmetrically and that it does allow for a negative relationship to exist between returns and future conditional volatility. (i.e. if $\gamma = 0$ and $\theta < 0$ then the change in the conditional variance is positive (negative) when $\varepsilon_t < 0$. $(\varepsilon_t > 0)$

The GED (Generalized Error Distribution (Harvey 1981b), Box and Tiao (1973)) is often used to model the innovation distribution. It is defined as follows:

$$f(x) = \frac{v \exp\left[\frac{-1}{2}\left|\frac{x}{\lambda}\right|^v\right]}{\lambda 2^{\left(1+\frac{1}{v}\right)}\Gamma\left(\frac{1}{v}\right)} \qquad where \quad -\infty < x < \infty \tag{4.30}$$

where $\lambda = \sqrt{2^{\left(\frac{-2}{v}\right)}\frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)}}$, $\Gamma(.)$ is the gamma function and $v > 0$ is a tail parameter to be estimated. The GED contains the normal distribution as a special case (i.e. when v=2), however heavy tailed distributions can be modelled by setting $v < 2$. Similarly thinner tailed distributions (than the normal) is modelled by setting $v > 2$

Part III

Model Selection

Chapter 5

Model Selection and Information Criteria

5.1 Introduction

Statistical inference is undertaken by making certain assumptions about the data generating process of a data set. Parzen (1982) states that statistical modelling is the process of fitting a model to a data set **without knowing what the true model is or might be**. In general the exact data generating process is unknown, indicating that we have to estimate the process by using statistical methods. Model selection techniques have been developed in order to aid researchers in finding a suitable approximation to the true data generating process of a data set.

Stone (1981) proposed that model selection criteria should obey Occam's Razor. According to Occam's Razor model selection techniques should ensure that **simpler models** are preferred to more **complex** ones. Different aspects of the model selection process have also been considered. Lindley (1968) notes that the measurement cost of acquiring information about different variables in a model should be taken into account when considering different models. This chapter focuses on the use of information criteria in order to undertake model selection in a time series and regression context. The chapter is organised as follows. Section 5.2 formalises how we can approximate the true data generating process of a data set by means of discrepancies. Section 5.3 discusses the use of different information criteria. Section 5.4 and section 5.5 discusses the use of variable selection in a regression and time series context respectively.

5.2 Model Selection based on Discrepancies

Model selection is the process of attempting to select one or a group of similar models from a finite number of competing models, that suitably describes the salient features of a given data set. Linhart and Zucchini (1986) defines the **operating model** as being the *"model that we use to think about the data"*. It is the best (possibly simple) representation of the true data generating process. In order to undertake statistical inference (model selection, parameter estimation, hypothesis testing, model checking and forecasting) , the true data generating process has to be approximated by some model or group of models. Linhart and Zucchini (1986) defines all approximating models as being part of a **approximating family** of models. The difference between the model proposed and the true model is defined as the **discrepancy**. (Haldane (1951), Sahler (1970), Robertson (1972), Durban and Knott (1972), Geisser (1974), Parr and Schucany (1980), Parr and De Wet (1981), Ponnapali (1976), Sakamoto and Akaike (1978a))

The discrepancy can be represented by some finite number. The larger the number, the greater the difference between the proposed model and the true model. Two types of discrepancies have been proposed in the literature. The **discrepancy due to approximation** is defined as being the discrepancy between the operating model and the approximating model that is nearest to it. The **discrepancy due to estimation** is the discrepancy between the fitted model and the approximating model. (Linhart and Zucchini (1986))

Model selection can now be formally separated into two stages. The first step entails the decision about the choice of an approximating family. As an example, consider multiple regression. Given a data set containing n observations from an unknown process, we could model the process by means of observing a number of related factors of the process and then represent the relationship by a linear combination of these factors. In this example the approximating family used is a linear approximation. Other approximating families could be used since the linear approximating family might not always be the most appropriate in all practical examples. The second step entails estimating the parameters of the model in order to compare competing models. Kullback and Leiber (1951) states that **any selection criteria should ensure that the final model selected should be as close as possible**

to the true data generating process of the data. Operationally this cannot be undertaken directly since the true data generating process is unknown in practice. The Kullback-Leiber statement suggests that the final model should be the model that minimises the **overall discrepancy.** This is often measured as being the sum of the two discrepancies mentioned previously.

5.3 Different Information Criteria

5.3.1 The Kullback-Leiber Distance

As noted above, model selection entails selecting a model from a finite number of competing models. Define $f(x)$ (continuous) as the density function of the **true data generating process** where $\mathbf{X}$ is a multivariate random variable. In reality $f(x)$ is unknown and has to be approximated by using a finite data set and the researchers prior beliefs about the data generating process. Define $g_i(x)$ (continuous) for $(i = 1, 2, ..., M)$ as being **a set of competing models** that approximate $f(x)$. In an attempt to measure the suitability of an approximating model the Kullback-Leiber distance was derived (Kullback and Leiber (1951)). The Kullback-Leiber distance, $KL(f(x), g_i(x))$ measures how well $g_i(x)$ approximates $f(x)$. It is defined as the expectation of $\log\left(\frac{f(x)}{g_i(x)}\right)$ under $f(x)$ such that

$$\begin{aligned} KL(f(x), g_i(x)) &= \int f(x) \log\left(\frac{f(x)}{g_i(x)}\right) dx \\ &= E(\log(f(x))) - E(\log(g_i(x))) \end{aligned} \tag{5.1}$$

$KL(f(x), g_i(x))$ is always non-negative and equals zero when $f(x) = g_i(x)$. From equation 5.1 it can be seen that the aim is to select $g_i(x)$ such that it is as close as possible to $f(x)$. Unfortunate since $f(x)$ is seldom known $KL(f(x), g_i(x))$ cannot be used directly when undertaking model selection.

5.3.2 Akaike's Information Criterion

From equation 5.1 it can be seen that $E(\log(f(x)))$ is equal to an unknown constant (c) that does not change for different $g_i(x)$. Model selection thus entails choosing $g_i(x)$ such that $-E(\log(g_i(x)))$ is minimised. Akaike (1973) showed that

$KL(f(x), g_i(x))$ could be approximated by

$$AIC = -2\log L\left(\hat{\theta}\right) + 2k \tag{5.2}$$

where $L\left(\hat{\theta}\right)$ is the maximum likelihood, $\hat{\theta}$ is the MLE of the k independent parameters of the model. Akaike (1973) stated that the model that minimises AIC should be preferred since it would theoretically be the closest model to the unknown model, $f(x)$. AIC is a compromise between the **"goodness of fit"** (the maximum likelihood) of the model and the **number of of free parameters** in the model. The second term in AIC acts as a penalty function which penalises high dimensional models more heavily than low dimensional models, ceteris paribus.

Sigiura (1978) and Sakamoto *et al.* (1986) found that AIC may perform poorly if there are too many parameters in the model in relation to the size of the sample size. Hurvich and Tsai (1989) refined AIC in order to take the sample size into consideration. AIC_c can provide a better approximation to the Kullback-Leiber distance than the AIC when the sample size is small and behaves like AIC when the sample size is *large*. Their criterion is defined as

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \tag{5.3}$$

Takeuchi (1976) found the Kullback-Leiber distance could be approximated by

$$TIC = -\log L\left(\hat{\theta}\right) + trace\left(J(\theta_0) I(\theta_0)^{-1}\right) \tag{5.4}$$

where $J(\theta_0)$ and $I(\theta_0)$ are the first and second partial derivatives of $E(\log(g_i(x)))$ with respect to θ evaluated at θ_0, the true parameters of the parameters of $g_i(x)$. If $g_i\left(x|\hat{\theta}\right)$ is a good approximation of $f(x)$ then $trace\left(J(\theta_0) I(\theta_0)^{-1}\right)$ is a good approximation of k such that AIC and TIC will select similar approximating models.

5.3.3 Dimension Consistent Information Criterion

Schwarz (1978) developed the Bayesian Information Criterion (BIC). The criterion is defined as

$$BIC = -2\log L\left(\hat{\theta}\right) + k\log(n) \tag{5.5}$$

where n is the number of observations in the data set. Model selection is undertaken by selecting the model that minimises BIC. BIC is known as a **dimension**

consistent information criterion. This means that as n becomes large, the probability of selecting a model of dimension smaller (underfitting) than the true dimension converges to zero and the probability of selecting a higher-dimensional models (overfitting) converges to zero.

Other dimension consistent information criteria include Rissanen's (1989) Minimum Description Length (MDL) and Hannan and Quinn's (1979) HQ criterion. MDL is equivalent to BIC while HQ has a penalty term equal to $c\log(\log(n))$ where $c > 2$ and n is the sample size.

5.3.4 Information Complexity (ICOMP)

Occums Razor suggests that a model selection criteria should ensure that **simpler** models are preferred to more **complex** ones. Simpler in this case suggests **parsimony** in terms of the number of variables included in a model, where as complex suggests a **high dimensional model**. Most of the above information criteria select a model by maximising the log likelihood of a model and penalising it by some scalar value. AIC penalises a model based on the number of variables in a model, where as AIC_c, BIC, MDL and HQ penalise a model as a function of both the sample size and the number of variables in the model. These penalty functions aim to penalise more complex models.

Numerous authors have investigated the **complexity** of statistical models leading to the development of numerous definitions of the complexity of a statistical model. (for example, "Kolmogorov Complexity" (Cover *et al.* (1989)), "Shannon complexity" (Rissanen (1989)) and "Stochastic Complexity " (Rissanen (1986))). Bozdogan and Bearse (1999) defines the complexity of a system as " *a measure of the degree of interdependency between the whole system and a simple enumerative composition of its subsystems or parts.*" This definition suggests that the complexity of a model is not based solely on the number of variables included in a model but rather on how they are **related** with each other. Bozdogan has used the above definition and derived a number of Information Complexity Criteria (ICOMP) used for model selection.

ICOMP uses a penalty function that is based on the **information complexity index** of Van Emden (1971). $ICOMP_o$ penalizes the **covariance complexity** of a estimated model. Models that have many **insignificant** parameter estimates as well

as well as **collinear** data sets (see section 5.4.1 for a definition) are penalized more heavily than models containing many **significant** parameter estimates based on a **orthoganal** data matrix. $ICOMP_o$ is defined as:

$$ICOMP_o = -2\log L\left(\hat{\theta}\right) + 2C_0\left(\hat{\Sigma}_{\hat{\theta}}\right) \tag{5.6}$$

where:

$$C_0\left(\hat{\Sigma}_{\hat{\theta}}\right) = \frac{1}{2}\sum_{j=1}^{p}\log\left(\sigma_j^2\right) - \frac{1}{2}\log\left|\hat{\Sigma}_{\hat{\theta}}\right| \tag{5.7}$$

and $\hat{\theta}$ is the estimated parameters of the model, k is the number of estimated free parameters in the model and $\hat{\Sigma}_{\hat{\theta}}$ is the estimated covariance matrix of the model parameters. Van Emden (1971) has shown that $ICOMP_o$ is not invariant under **orthonormal transformations**. This limits its use as a model selection criterion. Bozdogan (1988, 1990, 1993, 1994) shows that the **maximal informational complexity** of $\hat{\Sigma}_{\hat{\theta}}$, $ICOMP(C_1(\hat{\Sigma}_{\hat{\theta}}))$, solves the above problem since it is invariant with respect to scalar multiplication and orthonormal transformation. $ICOMP(C_1(\hat{\Sigma}_{\hat{\theta}}))$ is defined as:

$$ICOMP(C_1(\hat{\Sigma}_{\hat{\theta}})) = -2\log L\left(\hat{\theta}_k\right) + 2C_1\left(\hat{\Sigma}_{\hat{\theta}}\right) \tag{5.8}$$

where:

$$C_1\left(.\right) = \frac{s}{2}\log\left(\frac{tr\left(.\right)}{s}\right) - \frac{1}{2}\log|.| \tag{5.9}$$

and s is the dimension of $\hat{\Sigma}_{\hat{\theta}}$. J Magnus (proof shown in appendix of Bozdogan and Haughton (1998)) shows that $C_1\left(\hat{\Sigma}_{\hat{\theta}}\right)$ is a monotonically increasing function of the dimension of Σ, indicating that higher dimensional models are penalized more heavily due to the increased complexity of the model.

Bozdogan and Bearse (1999) show that ICOMP can be viewed as an approximation to the sum of two Kullback-Leiber measures. $ICOMP(IFIM)$ is defined as

$$ICOMP(IFIM) = -2\log L\left(\hat{\theta}\right) + 2C_1\left(\hat{F}^{-1}\right) \tag{5.10}$$

where $C_1\left(\hat{F}^{-1}\right)$ is the maximal information complexity of $\hat{F}^{-1}$, the inverse of the estimated Fisher Information matrix (Rice (1995)). $ICOMP(IFIM)$ is related to the geometric and arithmetic means of the eigenvalues of $IFIM$ such that:

$$ICOMP(IFIM) = -2\log L\left(\hat{\theta}\right) + s\log\left(\frac{\lambda_a}{\lambda_g}\right) \tag{5.11}$$

where λ_a is the arithmetic average of the eigenvalues of $\hat{F}^{-1}$, λ_g is the geometric average of the eigenvalues of $\hat{F}^{-1}$ and $s = \dim\left(\hat{F}^{-1}\right)$. The first component of $ICOMP(IFIM)$ is a measure of the **fit** of the model where as the second component measures the **complexity of the estimated inverse of the Fisher Information matrix**. The complexity of $\hat{F}^{-1}$ is measured by using the diagonal and the off-diagonal elements of $\hat{F}^{-1}$. The diagonal elements of $\hat{F}^{-1}$ is the estimated variances of the parameters of the model and the off-diagonal elements are the covariances between the parameters. From $C_1(.)$ it can be seen that in order to minimise $ICOMP(IFIM)$, a model should have small estimated parameter variances and $\log\left|\hat{F}^{-1}\right|$ should be large (the design matrix should preferably be orthogonal). Bozdogan and Bearse (1999) state that $ICOMP(IFIM)$ chooses simpler models than more complex over-specified models.

5.4 Variable and Model Selection in Regression

Variable selection procedures are mechanical rules that allow the researcher to decide which variables should be included in the linear model. The aim is to select important or relevent variables and to discard irrelevant ones. The following section discusses different variable selection techniques. For completeness sake a brief summary of the different regression techniques is provided next.

5.4.1 Regression Models and its Extensions

The Linear Regression Model

The linear regression model is defined as

$$Y = X\beta + e$$

where:

1. Y is a $n * 1$ matrix containing an observed response variable of interest,
2. X is a $n * k$ matrix containing p fixed independent variables and a column of ones if a intercept term is entertained such that $k = p + 1$
3. β is a $k * 1$ vector containing the beta coefficients and
4. e is assumed to be a normal random variable with mean zero and fixed variance σ^2.

The effects of Collinearity

When one utilises OLS multiple regression in order to model Y as a linear combination of p indepent variables, $\hat{\beta} = (X'X)^{-1} X'Y$ is used as an estimate of the beta coefficients in the model $Y = X\beta + e$. This estimate is an unbiased estimate of β and it can be shown that $\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$ (Wetherill (1986)) such that an unbiased estimate of σ^2 is equal to $s^2 = \frac{1}{n-k} e'e$. However,when $X'X$ is singular or near-singular one cannot perform the inversion of $X'X$ and the normal equations $X'X\beta = X'Y$ do not have a unique solution. This problem occurs since there exists at least one exact linear relationship or near linear relationship between some of the columns of the X matrix and it is said that the design matrix, X is ill-conditioned.

Thiart (1990) stresses however that "*Collinearity can not be described in simple*

terms as being present or absent. Rather, what is important is the degree and what effect this degree can have on the regression model." When collinearity is severe the standard errors of the regression coefficients are inflated thus making hypothesis testing, prediction and the interpretation of the OLS coefficients very difficult. The coefficients are often unstable and may have the incorrect sign. Variance inflation occurs since $var\left(\hat{\beta}\right) = s^2\left(X'X\right)^{-1}$ and thus when $|(X'X)|$ is close to zero $(X'X)^{-1}$ is large causing $var\left(\hat{\beta}\right)$ to be large as well. In the extreme case $|(X'X)| = 0$ implying that $var\left(\hat{\beta}\right) = \infty$.

Collinearity is important in regression analysis and in the following section the sources of collinearity and methods of identifying collinearities is briefly highlighted.

Sources of collinearity

Collinearity can be induced by a number of factors, namely:

(a) **Physical constraints on the model**

Collinearity could be induced if the experimentor constrains the design matrix of the model. This is often found in compositional data where the sum of the columns has to sum to one. Another example of this occurs in chemical analyses where the sum of certain constituents in a solution has to sum to some constant.

(b) **An over-defined model**

A model is said to be over-defined when the design matrix is such that there are more explanatory variables than observations. This often occurs when undertaking medical research where many observations are taken on each patient. Note however that when a model is over-defined one is unable to solve for β in the normal equations since there are infinitely many solutions to $X'X\beta = X'Y$.

(c) **Incorrect Sampling technique**

Collinearity should seldom be a problem in control experiments since the choice of treatment levels should be set up such that the factors are nearly independent, however in many situations data is not gathered by undertaking any experiments. Incorrect sampling could occur if the experimenter only samples from a small subset of the X variable space.

Detecting Collinearity

Collinearity can be detected by using the following diagnostics.

1. Investigate small singular values
2. Investiage large condition indices and large condition numbers
3. Variance decompostions
4. Mixed Condition index
5. Variance inflation factors

A detailed discussion can be found in Belsely Kuh and Welsch (1989) and will not be dealt with in this thesis.

The Standardised Linear Regression Model

The standardised regression model is defined as follows:

$$Y^* = X^*\beta^* + e^*$$

where:

1. Y^* is a $n * 1$ matrix such that the $i'th$ element of Y^* is equal to $y_i^* = \frac{(y_i - \overline{y})}{\sqrt{n-1}s_y}$ where $\overline{y}$ and s_y^2 is equal to the sample mean and variance of Y,
2. X^* is a $n*p$ matrix such that the $(ij)'\,th$ element of X^* is equal to $x_{ij}^* = \frac{(x_{ij} - \overline{x}_j)}{\sqrt{n-1}s_{x_j}}$ where $\overline{x}_j$ and $s_{x_j}^2$ is equal to the sample mean and variance of x_j

The ordinary least squares estimate of β^* is equal to

$$\hat{\beta}^* = (X^{*\prime}X^*)^{-1} X^{*\prime}Y^* \tag{5.12}$$

such that the covariance structure of the estimate is equal to

$$Var\left(\hat{\beta}^*\right) = \sigma^2 (X^{*\prime}X^*)^{-1}$$

where σ^2 can be estimated by $s^2 = \frac{1}{n-p}e^{*\prime}e^*$ and

$$Cov\left(\hat{\beta}_i^*, \hat{\beta}_j^*\right) = \sigma^2 (X^{*\prime}X^*)_{ij}^{-1}$$

The linear regression model and the standardised model are related. The beta estimates and the variance of these estimates are related as follows:

$$\begin{aligned} \beta_j &= \frac{s_y}{s_{x_j}}\beta_j^* \qquad \text{for } j = 1, 2, ..., p \\ \beta_0 &= \overline{y} - \beta_1\overline{x}_1 - \beta_2\overline{x}_2 - ... - \beta_p\overline{x}_p \end{aligned}$$

and

$$Var\left(\beta_j\right) = \frac{s_y^2}{s_{x_j}^2}\sigma_{CS}^2 (X^{*\prime}X^*)_{jj}^{-1} \qquad \text{for all } j = 1, 2, .., p$$

$$\begin{aligned} Var\left(\beta_0\right) &= Var\left(\overline{y} - \beta_1\overline{x}_1 - \beta_2\overline{x}_2 - ... - \beta_p\overline{x}_p\right) \\ &= Var\left(\overline{y} + \beta_1\overline{x}_1 + \beta_2\overline{x}_2 + ... + \beta_p\overline{x}_p\right) \\ &= \frac{\sigma^2}{n} + \sum_{i=1}^{p}\overline{x}_i^2 Var\left(\beta_i\right) + \sum_{i\neq j}^{p}\sum_{i\neq j}^{p}\overline{x}_i\overline{x}_j Cov\left(\beta_i, \beta_j\right) \\ &= \frac{\sigma^2}{n} + \sum_{i=1}^{p}\overline{x}_i^2\frac{s_y^2}{s_{x_i}^2}\sigma^2 (X^{*\prime}X^*)_{ii}^{-1} + 2\sum_{i\neq j}^{p}\overline{x}_i\overline{x}_j Cov\left(\beta_i, \beta_j\right) \end{aligned}$$

We know that

$$Cov\left(\hat{\beta}_i^*, \hat{\beta}_j^*\right) = \sigma^2 \left(X^{*\prime} X^*\right)_{ij}^{-1} \qquad \text{for all } i \neq j$$

and

$$\beta_i \beta_j = \left(\frac{s_y^2}{s_{x_i} s_{x_j}}\right) \beta_i^* \beta_j^*$$

such that

$$Cov\left(\beta_i, \beta_j\right) = \frac{s_y^2}{s_{x_i} s_{x_j}} \sigma^2 \left(X^{*\prime} X^*\right)_{ij}^{-1} \qquad \text{for all } i \neq j.\ i, j = 1, 2, .., p$$

and

$$\begin{aligned}
Cov\left(\beta_0, \beta_i\right) &= Cov\left(\overline{y} - \beta_1 \overline{x}_1 - \beta_2 \overline{x}_2 - ... - \beta_p \overline{x}_p, \beta_i\right) \qquad \text{for all } i = 1, 2, ..., p \\
&= Cov\left(-1\left(\beta_1 \overline{x}_1 + \beta_2 \overline{x}_2 + ... + \beta_p \overline{x}_p\right), \beta_i\right) \\
&= -\sum_{j=1}^{p} \overline{x}_j Cov\left(\beta_i, \beta_j\right) \\
&= -\sum_{j=1}^{p} \overline{x}_j \frac{s_y^2}{s_{x_i} s_{x_j}} \sigma^2 \left(X^{*\prime} X^*\right)_{ij}^{-1}
\end{aligned}$$

5.4.2 Standard Regression Model Selection Techniques

Model building entails selecting those variables that are deemed important to the area under investigation. In this section it is assumed that variable selection and model selection are equivalent processes. Rawlings *et al.* (1998) stresses that the elimination of variables from the model is dependent on the aims of the study. It is stressed that variable selection procedures are reletively unimportant if the researcher's aim is to provide a simple description of the behaviour of the response variable in a particular data set. Draper and Smith (1966) adds that variable selection should be undertaken so as to provide a linear model that is *"useful for prediction purposes and includes as many variables as possible so as to provide adequate fitted values for a data set."* It is however stressed that researchers should coinsider the cost of acquiring information about the variables to be included in the final model. In general variable selection entails making a compromise between the last two points since the monitoring of many variables may be too expensive. Miller (1990) notes the importance of finding a small subset of variables that provides adequate fit and precision.

The following regression variable selection techniques are the most popular:

(1) All Possible Regressions, (2) Stepwise Procedures and (3) Information criteria such as AIC and BIC.

All Possible Regressions

Consider the linear regression model: $Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + e$. There exists $2^p - 1$ different regression equation equations that can be undertaken with the p explanatory variables. The **all possible regression** procedure entails fitting all $2^p - 1$ equations and then judging which regression equation or group of regression equations best fit the data set based on some criterion (R^2, R^2_{adj}, MSE or C_p). Often researchers fit all one variable models and then choose the best of of these models. All two variable models are then fitted in order to select the best two variable model. This process is then continued for all subset sizes until p potential models have been found. At this stage the researcher then compares the different regression equations in order to put forward one or more competing regression equations.

(a) The R^2 Criterion

The coefficient of multiple determination, R^2, defined as

$$R^2 = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2} \tag{5.13}$$

measures the amount of variation explained by the linear regression. When used for model selection, the aim is to select a model that maximises the R^2 statistic. Strict application of this criterion would ensure that the maximum R^2 model would contain all explanantory variables since the statistic cannot decrease with the inclusion of new variables into the regression equation. A visual graph of R^2 against the number of variables considered might be appropriate in order to judge the marginal increase in R^2 by the addition of new variables into the regression equation. The final model selected under this criterion would thus be the model for which R^2 has stabalized close to its maximum.

(b) The Adjusted R^2 Criterion

The adjusted R^2 statistic, R^2_{adj}, defined as

$$R^2_{adj} = 1 - \frac{n-1}{n-p}\left(1 - R^2\right) \tag{5.14}$$

takes account of the number of explanatory variables included in the regression model. R^2_{adj} does not need to increase as the number of variables increase since the increase in the R^2 statistic is adjusted by the increase in the number of variables in the new regression equation. As new variables enter into the regression equation, R^2_{adj} tends to stabalize. Rawlings *et al.* (1998) states that the simplest model with R^2_{adj} near to this stabalized value should be chosen.

(c) The Residual Mean Square Criterion

The residual mean square (MSE) is often used as an estimate of the residual variance, σ^2. Draper and Smith (1966) show that σ^2 is expected to decrease as more important variables enter into the regression equation such that MSE will tend to stabalize as the number of variables included in the equation becomes large. In many applications the chosen model is the one that minimises the MSE.

(d) Mallows C_p Criterion

The Mallows C_p criterion (Mallows (1964)) is defined as

$$C_p = \frac{p\sum \hat{e}_i^2}{s^2} + 2p - n \tag{5.15}$$

where $\sum \hat{e}_i^2$ is the residual sums of squares from the p variable model, s^2 is the estimate of σ^2 based on all of the explanatory variables and n is the sample size.

Assuming that a p parameter model is appropriate then $E(C_p) = p$. It follows that the plot of C_p versus p will indicate potentially adequate models. Such models will be close to the $C_p = p$ line. Draper and Smith (1981) suggests that models with low C_p with a value close to p should be preferred to models with higher C_p values.

Note: It is often found that different models with different amounts of variables included are very similar. (Miller (1990)) In such cases the **above criteria** could be used in order to determine the best model for each subset size. It is then the responsibility of the researcher to suggest one final model or a small group of similar models that clients could use.

Stepwise Regression Procedures

Stepwise procedures might be preferred to the All Regressions procedure due to the amount of computation required in fitting all possible regression models. Stepwise procedures use partial F tests in order to investgate whether or not a variable should be added or deleted form a regression equation. These techniques require the user to specify two F statistics called F-to-enter (or F_{in}) and F-to-leave (or F_{out}). F_{in} is usually set equal to a value between 1 and 4 while F_{out} is often set equal to a value slightly smaller than F_{in}.

(a) Backward Selection

The Backward elimination procedure starts off by fitting the regression equation containing all of the variables considered and then searches for which variables to eliminate from the regression equation. The procedure consists of the following four steps:

1. Decide upon a value for F_{out}.
2. Fit the regression equation containing all of the variables.

3. Calculate the partial F-test value for each variable as though it were the last variable to enter the regression equation. Note that the partial F-test value for each variable is equal to the square of the t-statistics of the beta coefficients such that the F-test value for the i^{th} variable is equal to

$$F_i = \frac{\hat{\beta}_i}{v_{ii}} = t_i^2 \sim F_{1,n-p-1} \tag{5.16}$$

where v_{ii} are the diagonal elements of the variance covariance matrix of the beta coefficients. (i.e. $s^2 (X'X)^{-1}$)

4. Now compare the smallest F_i values with F_{out}.

   (a) If this value is smaller than F_{out} the associated variable is deleted from the regression equation and the process is repeated by considering only the remaining variables.

   (b) If the value is greater than F_{out} the process is stopped and the final model has been found.

Draper and Smith (1966) suggest that the above procedure can provide satisfactory results although cautions to the use of the procedure if the X matrix is ill conditioned. In this regard Troskie (1999) notes that *"Such collinearities can have disastrous effects on the OLS and MLE estimates. It is well known, that because of collinearities, that the backward procedure can give entirely different results from the forward selection procedure."* Forward Selection is discussed next.

(b) Forward Selection

The Forward Selection procedure is the opposite of the Backward elimination procedure. The procedure starts off by including the variable that exhibits the highest correlation with the response variable (Y) and then searches for which variables to include in the regression equation by examining the F-test values of the variables not already in the equation. The procedure consists of the following three steps:

1. Decide upon a value for F_{in}. As stated above F_{in} is usually set equal to a value between 1 and 4, since the value 4 corresponds with a t-statistic value of 2.

2. Determine which variable is most correlated with the response variable, say X_1 then fit the regression model: $Y = \beta_0 + \beta_1 X_1 + e$. If this regression is not significant the procedure stops and the response variable can only be modelled

by its mean, $\overline{Y}$.

3. Fit all two variable models containing X_1 and then calculate the partial F-test value for each variable to enter the regression equation given that X_1 is already in the model. (Once again this is simply equal to the square of the t-statistic of the beta coefficient of the new variable that enters the equation.)

   (a) If the variable with the largest F_i value is greater than F_{in}, the variable is included into the regression equation and the process is continued by considering all three- four, five, ... p-variable equations containing the previous two-, three-, four-, ... $p-1$-variables respectively. The procedure is continued by adding a new variable to the regression equation until

   (b) The largest F_i is smaller than F_{in}.

(c) Stepwise Regression

Note that when undertaking both forward- and backward selection that variables enter and leave the model one at a time. Once a variable has entered into the regression equation they may not be deleted. Similarly in the backward elimination procedure once a variable has been deleted from the equation it cannot re-enter the model. Both procedures do not consider the effect that the inclusion or deletion of a variable has on the other variables in the model. In this regard it should be noted that a variable added early on in the procedure might become insignificant when other variables enter the equation. Similarly, in the backward elimination procedure, a variable can become significant once a number of variables have left the model. Stepwise regression uses a combination of forward selection and backward elimination in order to solve the above problem.

Stepwise regression consists of the following steps:

1. Decide upon a value for F_{in} and F_{out}.

2. Determine which variable is most correlated with the response variable, say X_1 then fit the regression model: $Y = \beta_0 + \beta_1 X_1 + e$. If this regression is not significant the procedure stops and the response variable can only be modelled by its mean, $\overline{Y}$.

3. Fit all two variable models containing X_1 and then calculate the partial F-test value for each variable to enter the regression equation given that X_1 is already

in the model.

(a) If the variable, say X_2, with the largest F_i value is greater than F_{in}, the variable is included into the regression equation.

(b) If the variable with the largest F_i value is smaller than F_{in}, the procedure stops.

4. Fit all three variable models containing X_1 and X_2 and then calculate the partial F-test value for each variable to enter the regression equation given that X_1 and X_2 is already in the model.

(a) If the variable, say X_3, with the largest F_i value is greater than F_{in}, the variable is included into the regression equation.

(b) If the variable with the smallest F_i value is smaller than F_{out}, the variable is deleted from the equation.

5. Step 4 is continued in this way by adding and deleting variables at each step until no more variables either enter or leave the regression equation.

AIC and BIC

In a regression context *eq* 5.2 and *eq* 5.5 can be rewritten as

$$AIC = n \ln \left(\sum \hat{e}_i^2 \right) + 2p - n \ln (n)$$

and

$$BIC = n \ln \left(\sum \hat{e}_i^2 \right) + p \ln (n) - n \ln (n)$$

which could now be used as a model selection procedure. The appropriate model selected is the model that minimises the AIC or the BIC measure.

Note:

It should be noted that all of the above procedures should be used as a guide in the selection of an appropriate regression model. Rawlings *et al.* (1998) states that *"no variable selection procedure can substitute for the insight of the researcher."* With reference to the C_p, Mallows (1973) comments that *"C_p cannot be expected to provide a single best equation."* Draper and Smith (1981) agrees with the above statement and

adds, *"Nor can any other selection procedure. All selection procedures are essentially methods for the orderly displaying and reviewing of the data. applied with common sense, they can provide useful results; applied thoughlessly, and/or mechanically, they may be useless or even misleading."*

5.4.3 ICOMP and Regression Model Selection

The Linear Regression Model

ICOMP related criteria have been successfully applied in the modelling of multivariate linear models. Bozdogan, Bearse and Schottmann (1997) used ICOMP in order to model the food consumption in the USA and the Netherlands by means of autoregressive distributed lag models. These techniques have also been applied to vector autoregressive and regression models. This section discusses the main results relating to ICOMP in a linear regression context.

In linear regression $\hat{\beta}_k$ (the regression coefficients) and $\hat{\varepsilon}_k$ (the estimated residuals) are independent of each other. Due to this result Bozdogan and Haughton (1998) defined the complexity of a regression model as the sum of the complexity of the covariance matrix of the beta coefficients and the covariance matrix of the residuals. The complexity of the estimated model can be quantified by means of $C_1(\hat{\Sigma}_{\text{model}})$. This statistic is defined as

$$\begin{aligned} C_1(\hat{\Sigma}_{\text{model}}) &= C_1\left(\hat{\Sigma}_{\hat{\beta}}\right) + C_1\left(\hat{\Sigma}_{\hat{\varepsilon}}\right) \\ &= \frac{s}{2}\log\left(\frac{tr\left((X'X)^{-1}\right)}{s}\right) - \frac{1}{2}\log\left|(X'X)^{-1}\right| \end{aligned} \tag{5.17}$$

Note that since $\hat{\Sigma}_{\hat{\varepsilon}}$ is assumed to be diagonal, $C_1\left(\hat{\Sigma}_{\hat{\varepsilon}}\right) = 0$ such that $C_1(\hat{\Sigma}_{\text{model}}) = C_1\left(\hat{\Sigma}_{\hat{\beta}}\right)$. Also note that $C_1(a\hat{\Sigma}) = C_1(\hat{\Sigma})$ where a is a constant. In general X represents the design matrix used in the analysis. More specifically X_k represents that the design matrix contains k explanatary variables.

