The University of Sydney

Copyright in relation to this thesis*

Under the Copyright Act 1968 (several provisions of which are referred to below), this thesis must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Under Section 35(2) of the Copyright Act 1968 'the author of a literary, dramatic, musical or artistic work is the owner of any copyright subsisting in the work'. By virtue of Section 32(1) copyright 'subsists in an original literary, dramatic, musical or artistic work that is unpublished' and of which the author was an Australian citizen, an Australian protected person or a person resident in Australia.

The Act, by Section 36(1) provides: 'Subject to this Act, the copyright in a literary, dramatic, musical or artistic work is infringed by a person who, not being the owner of the copyright and without the licence of the owner of the copyright, does in Australia, or authorises the doing in Australia of, any act comprised in the copyright'.

Section 31(1)(a)(i) provides that copyright includes the exclusive right to 'reproduce the work in a material form'. Thus, copyright is infringed by a person who, not being the owner of the copyright, reproduces or authorises the reproduction of a work, or of more than a reasonable part of the work, in a material form, unless the reproduction is a 'fair dealing' with the work 'for the purpose of research or study' as further defined in Sections 40 and 41 of the Act.

Section 51(2) provides that 'Where a manuscript, or a copy of a thesis or other similar literary work that has not been published is kept in a library of a university or other similar institution or in an archives, the copyright in the thesis or other work is not infringed by the making of a copy of the thesis or other work by or on behalf of the officer in charge of the library or archives if the copy is supplied to a person who satisfies an authorized officer of the library or archives that he requires the copy for the purpose of research or study'.

*Thesis' includes 'treatise', dissertation and other similar productions.
Bootstrapping in Vector Autoregressive Models

by

Jae H KIM

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Econometrics
Faculty of Economics
University of Sydney
1997
ABSTRACT

The bootstrap method has been used widely in statistical analyses of econometric and time series models. Its main attraction lies in small sample situations where asymptotic methods can often provide a poor approximation to the unknown sampling distribution of a statistic. The basic idea is that the observed data contains useful information about the population characteristics and resampling from it can yield a good approximation to the sampling distribution. On this basis, the bootstrap can provide statistics with superior small sample performances to those obtainable from asymptotic approximations.

Vector autoregressive (VAR) models are widely used in time series econometrics. Their structure often leads to a proliferation of parameters to be estimated. This may have implications for the applicability of the associated asymptotic theories used as bases for inferences in the small and moderate sample sizes encountered in many applications. The bootstrap method provides a natural alternative, and this thesis applies the bootstrap to small sample analyses of VAR models. Several fundamental issues - parameter estimation, (point and interval) forecasting and order selection - are dealt with. For the problem of VAR order selection, recently introduced bootstrap methods, such as the moving block and stationary bootstraps, are employed. A sensible bootstrap method should possess the property of asymptotic validity, which is an assurance that bootstrap statistics converge to the true population values in large samples. In this thesis, asymptotic validity of the bootstrap in various facets of VAR modelling is established.
Throughout the thesis, the bootstrap method is employed mainly in the following contexts of VAR modelling: firstly, to estimate small sample properties which cannot be captured by asymptotic methods (e.g., biases in parameter estimators and forecasts); secondly, to evaluate the reliability of asymptotic approximations in small samples (e.g., asymptotic prediction regions); and thirdly, to enhance small sample performances of a method which possesses asymptotically desirable properties (e.g. the use of VAR order selection criteria). It is found in general that the bootstrap provides a useful tool in estimating small sample properties of VAR models. It can sometimes provide a superior alternative to the asymptotic method in evaluating small sample properties of VAR. It is also found that improvement can be made by employing the bootstrap in small sample performances of methods, such as VAR order selection, whose attraction rests on asymptotic theories. Although some results merely extend the previous findings in the univariate AR case, many useful results which provide new insights are also found in this thesis.

Rapid decline of computation cost has been experienced in recent years. This, coupled with small sample performances of the bootstrap found to be desirable in many applications including VAR, will make the use of the bootstrap more widespread and popular in many aspects of applied econometric and time series analyses.
ACKNOWLEDGEMENTS

My sincere thanks to Dr. M. D. Smith and Professor A. R. Tremayne for their first-rate supervision and guidance. Although a situation beyond his control prematurely ended Professor Tremayne's tenure as supervisor, his early endeavours contributed substantially to this thesis. Dr. Smith had been associate supervisor before he kindly took over the role of sole supervisor at the final stage of the thesis. He made many invaluable suggestions which, in many ways, led to substantial improvements of the thesis. Without his enthusiasm and effort, the completion of the thesis may not have been possible.

I would also like to thank Associate Professors R. Bartels and D. G. Fiebig and Professor A. D. Woodland for their warm supports and encouragements in numerous ways. Computing and other facilities provided by the Faculty of Economics and the Department of Econometrics of the University of Sydney are gratefully acknowledged. Professor T. Breusch made valuable comments, as a discussant of the paper I presented at the 1994 PhD Conference in Canberra, which substantially improved the content of the thesis.

