Model selection for cointegrated relationships in small samples

by

Wei He

Submitted in fulfillment of the requirements for the degree of

Magister Scientiae

in the

Faculty of Science

at the Nelson Mandela Metropolitan University

November 2008

Supervisor: Mr G. D. Sharp

Co-supervisor: Mr J. Hugo

Abstract

Vector autoregression models have become widely used research tools in the analysis of macroeconomic time series. Cointegrated techniques are an essential part of empirical macroeconomic research. They infer causal long-run relationships between nonstationary variables. In this study, six information criteria were reviewed and compared. The methods focused on determining the optimum information criteria for detecting the correct lag structure of a two-variable cointegrated process.

Key words

Unit root test, VAR model, Cointegration test, Information criteria

Contents

Chapter 3

Chapter 4

Chapter 5

Acknowledgements

I would like to express my gratitude and appreciation to the following persons, without whose help this study would not have been possible:

- My father and mother for their endless love and support.

- Mr Sharp, my supervisor, for his constant encouragement and professional attendance to detail. I am privileged to have been your student. My appreciation is extended to my co-supervisor, Mr Hugo, for his guidance and help.

- Cathy Logie, for the time spends proofreading this study and editing the language.

- My friends, thanks for believing in me and for your friendship.

Abbreviations list

Akaike information criterion	AIC
Augmented Dickey-Fuller test	ADF
Autoregressive	AR
Autoregressive distributed lag	ARDL
Autoregressive integrated moving average	ARIMA
Autoregressive moving average	ARMA
Bayesian information criterion	BIC
Cointegrated vector autoregressive	CVAR
Corrected Akaike's information criterion	AICc
Corrected Modified Akaike's information criterion	MAICc
Data generating process	DGP
Error-correction model	ECM
Final prediction error	FPE
Hannan-Quinn information criterion	HQ
"Improved" Akaike information criterion	AICi
Information criterion	IC
Independently and identically distributed	i.i.d
Kwaitowski, Phillips, Schmidt and Shin test	KPSS
Likelihood ratio test	LR
— 1% level of significance for each likelihood ratio test	LR1
— 5% level of significance for each likelihood ratio test	LR5
Linearized bi-exponential model	LinBiExp
Log-periodic power law	LPPL
Mean squared forecast errors	MSFE
Modified Akaike information criterion	MAIC

Modified AICc criterion	AICu
Modified Information Criteria	MIC
Monetary aggregate M1	M1
Monetary aggregate M2	M2
Monetary aggregate M3	M3
Moving average	MA
North American Free Trade Agreement	NAFTA
Ordinary least squares	OLS
Partial autocorrelations	PAC
Posterior Information Criterion	PIC
Phillips and Perron test	PP
Purchasing power parity	PPP
Schwarz Bayesian criterion	SBC
Schwarz information criterion	SIC
Sequential Likelihood Ratio	SLR
South African	SA
Stock-Recruitment	SR
The C_p of Hosmer et al. (1989) variable selection criterion	C_p
The criterion of autoregressive transfer function	CAT
The C_p^* of Pregibon (1979) variable selection criterion	C_p^*
The Shibata (1980) variable selection criterion	Shibata
United States of America	USA
Vector autoregressive	VAR
Vector error correction	VEC

Chapter 1

Introduction

Global economic growth has resulted in an increasing number of applied economists paying attention to the analysis of economic data. The data under consideration for this dissertation is commonly referred to as time series data. Typical time series analysis requires that the process is a stationary process. In practice, economic time series data is not always stationary, but methods have been proposed to correct the data in such a way that an easy analysis is possible. The simplest method is simply differencing the data to remove the nonstationary component. Given a system of equations, an alternative method of analysis for nonstationary data is possible provided the system is cointegrated. If multiple series are cointegrated, they have long-run properties. This long-run property provides a practical interpretation for macroeconomists and statisticians to analyse nonstationary data.

The difference between a stationary and nonstationary time series is whether or not the series has a unit root. The presence of a unit root in a series has an important impact on the understanding of economic scenarios. For example, financial market variables such as future contracts, stock prices (Kim, 2003), dividends (Engsted, 2006), spot and forward exchange rates (Zhang and Lowinger, 2005), and even aggregate series like real consumption (Phillips and Perron, 1988), due to their unit root status, are analysed in a cointegrated framework rather than by using single equation nonstationary models.

There are several methods to test if a series has a unit root; some of the most common methods include the Dickey-Fuller test (Dickey and Fuller, 1979), the Augmented Dickey-Fuller (ADF) test, Kwaitowski, Phillips, Schmidt and Shin (KPSS) test and

the Phillips-Perron (PP) test. These methods are routine in econometric software packages such as EViews and Stata. The most common test for detecting the presence of unit roots in applied work is the ADF test.

Much of the econometric literature of the last two decades has focused on testing the unit root hypothesis in economic time series. The characterization of a real exchange rate series as random in nature has been questioned in recent times by the application of new statistical tools. Several papers on exchange rate studies have reported evidence that the real exchange rates follow a unit root process. Assaf (2006) demonstrated the failure of purchasing power parity (PPP) in the presence of regime shifts in the series and provided new evidence on the stationarity of bilateral real exchange rates, after allowing for regime changes. An ADF type test for unit roots was applied to real exchange rate series by allowing for a level shift in the data generating process (DGP).

Cointegration has become an important concept in time series econometrics since the publication of papers by Granger (1983) and Engle and Granger (1987). The cointegrating relations of nonstationary series are used as a tool for discussing the existence of long-run economic relations. Engle and Granger (1987) presented a representation theorem based on Granger (1983), which connected the moving average (MA), autoregressive (AR) and error correction representations for cointegrated systems.

The analysis of cointegrated series may have been defined by Engle and Granger (1987) but the standard method of analysis was formalized by Soren Johansen. The likelihood method, as provided by Johansen (1988) and Johansen and Juselius (1990), applies maximum likelihood estimation to the vector autoregressive (VAR) model, and considers the linear relationship among two or more nonstationary series. The method not only provides estimates of the model parameters and their standard errors, but also determines the numbers of cointegration vectors in the model. Johansen's

method is available in econometric software packages such as Shazam, Cats for Rats and EViews (Liu, 2007).

There are many applications of cointegration that have been published in econometric journals, conference proceedings and academic texts. Typical journals with papers include the "Journal for Studies in Economics and Econometrics", "Studies in Nonlinear Dynamics and Econometrics" and "Econometrica" which contain many papers where the method is applied and interpreted. The text books by Harris (1995) and Lütkepohl (2005) provide detailed discussions on the methods for the application of cointegration and the analysis of multi-equation nonstationary series.

One drawback of the cointegration analysis method is the fact that it relies on the pre-determined lag length of the cointegrated model prior to estimation and interpretation. Several methods regarding lag length selection have been proposed. The aim of this dissertation's is to analyse existing methods in an attempt to provide clarity on the most appropriate selection technique for a two-variable model.

In this study the literature and the econometric theory is reviewed as a background for the analysis in the forthcoming chapters. Chapter 2 provides the historical literature review. Chapter 3 discusses the methodology of the study whilst chapter 4 discusses model selection. In chapter 5 the parameter choice, sample size and theory of the simulation models are shown whilst chapter 6 shows the results of the simulation exercise. Chapter 7 concludes and discusses an opportunity for future research.

Chapter 2

Historical Literature Review

In recent years researchers have focused on the study of cointegrated vector autoregressive (CVAR) models. The link between economic theory and empirical modelling lies in the field of statistical inference. From a statistical argument, the VAR models and the vector error correction (VEC) models offer a number of advantages in the general framework for addressing empirical questions in macroeconomics. VAR models are widely used in analyses and economic forecasting.

Lima, Luduvice and Tabak (2004) studied a VAR/VEC model to forecast medium-term interest rates for the Brazilian economy. The empirical results suggested that VAR/VEC models performed poorly in forecasting movements in interest rates in terms of forecasting accuracy, but these models were helpful in predicting the direction of movements of medium-term interest rates. It seemed that the use of VAR/VEC models in monetary policy modeling could help policymakers to create scenarios for the term structure of interest rates in the future.

Johansen and Swensen (1999) considered the testing of restrictions implied by the rational expectations hypothesis in a CVAR model for $I(1)$ variables. They developed the likelihood ratio (LR) tests when the restrictions implied by the rational expectations hypothesis had one-step-ahead observations only and the coefficients were known, and explicit parameterization of the restrictions was found. They also provided an application to a present value model.

An important part in the analysis of VAR models is the determination of the lag length of the VAR model. The majority of papers reviewed indicated that the lag length of the VAR model is determined using an information criterion. In this study, six well

known information criteria (IC) were used in order to determine the optimum IC for detecting the correct lag structure of a two-variable cointegrated process. These are Akaike information criterion (AIC), Corrected Akaike's information criterion (AICc), Schwarz information criterion (SIC), Hannan-Quinn information criterion (HQ), Modified Akaike's information criterion (MAIC) and Corrected modified Akaike's information criterion (MAICc).

Gonzalo and Pitarakis (1998) identified both the theoretical and applied properties of the model selection problem based on the estimation of the cointegrating rank of a multivariate time series model. They found that the model selection procedure improves the estimation of the process for under parameterised CVAR models.

The information criteria mentioned previously have been used by several researchers. In butterfly studies, performed by Pickens (2007) using a model selected with AIC, the results indicated that both sites were primarily influenced by density dependence during the summer period and by weather conditions during the winter period. In the engineering field, Lee (2007) developed a dynamic prediction model of a highway pavement contractor's quality-based performance using a panel data analysis. The panel data's modeling performance was measured by AIC. In Mathematical Biosciences, Peter (2007) showed that model selection criteria such as AIC and SIC indicate the superiority of the bilinear model compared to adequate less parameterised alternatives such as linear, parabolic, exponential growth or classical growth models.

Sin and White (1996) considered penalized likelihood criteria such as AIC, SIC and HQ for selecting models of dependent processes. In particular, AIC and SIC frequently determined the correct order of an autoregressive process before an ADF test was performed for the DGP's of nonstationary AR (k) processes.

Wang and Liu (2006) used the maximum likelihood method to fit the six statistical fish stock-recruitment (SR) models on six sets of simulated SR data. They concluded

that the AIC and Bayesian information criterion (BIC) were both robust in selecting the most suitable SR relationships.

Although the more well known IC are accepted universally, new model selection criteria continue to be developed and in many cases improve on the existing IC. Although these new IC often yield better results for specific cases, however, the effectiveness of the new IC often need further testing under different conditions. Based on the work of Hurvich, Shumway and Tsai (1990), a new model selection criterion was developed by Bengtsson and Cavanaugh (2006) for the linear state-space model. Their simulation results showed that the "improved" Akaike information criterion (AICi) performed effectively as a model selection criterion in a small sample setting. They also indicated that AICi estimated the Kullback-Leibler information with less bias than traditional AIC and AICc.

When building good models, an important aspect that must be considered is the appropriate model size. To address this problem, several information criteria have been proposed. These methods are covered extensively in literature. In this study, the aim was to compare some of the well known information criteria using simulation analysis for cointegrated bivariate models in small sample cases.

2.1 Review of selected publications using unit root tests

In time series models, the methods of testing for the presence of a unit root have been published extensively in the econometrics literature. The Dickey-Fuller test, which was originally developed for autoregressive representation of known order, remains valid for a general ARIMA $(p,1,q)$ (autoregressive integrated moving average) process in which p and q are of unknown orders. The ADF examines the null hypothesis of an ARIMA$(p,1,0)$ process against the stationary ARIMA$(p+1,0,0)$ alternative. An alternative procedure for testing for the presence of a unit root was proposed by Phillips and Perron (1988). This approach is non-parametric with respect to nuisance parameters and allows for a wide range of time series models in which there are unit roots.

The ADF test is the most common approach used to test for the stationarity of a series. Lopez (1997) analyzed the asymptotic distribution of the ADF t-statistic in autoregressive moving average (ARMA) models under the alternative hypothesis. The main result of this paper stated that the asymptotic power of the ADF test in the ARMA case was not affected by the distance between the null and the alternative hypothesis and just depended on the ratio of the sample size T and the order of the autoregression p.

Narayan and Smyth (2007) provided evidence on the random walk hypothesis in stock price indices of the G7 countries. Their interpretations were supported by testing for the presence of a unit root in stock prices using the ADF and the PP tests. They used unit root tests which allowed for structural breaks in the trend component. Their economic interpretation was that the second break in stock price had a detrimental effect on stock prices movements for the G7 countries.

In order to identify the effect of the variability of money growth on the velocity of circulation of money, Baliamoune-Lutz and Haughton (2004) tested Friedman's hypothesis that increased variability in the growth of money supply caused velocity to decline. In their study they used the PP unit root test to determine the order of integration for an Egyptian dataset. Their study progressed to test for cointegration which is discussed in sub-section 2.3.

Aggarwal and Kyaw (2005) examined integration of the three participating equity markets before and after the passing of North American Free Trade Agreement (NAFTA) in 1993. They considered daily, weekly, and monthly data. To support their empirical analysis, they used three unit root procedures, the ADF test, the PP test and the KPSS test. The unit root tests indicated that stock prices were nonstationary but stock returns were generally stationary for all three markets for all three periods. They followed this analysis with tests for cointegation, the discussion of which is included in sub-section 2.3.

For the analysis of joint confirmation of the integration order, new critical values need to be calculated for unit autoregressive root and unit moving average root tests, instead of conventionally used separate critical values. The ADF–KPSS test of the joint confirmation hypothesis of unit autoregressive roots was employed by Keblowski and Welfe (2004). They calculated the critical values of the test for small samples and investigated the power of the test, including the case of structural change. The results led to the conclusion that detecting a potential structural break and inclusion of appropriate dummies in models of statistics could be treated as a successful strategy in inference on unit autoregressive roots when the ADF–KPSS test was applied.

Many unit root tests are based on a parametric model, for example, the Dickey–Fuller test. An alternative approach was suggested by Phillips and Perron (1988). They proposed nonparametric tests for detecting the presence of a unit root in quite general

time series models. The tests accommodated models with a fitted drift and a time trend so that they might be used to discriminate between unit root nonstationarity and stationarity about a deterministic trend.

2.2 Review of selected publications using vector auto-regressive models and vector error correction models

During the last two decades VAR models have become important research tools in multivariate time series analysis. VAR models are easy to estimate and can account for relatively complex dynamic phenomena. VAR models have been used to model important features and applications, such as forecasting, causality analysis, impulse responses and cointegration. The cointegrated approach can be used to estimate long-run relationships among nonstationary variables. If the dynamic models are presented as long-run relations, it is often possible and desirable to convert the dynamic models into an error-correction model (ECM). These ECMs contain information on both the short-run and long-run properties of the model. The VEC models specify the short-run dynamics of each variable in the system, and in a framework that anchors the dynamics to long-run equilibrium relationships suggested by economic theory.

Following the analysis of Johansen's (1991, 1995) cointegrated systems, Pesaran, Shin, and Smith (2000) generalized the existing cointegration analysis literature in two respects. Firstly, the problem of efficient estimation of VEC models containing exogenous $I(1)$ variables was examined. Secondly, efficient estimation of VEC models when the short-run dynamics might differ within and between equations were also considered.

The standard autoregressive distributed lag (ARDL) model is widely used in estimating energy demand relationships in a time-series context. Bentzen and Engsted (2001) applied the ARDL approach to estimate a demand relationship for Danish

residential energy consumption, and the ARDL estimates were compared to the estimates obtained using cointegration techniques and ECMs. It was concluded that both quantitatively and qualitatively, the ARDL approach and the cointegration/ ECM approach gave very similar results.

In order to illustrate how conflicting results can be obtained by applying different testing approaches in small samples with bivariate models, Zachariadis (2007) tested for the existence of causality between energy and economy variable pairs in the G7 countries, using aggregate and sectoral data and three different modern econometric methods to test for Granger causality: a VEC model, an ARDL model and a VAR model. The results, underlined the importance of utilising as large a sample size as possible and the use of multivariate models, which were closer to economic theory. The econometric interpretations accommodated several mechanisms and causality channels and provided a better representation of real-world interactions between energy use and economic growth.

Hecq, Palm and Urbain (2006) considered common cyclical features in an n-dimensional CVAR model. They discussed two specific reduced rank VEC models, namely “strong form” and “weak form”. They also suggested asymptotic tests for each form and examined the small sample properties of these tests by Monte Carlo simulations.

Cubadda (2007) offered an approach for simultaneously modelling differing forms of common cyclical features among $I(1)$ time series. Some iterative procedures were proposed for testing and imposing diverse forms of common features upon a CVAR model. The empirical application revealed that the new methods provided a model of the US business cycle indicators that was considerably more parsimonious than those obtained using the pre-existing concepts of common features.

2.3 Review of selected publications using cointegration

The publication of papers by Granger (1983) and Engle and Granger (1987), introducing the concept of cointegrated nonstationary time series variables has generated huge volumes of empirical research in econometric journals. Cointegration implies that certain linear combinations of nonstationary variables are integrated of lower order than the process itself.

Since Engle and Granger (1987), the notion of cointegration has become one of the more important concepts in time series econometrics. The most widely used test in econometric applications of cointegrated series is the LR test of Johansen (1991) which is based on the rank of a cointegrated matrix using parametric estimation of a VAR model.

Shintani (2001) proposed a full nonparametric test for the cointegrated rank. The new test did not require estimation of a VAR model. The test exploited the well-known fact that the number of cointegrating vectors was identical to the degree of degeneration in the space spanned by the sample moment matrix in the limit. With an appropriate standardization, the test statistics were shown to have a nuisance parameter free limiting distribution and to be consistent under reasonable conditions. Monte Carlo experiments suggested that the finite sample performance of the test was satisfactory.

Selected examples of empirical research of cointegration analysis include the work of Baliamoune-Lutz and Haughton (2004) and Aggarwal and Kyaw (2005) which were mentioned in sub-section 2.2. Both papers use cointegration analysis to support their empirical analysis. Baliamoune-Lutz and Haughton (2004) determined that the variable's velocity and anticipated as well as unanticipated variability in the money supply were cointegrated, and that the variability of money growth drove velocity in the case of monetary aggregate M2 (M2) but not in the case of monetary aggregate

M1 (M1). Aggarwal and Kyaw (2005) showed that daily, weekly and monthly equity prices in the three NAFTA countries were cointegrated only for the post-NAFTA period. There seemed to be greater financial integration after NAFTA resulting in less opportunity for international portfolio diversification during the post-1994 period. The results had interesting implications for investors, managers, policymakers and scholars interested in portfolio diversification for NAFTA countries.

Kellard (2006) examined whether the conventional test of the order of integration of the forecast error was subject to serious bias. The ADF unit root test, the Engle and Granger test (1987) and the reduced rank method of Johansen (1995) were applied to simulated data. Johansen tests were found to be robust to differences in variable magnitudes and it was suggested for market efficiency assessment.