A second approach is to measure the complexity of a model by using the estimated inverse of the Fischer Information matrix (IFIM) with respect to all of the parameters of the model. This includes both the beta coefficients and the variance of the error

term. For the **linear regression model** IFIM is equal to:

$$\hat{F}^{-1} = \begin{pmatrix} \sigma^2 (X_k' X_k)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix} \tag{5.18}$$

Recall that $ICOMP(IFIM)$ is defined as:

$$ICOMP(IFIM) = -2\log L\left(\hat{\theta}_k\right) + 2C_1\left(\hat{F}^{-1}\right) \tag{5.19}$$

Bozdogan (1998) defines $ICOMP_{a_n}$, $ICOMP(IFIM_{a_n})$ and $ICOMP(COV_{a_n})$ as follows:

$$ICOMP_{a_n} = -2\log L\left(\hat{\theta}\right) + 2a_n C_1\left((X'X)^{-1}\right)$$

$$ICOMP(IFIM_{a_n}) = -2\log L\left(\hat{\theta}\right) + 2a_n C_1\left(\hat{F}^{-1}\right)$$

$$ICOMP(COV_{a_n}) = -2\log L\left(\hat{\theta}\right) + 2a_n C_1\left(\hat{Q}\right)$$

where a_n is some positive number and

$$Q = \begin{pmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & 2\sigma^4 \left(\frac{n-k-1}{n^2}\right) \end{pmatrix}$$

Since Q is asymptotically equivalent to $IFIM$ the model selected should be vary similar.

Bozdogan and Haughton (1998) show that in regression models where at least one independent variable is missing in the model, AIC, BIC and ICOMP are all asymptotically consistent. In the same paper it was also shown by means of a simulation study that with finite sample sizes that the ICOMP criterion selected the model that minimised the Kullback-Leiber distance between the true model and the fitted model more often than both AIC and BIC. This result suggests that in finite sample sizes ICOMP is a superior criterion to both AIC and BIC. Notice however that as the sample size increases, ICOMP tends to resemble both AIC and BIC.

Part IV

Simulation Studies

Chapter 6

ICOMP Simulation Studies

This chapter investigates the the behaviour of Bozdogan's information complexity (ICOMP) criteria when undertaking ordinary least squares regression and time series analysis.

6.1 Regression Simulation Study

6.1.1 Design of the multicollinearity study

Bozdogan and Haughton (1998) derived the asymptotic results relating to ICOMP and investigated the finite sample properties of various ICOMP measures. The results were based on the assumption that the design matrix (X) is of **full rank**. In many practical applications this is not the case, causing multicollinearity in certain regression applications. Bozdogan and Haughton (1998) state that ICOMP gaurds against collinearity. The following simulation study investigates the behaviour of ICOMP ($ICOMP\left(\hat{\Sigma}_{\hat{\beta}}\right)$, denoted C1COV is this study), AIC and BIC under various collinearity-, sample size- and residual variance- levels. The simulation study extends Bozdogan and Haughton (1998).

In section 6.1.2 we describe how to generate the design matrix $X = (X_1, X_2, ..., X_5)$ to achieve a given correlation structure among the five explanatory variables and how to generate parameter vectors $\beta_{\max}$ and $\beta_{\min}$ to yield maximum and minimum variability for $X\beta$ respectively.

Various different conditions are considered namely:

· Three different sample size levels: 50, 100, 1000

· Three different collinearity levels: high, moderate and low

· Three different error variance levels: 0.25, 2.5, 5

For each of the various conditions, 100 replications are considered. Following Bozdogan and Haughton (1998), four mispecified models, $(\beta_1 X_1)$, $(\beta_1 X_1 + \beta_2 X_2)$, $(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$, and $(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$, are considered and the frequency of selecting each of the four models are investigated. The frequency that the three information criteria agrees with the KL decision is also investigated. In the regression model the Kullback-Leiber distance is defined as

$$KL = \frac{n}{2}\log\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) + \frac{n}{2}\left(\frac{\hat{\sigma}^2}{\sigma^2} - 1\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((X\beta)_i - \left(X\hat{\beta}\right)_i\right)^2$$

where n is the sample size, σ^2 and $\hat{\sigma}^2$ is the true and estimated residual variance respectively, β and $\hat{\beta}$ is the true and estimated beta coefficients respectively and $(X\beta)_i$ is the i^{th} fitted value from the linear regression model.

6.1.2 Generation of the X matrix and the Beta coefficients

The simulation study follows that of McDonald and Galaraneau (1995), Wichern and Churchill (1978) and Bozdogan and Haughton (1998). A five parameter model with 50, 100 and 1000 observations is generated. The explanatory variables are generated from the relationships:

$$X_i = \begin{cases} \sqrt{1-\alpha_1^2}Z_i + \alpha_1 Z_6 & for \quad i = 1,2,3,\ j = 1,2,...100 \\ \sqrt{1-\alpha_2^2}Z_i + \alpha_2 Z_6 & for \quad i = 4,5,\ j = 1,2,...100 \end{cases}$$

where the six Z variables are independent $N(0,1)$ random variables. The Z variables could have been generated by utilising the **nrnd** function found in Eviews 3.1 ensuring that a random seed is used for each Z variable. In this study however the Z variables were generated by using the Cholesky decomposition found in Eviews 3.1.

Since the Z variables are independent $N(0,1)$ random variables, the correlation matrix between the Z variables is equal to an identity matrix, Λ. The Cholesky decomposition factorizes Λ such that $\Lambda = AA'$ where A is a lower triangular matrix matrix. The Z_i values are then equal to $Z_i = AB_i$ where $B \sim iid\ N(0,1)$. All simulations were undertaken using Eviews 3.1. The code used is attached in Appendix 1 and 2 of this chapter.

α_1 and α_2 control the composition of the elements in the correlation matrix of the

design matrix as well as the degree of collinearity among the variables in the design matrix. Three different combinations of (α_1, α_2) are considered in order to simulate a high (0.99,0.99), moderate (0.9,0.9) and low (0.7,0.3) level of collinearity. These combinations are presented in table 1 below. The table also displays elements of the correlations between the X matrix as well as the condition number of the correlation matrix of the X variables.

Two coefficient vectors are used in this study, denoted $\beta_{\max}$ and $\beta_{\min}$ respectively. One is the normalised eigenvector corresponding to the largest eigenvector of $X'X$ and the other is the eigenvector corresponding to the smallest eigenvector of $X'X$. Appendix 3 and appendix 4 displays the eigenvectors associated with the different collinearity levels and sample sizes for $\beta_{\max}$ and $\beta_{\min}$ respectively. The Y vector used in this study was generated as $Y = X\beta + e$ where β could be $\beta_{\min}$ or $\beta_{\max}$ and $e \sim iid\ N(0, \sigma^2)$ for some value of σ^2.

Table 1: Correlations between the X variables and the
Square of the *Condition numbers of Correlation Matrix*

	Elements in the X Correlation Matrix			(Condition Number)^2 n		
Collinearity Level	α_1^2	$\alpha_1\alpha_2$	α_2^2	50	100	1000
Case 1: High	0.98	0.98	0.98	438	404	271
Case 2: Moderate	0.81	0.81	0.81	38	37	25
Case 3: Low	0.49	0.21	0.09	7	7	5

From the above table it can be seen that in the first case that the collinearity between variables 1-3 and variable 4 and 5 is high, while collinearities between the same variables is moderate in the second case and low in the third case.

Due to the different collinearity-, variance-, sample size- levels and the chose of beta matrix employed, 54 different scenarios will be considered. Appendix 5 displays all of the different scenarios evaluated in a tabular format. The **agreement percentage** is defined as the percentage of the time that a information criteria selects the same model as the Kullback-Leiber criterion. The following questions will be investigated in this simulations study:

1. For a given beta matrix, collinearity level and fixed sample size, **what happens to the agreement percentage as σ^2 (the residual variance) increases.**
2. For a given beta matrix, collinearity level and fixed residual variance, **what happens to the agreement percentage as n (the sample size) increases.**
3. For a given beta matrix, fixed residual variance and fixed sample size, **what happens to the agreement percentage as the level of collinearity increases.**

The results will first be evaluated for $\beta_{\max}$ and then for $\beta_{\min}$. Appendix 6 through 8 displays the number of times the different information criteria selects each model under $\beta_{\max}$ while appendix 12 through 14 displays the number of times the different information criteria selects each model under $\beta_{\min}$. Appendix 9 through 11 and appendix 15 through 17 displays the number of times that the model selected by the different information criteria agrees with the model selected by the KL criterion when $\beta_{\max}$ and $\beta_{\min}$ is used as the true coefficients in the regression model respectively.

6.1.3 The Multicollinearity Simulation Results

Simulation Results for $\beta_{\max}$

Table 2 below summarises all of the results in this section. It displays the number of times that the different information criteria selects a model that **agrees** with the model selected by the KL distance under the different conditions.

High Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	9	29	62
100	32	65	90
1000	100	100	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	18	20	15
100	13	14	11
1000	17	64	89

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	41	24	26
100	20	17	15
1000	2	42	67

Moderate Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	94	98	100
100	100	100	100
1000	100	100	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	11	27	66
100	30	62	87
1000	100	100	100

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	5	9	36
100	6	31	67
1000	91	100	100

Low Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	99	99	100
100	100	100	100
1000	100	100	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	25	50	86
100	45	78	96
1000	100	100	100

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	3	18	51
100	15	48	83
1000	97	100	100

Table 2: Frequency of agreement with KL model

The effect σ^2 has on the agreement percentages

1. **High Collinearity, Fixed n**

In this section we investigate the behaviour of AIC, BIC and C1COV under different fixed sample sizes and **increasing residual variance levels**. The following findings can be made when analysing the data under the assumption that the design matrix is highly collinear and $\beta_{\max}$ is used as the true beta coefficients.

From table 2 above it can be seen that when $\sigma^2 = 0.25$ ($n = 50$), C1COV agrees with the KL decisions most often (62%). Both AIC and BIC agree less often with the KL decision when $\sigma^2 = 0.25$ ($n = 50$). As σ^2 increases (and the sample size is held constant at 50), the agreement percentages for C1COV and BIC decreases while the agreement percentages for AIC increases such that when $\sigma^2 = 5$ (and $n = 50$) BIC agrees with the KL decisions most often.

When $\sigma^2 = 0.25$ ($n = 100$), C1COV, AIC and BIC agrees with the KL decision 90%, 65% and 32% of the time respectively. As σ^2 increases (and the sample size is held constant at 100), the agreement percentages for all of the information criteria decreases.

When $n = 1000$ C1COV agrees with the KL decision most often for all of the residual variance levels. When $\sigma^2 = 0.25$ all of the information criteria selects the KL model 100% of the time. As σ^2 increases (and the sample size is held constant at 1000) the agreement percentages for all of the information criterion decreases with C1COV decreasing at the slowest rate.

2. **Moderate Collinearity, Fixed n**

In this section we investigate the behaviour of AIC, BIC and C1COV under different fixed sample sizes and **increasing residual variance levels.** The following findings can be made when analysing the data under the assumption that the design matrix is moderately collinear and $\beta_{\max}$ is used as the true beta coefficients.

With reference to table 2 above, it can be seen that when $\sigma^2 = 0.25$ ($n = 50$) that all of the information criteria agree with the KL decisions more than 90% of the time with C1COV's agreement percentage equal to 100%. As σ^2 increases (and the sample size is held constant at 1000), the agreement percentages for all of the information criteria decreases **monotonically** such that for all residual variance levels, C1COV agree more often with KL decision than both AIC and BIC.

When $\sigma^2 = 0.25$ ($n = 100$) all of the information criteria agree with the KL decisions 100% of the time. As σ^2 increases (and the sample size is held constant at 100), the agreement percentages for all of the information criteria decreases **monotonically** such that for all residual variance levels, C1COV agree more often with KL decision than both AIC and BIC. The agreement percentage for C1COV is equal to 87% when $\sigma^2 = 2.5$ and 67% when $\sigma^2 = 5$. These totals are greater than the agreement percentages of C1COV for all of the sample sizes when the collinearity in the design matrix is high.

When $n = 1000$ all of the information criteria agree with the KL decisions 100% of the time except for BIC when $\sigma^2 = 5$. These results suggest that when $n = 1000$

all of the information criteria behave according to their asymptotic results such the model selected is always $\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$.

3. **Low Collinearity, Fixed n**

In this section we investigate the behaviour of AIC, BIC and C1COV under different fixed sample sizes and **increasing residual variance levels**. The following findings can be made when analysing the data under the assumption that the design matrix has low levels of collinearity and $\beta_{\max}$ is used as the true beta coefficients.

With reference to table 2 above, it can be seen that when $\sigma^2 = 0.25$ ($n = 50$) all of the information criteria agree with the KL decisions at least 99% of the time. As σ^2 increases (and the sample size is held constant at 50), the agreement percentages for all of the information criteria decreases **monotonically** such that for all residual variance levels, C1COV agree more often with KL decision than both AIC and BIC.

When $\sigma^2 = 0.25$ ($n = 100$) all of the information criteria agree with the KL decisions 100% of the time. As σ^2 increases, the agreement percentages for all of the information criteria decreases such that the agreement percentage for C1COV is equal to 96% when $\sigma^2 = 2.5$ and 83% when $\sigma^2 = 5$. Note also that C1COV agrees with the KL more often than AIC and BIC for all levels of σ^2.

When $n = 1000$ all of the information criteria agree with the KL decisions 100% of the time except for BIC when $\sigma^2 = 5$. These results suggest that when $n = 1000$ all of the information criteria behave according to their asymptotic results such the model selected is always $\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$.

The effect n has on the agreement percentages

In this section we investigate the behaviour of AIC, BIC and C1COV under different fixed sample sizes and **increasing sample size**. The following findings can be made when analysing the data under the assumption that $\beta_{\max}$ is used as the true beta coefficients.

With reference to table 2 above it can be seen that for all of the residual variance levels and when the collinearity is at most moderate that as the sample size increases, the agreement percentages of all three information criteria **increases monotonically**

and that C1COV agree with the KL decision more often than both AIC and BIC. When the design matrix is highly collinear and the residual variance level is equal to 2.5 or 5, the agreement percentages does not increase monotonically as the sample size increases, however when $\sigma^2 = 0.25$ this does occur.

The effect that the collinearity level has on the agreement percentages

In this section we investigate the behaviour of AIC, BIC and C1COV under different fixed sample sizes and **increasing collinearity levels**. The following findings can be made when analysing the data under the assumption that $\beta_{\max}$ is used as the true beta coefficients.

With reference to table 2 above it can be seen that for all residual variance and sample size levels, as the design matrix becomes more collinear, the agreement percentage for C1COV **decreases monotonically**. The same can be said for AIC when $\sigma^2 = 0.25$ and $\sigma^2 = 2.5$. When $n = 1000$ the agreement percentages for AIC and BIC decreases monotonically for all residual variance levels as the collinearity levels increases, however when the $\sigma^2 = 2.5$ this decrease is not monotonic. When the design matrix is highly collinear and $n = 50$ or $n = 1000$, BIC agrees with the KL decision more often than both AIC and C1COV. Notice however in the second case that BIC only agrees with the KL 20% of the time while the agreement percentage for AIC is 17% and C1COV is 15%.

Concluding Remarks

The following conclusions can be made when $\beta_{\max}$ is used as the true coefficients in the regression model.

1. From the above simulation results it can be concluded that as the residual variance increases, the agreement percentages of all of the information criteria decreases. This empirical result holds for all of the sample sizes and collinearity levels considered. The ICOMP based information criterion agrees with the KL decision more often than AIC and BIC for all of the collinearity levels when $n = 100$ and $n = 1000$. This result also holds when $n = 25$ and the design matrix is at worst moderately collinear. When $n = 50$ and the design matrix is highly collinear none of the information criteria behave similar to KL when

$\sigma^2 = 2.5$ or $\sigma^2 = 5$.

2. It can be concluded that for all of the residual variance levels and when the collinearity is at most moderate, that as the sample size increases, the agreement percentages of all three information criteria increases monotonically and that C1COV agree with the KL decision more often than both AIC and BIC. When the collinearity level is high, C1COV agree with the KL decision more often than both AIC and BIC when the sample size is equal to 1000 (for all residual variance levels).

3. In general it can be concluded that as the collinearity levels in the design matrix increases, that the agreement percentages for all of the information criteria decreases monotonically and that C1COV agrees with the KL model selected more often than both AIC and BIC. When the design matrix is highly collinear and the sample size is at most equal to 100, all of the information criteria behave similarly such that all of the agreement percentages are less than 50%.

Simulation Results for β_{min}

Table 3 below summarises all of the results in this section. It displays the number of times that the different information criteria selects a model that agrees with the model selected by the KL distance under the different conditions when β_{min} is used as the true beta coefficients. This table can analysed similar to the β_{max} case analysed previously.

The main conclusions arising from the simulation results are as follows:

1. From table 3 below it can be seen that when the collinearity is low and the sample size is fixed at 50, 100 and 1000 respectively, as the residual variance increases, the agreement percentages for all of the information criteria decreases monotonically such that C1COV agrees more often with the model selected by the KL criterion than both AIC and BIC.

2. The same cannot be said when the design matrix is moderately or highly collinear. When the collinearity in the design matrix is moderate, the agreement percentages for C1COV and BIC does decrease monotonically as σ^2 increases for all sample sizes considered. AIC behaves differently. When $n = 1000$, it behaves similar to C1COV and BIC, however when $n = 100$ or $n = 1000$, the

agreement percentage for AIC decreases when σ^2 is increased from 0.25 to 2.5, and increases when σ^2 is increased from 2.5 to 5. Note however that the agreement percentages are less than the levels observed when $\sigma^2 = 0.25$. C1COV selects the KL model more often than both AIC and BIC for all residual variance levels considered except when $\sigma^2 = 5$ and $n = 50$ in which case AIC had the largest agreement percentage.

3. When the design matrix is highly collinear, the three information criteria behave differently. When $n = 50$ and $n = 100$ the agreement percentages for all of the information criteria **increases**. When $\sigma^2 = 0.25$ the agreement percentages for C1COV is the largest when $n = 50$ and $n = 100$, however when $\sigma^2 = 2.5$ and $\sigma^2 = 5$ AIC has the largest agreement percentage for the same sample size levels.

4. In general when the collinearity levels are at most moderate, as the sample size increases, so does the agreement percentages for all of the information criteria. This does not however occur when the design matrix is highly collinear.

High Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	21	10	43
100	19	5	44
1000	100	90	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	56	79	63
100	20	42	25
1000	35	4	59

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	61	81	64
100	38	59	41
1000	13	24	17

Moderate Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	100	99	100
100	99	90	99
1000	100	100	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	24	13	56
100	16	6	51
1000	100	90	100

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	24	48	19
100	13	9	25
1000	91	52	99

Low Collinearity

$\sigma^2 = 0.25$

n	BIC	AIC	C1COV
50	98	98	99
100	100	99	100
1000	100	100	100

$\sigma^2 = 2.5$

n	BIC	AIC	C1COV
50	31	11	69
100	50	16	83
1000	93	64	99

$\sigma^2 = 5$

n	BIC	AIC	C1COV
50	16	9	49
100	18	3	60
1000	75	39	95

Table 3: Frequency of agreement with KL model

6.2 ICOMP and Time Series

A stationary $ARMA(p,q)$ model is defined as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} ... - \theta_q \varepsilon_{t-q} \tag{6.1}$$

where $\{\varepsilon_t\}$ is a white noise sequence with unknown variance σ^2 and $c > 0$. Parameter estimation can be undertaken by maximising the conditional log likelihood function

$$l(\underline{\boldsymbol{\theta}}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=p+1}^{T} \frac{\varepsilon_t^2}{2\sigma^2} \tag{6.2}$$

by iterating on $\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - ... - \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} ... + \theta_q \varepsilon_{t-q}$ where the first p observations, $y_1, ..., y_p$ and $\varepsilon_1 = ... = \varepsilon_p = 0$ are used (Box-Jenkins (1970)) and $\underline{\boldsymbol{\theta}} = (c, \phi_1, \phi_2, ..., \phi_p, \theta_1, ..., \theta_q)$. Estimation results undertaken by using the conditional log likelihood funtion and the exact likelihood function are very similar in most cases. (Hamilton (1994)) This section will investigate the suitability of using $ICOMP$ in order to undertake model selection in a time series context.

Recall that $ICOMP\,(IFIM) = -2\log L\left(\hat{\underline{\boldsymbol{\theta}}}\right) + 2C_1\left(\hat{F}^{-1}\right)$. Equation 6.2 evaluated at $\hat{\underline{\boldsymbol{\theta}}}$ can be used as an estimate of the maximised log likelihood. Define $\gamma_0 = E\,(y_t^2)$ and $\gamma_i = E\,(y_t y_{t-i})$ for all t and $i > 0$ such that F is equal to

$$F_{(p+q+2)*(p+q+2)} = \begin{pmatrix} \frac{T-p}{\sigma^2} & A & 0 & 0 \\ A & B & C & 0 \\ 0 & C & D & 0 \\ 0 & 0 & 0 & \frac{T-p}{2(\sigma^2)^2} \end{pmatrix} \tag{6.3}$$

where A is a $p*1$ matrix containing the elements of $E\left(-\frac{\delta^2 l(\underline{\boldsymbol{\theta}})}{\delta\phi_i \delta c}\right)$ for all $i = 1, 2, ..., p$, B is a $p * p$ matrix containing the elements of $E\left(-\frac{\delta^2 l(\underline{\boldsymbol{\theta}})}{\delta\phi_i \delta\theta_j}\right)$ for all $i = 1, 2, ..., p$, $j = 1, 2, ..., q$, C is a $q * p$ matrix containing the elements of $E\left(-\frac{\delta^2 l(\underline{\boldsymbol{\theta}})}{\delta\phi_i \delta\phi_j}\right)$ for all $i = 1, 2, ..., p$, $j = 1, 2, ..., q$ and D is a $q * q$ matrix containing the elements of $E\left(-\frac{\delta^2 l(\underline{\boldsymbol{\theta}})}{\delta\theta_i \delta\theta_j}\right)$ for all $j = 1, 2, ..., q$. The exact equations can be found in appendix 18 of this chapter. F^{-1} can be rewritten as

$$F^{-1}_{(p+q+2)*(p+q+2)} = \begin{pmatrix} \hat{\Sigma}_{\hat{\theta}} & 0 \\ 0 & \frac{2\sigma^4}{T-p} \end{pmatrix}$$

where $\hat{\Sigma}_{\hat{\theta}}$ is the estimated covariance matrix of the model parameters.

Recall that $ICOMP(vanEmden) = -2\log L\left(\hat{\theta}_k\right) + 2C_0\left(\hat{\Sigma}_{\hat{\theta}}\right)$

$$\hat{\Sigma}_{\hat{\theta}(p+q+1)*(p+q+1)} = \begin{pmatrix} \frac{T-p}{\sigma^2} & A & 0 \\ A & B & C \\ 0 & C & D \end{pmatrix}$$

As an example consider an $AR(1)$ process then

$$F = \begin{pmatrix} \frac{T-1}{\sigma^2} & \frac{T-1}{\sigma^2}\left(\frac{c}{1-\phi_1}\right) & 0 \\ \frac{T-1}{\sigma^2}\left(\frac{c}{1-\phi_1}\right) & \frac{T-1}{\sigma^2}\gamma_0 & 0 \\ 0 & 0 & \frac{T-1}{2\sigma^4} \end{pmatrix}$$

$$F^{-1} = \begin{pmatrix} -\frac{\sigma^2\gamma_0(-1+\phi_1)^2}{T-1\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & -\frac{(-1+\phi_1)c\sigma^2}{(T-1)\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & 0 \\ -\frac{(-1+\phi_1)c\sigma^2}{(T-1)\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & -\frac{\sigma^2(-1+\phi_1)^2}{T-1\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & 0 \\ 0 & 0 & \frac{2\sigma^4}{T-1} \end{pmatrix}$$

$$\hat{\Sigma}_{\hat{\theta}} = \begin{pmatrix} -\frac{\sigma^2\gamma_0(-1+\phi_1)^2}{T-1\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & -\frac{(-1+\phi_1)c\sigma^2}{(T-1)\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} \\ -\frac{(-1+\phi_1)c\sigma^2}{(T-1)\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} & -\frac{\sigma^2(-1+\phi_1)^2}{T-1\left(-\gamma_0+2\gamma_0\phi_1-\gamma_0\phi_1^2+c^2\right)} \end{pmatrix}$$

Van Emden's complexity measure is defined as

$$\begin{aligned} C_0\left(\hat{\Sigma}_{\hat{\theta}}\right) &= \frac{1}{2}\log\left(tr\left(\hat{\Sigma}_{\hat{\theta}}\right)\right) - \frac{1}{2}\log\left|\hat{\Sigma}_{\hat{\theta}}\right| \\ &= \frac{1}{2}\log\left(\frac{-\sigma^2\left(-1+\phi_1\right)\left(\sigma^2+1-\phi_1\right)}{\sigma^2\phi_1-T-\sigma^2+c^2}\right) - \frac{1}{2}\log\left(\frac{-\sigma^4\left(-1+\phi_1\right)^2}{\left(T-1\right)^2\left(\sigma^2\phi_1-\sigma^2+c^2\right)}\right) \end{aligned}$$

and the maximal informational complexity of $C_0\left(\hat{\Sigma}_{\hat{\theta}}\right)$ is equal to

$$\begin{aligned} C_1\left(F^{-1}\right) &= \frac{3}{2}\log\left(\frac{tr\left(F^{-1}\right)}{3}\right) - \frac{1}{2}\log\left|F^{-1}\right| \\ &= \frac{3}{2}\log\left(\frac{\frac{2\sigma^4}{T-1} + tr\left(\hat{\Sigma}_{\hat{\theta}}\right)}{3}\right) - \frac{1}{2}\log\left(\frac{2\sigma^4}{T-1}\left|\hat{\Sigma}_{\hat{\theta}}\right|\right) \end{aligned}$$

This result can be generalised for an $ARMA(p,q)$ model such that

$$C_1\left(F^{-1}\right) = \frac{p+q+2}{2}\log\left(\frac{tr\left(\hat{\Sigma}_{\hat{\theta}} + \frac{2\sigma^4}{T-p}\right)}{p+q+2}\right) - \frac{1}{2}\log\left|\frac{2\sigma^4}{T-p}\hat{\Sigma}_{\hat{\theta}}\right|$$

In small samples $\hat{F}^{-1}$ can be approximated by means of the analytical derivatives of the log likelihood function. As $T \rightarrow \infty$ these analystical derivatives tend towards the true covariance matrix.

6.3 Time Series Simulation Study

Bozdogan (1998) has derived the asymptotic results as well as the finite sample results relating to ICOMP type information criteria in a regression context. The following simulation study will investigate the behaviour of ICOMP and other information criteria in a **time series** context. The study entails simulating stationary ARMA and GARCH models 1000 times and then fitting different time series models to the simulated series. Different series will be considered by changing the size of the residual variance. We consider all subset models that have at most three AR, MA, ARCH and GARCH terms. We then investigate the **frequency of selecting** the different models by each of the following information criteria: maximum log likelihood, AIC, BIC, ICOMP(IFIM), ICOMP(VanEmden) and COMP (defined as $2C_1\left(\hat{F}^{-1}\right)$ where $\hat{F}^{-1} = \begin{pmatrix} \hat{\Sigma}_{\hat{\theta}} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{T-p} \end{pmatrix}$ and $\hat{\Sigma}_{\hat{\theta}}$ is the estimated covariance matrix of the estimated coefficients of the mean equation).

The simulation study will be split into two parts. The ARMA models will first be investigated followed by the GARCH models. Seven different ARMA models are considered. Two different sample size- and residual variance- levels are employed when simulating the ARMA models, namely 250 and 1000 observations and a residual variance of 1 and 25. The time series are simulated using Eviews 3.1. See appendix 19 and 20 of this chapter for the Eviews code. The GARCH models investigated have a sample size of 1000 observations and a residual variance of 1. The sample size- and residual variance- levels were selected so as to investigate the properties of the different information criteria under different scenarios.

Seven different ARMA models are considered. Their formulations are as follows:

1. A pure AR(1) model with $\phi_1 = 0.8$
2. MA(1) with $\theta_1 = 0.8$
3. ARMA(1,1) with $\phi_1 = 0.8$ and $\theta_1 = 0.8$
4. A autoregressive model with a significant third term and parameter $\phi_3 = 0.8$.
5. A moving average model with a significant third term and parameter $\theta_3 = 0.8$.
6. A pure AR(3) model with $\phi_1 = 0.1$, $\phi_2 = 0.2$ and $\phi_3 = 0.3$

7. A pure MA(3) model with $\phi_1 = -0.1$, $\phi_2 = -0.2$ and $\phi_3 = -0.3$.

Five different GARCH models are considered. Their formulations are as follows:

1. ARCH(1) with $\alpha_0 = 2$ and $\alpha_1 = 0.2$.
2. ARCH(2) with $\alpha_0 = 2$, $\alpha_1 = 0.2$ and $\alpha_2 = 0.2$.
3. GARCH(1,1) with $\alpha_0 = 2$, $\alpha_1 = 0.2$ and $\beta_1 = 0.25$.
4. GARCH(2,1) with $\alpha_0 = 2$, $\alpha_1 = 0.2$, $\alpha_2 = 0.2$ and $\beta_1 = 0.25$.
5. GARCH(2,2) with $\alpha_0 = 2$, $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, $\beta_1 = 0.25$ and $\beta_2 = 0.25$

In the following section we investigate the properties of the different information criteria when the true models are stationary ARMA models consisting of 1000 observations and the residual variance of the time series is assumed to be normally distributed and equal to 1. The results from the 1000 replications are displayed in tables 5 through 11 below.

MODEL AR(1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)		**152**	**801**		**888**	**801**
ar(2)						
ar(3)						
ma(1)						
ma(2)						
ma(3)						
ar(1) ar(2)		13	4		1	
ar(1) ar(3)		28	20	20	50	
ar(1) ma(1)		11	5		2	22
ar(1) ma(2)		18	7		2	46
ar(1) ma(3)		27	17		6	125
ar(2) ar(3)						
ar(2) ma(1)		93	52		26	
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)						
ar(3) ma(2)						
ar(3) ma(3)						
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)						
ar(1) ar(2) ar(3)				4		
ar(1) ar(2) ma(1)	2	38	10	1		
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		3				
ar(1) ar(3) ma(1)		2		5		
ar(1) ar(3) ma(2)		2		10		
ar(1) ar(3) ma(3)		11	2	64	3	
ar(1) ma(1) ma(2)		1				
ar(1) ma(1) ma(3)		4				2
ar(1) ma(2) ma(3)		3				4
ar(2) ar(3) ma(1)	1	22	2	169	3	
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)						
ar(2) ma(1) ma(2)		16	3			
ar(2) ma(1) ma(3)		16	1		1	
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		60	9	160	14	
ar(3) ma(1) ma(3)						
ar(3) ma(2) ma(3)						
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)	9	17	3	19		
ar(1) ar(2) ar(3) ma(2)	32	72	8	78	4	
ar(1) ar(2) ar(3) ma(3)		1		14		
ar(1) ar(2) ma(1) ma(2)	1	6	2	1		
ar(1) ar(2) ma(1) ma(3)	7	24	6	3		
ar(1) ar(2) ma(2) ma(3)		2		1		
ar(1) ar(3) ma(1) ma(2)	12	26	8	16		
ar(1) ar(3) ma(1) ma(3)		1		10		
ar(1) ar(3) ma(2) ma(3)				46		
ar(1) ma(1) ma(2) ma(3)				1		
ar(2) ar(3) ma(1) ma(2)	8	22	6	18		
ar(2) ar(3) ma(1) ma(3)	10	11		72		
ar(2) ar(3) ma(2) ma(3)						
ar(2) ma(1) ma(2) ma(3)		8				
ar(3) ma(1) ma(2) ma(3)	8	13		105		
ar(1) ar(2) ar(3) ma(1) ma(2)	136	115	16	76		
ar(1) ar(2) ar(3) ma(1) ma(3)	35	6		3		
ar(1) ar(2) ar(3) ma(2) ma(3)	139	74	8	35		
ar(1) ar(2) ma(1) ma(2) ma(3)	3	3				
ar(1) ar(3) ma(1) ma(2) ma(3)	37	11	5	20		
ar(2) ar(3) ma(1) ma(2) ma(3)	48	15		16		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	512	53	5	33		

Table 5: The AR(1) Simulation Results ($n = 250, \sigma^2 = 1$)

MODEL-MA(1)	LOGL	AIC	SCHARZ	ICOMP (IFIM)	ICOMP (VEMD)	COMP
ar(1)						1000
ar(2)						
ar(3)						
ma(1)		**149**	**833**		**905**	
ma(2)						
ma(3)						
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)		13	13		5	
ar(1) ma(2)		43	19		12	
ar(1) ma(3)						
ar(2) ar(3)						
ar(2) ma(1)		27	13		5	
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)		36	15	43	58	
ar(3) ma(2)						
ar(3) ma(3)						
ma(1) ma(2)		12	5		3	
ma(1) ma(3)		17	13			
ma(2) ma(3)						
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)		1				
ar(1) ar(2) ma(2)		10				
ar(1) ar(2) ma(3)		11	1			
ar(1) ar(3) ma(1)		2	1	45	1	
ar(1) ar(3) ma(2)		19	3	156	5	
ar(1) ar(3) ma(3)						
ar(1) ma(1) ma(2)	2	31	7	2		
ar(1) ma(1) ma(3)		4				
ar(1) ma(2) ma(3)		12	3	1		
ar(2) ar(3) ma(1)		7		96	3	
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)						
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)						
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		5		17	1	
ar(3) ma(1) ma(3)		3	1	37	1	
ar(3) ma(2) ma(3)						
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)	1	3		46		
ar(1) ar(2) ar(3) ma(2)	1	4		14	1	
ar(1) ar(2) ar(3) ma(3)	10	11		164		
ar(1) ar(2) ma(1) ma(2)	3	6	1			
ar(1) ar(2) ma(1) ma(3)	16	72	7	12		
ar(1) ar(2) ma(2) ma(3)	16	69	16	10		
ar(1) ar(3) ma(1) ma(2)	10	22	6	32		
ar(1) ar(3) ma(1) ma(3)	1	3	1	4		
ar(1) ar(3) ma(2) ma(3)	23	17	2	62		
ar(1) ma(1) ma(2) ma(3)	3	4				
ar(2) ar(3) ma(1) ma(2)				25		
ar(2) ar(3) ma(1) ma(3)				12		
ar(2) ar(3) ma(2) ma(3)						
ar(2) ma(1) ma(2) ma(3)	2	16	1	2		
ar(3) ma(1) ma(2) ma(3)				17		
ar(1) ar(2) ar(3) ma(1) ma(2)	21	8		2		
ar(1) ar(2) ar(3) ma(1) ma(3)	66	32	3	31		
ar(1) ar(2) ar(3) ma(2) ma(3)	92	34	2	39		
ar(1) ar(2) ma(1) ma(2) ma(3)	128	169	17	33		
ar(1) ar(3) ma(1) ma(2) ma(3)	27	7		1		
ar(2) ar(3) ma(1) ma(2) ma(3)	42	15	2	39		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	536	106	15	58		