My deepest gratitude, by all means, is directed to my family, especially to my parents to whom this thesis is dedicated.
# TABLE OF CONTENTS

**ABSTRACT**

**ACKNOWLEDGEMENTS**

**LIST OF DEFINITIONS, THEOREMS AND LEMMAS**

1. **INTRODUCTION**
   
   1.1 Motivation
   
   1.2 Literature Review
      
      1.2.1 An Introduction to the Bootstrap
      
      1.2.2 Applications to Time Series Econometrics
      
      1.2.3 Bootstrap Methods for Dependent Data
   
   1.3. Concluding Remarks

2. **VAR MODELLING AND BOOTSTRAPPING**
   
   2.1 Introduction
   
   2.2 VAR Modelling
   
   2.3 Backward Representation of VAR Models
   
   2.4 Bootstrapping in VAR Models
      
      2.4.1 Bootstrapping with Resampling Residuals
      
      2.4.2 Bootstrapping with Resampling Blocks
   
   2.5 Concluding Remarks
3. BIASES OF VAR PARAMETER ESTIMATORS

3.1 Introduction .................................................. 73
3.2 Response Surfaces ............................................. 77
3.3 Experimental Design for Response Surfaces .............. 84
3.4 Constructing Response Surfaces ............................ 85
3.5 Some Out-of-Sample Forecasting Results .................. 93
3.6 Effect of Unknown Intercepts ............................... 94
3.7 Biases Under Non-Normal Innovations ..................... 96
3.8 Concluding Remarks .......................................... 101

Appendix 3.1 Estimated Response Surfaces for VAR(1) Models 104
Appendix 3.2 Estimated Response Surfaces for VAR(2) Models 106
Appendix 3.3 Tables ............................................... 108
Appendix 3.4 Figures ............................................. 111

4. VAR FORECASTING AND BOOTSTRAPPING

4.1 Introduction .................................................. 117
4.2 Tests for Skewness and Kurtosis ............................ 123
4.3 Resampling Backward Residuals ............................. 126
4.4 VAR Forecast Biases ......................................... 129
4.5 Variability of VAR Forecasts ............................... 133
4.6 Effect of Final Observations on Forecast Variability .... 135
4.7 Normality of VAR Forecast Error Distributions .......... 137
4.8 Concluding Remarks .......................................... 140

Appendix 4.1 Tables ............................................... 142
Appendix 4.2 Figures ............................................. 150
5. BOOTSTRAPPING PREDICTION REGIONS OF VAR MODELS

5.1 Introduction 157
5.2 Prediction Regions for VAR Forecasts 160
5.3 Experimental Design 164
5.4 Performance of Asymptotic Prediction Regions 167
5.5 Improved Prediction Cubes for VAR Forecasts 170
5.6 Concluding Remarks 177
Appendix 5.1 Tables 180
Appendix 5.2 Figures 183

6. VAR ORDER SELECTION AND BOOTSTRAPPING

6.1 Introduction 189
6.2 Experimental Design 198
6.3 Results and Discussions 200
6.4 Applications to Economic data 208
6.5 Concluding Remarks 211
Appendix 6.1 Tables 212
Appendix 6.2 Figures 218

7. CONCLUSION 224

REFERENCES 229
# List of Definitions, Theorems and Lemmas

## Definitions

<table>
<thead>
<tr>
<th>2.1 Similarity of Two Matrices</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Strict Stationarity</td>
<td>66</td>
</tr>
<tr>
<td>2.3 Ergodicity</td>
<td>66</td>
</tr>
</tbody>
</table>

## Theorems

<table>
<thead>
<tr>
<th>2.1 Asymptotic Properties of VAR Parameter Estimators</th>
<th>37</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Asymptotic Properties of Backward VAR Parameter Estimators</td>
<td>47</td>
</tr>
<tr>
<td>2.3 Asymptotic Properties of Bootstrap VAR Parameter Estimators</td>
<td>57</td>
</tr>
<tr>
<td>2.4 Asymptotic Properties of Bootstrap VAR Forecasts</td>
<td>58</td>
</tr>
<tr>
<td>2.5 Ergodic Theorem</td>
<td>67</td>
</tr>
<tr>
<td>2.6 Asymptotic Properties Bootstrap Autocovariances</td>
<td>68</td>
</tr>
<tr>
<td>2.7 Asymptotic Properties Bootstrap VAR Parameter Estimators</td>
<td>70</td>
</tr>
<tr>
<td>3.1 Correlation between Two Chi-Squared Random Variables</td>
<td>99</td>
</tr>
<tr>
<td>6.1 Asymptotic Properties of Bootstrap VAR Order Selection</td>
<td>195</td>
</tr>
</tbody>
</table>

## Lemmas

<table>
<thead>
<tr>
<th>2.1 Relationship between Forward and Backward VAR Innovations</th>
<th>44</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Asymptotic Properties of VAR Parameter Estimators</td>
<td>50</td>
</tr>
<tr>
<td>6.1 Asymptotic Consequence of Under-Estimating a VAR Model</td>
<td>192</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 Motivation

Economists have been particularly concerned with time series modelling of economic variables since the pioneering work of Box & Jenkins (1976). Many economic variables are generated as time series being observed at discrete and equi-spaced intervals of time. An economic time series can be expressed as a function of past history including its own past, mainly because economic agents make their decisions based on an information set whose content consists largely of past history. In this context, the autoregressive moving average (ARMA) model introduced by Box & Jenkins (1976) has been widely used in modelling economic time series. For example, Nelson (1972) and Naylor et al. (1972) applied ARMA modelling to a number of macroeconomic time series and found that ARMA models can provide superior forecasts to those obtainable from large scale econometric models; see, for details of large scale econometric models, Klein (1986) and Wallis (1988).