Engsted (2006) used Johansen's CVAR model to analyze stock prices and dividends in the US for the period 1871 to 2000. He found evidence that prices and dividends had a common $I(1)$ trend, meaning that they were cointegrated.

The South African (SA) monetary aggregates M3 (M3) demand consisting of real money, the opportunity cost of holding money and the inflation rate was analysed by Todani (2005). The model using quarterly data for the period 1980 to 2003 was estimated using CVAR. Todani (2005) concluded that notwithstanding the economic developments that had taken place in South Africa, which included financial liberalization and integration into the world economy, opening up of the economy and so forth, a stable long-run cointegrated money demand relation was still prevalent in the SA economy.

Pelipas (2006) analyzed the demand for nominal and real money balances in Belarus on the basis of the quarterly data from 1992 to 2003. He used cointegration analysis and derived stable long-run and short-run money demand functions. From the

cointegration analysis, it was shown that there existed a long-run function for nominal money balances and real money balances. Long-run demand for nominal money was determined by consumer prices, real industrial production, nominal exchange rates and the refinancing rate.

The development of forecasting models for short- and long-term interest rates has been studied for many years. The link between short- and medium-term interest rates is given by the Null Hypothesis which suggests that monetary policy affects medium-term interest rates by directly influencing short-term rates and by altering market expectations of future short-term rates. Engsted and Tanggaard (1994) used Johansen's cointegration analysis to test for the cointegration implications of the Null Hypothesis on a sample of the United States (US) discount yields and evidence was found suggesting that cointegration implications of the Null Hypothesis generally seemed to hold.

Exchange rate studies are receiving more and more attention in the applied field. Kim (2003) used Johansen's cointegration method to investigate the existence of short-run dynamics and long-run equilibrium relations between the aggregate stock price and relevant macroeconomic variables. They concluded that the stock price has long-run equilibrium relationships with selective determinant variables. To be more specific, the stock price was positively related to industrial production whilst being negatively related to interest rates, exchange rates and inflation.

The monetary model of exchange rate determination is a useful tool for understanding fluctuations in exchange rates. The model suggests the existence of a strong link between nominal exchange rates and a set of macroeconomic fundamentals (i.e. real income, money supply, interest rate and the inflation rate). Cointegration methods are often used in testing the monetary model of exchange rate determination. As an example, Zhang and Lowinger (2005) applied Johansen's cointegration methodology to test the long-run relationship between the exchange rate and certain

macroeconomic variables. Their study used quarterly data from Germany, Japan, the United Kingdom and the United States of America for the period 1973 to 1999. In an earlier study, Diamandis, Georgoutsos and Kouretas (1996) examined the Canadian and United States dollar exchange rate for the duration of the floating exchange rate period. By using the multivariate cointegration methodology, it was shown that an unrestricted monetary model provided a valid explanation of the long-run nominal Canadian and the United States dollar exchange rate.

Growth studies on whether or not permanent changes in economic fundamentals affect the long-run growth rate is of major empirical interest. Lau (2007) observed that the estimated coefficients of the error-correction terms and the associated t-statistics for the models with different lags remained relatively unchanged in France, Japan and the United Kingdom. He also showed that the natural log per capita output and the natural log per capita investment were cointegrated.

2.4 Review of selected publications using Akaike information criterion

The application of the AIC is widespread as can be seen from the diversity of papers reviewed below. The papers reviewed represent only a sample from many publications and provide an introduction to the benefits of IC methods.

In a study on the population dynamics of a butterfly species, Pickens (2007) demonstrated the effects of density-dependent factors and weather on separate life stages to identify critical life-history parameters. The AIC was used to lend support to the modelling results indicating that sites were primarily influenced by density dependence during the summer period and by weather conditions during the winter period. In a similar zoological study, Davros, Debinski, Reeder and Hohman (2006) used multiple regressions and AIC to model habitat width and landscape-level factors which influenced the abundance and diversity of the butterfly community in south-

western Minnesota, USA.

The use of AIC was not only restricted to the biological discipline; Lee (2007) developed a dynamic prediction model of a highway pavement contractor's quality-based performance using panel data analysis. Several random effects models were first developed using in-sample specification. The AIC was then used to measure the random effects models modeling performances. Out of sample specifications validated the developed random effects models by comparing out of sample forecasting accuracies. Conclusions reached showed that the asphaltic concrete pavement quality of construction could be predicted based on the contractor's past quality-based performance as well as other construction parameters.

Zhou and Sornette (2006) presented a general methodology to incorporate fundamental economic factors into the theory of herding, which was developed as part of their work to describe bubbles and antibubbles. They considered five factors. These factors were interest rate, interest spread, historical volatility, implied volatility and exchange rates. Standard AIC and Wilks tests were used to compare the five different proposed factor models with the first-order log-periodic power law (LPPL) and second-order LPPL.

2.5 Review of selected publications using Scharwz information criterion

Swanson (1998) used a variety of rolling window and increasing window estimation techniques to investigate the extent to which fluctuations in money stock anticipate fluctuations in real output. They also examined the impact of earlier findings on specifying the number of lags in regression models by using model selection criteria. Based on their empirical analysis, the finding indicated that the relation between income, money, prices and interest rates were stable, as long as sufficient data were used. The systems were particularly stable when the SIC was used to select the lag

order.

The identification of phone boundaries in continuous speech is an important problem in areas of speech synthesis and recognition. Almpanidis and Kotropoulos (2008) presented a text-independent automatic phone segmentation algorithm based on the BIC which was mentioned in sub-section 4.3. BIC as a model selection criterion was used to undertake the decision, as to whether the data in the large segment were more accurately described by a single Gaussian or a two-segment representation.

2.6 Review of selected publications using Corrected Akaike's information criterion

Hurvich and Tsai (1989) derived a small sample bias correction criterion, AICc, for the selection of extended quasi-likelihood models. Hurvich and Tsai (1995) used AICc for the selection of extended quasi-likelihood models. Their results showed that AIC, C_p (C_p variable selection criterion) of Hosmer, Jovanovic and Lemeshow (1989) and C_p^* of Pregibon (1979) tended to over fit the model. These authors concluded that as the dimension of the candidate model increased in comparison to the sample size, the three criteria became negatively biased estimates of the expected Kullback-Leibler information and that the AICc provided better model selections in all cases.

AIC and AICc are both designed as estimators of the expected Kullback-Leibler discrepancy between the model generating the data and a fitted candidate model. Cavanaugh (1997) mentioned that the derivations of the AIC and AICc proceeded along very different lines and that it was difficult to reconcile how AICc improved upon the approximations leading to AIC. The author presented a derivation which unifies the justifications of AIC and AICc in the linear regression framework.

A bias-corrected version of AIC, AICc and AICu (modified AICc criterion) was

applied to causal models by Lee and Tsai (1998). They showed that the performance of a global procedure with four selection criteria AIC, AICc, AICu and FPE (final prediction error) could be easily explored by means of simulations. It was concluded that AICc and AICu provided better model order choices than AIC and FPE. They recommended using both AICc and AICu, with the global selection procedure, to select the appropriate models.

2.7 Review of selected publications using Modified Akaike information criterion

Ng and Perron (2001) derived a class of Modified Information Criteria (MIC) with a penalty factor that was sample dependent. They motivated that the IC was more robust when there were negative moving-average errors.

Qu and Perron (2006) extended the work of Ng and Perron (2001) and suggested a modified Akaike information criteria, MAIC. The results showed that MAIC was more powerful when there were moving average errors negatively correlated and with a finite VAR system.

2.8 Review of selected publications using more than one information criterion

Granger and Jeon (2004) compared the forecasting performance of the AIC and the BIC. It was found that on average, BIC generally fitted more parsimonious AR models and out-forecasted AIC for progressively updated models. The forecasting performance of combining AR(4), AIC and BIC were similar. To summarise, BIC performed better than AIC in the sense of minimizing the mean squared forecast errors (MSFE).

Ng and Perron (2005) considered model selection of an autoregression order using IC.

Their simulation results showed that the lag lengths selected by both AIC and SIC were sensitive to model parameters in finite samples. The methods that gave the most precise estimation were those that held the effective sample size fixed across all models to be compared.

Wang and Bessler (2005) conducted Monte Carlo simulations to evaluate the use of AIC and SIC as an alternative to various probability-based tests for determining the cointegrating rank in multivariate analysis. Results showed that AIC had an advantage over trace tests for cointegrated models with small samples, but did not perform as well with large samples. SIC showed better large sample results than both the AIC and the trace test, even when the series were close to nonstationarity or when large negative moving average components were contained.

Granger and Jeon (2006) considered autoregressive models fitted to 215 US macro series, with lags chosen by either the BIC or AIC. Their results indicated that BIC out performed AIC in both univariate and bivariate models as the estimated parameters for the AIC fitted model gave spurious roots.

Peter (2007) used both AIC and BIC to support a linearized bi-exponential model (LinBiExp) to describe bilinear-type data. The LinBiExp model proposed was preferred to less parameterised alternatives such as the linear, the parabolic, the exponential or the classical growth (e.g. logistic, Gompertz, Weibull, and Richards) models. The proposed LinBiExp model was a versatile and useful five-parameter bilinear functional form that was convenient to implement, was suitable for full optimization and provided intuitive and easily interpretable parameters.

Gonzalo and Pitarakis (2002) used model selection criteria such as AIC, SIC, HQ and a LR test based on an general to specific testing strategies for lag length estimation in VAR models. They showed that the model selection criteria for choosing an optimal lag length could be extremely sensitive to factors such as the system dimension and

the preset upper bound. Their results strongly pointed in favor of an AIC-based approach for selecting the lag length in large dimensional systems.

In an autoregressive simulation study by Liew (2004), five criteria AIC, SIC, HQ, FPE and BIC were assessed. The probability of selecting the correct simulated model is computed for several sample sizes and results were compared. It was found that the five criteria correctly selected the lag length at least half of the time in small samples, whilst this performance increased substantially as the sample size increased. In large sized samples ($T \geq 120$), HQ was found to outperform the rest in correctly identifying the true lag length. In contrast, AIC and FPE were superior to other criteria in the case of small samples ($T \leq 60$) and minimized the chance of under estimation. The problem of over estimation was negligible in all cases.

Lütkepohl (1985) conducted an extensive Monte Carlo study analysing the properties of selection criteria in bivariate and trivariate stationary VAR models. The study compared twelve different criteria for estimating the order of an AR process in small samples. The results showed that for small samples, SIC estimated the order correctly more often than the other criteria assessed. As the sample size increased, the selection capability of the HQ criterion improved. The results also showed that CAT (the criterion of autoregressive transfer function) by Parzen (1974), FPE, AIC, and Shibata (the Shibata criterion) of Shibata (1980) had a tendency to overestimate the AR lag order. For a large sample, all four criteria provided very similar results whilst LR5 (5% level of significance for each likelihood ratio test), LR1 (1% level of significance for each likelihood ratio test) and PAC (partial autocorrelations) provided poor forecasts for small samples. The study concluded that for finite order AR processes none of the criteria studied had a strong tendency for underestimating the true order.

Ivanov and Kilian (2005) compared six selection criteria commonly used in applied work and provided recommendations about selecting the lag order to construct accurate impulse response estimates. The six criteria were AIC, SIC, HQ, the general-

to-specific sequential Likelihood Ratio (SLR) test, a small-sample correction to that SLR test and the specific-to-general sequential Portmanteau test. It was noted that the results differed if the objective of estimating the VAR model was forecasting or the construction of variance decompositions. The results showed that for monthly VAR models, the AIC had more accurate estimates for realistic sample sizes. For quarterly VAR models, the HQ appeared to be the most accurate criterion with exception of sample sizes smaller than 120, for which the SIC was more accurate. For persistence profiles based on quarterly VEC models with known cointegrating vector, SIC was the most accurate criterion for all realistic sample sizes.

Koreisha and Pukkila (1990) contrasted the performance of various criteria used for identifying the order of VAR processes when the component series was large. Their simulation results showed that the performances of AIC, BIC and HQ were highly dependent on the number of nonzero elements in the polynomial matrices of the autoregressive parameters and the permitted upper limit of the order used in testing the autoregressive structure. They proposed a new approach which was less dependent on component size and on the number of nonzero values in the parameter matrices of the model.

Chao and Phillips (1999) derived the Posterior Information Criterion (PIC) for partially nonstationary VAR processes with reduced rank structure. In a Monte Carlo study, they evaluated the finite sample performance of PIC relative to the existing model selection procedures, AIC and BIC. The results showed that PIC performed well and in some cases even better than the more established methods.

Baltagi and Wang (2006) re-examined 165 published data sets and used the model selection approach in testing for cointegration. They found that AIC, HQ, SIC and PIC provided similar results on cointegrating relationships. This confirmed the results obtained in the original studies in about 50% of the cases. A low percentage of correspondence between different testing procedures suggested that caution should be

taken when testing and interpreting cointegrating relationships. The empirical results indicated that the model selection approach (especially SIC and PIC) could be a useful complement to the widely used parametric tests in cointegration analysis for applied researchers.

Chapter 3

Methodology

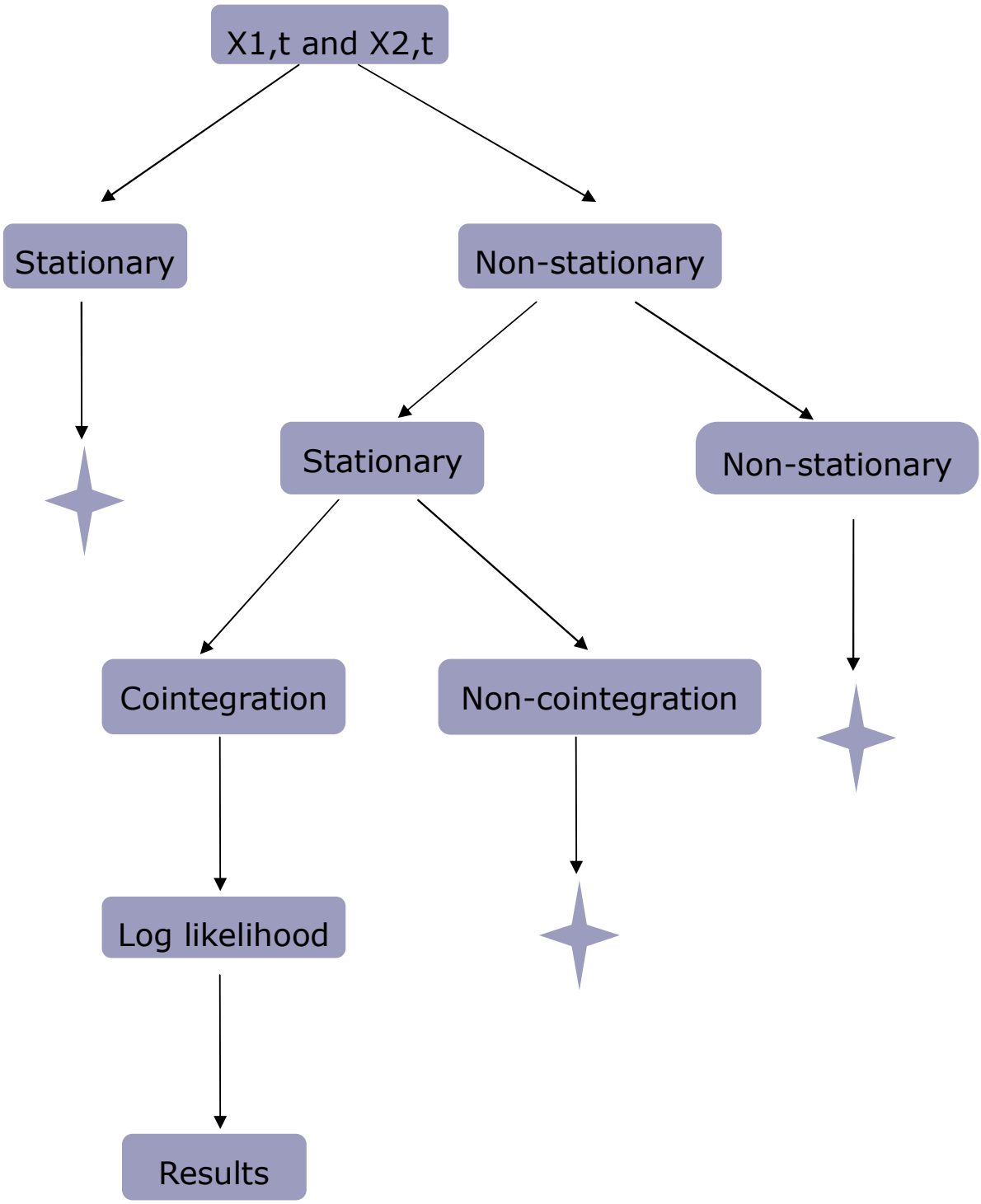
The objective of this study was to determine the model selection criterion which most correctly estimated the true lag for VEC models. The methodology of this study was based on goodness of fit assessment of theoretical nonstationary bivariate VAR models of varying length and parameter starting values. Models were simulated and the percentage of correctly selected models for each of six information criteria was recorded and inferences based on these results were then made. Four lag length bivariate models were assessed, starting with the simplest model from lag length one and progressing to the more complex model with lag length four.

The models were analysed sequentially. The ADF test was run to determine the order of integration for each series in the bivariate cointegrated models. This was followed by running Johansen's trace test for cointegration relationships. The results were then compared for those models that met all theoretical specifications and for those that deviated from theoretical specifications.

In this study, the lag length terms were restricted to four, as models with longer lags tend to be more difficult to estimate and are not that common in practice. Simulated data for each lag length model was compared for sample sizes of 30 with increments of 10 up to and including 100. In total therefore, for each model and sample of size n (n = 30, 40, …, 100), 5100 series were simulated. In each case, the initial value y_0 was set to zero. To minimise the influence of adverse starting values, the first 300 observations from each series were discarded. In all cases, these series were simulated using the random number generator in Excel and were recorded on the accompanying DVDs.

The flow chart, shown in Figure 3.1, reflects the procedure that was followed to determine whether the simulated bivariate VAR model was cointegrated.

Figure 3.1: Analysis flow chart



3.1 Theory

The time sequence $\{y_t\}$ is a weakly stationary series if the mean, variance and autocorrelations are time invariant. This process has a finite mean and variance and is covariance weakly stationary for all t and $t-s$ if

$$E(y_t) = E(y_{t-s}) = \mu,$$

$$\text{var}(y_t) = \text{var}(y_{t-s}) = \sigma_y^2,$$

$$\text{cov}(y_t, y_{t-s}) = \text{cov}(y_{t-j}, y_{t-j-s}) = \gamma_s,$$

where μ, σ_y^2 and γ_s are all constants (Enders, 2004).

