LAG LENGTH SELECTION FOR VECTOR ERROR CORRECTION MODELS

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

of

RHODES UNIVERSITY

by

GARY DAVID SHARP

March 2010

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACF	autocorrelation function
AIC	Akaike's information criterion
AICC	Hurvich and Tsai's corrected Akaike's information criterion
AICCBD	Brockwell and Davis's corrected Akaike's information criterion
$\mathrm{ARCH}(p)$	an autoregressive conditional heteroscedastic process
ARIMA	autoregressive integrated moving average
$\mathrm{AR}(1)$	first order autoregressive process
$\mathrm{AR}(p)$	an autoregressive process with *p* terms,
$\mathrm{ARIMA}(p,d,q)$	an autoregressive integrated moving average process with three components, an autoregressive component with *p* terms, a $d^{th}-$order of integration component, and a moving average component with *q* terms.
BIC	Schwarz's Bayesian information criterion
$\mathrm{CI}(d,b)$	a nonstationary, cointegrated process where *d* represents the order of integration of the nonstationary processes and *b* represents the number of stationary linear combinations between the nonstationary processes
ECM	an error correction model
FPE	Akaike's final prediction error
HQIC	Hannan-Quinn's information criterion
HQICC	McQuarrie and Tsai's corrected Hannan-Quinn's information criterion
IC	information criteria
$\mathrm{I}(d)$	The number of times (d) the series was differenced in order for the series to be stationary
LCIC	Gonzalo and Pitarakis's linear combination information criterion

MAIC	Qu and Perron's modified Akaike's information criterion
MAICC	Sharp and Radloff's modified corrected Akaike's information criterion
$\mathrm{MA}(q)$	a moving average process with q terms
MBIC	Schwarz's Bayesian information criterion
MHQIC	Hannan-Quinn's information criterion
MS	In reference to the data sets that did meet specification
NMS	In reference to the data sets that did not meet specification
PACF	partial autocorrelation function
ShibIC	Shibata's information criterion
VAR	a vector autoregressive process
VARMA	a vector autoregressive moving average process
$\mathrm{VAR}(p)$	a VAR with k variables and p lagged terms process
VEC	a vector error correction process
$\mathrm{VEC}(p-1)$	a VEC model with $(p-1)$ lag terms
$3d-\mathrm{VAR}(1)$	a VAR with 3 variables and 1 lagged terms process
$2d-\mathrm{VAR}(1)$	a VAR with 2 variables and 1 lagged terms process

ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis adviser, Prof. S. Radloff. Your constant support and encouragement is much appreciated.

Throughout the duration of this study I have had encouragement, advice and the guidance of many people, I would like to thank:

- Colleagues: Prof. John Gonsalves, David Friskin, Chris Parsons, Johan Hugo and Prof. Stephen Hosking.
- Colleagues from the Department of Statistics at Nelson Mandela Metropolitan University (NMMU), colleagues from the Department of Statistics at Rhodes University (RU), and colleagues from the Department of Statistics at the University of Stellenbosch (US).
- Post-graduate students of NMMU: Limin Liu, Wei He and Chris Liao.
- The NMMU Research Office for financial support.
- The NMMU inter library loans staff for attending to my many requests.
- The National Research Foundation (NRF) for financial support in the guise of Thuthuka (Researcher in Training) grants for 2006 – 2009.
- Participants at the model selection meeting: CPM 021, at the 56th Session of the ISI, Lisboa, Portugal, 2007.
- The comments from the external examiners which improved considerably the final product of this thesis.
- Family and friends, who all played a part, however small.

My final acknowledgements go to the two most important people in my life, my wife, Catherine, your love and support have been fantastic, without you there is a good *probability* this may never have been completed, and my daughter, Sarah, we can now spend a lot more time together.

ABSTRACT

This thesis investigates the problem of model identification in a Vector Autoregressive framework. The study reviews the existing research, conducts an extensive simulation based analysis of thirteen information theoretic criterion (IC), one of which is a novel derivation. The simulation exercise considers the evaluation of seven alternative error restricted vector autoregressive models with four different lag lengths. Alternative sample sizes and parameterisations are also evaluated and compared to results in the existing literature.

The results of the comparative analysis provide strong support for the efficiency based criterion of Akaike and in particular the selection capability of the novel criterion, referred to as a modified corrected Akaike information criterion, demonstrates useful finite sample properties.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The econometric modelling of time series data has seen tremendous growth in recent years. The advancements made in the analysis of times series models over the last three decades are partly due to the developments of theoretical models and partly due to the improvements in computational ability. In earlier years the analysis of time series models was severely restricted by the time available to perform repetitive calculations, but with the advances made in software development most of the models developed in the early 1970s, 1980s and 1990s have become standard in statistical software packages.

Earlier texts such as Box and Jenkins (1976), Judge, Hill, Griffiths, Lütkepohl and Lee (1988), Kendall and Ord (1990) and Bowerman and O'Connell (1993) emphasised the autoregressive integrated moving average (ARIMA) models based on the Box-Jenkins methodology. The ARIMA theory was developed on the basis that the series under consideration were stationary or easily transformable into stationary series. More recent texts such as Hamilton (1994), Shumway and Stoffer (2000), Štulajter (2002), Fan and Yao (2003), Harris and Sollis (2003), Enders (2004), Brüggemann (2004) and Lütkepohl (2005) have expanded on those initial concepts and extended the theory and empirical research into multivariate time series analysis, non-linear time series analysis, model selection for time series analysis and applied modelling and forecasting.

Over the last two decades, one of the time series modelling research directions has been the development of the theory of cointegrated time series modelling based primarily on the original work of Granger and Weiss (1983) and the seminal paper of Engle and Granger (1987). The theory was further developed by Johansen (1988, 1991), Stock and Watson (1991), Johansen and Juselius (1992) and Pesaran and Shin (1997) to mention just a few.

The purpose of this research study is to address a contradiction identified in the literature when applied researchers use the theory of cointegration to model nonstationary time series systems.

1.2 Objectives and Contributions of the Present Study

Cointegrated modelling requires that the series under investigation be nonstationary, and ever since Dickey and Fuller (1979) developed the initial theory and methodology for the stationarity testing of a time series, the analysis of nonstationary time series data has generated considerable research interest. One of the considerations in cointegrated modelling that has yet to be resolved is the determination of the appropriate lag length of the autoregressive representation of a cointegrated system. The lag structure of the model has a theoretical implication as estimation is influenced by the model's dimension whilst the practical implication lies in the interpretations and significance of the parameters estimated.

The motivation for this research began after reading several application based cointegrated studies. The studies were all very interesting from an application perspective but the determination of the model structure was in many cases poorly motivated. The problem with determining model structure became more apparent when approached by an economist to assist with the cointegrated modelling of their data. Both research articles and reference texts provided alternative methods for deciding on lag structure but in most cases none where committed to an individual or consistent preferential method. Further investigation confirmed that there was little uniformity in deciding model structure despite the extensive use of the modelling paradigm.

In an attempt to address this lack of uniformity it was decided to undertake an extensive Monte Carlo simulation modelling exercise which it was hoped would provide more clarity as to a method for deciding on the model structure. The simulation exercise required an extensive literature review in two directions. The primary direction was to identify

previous attempts at model selection for vector autoregressive represented systems whilst the secondary direction was to identify suitable software for the simulation exercise.

The objectives of this research are to

- review lag length selection methods for cointegrated systems,
- identify contradictions in the determination of lag length in the published literature, and
- summarise information theoretic criteria used for lag length selection.

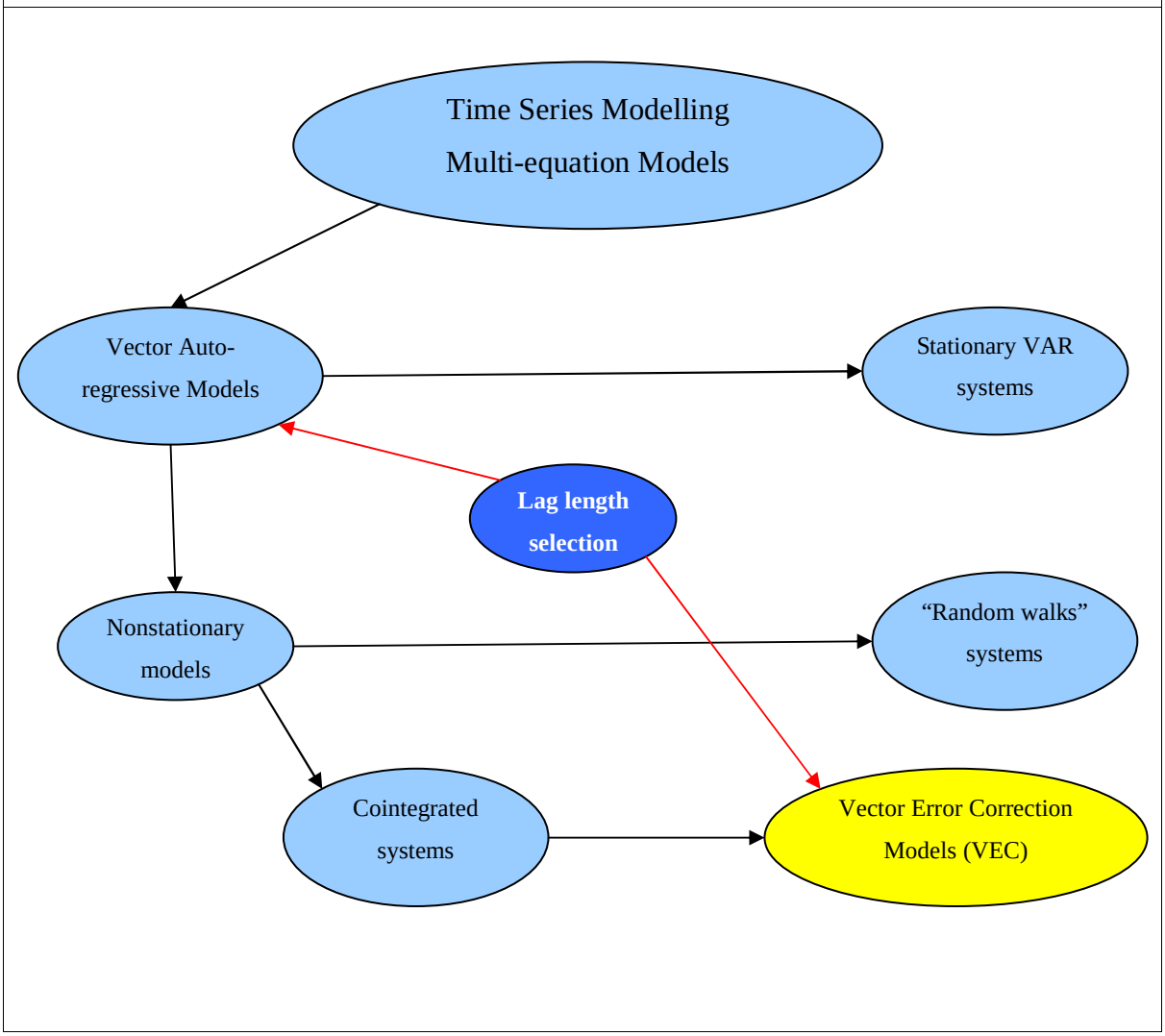
This research contributes to the body of knowledge by

- summarising the literature on model selection for autoregressive represented cointegrated systems,
- identifying and unifying the notation of information theoretic criteria used for determination of model structure for vector autoregressive systems which are constrained by the cointegrated relationship of the nonstationary variables,
- defining theoretical models appropriate for simulating cointegrated systems,
- defining an alternative information theoretic criterion, a modified corrected Akaike information criterion, denoted MAICC, and
- ultimately provide practitioners with a better method for determining the number of lag terms and hence dimension of their cointegrated models.

1.3 Outline of the Research

This research addresses a shortcoming identified in the empirical literature in the modelling of error restricted vector autoregressive (VAR) models. The flowchart in Figure 1.1 provides a schematic of where these models fit in a multi-equation time series framework. These error restricted VAR models are usually referred to as vector error correction (VEC) models and are shown in the yellow oval.

Figure 1.1: Flowchart of lag length selection of multi-equation models



The link between VAR and VEC models is two fold. VEC models are special cases of VAR models only applicable to systems classified as nonstationary and cointegrated. In addition, the lag length structure of a VEC model is directly related to the lag length structure of a VAR model. This link is highlighted by the red arrows which connect the lag length selection oval highlighted in dark blue.

These models have been used extensively in empirical studies over the last decade and theoretical refinements are continuously being added to the literature. The emphasis of this study is the selection of the lag structure of a VEC model.

Chapter 2 introduces the theoretical concepts of VEC models defined in a VAR framework. Included are the definitions of stationary and nonstationary systems, cointegration and model selection. Graphical illustrations provide descriptive measures of stationary and nonstationary data for both correlated and uncorrelated systems.

An introductory literature review emphasises the theoretical research into cointegrated modelling and the subsequent empirical applications. The literature review introduces the notion of model selection and discusses the current methods from an application and a definition perspective. The chapter concludes with the definitions of the thirteen information criteria (IC) assessed for model selection in this thesis.

Chapter 3 outlines the methodology and the models used in this study. The methodology is illustrated as a flowchart and highlights the sequence of events in the empirical analysis of the simulated data. All models are defined, their theoretical underpinnings shown and the simulation models explained. The chapter concludes with the justification of the parameters used for the simulation models.

Chapter 4 discusses the results and interpretations of this study. The section starts with the stepwise sequential analysis of a simulated dataset which follows the analytical routines given in the methodology. The illustration of a complete series is complemented by the computerised outputs from the software used in this study. The selection capabilities of the estimated criteria are evaluated from several arguments and ranking systems proposed for each of the assessment methods. The chapter concludes by identifying the better performing criteria for the selection of VEC models with an estimated error restriction constraint but an unknown lag structure.

Chapter 5 discusses the validity of the software and illustrates how the software was used in the analysis. The validity of the software is assessed by comparing the results with a published source. A stepwise illustration of the assessment of a published dataset is then shown and the results of Chapter 4 are incorporated into the analysis. The chapter concludes by proposing appropriate criteria for the selection of lag length of cointegrated models.

Chapter 6 reviews the results of similar studies identified in the literature and highlights their pitfalls and limitations. The results of this research were compared with these studies and the research contributions discussed. The comparative discussion emphasises equivalence in terms of model dimension whilst simultaneously referring to the differences encountered in the literature. The chapter concludes with recommendations and suggestions for further research.

CHAPTER 2

THEORETICAL CONSIDERATIONS

2.1 Stationary and Nonstationary Time Series

The definition of stationarity is dependent on model dimension. For the purposes of this study the single equation definition by Enders $(2004:53)$ was used to introduce concepts and notation and this was followed by Brockwell and Davis's $(2002:224)$ multi-equation definition. The single equation model is given by letting x_t, $t=1, 2, ..., T$, denote the time series under consideration.

The series, x_t, is (covariance) stationary if the process has

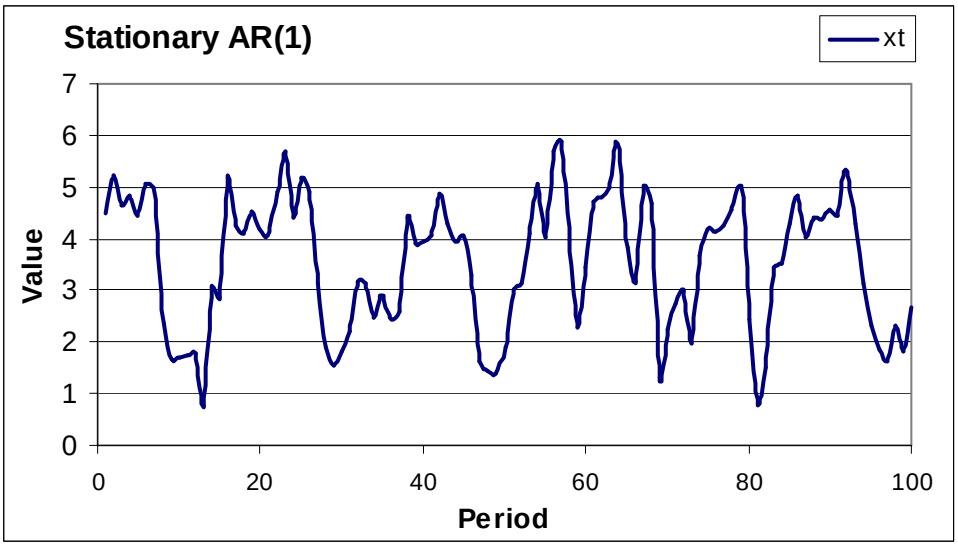
- a constant mean, $E(x_t)=E(x_s)=\mu \quad \forall_{t\neq s}$,
- a constant, finite variance, $Var(x_t)=Var(x_s)=\sigma^2<\infty \quad \forall_{t\neq s}$, and
- a finite covariance, $Cov(x_t, x_s)=\gamma_{t-s}$ and $Cov(x_{t+i}, x_{s+i})=\gamma_{t-s} \quad \forall_i$.

Series that do not satisfy these criteria are said to be nonstationary and the following first order process provides an illustration of the difference between a stationary and a nonstationary process. Consider the first order autoregressive, $\mathrm{AR}(1)$, process $x_t=a_0+a_1x_{t-1}+\varepsilon_t$, where a_0 is a constant term and for ease of exposition is set equal to zero. The process is stationary if the absolute value of the parameter $|a_1|<1$ and is nonstationary if $a_1=1$ (Engle & Granger, 1991; Lütkepohl, 2005). In the case of $a_1=1$ and $a_0=0$, the process is referred to as a random walk without drift.

Figures 2.1 and 2.2 show two simulated $\mathrm{AR}(1)$ processes. The processes are constructed by generating 300 observations for a series whose error terms are simulated from a standard normal distribution with a mean of zero and standard deviation of one. These observations are used in the $\mathrm{AR}(1)$ process to obtain the 300 observations of the series x_t with the process initialised by setting $x_0 = 0$. To minimise the initial condition effect and to improve randomness, the first 200 observations were omitted and the remaining 100 observations were used.

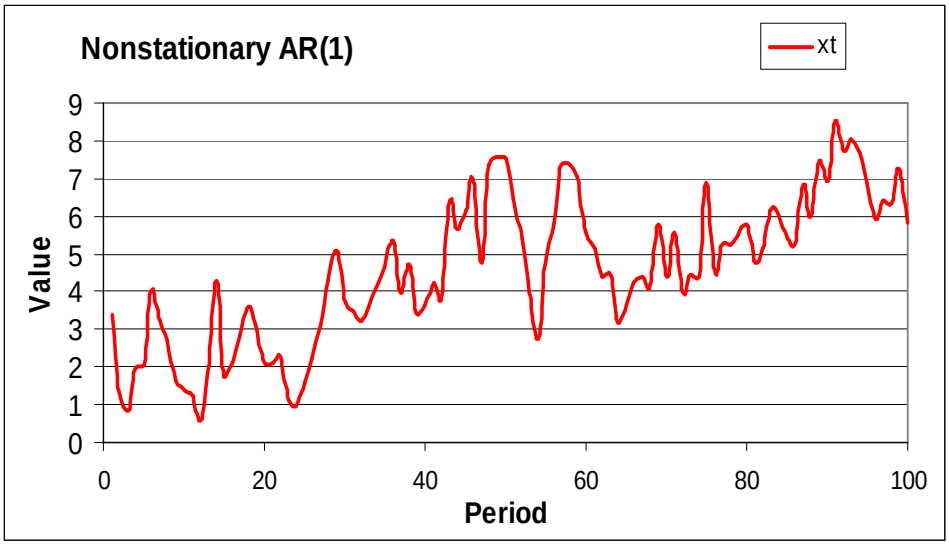
A visual inspection of the series $x_t = 0.8x_{t-1} + \varepsilon_t$ in Figure 2.1 indicates that the process mean and variance seem stable implying that the process is stationary. On the contrary, the series $x_t = x_{t-1} + \varepsilon_t$ in Figure 2.2 shows a changing mean and fluctuating variance, indicating that the series may be nonstationary. Although a visual inspection is satisfactory at this illustrative stage, inferential procedures are preferred and this study applied the more accepted inferential methods developed by Dickey and Fuller $(1979, 1981)$.

Figure 2.1: $\mathrm{AR}(1)$ plot for a stationary process



Stationary process simulated by the model $x_t = 0.8x_{t-1} + \varepsilon_t$

Figure 2.2: AR(1) plot for a nonstationary process



Nonstationary process simulated by the model $x_t = x_{t-1} + \varepsilon_t$

The multi-equation model is defined for a *k*-dimensional vector $\mathbf{x_{it}}$, where $i = 1, 2, ..., k$ and $t = 1, 2, ..., T$ denote the vector time series under consideration. Similar to the single equation case, Brockwell and Davis (2002: 224) define the (covariance) stationary process for the multi-equation model using the vector series, $\mathbf{x_{it}}$. The vector is (covariance) stationary if the first and second moments are time invariant, i.e.

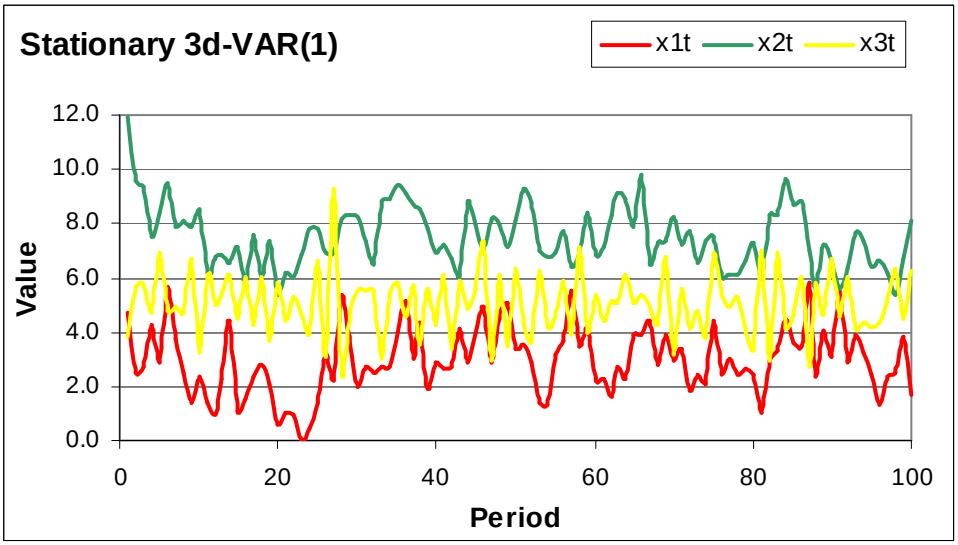
- if the mean vector is independent of time, $E(\mathbf{x_{it}}) = E(\mathbf{x_{is}}) = \boldsymbol{\mu}_\mathbf{i} \quad \forall_{t \neq s}$, and
- the covariance matrices $Cov(\mathbf{x_{it}}, \mathbf{x_{is}}) = Cov(\mathbf{x_{i(t+j)}}, \mathbf{x_{i(s+j)}}) = \boldsymbol{\Gamma}_{\mathbf{t,s}} \quad \forall_j$ are independent of time.

To describe this process, based on the Wold representation theorem, a stationary time series admits an infinite moving average representation which under certain conditions can be approximated by a finite order VAR with *k* variables and *p* lagged terms. This process, denoted as a *k-dimensional* VAR(p) process is a multi-equation model, which is easiest represented in matrix notation. The Figures 2.3, 2.4 and 2.5 illustrate three simulated $3d - \text{VAR}(1)$ processes where the term 3*d* shows that there are 3 variables in the model.

Following the methodology of the AR processes shown in Figures 2.1 and 2.2, the VAR processes are constructed by generating 300 observations from a series whose error terms are simulated from a standard normal distribution with a mean of zero and standard deviation of one. These observations are then used in the VAR processes to obtain 300 observations of each series $\mathbf{x}_{it}$ with the process initialised by setting $\mathbf{x}_{i0} = \mathbf{0}$. To minimise the initial condition effect and to improve randomness, the first 200 observations were omitted and the remaining 100 observations were used.

A visual inspection of Figure 2.3 lends support to the expectation that the process is stationary as each individual series appears stable with constant mean and variance. The series is simulated such that the coefficients of the lag terms ensure that each single equation is stationary.

Figure 2.3: $3d-\mathrm{VAR}(1)$ plot for a stationary process



Stationary VAR process simulated by the model

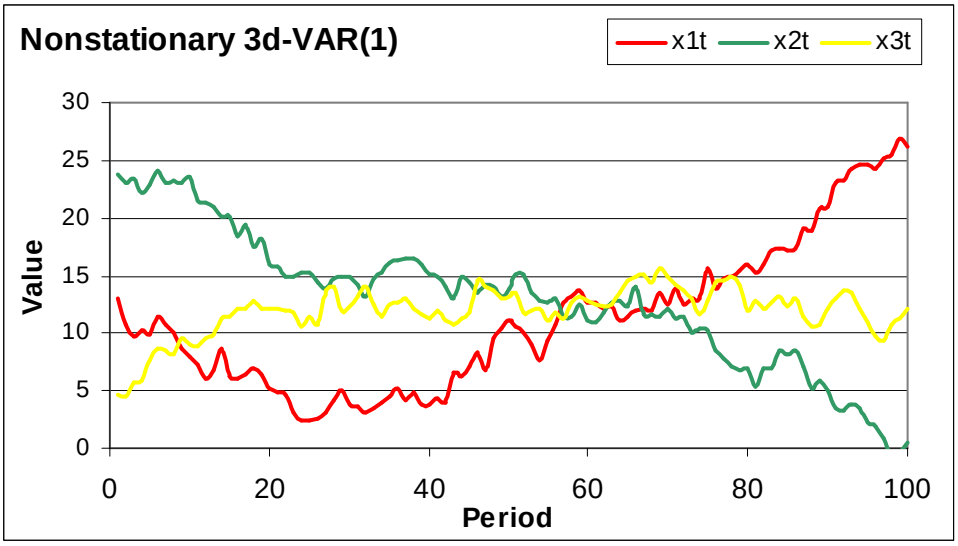
$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \\ 5 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0 & 0.6 & 0 \\ 0 & 0 & -0.7 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

On the contrary, the series in Figures 2.4 and 2.5 lend support to the expectation that the processes are nonstationary as each individual series appears less stable with a mean that is

time dependent and a variance that is time invariant. Although a visual inspection is satisfactory at this introductory stage, inferential procedures are preferred and this study applies the more universally practised inferential methods developed for unit root testing.

The process modelled in Figure 2.4 is the simplest of the nonstationary processes in that the coefficient of the lagged term for each single equation ensures that each individual equation is a random walk. The coefficients chosen for the model in Figure 2.5 ensure interdependencies of the equations whilst in Figure 2.4 the equations are all independent. This interdependency is the more interesting process and is referred to as a nonstationary process, one whose characteristics are the emphasis of this study. A more detailed description, with definitions and literature references are provided elsewhere in this chapter.

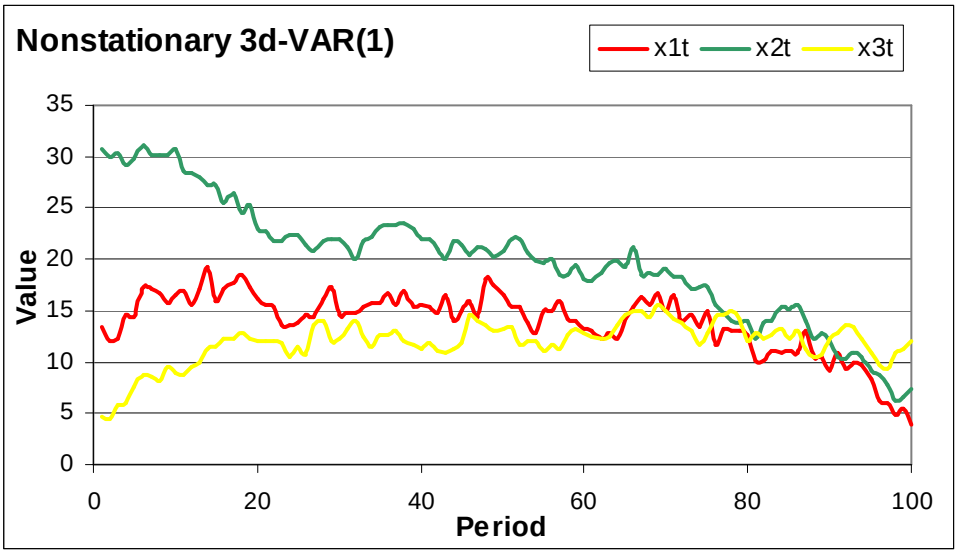
Figure 2.4: $3d-\mathrm{VAR}(1)$ plot for a nonstationary process



Nonstationary VAR process simulated by the model

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} = \begin{pmatrix} 20 \\ 30 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

Figure 2.5: $3d-\text{VAR}(1)$ plot for a nonstationary process



Nonstationary VAR process simulated by the model

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} = \begin{pmatrix} 15 \\ 37 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

The Box-Jenkins approach to analysing nonstationary processes requires the differencing of the nonstationary series in order to transform the process into a stationary series and then follow their estimation and forecasting methodology. Although this method is still used, the development of the unit root theory and testing by Dickey and Fuller $(1979, 1981)$ with further developments by Phillips and Perron (1988) and the subsequent theory of cointegration by Engle and Granger (1987) and Johansen $(1988, 1991)$ has meant that nonstationary processes can now be analysed in a cointegrated (regression) framework.

Prior to this development the inclusion of nonstationary variables in a regression framework resulted in the well-documented spurious regression problem. The spurious regression problem, in the terminology of Granger and Newbold (1974), occurs when a causal relationship between nonstationary variables in a regression framework is inferred

whilst the error terms of the model are also nonstationary. The assumptions of the linear model are that the error terms are white noises with no autocorrelation. Given that the error terms are nonstationary implies that autocorrelation is present and the assumption is violated.

2.2 The Analysis of Time Series Data

The earlier analytical time series methods focused on estimating and forecasting of the ARIMA models presented in Box and Jenkins (1976). In this section a brief summary of the methods and ideas of the ARIMA models are discussed as they provide a background for what is to follow. More detailed information on these models is available in several time series texts; for a more theoretical approach see for example Box and Jenkins (1976) or Kendall and Ord (1990) while for the practical aspects refer to Bowerman and O'Connell (1993). A more recent text by Enders (2004) provides a less dated approach to time series modelling whilst still retaining sufficient theoretical underpinnings to compliment the practical direction followed in recent years.

The estimation and forecasting methods presented by Box and Jenkins (1976) require that the series be stationary. In cases where the series was stationary, the order of integration was defined as zero and denoted $\mathrm{I}(0)$. Nonstationary processes were differenced until the resulting (differenced) series was stationary. The number of times (d) the series was differenced in order for the series to be stationary was defined as the order of integration, denoted as $\mathrm{I}(d)$. These concepts lead to the defining of ARIMA models.

An ARIMA model consists of three components, an autoregressive $(\mathrm{AR}(p))$ component with p terms, a $d^{th}-$order of integration $(\mathrm{I}(d))$ component, and a moving average, $(\mathrm{MA}(q))$ component with q terms. The model is represented as an $\mathrm{ARIMA}(p,d,q)$ model.

A typical $\text{ARIMA}(p, 0, q)$ model, where the order of integration is zero, is simplified to an $\text{ARMA}(p, q)$ model and is given by

$$x_t = a_0 + \underbrace{a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p}}_{p \text{ autoregressive components}} + \varepsilon_t + \underbrace{b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2} + \ldots + b_q \varepsilon_{t-q}}_{q \text{ moving average components}},$$

a process with *p* lag terms and *q* moving average terms.

This model can be written as two general sub-models, an $\text{AR}(p)$ model and a $\text{MA}(q)$ model. The models illustrated in Figures 2.1 and 2.2 are representations of $\text{AR}(1)$ models with coefficients $a_0 = 0$ and a_1 model dependent.

Stationarity restrictions for the $\text{AR}(1)$ model are easily established. Enders $(2004:54-55)$ derives the necessary and sufficient conditions for the stationarity $\text{AR}(1)$ model using solutions to difference equations. This is a simple yet tedious procedure where the $\text{AR}(1)$ process is written as an infinite MA model which yields finite first and second moments provided the determinant of the coefficient of the lagged term, $|a_1|$, is less than one. It is less easy to establish these restrictions for higher order models with *p* lag terms but Enders $(2004:59)$ shows that all finite-order MA processes will always be stationary and all higher order AR processes will be stationary if the characteristic roots of the homogenous equation all lie inside the unit circle. This simplifies to the restriction that $1 - \sum_{i=1}^{p} a_i > 0$. Given that this study emphasises the multi-equation $\text{VAR}(p)$ framework, the stationarity restrictions for these cases are left for discussion later in this chapter.

The analytical methodology of Box and Jenkins (1976) provides for the estimation of the autocorrelation function (ACF) and partial autocorrelation function (PACF) from the estimated ARIMA model. Characteristics of the estimated ACF and PACF are then compared to the characteristics of the theoretical ACF and PACF to determine the order of

the model. The assumptions that the series is stationary and that the errors are normally distributed made in Box and Jenkins (1976), then allows for significance testing of a group of sample autocorrelations. The Box and Pierce (1970) Q-statistic, originally used for this purpose has since been surpassed by the Ljung and Box (1978) Q-statistic which is now used for the significance tests of sample autocorrelations. Forecasting with the estimated model then follows. In summary, the Box-Jenkins methodology requires the tentative identification of the model using sample ACFs and sample PACFs. This is followed by the estimation of the tentative models, then checking model parameter estimates using Q-statistics and standard checks on the assumption of error normality and thereafter followed by forecasting if required.

The singular most distinct drawback to the Box-Jenkins methodology is the requirement that the series under consideration must be stationary. This drawback has been addressed to a large extent by the cointegrated analytical methodology proposed by Robert F. Engle III and Clive W.J. Granger, Economics Nobel Laureates 2003, and extended to the multi-equation systems developed to a large extent by Johansen (1988, 1991, 1992, 1995, 2005).

2.3 Unit Root Testing

The development of unit root theory, initially proposed by Dickey and Fuller (1979, 1981), has spawned a generation of unit root research. Unit root theory is the cornerstone to the methodology used for testing the stationarity or nonstationarity of a time series. Now that many of the procedures are standard offerings in econometric software packages, they have become routine tools for time series analysts.

The original research emphasised inferential methods for testing of nonstationary series that were first order difference stationary. The test procedures were developed for models with and without intercept terms, trend terms, structural breaks in the series and other fixed regressor terms. Developing inferential procedures for higher order series and addressing concerns relating to the power of the original inferential tests followed this original research. A more recent addition to the literature (Muller & Elliot, 2003) attempts to

determine the optimal unit root test and asserts that there is little value in trying to determine other unit root test statistics as the available procedures are near optimal and leave little opportunity for further exploitation.

Assuming this an acceptable conclusion, we applied arguably the most commonly used unit root test in the analysis of the data that was simulated for the theoretical models under consideration. The inferential procedure used in the analysis of the simulated series is the *augmented Dickey-Fuller* (ADF) unit root test. The ADF test used in this research was selected more for convenience than any other reason. The test is a standard routine in EViews 5.1, the software package used extensively throughout this study. The procedure followed in EViews 5.1 is described briefly as this was the routine used when performing some of the sequential stepwise routines of the methodology for this study.

The ADF test constructs a model with higher order lag terms and tests the significance of the parameter estimates using a non-standard *t-test*. The model used for this routine is $\Delta x_t = \alpha_1 x_{t-1} + \beta_1 \Delta x_{t-1} + \beta_2 \Delta x_{t-2} + ... + \beta_{p-1} \Delta x_{t-p+1} + \varepsilon_t$, where the *t-test* checks significance of the α_1 term. There are variations of this model which can account for a constant term and/or a trend term but the simulated models assessed in this study omitted constants and trends and thus no further explanation is provided. The interested reader may refer to Enders $(2004:182)$ which provides details of additional Dickey-Fuller tests.

2.4 Cointegration

The concept of a cointegrated time series is introduced at this stage. A brief introduction to some of the lag length selection techniques available to practitioners is given. Included are some of the concerns of these applications with respect to the cointegrated model. The introduction is brief and the detailed theoretical concepts are left to the relevant section of this chapter.

The argument for an error correction term in an autoregressive model by Engle and

Granger (1987) was the catalyst for extensive theoretical and empirical research. Additional theoretical developments were provided by Johansen (1988, 1991, 1992, 1995, 2005), MacKinnon (1991, 1996) and Stock and Watson (1988, 1991, 1993), whilst empirical research provided phenomenal growth as illustrated by the abundance of literature in applied econometric time series. Acceptance of an error correction term in econometric modelling has meant cointegration analysis has become a standard procedure worldwide. Commercial econometric software packages such as EViews 5.1 and STATA have pre-programmed cointegration routines, these simple to use software packages have allowed for application-based studies to become almost routine.

Empirical research began as early as the mid 1980s when Hall (1986) applied the single-equation Engle and Granger two-step procedure to construct an aggregate wage determination model. Since then multi-equation models have become common as illustrated by the foreign exchange rate market efficiency models analysed by Kellard, Newbold and Rayner (2001) and Ferre and Hall (2002) and the monetary system model analysed by Krolzig (2003).

In Southern Africa, researchers have used cointegration analysis for several studies. Gumede (2000) models import demand for several economic sectors, de Wet (2000) investigates purchasing power parity equilibrium, Fedderke and Joao (2001) examine the relationship between futures and spot markets, Leng (2002) revisits the futures and spot markets and Viljoen (2003) investigates efficiency of grain commodities. These are just a few of the empirical cointegrated studies published.

2.5 Cointegrated Models

The pre-cointegration method of analysis for nonstationary time series data required that one transforms the series by differencing until a stationary series resulted (Barr & Kantor, 2002). The existence of the error correction term as shown by Engle and Granger (1987) has provided time series analysts with an alternative method for the analysis of these variables.

The original definition of cointegration as a single equation model was propositioned by Engle and Granger (1987), an approach that has subsequently been surpassed by the systems of equations (multi-equation) approach advocated by Johansen (1988, 1991, 1995). This section introduces the underlying principles of cointegration and provides the notation that is used in the forthcoming sections.

The theory of cointegration is introduced with elementary concepts and extended to the more advanced ideas. Consider the $\mathrm{AR}(1)$ process, $x_t = a_1 x_{t-1} + \varepsilon_t$, illustrated graphically in Figure 2.1. The process is stable if $|a_1| < 1$ and defined as stationary (Lütkepohl, 2005). In the case where $|a_1| > 1$ the process is explosive and considered to be of little importance to econometric analysis. The remaining scenario, when $|a_1| = 1$ results in the well documented random walk. The random walk is a nonstationary process as the process variance tends to infinity (Engle & Granger, $1991:67$). Models such as the random walk can be differenced, $\Delta x_t = x_t - x_{t-1}$, resulting in a stationary differenced process. Processes that require differencing for stationarity are called nonstationary processes. A nonstationary process that when differenced *d* times results in a stationary process, is said to be a nonstationary process with order of integration, $\mathrm{I}(d)$. We omit further reference to explosive nonstationary processes as we only consider nonstationary processes that can be integrated into stationary processes.

Engle and Granger (1987) developed the theory that there exists the special case where linear combinations of nonstationary processes are stationary. They defined this linear combination of nonstationary processes as cointegration and used the notation $\mathrm{CI}(d, b)$, where *d* represents the order of integration of the nonstationary processes and *b* represents the number of stationary linear combinations between the nonstationary processes. Consider two $\mathrm{I}(1)$ processes, x_{1t} and x_{2t}, if there exists a linear combination of the two processes such that the linear combination is $\mathrm{I}(0)$, the two $\mathrm{I}(1)$ processes are considered to be $\mathrm{CI}(1,1)$.

In the notation of Engle and Granger (1987), they considered the vector $\mathbf{x_t} = (x_{1t}, x_{2t}, ..., x_{nt})^{\mathbf{t}}$ and defined that the *n* components of $\mathbf{x_t}$ were $\mathrm{CI}(d, b)$, if

- all *n* components of $\mathbf{x_t}$ are $\mathrm{I}(d)$, and
- there exists a non-zero vector $\boldsymbol{\alpha}$ such that $\mathbf{y_t} = \boldsymbol{\alpha}^{\mathbf{t}}\mathbf{x_t} \sim \mathrm{I}(d-b)$ with $b > 0$. The vector $\boldsymbol{\alpha}$ is called the cointegrating vector.