Table 6: The MA(1) Simulation Results $(n = 250, \sigma^2 = 1)$

MODEL AR(3)	LOGL	AIC	SCHARZ	ICOMP (IFIM)	ICOMP (VEMD)	COMP
ar(1)						
ar(2)						
ar(3)		**462**	**924**	**2**	**972**	**818**
ma(1)						
ma(2)						
ma(3)						
ar(1) ar(2)						
ar(1) ar(3)		71	10	1	5	
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)						
ar(2) ar(3)		70	12	5	2	
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)		60	11	9	4	70
ar(3) ma(2)		50	13	5	5	66
ar(3) ma(3)		97	22	8	8	30
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)						
ar(1) ar(2) ar(3)		12	2	28	1	
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)						
ar(1) ar(3) ma(1)		25	1	10		
ar(1) ar(3) ma(2)		5		31		
ar(1) ar(3) ma(3)		13		4		
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)						
ar(1) ma(2) ma(3)						
ar(2) ar(3) ma(1)		10	1	35	1	
ar(2) ar(3) ma(2)		31	3	5	1	
ar(2) ar(3) ma(3)		14		1		
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)						
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		12		81		13
ar(3) ma(1) ma(3)		13	1	34	1	2
ar(3) ma(2) ma(3)		13		27		1
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)		2		12		
ar(1) ar(2) ar(3) ma(2)		4		19		
ar(1) ar(2) ar(3) ma(3)		1		12		
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)						
ar(1) ar(2) ma(2) ma(3)						
ar(1) ar(3) ma(1) ma(2)		6		58		
ar(1) ar(3) ma(1) ma(3)		3		18		
ar(1) ar(3) ma(2) ma(3)		2		36		
ar(1) ma(1) ma(2) ma(3)						
ar(2) ar(3) ma(1) ma(2)		1		47		
ar(2) ar(3) ma(1) ma(3)		1		45		
ar(2) ar(3) ma(2) ma(3)		7		14		
ar(2) ma(1) ma(2) ma(3)						
ar(3) ma(1) ma(2) ma(3)		2		124		
ar(1) ar(2) ar(3) ma(1) ma(2)		3		29		
ar(1) ar(2) ar(3) ma(1) ma(3)				30		
ar(1) ar(2) ar(3) ma(2) ma(3)		4		33		
ar(1) ar(2) ma(1) ma(2) ma(3)						
ar(1) ar(3) ma(1) ma(2) ma(3)		1		99		
ar(2) ar(3) ma(1) ma(2) ma(3)	1	3		104		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	999	2		34		

Table 8: The AR(3) Simulation Results* $(n = 250, \sigma^2 = 1)$

MODEL MA(3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						
ar(2)						
ar(3)						998
ma(1)						
ma(2)						
ma(3)		**375**	**875**		**912**	
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)		47	13		6	1
ar(2) ar(3)						
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)		52	11		7	
ar(3) ma(1)						
ar(3) ma(2)						1
ar(3) ma(3)		94	35	7	61	
ma(1) ma(2)						
ma(1) ma(3)		95	31		4	
ma(2) ma(3)		67	21		6	
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		10		1		
ar(1) ar(3) ma(1)						
ar(1) ar(3) ma(2)						
ar(1) ar(3) ma(3)		13		32		
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)		31	1	1	1	
ar(1) ma(2) ma(3)		11	1	2		
ar(2) ar(3) ma(1)						
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)		10		23	1	
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)		14		1	1	
ar(2) ma(2) ma(3)		31	1			
ar(3) ma(1) ma(2)						
ar(3) ma(1) ma(3)		33	2	12		
ar(3) ma(2) ma(3)	1	34	4	7	1	
ma(1) ma(2) ma(3)		18	1			
ar(1) ar(2) ar(3) ma(1)						
ar(1) ar(2) ar(3) ma(2)						
ar(1) ar(2) ar(3) ma(3)	2	1		275		
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)	1	2	1			
ar(1) ar(2) ma(2) ma(3)		2		1		
ar(1) ar(3) ma(1) ma(2)						
ar(1) ar(3) ma(1) ma(3)		3		30		
ar(1) ar(3) ma(2) ma(3)		3		54		
ar(1) ma(1) ma(2) ma(3)		5	1	1		
ar(2) ar(3) ma(1) ma(2)						
ar(2) ar(3) ma(1) ma(3)	1	1		80		
ar(2) ar(3) ma(2) ma(3)	2	10		20		
ar(2) ma(1) ma(2) ma(3)		9		2		
ar(3) ma(1) ma(2) ma(3)	2	6	1	21		
ar(1) ar(2) ar(3) ma(1) ma(2)						
ar(1) ar(2) ar(3) ma(1) ma(3)	23	1		131		
ar(1) ar(2) ar(3) ma(2) ma(3)	19	3		159		
ar(1) ar(2) ma(1) ma(2) ma(3)	1	8		2		
ar(1) ar(3) ma(1) ma(2) ma(3)	18	1	1	29		
ar(2) ar(3) ma(1) ma(2) ma(3)	25	7		40		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	905	3		69		

Table 9: The MA(3) Simulation results* $(n = 250, \sigma^2 = 1)$

MODEL-AR(1,2,3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						8
ar(2)			1		1	53
ar(3)		2	48		98	435
ma(1)						
ma(2)						12
ma(3)			21		6	264
ar(1) ar(2)						
ar(1) ar(3)		4	31		38	
ar(1) ma(1)		1	46		14	
ar(1) ma(2)						
ar(1) ma(3)		3	15		9	11
ar(2) ar(3)		98	352	2	459	
ar(2) ma(1)						
ar(2) ma(2)			2		1	
ar(2) ma(3)		48	111		32	129
ar(3) ma(1)			2		5	
ar(3) ma(2)		33	71	5	109	87
ar(3) ma(3)			5	1	4	
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)		7	17		11	
ar(1) ar(2) ar(3)		**50**	**23**	**180**	**41**	
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)		1				
ar(1) ar(2) ma(3)		38	22	8	3	
ar(1) ar(3) ma(1)		3	5	1	9	
ar(1) ar(3) ma(2)		41	27	119	46	1
ar(1) ar(3) ma(3)	1	4		1		
ar(1) ma(1) ma(2)		19	6	2	1	
ar(1) ma(1) ma(3)		30	30	3	7	
ar(1) ma(2) ma(3)		8	2	5		
ar(2) ar(3) ma(1)		67	38	190	49	
ar(2) ar(3) ma(2)		14	3	6	8	
ar(2) ar(3) ma(3)		19	6	8	3	
ar(2) ma(1) ma(2)	1	4	5			
ar(2) ma(1) ma(3)		23	11	2	1	
ar(2) ma(2) ma(3)		59	34	4	8	
ar(3) ma(1) ma(2)		26	18	58	19	
ar(3) ma(1) ma(3)		1	1		2	
ar(3) ma(2) ma(3)		12	6	3	5	
ma(1) ma(2) ma(3)		4	2	3	1	
ar(1) ar(2) ar(3) ma(1)		16		15	1	
ar(1) ar(2) ar(3) ma(2)		3	1	3		
ar(1) ar(2) ar(3) ma(3)		1		15		
ar(1) ar(2) ma(1) ma(2)	2	2				
ar(1) ar(2) ma(1) ma(3)		23		2		
ar(1) ar(2) ma(2) ma(3)	1	37	2	2		
ar(1) ar(3) ma(1) ma(2)	1	14		8	1	
ar(1) ar(3) ma(1) ma(3)	11	18	7	10	3	
ar(1) ar(3) ma(2) ma(3)		15		19		
ar(1) ma(1) ma(2) ma(3)		10		9		
ar(2) ar(3) ma(1) ma(2)		6	1	9	3	
ar(2) ar(3) ma(1) ma(3)		4		9		
ar(2) ar(3) ma(2) ma(3)	5	13	4	17		
ar(2) ma(1) ma(2) ma(3)	1	38	7	3		
ar(3) ma(1) ma(2) ma(3)		25		28	2	
ar(1) ar(2) ar(3) ma(1) ma(2)	18	15	1	24		
ar(1) ar(2) ar(3) ma(1) ma(3)	12	33	3	38		
ar(1) ar(2) ar(3) ma(2) ma(3)	32	34	3	71		
ar(1) ar(2) ma(1) ma(2) ma(3)	5	5				
ar(1) ar(3) ma(1) ma(2) ma(3)	19	13	3	26		
ar(2) ar(3) ma(1) ma(2) ma(3)	25	15	4	34		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	866	41	3	57		

Table 10: The pure AR(3) Simualtion Results $(n = 250, \sigma^2 = 1)$

MODEL-MA(1,2,3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)			3		5	28
ar(2)		1	2		2	58
ar(3)			18		99	282
ma(1)						5
ma(2)			5		8	101
ma(3)		3	94		59	474
ar(1) ar(2)			3		3	
ar(1) ar(3)			3		18	
ar(1) ma(1)			2		1	
ar(1) ma(2)						
ar(1) ma(3)		8	32		20	2
ar(2) ar(3)		7	10	1	32	
ar(2) ma(1)						
ar(2) ma(2)		1				
ar(2) ma(3)		50	90	3	50	24
ar(3) ma(1)			2		7	
ar(3) ma(2)		23	50	5	197	20
ar(3) ma(3)		1	10		13	
ma(1) ma(2)						
ma(1) ma(3)		3	21		11	
ma(2) ma(3)		133	362	2	223	
ar(1) ar(2) ar(3)				10	1	
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		18	8	9	2	
ar(1) ar(3) ma(1)	6	7	10			
ar(1) ar(3) ma(2)		7	5	69	13	6
ar(1) ar(3) ma(3)		3	2	1	11	
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)		4	2		1	
ar(1) ma(2) ma(3)		68	45	37	25	
ar(2) ar(3) ma(1)		1		25	3	
ar(2) ar(3) ma(2)		9	2	3	17	
ar(2) ar(3) ma(3)		19	11	11	20	
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)		53	31	24	16	
ar(2) ma(2) ma(3)		47	28		9	
ar(3) ma(1) ma(2)		8	6	101	26	
ar(3) ma(1) ma(3)		2	3			
ar(3) ma(2) ma(3)		23	18	6	31	
ma(1) ma(2) ma(3)		**126**	**73**	**67**	**43**	
ar(1) ar(2) ar(3) ma(1)		5	1	7	2	
ar(1) ar(2) ar(3) ma(2)	1	12	3	26	5	
ar(1) ar(2) ar(3) ma(3)	1	4		60	2	
ar(1) ar(2) ma(1) ma(2)	1	7	2	1		
ar(1) ar(2) ma(1) ma(3)	1	22	4	2		
ar(1) ar(2) ma(2) ma(3)		22	4			
ar(1) ar(3) ma(1) ma(2)		14	3	15	2	
ar(1) ar(3) ma(1) ma(3)	12	19	7	9	1	
ar(1) ar(3) ma(2) ma(3)	1	5		53	1	
ar(1) ma(1) ma(2) ma(3)		22	1		1	
ar(2) ar(3) ma(1) ma(2)		11	2	10	1	
ar(2) ar(3) ma(1) ma(3)	3	23	5	59	4	
ar(2) ar(3) ma(2) ma(3)	1	20	2	22	2	
ar(2) ma(1) ma(2) ma(3)		15	1			
ar(3) ma(1) ma(2) ma(3)		29	2	52	9	
ar(1) ar(2) ar(3) ma(1) ma(2)	39	31	2	95	2	
ar(1) ar(2) ar(3) ma(1) ma(3)	44	20	2	27		
ar(1) ar(2) ar(3) ma(2) ma(3)	26	23	1	68		
ar(1) ar(2) ma(1) ma(2) ma(3)	3	13	1			
ar(1) ar(3) ma(1) ma(2) ma(3)	16	25	3	52	1	
ar(2) ar(3) ma(1) ma(2) ma(3)	18	17	2	33	1	
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	827	16	1	35		

Table 11: The pure MA(3) Simualtion Results $(n = 250, \sigma^2 = 1)$

6.3.1 ARMA Simulation Results - $\sigma^2 = 1$

In this section we investigate the properties of the different information criteria when the true models are stationary ARMA models consisting of 1000 observations and the residual variance of the time series is assumed to be normally distributed and equal to 1. The results from the 1000 replications are displayed in tables 5 through 11 above.

The main conclusions arising from the above results when the true model is pure AR(1) model with $\phi_1 = 0.8$ are as follows:

1. BIC and ICOMP(Van Emden) selects the AR(1) model more often than AIC with the frequency counts being 801 and 888 respectively.
2. ICOMP(IFIM) never selects the correct model. Notice however that this criterion selects an AR(3)-MA(1)-MA(2), 160 times and an AR(3)-MA(1)-MA(2)-MA(3), 105 times. The AR(1) model can be represented as an infinite MA process. These two models would be roughly equivelent to an AR(1) if the AR component and the MA terms greater than two two is not significantly different from zero.
3. $2C_1\,(F^{-1})$ selects the correct model 801 out of 1000 times.
4. AIC selects the correct model only 152 times out of 1000.
5. The criterion based on maximising the log likelihood function never selects the correct model. It overfits the data more often than the other information criteria since models with more parameters are preferred to models containing few parameters. This can be seen since the LOGL criterion selects the most complex model (AR(1) AR(2) AR(3) MA(1) MA(2) MA(3)) 512 times out of the thousands replications.

The main conclusions arising from the above results when the true model is a pure MA(1) with $\theta_1 = 0.8$ are as follows:

1. BIC and ICOMP(Van Emden) selects the MA(1) model more often than AIC with the frequency counts being 833 and 905 respectively.
2. ICOMP(IFIM) and COMP never selects the correct model. ICOMP(IFIM) does however select the the AR(1)-AR(3)-MA(2), 156 times and the AR(1)-AR(2)-AR(3)-MA(3) 164 times.

6.3.1 ARMA Simulation Results - $\sigma^2 = 1$

In this section we investigate the properties of the different information criteria when the true models are stationary ARMA models consisting of 1000 observations and the residual variance of the time series is assumed to be normally distributed and equal to 1. The results from the 1000 replications are displayed in tables 5 through 11 above.

The main conclusions arising from the above results when the true model is pure AR(1) model with $\phi_1 = 0.8$ are as follows:

1. BIC and ICOMP(Van Emden) selects the AR(1) model more often than AIC with the frequency counts being 801 and 888 respectively.
2. ICOMP(IFIM) never selects the correct model. Notice however that this criterion selects an AR(3)-MA(1)-MA(2), 160 times and an AR(3)-MA(1)-MA(2)-MA(3), 105 times. The AR(1) model can be represented as an infinite MA process. These two models would be roughly equivelent to an AR(1) if the AR component and the MA terms greater than two two is not significantly different from zero.
3. $2C_1\,(F^{-1})$ selects the correct model 801 out of 1000 times.
4. AIC selects the correct model only 152 times out of 1000.
5. The criterion based on maximising the log likelihood function never selects the correct model. It overfits the data more often than the other information criteria since models with more parameters are preferred to models containing few parameters. This can be seen since the LOGL criterion selects the most complex model (AR(1) AR(2) AR(3) MA(1) MA(2) MA(3)) 512 times out of the thousands replications.

The main conclusions arising from the above results when the true model is a pure MA(1) with $\theta_1 = 0.8$ are as follows:

1. BIC and ICOMP(Van Emden) selects the MA(1) model more often than AIC with the frequency counts being 833 and 905 respectively.
2. ICOMP(IFIM) and COMP never selects the correct model. ICOMP(IFIM) does however select the the AR(1)-AR(3)-MA(2), 156 times and the AR(1)-AR(2)-AR(3)-MA(3) 164 times.

The main conclusions arising from the above results when the true model is a pure ARMA(1,1) with $\phi_1 = 0.8$ and $\theta_1 = 0.8$ are as follows:

1. BIC and ICOMP(Van Emden) selects the correct model more often than AIC with the frequency counts being 861 and 922 respectively.
2. COMP selects the correct model most often while LOGL and ICOMP(IFIM) never selects the correct model.

The main conclusions arising from the above results when the true model is a autoregressive model with only a significant third term and parameter $\phi_3 = 0.8$. are as follows:

1. BIC and ICOMP(Van Emden) selects the correct model more often than AIC with the frequency counts being 924 and 972 respectively.
2. LOGL selects the most complicated model 999 times out of 1000.
3. ICOMP(IFIM) selects the correct model twice where as COMP selects the correct model 818 times out of 1000.

The main conclusions arising from the above results when the true model is a moving average model with only a significant third term and parameter $\theta_3 = 0.8$. are as follows:

1. BIC and ICOMP(Van Emden) selects the correct model more often than AIC with the frequency counts being 875 and 912 respectively.
2. LOGL, COMP and ICOMP(IFIM) never selects the correct model.
3. ICOMP(IFIM) tend to overfit the time series.

The main conclusions arising from the above results when the true model is a pure AR(3) model with $\phi_1 = 0.1$, $\phi_2 = 0.2$ and $\phi_3 = 0.3$ are as follows:

1. LOGL and COMP never selects the correct model. LOGL selects the most complex model 866 times while COMP selects AR(3) 435 times.
2. ICOMP(IFIM) selects the correct model most often. Notice however that the criterion selects the correct model only 180 times.
3. BIC and ICOMP(VEMD) selects AR(2)-AR(3) most often with the frequency counts being 352 and 459 respectively indicating that ICOMP(VEMD) selects

the best approximating model (the best model excluding the correct model) to the correct model most often.

The main conclusions arising from the above results when the true model is a pure MA(3) model with $\phi_1 = -0.1$, $\phi_2 = -0.2$ and $\phi_3 = -0.3$. are as follows:

1. LOGL and COMP never selects the correct model. LOGL selects the most complex model 827 times while COMP selects MA(3) 474 times.
2. AIC selects the correct model most often. Notice however that the frequency of selecting the correct model is small (126 out of 1000).
3. BIC and ICOMP(VEMD) selects MA(2)-MA(3) most often with the frequency counts being 362 and 223 respectively indicating that BIC selects the best approximating model (the best model excluding the correct model) to the correct model most often.

The following conclusions can be made from the above simulation study:

1. None of the information criteria selected selected the true model most often for all of the models investigated.
2. When the ARMA model consists of one parameter to estimate, ICOMP(VEMD) selects the true model most often.
3. ICOMP(VEMD) selects the true model most often when the true model is ARMA(1,1).
4. ICOMP(IFIM) does not select the true model most often for all of the models considered, however it does select the corect model model often when the true model is a pure AR(3) model.
5. The frequency counts associated with BIC and ICOMP(VEMD) were similar for all of the models considered.
6. When the true model is a pure AR(3) model, BIC and ICOMP(VEMD) selects AR(2)-AR(3) most often with the frequency counts being 352 and 459 respectively indicating that ICOMP(VEMD) selects the best approximating model (the best model excluding the correct model) to the correct model most often.
7. When the true model is a pure MA(3) model, BIC and ICOMP(VEMD) selects MA(2)-MA(3) most often with the frequency counts being 362 and 223

respectively indicating that BIC selects the best approximating model (the best model excluding the correct model) to the correct model most often.

8. The maximum log likelihood criterion tends to overfit the data by selecting complex models more often.

6.3.2 ARMA Simulation Results - $\sigma^2 = 25$

In this section we investigate the properties of the different information criteria when the true models are stationary ARMA models consisting of 1000 observations and the residual variance of the time series is assumed to be normally distributed and equal to 25. The results from the 1000 replications are displayed in tables 55 through 60. in appendix 21 through 27. These tables can analysed similar to the $(n = 250, \sigma^2 = 1)$ case analysed previously. The results can be summarised in table 12 below. The table displays the seven models and the number of times that AIC, BIC, ICOMP(IFIM) and ICOMP(VEMD) selects the correct model and close approximations of the correct model.

Correct Model	Selected Model	AIC	BIC	ICOMP(IFIM)	ICOMP(VEMD)
ar(1)	ar(1)	195	943	873	972
	Total	**195**	**943**	**873**	**972**
ma(1)	ma(1)	210	948	2	792
	ar(3) ma(1)	52	9	889	189
	Total	**262**	**957**	**891**	**981**
ar(1) ma(1)	ar(3)	24	180	407	900
	ar(1) ma(1)	32	28	0	0
	ar(3) ma(3)	104	103	492	77
	Total	**160**	**311**	**899**	**977**
ar(3)	ar(3)	441	963	920	981
	Total	**441**	**963**	**920**	**981**
ma(3)	ma(3)	426	959	907	977
	Total	**426**	**959**	**907**	**977**
ar(1) ar(2) ar(3)	**ar(2) ar(3)**	12	200	94	236
	ar(1) ar(2) ar(3)	338	466	373	434
	ar(2) ar(3) ma(1)	207	205	223	206
	Total	**557**	**871**	**690**	**876**
ma(1) ma(2) ma(3)	**ma(2) ma(3)**	17	198	90	226
	ar(1) ma(2) ma(3)	200	230	232	240
	ma(1) ma(2) ma(3)	337	456	362	418
	Total	**554**	**884**	**684**	**884**

Table 12: ARMA Simulation Results - $\sigma^2 = 25$

The following conclusions can be made from the above simulation study:

1. None of the information criteria selected selected the true model most often for all of the models investigated.
2. ICOMP(IFIM) selects the correct model for all of the models considered more often than when the sample size is 250 and the residual variance is one.

3. AIC is not a good predictor of the correct model for all of the models considered since the criterion is unable to accurately select the correct time series model.

4. For all of the one variable models considered, BIC, ICOMP (IFIM) and ICOMP (VEMD) performs better (i.e. selects the correct model more often) than AIC and the maximum log likelihood criterion. BIC in turn performs better than ICOMP (IFIM), however ICOMP (VEMD) performs better than BIC when the true model considered is either an AR(1), a MA or a AR with a significant third lag.

5. The performance of the different information criteria are varied when the true model has more than one variable to estimate.

   (a) When the true model is an ARMA(1,1), BIC, ICOMP(IFIM) and ICOMP (VEMD) all perform poorly and selects the correct model less than 5% of the time. ICOMP (IFIM) and ICOMP(VEMD) behaves similarly and selects models containing an AR(3) term at least 89% of the time.

   (b) When the true model is a pure AR(3), BIC and ICOMP (VEMD) behaves similarly since the correct model is selected 466 and 434 times respectively, the AR(2)-AR(3) model is selected 200 and 236 times respectively and the AR(2)-AR(3)-MA(1) model is selected 205 and 206 times respectively.

   (c) When the true model is a pure MA(3), BIC and ICOMP (VEMD) behaves similarly since the correct model is selected 456 and 418 times respectively, the MA(2)-MA(3) model is selected 198 and 226 times respectively and the MA(2)-MA(3)-AR(1) model is selected 230 and 240 times respectively.

6.3.3 GARCH Simulation Results - $\sigma^2 = 1$

In this section we investigate the properties of the different information criteria when the true models are stationary GARCH models consisting of 1000 observations and the residual variance of the time series is assumed to be normally distributed and equal to 1. The simulation results are presented in table 13 below. Take note that the Fischer's information matrix is estimated by $\hat{F} = \begin{pmatrix} \hat{\Sigma}_{\hat{\theta}} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{T} \end{pmatrix}$ where $\hat{\Sigma}_{\hat{\theta}}$ is the estimated covariance matrix of the estimated coefficients of the variance equation)

The following conclusions can be made from the above simulation study:

1. None of the information criteria selected the true model most often for all of the models investigated.
2. BIC and ICOMP(VEMD) behaves similarly for all of the GARCH models considered and selected the correct model more often than the other information criteria when the true model was an ARCH(1), ARCH(2) and GARCH(1,1) process.
3. As the number of parameters increased, AIC performed relatively better than all of the information criteria investigated however this property was only observed for the GARCH(2,1) model.
4. The maximum loglikelihood criteria perfroms poorly and tends to select over-specified models
5. The performance of the different information criteria are varied when the volatiltiy models has more than one GARCH term.
   (a) When the true model is a GARCH(2,1) and a GARCH(2,2), none of the information criteria selects the correct model more than 20% of the time. Notice however that in these cases that the information criteria selects models that **approximates** the true model. In the GARCH(2,1) case, ICOMP(VEMD) selects the ARCH(2,0) model 545 times and BIC selects the model 493 times. ICOMP(IFIM) behaves similarly and selects the ARCH(3,0) model 456 times. In the GARCH(2,2) case, ICOMP(VEMD) selects the GARCH(1,1) model 686 times and BIC selects the model 791 times.

6.3.4 FINAL NOTE

From the above time series simulation results it can be seen that none of the information criteria considered selects the correct model under all of the different conditions. It is suggested that no single information criteria should be used independently of other information criteria. Any modelling decision should involve a careful examination of all of the different models proposed taking into consideration not only technical issues such as model fit, tests for significanct parameters and correlated errors but we should also consider the economic plausibility of any proposed model.

MODEL-ARCH(1,1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
1,0		355	980	796	980	1000
2,0		34	6	99	3	
3,0	6	23	1	42		
1,1		27	6	1	3	
1,2		30		11	1	
1,3	19	41	2	20	1	
2,1	7	115	2	8	2	
2,2	16	22		1		
2,3	99	28		2	1	
3,1	21	31	2	2	1	
3,2	219	176	1	11	3	
3,3	613	118		7	5	

MODEL-ARCH(2,0)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
1,0						869
2,0		543	970	865	972	131
3,0	1	58	5	52		
1,1			3		1	
1,2		4	1	3	3	
1,3		8	1	7	2	
2,1	1	66	16	13	7	
2,2	15	80		21		
2,3	79	50		24	3	
3,1	55	115	3	1		
3,2	127	49	1	6	4	
3,3	722	47		8	8	

MODEL-GARCH(2,1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
1,0						994
2,0		137	493	270	545	6
3,0		129	85	456	117	
1,1		36	255	32	184	
1,2	1	87	69	98	22	
1,3	13	62	7	33	1	
2,1		197	80	32	126	
2,2	8	47	3	35		
2,3	67	62		28	2	
3,1	17	81	3	8	1	
3,2	108	63	2	7	2	
3,3	786	99	3	1		

MODEL-GARCH(2,2)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
1,0						995
2,0			2	1	2	5
3,0		1		2	1	
1,1		185	791	239	686	
1,2		95	42	227	20	
1,3	5	69	4	125	2	
2,1		205	144	176	248	
2,2	9	108	12	87	35	
2,3	92	64	1	68	3	
3,1		70	4	63	2	
3,2	61	51		9		
3,3	833	171		5	1	

MODEL-GARCH(1,1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
1,0		2	73	16	76	1000
2,0		73	135	299	181	
3,0		39	3	205	5	
1,1		344	768	270	718	
1,2		35	11	103	13	
1,3	5	43	2	74	2	
2,1		35	4	6	1	
2,2	15	77		3		
2,3	93	43				
3,1	11	45	2	15		
3,2	94	69	1	4	1	
3,3	782	195	1	5	3	

Table 13: GARCH Simulation Results - $\sigma^2 = 1$

6.4 List of Appendices

Appendix 1: 1: Eviews Multicollinearity Code

Appendix 2: Eviews Multicollinearity Subroutines

Appendix 3: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix for various sample sizes

Appendix 4: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix for various sample sizes

Appendix 5: A Display of the different Scenarios considered

Appendix 6: Frequency of Selecting Different models $(\beta_{\max}, n = 50)$

Appendix 7: Frequency of Selecting Different models $(\beta_{\max}, n = 100)$

Appendix 8: Frequency of Selecting Different models $(\beta_{\max}, n = 1000)$

Appendix 9: Frequency with which Models selected agree with KL Decisions $(\beta_{\max}, n = 50)$

Appendix 10: Frequency with which Models selected agree with KL Decisions $(\beta_{\max}, n = 100)$

Appendix 11: Frequency with which Models selected agree with KL Decisions $(\beta_{\max}, n = 1000)$

Appendix 12: Frequency of Selecting Different models $(\beta_{\min}, n = 50)$

Appendix 13: Frequency of Selecting Different models $(\beta_{\min}, n = 100)$

Appendix 14: Frequency of Selecting Different models $(\beta_{\min}, n = 1000)$

Appendix 15: Frequency with which Models selected agree with KL Decisions $(\beta_{\min}, n = 50)$

Appendix 16: Frequency with which Models selected agree with KL Decisions $(\beta_{\min}, n = 100)$

Appendix 17: Frequency with which Models selected agree with KL Decisions $(\beta_{\min}, n = 1000)$

Appendix 18: Time series and ICOMP

Appendix 19: Eviews Time Series Simulation Code

Appendix 20: Eviews Time Series Simulation Subroutines

Appendix 21: The AR(1) Simulation Results $(n = 1000, \sigma^2 = 25)$

Appendix 22: The MA(1) Simulation Results $(n = 1000, \sigma^2 = 25)$

Appendix 23: The ARMA(1,1) Simulation Results $(n = 1000, \sigma^2 = 25)$

Appendix 24: The AR(3)* Simulation Results $(n = 1000, \sigma^2 = 25)$

Appendix1: Eviews Multicollinearity Code

The following program is used in order to undertake the regression simulation study. In the program comments are made by using a single exclamation mark. i.e " ' " The following program requires a few requirements before it will run properly. They are:

1. Store the program found in Appendix 2 of this chapter as c:\ allan\stats\sta500w \REGRESSION\ regressiondespSUB2.prg. You could change the location but you then should ensure that the first line in this program has the correct location of the file specified by the *include* command.
2. !n=total number of observations. either 50, 100 or 1000
3. !p=number of x variables excluding the y intercept. 5 was used in this case.
4. !runs= the number of simulation runs that is undertaken. 100 was used in this study.
5. !betavalue={0=betamin,1=betamax} The beta coefficient used.
6. !alpha values stipulates the collinearity levels. i.e. !alpha1={0.99, 0.9 or 0.7} and !alpha2={0.99, 0.9 or 0.3}
7. Create a blank workfile named *BOZSIM50* from which to run the simulation study in. Sample sizes of 100 and 1000 should also be run in a workfile with the same name. You should just change the sample size as required.
8. Create a new database and store it in the c drive (dbcreate db_name e.g dbcreate allan). This database should be the default database in use for Eviews since the program does not explicitly specify which database should be used.

```
'include file location \program filename
include c:\allan\stats\sta500w\REGRESSION\regressiondespSUB2.prg
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
'!n=total number of observations
'!p=number of x variables excluding the y intercept
'!runs= the number of simulation runs that is undertaken
'!betavalue={0=betamin,1=betamax}
'the !alpha values stipulates the collinearity levels
'!alpha1={0.99,0.9,0.7}
```

```
'!alpha2={0.99,0.9,0.3}
''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
'INPUTS
!numobs=115 'ridge=101 pc=4 ols=5 stand=5 = 115
!numOLS=5
!p=5
!n=50
!runs=100
!betavalue=1 ' 0 = betamin 1=betamax
!vare=5
!ALPHA1=.7
!ALPHA2=.3
!k=1+!p
''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
WORKFILE BOZDOGANRESULTS U 1 !numobs
CALL MAKESERIES
WORKFILE BOZSIM50
CALL MAKEMATRICES
WORKFILE BOZDOGANRESULTS
CALL MAKESERIES2
WORKFILE BOZSIM50
'GENERATE THE X MATRICES
CALL XMATRICES
CALL XSMATRICES
CALL BETAVECTORS
CALL SIMULATIONR
WORKFILE BOZDOGANRESULTS
CALL RESULT
```

Appendix 2: Eviews Multicollinearity Subroutines

SUBROUTINE BETAVECTORS

'The following subroutine generates the beta vectors used in the regression simulation study

```
%model=”xs1 xs2 xs3 xs4 xs5”
group xsgrp %model
matrix xsmat=@convert(xsgrp)
matrix xtxsmat=@transpose(xsmat)*xsmat
vector xtxsSing
matrix xtxsV
matrix xtxsU=@svd(xtxsmat,xtxsSing,xtxsV)
'to calculate the condition numbers SQUARED
”””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””
vector xsSing
matrix xsV
matrix xsU=@svd(xsmat,xsSing,xsV)
'note that the eigenvalues are arranged from small to big
'the eigenvectors associated with eigenvalues are in the columns
”””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””
sym xtxsmatsym
xtxsmatsym=xtxsmat
matrix eigenvector = @eigenvectors(xtxsmatsym)
vector eigenvalues = @eigenvalues(xtxsmatsym)
vector betamax=@columnextract(eigenvector,5)
vector betamin=@columnextract(eigenvector,1)
vector betaave=0.5*(betamin+betamax)
”””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””””
if !betavalue=0 then
    VECTOR BETA=BETAMIN
else
    VECTOR BETA=BETAMAX
endif
VECTOR BETAstd=beta/@sqrt(!vare)
```

```
    VECTOR(!p+1) BETAmseRIDGE
    BETAmseRIDGE(1)=0
    for !q=1 to !p
        BETAmseRIDGE(!q+1)=BETA(!q)
    next !q
ENDSUB

SUBROUTINE DELETES
'The following subroutine deletes unwanted items
FOR !A= 1 TO !numOLS-1
    DELETE zgrp{!a}
    DELETE zMAT{!a}
    DELETE olsgrp{!a}
    DELETE OLSmat{!a}
    DELETE BETAOLS{!A}
    DELETE COVOLS{!A}
    DELETE EMATOLS{!A}
    DELETE VAROLS{!A}
    DELETE BETASTAND{!A}
    DELETE COVSTAND{!A}
    DELETE EMATSTAND{!A}
    DELETE VARSTAND{!A}
    DELETE R{!A}
    DELETE ROLS{!A}
    DELETE RINV{!A}
    DELETE ROLSINV{!A}
NEXT !A
ENDSUB

SUBROUTINE ICOMPols
    regtab(1,5)="MODEL NAME"
    regtab(1,6)="LOGL"
    regtab(1,7)="AIC"
    regtab(1,8)="BIC"
```

regtab(1,9)=”C1”

regtab(1,10)=”V EMD”

regtab(1,11)=”TMSE”