On the contrary, a nonstationary time series implies that the mean, variance and/or the covariance of the series change over time. The easiest way to analyse a nonstationary time series is to convert the series to a stationary series and apply the Box-Jenkins methodology. The series is converted by differencing until stationary. If a series differenced d times before it becomes stationary, then the series is defined to be integrated of order d and denoted as $I(d)$. This means that the series contains d unit roots. As an illustration, a series with d =1 means that the process is stationary after differencing once. This is denoted as $I(1)$ and means that the series contains one unit root.

3.2 Unit root test

When a regression model is estimated using least squares and the data come from a nonstationary time series, then the regression is said to be spurious. To address this problem, prior to estimation, the data is tested for the presence of unit roots, i.e. whether or not the series is stationary. There are several ways of detecting the presence of a unit root. Some of the most common methods include the Dickey-Fuller test, ADF test, KPSS test and the PP test. From literature it is clear that the most commonly used method is the ADF test. For this reason, the study adopted the same approach throughout the unit root analysis procedure.

3.2.1 Dickey-Fuller Test

The simplest nonstationary time series is known as a random walk. The Dickey-Fuller test involves testing for the presence of a random walk and is based on estimates from an augmented autoregression. The test is valid for stationary and invertible ARMA noise functions of unknown order provided that the lag k, is chosen by sample size, to satisfy lower and upper bound conditions.

Dickey and Fuller (1979) proposed test statistics for the unit root hypothesis for an observed time series which can be obtained from three different processes. The null hypothesis states that the series contains a unit root against the alternative hypothesis that there is no unit root present. These hypotheses imply that failure to reject the null hypothesis provides evidence of a nonstationary series, whilst rejection of the null hypothesis provides evidence of a stationary series. Given that ε_t is independently and identically distributed ($i.i.d$) with mean zero and variance σ^2, the processes are as follows for T observations y_1, y_2 … y_T:

$$y_t = a_1 y_{t\text{-}1} + \varepsilon_t \ , \ \ t = 1, 2, \ldots$$

$$y_t = a_0 + a_1 y_{t\text{-}1} + \varepsilon_t \ , \ \ t = 1, 2, \ldots$$

$$y_t = a_0 + a_1 y_{t\text{-}1} + a_2 t + \varepsilon_t \ , \ \ t = 1, 2, \ldots$$

The ordinary least squares (OLS) estimator for a_1 is $\hat{a}_1$. This is a consistent estimator. The difference between the three equations concerns the presence of the deterministic elements a_1 and a_2. The constant term is a_0, and t represents the time trend.

To better understand this method, the first equation is taken as an example. To test the null hypothesis that a_1=1, the statistic is labeled as t. If the statistic t is greater than the critical value such as at the 5% level of significance, then the null hypothesis is not rejected and it should be concluded that the process is nonstationary.

3.2.2 Augmented Dickey-Fuller Test

To overcome the problem of autocorrelation in the Dickey-Fuller test, two approaches have been proposed to modify the test. The first method is a parametric approach commonly referred to as the ADF test, whilst an alternative is the nonparametric approach referred to as the PP test.

Dickey and Fuller (1981) extended the Dickey-Fuller test to an AR process of known order containing no more than one unit root. The ADF test is a unit root testing procedure that employs augmented autoregression estimates. The null hypothesis of this test statistic states that the series is nonstationary, whilst the alternative states the series is stationary. The original test assumed that the data are generated by a finite autoregression model of order p and therefore, the test was based on fitting an autoregression model of the same order. The constant term is a_0, t represents the

trend and ε_t is *i.i.d* with mean equal to zero and variance σ^2. The processes are as follows:

$$y_t = a_1 y_{t\text{-}1} + \sum_{t=2}^{\rho} \beta_i y_{t\text{-}i+1} + \varepsilon_t ,$$

$$y_t = a_0 + a_1 y_{t\text{-}1} + \sum_{t=2}^{\rho} \beta_i y_{t\text{-}i+1} + \varepsilon_t ,$$

$$y_t = a_0 + a_1 y_{t\text{-}1} + a_2 t + \sum_{t=2}^{\rho} \beta_i y_{t\text{-}i+1} + \varepsilon_t .$$

An alternative test, as mentioned before, is the PP test. In this study, the ADF test was used instead of the PP test to determine the order of integration of each series in the three cointegrated bivariate models.

3.3 Cointegration

Since the defining of cointegration by Granger (1983), Engle and Granger (1987) and Johansen (1988), researchers have concentrated on determining the number of cointegration relations in a system of time series variables.

Engle and Granger (1987) showed that a linear combination of two or more nonstationary series may be stationary. If such a stationary linear combination exists, the non-stationary time series was defined to be cointegrated.

3.3.1 Johansen Maximum Likelihood Method

The initial multivariate procedure for assessing long-run relationships was developed by Johansen (1988). The method uses a maximum likelihood procedure to determine the presence of cointegrating vectors in a set of nonstationary time series. The null hypothesis states that there is no cointegration amongst the series whilst the

alternative states that there is at least one cointegrated series. If the null hypothesis is rejected, i.e. there is evidence of unit root presence, the procedure follows a sequential route by then testing the null hypothesis of a single cointegrated relationship versus at least two. The process ends when one is no longer able to reject the null hypothesis.

The procedure follows the following steps:

Step 1: Pretest all variables to assess their order of integration.

Step 2: Estimate the model and determine the rank of, the parameter matrix $\mathbf{\Pi}$.

Step 3: Analyze the normalized cointegrating vectors and speed of adjustment coefficients.

Step 4: Innovation accounting and causality tests on the error-correction model.

This procedure applies maximum likelihood estimation to the VAR model, and considers the relationships among more than two variables. The method also estimates the unknown parameters and identifies the numbers of cointegrating vectors.

If a series were to return to a long-run equilibrium, the movement of at least some variables must respond to the magnitude of the error-correction. In other words, if two variables y and x were cointegrated, then the relationship between the two could be expressed as an ECM. Cointegration provides evidence of a long-run relationship between these variables, whilst the ECM provides evidence of the short-run relationship. Johansen provided the following VEC model:

$$\Delta y_t = \mathbf{\Pi} y_{t\text{-}1} + \mathbf{\Gamma_1} \Delta y_{t\text{-}1} + \ldots + \mathbf{\Gamma_{p\text{-}1}} \Delta y_{t\text{-}p+1} + \varepsilon_t$$

where $\mathbf{\Pi = -(I - A_1 - \ \ldots \ -A_p)}$ and $\mathbf{\Gamma_i = -(A_{i+1} + \ \ldots \ +A_p)}$ for $i = 1, \ldots, p\text{-}1$, are $K \times K$ coefficient matrices.

The parameter matrix $\mathbf{\Pi}$ indicates whether a vector of y_t has a long-run dynamic relationship or not. If the variables y_t are cointegrated, the matrix $\mathbf{\Pi}$ has a reduced rank $rk(\mathbf{\Pi}) = r$ and can be decomposed as $\mathbf{\Pi} = \alpha\beta'$, where α and β are $k \times r$ matrices. The rank r of $\mathbf{\Pi}$ is called the cointegrating rank. The matrix β contains the coefficients of the cointegration relations and it is called the matrix of cointegration vectors.

If the rank of $\mathbf{\Pi}$ equals the number of variables n (i.e. if $\mathbf{\Pi}$ is of full rank), the long-run equilibrium is given by n independent equations. If the rank of $\mathbf{\Pi}$ is zero, y_t are unit root processes and there are no error corrections and cointegration relationships. If the rank of $\mathbf{\Pi}$ is between zero and the number n, i.e. is $0 < r < n$, then cointegration vectors are present.

To determine the number of cointegrating vectors, Johansen derived two tests: the trace test and the maximal eigenvalue test. In this study, only the trace test will be considered to determine the number of cointegrating vectors. The maximal eigenvalue test will not be illustrated.

3.3.2 Trace statistic

The null hypothesis of the trace test states that the number of distinct cointegrating vectors is less than or equal to r against a general alternative. The distribution of the λ_{trace} test statistic is given by

$$\lambda_{trace} = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i)$$

where $\hat{\lambda}_i$ = estimated values of the characteristic roots,

$r = 0, 1, 2, \ldots, n-2, n-1$, and

n = number of characteristic roots.

T = number of observations used

By comparison, although the OLS method is robust, Johansen's procedure has been developed to avoid defects. Johansen's procedure is fast becoming an essential tool for applied economists to estimate time series models.

Chapter 4

Model selection criteria

In this chapter, the model selection criteria are shown. The criteria used have been sourced in literature and are specific to high dimensional, nonstationary processes with error correction terms.

Sclove (1994) describes model selection as the choice of selecting the best model from a set of candidate models. The principle of model selection is based on a well-justified criterion of choosing the "best" model and that criterion should be based on relevant theory. The criterion must be estimable from the data for each fitted model and the criterion must fit into a general statistical inference framework. An information theoretic criterion meets these requirements by minimizing the Kullback-Leibler distance as first proposed by Akaike (1974).

As mentioned, this research considers model selection criteria for selecting the order of a nonstationary VAR process with p lag terms and focuses on six criteria identified in literature for lag length selection of VAR processes. The six criteria which will be used are AIC, SIC, HQ, AICc, MAIC and MAICc.

These criteria will be used to analyze simulated data from a theoretical cointegrated model and will provide insight into the sensitivity of each information criterion. The criterion which identifies the corrected model most often will be identified as the most appropriate criterion.

The model selection criteria which will be used in the assessment process follows the general form shown in the equations that follow. The formulae for the general form are functions of the log-likelihood estimates plus penalty functions for the number of

parameters in the model. The model with the smallest IC estimate will be chosen as the best fitting model. The equations are:

$$IC(p) = \ln\left|\hat{\Sigma}\right| + C_T \frac{p}{T},$$

where $\hat{\Sigma}$ is an approximation of the residual covariance matrix associated with the fitted VAR(p) model,

C_T is a deterministic penalty term,

T denotes the number of observations used for estimation, and

p denotes the lag order.

The second term on the right hand side of the above equation, is often referred to as the penalty term, includes C_T which is a function of the sample size T. C_T varies with the choice of different information criteria. k denotes the number of equations in the VAR model and is given by the following:

AIC was proposed by Akaike (1973) and uses

$$C_T = 2k^2.$$

SIC was proposed by Schwarz (1978) and uses

$$C_T = k^2 \ln T.$$

HQ was proposed by Hannan and Quinn (1979) and uses

$$C_T = 2k^2 \ln\ln T.$$

Some information criteria possess different statistical properties such as consistency and efficiency. Lütkepohl (2005) showed that AIC was inconsistent, but SIC and HQ were consistent. Hurvich and Tsai (1993) derived a finite sample equivalent of the AIC called AICc. They showed that the AICc was asymptotically efficient despite its inconsistency property. The simulation results of Hurvich and Tsai (1993) showed that AICc was able to compete well with SIC at selecting the true order of a model in small samples.

4.1 Akaike information criterion

Akaike (1973) derived the AIC as an approximately unbiased estimator of the Kullback-Leibler information. Although this criterion has been widely used by researchers, Hurvich and Tsai (1989) illustrated that AIC overfits when the sample sizes are small or when the number of fitted parameters are a moderate to large fraction of the sample size.

The AIC function that has been used in VAR analysis is given by Ivanov and Kilian (2005), Lütkepohl (1985) and Gonzalo and Pitarakis (1998) as

$$AIC(p) = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 p}{T}$$

where k = the number of equations in the VAR model,

T = the effective sample size,

p = the number of lag terms in the model, and

$\hat{\Sigma}$ = the estimated covariance matrix of the fitted VAR(p) model.

4.2 Corrected Akaike's information criterion

Although AIC has proved to be a widely accepted method, it does have limitations. In the regression framework, Hurvich and Tsai (1989) showed that when the sample size was small, or when the number of fitted parameters was a moderate to large fraction of the sample size, AIC overfits the model. They subsequently proposed a finite sample corrected version, AICc, which was less unbiased and consequently tended to select better models than AIC.

Hurvich and Tsai (1989) found that the bias reduction of AICc, when compared to AIC, was noticeable, so too was the improvement in the IC's ability for selection of the model.

Originally developed for linear regression, AICc has been extended to a number of additional frameworks, including ARMA modeling (Hurvich et al., 1990), VAR modeling (Hurvich and Tsai, 1993), and multivariate regression modeling (Bedrick and Tsai, 1994). The IC used in the VAR framework is given as

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

where k, T, p and $\hat{\Sigma}$ are as previously defined.

4.3 Schwarz information criterion

Schwarz (1978) derived SIC by approaching model selection from a Bayesian perspective. The criterion developed by Schwarz (1978) is referred to in literature in three different ways. Some researchers refer to the Schwarz information criteria as SIC, others call it the BIC, whilst there are some who combine the terms and refer to the criterion as the Schwarz Bayesian criterion (SBC). Throughout this study we standardize on the notation and use SIC.

Schwarz (1978) derived the original criterion in the linear models framework. The IC has been extended to the VAR framework and in this study the IC given by Ivanov and Kilian (2005), Lütkepohl (1985) and Gonzalo and Pitarakis (1998) was used. The IC used is defined as

$$SIC(p) = \ln\left|\hat{\Sigma}\right| + \frac{k^2 p \ln(T)}{T}$$

where k, T, p and $\hat{\Sigma}$ are as previously defined.

4.4 Hannan-Quinn information criterion

Hannan and Quinn (1979) based their derivation of HQ on the approach adopted by Shibata for developing the Shibata's information criterion. They provided a procedure where the method was strongly consistent for estimating the order of an autoregressive model. The IC derived is often referred to in literature as HQIC, for this study the notation used is HQ.

The autoregressive framework was extended to a VAR framework and in this study the function given in the studies of Ivanov and Kilian (2005), Lütkepohl (1985) and Gonzalo and Pitarakis (1998) was used. The IC function is given as

$$HQ(p) = \ln\left|\hat{\Sigma}\right| + \frac{2k^2 p \ln\ln(T)}{T}$$

where k, T, p and $\hat{\Sigma}$ are as previously defined.

4.5 Modified Akaike's information criterion

Ng and Perron (2001) showed that the BIC and AIC were insufficiently flexible when applied to models with unit roots. They derived a modification to Akaike's IC which was called MAIC. The MAIC was shown to be more robust when there were negative moving-average error terms in the autoregressive process. They also showed that the key distinction between MAIC and standard ICs was that the former took into account the fact that the bias in the estimate of the sum of the autoregressive coefficients was dependent on the lag term p.

Qu and Perron (2006) extended the work of Ng and Perron (2001) to the nonstationary VAR framework. The idea was to include a restriction term for the cointegration presence in the VAR model. In their study, it was stated that the extra term in the MAIC provided only a finite sample adjustment.

Qu and Perron (2006) proposed a MAIC, as follows,

$$MAIC(p) = \ln\left|\hat{\Sigma}\right| + \frac{2(\Gamma_p(r) + kp^2)}{T}$$

where $\Gamma_k(r)$ is the likelihood ratio statistic test of r against more than r cointegrating vectors,

r denotes the number of cointegrating vectors, and

k, T, p and $\hat{\Sigma}$ are as previously defined.

4.6 Corrected modified Akaike's information criterion

A modification of Hurvich and Tsai's AICc and Qu and Perron's MAIC gives the last IC analysed in this study. This IC was used by Sharp and Radloff (2007) and is referred to as the corrected modified Akaike's IC, MAICc. The formula for MAICc is given as

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

where $\Gamma_k(r)$, k, T, p and $\hat{\Sigma}$ are as previously defined.

Chapter 5

Vector autoregressive models

The ARMA models are a popular class of linear time series models. These models consist of purely AR and purely MA models as special cases. ARMA models are frequently used to model linear dynamic structures, to depict linear relationships among lagged variables, and to serve as vehicles for linear forecasting. In this study, only the bivariate VAR model with an error correction component was considered.

In the bivariate system assessed, no intercept and trend terms were included and the model was limited to the case where there were a maximum of four lag terms in VAR representation. This study assessed three models for each lag term model and then compared the results. The lag restriction was motivated by the methodology of Liew (2004) who limited the models to four lag terms.

The presentations of the models that follow were adopted from the study of Liu (2007) who evaluated a simulated VEC model for four ICs: AIC, SIC, HQ and FPE.

The models that follow this introduction include three different parameterizations for each of four models of varying lag lengths. The general case theoretical models are shown first which is followed by the defined parameterized models. In total, twelve defined parameterized models will be given.

5.1 Lag length one models

The two dimensional VAR model with a lag length of one is represented below.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + \varepsilon_{1,t}$$

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + \varepsilon_{2,t}$$

where a_{ij} denotes the coefficient of the ith equation for the jth variable

(where i = 1, 2 and j = 1, 2),

$x_{j,t}$ denotes the jth variable at time period t, and

$\varepsilon_{i,t}$ denotes the error term of the ith equation at time period t.

Assuming the bivariate VAR model is cointegrated and that each single equation is first order stationary, the model can be re-written as a cointegrated model by subtracting $x_{j,t-1}$ from both sides of each equation i.e.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + \varepsilon_{1,t}$$

which is re-written as

$$x_{1,t} - x_{1,t-1} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + \varepsilon_{1,t}$$

and

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + \varepsilon_{2,t}$$

is re-written as

$$x_{2,t} - x_{2,t-1} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + \varepsilon_{2,t}\,.$$

Thus for the cointegrated model we have

$$\Delta x_{1,t} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + \varepsilon_{1,t}\text{, and}$$

$$\Delta x_{2,t} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + \varepsilon_{2,t}$$

which can be re-written in matrix form as

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

This can be represented in matrix notation and simplified to

$$\Delta x_t = (A_1 \text{ - } I)x_{t\text{-}1} + E_t\,.$$

Simplifying, the notation further we have

$$\begin{aligned} \Delta x_t &= (A_1 \text{ - } I)x_{t\text{-}1} + E_t \\ &= \text{-}(I \text{ - } A_1)x_{t\text{-}1} + E_t \\ &= \mathbf{\Pi} x_{t\text{-}1} + E_t \end{aligned}$$

where $\mathbf{\Pi} = \text{-}(I \text{ - } A_1)$, $A_1 = \begin{bmatrix} a_{11}\ a_{12} \\ a_{21}\ a_{22} \end{bmatrix}$, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $E_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$ and $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{\mathbf{11}}\ \boldsymbol{\pi}_{\mathbf{12}} \\ \boldsymbol{\pi}_{\mathbf{21}}\ \boldsymbol{\pi}_{\mathbf{22}} \end{bmatrix}$.

This model assumed that the bivariate model was cointegrated, which meant that the coefficient matrix $\mathbf{\Pi}$ has rank r, where $0 < r < k$, with k denoting the number of equations in the multivariate model (Enders, 2004). The rank r is the number of cointegrating relationships between the variables.