The usefulness of ARMA modelling, however, is somewhat restricted by the fact that it is a univariate approach. That is, a time series is modelled only as a linear function of its own past, totally ignoring possible interactions with other series. The inter-dependent nature of economic time series suggests that explicit modelling of these interactions may broaden our understanding of data generation and improve reliability of forecasts. This necessitates the use of multivariate time series models, in which at least two time series, as a system, are generated as a function of
past history including the past of their own. A general class of linear multiple time series models can be found in the class of the vector ARMA (VARMA) models developed by Tiao & Box (1981) which is a multivariate extension of univariate ARMA models; see, for details of VARMA modelling, Lütkepohl (1991) and Reinsel (1993). Its structure allows explicit modelling of dynamic interactions which may be present among all variables in the system.

In this thesis, we are wholly interested in a subset of the VARMA class, called the vector autoregression (VAR), which is a widely used tool in multiple time series econometrics; see, for example, Sims (1980). The VAR involves a linear relationship between current and past values of the variables in the system and is useful in specifying a dynamic relationship among jointly related variables. Each variable is expressed as a linear function of lagged values of all variables in the system. It possesses a unique parameterisation and can readily be estimated by well-known estimation methods such as least squares and Yule-Walker. It may often provide a useful first approximation to a more general model such as VARMA. However, a VARMA model requires ad hoc restrictions on parameters to ensure uniqueness, and non-linear techniques should be employed for parameter estimation. A further attraction of VAR is that forecasts can be generated simply as a linear combination of parameter estimators and past values (or forecasts).

Care should be taken not to over-specify a VAR because the number of parameters proliferates rapidly as system dimensionality and lag length increase. For example, a VAR model of order two with five variables in the system has fifty-five free parameters to be estimated (including five
intercept terms), while a VAR of order three with the same number of variables has eighty. A typical consequence of fitting too many parameters is poor post-sample forecasting performances, as pointed out by Litterman (1986a) and McNees (1986). Efforts directed to overcome the problem of over-parameterisation include: the subset VAR model of Penn & Terrell (1982, 1984); reduced rank VAR of Velu et al. (1986) and Ahn & Reinsel (1988); and Bayesian VAR of Litterman (1986b). These models impose parameter restrictions in an attempt to decrease the number of VAR parameters to be estimated. For example, in Bayesian VAR, the parameter associated with the variable lagged j periods is assumed to follow a normal distribution with mean zero and variance inversely proportional to j. In this thesis, our interest is placed exclusively on the unrestricted VAR because the restricted VAR models often require complicated estimation methods. Investigations into restricted VAR modelling are the subject of future research.

There have been numerous studies into the forecasting performance of VAR models. Favourable performance of these against other time series and/or large scale econometric models include: Fackler & Krieger (1986), Litterman (1986b), McNees (1986), Trevor & Thorp (1988), Zapata & Garcia (1990), Park (1990) and Liu et al. (1994). In particular, Fackler & Krieger (1986) favoured VAR forecasts over univariate ARMA and VARMA; Litterman (1986b) and McNees (1986) provided evidence of Bayesian VAR forecasts performing better than forecasts obtained from large scale econometric models; and Trevor & Thorp (1988) provided evidence of VAR forecasts being superior to those based on newspaper surveys of private sector economists. It seems that there is a consensus in the validity and effectiveness of VAR forecasting. However, the use of VAR for what Lütkepohl (1991) called the
"structural analysis" - innovation accounting, exogeneity testing, variance decompositions and impulse response analysis - is still controversial. Empirical studies on this issue can be found in Sims (1980, 1982), Gordon & King (1982), Reagan & Sheehan (1985) and Backus (1986). The latter argued that structural analyses using VAR can provide a sound foundation for building econometric models. Skeptical views on using VAR as a means of conducting structural analyses among macroeconomic variables are expressed by Leamer (1985), Cooley & LeRoy (1985) and Runkle (1987). Their main argument is that VAR does not model specific relationships among macroeconomic variables based on economic theory and that innovations cannot be uniquely identified with a particular variable; they argue against the use of VAR as a means of performing macroeconomic analyses. This controversy is, however, not examined in this thesis.