’TMSE/TMSE(OLS FULL MODEL (USING STAND MATRIX))

regtab(1,12)=”R=mset(BIASED)/mset(OLS)”

regtab(1,13)=”C1(CORR)”

regtab(1,14)=”V EMD(CORR)”

if @det(Cov)>0 then

’C(1) BASED ON THE COVARIANCE MATRIX

”””””””’““““““““““““““““““““““““““““““““““““

scalar c1= 0.5*!pmodel*log(@trace(Cov)/ !pmodel) -0.5* log(@det(Cov))

regtab(1+!count,9)=-2*loglik+2*c1

C1covMols(!count)=-2*loglik+2*c1

’IFIMvanEMDEN BASED ON THE COVARIANCE MATRIX

’““”””””””””””’

scalar ve= 0.5*log(@trace(Cov)) - 0.5*log(@det(Cov))

VEcovMols(!count)=-2*loglik +2*ve

else

C1covMols(!count)=1000000

VEcovMols(!count)=1000000

endif

If !a=!numOLS then

C1covMols(!count)=1000000000000000000

endif

ENDSUB

SUBROUTINE ICOMPpcreg

if GEOEigeng{!b}>0 then

’ICOMPC(1)

scalar c1= 0.5*!p*log(ARITHEigeng{!b}/ GEOEigeng{!b})

C1covMpc(!count)=-2*loglik+2*c1

’ICOMPIFIMvanEMDEN

else

C1covMpc(!count)=1000000

endif

ENDSUB

SUBROUTINE ICOMPridge

if @det(Cov)>0 then

'ICOMPC(1) BASED ON THE COVARIANCE MATRIX

'""

scalar c1= 0.5*(!p+1)*log(@trace(Cov)/ (!p+1)) -0.5* log(@det(Cov))

regtab(1+!count,9)=-2*loglik+2*c1

C1covMridge(!count)=-2*loglik+2*c1

'ICOMPIFIMvanEMDEN BASED ON THE COVARIANCE MATRIX

'""

scalar ve= 0.5*log(@trace(Cov)) - 0.5*log(@det(Cov))

VEcovMridge(!count)=-2*loglik +2*ve

else

regtab(1+!count,9)=1000000

C1covMridge(!count)=1000000

VEcovMridge(!count)=1000000

endif

ENDSUB

SUBROUTINE ICOMPstand

if @det(Cov)>0 then

'C(1) BASED ON THE COVARIANCE MATRIX

'"""

scalar c1= 0.5*!pmodel*log(@trace(Cov)/ !pmodel) -0.5* log(@det(Cov))

C1covMstand(!count)=-2*loglik+2*c1

'ICOMPIFIMvanEMDEN BASED ON THE COVARIANCE MATRIX

'""

scalar ve= 0.5*log(@trace(Cov)) - 0.5*log(@det(Cov))

VEcovMstand(!count)=-2*loglik +2*ve

else

C1covMstand(!count)=1000000

VEcovMstand(!count)=1000000

```
    endif
    If !a=!numOLS then
        C1covMstand(!count)=1000000
    endif
ENDSUB

SUBROUTINE MAKEMATRICES
    FOR %0 OLS STAND RIDGE PC
        MATRIX(!numobs,1) LOGLM{%0}=-10000000000000000000
        MATRIX(!numobs,1) AICM{%0}=100000000000000000
        MATRIX(!numobs,1) SCM{%0}=100000000000000000
        MATRIX(!numobs,1) KLM{%0}=100000000000000000
        MATRIX(!numobs,1) C1covM{%0}=100000000000000000
        MATRIX(!numobs,1) VEcovM{%0}=100000000000000000
    NEXT %0
ENDSUB

SUBROUTINE MAKESERIES
    TABLE(200,10) RESULTS
    RESULTS(1,2)="LOGL"
    RESULTS(1,3)="AIC"
    RESULTS(1,4)="SC"
    RESULTS(1,5)="KL"
    RESULTS(1,6)="C1COV"
    RESULTS(1,15)="TMSE"
    RESULTS(1,17)="PC delete"
    RESULTS(1,18)="avePC-TMSE"
    RESULTS(1,19)="OLS TMSE"
    RESULTS(1,20)="avePC-RelEff"
    SERIES ORDER
    FOR !Z=1 TO !numobs
        ORDER(!Z)=!Z
    NEXT !z
    for %0 LOGL AIC SC KL C1cov VEcov
```

```
        for %1 ols stand ridge pc
        SERIES {%0}{%1}
        SERIES {%0}{%1}count=0
        SERIES {%0}{%1}count2=0
        next %1
    next %0
    series TMSEridgecount=0
    series TMSEridgecount2=0
    series TMSEcountpc=0
    series TMSEcount2pc=0
    matrix(!runs,1) zzcount
    matrix(!runs,1) zzcountpc
```

ENDSUB

SUBROUTINE MAKESERIES2

```
     scalar aveTMSE=0
    scalar aveTSTAND=0
    scalar aveTMSEpc=0
    scalar aveTSTANDpc=0
```

ENDSUB

SUBROUTINE OLS

```
     !count=!count+1
    regtab(1+!count,5)="OLS " + %modelOLS
    matrix Rols{!a}=@transpose(OLSmat{!a})*OLSmat{!a}
    matrix Rolsinv{!a}=@inverse(Rols{!a})
    matrix Betaols{!a}=Rolsinv{!a}*@transpose(OLSmat{!a})*ymat
    matrix Ematols{!a}=ymat-OLSmat{!a}*Betaols{!a}
    matrix Varols{!a}=@transpose(Ematols{!a})*Ematols{!a}/(!n-!pmodel)
    'the beta covariance matrix of the standardised betas
    matrix Covols{!a}=Varols{!a}(1,1)*Rolsinv{!a}
    'THE CORRELATION MATRIX OF THE BETA ESTIMATES
    '""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
    vector CovOLSdgv{!a}=@getmaindiagonal(CovOLS{!a})
```

```
matrix CovOLSdiag{!a}= @makediagonal(CovOLSdgv{!a})
matrix STDOLSdiag{!a}=@sqrt(@inverse(CovOLSdiag{!a}))
matrix CorrOLS{!a}=STDOLSdiag{!a}*CovOLS{!a}*STDOLSdiag{!a}
matrix Corr=CorrOLS{!a}
matrix Cov=Covols{!a}
if !a=!numOLS then
    matrix mseOLS=@TRACE((BetaOLS{!a}-BETA)*@transpose(BetaOLS{!a}-BETA))
    regtab(1+!count,11) = mseOLS(1,1)
    regtab(1+!count,12) = 1
endif
'using the MLE estimates of the variance
''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
scalar loglik=-0.5*!n*(log(2*@acos(-1)))-0.5*!n*log(Varols{!a}(1,1))-0.5*(!n-!pmodel)
regtab(1+!count,6) = loglik
regtab(1+!count,7) = -2*loglik+2*!pmodel
regtab(1+!count,8) = -2*loglik+!pmodel*log(!n)
LOGLMols(!COUNT)=loglik
AICMols(!COUNT)= -2*loglik+2*!pmodel
SCMols(!COUNT)= -2*loglik+!pmodel*log(!n)
MATRIX MEANDIFFo{!A}=truemean-OLSmat{!a}*Betaols{!a}
matrix Errorsqols{!a}=@transpose(MEANDIFFo{!A})*(MEANDIFFo{!A})
KLMols(!COUNT)=0.5*!n*log(!vare/Varols{!a}(1,1))+
0.5*!n*(Varols{!a}(1,1)/!vare-1)+0.5*Errorsqols{!a}(1,1)/!vare
If !a=!numOLS then
    LOGLMols(!COUNT)=-1000000000000000000
    AICMols(!COUNT)=1000000000000000000
    SCMols(!COUNT)=1000000000000000000
    KLMols(!COUNT)=1000000000000000000
endif
CALL ICOMPols
ENDSUB

SUBROUTINE MINPROG
for %0 ols stand ridge pc
```

```
    SORT KL{%0} 'SORT KL NOW CHOOSE THE SMALLEST KL
    KL{%0}count(1)=KL{%0}count(1)+1
    RESULTS(ORDER(1)+1,5)=KL{%0}count(1)
    scalar KLyes{%0}=ORDER(1)
    SORT ORDER
next %0
for %0 ols stand ridge pc
    SORT(D) LOGL{%0} 'SORT LOGL NOW CHOOSE THE LARGEST LOGL
    LOGL{%0}count(1)=LOGL{%0}count(1)+1
    RESULTS(ORDER(1)+1,2)=LOGL{%0}count(1)
    scalar LOGLyes{%0}=ORDER(1)
    if LOGLyes{%0}=KLyes{%0} then
        LOGL{%0}count2(1)=LOGL{%0}count2(1)+1
        RESULTS(ORDER(1)+1,7)=LOGL{%0}count2(1)
    endif
    SORT ORDER
next %0
for %0 ols stand ridge pc
    SORT AIC{%0} 'SORT LOGL NOW CHOOSE THE SMALLEST LOGL
    AIC{%0}count(1)=AIC{%0}count(1)+1
    RESULTS(ORDER(1)+1,3)=AIC{%0}count(1)
    scalar AICyes{%0}=ORDER(1)
    if AICyes{%0}=KLyes{%0} then
        AIC{%0}count2(1)=AIC{%0}count2(1)+1
        RESULTS(ORDER(1)+1,8)=AIC{%0}count2(1)
    endif
    SORT ORDER
next %0
for %0 ols stand ridge pc
    SORT SC{%0} 'SORT SC NOW CHOOSE THE SMALLEST SC
    SC{%0}count(1)=SC{%0}count(1)+1
    RESULTS(ORDER(1)+1,4)=SC{%0}count(1)
    scalar SCyes{%0}=ORDER(1)
```

```
    if SCyes{%0}=KLyes{%0} then
        SC{%0}count2(1)=SC{%0}count2(1)+1
        RESULTS(ORDER(1)+1,9)=SC{%0}count2(1)
    endif
    SORT ORDER
next %0
for %0 ols stand pc
    SORT C1cov{%0} 'SORT C1cov NOW CHOOSE THE SMALLEST C1cov
    C1cov{%0}count(1)=C1cov{%0}count(1)+1
    RESULTS(ORDER(1)+1,6)=C1cov{%0}count(1)
    scalar C1covyes{%0}=ORDER(1)
    if C1covyes{%0}=KLyes{%0} then
        C1cov{%0}count2(1)=C1cov{%0}count2(1)+1
        RESULTS(ORDER(1)+1,10)=C1cov{%0}count2(1)
    endif
    SORT ORDER
next %0
for %0 ridge
    SORT C1cov{%0} 'SORT C1cov NOW CHOOSE THE SMALLEST C1cov
    C1cov{%0}count(1)=C1cov{%0}count(1)+1
    scalar RIDGEyes{%0}=ORDER(1)
    matrix zzcount(!simulruns)=order(1)
    scalar aveTMSE=aveTMSE+SaveMSER(1)
    scalar aveTSTAND=aveTSTAND+SaveSTAND(1)
    RESULTS(2,13)=aveTMSE
    RESULTS(2,14)=aveTSTAND
    RESULTS(ORDER(1)+1,6)=C1cov{%0}count(1)
    scalar C1covyes{%0}=ORDER(1)
    if C1covyes{%0}=KLyes{%0} then
        C1cov{%0}count2(1)=C1cov{%0}count2(1)+1
        RESULTS(ORDER(1)+1,10)=C1cov{%0}count2(1)
    endif
    SORT ORDER
```

```
next %0
```

ENDSUB

SUBROUTINE PCREG

```
matrix(!numobs,1) aveMSERPC' the average total mean square error
aveMSERPC=1000000000
matrix(!numobs,1) aveSTANDpc' the average TMSE/TMSE(OLS=STAND VARIABLES)
aveSTANDpc=1000000000
matrix(!n,!p) U
vector DS
matrix(!p,!p) V
matrix U=@svd(Xsmat,DS,V)
matrix DSING=@makediagonal(DS)
matrix ZPCcomps=Xsmat*V
matrix PCBetadelta=@inverse(DSING*DSING)*@transpose(ZPCcomps)*ysmat
matrix PCBeta=V*PCBetadelta
for !b=1 to !p-1
    matrix(!p+1,1) BetaPCBACK{!b}
    matrix(!p+1,1+!p) BetaPCBc{!b}
    matrix(!p,!p) vbetaPCrest{!b}
    !count=1+!count
    regtab(1+!count,5)=" PC del("+@str(!b)+") "+%model
    !pmodel=!p-!b
    matrix PCBetadeltag{!b}=@subextract(PCBetadelta,1,1,!pmodel,1)
    matrix Vg{!b}=@subextract(V,1,1,!p,!pmodel)
    matrix Vs{!b}=@subextract(V,1,1+!pmodel,!p,!p)
    'the standardised betas
    matrix PCBetag{!b}=Vg{!b}*PCBetadeltag{!b}
    matrix DSING{!b}=@subextract(DSING,1,1,!pmodel,!pmodel)
    matrix Ematg{!b}=ysmat-Xsmat*PCBetag{!b}
    matrix VarPCBeta{!b}=@transpose(Ematg{!b})*Ematg{!b}/(!n-!p) 'the s squared es-
timate
    'the pc covariance matrix of the betas
matrix CovPCBetag{!b}=VarPCBeta{!b}(1,1)*Vg{!b}*
```

@inverse(DSING{!b}*DSING{!b})*@transpose(Vg{!b})

matrix MSETPCBetag{!b}=@trace((PCBetag{!b}-BETAstd)*

@transpose((PCBetag{!b}-BETAstd)))

'THE CORRELATION MATRIX OF THE BETA ESTIMATES

vector CovPCDGV{!b}=@getmaindiagonal(CovPCBetag{!b})

matrix CovPCDIAG{!b}= @makediagonal(CovPCDGV{!b})

matrix STDCovPCDIAG{!B}=@sqrt(@inverse(CovPCDIAG{!b}))

matrix CorrPCG{!B}=STDCovPCDIAG{!B}*CovPCBetag{!b}* STDCovPCDIAG{!B}

matrix Corr=CorrPCG{!B}

'ridge estimates and COV matrix back in terms of the ORIGINAL VARIABLES

'TRANSFORMING BACK

'beta estimates = BetaPCBACK{!b}

'covariance matrix = BetaPCBc{!b}

matrix BetaPCBACK{!b}(1,1)=@mean(y)

for !q=1 to !p

matrix BetaPCBACK{!b}(!q+1,1)=@stdev(y)*PCbetag{!b}(!q,1)/@stdev(x{!q})

matrix BetaPCBACK{!b}(1,1)=BetaPCBACK{!b}(1,1)-BetaPCBACK{!b}(!q+1,1)*@mean(x{!q})

next !q

for !w=1 to !p

for !r=1 to !p

matrix vbetaPCrest{!b}(!w,!r)=CovPCBetag{!b}(!w,!r)*@stdev(y)*

@stdev(y)/(@stdev(x{!w})*@stdev(x{!r}))

next !r

next !w

matrix EmatPCB{!b}=ymat-xmatINT*BetaPCBACK{!b}

matrix VarPCB{!b}=@transpose(EmatPCB{!b})*EmatPCB{!b}/(!n-!p-1)

MATRIX vbetaPC0{!b}=@transpose(xmean)*vbetaPCrest{!b}*xmean

vbetaPC0{!b}(1,1)=vbetaPC0{!b}(1,1)+ VarPCBeta{!b}(1,1)/!n

BetaPCBc{!b}(1,1)=vbetaPC0{!b}(1,1)

matplace(BetaPCBc{!b},vbetaPCrest{!b},2,2)

for !q=1 to !p

rowvector v{!q}betaPCrestR{!b} = @rowextract(vbetaPCrest{!b},!q)

matrix v{!q}betaPC0j{!b}=-1*v{!q}betaPCrestR{!b}*xmean

```
        BetaPCBc{!b}(!q+1,1)=v{!q}betaPC0j{!b}(1,1)
        BetaPCBc{!b}(1,!q+1)=v{!q}betaPC0j{!b}(1,1)
    next !q
    'TO CALCULATE THE SINGULAR VALUES OF CovPCBetag{!b} BetaPCBc{!b}
    sym CovPCBetagSYM{!b}
    CovPCBetagSYM{!b}=BetaPCBc{!b} 'CovPCBetag{!b}
    matrix EigenCovPCg{!b}=@eigenvalues(CovPCBetagSYM{!b})
    scalar ARITHEigeng{!b}=@mean(EigenCovPCg{!b})
    matrix(1,1) PRODpcg{!b}
    PRODpcg{!b}(1,1)=1
    for !q=1 to !p
        PRODpcg{!b}(1,1)=PRODpcg{!b}(1,1)*EigenCovPCg{!b}(!q,1)
    next !q
    scalar GEOEigeng{!b}=@abs(PRODpcg{!b}(1,1))^(1/!p)
    'THE MATRIX TO USE FOR THE ICOMP
    matrix Cov=BetaPCBc{!b} ' CovPCbetag{!b}
    scalar loglik=-0.5*!n*(log(2*@acos(-1)))-0.5*!n*log(VarPCB{!b}(1,1))-0.5*(!n-!p)
matrix msePC{!b}=@TRACE((BetaPCBACK{!b}-
BETAmseRIDGE)*@transpose(BetaPCBACK{!b}-BETAmseRIDGE))
    matrix aveMSERPC(!count)=msePC{!b}(1,1)
    matrix zcheckPC(!count,!simulruns)=msePC{!b}(1,1)
    matrix aveSTANDpc(!count)=mseRIDGE1(1,1)
    scalar aveTMSEPC{!B}=aveTMSEPC{!B}+aveMSERPC(!count)
    scalar aveTMSESTAND{!B}=aveTMSESTAND{!B}+aveSTANDpc(!count)
    LOGLMpc(!COUNT)=loglik
    AICMpc(!COUNT)= -2*loglik+2*!pmodel
    SCMpc(!COUNT)= -2*loglik+!pmodel*log(!n)
    MATRIX MEANDIFFp{!b}=truemeanSTD-Xsmat*PCBetag{!b}
    matrix Errorsqpcg{!b}=@transpose(MEANDIFFp{!b})*(MEANDIFFp{!b})
    KLMpc(!COUNT)= 0.5*!n*log(1/VarPCBeta{!b}(1,1))+
0.5*!n*( VarPCBeta{!b}(1,1)/1 -1)+0.5*Errorsqpcg{!b}(1,1)/1
    CALL ICOMPpcreg
next !b
```

```
ENDSUB


SUBROUTINE RESULT
'The following subroutine generates the output in a table
    results(1,7)="LOGL"
    results(1,8)="AIC"
    results(1,9)="SC"
    results(1,10)="C1cov"
    show results
    'for the ridge regression
    scalar aveTMSE=aveTMSE
    scalar aveRM=aveTMSE/aveTSTAND
    RESULTS(2,13)=aveTMSE
    RESULTS(2,14)=aveRM
    results(2,15)=aveTSTAND
    'for the principal components
   for !b=1 to !p-1
        RESULTS(1+!b,17)=!b
        RESULTS(1+!b,18)=aveTMSEPC{!B}
        RESULTS(1+!b,19)=aveTMSESTAND{!B}
        RESULTS(1+!b,20)=aveTMSEPC{!B}/aveTMSESTAND{!B}
   NEXT !b
ENDSUB


SUBROUTINE RIDGE
    !totRIDGEnum=101
   matrix(!numobs,1) aveMSER' the average total mean square error
   aveMSER=1000000000
   matrix(!numobs,1) aveSTAND' the average TMSE/TMSE(OLS=STAND VARIABLES)
   aveSTAND=1000000000
   for !k=1 to !totRIDGEnum
       matrix(!p+1,1) BetaridgeBACK{!k}
       matrix(!p+1,1+!p) BetaridgeBc{!k}
       matrix(!p,!p) vbetarest{!k}
```

```
!count=!count+1
!kridge=0+0.01*(!k-1)
regtab(1+!count,5)="RIDGE (k="+@str(!kridge)+") "+%model
if !k=101 then
    'THE HOERL KENNARD BALDWIN ESTIMATE OF K
    '""""""""""""""""""""""""""""""""""""""""""""""""""""""
    matrix khlden=@transpose(Betaridge1)*Betaridge1
    !kridge=!p*Varridge1(1,1)/khlden(1,1)
    regtab(1+!count,5)="RIDGE (Hoerl Kennard) "+%model
endif
matrix Ident{!k}=@identity(!kmodel-1)
matrix W{!k}=@inverse(R{!a}+!kridge*Ident{!k})
matrix Betaridge{!k}=W{!k}*R{!a}*Betastand{!a}
matrix Ematridge{!k}=ysmat-zmat{!a}*Betaridge{!k}
matrix Varridge{!k}=@transpose(Ematridge{!k})*Ematridge{!k}/(!n-!pmodel)
'the beta covariance matrix of the standardised betas
matrix Covridge{!k}=VarRIDGE{!k}(1,1)*W{!k}*R{!a}*W{!k}
'THE CORRELATION MATRIX OF THE BETA ESTIMATES
'""""""""""""""""""""""""""""""""""""""""""""""""""""""""
vector CovridgeDGV{!k}=@getmaindiagonal(Covridge{!k})
matrix CovridgeDIAG{!k}= @makediagonal(CovridgeDGV{!k})
matrix STDridgeDIAG{!k}=@sqrt(@inverse(CovridgeDIAG{!k}))
matrix Corrridge{!k}=STDridgeDIAG{!k}*Covridge{!k}*STDridgeDIAG{!k}
matrix Corr=CorrRidge{!k}
'ridge estimates and COVARIANCE matrix back in terms
'of the ORIGINAL VARIABLES
'TRANSFORMING BACK
'beta estimates = BetaridgeBACK{!k}
'covariance matrix = BetaridgeBc{!k}
matrix BetaridgeBACK{!k}(1,1)=@mean(y)
for !q=1 to !p
matrix BetaridgeBACK{!k}(!q+1,1)=@stdev(y)*Betaridge{!k}(!q,1)/@stdev(x{!q})
matrix BetaridgeBACK{!k}(1,1)=BetaridgeBACK{!k}(1,1)-
```

```
BetaridgeBACK{!k}(!q+1,1)*@mean(x{!q})
    next !q
        for !w=1 to !p
            for !r=1 to !p
    matrix vbetarest{!k}(!w,!r)=Covridge{!k}(!w,!r)*@stdev(y)*@stdev(y)
/(@stdev(x{!w})*@stdev(x{!r}))
            next !r
        next !w
        matrix EmatridgeB{!k}=ymat-xmatINT*BetaridgeBACK{!k}
        matrix VarridgeB{!k}=@transpose(EmatridgeB{!k})*EmatridgeB{!k}/(!n-!p-1)
        matrix vbeta0{!k}=@transpose(xmean)*vbetarest{!k}*xmean
        vbeta0{!k}(1,1)=vbeta0{!k}(1,1)+ VarridgeB{!k}(1,1)/!n
        BetaridgeBc{!k}(1,1)=vbeta0{!k}(1,1)
        matplace(BetaridgeBc{!k},vbetarest{!k},2,2)
        for !q=1 to !p
            rowvector v{!q}betarestR{!k} = @rowextract(vbetarest{!k},!q)
            matrix v{!q}beta0j{!k}=-1*v{!q}betarestR{!k}*xmean
            BetaridgeBc{!k}(!q+1,1)=v{!q}beta0j{!k}(1,1)
            BetaridgeBc{!k}(1,!q+1)=v{!q}beta0j{!k}(1,1)
        next !q
        'THE MATRIX TO USE FOR THE ICOMP
        '""""""""""""""""""""""""""""""""""""""""""
        matrix Cov=BetaridgeBc{!k}
        scalar loglik=-0.5*(!n*(log(2*@acos(-1)))-!n*log(VarRIDGEB{!k}(1,1))-(!n-!p-1))
        regtab(1+!count,6) = loglik
        regtab(1+!count,7) = -2*loglik+2*(!p+1)
        regtab(1+!count,8) = -2*loglik+(!p+1)*log(!n)
    matrix mseRIDGE{!k}=@TRACE((BetaridgeBACK{!k}-
BETAmseRIDGE)*@transpose(BetaridgeBACK{!k}-BETAmseRIDGE))
        regtab(1+!count,11) = mseRidge{!k}(1,1)
        matrix aveMSER(!count)=mseRIDGE{!k}(1,1)
        matrix zcheck(!count,!simulruns)=mseRIDGE{!k}(1,1)
        matrix zcheckstand(!count,!simulruns)=mseRIDGE1(1,1)
```

```
        matrix aveSTAND(!count)=aveMSER(1+2*!p)
        'MSETotal(ridge)/MSETotal(OLS=STAND VARS)
        '""""""""""""""""""""""""""""""""""""""""""""""""'
        regtab(1+!count,12) = mseRidge{!k}(1,1)/mseSTAND(1,1)
        LOGLMridge(!COUNT)=loglik
        AICMridge(!COUNT)= -2*loglik+2*(!p+1)
        SCMridge(!COUNT)= -2*loglik+(!p+1)*log(!n)
        MATRIX MEANDIFFr{!k}=truemeanSTD-zmat{!a}*Betaridge{!k}
        matrix Errorsqridge{!k}=@transpose(MEANDIFFr{!k})*(MEANDIFFr{!k})
    KLMridge(!COUNT)= 0.5*!n*log(1/Varridge{!k}(1,1))+
0.5*!n*( Varridge{!k}(1,1)/1 -1)+0.5*Errorsqridge{!k}(1,1)/1
        CALL ICOMPridge
    next !k
ENDSUB
```

SUBROUTINE SIMULATIONR

'The following subroutine undertakes most of the simulation aspects of the regression simulation study

```
    !countrun=0
    matrix(!numobs,!runs) zcheck
    matrix(!numobs,!runs) zcheckstand
    matrix(!numobs,!runs) zcheckPC
    matrix(!runs,1) zzcount
    matrix(!runs,1) zzcountPC
    FOR !B=1 TO !P-1
        scalar aveTMSEPC{!B}=0
        scalar aveTMSESTAND{!B}=0
    next !B
    for !simulruns=1 to !runs
        scalar arun=!simulruns
        !countrun=!countrun+1
        !count=0
        rndseed 7+!countrun
        series e=nrnd*@sqrt(!vare)
```

```
stom(e,emat)
vector ymat=emat+xmat*BETA
vector truemean=xmat*BETA
vector truemeanSTD=xsmat*BETAstd
mtos(ymat,y)
'standardise the y var
series ys=(y-@mean(y))/(@sqrt(!n-1)*@stdev(y))
stom(ys,ysmat)
for !a=1 to !numOLS
    'the stand model
    %model=regtab(1+!a,1)
    'the ols model
    %modelols=regtab(1+!a,3)
    !kmodel=1+@val(regtab(1+!a,2))
    !pmodel=!kmodel-1
    'THE NORMAL VARIABLES-NOT STAND
    group olsgrp{!a} %modelols
    matrix OLSmat{!a}=@convert(olsgrp{!a})
    'THE STAND VARIABLES
    group zgrp{!a} %model
    matrix zmat{!a}=@convert(zgrp{!a})
    CALL OLS
    CALL STAND
    if !a=!numOLS then
        CALL RIDGE
        CALL PCREG
    endif
next !a
'WANT TO FIND THE MINIMUM NOW
CALL TRY
CALL MINPROG
WORKFILE BOZSIM50
CALL DELETES
```

```
next !simulruns
ENDSUB

SUBROUTINE STAND
!count=!count+1
regtab(1+!count,5)="STAND "+ %model
matrix R{!a}=@transpose(zmat{!a})*zmat{!a}
matrix Rinv{!a}=@inverse(R{!a})
matrix Betastand{!a}=Rinv{!a}*@transpose(zmat{!a})*ysmat
matrix Ematstand{!a}=ysmat-zmat{!a}*Betastand{!a}
matrix Varstand{!a}=@transpose(Ematstand{!a})*Ematstand{!a}/(!n-!pmodel)
'the beta covariance matrix of the standardised betas
'""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
matrix Covstand{!a}=Varstand{!a}(1,1)*Rinv{!a}
'THE CORRELATION MATRIX OF THE BETA ESTIMATES
'""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
vector Covstanddgv{!a}=@getmaindiagonal(Covstand{!a})
matrix Covstanddiag{!a}= @makediagonal(Covstanddgv{!a})
matrix STDstanddiag{!a}=@sqrt(@inverse(Covstanddiag{!a}))
matrix Corrstand{!a}=STDstanddiag{!a}*Covstand{!a}*STDstanddiag{!a}
matrix Corr=Corrstand{!a}
matrix Cov=Covstand{!a}
'using the MLE estimates of the variance
scalar loglik=-0.5*!n*(log(2*@acos(-1)))-0.5*!n*log(Varstand{!a}(1,1))-0.5*(!n-!pmodel)
if !a=!numOLS then
    matrix mseSTAND=@TRACE((Betastand{!a}-BETAstd)*@transpose(Betastand{!a}-
BETAstd))
    regtab(1+!count,11) = mseSTAND(1,1)
    regtab(1+!count,12) = 1
endif
LOGLMstand(!COUNT)=loglik
AICMstand(!COUNT)= -2*loglik+2*!pmodel
SCMstand(!COUNT)= -2*loglik+!pmodel*log(!n)
MATRIX MEANDIFFs{!A}=truemeanSTD-zmat{!a}*Betastand{!a}
```

```
matrix Errorsqstand{!a}=@transpose(MEANDIFFs{!A})*(MEANDIFFs{!A})
KLMstand(!COUNT)= 0.5*!n*log(1/Varstand{!a}(1,1))+
0.5*!n*( Varstand{!a}(1,1)/1 -1)+0.5*Errorsqstand{!a}(1,1)/1
If !a=!numOLS then
    LOGLMstand(!COUNT)=-1000000000000000000
    AICMstand(!COUNT)=1000000000000000000
    SCMstand(!COUNT)=1000000000000000000
    KLMstand(!COUNT)=1000000000000000000
endif
CALL ICOMPstand
ENDSUB

SUBROUTINE TRY
FOR %0 OLS STAND RIDGE PC
    STORE LOGLM{%0}
    STORE AICM{%0}
    STORE SCM{%0}
    STORE KLM{%0}
    STORE C1covM{%0}
    STORE VEcovM{%0}
NEXT %0
for !b=1 to !p-1
    store aveTMSEPC{!B}
    store aveTMSESTAND{!B}
NEXT !b
STORE aveMSER
STORE aveSTAND
STORE ZCHECK
STORE ZCHECKSTAND
STORE ZCHECKPC
STORE AVESTANDPC
STORE aveMSERpc
STORE aveSTANDpc
WORKFILE BOZDOGANRESULTS
```

```
FOR %0 OLS STAND RIDGE PC
    FETCH LOGLM{%0}
    FETCH AICM{%0}
    FETCH SCM{%0}
    FETCH KLM{%0}
    FETCH C1covM{%0}
    FETCH VEcovM{%0}
    SERIES LOGL{%0}
    SERIES AIC{%0}
    SERIES SC{%0}
    SERIES KL{%0}
    SERIES C1cov{%0}
    SERIES VEcov{%0}
    MTOS(LOGLM{%0},LOGL{%0})
    MTOS(AICM{%0},AIC{%0})
    MTOS(SCM{%0},SC{%0})
    MTOS(KLM{%0},KL{%0})
    MTOS(C1covM{%0},C1cov{%0})
    MTOS(VEcovM{%0},VEcov{%0})
NEXT %0
for !b=1 to !p-1
    FETCH aveTMSEPC{!B}
    FETCH aveTMSESTAND{!B}
NEXT !b
'FOR THE RIDGE REGRESSION
"""""""""""""""""""""""""""""""""""""""""""""""""""
FETCH aveMSER
FETCH aveSTAND
series SaveMSER
series SaveSTAND
MTOS(aveMSER,SaveMSER)
MTOS(aveSTAND,SaveSTAND)
'for the PC
```

```
""""""""""""
FETCH aveMSERpc
FETCH aveSTANDpc
series SaveMSERpc
series SaveSTANDpc
MTOS(aveMSERpc,SaveMSERpc)
MTOS(aveSTANDpc,SaveSTANDpc)
FETCH ZCHECK
FETCH ZCHECKSTAND
FETCH ZCHECKPC
FETCH AVESTANDPC
RESULTS(1,11)="aveTMSE"
RESULTS(1,12)="aveRM"
RESULTS(1,13)="RidgeaveTMSE"
RESULTS(1,14)="RidgeaveRatio"
for !Q=1 to !totRIDGEnum
    results(11+!q,11)=aveMSER(10+!q)
    results(11+!q,12)=aveSTAND(10+!q)
next !Q
```

ENDSUB

SUBROUTINE XMATRICES

'the following subroutine generates the X MATRICES used in the regression simulation study

'See Bozdogan (1998),Wichern,Churchhil (1978), Mcdonald and Galarneau (1975)

```
matrix CORRS=@identity(6)
'cholesKy decomposition of CORRS
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
sym CORRSsym
CORRSsym=CORRS
matrix A=@cholesky(CORRSsym)
for !q=1 to 6
    series z{!q}
next
```

```
smpl 1 !p+1
'the original Z~N(0,1) variables
for !i=1 to !n
    rndseed !i
    series CholesY=nrnd
    stom(CholesY,CholesYmat)
    matrix CholesXmat=A*CholesYmat
    z1(!i)=CholesXmat(1)
    z2(!i)=CholesXmat(2)
    z3(!i)=CholesXmat(3)
    z4(!i)=CholesXmat(4)
    z5(!i)=CholesXmat(5)
    z6(!i)=CholesXmat(6)
next !i
smpl 1 !n
''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
For !z=1 to 3
    series x{!z}=@sqrt(1-!ALPHA1^2)*z{!z}+(!ALPHA1)*z6
next !z
for !z=4 to 5
    series x{!z}=@sqrt(1-!ALPHA2^2)*z{!z}+(!ALPHA2)*z6
next !z
series int=1
group g1 x1 x2 x3 x4 x5
matrix xmat=g1
group g2 INT x1 x2 x3 x4 x5
matrix xmatINT=g2
matrix xtxmat=@transpose(xmat)*xmat
matrix xtxinv=@inverse(xtxmat)
scalar q=@det(xtxinv)
```

ENDSUB

SUBROUTINE XSMATRICES

'the following subroutine standardises the X MATRICES used in the regression simulation study

```
matrix(!p,1) xmean
matrix(!p,1) xstd
for !a=1 to !p
    series xs{!a}=(x{!a}-@mean(x{!a}))/(@sqrt(!n-1)*@stdev(x{!a}))
    xmean(!a,1)=@mean(x{!a})
    xstd(!a,1)=@stdev(x{!a})
next !a
```

ENDSUB

Appendix 3: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix for various sample sizes