Three models with different parameterization for the VAR (1) model are illustrated below.

Model 1

The series $x_{1,t}$ and $x_{2,t}$ was constructed as

$$x_{1,t} = 0.9x_{1,t-1} + 0.4x_{2,t\text{-}1} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 1x_{2,t\text{-}1} + \varepsilon_{2,t}$$

where $x_{1,0} = 0$, $x_{2,0} = 0$,

$\varepsilon_{1,t} \sim N(0,1)$ and $\varepsilon_{2,t} \sim N(0,1)$.

Using the coefficients given above, the equations were restricted to provide a cointegrated series with one cointegrating equation. This is seen by substituting the values into the error correction model and observing that the error correction parameter matrix, $\mathbf{\Pi}$ has rank of one.

$$x_{1,t} - x_{1,t-1} = (0.9x_{1,t-1} - 1x_{1,t-1}) + 0.4x_{2,t\text{-}1} + \varepsilon_{1,t}$$

$$\Delta x_{1,t} = (0.9 - 1)x_{1,t-1} + 0.4x_{2,t\text{-}1} + \varepsilon_{1,t}$$

and

$$x_{2,t} - x_{2,t-1} = 0x_{1,t-1} + (1x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + \varepsilon_{2,t}$$

$$\Delta x_{2,t} = 0x_{1,t-1} + (1-1)x_{2,t\text{-}1} + \varepsilon_{2,t}\,.$$

Re-written in matrix form this is

$$\begin{bmatrix} \Delta x_{1,\mathrm{t}} \\ \Delta x_{2,\mathrm{t}} \end{bmatrix} = \left[\begin{pmatrix} 0.9 & 0.4 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{bmatrix} x_{1,\mathrm{t}-1} \\ x_{2,\mathrm{t}\text{-}1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,\mathrm{t}} \\ \varepsilon_{2,\mathrm{t}} \end{bmatrix}.$$

This can be written in matrix notation and simplified to give

$$\Delta x_t = \left[\begin{pmatrix} (0.9\text{ - }1) & 0.4 \\ 0 & (1\text{ - }1) \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= \left[\begin{pmatrix} \text{-}0.1 & 0.4 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{-0.1} & \mathbf{0.4} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

The other two lag length one simulated models follow the same procedures.

Model 2

$$x_{1,t} = 0.65x_{1,t-1} + 0.3x_{2,t\text{-}1} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 1x_{2,t\text{-}1} + \varepsilon_{2,t}\,.$$

Written in matrix notation and simplifying, we get

$$\Delta x_t = \left[\begin{pmatrix} (0.65-1) & 0.3 \\ 0 & (1-1) \end{pmatrix}\right]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$
$$= \left[\begin{pmatrix} -0.35 & 0.3 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{-0.35} & \mathbf{0.3} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

Model 3

$$x_{1,t} = 1x_{1,t-1} + 0x_{2,t-1} + \varepsilon_{1,t}$$

$$x_{2,t} = 0.3x_{1,t-1} + 0.5x_{2,t-1} + \varepsilon_{2,t}.$$

Written in matrix notation and simplifying, we get

$$\Delta x_t = \left[\begin{pmatrix} (1-1) & 0 \\ 0.3 & (0.5-1) \end{pmatrix}\right]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$
$$= \left[\begin{pmatrix} 0 & 0 \\ 0.3 & -0.5 \end{pmatrix}\right]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0.3} & \mathbf{-0.5} \end{bmatrix}$.

5.2 Lag length two models

The two dimensional VAR model with a lag length of two is represented below.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + \varepsilon_{1,t}$$

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + \varepsilon_{2,t}$$

where a_{ij}, $x_{j,t}$ and $\varepsilon_{i,t}$ are as denoted for the autoregressive model of lag length one and b_{ij} denotes the coefficient of the ith equation for the jth variable (where i = 1, 2 and j = 1, 2).

Assuming that the bivariate model is cointegrated and that each single equation is first order stationary, the model can be re-written as a cointegrated model by subtracting $x_{j,t-1}$ from both sides of each equation i.e.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + \varepsilon_{1,t}$$

is re-written as

$$x_{1,t} - x_{1,t-1} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + \varepsilon_{1,t}$$

and

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + \varepsilon_{2,t}$$

is re-written as

$$x_{2,t} - x_{2,t-1} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + \varepsilon_{2,t}\,.$$

Thus for the cointegrated model we have

$$\Delta x_{1,t} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + \varepsilon_{1,t}\text{, and}$$

$$\Delta x_{2,t} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + \varepsilon_{2,t}$$

which can be re-written in matrix form as

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

This can be represented in matrix notation and simplified to

$$\Delta x_t = (A_1 - I)x_{t\text{-}1} + A_2 x_{t\text{-}2} + E_t\,.$$

Now, a term $A_2 x_{t-1}$ is added to and subtracted from the right-hand-side of the equation. This can then be simplified as illustrated below.

$$\begin{aligned} \Delta x_t &= (A_1 - I)x_{t\text{-}1} + A_2 x_{t\text{-}2} + (A_2 x_{t\text{-}1} - A_2 x_{t\text{-}1}) + E_t \\ &= (A_1 + A_2 - I)x_{t\text{-}1} + A_2 x_{t\text{-}2} - A_2 x_{t\text{-}1} + E_t \\ &= -(I - A_1 - A_2)x_{t\text{-}1} - A_2(x_{t\text{-}1} - x_{t\text{-}2}) + E_t \\ &= -(I - A_1 - A_2)x_{t\text{-}1} - A_2 \Delta x_{t\text{-}1} + E_t \end{aligned}$$

where $A_1 = \begin{bmatrix} a_{11}\ a_{12} \\ a_{21}\ a_{22} \end{bmatrix}$, $A_2 = \begin{bmatrix} b_{11}\ b_{12} \\ b_{21}\ b_{22} \end{bmatrix}$, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $E_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$.

Simplifying, the notation further we have

$$\Delta x_t = \mathbf{\Pi} x_{t\text{-}1} + \mathbf{\Pi_1} \Delta x_{t\text{-}1} + E_t$$

where $\mathbf{\Pi} = \text{-}(I \text{ - } A_1 \text{ - } A_2)$, $\mathbf{\Pi_1} = \text{-}A_2$ and $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix}$.

This model assumed that the bivariate model was cointegrated, which meant that the coefficient matrix $\mathbf{\Pi}$ has rank r, where $0 < r < k$, with k denoting the number of equations in the multivariate model (Enders, 2004). The rank r is the number of cointegrating relationships between the variables.

Three models with different parameterization for the VAR(2) model are illustrated below.

Model 1

$$x_{1,t} = -0.5x_{1,t-1} + 0.2x_{2,t\text{-}1} + 0.1x_{1,t\text{-}2} + 0.1x_{2,t\text{-}2} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 0.7x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.3x_{2,t\text{-}2} + \varepsilon_{2,t}$$

where $x_{1,0} = 0$, $x_{2,0} = 0$,

$\varepsilon_{1,t} \sim N(0,1)$ and $\varepsilon_{2,t} \sim N(0,1)$.

Using the coefficients given above, the equations are restricted to provide a cointegrated series with one cointegrating equation. This can be seen by substituting the values into the error correction model and observing that the error correction parameter matrix, $\mathbf{\Pi}$ has rank of one.

$$x_{1,t} - x_{1,t-1} = (-0.5x_{1,t-1} - 1x_{1,t-1}) + 0.2x_{2,t\text{-}1} + 0.1x_{1,t\text{-}2} + 0.1x_{2,t\text{-}2} + \varepsilon_{1,t}$$

$$\Delta x_{1,t} = (-0.5 - 1)x_{1,t-1} + 0.2x_{2,t\text{-}1} + 0.1x_{1,t\text{-}2} + 0.1x_{2,t\text{-}2} + \varepsilon_{1,t}$$

and

$$x_{2,t} - x_{2,t-1} = 0x_{1,t-1} + (0.7x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + 0x_{1,t\text{-}2} + 0.3x_{2,t\text{-}2} + \varepsilon_{2,t}$$

$$\Delta x_{2,t} = 0x_{1,t-1} + (0.7 - 1)x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.3x_{2,t\text{-}2} + \varepsilon_{2,t}\,.$$

Written in matrix notation and simplifying, we get

$$\Delta x_t = -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.5 & 0.2 \\ 0 & 0.7 \end{pmatrix} - \begin{pmatrix} 0.1 & 0.1 \\ 0 & 0.3 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0.3 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= -\left[\begin{pmatrix} 1.4 & -0.3 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0.3 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = -\begin{bmatrix} \mathbf{1.4} & \mathbf{-0.3} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

The other two lag length two simulated models follow the same procedures.

Model 2

$$x_{1,t} = -0.75x_{1,t-1} + 0.4x_{2,t\text{-}1} + 0.2x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 0.8x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + \varepsilon_{2,t}\,.$$

Written in matrix notation and simplifying, we get

$$\Delta x_t = -\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}-0.75 & 0.4\\ 0 & 0.8\end{pmatrix} - \begin{pmatrix}0.2 & 0.2\\ 0 & 0.2\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \begin{bmatrix}0.2 & 0.2\\ 0 & 0.2\end{bmatrix}\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

$$= -\left[\begin{pmatrix}1.55 & -0.6\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \begin{bmatrix}0.2 & 0.2\\ 0 & 0.2\end{bmatrix}\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix}\boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12}\\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22}\end{bmatrix} = -\begin{bmatrix}\mathbf{1.55} & \mathbf{-0.6}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$.

Model 3

$$x_{1,t} = 0.6x_{1,t-1} - 0.3x_{2,t-1} + 0.4x_{1,t-2} + 0.3x_{2,t-2} + \varepsilon_{1,t}$$

$$x_{2,t} = 0.45x_{1,t-1} + 0x_{2,t-1} + 0.15x_{1,t-2} + 0x_{2,t-2} + \varepsilon_{2,t}\ .$$

written in matrix notation and simplifying, we get

$$\Delta x_t = -\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}0.6 & -0.3\\ 0.45 & 0\end{pmatrix} - \begin{pmatrix}0.4 & 0.3\\ 0.15 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \begin{bmatrix}0.4 & 0.3\\ 0.15 & 0\end{bmatrix}\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

$$= -\left[\begin{pmatrix}0 & 0\\ -0.6 & 1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \begin{bmatrix}0.4 & 0.3\\ 0.15 & 0\end{bmatrix}\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix}\boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12}\\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22}\end{bmatrix} = -\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{-0.6} & \mathbf{1}\end{bmatrix}$.

5.3 Lag length three models

The two dimensional VAR model with a lag length of three is represented below.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + \varepsilon_{1,t}$$

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + \varepsilon_{2,t}$$

where a_{ij}, b_{ij}, $x_{j,t}$ and $\varepsilon_{i,t}$ are as denoted for autoregressive models of lag length two and c_{ij} denotes the coefficient of the ith equation for the jth variable (where i = 1, 2 and j = 1, 2).

Assuming the bivariate model is cointegrated and that each single equation is first order stationary, the model can be re-written as a cointegrated model by subtracting $x_{j,t-1}$ from both sides of each equation i.e.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + \varepsilon_{1,t}$$

which is re-written as

$$x_{1,t} - x_{1,t-1} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2}$$
$$+c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + \varepsilon_{1,t}$$

and

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + \varepsilon_{2,t}$$

is re-written as

$$x_{2,t} - x_{2,t-1} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2}$$
$$+c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + \varepsilon_{2,t}\,.$$

Thus for the cointegrated model we have

$$\Delta x_{1,t} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2}$$
$$+c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + \varepsilon_{1,t}\text{, and}$$
$$\Delta x_{2,t} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2}$$
$$+c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + \varepsilon_{2,t}$$

which can be re-written in matrix form as

$$\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} = \left[\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} x_{1,\text{t-}2} \\ x_{2,\text{t-}2} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} x_{1,t-3} \\ x_{2,t-3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = (A_1 - I)x_{t\text{-}1} + A_2x_{t\text{-}2} + A_3x_{t\text{-}3} + E_t\,.$$

Now, the terms A_2x_{t-1}, A_3x_{t-1} and A_3x_{t-2} are added to and subtracted from the right-hand-side of the equation. This is then simplified as illustrated below.

$$\Delta x_t = (A_1 - I)x_{t\text{-}1} + A_2x_{t\text{-}2} + A_3x_{t\text{-}3} + E_t$$
$$= (A_1 - I)x_{t\text{-}1} + A_2x_{t\text{-}2} + A_3x_{t\text{-}3} + (A_2x_{t\text{-}1} - A_2x_{t\text{-}1}) + (A_3x_{t\text{-}1} - A_3x_{t\text{-}1}) + E_t$$

$$= (A_1 + A_2 + A_3 - I)x_{t-1} + A_2x_{t-2} + A_3x_{t-3} - A_2x_{t-1} - A_3x_{t-1} + E_t$$

$$= -(I - A_1 - A_2 - A_3)x_{t-1} - (A_2x_{t-1} - A_2x_{t-2}) + (A_3x_{t-3} - A_3x_{t-1}) + E_t$$

$$= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2\Delta x_{t-1} - A_3x_{t-1} + A_3x_{t-3} + (A_3x_{t-2} - A_3x_{t-2}) + E_t$$

$$= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2\Delta x_{t-1} - (A_3x_{t-1} - A_3x_{t-2}) - (A_3x_{t-2} - A_3x_{t-3}) + E_t$$

$$= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2(\Delta x_{t-1}) - A_3(\Delta x_{t-1}) - A_3(\Delta x_{t-2}) + E_t$$

$$= -(I - A_1 - A_2 - A_3)x_{t-1} - (A_2 + A_3)\Delta x_{t-1} - A_3(\Delta x_{t-2}) + E_t$$

where $A_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $A_2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, $A_3 = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $E_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$.

Simplifying the notation further we have

$$\Delta x_t = \mathbf{\Pi} x_{t-1} - \mathbf{\Pi_1}(\Delta x_{t-1}) - \mathbf{\Pi_2}(\Delta x_{t-1}) - \mathbf{\Pi_2}(\Delta x_{t-2}) + E_t$$

$$= \mathbf{\Pi} x_{t-1} - (\mathbf{\Pi_1} + \mathbf{\Pi_2})\Delta x_{t-1} - \mathbf{\Pi_2}(\Delta x_{t-2}) + E_t$$

where $\mathbf{\Pi} = -(I - A_1 - A_2 - A_3)$, $\mathbf{\Pi_1} = A_2$, $\mathbf{\Pi_2} = A_3$ and $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi_{11}} & \boldsymbol{\pi_{12}} \\ \boldsymbol{\pi_{21}} & \boldsymbol{\pi_{22}} \end{bmatrix}$.

This model assumed that the bivariate model was cointegrated, which meant that the coefficient matrix $\mathbf{\Pi}$ has rank r, where $0 < r < k$, with k denoting the number of equations in the multivariate model (Enders, 2004). The rank r is the number of cointegrating relationships between the variables.

Three models with different parameterization for the VAR (3) model are illustrated below.

Model 1

$$x_{1,t} = -0.8x_{1,t-1} + 0.35x_{2,t\text{-}1} + 0.3x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0.2x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 0.7x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{2,t}$$

where $x_{1,0} = 0$, $x_{2,0} = 0$,

$\varepsilon_{1,t} \sim N(0,1)$ and $\varepsilon_{2,t} \sim N(0,1)$.

Using the coefficients given above, the equations were restricted to provide a cointegrated series with one cointegrating equation. This is seen by substituting the values into the error correction model and observing that the error correction parameter matrix, $\mathbf{\Pi}$ has rank of one, i.e.

$$x_{1,t} - x_{1,t-1} = (-0.8x_{1,t-1} - 1x_{1,t-1}) + 0.35x_{2,t\text{-}1} + 0.3x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0.2x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{1,t}$$

$$\Delta x_{1,t} = (-0.8 - 1)x_{1,t-1} + 0.35x_{2,t\text{-}1} + 0.3x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0.2x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{1,t}$$

and

$$x_{2,t} - x_{2,t-1} = 0x_{1,t-1} + (0.7x_{2,t-1} - 1x_{2,t-1}) + 0x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{2,t}$$

$$\Delta x_{2,t} = 0x_{1,t-1} + (0.7 - 1)x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2} + 0x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{2,t}.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.8 & 0.35 \\ 0 & 0.7 \end{pmatrix} - \begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.2 \end{pmatrix} - \begin{pmatrix} 0.2 & 0.1 \\ 0 & 0.1 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.2 & 0.1 \\ 0 & 0.1 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$-\begin{bmatrix} 0.2 & 0.1 \\ 0 & 0.1 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= -\left[\begin{pmatrix} 1.3 & -0.65 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \left[\begin{pmatrix} 0.5 & 0.3 \\ 0 & 0.3 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$-\begin{bmatrix} 0.2 & 0.1 \\ 0 & 0.1 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = -\begin{bmatrix} \mathbf{1.3} & \mathbf{-0.65} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

The other two lag length three simulated models follow the same procedures.

Model 2

$$x_{1,t} = 0.65x_{1,t-1} + 0x_{2,t\text{-}1} + 0.2x_{1,t\text{-}1} + 0x_{2,t\text{-}1} + 0.15x_{1,t\text{-}3} + 0x_{2,t\text{-}3} + \varepsilon_{1,t}$$

$$x_{2,t} = -0.4x_{1,t-1} + 0.2x_{2,t\text{-}1} + 0.2x_{1,t\text{-}2} + 0.15x_{2,t\text{-}2} + 0.1x_{1,t\text{-}3} + 0x_{2,t\text{-}3} + \varepsilon_{2,t}\,.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.65 & 0 \\ -0.4 & 0.2 \end{pmatrix} - \begin{pmatrix} 0.2 & 0 \\ 0.2 & 0.15 \end{pmatrix} - \begin{pmatrix} 0.15 & 0 \\ 0.1 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.2 & 0 \\ 0.2 & 0.15 \end{pmatrix} + \begin{pmatrix} 0.15 & 0 \\ 0.1 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$-\begin{bmatrix} 0.15 & 0 \\ 0.1 & 0 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= -\left[\begin{pmatrix} 0 & 0 \\ 0.1 & 0.65 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \left[\begin{pmatrix} 0.35 & 0 \\ 0.3 & 0.15 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$- \begin{bmatrix} 0.15 & 0 \\ 0.1 & 0 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = -\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0.1} & \mathbf{0.65} \end{bmatrix}$.