In many aspects of econometric and time series modelling, including VAR, statistical analyses depend heavily on large sample asymptotic theory. Asymptotic formulae provide approximations to unknown properties of the true sampling distribution. The use of asymptotic approximations in small samples can be justified when they represent properties of the true sampling distribution reasonably well. However, the reliability of asymptotic approximations in small samples must be questioned, because, even in simple models, these approximations can often be poor, in which case statistical inference and forecasting may be misleading. A well-known example is the portmanteau statistic of Box & Pierce (1970). This statistic is used to test whether innovations of a time series model follow a white noise process. The test statistic asymptotically follows a chi-squared distribution, yet the asymptotic distribution can sometimes provide a poor approximation in small samples, as found by Davies et al. (1977). A further
example is a prediction interval constructed from an estimated econometric or time series model. The forecast error distribution generated from an estimated model follows a normal asymptotically, hence providing a basis for using a normal approximation in constructing a prediction interval. However, it may often be the case that a normal distribution fails to provide a reasonable approximation to the true sampling distribution in small samples; Phillips (1979) analyses the univariate first-order AR case. In this case, prediction intervals constructed based on the normality assumption may provide misleading assessment of future uncertainty in small samples.

Efron (1979) has proposed the bootstrap which can be used as a small sample alternative to the asymptotic method in approximating the sampling distribution. This method involves resampling the observed data, based on the idea that the observed data is a good representation of the underlying population. It is particularly useful when the sampling distribution is unknown or analytically intractable. It is a computer-intensive method which replaces analytical difficulties with (often) a considerable amount of computation, and has gained dramatically in popularity as computational costs have plummeted. Detailed illustrations of the bootstrap principle can be found in Hall (1992; Chapter 1) and Efron & Tibshirani (1993; Chapter 6). The bootstrap distribution is obtained by repeated resampling and it can be used as an alternative to the asymptotic distribution. Using the bootstrap distribution, small sample properties of the statistic of interest can be estimated, which are called bootstrap statistics (or estimators).

In his influential paper, Freedman (1984) showed that the bootstrap
method can be applied to a general form of econometric model which includes the VAR as a special case. Residuals from an estimated econometric model are resampled to generate bootstrap distributions for parameter estimators. Freedman (1984) also stressed that a sensible bootstrap procedure should possess the property of asymptotic validity, which ensures that bootstrap statistics converge (in distribution) to the true distribution in large samples. Under this proviso, the use of the bootstrap as a means of estimating small sample properties of an econometric equation can be justified. The bootstrap method has been widely used in econometric and time series models. For example, Freedman & Peters (1984a, 1984b) evaluated small sample properties of parameter estimates in simultaneous equation models by bootstrapping; and Thombs & Schucany (1990) and Masarotto (1990) proposed the use of bootstrap prediction intervals in univariate AR models in small samples. A detailed review of applications appears later in Section 1.2. It should be noted that application of the bootstrap method to VAR modelling has attracted relatively little attention in the literature: Runkle (1987) considered the use of the bootstrap in VAR in order to construct confidence intervals for variance decompositions and impulse responses; and Griffiths & Lütkepohl (1990) conducted Monte Carlo experiments to compare small sample performances of asymptotic and bootstrap standard errors and confidence intervals for impulse responses. However, investigations into fundamental issues such as estimation, forecasting and order selection of VAR using the bootstrap are to the best of my knowledge yet to appear in the literature.

This thesis employs the bootstrap method as a means of investigating several fundamental issues of VAR modelling in small samples. Issues concerning estimation, forecasting and order selection of VAR models are
examined, as mentioned earlier. Small samples here indicate those frequently encountered in time series analyses with economic data, especially those in the range of 30 to 100 data points. As we shall see, the bootstrap is used to evaluate reliability of asymptotic approximations in small samples or as means of improving small sample performances of the asymptotic method in VAR modelling. It should be noted that only stationary VAR models are investigated in this thesis. This is mainly because little is known of the asymptotic properties of the bootstrap method when the underlying process is non-stationary. This point will be discussed in more detail in Section 1.2.

In all, this thesis consists of seven chapters. A review of relevant literature appears in the remaining section of this chapter. Plans and synopses of Chapters 2 to 6 are given below. Conclusions are drawn in Chapter 7. Chapter 2 provides a review of issues in VAR modelling such as estimation, forecasting and order selection. The associated asymptotic theories are presented and their bootstrap counterparts are introduced. Large sample properties of these bootstrap methods are examined to establish their asymptotic validity in various facets of VAR modelling. In Chapter 3, the bootstrap method is employed to investigate the issues related to VAR estimation. The major focus is to find the determinants of small sample biases of VAR parameter estimators. We also attempt to construct response surfaces for biases of VAR parameter estimators. In Chapters 4 and 5, VAR forecasting issues are investigated by using the bootstrap method. Chapter 4 investigates small sample properties of point forecasts - properties such as biases, variability and normality of VAR forecasts are focused upon. In Chapter 5, small sample properties of prediction regions based on asymptotic formulae are examined and compared
to those of their bootstrap counterparts. In an attempt to overcome the problem of over-estimation of future uncertainty by VAR prediction regions, an alternative to the conventional VAR prediction region is proposed and its small sample properties are examined. Chapter 6 applies the bootstrap method to the problem of VAR order selection. Order selection criteria (e.g., AIC of Akaike, 1974, and BIC of Schwarz, 1978) are often employed to estimate the unknown VAR order. Although these criteria possess desirable asymptotic properties, their small sample performances can sometimes be poor in terms of choosing the true order; see Mills & Prasad (1992) and Lütkepohl (1985). In an attempt to improve small sample performances of VAR order selection, a bootstrap method is proposed and its small sample properties are evaluated.