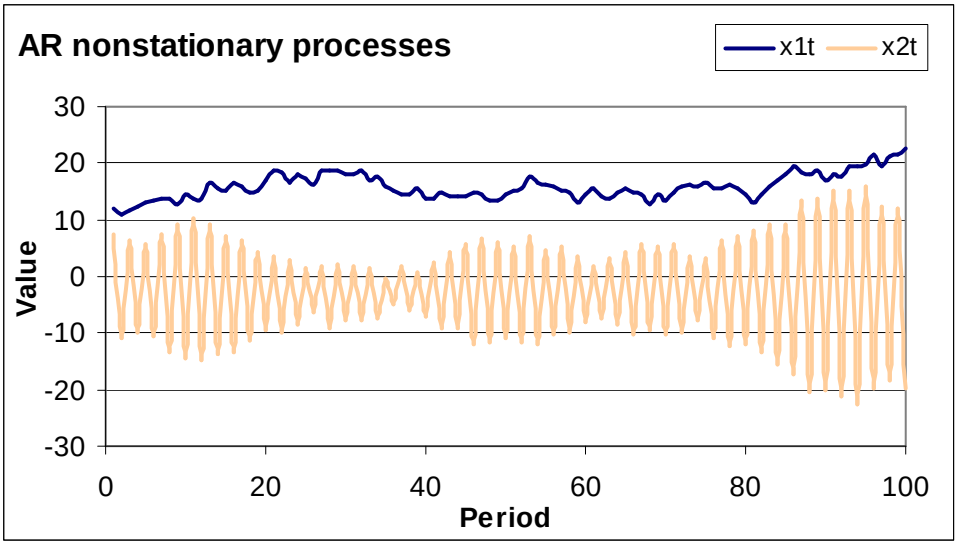
In conclusion it was shown that if $\mathbf{x_t}$ is cointegrated, there exists an error correction model (ECM), $\mathbf{\Delta x_t} = \mathbf{\Pi x_{t-1}} + \text{other } \mathbf{A_i \Delta x_{t-i}} \text{ terms} + \boldsymbol{\varepsilon_t}$, where all terms in the ECM are $\mathrm{I}(d-b)$. The existence of the error correction term, $\mathbf{\Pi x_{t-1}}$, shows that cointegrated variables are influenced by the extent of deviation from their long run equilibrium.

To demonstrate these concepts consider the two $\mathrm{I}(1)$ processes, x_{1t} and x_{2t}, which in the notation of Engle and Granger (1987) are represented by, $y_t = \alpha_1 x_{1t} + \alpha_2 x_{2t}$. If y_t is an $\mathrm{I}(0)$ process, x_{1t} and x_{2t}, have one cointegrating relationship with the cointegrating vector, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^{\mathbf{t}}$. It is worth noting, that the number of cointegrating relationships is always less than the number, *n*, of $\mathrm{I}(d)$ processes. In this example, there are two $\mathrm{I}(1)$ processes, therefore the number of cointegrating relationships must be less than two.

Figures 2.6 and 2.7 show two pairs of simulated $\mathrm{AR}(1)$ nonstationary processes. The nonstationary processes were constructed by generating 300 observations and omitting the first 200 observations. The remaining 100 data observations were then used for modelling. The two processes in Figure 2.6 are both $\mathrm{I}(1)$ and if differenced are $\mathrm{I}(0)$. Intuitively one can see that the series are nonstationary as the x_{1t} process illustrates that the mean changes over time and that the x_{2t} process illustrates a non constant variation over time. Intuitively the series are not cointegrated as they do not "move" together. By construction, the series move in a different manner which illustrates that $\mathrm{I}(1)$ series do not necessarily have linear combinations that are $\mathrm{I}(0)$. Rather the norm is that the linear combination of two $\mathrm{I}(1)$

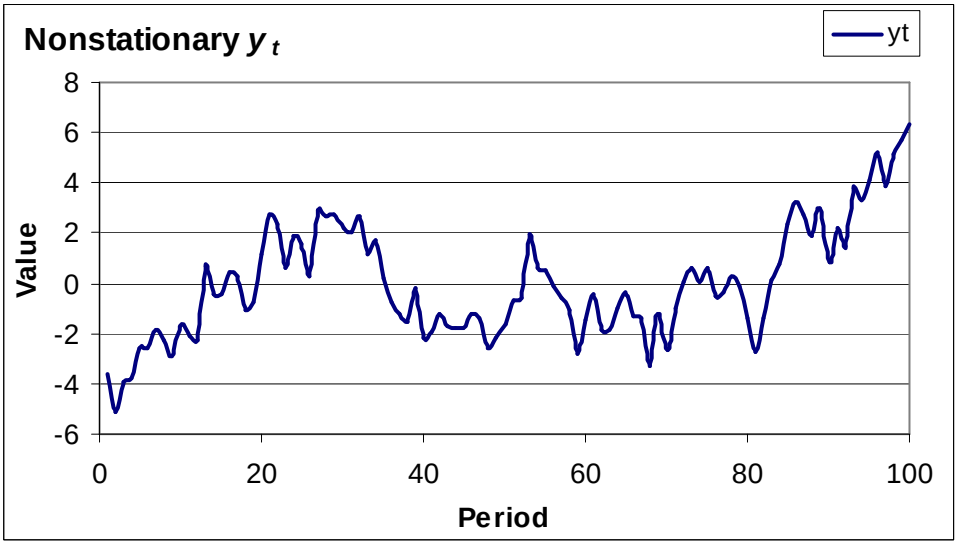
processes is itself $I(1)$.

Figure 2.6: Two AR plots for nonstationary processes



Nonstationary processes simulated by the models;

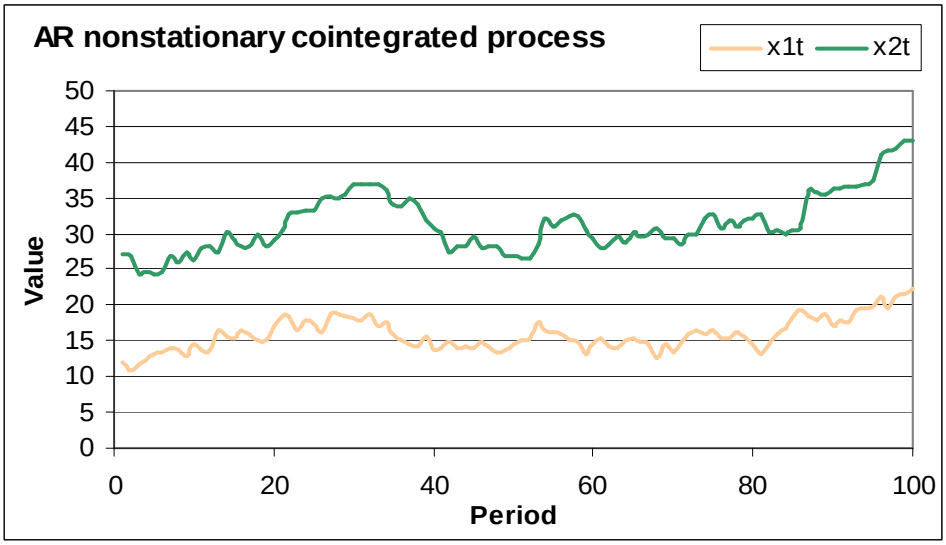
$x_{1t} = x_{1t-1} + \varepsilon_t$ and $x_{2t} = -0.35x_{1t-1} - x_{2t-1} + \varepsilon_t$



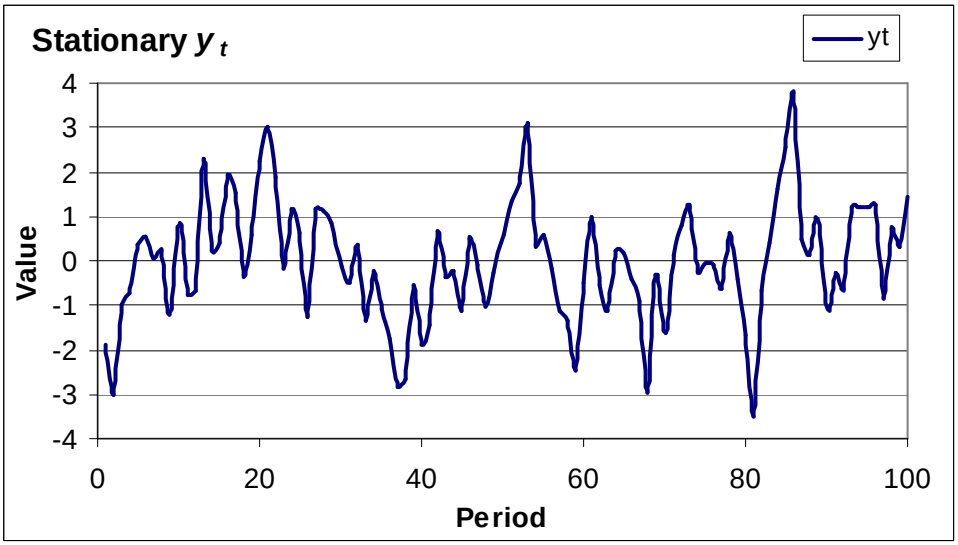
A nonstationary plot of the linear combination of $y_t = \alpha_1 x_{1t} + \alpha_2 x_{2t}$

In Figure 2.7, it appears that the two $I(1)$ processes "move" together, as x_{1t} decreases, so to does x_{2t} and as x_{1t} increases, so to does x_{2t}. It seems intuitive that the two $I(1)$ processes are correlated implying a co-dependency.

Figure 2.7: Two nonstationary processes plotted for the same period



$x_{1t} = x_{1t-1} + \varepsilon_t$ and $x_{2t} = 0.6x_{1t-1} + 0.7x_{2t-1} + \varepsilon_t$



A stationary plot of the linear combination of $y_t = \alpha_1 x_{1t} + \alpha_2 x_{2t}$

Given that the individual processes are nonstationary and the extent of the deviation between the two processes is constant, the joint bivariate process is cointegrated and their linear combination is stable such that the combination is stationary, i.e. $\mathrm{I}(0)$. These plots are used for intuitive illustrations; inferential procedures were used for the analytical problems encountered in this study.

Despite the seminal work of Engle and Granger (1987), there are several distinct disadvantages to their proposal. Most notably are the choice of the independent variable in their two-step estimation procedure and the detection of the number of cointegrating equations in models with more than two variables. These shortcomings were addressed by Johansen (1988, 1991) and Stock and Watson (1988) who considered cointegration in a system of equations and developed the theory that is used in more recent empirical studies.

The multi-equation theory of cointegration is introduced with elementary concepts and generalised to more advanced ideas. The introduction begins by considering a linear two-*dimensional* VAR process with one lagged term, hereafter denoted as a $2d-\mathrm{VAR}(1)$ process. Additional information of VAR processes are provided in Hamilton (1994), this thesis begins the discussion from the simplest model to consolidate notation and understanding.

The process is defined as the $2d-\mathrm{VAR}(1)$ model and in matrix notation is written as $\mathbf{x_t = A_1 x_{t-1} + \varepsilon_t}$ where the term $\mathbf{x_t}$ denotes a two *dimensional* vector with elements in two rows and one column defined as a (2x1) vector and represented as, $\mathbf{x_t} = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$. The lagged term, $\mathbf{x_{t-1}}$ denotes a (2x1) vector, and is represented as, $\mathbf{x_{t-1}} = \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix}$.

The $2d-\mathrm{VAR}(1)$ model is also referred to as a bivariate $\mathrm{VAR}(1)$ model and when written in full is given by $\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} a_{1,11}x_{1t-1} + a_{1,12}x_{2t-1} + \varepsilon_{1t} \\ a_{1,21}x_{1t-1} + a_{1,22}x_{2t-1} + \varepsilon_{2t} \end{pmatrix}$.

This model has the (2x2) parameter matrix $\mathbf{A_1} = \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix}$.

Assuming that both processes, denoted by x_{1t} and x_{2t}, are first order difference stationary i.e. $\mathrm{I}(1)$ and that they are cointegrated then, following Engle and Granger (1987), this

model can be represented as an ECM by subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix}$ from both sides of the equation and rearranging into ECM format. All individual components of the ECM will now be $\mathrm{I}(0)$ processes, with a parameterised error correction term. This representation is shown below.

Subtract $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix}$ from both sides of the equation:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} - \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Re-arrange the RHS by grouping $\mathbf{x_{t-1}}$:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = -\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$= -\begin{pmatrix} 1-a_{1,11} & 0-a_{1,12} \\ 0-a_{1,21} & 1-a_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

Finally simplify the representation by writing the model in matrix notation as $\mathbf{\Delta x_t = -\left(I - A_1\right) x_{t-1} + \varepsilon_t = \Pi x_{t-1} + \varepsilon_t}$, where $\mathbf{\Pi = -\left(I - A_1\right)}$.

The Johansen (1988) methodology considers three cases for the rank of the parameter matrix $\mathbf{\Pi}$. This is shown by considering the following scenarios. Consider all three simple cases when elements $a_{1,12} = a_{1,21} = 0$, then $\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = -\begin{pmatrix} 1-a_{1,11} & 0 \\ 0 & 1-a_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$.

If $\mathbf{\Pi}$ has rank 0, then $a_{1,11} = a_{1,22} = 1$. This restriction ensures variables, x_{1t} and x_{2t}, in the system are random walk processes and are nonstationary. In this case there is no evidence of cointegration, the two variables are independent and no linear combination will be stationary.

This is seen by rewriting the $2d-\mathrm{VAR}(1)$ as $\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1x_{1t-1} + \varepsilon_{1t} \\ 1x_{2t-1} + \varepsilon_{2t} \end{pmatrix}$.

If $\mathbf{\Pi}$ has (full) rank 2, then both $a_{1,11}$ and $a_{1,22} \neq 1$. Excluding explosive processes, both $a_{1,11}$ and $a_{1,22} < |1|$. These restrictions ensure that variables, x_{1t} and x_{2t}, in the system of equations are stationary. If they are stationary, they are $\mathrm{I}(0)$ and not $\mathrm{I}(1)$ processes and by definition the system of equations is not cointegrated.

This is seen by rewriting the $2d-\mathrm{VAR}(1)$ as

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} a_{1,11} & 0 \\ 0 & a_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} a_{1,11}x_{1t-1} + \varepsilon_{1t} \\ a_{1,22}x_{2t-1} + \varepsilon_{1t} \end{pmatrix}.$$

The interesting case is if the rank of $\mathbf{\Pi}$ is between zero and two, i.e $\mathrm{rank}(\mathbf{\Pi}) = 1$, then there exists a linear combination of the columns and rows of $\mathbf{\Pi}$ that ensure that either Δx_{1t} or Δx_{2t} can be written as a linear combination of the other. This is illustrated by considering the $2d-\mathrm{VAR}(1)$ model where $\mathbf{\Pi} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}$.

Given that the rank of $\mathbf{\Pi} = 1$, let $\pi_{11} = \pi_{12} = \pi_{21} = 0$, then $\pi_{22} \neq 0$. If $\pi_{11} = 0$ then $a_{1,11} = 1$ and the series x_{1t} is a random walk, $\mathrm{I}(1)$ process. If $\pi_{22} \neq 0$ and explosive processes are excluded, then $a_{1,22} < |1|$. Then x_{2t} is a stationary process. This would indicate that we have one $\mathrm{I}(1)$ process and one $\mathrm{I}(0)$ process and thus a system with no cointegration. Now consider the case when either of π_{12} or π_{21} or both ≠ 0. To ensure rank $\mathbf{\Pi} = 1$, there exists $a_{1,12}$ or $a_{1,21}$ or both ≠ 1, which ensures that there is a linear combination between x_{1t} and x_{2t}. This linear combination between x_{1t} and x_{2t} ensures that the $\mathrm{I}(0)$ process of x_{2t} is dominated by the $\mathrm{I}(1)$ process, x_{1t-1}, and is thus $\mathrm{I}(1)$ and that the two $\mathrm{I}(1)$ processes are

related in such a manner that their relationship results in a lower order integrated process, i.e. exhibits the evidence of cointegration as defined by Engle and Granger (1987).

The $2d-\mathrm{VAR}(1)$ discussion is generalised to the multi-equation theory of cointegration by considering a *k-dimensional* VAR process with one lagged term, denoted as a $\mathrm{VAR}(1)$ process. The linear $\mathrm{VAR}(1)$ model is defined as, $\mathbf{x_t = A_1 x_{t-1} + \varepsilon_t}$ where the term $\mathbf{x_t}$ represents a *k-dimensional* vector for a multivariate series. In vector notation this (*k*x1) vector is represented as, $\mathbf{x_t} = (x_{1t}\ x_{2t} \cdots x_{kt})^{\mathbf{t}}$. The lagged term, $\mathbf{x_{t-1}}$ is also a (*k*x1) vector and is shown as the lagged term of the (*k*x1) vector $\mathbf{x_t}$. This is represented as $\mathbf{x_{t-1}} = (x_{1t-1}\ x_{2t-1} \cdots x_{kt-1})^{\mathbf{t}}$.

The error term, $\boldsymbol{\varepsilon_t}$, is a (*k*x1) error vector represented as $\boldsymbol{\varepsilon_t} = (\varepsilon_{1t}\ \varepsilon_{2t} \cdots \varepsilon_{kt})^{\mathbf{t}}$. The error term, $\boldsymbol{\varepsilon_t}$, is assumed to be a vector white noise process. That is, it is assumed that $E(\boldsymbol{\varepsilon_t}) = \mathbf{0}$, the covariance matrix is finite, $E(\boldsymbol{\varepsilon_t}\ \boldsymbol{\varepsilon_t^t}) = \boldsymbol{\Sigma_\varepsilon}$ and there is no autocorrelation in the error structure, $E(\boldsymbol{\varepsilon_t}\,\boldsymbol{\varepsilon_s^t}) = \mathbf{0} \quad \forall t \neq s$. In addition, the covariance matrix $\boldsymbol{\Sigma_\varepsilon}$ is, unless stated, always assumed to be non-singular.

The $\mathrm{VAR}(1)$ model is given by

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix} = \begin{pmatrix} a_{1,11}x_{1t-1} + a_{1,12}x_{2t-1} + \cdots + a_{1,1k}x_{kt-1} \\ a_{1,21}x_{1t-1} + a_{1,22}x_{2t-1} + \cdots + a_{1,2k}x_{kt-1} \\ \vdots \\ a_{1,k1}x_{1t-1} + a_{1,k2}x_{2t-1} + \cdots + a_{1,kk}x_{kt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}.$$

This model can be written as an ECM by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix}$ from both sides of the equation and rearranging:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{pmatrix} - \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} - \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}$$

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \vdots \\ \Delta x_{kt} \end{pmatrix} = -\left\{ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} - \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = -\left(I - A_1\right) x_{t-1} + \varepsilon_t}$$

$$\mathbf{\Delta x_t = \Pi x_{t-1} + \varepsilon_t} \qquad \text{where } \mathbf{\Pi = -\left(I - A_1\right)}.$$

Analogous to the discussion for the bivariate process consider the rank of $\mathbf{\Pi}$. If $\mathbf{\Pi}$ has full rank *k*, then all variables in the system are stationary and no cointegration exists. If the rank of $\mathbf{\Pi}$ is 0, then all variables in the system are nonstationary random walks (excluding explosive cases) and no cointegration exists. If the rank of $\mathbf{\Pi}$ is *r*, where $0 < r < k$, then the system is said to have *r* cointegrating vectors.

The $\mathrm{VAR}(1)$ model is easily generalized to include additional lagged terms. Consider the *k-dimensional* VAR process with *p* lagged term, hereafter denoted as a $\mathrm{VAR}(p)$ process. The linear $\mathrm{VAR}(p)$ model is defined as, $\mathbf{x_t = A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + \varepsilon_t}$ where the term $\mathbf{x_t}$ represents a *k-dimensional* vector for a multivariate series. A further generalisation can be made by adding other deterministic terms, in this case define the linear $\mathrm{VAR}(p)$

model as, $\mathbf{x_t = A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + \Phi D_t + \varepsilon_t}$ where the term $\mathbf{D_t}$, is a matrix of deterministic terms (which could include an intercept term and/or other fixed regressors). The model parameters are the p (kxk) matrices $\mathbf{A_1}$, $\mathbf{A_2}$, …, $\mathbf{A_p}$ and $\mathbf{\Phi}$ (whose dimension depends on $\mathbf{D_t}$) the parameter matrix of the deterministic terms.

For ease of exposition this study assumed a zero intercept term and zero fixed regressor terms for all models. This simplification makes little difference to the findings of the study and leaves an option for further research.

The *k-dimensional* $\mathbf{A_i}$ matrices are represented as

$$\mathbf{A_1} = \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix}, \quad \cdots, \quad \mathbf{A_p} = \begin{pmatrix} a_{p,11} & a_{p,12} & \cdots & a_{p,1k} \\ a_{p,21} & a_{p,22} & \cdots & a_{p,2k} \\ \vdots & & & \vdots \\ a_{p,k1} & a_{p,k2} & \cdots & a_{p,kk} \end{pmatrix}.$$

The complete $\mathrm{VAR}(p)$ model is represented as,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} + \cdots + \begin{pmatrix} a_{p,11} & a_{p,12} & \cdots & a_{p,1k} \\ a_{p,21} & a_{p,22} & \cdots & a_{p,2k} \\ \vdots & & & \vdots \\ a_{p,k1} & a_{p,k2} & \cdots & a_{p,kk} \end{pmatrix} \begin{pmatrix} x_{1t-p} \\ x_{2t-p} \\ \vdots \\ x_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}.$$

This model can be written as an ECM by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix}$ from both sides of the equation and rearranging:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{pmatrix} - \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix} - \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{2,11} & a_{2,12} & \cdots & a_{2,1k} \\ a_{2,21} & a_{2,22} & \cdots & a_{2,2k} \\ \vdots & & & \vdots \\ a_{2,k1} & a_{2,k2} & \cdots & a_{2,kk} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ \vdots \\ x_{kt-2} \end{pmatrix} + \cdots + \begin{pmatrix} a_{p,11} & a_{p,12} & \cdots & a_{p,1k} \\ a_{p,21} & a_{p,22} & \cdots & a_{p,2k} \\ \vdots & & & \vdots \\ a_{p,k1} & a_{p,k2} & \cdots & a_{p,kk} \end{pmatrix} \begin{pmatrix} x_{1t-p} \\ x_{2t-p} \\ \vdots \\ x_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}$$

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \vdots \\ \Delta x_{kt} \end{pmatrix} = -\left\{ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} - \begin{pmatrix} a_{1,11} & a_{1,12} & \cdots & a_{1,1k} \\ a_{1,21} & a_{1,22} & \cdots & a_{1,2k} \\ \vdots & & & \vdots \\ a_{1,k1} & a_{1,k2} & \cdots & a_{1,kk} \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ \vdots \\ x_{kt-1} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{2,11} & a_{2,12} & \cdots & a_{2,1k} \\ a_{2,21} & a_{2,22} & \cdots & a_{2,2k} \\ \vdots & & & \vdots \\ a_{2,k1} & a_{2,k2} & \cdots & a_{2,kk} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ \vdots \\ x_{kt-2} \end{pmatrix} + \cdots + \begin{pmatrix} a_{p,11} & a_{p,12} & \cdots & a_{p,1k} \\ a_{p,21} & a_{p,22} & \cdots & a_{p,2k} \\ \vdots & & & \vdots \\ a_{p,k1} & a_{p,k2} & \cdots & a_{p,kk} \end{pmatrix} \begin{pmatrix} x_{1t-p} \\ x_{2t-p} \\ \vdots \\ x_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = -\left(I - A_1\right) x_{t-1} + A_2 x_{t-2} + ... + A_p x_{t-p} + \varepsilon_t}$$

(iii) converting the RHS terms to stationary processes by adding and subtracting $\mathbf{A_2 x_{t-1}}$, $\mathbf{A_3 x_{t-1}}$, …, $\mathbf{A_p x_{t-1}}$ and $\mathbf{A_3 x_{t-2}}$, $\mathbf{A_4 x_{t-2}}$, …, $\mathbf{A_p x_{t-2}}$ and … and $\mathbf{A_p x_{t-p+1}}$:

$$\Delta \mathbf{x}_t = -\left(\mathbf{I} - \mathbf{A}_1\right)\mathbf{x}_{t-1}$$

$$+\left(\mathbf{A}_2\mathbf{x}_{t-1} + \mathbf{A}_3\mathbf{x}_{t-1} + \mathbf{A}_4\mathbf{x}_{t-1} + \ldots + \mathbf{A}_p\mathbf{x}_{t-1} - \mathbf{A}_2\mathbf{x}_{t-1} - \mathbf{A}_3\mathbf{x}_{t-1} - \mathbf{A}_4\mathbf{x}_{t-1} - \ldots - \mathbf{A}_p\mathbf{x}_{t-1}\right)$$

$$+\left(\mathbf{A}_3\mathbf{x}_{t-2} + \mathbf{A}_4\mathbf{x}_{t-2} + \ldots + \mathbf{A}_p\mathbf{x}_{t-2} - \mathbf{A}_3\mathbf{x}_{t-2} - \mathbf{A}_4\mathbf{x}_{t-2} - \ldots - \mathbf{A}_p\mathbf{x}_{t-2}\right)$$

$$+ \ldots + \left(\mathbf{A}_p\mathbf{x}_{t-p+1} - \mathbf{A}_p\mathbf{x}_{t-p+1}\right)$$

$$+ \mathbf{A}_2\mathbf{x}_{t-2} + \mathbf{A}_3\mathbf{x}_{t-3} + \ldots + \mathbf{A}_p\mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t$$

(iv) grouping the RHS terms as lagged terms and simplifying:

$$\Delta \mathbf{x}_t = -\left(\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \mathbf{A}_3 - \mathbf{A}_4 - \ldots - \mathbf{A}_p\right)\mathbf{x}_{t-1}$$

$$- \mathbf{A}_2\left(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}\right) - \mathbf{A}_3\left(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}\right) - \mathbf{A}_4\left(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}\right) - \ldots - \mathbf{A}_p\left(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}\right)$$

$$- \mathbf{A}_3\left(\mathbf{x}_{t-2} - \mathbf{x}_{t-3}\right) - \mathbf{A}_4\left(\mathbf{x}_{t-2} - \mathbf{x}_{t-3}\right) - \ldots - \mathbf{A}_p\left(\mathbf{x}_{t-2} - \mathbf{x}_{t-3}\right)$$

$$- \ldots - \mathbf{A}_p\left(\mathbf{x}_{t-p+1} - \mathbf{x}_{t-p}\right) + \boldsymbol{\varepsilon}_t.$$

With some simple re-arranging, the ECM is written as

$$\Delta \mathbf{x}_t = \boldsymbol{\Pi}\mathbf{x}_{t-1} - \sum_{i=2}^{p} \mathbf{A}_i\Delta\mathbf{x}_{t-1} - \sum_{i=3}^{p} \mathbf{A}_i\Delta\mathbf{x}_{t-2} - \ldots - \mathbf{A}_p\Delta\mathbf{x}_{t-p+1} + \boldsymbol{\varepsilon}_t$$

where $\boldsymbol{\Pi} = -\left(\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \ldots - \mathbf{A}_p\right)$.

In conclusion, if a nonstationary $\mathrm{VAR}(p)$ process can be described as a stationary VEC model with $(p-1)$ lag terms then according to the Engle and Granger (1987) representation theorem the $\mathrm{VAR}(p)$ process has at least one cointegrating relationship between the *k* variables.

The $\mathrm{VEC}(p-1)$ model is defined as $\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{j=2}^{p} \Gamma_j \Delta x_{t-j+1} + \varepsilon_t}$. Where $\mathbf{x_{t-1}}$ is the $k-dimensional$ error correction term with parameter matrix $\mathbf{\Pi = -\left(I - \sum_{i=1}^{p} A_i \right) \neq 0}$ and the stationary differenced lag terms parameterised by $\mathbf{\Gamma_j = \sum_{i=j}^{p} A_i}$ (for $j = 2, 3, ..., p$).

Analogous to the discussion for the $2d-\mathrm{VAR}(1)$ process, the *k-dimensional* $\mathrm{VAR}(p)$ process or alternatively the *k-dimensional* $\mathrm{VEC}(p-1)$ process can follow one of three possible cases depending on the rank of $\mathbf{\Pi}$. If $\mathbf{\Pi}$ has full rank *k*, then all variables in the system are stationary and by definition no cointegration exists. If rank of $\mathbf{\Pi}$ is 0, then all variables in the system are nonstationary but independent and no cointegration exists. If rank of $\mathbf{\Pi}$ is *r*, where $0 < r < k$ and all variables in the system are nonstationary with the same order of integration, then the system is said to have *r* cointegrating vectors.

2.6 Lag Length Selection in Cointegrated Models

To estimate the ECM for inferential and interpretive purposes requires the selection of the most appropriate number of differenced, lagged terms in the model. In several econometric models, the lag length choice can be justified on an economic theoretical basis but for those models that lack a theoretical basis, the selection of the number of differenced lag terms to include in the model is left to the analyst.

A review of the literature reveals that a few attempts have been made to determine the most appropriate method for selecting the number of differenced lag terms in cointegrated models. The results from these attempts have yet to become established practices as they have been cautious in their recommendations. Rather the empirical cointegration studies have applied the methods previously developed for ARMA and/or VAR models. This may or may not be the correct approach but it is worthwhile noting that despite the volume of

research in this area, consensus has yet to be reached on a unanimous approach. One may argue that the methods developed for stationary VAR models are adequate for VEC models, although Qu and Perron (2006) demonstrate that the additional restriction on the VAR provides theoretical differences between the IC for VAR and VEC models. Examples of this distinct lack of consistency in the published empirical studies undertaken for determining the most appropriate lag length of their cointegrated model are provided next.

Enders (2004) and Davidson and MacKinnon (2004) suggest that a likelihood ratio, χ^2 test be used to compare a complete model versus a restricted model. Several researchers have opted to use one (or more) of the information criteria identified in the ARMA and/or VAR modelling paradigm. Gumede (2000) analyses several single equation econometric models across eleven sectors. Sufficient evidence is provided to lend support to the claim that the variables in the models analysed are cointegrated. What is not provided is the reason or method followed to justify the inclusion of several lagged variables in the models. Other than concluding, "*The significance of the coefficient of lagged imports basically suggests that the last period imports influence current period imports, which is appealing*", no theoretical or inferential reason is provided for the inclusion of lagged terms. The comments, "*significance of*" and "*basically suggests*" could be interpreted as the variables coefficient is significant at some pre-determined significance level but how he decided to include or exclude lagged variables is unknown.

In the efficient market study of Fedderke and Joao (2001), the conclusion of a single cointegrated equation is reached using two analytical approaches. The ECM established has several lag terms and the justification provided for the inclusion of the lagged terms is based on Akaike's Information Criterion (AIC). In the efficient market study of Kellard, Newbold and Rayner (2001) the authors conclude that several of their models have cointegrated variables. The inclusion of lag terms in their models is justified using Schwarz's Information Criterion (BIC).

In a study by Leng (2002), the ECM estimated provides for three lag terms. The author chooses a model with three lag terms and motivates his choice based on the impact this has on reducing the effect of heteroscedasticity. If the author were to follow one of either the

Schwarz or Hannan-Quinn Information Criterion (HQIC) approaches to model selection, his results would have specified the inclusion of a single lag term.

This inconsistency in selecting the number of differenced lag terms for inclusion in the final cointegrated model is an important research opportunity, one which this thesis attempts to address. The results of this study address this inconsistency and provide a platform from which researchers can apply cointegrated modelling more uniformly.

2.7 Lag Length Selection Methods

The selection of a set of variables to be used in any model is an occupational hazard for a statistician. With the phenomenal growth in computational capacity comes the opportunity to investigate and analyse large datasets. Data mining is a buzzword that is common in the statistical literature and the debate over dimensionality of a model has raged back and forth from the "*Keep it sophisticatedly simple*" (KISS) as so elegantly put by Zellner (2003) to the concept of model complexity and selecting the best fitting model as argued by Bozdogan (2003).

Model selection strategies have been used extensively for the determination of lag length in a regression framework. Two distinct strategies have been followed in the literature, model selection based on likelihood ratio tests for the comparison of a model nested within a model and model selection based on theoretic information criteria (IC). The likelihood ratio test approach is advocated in the texts of Enders (2004) and Davidson and MacKinnon (2004) and is generally used in conjunction with a stepwise strategy to identify significant variables in a model.

The stepwise strategies have received considerable recognition in the literature and a scan in many time series and/or regression texts will convince anyone of the benefits of this approach, see for example, Mendenhall and Sincich (2003) or Bowerman and O'Connell (1993). Unfortunately though there are also many who emphasise the limitations of the stepwise strategies, see for example Montgomery, Peck and Vining (2001) and Bozdogan

(2003). Arguably the biggest concern to the stepwise strategy is the lack of any theoretical rationalisation for the sequence of variable selection. Burnham and Anderson $(2002:35-36)$ argue that likelihood ratio tests based on arbitrary (subjective) significance levels are problematic if many tests are to be made and perhaps more importantly likelihood ratio tests of models which are not nested within a model are not possible. In light of the limitations of the stepwise/likelihood approach it is worthwhile reconsidering the IC approach developed by Akaike (1973). The IC approach does not rely on some predefined significance level, nor does it matter whether or not a model is nested within another. The logic is to fit several competing models to a dataset and the IC will select the model which fits the data best.

Information theoretic criteria have come a long way since Akaike first used the relationship between the expected log-likelihood of a model and the sufficiency theorem of Kullback-Leibler to derive the well known information criterion, commonly referred to as AIC. The Akaike logic to IC has provided the inspiration for the developments of several model selection criteria. The IC framework has developed along two approaches; one that a true model exists and the IC attempts to determine the true model. These IC are asymptotically consistent, assume that the true model exists and is included in the group of candidate models assessed. The second method develops IC that are asymptotically efficient and assumes that the true model does not exist but identifies the most parsimonious model from the group of candidate models assessed. Akaike's was the latter of the two and was motivated by the desire to develop a measure that could be used to compare models of different dimensions trading off the fit of the model with respect to the number of parameters included in the model.

This study assesses the ability to select the correct lag structure of a cointegrated model using thirteen different criteria. These criteria have been developed to ensure they are either efficient or consistent estimators for model selection. The criteria used in this study are listed in Table 2.1. All criteria notation and references follow in Section 2.8 under the relevant IC. As an example, the reference to the term $\tau_{p-1}(r)$ for criterion 2 is found in Section 2.8 under criterion MAIC (on page 53).

Criterion #	IC definition in $\mathrm{VEC}(p-1)$ framework
1	$\mathrm{AIC}^{\mathrm{VEC}(p\text{-}1)} = \ln\left\lvert\hat{\boldsymbol{\Sigma}}\right\rvert + 2\dfrac{k^2(p-1)}{T}$
2	$\mathrm{MAIC}^{\mathrm{QP\text{-}VEC}(p\text{-}1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{2\left[\tau_{p-1}(r) + k^2(p-1)\right]}{T}.$
3	$\mathrm{AICC}^{\mathrm{HT\text{-}VEC}(p\text{-}1)} = \ln\left\lvert\hat{\boldsymbol{\Sigma}}\right\rvert + \dfrac{\left(kT + k^2(p-1)\right)}{\left(T - \left(k(p-1)+k+1\right)\right)}.$
4	$\mathrm{MAICC}^{\mathrm{SR\text{-}VEC}(p\text{-}1)} = \ln\left\lvert\hat{\boldsymbol{\Sigma}}\right\rvert + \dfrac{kT}{\left(T - \left(k(p-1)+k+1\right)\right)} + \dfrac{k^2(p-1)}{T} + \dfrac{\tau_{p-1}(r)}{T}$
5	$\mathrm{AICCBD}^{\mathrm{VEC}(p\text{-}1)} = \ln\left\lvert\hat{\boldsymbol{\Sigma}}\right\rvert + \dfrac{2k^2(p-1)+2}{\left(T - \left(k^2(p-1)+2\right)\right)}$
6	$\mathrm{BIC}^{\mathrm{VEC}(p\text{-}1)} = \ln\left\lvert\hat{\boldsymbol{\Sigma}}\right\rvert + \dfrac{k^2(p-1)\ln T}{T}$
7	$\mathrm{HQIC}^{\mathrm{VEC}(p-1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{2k^2(p-1)\ln\ln T}{T}$
8	$\mathrm{HQICC}^{\mathrm{VEC}(p-1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{2k^2(p-1)\ln(\ln T)}{T - \left(k(p-1)+k+1\right)}$
9	$\mathrm{LCIC}^{\mathrm{VEC}(p-1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{k^2(p-1)}{T}\left((\ln T + 2\ln\ln T)/2\right)$
10	$\ln\left(\mathrm{FPE}^{\mathrm{VEC}(p-1)}\right) = \ln\left\lvert\hat{\Sigma}\right\rvert + k\ln\left(\dfrac{T + k(p-1)}{T - k(p-1)}\right)$
11	$\mathrm{ShibIC}^{\mathrm{VEC}(p-1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + k\ln\left(1 + \dfrac{2\left(k(p-1)+1\right)}{T}\right)$
12	$\mathrm{MBIC}^{\mathrm{QP\text{-}VEC}(p\text{-}1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{\left(\tau_{p-1}(r) + k^2(p-1)\right)\ln T}{T}$
13	$\mathrm{MHQIC}^{\mathrm{QP\text{-}VEC}(p-1)} = \ln\left\lvert\hat{\Sigma}\right\rvert + \dfrac{2\left(\tau_{p-1}(r) + k^2(p-1)\right)\ln(\ln T)}{T}$

Table 2.1: Complete list of criteria assessed in this study

Akaike's criterion has been derived in a multi-equation framework, and then bias corrected to improve capability for small sample cases (Hurvich & Tsai, 1993). This was extended to the case of error restricted nonstationary VAR models (Qu & Perron, 2006) and in this study corrected to improve selection for small samples. To provide a more detailed comparative study, several other criteria used elsewhere in the literature were included for assessment.

The majority of the criteria extracted from the literature have been defined in terms of *k-dimensional* $\mathrm{VAR}(p)$ processes. This study restricts the $\mathrm{VAR}(p)$ process to a $(p-1)$ error restricted process denoted as $\mathrm{VEC}(p-1)$. The criteria evaluated are therefore assessed given that the VEC model has $(p-1)$ lag terms rather than the traditional p lag terms in a VAR framework.

An additional problem with model selection in a VAR framework was that the number of observations changed as the lag structure changed. This problem was addressed by Ng and Perron (2005) who emphasised the benefits of holding the sample size fixed. This was particularly important in cases where the number of observations changed subject to the number of lag terms included in a model. The best results for IC model selection were obtained when the number of observations for each model were kept the same. This result was supported by Qu and Perron (2006) and thus the same approach was adopted in this study. In all cases, the number of observations used in the assessment of the IC is denoted by T whilst the actual number of observations available (although not always used) is N with $T \leq N$.

Most of the criteria given in the forthcoming section were developed or modified assuming a multi-equation, $kd-\mathrm{VAR}(p)$ time series framework. The criteria were developed over the last decade or two and in many cases applied to simulation and/or empirical studies. In several cases the IC had been derived by the author(s) and tested individually. Several attempts to compare some of the well known IC simultaneously have been made and these studies are discussed both in the literature review and compared to the simulation results

of this study. Simulation exercises have been done in several frameworks, linear regression, ARMA time series, multi-equation VAR models and recently in a VEC system of equations.

The exercises most relevant to this study are the studies by Lütkepohl (1985), Hurvich and Tsai (1993), Koreisha and Pukkila (1993), Gonzalo and Pitarakis (1998, 2002), Kadilar and Erdemir (2002), and Qu and Perron (2006, 2007). The studies of Ng and Perron (2001, 2005), Ivanov and Kilian (2005), Brüggemann, Lütkepohl and Saikkonen (2006) and Baltagi and Wang (2007) also have relevance. Selected results of published studies are tabulated in Chapter 6. The summarised results are used as a reference point to compare the results of this study.

After the theoretical IC publications in the 1970s and early 1980s, Lütkepohl (1985) undertook the first VAR modelled Monte Carlo experimental study to compare criteria performances. The study was used to determine how frequently a criterion chose the correct theoretical bivariate and trivariate stationary VAR models from data simulated with 1000 replications per model. The results of the study showed that the likelihood ratio tests performed poorly; in particular for small sample sizes overestimating the VAR order was the norm. The Lütkepohl (1985) study supported a preference for an IC analytical approach, an additional motivation for the methodology followed in this study. From a results perspective, the IC of Schwarz and Hannan-Quinn produced the better results, followed by AIC and Akaike's final prediction error (FPE). Of particular interest to the current study was the observation that for the $3d-\mathrm{VAR}(1)$ model, the FPE outperformed AIC for the smaller sample sizes. It was concluded that as the sample size increased the performances of the IC improved and that the differences in performances between IC decreased. Lütkepohl's (1985) conclusions were reached after the evaluation results of a $\mathrm{VAR}(1)$ model were completed. Numerical results for this model were not reported but criteria were ranked according to performance. The best performer was BIC followed by HQIC, FPE, AIC and finally ShibIC.