Model 3

$$x_{1,t} = -0.7x_{1,t-1} + 0.3x_{2,t\text{-}1} + 0.2x_{1,t\text{-}2} + 0.15x_{2,t\text{-}2} + 0.1x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + \varepsilon_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 0.5x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.5x_{2,t\text{-}2} + 0x_{1,t\text{-}3} + 0x_{2,t\text{-}3} + \varepsilon_{2,t}\,.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.7 & 0.3 \\ 0 & 0.5 \end{pmatrix} - \begin{pmatrix} 0.2 & 0.15 \\ 0 & 0.5 \end{pmatrix} - \begin{pmatrix} 0.1 & 0.1 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix}$$

$$- \left[\begin{pmatrix} 0.2 & 0.15 \\ 0 & 0.5 \end{pmatrix} + \begin{pmatrix} 0.1 & 0.1 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$- \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= -\left[\begin{pmatrix} 1.4 & -0.55 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \left[\begin{pmatrix} 0.3 & 0.25 \\ 0 & 0.5 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$- \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = -\begin{bmatrix} \mathbf{1.4} & \mathbf{-0.55} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

In the last section, the bivariate VAR models are extended to lag length four models.

5.4 Lag length four models

The two dimensional VAR model with a lag length of four is represented below.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3}$$
$$+d_{11}x_{1,t\text{-}4} + d_{12}x_{2,t\text{-}4} + \varepsilon_{1,t}$$
$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3}$$
$$+d_{21}x_{1,t\text{-}4} + d_{22}x_{2,t\text{-}4} + \varepsilon_{2,t}$$

where a_{ij}, b_{ij}, c_{ij} $x_{j,t}$ and $\varepsilon_{i,t}$ are as denoted for autoregressive model of lag length three and d_{ij} denotes the coefficient of the ith equation for the jth variable (where i = 1, 2 and j = 1, 2).

Assuming the bivariate model is cointegrated and that each single equation is first order stationary, the model can be re-written as a cointegrated model by subtracting $x_{j,t-1}$ from both sides of each equation i.e.

$$x_{1,t} = a_{11}x_{1,t-1} + a_{12}x_{2,t\text{-}1} + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2} + c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3}$$
$$+d_{11}x_{1,t\text{-}4} + d_{12}x_{2,t\text{-}4} + \varepsilon_{1,t}$$

which is re-written as

$$x_{1,t} - x_{1,t-1} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2}$$

$$+c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + d_{11}x_{1,t\text{-}4} + d_{12}x_{2,t\text{-}4} + \varepsilon_{1,t}$$

and

$$x_{2,t} = a_{21}x_{1,t-1} + a_{22}x_{2,t\text{-}1} + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2} + c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3}$$

$$+d_{21}x_{1,t\text{-}4} + d_{22}x_{2,t\text{-}4} + \varepsilon_{2,t}$$

which is re-written as

$$x_{2,t} - x_{2,t-1} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2}$$

$$+c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + d_{21}x_{1,t\text{-}4} + d_{22}x_{2,t\text{-}4} + \varepsilon_{2,t}\,.$$

Thus, for the cointegrated model we have

$$\Delta x_{1,t} = (a_{11}x_{1,t-1} - 1x_{1,t-1}) + (a_{12}x_{2,t\text{-}1} - 0x_{2,t\text{-}1}) + b_{11}x_{1,t\text{-}2} + b_{12}x_{2,t\text{-}2}$$

$$+c_{11}x_{1,t\text{-}3} + c_{12}x_{2,t\text{-}3} + d_{11}x_{1,t\text{-}4} + d_{12}x_{2,t\text{-}4} + \varepsilon_{1,t}$$

$$\Delta x_{2,t} = (a_{21}x_{1,t-1} - 0x_{1,t-1}) + (a_{22}x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + b_{21}x_{1,t\text{-}2} + b_{22}x_{2,t\text{-}2}$$

$$+c_{21}x_{1,t\text{-}3} + c_{22}x_{2,t\text{-}3} + d_{21}x_{1,t\text{-}4} + d_{22}x_{2,t\text{-}4} + \varepsilon_{2,t}$$

which can be re-written in matrix form as

$$\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} = \left[\begin{pmatrix} a_{11}\ a_{12} \\ a_{21}\ a_{22} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11}\ b_{12} \\ b_{21}\ b_{22} \end{bmatrix} \begin{bmatrix} x_{1,\text{t-}2} \\ x_{2,\text{t-}2} \end{bmatrix} + \begin{bmatrix} c_{11}\ c_{12} \\ c_{21}\ c_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-3} \\ x_{2,t-3} \end{bmatrix} +$$

$$\begin{bmatrix} d_{11}\ d_{12} \\ d_{21}\ d_{22} \end{bmatrix} \begin{bmatrix} x_{1,t-4} \\ x_{2,t-4} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = (A_1 - I)x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + A_4 x_{t-4} + E_t .$$

Now, the terms $A_2 x_{t-1}$, $A_3 x_{t-1}$, $A_4 x_{t-1}$, $A_3 x_{t-2}$, $A_4 x_{t-2}$ and $A_4 x_{t-3}$ are added to and subtracted from the right-hand-side of the equation. This is then simplified as illustrated below.

$$\begin{aligned}
\Delta x_t &= (A_1 - I)x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + A_4 x_{t-4} + E_t \\
&= (A_1 - I)x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + A_4 x_{t-4} + (A_2 x_{t-1} - A_2 x_{t-1}) + (A_3 x_{t-1} - A_3 x_{t-1}) \\
&\quad +(A_4 x_{t-1} - A_4 x_{t-1}) + E_t \\
&= (A_1 + A_2 + A_3 + A_4 - I)x_{t-1} - (A_2 x_{t-1} - A_2 x_{t-2}) - (A_3 x_{t-1} - A_3 x_{t-3}) \\
&\quad -(A_4 x_{t-1} - A_4 x_{t-4}) + E_t \\
&= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2(\Delta x_{t-1}) - A_3 x_{t-1} + A_3 x_{t-3} - A_4 x_{t-1} + A_4 x_{t-4} \\
&\quad +(A_3 x_{t-2} - A_3 x_{t-2}) + (A_4 x_{t-2} - A_4 x_{t-2}) + (A_4 x_{t-3} - A_4 x_{t-3}) + E_t \\
&= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2(\Delta x_{t-1}) - (A_3 x_{t-1} - A_3 x_{t-2}) - (A_3 x_{t-2} + A_3 x_{t-3}) \\
&\quad -(A_4 x_{t-1} - A_4 x_{t-2}) - (A_4 x_{t-2} - A_4 x_{t-3}) - (A_4 x_{t-3} - A_4 x_{t-4}) + E_t \\
&= -(I - A_1 - A_2 - A_3)x_{t-1} - A_2(\Delta x_{t-1}) - A_3(\Delta x_{t-1}) - A_3(\Delta x_{t-2}) \\
&\quad -A_4(\Delta x_{t-1}) - A_4(\Delta x_{t-2}) - A_4(\Delta x_{t-3}) + E_t \\
&= -(I - A_1 - A_2 - A_3 - A_4)x_{t-1} - (A_2 + A_3 + A_4)\Delta x_{t-1} - (A_3 + A_4)\Delta x_{t-2} \\
&\quad -A_4(\Delta x_{t-3}) + E_t
\end{aligned}$$

where $A_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $A_2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, $A_3 = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$, $A_4 = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$,

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } E_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}.$$

Simplifying, the notation further we have

$$\Delta x_t = \mathbf{\Pi} x_{t\text{-}1} - (\mathbf{\Pi_1} + \mathbf{\Pi_2} + \mathbf{\Pi_3})\Delta x_{t\text{-}1} - (\mathbf{\Pi_2} + \mathbf{\Pi_3})\Delta x_{t\text{-}2} - \mathbf{\Pi_3}(\Delta x_{t\text{-}2}) + E_t$$

where $\mathbf{\Pi} = -(I - A_1 - A_2 - A_3 - A_4)$, $\mathbf{\Pi_1} = A_2$, $\mathbf{\Pi_2} = A_3$, $\mathbf{\Pi_3} = A_4$ and

$$\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix}.$$

This model assumed that the bivariate model was cointegrated, which meant that the coefficient matrix $\mathbf{\Pi}$ has rank r , where $0 < r < k$, with k denoting the number of equations in the multivariate model (Enders, 2004). The rank r is the number of cointegrating relationships between the variables.

Three models with different parameterization for the VAR(4) model are illustrated below.

Model 1

$$\begin{aligned} x_{1,t} &= -0.4x_{1,t\text{-}1} + 0.25x_{2,t\text{-}1} + 0.25x_{1,t\text{-}2} - 0.2x_{2,t\text{-}2} \\ &\quad +0.15x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + 0.15x_{1,t\text{-}4} + 0.1x_{2,t\text{-}4} + \varepsilon_{1,t} \\ x_{2,t} &= 0x_{1,t-1} + 0.35x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.25x_{2,t\text{-}2} \\ &\quad +0x_{1,t\text{-}3} + 0.2x_{2,t\text{-}3} + 0x_{1,t\text{-}4} + 0.2x_{2,t\text{-}4} + \varepsilon_{2,t} \end{aligned}$$

where $x_{1,0} = 0$, $x_{2,0} = 0$,

$\varepsilon_{1,t} \sim N(0,1)$ and $\varepsilon_{2,t} \sim N(0,1)$.

Using the coefficients given above, the equations are restricted to provide a cointegrated series with one cointegrating equation. This is seen by substituting the values into the error correction model and observing that the error correction parameter matrix, $\mathbf{\Pi}$, has rank of one, i.e.

$$x_{1,t} - x_{1,t-1} = (-0.4x_{1,t-1} - 1x_{1,t-1}) + 0.25x_{2,t\text{-}1} + 0.25x_{1,t\text{-}2} - 0.2x_{2,t\text{-}2}$$

$$+0.15x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + 0.15x_{1,t\text{-}4} + 0.1x_{2,t\text{-}4} + \varepsilon_{1,t}$$

$$\Delta x_{1,t} = (-0.4 - 1)x_{1,t-1} + 0.25x_{2,t\text{-}1} + 0.25x_{1,t\text{-}2} - 0.2x_{2,t\text{-}2}$$

$$+0.15x_{1,t\text{-}3} + 0.1x_{2,t\text{-}3} + 0.15x_{1,t\text{-}4} + 0.1x_{2,t\text{-}4} + \varepsilon_{1,t}$$

and

$$x_{2,t} - x_{2,t-1} = 0x_{1,t-1} + (0.35x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + 0x_{1,t\text{-}2} + 0.25x_{2,t\text{-}2}$$

$$+0x_{1,t\text{-}3} + 0.2x_{2,t\text{-}3} + 0x_{1,t\text{-}4} + 0.2x_{2,t\text{-}4} + \varepsilon_{2,t}$$

$$\Delta x_{2,t} = 0x_{1,t-1} + (0.35x_{2,t\text{-}1} - 1x_{2,t\text{-}1}) + 0x_{1,t\text{-}2} + 0.25x_{2,t\text{-}2}$$

$$+0x_{1,t\text{-}3} + 0.2x_{2,t\text{-}3} + 0x_{1,t\text{-}4} + 0.2x_{2,t\text{-}4} + \varepsilon_{2,t}\ .$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.4 & 0.25 \\ 0 & 0.35 \end{pmatrix} - \begin{pmatrix} 0.25 & -0.2 \\ 0 & 0.25 \end{pmatrix} - \begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix} - \begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.25 & (-0.2) \\ 0 & 0.25 \end{pmatrix} + \begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix} + \begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}3} - x_{1,t\text{-}4} \\ x_{2,t\text{-}3} - x_{2,t\text{-}4} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$= -\left[\begin{pmatrix} 0.85 & -0.25 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} \\ x_{2,t\text{-}1} \end{bmatrix} - \left[\begin{pmatrix} 0.55 & 0 \\ 0 & 0.65 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}1} - x_{1,t\text{-}2} \\ x_{2,t\text{-}1} - x_{2,t\text{-}2} \end{bmatrix}$$

$$-\left[\begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.4 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}2} - x_{1,t\text{-}3} \\ x_{2,t\text{-}2} - x_{2,t\text{-}3} \end{bmatrix} - \left[\begin{pmatrix} 0.15 & 0.1 \\ 0 & 0.2 \end{pmatrix}\right]\begin{bmatrix} x_{1,t\text{-}3} - x_{1,t\text{-}4} \\ x_{2,t\text{-}3} - x_{2,t\text{-}4} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} \end{bmatrix} = -\begin{bmatrix} \mathbf{0.85} & \mathbf{-0.25} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

The other two lag length four simulated models follow the same procedures.

Model 2

$$x_{1,t} = -0.3x_{1,t\text{-}1} + 0.2x_{2,t\text{-}1} + 0.2x_{1,t\text{-}1} + 0.2x_{2,t\text{-}1}$$

$$+0.2x_{1,t\text{-}3} + 0.15x_{2,t\text{-}3} + 0.1x_{1,t\text{-}4} + 0.1x_{2,t\text{-}4} + e_{1,t}$$

$$x_{2,t} = 0x_{1,t\text{-}1} + 0.5x_{2,t\text{-}1} + 0x_{1,t\text{-}2} + 0.2x_{2,t\text{-}2}$$

$$+0x_{1,t\text{-}3} + 0.2x_{2,t\text{-}3} + 0x_{1,t\text{-}4} + 0.1x_{2,t\text{-}4} + e_{2,t}\,.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}-0.3 & 0.2\\ 0 & 0.5\end{pmatrix} - \begin{pmatrix}0.2 & 0.2\\ 0 & 0.2\end{pmatrix} - \begin{pmatrix}0.2 & 0.15\\ 0 & 0.2\end{pmatrix} - \begin{pmatrix}0.1 & 0.1\\ 0 & 0.1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.2 & 0.2\\ 0 & 0.2\end{pmatrix} + \begin{pmatrix}0.2 & 0.15\\ 0 & 0.2\end{pmatrix} + \begin{pmatrix}0.1 & 0.1\\ 0 & 0.1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.2 & 0.15\\ 0 & 0.2\end{pmatrix} + \begin{pmatrix}0.1 & 0.1\\ 0 & 0.1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-2} - x_{1,t-3}\\ x_{2,t-2} - x_{2,t-3}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.1 & 0.1\\ 0 & 0.1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-3} - x_{1,t-4}\\ x_{2,t-3} - x_{2,t-4}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

$$= -\left[\begin{pmatrix}0.8 & -0.65\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \left[\begin{pmatrix}0.5 & 0.45\\ 0 & 0.5\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.3 & 0.25\\ 0 & 0.3\end{pmatrix}\right]\begin{bmatrix}x_{1,t-2} - x_{1,t-3}\\ x_{2,t-2} - x_{2,t-3}\end{bmatrix} - \left[\begin{pmatrix}0.1 & 0.1\\ 0 & 0.1\end{pmatrix}\right]\begin{bmatrix}x_{1,t-3} - x_{1,t-4}\\ x_{2,t-3} - x_{2,t-4}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix}\boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12}\\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22}\end{bmatrix} = -\begin{bmatrix}\mathbf{0.8} & \mathbf{-0.65}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$.

Model 3

$$x_{1,t} = 0.5x_{1,t-1} + 0.3x_{2,t-1} + 0.2x_{1,t-2} + 0.15x_{2,t-2}$$

$$+0.1x_{1,t-3} + 0.1x_{2,t-3} + 0x_{1,t-4} + 0x_{2,t-4} + e_{1,t}$$

$$x_{2,t} = 0x_{1,t-1} + 0.4x_{2,t-1} + 0x_{1,t-2} + 0.35x_{2,t-2}$$

$$+0x_{1,t-3} + 0.25x_{2,t-3} + 0x_{1,t-4} + 0x_{2,t-4} + e_{2,t}.$$

This can be represented in matrix notation and simplified to

$$\Delta x_t = -\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}0.5 & 0.3\\ 0 & 0.4\end{pmatrix} - \begin{pmatrix}0.2 & 0.15\\ 0 & 0.35\end{pmatrix} - \begin{pmatrix}0.1 & 0.1\\ 0 & 0.25\end{pmatrix} - \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.2 & 0.15\\ 0 & 0.35\end{pmatrix} + \begin{pmatrix}0.1 & 0.1\\ 0 & 0.25\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.1 & 0.1\\ 0 & 0.25\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-2} - x_{1,t-3}\\ x_{2,t-2} - x_{2,t-3}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-3} - x_{1,t-4}\\ x_{2,t-3} - x_{2,t-4}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

$$= -\left[\begin{pmatrix}0.2 & -0.55\\ 0 & 0\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1}\\ x_{2,t-1}\end{bmatrix} - \left[\begin{pmatrix}0.3 & 0.25\\ 0 & 0.6\end{pmatrix}\right]\begin{bmatrix}x_{1,t-1} - x_{1,t-2}\\ x_{2,t-1} - x_{2,t-2}\end{bmatrix}$$

$$-\left[\begin{pmatrix}0.1 & 0.1\\ 0 & 0.25\end{pmatrix}\right]\begin{bmatrix}x_{1,t-2} - x_{1,t-3}\\ x_{2,t-2} - x_{2,t-3}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\end{bmatrix}$$

where $\mathbf{\Pi} = \begin{bmatrix}\boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12}\\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22}\end{bmatrix} = -\begin{bmatrix}\mathbf{0.2} & \mathbf{-0.55}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$.

This chapter illustrated the theoretical general case models and the subsequent parameterized simulated models. All the models have a zero intercept term and zero trend term. Two dimensional VAR models with lag terms of one, two, three and four have been illustrated in this chapter. These VAR models can be represented in matrix notation as error correction models with lag terms of zero, one, two and three. The parameterization of the subsequent simulated models ensure that the rank r of the error correction matrix, $\mathbf{\Pi}$, is less than the number of equations k in the bivariate VAR model.

Chapter 6

Results

The results of the simulation assessment of the twelve theoretical models shown in chapter 5 are given in this chapter. Six model information criteria were used in this assessment, AIC, SIC, HQ, AICc, MAIC and MAICc. The results are shown tabulated and graphically as both methods are useful for the discussions that follow.

The simulation results are grouped into two distinct groups. The first group is defined as the group that "meets specification" whilst the second group is defined as the group that "does not meet specification". The motivation for the independent groupings is that we would like to assess the "ideal case" data separately from the "questionable case" data. The "questionable case" data are those data which did not meet specification according to a pre-defined level of significance. In this study, the pre-defined level of significance for all hypothesis tests was set at 5%. The assessment is addressed as three questions:

- *How do the information criteria perform as the sample size changes?*
- *How do the information criteria perform in small sample sizes?*
- *How do the information criteria perform as the models lag terms increase?*

These questions are addressed in the sections that follow. All tabulated results are reported as percentages; the value is the percentage determined by the number of times that the IC selected the model with the correct lag structure. In the discussion, reference is made to the average performance of the IC; this term is the average percentage determined by averaging the number of times that the IC selected the model for the three different parameterized models.

6.1 How do the information criteria perform as the sample size changes?

To answer this question the results obtained by using the VEC model for sample sizes between 30 and 100 both inclusive are summarized. In addition what happens when the parameters of the three simulated models change is considered by subdividing into the two groups “meets specification” and “does not meet specification”.