Table 14: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 50, \beta_{\max})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} 0.4479 & 0.4474 & 0.4473 & 0.4477 & 0.4456 \end{pmatrix}$	438
Moderate	$\begin{pmatrix} 0.4540 & 0.4490 & 0.4482 & 0.4532 & 0.4312 \end{pmatrix}$	38
Low	$\begin{pmatrix} 0.5300 & 0.5327 & 0.4925 & 0.3821 & 0.2165 \end{pmatrix}$	7

Table 15: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 100, \beta_{\max})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} 0.4479 & 0.4471 & 0.4474 & 0.4469 & 0.4467 \end{pmatrix}$	404
Moderate	$\begin{pmatrix} 0.4556 & 0.4440 & 0.4487 & 0.4451 & 0.4426 \end{pmatrix}$	37
Low	$\begin{pmatrix} 0.5497 & 0.4964 & 0.5015 & 0.3237 & 0.3085 \end{pmatrix}$	7

Table 16: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 1000, \beta_{\max})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} 0.4472 & 0.4472 & 0.4473 & 0.4472 & 0.4472 \end{pmatrix}$	271
Moderate	$\begin{pmatrix} 0.4570 & 0.4462 & 0.4478 & 0.4473 & 0.4479 \end{pmatrix}$	25
Low	$\begin{pmatrix} 0.5163 & 0.5160 & 0.5135 & 0.3139 & 0.3239 \end{pmatrix}$	5

Appendix 4: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix for various sample sizes

Table 17: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 50, \beta_{\min})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} -0.5543 & -0.4386 & 0.3382 & 0.6203 & 0.0350 \end{pmatrix}$	438
Moderate	$\begin{pmatrix} -0.5102 & -0.4639 & 0.2968 & 0.6604 & 0.0175 \end{pmatrix}$	38
Low	$\begin{pmatrix} -0.4180 & -0.4917 & 0.6895 & 0.30 & 0.135 \end{pmatrix}$	7

Table 18: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 100, \beta_{\min})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} 0.8775 & -0.2203 & -0.0714 & -0.2513 & -0.3366 \end{pmatrix}$	404
Moderate	$\begin{pmatrix} 0.8715 & -0.1805 & -0.0863 & -0.270 & -0.3576 \end{pmatrix}$	37
Low	$\begin{pmatrix} 0.8174 & -0.1683 & -0.4588 & -0.1985 & -0.2317 \end{pmatrix}$	7

Table 19: Eigenvalues of $(X'X)$ and the Square of the Condition numbers of the correlation matrix $(n = 1000, \beta_{\min})$

Collinearity Level	*Beta Vector*	λ_1/λ_5
High	$\begin{pmatrix} -0.3038 & -0.4360 & 0.6111 & 0.4741 & -0.3454 \end{pmatrix}$	271
Moderate	$\begin{pmatrix} -0.2631 & -0.3965 & 0.5806 & 0.5045 & -0.4266 \end{pmatrix}$	25
Low	$\begin{pmatrix} -0.30 & -0.5401 & 0.7711 & 0.1525 & -0.0334 \end{pmatrix}$	5

Appendix 5: A Display of the different Scenarios considered

BETA MIN

ALPHAS	(0.99,0.99)		
	n		
sigma	50	100	1000
0.25	1	10	19
2.5	2	11	20
5	3	12	21

BETA MAX

ALPHAS	(0.99,0.99)		
	n		
sigma	50	100	1000
0.25	28	37	**46**
2.5	29	38	**47**
5	30	39	**48**

ALPHAS	(0.9,0.9)		
	n		
sigma	50	100	1000
0.25	**4**	13	22
2.5	5	14	23
5	6	15	24

ALPHAS	(0.9,0.9)		
	n		
sigma	50	100	1000
0.25	31	40	**49**
2.5	32	**41**	**50**
5	33	42	**51**

ALPHAS	(0.7,0.3)		
	n		
sigma	50	100	1000
0.25	7	16	25
2.5	8	17	26
5	9	18	27

ALPHAS	(0.7,0.3)		
	n		
sigma	50	100	1000
0.25	34	**43**	**52**
2.5	35	**44**	**53**
5	36	**45**	**54**

Model 1: $\sigma^2 = 0.25$, $n = 50$, $\beta_{\min}$, $\alpha_1 = \alpha_2 = 0.99$

Appendix 6: Frequency of Selecting Different models ($\beta_{\max}, n = 50$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25	3	24	37	36
	2.5	59	19	10	12
	5	65	14	11	10
AIC	0.25	17	40	28	15
	2.5	83	12	2	3
	5	89	8	2	1
BIC	0.25		5	18	77
	2.5	31	43	20	6
	5	50	34	13	3
C1 COV	0.25	3	2	24	71
	2.5	65	3	13	19
	5	66	2	14	18

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25			2	98
	2.5	6	28	32	34
	5	27	26	23	24
AIC	0.25			6	94
	2.5	25	38	24	13
	5	52	30	10	8
BIC	0.25				100
	2.5	1	6	19	74
	5	2	19	28	51
C1 COV	0.25				100
	2.5		4	24	72
	5	8	9	26	57

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25			1	99
	2.5	2	16	28	54
	5	12	30	24	34
AIC	0.25			1	99
	2.5	7	42	22	29
	5	35	38	13	14
BIC	0.25				100
	2.5		2	9	89
	5	1	10	24	65
C1 COV	0.25				100
	2.5		1	6	93
	5		3	21	76

Tables 20-22

Appendix 7: Frequency of Selecting Different models ($\beta_{\max}, n = 100$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25		8	25	67
	2.5	52	21	13	14
	5	70	15	7	8
AIC	0.25	6	23	37	34
	2.5	87	10	2	1
	5	89	8	2	1
BIC	0.25			6	94
	2.5	18	36	21	25
	5	29	44	16	11
C1 COV	0.25	1	1	6	92
	2.5	52	5	9	34
	5	68	4	6	22

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5	1	8	27	64
	5	11	17	33	39
AIC	0.25				100
	2.5	11	25	32	32
	5	39	31	18	12
BIC	0.25				100
	2.5			6	94
	5	1	5	19	75
C1 COV	0.25				100
	2.5		3	8	89
	5	1	5	14	80

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5		5	16	79
	5	6	11	31	52
AIC	0.25				100
	2.5	5	16	33	46
	5	29	31	20	20
BIC	0.25				100
	2.5			5	95
	5		3	9	88
C1 COV	0.25				100
	2.5			1	99
	5		2	7	91

Appendix 8: Frequency of Selecting Different models ($\beta_{\max}, n = 1000$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5		5	30	65
	5	1	17	34	48
AIC	0.25				100
	2.5	2	49	32	17
	5	26	59	12	3
BIC	0.25				100
	2.5		1	8	91
	5		3	17	80
C1 COV	0.25				100
	2.5			10	90
	5	2	3	20	75

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5				100
	5				100
AIC	0.25				100
	2.5				100
	5			9	91
BIC	0.25				100
	2.5				100
	5				100
C1 COV	0.25				100
	2.5				100
	5				100

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5				100
	5				100
AIC	0.25				100
	2.5				100
	5			3	97
BIC	0.25				100
	2.5				100
	5				100
C1 COV	0.25				100
	2.5				100
	5				100

Appendix 9: Frequency with which Models selected agree with KL Decisions ($\beta_{\max}, n = 50$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25		5	18	77
	2.5	31	43	20	6
	5	50	34	13	3
AIC	0.25		2	4	23
	2.5	15	5		
	5	24			
BIC	0.25		1	2	6
	2.5	18			
	5	41			
C1 COV	0.25			6	56
	2.5	15			
	5	26			

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5	1	6	19	74
	5	2	19	28	51
AIC	0.25				98
	2.5	1	2	3	21
	5	1	3	2	3
BIC	0.25				94
	2.5	1	3	2	5
	5	2	3		
C1 COV	0.25				100
	2.5		2	7	57
	5		1	8	27

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5		2	9	89
	5	1	10	24	65
AIC	0.25				99
	2.5			3	47
	5	1	1	4	12
BIC	0.25				99
	2.5			3	22
	5	1	1	1	
C1 COV	0.25				100
	2.5			2	84
	5			3	48

Appendix 10: Frequency with which Models selected agree with KL Decisions ($\beta_{\max}, n = 100$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25			6	94
	2.5	18	36	21	25
	5	29	44	16	11
AIC	0.25			2	63
	2.5	10	4		
	5	15	2		
BIC	0.25			2	30
	2.5	12	1		
	5	20			
C1 COV	0.25			2	88
	2.5	9	1		1
	5	15			

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5			6	94
	5	1	5	19	75
AIC	0.25				100
	2.5			2	60
	5		1	5	25
BIC	0.25				100
	2.5			2	28
	5	1		2	3
C1 COV	0.25				100
	2.5			2	85
	5			5	62

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5			5	95
	5		3	9	88
AIC	0.25				100
	2.5			2	76
	5		1	2	45
BIC	0.25				100
	2.5			2	43
	5		1	1	13
C1 COV	0.25				100
	2.5			1	95
	5		1	1	81

Appendix 11: Frequency with which Models selected agree with KL Decisions ($\beta_{\max}, n = 1000$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5		1	8	91
	5		3	17	80
AIC	0.25				100
	2.5			4	60
	5			6	36
BIC	0.25				100
	2.5		1	3	13
	5		2		
C1 COV	0.25				100
	2.5			4	85
	5			6	61

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5				100
	5				100
AIC	0.25				100
	2.5				100
	5				100
BIC	0.25				100
	2.5				100
	5				91
C1 COV	0.25				100
	2.5				100
	5				100

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5				100
	5				100
AIC	0.25				100
	2.5				100
	5				100
BIC	0.25				100
	2.5				100
	5				97
C1 COV	0.25				100
	2.5				100
	5				100

Appendix 12: Frequency of Selecting Different models ($\beta_{\min}, n = 50$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25	44	5	18	33
	2.5	69	12	8	11
	5	74	12	6	8
AIC	0.25	85	1	4	10
	2.5	94	3	1	2
	5	96	3	1	
BIC	0.25	22	2	5	71
	2.5	84	7	6	3
	5	84	10	4	2
C1 COV	0.25	36		6	58
	2.5	76	1	5	18
	5	77	2	4	17

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5	47	6	14	33
	5	62	8	12	18
AIC	0.25			1	99
	2.5	87	2	2	9
	5	88	3	2	7
BIC	0.25				100
	2.5	25	1	4	70
	5	59	2	9	30
C1 COV	0.25				100
	2.5	14	3	7	76
	5	21	4	17	58

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25			2	98
	2.5	11		44	45
	5	29	7	34	30
AIC	0.25			2	98
	2.5	41	2	37	20
	5	69	3	19	9
BIC	0.25				100
	2.5	2		23	75
	5	9	1	37	53
C1 COV	0.25			1	99
	2.5			16	84
	5	2		26	72

Tables 38-40

Appendix 13: Frequency of Selecting Different models ($\beta_{\min}, n = 100$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25	19	51	8	22
	2.5	67	22	5	6
	5	73	18	5	4
AIC	0.25	60	34	3	3
	2.5	92	7	1	
	5	94	5	1	
BIC	0.25	4	23	9	64
	2.5	50	34	8	8
	5	65	23	9	3
C1 COV	0.25	25	9	4	62
	2.5	69	6	5	20
	5	74	5	4	17

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25		1		99
	2.5	32	41	7	20
	5	52	28	7	13
AIC	0.25		8	2	90
	2.5	71	24	2	3
	5	81	15	3	1
BIC	0.25				100
	2.5	8	20	9	63
	5	25	31	14	30
C1 COV	0.25			1	99
	2.5	8	15	9	68
	5	21	13	5	61

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5	4	14	37	45
	5	27	19	29	25
AIC	0.25			1	99
	2.5	42	19	24	15
	5	74	17	7	2
BIC	0.25				100
	2.5		1	12	87
	5	6	5	25	64
C1 COV	0.25				100
	2.5			20	80
	5		8	20	72

Appendix 14: Frequency of Selecting Different models ($\beta_{\min}, n = 1000$)

High Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5	25	2	33	40
	5	50	3	25	22
AIC	0.25			10	90
	2.5	93		5	2
	5	99	1		
BIC	0.25				100
	2.5	6		20	74
	5	25	2	34	39
C1 COV	0.25				100
	2.5	14	1	14	71
	5	43	1	14	42

Moderate Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5				100
	5			8	92
AIC	0.25				100
	2.5			10	90
	5	16		31	53
BIC	0.25				100
	2.5				100
	5			1	99
C1 COV	0.25				100
	2.5				100
	5				100

Low Collinearity

	σ^2	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	0.25				100
	2.5			6	94
	5			25	75
AIC	0.25				100
	2.5	16		35	65
	5	16		61	39
BIC	0.25				100
	2.5			1	99
	5			6	94
C1 COV	0.25				100
	2.5				100
	5			3	97

Appendix 15: Frequency with which Models selected agree with KL Decisions ($\beta_{\min}, n = 50$)

sigma=0.25 High Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	2	22	5	71
	100	23	4	9	64
	1000				100
AIC	50		2		19
	100	7	2		10
	1000				100
BIC	50		9		1
	100	3	2		
	1000				90
C1 COV	50		3		40
	100	1	1	2	40
	1000				100

sigma=2.5 High Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	7	84	6	3
	100	34	50	8	8
	1000		6	20	74
AIC	50		56		
	100		20		
	1000		2	5	28
BIC	50		79		
	100		42		
	1000		4		
C1 COV	50		63		
	100		25		
	1000		1	3	55

sigma=5 High Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	10	84	4	2
	100	23	65	9	3
	1000	2	25	34	39
AIC	50		61		
	100		38		
	1000		2	8	3
BIC	50		81		
	100		59		
	1000		24		
C1 COV	50		64		
	100		41		
	1000		3	5	9

Appendix 16: Frequency with which Models selected agree with KL Decisions ($\beta_{\min}, n = 100$)

Sigma=0.25 Moderate Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50				100
	100				100
	1000				100
AIC	50				100
	100				99
	1000				100
BIC	50				99
	100				90
	1000				100
C1 COV	50				100
	100				99
	1000				100

Sigma=2.5 Moderate Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	1	25	4	70
	100	20	8	9	63
	1000				100
AIC	50		4		20
	100	3	5		8
	1000				100
BIC	50		13		
	100	1	5		
	1000				90
C1 COV	50		1		55
	100	3	1	2	45
	1000				100

Sigma=5 Moderate Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	2	59	9	30
	100	31	25	14	30
	1000			1	99
AIC	50		24		
	100	4	9		
	1000				91
BIC	50		48		
	100		9		
	1000				52
C1 COV	50		4		15
	100	2	4	2	17
	1000				99

Tables 50-52

Appendix 17: Frequency with which Models selected agree with KL Decisions ($\beta_{\min}, n = 1000$)

Sigma=0.25 Low Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50				100
	100				100
	1000				100
AIC	50				98
	100				100
	1000				100
BIC	50				98
	100				99
	1000				100
C1 COV	50				99
	100				100
	1000				100

Sigma=2.5 Low Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50		2	23	75
	100	1		12	87
	1000			1	99
AIC	50			5	26
	100			8	42
	1000				93
BIC	50			5	6
	100			4	12
	1000				64
C1 COV	50			5	64
	100			8	75
	1000				99

Sigma=5 Low Collinearity

	n	OLS x1	OLS x1 x2	OLS x1 x2 x3	OLS x1 x2 x3 x4
KL	50	1		37	53
	100	5	6	25	64
	1000			6	94
AIC	50		1	9	6
	100		1	5	12
	1000			3	72
BIC	50		2	7	
	100		3		
	1000			3	36
C1 COV	50			11	38
	100	1		8	51
	1000			2	93

Tables 53-55

Appendix 18: Time series and ICOMP

$$A' = \left(\frac{T-p}{\sigma^2}\left(\frac{c}{1-\phi_1-\ldots-\phi_p}\right) \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \frac{T-p}{\sigma^2}\left(\frac{c}{1-\phi_1-\ldots-\phi_p}\right) \right)' \tag{6.1}$$

$$B_{p*p}= \begin{pmatrix} \frac{T-p}{\sigma^2}\gamma_0 & & & & & \\ \frac{T-p}{\sigma^2}\gamma_1 & \ldots & & & & \\ \ldots & \ldots & \ldots & & & \\ \ldots & \ldots & \ldots & \ldots & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \\ \frac{T-p}{\sigma^2}\gamma_{|p-1|} & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_1 & \frac{T-p}{\sigma^2}\gamma_0 \end{pmatrix} \tag{6.2}$$

$$C_{q*p} = \begin{pmatrix} \frac{T-p}{\sigma^2}\gamma_0 & \frac{T-p}{\sigma^2}\gamma_1 & \frac{T-p}{\sigma^2}\gamma_2 & & & \frac{T-p}{\sigma^2}\gamma_\Phi \\ \frac{T-p}{\sigma^2}\gamma_1 & \ldots & \frac{T-p}{\sigma^2}\gamma_1 & \ldots & & \\ \frac{T-p}{\sigma^2}\gamma_2 & \frac{T-p}{\sigma^2}\gamma_1 & \ldots & \ldots & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_1 \\ \frac{T-p}{\sigma^2}\gamma_\Phi & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_1 & \frac{T-p}{\sigma^2}\gamma_0 \\ \frac{T-p}{\sigma^2}\gamma_{\Phi+1} & \frac{T-p}{\sigma^2}\gamma_\Phi & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_1 \\ \ldots & \frac{T-p}{\sigma^2}\gamma_{\Phi+1} & \ldots & \ldots & \ldots & \ldots \\ \frac{T-p}{\sigma^2}\gamma_\Delta & \ldots & \frac{T-p}{\sigma^2}\gamma_{\Phi+1} & \frac{T-p}{\sigma^2}\gamma_\Phi & \ldots & \frac{T-p}{\sigma^2}\gamma_\square \end{pmatrix} \tag{6.3}$$

* $\Phi = \min(p,q) - 1$
* $\Delta = \max(p,q) - 1$
* $\square = |p-q|$

$$D_{q*q} = \begin{pmatrix} T-p & & & & & & \\ \frac{T-p}{\sigma^2}\gamma_1 & \ldots & & & & & \\ \frac{T-p}{\sigma^2}\gamma_2 & \ldots & \ldots & & & & \\ \ldots & \ldots & \ldots & \ldots & & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\ \frac{T-p}{\sigma^2}\gamma_{|q-1|} & \ldots & \ldots & \ldots & \frac{T-p}{\sigma^2}\gamma_2 & \frac{T-p}{\sigma^2}\gamma_1 & T-p \end{pmatrix}$$

Note that B and D are symmetric.

Appendix 19: Eviews Time Series Simulation Code

```
include c:\allan\stats\sta500w\eviews\programs\
icompsimulation\INPUTLINESSUBAR1.prg
'!t=the length of the time series
'2 different lengths are considered namely 250 and 1000
'!totalrun=the total number of different models considered
'You should change the tyype of time series model
'by changing the "Y = " line
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
SMPL 1 5000
!T=250-2+1
!obssize=5000
!TOTALRUN=63
CALL START1 'THE TABLE AND MAKING THE INITIAL SERIES
CALL START2 'MAKES THE LIST OF POSSIBLE MODELS
FOR !AA=1 TO 1000
SMPL 1 5000
rndseed !aa
SERIES E=NRND
SERIES Y=0
Y(2)=1
SMPL 2 5000
Y=0.8*Y(-1)+e
CALL START3 'ESTIMATION ETC
CALL MINPROG
NEXT !AA
```

Appendix 20: Eviews Time Series Simulation Subroutines

```
SUBROUTINE TABLES
If @isobject("TRY") then
delete TRY
delete TRY2
endif
table(100,2) try
table(100,2) try2
setcolwidth(TRY,1,25)
setcolwidth(TRY2,1,25)
'THE FIRST TABLE GIVES THE RESULTS INCLUDING THE CORRECT MODEL
TRY(1,1)="MODEL"
TRY(1,2)="LOGL"
TRY(1,3)="AIC"
TRY(1,4)="SCHARZ"
' TRY(1,5)="KULL-LEIB"
TRY(1,6)="ICOMP(IFIM)"
TRY(1,7)="ICOMP(VEMD)"
TRY(1,8)="COMP"
' TRY(1,9)="Dur Watson"
TRY(1,9)="COMP(C1)"
TRY(1,10)="Q*(5)"
TRY(1,11)="Q*(10)"
TRY(1,12)="Q*(15)"
'THE SECOND TABLE GIVES THE RESULTS EXCLUDING THE CORRECT MODEL
TRY2(1,1)="MODEL"
TRY2(1,2)="LOGL"
TRY2(1,3)="AIC"
TRY2(1,4)="SCHARZ"
' TRY2(1,5)="KULL-LEIB"
TRY2(1,6)="ICOMP(IFIM)"
TRY2(1,7)="ICOMP(VEMD)"
TRY2(1,8)="COMP"
```

```
TRY2(1,9)="COMP(C1)"
ENDSUB
"""""""""""""""""""""""""""""""""""""""""
SUBROUTINE MAKESERIES
SERIES ORDER
FOR !Z=1 TO !obssize
ORDER(!Z)=!Z
NEXT
SERIES pfind
SERIES LOGL
SERIES LOGLcount
SERIES LOGL2count
SERIES AIC
SERIES AICcount
SERIES AIC2count
SERIES SC
SERIES SCcount
SERIES SC2count
SERIES KULLBACK
SERIES KULLBACKcount
SERIES KULLBACK2count
SERIES ICOMPIFIM
SERIES ICOMPIFIMcount
SERIES ICOMPIFIM2count
SERIES ICOMPC1
SERIES ICOMPC1count
SERIES ICOMPC12count
SERIES ICOMPVE
SERIES ICOMPVEcount
SERIES ICOMPVE2count
SERIES COMP
SERIES COMPcount
SERIES COMP2count
```

```
LOGL=-1000000000000000

AIC=1000000000000000

SC=1000000000000000

KULLBACK=1000000000000000

ICOMPIFIM=1000000000000000

ICOMPVE=1000000000000000

ICOMPC1=1000000000000000

COMP=1000000000000000

LOGLcount=0

AICcount=0

SCcount=0

KULLBACKcount=0

ICOMPIFIMcount=0

ICOMPVEcount=0

ICOMPC1count=0

COMPcount=0

LOGL2count=0

AIC2count=0

SC2count=0

KULLBACK2count=0

ICOMPIFIM2count=0

ICOMPVE2count=0

ICOMPC12count=0

COMP2count=0

ENDSUB

""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

SUBROUTINE ESTIMATION

SMPL 2 250

equation a{!BB}.LS Y %INPUT

if !aa=1 then

scalar ascSEVEN{!bb}=0

scalar ascten{!bb}=0

scalar ascfifteen{!bb}=0
```

```
endif
freeze(afr{!bb}) a{!bb}.correl(15)
if afr{!bb}(14,7)>0.05 then
ascSEVEN{!bb}=ascSEVEN{!bb}+1
endif
if afr{!bb}(17,7)>0.05 then
ascten{!bb}=ascten{!bb}+1
endif
if afr{!bb}(22,7)>0.05 then
ascfifteen{!bb}=ascfifteen{!bb}+1
endif
delete afr{!bb}
SMPL 1 5000
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE INFCRIT
matrix ccv{!bb}=a{!bb}.@coefcov 'COVARIANCE OF THE BETAS
MATRIX(!SIZE+1,!SIZE+1) CCVM{!bb}=0
CALL FINDP 'DETERMINES THE VALUE OF !P
ccvm{!bb}(1+!size,1+!size)=(2*a{!bb}.@se^4) /(!T-!p ) 'the covariance of sigma squared
matplace(ccvm{!bb},ccv{!bb},1,1) 'covariance of all the estimates
MATRIX IFIM{!bb}=ccvm{!bb} 'INVERSE OF THE INFORMATION MATRIX
SERIES FITS2=0
FITS2=(Y-RESID)^2
LOGL(!bb)=a{!bb}.@logl
AIC(!bb)=a{!bb}.@aic
SC(!bb)=a{!bb}.@sc
'
KULLBACK(!bb)=0.5*(!T-!P)*LOG(1/(a{!bb}.@se^2))+
0.5*(!T-!P)*((a{!bb}.@se^2)-1)+0.5*(@SUM(FITS2))
'THE KULLBACK-LEIBER DISTANCE
ICOMPIFIM(!bb) = -2*a{!bb}.@logl+ (1+!SIZE)*log( @trace(IFIM{!bb})/ (1+!SIZE) ) - log( @det(IFIM{!bb})
) 'ICOMPIFIM
```

ICOMPVE(!bb) = -2*a{!bb}.@logl + log(@trace(CCV{!bb})) - log(@det(CCV{!bb})) 'ICOMPIFIMvanEMDEN

COMP(!bb) = (1+!SIZE)*log(@trace(IFIM{!bb})/ (1+!SIZE)) - log(@det(IFIM{!bb})) 'COMPLEXITY FOR IFIM

ICOMPC1(!bb) = -2*a{!bb}.@logl + (!SIZE)*log(@trace(CCV{!bb})/ (!SIZE)) - log(@det(CCV{!bb})) 'ICOMPIFIM C(1)

ENDSUB

""

SUBROUTINE INFCRITmodel

TRY(!count,2)=a{!count}.@logl

TRY(!count,3)=a{!count}.@aic

TRY(!count,4)=a{!count}.@sc

matrix ccv{!count}=a{!count}.@coefcov 'COVARIANCE OF THE BETAS

MATRIX(!SIZE+1,!SIZE+1) CCVM{!count}=0

CALL FINDP 'DETERMINES THE VALUE OF !P

ccvm{!count}(1+!size,1+!size)=(2*a{!count}.@se^4) /(!T-!p) 'the covariance of sigma squared

matplace(ccvm{!count},ccv{!count},1,1) 'covariance of all the estimates

MATRIX IFIM{!COUNT}=ccvm{!count} 'INVERSE OF THE INFORMATION MATRIX

SERIES FITS2=0

FITS2=(Y-RESID)^2

' TRY(!COUNT,5)=0.5*(!T-!P)*LOG(1/(a{!count}.@se^2))+0.5*(!T-!P)*((a{!count}.@se^2)-1)+0.5*(@SUM(FITS2)) 'THE KULLBACK-LEIBER DISTANCE

TRY(!count,6) = -2*a{!count}.@logl + (1+!SIZE)*log(@trace(IFIM{!count})/ (1+!SIZE)) - log(@det(IFIM{!count})) 'ICOMPIFIM

TRY(!count,7) = -2*a{!count}.@logl + log(@trace(CCV{!count})) - log(@det(CCV{!count})) 'ICOMPIFIMvanEMDEN

TRY(!count,8) = (1+!SIZE)*log(@trace(IFIM{!count})/ (1+!SIZE)) - log(@det(IFIM{!count})) 'COMPLEXITY FOR IFIM

TRY(!count,9) = -2*a{!count}.@logl + (!SIZE)*log(@trace(CCV{!count})/ (!SIZE)) - log(@det(CCV{!count})) 'ICOMPIFIM C(1)

ENDSUB

""

SUBROUTINE FINDP

```
pfind=0
!pcount=0
FOR !J=1 TO !SIZE-1
if @MID(%INPUT,1+6*(!J-1),1)="a" then
!pcount=!pcount+1
pfind(!pcount)=@VAL(@MID(%INPUT,4+6*(!J-1),1))
endif
NEXT !J
!P=@MAX(PFIND)
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE DELETES
DELETE a{!bb}
DELETE ccv{!bb}
DELETE CCVM{!bb}
DELETE IFIM{!bb}
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE MINPROG
SMPL 1 5000
SORT(D) LOGL 'SORT LOGL NOW CHOOSE THE LARGEST LOGL
LOGLcount(1)=LOGLcount(1)+1
LOGL2count(2)=LOGL2count(2)+1
TRY(ORDER(1)+1,2)=LOGLcount(1)
TRY2(ORDER(2)+1,2)=LOGL2count(2)
SORT ORDER
SORT AIC 'SORT AIC NOW CHOOSE THE SMALLEST AIC
AICcount(1)=AICcount(1)+1
AIC2count(2)=AIC2count(2)+1
TRY(ORDER(1)+1,3)=AICcount(1)
TRY2(ORDER(2)+1,3)=AIC2count(2)
SORT ORDER
SORT SC 'SORT SC NOW CHOOSE THE SMALLEST SC
```

```
SCcount(1)=SCcount(1)+1

SC2count(2)=SC2count(2)+1

TRY(ORDER(1)+1,4)=SCcount(1)

TRY2(ORDER(2)+1,4)=SC2count(2)

SORT ORDER

' SORT KULLBACK 'SORT KULLBACK NOW CHOOSE THE SMALLEST KULLBACK

' KULLBACKcount(1)=KULLBACKcount(1)+1

' KULLBACK2count(2)=KULLBACK2count(2)+1

' TRY(ORDER(1)+1,5)=KULLBACKcount(1)

' TRY2(ORDER(2)+1,5)=KULLBACK2count(2)

' SORT ORDER

SORT ICOMPIFIM 'SORT ICOMPIFIM NOW CHOOSE THE SMALLEST ICOMPIFIM

ICOMPIFIMcount(1)=ICOMPIFIMcount(1)+1

ICOMPIFIM2count(2)=ICOMPIFIM2count(2)+1

TRY(ORDER(1)+1,6)=ICOMPIFIMcount(1)

TRY2(ORDER(2)+1,6)=ICOMPIFIM2count(2)

SORT ORDER

SORT ICOMPVE 'SORT ICOMPVE NOW CHOOSE THE SMALLEST ICOMPVE

ICOMPVEcount(1)=ICOMPVEcount(1)+1

ICOMPVE2count(2)=ICOMPVE2count(2)+1

TRY(ORDER(1)+1,7)=ICOMPVEcount(1)

TRY2(ORDER(2)+1,7)=ICOMPVE2count(2)

SORT ORDER

SORT COMP 'SORT COMP NOW CHOOSE THE SMALLEST COMP

COMPcount(1)=COMPcount(1)+1

COMP2count(2)=COMP2count(2)+1

TRY(ORDER(1)+1,8)=COMPcount(1)

TRY2(ORDER(2)+1,8)=COMP2count(2)

SORT ORDER

SORT ICOMPC1 'SORT ICOMPC1 NOW CHOOSE THE SMALLEST ICOMPVE

ICOMPC1count(1)=ICOMPC1count(1)+1

ICOMPC12count(2)=ICOMPC12count(2)+1

TRY(ORDER(1)+1,9)=ICOMPC1count(1)
```

```
TRY2(ORDER(2)+1,9)=ICOMPC12count(2)

SORT ORDER

ENDSUB

""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

SUBROUTINE FINALDELETE

delete LOGL

delete LOGLcount

delete LOGL2count

delete AIC

delete AICcount

delete AIC2count

delete SC

delete SCcount

delete SC2count

delete KULLBACK

delete KULLBACKcount

delete KULLBACK2count

delete ICOMPIFIM

delete ICOMPIFIMcount

delete ICOMPIFIM2count

delete ICOMPVE

delete ICOMPVEcount

delete ICOMPVE2count

delete ICOMPC1

delete ICOMPC1count

delete ICOMPC12count

delete COMP

delete COMPcount

delete COMP2count

delete fits2

delete order

delete pfind

delete c
```

```
'delete y
ENDSUB
"""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP1
!count=1
!SIZE=1 'SIZE OF THE COVARIANCE MATRIX
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!count=!count+1
%input= %0
try(!count,1)=%input
try2(!count,1)=!SIZE
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
next %0
ENDSUB
"""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP2
!a=0
!SIZE=2
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!a=!a+1
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
if (%0<>%1 and !a<!b )then
!count=!count+1
%input= %0 +" "+ %1
try(!count,1)=%input
try2(!count,1)=!SIZE
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
```

```
endif
next %1
next %0
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP3
!SIZE=3
!a=0
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!a=!a+1
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
!c=0
for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!c=!c+1
if (%0<>%1 and !a<!b and !b<!c ) then
!count=!count+1
%input= %0 +" "+ %1+" " +%2
try(!count,1)=%input
try2(!count,1)=!SIZE
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %2
next %1
next %0
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP4
!SIZE=4
```

```
!a=0
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!a=!a+1
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
!c=0
for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!c=!c+1
!d=0
for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!d=!d+1
if (%0<>%1 and !a<!b and !b<!c and !c<!d ) then
!COUNT=!COUNT+1
%INPUT= %0 +" "+ %1+" " +%2+" " +%3
try(!COUNT,1)=%INPUT
try2(!count,1)=!SIZE
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %3
next %2
next %1
next %0
ENDSUB
""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP5
!SIZE=5
!a=0
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!a=!a+1
```

```
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
!c=0
for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!c=!c+1
!d=0
for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!d=!d+1
!e=0
for %4 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!e=!e+1
if (%0<>%1 and !a<!b and !b<!c and !c<!d and !d<!e ) then
!COUNT=!COUNT+1
%INPUT= %0 +" "+ %1+" " +%2+" "+%3+ " "+%4
try(!COUNT,1)=%INPUT
try2(!count,1)=!SIZE
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %4
next %3
next %2
next %1
next %0
ENDSUB
''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
SUBROUTINE GROUP6
!SIZE=6
!a=0
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
```

```
!a=!a+1
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
!c=0
for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!c=!c+1
!d=0
for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!d=!d+1
!e=0
for %4 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!e=!e+1
!f=0
for %5 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!f=!f+1
if (%0<>%1 and !a<!b and !b<!c and !c<!d and !d<!e and !e<!f ) then
!COUNT=!COUNT+1
%INPUT= %0 +" "+ %1+" " +%2+" "+%3+ " "+%4+" "+%5
try(!COUNT,1)=%INPUT
try2(!count,1)=!SIZE
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %5
next %4
next %3
next %2
next %1
next %0
ENDSUB
```

```
""""""""""""""""""""""""""""""""""""""""

SUBROUTINE GROUP7

!SIZE=7

!a=0

for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!a=!a+1

!b=0

for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!b=!b+1

!c=0

for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!c=!c+1

!d=0

for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!d=!d+1

!e=0

for %4 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!e=!e+1

!f=0

for %5 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!f=!f+1

!g=0

for %6 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!g=!g+1

if (%0<>%1 and !a<!b and !b<!c and !c<!d and !d<!e and !e<!f and !f<!g ) then

!COUNT=!COUNT+1

%INPUT= %0 +" "+ %1+" " +%2+" "+%3+ " "+%4+" "+%5+" "+%6

try(!COUNT,1)=%INPUT

try2(!count,1)=!SIZE

'

' CALL ESTIMATION

' CALL INFCRIT

' CALL DELETES
```

```
endif

next %6

next %5

next %4

next %3

next %2

next %1

next %0

ENDSUB

"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

SUBROUTINE GROUP8

!SIZE=8

!a=0

for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!a=!a+1

!b=0

for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!b=!b+1

!c=0

for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!c=!c+1

!d=0

for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!d=!d+1

!e=0

for %4 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!e=!e+1

!f=0

for %5 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!f=!f+1

!g=0

for %6 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)

!g=!g+1
```

```
!h=0
for %7 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!h=!h+1
if (%0<>%1 and !a<!b and !b<!c and !c<!d and !d<!e and !e<!f and !f<!g and !g<!h ) then
!COUNT=!COUNT+1
%INPUT= %0 +" "+ %1+" " +%2+" "+%3+ " "+%4+" "+%5+" "+%6+" " +%7
try(!COUNT,1)=%INPUT
try2(!count,1)=!SIZE
'
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %7
next %6
next %5
next %4
next %3
next %2
next %1
next %0
ENDSUB
""""""""""""""""""""""""""""""""""""""""""""""""""
SUBROUTINE GROUP9
!SIZE=9
!a=0
for %0 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!a=!a+1
!b=0
for %1 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!b=!b+1
!c=0
```

```
for %2 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!c=!c+1
!d=0
for %3 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!d=!d+1
!e=0
for %4 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!e=!e+1
!f=0
for %5 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!f=!f+1
!g=0
for %6 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!g=!g+1
!h=0
for %7 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!h=!h+1
!i=0
for %8 ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)
!i=!i+1
if (%0<>%1 and !a<!b and !b<!c and !c<!d and !d<!e and !e<!f and !f<!g and !g<!h and !h<!i ) then
!COUNT=!COUNT+1
%INPUT= %0 +" "+ %1+" " +%2+" "+%3+ " "+%4+" "+%5+" "+%6+" " +%7+" " +%8
try(!COUNT,1)=%INPUT
try2(!count,1)=!SIZE
'
'
' CALL ESTIMATION
' CALL INFCRIT
' CALL DELETES
endif
next %8
next %7
```

```
next %6

next %5

next %4

next %3

next %2

next %1

next %0

ENDSUB

”””””””””””””””””””””””””””””””””””””””””””””””

SUBROUTINE GROUP10

!SIZE=10

%INPUT=”ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)”

!COUNT=!COUNT+1

try(!COUNT,1)=%INPUT

try2(!count,1)=!SIZE

ENDSUB

””””””””””””””””””””””””””””””””””””””””

SUBROUTINE START1

CALL TABLES

CALL MAKESERIES

ENDSUB

””””””””””””””””””””””””””””””””””””””””””””””

SUBROUTINE START2

CALL GROUP1

CALL GROUP2

CALL GROUP3

CALL GROUP4

CALL GROUP5

CALL GROUP6

' CALL GROUP7

' CALL GROUP8

' CALL GROUP9

' CALL GROUP10
```

```
ENDSUB

"""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

SUBROUTINE START3

FOR !BB=1 TO !TOTALRUN

%INPUT=try(!BB+1,1)

!SIZE=@VAL(try2(!BB+1,1))

CALL ESTIMATION

CALL INFCRIT

CALL DELETES

NEXT !BB

ENDSUB
```