The simulation study by Koreisha and Pukkila (1993) has relevance to this research. In the

study, extensive assessments of both $3d-\mathrm{VAR}(1)$ and $3d-\mathrm{VAR}(2)$ stationary models were undertaken. Sample sizes for the simulation were $T=50$ and $T=100$, whilst two alternative model parameterisations were used. The study assessed the selection performance of AIC, BIC and HQIC. The results for the $3d-\mathrm{VAR}(1)$ model showed that as the sample size increased the performances of all three criteria improved, a finding consistent with Lütkepohl (1985). Also noted was the influence of model parameter on criteria performance. As an illustration, the selection performance of BIC for the $3d-\mathrm{VAR}(1)$ model with $T=50$ was 60% for the first parameterisation and 100% for the second parameterisation. Koreisha and Pukkila (1993) were conservative in their model selection recommendations. Rather than advocate a single criterion, they noted that increasing the sample sized improved selection performances, especially those of BIC and HQIC. In addition they cautioned against the advocacy of the asymptotic property of consistency. Interpretations of these comments could imply a resistance to the consistent estimators of BIC and HQIC. Koreisha and Pukkila (1993) placed emphasis on how dependent criteria performances were on the number of variables in the model, the upper limit of the lag structure chosen for assessment and the number of non zero coefficient elements in the model. Summarised results from the Koreisha and Pukkila (1993) study are given in Chapter 6 in order to compare with the results of the current study.

The theoretical derivations of a small sample bias correction for AIC have been considered to a large extent by Chih-Ling Tsai and other co-authors. Hurvich and Tsai (1989) proposed a small sample less bias corrected Akaike information criterion (AICC) for linear regression and ARMA time series models. A small simulation exercise of 100 replications was used to compare AICC with amongst others AIC, FPE, BIC and HQIC. The results from the study reported that AICC outperformed the other IC for sample sizes of $T=10$ and $T=20$. The follow-up study reported by Hurvich, Shumway and Tsai (1990) proposed an unbiased small sample correction for ARMA time series models. The $\mathrm{AR}(2)$ simulation study undertaken provided satisfactory evidence to the researchers that AICC was a better performer than both AIC and BIC. This criterion was extended to consider $\mathrm{VAR}(p)$ processes when Hurvich and Tsai (1993) reported the results of 100 replication

simulation exercises for stationary $2d-\mathrm{VAR}(1)$ and $2d-\mathrm{VAR}(2)$ models. The sample size for both simulated VAR models was $T=40$ with a maximum lag length cutoff of six. In addition to comparing AICC with AIC and BIC, they also evaluated the performance of the criterion (AICCBD) proposed by Brockwell and Davis (1991, 432). The results from the study reported that AICC and AICCBD outperformed both AIC and BIC. These results have implications for the current research as a small sample bias corrected criterion in a VEC framework is suggested using the arguments of Hurvich and Tsai (1993).

Additional developments to AICC have been covered in a multivariate regression framework (Bedrick & Tsai, 1994), a quasi-likelihood model (Hurvich & Tsai, 1995) and semiparametric models (Simonoff & Tsai, 1999). The first two studies compared several criteria and reported results that showed that AICC outperformed the more commonly used criteria. In the quasi-likelihood study Hurvich and Tsai (1995) used the AICC for a logistic regression model. The practical illustration used in the study considered data from the space shuttle Challenger prior to the 1986 accident and showed that AICC, unlike AIC, selected the same significant variables in the model as in the detailed findings of Dalal, Fowlkes and Hoadley (1989). The analysis used 23 observations, a small sample. The third study considered AICC for the modelling of practical problems. Examples modelled included the ranking of academic institutions and a mileage versus vehicle horsepower exercise. The Simonoff and Tsai (1999) study was a practical exercise, illustrating in a semiparametric setting, the use of AICC rather than a comparative study.

Whilst evaluating a criterion based alternative to Johansen's cointegration likelihood ratio test, Gonzalo and Pitarakis (1998) simultaneously assessed model selection performances of four criteria. They proposed a new criterion, a linear combination of Schwarz's and Hannan-Quinn's criteria (LCIC) and evaluated this criterion with the selection performances of AIC, BIC and HQIC. The simulation exercise used a $3d-\mathrm{VAR}(1)$ model with sample sizes of $T=150, 250, ..., 650$ and several parameter alternatives. The parameter values ranged between 0.60 and 1.00 both inclusive. The parameter range implied that both stationary and nonstationary processes for the $3d-\mathrm{VAR}(1)$ model were evaluated. The results of the simulation exercise provided the researchers with evidence

supporting the use of BIC as the preferred model selector with LCIC and HQIC as alternatives. There was less support for the use of AIC which was reported as having overfitted the model more frequently than the other IC. To compare their results with the current study, summarised results from the Gonzalo and Pitarakis (1998) study are given in Chapter 6.

Gonzalo and Pitarakis (2002) continued the research concentrating more on model selection rather than cointegration test comparisons. Unlike the previous study, the conclusions reached were noticeably different. Given that the primary reason for the follow-up study was assessment of criteria performances for model selection the study was more comprehensive. The study considered multi-equation models of dimensions up to and including 10 variables. Only three criteria were assessed; a noticeable omission was any reference to LCIC, the criterion defined in the 1998 study. Gonzalo and Pitarakis (2002) only considered the performances of AIC, BIC and HQIC in a VAR framework. Of particular relevance to the current research were the performance capabilities of the IC for the stationary $3d-\mathrm{VAR}(2)$ simulated model. The results for this model strongly advocated the use of AIC as the preferred model selector. In general the results showed that both BIC and HQIC were good selectors for $\mathrm{VAR}(1)$ models but poor selectors for models with more than one lag term. It was these observations that presumably lead Gonzalo and Pitarakis (2002) to reconsider the recommendations of the 1998 study. The conclusions reached were that as the models' dimensions increased in lag structure, the best selector was AIC with both BIC and HQIC underfitting. The exception to this was the $\mathrm{VAR}(1)$ model but collectively AIC was the best performer. The summarised results from the Gonzalo and Pitarakis (2002) study are given in Chapter 6 in order to compare with the results of the current study.

An assessment of criteria performance for VAR and seasonal VAR models was undertaken by Kadilar and Erdemir (2002). Four criteria, AIC, BIC, HQIC and Shibata's criterion (ShibIC) were evaluated for simulated models with a sample sizes of $T=100$. The exercise included assessment of $2d$ and $3d$ VAR models with either one or two lag terms.

The study followed the parameterisation approach of Koreisha and Pukkila (1993) and the results provided support for the choice of BIC as the preferred model selector. This recommendation seemed to contradict the comments of Koreisha and Pukkila (1993) which were more conservative in their recommendations. The researchers concluded by ranking the criteria in preference order, BIC followed by HQIC, AIC and lastly ShibIC.

In an interesting study by Ivanov and Kilian (2005), empirical data were obtained from various sources and used to construct simulated series. The primary focus of the study was lag order selection for impulse response analysis whilst criteria performance was of secondary importance. The concept was interesting in that they took the empirical data and fitted the original model proposed by the data sources. The estimated parameters were then used as "theoretical" parameters in a Monte Carlo simulation. The IC were then evaluated to determine their ability to select the model from which the data were simulated. The study was extensive, monthly VAR models with sample sizes of $T = (240, 300, 360, 480, 600)$ were fitted, quarterly VAR models with sample sizes $T = (80, 100, 120, 160, 200)$ were fitted as were quarterly VEC models. Three criteria AIC, BIC and HQIC were assessed simultaneously using likelihood sequential testing procedures.

Several findings were reported, in particular, the IC outperformed the sequential testing procedures, a result consistent with Lütkepohl (1985). Contrasting performance results for the criteria were reported. AIC performed better that SIC and HQIC for the monthly VAR processes, SIC provided the best performance for quarterly VAR processes with sample sizes $T \leq 120$ whilst HQIC performed better for $T > 120$. This inconsistency in performance capability illustrates the difficulties analysts can expect when trying to determine the lag structure for multi-equation modelling.

Two cautionary notes were mentioned, the results showed that all criteria underfitted the model, most notably for $T = 240$ and that their results were limited by the simulation design used. Of some concern was the use of the data from some of the empirical studies which originally fitted particularly large models. As an example, it was stated that Leeper

(1997) used a six-dimensional model with eighteen lags. Assuming all parameters are fitted, this would require a minimum of $k^2 p = 6^2 \times 18 = 648$ estimated parameters, clearly not a parsimonious model. In conclusion, the study's intention was not to see which IC selected the lag length correctly, rather it was used to provide support for impulse response analysis and the conclusions reached were an aside issue.

Studies by Qu and Perron (2006, 2007) extended the modified information criteria (MIC) work of Ng and Perron (2001) to the multivariate framework. They considered vector autoregressive moving average (VARMA) processes and derived a modified Akaike information criteria (MAIC) which included an additional term in the penalty function. They justified the use of the statistic by motivating that the model was a constrained VAR with an error correction term. They cautioned that the use was restricted to cointegrated models and was quite likely inappropriate in other scenarios. The results of a simulation study for a bivariate model provided evidence that the MAIC performed the same as or better than both AIC and BIC. The results indicated that both AIC and BIC underfitted, particularly when a negative MA component was included in the model. The study concluded with a tri-variate VARMA design based on an empirical study by Yap and Reinsel (1995). The results obtained supported the claim that MAIC outperformed AIC and BIC. To compare their results with the current study, summarised results from the Qu and Perron (2006) study are give in Chapter 6.

It is worth noting that simulation based model selection is not restricted to VAR models, there have been several recent publications of model selection using simulation exercises in different frameworks. Included are a select few of the publications reviewed and the context of the study.

In a theoretical study, Ng and Perron (2001) considered lag length selection for unit root tests. They argued that the existing IC methods underfitted AR and MA models when the roots of the models were close to unity. This adversely affected the performance of existing unit root tests concluding with more rejections of the unit root hypotheses. They derived a class of MIC which included an additional penalty term for the well known AIC and BIC. Their empirical analysis concluded that the MIC performed better than the well

known criteria but cautioned that these results were only useful in the context of unit root testing and unsuitable for other scenarios.

In an AR study by Liew (2004), the performances of five IC were investigated. The study considered an $\mathrm{AR}(4)$ process with sample sizes $T=(30, 60, 120, 240, 480, 960)$ and concluded that for lower sample sizes, i.e. $T=30$ and $T=60$, AIC and FPE outperformed the other criteria evaluated. For larger samples, HQIC was the best performer and that overfitting the model was negligible with AIC and FPE having the least probability of under estimation amongst all criteria compared. The limitations of this study were that the conclusions were based on comparisons generated from $1\,000$ replications per model and the frequency of correct decisions for the IC were similar, for example despite the claim that AIC outperformed the other criteria at smaller *T*, the frequency of correct decisions for all IC fell between 53% and 57%. This range was arguably not big enough to justify fully a criterion preference.

Ng and Perron (2005) augmented the 2001 study with an extensive simulation exercise to compare stationary $\mathrm{ARMA}(p, q)$ and $\mathrm{ARCH}(p)$ processes. The study analysed the performances of the Akaike and the Schwarz criteria to assess how fixing the number of observations influenced their ability. The study used 10 variants of the IC published in different sources. As an example, the penalty function of the AIC in some texts was defined in relation to the total sample size (N), whilst in other sources it was defined in relation to the size of the data used for the fitted model (T). Ng and Perron's (2005) study was comprehensive, $5\,000$ replications were simulated per model for sample sizes $T=(100, 250, 500, 1000)$ and twelve AR processes of lag lengths $p=(0, 1, 2, 3, 4, 8)$, eight truncated $\mathrm{ARMA}(1,1)$ processes and five ARCH processes were assessed. They concluded that the AIC overfitted low order AR models whilst the BIC when compared to AIC underfitted higher order models. In conclusion, researchers where cautioned against using IC unless the criteria was defined for the model evaluated. They concluded that the formulation of the criterion affected the IC performance and urged that the effective number of observations be fixed when comparing models. This current study accepts this

recommendation and fixes the effective number of observations (T) whilst evaluating the simulation datasets.

Hafidi and Mkhadri (2006) considered a bias correction for AIC in univariate AR, linear regression and multivariate linear regression models. The criterion defined as a biased corrected AIC (KIC) was examined in a simulation exercise and the results showed that KIC outperformed AIC, BIC and HQIC. The simulation exercise by Hafidi and Mkhadri (2006) was extensive and included linear regression models with $T = 20 \text{ and } 30$, multivariate regression models with $T = 20 \text{ and } 35$ and $\text{AR}(2)$ models with $T = 23 \text{ and } 35$.

It is also worth noting that model selection is not the only area of recent research for VAR modelling. There have been recent publications of alternative research investigations undertaken, included is a review of three recent publications and the context of the study.

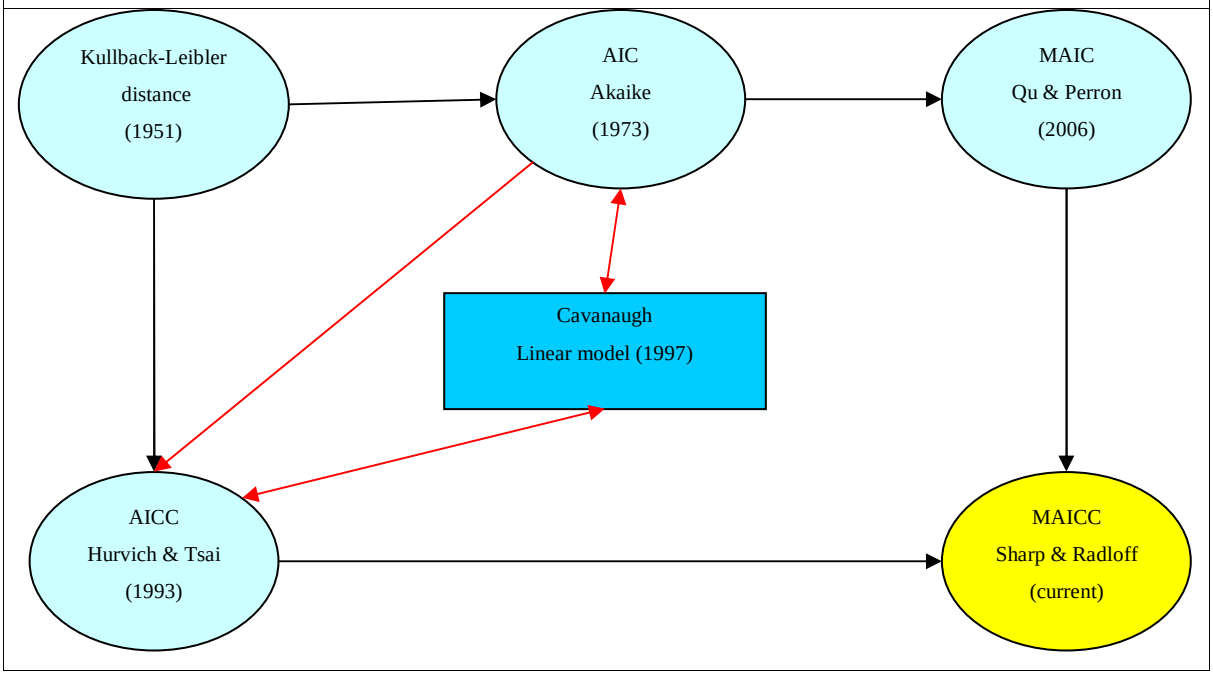
Residual analysis is a diagnostic tool analysts use to check the assumptions of the model. In a VEC framework Brüggemann *et al.* (2006) derive the asymptotic distribution of a Lagrange multiplier (LM) residual autocorrelation test. A simulation study with a $3d-\text{VEC}$ model with one cointegration relationship was used to illustrate the effectiveness of the LM test as compared to the usual portmanteau test. Baltagi and Wang (2007) revisit the proposal of an IC approach for the testing of cointegration rank. Four criteria, AIC, BIC, HQIC and a posterior information criterion (PIC) were used in the assessment of 165 datasets. Although far from conclusive, the results of the criteria based method were promising when compared to the Johansen trace statistic. Agreements in cointegration specifications between IC and the trace statistic ranged between $23\% \text{ and } 71\%$.

Typically bootstrap methods are used to find interval estimates for test statistics, however Demiralp, Hoover and Perez (2008) evaluated a bootstrap method for identifying a VAR. A simulation exercise was used to illustrate the successfulness of the method. This brief review of recent publications illustrates opportunities available for research in a VAR

equation system.

In summary, the majority of studies have considered model selection for stationary VAR systems. Limited work on VEC modelling has been done by Brüggemann (2004) who considers model reduction techniques using sequential stepwise routines. The reference to limited work is based on the reason that the Brüggemann (2004) work is primarily targeting stationary VAR models with limited extension to the VEC framework. The most relevant and current research is the work of Qu and Perron (2006) who consider the asymptotic properties of a modified AIC which includes a penalty term for the error correction restriction. This study extends on their work whilst simultaneously incorporating simulation design concepts from Koreisha and Pukkila (1993) and Hurvich and Tsai (1993).

Figure 2.8: Flowchart of IC developments



This thesis considers IC model selection methods for nonstationary cointegrated models defined in a VEC framework. This research study extends the research of Hurvich and Tsai (1993) and Qu and Perron (2006) in that the proposed criterion for VEC models is a derivative of the two methodologies. The flowchart in Figure 2.8 provides a schematic of

where these criteria fit in relation to a time frame. The yellow oval represents the criterion which is a combination of Hurvich and Tsai's (1993) small sample bias corrected criterion and Qu and Perron's (2006) modified criterion.

2.8 Information Criteria Used in Present Study

The thirteen criteria defined in Table 2.1 will now be discussed. This section provides the details of the criteria with selected derivations in a VEC system of equations. Notation is consistent with notation used previously in this study with the *k-dimensional* $\mathrm{VAR}(p)$ model defined as $\mathbf{x_t} = \mathbf{A_1 x_{t-1}} + \mathbf{A_2 x_{t-2}} + \cdots + \mathbf{A_p x_{t-p}} + \boldsymbol{\varepsilon}_\mathbf{t}$, where the vector $\mathbf{x_t}$ denotes a kd-$\mathrm{VAR}(p)$ process with zero mean. The notation is simplified by writing the model as $\mathbf{x_t} = \sum_{i=1}^{p} \mathbf{A_i x_{t-i}} + \boldsymbol{\varepsilon}_\mathbf{t}$. When cointegration exists within the system of equations, this model is a kd-$\mathrm{VEC}(p-1)$ process given in ECM form as $\boldsymbol{\Delta x_t} = \boldsymbol{\Pi x_{t-1}} - \sum_{i=2}^{p} \boldsymbol{\Gamma}_\mathbf{i} \boldsymbol{\Delta x_{t-i+1}} + \boldsymbol{\varepsilon}_\mathbf{t}$.

Suppose the VAR process $\mathbf{x_t^t} = (\mathbf{x_{1t}}, \mathbf{x_{2t}}, \cdots, \mathbf{x_{kt}})$ generates a set of data with *N* observations for each of the *k* variables of the model. The *N* observations may be written as a $(k\mathrm{x}N)$ observation matrix which is a function of the $(k\mathrm{x}k)$ parameter matrices, $\mathbf{A_i}$ and their respective $(k\mathrm{x}N)$ lag matrices $\mathbf{x_{t-i}}$ for $i = 1, 2, ..., p$. The *k* first order lag variables are represented as $\mathbf{x_{t-1}^t} = (\mathbf{x_{1t-1}}, \mathbf{x_{2t-1}}, \cdots, \mathbf{x_{kt-1}})$, with the *k* *i*th order lag variable as $\mathbf{x_{t-i}^t} = (\mathbf{x_{1t-i}}, \mathbf{x_{2t-i}}, \cdots, \mathbf{x_{kt-i}})$ and the *k* *p*th order lag variables as $\mathbf{x_{t-p}^t} = (\mathbf{x_{1t-p}}, \mathbf{x_{2t-p}}, \cdots, \mathbf{x_{kt-p}})$. Given that a process $\mathbf{x_t}$ generates a set of data with *N* observations starting from period 1 to period *N*, the process $\mathbf{x_{t-1}}$ generates a set of *N* observations starting from period 0 to period $N-1$ and similarly the process $\mathbf{x_{t-p}}$ generates a set of *N* observations starting from period $-p+1$ to period $T = N - p$.

Hurvich and Tsai (1993) assumed that all pre-sample values are zero, a simplifying

assumption first suggested by Akaike (1974). In practice, this assumption is easy to implement as the pre-sample values necessary for the *p* lagged terms are usually obtained by omitting the first *p* observations from x_{it} in order to obtain *T* numerical values for the x_{it-p} observations.

The derivations that follow use the notation that all IC are defined with *T* as the fixed number of observations, *k* as the number of variables and *p* as the number of lagged dependent terms in the VAR model. A strong assumption for all criteria derived is that the true unknown model is nested within the set of approximating models. This assumption means that the true model has a lag length (p_0) no greater than the maximum lag length (p) of the approximating models, i.e. that $p \geq p_0$. The criteria are used as a measure of the goodness of fit of the model similar to the well known χ^2 goodness of fit methods used for contingency tables. The unbiased estimated residual covariance matrix is defined as $\hat{\Sigma}$ which is adjusted for degrees of freedom.

Criterion 1: Akaike's Information Criterion (AIC)

Akaike uses a second order Taylor expansion as an approximately unbiased estimator of the Kullback-Leibler distance (Hurvich & Tsai, 1993). The efficiency methodology proposed by Akaike (1974) was used for selecting the true model from several competing models by selecting the model which minimises the estimated criterion value. A useful derivation of AIC is given by Burnham and Anderson (2002: 362 - 368) but we follow the derivation of Qu and Perron (2006) which is specific to VEC models whilst simultaneously drawing on theoretical comments from Cavanaugh (1997). In most cases the derivation differs from Qu and Perron (2006) only in notation to ensure continuity of notation with this study. The notational differences are noticeable in that Qu and Perron (2006) defined a $\mathrm{VEC}(p)$ model rather than constructing the ECM from a $\mathrm{VAR}(p)$.

This proof considers the $\mathrm{VAR}(p)$ model as the $\mathrm{VEC}(p-1)$ model,

$\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{i=2}^{p} \Gamma_i \Delta x_{t-i+1} + \varepsilon_t}$. Qu and Perron (2006) reparameterise the model as the multivariate linear model and then derive the expected Kullback-Leibler distance of the true unknown model.

Derivation of AIC

Step 1: Define the linear transformation

Let $\mathbf{w_t} = \left(\mathbf{\Delta x_{t-1}^t, \Delta x_{t-2}^t, \cdots, \Delta x_{t-p+1}^t}\right)^{\mathbf{t}}$ and $\mathbf{-\Gamma} = \left(\mathbf{\Gamma_2, \Gamma_3, \cdots, \Gamma_p}\right)$ where $\mathbf{w_t}$ is a $\left(k\left(p-1\right)\text{x}1\right)$ vector and $\mathbf{\Gamma}$ is a $\left(k\text{x}k\left(p-1\right)\right)$ matrix, then the $\text{VEC}\left(p-1\right)$ is equivalent to $\mathbf{\Delta x_t = \Pi x_{t-1} + \Gamma w_t + \varepsilon_t}$ where $\mathbf{x_{t-1}}$ is the nonstationary I(1) process, and both $\mathbf{\Delta x_t}$ and $\mathbf{w_t}$ are stationary I(0) processes. Qu and Perron (2006) projected $\mathbf{x_{t-1}}$ onto the range space of $\mathbf{w_t}$ by defining the terms $\mathbf{x_{t-1}^*}$ and $\mathbf{\Gamma^*}$ as $\mathbf{x_{t-1}^* = x_{t-1} - \sum_{s=2}^{T} x_{s-1} w_s^t \left(\sum_{s=2}^{T} w_s w_s^t\right)^{-1} w_t}$ and $\mathbf{\Gamma^* = \Gamma + \Pi \sum_{t=2}^{T} x_{t-1} w_t^t \left(\sum_{t=2}^{T} w_t w_t^t\right)^{-1}}$. The $\text{VEC}\left(p-1\right)$ model is simplified with orthogonal regressors $\mathbf{x_{t-1}^*}$ and $\mathbf{w_t}$ as $\mathbf{\Delta x_t = \Pi x_{t-1}^* + \Gamma^* w_t + \varepsilon_t}$. The dimensions of the matrix $\mathbf{\Gamma^*}$ are $\left(k\text{x}k\left(p-1\right)\right)$ whilst the dimensions of the vector $\mathbf{x_{t-1}^*}$ are $\left(k\text{x}1\right)$.

Reparameterising the $\text{VEC}\left(p-1\right)$ model with $\mathbf{\beta} = \left(\mathbf{\Pi, \Gamma^*}\right)$, $\mathbf{\theta} = \left(\mathbf{\beta, \Sigma}\right)$ and $\mathbf{z_{t-1}} = \left(\left(\mathbf{x_{t-1}^*}\right)^{\mathbf{t}}, \mathbf{w_t^t}\right)^{\mathbf{t}}$ gives $\mathbf{\Delta x_{t-1} = \beta z_{t-1} + \varepsilon_t}$. The dimension of the parameter matrix $\mathbf{\beta}$ is $\left(k\text{x}k\left(p-1\right)\right)$ whilst the dimension of the data vector $\mathbf{z_{t-1}}$ is $\left(k\left(p-1\right)\text{x}1\right)$. Given this multivariate linear model the log likelihood function of the $\text{VEC}\left(p-1\right)$ model is $\ln L\left(\mathbf{\theta; x}\right) = -\frac{Tk}{2}\ln\left(2\pi\right) - \frac{T}{2}\ln\left|\mathbf{\Sigma}\right| - \frac{1}{2}\sum_{t=2}^{T}\left(\mathbf{\Delta x_t - \beta z_{t-1}}\right)^{\mathbf{t}} \mathbf{\Sigma^{-1}} \left(\mathbf{\Delta x_t - \beta z_{t-1}}\right)$ where $\mathbf{\theta} = \left(\mathbf{\beta, \Sigma}\right)$.

The model selection choice is based on the distance between two models, the true unknown model and the approximating model, the smaller the distance, the better the model fit, a perfect fit should yield a distance of zero. This distance, referred to as the Kullback-Leibler distance, is defined as $KLd = E_0\left\{l(\boldsymbol{\theta_0};\mathbf{x}) - l(\boldsymbol{\theta};\mathbf{x})\right\}$ where $l(\boldsymbol{\theta_0};\mathbf{x})$ is the likelihood function of the true unknown model, $l(\boldsymbol{\theta};\mathbf{x})$ is the likelihood function of the approximating model and E_0 is the expectation of the function taken with respect to the true unknown model. The first term is a constant for all model comparisons and thus minimising *KLd* is equivalent to maximising $E_0\left\{l(\boldsymbol{\theta};\mathbf{x})\right\}$.

In reality the likelihood of the approximating model is never known but can be estimated for the likelihood function for a set of estimated parameters. The parameters are estimated using maximum likelihood which provides us with a set of estimated likelihood functions $l(\boldsymbol{\theta};\mathbf{x})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. Therefore the model selection choice becomes a problem of comparing the distance between the true unknown model and the estimated approximating model given by $KLd \approx E_0\left\{l(\boldsymbol{\theta_0};\mathbf{x}) - l(\boldsymbol{\theta};\mathbf{x})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\} = E_0\left\{l(\boldsymbol{\theta};\mathbf{x})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\}$. AIC is defined as a quantity, whose expectation is twice the function to be minimised, i.e. $E_0(\text{AIC}) = 2E_0\left[E_0\left\{l(\boldsymbol{\theta};\mathbf{x})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\}\right] + o(1)$. The likelihood function in its natural log form is used to simplify the equations that follow.

Step 2: Derivation of the IC

$$E_0\left[E_0\left\{l(\boldsymbol{\theta};\mathbf{x})\right\}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]$$

$$= \underset{(i)}{E_0\left\{l\left(\hat{\boldsymbol{\theta}};\mathbf{x}\right)\right\}} + \underset{(ii)}{\left[E_0\left\{l(\boldsymbol{\theta_0};\mathbf{x})\right\} - E_0\left\{l\left(\hat{\boldsymbol{\theta}};\mathbf{x}\right)\right\}\right]} + \underset{(iii)}{\left[E_0\left\{E_0\left\{l(\boldsymbol{\theta};\mathbf{x})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\}\right\} - E_0\left\{l(\boldsymbol{\theta_0};\mathbf{x})\right\}\right]}.$$

Now consider the results of a Taylor expansion around $\hat{\boldsymbol{\theta}}$ for term (ii) of AIC.

$$E_0\left\{l\left(\boldsymbol{\theta}_0;\mathbf{x}\right)\right\}-E_0\left\{l\left(\hat{\boldsymbol{\theta}};\mathbf{x}\right)\right\}$$

$$=-\frac{1}{2}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\left(\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}^{t}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)^{t}\right)\right\}+o_{p-1}(1/T)+\text{linear term}$$

$$\approx-\frac{1}{2}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)\underbrace{\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)}_{(iv)}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)^{t}\right)\right\}-\frac{1}{2}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)\underbrace{\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)}_{(v)}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)^{t}\right)\right\}.$$

The orthogonality of $\mathbf{w}_t$ and $\mathbf{x}^{*}_{t-1}$ simplifies the expression by ensuring the middle term $\left(\mathbf{w}_t\ \mathbf{x}^{*}_{t-1}\right)$ is zero. The first order linear term is zero and the remainder term is negligible for large *T* allowing for the approximation above. The expectation of term (*iv*) converges to a central chi-squared distributed random variable with $k^2\left(p-1\right)$ degrees of freedom, since $\mathbf{w}_t$ is stationary, i.e. $\lim_{T\to\infty}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)^{t}\right)\right\}=k^2\left(p-1\right)$. We consider term (*v*) later for now consider the expectation of term (*iii*) of the AIC.

Consider the results of a Taylor expansion around $\hat{\boldsymbol{\theta}}$ for term (iii) of AIC.

$$E_0\left\{E_0\left(l\left(\boldsymbol{\theta};\mathbf{x}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)\right\}-E_0\left\{l\left(\boldsymbol{\theta}_0;\mathbf{x}\right)\right\}$$

$$=-\frac{1}{2}E_0\left\{tr\left(\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\left(E_0\left\{\sum_{t=2}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}^{t}\right\}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)^{t}\right)\right\}+o_{p-1}(1/T)+\text{linear term}$$

$$\approx-\frac{1}{2}E_0\left\{tr\left(\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)\underbrace{\left(E_0\left\{\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right\}\right)}_{(vi)}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)^{t}\right)\right\}$$

$$-\frac{1}{2}E_0\left\{tr\left(\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)\underbrace{\left(E_0\left\{\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right\}\right)}_{(vii)}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)^{t}\right)\right\}.$$

Again, the orthogonality of $\mathbf{w}_t$ and $\mathbf{x}^*_{t-1}$ simplifies the expression by ensuring the middle term $\left(\mathbf{w}_t\ \mathbf{x}^*_{t-1}\right)$ is zero. The first order linear term is zero and the remainder term is negligible for large *T* allowing for the approximation above. The expectation of term (*vi*) converges to a central chi-squared distributed random variable with $k^2(p-1)$ degrees of freedom, since $\mathbf{w}_t$ is stationary, i.e.

$$\lim_{T\to\infty} E_0\left\{tr\left(\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\Gamma}}^* - \boldsymbol{\Gamma}_0^*\right)\left(E_0\left\{\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^t\right\}\right)\left(\hat{\boldsymbol{\Gamma}}^* - \boldsymbol{\Gamma}_0^*\right)^t\right)\right\} = k^2(p-1).$$

Now consider terms (*v*) the remainder from term (*ii*) and term (*vii*) the remainder from term (*iii*). Let **A** be the sum of terms given by,

$$\mathbf{A} = tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_0\right)\underbrace{\left(\sum_{t=2}^{T}\mathbf{x}^*_{t-1}\mathbf{x}^{*t}_{t-1}\right)}_{(v)}\left(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_0\right)^t\right)$$
$$+ tr\left(\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_0\right)\underbrace{\left(E_0\left\{\sum_{t=2}^{T}\mathbf{x}^*_{t-1}\mathbf{x}^{*t}_{t-1}\right\}\right)}_{(vii)}\left(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}_0\right)^t\right).$$

The process $\mathbf{x}^*_{t-1}$ is a sum of two components, the nonstationary $\mathbf{x}_{t-1}$ component and the stationary $\mathbf{w}_t$ component. The nonstationary component dominates the stationary component (Engle & Granger, 1991: 6) hence $\mathbf{x}^*_{t-1}$ is considered a nonstationary process. The $\lim_{T\to\infty} T^{-2}\sum_{t=2}^{T}\mathbf{x}^*_{t-1}\mathbf{x}^{*t}_{t-1}$ is a random variable independent of the lag length $(p-1)$ and thus the sum of the expectation of terms (*v*) and (*vii*) is a function of a vector Wiener process and is a constant. Therefore we have $E_0(\mathrm{AIC}) \approx 2E_0\left\{l\left(\hat{\boldsymbol{\theta}};\mathbf{y}\right)\right\} - C - 2k^2(p-1)$ which requires that the quantity to minimise is given as $\mathrm{AIC} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + 2\dfrac{k^2(p-1)}{T}$.

AIC has been used extensively in both simulation and empirical studies. AIC was initially used in single equation regression and time series models but has been extended to multivariate time series models by Lütkepohl (1985, 2005), Brockwell and Davis (2002),

Gonzalo and Pitarakis (2002), Brüggemann (2004), Ng and Perron (2005), Ivanov and Kilian (2005) and Qu and Perron (2006, 2007).

Criterion 2: Qu and Perron's Modified Akaike's Information Criterion (MAIC)

The disadvantage of the derivation of AIC is the asymptotic limit of the $\mathbf{x}_{t-1}^{*}$ nonstationary dominated process. Qu and Perron (2006, 2007) emphasise the finite sample dependency of term **A** (defined in criterion 1) on the lag structure of the model. Rather than accepting the asymptotic limit approximation, they partition the $\mathbf{x}_{t-1}^{*}$ nonstationary dominated process into the stationary and nonstationary components. These components are then approximated separately providing an alternative model selection criterion to the AIC derived as criterion 1.

Qu and Perron (2006, 2007) considered that if evidence of cointegration existed, the VEC model had a restricted structure for the error correction parameter, $\mathbf{\Pi}$. In particular, if the systems of equations were cointegrated, Johansen (1988) had shown that the term $\mathbf{\Pi}$ had reduced rank and could be written as the product of two matrices which were denoted as $\boldsymbol{\alpha}_{SJ}$ and $\boldsymbol{\beta}_{SJ}$. The rank of $\mathbf{\Pi}$ was determined by the number of cointegrating vectors which in turn were determined by the column rank of $\boldsymbol{\beta}_{SJ}$. The column rank of $\boldsymbol{\beta}_{SJ}$ was then defined as *r* cointegrating columns. Using the *r* cointegrating vector restriction, Qu and Perron (2006) derived a new IC called the modified AIC and abbreviated as MAIC.

The derivation given by Qu and Perron (2006) partitions the process $\mathbf{x}_{t-1}^{*}$ into its stationary cointegrated and nonstationary non-cointegrated components. The cointegration rank of $\mathbf{x}_{t-1}^{*}$ was determined by the *r* cointegrating vectors of $\boldsymbol{\beta}_{SJ}$ allowing one to rearrange the terms (v) and (vii) from **A** (of criterion 1) into the sums of stationary and nonstationary components.

Derivation of MAIC

From criterion 1,

$$\mathbf{A} = tr\left(\hat{\mathbf{\Sigma}}^{-1} \left(\hat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right) \underbrace{\left(\sum_{t=2}^{T} \mathbf{x}^*_{t-1} \mathbf{x}^{*t}_{t-1} \right)}_{(v)} \left(\hat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right)^t \right)$$
$$+ tr\left(\mathbf{\Sigma}_0^{-1} \left(\hat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right) \underbrace{\left(E_0 \left\{ \sum_{t=2}^{T} \mathbf{x}^*_{t-1} \mathbf{x}^{*t}_{t-1} \right\} \right)}_{(vii)} \left(\hat{\mathbf{\Pi}} - \mathbf{\Pi}_0 \right)^t \right).$$

There exists an invertible matrix $\mathbf{P}$, with $\mathbf{Q} = \mathbf{P}^{-1}$ such that $\mathbf{Q}\left(\mathbf{I_k} + \mathbf{\Pi}\right)\mathbf{P} = diag\left[\mathbf{I_{k-r}}, \mathbf{\Lambda_r}\right] = \mathbf{J}$. Thus $\mathbf{\Pi} = \mathbf{P}\left(\mathbf{J} - \mathbf{I_n}\right)\mathbf{Q} = \mathbf{P_2}\left(\mathbf{\Lambda_r} - \mathbf{I_r}\right)\mathbf{Q_2^t}$ where $\mathbf{P} = \left[\mathbf{P_1}, \mathbf{P_2}\right]$ and $\mathbf{Q^t} = \left[\mathbf{Q_1}, \mathbf{Q_2}\right]$ with $\mathbf{P_1}$ and $\mathbf{Q_1}$ the $\left(k \mathrm{x} \left(k - r\right)\right)$ matrices, so that $\mathbf{v}^*_{1t-1} = \mathbf{Q}_1^t \mathbf{x}^*_{t-1}$ is nonstationary and non-cointegrated and $\mathbf{v}^*_{2t-1} = \mathbf{Q}_2^t \mathbf{x}^*_{t-1}$ is stationary. This means that $\mathbf{\Pi P_1} = \mathbf{0}$, and $\mathbf{\beta}^t_{SJ}\mathbf{P_1} = \mathbf{0}$ for $\mathbf{\Pi} = \mathbf{\alpha}_{SJ}\mathbf{\beta}^t_{SJ}$ where $\mathbf{\alpha}_{SJ}$ has full rank.

Defining a scaling matrix, $\mathbf{D} = \begin{bmatrix} \frac{\mathbf{I_{k-r}}}{T} & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{I_r}}{T^{1/2}} \end{bmatrix}$ with the first block corresponding to $\mathbf{v}^*_{1t-1} = \mathbf{Q}_1^t \mathbf{x}^*_{t-1}$ and the second to $\mathbf{v}^*_{2t-1} = \mathbf{Q}_2^t \mathbf{x}^*_{t-1}$, then partitioning term (v) into component systems gives

$$tr\left(\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)^{t}\right)$$

$$=tr\left(\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)\mathbf{P}\mathbf{D}^{-1}\left(\mathbf{D}\sum_{t=2}^{T}\mathbf{Q}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\mathbf{Q}^{t}\mathbf{D}\right)\mathbf{D}^{-1}\mathbf{P}^{t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)^{t}\right)$$

$$=tr\left(\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)\mathbf{P}_1\underbrace{\left(\sum_{t=2}^{T}\mathbf{v}^{*}_{1,t-1}\mathbf{v}^{*t}_{1,t-1}\right)}_{(viii)}\mathbf{P}_1^{t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)^{t}\right)$$

$$+tr\left(\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)\mathbf{P}_2\underbrace{\left(\sum_{t=2}^{T}\mathbf{v}^{*}_{2,t-1}\mathbf{v}^{*t}_{2,t-1}\right)}_{(ix)}\mathbf{P}_2^{t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)^{t}\right)+o_{p-1}(v).$$

This equation has three components, the $(k-r)$ nonstationary non-cointegrated components of term $(viii)$, the *r* stationary components of term (ix) and the orthogonal $(k-r)$ nonstationary and the *r* stationary components of term $o_{p-1}(v)$. The *r* stationary components of term (ix) converges asymptotically to a chi-squared distribution with (rk) degrees of freedom, i.e.

$$\lim_{T\to\infty}\left\{tr\left(\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)\mathbf{P}_2\left(\sum_{t=2}^{T}\mathbf{v}^{*}_{2,t-1}\mathbf{v}^{*t}_{2,t-1}\right)\mathbf{P}_2^{t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi}_0\right)^{t}\right)\right\}\sim\chi^2_{rk}\text{, i.e. }E_0\left(\chi^2_{rk}\right)=rk\,.$$

Given that term (ix) is independent of the $(p-1)$ lag structure of the model, this term is a constant in the IC and can be ignored from further calculations. The term $o_{p-1}(v)$ is orthogonal and can also be omitted. Now consider partitioning the term (vii) from **A** into the three component system given by

$$tr\left(\mathbf{\Sigma_0^{-1}}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)\left(E_0\left\{\sum_{t=2}^{T}\mathbf{x^*_{t-1}x^{*t}_{t-1}}\right\}\right)\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)^{\mathbf{t}}\right)$$

$$=tr\left(\mathbf{\Sigma_0^{-1}}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)\mathbf{P}\mathbf{D}^{-1}\left(\mathbf{D}E_0\left\{\sum_{t=2}^{T}\mathbf{Q}\mathbf{x^*_{t-1}x^{*t}_{t-1}}\mathbf{Q^t}\right\}\mathbf{D}\right)\mathbf{D}^{-1}\mathbf{P^t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)^{\mathbf{t}}\right)$$

$$=tr\left(\mathbf{\Sigma_0^{-1}}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)\mathbf{P_1}\left(\underbrace{E_0\sum_{t=2}^{T}\mathbf{v^*_{1,t-1}v^{*t}_{1,t-1}}}_{(x)}\right)\mathbf{P_1^t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)^{\mathbf{t}}\right)$$

$$+tr\left(\mathbf{\Sigma_0^{-1}}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)\mathbf{P_2}\left(\underbrace{E_0\sum_{t=2}^{T}\mathbf{v^*_{2,t-1}v^{*t}_{2,t-1}}}_{(xi)}\right)\mathbf{P_2^t}\left(\hat{\mathbf{\Pi}}-\mathbf{\Pi_0}\right)^{\mathbf{t}}\right)+o_{p-1}(vii).$$

Similarly to the previous components system, the orthogonal expectation term $o_{p-1}(vii)$ is zero and can be omitted. Term (xi) converges asymptotically to a chi-squared distribution with (rk) degrees of freedom leaving term (x) for further consideration. Qu and Perron (2006) imposed the cointegrated restriction on the $\mathrm{VEC}(p-1)$ model and replaced the unknown parameters, $\mathbf{\Pi_0}=\boldsymbol{\alpha_0}\boldsymbol{\beta_0^t}$ and $\mathbf{\Sigma_0}$ with their respective maximum likelihood estimators (MLE). Grouping the remaining terms from **A**, gives

$\mathbf{A}=tr\left(2\hat{\mathbf{\Sigma}}^{-1}\left(\hat{\mathbf{\Pi}}-\tilde{\mathbf{\Pi}}_{\mathbf{0}}\right)\mathbf{P_1}\left\{\sum_{t=2}^{T}\mathbf{v^*_{1,t-1}v^{*t}_{1,t-1}}\right\}\mathbf{P_1^t}\left(\hat{\mathbf{\Pi}}-\tilde{\mathbf{\Pi}}_{\mathbf{0}}\right)^{\mathbf{t}}\right)$ providing $p\geq p_0$.