6.1.1 “Meets specification” group

Table 6.1 contains the results of the six IC performances for the three models with zero lag terms in the error correction representation.

Three important observations are possible from the table:

1. The changing parameter of the three models makes little difference to the selection capability of the six IC. In Table 6.1, the results of IC selecting the correct lag for VEC(0) models with varying parameterization from sample size 30 to sample size 100 show that there is no clear difference as sample size changes. The idea of changing parameters in the three models seems to make no clear difference. The lag length selection capability of AIC, SIC, HQ, AICc, MAIC and MAICc were not effected by changing the parameters of the three models.

2. All six IC show good selection performances for the simulated results. As the sample size increases from 30 to 100, SIC selects the correct lag length more often than the other five criteria for all three VEC(0) models. Based on these observed results, SIC would be the recommended criteria for the two-dimensional VEC (0) model.

3. The simulation results show that the six IC all select the correct VEC (0) model quite well. In summary, as the sample size increases, HQ selects the model almost as well as SIC. The average correct lag selection for all sample sizes by HQ is 95% in models one and three and 95.2% in model two. The average correct selection percentages of AIC and MAIC are close for the VEC (0) model. The average correct selection for all sample sizes by AIC is 83.9% in model one, 84.7% in model two and 83.8% in model three. The average correct selection for all sample sizes by MAIC is 84% in model one, 84.7% in model two and 83.9% in model three. AICc and MAICc both have acceptable performances for the smaller sample sizes. The average correct selection percentages of MAICc is just below that for SIC and HQ for VEC (0) models. The average correct selection for all sample sizes by MAICc is 91.4% in model one, 91.6% in model two and 91.2% in model three.

Theoretically Lütkepohl (2005) states that if the dimensions of the process is greater than one, then both HQ and SIC are strongly consistent. The simulation results from the VEC (0) model in this study indicated that in general SIC estimated the correct order more often than the other IC for the small sample sizes used. The average correct selection percentage of SIC for all sample sizes was approximately 99% for all three VEC (0) models. As the sample size increased, HQ surpassed the SIC capabilities.

Table 6.1: The correct lag selected by various criteria of VEC(0) model with small sample sizes

(a) Model one

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	76.62	84.01	97.54	86.72	76.83	94.53
Sample 40	82.53	82.65	98.89	92.63	82.53	92.86
Sample 50	83.31	80.96	99.26	94.72	83.57	91.73
Sample 60	84.93	80.83	99.56	96.34	84.71	91.04
Sample 70	85.94	80.56	99.52	96.99	86.22	91.09
Sample 80	85.66	78.72	99.74	97.17	85.66	90.35
Sample 90	86.27	78.82	99.76	97.86	86.41	90.20
Sample 100	86.15	78.65	99.83	97.55	85.98	89.61

(b) Model two

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	78.53	84.11	97.01	87.05	78.4	93.72
Sample 40	83.28	83.57	98.84	93.76	83.08	93.05
Sample 50	84.48	81.85	99.44	95.23	84.48	92.00
Sample 60	84.83	81.06	99.65	96.08	85.15	91.26
Sample 70	85.20	79.88	99.68	96.91	85.58	90.35
Sample 80	86.08	79.80	99.83	96.95	86.40	90.79
Sample 90	86.72	80.02	99.87	97.57	86.60	90.38
Sample 100	88.08	80.29	99.83	98.08	87.91	91.23

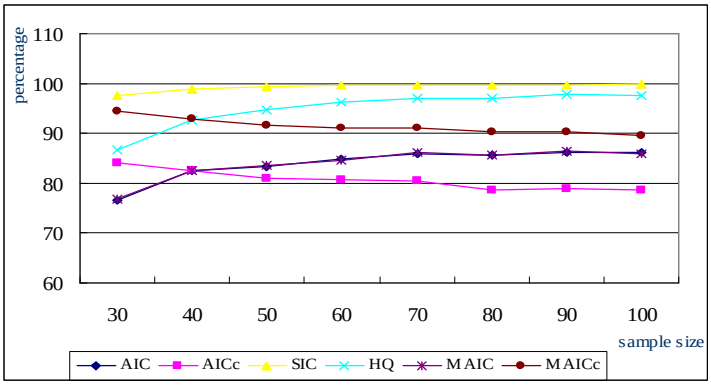
(c) Model three

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	77.78	84.33	97.02	87.54	77.88	94.06
Sample 40	82.03	82.24	98.66	92.71	82.56	92.01
Sample 50	82.97	79.95	99.48	94.45	82.88	91.03
Sample 60	84.93	81.10	99.52	96.67	85.17	91.92
Sample 70	85.09	80.08	99.69	96.74	85.04	90.79
Sample 80	85.70	79.70	99.67	97.12	85.44	90.35
Sample 90	85.96	79.45	99.70	96.92	85.53	89.60
Sample 100	86.29	79.73	99.85	97.54	86.38	89.79

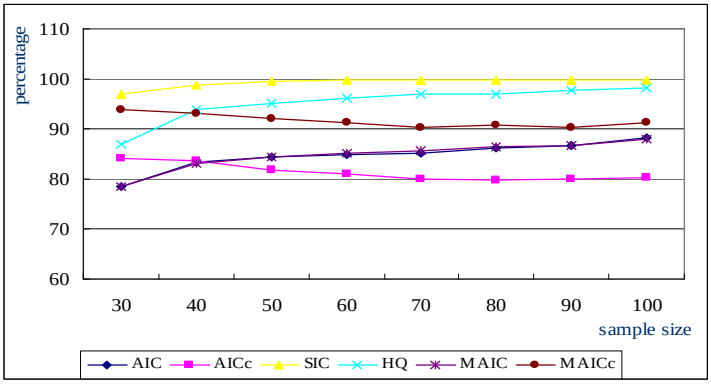
Figure 6.1 is a graphical representation of the six IC selection performances for the three VEC (0) models. The selection performances are considered as the sample size increases from 30 to 100. The results displayed lend support to the theory that as sample size increases so too do the IC's selection capabilities. The graphical displays show that the percentages of correct selection improve as sample size increases, except for AICc and MAICc criteria. This observation is due to the upward trend of the plots and is noticeable for all three VEC (0) models.

Figure 6.1: IC selection performance of VEC (0) for three models

(a) Model 1 — Lag length 0



(b) Model 2 — Lag length 0



(b) Model 3 — Lag length 0

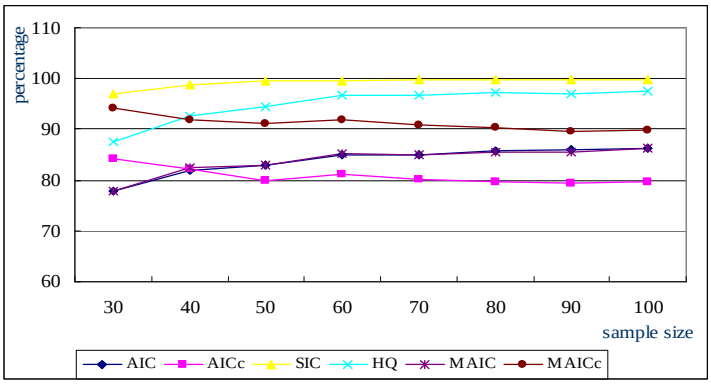


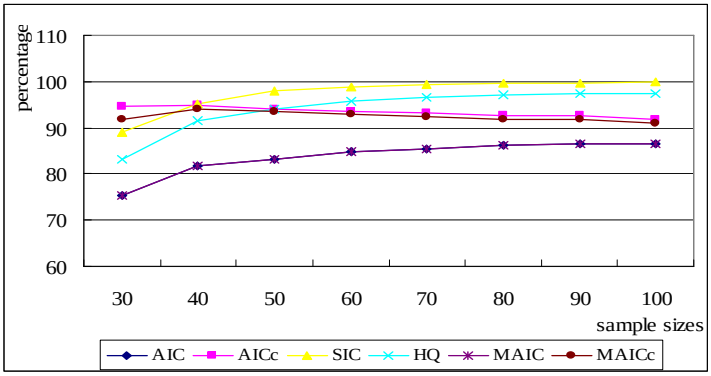
Figure 6.2 is a graphical representation of the average IC selection performances for the three parameterization models with lag lengths one to three, i.e. VEC (1) to VEC (3). The selection performances are considered as the sample size increases from 30 to 100.

In Figure 6.1 it was concluded that changing the parameters of the model made little impact on IC selection capabilities, hence in Figure 6.2, the average correct selection percentages for the different parameterized models are shown.

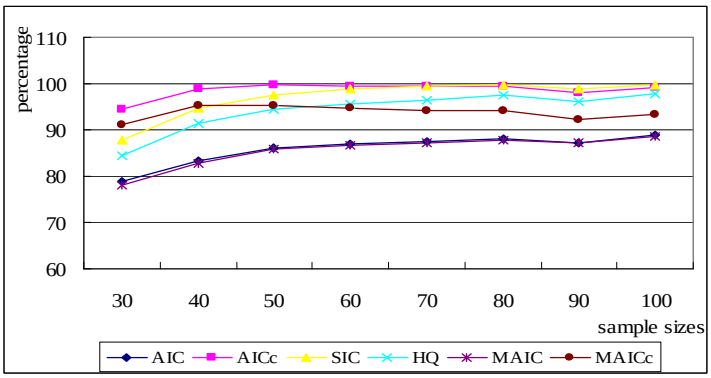
The results of the six IC selection capabilities for the VEC (3) model are quite different from the results of the VEC (1) and VEC (2) models. For the VEC (3) model, the graphical display illustrates that for all six IC their selection performance improved as sample size increased. In particular, the SIC, AICc and MAICc show a clear improvement in their performances.

Figure 6.2: The average IC correct lag selection for VEC (1) to VEC (3)

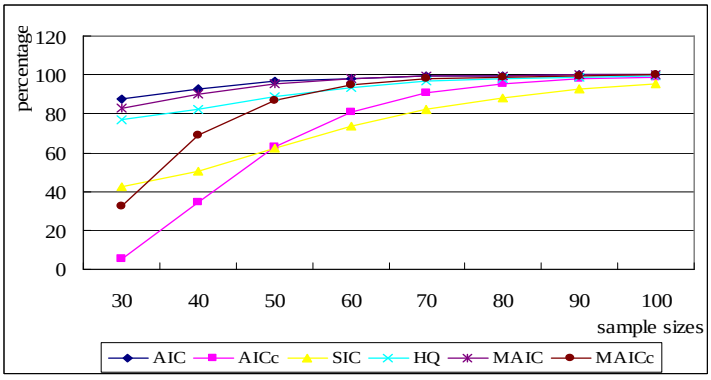
Average — VEC (1)



Average —VEC (2)



Average —VEC (3)



6.1.2 “Does not meet specification” group

The results already discussed have only considered the selection performances of the IC when considering the “ideal” datasets, i.e. the datasets that met model specification. This section briefly considers the “not ideal” datasets, i.e. the datasets which did not meet model specification. The results are shown graphically for ease of interpretation and comparison with the previous section.

Table 6.2 contains the results of the six IC performances for the three models with zero lag terms in the error correction representation. Surprisingly, the results are similar to those obtained for the “meet specification” group. The different parameterized models show little difference in the selection capabilities of the six IC. In general, the lag length selection capabilities of the IC are similar to those in the previous section.

Two observations can be concluded:

1. In Table 6.2, the results of the six IC selection capabilities for the VEC(0) models with sample sizes 30 to 100 provide evidence that there are no clear differences as sample size increases. The lag length selection capabilities of the six IC are also not adversely influenced by changing the parameters of the model.

2. SIC selects the correct lag length more often than the other five criteria for the VEC (0) model. From the simulated results, the best performing IC for model selection of the VEC(0) models is SIC. This is consistent for sample sizes 30 to 100. The average correct selection by AIC is 82.5%, AICc is 79.4%, SIC is 99%, HQ is 94.2%, MAIC is 77.5% and MAICc is 88.6%.

Table 6.2: The correct lag selected by various criteria of VEC(0) model with small sample sizes

(a) Model one

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	75.46	83.19	97.12	85.82	72.26	91.83
Sample 40	78.79	80.93	97.62	91.15	74.57	88.65
Sample 50	80.10	77.31	99.15	93.23	76.29	87.38
Sample 60	83.68	80.90	99.40	95.82	78.81	89.55
Sample 70	86.84	80.92	99.67	97.04	80.15	89.69
Sample 80	85.36	79.91	99.47	97.01	81.41	89.00
Sample 90	84.97	78.06	99.89	96.10	80.07	87.75
Sample 100	86.29	78.32	99.79	97.34	83.10	89.48

(b) Model two

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	77.64	83.98	97.24	86.89	76.26	93.44
Sample 40	79.64	79.71	98.54	92.45	76.92	90.93
Sample 50	80.30	77.18	99.00	92.02	75.69	87.91
Sample 60	84.80	79.92	99.81	96.88	75.44	89.08
Sample 70	82.00	76.31	99.54	94.31	77.68	87.02
Sample 80	86.03	79.66	100.00	97.06	79.41	87.25
Sample 90	83.09	77.62	99.68	96.94	80.52	86.96
Sample 100	86.32	77.86	99.75	97.26	80.85	88.56

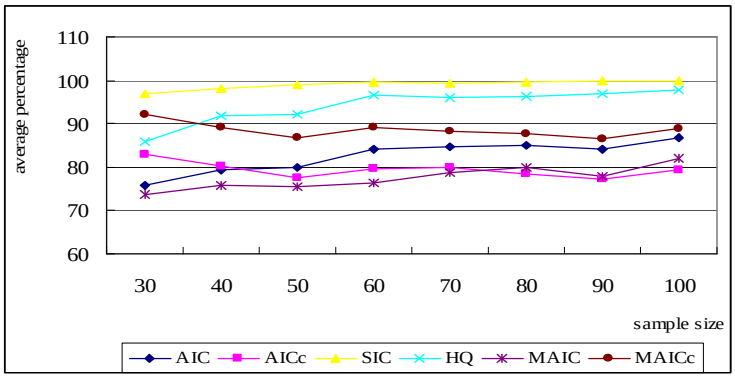
(c) Model three

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	74.52	81.46	96.38	85.32	72.29	91.08
Sample 40	79.65	80.16	97.76	91.66	75.48	88.10
Sample 50	79.70	77.99	98.67	91.27	74.19	85.39
Sample 60	83.90	78.09	99.81	97.19	75.09	89.14
Sample 70	85.58	82.54	99.05	96.77	78.18	88.24
Sample 80	83.54	75.95	99.10	95.30	78.84	86.80
Sample 90	83.90	75.66	100.00	97.94	72.85	84.46
Sample 100	87.55	81.73	99.80	98.80	82.53	88.96

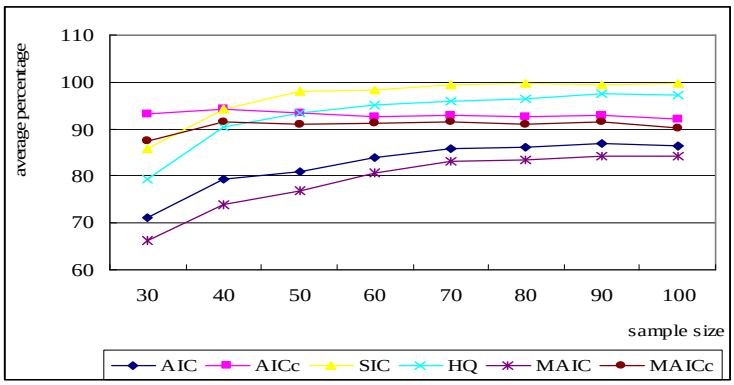
Figure 6.3 is an illustration of the lag length selection performances of the six IC for the average of the three parameterizations for VEC (0) to VEC (3) models. SIC dominates the selection performances for the VEC (0) and VEC (1) models, has reasonable lag length selection capabilities for VEC (2) whilst under performs for the VEC (3) model. AICc and MAICc show clear improvement in the performances for the VEC (3) model whilst AIC, HQ and MAIC perform adequately for the VEC (0) to VEC (3) models.

Figure 6.3: The average corrected lag selected by IC for VEC(0) to VEC(3) models

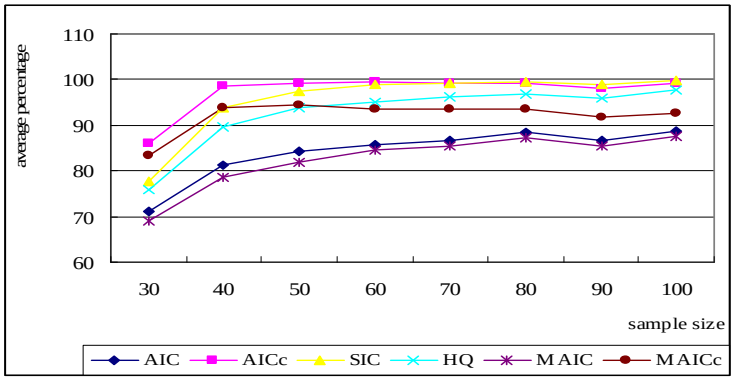
Average — VEC (0)



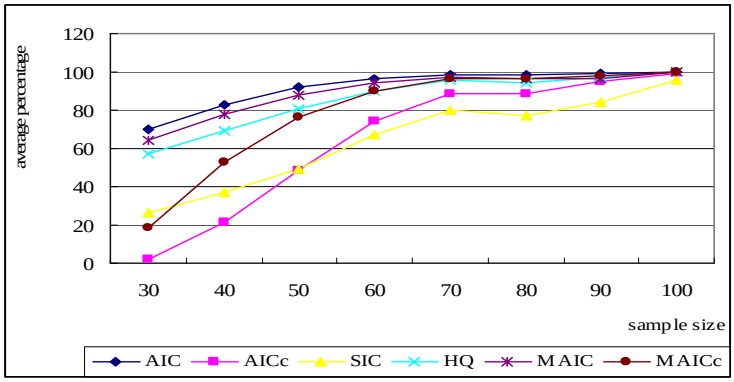
Average —VEC (1)



Average —VEC (2)



Average —VEC (3)



6.2 How do the information criteria perform in small sample sizes?

To answer this question the results were summarized by VEC model for sample size 30, sample size 40 and sample size 50. In addition we consider how well the six IC performed with three small samples was considered by subdividing into the two groups “meets specification” and “does not meet specification”.

6.2.1 “Meets specification” group

To compare the performances of the six IC in small samples, the lag selection performances were considered for lag lengths zero to lag length three. The average correct lag selection for the IC for all sample sizes 30 to 100 and for the smaller sample sizes 30 to 50 were calculated and are shown in Table 6.3 below. The values highlighted in the table are the correct lag selection percentages for the sample size 30 to 50 for each of the IC.