MODEL-ar(1)	LOGL	AIC	SCHARZ	ICOMP (IFIM)	ICOMP (VEMD)	COMP
ar(1)		**195**	**943**	**873**	**972**	**1000**
ar(2)						
ar(3)						
ma(1)						
ma(2)						
ma(3)						
ar(1) ar(2)		9	1	1		
ar(1) ar(3)		19	3	7	3	
ar(1) ma(1)		12	2	20	3	
ar(1) ma(2)		24	5	11	3	
ar(1) ma(3)		29	8	22	5	
ar(2) ar(3)						
ar(2) ma(1)		98	24	36	13	
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)						
ar(3) ma(2)						
ar(3) ma(3)						
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)						
ar(1) ar(2) ar(3)		1				
ar(1) ar(2) ma(1)	1	46	3	9		
ar(1) ar(2) ma(2)		1				
ar(1) ar(2) ma(3)		1				
ar(1) ar(3) ma(1)						
ar(1) ar(3) ma(2)						
ar(1) ar(3) ma(3)		6		3		
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)		1		1		
ar(1) ma(2) ma(3)		4				
ar(2) ar(3) ma(1)		9		1		
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)						
ar(2) ma(1) ma(2)		8		1		
ar(2) ma(1) ma(3)		24	1	1	1	
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		61	2	6		
ar(3) ma(1) ma(3)						
ar(3) ma(2) ma(3)						
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)	8	13				
ar(1) ar(2) ar(3) ma(2)	45	116	2	3		
ar(1) ar(2) ar(3) ma(3)						
ar(1) ar(2) ma(1) ma(2)	4	9	1			
ar(1) ar(2) ma(1) ma(3)	5	30				
ar(1) ar(2) ma(2) ma(3)						
ar(1) ar(3) ma(1) ma(2)	3	16	1	1		
ar(1) ar(3) ma(1) ma(3)						
ar(1) ar(3) ma(2) ma(3)						
ar(1) ma(1) ma(2) ma(3)						
ar(2) ar(3) ma(1) ma(2)	4	20		1		
ar(2) ar(3) ma(1) ma(3)	2	3				
ar(2) ar(3) ma(2) ma(3)						
ar(2) ma(1) ma(2) ma(3)	1	5				
ar(3) ma(1) ma(2) ma(3)	2	17		1		
ar(1) ar(2) ar(3) ma(1) ma(2)	112	105	2	1		
ar(1) ar(2) ar(3) ma(1) ma(3)	26	5				
ar(1) ar(2) ar(3) ma(2) ma(3)	139	40				
ar(1) ar(2) ma(1) ma(2) ma(3)	18	4				
ar(1) ar(3) ma(1) ma(2) ma(3)	31	7		1		
ar(2) ar(3) ma(1) ma(2) ma(3)	33	8				
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	566	54	2			

Appendix 21: Table 56: The AR(1) Simulation Results $(n = 1000, \sigma^2 = 25)$

MODEL-MA(1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						1000
ar(2)						
ar(3)						
ma(1)		**210**	**948**	**2**	**792**	
ma(2)						
ma(3)						
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)		15	2			
ar(1) ma(2)		60	16	1	5	
ar(1) ma(3)						
ar(2) ar(3)						
ar(2) ma(1)		36	4		2	
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)		52	9	889	189	
ar(3) ma(2)						
ar(3) ma(3)						
ma(1) ma(2)		16	2		1	
ma(1) ma(3)		37	3		1	
ma(2) ma(3)						
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)		2				
ar(1) ar(2) ma(2)		9				
ar(1) ar(2) ma(3)		21				
ar(1) ar(3) ma(1)		1		18	3	
ar(1) ar(3) ma(2)		18	2	37	4	
ar(1) ar(3) ma(3)						
ar(1) ma(1) ma(2)		27	3			
ar(1) ma(1) ma(3)						
ar(1) ma(2) ma(3)		7				
ar(2) ar(3) ma(1)		4		15		
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)						
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)						
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		4		3		
ar(3) ma(1) ma(3)		18		15	3	
ar(3) ma(2) ma(3)						
ma(1) ma(2) ma(3)		2				
ar(1) ar(2) ar(3) ma(1)		1		1		
ar(1) ar(2) ar(3) ma(2)		4		2		
ar(1) ar(2) ar(3) ma(3)	4	3		7		
ar(1) ar(2) ma(1) ma(2)		4				
ar(1) ar(2) ma(1) ma(3)	6	55	2			
ar(1) ar(2) ma(2) ma(3)	4	49	1			
ar(1) ar(3) ma(1) ma(2)	7	24	1	2		
ar(1) ar(3) ma(1) ma(3)						
ar(1) ar(3) ma(2) ma(3)	3	5				
ar(1) ma(1) ma(2) ma(3)		1				
ar(2) ar(3) ma(1) ma(2)	1	1				
ar(2) ar(3) ma(1) ma(3)						
ar(2) ar(3) ma(2) ma(3)						
ar(2) ma(1) ma(2) ma(3)	1	33				
ar(3) ma(1) ma(2) ma(3)		1				
ar(1) ar(2) ar(3) ma(1) ma(2)	15	1				
ar(1) ar(2) ar(3) ma(1) ma(3)	54	27		1		
ar(1) ar(2) ar(3) ma(2) ma(3)	96	16		3		
ar(1) ar(2) ma(1) ma(2) ma(3)	48	149	7			
ar(1) ar(3) ma(1) ma(2) ma(3)	21					
ar(2) ar(3) ma(1) ma(2) ma(3)	61	11		1		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	679	76		3		

Appendix 22: Table 57: The MA(1) Simulation Results $(n = 1000, \sigma^2 = 25)$

MODEL-ARMA(1,1)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)		12	108		5	14
ar(2)		7	88		4	20
ar(3)		24	180	407	900	172
ma(1)		7	111		2	116
ma(2)		10	100		3	186
ma(3)		18	156			492
ar(1) ar(2)						
ar(1) ar(3)				1		
ar(1) ma(1)	**1**	**32**	**28**			
ar(1) ma(2)						
ar(1) ma(3)						
ar(2) ar(3)		3		3		
ar(2) ma(1)						
ar(2) ma(2)		62	63	3		
ar(2) ma(3)						
ar(3) ma(1)		1		6	2	
ar(3) ma(2)		1		2	1	
ar(3) ma(3)	2	104	103	492	77	
ma(1) ma(2)						
ma(1) ma(3)		2				
ma(2) ma(3)		2				
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)		4				
ar(1) ar(2) ma(2)		14	2			
ar(1) ar(2) ma(3)						
ar(1) ar(3) ma(1)	4	13	1	10		
ar(1) ar(3) ma(2)						
ar(1) ar(3) ma(3)	4	24	2	9	2	
ar(1) ma(1) ma(2)		4				
ar(1) ma(1) ma(3)	1	12	1			
ar(1) ma(2) ma(3)						
ar(2) ar(3) ma(1)						
ar(2) ar(3) ma(2)		22		16		
ar(2) ar(3) ma(3)	7	32	4	7	1	
ar(2) ma(1) ma(2)		14	1			
ar(2) ma(1) ma(3)						
ar(2) ma(2) ma(3)		19	5			
ar(3) ma(1) ma(2)						
ar(3) ma(1) ma(3)	3	28	7	12	2	
ar(3) ma(2) ma(3)	1	19	5	15		
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)	3			1		
ar(1) ar(2) ar(3) ma(2)	16	2				
ar(1) ar(2) ar(3) ma(3)	14	4				
ar(1) ar(2) ma(1) ma(2)	8	53	10			
ar(1) ar(2) ma(1) ma(3)	1	2	1			
ar(1) ar(2) ma(2) ma(3)	2	16	1			
ar(1) ar(3) ma(1) ma(2)	2	3				
ar(1) ar(3) ma(1) ma(3)	43	133	9	9	1	
ar(1) ar(3) ma(2) ma(3)	9	3				
ar(1) ma(1) ma(2) ma(3)						
ar(2) ar(3) ma(1) ma(2)	13	5	1	1		
ar(2) ar(3) ma(1) ma(3)	7	6				
ar(2) ar(3) ma(2) ma(3)	17	87	6	4		
ar(2) ma(1) ma(2) ma(3)		4				
ar(3) ma(1) ma(2) ma(3)	8	7				
ar(1) ar(2) ar(3) ma(1) ma(2)	38	16	1			
ar(1) ar(2) ar(3) ma(1) ma(3)	63	30	4			
ar(1) ar(2) ar(3) ma(2) ma(3)	64	15				
ar(1) ar(2) ma(1) ma(2) ma(3)	11	25				
ar(1) ar(3) ma(1) ma(2) ma(3)	71	29	1			
ar(2) ar(3) ma(1) ma(2) ma(3)	58	17				
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	529	53	1	2		

Appendix 23: Table 58: The ARMA(1,1) Simulation Results ($n = 1000, \sigma^2 = 25$)

MODEL-AR(3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						
ar(2)						
ar(3)		**441**	**963**	**920**	**981**	**1000**
ma(1)						
ma(2)						
ma(3)						
ar(1) ar(2)						
ar(1) ar(3)		60	8	8	1	
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)						
ar(2) ar(3)		78	9	11	7	
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)						
ar(3) ma(1)		62	7	16	3	
ar(3) ma(2)		72	5	13	3	
ar(3) ma(3)		91	7	24	5	
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)						
ar(1) ar(2) ar(3)		17				
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)						
ar(1) ar(3) ma(1)		32				
ar(1) ar(3) ma(2)		11				
ar(1) ar(3) ma(3)		13				
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)						
ar(1) ma(2) ma(3)						
ar(2) ar(3) ma(1)		12				
ar(2) ar(3) ma(2)		24	1	2		
ar(2) ar(3) ma(3)		13		1		
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)						
ar(2) ma(2) ma(3)						
ar(3) ma(1) ma(2)		13		2		
ar(3) ma(1) ma(3)		13		2		
ar(3) ma(2) ma(3)		13				
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)		2				
ar(1) ar(2) ar(3) ma(2)		5				
ar(1) ar(2) ar(3) ma(3)		3				
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)						
ar(1) ar(2) ma(2) ma(3)						
ar(1) ar(3) ma(1) ma(2)		2				
ar(1) ar(3) ma(1) ma(3)		6				
ar(1) ar(3) ma(2) ma(3)		3				
ar(1) ma(1) ma(2) ma(3)						
ar(2) ar(3) ma(1) ma(2)		4		1		
ar(2) ar(3) ma(1) ma(3)		1				
ar(2) ar(3) ma(2) ma(3)		4				
ar(2) ma(1) ma(2) ma(3)						
ar(3) ma(1) ma(2) ma(3)		2				
ar(1) ar(2) ar(3) ma(1) ma(2)	2					
ar(1) ar(2) ar(3) ma(1) ma(3)						
ar(1) ar(2) ar(3) ma(2) ma(3)		1				
ar(1) ar(2) ma(1) ma(2) ma(3)						
ar(1) ar(3) ma(1) ma(2) ma(3)		1				
ar(2) ar(3) ma(1) ma(2) ma(3)		1				
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	998					

Appendix 24: Table 59: The AR(3) Simulation Results* ($n = 1000, \sigma^2 = 25$)

MODEL-MA(3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						
ar(2)						
ar(3)						1000
ma(1)						
ma(2)						
ma(3)		**426**	**959**	**907**	**977**	
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)		71	5	19	4	
ar(2) ar(3)						
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)		72	5	12	3	
ar(3) ma(1)						
ar(3) ma(2)						
ar(3) ma(3)		95	6	27	3	
ma(1) ma(2)						
ma(1) ma(3)		66	9	9	4	
ma(2) ma(3)		70	16	20	9	
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		11				
ar(1) ar(3) ma(1)						
ar(1) ar(3) ma(2)						
ar(1) ar(3) ma(3)		13				
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)		34		1		
ar(1) ma(2) ma(3)		13				
ar(2) ar(3) ma(1)						
ar(2) ar(3) ma(2)						
ar(2) ar(3) ma(3)		10				
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)		17				
ar(2) ma(2) ma(3)		22		2		
ar(3) ma(1) ma(2)						
ar(3) ma(1) ma(3)		16		2		
ar(3) ma(2) ma(3)		17				
ma(1) ma(2) ma(3)		10				
ar(1) ar(2) ar(3) ma(1)						
ar(1) ar(2) ar(3) ma(2)						
ar(1) ar(2) ar(3) ma(3)		2		1		
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)		3				
ar(1) ar(2) ma(2) ma(3)		9				
ar(1) ar(3) ma(1) ma(2)						
ar(1) ar(3) ma(1) ma(3)		7				
ar(1) ar(3) ma(2) ma(3)		1				
ar(1) ma(1) ma(2) ma(3)		2				
ar(2) ar(3) ma(1) ma(2)						
ar(2) ar(3) ma(1) ma(3)		2				
ar(2) ar(3) ma(2) ma(3)		3				
ar(2) ma(1) ma(2) ma(3)		3				
ar(3) ma(1) ma(2) ma(3)		2				
ar(1) ar(2) ar(3) ma(1) ma(2)						
ar(1) ar(2) ar(3) ma(1) ma(3)	2	1				
ar(1) ar(2) ar(3) ma(2) ma(3)	2					
ar(1) ar(2) ma(1) ma(2) ma(3)	2	1				
ar(1) ar(3) ma(1) ma(2) ma(3)		1				
ar(2) ar(3) ma(1) ma(2) ma(3)	2					
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	992					

Appendix 25: Table 60: The MA(3) Simulation Results* $(n = 1000, \sigma^2 = 25)$

MODEL-AR(1,2,3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						
ar(2)						8
ar(3)						977
ma(1)						
ma(2)						
ma(3)						15
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)						
ar(2) ar(3)		12	200	94	236	
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)			3	1	3	
ar(3) ma(1)						
ar(3) ma(2)		1	8	5	8	
ar(3) ma(3)						
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)						
ar(1) ar(2) ar(3)		**338**	**466**	**373**	**434**	
ar(1) ar(2) ma(1)						
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		3	6	5	6	
ar(1) ar(3) ma(1)						
ar(1) ar(3) ma(2)		44	55	55	55	
ar(1) ar(3) ma(3)						
ar(1) ma(1) ma(2)			3	1	2	
ar(1) ma(1) ma(3)		1	2	2	2	
ar(1) ma(2) ma(3)		1	1	1	1	
ar(2) ar(3) ma(1)		207	205	223	206	
ar(2) ar(3) ma(2)		2	8	11	7	
ar(2) ar(3) ma(3)		2	4	4	1	
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)			1		1	
ar(2) ma(2) ma(3)		3	5	5	4	
ar(3) ma(1) ma(2)		7	6	6	6	
ar(3) ma(1) ma(3)						
ar(3) ma(2) ma(3)		1	2	2	2	
ma(1) ma(2) ma(3)						
ar(1) ar(2) ar(3) ma(1)		33	1	19	1	
ar(1) ar(2) ar(3) ma(2)		18	1	6	1	
ar(1) ar(2) ar(3) ma(3)		43		16		
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)		12	3	10	3	
ar(1) ar(2) ma(2) ma(3)		35	2	22	2	
ar(1) ar(3) ma(1) ma(2)		60	5	50	5	
ar(1) ar(3) ma(1) ma(3)						
ar(1) ar(3) ma(2) ma(3)		44	6	18	5	
ar(1) ma(1) ma(2) ma(3)						
ar(2) ar(3) ma(1) ma(2)		34	3	18	3	
ar(2) ar(3) ma(1) ma(3)		28		18	2	
ar(2) ar(3) ma(2) ma(3)				5		
ar(2) ma(1) ma(2) ma(3)		13	3	12	3	
ar(3) ma(1) ma(2) ma(3)		11	1	6	1	
ar(1) ar(2) ar(3) ma(1) ma(2)	3	5		1		
ar(1) ar(2) ar(3) ma(1) ma(3)		7		3		
ar(1) ar(2) ar(3) ma(2) ma(3)	2	7		2		
ar(1) ar(2) ma(1) ma(2) ma(3)	3	3				
ar(1) ar(3) ma(1) ma(2) ma(3)	6	7		5		
ar(2) ar(3) ma(1) ma(2) ma(3)	1	6				
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	985	12		1		

Appendix 26: Table 61: The pure AR(3) Simualtion Results $(n = 1000, \sigma^2 = 25)$

MODEL-MA(1,2,3)	LOGL	AIC	SCHARZ	ICOMP(IFIM)	ICOMP(VEMD)	COMP
ar(1)						
ar(2)						4
ar(3)						735
ma(1)						
ma(2)						1
ma(3)						260
ar(1) ar(2)						
ar(1) ar(3)						
ar(1) ma(1)						
ar(1) ma(2)						
ar(1) ma(3)						
ar(2) ar(3)						
ar(2) ma(1)						
ar(2) ma(2)						
ar(2) ma(3)			4	1	4	
ar(3) ma(1)						
ar(3) ma(2)			1	1	2	
ar(3) ma(3)						
ma(1) ma(2)						
ma(1) ma(3)						
ma(2) ma(3)		17	198	90	226	
ar(1) ar(2) ar(3)						
ar(1) ar(2) ma(1)			4	1	4	
ar(1) ar(2) ma(2)						
ar(1) ar(2) ma(3)		4	9	9	12	
ar(1) ar(3) ma(1)						
ar(1) ar(3) ma(2)						
ar(1) ar(3) ma(3)						
ar(1) ma(1) ma(2)						
ar(1) ma(1) ma(3)			1	1	1	
ar(1) ma(2) ma(3)		200	230	232	240	
ar(2) ar(3) ma(1)						
ar(2) ar(3) ma(2)		2	3	4	2	
ar(2) ar(3) ma(3)		1		1		
ar(2) ma(1) ma(2)						
ar(2) ma(1) ma(3)		35	57	40	57	
ar(2) ma(2) ma(3)		2	1	11		
ar(3) ma(1) ma(2)		2	8	5	8	
ar(3) ma(1) ma(3)						
ar(3) ma(2) ma(3)		3		6		
ma(1) ma(2) ma(3)		**337**	**456**	**362**	**418**	
ar(1) ar(2) ar(3) ma(1)		1	2	1	2	
ar(1) ar(2) ar(3) ma(2)		16	3	8	3	
ar(1) ar(2) ar(3) ma(3)		17	4	17	3	
ar(1) ar(2) ma(1) ma(2)						
ar(1) ar(2) ma(1) ma(3)		57	2	49	3	
ar(1) ar(2) ma(2) ma(3)		40	2	26	2	
ar(1) ar(3) ma(1) ma(2)		15	2	12	3	
ar(1) ar(3) ma(1) ma(3)			1		1	
ar(1) ar(3) ma(2) ma(3)		32	4	22	3	
ar(1) ma(1) ma(2) ma(3)		25		26		
ar(2) ar(3) ma(1) ma(2)		34	3	22	3	
ar(2) ar(3) ma(1) ma(3)		54	3	24	2	
ar(2) ar(3) ma(2) ma(3)						
ar(2) ma(1) ma(2) ma(3)		17		4		
ar(3) ma(1) ma(2) ma(3)		42	2	9	1	
ar(1) ar(2) ar(3) ma(1) ma(2)	4	1				
ar(1) ar(2) ar(3) ma(1) ma(3)	4	10		5		
ar(1) ar(2) ar(3) ma(2) ma(3)	6	2				
ar(1) ar(2) ma(1) ma(2) ma(3)	2	7		3		
ar(1) ar(3) ma(1) ma(2) ma(3)	4	8		4		
ar(2) ar(3) ma(1) ma(2) ma(3)		5		1		
ar(1) ar(2) ar(3) ma(1) ma(2) ma(3)	980	14		3		

Appendix 27: Table 62: The pure MA(3) Simualtion Results $(n = 1000, \sigma^2 = 25)$

Part V

Applications

Chapter 7

Volatility Modelling

7.1 Introduction

This chapter examines the stability of the correlation and the covariance structure of share returns of the largest market capitalization companies listed on the JSE. These companies comprise of shares included in the Satrix40 , Satrix-Indi and Satrix-Findi indices (as at Nov 2002) that have a price history back to at least January 1990. Satrix indices serves as popular benchmark portfolios for asset- and unit trust-managers who wish to replicate the above indices as their holding portfolios. The Satrix40, listed in November 2000, tracks the performance of the Alsi 40. Similarly Satrix-Indi and Satrix-Findi tracks the performance of the Industrial and The Industrial and Financial Index respectively. Satrix indices are revised every three months (every quarter) based on the market capitalization of the respective companies and thus the weights as well as the companies included in the index can change. These indices provides investors with the potential to track market indices at a relatively low cost as well as diversifying their respective portfolios.

The stability of the covariance matrix of asset returns is important to investors who utilize a mean-variance approach when undertaking portfolio construction since the constructed portfolios will only be representative of future optimal portfolios if the covariance matrix is relatively stationary. This implies that if the covariance of asset returns is not stable that one might have to estimate covariance matrices using relatively short periods of returns.

Numerous international studies have examined the stability of the correlation- and covariance-matrix for share returns. Kaplanis (1988) examined the potential benefits of diversifying across international markets. Equity returns from ten countries

namely, US, Canada, UK, Belgium, France, Germany, Italy, Netherlands, Switzerland and Japan, were investigated over the period 1967 to 1982. The co-movements of both nominal and real returns were investigated however the results where only presented from the perspective of UK investors. Tests developed by Jenrich (1970) and Box (1949) were used in order to test the stability of both the correlation- and the covariance-matrix over four equal sub-periods. The hypothesis that the correlation matrix is stable over any two time periods could not be rejected at the 95% level for most of the sub-periods investigated, however the hypothesis of covariance stability could be rejected at the 95% level for most of the sub-periods. The instability of the covariance matrix was attributed to the time varying nature of a few of the variances of the different countries. It was thus found that empirical evidence supported the hypothesis that the correlation structure is stable over time, however the support for the stability of the covariance matrix was found to be much weaker.

Eichholtz (1995) investigated the covariance structure of international property (Belgium, France, The UK, Australia, Japan, Singapore, Canada, and the USA) share returns during the period February 1973 to May 1993. Jenrich (1970) tests were also used in order to test the stability of the covariance matrix of international property share returns. It was found that the covariance matrix was not stable over time where as the correlation matrix was stable between certain time periods and unstable between others. Eichholtz (1996) further emphasizes that *"the instability in the covariance matrix limits the use of standard portfolio models to determine the allocation of international real estate security investments."*

Grubel and Fadner (1971), Lessard (1973), Levy and Sarnat (1970), Ripley (1973) and Solnik (1974) have shown that international diversification is beneficial to investors since it reduces the risk of their portfolios due to the fact that international markets were uncorrelated. With the development of new technology, increased capital mobility and the lifting of exchange controls between various countries, many researchers felt that international markets were becoming increasingly integrated thereby reducing the benefits of diversifying share portfolios across different international markets. Hillard (1979) investigated ten countries during the energy crisis of 1973 and 1974 and found that stock markets on the same continent moved together, where as distant markets were generally unrelated. The following section highlights

the different research that has been undertaken in the international community in order to understand how different stock markets are inter-related.

7.2 International Studies

Von Furstenberg and Jeon (1989) used Vector Auto-Regressive models over the period 1986-1988 to investigate the correlations between the USA, the UK, Germany and Japan. They concluded that international markets were becoming more dependent on one another after the market crash of 1987 although they failed to find any significant industry factors that impacted upon the correlation matrix between the countries investigated. Eun and Shim (1989) and Schollhammer and Sand (1985) found that most foreign markets responded to the American market within one to two days, thus indicating that international stock markets were becoming more integrated.

King and Wadhwani (1990) and Bertero and Mayer (1990) used high frequency data surrounding the 1987 stock market crash and showed that international correlation increased during periods of stock market crisis. This hypothesis originated from the fact that most international markets experienced significant losses during the 1987 stock market crash thereby leading researchers to believe that international markets were becoming increasingly correlated and more integrated such that events occurring on different markets would impact other markets simultaneously or in a short period of time after the occurrence of the event.

King and Wadhwani (1990) investigated intra-day data from the stock markets in London, New York and Tokyo from July 1987 to February 1988. It was found that the London market, which opens before the New York market, experienced an increase in volatility around the time that the New York stock market opened each day. This suggested that traders in London were digesting news specific to the New York market and acting upon it when the New York market opened indicating that these two markets were not totally uncorrelated.

Koch and Koch (1991) examined the correlation matrix of eight markets using daily data for three separate years (1972, 1980 and 1987). They utilized dynamic simultaneous equation models in order to model the contemporaneous as well as the lead/lag relationships across different stock markets. It was found that international

markets were efficient since stock markets responded quickly to price adjustments in other international markets. This indicated that international stock markets were dependent on one another.

King *et al.* (1992) utilized a multifactor GARCH process in order to model the links between international stock markets. They claimed that the interdependence between markets was very small and stressed that the use of economic variables was unlikely to explain the variation in the correlation matrix between different international stock markets. Ratner (1992) however showed that the international correlations remained constant over the period 1973-1989.

Arshanapalli and Doukas (1993) undertook an empirical investigation into the dynamic linkages between international markets during the October 1987 crash. Daily closing prices for five indices were used to undertake Granger-Engle (1987) two-step cointegration and error-correction methods. It was found that international markets had become increasingly inter-related after the 1987 crash. Evidence in support of a significant relationship between the US and European markets was found, where as the evidence in support of a strong relationship between the US market and that of Japan was found to be lacking.

Erb *et al.* (1994) have shown that correlations tended to vary over time according to the phase of the business cycle. It was shown that the correlation between different countries was highest when the business cycles were out-of-phase (i.e. both countries were experiencing low growth rates and the market was contracting). Ragunathann *et al.* (1997) found similar results when investigating the correlation between the USA and the Australian market. They suggested that the US business cycle had a larger impact on the Australian business cycle since the correlations between the two countries were high whenever the US market contracted.

Ragunathann and Mitchell (1997) extended the research done by Erb *et al.* (1994). Multivariate GARCH models were used in order to model the time varying correlations of eighteen Morgan Stanley Capital International (MSCI) country indices over the period January 1970 to May 1995. They were unable to reject the hypothesis of a constant correlation between the different markets however they found that the pairwise correlation between Germany, Hong Kong, Norway and the US index were time

varying. Note however that these results were dependent on the index against which the countries were being compared to. For example, when the correlations were compared relative to the world index, seven of the countries' pairwise correlations were found to be time varying compared to four when the correlations were compared to the US index.

Ragunathann and Mitchell (1997) further investigated whether or not the cross-correlations in the 1980's and the 1990's were significantly different from those in the 1970's by regressing the time varying correlations against dummy variables for the 1980's, the 1990's and the October 1987 crash. It was found that relative to the MSCI US index that there existed a positive time trend for Hong Kong and a slight negative time trend for the US/World Index. It was also found that the correlations between the MSCI US and the two indices were significantly different in the 1980's compared to the 1970's thus indicating that the correlations have changed over time. Similar results were found when the comparisons were made relative to the World Index. Note however that the 1987 stock market crash was found not to have any impact on the time varying correlations between the different countries even though from a graphical point of view it could be seen that the time varying correlations increased for all countries during the 1987 crisis.

Longin and Solnik (1995) examined the monthly excess equity returns for seven major countries (Germany, France, UK, Switzerland, Japan, Canada and USA) over the period 1960 to 1990. They concluded that the international correlation- and covariance- matrix was unstable over the entire period investigated. A multivariate GARCH(1,1) model with a constant conditional covariance structure was able to capture the evolution of the conditional covariance matrix over time. Their research also indicated that the correlation between the major international markets had increased over the thirty years investigated. They also noted the asymmetric nature of the international correlation matrix. Gourieroux and Monfort's (1992) threshold GARCH model were used to model the asymmetry observed in the conditional correlation matrix. It was tentatively concluded that correlations seemed to increase during periods of *"high turbulence"* and that there existed a positive relationship between the correlations and the variances of different countries.

Solnik *et al.* (1996) studied the correlations across international bond markets

as well as across stock markets over the period 1958 to 1995 for shares, and 1959 to 1995 for bonds. Their study focused on the USA, Germany, France, the UK, Switzerland and Japan. It was found that international correlations for both shares and bonds fluctuated over time. Evidence was also found that international correlations increased significantly during periods of high market volatility thus highlighting the fact that international markets were beginning to become increasingly correlated. The correlation between the USA share returns and the countries investigated had generally increased over the period investigated however the correlation did not increase significantly over the period 1985 to 1986. It was however found that the French and the German market experienced increased correlation due to their participation in the European Union. It was also found that share- and bond- returns were slightly correlated. The authors suggested that this indicated that domestic factors still strongly affected local asset prices.

The stability of the covariance matrix of asset returns is important to investors who utilizes a mean-variance approach when undertaking portfolio construction since the constructed portfolios will only be representative of future optimal portfolios if the covariance matrix is relatively stationary. This implies that if the covariance of asset returns is not stable that one might have to estimate covariance matrices using relatively short periods of returns. In this chapter we test for the stability of the correlation-and the covariance-matrix. The chapter proceeds as follows: Section 7.2 discusses the Jenrich (1970) methodology utilized. Section 7.3 describes the Share returns used during the course of this chapter. Section 7.4 presents the results concerning the stability of the correlation-and the covariance-matrix of South African Share returns. Section 7.5 investigates whether or not the conditional variances of South African Share returns can be modelled using different volatility models and concluding remarks are given in Section 7.6.

7.3 The Jenrich (1970) methodology

The methodology developed by Jenrich (1970) is used in order to test the stability of the unconditional correlation- and the covariance-matrix of South African share returns over the period January 1990 to December 2000. The following Jenrich χ^2 test statistic used in order to test the hypothesis that two correlation matrices are

equal is as follows:

$$T_{corr} = \frac{1}{2}tr\left(Z^2\right) - diag'\left(Z\right)S^{-1}diag\left(Z\right) \tag{7.1}$$

where R_1 and R_2 are the two sample autocorrelation matrices that we wish to compare.

- $R = \frac{n_1R_1+n_2R_2}{n_1+n_2}$ where n_1 and n_2 are the sample sizes on which R_1 and R_2 are based on.
- $c = \frac{n_1n_2}{n_1+n_2}$
- $Z = \sqrt{c}R^{-1}\left(R_1 + R_2\right)$
- and $S = \delta_{ij} + r^{ij}r_{ij}$ where $\delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & if \quad i \neq j \end{cases}$, (i.e δ_{ij} is the Kronecker delta) , r_{ij} are the (i,j)'th elements of R and r^{ij} are the (i,j)'th elements of R^{-1}

T_{corr} is assymptotically χ^2 distributed with $\frac{p(p-1)}{2}$ degrees of freedom, where p is the dimension of the correlation matrices. The hypothesis $H_0 : R_1 = R_2$ is thus rejected if $T_{corr} > \chi^2_{\frac{p(p-1)}{2}}$ for a given significance level. Note that the small sample properties of T_{corr} is not known.