1.2 Literature review

This section provides a review of the bootstrap method and its application to time series econometrics. Since proposed by Efron (1979), the bootstrap has found many applications in statistics and econometrics. In line with this, a number of surveys on the bootstrap method and its applications have appeared: they include Efron & Tibshirani (1986), Babu & Rao (1993), Jeong & Maddala (1993), Vinod (1993) and Li & Maddala (1996). Two recent monographs by Hall (1992) and Efron & Tibshirani (1993) provide detailed discussions on theoretical issues and wide coverages of applications of the bootstrap. Since these up-to-date reviews are already available in the literature, the review provided here is given as an introduction to the bootstrap method and a survey of applications mainly in the context of the present thesis.
1.2.1 An introduction to the bootstrap

The bootstrap method involves pseudo-random resampling of the original data, based on the idea that the observed data contains useful representative information on the population characteristics of the data generation mechanism. Resampling is a process of forming artificial data by drawing samples from the original data with replacement. A large number of artificial data sets can be obtained by repeated resampling, and information about the population contained in the original data can be extracted by exploiting these artificial data sets. From each artificial data set, the statistic of interest can be calculated. The collection of these statistics forms the bootstrap distribution of the statistic, which can be used to approximate its sampling distribution. From the bootstrap distribution, properties such as bias, standard error and confidence intervals of the sampling distribution can be estimated. These bootstrap estimators can be used as an alternative approximation device to, for example, asymptotic approximations. In performing resampling, the original data and observed statistic are treated as if they are the population and unknown parameter respectively. That is, artificial data sets are constructed from the original data set, the original data set being treated as if it was the population of interest. Hence, the bootstrap method can also be thought of as a Monte Carlo experiment in which random quantities are drawn from an empirical distribution rather than an assumed distribution. For more details of the bootstrap principle, see Hall (1992; Chapter 1) and Efron & Tibshirani (1993; Chapters 3 to 6).
A simple illustration of the bootstrap will fix ideas. We consider the one-sample situation where the observations are generated from a completely unspecified probability distribution function \( F \) defined on the real line \( \mathbb{R} \). Consider the random variable \( X \) which is generated as

\[
X_i \sim \text{i.i.d. } F,
\]

where i.i.d. indicates independent identically distributed random sampling. Let \( x \) denote a realisation of \( X \) with \( x_i \) being a realised value of \( X_i \). Let \( \bar{X} = (X_1, \ldots, X_n) \) denote any random sample of size \( n \) obtained from \( X \); \( \bar{x} = (x_1, \ldots, x_n) \) denotes a realisation of \( \bar{X} \). Suppose that \( \theta \) is a parameter (of arbitrary dimension), that is,

\[
\theta = \theta(F),
\]

where the notation \( \theta(F) \) indicates that \( \theta \) depends on the distribution \( F \); its value is obtained by applying some numerical or mathematical evaluation procedure. For example, the mean and variance of the distribution are expressed as

\[
\mu(F) = E_F(X) = \int_{\mathbb{R}} X \, F(X) \, dX; \text{ and}
\]

(1.1)

\[
\sigma^2(F) = E_F(X^2) - [E_F(X)]^2 = \int_{\mathbb{R}} [X - E_F(X)]^2 \, F(X) \, dX.
\]

If the sample mean \( \bar{X} \) is used to estimate \( \mu(F) \), its standard error can be written as \( \sigma(\bar{X}; F) = [\sigma^2(F)/n]^{1/2} \). The value of \( \sigma^2(F) \) is unknown and has to be estimated. The estimated standard error can be calculated as
(1.2) \( \hat{\sigma}(X; F) = \left[ \hat{\sigma}^2 / n \right]^{1/2} \),

where \( \hat{\sigma}^2 = 1/(n-1) \sum_{i=1}^{n} (x_i - \bar{x})^2 \), the unbiased estimate of \( \sigma^2(F) \).

The empirical distribution \( \hat{F} \) of \( F \) can be defined as

\[ \hat{F} \text{: probability mass } 1/n \text{ on } x_i, \]

which indicates that each \( x_i \) is assigned with probability \( 1/n \). The bootstrap estimator of \( \sigma(X; F) \), an alternative to (1.2), can be obtained by replacing \( F \) with \( \hat{F} \). The bootstrap estimates of (1.1) can be expressed as

\[ \mu(\hat{F}) = E_{F}^*(x) = 1/n \sum_{i=1}^{n} x_i = \bar{x}; \text{ and} \]

\[ \sigma^2(\hat{F}) = E_{F}^*(x^2) - [E_{F}^*(x)]^2 = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2, \]

and the bootstrap standard error can be written as

(1.4) \[ \sigma(\hat{X}; \hat{F}) = \left[ \sigma^2(\hat{F}) / n \right]^{1/2}. \]

It can be seen that the estimated standard error (1.2) and its bootstrap counterpart (1.4) are slightly different, although the difference becomes negligible as \( n \) increases.