Three observations can be concluded:

1. The different parameterizations of the three models indicate little effect on the selection capability of the six IC for the smaller samples. In Table 6.3, the results show that each IC has similar selection capabilities for lag length zero to lag length two. SIC outperforms the other five criteria, HQ performs just less than SIC whilst AIC and MAIC show similar performances. It is worth noting that AICc and MAICc perform well in the three smaller samples.

2. The performances of SIC, HQ, AIC and MAIC as measured by their average correct selection in smaller samples is lower than the average for all sample sizes for the VEC (0) to VEC (3) models. The performances of AICc and MAICc in smaller samples is better than the average for all sample sizes for the VEC (0) and

VEC (1) model. In general, the selection performances of AICc and MAICc are better for the smaller sampler than AIC, HQ and MAIC for the VEC (0) to VEC (2) models.

3. The performance SIC as measured by the average correct selection for all samples (30 to 100) is better than all other IC for VEC (0) to VEC (2) models. This pattern only changes for VEC (3) model where AIC is the best performer. The SIC performance for the VEC (0) is 99.2%, VEC (1) is 97.5%, VEC (2) is 97.1% and drops to 73.4% for the VEC (3) model. When compared to the smaller samples (30 to 50), these estimates decrease to 98.6%, 95.3%, 97.2%, and 51.6%, respectively. This is a clear indication that as sample size increases so too does the selection performance of SIC. There is one anomaly with this observation. The performances of MAICc and AICc for VEC (0) to VEC (2) models are better for the smaller samples (30 to 50) than for the whole sample (30 to 100). This result is not evident for the VEC (3) model.

Table 6.3: The average correct lag selected by IC of three models for lag length zero to lag length two

(a) Lag length zero

model	sample	AIC	AICc	SIC	HQ	MAIC	MAICc
model 1	30 to 100	83.93	80.65	99.26	95.00	83.99	91.43
	30 to 50	80.82	82.54	98.56	91.36	80.98	93.04
model 2	30 to 100	84.65	81.32	99.27	95.20	84.70	91.60
	30 to 50	82.10	83.18	98.43	92.01	81.99	92.92
model 3	30 to 100	83.84	80.82	99.20	94.96	83.86	91.19
	30 to 50	80.93	82.17	98.39	91.57	81.11	92.37

(b) Lag length one

model	sample	AIC	AICc	SIC	HQ	MAIC	MAICc
model 1	30 to 100	83.91	93.70	98.05	94.50	84.00	92.79
	30 to 50	80.67	95.01	95.32	90.28	80.79	94.05
model 2	30 to 100	83.97	93.86	98.97	94.60	84.07	93.25
	30 to 50	80.65	95.41	97.84	90.72	80.68	94.94
model 3	30 to 100	83.24	92.96	95.39	93.45	83.20	91.34
	30 to 50	79.06	93.27	89.02	87.85	78.86	90.30

(c) Lag length two

model	sample	AIC	AICc	SIC	HQ	MAIC	MAICc
model 1	30 to 100	85.39	99.05	98.33	94.08	85.25	94.50
	30 to 50	82.78	99.76	97.23	90.66	82.62	96.68
model 2	30 to 100	86.06	97.74	95.00	94.12	85.48	92.56
	30 to 50	82.67	94.99	87.85	89.27	81.30	90.06
model 3	30 to 100	86.16	99.01	97.88	94.52	85.97	94.26
	30 to 50	82.87	98.23	94.89	90.43	82.69	94.78

(d) Lag length three

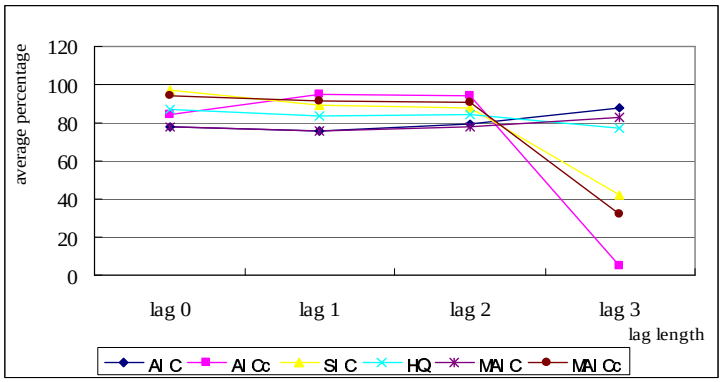
model	sample	AIC	AICc	SIC	HQ	MAIC	MAICc
model 1	30 to 100	96.81	68.37	69.17	91.19	95.61	84.01
	30 to 50	92.72	31.62	46.93	82.09	89.88	61.03
model 2	30 to 100	95.90	68.23	67.44	89.60	94.91	82.89
	30 to 50	90.32	29.98	43.99	78.73	87.96	58.40
model 3	30 to 100	97.50	76.16	83.63	94.69	96.48	87.81
	30 to 50	93.63	41.64	64.00	87.16	91.03	68.42

Figure 6.4 is a graphical representation of the average IC selection performances for the four error correction models VEC (0) to VEC (3). The selection performances are illustrated for sample sizes 30, 40 and 50.

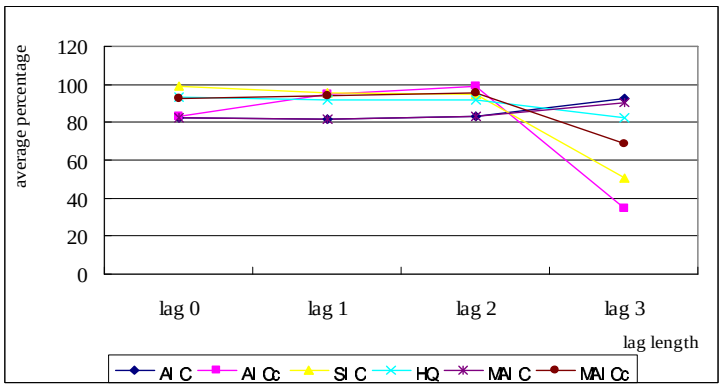
The graphs clearly illustrate that the lag three model is the most difficult to select correctly. When the sample size is 30, AICc selection capability is a paltry 4.6% whilst MAICc is a low 32.4%. As sample size increases from 30 to 50, all six IC improve their performances.

Figure 6.4: The average IC correct lag selection for VEC (0) to VEC (3) with small sample size 30, 40 and 50

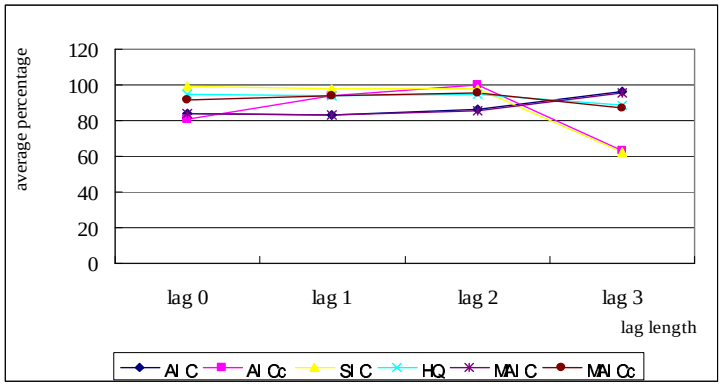
(a) Sample size 30



(b) Sample size 40



(c) Sample size 50

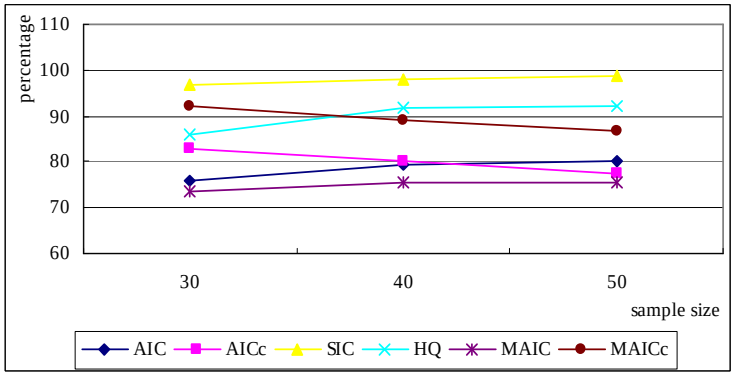


6.2.2 "Does not meet specification" group

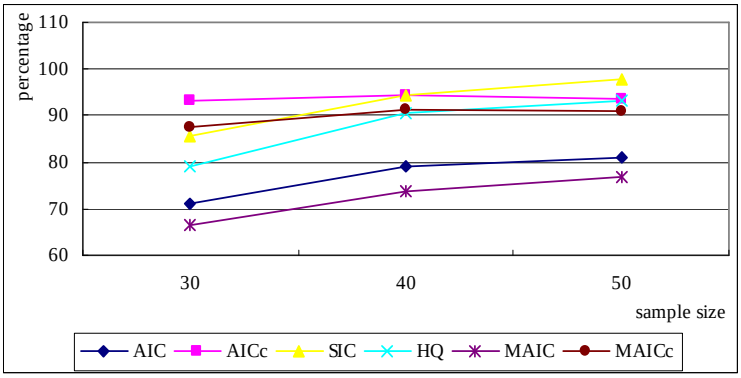
Figure 6.5 illustrates graphically the average selection performances of the six IC for the four error correction models with sample sizes 30, 40 and 50. The plots shown are selection performances as a function of sample size, the general increasing trend supports the previous findings in 6.2.1.

Figure 6.5: The average IC correct lag selection for VEC (0) to VEC (3) with small sample size 30, 40 and 50

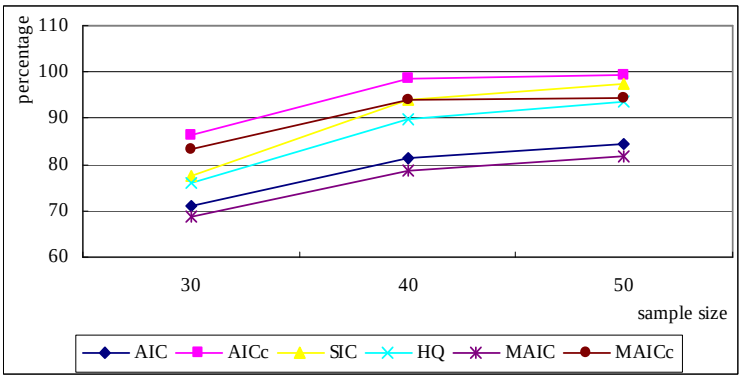
(a) Lag length zero



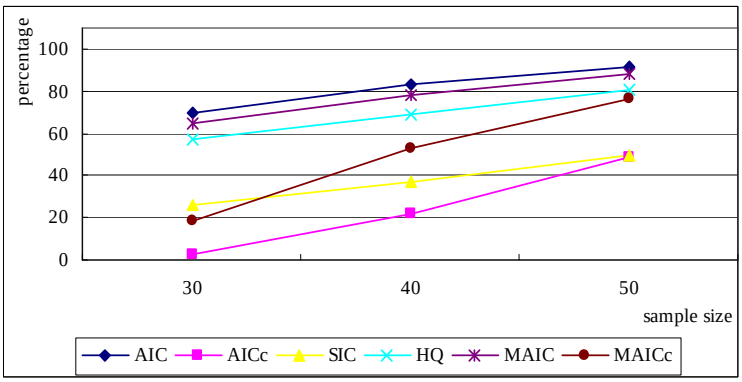
(b) Lag length one



(c) Lag length two



(d) Lag length three



6.3 How do the information criteria perform as the models' lag lengths increase?

To answer this question the results are summarized by VEC model for sample sizes increasing from 30 to 100. In addition, how consistently the six IC perform in the twelve cointegrated models was considered by subdividing into the two groups "meets specification" and "does not meet specification".

6.3.1 "Meets specification" group

Figure 6.4 shows the average correct lag selection for the six IC of the three parameterizations for VEC (0) to VEC (3) models. The sample sizes are used are 60 to 100 and the smaller sample size 30 to 50.

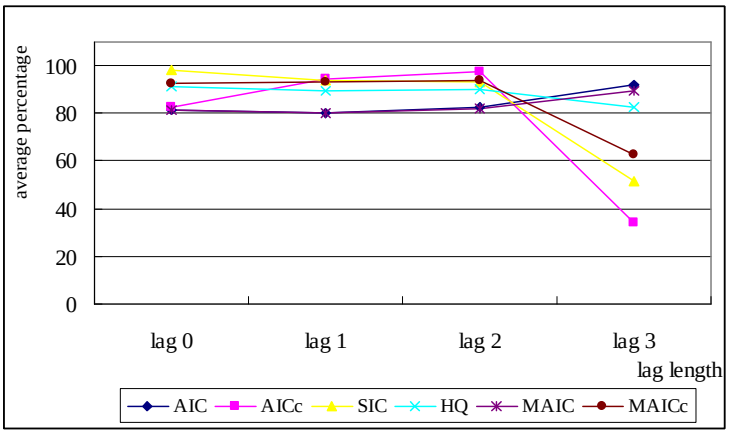
Two observations can be concluded:

1. When considering the smaller samples performances, it is noticeable that the lag length three model is the most difficult to determine. This is logical because as the number of lag terms increase in a model, degrees of freedom are lost, estimation becomes more difficult as parameter values are more likely to be close to zero and therefore considered insignificant.

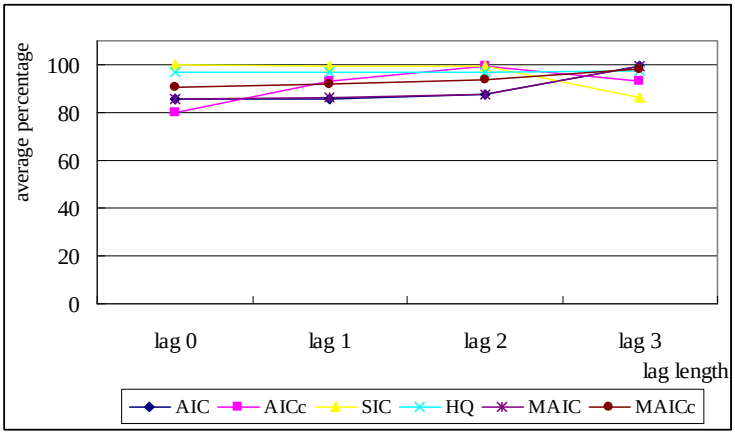
2. The larger sample (60-100) performance is promising. Nearly all IC performances exceed 80% for all four VEC models. The general exception to this is SIC for the VEC (3) with samples sizes 60 and 70. In many cases the selection performances are excellent, exceeding 90%.

Figure 6.4: The average IC correct lag selection for VEC (0) to VEC (3)

(a) Sample sizes 30-50



(b) Sample sizes 60-100



The summary in Table 6.5 provides the complete assessment of the first parameterized VEC (3) model according to sample size.

The results show that AICc, SIC and MAICc deliver poor performances in sample sizes 30 and 40. The AICc, SIC and MAICc tend to under-estimate the correct lag structure whilst AIC does well with 89.2% estimate. MAICc improves considerable for sample sizes of 50 and above whilst SIC's improvement is more noticeable for

sample sizes of 70 and above. The VEC (3) model is the most difficult of the models assessed to determine. All six IC perform better in selecting the correct lag length as the sample size increases.

Table 6.5: Tabulated percentages by IC for first parameterized VEC (3) model

		AIC	AICc	SIC	HQ	MAIC	MAICc
sample 30	Lag 0	3.33	26.79	40.83	8.39	4.52	32.74
	Lag 1	5.06	51.90	16.85	9.64	6.96	28.75
	Lag 2	2.44	16.67	2.50	3.10	3.87	6.13
	Lag 3	89.17	4.64	39.82	78.87	84.64	32.38
sample 40	Lag 0	0.85	5.70	29.67	5.39	1.24	7.29
	Lag 1	4.00	37.82	20.33	9.81	5.20	18.74
	Lag 2	2.56	24.28	4.03	3.96	3.30	7.72
	Lag 3	92.59	32.20	45.97	80.84	90.26	66.25
sample 50	Lag 0	0.12	0.85	19.87	1.83	0.15	1.00
	Lag 1	1.46	17.86	20.02	7.18	2.31	8.31
	Lag 2	2.01	23.25	5.11	4.44	2.80	6.24
	Lag 3	96.41	58.03	54.99	86.55	94.74	84.45
sample 60	Lag 0	0	0.06	10.00	0.72	0.03	0.14
	Lag 1	0.75	7.60	16.85	3.67	0.91	2.49
	Lag 2	1.22	16.71	6.33	3.31	1.66	3.73
	Lag 3	98.04	75.64	66.82	92.29	97.40	93.65
sample 70	Lag 0	0	0	4.60	0.11	0	0
	Lag 1	0.11	2.35	11.82	1.62	0.19	0.49
	Lag 2	0.76	10.04	6.30	2.73	0.92	2.03
	Lag 3	99.13	87.61	77.27	95.54	98.89	97.48
sample 80	Lag 0	0	0	1.87	0	0	0
	Lag 1	0.11	0.87	8.47	0.95	0.11	0.33
	Lag 2	0.52	5.48	5.29	1.47	0.62	1.06
	Lag 3	99.38	93.65	84.36	97.58	99.27	98.62
sample 90	Lag 0	0	0	0.40	0	0	0
	Lag 1	0	0.38	5.29	0.43	0	0.03
	Lag 2	0.11	2.94	3.86	0.94	0.16	0.38
	Lag 3	99.89	96.68	90.45	98.62	99.84	99.60
sample 100	Lag 0	0	0	0.14	0	0	0
	Lag 1	0	0.11	2.93	0.22	0	0
	Lag 2	0.14	1.42	3.26	0.56	0.20	0.33
	Lag 3	99.86	98.47	93.67	99.22	99.80	99.67

6.3.2 "Does not meet specification" group

The Table 6.6 shows that the average correction selection of the six IC for the four error correction models with sample sizes 30 to 100.