The term in **A** is now asymptotically equivalent to twice the likelihood ratio test of Johansen (1991) used for testing the null hypothesis of *r* cointegrating vectors. The LR test is given by $\tau_{p-1}(r)=-T\sum_{j=r+1}^{k}\ln(1-\hat{\lambda}_j)$. Replacing the constant term in AIC with this additional penalty term provides the modified AIC proposed by Qu and Perron (2006, 2007). The IC of the $\mathrm{VEC}(p-1)$ model is given as

$$\mathrm{MAIC}^{\text{QP-VEC}(p\text{-}1)}=\ln\left|\hat{\Sigma}\right|+\frac{2\left[\tau_{p-1}(r)+k^2(p-1)\right]}{T}.$$

To determine the lag order of a $\mathrm{VEC}(p-1)$ model Qu and Perron (2006) proposed the use of the rank restricted MAIC in lieu of the traditional AIC.

Criterion 3: Hurvich and Tsai's Corrected Akaike's Information Criteria (AICC)

One of the lesser known IC and a noticeable omission from many simulation studies, is the small sample bias corrected IC derived by Hurvich and Tsai (1993). The IC is derived from a $\mathrm{VAR}(p)$ model and abbreviated as AICC. The unbiased corrected IC is an extension of the univariate equivalent for ARMA models given by Hurvich and Tsai (1989). The IC proof in this study is adjusted for assessment as a $\mathrm{VEC}(p-1)$ model. The IC was recommended by Burnham and Anderson (2002) when the sample size is small, they suggested the use of AICC when the ratio $T/k^2 p < 40$. Additional references to Hurvich and Tsai's AICC for a $\mathrm{VAR}(p)$ model included the studies by Cavanaugh (1997), Kilian (2001) and the comments by Burnham and Anderson (2002: 425).

The presentation of the IC given here follows the derivation by Hurvich and Tsai (1993) whilst notation is consistent with that used elsewhere in this study. The derivation begins by considering the $\mathrm{VAR}(p)$ process as two models, one a true model using the notation, $\mathrm{VAR}(p_o)$ and the other, a set of approximating models $\mathrm{VAR}(p)$. Given the case with T observations x_{it} where $i = 1, 2, ..., k$ and $t = 1, 2, ..., T$, the data matrix of the true model is given by $\mathbf{Y} = \mathbf{X_0}\boldsymbol{\beta_0} + \mathbf{u}$ whilst the data matrix of the approximating model is given by $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$. Let the dimension of $\mathbf{Y}$ be the $(T\mathrm{x}k)$ matrix $\mathbf{x_t^t}$, $\mathbf{X_0}$ be the $(T\mathrm{x}kp_o)$ matrix encapsulating the true models lag terms, $\mathbf{X}$ be the $(T\mathrm{x}kp)$ matrix encapsulating the candidate models lag terms, $\boldsymbol{\beta_0}$ the $(kp_o\mathrm{x}k)$ matrix of parameter terms of the true model, $\boldsymbol{\beta}$ the $(kp\mathrm{x}k)$ matrix of parameter terms of the candidate models and $\mathbf{u}$ and $\mathbf{v}$ the respective

(kxk) covariance matrices $\boldsymbol{\Sigma_0}$ and $\boldsymbol{\Sigma}$.

Using the measure defined previously, $KLd = E_0\{l(\boldsymbol{\theta_0;x}) - l(\boldsymbol{\theta;x})\}$ where $l(\boldsymbol{\theta_0;x})$ is the likelihood function of the true model denoted by Hurvich and Tsai (1993) as $l(\boldsymbol{\beta_0, \Sigma_0; Y})$ and $l(\boldsymbol{\theta;x})$ is the likelihood function of the approximating model denoted by Hurvich and Tsai (1993) as $l(\boldsymbol{\beta, \Sigma; Y})$. Recall, the equations are simplified by using the likelihood function in its natural log form and restated by $KLd \approx E_0\left\{l(\boldsymbol{\beta, \Sigma; Y})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}}\right\}$ with $E_0(\text{AIC}) = 2E_0\left[E_0\left\{l(\boldsymbol{\beta, \Sigma; Y})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}}\right\}\right]$.

The log likelihood function for $\text{VAR}(p)$, the approximating model, is given as $l(\boldsymbol{\beta, \Sigma; Y}) = -\frac{Tk}{2}\ln 2\pi - \frac{T}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\left[\sum(\mathbf{Y - X\beta})^{\mathbf{t}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{Y - X\beta})\right]$. To simplify the likelihood multiply the function by the constant $-\frac{2}{T}$ and for ease of exposition, ignore the constant term, $-\frac{Tk}{2}\ln 2\pi$. The likelihood function is then given by $-\frac{2}{T}l(\boldsymbol{\beta, \Sigma; Y}) = \ln|\boldsymbol{\Sigma}| + \frac{1}{T}\left[\sum(\mathbf{Y - X\beta})^{\mathbf{t}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{Y - X\beta})\right]$.

Resuming the assumption that $p \geq p_o$, it is possible to obtain a matrix $\boldsymbol{\beta}^*$ of dimension $\left((kp_o + (p - p_o)k)\text{x}k\right)$ which nests $\boldsymbol{\beta_0}$ the parameter matrix of the true model. Thereby matrix $\boldsymbol{\beta}^* = \left(\boldsymbol{\beta}_{\mathbf{0}}^{\mathbf{t}}, \mathbf{0}^{\mathbf{t}}\right)^t$ where $\mathbf{0}$ is the $\left(((p - p_0)k)\text{x}k\right)$ matrix of zero elements. Letting $\mathbf{Y = X\beta^* + u}$ allows the re-parameterisation of the likelihood function in terms of $\mathbf{X\beta^*}$ and taking expectations on both sides gives

$$E_0\left[-\frac{2}{T}l(\boldsymbol{\beta, \Sigma; Y})\right] = E_0\left[\ln|\boldsymbol{\Sigma}| + \frac{1}{T}\left[\sum(\mathbf{X\beta^* + u - X\beta})^{\mathbf{t}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{X\beta^* + u - X\beta})\right]\right].$$

Using the knowledge that trace and expectation are transposable operators,

$$E_0\left[-\frac{2}{T}l(\boldsymbol{\beta},\boldsymbol{\Sigma};\mathbf{Y})\right]=\ln|\boldsymbol{\Sigma}|+E_0\frac{1}{T}\left[tr\left(\mathbf{X}\boldsymbol{\beta}^{*}+\mathbf{u}-\mathbf{X}\boldsymbol{\beta}\right)\left(\boldsymbol{\Sigma}^{-1}\right)\left(\mathbf{X}\boldsymbol{\beta}^{*}+\mathbf{u}-\mathbf{X}\boldsymbol{\beta}\right)^{t}\right]$$

$$=\ln|\boldsymbol{\Sigma}|+E_0\frac{1}{T}\left[tr\left(\mathbf{u}\boldsymbol{\Sigma}^{-1}\mathbf{u}^{t}\right)\right]+E_0\frac{1}{T}\left[tr\left(\mathbf{X}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)^{t}\mathbf{X}^{t}\right)\right]$$

$$=\ln|\boldsymbol{\Sigma}|+tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathbf{0}}\right)+\frac{1}{T}tr\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)\right).$$

To estimate this equation, replace all the unknown parameters with their maximum likelihood estimators. The resulting expectation, assuming $p \geq p_0$ is

$$E_0\left[l\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}};\mathbf{Y}\right)\right]=\ln|\hat{\boldsymbol{\Sigma}}|+tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\mathbf{0}}\right)+\frac{1}{T}tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right).$$

Now taking expectations under the true model the IC is given by

$$\mathrm{IC}=E_0\left[E_0\left[l\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}};\mathbf{Y}\right)\right]\right]$$

$$=\underset{(i)}{E_0\left[\ln|\hat{\boldsymbol{\Sigma}}|\right]}+\underset{(ii)}{E_0\left[tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\mathbf{0}}\right)\right]}+\underset{(iii)}{E_0\left[\frac{1}{T}tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right)\right]}.$$

Considering term (ii) of the IC, Hurvich and Tsai (1993) used results from Wei (1990: 354) that

$$vec\left(\hat{\boldsymbol{\beta}}\right)\sim MN\left(vec\left(\boldsymbol{\beta}^{*}\right),\boldsymbol{\Sigma}_{\mathbf{0}}\otimes\left\{\mathbf{X}^{t}\mathbf{X}\right\}^{-1}\right) \text{ and } \hat{\boldsymbol{\Sigma}}\sim W_{(T-kp)}\left(\boldsymbol{\Sigma}_{\mathbf{0}}\right).$$

Then $E_0\left(\hat{\boldsymbol{\Sigma}}^{-1}\right)\approx\frac{T}{\left(T-\{kp+k+1\}\right)}\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}$ and $E_0\left[tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_{\mathbf{0}}\right)\right]\approx\frac{Tk}{(T-(kp+k+1))}$.

Considering term (iii) of the IC, Hurvich and Tsai (1993) showed that

$$\frac{1}{T}E_0\left[tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right)\right]$$

$$=\frac{1}{T}tr\left[E_0\left(\hat{\boldsymbol{\Sigma}}^{-1}\right)E_0\left(\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right)\right]$$

$$\approx\frac{1}{\left(T-\{kp+k+1\}\right)}tr\left[\boldsymbol{\Sigma}_0^{-1}\left(\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right)\right]$$

$$=\frac{1}{\left(T-\{kp+k+1\}\right)}\left[\left(vec\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)^{t}\left\{\boldsymbol{\Sigma}_0^{-1}\otimes E_0\left[\mathbf{X}^{t}\mathbf{X}\right]\right\}vec\left(\boldsymbol{\beta}^{*}-\hat{\boldsymbol{\beta}}\right)\right)\right]$$

$$=\frac{k^2p}{\left(T-\{kp+k+1\}\right)}.$$

Consequently

$$\mathrm{IC}\approx\underset{(i)}{E_0\left[\ln\left|\hat{\boldsymbol{\Sigma}}\right|\right]}+\underset{(ii)}{\frac{kT}{\left(T-\{kp+k+1\}\right)}}+\underset{(iii)}{\frac{k^2p}{\left(T-\{kp+k+1\}\right)}}$$

$$=\ln\left|\hat{\boldsymbol{\Sigma}}\right|+\frac{\left(kT+k^2p\right)}{\left(T-\{kp+k+1\}\right)}.$$

Therefore Hurvich and Tsai (1993) proposed the use of $\mathrm{AICC}^{\text{HT-VAR}(p)}=\ln\left|\hat{\boldsymbol{\Sigma}}\right|+\frac{\left(kT+k^2p\right)}{\left(T-\left(kp+k+1\right)\right)}$ for small sample $\mathrm{VAR}(p)$ processes. Given that there are p lag terms in the VAR representation and $(p-1)$ in VEC representation, the IC used in the analysis for this study is defined as

$$\mathrm{AICC}^{\text{HT-VEC}(p\text{-}1)}=\ln\left|\hat{\boldsymbol{\Sigma}}\right|+\frac{\left(kT+k^2\left(p-1\right)\right)}{\left(T-\left(k\left(p-1\right)+k+1\right)\right)}.$$

Criterion 4: A Modified Corrected Akaike Information Criterion (MAICC)

This study proposes to combine the benefits of the small sample correction of Hurvich and Tsai's (1993) AICC and the cointegrated restricted penalty of Qu and Perron's (2006) MAIC. The derivation that accompanies this section is based on the results of the previous three derivations for AIC, MAIC and AICC.

Consider the results of the Hurvich and Tsai's derivation of the AICC.

$$E_0\left[-\frac{2}{T}l(\boldsymbol{\beta},\boldsymbol{\Sigma};\mathbf{Y})\right]=\ln|\boldsymbol{\Sigma}|+E_0\frac{1}{T}\left[tr\left(\mathbf{X}\boldsymbol{\beta}^{*}+\mathbf{u}-\mathbf{X}\boldsymbol{\beta}\right)\left(\boldsymbol{\Sigma}^{-1}\right)\left(\mathbf{X}\boldsymbol{\beta}^{*}+\mathbf{u}-\mathbf{X}\boldsymbol{\beta}\right)^{t}\right]$$

$$=\ln|\boldsymbol{\Sigma}|+E_0\frac{1}{T}\left[tr\left(\mathbf{u}\boldsymbol{\Sigma}^{-1}\mathbf{u}^{t}\right)\right]+E_0\frac{1}{T}\left[tr\left(\mathbf{X}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)^{t}\mathbf{X}^{t}\right)\right].$$

Using Qu and Perron's (2006) argument of an error correction restriction allowing for the partitioning of the design matrix into two components, one for the stationary (lagged) terms and one for the nonstationary (cointegrated) terms changes the term $E_0\frac{1}{T}\left[tr\left(\mathbf{X}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}\right)^{t}\mathbf{X}^{t}\right)\right]$ into two components.

$$E_0\left[-\frac{2}{T}l(\boldsymbol{\beta},\boldsymbol{\Sigma};\mathbf{Y})\right]=\ln|\boldsymbol{\Sigma}|+E_0\frac{1}{T}\left[tr\left(\mathbf{u}\boldsymbol{\Sigma}^{-1}\mathbf{u}^{t}\right)\right]$$

$$+E_0\frac{1}{T}\left\{tr\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Gamma}^{*}-\boldsymbol{\Gamma}_{\mathbf{0}}^{*}\right)\underbrace{\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)}_{(iv)}\left(\boldsymbol{\Gamma}^{*}-\boldsymbol{\Gamma}_{\mathbf{0}}^{*}\right)^{t}\right)\right\}+\frac{1}{T}E_0\left\{tr\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Pi}-\boldsymbol{\Pi}_{\mathbf{0}}\right)\underbrace{\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)}_{(v)}\left(\boldsymbol{\Pi}-\boldsymbol{\Pi}_{\mathbf{0}}\right)^{t}\right)\right\}.$$

$$=\ln|\boldsymbol{\Sigma}|+tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{\mathbf{0}}\right)+\frac{1}{T}\left\{tr\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Gamma}^{*}-\boldsymbol{\Gamma}_{\mathbf{0}}^{*}\right)E_0\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)\left(\boldsymbol{\Gamma}^{*}-\boldsymbol{\Gamma}_{\mathbf{0}}^{*}\right)^{t}\right)\right\}$$

$$+\frac{1}{T}\left\{tr\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Pi}-\boldsymbol{\Pi}_{\mathbf{0}}\right)E_0\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)\left(\boldsymbol{\Pi}-\boldsymbol{\Pi}_{\mathbf{0}}\right)^{t}\right)\right\}.$$

To estimate this equation, replace all the unknown parameters with their maximum likelihood estimators. The resulting expectation is

$$E_0\left[-\frac{2}{T}l\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}};\mathbf{Y}\right)\right]=\ln\left|\hat{\boldsymbol{\Sigma}}\right|+tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)+\frac{1}{T}\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)\underset{(iv)}{E_0\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)^{t}\right)\right\}$$

$$+\frac{1}{T}\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)\underset{(v)}{E_0\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)^{t}\right)\right\}.$$

Now taking expectations under the true model, the IC is given by

$$\mathrm{IC}=E_0\left[E_0\left[l\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\Sigma}};\mathbf{Y}\right)\right]\right]=\underset{(i)}{E_0\left[\ln\left|\hat{\boldsymbol{\Sigma}}\right|\right]}+\underset{(ii)}{E_0\left[tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)\right]}$$

$$+\frac{1}{T}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)\underset{(iv)}{E_0\left(\sum_{t=2}^{T}\mathbf{w}_t\mathbf{w}_t^{t}\right)}\left(\hat{\boldsymbol{\Gamma}}^{*}-\boldsymbol{\Gamma}_0^{*}\right)^{t}\right)\right\}$$

$$+\frac{1}{T}E_0\left\{tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)\underset{(v)}{E_0\left(\sum_{t=2}^{T}\mathbf{x}^{*}_{t-1}\mathbf{x}^{*t}_{t-1}\right)}\left(\hat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}_0\right)^{t}\right)\right\}.$$

Components of this IC have been considered previously,

The expectation of term (ii) is $E_0\left[tr\left(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\right)\right]\approx\frac{Tk}{\left(T-\left(k\left(p-1\right)+k+1\right)\right)}$.

The expectation of term (iv) is $\frac{k^2\left(p-1\right)}{T}$.

The expectation of term (v) is $\frac{\tau_{p-1}\left(r\right)}{T}$.

Combining all four terms of the IC, gives

$$\mathrm{IC}\approx\ln\left|\hat{\boldsymbol{\Sigma}}\right|+\frac{Tk}{\left(T-\left(k\left(p-1\right)+k+1\right)\right)}+\frac{k^2\left(p-1\right)}{T}+\frac{\tau_{p-1}\left(r\right)}{T}.$$

Therefore this study proposes the corrected small sample modified IC, defined as MAICC, as an alternative selection criterion for $\mathrm{VEC}(p-1)$ models.

$$\mathrm{MAICC}^{\text{SR-VEC}(p\text{-}1)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + \frac{kT}{\left(T-\left(k(p-1)+k+1\right)\right)} + \frac{k^2(p-1)}{T} + \frac{\tau_{p-1}(r)}{T}.$$

The criteria that follow have been defined and used by researchers in other multivariate simulation studies. In the majority of cases the criteria have been stated without proof but have been used to measure IC performance. Like those studies, these IC have been included for comparative reasons despite the lack of provision of a theoretical derivation. In each case, the IC is defined and reference source stated.

Criterion 5: Brockwell and Davis Corrected AIC (AICCBD)

In the study of Hurvich and Tsai (1993) a comparison of AICC was made with the multivariate IC of Brockwell and Davis (1991). No theoretical justification is given in either reference but for comparative interest purposes the IC was used in this study. The Brockwell and Davis IC used by Hurvich and Tsai (1993) follows the definition for a $\mathrm{VAR}(p)$ model and was given as $\mathrm{AICCBD}^{\mathrm{VAR}(p)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + \dfrac{2k^2p+2}{\left(T-\left(k^2p+2\right)\right)}$.

This study uses the Brockwell and Davis IC in a $\mathrm{VEC}(p-1)$ framework and states the definition as $\mathrm{AICCBD}^{\mathrm{VEC}(p\text{-}1)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + \dfrac{2k^2(p-1)+2}{\left(T-\left(k^2(p-1)+2\right)\right)}$.

Criterion 6: Schwarz's Bayesian Information Criterion (BIC)

Schwarz (1978) derived an IC by treating the model selection problem from a Bayesian perspective. The criterion has been applied in empirical studies in both univariate and multivariate time series setting. The AIC and BIC are arguably the most popular of the IC and much of the subsequent developments were based upon the underlying principles of their derivations. In the $\mathrm{VAR}(p)$ modelling framework the IC is defined in Hurvich and Tsai (1993), Brüggemann (2004) and Lütkepohl (2005) as $\mathrm{BIC}^{\mathrm{VAR}(p)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + \frac{k^2 p \ln T}{T}$. This study uses the $\mathrm{VEC}(p-1)$ framework and states the definition as $\mathrm{BIC}^{\mathrm{VEC}(p\text{-}1)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + \frac{k^2 (p-1) \ln T}{T}$.

Criterion 7: Hannan and Quinn's Information Criterion (HQIC)

Although AIC and BIC appear to be the more popular model selection methods used when reporting empirical studies, a third IC, derived by Hannan and Quinn (1979) using "*the law of the iterated logarithm*", is also often reported. In this study, modelling estimation is performed in the Econometric software, EViews 5.1. As an indication of HQIC's importance in empirical assessments, it is worth noting that the outputs of all VEC estimations in EViews 5.1 include an IC results summary for the three criteria AIC, BIC and HQIC. Hannan and Quinn (1979) analysed the performance of HQIC and reported that HQIC outperforms AIC for larger size samples but under parameterises the models, relative to AIC, for smaller sample sizes. In the $\mathrm{VAR}(p)$ modelling framework the IC is defined in Gonzalo and Pitarakis (1998) and Ivanov and Kilian (2005) as $\mathrm{HQIC}^{\mathrm{VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 p \ln\ln T}{T}$. This study uses the $\mathrm{VEC}(p-1)$ framework and states the definition as $\mathrm{HQIC}^{\mathrm{VEC}(p-1)} = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 (p-1) \ln\ln T}{T}$.

Criterion 8: Hannan and Quinn's Small Sample Corrected Information Criterion (HQICC)

An interesting variant of the definition of HQICC was found in McQuarrie and Tsai (1998: 206). The IC, considered as a small sample corrected IC, was stated without a theoretical justification. McQuarrie and Tsai (1998: 206) defined the IC in the $\mathrm{VAR}(p)$ modelling framework as $\mathrm{HQICC}^{\mathrm{VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 p \ln(\ln T)}{T-(kp+k+1)}$. In an effort to be as inclusive as possible but without exaggerating, this study assessed the performance of the HQIC variant using the criterion in the $\mathrm{VEC}(p-1)$ framework. The function used in this study is defined as $\mathrm{HQICC}^{\mathrm{VEC}(p-1)} = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 (p-1) \ln(\ln T)}{T-(k(p-1)+k+1)}$.

Criterion 9: Gonzalo and Pitarakis's Information Criterion (LCIC)

An interesting variation from the theoretical IC derivations was the study of Gonzalo and Pitarakis (1998). The researchers motivated their IC following the arguments of Zhang (1992) that in most cases the penalty term of the IC falls within the interval [1.5, 5.0]. Gonzalo and Pitarakis (1998) suggested the application of a linear function of the penalty term of two well known IC. As a model selection technique the suggestion has merit in that combining the benefits of IC could provide a useful selection strategy. The obvious drawback to the study was the lack of a theoretical foundation for the method and the decision to only consider equally weighted linear combinations of two IC. In the Gonzalo and Pitarakis (1998) study, the researchers used a linear combination of the BIC and HQIC.

Gonzalo and Pitarakis (1998) defined the IC in the $\mathrm{VAR}(p)$ modelling framework as $\mathrm{LCIC}^{\mathrm{VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{k^2 p}{T}\left((\ln T + 2\ln\ln T)/2\right)$. This study uses the IC in the

$\mathrm{VEC}(p-1)$ framework and states the definition as

$$\mathrm{LCIC}^{\mathrm{VEC}(p-1)} = \ln\left|\hat{\Sigma}\right| + \frac{k^2(p-1)}{T}\left((\ln T + 2\ln\ln T)/2\right).$$

Criterion 10: Final Prediction Error (FPE)

The mean square error (MSE) and mean absolute deviation (MAD) methods are often used to compare the prediction capabilities of models. Akaike's (1969) MSE measure lead to the development of the final prediction error (FPE) criterion. Several researchers (see McQuarrie & Tsai, 1998; Liew, 2004) have used FPE in comparative studies with traditional IC. This study complements those studies by including the model selection capabilities of FPE in the $\mathrm{VEC}(p-1)$ framework. McQuarrie and Tsai (1998: 204) defined the IC in the $\mathrm{VAR}(p)$ modelling framework as $\mathrm{FPE}^{\mathrm{VAR}(p)} = \left|\hat{\Sigma}\right|\left(\frac{T+kp}{T-kp}\right)^k$. This study uses the natural log of FPE in the $\mathrm{VEC}(p-1)$ framework and states the definition as

$$\ln\left(\mathrm{FPE}^{\mathrm{VEC}(p-1)}\right) = \ln\left|\hat{\Sigma}\right| + k\ln\left(\frac{T+k(p-1)}{T-k(p-1)}\right).$$

Criterion 11: Shibata's Information Criterion (ShibIC)

In the study by Lütkepohl (1985), one of the lesser known IC developed by Shibata (1980) was assessed. The inclusion of Shibata was motivated not by the expectation that it would provide better results when compared to AIC, rather as it was expected to provide different results to AIC and in particular differences when used in small sample assessments. Given that Lütkepohl's study simulated and assessed stationary VAR processes, the IC was included in this assessment for the cointegrated nonstationary processes.

It is worth noting the comment in Shibata's (1989) paper, that if the objective of a study was to determine the correct model from a class of competing models for which there was no clear dominant model selection procedure, perhaps the most appropriate model selection procedure will be to develop simulation models and generate approximating selection results in a tabulated form. With the computational capabilities of personal computers, this comment may be a reality in the near future. In EViews 5.1, at the click of a button, analysts are already provided with tabulated results of estimated criteria for AIC, BIC and HQIC for some pre-determined lag specification.

Lütkepohl (1985) defined the IC in the $\text{VAR}(p)$ modelling framework as $\text{ShibIC}^{\text{VAR}(p)} = \left|\hat{\Sigma}\right|\left(1+\frac{2(pk+1)}{T}\right)^{k}$. This study uses the natural log of Shibata's IC in the $\text{VEC}(p-1)$ framework and states the definition as

$$\text{ShibIC}^{\text{VEC}(p-1)} = \ln\left|\hat{\Sigma}\right| + k\ln\left(1+\frac{2(k(p-1)+1)}{T}\right).$$

Criterion 12: Modified Schwarz's Bayesian Information Criteria (MBIC)

In the extension of the work by Ng and Perron (2001), two additional IC were defined in the paper by Qu and Perron (2006). Not only did Qu and Perron (2006, 2007) derive MAIC (criterion 2), they derive a class of IC which considers the error correction restriction in the VAR framework. This class of IC included a modified BIC and modified HQIC. Both criteria were assessed in this study and the results were compared to the Qu and Perron study.

Qu and Perron (2007) defined the IC in the $\text{VAR}(p)$ modelling framework as $\text{MBIC}^{\text{QP-VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{\left(\tau_{p-1}(r)+k^{2}p\right)\ln T}{T}$. This study uses the IC in the $\text{VEC}(p-1)$

framework and states the definition as $\text{MBIC}^{\text{QP-VEC}(p\text{-}1)} = \ln\left|\hat{\Sigma}\right| + \frac{\left(\tau_{p-1}(r) + k^2\left(p-1\right)\right)\ln T}{T}$.

Criterion 13: Modified Hannan and Quinn's Information Criteria (MHQIC)

The last modified IC derived by Qu and Perron (2006) was the Hannan-Quinn multivariate criterion. Qu and Perron (2007) defined the IC in the $\text{VAR}\left(p\right)$ modelling framework as $\text{MHQIC}^{\text{QP-VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{2\left(\tau_{p-1}(r) + k^2 p\right)\ln\left(\ln T\right)}{T}$. This study uses the IC in the $\text{VEC}\left(p-1\right)$ framework and states the definition as

$$\text{MHQIC}^{\text{QP-VEC}(p-1)} = \ln\left|\hat{\Sigma}\right| + \frac{2\left(\tau_{p-1}(r) + k^2\left(p-1\right)\right)\ln\left(\ln T\right)}{T}.$$

In summary, this chapter introduced the theory of Engle and Granger's (1987) definition of cointegration and provided the theoretical foundation of cointegration in the multivariate autoregressive framework. Extensive use of simulation modelling is presented and illustrative examples of models explained. This chapter also discussed the objectives of the study and provided motivation based on the contradicting results observed in the literature. The literature review in this chapter covered both national and international publications and for the sake of brevity has been summarised to emphasise the studies objective. In conclusion, this chapter closed with the theoretical foundations of model selection from an informational criterion perspective. The IC assessed in this study are clearly defined and for each, a motivation for their use is provided.

CHAPTER 3

METHODOLOGY OF PRESENT STUDY

3.1 Methodology

Nine simulation models of different lag structure and sample sizes were used in this study. To simplify the analysis all models were restricted to a variable dimension of three $(k=3)$. This study simulated $5\,000$ datasets for each variable which gave a total of $15\,000$ series per model. The simulated data were used to determine the best fitted model. This was done by estimating six VEC models with lag lengths of $0, 1, ..., 5$ for each dataset. For each dataset the best fitted model was identified by determining the minimum criterion estimate of the six VEC models. The model selected for each criterion per dataset was then recorded and the frequency of selections summarised by lag structure. The criterion which selected the correct model most often was then considered the best criterion for the analysis.

To observe the influence of sample size (N) on the ability of the IC to select correctly, this study let $N = 40, 100$ and 200. These sample sizes were categorised into three groups, small samples $(N=40)$, common samples $(N=100)$ and large samples $(N=200)$. The effective fixed sample sizes (T) of 34, 94 and 194 were used to estimate the six VEC models. Sample sizes of approximately 100 were common in the literature and it was decided that knowledge of criteria model selection ability between the range of 40 and 200 would be sufficient for cointegration practitioners. The classification that $(N=200)$ is a large sample is debatable as asymptotically $N \to \infty$ is the ideal scenario. However it is reasoned that in a practical setting, the literature review reveals that sample sizes in many studies are less than 200, hence the classification definition.

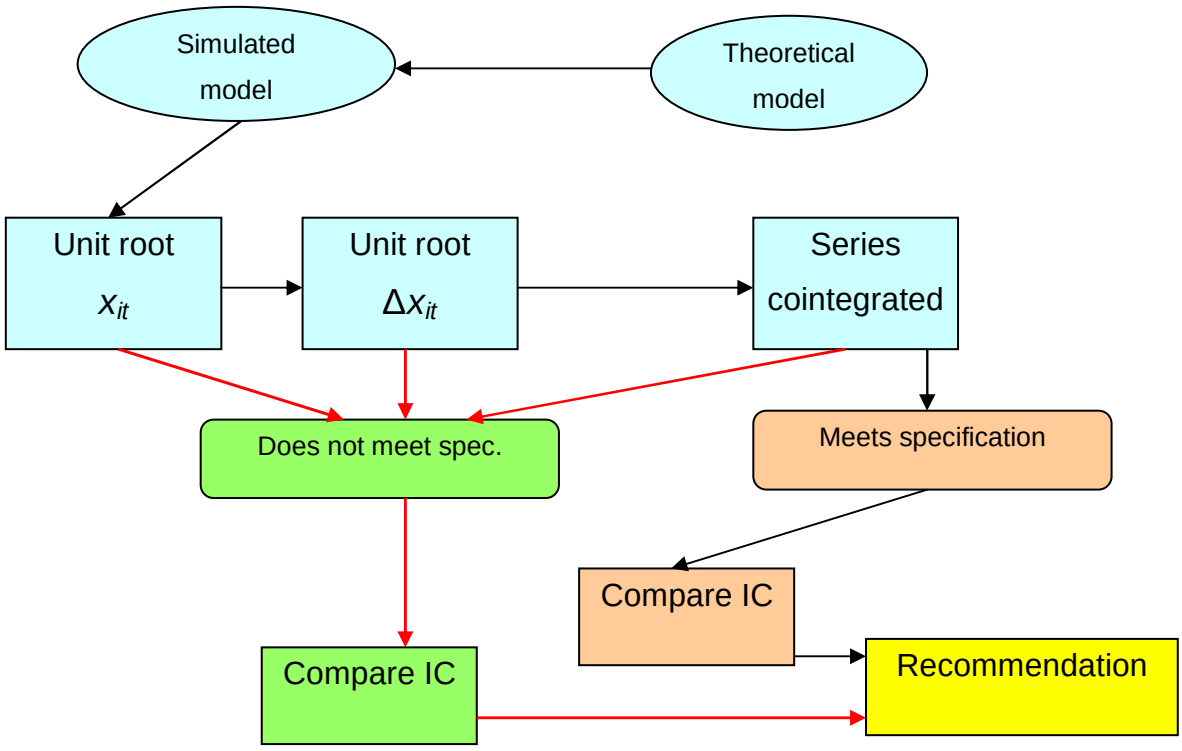
In the Monte Carlo experiment the VAR series were generated by setting the initial value x_{i0} to zero and creating $N+100$ observations. To minimise the effect of the initial condition the first 100 observations were discarded. This approach to simulating data was advocated in the literature, see for example Enders (2004), Liew (2004) and Cheung and Lai (1993). To determine the impact of lag structure, this study constructed models with four different but sequential lag terms. The constructed VAR models for the study had lag lengths $1, 2, 3$ and 4 which were evaluated as VEC models of lag length $0, 1, 2$ and 3, respectively. The choice of coefficients of the VAR model ensured that the series were nonstationary and cointegrated. The lag length structure of the simulation models were restricted to the lag lengths of those found frequently, but not exclusively, in the empirical literature. To compare the influence of parameterisations on criteria performances, two simulation models were compared keeping dimension and sample size fixed. The model used for this comparison was the $3d-\mathrm{VAR}(2)$ model with samples of size $T=94$.

The simulated data were generated from a pre-determined theoretical model. The data obtained were then tested to determine whether or not they satisfied the inferential routines for cointegrated models. The analytical procedure of the statistical routines is shown in Figure 3.1. The inferential analysis followed a two step procedure. Step 1 provided for stationarity testing using the ADF test, whilst step 2 used the Johansen trace statistic to test the cointegrating relationships of the variables.

Data that conformed to the theoretical specifications of the pre-determined model, i.e., that all variables in the model were first order stationary and that the model had one cointegrating relationship were classified as "*meets specification*" (MS) data. In the case where one or more of the theoretical specifications were inferentially insignificant, the data were classified as does "*does not meet specification*" (NMS). The data sets that did not meet specification were analysed separately as the *priori* expectations were that the results from these series would adversely impact on the criterion's ability to select the correct model. This justification seemed reasonable as empirical data that fails the usual inferential analysis would no longer be considered cointegrated and VEC modelling would be terminated. Merely for comparison purposes both the MS and NMS data were analysed

and results reported.

Figure 3.1: Flowchart of analysis procedure



Stepwise routine of data evaluation for MS and NMS data

3.2 Parameterised Simulation Models

Models 1 - 3: Vector Autoregressive Models with One Lag Term

The datasets for three $\mathrm{VAR}(1)$ models were simulated by setting the intercept terms equal to zero. Sample sizes of $N = 40, 100$ and 200 were simulated with the coefficients of the lagged terms chosen to ensure that each single equation was first order difference stationary, the systems of equations were nonstationary and the multi-equation models were cointegrated with one cointegration relationship. Each single equation's error

terms were simulated from a standard normal distribution, with mean equal to zero and variance equal to one.

The $\mathrm{VAR}(1)$ theoretical models were given by

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

These models were represented as ECMs by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix}$ from both sides of the equation:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \Delta x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} - \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = A_0 - (I - A_1) x_{t-1} + \varepsilon_t}$$

$$\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} + \varepsilon_t}$$

where $\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\mathbf{A_1} = \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix}$ and $\mathbf{\Pi = -(I - A_1)}$.

The data were simulated for the $\mathrm{VAR}(1)$ models with the coefficients,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

$$\begin{pmatrix}\Delta x_{1t}\\ \Delta x_{2t}\\ \Delta x_{3t}\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}-\left\{\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\begin{pmatrix}0.4&0.4&0.5\\0.0&1.0&0.0\\0.0&0.0&1.0\end{pmatrix}\right\}\begin{pmatrix}x_{1t\text{-}1}\\x_{2t\text{-}1}\\x_{3t\text{-}1}\end{pmatrix}+\begin{pmatrix}\varepsilon_{1t}\\\varepsilon_{2t}\\\varepsilon_{3t}\end{pmatrix}$$

$$\begin{pmatrix}\Delta x_{1t}\\ \Delta x_{2t}\\ \Delta x_{3t}\end{pmatrix}=\begin{pmatrix}-0.6&0.4&0.5\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}x_{1t-1}\\x_{2t-1}\\x_{3t-1}\end{pmatrix}+\begin{pmatrix}\varepsilon_{1t}\\\varepsilon_{2t}\\\varepsilon_{3t}\end{pmatrix}.$$

Written in matrix notation, $\mathbf{\Delta x_t = \Pi x_{t\text{-}1} + \varepsilon_t}$, where the rank of $\mathbf{\Pi}$, determined the number of cointegrating relationships, with $\mathbf{\Pi}=\begin{pmatrix}-0.6&0.4&0.5\\0&0&0\\0&0&0\end{pmatrix}$.

To confirm that the pre-determined coefficients of the models satisfied theoretical specifications, i.e. that all variables were nonstationary, differenced variables were stationary, the VAR process was nonstationary and one cointegrating relationship existed within the model, consider the following discussions.

The variables x_{2t} and x_{3t} were random walk processes which were nonstationary and when differenced were stationary. The variable x_{1t}, was a linear combination of the variables $x_{2t\text{-}1}$ and $x_{3t\text{-}1}$, both themselves random walk processes. The random walk $\mathrm{I}(1)$ processes dominated the lower ordered process (Engle & Granger, 1991: 6) ensuring that x_{1t} was also an $\mathrm{I}(1)$ nonstationary process and, when differenced, was stationary.

The $\mathrm{VAR}(1)$ process is a stable process if the eigenvalues of $\mathbf{A_1}$ have modulus less than one. This condition means that the process is stable if the reverse characteristic polynomial has no roots in or on the complex unit circle (Lütkepohl, $2005{:}15-16$). This requirement is equivalent to the $\det(\mathbf{I_k} - \mathbf{A_1}z) \neq 0$ for $|z| \leq 1$. To determine whether or not the process was nonstationary required evaluating the reverse characteristic polynomial of the process. The $\mathrm{VAR}(1)$ process was stationary if the roots of the reverse characteristic polynomial

were all greater than unity. The roots of the reverse characteristic polynomial were obtained by finding the roots for the determinant $L = \det(\mathbf{I}_k - \mathbf{A_1} z)$. The process would be stationary if $|z| > 1$ for all roots.

Consider the determinant, L where

$$L = \det(\mathbf{I}_k - \mathbf{A_1} z)$$

$$= \det\left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} z\right]$$

$$= \det\left[\begin{pmatrix} 1-a_{11}z & -a_{12}z & -a_{13}z \\ 0 & 1-a_{22}z & 0 \\ 0 & 0 & 1-a_{33}z \end{pmatrix}\right]$$

$$= 1-(a_{11}+a_{22}+a_{33})z+(a_{11}a_{22}+a_{11}a_{33}+a_{22}a_{33})z^2-a_{11}a_{22}a_{33}z^3$$

$$= 1-2.4z+1.8z^2-0.4z^3.$$

The roots of this polynomial are $|z_1| = 1.00$, $|z_2| = 1.00$ and $|z_3| = 2.50$. Given that at least one root is not greater than unity, i.e. the first and second roots equal unity, the $\mathrm{VAR}(1)$ process is nonstationary.

Johansen (1995) showed that the rank of $\mathbf{\Pi}$ determined the number of cointegrating relationships. Given that $\mathbf{\Pi} = \begin{pmatrix} -0.6 & 0.4 & 0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, the rank of $\mathbf{\Pi}$ is one. Thus the $\mathrm{VAR}(1)$ model provided for one cointegrating relationship between the variables x_{1t}, x_{2t} and x_{3t}.

To summarise, these checks confirmed that the pre-determined coefficients of the $\mathrm{VAR}(1)$ process satisfied the theoretical specifications for a model defined as a $\mathrm{VEC}(0)$ process

with one cointegrating relationship.

Models 4 - 6: Vector Autoregressive Models with Two Lag Terms

The datasets for three $\mathrm{VAR}(2)$ models were simulated by setting the intercept terms equal to zero. Sample sizes of $N = 40, 100$ and 200 were simulated with the coefficients of the lagged terms chosen to ensure that each single equation was first order difference stationary, the systems of equations were nonstationary and the multi-equation models were cointegrated with one cointegration relationship. Each single equation's error terms were simulated from a standard normal distribution, with mean equal to zero and variance equal to one.