In the simple case of mean and variance of \( \bar{X} \), the bootstrap estimators can be obtained analytically as in (1.3). However, in general, analytical
calculations are intractable. Suppose the parameter \( \theta \) is estimated by using a statistic \( s(X) \), i.e. \( \hat{\theta} = s(X) \). Suppose we want to estimate the bias and standard error of \( \hat{\theta} \), which can be written as

\[
\beta(\hat{\theta}; F) = \mathbb{E}_F[s(X)] - \theta(F); \text{ and}
\]

\[
\sigma(\hat{\theta}; F) = \{\mathbb{E}_F[s(X)^2] - \mathbb{E}_F[s(X)]^2\}^{1/2}.
\]

Their bootstrap counterparts can be written as

\[
\beta(\hat{\theta}; \hat{F}) = \mathbb{E}_{\hat{F}}[s(x)] - \theta(\hat{F}); \text{ and}
\]

(1.5)

\[
\sigma(\hat{\theta}; \hat{F}) = \{\mathbb{E}_{\hat{F}}[s(x)^2] - \mathbb{E}_{\hat{F}}[s(x)]^2\}^{1/2}.
\]

These bootstrap estimators can be obtained by implementing numerical evaluation by means of Monte Carlo simulation; that is, by drawing pseudo-random samples repeatedly from the empirical distribution. The bootstrap algorithm can be described as follows:

(i) Obtain a bootstrap sample \( \{x^*_1\}_{i=1}^b \) where \( x^*_i \) is a i.i.d. pseudo-random drawing from \( \hat{F} \) with replacement.

(ii) From \( x^* = (x^*_1, \ldots, x^*_n) \), obtain sample statistic \( \hat{\theta}^* = s(x^*) \)

(iii) Repeat (i) to (ii) \( b \) times to obtain \( \{\hat{\theta}^*_1\}_{1=1}^b \).

The estimators of bootstrap bias and standard error, denoted respectively
as $\hat{\beta}_b$ and $\hat{\sigma}_b$, can be calculated as

$$
\hat{\beta}_b = \bar{\theta}^* - \theta(\hat{F}); \text{ and } \hat{\sigma}_b = \left( \frac{\sum_{i=1}^{b} (\hat{\theta}_i^* - \bar{\theta}^*)^2}{(b-1)} \right)^{1/2},
$$

where $\bar{\theta}^* = 1/b \sum_{i=1}^{b} \hat{\theta}_i^*$. It can be seen that, as $b \to \infty$, $\hat{\beta}_b$ and $\hat{\sigma}_b$ converge to bootstrap estimators in (1.5); see, for details, Efron & Tibshirani (1986, 1993).

In fact, the bootstrap distribution $\{\hat{\theta}_i^*\}_{i=1}^{b}$ can be used as an approximation to the sampling distribution of $\hat{\theta}$. That is, the distribution $(\hat{\theta} - \theta)$ can be approximated by the bootstrap distribution $(\hat{\theta}_i^* - \bar{\theta})_{i=1}^{b}$. In this context, statistics which measure skewness and kurtosis can be applied to $(\hat{\theta}_i^* - \bar{\theta})_{i=1}^{b}$ in order to examine the possibility of the underlying sampling distribution being asymmetric or platykurtic with respect to the standard normal. Furthermore, the bootstrap distribution can be used to construct a confidence interval for $\theta$. Let $\hat{s}(F;\tau) = s(X;\tau)$ be the true 100th percentiles of the distribution of $\hat{\theta}$. Although the bootstrap 100th percentile can be defined as $\hat{s}(\hat{F};\tau) = s(X;\tau)$, it cannot be obtained mainly due to analytical intractability. Alternatively, the bootstrap confidence interval for $\theta$ can be constructed by what is called the percentile method (see, Efron, 1993, Chapter 13) in which the percentiles of the bootstrap distribution are used to define confidence limits. Let $\hat{G}(s)$ be the bootstrap cumulative distribution function defined as

$$
\hat{G}(s) = 1/b \sum_{i=1}^{b} I(\hat{\theta}_i^* < s),
$$

where $I(\cdot)$ is the indicator function which takes the value one if the
condition in the bracket is true and zero otherwise. The bootstrap confidence interval for \( \theta \) with probability content \((1-\alpha)\) can be constructed by taking the interval

\[
\{\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1-\alpha/2)\}.
\]

It can be shown that, as \( b \to \infty \), the bootstrap confidence interval converge to \([s(x; \alpha/2), s(x; 1-\alpha/2)]\); see Efron & Tibshirani (1986; p62).

The conventional method of constructing a confidence interval is to find an interval based on the asymptotic distribution of the parameter estimator; this distribution is often a normal. However, constructing confidence intervals based on the normality assumption can be misleading especially when the underlying sampling distribution is non-normal. It is mainly because a normal approximation is unable to capture properties such as asymmetry and platykurtosis that may be present in the underlying sampling distribution. For example, a confidence interval constructed using a normal distribution is always symmetric around the observed statistic even when the underlying sampling distribution is asymmetric. The bootstrap method in this case provides a more sensible assessment than the conventional one as the bootstrap confidence interval is, in general, asymmetric around the observed statistic. Hence, the advantage of the bootstrap confidence interval over its conventional counterpart is that it can capture more detailed features of the underlying sampling distribution. Further refinements of the percentile method, called the BC and BCₐ methods by Efron (1981, 1987) adjust for median-bias present in the underlying distribution. Efron (1981, 1987) gives an example to demonstrate that confidence intervals based on these methods perform better than those based
on the percentile method. However, the effectiveness of these methods is yet to be substantiated by empirical evidence in the context of general econometric and time series models.