The test statistic in order to test the null hypothesis that two covariance matrices are equal is:

$$T_{cov} = \frac{1}{2}tr\left(Z^2\right) \tag{7.2}$$

where T_{cov} is assymptotically χ^2 distributed with $\frac{p(p+1)}{2}$ degrees of freedom. If two covariance matrices or two correlation matrices are found to be significantly different from one another then it can be concluded that the correlation- and the covariance-matrices might not generated by a stationary distribution over time. This might suggest that the correlation- and the covariance- matrices could be time varying.

7.4 Description of the Data

This chapter examines the stability of the correlation and the covariance structure of share returns of the largest market capitalization companies listed on the JSE. The analysis utilizes thirty four return series that had a price history during the period January 1990 to December 2000. Appendix one provided a list of the series included in the analysis.

Sample statistics of the companies under investigation for the full period are included in appendix two. The mean weekly continuously compounded return varies significantly across the different shares with Didata generating the highest return of 0.9732% per week and Durban Deep generating the lowest return of -0.3074% per week. The riskiest shares have been the resources companies. Durban Deep has the highest standard deviation of 7.9664%. Other risky resources shares in the sample were Harmony (with a standard deviation of 6.6828%), Gold Fields (with a standard deviation of 5.1305%), Anglos Gold (with a standard deviation of 5.1175%) and Implats (with a standard deviation of 4.8619%). None of the share returns show any indication of exessive skewness although most do exhibit excessive kurtosis, indicating that the normal distribution might not accurately reflect the return generating process of most of the share return series in the data set examined. This is further emphasized by the fact that the Jarque-Bera statistic of all of the series considered is greater than $\chi_2^{2(0.05)} = 5.991$ where the null hypothesis of the test is that of normality.

The correlations between the series are very different and range between -0.109 to 0.606. Most of the shares have a correlation of at least 0.2. This should be expected since all of the shares are impacted upon by domestic factors. This fact is further emphasized when one examines the correlations between the companies and the weekly continuously compounded returns of the JSE in Appendix one. It can be seen that all of the shares have a correlation of at least 0.4 with the JSE (except for two resources companies namely Durban Deep and Harmony). The diversification potential is significantly reduced due to this fact since most of the companies are inter-related with few having either small positive- or negative-correlation.

7.5 The Stability of Correlations and Covariances

Jenrich (1970) tests are used in order to determine whether or not the covariance- and the correlation- matrix are constant over time. The period under investigation (Jan 1990 to Dec 2000) is split into three, four and fives year sub-periods. Consecutive periods as well as non-consecutive sub-periods are examined.

The null hypothesis that the correlation matrix is stationary (stable) over time is first tested. Table 1 below displays the results of the Jenrich tests for the stability of the correlation matrix over sub-periods that are three years in length. It consists

of six sets of results such that each are associated with different starting periods. For example, set one compares the results found by using Jan 1990-Dec 1992 as the starting period of the analysis. The tests indicate that the correlation matrix is not stable over any of the periods investigated when one uses a three year returns history since the Jenrich χ^2 statistic (shown in the third column of the tables) is greater than the cut-off value of 628.5247 $\left(\chi^{2(0.05)}_{(561)}\right)$. The test statistic of consecutive sub-periods is found to be smaller than that of non-consecutive sub-periods. For example, if one compares the Jenrich test statistic for Jan 1990-Dec 1992 and Jan 1993-Dec 1995 with that any of the Jan 1990-Dec 1992 and Jan 1994-Dec 1996 it can be seen that the former has a lower test statistic than the latter.

From table 1 below it can be seen that the test statistic of consecutive sub-periods is smaller than that of non-consecutive sub-periods and that this result is not influenced by using a different starting period for the analysis.

3 Year Periods Compared		
Period 1	**Period 2**	**Jenrich Chi-square**
Jan 1990-Dec 1992	Jan 1993-Dec 1995	740.845
Jan 1990-Dec 1992	Jan 1994-Dec 1996	747.732
Jan 1990-Dec 1992	Jan 1995-Dec 1997	799.427
Jan 1990-Dec 1992	Jan 1996-Dec 1998	891.174
Jan 1990-Dec 1992	Jan 1997-Dec 1999	922.252
Jan 1990-Dec 1992	Jan 1998-Dec 2000	922.950
Jan 1991-Dec 1993	Jan 1994-Dec 1996	705.220
Jan 1991-Dec 1993	Jan 1995-Dec 1997	766.534
Jan 1991-Dec 1993	Jan 1996-Dec 1998	904.325
Jan 1991-Dec 1993	Jan 1997-Dec 1999	938.189
Jan 1991-Dec 1993	Jan 1998-Dec 2000	928.116
Jan 1992-Dec 1994	Jan 1995-Dec 1997	753.805
Jan 1992-Dec 1994	Jan 1996-Dec 1998	872.854
Jan 1992-Dec 1994	Jan 1997-Dec 1999	909.019
Jan 1992-Dec 1994	Jan 1998-Dec 2000	894.745
Jan 1993-Dec 1995	Jan 1996-Dec 1998	870.744
Jan 1993-Dec 1995	Jan 1997-Dec 1999	892.889
Jan 1993-Dec 1995	Jan 1998-Dec 2000	867.823
Jan 1994-Dec 1996	Jan 1997-Dec 1999	797.205
Jan 1994-Dec 1996	Jan 1998-Dec 2000	791.199
Jan 1995-Dec 1997	Jan 1998-Dec 2000	867.853
Cut off value	628.525	

Table 1: Jenrich (Correlation matrix) Test Statistics for all three year sub-periods

The results of the other Jenrich (correlation matrix) tests are shown in appendices

three to five. It was found that the correlation matrix is not stable over time for all of the sub-periods compared. These results are different to the results of Kaplanis (1988) and Longin and Solnik (1993) although their research focused on international diversification/investment where as this research focuses only on South African shares. It was also found that in general consecutive sub-periods have lower test statistics than non-consecutive sub-periods. This occurs for all of the period lengths considered, further emphasizing that the correlation matrix does not appear to be stable over time. Different starting periods were examined in order to identify whether or not this might have any impact on the results. It was found that the starting period does not change the outcome of the tests. This indicates that the correlation matrix could be time varying.

The instability of the correlation matrix could be attributed to the fact that the South African market experienced major political developments during the 1990's which could have impacted on the returns generated by shares listed on the JSE. By 1989 South Africa experienced negative growth rates and by 1989 the country's gold and foreign reserves had been depleted. Increasing pressure from the international community and an ailing economy forced the De Klerk government to alter its political stance. The ANC won the first democratic elections held on 27 April 1994 by gaining 62.5% of the vote. South Africa rejoined the international political community when it joined the Organisation of African Unity in May 1994, rejoined the Common Wealth in June 1994 as well as becoming a member of the United Nations in the same month.

The opening of the South African market to international investment could also have attributed to the instability of the correlation matrix. South Africa is classified as an emerging market and thus international market crisis such as the collapse of the Mexican Peso in 1994-1995 and The East Asian financial collapse of 1997-1998, the Nasdaq meltdown of 2000 and the Russian,Turkish and the Argentinean defaults all had negative impacts on the South African economy.

The Jenrich test results for the stability of the covariance matrix are shown in appendices six to eight. It is found that the covariance matrix does not appear to be stable over time. This conclusion is reached even if one considers different starting periods in order to undertake the tests.

Since the covariance-and the correlation- matrix is found to be non-stationary

over the period investigated, it suggests that we could attempt to model the way that they change over time. The correlation matrix could be non-stationary due to time varying covariances or due to time varying variances. The above tests indicate that the covariance matrix is non-stationary. A first attempt in order to understand the time varying nature of the correlation matrix is to investigate whether or not the conditional variances of the different share return series are time varying by comparing the modelling performance of several *univariate* conditional variance models, within the GARCH class of models. The following specifications of GARCH models are analysed: ARCH, GARCH. Two different specifications of the conditional distribution are considered namely, normal and standardised students t. Model selection is undertaken by considering Schwarz and ICOMP.

7.6 Garch Modelling

As a first attempt, the weekly log returns (r_t) is modelled by identifying an AR process for the mean of of r_t where as GARCH(p,q) models are used to model the conditional variance equation. The joint estimation undertaken is of the following form:

$$r_t = \phi_0 + \sum_{i=1}^{o^*} \phi_i r_{t-1} + \sigma_t \varepsilon_t \tag{7.3}$$

where $r_t^* = r_t - \phi_0 + \sum_{i=1}^{o^*} \phi_i r_{t-1}$, ε_t is a random variable with mean 0 and unit variance In order to limit the amount of autoregressive and conditional variance terms, o^*, p and q are allowed to have a maximum of 5,1 and 1. The conditional variance equation is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^{*2} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{7.4}$$

Joint estimation has to be undertaken since the autoregression orders generally change when one uses a two-step approach of first estimating an AR process for the mean of r_t and then fitting GARCH models to the residuals. Parameter estimation is undertaken by utilizing the Eviews 3 package. Model selection is undertaken by utilizing Schwarz(1978) and ICOMP in conjuction with the correllogram and Ljung-Box tests of the squared standardized residuals.

and kurtosis of the series with those from the normal distribution. The statistic is computed as:

$$JB = \frac{n-k}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

where S is the skewness, K is the kurtosis, and k represents the number of estimated coefficients used to create the series. Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is χ^2 distributed with 2 degrees of freedom.

The normality assumption is rejected for all of the models. In the following section we attempt to model the excess kurtosis observed by using a heavy tailed distribution.

Model Selected by minimising BIC					Model Selected by minimising ICOMP				
	Coefficient	Std. Error	z-Statistic	Prob.		Coefficient	Std. Error	z-Statistic	Prob.
Mean Equation					**Mean Equation**				
C	0.246	0.132	1.861	0.063	C	0.229	0.151	1.517	0.129
AR(1)	0.404	0.053	7.574	0.000	AR(1)	0.361	0.044	8.204	0.000
AR(2)	-0.098	0.061	-1.604	**0.109**	AR(4)	0.084	0.041	2.055	0.040
AR(3)	-0.043	0.064	-0.673	**0.501**					
AR(4)	0.110	0.057	1.929	**0.054**					
AR(5)	-0.023	0.052	-0.436	**0.663**					
Variance Equation					**Variance Equation**				
C	0.394	0.107	3.695	0.000	C	0.361	0.103	3.495	0.001
ARCH(1)	0.095	0.021	4.428	0.000	ARCH(1)	0.087	0.021	4.203	0.000
GARCH(1)	0.803	0.047	17.047	0.000	GARCH(1)	0.820	0.046	17.692	0.000

Table 2: Model Selection, BIC compared to ICOMP

substituting the zeros with the return based on the last trade made during that week if the share price at the end of the week is the same as the price at the beginning of the week. Note however that this procedure has not been undertaken here.

The estimated v parameters range between 2.5 to 11.24. The average estimate of v (excluding shares with more than 10% of zeros) is 6 with a standard deviation of 2.02. This indicates that on average the share returns departed from normality. The return series had higher peaks and fatter tails than the normal distribution.

In the following section we test the validity of the error distribution assumption.

Goodness of Fit tests

In order to check the distribution assumption we constructed the p-p plots (probability -probability plots) associated with the standardised residuals of each of the models under the assumption that the error distribution followed a standardised t distribution. Kolmogorov-Smirnov test were also used in order to test the distribution assumption.

Assume that we observe n observations from a cumalative distribution function $F(x)$. A **p-p plot** is constructed by plotting $F(x_k)$ versus $\frac{k}{n+1}$ for all observations where x_k is the k^{th} order statistic. If data is drawn from $F(x)$ the points should plot along a 45 degree line through the origin. Departures from the straight indicates that the data was generated by some other distribution. The cumalative distribution function of a standardised t distribution with v degrees of freedom is equal to:

$$F(x^*) = \int_{-\infty}^{x^*\sqrt{\frac{v}{v-2}}} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi}} \left(1+\frac{y^2}{v}\right)^{-\frac{v+1}{2}} dy$$

These values were used in order to calculate the Kolmogorov-Smirnov test statistic.

The Kolmogorov-Smirnov test statistic:

$$D = \max\left\{\left|F(x_i) - \frac{2i-1}{2n}\right|\right\} \text{ for } i = 1, 2, ..n$$

is used to test:

H_o : The data follows a standardised t distribution with v degrees of freedom

H_1 : The data follows some other distribution

The test statistic compares the cumalative distribution of each observation under the null hypothesis, to the cumalants of a uniform distribution. The critical points used for this test are based on the asymptotic distribution of D as tabulated in Miller (1956). The right tail probability for a two-sided test are 0.0688 (1%), 0.0574 (5%) and 0.0515 (10%).

The estimated D statistics are displayed in appendix 14. The results indicate that the distribution assumption cannot be rejected at the 1%, 5% or 10% levels for majority of the share return series. The null hypothesis is however rejected for Anggold (10% level), Bidvest, Firstrand and Libhold all at the 5% level.

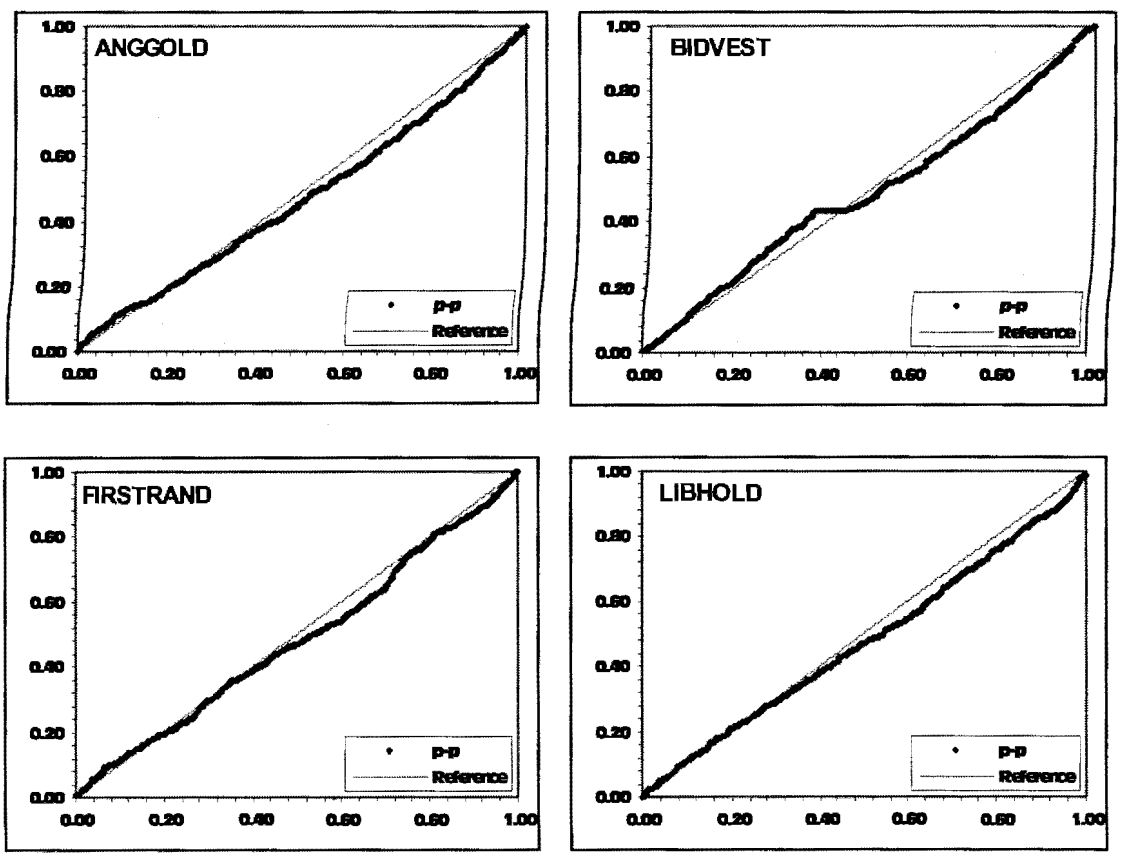


Figure 1: p-p plots of Anggold, Bidvest, Firstrand and Libhold

The same conclusion can be drawn by examining the p-p plots associated with Anggold, Bidvest, Firstrand and Libhold. Figure 1 above displays the p-p plots for the four shares. It can be seen that there appears to be significant departures from the 45% reference line, indicating that the standardised t distribution does not adequately these series.

7.6.2 Conclusions

This chapter investigated the stabality of the covariance- and the correlation-matrix of thirty three companies as well as the JSE All share return index. It was found that both the covariance- and the correlation-matrix does appear to change over time. An attempt was made to model the conditional variance of these series by utilising GARCH models. It was found that the normality assumption for the error distribution was not appropriate and that in certain cases a standardised t distribution was more appropriate.

7.7 Aspects warranting further study

Covariance and correlation matrix instability could be caused due to time varying variances and covariances. This chapter only set about to model the time varying conditional variances of the series. Both aspects could be modelled by using a multivariate approach.

7.8 List of Appendices

Appendix 1: The data included in the analysis as well as their correlation with the JSE All share index

Appendix 2: Weekly Summary statistics

Appendix 3: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of three years

Appendix 4: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of four years

Appendix 5: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of five years

Appendix 6: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of three years

Appendix 7: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of four years

Appendix 8: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of five years

Appendix 9: Models Selection undertaken by minimising BIC

Appendix 10: Models Selection undertaken by minimising ICOMP

Appendix 11: A comparison of the number of insignificant parameters when using BIC and ICOMP

Appendix 12: Final Model under the normality assumption

Appendix 13 (a): Parameter Estimation under the standardised t distribution

Appendix 13 (b): Parameter Estimation under the standardised t distribution

Appendix 14: Goodness of Fit tests

Appendix 1: The data included in the analysis as well as their correlation with the JSE All share index

SERIES INCLUDED			
ABI	0.451	JOHNNIC	0.644
ABSA	0.581	LIBERTY	0.593
ANGGOLD	0.427	LIBHOLD	0.574
ANGLOS	0.702	METCASH	0.331
BARWORLD	0.574	NAMPAK	0.541
BIDVEST	0.457	NEDCOR	0.611
COROHOLD	0.421	PICKNPAY	0.441
DBN DEEP	0.25	REMGRO	0.597
DIDATA	0.465	REUNERT	0.41
FIRST RAND	0.484	RICHEMONT	0.477
GFIELDS	0.233	SAB	0.618
HARMONY	0.305	SANTAM	0.404
IMPERIAL	0.442	SAPPI	0.483
IMPLATS	0.441	SASOL	0.474
INVESTEC	0.504	STANBIC	0.53
ISCOR	0.446	TIGERBRAND	0.504
JSE	1	TONGAAT	0.47

Appendix 3: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of three years

3 Year Periods Compared		
Period 1	Period 2	Jenrich Chi-square
Jan 1990-Dec 1992	Jan 1993-Dec 1995	740.845
Jan 1990-Dec 1992	Jan 1994-Dec 1996	747.732
Jan 1990-Dec 1992	Jan 1995-Dec 1997	799.427
Jan 1990-Dec 1992	Jan 1996-Dec 1998	891.174
Jan 1990-Dec 1992	Jan 1997-Dec 1999	922.252
Jan 1990-Dec 1992	Jan 1998-Dec 2000	922.950
Jan 1991-Dec 1993	Jan 1994-Dec 1996	705.220
Jan 1991-Dec 1993	Jan 1995-Dec 1997	766.534
Jan 1991-Dec 1993	Jan 1996-Dec 1998	904.325
Jan 1991-Dec 1993	Jan 1997-Dec 1999	938.189
Jan 1991-Dec 1993	Jan 1998-Dec 2000	928.116
Jan 1992-Dec 1994	Jan 1995-Dec 1997	753.805
Jan 1992-Dec 1994	Jan 1996-Dec 1998	872.854
Jan 1992-Dec 1994	Jan 1997-Dec 1999	909.019
Jan 1992-Dec 1994	Jan 1998-Dec 2000	894.745
Jan 1993-Dec 1995	Jan 1996-Dec 1998	870.744
Jan 1993-Dec 1995	Jan 1997-Dec 1999	892.889
Jan 1993-Dec 1995	Jan 1998-Dec 2000	867.823
Jan 1994-Dec 1996	Jan 1997-Dec 1999	797.205
Jan 1994-Dec 1996	Jan 1998-Dec 2000	791.199
Jan 1995-Dec 1997	Jan 1998-Dec 2000	867.853
Cut off value	628.525	

Appendix 4: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of four years

4 Year Periods Compared		
Period 1	Period 2	Jenrich Chi-square
Jan 1990-Dec 1993	Jan 1994-Dec 1997	823.004
Jan 1990-Dec 1993	Jan 1995-Dec 1998	1003.738
Jan 1990-Dec 1993	Jan 1996-Dec 1999	1021.186
Jan 1990-Dec 1993	Jan 1997-Dec 2000	1034.023
Jan 1991-Dec 1994	Jan 1995-Dec 1998	974.644
Jan 1991-Dec 1994	Jan 1996-Dec 1999	993.938
Jan 1991-Dec 1994	Jan 1997-Dec 2000	974.192
Jan 1992-Dec 1995	Jan 1996-Dec 1999	966.762
Jan 1992-Dec 1995	Jan 1997-Dec 2000	944.407
Jan 1993-Dec 1996	Jan 1997-Dec 2000	890.766
Cut off value	628.525	

Appendix 5: Results of the Jenrich tests for the stability of the correlation matrix over sub periods of five years

5 Year Periods Compared		
Period 1	**Period 2**	**Jenrich Chi-square**
Jan 1990-Dec 1994	Jan 1995-Dec 1999	1071.089
Jan 1990-Dec 1994	Jan 1996-Dec 2000	1060.305
Jan 1991-Dec 1995	Jan 1996-Dec 2000	1020.746
Cut off value	**628.525**	

Appendix 6: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of three years

3 Year Periods Compared		
Period 1	**Period 2**	**Jenrich Chi-square**
Jan 1990-Dec 1992	Jan 1993-Dec 1995	925.143
Jan 1990-Dec 1992	Jan 1994-Dec 1996	990.438
Jan 1990-Dec 1992	Jan 1995-Dec 1997	1054.055
Jan 1990-Dec 1992	Jan 1996-Dec 1998	1267.625
Jan 1990-Dec 1992	Jan 1997-Dec 1999	1404.564
Jan 1990-Dec 1992	Jan 1998-Dec 2000	1484.700
Jan 1991-Dec 1993	Jan 1994-Dec 1996	970.984
Jan 1991-Dec 1993	Jan 1995-Dec 1997	1026.309
Jan 1991-Dec 1993	Jan 1996-Dec 1998	1298.640
Jan 1991-Dec 1993	Jan 1997-Dec 1999	1448.796
Jan 1991-Dec 1993	Jan 1998-Dec 2000	1522.663
Jan 1992-Dec 1994	Jan 1995-Dec 1997	952.809
Jan 1992-Dec 1994	Jan 1996-Dec 1998	1246.969
Jan 1992-Dec 1994	Jan 1997-Dec 1999	1410.550
Jan 1992-Dec 1994	Jan 1998-Dec 2000	1500.029
Jan 1993-Dec 1995	Jan 1996-Dec 1998	1269.777
Jan 1993-Dec 1995	Jan 1997-Dec 1999	1463.791
Jan 1993-Dec 1995	Jan 1998-Dec 2000	1575.262
Jan 1994-Dec 1996	Jan 1997-Dec 1999	1401.572
Jan 1994-Dec 1996	Jan 1998-Dec 2000	1513.444
Jan 1995-Dec 1997	Jan 1998-Dec 2000	1450.920
Cut off value	**664.486**	

Appendix 7: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of four years

4 Year Periods Compared		
Period 1	**Period 2**	**Jenrich Chi-square**
Jan 1990-Dec 1993	Jan 1994-Dec 1997	1093.8729
Jan 1990-Dec 1993	Jan 1995-Dec 1998	1379.3006
Jan 1990-Dec 1993	Jan 1996-Dec 1999	1536.467
Jan 1990-Dec 1993	Jan 1997-Dec 2000	1649.9947
Jan 1991-Dec 1994	Jan 1995-Dec 1998	1343.5118
Jan 1991-Dec 1994	Jan 1996-Dec 1999	1534.6374
Jan 1991-Dec 1994	Jan 1997-Dec 2000	1645.8919
Jan 1992-Dec 1995	Jan 1996-Dec 1999	1582.0775
Jan 1992-Dec 1995	Jan 1997-Dec 2000	1716.7433
Jan 1993-Dec 1996	Jan 1997-Dec 2000	1629.9616
Cut off value	**664.4855998**	

Appendix 8: Results of the Jenrich tests for the stability of the covariance matrix over sub periods of five years

5 Year Periods Compared		
Period 1	**Period 2**	**Jenrich Chi-square**
Jan 1990-Dec 1994	Jan 1995-Dec 1999	1557.642
Jan 1990-Dec 1994	Jan 1996-Dec 2000	1689.976
Jan 1991-Dec 1995	Jan 1996-Dec 2000	1781.376
Cut off value	**664.486**	

Appendix 9: Models Selection undertaken by minimising BIC

Model Selection undertaken by minimising BIC

Series	Model	BIC	ICOMP
ABI	arch(1,1) ar(1) ar(2) ar(5)	5.24	2942.36
ABSA	arch(1,1) ar(1) ar(2) ar(4) ar(5)	5.57	3128.95
ANGGOLD	arch(1,1) ar(1) ar(2) ar(5)	6.06	3423.63
ANGLOS	arch(1,1) ar(1) ar(2) ar(5)	5.53	3108.45
BARWORLD	arch(1,1) ar(1) ar(2) ar(5)	5.43	3052.16
BIDVEST	arch(1,1) ar(1) ar(3) ar(5)	5.06	2839.64
COROHLD	arch(1,1) ar(1) ar(2) ar(4) ar(5)	6.13	3469.20
DBNDEEP	**arch(1,0)** ar(1) ar(2) ar(5)	6.72	3803.52
DIDATA	arch(1,1) ar(1) ar(2) ar(5)	5.82	3278.41
FIRSTRAND	arch(1,1) ar(1) ar(5)	5.74	3239.04
GFIELDS	arch(1,1) ar(1) ar(2) ar(4) ar(5)	6.06	3406.59
HARMONY	arch(1,1) ar(1) ar(5)	6.58	3715.70
IMPERIAL	arch(1,1) ar(1) ar(2) ar(4) ar(5)	5.42	3046.80
IMPLATS	arch(1,1) ar(1) ar(2) ar(5)	5.92	3336.32
INVSTEC	arch(1,1) ar(1) ar(2) ar(3) ar(5)	5.19	2910.36
ISCOR	arch(1,1) ar(1) ar(2) ar(5)	5.81	3268.19
JOHNNIC	arch(1,1) ar(1) ar(2) ar(5)	5.73	3227.47
JSEW	arch(1,1) ar(1) ar(2) ar(3) ar(4) ar(5)	4.22	2350.67
LIBERTY	arch(1,1) ar(1) ar(2) ar(5)	5.36	3011.01
LIBHOLD	arch(1,1) ar(1) ar(2) ar(5)	5.05	2832.61
METCASH	arch(1,1) ar(1) ar(2) ar(5)	5.88	3313.25
NAMPAK	arch(1,1) ar(1) ar(2) ar(4) ar(5)	5.36	3004.86
NEDCOR	arch(1,1) ar(1) ar(2) ar(3) ar(5)	5.36	3008.69
PICKNPAY	arch(1,1) ar(1) ar(2) ar(3) ar(4) ar(5)	5.52	3095.28
REMGRO	arch(1,1) ar(1) ar(2) ar(3) ar(5)	5.58	3138.72
REUNERT	arch(1,1) ar(1) ar(3) ar(5)	5.74	3235.44
RICHEMONT	arch(1,1) ar(1) ar(2) ar(3) ar(4) ar(5)	5.19	2904.44
SABREWS	arch(1,1) ar(1) ar(2) ar(5)	5.21	2924.03
SANTAM	**arch(1,0)** ar(1) ar(2) ar(5)	5.45	3067.95
SAPPI	arch(1,1) ar(1) ar(2) ar(5)	5.83	3279.91
SASOL	**arch(1,0)** ar(1) ar(3) ar(5)	5.74	3236.18
STANBIC	arch(1,1) ar(1) ar(2) ar(5)	5.28	2961.46
TIGBRANDS	arch(1,1) ar(1) ar(2) ar(3) ar(5)	5.42	3043.10
TONGAAT	arch(1,1) ar(1) ar(2) ar(4) ar(5)	5.63	3162.90

Appendix 10: Models Selection undertaken by minimising ICOMP

Model Selection undertaken by minimising ICOMP

Series	Model	SCHARZ	ICOMP
ABI	arch(1,1) ar(1) ar(2)	5.22	2955.58
ABSA	arch(1,1) ar(1)	5.54	3143.61
ANGGOLD	arch(1,1) ar(1) ar(2)	6.05	3439.55
ANGLOS	arch(1,1) ar(1)	5.52	3140.54
BARWORLD	arch(1,1) ar(1)	5.42	3077.80
BIDVEST	arch(1,1) ar(1)	5.03	2857.72
COROHLD	arch(1,1) ar(1) ar(4)	6.13	3473.78
DBNDEEP	**arch(1,0)** ar(1) ar(2)	6.71	3818.71
DIDATA	arch(1,1) ar(1) ar(2) ar(5)	5.82	3278.41
FIRSTRAND	arch(1,1) ar(1)	5.73	3260.29
GFIELDS	arch(1,1) ar(1) ar(2)	6.05	3428.57
HARMONY	arch(1,1) ar(1)	6.57	3751.44
IMPERIAL	arch(1,1) ar(1) ar(2)	5.40	3056.88
IMPLATS	arch(1,1) ar(1) ar(2)	5.91	3350.72
INVSTEC	arch(1,1) ar(1) ar(2)	5.17	2924.23
ISCOR	arch(1,1) ar(1) ar(2)	5.80	3285.06
JOHNNIC	arch(1,1) ar(1)	5.72	3251.86
JSEW	arch(1,1) ar(1) ar(4)	4.20	2361.85
LIBERTY	arch(1,1) ar(1) ar(5)	5.35	3014.93
LIBHOLD	arch(1,1) ar(1) ar(2) ar(5)	5.05	2832.61
METCASH	arch(1,1) ar(1)	5.86	3334.58
NAMPAK	arch(1,1) ar(1)	5.34	3029.83
NEDCOR	arch(1,1) ar(1)	5.33	3026.29
PICKNPAY	arch(1,1) ar(1) ar(2)	5.51	3122.00
REMGRO	arch(1,1) ar(2)	5.57	3168.96
REUNERT	arch(1,1) ar(1)	5.72	3255.84
RICHEMONT	**arch(1,0)** ar(1) ar(3)	5.16	2914.64
SABREWS	arch(1,1) ar(1) ar(5)	5.20	2926.80
SANTAM	**arch(1,0)** ar(1) ar(2)	5.44	3079.59
SAPPI	arch(1,1) ar(1) ar(2)	5.81	3293.72
SASOL	arch(1,1) ar(1) ar(3)	5.73	3249.67
STANBIC	arch(1,1) ar(1) ar(2)	5.26	2973.38
TIGBRANDS	arch(1,1) ar(1) ar(2) ar(3)	5.41	3052.23
TONGAAT	arch(1,1) ar(1)	5.61	3191.64

Appendix 11: A comparison of the number of insignificant parameters when using BIC and ICOMP

Number of insignificant autoregressive parameters and the insignificant lag					
	BIC			ICOMP	
ABI	1	ar(5)		0	
ABSA	3	ar(2) ar(4) ar(5)		0	
ANGGOLD	1	ar(5)		0	
ANGLOS	1	ar(5)		0	
BARWORLD	2	ar(2) ar(5)		0	
BIDVEST	2	ar(3) ar(5)		0	
COROHLD	1	ar(5)		0	
DBNDEEP	1	ar(5)		0	
DIDATA	1	ar(5)		1	**ar(5)**
FIRSTRAND	1	ar(5)		0	
GFIELDS	2	ar(4) ar(5)		0	
HARMONY	1	ar(5)		0	
IMPERIAL	2	ar(4) ar(5)		1	**ar(2)**
IMPLATS	1	ar(5)		0	
INVSTEC	2	ar(3) ar(5)		0	
ISCOR	2	ar(2) ar(5)		1	**ar(2)**
JOHNNIC	2	ar(2) ar(5)		0	
JSEW	4	ar(2) ar(3) ar(4) ar(5)		0	
LIBERTY	1	ar(5)		1	**ar(5)**
LIBHOLD	0			0	
METCASH	2	ar(2) ar(5)		0	
NAMPAK	2	ar(4) ar(5)		0	
NEDCOR	3	ar(2) ar(3) ar(5)		0	
PICKNPAY	2	ar(3) ar(5)		0	
REMGRO	3	ar(1) ar(3) ar(5)		0	
REUNERT	2	ar(3) ar(5)		0	
RICHEMONT	4	ar(2) ar(3) ar(4) ar(5)		0	
SABREWS	2	ar(2) ar(5)		1	**ar(5)**
SANTAM	1	ar(5)		0	
SAPPI	1	ar(5)		0	
SASOL	1	ar(5)		0	
STANBIC	1	ar(5)		0	
TIGBRANDS	1	ar(5)		0	
TONGAAT	2	ar(4) ar(5)		0	