The $\mathrm{VAR}(2)$ theoretical models were given by

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

These models were represented as ECMs by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix}$ from both sides of the equation:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \Delta x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \left\{ \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = A_0 - (I - A_1) x_{t-1} + A_2 x_{t-2} + \varepsilon_t}$$

(iii) adding $\mathbf{A_2 x_{t-1} - A_2 x_{t-1}}$ to the RHS and simplifying:

$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2)x_{t-1} - A_2 x_{t-1} + A_2 x_{t-2} + \varepsilon_t}$$
$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2)x_{t-1} - A_2 \Delta x_{t-1} + \varepsilon_t}$$
$$\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} + \varepsilon_t}$$

$$\text{where } \mathbf{\Pi = -(I - A_1 - A_2)} = -\left\{\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}a_{1,11}&a_{1,12}&a_{1,13}\\a_{1,21}&a_{1,22}&a_{1,23}\\a_{1,31}&a_{1,32}&a_{1,33}\end{pmatrix} - \begin{pmatrix}a_{2,11}&a_{2,12}&a_{2,13}\\a_{2,21}&a_{2,22}&a_{2,23}\\a_{2,31}&a_{2,32}&a_{2,33}\end{pmatrix}\right\}.$$

The data were simulated for the $\mathrm{VAR}(2)$ models with the coefficients,

$$\begin{pmatrix}x_{1t}\\x_{2t}\\x_{3t}\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}0.4&0.4&0.5\\0.0&0.8&0.0\\0.0&0.0&0.6\end{pmatrix}\begin{pmatrix}x_{1t-1}\\x_{2t-1}\\x_{3t-1}\end{pmatrix} + \begin{pmatrix}0.2&-0.2&0.0\\0.0&0.2&0.0\\0.0&0.0&0.4\end{pmatrix}\begin{pmatrix}x_{1t-2}\\x_{2t-2}\\x_{3t-2}\end{pmatrix} + \begin{pmatrix}\varepsilon_{1t}\\\varepsilon_{2t}\\\varepsilon_{3t}\end{pmatrix}.$$

Written in matrix notation, $\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} + \varepsilon_t}$, with

$$\mathbf{\Pi} = -\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}0.4&0.4&0.5\\0&0.8&0\\0&0&0.6\end{pmatrix} - \begin{pmatrix}0.2&-0.2&0\\0&0.2&0\\0&0&0.4\end{pmatrix}\right)$$

$$= \begin{pmatrix}-0.4&0.2&0.5\\0&0&0\\0&0&0\end{pmatrix}.$$

Using the arguments given for the $\mathrm{VAR}(1)$ process, the $\mathrm{VAR}(2)$ models theoretical specifications were confirmed.

The variables x_{2t} and x_{3t} are random walk processes which are nonstationary and when differenced are stationary. The variable x_{1t} is an $\mathrm{I}(1)$ dominated nonstationary process and when differenced is stationary.

The $\mathrm{VAR}(2)$ process is nonstationary if the roots of the reverse characteristic polynomial are all greater than unity (Lütkepohl, $2005{:}16$). The roots of the reverse characteristic

polynomial were obtained by finding the roots for the determinant $L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2)$. The process is stationary if $|z| > 1$ for all roots.

Consider the determinant, L where

$$L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2)$$

$$= \det\left[\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} - \begin{pmatrix}a_{1,11} & a_{1,12} & a_{1,13}\\ a_{1,21} & a_{1,22} & a_{1,23}\\ a_{1,31} & a_{1,32} & a_{1,33}\end{pmatrix}z - \begin{pmatrix}a_{2,11} & a_{2,12} & a_{2,13}\\ a_{2,21} & a_{2,22} & a_{2,23}\\ a_{2,31} & a_{2,32} & a_{2,33}\end{pmatrix}z^2\right]$$

$$= \det\begin{bmatrix}1-a_{1,11}z-a_{2,11}z^2 & -a_{1,12}z-a_{2,12}z^2 & -a_{1,13}z-a_{2,13}z^2\\ -a_{1,21}z-a_{2,21}z^2 & 1-a_{1,22}z-a_{2,22}z^2 & -a_{1,23}z-a_{2,23}z^2\\ -a_{1,31}z-a_{2,31}z^2 & -a_{1,32}z-a_{2,32}z^2 & 1-a_{1,33}z-a_{2,33}z^2\end{bmatrix}$$

$$= \det\begin{bmatrix}1-0.4z-0.2z^2 & -0.4z+0.2z^2 & -0.5z\\ 0 & 1-0.8z-0.2z^2 & 0\\ 0 & 0 & 1-0.6z-0.4z^2\end{bmatrix}$$

$$= 1-1.8z+0.24z^2+0.768z^3-0.072z^4-0.12z^5-0.016z^6.$$

The roots of this polynomial are $|z_1| = 5$, $|z_2| = 3.44949$, $|z_3| = 2.5$, $|z_4| = 1$, $|z_5| = 1$ and $|z_6| = 1.44949$. Given that at least one root is not greater than unity, i.e. the fourth and fifth roots equal unity, the $\mathrm{VAR}(2)$ process is nonstationary.

Given that $\mathbf{\Pi} = \begin{pmatrix}-0.4 & 0.2 & 0.5\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$, the rank of $\mathbf{\Pi}$ is one. Thus the $\mathrm{VAR}(2)$ model provided for one cointegrating relationship between the variables x_{1t}, x_{2t} and x_{3t}. To summarise, the pre-determined coefficients of the $\mathrm{VAR}(2)$ process satisfied the theoretical specifications for a model defined as a $\mathrm{VEC}(1)$ process with one cointegrating relationship.

Model 7: Vector Autoregressive Models with Two Lag Terms and Different Coefficients

The datasets for the alternatively parameterised $\mathrm{VAR}(2)$ model were simulated by setting the intercept terms equal to zero. Fixed sample sizes of $N = 100$ were simulated with the coefficients of the lagged terms chosen to ensure that each single equation was first order difference stationary, the systems of equations were nonstationary and the multi-equation models were cointegrated with one cointegration relationship. Each single equation's error terms were simulated from a standard normal distribution, with mean equal to zero and variance equal to one.

The data were simulated for the $\mathrm{VAR}(2)$ model with the coefficients,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.6 \\ 0.2 & 0.8 & -0.15 \\ 0.0 & 0.0 & 0.6 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 & 0.0 \\ -0.1 & 0.15 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix}.$$

Written in matrix notation, $\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} + e_t}$, with

$$\mathbf{\Pi} = -\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.4 & 0.4 & 0.6 \\ 0.2 & 0.8 & -0.15 \\ 0 & 0 & 0.6 \end{pmatrix} - \begin{pmatrix} 0.2 & -0.2 & 0 \\ -0.1 & 0.15 & 0 \\ 0 & 0 & 0.4 \end{pmatrix} \right)$$

$$= \begin{pmatrix} -0.4 & 0.2 & 0.6 \\ 0.1 & -0.05 & -0.15 \\ 0 & 0 & 0 \end{pmatrix}.$$

Using the previous arguments given for the $\mathrm{VAR}(2)$ process, the models' theoretical specifications were confirmed.

The variables x_{2t} and x_{3t} are random walk processes which are nonstationary and when differenced are stationary. The variable x_{1t} is an $\mathrm{I}(1)$ dominated nonstationary process and when differenced is stationary.

The $\mathrm{VAR}(2)$ process is nonstationary if the roots of the reverse characteristic polynomial are all greater than unity (Lütkepohl, $2005:16$). The roots of the reverse characteristic polynomial were obtained by finding the roots for the determinant $L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2)$. The process is stationary if $|z| > 1$ for all roots.

Consider the determinant, L where

$$L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2)$$

$$= \det\left[\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} - \begin{pmatrix}a_{1,11} & a_{1,12} & a_{1,13}\\ a_{1,21} & a_{1,22} & a_{1,23}\\ a_{1,31} & a_{1,32} & a_{1,33}\end{pmatrix}z - \begin{pmatrix}a_{2,11} & a_{2,12} & a_{2,13}\\ a_{2,21} & a_{2,22} & a_{2,23}\\ a_{2,31} & a_{2,32} & a_{2,33}\end{pmatrix}z^2\right]$$

$$= \det\left[\begin{pmatrix}1-a_{1,11}z-a_{2,11}z^2 & -a_{1,12}z-a_{2,12}z^2 & -a_{1,13}z-a_{2,13}z^2\\ -a_{1,21}z-a_{2,21}z^2 & 1-a_{1,22}z-a_{2,22}z^2 & -a_{1,23}z-a_{2,23}z^2\\ -a_{1,31}z-a_{2,31}z^2 & -a_{1,32}z-a_{2,32}z^2 & 1-a_{1,33}z-a_{2,33}z^2\end{pmatrix}\right]$$

$$= \det\left[\begin{pmatrix}1-0.4z-0.2z^2 & -0.4z+0.2z^2 & -0.6z\\ -0.2z+0.1z^2 & 1-0.8z-0.15z^2 & 0.15z\\ 0 & 0 & 1-0.6z-0.4z^2\end{pmatrix}\right]$$

$$= 1-1.8z+0.21z^2+0.846z^3-0.126z^4-0.126z^5-0.004z^6.$$

The roots for this polynomial are $|z_1| = 30.23$, $|z_2| = 2.50$, $|z_3| = 2.24$, $|z_4| = 1.00$, $|z_5| = 1.00$ and $|z_6| = 1.47$. Given that at least one root is not greater than unity, i.e. the fourth and fifth roots equal unity, the $\mathrm{VAR}(2)$ process was nonstationary.

Given that $\mathbf{\Pi} = \begin{pmatrix} -0.4 & 0.2 & 0.6 \\ 0.1 & -0.05 & -0.15 \\ 0 & 0 & 0 \end{pmatrix}$, the rank of $\mathbf{\Pi}$ was one. Thus the $\mathrm{VAR}(2)$ model provided for one cointegrating relationship between the variables x_{1t}, x_{2t} and x_{3t}. To summarise, the pre-determined coefficients of the $\mathrm{VAR}(2)$ process satisfied the theoretical specifications for a model defined as a $\mathrm{VEC}(1)$ process with one cointegrating relationship.

Model 8: Vector Autoregressive Models with Three Lag Terms

The datasets for the $\mathrm{VAR}(3)$ model were simulated by setting the intercept terms equal to zero. Sample sizes of $N = 100$ were simulated with the coefficients of the lagged terms chosen to ensure that each single equation was first order difference stationary, the systems of equations were nonstationary and the multi-equation models were cointegrated with one cointegration relationship. Each single equation's error terms were simulated from a standard normal distribution, with mean equal to zero and variance equal to one.

The $\mathrm{VAR}(3)$ theoretical model was given by

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{3,11} & a_{3,12} & a_{3,13} \\ a_{3,21} & a_{3,22} & a_{3,23} \\ a_{3,31} & a_{3,32} & a_{3,33} \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

This model can be represented as an ECM by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix}$ from both sides of the equation:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \\ \Delta x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \left\{ \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{3,11} & a_{3,12} & a_{3,13} \\ a_{3,21} & a_{3,22} & a_{3,23} \\ a_{3,31} & a_{3,32} & a_{3,33} \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = A_0 - (I - A_1) x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \varepsilon_t}$$

(iii) adding $\mathbf{A_2 x_{t-1} - A_2 x_{t-1}}$, $\mathbf{A_3 x_{t-1} - A_3 x_{t-1}}$ and $\mathbf{A_3 x_{t-2} - A_3 x_{t-2}}$ to RHS and simplifying:

$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2 - A_3) x_{t-1} - A_2 x_{t-1} + A_2 x_{t-2} - A_3 x_{t-1} + A_3 x_{t-2} - A_3 x_{t-2} + A_3 x_{t-3} + \varepsilon_t}$$

$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2 - A_3) x_{t-1} - A_2 \Delta x_{t-1} - A_3 \Delta x_{t-1} - A_3 \Delta x_{t-2} + \varepsilon_t}$$

$$\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} - A_3 \Delta x_{t-1} - A_3 \Delta x_{t-2} + \varepsilon_t}$$

where $\mathbf{\Pi = -(I - A_1 - A_2 - A_3)}$

$$= -\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} - \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} - \begin{pmatrix} a_{3,11} & a_{3,12} & a_{3,13} \\ a_{3,21} & a_{3,22} & a_{3,23} \\ a_{3,31} & a_{3,32} & a_{3,33} \end{pmatrix} \right\}.$$

The data were simulated for the $\mathrm{VAR}(3)$ model with the coefficients,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 0.6 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix} + \begin{pmatrix} -0.2 & 0.15 & 0.1 \\ 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

Written in matrix notation, $\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} - A_3 \Delta x_{t-1} - A_3 \Delta x_{t-2} + \varepsilon_t}$, with

$$\mathbf{\Pi} = -\left(\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} - \begin{pmatrix}0.4 & 0.4 & 0.5\\ 0.0 & 0.6 & 0.0\\ 0.0 & 0.0 & 0.4\end{pmatrix} - \begin{pmatrix}0.2 & -0.2 & 0.0\\ 0.0 & 0.2 & 0.0\\ 0.0 & 0.0 & 0.3\end{pmatrix} - \begin{pmatrix}-0.2 & 0.15 & 0.1\\ 0.0 & 0.2 & 0.0\\ 0.0 & 0.0 & 0.3\end{pmatrix}\right)$$

$$= \begin{pmatrix}-0.6 & 0.35 & 0.6\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.$$

Using the arguments given for the $\mathrm{VAR}(1)$ process, the $\mathrm{VAR}(3)$ model's theoretical specifications were confirmed.

The variables x_{2t} and x_{3t} are random walk processes which are nonstationary and when differenced are stationary. The variable x_{1t} is an $\mathrm{I}(1)$ dominated nonstationary process and when differenced is stationary.

The $\mathrm{VAR}(3)$ process is nonstationary if the roots of the reverse characteristic polynomial are all greater than unity (Lütkepohl, $2005:16$). The roots of the reverse characteristic polynomial were obtained by finding the roots for the determinant $L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2 - \mathbf{A}_3 z^3)$. The process would be stationary if $|z| > 1$ for all roots.

Consider the determinant, *L* where

$L = \det(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2 - \mathbf{A}_3 z^3)$

$$= \det\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}a_{1,11}&a_{1,12}&a_{1,13}\\a_{1,21}&a_{1,22}&a_{1,23}\\a_{1,31}&a_{1,32}&a_{1,33}\end{pmatrix}z - \begin{pmatrix}a_{2,11}&a_{2,12}&a_{2,13}\\a_{2,21}&a_{2,22}&a_{2,23}\\a_{2,31}&a_{2,32}&a_{2,33}\end{pmatrix}z^2 - \begin{pmatrix}a_{3,11}&a_{3,12}&a_{3,13}\\a_{3,21}&a_{3,22}&a_{3,23}\\a_{3,31}&a_{3,32}&a_{3,33}\end{pmatrix}z^3\right]$$

$$= \det\left[\begin{pmatrix}1-a_{1,11}z-a_{2,11}z^2-a_{3,11}z^3 & -a_{1,12}z-a_{2,12}z^2-a_{3,12}z^3 & -a_{1,13}z-a_{2,13}z^2-a_{3,13}z^3\\ -a_{1,21}z-a_{2,21}z^2-a_{3,21}z^3 & 1-a_{1,22}z-a_{2,22}z^2-a_{3,22}z^3 & -a_{1,23}z-a_{2,23}z^2-a_{3,23}z^3\\ -a_{1,31}z-a_{2,31}z^2-a_{3,31}z^3 & -a_{1,32}z-a_{2,32}z^2-a_{3,32}z^3 & 1-a_{1,33}z-a_{2,33}z^2-a_{3,33}z^3\end{pmatrix}\right]$$

$$= \det\left[\begin{pmatrix}1-0.4z-0.2z^2+0.2z^3 & -0.4z+0.2z^2-0.15z^3 & -0.5z-0.1z^3\\ 0 & 1-0.6z-0.2z^2-0.2z^3 & 0\\ 0 & 0 & 1-0.4z-0.3z^2-0.3z^3\end{pmatrix}\right]$$

$$= 1-1.4z-0.06z^2+0.264z^3+0.268z^4-0.012z^5-0.1z^6+0.016z^7+0.012z^8+0.012z^9.$$

The roots of this polynomial are $|z_1|=1.76$, $|z_2|=2.24$, $|z_3|=2.24$, $|z_4|=1.83$, $|z_5|=1.83$, $|z_6|=1.00$, $|z_7|=1.00$, $|z_8|=1.69$ and $|z_9|=1.69$. Given that at least one root is not greater than unity, i.e. the sixth and seventh roots equal unity, the $\mathrm{VAR}(3)$ process is nonstationary.

Given that $\mathbf{\Pi} = \begin{pmatrix}-0.6&0.35&0.6\\0&0&0\\0&0&0\end{pmatrix}$, the rank of $\mathbf{\Pi}$ was one. Thus the $\mathrm{VAR}(3)$ model provided for one cointegrating relationship between the variables x_{1t}, x_{2t} and x_{3t}. To summarise, the pre-determined coefficients of the $\mathrm{VAR}(3)$ process satisfied the theoretical specifications for a model defined as a $\mathrm{VEC}(2)$ process with one cointegrating relationship.

Model 9: Vector Autoregressive Models with Four Lag Terms

The datasets for the $\mathrm{VAR}(4)$ model were simulated by setting the intercept terms equal to zero. Sample sizes of $N=100$ were simulated with the coefficients of the lagged terms chosen to ensure that each single equation was first order difference stationary, the systems of equations were nonstationary and the multi-equation models were cointegrated with one cointegration relationship. Each single equation's error terms were simulated from a standard normal distribution, with mean equal to zero and variance equal to one.

The $\mathrm{VAR}(4)$ theoretical model was given by

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} a_{10} \\ a_{20} \\ a_{30} \end{pmatrix} + \begin{pmatrix} a_{1,11} & a_{1,12} & a_{1,13} \\ a_{1,21} & a_{1,22} & a_{1,23} \\ a_{1,31} & a_{1,32} & a_{1,33} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} & a_{2,13} \\ a_{2,21} & a_{2,22} & a_{2,23} \\ a_{2,31} & a_{2,32} & a_{2,33} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix}$$

$$+ \begin{pmatrix} a_{3,11} & a_{3,12} & a_{3,13} \\ a_{3,21} & a_{3,22} & a_{3,23} \\ a_{3,31} & a_{3,32} & a_{3,33} \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} a_{4,11} & a_{4,12} & a_{4,13} \\ a_{4,21} & a_{4,22} & a_{4,23} \\ a_{4,31} & a_{4,32} & a_{4,33} \end{pmatrix} \begin{pmatrix} x_{1t-4} \\ x_{2t-4} \\ x_{3t-4} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

This model can be represented as an ECM by

(i) subtracting $\begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix}$ from both sides of the equation:

$$\begin{pmatrix}\Delta x_{1t}\\ \Delta x_{2t}\\ \Delta x_{3t}\end{pmatrix}=\begin{pmatrix}a_{10}\\ a_{20}\\ a_{30}\end{pmatrix}+\left\{\begin{pmatrix}a_{1,11} & a_{1,12} & a_{1,13}\\ a_{1,21} & a_{1,22} & a_{1,23}\\ a_{1,31} & a_{1,32} & a_{1,33}\end{pmatrix}-\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right\}\begin{pmatrix}x_{1t-1}\\ x_{2t-1}\\ x_{3t-1}\end{pmatrix}+\begin{pmatrix}a_{2,11} & a_{2,12} & a_{2,13}\\ a_{2,21} & a_{2,22} & a_{2,23}\\ a_{2,31} & a_{2,32} & a_{2,33}\end{pmatrix}\begin{pmatrix}x_{1t-2}\\ x_{2t-2}\\ x_{3t-2}\end{pmatrix}$$

$$+\begin{pmatrix}a_{3,11} & a_{3,12} & a_{3,13}\\ a_{3,21} & a_{3,22} & a_{3,23}\\ a_{3,31} & a_{3,32} & a_{3,33}\end{pmatrix}\begin{pmatrix}x_{1t-3}\\ x_{2t-3}\\ x_{3t-3}\end{pmatrix}+\begin{pmatrix}a_{4,11} & a_{4,12} & a_{4,13}\\ a_{4,21} & a_{4,22} & a_{4,23}\\ a_{4,31} & a_{4,32} & a_{4,33}\end{pmatrix}\begin{pmatrix}x_{1t-4}\\ x_{2t-4}\\ x_{3t-4}\end{pmatrix}+\begin{pmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\\ \varepsilon_{3t}\end{pmatrix}.$$

(ii) simplifying and writing in matrix notation:

$$\mathbf{\Delta x_t = A_0 - (I - A_1)x_{t-1} + A_2x_{t-2} + A_3x_{t-3} + A_4x_{t-4} + \varepsilon_t}$$

(iii) adding $\mathbf{A_2x_{t-1} - A_2x_{t-1}}$, $\mathbf{A_3x_{t-1} - A_3x_{t-1}}$, $\mathbf{A_3x_{t-2} - A_3x_{t-2}}$, $\mathbf{A_4x_{t-1} - A_4x_{t-1}}$, $\mathbf{A_4x_{t-2} - A_4x_{t-2}}$ and $\mathbf{A_4x_{t-3} - A_4x_{t-3}}$ to the RHS and simplifying:

$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2 - A_3 - A_4)x_{t-1} - A_2x_{t-1} + A_2x_{t-2} - A_3x_{t-1} + A_3x_{t-2} - A_3x_{t-2} + A_3x_{t-3}}$$
$$\mathbf{- A_4x_{t-1} + A_4x_{t-2} - A_4x_{t-2} + A_4x_{t-3} - A_4x_{t-3} + A_4x_{t-4} + \varepsilon_t}$$
$$\mathbf{\Delta x_t = A_0 - (I - A_1 - A_2 - A_3 - A_4)x_{t-1} - A_2\Delta x_{t-1} - A_3\Delta x_{t-1} - A_3\Delta x_{t-2}}$$
$$\mathbf{- A_4\Delta x_{t-1} - A_4\Delta x_{t-2} - A_4\Delta x_{t-3} + \varepsilon_t}$$
$$\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2\Delta x_{t-1} - A_3\Delta x_{t-1} - A_3\Delta x_{t-2} - A_4\Delta x_{t-1} - A_4\Delta x_{t-2} - A_4\Delta x_{t-3} + \varepsilon_t}$$

where $\mathbf{\Pi = -(I - A_1 - A_2 - A_3 - A_4)}$

$$=-\left\{\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}-\begin{pmatrix}a_{1,11} & a_{1,12} & a_{1,13}\\ a_{1,21} & a_{1,22} & a_{1,23}\\ a_{1,31} & a_{1,32} & a_{1,33}\end{pmatrix}-\begin{pmatrix}a_{2,11} & a_{2,12} & a_{2,13}\\ a_{2,21} & a_{2,22} & a_{2,23}\\ a_{2,31} & a_{2,32} & a_{2,33}\end{pmatrix}\right.$$

$$\left.-\begin{pmatrix}a_{3,11} & a_{3,12} & a_{3,13}\\ a_{3,21} & a_{3,22} & a_{3,23}\\ a_{3,31} & a_{3,32} & a_{3,33}\end{pmatrix}-\begin{pmatrix}a_{4,11} & a_{4,12} & a_{4,13}\\ a_{4,21} & a_{4,22} & a_{4,23}\\ a_{4,31} & a_{4,32} & a_{4,33}\end{pmatrix}\right\}.$$

The data were simulated for the $\mathrm{VAR}(4)$ with the coefficients,

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 & 0.1 \\ 0.0 & 0.25 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix}$$

$$+ \begin{pmatrix} -0.15 & 0.15 & 0.2 \\ 0.0 & 0.15 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} 0.1 & -0.1 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} \begin{pmatrix} x_{1t-4} \\ x_{2t-4} \\ x_{3t-4} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

Written in matrix notation,

$\mathbf{\Delta x_t = A_0 + \Pi x_{t-1} - A_2 \Delta x_{t-1} - A_3 \Delta x_{t-1} - A_3 \Delta x_{t-2} - A_4 \Delta x_{t-1} - A_4 \Delta x_{t-2} - A_4 \Delta x_{t-3} + \varepsilon_t}$, with

$$\mathbf{\Pi} = -\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} - \begin{pmatrix} 0.2 & -0.2 & 0.1 \\ 0.0 & 0.25 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \right.$$

$$\left. - \begin{pmatrix} -0.15 & 0.15 & 0.2 \\ 0.0 & 0.15 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} - \begin{pmatrix} 0.1 & -0.1 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} \right\}$$

$$\mathbf{\Pi} = \begin{pmatrix} -0.45 & 0.25 & 0.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Using the arguments given for the $\mathrm{VAR}(1)$ process, the $\mathrm{VAR}(4)$ model's theoretical specifications were confirmed.

The variables x_{2t} and x_{3t} are random walk processes which are nonstationary and when differenced are stationary. The variable x_{1t} is an $\mathrm{I}(1)$ dominated nonstationary process and when differenced is stationary.

The $\mathrm{VAR}(4)$ process is nonstationary if the roots of the reverse characteristic polynomial are all greater than unity (Lütkepohl, $2005:16$). The roots of the reverse characteristic polynomial were obtained by finding the roots for the determinant $L = \det\left(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2 - \mathbf{A}_3z^3 - \mathbf{A}_4z^4\right)$. The process is stationary if $|z| > 1$ for all roots.

Consider the determinant, L where

$$L = \det\left(\mathbf{I}_k - \mathbf{A_1}z - \mathbf{A_2}z^2 - \mathbf{A}_3z^3 - \mathbf{A}_4z^4\right)$$

$$= \det\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}a_{1,11}&a_{1,12}&a_{1,13}\\a_{1,21}&a_{1,22}&a_{1,23}\\a_{1,31}&a_{1,32}&a_{1,33}\end{pmatrix}z - \begin{pmatrix}a_{2,11}&a_{2,12}&a_{2,13}\\a_{2,21}&a_{2,22}&a_{2,23}\\a_{2,31}&a_{2,32}&a_{2,33}\end{pmatrix}z^2\right.$$

$$\left. - \begin{pmatrix}a_{3,11}&a_{3,12}&a_{3,13}\\a_{3,21}&a_{3,22}&a_{3,23}\\a_{3,31}&a_{3,32}&a_{3,33}\end{pmatrix}z^3 - \begin{pmatrix}a_{4,11}&a_{4,12}&a_{4,13}\\a_{4,21}&a_{4,22}&a_{4,23}\\a_{4,31}&a_{4,32}&a_{4,33}\end{pmatrix}z^4\right]$$

$$= \det\left[\begin{pmatrix}1-a_{1,11}z-a_{2,11}z^2-a_{3,11}z^3-a_{4,11}z^4 & \cdots & -a_{1,13}z-a_{2,13}z^2-a_{3,13}z^3-a_{4,11}z^4\\ -a_{1,21}z-a_{2,21}z^2-a_{3,21}z^3-a_{4,21}z^4 & \cdots & -a_{1,23}z-a_{2,23}z^2-a_{3,23}z^3-a_{4,21}z^4\\ -a_{1,31}z-a_{2,31}z^2-a_{3,31}z^3-a_{4,31}z^4 & \cdots & 1-a_{1,33}z-a_{2,33}z^2-a_{3,33}z^3-a_{4,31}z^4\end{pmatrix}\right]$$

$$= \det\left[\begin{pmatrix}1-0.4z-0.2z^2+0.15z^3-0.1z^4 & -0.4z+0.2z^2-0.15z^3+0.1z^4 & -0.5z-0.1z^2-0.2z^3\\ 0 & 1-0.5z-0.25z^2-0.15z^3-0.1z^4 & 0\\ 0 & 0 & 1-0.4z-0.3z^2-0.15z^3-0.15z^4\end{pmatrix}\right]$$

$$= 1-1.3z-0.19z^2+0.42z^3-0.185z^4+0.261z^5+0.0465z^6$$

$$-0.039z^7+0.015625z^8-0.01975z^9-0.006375z^{10}-0.0015z^{11}-0.0015z^{12}.$$

The roots of this polynomial are $|z_1| = 2.25$, $|z_2| = 1.82$, $|z_3| = 1.55$, $|z_4| = 2.11$, $|z_5| = 2.11$,

$|z_6| = 1.92$, $|z_7| = 1.92$, $|z_8| = 2.10$, $|z_9| = 2.10$, $|z_{10}| = 1.00$, $|z_{11}| = 1.00$ and $|z_{12}| = 1.46$. Given that at least one root is not greater than unity, i.e. the tenth and eleventh roots equal unity, the $\mathrm{VAR}(4)$ process is nonstationary.

Given that $\mathbf{\Pi} = \begin{pmatrix} -0.45 & 0.25 & 0.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, the rank of $\mathbf{\Pi}$ was one. Thus the $\mathrm{VAR}(4)$ model provided for one cointegrating relationship between the variables x_{1t}, x_{2t} and x_{3t}. To summarise, the pre-determined coefficients of the $\mathrm{VAR}(4)$ process satisfied the theoretical specifications for a model defined as a $\mathrm{VEC}(3)$ process with one cointegrating relationship.

This chapter included a complete description of the methodology followed in this study. The simulation models were defined and VEC model requirements confirmed. The determinants for each model were solved using Mathematica 6, the code and results are given in Appendix 1. The roots of the polynomials were solved using R2.5.1, the code and results are given in Appendix 2. The rank of $\mathbf{\Pi}$ was determined using EViews 5.1.

CHAPTER 4

RESULTS AND INTERPRETATIONS

4.1 An Illustration of a Single Replication

To illustrate how each model was analysed consider the first replication of model 1, the $\mathrm{VAR}(1)$ model with $N = 40$ observations per series. This example illustrates the methodology and shows the inferential routine that was followed for every replication for all nine models. Much of the analysis was automated using EViews 5.1 and Excel 2000. The program routines are provided in Appendix 3 and included on the accompanying DVD.

Table 4.1 provides an extract of the data for this illustration. The first column shows the time period, *t*, The second to fourth columns show the $N = 40$ data observations for the three variables, x_{11t}, x_{12t} and x_{13t}, of the VAR model. The fifth to the seventh columns show the differenced data, Δx_{11t}, Δx_{12t} and Δx_{13t}, necessary for the VEC models.

t	x_{11t}	x_{12t}	x_{13t}	Δx_{11t}	Δx_{12t}	Δx_{13t}
1	-22.16	-5.89	-24.00	-	-	-
2	-23.71	-6.11	-24.30	-1.55	-0.22	-0.29
3	-24.19	-5.73	-24.27	-0.48	0.38	0.03
:	:	:	:	:	:	:
38	-33.61	-11.94	-32.23	-0.09	-0.22	-1.61
39	-33.61	-12.33	-32.82	0.00	-0.38	-0.59
40	-35.38	-12.13	-34.61	-1.77	0.19	-1.79

Table 4.1: Dataset number one (of 5 000) for $\mathrm{VAR}(1)$ with $N = 40$

Throughout the analysis the notation used to represent a data point observation is x_{ijt}, were i denotes the replication number $(i = 1, 2, ..., 5\,000)$, j denotes the variable number $(j = 1, 2, 3)$ and t denotes the time period $(t = 1, 2, ..., N)$. In all analyses, the value of T is fixed as the number of observations used for estimation.

Six VAR models, with lag lengths 1 to 6 were used for each replication to compare the criteria. That is, for the $\text{VAR}(1)$ model's simulated data of $5\,000$ series; $30\,000$ models were estimated, $5\,000$ for each VEC model with lag lengths 0 to 5 both inclusive. The estimated likelihood functions for each model were used to calculate the respective criterion estimates. Thirteen criteria were compared which required a total of $390\,000$ criterion estimates per model.

The lag structure of the VEC models creates the problem of different sample sizes for the analysis. To compare models, the number of observations for each estimated VEC model must be the same, i.e. fixed (Ng & Perron, 2005). As an example consider the case for comparing the $\text{VEC}(0)$ and $\text{VEC}(5)$ models starting with $N = 40$ observations. The $\text{VEC}(0)$ model only requires one observation less than the available number of observations. This model loses a single observation because of the differenced data. However the $\text{VEC}(5)$ model loses an observation for differencing and five observations for the lag structure, i.e. a total of six observations are lost. To ensure a fixed number of observations per model the datasets for each model were fixed at $T = N - p$. Consider the datasets with $N = 40$, the number of observations were fixed at $T = 40 - 6 = 34$. To estimate the $\text{VEC}(0)$ model, 35 observations from t_6 to t_{40} were used to create 34 differenced data points. To estimate the $\text{VEC}(5)$ model, 40 observations from t_1 to t_{40} were used to create 34 differenced data points. The same data reduction system was used for all the VEC models.

The lag structures for the six estimated models are shown in Table 4.2. The lag structure of the VAR is shown as an ECM, as the model was estimated as a VEC model. The lag

structure of the ECM is the number of differenced lag terms defined as the Δx_{i-j} term with $j = 1, 2, ..., 5$. In Table 4.2 the equivalent representations of the VAR and VEC models are given. The $\mathrm{VAR}(1)$ model or equivalent $\mathrm{VEC}(0)$ model has no Δ terms on the right-hand side (RHS) hence there are zero lag terms in the VEC model. The $\mathrm{VAR}(2)$ model or equivalent $\mathrm{VEC}(1)$ model has one Δx_{i-j} term on the RHS hence there is one lag term in the VEC model. Similar interpretations apply to the remaining VEC models, the number of lag terms are shown in the last column of Table 4.2.

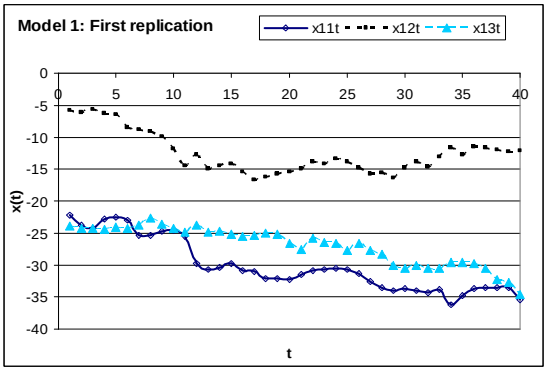
Model	Lag Structure of ECM	No. of Lag Terms
$\mathrm{VAR}(1) \equiv \mathrm{VEC}(0)$	$\mathbf{\Delta x_t = \Pi x_{t-1} + \varepsilon_t}$	0
$\mathrm{VAR}(2) \equiv \mathrm{VEC}(1)$	$\mathbf{\Delta x_t = \Pi x_{t-1} - A_2 \Delta x_{t-1} + \varepsilon_t}$	1
$\mathrm{VAR}(3) \equiv \mathrm{VEC}(2)$	$\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{i=2}^{3} A_i \Delta x_{t-1} - A_3 \Delta x_{t-2} + \varepsilon_t}$	2
$\mathrm{VAR}(4) \equiv \mathrm{VEC}(3)$	$\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{i=2}^{4} A_i \Delta x_{t-1} - \sum_{i=3}^{4} A_i \Delta x_{t-2} - A_4 \Delta x_{t-3} + \varepsilon_t}$	3
$\mathrm{VAR}(5) \equiv \mathrm{VEC}(4)$	$\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{i=2}^{5} A_i \Delta x_{t-1} - \sum_{i=3}^{5} A_i \Delta x_{t-2} - ... - A_5 \Delta x_{t-4} + \varepsilon_t}$	4
$\mathrm{VAR}(6) \equiv \mathrm{VEC}(5)$	$\mathbf{\Delta x_t = \Pi x_{t-1} - \sum_{i=2}^{6} A_i \Delta x_{t-1} - \sum_{i=3}^{6} A_i \Delta x_{t-2} - ... - A_6 \Delta x_{t-5} + \varepsilon_t}$	5

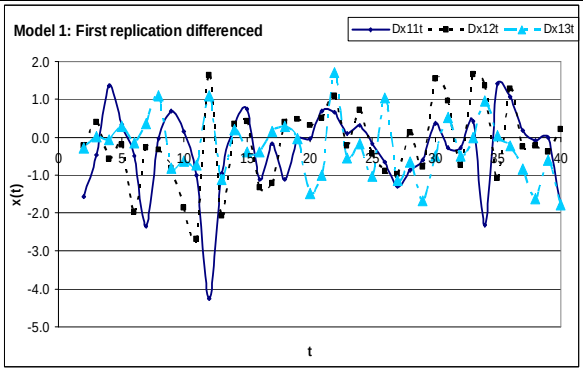
Table 4.2: Error restricted $\mathrm{VAR}(p)$ model represented as $\mathrm{VEC}(p-1)$ model

A preliminary analysis procedure included a graphical check for the stationarity of the data. A plot of the data from Table 4.1 is provided in Figure 4.1. On the left-hand side (LHS) a plot of the series x_{11t}, x_{12t} and x_{13t} is shown, there is some graphical evidence to suspect that the data may be nonstationary. All three series appear to be moving in a decreasing manner. There is some evidence that the series is cointegrated, the deviation between the series appears “constant”, although this visual inspection is questionable.

On the RHS a plot of the differenced series is shown. All three series appear to be fluctuating randomly about a central point, providing some evidence of stationarity. Visual inspections are best used for illustrative purposes, this study followed the inferential routines discussed in the methodology. All inferential methods during these routines used a significance level of 5% unless otherwise stated.

Figure 4.1: Nonstationary and stationary plots of the data from Table 4.1, the first replication for the $\mathrm{VAR}(1)$, $N = 40$ model





$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

The methodology required that each individual series is $\mathrm{I}(1)$ and that the number of cointegrating equations in the three variable model is one. The results of the unit root analysis for the series x_{11t} are shown in Table 4.3. The *p*-value of the test equals 0.9444, giving evidence to support the null hypothesis and conclude that the series x_{11t} has a unit root and is nonstationary. The same analytical procedure was followed for x_{12t} and x_{13t}.

Null Hypothesis: x_{11t} has a unit root

Exogenous: None

Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.259000	0.9444
Test critical values:	1% level	-2.627238	
	5% level	-1.949856	
	10% level	-1.611469	

Table 4.3: Augmented Dickey-Fuller test for a unit root

The analysis continued with the unit root test of the differenced series. The ADF test assessed the differenced term, Δx_{11t}, with one lag term included and intercept and trend components excluded. The results of the analysis are shown in Table 4.4. The *p*-value of the test equals 0.0001, giving sufficient evidence to support the alternative hypothesis that the series Δx_{11t} has no unit root and is stationary. This analysis confirmed that the x_{11t} series was $\mathrm{I}(1)$ and was therefore ready to be included in the MS group. The same procedure was followed for the x_{12t} and x_{13t} series. Results for these variables are included in the Appendix 4. Both x_{12t} and x_{13t} satisfied the inferential requirements to be declared $\mathrm{I}(1)$ processes.

Null Hypothesis: Δx_{11t} has a unit root

Exogenous: None

Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.430794	0.0001
Test critical values:	1% level	-2.628961	
	5% level	-1.950117	
	10% level	-1.611339	

Table 4.4: Augmented Dickey-Fuller test for a unit root

The number of cointegrating equations in the three variable system of model 1 was determined using the Johansen trace statistic test. Theoretical specification required that the intercept and trend terms be excluded and that the model has one lag term. The results of the analysis are given in Table 4.5. The *p*-value of the test shows the existence of one cointegrating equation. This was considered sufficient statistical evidence to conclude that the simulated data for replication one of model 1 was $\mathrm{CI}(1,1)$ and thus the data were included in the MS group. The same procedure was followed for the remaining $4\,999$ replications of the model.

Sample (adjusted): 3 40

Included observations: 38 after adjustments

Trend assumption: No deterministic trend

Series: x_{11t} x_{12t} x_{13t}

Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.527115	37.01330	24.27596	0.0008
At most 1	0.149721	8.554968	12.32090	0.1968
At most 2	0.061000	2.391706	4.129906	0.1440

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.527115	28.45833	17.79730	0.0009
At most 1	0.149721	6.163262	11.22480	0.3315
At most 2	0.061000	2.391706	4.129906	0.1440

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level

Table 4.5: Johansen test for cointegrating rank in multivariate model

This routine was used to group the data into two distinct groups. The groups were labeled MS and NMS. The MS group is the database of all simulated data that provided sufficient statistical evidence to support the theoretical specifications of the model. The NMS group is the database of all simulated data that provided insufficient statistical evidence to support the theoretical specification of the model. Excluded from the body of the text is an illustration of an example of a replication classified NMS, an example of this scenario is provided in the Appendix 5.

Continuing with the analysis of replication one, the data were arranged so that $T = 34$ observations were available for the estimation of the six VEC models. The model estimated required that the data had one cointegrating equation, no intercept or trend terms and was a VEC(0) model with no lag terms. The estimated output for the model is shown in Table 4.6 and the log-likelihood estimated for replication one of model 1 is $\ln L\left(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}; \mathbf{X}\right)^{\text{VEC}(0)} = -134.0186$.