In implementing the bootstrap, the choice of the number of bootstrap iterations \( b \) must be addressed. According to Efron (1987; Sec. 9), bias and variance calculations require between 50 and 200 bootstrap iterations, while confidence interval calculations require between 1000 to 2000 iterations. More efficient resampling methods have been devised which can significantly reduce the number of bootstrap iterations and enhance precision at the same time. Since a detailed review of these resampling methods is given by Hall (1992; Appendix II), only a brief summary is provided here. The balanced sampling method, proposed by Davidson et al. (1986), involves augmentation of the original data \( \{x_t\}_{t=1}^n \) and resampling from the augmented sample. The method begins by concatenating \( n \) copies of \( \{x_t\}_{t=1}^n \) to form an augmented data set of length \( nm \); then \( n \) observations are selected from the \( nm \) pool without replacement. The importance resampling method of Johns (1988) attempts to gain greater precision by increasing the resampling probabilities of samples that are of most importance. The importance can be determined by using the likelihood ratio from a suitably chosen alternative distribution. Hall (1989) proposed an antithetic sampling method for the bootstrap. Let \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) be different estimators for the unknown parameter \( \theta \). Provided \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) are negatively correlated, \( \hat{\theta} = 0.5 (\hat{\theta}_1 + \hat{\theta}_2) \) has a smaller variance than either \( \hat{\theta}_1 \) or \( \hat{\theta}_2 \). This principle can be used in resampling to enhance precision of bootstrap estimators. These efficient resampling methods, however, have not yet been used widely in bootstrapping econometric models. Some of them are hard to implement for econometric or time series models such as a VAR; and it seems
that fairly accurate bootstrap estimators can still be obtained by employing a high enough number of bootstrap iterations that can easily be handled by modern computers. Hence, in this thesis, these efficient resampling methods are not adopted; only the standard uniform resampling method is used throughout.

1.2.2 Applications to time series econometrics

The theoretical justification of applying the bootstrap to time series and econometric models can be found in Freedman (1984). Its precursor is Freedman (1981) where bootstrapping least squares (LS) estimators of a regression model is dealt with. Only the former is discussed here since the latter is its special case. Freedman (1984) considered a general linear model of the form

\[(1.6) \quad Y_t = Y_{t-1}B + X_tC + \varepsilon_t, \quad t = 1, \ldots, n,\]

where \(Y_t\) is a \((1 \times a)\) row vector of endogenous variables; \(X_t\) is a \((1 \times b)\) vector of fixed exogenous variables; and \(\varepsilon_t\) is an \((1 \times a)\) vector of \(i.i.d.\) disturbances with zero means; \(A, B\) and \(C\) are unknown coefficient matrices of the size \((a \times a), (a \times a)\) and \((b \times a)\) respectively. Equation (1.6) can be interpreted as a simultaneous equation system with dynamic structure. Note that many econometric and time series models are special cases of (1.6). For example, if \(A\) and \(B\) are null matrices, it takes the form of a multiple regression equation with fixed regressors. If \(A\) is a null matrix, then it is a multiple linear regression equation with lagged dependent variables. If \(A\) and \(C\) are null matrices, then it takes the form of a VAR(1)
model. Since univariate AR(p) and VAR(p) models can be expressed in a 
VAR(1) form, it can be seen that the class of univariate AR and VAR models 
is also a special case of (1.6). It can also be seen that seemingly 
unrelated regressions (SUR) of Zellner (1962) is a special case.

\( Y_t \) is assumed to be (second-order) stationary so that it can be 
expressed as an infinite sum of disturbances combined with an absolutely 
summable sequence (see, for definition, Section A.9.1. of Lütkepohl, 1991) 
as

\[
Y_t = \sum_{s=0}^{\infty} \xi_{t-s},
\]

where \( \xi_{t-s} = (\epsilon_{t-s} + X_{t-s} C) (I - A)^{-1} [B (I - A)^{-1}]^s \) and the matrix \((I - A)\) 
is assumed to be invertible. It is also assumed that appropriate zero 
restrictions are imposed on \( A \) and \( B \) for identification of the simultaneous 
equations system.

Freedman (1984) considered two-stage LS (2SLS) estimation of (1.5) and 
used the bootstrap to approximate the sampling distributions of these 
estimators. Let \( \hat{A}, \hat{B} \) and \( \hat{C} \) be the 2SLS parameter estimators. Residuals are 
generated as

\[
\hat{\epsilon}_t = Y_t - [Y_t \hat{A} + Y_{t-1} \hat{B} + X_t \hat{C}].
\]