Table 6.6: The average correct lag selected by various criteria for VEC(0) to VEC(3) models with small sample sizes

(a) Lag length zero

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	75.87	82.88	96.91	86.01	73.60	92.12
Sample 40	79.36	80.27	97.97	91.75	75.66	89.23
Sample 50	80.03	77.49	98.94	92.17	75.39	86.89
Sample 60	84.13	79.64	99.67	96.63	76.45	89.26
Sample 70	84.81	79.92	99.42	96.04	78.67	88.32
Sample 80	84.98	78.51	99.52	96.46	79.89	87.68
Sample 90	83.99	77.11	99.86	96.99	77.81	86.39
Sample 100	86.72	79.30	99.78	97.80	82.16	89.00

(b) Lag length one

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	71.15	93.08	85.70	79.25	66.34	87.37
Sample 40	79.21	94.20	94.28	90.54	73.93	91.39
Sample 50	80.92	93.45	97.94	93.34	76.97	90.92
Sample 60	83.87	92.69	98.26	95.17	80.66	91.31
Sample 70	85.73	93.00	99.48	96.00	82.98	91.64
Sample 80	86.10	92.49	99.75	96.43	83.32	91.05
Sample 90	86.91	92.76	99.49	97.44	84.12	91.49
Sample 100	86.47	92.09	99.64	97.24	84.24	90.03

(c) Lag length two

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	71.02	86.19	77.69	76.02	68.92	83.27
Sample 40	81.39	98.56	93.87	89.70	78.51	93.87
Sample 50	84.33	99.32	97.50	93.75	81.78	94.30
Sample 60	85.65	99.52	98.83	95.03	84.48	93.62
Sample 70	86.52	99.32	99.21	96.31	85.57	93.51
Sample 80	88.31	99.23	99.60	96.85	87.19	93.52
Sample 90	86.77	98.16	98.94	96.08	85.45	91.64
Sample 100	88.71	99.10	99.78	97.61	87.67	92.78

(d) Lag length three

	AIC	AICc	SIC	HQ	MAIC	MAICc
Sample 30	70.03	2.31	26.31	57.27	64.24	18.45
Sample 40	83.09	21.74	36.86	69.11	78.19	52.64
Sample 50	91.88	48.56	49.51	80.46	88.00	76.10
Sample 60	96.72	73.94	67.40	89.80	94.13	90.35
Sample 70	98.87	88.92	80.23	95.39	97.48	96.37
Sample 80	98.53	88.80	76.96	94.28	96.40	96.21
Sample 90	99.47	94.75	84.18	97.08	96.60	97.73
Sample 100	99.98	99.31	95.65	99.71	99.83	99.91

The simulation results of the VEC (0) to VEC (2) show that the six IC perform well in selecting the correct lag length for the three models. As in the "meets specification assessment", the results for the VEC(3) model are quite different when compared to VEC (0) to VEC (2). The summary in Table 6.7 is similar to that of Table 6.5 except that the SIC performance is worse for the smaller samples. The results indicate that SIC tends to under-fit the model more frequently than other IC. The six IC perform better in selecting the correct lag structure length as the sample size increases.

Table 6.7: Tabulated percentages by IC for first parameterized VEC (3) model

		AIC	AICc	SIC	HQ	MAIC	MAICc
sample 30	Lag 0	6.75	24.94	50.23	15.58	7.63	35.20
	Lag 1	11.84	48.57	19.74	16.75	13.98	33.33
	Lag 2	13.51	24.85	8.60	13.39	15.82	16.05
	Lag 3	67.89	1.64	21.43	54.27	62.57	15.41
sample 40	Lag 0	1.59	4.96	32.99	6.38	1.94	7.57
	Lag 1	6.66	34.89	22.76	13.52	8.37	22.36
	Lag 2	10.67	41.24	11.14	13.48	13.08	20.58
	Lag 3	81.09	18.91	33.11	66.61	76.61	49.48
sample 50	Lag 0	0.06	0.61	22.05	1.65	0.17	0.88
	Lag 1	2.21	17.20	21.28	8.16	3.36	8.43
	Lag 2	7.06	39.91	14.28	13.34	8.88	17.36
	Lag 3	90.68	42.28	42.39	76.85	87.60	73.32
sample 60	Lag 0	0	0	11.15	0.54	0.07	0.20
	Lag 1	0.61	6.89	16.28	4.12	1.28	3.04
	Lag 2	3.65	26.42	12.43	8.04	4.73	8.65
	Lag 3	95.74	66.69	60.14	87.30	93.92	88.11
sample 70	Lag 0	0	0	3.42	0.14	0	0
	Lag 1	0.28	2.85	11.25	2.21	0.57	1.14
	Lag 2	1.35	11.18	7.62	3.92	2.21	3.42
	Lag 3	98.36	85.97	77.71	93.73	97.22	95.44
sample 80	Lag 0	0	0	16.35	0.68	0.14	0.14
	Lag 1	0.61	4.93	18.18	5.14	1.15	1.55
	Lag 2	3.65	21.35	12.84	8.45	8.31	8.31
	Lag 3	95.74	73.72	52.64	85.74	90.41	90.00
sample 90	Lag 0	0	0	0.36	0	0	0
	Lag 1	0	0.07	4.73	0.22	0	0.07
	Lag 2	0.29	2.87	4.09	1.22	0.43	0.79
	Lag 3	99.71	97.06	90.82	98.57	99.57	99.14
sample 100	Lag 0	0	0	0.20	0	0	0
	Lag 1	0.07	0.13	2.58	0.13	0.07	0.07
	Lag 2	0	0.79	2.12	0.26	0.07	0.07
	Lag 3	99.93	99.07	95.11	99.60	99.87	99.87

The summary in Table 6.8 provides the averages of selection performances for all models by their lag structure. The results are summarized for each IC assessed in this study. The results show that SIC performs well for VEC (0) to VEC (2) models but less impressively for VEC (3) models. Overall HQ is the best performer with an average of 91.37% correct selections followed by SIC with 89.04% and MAICc with 87.45%. In general, AIC and HQ are the recommended methods of assessment.

Table 6.8: The average correct lag length selected by IC of three models for VEC(0) to VEC(3) with small sample sizes

	AIC	**AICc**	**SIC**	**HQ**	**MAIC**	**MAICc**
lag 0	82.49	79.39	99.01	94.23	77.45	88.61
lag 1	82.54	92.97	96.82	93.18	79.07	90.65
lag 2	84.09	97.43	95.68	92.67	82.45	92.06
lag 3	92.32	64.79	64.64	85.39	89.36	78.47
average	85.36	83.64	89.04	91.37	82.08	87.45

Chapter 7

Conclusion and Further work

7.1 Conclusion

In this study, six model selection criteria were used to determine the number of lag lengths in twelve bivariate cointegrated models. Data was simulated by Excel for sample sizes of 30, 40, 50, 60, 70, 80, 90 and 100. To reduce the impact of starting values, the first three hundred observations were omitted. The data were then used to test for unit roots and cointegration for each sample size. The number of observations was reduced by four in all samples (i.e. sample size 26, 36, 46, 56, 66, 76, 86 and 96) to account for at most four VAR lag terms. In this way the same number of observations was used for each lag length model during evaluation.

The selection capabilities of the six information criteria were compared and which criterion most correctly estimated the true lag of the models was determined. In the VEC (0), VEC (1) and VEC (2) models, SIC selected the correct lag length most often. In the VEC (3) model, SIC was less accurate than other criteria. For smaller sample sizes, AICc, SIC and MAICc under-estimated the correct lag for the VEC (3) model quite often. The criterion HQ was consistent and the order was estimated correctly more often when the sample size increased. The criteria AIC and MAIC were consistent performers for the lag lengths in all small sample sizes, particularly in higher order.

The results of this study lend support to the results of Lütkepohl (1985) and Liu (2007). In the finding of Lütkepohl (1985), SIC had the best performance for three sample sizes (i.e. 40, 100 and 200). Liu (2007) found that AIC outperformed the other criteria in higher orders.

7.2 Further work

Model selection criteria were used to determine the correct lag structure in this study. There are few ways to follow the study of model selection in CVAR models.

This study can be continued by considering other selection criteria methods, i.e. Gonzalo and Pitarakis (2002) IC. The models dimensions can be increased in two ways, the lag structure extended and the number of equations or variables in the model can be increased. All these would be of use to practicing economists who use VAR/VEC modeling.

Secondly, the sample size can be extended. In this study, we focused on the small sample sizes, sample size 30 to sample size100. Liu (2007) simulated sample sizes of 30, 60, 120 and 240 observations. Liew (2004) used sample sizes of 30, 60, 120, 240, 480 and 960.

References

Aggarwal, R. & Kyaw, N. A. (2005). Equity market integration in the NAFTA region: Evidence from unit root and cointegration tests. International Review of Financial Analysis, vol. 14, pp. 393-406.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Cszaki (Eds) 2nd International Symposium on Information Theory, Budapest: Akademiai Kiado, pp. 267-281.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transations on Automatic Control, vol. 19, pp. 716-723.

Almpanidis, G. & Kotropoulos, C. (2008). Phonemic segmentation using the generalized Gamma distribution and small sample Bayesian information criterion. Speech Communication, vol. 50, pp. 38-55.

Assaf, A. (2006). Nonlinear trend stationarity in real exchange rates: Evidence from nonlinear adf tests. Annals of Economics and Finance, vol. 2, pp. 283–294.

Baliamoune-Lutz, M. & Haughton, J. (2004). Velocity effects of increased variability in monetary growth in Egypt: A test of Friedman's velocity hypothesis. African Development Review, vol.16, no.1, pp. 36-52.

Baltagi, B. H. & Wang, Z. J. (2006). Testing for cointegrating rank via model selection: evidence from 165 data sets. Center for Policy Research, Maxwell School of Citizenship and Public Affairs Syracuse University, Working Paper, no. 83.

Bedrick, E. J. & Tsai, C. L. (1994). Model selection for multivariate regression in small samples. Biometrics, vol. 50, pp. 226-231.

Bengtsson, T. & Cavanaugh, J. E. (2006). An improved Akaike information criterion for state-space model selection. Computational Statistics & Data Analysis, vol. 50, pp. 2635-2654.

Bentzen, J. & Engsted, T. (2001). A revival of the autoregressive distributed lag model in estimating energy demand relationships. Energy, vol. 26, pp. 45–55.

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. Statistics & Probability Letters, vol. 33, pp. 201-208.

Chao, J. C. & Phillips, P. C. B. (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. Journal of Econometrics, vol. 91, pp. 227-271.

Cubadda, G. (2007). A unifying framework for analysing common cyclical features in cointegrated time series. Computational Statistics & Data Analysis, vol. 52, pp. 896 – 906.

Davros, N. M., Debinski, D. M., Reeder, K. F. & Hohman, W. L. (2006). Butterflies and continuous conservation reserve program filter strips: Landscape considerations. Wildlife Society Bulletin, vol. 34, no. 4, pp. 936-943.

Diamandis, P. F., Georgoutsos, D. A. & Kouretas, G. P. (1996). “Cointegration tests of the monetary exchange rate model: the Canadian-US dollar, 1970-1994.” International Economic Journal, vol. 20, no. 10, pp. 83-97.

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, vol. 74, pp. 427-431.

Dickey, D. A. & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, vol. 49, pp. 1057–1072.

Enders, W. (2004). Applied Econometric Time Series. 2nd Edition. John Wiley and Sons Inc.

Engle, R. F. & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. Econometrica, vol. 55, pp. 251-276.

Engsted, T. (2006). Explosive bubbles in the cointegrated VAR model. Finance Research Letters, vol. 3, no. 2, pp. 154-162.

Engsted, T. & Tanggaard, C. (1994). Cointegration and the US term structure. Journal of Banking & Finance, vol. 18, no. 1, pp. 167-181.

Gonzalo, J. & Pitarakis, J. (1998). Specification via model selection in vector error correction models. Economics Letters, vol. 60, pp. 321-328.

Gonzalo, J. & Pitarakis, J. (2002). Lag length estimation in large dimensional systems. Journal of Time Series Analysis, vol. 23, no. 4, pp. 401-423.

Granger, C. W. J. (1983). Cointegrated variables and error-correcting models. Discussion Paper 83-13, Department of Economics, University of California at San Diego.

Granger, C. W. J. & Jeon, Y. (2004). Forecasting performance of information criteria

with many macro series. Journal of Applied Statistics, vol. 31, no. 10, pp.1227–1240.

Granger, C. W. J. & Jeon, Y. (2006). Dynamics of model overfitting measured in terms of autoregressive roots. Journal of Time Series Analysis, vol. 27, no. 3, pp. 347–365.

Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society, series B, vol. 41, pp. 190-195.

Harris, R. (1995). Cointegration analysis in econometric modelling. 1st Edition, Prentice Hall, Pearson Education Limited.

Hecq, A., Palm, F. C. & Urbain, J. P. (2006). Common cyclical features analysis in VAR models with cointegration. Journal of Econometrics, vol. 132, pp. 117–141.

Hosmer, D. W., Jovanovic, B. & Lemeshow, S. (1989). Best subsets logistic regression. Biometrics, vol. 45, pp. 1265-1270.

Hurvich, C. M., Shumway, R. & Tsai, C. L. (1990). Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. Biometrika, vol. 77, no. 4, pp. 709-719.

Hurvich, C. M. & Tsai, C. L. (1989). Regression and time series model selection in small samples. Biometrika, vol. 76, pp. 297–307.

Hurvich, C. M. & Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive selection. Journal of Time Series Analysis, vol. 14, no. 3, pp. 271-279.

Hurvich, C. M. & Tsai, C. L. (1995). Model selection for extended quasi-likelihood

models in small sample. Biometrics, vol. 51, pp. 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, vol. 9, no. 1, article 2.

Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, vol. 12, pp. 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, vol. 59, pp. 1551-1580.

Johansen, S. (1995). Likelihood – based inference in cointegrated vector autoregressive models. Working Paper, Oxford University.

Johansen, S. & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. Oxford Bulletin of Economics and Statistics, vol. 52, pp. 169-210.

Johansen, S. & Swensen, A. R. (1999). Testing exact rational expectations in cointegrated vector autoregressive models. Journal of Econometrics, vol. 93, pp. 73-91.

Keblowski, P. & Welfe, A. (2004). The ADF–KPSS test of the joint confirmation hypothesis of unit autoregressive root. Economics Letters, vol. 85, no. 2, pp. 257-263.

Kellard, N. (2006). On the robustness of cointegration test when assessing market efficiency. Finance Research Letters, no. 3, pp. 57-64.

Kim, K. H. (2003). Dollar exchange rate and stock price: Evidence from multivariate

cointegration and error correction model. Review of Financial Economics, vol. 12, pp. 301-313.

Koreisha, S. G. & Pukkila, T. (1990). Determining the order of a vector autoregression when the number of component series is large. Journal of Time Series Analysis, vol. 14, no. 1, pp. 47-69.

Lau, S.-H. P. (2007). Using an error-correction model to test whether endogenous long-run growth exists. Journal of Economic Dynamics & Control, vol. 32, pp. 648–676.

Lee, D. G. (2007). Dynamic prediction model of as-built roughness in asphaltic concrete pavement construction. Journal of Transportation Engineering, vol. 133, no. 2, pp. 90-95.

Lee, S. Y. & Tsai, C. L. (1998). Model selection for causal models: the global procedure with AICc and AICu. Global Finance Journal, vol. 9, no. 2, pp. 205-223.

Liew, V. K. S. (2004). Which lag length selection criteria should we employ? Economic Bulletin, vol. 3, no. 33, pp. 1-9.

Lima, E. J. A., Luduvice, F. & Tabak, B. M. (2004). Forecasting interest rates: An application for Brazil. Working Paper, Research Department, Banco Central do Brasil.

Liu, L. (2007). Consistent testing for lag length in cointegrated relationships. MSc dissertation, Nelson Mandela Metropolitan University.

Lopez, J. H. (1997). The power of the ADF test. Economics Letters, vol. 57, no. 1, pp. 5-10.

Lütkepohl, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive process. Journal of Time Series Analysis, vol. 6, no.1, pp. 35-52.

Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer Verlag: Berlin.

Narayan, P. K. & Smyth, R. (2007). Mean reversion versus random walk in G7 stock prices evidence from multiple trend break unit root tests. International Financial Markets, Institutions and Money, vol. 17, pp. 152–166.

Ng, S. & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. Econometrica, vol. 69, pp. 1519-1554.

Ng, S. & Perron, P. (2005). Practitioners' corner: a note on the selection of time series models. Oxford Bulletin of Economics and Statistics, vol. 67, no.1, pp. 115-134.

Parzen, E. (1974). Some recent advances in time series modeling. IEEE Transactions on Automatic Control, AC-19, pp. 723-730.

Pelipas, I. (2006). Money demand and inflation in Belarus: Evidence from cointegrated VAR. Research in International Business and Finance, vol. 20, no. 2, pp. 200-214.

Pesaran, M. H., Shin, Y. & Smith, R. J. (2000). Structural analysis of vector error correction models with exogenous $I(1)$ variables. Journal of Econometrics, vol. 97, pp. 293-343.

Peter, B. (2007). A general bilinear model to describe growth or decline time profiles. Mathematical Biosciences, vol. 205, no. 1, pp. 108-136.

Phillips, P. C. B. & Perron, P. (1988). Testing for unit roots in time series regression. Biometrika, vol. 75, pp. 335-346.

Pickens, B. A. (2007). Understanding the population dynamics of a rare, polyvoltine butterfly. Journal of Zoology, vol. 273, no. 3, pp. 229-236.

Pregibon, D. (1979). Data analytic methods for generalized linear model. PHD Thesis, University of Toronto.

Qu, Z. J. & Perron, P. (2006). modified information criterion for cointegration tests based on a VAR approximation. Working Papers Series WP2006-011, Department of Economics, Boston University.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, vol. 6, no. 2, pp. 461-464.

Sclove, S. L. (1994). Some aspects of model-selection criteria. In H. Bozdogan (Ed), Proceedings of the first US/Japan conference on the frontiers of statistical modeling: an informational approach, Netherlands: Kluwer Academic Publishers, pp. 37-67.

Sharp, G. & Radloff, S. (2007). Lag length selection for error corrected model. Paper presented at the 56th section of International Statistical Institute conference, Lisbon, Portugal.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. The Annals of Statistics, vol. 8, pp. 147-164.

Shintani, M. (2001). A simple cointegrating rank test without vector autoregression Journal of Econometrics, vol. 105, pp. 337–362.

Sin, C. Y. & White, H. (1996). Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics, vol. 71, pp. 207-225.

Swanson, N. R. (1998). Money and output viewed through a rolling window. Journal of Monetary Economics, vol. 41, pp. 455-473.

Todani, K. R. (2005). A cointegrated VAR model of M3 demand in South Africa. Working Paper, South African Reserve Bank, Research Department.

Wang, Z. J. & Bessler, D. A. (2005). A Monte Carlo study on the selection of cointegrating rank using information criteria. Econometric Theory, vol. 21, no. 3, pp. 593 – 620.

Wang, Y. J. & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock-recruitment relationships. Fisheries Research, vol. 77, pp. 220-225.

Zachariadis, T. (2007). Exploring the relationship between energy use and economic growth with bivariate models: New evidence from G-7 countries. Energy Economics, vol. 29, no. 6, pp. 1233-1253.

Zhang, S. D. & Lowinger, T. C. (2005). Cointegration in a monetary model of exchange rate determination. ASBBS E-Journal, vol. 1, no. 1, pp. 1-18.

Zhou, W. X. & Sornette, D. (2006). Fundamental factors versus herding in the 2000-2005 US stock market and prediction. Physica A, vol. 360, pp. 459-482.