Vector Error Correction Estimates
Sample (adjusted): 2 35
Included observations: 34 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Error Correction:	$\Delta x_{11t\text{-}5}$	$\Delta x_{12t\text{-}5}$	$\Delta x_{13t\text{-}5}$
CointEq1	-0.493623	0.041549	0.043008
	(0.08793)	(0.11300)	(0.09539)
	[-5.61390]	[0.36769]	[0.45086]

Determinant resid covariance (dof adj.)	0.582382
Determinant resid covariance	0.532492
Log likelihood	-134.0186
Akaike information criterion	8.236385
Schwarz criterion	8.505743

Table 4.6: Output of the estimation of the $\text{VEC}(0)$ model

Table 4.7 shows the results of the estimated log-likelihood for the $\text{VEC}(5)$ model with the dataset from the first replication. The models, $\text{VEC}(0)$ and $\text{VEC}(5)$, are used to illustrate the methods of estimation. The $\text{VEC}(1)$ to $\text{VEC}(4)$ models followed the same procedure but used a different number of starting observations but the same number, $T = 34$, of fixed estimation observations. The estimation procedure for the $\text{VEC}(5)$ model required that the

data had one cointegrating equation, no intercept or trend terms and five lag terms. The estimated output for the model is shown and the estimated log-likelihood for replication one of model 1 is $\ln L\left(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}; \mathbf{X}\right)^{\mathrm{VEC}(5)} = -89.4162$.

Vector Error Correction Estimates

Sample (adjusted): 7 40

Included observations: 34 after adjustments

Standard errors in ( ) & t-statistics in [ ]

Error Correction:	Δx_{11t}	Δx_{12t}	Δx_{13t}
CointEq1	-0.440086	0.207077	0.996112
	(0.26423)	(0.36151)	(0.27514)
	[-1.66554]	[0.57281]	[3.62038]
D(X11T(-1))	0.086633	0.246676	-0.809456
D(X11T(-2))	0.250518	-0.529697	-0.935089
D(X11T(-3))	0.087501	0.097275	-0.406740
D(X11T(-4))	0.101828	-0.031736	-0.051447
D(X11T(-5))	0.279043	0.022864	-0.431550
D(X12T(-1))	0.170617	0.314915	0.854390
D(X12T(-2))	0.367975	-0.218612	0.704821
D(X12T(-3))	0.122617	0.523011	0.761021
D(X12T(-4))	-0.622212	0.024856	0.512090
D(X12T(-5))	-0.221837	0.308501	0.021342
D(X13T(-1))	-0.146244	-0.221263	0.513616
D(X13T(-2))	0.140214	0.158223	0.754749
D(X13T(-3))	0.082780	-0.610340	0.457827
D(X13T(-4))	0.226216	0.084794	0.821923
D(X13T(-5))	0.125977	-0.323462	0.453575

Determinant resid covariance (dof adj.)	0.260309
Determinant resid covariance	0.038625
Log likelihood	-89.41621
Akaike information criterion	8.259777
Schwarz criterion	10.54932

Table 4.7: Output of the estimation of the $\mathrm{VEC}(5)$ model

The estimated likelihood values were then used to determine the estimated residual covariance and inserted into the criteria functions defined in Chapter 2. Continuing the illustration, the value of AIC was estimated for both the $\text{VEC}(0)$ and $\text{VEC}(5)$ models. The $\text{VAR}(p)$ estimated likelihood function is given in the users guide (EViews 5: 708) and defined in log form as $\ln L(\mathbf{A}, \boldsymbol{\Sigma}, \mathbf{X}) = -\frac{Tk}{2}(1+\ln 2\pi) - \frac{T}{2}\ln\left|\hat{\Sigma}\right|$, where $\left|\hat{\Sigma}\right| = \det\left[\sum_t \hat{\varepsilon}\hat{\varepsilon}' / T\right]$ and $k = \text{number of equations}$.

Rearranging to obtain $\ln\left|\hat{\Sigma}\right|$ gives

$$\ln\left|\hat{\Sigma}\right| = -\frac{2}{T}\left(\ln L\left(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}, \mathbf{X}\right) + \frac{Tk}{2}\left(1+\ln\left(2\pi\right)\right)\right)$$

$$= -\frac{2}{34}\left(\left(-134.0186\right) + \frac{(34)(3)}{2}\left(1+\ln\left(2\pi\right)\right)\right) = -0.6314.$$

Now substituting this estimate into the criterion estimate for the zero lag model gives

$$\text{AIC}^{\text{VEC}(0)} = \ln\left|\hat{\boldsymbol{\Sigma}}\right| + 2\frac{k^2\left(p-1\right)}{T}$$

$$= -0.6314 + 2\frac{\left(\left(3^2\right)\times(0)\right)}{34}$$

$$= -0.6314.$$

The estimated AIC value for the $\text{VEC}(5)$ model follows the same reasoning, the estimate is shown below:

$$\ln\left|\hat{\Sigma}\right| = -\frac{2}{34}\left(\left(-89.4162\right) + \frac{(34)(3)}{2}\left(1+\ln\left(2\pi\right)\right)\right) = -3.25506$$

$$\text{AIC}^{\text{VEC}(5)} = -3.25506 + \frac{2\left(\left(3^2\right)\times(5)\right)}{34} = -0.60800.$$

This illustration of the estimated criteria values for the first replication of the $\mathrm{VAR}(1)$ model with $N = 40$ is comprehensive. Similar calculations were computed for the other VEC models. Table 4.8 shows the criteria estimated for replication one. By design the criterion value which determines the best fitting model is the one with the smallest value. The last column identifies the model selected for each criterion. The results for replication one showed that the majority of criteria identified the $\mathrm{VEC}(0)$ model as the best fitted model. This procedure was followed for all $5\,000$ replications.

Criterion	VEC(0)	VEC(1)	VEC(2)	VEC(3)	VEC(4)	VEC(5)	Model selected
AIC	-0.6314	-0.5978	-0.2460	-0.3553	-0.3094	-0.6080	VEC(0)
MAIC	-0.2021	-0.1585	0.1396	0.1246	-0.3094	-0.6080	VEC(5)
AICC	2.7686	2.9839	3.6951	4.1993	5.2396	6.5449	VEC(0)
MAICC	2.2675	2.2714	2.6049	2.6078	2.5548	2.5033	VEC(0)
AICCBD	-0.5689	-0.2576	1.4094	9.2565	-20.9271	-10.3320	VEC(4)
BIC	-0.6314	-0.1937	0.5620	0.8568	1.3067	1.4122	VEC(0)
HQIC	-0.6314	-0.4600	0.0295	0.0581	0.2417	0.0809	VEC(0)
HQICC	-0.6314	-0.2870	0.5855	1.2972	2.6140	4.3065	VEC(0)
LCIC	-0.6314	-0.3269	0.2958	0.4575	0.7742	0.7466	VEC(0)
FPE	-0.6314	-0.5964	-0.2348	-0.3165	-0.2143	-0.4129	VEC(0)
ShibIC	-0.5444	-0.5656	-0.3335	-0.6117	-0.7735	-1.3110	VEC(5)
MBIC	0.1256	0.5807	1.2420	1.7030	1.3067	1.4122	VEC(0)
MHQIC	-0.0903	0.0936	0.5155	0.6629	0.2417	0.0809	VEC(0)

Table 4.8: Estimated criteria for replication one of $\mathrm{VAR}(1)$ with $N = 40$

The section that follows summarises the results of the analysis and discusses the advantages and disadvantages of several criterion ranking systems.

4.2 Introduction to the Assessment Results

The results of the simulations for the models and IC performances are reported in the accompanying tables. The tabulated results are the percentage of correct selections by the criterion. The tabulated results in section 4.3 to 4.5 include models 1 to 6 and models 8 to 9 as defined in chapter 3. The discussion of the results for model 7 of chapter 3 is withheld until section 4.6.

To simplify interpretations, this study defines six performance capability categories. These categories are shown in Table 4.9 and are rated according to the criterion's ability to correctly identify a model from the simulated data of the theoretically defined VEC model.

The first performance rating is defined as *excellent* and is assigned to criteria that select the correct model (95% – 100%] of the time. The second performance rating is defined as *very good* and is assigned to criteria that select the correct model (90% – 95%] of the time. The third performance rating is defined as *good* and is assigned to criteria that select the correct model (75% – 90%] of the time. The fourth performance rating is defined as *acceptable* and is assigned to criteria that select the correct model (60% – 75%] of the time. The fifth performance rating is defined as *poor* and is assigned to criteria that select the correct model (40% – 60%] of the time. The sixth and last performance rating is defined as *unacceptable* and is assigned to criteria that select the correct model [0% – 40%] of the time. These subjective ratings provide a method for comparing the overall capability of the criteria whilst still allowing individual comparisons.

No.	Percentage of correct classifications (%)	Performance Rating
1	(95 - 100]	Excellent
2	(90 - 95]	Very Good
3	(75 - 90]	Good
4	(60 - 75]	Acceptable
5	(40 - 60]	Poor
6	[0 - 40]	Unacceptable

Table 4.9: Performance capability categories

The IC results are tabulated and discussed sequentially. The MS results are given first, this is followed by the NMS results and overall summaries conclude the discussions.

4.3 How do the Information Criteria Perform Individually?

To answer this question, each criterion's performance was assessed for the MS and NMS data.

Meets Specification Data

Consider the results of the percentage of correct classifications of AIC in Table 4.10. Results are shown for eight models. The results for the 2^{nd} $\mathrm{VAR}(2)$ model with the alternative parameterisations are withheld from this summary as they are discussed separately when comparing parameterisation influence with the $\mathrm{VAR}(2)$ model defined as model 5 in Chapter 3.

VAR order	N = 40	N = 100	N = 200
1	78.2	94.0	95.5
2	33.2	83.7	87.2
3		77.7	
4		17.5	

Table 4.10: AIC percentage of correct classifications for the meet specification database

As the sample size increased from $N = 40$ to $N = 200$, the performance capability of AIC improved. This was observed for both the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models. In particular for the $\mathrm{VAR}(1)$ model, when the sample size was 40, AIC's performance capability was approximately 78% whilst when the sample size was 200 the performance capability was in excess of 95%. In general, AIC selected the correct model from which the data were simulated in excess of 80% of the time for the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models with sample sizes of 100 and 200. The AIC performance capabilities for the $\mathrm{VAR}(2)$, $N = 40$, and $\mathrm{VAR}(4)$, $N = 100$ models were unacceptable, with both cases below 40%. Using the performance ratings, AIC was classified as a good performer or better for six of the eight tabulated results.

The percentages of correct classifications of MAIC are given in Table 4.11. The results obtained were similar to those of AIC in that as the sample size increased the performance capability of the IC improved. The IC performance capabilities for the $\mathrm{VAR}(2)$, $N = 40$ and $\mathrm{VAR}(4)$, $N = 100$ models were also unacceptable, with both cases below 40% whilst the performance ratings were similar to those of AIC. MAIC was classified as a good performer or better for five of the eight tabulated results.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	78.1	92.8	95.6
2	32.1	82.6	84.3
3		74.5	
4		17.8	

Table 4.11: MAIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of AICC are given in Table 4.12. The influence of sample size was less noticeable on this criterion's performance than on AIC and MAIC. The performance ratings for the $\mathrm{VAR}(1)$ models were superior to those for AIC and MAIC. The IC performance rating for the $\mathrm{VAR}(1)$ models were classified as excellent for all the models, irrespective of the sample size. A drawback to AICC was the unacceptable identification of the $\mathrm{VAR}(4)$, $N = 100$, model. This model was selected 5.9% of the time, evidence that when the number of lag terms increased above three, AICC struggled to identify the model correctly. The IC was classified as a good performer or better for five of the eight results tabulated.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	98.1	96.4	96.8
2	21.5	84.4	87.8
3		73.2	
4		5.9	

Table 4.12: AICC percentage of correct classifications for the meet specification database

The percentages of correct classifications of MAICC are given in Table 4.13. The results were similar to AIC's in that as the sample size increased the performance capability of the IC improved. The IC performance capability for the $\mathrm{VAR}(4)$, $N = 100$, model was also unacceptable, with a selection of 15.8% whilst the performance ratings were similar to

those of AIC. MAICC was classified as a good performer or better for six of the eight results tabulated.

VAR order	$N = 40$	$N = 100$	$N = 200$
1	81.7	92.8	95.3
2	43.3	85.3	87.8
3		77.9	
4		15.8	

Table 4.13: MAICC percentage of correct classifications for the meet specification database

The percentages of correct classifications of AICCBD are given in Table 4.14. The results differ from AIC's for models with sample sizes of $N = 40$. The criterion was unable to identify the models when the sample sizes were small. AICCBD's performance capability improved considerably as the sample size increased, the IC was classified as a good performer or better for the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models with $N \geq 100$. The inability to identify three models was a limitation of the criterion's capability. All three unacceptable ratings had identifications of less than 1%.

VAR order	$N = 40$	$N = 100$	$N = 200$
1	0.0	97.3	97.1
2	0.0	84.1	87.4
3		58.1	
4		0.3	

Table 4.14: AICCBD percentage of correct classifications for the meet specification database

The percentages of correct classifications of BIC are given in Table 4.15. The results were noticeably different from the criteria already discussed. The performance rating for the low order $\mathrm{VAR}(1)$ model was excellent. Unfortunately that was the extent of this criterion's ability. The higher order VAR models were selected at an unacceptable level, even for the larger sized samples. These results indicated that lag order dimension had a substantial

influence on BIC's selection capabilities.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	99.7	100.0	100.0
2	4.8	18.8	23.0
3		6.6	
4		0.0	

Table 4.15: BIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of HQIC are given in Table 4.16. The results differ from those of AIC in that performance capabilities for the $\mathrm{VAR}(2)$ models were of lower ratings. When compared to BIC, HQIC's performance capabilities for the $\mathrm{VAR}(2)$ models were better. The results of HQIC's performance capability indicated that the criterion was a compromise between AIC and BIC. HQIC's results for the low order $\mathrm{VAR}(1)$ model were excellent but unfortunately similar results were not observed for the higher order models.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	96.1	99.8	100.0
2	25.9	62.9	66.8
3		48.8	
4		1.1	

Table 4.16: HQIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of HQICC are given in Table 4.17. The results were similar to those for HQIC in that the performance capability of the low order $\mathrm{VAR}(1)$ models was excellent with capability decreasing as lag order increased. Unacceptable performance ratings were obtained for three of the eight models assessed indicating a weaker performance rating than that for HQIC and much weaker performance rating than that for AIC.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	99.4	99.9	100.0
2	10.1	54.6	59.0
3		32.3	
4		0.1	

Table 4.17: HQICC percentage of correct classifications for the meet specification database

The percentages of correct classifications of LCIC are given in Table 4.18. LCIC was derived as a linear combination of HQIC and BIC and *priori* expectation was to capture the benefits of each criterion's strengths. The results were different from those for AIC in that the IC was an excellent performer for the low order $\mathrm{VAR}(1)$ model but capability for higher order models followed the results of HQIC and decreased rapidly. LCIC was classified as an unacceptable performer for four of the eight results tabulated, a disappointing result for a criterion expected to benefit from the strengths from which it was derived.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	99.1	100.0	100.0
2	12.8	37.9	41.9
3		21.8	
4		0.1	

Table 4.18: LCIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of FPE are given in Table 4.19. The results are similar to AIC in that the performance rating of the IC was classified as a good performer or better for six of the eight results tabulated. Unacceptable ratings were obtained for the $\mathrm{VAR}(2)$, $N = 40$, and $\mathrm{VAR}(4)$, $N = 100$, models, an observation consistent with almost all criteria assessed. The results of FPE were promising and further comparisons with the better performing IC follow.

VAR order	N = 40	N = 100	N = 200
1	82.9	94.0	95.5
2	36.4	83.8	87.3
3		77.9	
4		17.1	

Table 4.19: FPE percentage of correct classifications for the meet specification database

The percentages of correct classifications of ShibIC are given in Table 4.20. The results for the larger sized samples are similar to those for AIC. The criterion performed unacceptably for the $N = 40$ sized samples with approximately 28% and 9% capability for the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models, respectively. These results indicated that ShibIC is a useful model selector for studies with larger sized samples but should be avoided when sample sizes are small.

VAR order	N = 40	N = 100	N = 200
1	28.1	90.8	94.9
2	8.8	79.0	80.9
3		70.9	
4		24.0	

Table 4.20: ShibIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of MBIC are given in Table 4.21. The results are similar to BIC in that the criterion was an excellent performer for low order models but an unacceptable performer for higher order models. Results were as expected as the IC was a derivative of BIC with the expected benefit of the error restriction constraint. Like BIC, the selection capabilities for the larger sample $\mathrm{VAR}(1)$ models were 100%, a perfect score. Unfortunately these perfect scores were not obtained for the higher order model selections where the selection of $\mathrm{VAR}(2)$ models were all rated as unacceptable.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	98.4	100.0	100.0
2	11.8	26.8	25.3
3		15.2	
4		0.1	

Table 4.21: MBIC percentage of correct classifications for the meet specification database

The percentages of correct classifications of MHQIC are given in Table 4.22. The results are similar to those for HQIC in that the criterion was a very good performer for low order models with a decreasing performance ability as lag order increased. There was some improvement in selection capability of IC when compared to HQIC but insufficient improvement when compared to AIC or the AIC derivatives.

VAR order	*N* = 40	*N* = 100	*N* = 200
1	93.4	99.7	99.9
2	26.3	63.3	65.2
3		49.6	
4		2.1	

Table 4.22: MHQIC percentage of correct classifications for the meet specification database

To provide a comprehensive comparison of the IC based on the performance rating, this study proposed a weighted ranking scale. Each performance capability was given a weight and the criteria were ranked according to their overall rating. A decreasing weight was assigned to decreasing performance capabilities, the weights are given in the first row of Table 4.23. The last column captures the rank of the criterion based on the weighted performance capability.

Weights	(6)	(5)	(4)	(3)	(1)	(0)	
	Performance capability						
Criterion	Excellent	Very Good	Good	Acceptable	Poor	Unacceptable	Rank
AIC	1	1	4	0	0	2	3
MAIC	1	1	3	1	0	2	5
AICC	3	0	2	1	0	2	1
MAICC	1	1	4	0	1	1	2
AICCBD	2	0	2	0	1	3	8
BIC	3	0	0	0	0	5	12
HQIC	3	0	0	2	1	2	6
HQICC	3	0	0	0	2	3	10
LCIC	3	0	0	0	1	4	11
FPE	1	1	4	0	0	2	3
ShibIC	0	2	2	1	0	3	8
MBIC	3	0	0	0	0	5	12
MHQIC	2	1	0	2	1	2	7

Table 4.23: Performance rating summary of IC for the meet specification database

In order of decreasing performance capability, the top four performing criteria were AICC, MAICC and AIC and FPE, with a joint ranking. The worst performing IC were jointly BIC and MBIC. The results of this summary showed that for the eight models assessed, the efficiency based criteria (Akaike criteria) performed better than the consistency based criteria (Bayesian criteria). It was also clear that the criteria had difficulty identifying the $\mathrm{VAR}(2)$, $N=40$, and $\mathrm{VAR}(4)$, $N=100$ models. With the exception of MAICC, all criteria had an unacceptable rating for selecting these models. It would however be flattering to claim that MAICC did much better than the other IC as the criterion only did marginally better with ratings of unacceptable for the $\mathrm{VAR}(4)$, $N=100$ model, and poor for the $\mathrm{VAR}(2)$, $N=40$ model.

The consistency based IC were very good low order model identifiers but struggled as the dimension of the lag term increased. This was obvious given the number of unacceptable ratings observed for these IC.

Does Not Meet Specification Data

The IC capability results for the NMS data were very similar to the IC capability results for the respective MS data. Rather than repeat the individual criterion discussions, the results for the individual IC for the NMS data are tabulated in Appendix 6. For comparison purposes, the performance rating summaries were obtained and are given in Table 4.24 below.

Weights	(6)	(5)	(4)	(3)	(1)	(0)	
	Performance capability						
Criterion	**Excellent**	**Very Good**	**Good**	**Acceptable**	**Poor**	**Unacceptable**	**Rank**
AIC	1	1	3	1	0	2	3
MAIC	0	2	2	2	0	2	5
AICC	3	0	2	1	0	2	1
MAICC	0	2	4	0	0	2	3
AICCBD	2	0	2	0	1	3	8
BIC	3	0	0	0	0	5	12
HQIC	2	1	0	2	1	2	5
HQICC	3	0	0	0	2	3	10
LCIC	3	0	0	0	1	4	11
FPE	1	1	4	0	0	2	2
ShibIC	0	2	2	1	0	3	8
MBIC	3	0	0	0	0	5	12
MHQIC	2	1	0	2	1	2	5

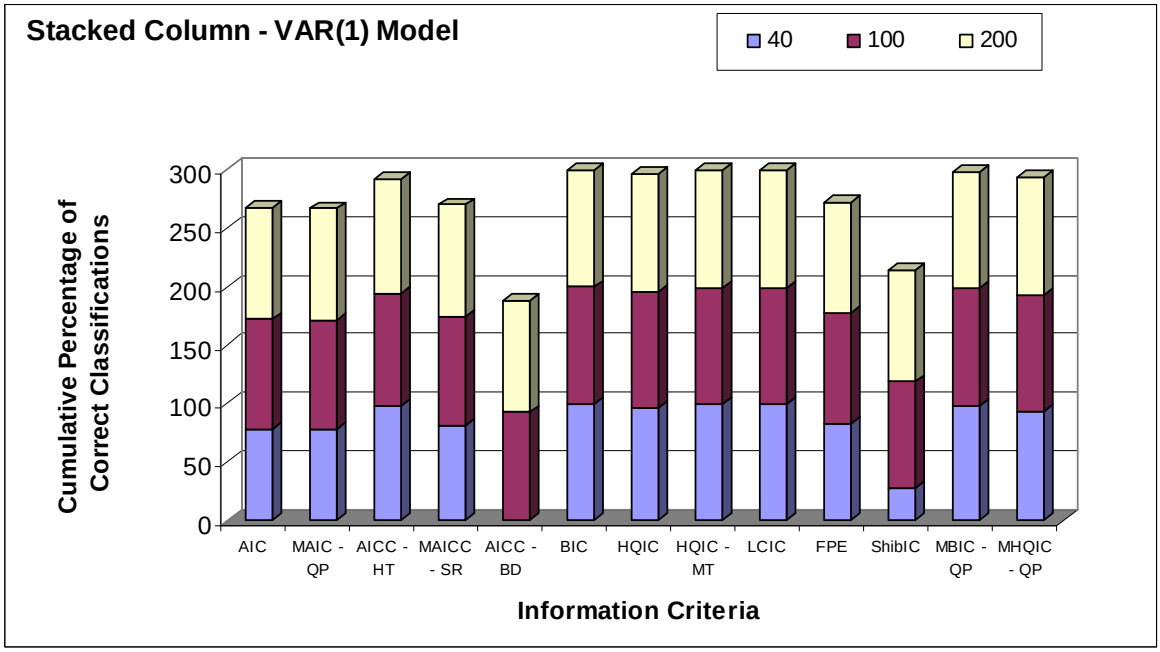
Table 4.24: Performance rating summary of IC for the does not meet specification database

A pattern similar to Table 4.23 was observed. The efficiency based criteria performed better than the consistency based criteria. Ranking orders showed nominal changes, with FPE and MAICC changing order, but in general the results were the same. The results showed that there was little benefit to partition the data into two groups as no discernible differences were observed.

4.4 How do the Information Criteria Perform as Sample Size Increases?

The discussion of this question began in the previous section, the general response was that as sample size increased, so too do the criteria's performance capabilities. In this section, the data are summarised graphically for the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models. Although the answer to this question is intuitive, the graphical results provided an opportunity for IC comparisons not previously observed. Given that the MS and NMS results were similar; this discussion is limited to the MS data only. The stacked column plots (Figures 4.2 and 4.3) for the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models are provided for discussion.

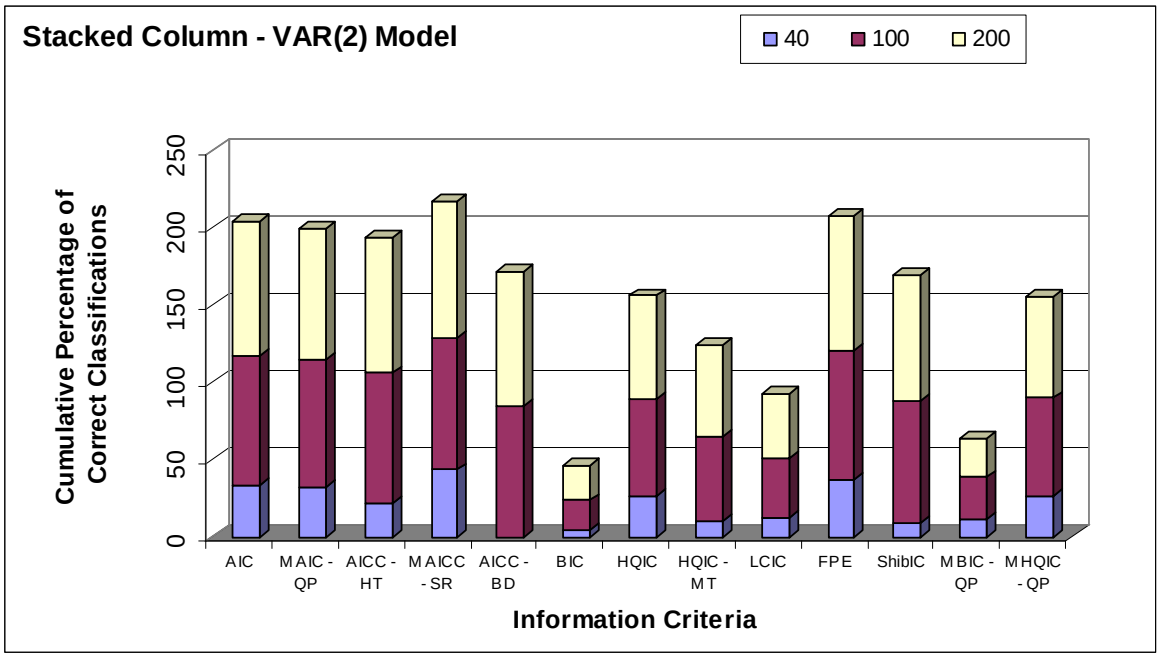
Figure 4.2: Criteria cumulative percentage of correct classifications for $\mathrm{VAR}(1)$ model



$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

The stacked columns in Figure 4.2 show the cumulative percentages of correct classifications of the IC performances for the $\mathrm{VAR}(1)$ models with different sample sizes. The graph indicates that identification for the $\mathrm{VAR}(1)$ models are good. Two exceptions were noted; the stacked columns for AICCBD and ShibIC are lower than those for the other IC showing that they were not as good as the other IC at identifying the low order models. Also noticeable are the exceptional performances of the Bayesian based criteria such as BIC, LCIC and MBIC and the Hannan-Quinn based criteria such as HQIC, HQICC and MHQIC.

Figure 4.3: Criteria cumulative percentage of correct classifications for $\mathrm{VAR}(2)$ model



$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.6 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 & 0.0 \\ 0.0 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}$$

The stacked columns in Figure 4.3 show the cumulative percentages of correct classifications of the IC performances for the $\mathrm{VAR}(2)$ models with different sample sizes.

The results displayed for these models were interesting. In general the Bayesian and Hannan-Quinn based criteria performed poorly. The dominant criteria for the $\mathrm{VAR}(2)$ models were Akaike based criteria with MAICC outperforming all other tested IC. A clear distinction between the top five performing IC and the rest was observed. This observation was an important result for this study as the results provided the first opportunity to recommend a model selection criterion for practitioners to consider when modelling ECMs.

This concludes the discussion on the IC performances as the sample size increases, the section that follows considers the outcomes of the IC performances as the dimension of the model increases with increasing lag order. The results of the forthcoming section were important for this study as the interpretations of the results have a direct bearing on the objective of identifying the most appropriate method for determining the lag structure of the cointegrated model.

4.5 How do the Information Criteria Perform as the Lag Length of the Model Increases?

Given the relevance of this question to this study, the results will be evaluated in detail. To answer this question, the four models with sample sizes of 100 were the most complete series and hence this assessment considered their results in particular. In addition this study considered the individual results of the MS and NMS data of the $N = 100$ modelled data.

The percentages of correct classifications for each model are summarised in Table 4.25, the results obtained were used to plot the stacked columns of the cumulative percentages of correct classifications shown in Figures 4.4 and 4.5. These stacked columns provide an excellent illustration of IC performance for the $\mathrm{VAR}(1)$ to $\mathrm{VAR}(4)$ models. The frequencies of the correct identification of the models for the MS and NMS data were then sorted to provide a performance ranking for the criteria. The performance ranking was based solely on the cumulative frequency of correct observations and provided an

alternative to the categorized ranking in section 4.3, Tables 4.23 and 4.24.

***N* = 100**	**MS Data**				**NMS Data**			
	VAR(1)	**VAR(2)**	**VAR(3)**	**VAR(4)**	**VAR(1)**	**VAR(2)**	**VAR(3)**	**VAR(4)**
AIC	94.0	83.7	77.7	17.5	95.5	84.6	79.9	17.6
MAIC	92.8	82.6	74.5	17.8	94.6	83.5	72.3	17.6
AICC	96.4	84.4	73.2	5.9	97.9	86.5	74.4	5.6
MAICC	92.8	85.3	77.9	15.8	94.7	85.9	78.8	16.2
AICCBD	97.3	84.1	58.1	0.3	98.3	87.2	59.0	0.4
BIC	100.0	18.8	6.6	0.0	100.0	19.2	7.3	0.0
HQIC	99.8	62.9	48.8	1.1	99.8	67.5	50.2	1.3
HQICC	99.9	54.6	32.3	0.1	99.9	59.3	33.8	0.1
LCIC	100.0	37.9	21.8	0.1	100.0	40.6	21.5	0.1
FPE	94.0	83.8	77.9	17.1	95.5	84.7	80.0	17.4
ShibIC	90.8	79.0	70.9	24.0	92.7	79.0	69.3	24.1
MBIC	100.0	26.8	15.2	0.1	100.0	28.4	17.8	0.2
MHQIC	99.7	63.3	49.6	2.1	99.6	69.8	52.5	1.8

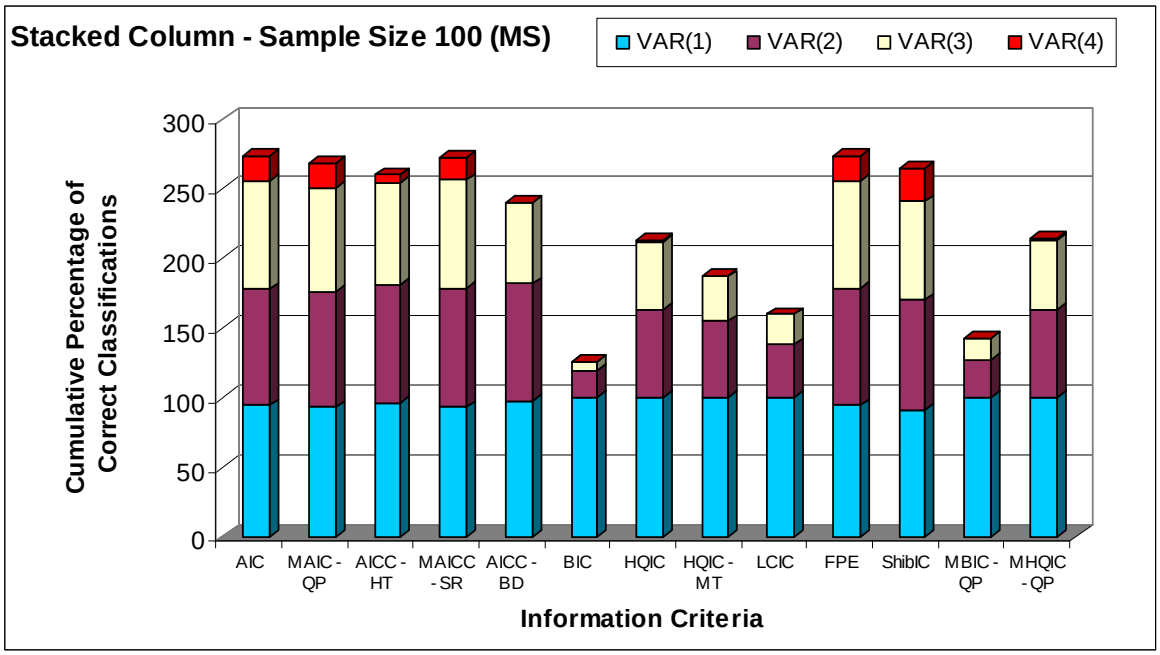
Table 4.25: Percentage of correct classifications of IC for samples of size $N = 100$

Highlighted cells in Table 4.25 identify the criteria which were the best and worst performers for the model identified by column label. The blue cells identify the best performing IC, whilst the yellow cells identify the worst performing IC. An interesting observation was that BIC and ShibIC lay claim to both labels depending on model lag structure. More importantly it was observed that the variability of performances for the $\mathrm{VAR}(1)$ model was considerably less than for the other models, indicating that identification of the $\mathrm{VAR}(1)$ model was relatively easy for most IC. The real problem came when trying to identify the higher order lag models. It is this problem that this study addresses and it then provides practitioners with a justification for their choice of model lag structure.

Meet Specification Data

The results of the cumulative percentages of correct classifications of the MS data are shown graphically in Figure 4.4.

Figure 4.4: Criteria cumulative percentage of correct classifications for the meet specification, $N = 100$ sized models



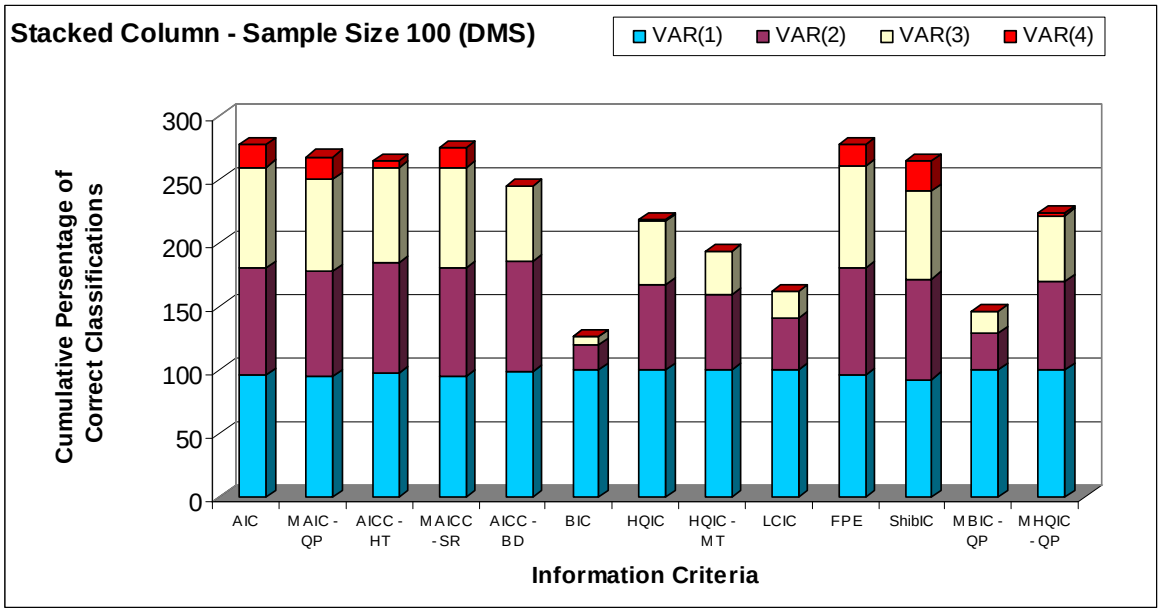
The graphical results in Figure 4.4 illustrated a clear distinction in performance capabilities of the IC evaluated. The percentages of correct classifications of the $\mathrm{VAR}(1)$ model, as observed by the height of the blue section of the stacked column, were approximately the same for each IC. Differences in performance capabilities were noticed as the models lag length was increased. The percentages of correct classifications of the $\mathrm{VAR}(2)$ model, as observed by the height of the purple section of the stacked column, provided the first noticeable difference in performance capabilities. There were distinct differences in the length of the purple columns. These differences were even more noticeable for the

$\mathrm{VAR}(3)$ and $\mathrm{VAR}(4)$ models as shown by the height of the yellow and red columns. The graphic clearly illustrates the difference in performance capability of the Akaike derived criteria versus the Bayesian and Hannan-Quinn derived criteria. For the models simulated with sample of size $N = 100$, the Akaike derived criteria were better performers than their Bayesan and Hannan-Quinn counterparts. This conclusion was reached as the cumulative frequency totals for the Akaike derived criteria were greater than those of their Bayesian and Hannan-Quinn equivalents.

Does Not Meet Specification Data

The results of the cumulative percentages of correct classifications for the NMS data are shown graphically in Figure 4.5.

Figure 4.5: Criteria cumulative percentage of correct classifications for does not meet specification, $N = 100$ sized models



There were marginal differences between the MS and NMS data sets, these marginal

differences were considered negligible. The general results and interpretations were the same, the Akaike derived criteria were much better performers than the Bayesian and Hannan-Quinn derived criteria for the models simulated with the samples of size $N = 100$. The results of the MS and NMS data sets indicated that for models with unknown lag structure, analysts using the Bayesian and Hannan-Quinn derived criteria would select the $\mathrm{VAR}(1)$ model well but would be less successful in the selecting of the correct higher order lag models. The inference is that if the model under examination has an unknown lag structure and the objective of the research is to determine the best fitting model then the Akaike based criteria are more likely to select the correct model.

The graphical displays in Figures 4.4 and 4.5 were illustrative for visual identification of poor performing criteria but numerical summaries were required for performance rankings. Table 4.26 summarises the cumulative frequencies of correct lag identification of the models (rather than percentages of correct classifications) for the $N = 100$ sized models. The data were summarised for the MS and NMS data separately and then combined to provide a cumulative total for the number of correct selections from $20\,000$ possible simulations.

The cumulative frequency results in Table 4.26 provided an opportunity for ranking the cumulative performance of the IC for models using sample sizes of $N = 100$. The last column in Table 4.26 gave the sequentially ranked IC based on the cumulative frequencies. The results obtained showed that there was little difference between the three top ranked IC based on their cumulative frequencies. An interesting observation was the poor overall performance of the Bayesian and Hannah-Quinn based criteria. The top ranked criterion, AIC, selected the correct model $13691\,(68.5\%)$ times out of a possible $20\,000$ times, whilst the bottom ranked criterion, BIC, only selected the correct model $6\,278\,(31.4\%)$ times. These low scoring percentages are an indication of the difficulty faced by analysts when determining the correct lag structures of the VAR models in VEC representations.

Criterion	Meet specification		Does not meet specification		Cumulative	
	Frequency	**Rank**	**Frequency**	**Rank**	**Frequency**	**Rank**
AIC	11128	1	2563	1	13691	1
FPE	11124	2	2563	1	13687	2
MAICC	11083	3	2542	3	13625	3
ShibIC	10768	4	2469	4	13237	4
AICC	10625	5	2411	5	13036	5
MAIC	10405	6	2354	6	12759	6
AICCBD	9816	7	2221	7	12037	7
MHQIC	8788	8	2028	8	10816	8
HQIC	8706	9	1982	9	10688	9
HQICC	7653	10	1746	10	9399	10
LCIC	6545	11	1462	11	8007	11
MBIC	5826	12	1317	12	7143	12
BIC	5143	13	1135	13	6278	13

Table 4.26: Frequency of correct lag identification and rank summary of assessment of models with $N = 100$

To explain the performances of the IC, the selection choices of the criteria for individual models were tabulated. Given the clear distinction between the poor and better performers, only the results of the top five ranked IC from Table 4.26 were summarised. The summarised data were tabulated as percentages of correct classifications for ease of comparison between criteria.

The results in Tables 4.27, 4.28 and 4.29 summarise the relative frequencies of the top five ranked IC for the $\mathrm{VAR}(2)$, $\mathrm{VAR}(3)$ and $\mathrm{VAR}(4)$ models with samples sized $N = 100$. The column highlighted in yellow identifies the model from which the data were simulated, given the similarity of results between MS and NMS, the results summarised were only from the MS database.

IC by lag model	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)	VAR(6)
AIC	9.5	83.7	5.3	0.9	0.5	0.1
FPE	9.5	83.8	5.3	0.9	0.4	0.1
MAICC	8.3	85.3	5.4	0.8	0.2	0.0
ShibIC	7.5	79.0	7.4	2.4	1.8	1.8
AICC	13.6	84.4	1.9	0.1	0.0	0.0

Table 4.27: Percentage of correct classifications summary for $\mathrm{VAR}(2)$, N = 100 model

The performances of the top five ranked IC for the $\mathrm{VAR}(2)$ model were relatively consistent. Four of the five results ranged between 83.7% and 85.3% with the fifth result a credible 79.0%. Two observations need mentioning, the selection of the data as a $\mathrm{VAR}(1)$ model, referred to in the literature as underfitting, occurred approximately 10% of the time whilst the selection of the data as a $\mathrm{VAR}(3)$ or higher order model, referred to as overfitting, occurred less than the underfitting. This observation indicated a bias towards underfitting. The noticeable exception was ShibIC which showed a bias towards overfitting. In general the selection performances of the top five ranked IC for the $\mathrm{VAR}(2)$, $N = 100$ model were reasonable.