Appendix 12: Final Model under the normality assumption

	Model Proposed	Jarque Bera	Kurtosis
ABI	arch(1,1) ar(1) ar(2)	1199.301	**10.023**
ABSA	arch(1,1) ar(1)	90.306	4.837
ANGGOLD	arch(1,1) ar(1) ar(2)	28.809	3.627
ANGLOS	arch(1,1) ar(1)	195.886	5.777
BARWORLD	arch(1,1) ar(1)	5.224	3.452
BIDVEST	arch(1,1) ar(1)	443.876	7.196
COROHLD	arch(1,1) ar(1) ar(4)	1295.893	**10.298**
DBNDEEP	arch(1,0) ar(1) ar(2)	87.371	4.786
DIDATA	arch(1,1) ar(1) ar(2)	1117.294	9.651
FIRSTRAND	arch(1,1) ar(1)	26.271	3.725
GFIELDS	arch(1,1) ar(1) ar(2)	28.368	3.745
HARMONY	arch(1,1) ar(1)	3082.250	**14.337**
IMPERIAL	arch(1,1) ar(1)	330.464	6.637
IMPLATS	arch(1,1) ar(1) ar(2)	95.850	4.986
INVSTEC	arch(1,1) ar(1) ar(2)	162.234	5.572
ISCOR	arch(1,1) ar(1)	45.566	4.372
JOHNNIC	arch(1,1) ar(1)	100.900	4.918
JSEW	arch(1,1) ar(1) ar(4)	1616.512	**10.869**
LIBERTY	arch(1,1) ar(1)	406.473	7.124
LIBHOLD	arch(1,1) ar(1) ar(2) ar(5)	181.596	5.721
METCASH	arch(1,1) ar(1)	1116.570	**9.608**
NAMPAK	arch(1,1) ar(1)	274.041	6.317
NEDCOR	arch(1,1) ar(1)	42.342	4.299
PICKNPAY	arch(1,1) ar(1) ar(2)	46.105	4.386
REMGRO	arch(1,1) ar(2)	84.661	4.839
REUNERT	arch(1,1) ar(1)	4357.875	**16.521**
RICHEMONT	arch(1,0) ar(1) ar(3)	1140.863	**9.769**
SABREWS	arch(1,1) ar(1)	12.879	3.698
SANTAM	arch(1,0) ar(1) ar(2)	265.318	6.304
SAPPI	arch(1,1) ar(1) ar(2)	98.445	5.021
SASOL	arch(1,1) ar(1) ar(3)	2162.387	**12.254**
STANBIC	arch(1,1) ar(1) ar(2)	104.855	5.034
TIGBRANDS	arch(1,1) ar(1) ar(2) ar(3)	443.233	7.088
TONGAAT	arch(1,1) ar(1)	628.752	8.088

Appendix 13 a: Parameter Estimation under the standardised t distribution

Key: The following table displays the mean and variance equations as well as their standard errors. The standard error are listed below the parameter estimates.

	c	AR 1	AR 2	AR 3	AR 4	AR 5		C*	ARCH 1	GARCH 1	V
ABI	0.198	0.330	-0.135					1.731	0.201	0.662	4.455
	-0.110	*0.042*	*0.043*					*0.645*	*0.076*	*0.091*	*0.728*
ABSA	0.272	0.179						0.900	0.094	0.853	6.686
	0.143	*0.041*						*0.367*	*0.032*	*0.045*	*1.791*
ANGGOLD	-0.305	0.264	-0.149					4.679	0.103	0.709	10.958
	0.206	*0.046*	*0.042*					*1.772*	*0.051*	*0.099*	*4.022*
ANGLOS	0.107	0.198						0.738	0.162	0.808	6.060
	0.134	*0.042*						*0.344*	*0.047*	*0.053*	*1.088*
BARWORLD	0.096	0.213						1.031	0.149	0.788	8.711
	0.141	*0.043*						*0.358*	*0.046*	*0.053*	*2.688*
BIDVEST	0.211	0.294						4.595	1.574	0.205	2.502
	0.085	*0.041*						*2.841*	*1.050*	*0.078*	*0.428*
COROHLD	0.418	0.331			-0.135			6.125	0.276	0.561	3.650
	0.169	*0.042*			*0.042*			*2.174*	*0.103*	*0.116*	*0.590*
DBNDEEP	-0.774	0.345	-0.097					29.472	0.685		4.066
	0.235	*0.045*	*0.033*					*4.979*	*0.189*		*0.922*
DIDATA	0.776	0.353	-0.186					2.682	0.172	0.701	5.536
	0.164	*0.044*	*0.044*					*1.016*	*0.055*	*0.076*	*1.057*
FIRSTRAND	0.387	0.274						2.586	0.256	0.637	6.517
	0.155	*0.045*						*1.013*	*0.077*	*0.097*	*2.324*
GFIELDS	-0.316	0.171	-0.133					9.223	0.186	0.440	11.240
	0.196	*0.046*	*0.044*					*3.429*	*0.077*	*0.173*	*2.594*
HARMONY	-0.278	0.221						1.671	0.093	0.876	4.444
	0.213	*0.039*						*0.766*	*0.033*	*0.038*	*0.671*
IMPERIAL	0.367	0.283						1.971	0.386	0.580	3.797
	0.115	*0.043*						*0.596*	*0.125*	*0.077*	*0.687*
IMPLATS	0.044	0.340	-0.097					10.749	0.302	0.268	4.471
	0.164	*0.046*	*0.042*					*2.939*	*0.112*	*0.155*	*0.920*
INVSTEC	0.430	0.322	-0.155					2.395	0.382	0.497	3.997
	0.109	*0.044*	*0.043*					*0.820*	*0.115*	*0.097*	*0.829*
ISCOR	-0.229	0.337						0.801	0.132	0.843	7.327
	0.157	*0.037*						*0.279*	*0.039*	*0.034*	*2.170*
JOHNNIC	0.232	0.243						1.338	0.161	0.785	5.594
	0.147	*0.042*						*0.533*	*0.047*	*0.056*	*1.237*

KEY

Estimate	2.000
Standard Error	*0.050*

Appendix 13 b: Parameter Estimation under the standardised t distribution

Key: The following table displays the mean and variance equations as well as their standard errors. The standard error are listed below the parameter estimates.

	c	AR 1	AR 2	AR 3	AR 4	AR 5		C*	ARCH 1	GARCH 1	V
JSEW	0.220	0.449			-0.154			0.296	0.095	0.831	4.952
	0.067	*0.043*			*0.042*			*0.127*	*0.033*	*0.053*	*0.753*
LIBERTY	0.260	0.212						0.603	0.136	0.832	5.830
	0.120	*0.040*						*0.213*	*0.040*	*0.041*	*1.001*
LIBHOLD	0.231	0.372	-0.135			0.046		0.465	0.171	0.799	5.831
	0.100	*0.045*	*0.046*			*0.043*		*0.158*	*0.047*	*0.043*	*1.143*
METCASH	0.169	0.295						2.084	0.266	0.687	4.350
	0.143	*0.041*						*0.815*	*0.074*	*0.069*	*0.723*
NAMPAK	0.183	0.258						0.171	0.087	0.920	3.460
	0.100	*0.035*						*0.081*	*0.031*	*0.020*	*0.652*
NEDCOR	0.365	0.211						1.117	0.160	0.759	9.123
	0.131	*0.044*						*0.416*	*0.047*	*0.060*	*2.971*
PICKNPAY	0.299	0.376	-0.164					0.797	0.065	0.883	6.117
	0.139	*0.040*	*0.042*					*0.418*	*0.028*	*0.045*	*1.814*
REMGRO	0.240		0.094					3.546	0.120	0.655	7.592
	0.151		*0.045*					*1.381*	*0.052*	*0.115*	*2.142*
REUNERT	0.208	0.375						1.339	0.442	0.644	3.334
	0.117	*0.040*						*0.505*	*0.133*	*0.064*	*0.498*
RICHEMONT	0.292	0.260		-0.102				7.538	0.247		5.481
	0.115	*0.047*		*0.039*				*0.799*	*0.091*		*1.017*
SABREWS	0.158	0.194						0.837	0.120	0.811	7.863
	0.121	*0.043*						*0.347*	*0.044*	*0.062*	*2.273*
SANTAM	0.401	0.429	-0.122					8.491	0.564		4.332
	0.122	*0.044*	*0.035*					*1.196*	*0.154*		*0.873*
SAPPI	0.080	0.341	-0.145					1.117	0.157	0.814	4.651
	0.152	*0.041*	*0.041*					*0.445*	*0.050*	*0.047*	*1.037*
SASOL	0.245	0.205		-0.091				0.908	0.151	0.810	5.210
	0.140	*0.041*		*0.043*				*0.406*	*0.048*	*0.050*	*0.840*
STANBIC	0.408	0.311	-0.102					1.045	0.228	0.716	4.819
	0.112	*0.044*	*0.045*					*0.376*	*0.064*	*0.061*	*1.126*
TIGBRANDS	0.182	0.301	-0.128	-0.080				1.378	0.249	0.686	4.835
	0.120	*0.044*	*0.046*	*0.045*				*0.453*	*0.064*	*0.057*	*0.963*
TONGAAT	0.214	0.282						0.273	0.071	0.928	3.413
	0.121	*0.035*						*0.149*	*0.027*	*0.021*	*0.561*

KEY

Appendix 14: Goodness of Fit tests

		Significance Level		
		1%	5%	10%
	D STAT	0.069	0.057	0.052
ABI	0.039			
ABSA	0.029			
ANGGOLD	0.056			**REJECT**
ANGLOS	0.031			
BARWORLD	0.024			
BIDVEST	0.066		**REJECT**	**REJECT**
COROHLD	0.038			
DBNDEEP	0.035			
DIDATA	0.032			
FIRSTRAND	0.058		**REJECT**	**REJECT**
GFIELDS	0.038			
HARMONY	0.034			
IMPERIAL	0.046			
IMPLATS	0.045			
INVSTEC	0.035			
ISCOR	0.041			
JOHNNIC	0.033			
JSEW	0.040			
LIBERTY	0.023			
LIBHOLD	0.060		**REJECT**	**REJECT**
METCASH	0.021			
NAMPAK	0.017			
NEDCOR	0.040			
PICKNPAY	0.027			
REMGRO	0.038			
REUNERT	0.023			
RICHEMONT	0.019			
SABREWS	0.034			
SANTAM	0.015			
SAPPI	0.030			
SASOL	0.023			
STANBIC	0.032			
TIGBRANDS	0.018			
TONGAAT	0.028			

Part VI

References

1. Akaike, H. (1969), "Fitting autoregressive models for prediction." Ann. Inst. Statist. Math, 21, pages 243-247

2. Akaike, H. (1973), "Information theory as an extension of the maximum likelihood principle. pages 267-281 in Second international symposium on information theory. B.N Petrov, B. and F Csaki (editors). Akademiai Kiado, Budapest.

3. Arshanapalli, B., and Doukas, J. (1993), "International Stock market Linkages: Evidence from the Pre- and Post October 1987 period", Journal of Banking and Finance, 68, pages 628-637

4. Attanasia, O.P. (1991), "Risk,time varying second moments and market efficiency," Review of Economic Studies, 58, pages 479-494

5. Attanasia, O.P., and Wadhwani, S. (1989), "Risk and the predictability of stock market returns," Unpublished manuscript, Department of Economics, Stanford University.

6. Ballie, R.T., and Bollerslev, T. (1989), "The message in Daily Exchange Rates: A conditional variance tale" Journal of Business and Economic Statistics 7, pages 297-305

7. Ballie, R.T., and DeGennaro (1990), "Stock returns and volatility," Journal of Financial and Quantitative Analysis, 25, pages 203-214

8. Bartlett, M.S. (1946), "On the theoretical specification of sampling properties of autocorrelated time series", Journal of Royal Statistical Society, B 8, pages 27-41

9. Belsely, D.A., Kuh, E., and Welsch, R.E. (1989), "Regression diagnostics: identifying influential data and sources of collinearity." Journal of Applied Econometrics, 4, pages 97-99

10. Berndt, E. K., Hall, B. H., Hall, R. E. and Hausman, J. A. (1974) "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, pages. 653-665.

11. Bertero, E., and Mayer, C. (1990), "Structure and Performance: Global Interdependence of Stock Markets around the Crash of October 1987", European Economic Review, September 1990, 34, pages 1155-1180

12. Black, F. (1976), "Studies of stock price volatility changes," in Proceedings of

the Business and Economic Statistics Section, American Statistical Association, pages 177-181

13. Black, F., and Scholes, M. (1973), "The pricing of options and corporate liabiliies," Journal of Political Economy 81, pages 637-659

14. Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity" Journal of Econometrics 31, pages 301-327

15. Bollerslev, T. (1987), "A conditional heteroskedastic time series model for speculative prices and rates of return" Review of Economics and Statistics 69, pages 542-547

16. Bollerslev, T., Engle, R.F., and Wooldridge, J.M (1988), "A capital asset pricing model with time varying covariances," Journal of Political Economy 96, pages 116-131

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

18. Bollerslev, T. and Wooldridge, J.M. (1991), "Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances," Econometric Reviews, 11, pages 143-172

19. Box, G.E.P., (1949), "A general distribution theory for a class of likelihood criteria", Biometrica, 36, pages 317-346

20. Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1970), "Time series analysis : forecasting and control"

21. Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994), "Time series analysis : forecasting and control", Prentice Hall

22. Box, G.E.P., and Pierce, D. (1970), "Distribution of residual auto-correlations in autoregressive-integrated moving average time series models." Journal of American Statistical Association, 64, pages 1509-1526

23. Box, G.E.P., and Tiao, G.C. (1973), "Baysian Inference in Statistical Analysis," (Reading, Mass.)

24. Box, G.E.P., and Tiao, G.C. (1975), "Intervention analysis with applications to

economic and enviromental problems," Journal of American Statistical Association70, pages 70-79

25. Box, G. & Cox, D. (1964). An Analysis of Transformations. Journal of the Royal Statistical

26. Society, Series B 26, pages 211-252.

27. Bozdogan, H. (1987) "Model Selection and Akaike's information criterion (AIC): the general theory and its analytical extensions." Psychometrika 52, pages 345-370

28. Bozdogan, H. (1988), "ICOMP: A new model selection criterion." In: H.H. Bock (Ed), Classification and related Methods of Data Analysis, North-Holland, Amsterdam, pages 599-608.

29. Bozdogan, H. (1990), "On the information-based measure of covariance complexity and its application on the evaluation of multivariate linear models." Communications in Statistical Thoery and Methods. 19(1), pages 221-278

30. Bozdogan, H. (1993), "Choosing the number of component clusters in the mixture model using a new informational complexity criterion of the fisher information matrix." In: O. Opitz, B Lausen and R Klar (Eds), Studies in classification, Data analysis, and Knowledge Organization, Springer, Heidelberg, pages 40-54

31. Bozdogan, H. (1994), "Mixture-model cluster analysis using a new informational complexity and model selection criteria." In: H Bozdogan (Ed), Multivariate Statistical Modelling, vol 2, Proceedings 1^{st} US/Japan Conference on the Frontiers of Statistical Modelling: An informational Approach, Kluwer Academic Publishers, the Netherlands, Dordrecht, pages 69-113

32. Bozdogan, H., Bearse, P.M., and Schottmann, A.M. (1997), "Empirical econometric modelling of food consumption using a new informational complexity approach." Journal of Applied Econometrics, Vol 12, pages 563-592

33. Bozdogan, H. (1998), "Akaike information criteria and recent developments in informational complexity." Invited paper to appear in: Journal Math Psychology

34. Bozdogan, H. and Haughton, D.M.A. (1998), "Informational complexity criteria for regression models." Computational Statistics and Datat analysis 28, pages

51-76

35. Bozdogan, H. and Bearse, P.M. (1999), "Model selection using information complexity with applications to vector autoregressive (VAR) models" Invited contributed paper for JSPI on Model Selection.

36. Breeden, D. (1979), "An Intertemporal Asset Pricing Model With Stochastic Consumption and Investment Opportunities", Journal of Financial Economics 7, pages 265-96

37. Brewer, E. and Room, A. (1999), Brewer's Dictionary of Phrase and Fable, 16th edition, Cassell & Co.

38. Brockwell, P.J., and Davis, R.A. (1991), "Time Series: Theory and Methods", Second Edition, Springer

39. Browne, M.W., and Rock, R.A. (1978), "The choice of additive constants in ridge regression," SA Statistical Journal, 12, pages 57-74

40. Campbell, J.Y. (1987), "Stock returns and the term structure," Journal of Financial Economics 18, pages 373-399

41. Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, factor structure and mean-variance analysis on large asset markets," Econometrica 51, pages 1281-1304

42. Chatfield, C. (1975), "The analysis of time Series: theory and practice," London: Chapman & Hall

43. Chatfield, C. (1996a), "The analysis of time Series," 5th edition London: Chapman & Hall

44. Chatfield, C. (1980), "The analysis if time series : An introduction," Chapman and Hall New York

45. Chatfield, C. (2001), "Time Series Forecasting, " Boca Raton: Chapman & Hall

46. Christie, A.A. (1982), "The stochastic behaviour of common stock variances: Value, Leverage and interest rate effects," Journal of Applied Econometrics, 10, pages 407-432

47. Cover, T.M., Gacs, P., and Gray, R.M. (1989), "Kolmogorov's contribution to information theory and algorithmic complexity." Annals of Probability, 17,

pages 840-865

48. Dezhbaksh, H. (1990), "The Inappropriate Use of Serial Correlation Tests in Dynamic Linear Models," Review of Economics and Statistics, 72, NO 1, pages 126-132.

49. Diebold, F.X., and Nerlove, M. (1989), "The dynamics of exchange rate volatility: A multivariate latent factor ARCH model", Journal of Applied Econometrics 4, pages 1-21

50. Draper, N.R., and Smith, H. (1966), Applied Regression Analysis, Wiley

51. Draper, N.R., and Smith, H. (1981), Applied Regression Analysis, Wiley

52. Dunne, T., and Clark, A.E. (2003), STA341F class notes

53. Durbin, J., and Knott, M. (1972), "Components of Cramer-von Mises statistics." International Journal of Royal Statistical Society. B. 34, pages 290-307

54. Eichholtz, P.M.A. (1996), "The stability of the covariances of international property share returns," The journal of real estate research, 2, pages 149-158

55. Engel, C., and Rodrigues, A.P. (1989), "Tests of international CAPM with time varying covariances," Journal of Applied Econometrics 4, pages 119-138

56. Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with estimates of the variance of UK inflation," Econometrica, 50: pages 987-1008

57. Engle, R.F. (1995), "ARCH SELECTED READINGS, advanced texts in Econometrics"

58. Engle, R.F., and Gonzalez Rivera, G. (1991), "Semiparametric ARCH Models", Journal of Business and Economic Statistics 9, pages 345-359.

59. Engle, R.F., Ito, T., and Lin, W. (1990a), "Meteor showers or heat waves? Heteroskedastic intra daily volatility in the foreign exchange market," Econometrica 58, pages 525-542

60. Engle, R.F., Ito, T., and Lin, W. (1990b), "Where does the meteor shower come from? The role of stochastic policy coordination," Unpublished manuscript (University of Wisconsin, Madison, WI)

61. Erb, C.B., Harvey, C.R., and Viskanta, T.E. (1994) ,"Forecasting international equity correlations", Financial Analysts Journal, 50, pages 32-45

62. Eun, C., and Shim, S. (1989), "International transmission of stock market movements," 24, pages 241-256

63. Fernandez, C., and Steel, M.F.J., (1998), "On the bayesian modelling of fat tails and skewness", Journal of the American Statistical Association, 93, pages 359-371

64. Fletcher, R. (1969), "Optimization", New York: Academic Press.

65. Fletcher, R. and Powell, M.J.D. (1963), "A rapidly convergent descent method for minimization.", Computer Journal, 6, pages 163-168.

66. Gallant, A.R., and Nychka, D.W. (1987), "Semi nonparametric maximum likelihood estimation," Econometrica 55, pages 363-390

67. Gallant, A.R., and Tauchen, G. (1989), "Semi nonparametric estimation of conditionally constrained heterogeneous processes: Asset Pricing Applications," Econometrica 57, pages 1091-1120

68. Gallant, A.R., Hsieh, D.A., and Tauchen, G. (1989), "On fitting a recalcitrant series: The pound/dollar exchange rate 1974-83," Unpublished manuscript , Department of Economics, Duke University, Durham, NC

69. Gallant, A.R., Rossi, P.E., and Tauchen, G. (1990), "Stock prices and volume," Unpublished manuscript , Department of Economics, Duke University, Durham, NC

70. Geisser, S. (1974), "A predictive approach to the random effects model." Biometrika, 61, pages 101-107

71. Glosten, L.R., Jagannathan, R., and Runkle, D. (1991), "On the Relation between the Expected Value and the Volatility of the Normal Excess Return on Stocks", Journal of Finance, 48, pages 1779-1801.

72. Gourieroux, C., and Monfort, A. (1992) ,"Qualitative Threshold ARCH models," Journal of Econometrics, April/May, 52, pages 159-199

73. Gourieroux, C. (1997), "ARCH Models and Financial Applications", Cambridge University Press

74. Grubel, H.G., and Fadner, K. (1971), "The interdependence of international equity markets," Journal of Finance, 26, pages 89-94

75. Haldane (1951), "A class of efficient estimates of a parameter. " Bulletin of International Statistical Institute, 33 , pages 231-248.

76. Hamilton, J. D. (1994a), "Time Series Analysis," Princeton University Press.

77. Hannan, E.J. and Quinn, B.G. (1979) "The determination of the order of an autoregression." Journal of the Royal Statistical Society, Series B 41, pages 190-195

78. Harvey, A. C. (1981a), "Time series models," Oxford: Philip Allan

79. Harvey, A. C. (1981b), "The econometric analysis of time series," Oxford: Philip Allan

80. Harvey, A. C. (1990), "The Econometric Analysis of Time Series", 2nd edition, MIT Press.

81. Harvey, A. C. (1993), "Time Series Models", 2nd edition, MIT Press.

82. Hemmerle, W.J. and Carey, M.B. (1983), "Explicit and constrained generalized ridge estimation." Technometrics, 20, pages 109-120

83. Hillard, J. (1979), "The relationship between equity indices on world exchange", Journal of finance, March, 34 , pages 103-114

84. Hoerl, A.E. (1962), "Application of ridge analysis to regression problems." Chemical engineering Progress, 58, No 3 ,May , pages 54-59

85. Hoerl, A.E., and Kennard, R.W. (1970a), "Ridge regression: Biased estimations for nonorthogonal Problems." Technometrics, 12, February 1970, pages 55-67

86. Hoerl, A.E., and Kennard, R.W. (1970b), "Ridge regression: Applications to nonorthogonal Problems." Technometrics, 12, No1 , February 1970, pages 69-82

87. Hoerl, A.E., Kennard, R.W., and Baldwin, K.F. (1975), "Ridge Regression: Some simulations." Communications in Statistics, 4, pages 105-123

88. Hsieh, D.A. (1989a), "Modelling heteroskedasticity in Daily Foreign Exchange rates" Journal of Business and Economic Statistics 7, pages 307-317

89. Hsieh, D.A. (1989b), "Testing for nonlinear dependence in daily foreign exchange rate changes" Journal of Business and Economic Statistics 62, pages 339-368

90. Hurvich, C.M., and Tsai, C.L. (1989) "Regression and Time series model selec-

tion in small samples. Biometrica 76, pages 297-307

91. Jenrich, R.I. (1970) , An Asymptotic χ^2 for the equality of two correlation matrices, Journal of the American Statistical Association, June 1970, 65, pages 904-912

92. Jorion, P. (1988), "On jump processes in the foreign exchange and stock markets." Review of Financial Studies 1, pages 427-445

93. Kaplanis, E.C. (1988), "Stability and Forecasting of the Comovement Measures of International stock Market Returns," Journal of International Money and Finance 7, pages 63-75

94. Karpoff, J.M. (1987), "The relation between price changes and trading volume: A survey," Journal of Financial and Quantitative Analysis 22, pages 399-408

95. King, M., and Wadhwani, S. (1990), "Transmission of Volatility between National Stock Markets", Review of Financial Studies", 1990, 3, pages 5-33

96. King, M., Sentana, E., and Wadhwani, S. (1992), "Volatility and Links between National Stock Markets", Working Paper

97. Koch, P.D., and Koch, T.W. (1991), "Evolution in dynamic Linkages across National Stock Indices", Journal of international Money and Finance, June,10, pages 231-251

98. Kullback, S., and Leibler,R. (1951) "On information and sufficiency." Annals of Mathematical Statistics 22, pages 79-86

99. Krishnaiah, P., and Rao, C. (1988), "Sampling," Handbook of statistics,

100. Lamoureux, C.G., and Lastrapes, W.D. (1990b), "Heteroskedasticity in stock return data: Volume versus GARCH effects." Journal of finance 45, pages 221-229

101. Lambert, P., and Laurent, S. (2001), "Modelling financial time series using GARCH-type models with a Skewed Student distribution for the innovations," Mimeo, Université de Liège

102. Lawless, J.F., and Wang, P. (1976), "A simulation study of ridge and other estimators." Communications in Statistical Theory and Methodology, A5, pages 307-323

103. Lessard, D.R. (1973), "International portfolio diersification multivariate analysis for a group of Latin American Countries," Journal of Finance, 28, pages 619-633

104. Levy, H., and Sarnat, M. (1970), "International Diversification of investment portfolios," American Economic Review, pages 668-675

105. Lewis-Beck, M.S. (1993), "Regression Analysis, International Handbooks of Quantitative Applications in the Social Sciences," Volume 2

106. Lindley, D.V. (1968), "The choice of variables in multiple regression" Journal of the Royal Statistical Society, Series B, 30, pages 31-36

107. Lindley, D.V., and Smith, A.F.M. (1972), "Bayes estimates for the linear model." Journal of the Royal Statistical Society, Series 34, No 1, pages 1-18

108. Linhart, H., and Zucchini, W. (1986), "Model Selection", John Wiley and Sons Inc.

109. Lintner, J. (1965), "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets" , Review of economics and statistics , February)

110. Ljung, G., and Box, G.E.P. (1979), "On a Measure of Lack of Fit in Time Series Models", Biometrika, 66, pages 265-270.

111. Longin, F., and Solnik, B. (1995) , "Is the correlation in international equity returns constant: 1960-1990?" , Journal of International Money and Finance , 14, pages 3-26

112. Lucas, R. (1978), "Asset prices in a production economy," Econometrica 46, pages 1429-1445

113. Mallows, C.L. (1964), "Choosing the variables in a linear regression: A Graphical aid", presented at the Central Regional meeting of the /institute of Mathematical Statistics, Manhattan, Kansas, May 7-9

114. Mallows, C.L. (1973), "Some comments on C_p," Technometrics, 15, pages 666-675

115. Mandelbrot, B. (1963), "The variation of certain speculative prices" Journal of Business 36, pages 394-419

116. Markowitz, H. (1952), "Portfolio Selection", Journal of Finance 7, pages 77-91

117. Marquardt, D.W. (1970), "Generalized inverses, ridge regression, biased linear estimation and non linear estimation." Technometrics, 12, No 3, August, pages 591-612

118. McDonald, G.L., and Galarneau, D.I. (1975), "A Monte Carlo evaluation of some ridge type estimators" Journal of the American Statistical Association, June , Volume 70, Number 350, theory and Methods section.

119. McCulloch, J.H. (1985), "Interest-risk sensitive deposit insurances premia-Stable ARCH estimates" Journal of Banking and finance 9, pages 137-156

120. Merton, R. (1973), "An Intertemporal Capital Asset Pricing Model", Econometrica 41, pages 867-87.

121. Milhoj, A. (1985), "The moment structure of ARCH processes ," Scandiavian Journal of Statistics 12, pages 281-292

122. Miller, L.H. (1956), "Table of Percentage Points of Kolmogorov Statistics," Journal of the American Statistical Association, vol 51, pages 111-121

123. Miller, A. (1990), "Subset Selection in regression" ,Chapman and Hall

124. Mossin, J. (1966), "Equilibrium in a capital asset market," Econometrica , October 1966

125. Nelson, D.B. (1990c), "Conditional heteroskedasticity in asset returns: A new approach." Econometrica 59, pages 347-370

126. Nelson, H.I., and Granger, W.J. (1979), "Experience with using the Box Cox transformation when forecasting economic time series," Journal Econometrics 10, pages 57-69

127. Newbold, P. (1988), "Predictors projecting linear trends plus seasonal dummies," The Statistician, 37, pages 111-127

128. Newhouse, J.P., and Oman, S.D. (1971), "An evaluation of Ridge estimators" Rand Report, No R-716-PR, April (1971) pages 1-28

129. Pagan, A.R., and Ullah, A. (1998), "The econometric analysis of models with risk terms," Journal of Applied Econometrics 3, pages 87-105

130. Palm, F.C., and Vlaar, P.J.G. (1997), "Simple Diagnostics Procedures for Mod-

elling Financial Time Series." Allegemeine Statistisches Archiv, 81, pages pages 85-101

131. Parr, W.C. and Schucany, W.R. (1980), "Minimum distance and Robust estimation" Journal of American Statistical Association, 75, pages 616-624

132. Parr, W.C., and De Wet, T. (1981), "On minimum Cramer-von Mises norm parameter estimation" Communications in Statistical Theory and Methods, A10, pages 1149-1166

133. Parzen, E. (1997), "Multiple time series: Determining the order of approximating autoregressive Schemes." In Krishnaiiah, ed, Multivariate Analysis IV. North-Holland, Amsterdam, pages 283-295.

134. Parzen, E. (1982), "Data modelling using quantile and density-quantile functions," In J.T de Oliveria and B. Epstein (Eds.), Some recent advances in statistics (pages 23-52). London: Academic Press

135. Pena, D., Tiao, G.C., and Tsay, R.S. (2001), "A course in Time Series Analysis", John Wiley and Sons, INC, Wiley Series in Probability and Statistics

136. Pindyck, R.S., and Rubinfeld, D.L. (1997) , Econometric Models and Economic Forecasts, Fourth Edition, The McGraw-Hill Companies

137. Ponnapalli, R. (1976), "Deficiencies of minimum discrepancy estimators." Can. Journal of Statistics, 4, pages 33-50

138. Priestley, M.B. (1993), "Developments in time series analysis," Chapman & Hall

139. Quenouille, M.H. (1949), "Approximate tests of correlation in time series" Journal of Royal Statistical Society, B 11, pages 68-84

140. Ragunathann, V., Faff, R., and Brooks (1997) ,"Correlations, Business Cycles and Integration", MIMEO, Royal Melbourne Institute of Technology.

141. Ragunathann, V., and Mitchell, H. (1997), "Modelling the Time-Varying Correlation between national stock market returns," Department of Economics and Fiance Royal Melbourne Institute of Technology, Melbourne - Centre in Finance in its series Papers with number 97-7

142. Ratner, M. (1992), "Portfolio Diversification and the Inter-temporal Stability

of International Indices," Global finance Journal, 1992, 3, pages 67-78

143. Rawlings, J.O., Pantula, S.G., and Dickey, D.A. (1998), "Applied Regression Analysis- a research tool"

144. Rice, J. (1995). "Mathematical Statistics and Data Analysis". Duxbury Press.

145. Ripley, D.M. (1974), "Systematic Elements in the Linkages of National Stock Market Indices," Review of Economics and Statistics, August, 55, pages 356-361

146. Rissanen, J. (1986) "Stochastic complexity and modelling" Annals of Statistics, 14, pages 1080-1100

147. Rissanen, J. (1989) "Stochastic complexity in statistical inquiry." World Scientific, Singapore

148. Robertson, C.A. (1972), "On minimum discrepancy estimators." Sankhya A, 34, pages 133-144

149. Ross, S. (1976), "Arbitrage Theory of capital asset pricing," Journal of econometric theory

150. Sahler, W. (1970), "Estimation by minimum-discrepancy methods." Metrika, 16, pages 85-106

151. Sakamoto, Y., Ishiguro, M., and Kitagawal, G. (1986) "Akaike information criterion statistics." KTK Scientific Publishers, Tokyo

152. Sakamoto, Y., and Akaike, H. (1978a), "Analysis of cross classified data by AIC." Ann. Inst. Statist. Math, 30, pages 185-197

153. Schollhammer, H., and Sand, O. (1985), "The interdependence among the stock markets of major European Countries and the United States: An empirical Investigation of Interrelationships among National Stock price movements," Management International Review, 25, pages 17-26

154. Schwarz, G. (1978), "Estimating the dimension of a model" Annals of Statistics 6, pages 466-464

155. Sharpe, W. (1964), "Capital asset prices: A Theory of market equilibrium," Journal of finance, September, pages 885-901.

156. Shiskin, J., Young, A.H., and Musgrave, J.C. (1967), "The X-11 variant of the census method 2 seasonal adjustment program," Technical paper number

15.(revised) Washington Bereau of the census.

157. Slutsky, E.E. (1937), "The summation of random causes as the source of cyclic precesses," ECONOMETRICA, 5, pages 105-106.

158. Sugiura, N., (1978) "Further analysis of the data by Akaike's information criterion and the finite corrections" Communications in Statistics, Theory and Methods , pages 13-26

159. Solnik, B. (1974), "Why not diversify internationally rather than domestically?," July/August, 30, pages 48-54

160. Solnik, B., Boucrelle, C., and Le Fur, Y. (1996), "International Market correlation and volatility" , Financial Analysts Journal , September/October 1996

161. Stone, M. (1981), "A theoretical view of the use of AIC." In Anderson (Ed), Time Series Analysis: Theory and Practice, Vol 4, pages 237-244, Amsterdam: North Holland

162. Stuart, A. (1994), "Kendall's advanced theory of statistics," Volume 3, Halsted Press

163. Takeuchi, K. (1976) "Distribution of informational statistics and a criterion of model fitting." Suri-Kagaku (Mathematic Sciences) 153, pages 12-18

164. Thiart, C. (1990), "Collinearity and consequences", Unpublished Msc Thesis, UCT

165. Troskie, C. (1999), Applied Regression course notes (UCT)

166. Troskie, C., and Chalton, D. (1996), "A Bayesian estimate for the constants in ridge regression." South African Statistical Journal (1996), 30, pages 119-137

167. Tsay, R.S., and Tiao, G.C. (1984), "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and non-stationary ARMA Models," Journal of the American Statistical Association, 79, pages 84-96

168. Tsay, R.S. (2002), "Analysis of Financial Time Series", Financial Econometrics, John Wiley and Sons, INC ,Wiley Series in Probability and Statistics

169. van den Honert, R. (1997), Intermediate Statistical Methods for Business and Economics

170. van Emden, M.H. (1971), "An analysis of complexity." Mathematical Centre Tracts, Amsterdam, 35

171. Vinod, H.D. (1976), "Application of ridge regression methods to a study of bell System scale economies." Journal of the American Statistical Association, 71, pages 835-841

172. Von Furstenberg, G., and Jeon, B.N. (1989) , "International Stock Prices Movements: Links and Messages," Brookings Papers on Economic Activity" , 1989, 1, pages 125-179

173. Weiss, A.A. (1984), "ARMA models with ARCH errors," Journal of Time Series Analysis 5, pages 129-143

174. Weiss, A.A. (1986), "Asymptotic theory for ARCH models: Estimation and testing," Econometric Theory, 2, pages 107-131

175. Wetherill, G.B. (1986), Regression Analysis with Applications

176. Wichern, D.W., and Churchill, G.A. (1978), "A Comparison of ridge estimators." Technometrics, vol 20, No 3, August

177. Zakoian, J.M. (1990), Threshold Heteroskedastic Models, manuscript, CREST, INSEE, Paris.