The performances of the top five ranked IC for the $\mathrm{VAR}(3)$ model, given in Table 4.28, were less consistent than those for the $\mathrm{VAR}(2)$ model. The spread of the results was greater than that of the $\mathrm{VAR}(2)$ model hinting that model identification was becoming more difficult at the higher order. The results ranged between 70% and 78%, performances considerably better than those for some of the IC not grouped in the top five. Underfitting by AICC was more noticeable than for the other IC whilst overfitting by ShibIC was once again observed. The performances of the top three ranked IC were similar, both in terms of underfitting and overfitting with some evidence of a bias towards underfitting.

IC by lag model	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)	VAR(6)
AIC	0.6	13.4	77.7	5.9	1.7	0.7
FPE	0.6	13.6	77.9	5.8	1.5	0.6
MAICC	0.4	14.5	77.9	5.7	1.2	0.3
ShibIC	0.4	8.7	70.9	9.6	4.9	5.5
AICC	1.4	24.0	73.2	1.4	0.0	0.0

Table 4.28: Percentage of correct classifications summary for $\mathrm{VAR}(3)$, $N = 100$ model

The performances of the top five ranked IC for the $\mathrm{VAR}(4)$ model are given in Table 4.29. These results were less flattering than their performances for the lower ordered models. The best performance, by ShibIC, was a lowly 24% highlighting the difficulty of model selection for the higher dimensional models. This difficulty was consistent for all IC, including those not ranked in the top five.

IC by lag model	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)	VAR(6)
AIC	0.2	29.4	48.9	17.5	3.0	0.9
FPE	0.2	29.8	49.3	17.1	2.8	0.7
MAICC	0.1	29.7	51.8	15.8	2.3	0.3
ShibIC	0.2	19.6	42.0	24.0	7.8	6.5
AICC	0.5	49.8	43.5	5.9	0.2	0.0

Table 4.29: Percentage of correct classifications summary for $\mathrm{VAR}(4)$, $N = 100$ model

A plausible reason for this difficulty was that as the model lag dimension increased, the models parameter values decreased which made parameter estimation more difficult. As an illustration, the parameter values for the $\mathrm{VAR}(4)$ model are given below:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.4 & 0.5 \\ 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{pmatrix} + \begin{pmatrix} 0.2 & -0.2 & 0.1 \\ 0.0 & 0.25 & 0.0 \\ 0.0 & 0.0 & 0.3 \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \\ x_{3t-2} \end{pmatrix}$$

$$+ \begin{pmatrix} -0.15 & 0.15 & 0.2 \\ 0.0 & 0.15 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \\ x_{3t-3} \end{pmatrix} + \begin{pmatrix} 0.1 & -0.1 & 0.0 \\ 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.15 \end{pmatrix} \begin{pmatrix} x_{1t-4} \\ x_{2t-4} \\ x_{3t-4} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}.$$

Consider the coefficients for the fourth lagged term, only four of a possible nine coefficients were non zero, whilst three of those four coefficients were $|a_{ij}| = 0.1$, a value close to zero. When estimating the confidence interval for a parameter, if the interval covers zero, standard practice implies that the variable corresponding to the parameter is insignificant. In this model, the close proximity of the parameters to zero would adversely influence the significance of the fourth lagged vector, $\mathbf{x_{t-4}}$, leading towards model underfitting.

Unfortunately the error corrected restriction of the VAR model usually ensures that as the lag dimension increases the coefficient values decrease. Exceptions to this would be models with small coefficient values for the lower order structure, a case not common in the literature. This argument applies to all higher order dimensional models, hence this study's assessment restriction is to a maximum lag structure of four.

Tables 4.30 and 4.31 summarise the cumulative frequencies of correct lag identification of the models (rather than percentages of correct classifications) for the $N = 40$ and $N = 200$ sized models, respectively. The data were summarised for the MS and NMS data for both the $\mathrm{VAR}(1)$ and $\mathrm{VAR}(2)$ models separately and then combined to provide a cumulative total for the number of correct selections from 10000 replications (R). Just as was done for the $N = 100$ based models, the cumulative frequency results provided an opportunity for ranking the cumulative performances of the IC for models using the sample sizes of $N = 40$ and $N = 200$. The last columns in Tables 4.30 and 4.31 give the ranking as determined by cumulative frequencies.

N = 40	MS		NMS		Total		Cumulative Total (R = 10000)	Rank
	VAR(1) (R = 3876)	VAR(2) (R = 1658)	VAR(1) (R = 1124)	VAR(2) (R = 3342)	VAR(1) (R = 5000)	VAR(2) (R = 5000)		
AIC	3031	551	827	880	3858	1431	5289	9
MAIC	3027	533	826	1022	3853	1555	5408	7
AICC	3803	357	1097	516	4900	873	5773	3
MAICC	3165	718	867	1271	4032	1989	6021	1
AICCBD	0	0	0	0	0	0	0	13
BIC	3866	79	1121	122	4987	201	5188	10
HQIC	95	430	1065	620	1160	1050	2210	11
HQICC	3854	167	1117	230	4971	397	5368	8
LCIC	3842	212	1114	293	4956	505	5461	5
FPE	3214	604	882	961	4096	1565	5661	4
ShibIC	1088	146	261	202	1349	348	1697	12
MBIC	3814	195	1106	313	4920	508	5428	6
MHQIC	3621	436	1035	795	4656	1231	5887	2

Table 4.30: Frequency of correct lag identification and rank summary of assessments of models with N = 40

The results seen in Table 4.30 showed that the top ranked criterion, MAICC, selected the correct model 6021 out of 10000 (60.2%) times whilst the bottom ranked criterion, AICCBD was unable to select the correct model at all. The inability of AICCBD to select the correct model for the $N = 40$ sized samples was a serious shortcoming of the criterion and was a result of excessive overfitting by the criterion. The low scoring frequency counts were an indication of the difficulties faced by analysts, confronted with small datasets, who need to determine the correct lag structure of the VEC/VAR model.

An interesting observation was the reasonable performances of the small sample correction criterion based on the Hurvich and Tsai (1993) methodology. Both MAICC and AICC were two of the better performers, with MAICC performing marginally better than AICC, an indication that the error restricted correction proposed by Qu and Perron (2006) improved the selection capability of Hurvich and Tsai's (1993) small sample bias correction.

N = 200	MS		NMS		Total		Cumulative Total (R = 10000)	Rank
	VAR(1) (R = 3876)	VAR(2) (R = 1658)	VAR(1) (R = 1124)	VAR(2) (R = 3342)	VAR(1) (R = 5000)	VAR(2) (R = 5000)		
AIC	3903	3611	866	738	4769	4349	9118	5
MAIC	3905	3490	859	710	4764	4200	8964	6
AICC	3957	3636	880	744	4837	4380	9217	2
MAICC	3896	3633	859	744	4755	4377	9132	3
AICCBD	3969	3619	885	747	4854	4366	9220	1
BIC	4086	951	914	162	5000	1113	6113	13
HQIC	4085	2766	914	561	4999	3327	8326	8
HQICC	4085	2443	914	500	4999	2943	7942	10
LCIC	4086	1733	914	343	5000	2076	7076	11
FPE	3903	3615	866	739	4769	4354	9123	4
ShibIC	3878	3350	860	685	4738	4035	8773	7
MBIC	4086	1047	914	205	5000	1252	6252	12
MHQIC	4081	2697	914	558	4995	3255	8250	9

Table 4.31: Frequency of correct lag identification and rank summary of assessments of models with N = 200

The results seen in Table 4.31 for $N = 200$ models contradicted some of the results for the $N = 40$ models. The best performing IC as identified by rank was AICCBD, the worst performer for the $N = 40$ case. Anomalies like this illustrate the difficulties of model selection and emphasise the dependency of criteria performances on sample size, estimation procedures and dimensions of the model both in terms of lag structure and variable number.

Surprisingly the small sample correction based criteria, MAICC and AICC were again two of the better performers, with MAICC performing marginally better than AICC, lending support to the claim that the error restricted correction proposed by Qu and Perron (2006) improved the selection capability of Hurvich and Tsai's (1993) criterion, albeit within a larger sample framework.

The results presented in this section addressed the question of how the criteria perform as the lag length of the models increased. The discussions included detailed descriptions of individual criterion performances and proposed ranking systems. In the closing summary

of this chapter, these ranking systems were compared and recommendations were made. The final question for this chapter considers the effects of changing the parameters of the theoretical models whilst ensuring the error correction restriction.

4.6 How do the Information Criteria Perform as the Model's Parameter Structure Changes?

This study considered the impact of the simulation results for two $\mathrm{VAR}(2)$ models (models 5 and 7 in Chapter 3), with sample sizes of $N=100$. The change in parameter values were intentionally kept small to allow for criteria performance comparisons. The parameter values of the models are shown below, the first model (model 5) was simulated with parameterisation:

$$x_{1t} = 0.4x_{1t-1} + 0.4x_{2t-1} + 0.5x_{3t-1} + 0.2x_{1t-2} - 0.2x_{2t-2} + 0.0x_{3t-2},$$

$$x_{2t} = 0.0x_{1t-1} + 0.8x_{2t-1} + 0.0x_{3t-1} + 0.0x_{1t-2} + 0.2x_{2t-2} + 0.0x_{3t-2}, \text{ and}$$

$$x_{3t} = 0.0x_{1t-1} + 0.0x_{2t-1} + 0.6x_{3t-1} + 0.0x_{1t-2} + 0.0x_{2t-2} + 0.4x_{3t-2}.$$

The second model (model 7) was simulated with parameterisation:

$$x_{1t} = 0.4x_{1t-1} + 0.4x_{2t-1} + 0.6x_{3t-1} + 0.2x_{1t-2} - 0.2x_{2t-2} + 0.0x_{3t-2},$$

$$x_{2t} = 0.2x_{1t-1} + 0.8x_{2t-1} - 0.15x_{3t-1} - 0.1x_{1t-2} + 0.15x_{2t-2} + 0.0x_{3t-2}, \text{ and}$$

$$x_{3t} = 0.0x_{1t-1} + 0.0x_{2t-1} + 0.6x_{3t-1} + 0.0x_{1t-2} + 0.0x_{2t-2} + 0.4x_{3t-2}.$$

The parameter changes (highlighted in red) primarily affected the x_{2t} equation with nominal changes to four of the six lag terms in the equation. The results of the IC performances with the simulated data are summarised as frequencies of correct lag identification and given in Table 4.32. A performance ranking for the cumulative totals is

shown in the last column. Considering the similarities of the frequencies of correct lag identification for the MS and NMS data, the results were summarised as cumulative frequencies for the combined MS and NMS data.

VAR(2) N = 100	Total		Δ = M5 – M7	Cumulative Total (R = 10000)	Rank
	Model 5 (R = 5000)	Model 7 (R = 5000)			
AIC	4193	4740	-547 (10.9%)	8933	5
MAIC	4138	4701	-563 (11.3%)	8839	6
AICC	4239	4844	-605 (12.1%)	9083	2
MAICC	4271	4744	-473 (9.5%)	9015	3
AICCBD	4232	4929	-697 (13.9%)	9161	1
BIC	943	3712	-2769 (55.4%)	4655	13
HQIC	3187	4897	-1710 (34.2%)	8084	9
HQICC	2772	4882	-2110 (42.2%)	7654	10
LCIC	1920	4564	-2644 (52.9%)	6484	11
FPE	4198	4743	-545 (10.9%)	8941	4
ShibIC	3951	4622	-671 (13.4%)	8573	7
MBIC	1356	3808	-2452 (49.0%)	5164	12
MHQIC	3225	4865	-1640 (32.8%)	8090	8

Table 4.32: Frequency of correct lag identification and rank of assessments of different parameterisation models

The frequency differences between model 5 (first parameterisation) and model 7 (second parameterisation), denoted as $\Delta = \text{M5} - \text{M7}$, are highlighted in yellow on Table 4.32. It is clear from these differences that the IC performances for model 7 are better than those for model 5. All the differenced results are negative showing that all observed correct model selection frequencies are better for model 7 than for model 5. These results demonstrate the difficulties faced with large dimensional modelling problems. In addition, these differences demonstrate that even for models which have small parameter differences, the identification of the correct lag structure using IC is highly volatile. The percentage change in correct identification ranged between a moderate 9.5% and a volatile 55.4%.

In conclusion, the IC performances were ranked using the cumulative totals of the two models. The selection performances for the majority of criteria were reasonable, this was observed by the 80.8% relative frequency for HQIC, the criteria ranked a lowly 9th. The

two exceptions were BIC (46.6%) and MBIC (51.6%), their selection performances were disappointing.

4.7 Results Summary

This section summarises the analysis of the results. Table 4.33 summarises the performance rankings of the results discussed so far. It should be noted that in many cases the column rankings are dependent assessments, i.e. part of the performance rating of the MS data was also captured in the $N = 100$ ranking, similarly part of the performance rating of the MS data was also captured in the different parameterisation assessment.

Rank by method	Performance Rating (MS)	Performance Rating (NMS)	N = 100	N = 40	N = 200	Different Parameterisation
AIC	3	3	1	9	5	5
MAIC	5	5	6	7	6	6
AICC	1	1	5	3	2	2
MAICC	2	3	3	1	3	3
AICCBD	8	8	7	13	1	1
BIC	12	12	13	10	13	13
HQIC	6	5	9	11	8	9
HQICC	10	10	10	8	10	10
LCIC	11	11	11	5	11	11
FPE	3	2	2	4	4	4
ShibIC	8	8	4	12	7	7
MBIC	12	12	12	6	12	12
MHQIC	7	5	8	2	9	8

Table 4.33: Performance ranking of IC in results chapter

The ranking systems used in this study illustrated marked differences in the performance capabilities of some of the criteria. The performances of five of the IC were consistently worse than those of other IC. These criteria are those which have either a Bayesian or Hannan-Quinn basis. In particular, the criteria BIC, MBIC, HQIC, HQICC and LCIC are often ranked in the bottom half of the IC list. Considering the poor performances of these IC, it is strongly recommended that their application to VEC modelling be applied only when it is necessary to underfit the lag structure of the model based on an economic

justification. Except in the large sampled $\mathrm{VAR}(1)$ models, these IC are unlikely to select the correct lag structure of the VEC model.

Some of the criteria were inconsistent in their capabilities, as an example, AICCBD was ranked last for the $N = 40$ data and first for the $N = 200$ data model. This volatility in identification is not recommended for model selection and the use of this criterion for model selection can best be described as cautionary. Also included in this category of cautionary use are ShibIC and MHQIC. Both exhibit volatility in their capabilities with ShibIC performing poorly in small samples and MHQIC only performing well in the small sample case.

The performances of AIC and MAIC were reasonable. The limitation of AIC was exposed by the low ranking for the $N = 40$ sized sample, whilst MAIC was a consistent performer for all ranked cases without performing exceptionally for any method. The use of these criteria for VEC modelling is justified on the basis that they are methods which have shown reasonable performances, are easily to implement in analysis routines and are theoretically justified.

The best performing criteria were AICC, MAICC and FPE. These criteria were the most consistent across the ranking systems and in several cases were the best performing IC for the ranking method. The benefits of the Qu and Perron (2007) error restricted term in the VEC model was captured by the MAICC whilst maintaining the bias correction of Hurvich and Tsai's (1993) small sample correction of Akaike's original derivation.

In conclusion, this chapter discussed the evaluation of the selection capabilities of thirteen informational criteria for nine simulation models. Assessments have been done in terms of the influence of sample size, parameterisation and lag structure. The results for these evaluations were discussed individually and performance rankings for the criteria obtained. The rankings are summarised and the performances of individual criteria were discussed.

CHAPTER 5

ILLUSTRATIONS OF TWO EXAMPLES

5.1 Introduction

In this chapter, the lag length selection criteria are applied to two published datasets. The purposes of this were two-fold, first to provide empirical evidence of the estimation procedures used in this study by reproducing published results and thereafter to analyse data in a VEC framework and propose a model based on the criteria's model selections.

5.2 Software Validation

EViews 5.1 was used for estimating the likelihood functions for the VEC models of the simulated data series. This section shows the computational outputs of the estimated routines and reconciliation of the results of the software with the results of an independent source. The data used for this purpose were taken from the text of Lütkepohl $(2005:145-148)$ who used the data to show estimation and lag length selection methods for a stationary *3d*-VAR model.

The dataset for this example were three quarterly variables, investment, income and expenditure, which were seasonally adjusted for the periods 1960Q1 to 1982Q4. The series had 92 quarterly periods although Lütkepohl (2005) used the first 76 observations for VAR modelling. The last 16 observations were withheld for forecasting comparisons and were excluded from the estimation process. To validate the software, the same 76 observations were used for this illustration. The full and partial datasets are provided in the Appendix 7. The three variables in the dataset were measured in billions of Deutsche Mark for West

Germany and defined as the quarterly fixed investment (x_{1t}), the quarterly disposable income (x_{2t}) and the quarterly consumption expenditures (x_{3t}) for the period. The time graphs of the complete dataset for the three variables are shown in Figures 5.1 and 5.2.

Figure 5.1: Period plot for fixed investment variables from Lütkepohl (2005: 77-78)

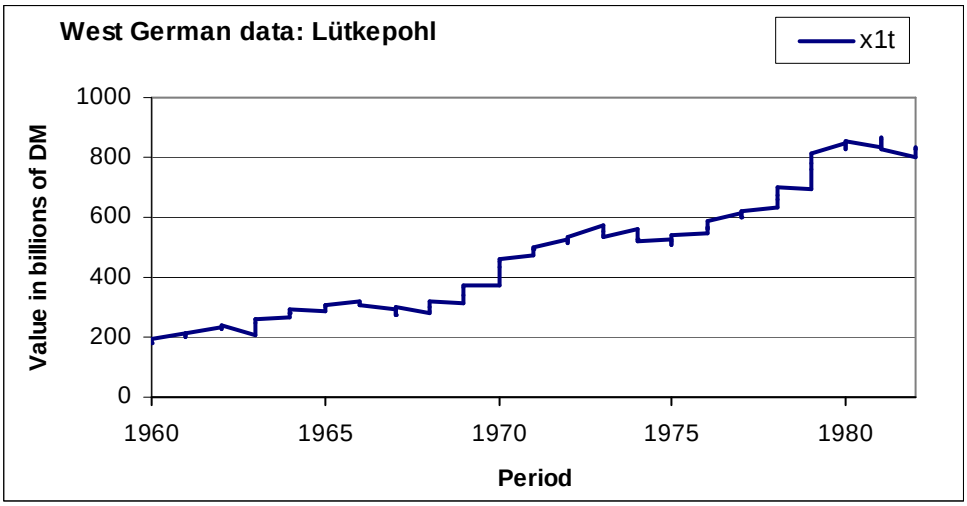


Illustration of trending x_{1t} series

Figure 5.2: Period plot for disposable income and consumption expenditures variables from Lütkepohl (2005: 77-78)

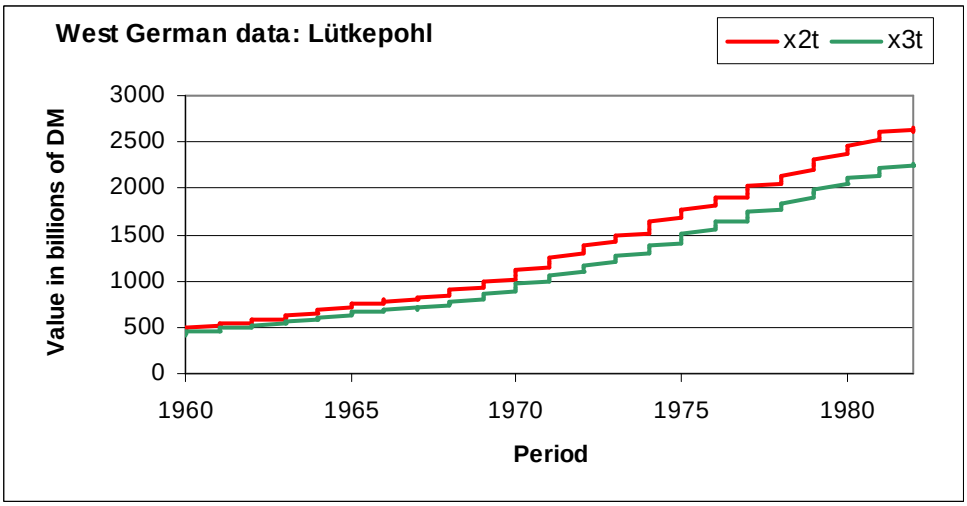


Illustration of trending x_{2t} and x_{3t} series

The graphical displays showed that the investment, income and expenditure variables have an increasing trend. For the purposes of this illustration, the trends are sufficient evidence to conclude that the series are nonstationary.

To model the series as a VAR process, the individual series were log transformed and then differenced to obtain a stationary series. The graphs of the transformed series are shown in Figures 5.3 and 5.4. A visual inspection lends support to the belief that the transformed series are stationary. As the purpose of this example was to reconcile the estimation results of EViews 5.1 with the estimation results of an independent source, the assumption of Lütkepohl (2005) that the transformed series were stationary was accepted.

Figure 5.3: Period plots for differenced logs of fixed investment variable from Lütkepohl (2005: 77-78)

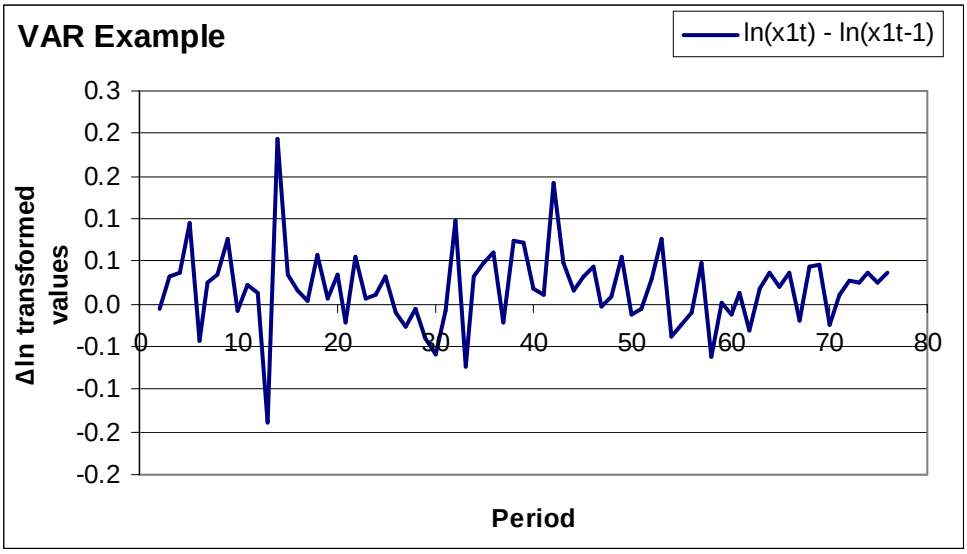


Illustration of stationary transformed series, $\ln(x_{1t}) - \ln(x_{1t-1})$

The illustrations in Figures 5.3 and 5.4 are replicates of the plots in Lütkepohl (2005: 79) and confirm that the series used for the validity process of this study are the same as the series already in the public domain.

Figure 5.4: Period plots for differenced logs of disposable income and consumption expenditures variables from Lütkepohl (2005: 77-78)

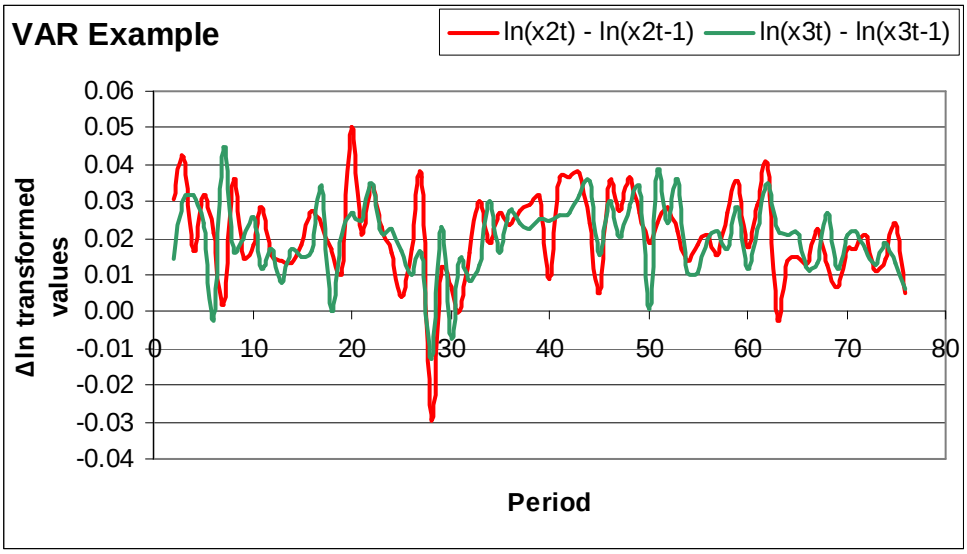


Illustration of stationary transformed series, $\ln(x_{2t}) - \ln(x_{2t-1})$ and $\ln(x_{3t}) - \ln(x_{3t-1})$

The analysis by Lütkepohl (2005) restricted the maximum lag length of the VAR process to four. In so doing, the partial dataset of 76 observations was reduced by an additional five observations, four of these observations were used for the lag length upper bound and the fifth was used for the differencing of the log variables. Therefore a total of 71 observations were available for estimation.

Lütkepohl (2005) estimated five VAR models starting with a $\text{VAR}(0)$ up to and including a $\text{VAR}(4)$ model. To estimate the $\text{VAR}(0)$ model, Lütkepohl (2005) included an intercept (constant) term, an approach followed in this example. Given the inclusion of an intercept term in the $\text{VAR}(0)$ model and to compare like models, Lütkepohl (2005) included intercept terms in all VAR models. The results of Lütkepohl (2005: 148) are shown in Table 5.1. Highlighted in yellow are the models selected by the four criteria used to demonstrate model selection.

VAR order	FPE (x10^{11})	AIC	HQIC	BIC
0	2.691	-24.42	-24.42	-24.42
1	2.500	-24.50	-24.38	-24.21
2	2.272	-24.59	-24.37	-24.02
3	2.748	-24.41	-24.07	-23.55
4	2.910	-24.36	-23.90	-23.21

Table 5.1: Criteria estimation results of Lütkepohl (2005: 148)

To validate the estimation routine used in this study, the $\mathrm{VAR}(1)$ model was estimated, the estimated likelihood function obtained and the AIC and BIC results computed. These results were then compared to the tabulated results in Table 5.1.

Reconciliation of definitions

To validate results required reconciling the definitions of Lütkepohl (2005) and the definitions used in this study. Lütkepohl (2005: 147) used the constant omitted definition of AIC given as $\mathrm{AIC}_{\mathrm{Lüt}}^{\mathrm{VAR}(p)} = \ln\left|\widehat{\boldsymbol{\Sigma}}\right| + \frac{2k^2 p}{T}$ whilst the definition of AIC used by EViews 5.1 is $\mathrm{AIC}_{\mathrm{Eviews}}^{\mathrm{VAR}(p)} = -\frac{2}{T}\ln L\left(\mathbf{A},\boldsymbol{\Sigma},\mathbf{X}\right) + \frac{2k^2 p}{T} + \frac{2kd}{T}$. Hence the direct comparison of numerical estimates would not validate results. To compare the results, the definition of Lütkepohl (2005) was rearranged to include the constant terms.

EViews 5.1 defines the $\mathrm{VAR}(p)$ likelihood function as $\ln L(\mathbf{A},\boldsymbol{\Sigma},\mathbf{X}) = -\frac{Tk}{2} - \frac{Tk}{2}\ln 2\pi - \frac{T}{2}\ln\left|\widehat{\boldsymbol{\Sigma}}\right|$. Therefore re-arranging the likelihood function in terms of the log determinant of the estimated error covariance matrix gives $\ln\left|\widehat{\boldsymbol{\Sigma}}\right| = -\frac{2}{T}\ln L(\mathbf{A},\boldsymbol{\Sigma},\mathbf{X}) - k - k\ln 2\pi.$ Substituting into $\mathrm{AIC}_{\mathrm{Lüt}}^{\mathrm{VAR}(p)}$ gives $\mathrm{AIC}_{\mathrm{Lüt\text{-}reconciled}}^{\mathrm{VAR}(p)} = -\frac{2}{T}\ln L\left(\mathbf{A},\boldsymbol{\Sigma},\mathbf{X}\right) - k - k\ln 2\pi + \frac{2k^2 p}{T}$. The use of this definition allowed for the comparison of numerical results.

Example

The estimation results of the $\text{VAR}(1)$ model with intercept term included are shown in Table 5.2. The estimated likelihood function is $\ln L(\hat{\mathbf{A}}, \hat{\mathbf{\Sigma}}, \mathbf{X}) = 576.4087$.

Vector Autoregression Estimates
Sample (adjusted): 2 72
Included observations: 71 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	LNDX1TV1	LNDX2TV1	LNDX3TV1
LNDX1TV1(-1)	-0.251427	0.034004	-0.004552
	(0.11969)	(0.03062)	(0.02691)
	[-2.10068]	[1.11055]	[-0.16912]
LNDX2TV1(-1)	0.305564	-0.089126	0.220840
	(0.51781)	(0.13247)	(0.11643)
	[0.59011]	[-0.67283]	[1.89676]
LNDX3TV1(-1)	0.636619	0.257615	-0.209026
	(0.61570)	(0.15751)	(0.13844)
	[1.03398]	[1.63557]	[-1.50985]
C	0.002440	0.016296	0.019258
	(0.01305)	(0.00334)	(0.00293)
	[0.18696]	[4.88153]	[6.56344]
Determinant resid covariance (dof adj.)		2.12E-11	
Determinant resid covariance		1.78E-11	
Log likelihood		**576.4087**	
Akaike information criterion		-15.89884	
Schwarz criterion		-15.51641	

Table 5.2: EViews 5.1 estimation output for $\text{VAR}(1)$ model

The Eviews 5.1 table shows the parameter estimates for the lagged terms denoted by (-1), whilst standard errors are shown in ( ) and t-statistics in [ ]. The log likelihood, AIC and BIC estimates are shown in the last column of the table.

Substituting the estimate of the likelihood function into the $\text{AIC}_{\text{Lüt-reconciled}}^{\text{VAR}(p)}$ gives

$$\text{AIC}_{\text{Lüt-reconciled}}^{\text{VAR}(p=1)} = -\frac{2}{71}(576.4087) - 3 - 3\ln 2\pi + \frac{2(3)^2(1)}{71} = -24.50.$$

This is the same result obtained by Lütkepohl (2005: 148) and shown in Table 5.1. Similar calculations are done for the other VAR models and the estimated results are given in Table 5.3. The numerical values are almost identical to Lütkepohl (2005: 148), there is a difference of 0.01 observed for the $\text{VAR}(2)$ estimate which is assumed to be a rounding difference and considered negligible; all other estimates are exact to the second decimal point as shown in Table 5.3.

VAR order	$\text{AIC}_{\text{Lüt}}^{\text{VAR}(p)}$	$\ln L(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}, \mathbf{X})$	$\text{AIC}_{\text{Lüt-reconciled}}^{\text{VAR}(p)}$	**Validated**
0	-24.42	564.784	-24.42	Yes
1	-24.50	576.409	-24.50	Yes
2	-24.59	588.859	-24.60	Yes
3	-24.41	591.237	-24.41	Yes
4	-24.36	598.457	-24.36	Yes

Table 5.3: Results of AIC estimates for Lütkepohl (2005) and EViews 5.1

As an additional confirmation of the estimation procedure used in this study the same method was used to compare the BIC estimates for the $\text{VAR}(1)$ model. The reconciled formula was derived by substituting the estimated covariance into the definition used by Lütkepohl (2005) and is given as

$$\text{BIC}_{\text{Lüt}}^{\text{VAR}(p)} = \ln\left|\hat{\Sigma}\right| + \frac{pk^2 \ln T}{T},$$

$$\text{BIC}_{\text{Lüt-reconciled}}^{\text{VAR}(p)} = -\frac{2}{T}\ln L\left(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}, \mathbf{X}\right) - k - k\ln 2\pi + \frac{pk^2 \ln T}{T}, \text{ and}$$

$$\text{BIC}_{\text{Lüt-reconciled}}^{\text{VAR}(p)} = -\frac{2}{71}(576.4087) - 3 - 3\ln 2\pi + \frac{(1)3^2 \ln 71}{71} = -24.2114.$$

The estimation results are given in Table 5.4. All five numerical values are identical to the

results of Lütkepohl (2005: 148). This reconciliation confirms the validity of the estimation routine used in EViews 5.1 for the VAR model.

VAR order	$\mathrm{BIC}_{\text{Lüt}}^{\mathrm{VAR}(p)}$	$\ln L(\hat{\mathbf{A}}, \hat{\boldsymbol{\Sigma}}, \mathbf{X})$	$\mathrm{BIC}_{\text{Lüt-reconciled}}^{\mathrm{VAR}(p)}$	Validated
0	-24.42	564.784	-24.42	Yes
1	-24.21	576.409	-24.21	Yes
2	-24.02	588.859	-24.02	Yes
3	-23.55	591.237	-23.55	Yes
4	-23.21	598.457	-23.21	Yes

Table 5.4: Results of BIC estimates for Lütkepohl (2005) and EViews 5.1

5.3 An Example in a VEC Framework

Given the validation evidence of the software capability this section continues with a second empirical example of the determination of the lag length of a cointegrated dataset. The dataset of four U.S. economic variables for this example was used by Lütkepohl (2005: 312) to model a four dimensional system from a Bayesian approach. This study used the same data and modelled the system in a cointegrated framework to compare the 13 criteria defined in Chapter 2. This comparison is restricted to a maximum lag length of four in the VEC framework.

The dataset, quarterly data for four U.S. economic variables from 1954 to 1987, each with 136 observations are defined as

x_{1t} = logarithm of the real money stock, M1,

x_{2t} = logarithm of GNP in billions of 1982 dollars,

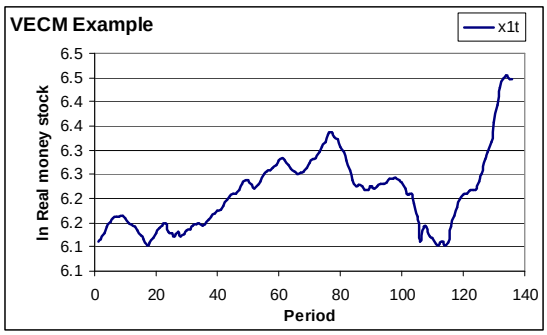
x_{3t} = discount interest rate on new issues of 91-day Treasury bills, and

x_{4t} = yield on long term $(20$ years$)$ Treasury bonds.

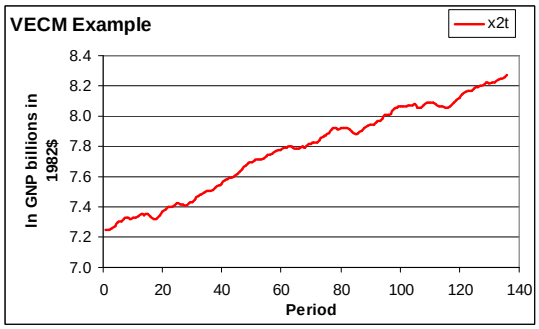
The plots of x_{1t} and x_{2t} are shown in Figure 5.5, whilst the plots of x_{3t} and x_{4t} are shown in Figure 5.6. Based on the graphical evidence the variables were assumed to be

nonstationary and of order $I(1)$. These assumptions were considered reasonable as the purpose of this section was to illustrate selection capabilities of criteria rather than to conduct an econometric assessment of the data.

Figure 5.5: Plots of x_{1t} = logarithm of the real money stock, M1, and x_{2t} = logarithm of GNP in billions of 1982 dollars

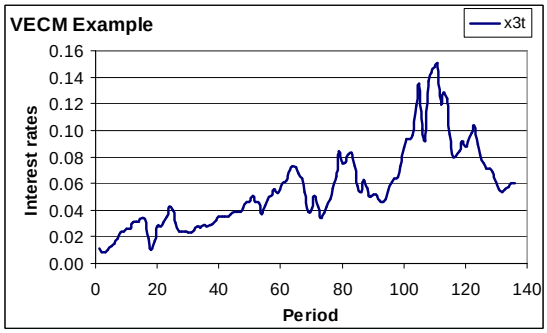


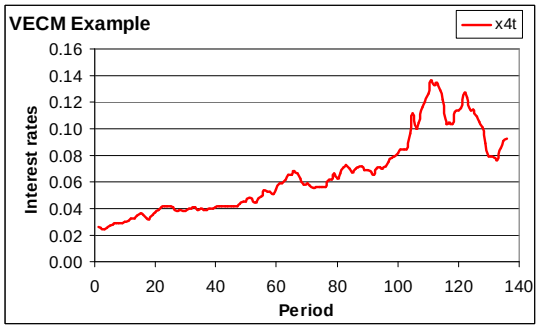
Series assumed nonstationary as mean changes over time.



Series assumed nonstationary as increasing trend observed.

Figure 5.6: Plots of x_{3t} = discount interest rate on new issues of 91-day Treasury bills, and x_{4t} = yield on long term $(20$ years$)$ Treasury bonds





Both series assumed nonstationary as mean changes over time with some evidence of volatility, thus variability changes with time.

The system of equations was tested for evidence of cointegration using the trace statistic and the computational results of the test are given in Table 5.5. These results demonstrated one of the difficulties when modelling a large dimensional cointegrated system; the

analysis produced conflicting results dependent on the structure of the model and the inferential test used. This study used the trace statistic to evaluate cointegrated simulation models without an intercept or trend term. For consistency this practice was followed and the conclusion reached was that there were three cointegrated relationships between the four variables. Assuming the no intercept and no trend model, Table 5.5 showed that the trace statistic inferred three cointegrated relationships. Accepting the conclusion of three cointegrated relationships and restricting the maximum lag length to four, the sample sizes were fixed (Ng & Perron, 2005) to 131 observations for each VEC model.

Sample: 1 136
Included observations: 131
Series: X1T X2T X3T X4T
Lags interval: 1 to 4

Selected (0.05 level*) Number of Cointegrating Relations by Model

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	**No Intercept**	Intercept	Intercept	Intercept	Intercept
	No Trend	No Trend	No Trend	Trend	Trend
Trace	**3**	2	1	1	1
Max-Eig	3	1	1	0	0

*Critical values based on MacKinnon-Haug-Michelis (1999)

Table 5.5: Cointegration assessment of U.S. data

The four VEC models were then estimated, the resulting likelihood functions determined and the equivalent trace statistics recorded. The log of the determinant of the estimated covariance matrices were calculated from the likelihood functions. To illustrate this procedure, extracts of these assessments for the $\mathrm{VEC}(1)$ model are given in Tables 5.6 and 5.7, the complete assessment results are included in the Appendix 8.

Vector Error Correction Estimates
Sample (adjusted): 3 133
Included observations: 131 after adjustments

Determinant resid covariance (dof adj.)	1.88E-18
Determinant resid covariance	1.51E-18
Log likelihood	**1944.394**
Akaike information criterion	-29.07472
Schwarz criterion	-28.19679

Table 5.6: Determination of likelihood estimate for $\mathrm{VEC}(1)$ model

Table 5.6 shows the estimated log-likelihood for the $\mathrm{VEC}(1)$ model, with $\ln L(\hat{\mathbf{A}}, \hat{\mathbf{\Sigma}}, \mathbf{X}) = 1944.39$. This estimate was used to determine the log determinant of the covariance estimate, i.e. $\ln\left|\hat{\Sigma}\right|$. As an example, consider the $\mathrm{VEC}(1)$ model with $\ln L(\hat{\mathbf{A}}, \hat{\mathbf{\Sigma}}, \mathbf{X}) = 1944.39$, the log determinant of the covariance estimate was calculated as

$$\ln\left|\hat{\mathbf{\Sigma}}\right| = -\frac{2}{T}\ln L(\hat{\mathbf{A}}, \hat{\mathbf{\Sigma}}, \mathbf{X}) - k - k\ln 2\pi$$

$$= -\frac{2}{T}(1944.39) - 4 - 4\ln 2\pi = -41.039.$$

This calculation was done for each model and the results summarised in Table 5.8. To estimate some of the criterion functions, an estimate of the trace statistic was required. To illustrate this procedure, extracts of this estimate for the $\mathrm{VEC}(1)$ model are given in Table 5.7. The trace statistic for three cointegrating relationships for the $\mathrm{VEC}(1)$ model was estimated as $\hat{\tau}_1(3) = 3.696597$.

Sample (adjusted): 3 133
Included observations: 131 after adjustments
Trend assumption: No deterministic trend
Series: X1T_3 X2T_3 X3T_3 X4T_3
Lags interval (in first differences): 1 to 1
Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.324202	83.99203	40.17493	0.0000
At most 1 *	0.132333	32.65816	24.27596	0.0035
At most 2 *	0.076084	14.06312	12.32090	0.0253
At most 3	0.027824	3.696597	4.129906	0.0647

Trace test indicates 3 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level

Table 5.7: Determination of trace statistic for $4d$-VEC(1) model with three cointegrated relationships

These results for the likelihood estimates, the log determinant covariance estimates and trace statistics are summarised in Table 5.8. Using these results the criterion estimates for model selection were calculated and the results are summarised in Table 5.9.

	VEC order			
Functions	**1**	**2**	**3**	**4**
Likelihood estimate	1944.4	1964.5	1976.2	1990.3
Covariance estimate	-41.039	-41.345	-41.525	-41.739
Trace statistic	3.6966	3.1448	3.2906	3.1608

Table 5.8: Estimated statistics for 4*d*-VEC model with 3 cointegrated relationships

The last column in the Table 5.9 identifies the VEC model selected on the basis of the minimum criterion function. The results are unsurprising in that the Bayesian and Hannan-Quinn based criteria select models of lower order than the Akaike based criteria. Given the propensity to underfit data by the Bayesian and Hannan-Quinn based criteria, this study would suggest that the most appropriate model for econometric based interpretation would be the $4d-\mathrm{VEC}(2)$ model with three cointegrated relationships. This decision would be based on the agreement between the model selection criteria of AIC, AICC, MAICC and FPE.

No.	Criterion	VEC order				Model Selected
		1	2	3	4	
1	**AIC**	-40.794	-40.856	-40.792	-40.762	$\mathrm{VEC}(2)$
2	**MAIC**	-40.738	-40.808	-40.742	-40.714	$\mathrm{VEC}(2)$
3	**AICC**	-36.612	-36.633	-36.507	-36.394	$\mathrm{VEC}(2)$
4	**MAICC**	-36.593	-36.636	-36.537	-36.463	$\mathrm{VEC}(2)$
5	**AICCBD**	-40.738	-40.664	-40.315	-39.739	$\mathrm{VEC}(1)$
6	**BIC**	-40.443	-40.154	-39.739	-39.358	$\mathrm{VEC}(1)$
7	**HQIC**	-40.652	-40.571	-40.364	-40.192	$\mathrm{VEC}(1)$
8	**HQICC**	-40.623	-40.486	-40.191	-39.896	$\mathrm{VEC}(1)$
9	**LCIC**	-40.547	-40.362	-40.051	-39.775	$\mathrm{VEC}(1)$
10	**FPE**	-40.794	-40.856	-40.790	-40.757	$\mathrm{VEC}(2)$
11	**ShibIC**	-40.744	-40.830	-40.801	-40.816	$\mathrm{VEC}(2)$
12	**MBIC**	-40.415	-40.130	-39.713	-39.334	$\mathrm{VEC}(1)$
13	**MHQIC**	-40.562	-40.495	-40.284	-40.115	$\mathrm{VEC}(1)$

Table 5.9: Criterion selection for the U.S. economic dataset

To conclude, this chapter provided sufficient evidence that the estimation routines in EViews 5.1, the software used for analysis in this study, is adequate for the study. The independent assessment of the VAR models by Lütkepohl (2005) lends support to this conclusion. A complete illustration of a publicly available dataset has been provided and the results of the criteria selections have been discussed.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

6.1 Introduction

The results of criteria selection performances for other VAR model simulation studies are reported in this chapter. The results are summarised and comparisons with the current study are discussed. Included in the discussion are the arguments for the many contradictory conclusions reached.

6.2 Discussion Including Results from Other Studies

Lütkepohl (1985) reported the results of a criteria model selection simulation exercise for a $3d-\mathrm{VAR}(1)$ stationary process for samples sized $T=40$ and $T=100$. The results were reported as rankings on a relative frequency $(\%)$ scale between 0 and 100. The results omitted to include the relative frequencies, so the results in Table 6.1 are approximations based on the position of the criteria on the scale.

Stationary $3d-\mathrm{VAR}(1)$ models		
Criteria	$T=40$	$T=100$
BIC	1 (~100%)	1 (~100%)
HQIC	2 (~99%)	1 (~100%)
FPE	3 (~85%)	3 (~92%)
AIC	4 (~80%)	3 (~92%)
ShibIC	5 (~18%)	5 (~87%)

Table 6.1: Summarised relative frequencies of correct lag identification of Lütkepohl (1985)

The results provided strong evidence that BIC was the preferred model selector for the $3d-\mathrm{VAR}(1)$ model for both the $T=40$ and $T=100$ sized samples. Based on these results, Lütkepohl (1985) advocated the use of BIC as the preferred model selector for VAR models.

The Lütkepohl (1985) study did not include the assessment of criteria performances for higher order VAR models. This limitation prevented the detection of underfitting by BIC and HQIC as observed in the current study. The approximated results of the $3d-\mathrm{VAR}(1)$ model were very close to the results of the $3d-\mathrm{VAR}(1)$ model of the current study. The rank sequences of the studies were identical, indicating a close relationship between the stationary and nonstationary VAR models.

The next study of relevance to this thesis was the theoretical developments of the AICC by Hurvich and Tsai (1993). The selection performances of criteria were assessed using two bivariate simulation models. The results of the relative frequencies of correct lag identification are summarised in Table 6.2. The results provided satisfactory evidence that AICC and AICCBD were the best selectors for the $2d-\mathrm{VAR}(1)$ and $2d-\mathrm{VAR}(2)$ models for the $T=40$ sized samples. Based on these results Hurvich and Tsai (1993) advocated the use of AICC as the preferred model selector for small sampled VAR models.

Stationary VAR models with $T=40$		
Criteria	$2d-\mathrm{VAR}(1)$	$2d-\mathrm{VAR}(2)$
AIC	0.87	0.43
AICC	0.94	0.71
AICCBD	0.97	0.79
BIC	0.99	0.68

Table 6.2: Summarised relative frequencies of correct lag identification of Hurvich and Tsai (1993)

Hurvich and Tsai (1993) used a $2d$ model to demonstrate criteria performances whilst the current study used a $3d$ model. The dimension incompatibility makes direct comparisons difficult but there are similarities worth mentioning. As expected the selection

performances of the criteria for the $\mathrm{VAR}(1)$ models were better than the $\mathrm{VAR}(2)$ models and performances for AIC were similar. The noticeable differences were the performances of AICCBD and BIC.

The large scale simulation study by Koreisha and Pukkila (1993) reported the results of criteria model selection for several simulation models. Of interest to this study were the results for the $3d-\mathrm{VAR}(1)$ and $3d-\mathrm{VAR}(2)$ models for samples sized $T=50$ and $T=100$. Their study also considered the selection performances of alternative model parameterisations defined as diagonal and triangular. The results of the relative frequencies of correct lag identification are summarised in Table 6.3. Although the results provided some evidence that BIC or HQIC were the preferred criteria model selectors, Koreisha and Pukkila (1993) were more circumspect with their recommendations. They chose not to advocate any particular criterion, rather they concluded that selection performances were influenced by model dimensions and the number of non-zero elements in the parameter matrices.

Stationary VAR models with alternative parameterisations					
		$T=50$		$T=100$	
Dimension	Criteria	Diagonal	Triangular	Diagonal	Triangular
$3d-\mathrm{VAR}(1)$	AIC	0.88	0.96	0.99	0.98
	BIC	0.60	1.00	1.00	1.00
	HQIC	0.83	1.00	1.00	1.00
$3d-\mathrm{VAR}(2)$	AIC	0.92	0.90	0.97	0.95
	BIC	1.00	1.00	1.00	1.00
	HQIC	0.99	0.98	1.00	1.00

Table 6.3: Summarised relative frequencies of correct lag identification of Koreisha and Pukkila (1993)

The evaluations of only three criteria were the shortcomings of the Koreisha and Pukkila (1993) study. When compared to the current study, the performances of BIC and HQIC for the $3d-\mathrm{VAR}(2)$ model were surprising. The near perfect performances of BIC and HQIC were a direct contradiction to the current study and caused some concern. Inspection of the parameter choices for the parameter matrix $\mathbf{A}_2$ revealed that the numerical values for the

Koreisha and Pukkila (1993) study were larger than this study providing easier identification of the 2nd order term. These results were consistent with the conclusions of Koreisha and Pukkila (1993) that selection performances were influenced by parameter matrices. The current study's alternative parameterisation for the $3d-\mathrm{VAR}(2)$ model supported this argument. The performances of AIC in Koreisha and Pukkila's (1993) study and the current study were similar.

The simulation study by Gonzalo and Pitarakis (1998) reported results of criteria model selection for simulation models with different parameterisations. Of interest to this study were the results for the $3d-\mathrm{VAR}(1)$ models with samples sized $T=150$ and other larger sized samples. The results of the relative frequencies of correct lag identification are summarised in Table 6.4 and 6.5. The results in Table 6.4 were obtained for the $\mathrm{VAR}(1)$ model by changing the parameter value of a_{11} in the coefficient matrix

$$\mathbf{A_1}=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}=\begin{pmatrix} 0.60 & 0 & 0 \\ 0 & 1.00 & 0 \\ 0 & 0 & 1.00 \end{pmatrix}.$$

The value of a_{11} ranged between 0.60 and 1.00 both inclusive, for models with different sample sizes. Although Gonzalo and Pitarakis (1998) concluded that no criterion was a "*clear overall performer*" they criticised the lack of variability in the AIC performances, whilst praising the performances of BIC, HQIC and LCIC. These comments typically lead one to discern that the results provided some evidence that BIC, HQIC and LCIC were the preferred criteria model selectors. An alternative interpretation of these results would be that the performances of BIC were highly volatile and influenced by parameter choice whilst AIC performances were more robust to different parameterisations.

Nonstationary $3d-\text{VAR}(1)$ models with $T=150$					
Criteria	$a_{11}=0.60$	$a_{11}=0.70$	$a_{11}=0.80$	$a_{11}=0.90$	$a_{11}=1.00$
AIC	0.64	0.64	0.64	0.55	0.47
BIC	0.97	0.76	0.23	0.02	1.00
HQIC	0.90	0.90	0.78	0.27	0.90
LCIC	0.96	0.92	0.52	0.10	0.98

Table 6.4: Summarised relative frequencies of correct lag identification of Gonzalo and Pitarakis (1998)

Gonzalo and Pitarakis (1998) failed to explain the poor performances of BIC, HQIC and LCIC for $a_{11}=0.90$ but were sufficiently cautious so as not to advocate BIC as the preferred model selector. The lack of higher order VAR simulation models was a limitation and probably played a huge role in the decision to criticise performances of AIC.

Kadilar and Erdemir (2002) reported results of criteria performances for the simulation exercises of $3d-\text{VAR}(1)$ and $3d-\text{VAR}(2)$ models with $T=100$. Four criteria were assessed and the performances are summarised in Table 6.5. The results reported provided strong evidence that BIC and HQIC were the best model selectors for sample sizes of $T=100$. Based on these results Kadilar and Erdemir (2002) praised the performances of BIC and HQIC without going so far as to recommend them as the preferred model selectors for VAR models.

Stationary VAR models with $T=100$		
Criteria	$3d-\text{VAR}(1)$	$3d-\text{VAR}(2)$
AIC	0.86	0.84
BIC	1.00	1.00
ShibIC	0.72	0.61
HQIC	1.00	0.99

Table 6.5: Summarised relative frequencies of correct lag identification of Kadilar and Erdemir (2002)

The study was almost a replication of the experiment undertaken by Koreisha and Pukkila (1993). The parameterisation choices were the same, so too were the sample sizes. The

inclusion of ShibIC in the assessments was the primary difference between the exercises. The results for BIC and HQIC in their study were almost identical to the Koreisha and Pukkila (1993) study. Although Kadilar and Erdemir (2002) discussed the bias corrected univariate criterion of Hurvich and Tsai (1989) they did not include an assessment of the multivariate VAR criterion of Hurvich and Tsai (1993). In all, a useful validation exercise but an opportunity lost to compare alternative criteria, for example AICC, or an alternative parametric structures.

The follow-up study by Gonzalo and Pitarakis (2002) reported preferences of criteria model selection for several stationary $kd-\mathrm{VAR}(2)$ $(k=2,3,...,10)$ simulation models. Relevant to this study were the results for the $3d-\mathrm{VAR}(2)$ models with samples sized $T=100,150$ and 200. The results summarised in Table 6.6 give criteria performances for models with different parameterisations values. Gonzalo and Pitarakis (2002) used two parameterisation methods, large values (close to 1) and small values (close to 0). Based on these results Gonzalo and Pitarakis (2002) strongly advocated the use of AIC as the preferred model selector for VAR models whilst emphasising the underfitting observed by BIC.

Stationary $3d-\mathrm{VAR}(2)$ models						
	$T=100$		$T=150$		$T=200$	
Criteria	Large	Small	Large	Small	Large	Small
AIC	0.26	0.00	1.00	0.08	1.00	0.23
BIC	0.00	0.00	0.05	0.00	0.68	0.00
HQIC	0.46	0.00	0.99	0.00	1.00	0.00

Table 6.6: Summarised relative frequencies of correct lag identification of Gonzalo and Pitarakis (2002)

The differences in the performances between large and small parameter values were noticeable. The results reported strongly supported Koreisha and Pukkila's (1993) conclusions that parameter choices had an influence on criteria performances. Unfortunately the exercise undertaken only considered the performances of three criteria with VAR models of maximum lag two, a limitation to an otherwise good study. The results of the current study were similar to those of the Gonzalo and Pitarakis (2002) study

with noticeable underfitting by BIC for higher order lag models with limited underfitting by AIC.

The theoretical derivation of MAIC by Qu and Perron (2007) was followed by a simulation exercise to compare its performance with AIC and BIC. The results of criteria performances for a $3d-\mathrm{VAR}(3)$ model are summarised in Table 6.7. The results provided evidence that MAIC outperformed both BIC and AIC as model selectors for error restricted VAR models. Based on these results Qu and Perron (2007) advocated the use of MAIC as the preferred model selector for error restricted VAR models.

Stationary $3d-\mathrm{VAR}(3)$ models with $T=200$		
Criteria	Intercept model	Intercept and trend model
MAIC	0.16	0.21
AIC	0.08	0.07
BIC	0.00	0.00

Table 6.7: Summarised relative frequencies of correct lag identification of Qu and Perron (2007)

The study reported the findings of criteria performances for two VEC models, a model which included an intercept term and a model which included both an intercept and trend term. The dimension incompatibility with this study makes direct comparisons difficult but worth mentioning are the poor performances of BIC for both studies. In general, the results showed poor performances for all three criteria assessed, the inclusion of criteria performances for the lower order models could have provided additional justification for the recommendations made.

6.3 Summary

The properties of cointegrated VAR model reduction techniques are still generally unknown. This thesis adds to the body of knowledge by considering criteria performances for these error restricted VAR models.

In general, the limitations of previous studies were that

- the majority of the studies only evaluated a few criteria, usually including AIC and BIC,
- the lag structures of the models were predominantly lower order models, in many cases restricted to 1st order with more recent studies including 2nd order models, and
- assessments of criteria performances were predominantly for unrestricted stationary VAR models.

The consensus of most simulation studies undertaken were that

- as the sample size increased, selection performances improved,
- as the dimension of the lag structure increased, selection performance decreased, and
- the parameterisation choice for models influenced the identification of the lag structure.

The current study's additions to the literature are

- the simultaneous evaluations of the performances of many criteria not previously assessed together,
- the evaluations of the performances of higher order lag models,
- defining a small sample bias correction criterion for error restricted VAR models,
- finding support for Gonzalo and Pitarakis's (2002) claim that BIC underfits whilst AIC is the better performer, and
- the tentative confirmation that the identification pattern for VEC models is similar to the identification pattern for VAR models.

The implications for practitioners are that

- the larger the sample, the better the chance of correct selection,
- they must be aware that underfitting will quite likely occur,
- higher order lag terms with parameters close to zero will unlikely be identified, and
- efficiency based selection criteria are cumulatively better model selectors.

6.4 Conclusions

This study has assessed the capabilities of thirteen IC to select the correct lag order of an ECM in a VAR framework. Based on simulation exercises the results of the study lend support for Akaike based informational criteria with one (or more) of AIC, AICC or MAICC as the preferred lag length selector for data based analysis.

The results contradict some of the previous simulation based exercises which in general have used stationary $\mathrm{VAR}(1)$ models for their simulations. The choice of the $\mathrm{VAR}(1)$ model has meant that BIC and HQIC have often been the preferred criterion choices, but as this study has shown, as the order increases, these criteria underfit considerably, strengthening the argument for the efficiency based preferences. These recommendations are not lone voices in the advocacy of Akaike based criterion choices, Hurvich and Tsai (1993) advocated AICC, Gonzalo and Pitarakis (2002) advocated AIC and Qu and Perron (2007) advocated MAIC, all directly related to AIC.

6.5 Recommendations and Further Work

An omission from this study was the lack of a simulation based example for the error restricted $\mathrm{VAR}(3)$ and $\mathrm{VAR}(4)$ models with $N = 200$. Given the performances of the IC in this study, this addition would add value by providing a comparison to the $N = 100$ sized models and the lower order $N = 200$ sized models. Although smaller sized sample simulation studies would be beneficial to practioners, the limitation encountered when faced by starting with a reasonable maximum possible lag length, as $k \leq 6$ which was used in this study, requires a large sample size for estimation of all the lagged term parameters and the covariance matrix. This is an area that would be of particular benefit to practitioners as sample size is an area that has received no attention in the literature.

Further investigation into the influences of parameter choices should be considered as the results from this study and that of Koreisha and Pukkila (1993) indicate that this is an area that could provide useful results for practitioners.

There are many areas for further research, the asymptotic properties of MAICC bear further investigation, the comparison of IC with sequential stepwise procedures would be useful as would the evaluations of models with intercept and trend components. This is an area open to extensive exploitation, hopefully the research continues to the extent of unit root testing as stated by Muller and Elliot (2003), that the available procedures are near optimal.

REFERENCES

Akaike, H. (1969) "Fitting autoregressive models for prediction", *Annals of the Institute of Statistical Mathematics*, **20**: 243 – 247.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle, 267 – 281. In: *2nd International Symposium on Information Theory*. Eds. Petrov, B.N. & Csaki, F. Akademia Kiado, Budapest, Hungary.

Akaike, H. (1974) "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, **19**(6): 716 – 723.

Baltagi, B.H. & Wang, Z. (2007) "Testing for cointegrating rank via model selection evidence from 165 data sets", *Empirical Economics*, **33**: 41 – 49.

Barr, G. & Kantor, B. (2002) "The South African Economy and its asset markets", *The South African Journal of Economics*, **70**(1): 53 – 77.

Bedrick, E.J. & Tsai, C.L. (1994) "Model selection for multivariate regression in small samples", *Biometrics*, **50** (March): 226 – 231.

Bowerman, B.L. & O'Connell, R.T. (1993) *Forecasting and Time Series*: An applied approach. Duxbury Thomson Learning, Australia.

Box, G.E.P. & Pierce, D.A. (1970) "Distribution of residual autocorrelations in autoregressive integrated moving average time series models", *Journal of the American Statistical Association*, **65** (December): 1509 – 1526.

Box, G.E.P. & Jenkins, G.M. (1976) *Time Series Analysis, forecasting and control*. Holden-Day Inc., San Francisco, California.

Bozdogan, H. (2003) *Statistical Data Mining and Knowledge Discovery*. Chapman & Hall/CRC Press, Boca Raton, Florida, U.S.A.

Brockwell, P.J. & Davis, R.A. (1991) *Time Series Theory and Methods*, 2nd edition. Springer-Verlag, New York.

Brockwell, P.J. & Davis, R.A. (2002) *Introduction to Time Series and Forecasting*, 2nd edition. Springer-Verlag, New York, U.S.A.

Brüggemann, R. (2004) *Model Reduction Methods for Vector Autoregressive Methods*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, Heidelberg, Germany.

Brüggemann, R., Lütkepohl, H. & Saikkonen, P. (2006) "Residual autocorrelation testing for vector error correction models", *Journal of Econometrics*, **134**: 579 – 604.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference*: A Practical Information-Theoretic Approach, 2nd edition. Springer Science + Business Media, New York, U.S.A.

Cavanaugh, J.E. (1997) "Unifying the derivations for the Akaike and corrected Akaike information criteria", *Statistics and Probability Letters*, **33**: 201 – 208.

Cheung, Y.W. & Lai, K.S. (1993) "Finite-sample sizes of Johansen's likelihood ratio tests for cointegration", *Oxford Bulletin of Economics and Statistics*, **55**(3): 313 – 328.

Dalal, S.R., Fowlkes, E.B. & Hoadley, B. (1989) "Risk analysis of the space shuttle: Pre-Challenger prediction of failure", *Journal of the American Statistical Association*, **84** (December): 945 – 957.

Davidson, R. & MacKinnon, J.G. (2004) *Econometric Theory and Methods*. Oxford University Press, New York.

Demiralp, S., Hoover, K.D. & Perez, S.J. (2008) "A bootstrap method for identifying and evaluating a structural vector autoregression", *Oxford Bulletin of Economics and Statistics*, **70**(4): 509 – 533.

de Wet, T. (2000) "Purchasing power parity in the long run: An empirical re-evaluation of the South African evidence", *Studies in Economics and Econometrics*, **24**(2): 19 – 34.

Dickey, D.A. & Fuller, W.A. (1979) "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, **74** (June): 427 – 431.

Dickey, D.A. & Fuller, W.A. (1981) "Likelihood ratio statistics for autoregressive time series with a unit root", *Econometrica*, **49**(4): 1057 – 1072.

Enders, W. (2004) *Applied Econometric Time Series*, 2nd edition. John Wiley and Sons Inc. Hoboken, New Jersey, U.S.A.

Engle, R.F. & Granger, C.W.J. (1987) "Cointegration and error correction: representation, estimation and testing", *Econometrica*, **55**(2): 251 – 276.

Engle, R.F. & Granger, C.W.J. (1991) *Long-Run Economic Relationships*: Readings in Cointegration. Oxford University Press Inc., New York, U.S.A.

Fan, J. & Yao, Q. (2003) *Nonlinear Time Series, Nonparametric and Parametric Methods*. Springer – Verlag, New York, U.S.A.

Fedderke, J. & Joao, M. (2001) "Arbitrage, cointegration and efficiency in financial markets in the presence of financial crises", *The South African Journal of Economics*, **69**(3): 366 – 384.

Ferre, M. & Hall, S.G. (2002) "Foreign exchange market efficiency and cointegration, *Applied Financial Economics*, **12**: 131 – 139.

Gonzalo, J. & Pitarakis, J. (1998) "Specification via model selection in vector error correction models", *Economics Letters*, **60**: 321 – 328.

Gonzalo, J. & Pitarakis, J. (2002) "Lag length estimation in large dimensional systems", *Journal of Time Series Analysis*, **23**(4): 401 – 423.

Granger, C.W.J. & Newbold, P. (1974) "Spurious regressions in econometrics", *Journal of Econometrics*, **26**: 1045 – 1066.

Granger, C.W.J. & Weiss, A.A. (1983) Time series analysis of error-correction models, 255 – 278. In: *Studies in Econometrics, Time Series and Multivariate Statistics*. Eds. Karlin, S., Amemiya, T., Goodman, L.A. & Anderson, T.W. Academic Press, New York, U.S.A.

Gumede, V. (2000) "Import demand elasticities for South Africa: A cointegration analysis", *Studies in Economics and Econometrics*, **24**(1): 21 – 37.

Hafidi, B. & Mkhadri, A. (2006) "A corrected Akaike criterion based on Kullback's symmetric divergence: applications in time series, multiple regression and multivariate regression", *Computational Statistics & Data Analysis*, **50**: 1524 – 1550.

Hall, S.G. (1986) “An application of the Granger & Engle two-step procedure to United Kingdom aggregate wage data”, *Oxford Bulletin of Economics and Statistics*, **48**(3): 229 – 239.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press, Princeton, New Jersey, U.S.A.

Hannan, E.J. & Quinn, B.G. (1979) “The determination of the order of an autoregression”, *Journal of the Royal Statistical Society*, Series B, **41**(2): 190 – 195.

Harris, R. & Sollis, R. (2003) *Applied Time Series Modelling and Forecasting*. John Wiley & Sons Inc., Chichester, West Sussex, U.K.

Hurvich, C.M. & Tsai, C.L. (1989) “Regression and time series model selection in small samples”, *Biometrika*, **76**(2): 297 – 307.

Hurvich, C.M., Shumway, R. & Tsai, C.L. (1990) “Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples”, *Biometrika*, **77**(4): 709 – 719.

Hurvich, C.M. & Tsai, C.L. (1993) “A corrected Akaike information criterion for vector autoregressive model selection”, *Journal of Time Series Analysis*, **14**(3): 271 – 279.

Hurvich, C.M. & Tsai, C.L. (1995) “Model selection for extended quasi-likelihood models in small samples”, *Biometrics*, **51** (September): 1077 – 1084.

Ivanov, V. & Kilian, L. (2005) “A practitioner’s guide to lag order selection for VAR impulse response analysis”, *Studies in Nonlinear Dynamics & Econometrics*, **9**(1): 1 – 33.

Johansen, S. (1988) “Statistical analysis of cointegration vectors”, *Journal of Economic Dynamics and Control*, **12**: 231 – 254.

Johansen, S. (1991) “Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models”, *Econometrica*, **59**(6): 1551 – 1580.

Johansen, S. (1992) “Determination of cointegration rank in the presence of a linear trend”, *Oxford Bulletin of Economics and Statistics*, **54**(3): 383 – 397.

Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Auto-regressive Models*. Oxford University Press, New York, U.S.A.

Johansen, S. (2005) “Interpretation of cointegrating coefficients in the cointegrated vector autoregressive model”, *Oxford Bulletin of Economics and Statistics*, **67**(1): 93 – 104.

Johansen, S. & Juselius, K. (1992) “Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP for the UK”, *Journal of Econometrics*, **53**: 211 – 244.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. & Lee, T.C. (1988) *Introduction to the Theory and Practice of Econometrics*, 2nd edition. John Wiley & Sons Inc., U.S.A.

Kadilar, C. & Erdemir, C. (2002) “Comparison of performance among information criteria in VAR and seasonal VAR models”, *Hacettepe Journal of Mathematics and Statistics*, **31**: 127 – 137.

Kellard, N., Newbold, P. & Rayner, T. (2001) “Evaluating currency market efficiency: are cointegration tests appropriate?”, *Applied Financial Economics*, **11**: 681 – 691.

Kendall, M. & Ord, J.K. (1990) *Time Series*, 3rd edition. Oxford University Press, New York, U.S.A.

Kilian, L. (2001) “Impulse response analysis in vector autoregressions with unknown lag order”, *Journal of Forecasting*, **20**: 161 – 179.

Koreisha, S.G. & Pukkila, T. (1993) “Determining the order of a vector autoregression when the number of component series is large”, *Journal of Time Series Analysis*, **14**(1): 47 – 69.

Krolzig, H.M. (2003) “General-to-specific model selection procedures for structural vector autoregressions”, *Oxford Bulletin of Economics and Statistics*, **65** (supplement): 769 – 801.

Leeper, E.M. (1997) “Narrative and VAR approaches to monetary policy: Common identification problems”, *Journal of Monetary Economics*, **40**(3): 641 – 657.

Leng, H.M.J. (2002) “The South African share index futures and share markets: efficiency and causality revisited”, *Studies in Economics and Econometrics*, **26**(3): 1 – 18.

Liew, V.K-S. (2004) “Which lag length selection criteria should we employ?”, *Economics Bulletin*, **33**(3): 1 – 9.

Ljung, G.M. & Box, G.E.P. (1978) “On a measure of a lack of fit in time series models”, *Biometrika*, **65**: 297 – 303.

Lütkepohl, H. (1985) “Comparison of criteria for estimating the order of a vector autoregressive process”, *Journal of Time Series Analysis*, **6**(1): 35 – 52.

Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, Heidelberg, Germany.

MacKinnon, J.G. (1991) Critical values for cointegration tests, 267 – 276. In: *Long-run Economic Relationships: Readings in Cointegration*. Eds. Engle, R.F. & Granger, C.W.J. Oxford University Press, Oxford, U.K.

MacKinnon, J.G. (1996) "Numerical distribution functions for unit root and cointegration tests", *Journal of Applied Econometrics*, **11**: 601 – 618.

MacKinnon, J.G., Haug, A.A. & Michelis, L. (1999) "Numerical distribution functions of likelihood ratio tests for cointegration", *Journal of Applied Econometrics*, **14**: 563 – 577.

McQuarrie, A.D.R. & Tsai, C.L. (1998) *Regression and Time Series Model Selection*. World Scientific Publishing Co. Pty. Ltd., Singapore.

Mendenhall, W. & Sincich, T. (2003) *A Second Course in Statistics*: Regression Analysis, 6th edition. Pearson Education Inc., U.S.A.

Montgomery, D.C., Peck, E.A. & Vining, G.G. (2001) *Introduction to Linear Regression Analysis*, 3rd edition. Wiley-Interscience, Texas, U.S.A.

Muller, U.K. & Elliot, G. (2003) "Tests for unit roots and the initial condition", *Econometrica*, **71**(4): 1269 – 1286.

Ng, S. & Perron, P. (2001) "Lag length selection and the construction of unit root tests with good size and power", *Econometrica*, **69**(6): 1519 – 1554.

Ng, S. & Perron, P. (2005) "A note on the selection of time series models", *Oxford Bulletin of Economics and Statistics*, **67**(1): 115 – 134.

Pesaran, M.H. & Shin, Y. (1997) An autoregressive distributed lag modelling approach to cointegration analysis. Revised working paper presented at the Symposium at the

Centennial of Ragnar Frisch, The Norwegian Academy of Science and Letters, Oslo, March, 1995.

Phillips, P.C.B. & Perron, P. (1988) "Testing for a unit root in time series regression", *Biometrika*, **75**(2): 335 – 346.

Qu, Z. & Perron, P. (2006) A modified information criterion for cointegration tests based on VAR approximation. *Working paper series WP2006-011*, Department of Economics, Boston University, Boston, U.S.A.

Qu, Z. & Perron, P. (2007) "A modified information criterion for cointegration tests based on a VAR approximation", *Econometric Theory*, **23**: 638 – 685.

Quantitative Micro Software, LLC. (2004) *EViews 5, User's Guide*. Irvine, California, U.S.A.

Schwarz, G. (1978) "Estimating the dimension of a model", *The Annals of Statistics*, **6**(2): 461 – 464.

Shibata, R. (1980) "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process", *The Annals of Statistics*, **8**(1): 147 – 164.

Shibata, R. (1989) Statistical aspects of model selection, 215 – 240. In: *From Data to Model*. Ed. Willems, J.C. Springer-Verlag, Berlin, Heidelberg, Germany.

Shumway, R.H. & Stoffer, D.S. (2000) *Time Series Analysis and its Applications*. Springer-Verlag, New York, U.S.A.

Simonoff, J.S. & Tsai, C.L. (1999) "Semiparametric and additive model selection using an improved Akaike information criterion", *Journal of Computational and Graphical Statistics*, **8**(1): 22 – 40.

Stock, J.H. & Watson, M.W. (1988) “Testing for common trends”, *Journal of the American Statistical Association*, **83** (December): 1097 – 1107.

Stock, J.H. & Watson, M.W. (1991) Variable trends in economic time series, 17 – 15. In: *Long-run Economic Relationships: Readings in Cointegration*. Eds. Engle, R.F. & Granger, C.W.J. Oxford University Press, Oxford, U.K.

Stock, J.H. & Watson, M.W. (1993) “A simple estimator of cointegrating vectors in higher-order integrated systems”, *Econometrica*, **61**(4): 783 – 820.

Štulajter, F. (2002) Predictions in Time Series using Regression Models. Springer – Verlag, New York, U.S.A.

Viljoen, C. (2003) Price discovery, price behaviour, and efficiency of selected grain commodities traded on the agricultural products division of the JSE securities exchange. *Unpublished PhD Thesis*, Department of Economics and Economic History, Rhodes University, Grahamstown, South Africa.

Wei, W.W.S. (1990) *Time Series Analysis, Univariate and Multivariate Methods*. Addison-Wesley Publishing Company, U.S.A.

Yap, S.F. & Reinsel, G.C. (1995) “Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model”, *Journal of the American Statistical Association*, **90** (March): 253 – 267.

Zellner, A. (2003) Econometric and Statistical Data Mining, Prediction and Policy-Making, 57 – 78. In: *Statistical Data Mining and Knowledge Discovery*. Ed. Bozdogan, H. United States of America: Chapman & Hall/CRC Press, Boca Raton Florida, U.S.A.

Zhang, P. (1992) “On the distributional properties of model selection criteria”, *Journal of the American Statistical Association*, **87** (September): 732 – 737.

APPENDICES

APPENDIX 1: Mathematica 6 code

```
Ao={{1,0,0},{0,1,0},{0,0,1}}
 A1={{0.4 z,0.4 z,0.5 z},{0,z,0},{0,0,z}}
V1=Ao-A1
Det[V1]
 {{0.4 z,0.4 z,0.5 z},{0,z,0},{0,0,z}}
 {{1-0.4 z,-0.4 z,-0.5 z},{0,1-z,0},{0,0,1-z}}
 1-2.4 z+1.8 z^2-0.4 z^3
 Clear all

 all Clear
 Ao={{1,0,0},{0,1,0},{0,0,1}}
A1={{0.4 z,0.4 z,0.5 z},{0,0.8 z,0},{0,0,0.6 z}}
A2={{0.2 z^2,-0.2 z^2,0},{0,0.2 z^2,0},{0,0,0.4 z^2}}
V2=Ao-A1-A2
 {{1,0,0},{0,1,0},{0,0,1}}
 {{0.4 z,0.4 z,0.5 z},{0,0.8 z,0},{0,0,0.6 z}}
 {{0.2 z^2,-0.2 z^2,0},{0,0.2 z^2,0},{0,0,0.4 z^2}}
 {{1-0.4 z-0.2 z^2,-0.4 z+0.2 z^2,-0.5 z},{0,1-0.8 z-0.2
z^2,0},{0,0,1-0.6 z-0.4 z^2}}
 Det[V2]
 1-1.8 z+0.24 z^2+0.768 z^3-0.072 z^4-0.12 z^5-0.016 z^6
 Clear all

 all Clear
 Ao={{1,0,0},{0,1,0},{0,0,1}}
A1={{0.4 z,0.4 z,0.6 z},{0.2 z,0.8 z,0.15 z},{0,0,0.6 z}}
A2={{0.2 z^2,-0.2 z^2,0},{-0.1 z^2,0.15 z^2,0},{0,0,0.4 z^2}}
V2b=Ao-A1-A2
Det[V2b]
 {{1,0,0},{0,1,0},{0,0,1}}
 {{0.4 z,0.4 z,0.6 z},{0.2 z,0.8 z,0.15 z},{0,0,0.6 z}}
 {{0.2 z^2,-0.2 z^2,0},{-0.1 z^2,0.15 z^2,0},{0,0,0.4 z^2}}
 {{1-0.4 z-0.2 z^2,-0.4 z+0.2 z^2,-0.6 z},{-0.2 z+0.1 z^2,1-0.8
z-0.15 z^2,-0.15 z},{0,0,1-0.6 z-0.4 z^2}}
 1-1.8 z+0.21 z^2+0.846 z^3-0.126 z^4-0.126 z^5-0.004 z^6
 clear all

 all clear
 Ao={{1,0,0},{0,1,0},{0,0,1}}
A1={{0.4 z,0.4 z,0.5 z},{0,0.6 z,0},{0,0,0.4 z}}
A2={{0.2 z^2,-0.2 z^2,0},{0,0.2 z^2,0},{0,0,0.3 z^2}}
```

```
A3={{-0.2 z^3,0.15 z^3,0.1 z^3},{0,0.2 z^3,0},{0,0,0.3 z^3}}
V3=Ao-A1-A2-A3
Det[V3]
 {{1,0,0},{0,1,0},{0,0,1}}
 {{0.4 z,0.4 z,0.5 z},{0,0.6 z,0},{0,0,0.4 z}}
 {{0.2 z^2,-0.2 z^2,0},{0,0.2 z^2,0},{0,0,0.3 z^2}}
 {{-0.2 z^3,0.15 z^3,0.1 z^3},{0,0.2 z^3,0},{0,0,0.3 z^3}}
 {{1-0.4 z-0.2 z^2+0.2 z^3,-0.4 z+0.2 z^2-0.15 z^3,-0.5 z-0.1
z^3},{0,1-0.6 z-0.2 z^2-0.2 z^3,0},{0,0,1-0.4 z-0.3 z^2-0.3 z^3}}
 1-1.4 z-0.06 z^2+0.264 z^3+0.268 z^4-0.012 z^5-0.1 z^6+0.016
z^7+0.012 z^8+0.012 z^9
 clear all

 all clear
 Ao={{1,0,0},{0,1,0},{0,0,1}}
A1={{0.4 z,0.4 z,0.5 z},{0,0.5 z,0},{0,0,0.4 z}}
A2={{0.2 z^2,-0.2 z^2,0.1 z^2},{0,0.25 z^2,0},{0,0,0.3 z^2}}
A3={{-0.15 z^3,0.15 z^3,0.2 z^3},{0,0.15 z^3,0},{0,0,0.15
z^3}}
A4={{0.1 z^4,-0.1 z^4,0},{0,0.1 z^4,0},{0,0,0.15 z^4}}
V4=Ao-A1-A2-A3-A4
Det[V4]
 {{1,0,0},{0,1,0},{0,0,1}}
 {{0.4 z,0.4 z,0.5 z},{0,0.5 z,0},{0,0,0.4 z}}
 {{0.2 z^2,-0.2 z^2,0.1 z^2},{0,0.25 z^2,0},{0,0,0.3 z^2}}
 {{-0.15 z^3,0.15 z^3,0.2 z^3},{0,0.15 z^3,0},{0,0,0.15 z^3}}
 {{0.1 z^4,-0.1 z^4,0},{0,0.1 z^4,0},{0,0,0.15 z^4}}
 {{1-0.4 z-0.2 z^2+0.15 z^3-0.1 z^4,-0.4 z+0.2 z^2-0.15 z^3+0.1 z^4,-
0.5 z-0.1 z^2-0.2 z^3},{0,1-0.5 z-0.25 z^2-0.15 z^3-0.1
z^4,0},{0,0,1-0.4 z-0.3 z^2-0.15 z^3-0.15 z^4}}
 1-1.3 z-0.19 z^2+0.42 z^3-0.185 z^4+0.261 z^5+0.0465 z^6-0.039
z^7+0.015625 z^8-0.01975 z^9-0.006375 z^10-0.0015 z^11-0.0015 z^12
```

APPENDIX 2: R2.5.1 code

VAR(1)

```
> g <- polynomial(c(1, -2.4, 1.8, -0.4))
> solve(g)
[1] 1.0-0i 1.0+0i 2.5+0i
```

VAR(2)

```
> g <- polynomial(c(1, -1.8, 0.24, 0.768, -0.072, -0.12, -0.016))
> solve(g)
[1] -5.00000+0i -3.44949+0i -2.50000+0i  1.00000-0i  1.00000+0i  1.44949+0i
```

VAR(2b)

```
g <- polynomial(c(1, -1.8, 0.21, 0.846, -0.126, -0.126, -0.004))
> solve(g)
[1] -30.228948  -2.500000  -2.244751   1.000000   1.000000   1.473699
```

VAR(3)

```
> g <- polynomial(c(1, -1.4, -0.06, 0.264, 0.268, -0.012, -0.1, 0.016, 0.012, 0.012))
> solve(g)
[1] -1.757279+0.000000i -1.000000-2.000000i -1.000000+2.000000i
[4] -1.000000-1.527525i -1.000000+1.527525i  1.000000+0.000000i
[7]  1.000000+0.000000i  1.378639-0.971937i  1.378639+0.971937i
```

VAR(4)

```
> g <- polynomial(c(1, -1.3, -0.19, 0.42, -0.185, 0.261, 0.0465, -0.039, 0.015625, -0.01975, -0.006375, -
0.0015, -0.0015))
> solve(g)
 [1] -2.251747+0.000000i -1.817428+0.000000i -1.551331+0.000000i -
0.124127-2.103708i -0.124127+2.103708i
 [6] -0.091286-1.913075i -0.091286+1.913075i  0.794828-1.943788i
0.794828+1.943788i  1.000000+0.000000i
[11]  1.000000+0.000000i  1.461675+0.000000i
```

APPENDIX 3: EViews codes

The data and computer routines are given in the accompanying DVD.

APPENDIX 4: Routine analysis of replication x_{12t} and x_{13t}

The tabulated results are given in the accompanying DVD.

APPENDIX 5: Inferential routine of NMS replication

The tabulated results are given in the accompanying DVD.

APPENDIX 6.1: Tabulated frequencies for individual criterion for NMS database

The tabulated results are given in the accompanying DVD.

APPENDIX 6.2: NMS summary for individual criteria

The tabulated results are given in the accompanying DVD.

APPENDIX 7: Datasets for validation and example

Complete datasets for software validation are given on the accompanying DVD.

APPENDIX 8: VEC(*p*) complete assessments

The tabulated results are given in the accompanying DVD.