Resampling should be conducted so that a bootstrap sample mimics the 
structure of the data generation mechanism. This can be achieved by 
resampling the residuals and generating artificial data based on (1.6). 
That is,
\[
y_t^* = (Y_{t-1}^* \hat{B} + X_t \hat{C} + \epsilon_t^*) (I - \hat{A})^{-1},
\]

where \( \epsilon_t^* \sim i.i.d. \{ \epsilon_t \}_{t=1}^n \) and \( Y_0^* = Y_0 \) is assumed to be known. Denote the bootstrap sample by \( Y^* = \{Y_t^*\}_{t=1}^n \). By repeating this process \( b \) times, \( b \) bootstrap samples, denoted \( \{Y_t^*\}_{t=1}^b \), are obtained. From each bootstrap sample, parameters can be estimated to yield the bootstrap distribution of parameter estimators \( \{\hat{A}_t^*\}_{t=1}^b, \{\hat{B}_t^*\}_{t=1}^b \) and \( \{\hat{C}_t^*\}_{t=1}^b \). These bootstrap distributions can be used to approximate the unknown sampling distributions of parameter estimators. For example, the distribution of \( \hat{A} - A \) can be approximated by \( \{\hat{A}_t^* - A\}_{t=1}^b \).

Freedman (1984) stressed that the following two properties should be satisfied for the bootstrap to be an appealing estimation and inferential method:

(i) the bootstrap gives the right answers in large samples; and

(ii) the bootstrap outperforms the conventional asymptotics in small samples.

Property (i) refers to the asymptotic validity of the bootstrap which ensures that the bootstrap exhibits similar large sample properties with those of the conventional asymptotics. Freedman (1984) has shown, under suitable conditions, that the bootstrap estimators of the model (1.5) are valid asymptotically; i.e., bootstrap 2SLS estimators for parameters A, B and C in equation (1.6) converge in conditional probability to their true values as the sample size increases. Property (ii) was not investigated by
Freedman (1984); small sample comparisons between the two will extremely be difficult mainly due to analytical intractability. Hence, it is not clear whether the bootstrap outperforms the conventional asymptotics in small samples. As pointed out by Peters & Freedman (1984), there is no guideline as to which circumstances favour the bootstrap. Subsequent studies on the bootstrap attempted to compare small sample properties of the two alternatives either empirically or by means of Monte Carlo simulations.

Freedman & Peters (1984a) applied the bootstrap to a system of SUR's designed to forecast regional demand for energy. They compared the bootstrap and asymptotic standard errors of parameter estimators when the method of generalised LS (GLS) is used for estimation. They found that asymptotic standard errors are too optimistic by factors of nearly three and concluded that using them would be inappropriate in small samples. This is supported by a mathematical proof in a simpler context; the use of the estimated variance-covariance matrix of the disturbances in calculating that of the GLS estimators is identified as a source of downward bias. Freedman & Peters (1984b) applied the bootstrap method to 2SLS and three-stage LS (3SLS) estimations of the Berndt-Wood (1975) model. Asymptotic standard errors of parameter estimators obtained from both estimation methods are found to be optimistic, but the extent of under-estimation is not as serious as in Freedman & Peters (1984a). It is, on the whole, concluded that asymptotic standard errors perform reasonably well. However, small sample biases of parameter estimators are found to be statistically significant, in contrast to the case of Freedman & Peters (1984a). This is evidence that, when asymptotic formulae are used, small sample biases present in parameter estimators can be under-estimated.
Monte Carlo simulations conducted in an attempt to examine property (ii) include: Stein (1985, 1987), Masarotto (1990) and Thombs & Schucany (1990) among others. The major concern of these authors is to consider the bootstrap as a means of constructing prediction intervals for econometric models. Stein (1985) evaluated small sample properties of bootstrap prediction intervals for regression models; the other authors examined univariate AR models. A prediction interval for an estimated econometric or time series model is constructed based on the assumption that the underlying distribution of forecasts follows a normal. However, due to additional uncertainty brought about by parameter estimation, the assumption of normality is highly unlikely to be tenable in small samples. In this case, the conventional interval based on a normal approximation may provide a misleading assessment of future uncertainty; a natural alternative is the prediction interval based on the bootstrap. In what follows, the details of Thombs & Schucany (1990) and Masarotto (1990) are discussed. The bootstrap procedures and the major findings of Stein (1985, 1987) are similar to those of these authors.

An interesting feature of Thombs & Schucany (1990) is that the property of AR forecasts being conditional on past observations is explicitly considered by employing the backward AR representation. Consider the AR(p) model of the form

\[(1.7) \quad y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + u_t^* \]

where \(u_t^* \sim i.i.d.(0, \sigma_u^2)\). The corresponding backward model can be written as
(1.8) \[ y_{t} = \alpha_1 y_{t+1} + \ldots + \alpha_p y_{t+p} + \varepsilon_t, \]

with \( \varepsilon_t \sim (0, \sigma_c^2) \) and \( \sigma_c^2 = \sigma_u^2 \). It should be noted that innovations for the backward model is i.i.d. only when the \( u_t \)'s are normal. In case of non-normal \( u_t \)'s, backward innovations \( \varepsilon_t \)'s are uncorrelated, but they may be dependent, as pointed out by Thombs & Schucany (1990). It can be shown that the two representations share the same autocorrelation structure, and parameter estimators based the LS method are also identical (see Box et al., 1994; Section 6.4). The bootstrap procedure suggested by Thombs & Schucany (1990) is as follows:

Estimate (1.7) by the LS method to obtain parameter estimates \( \hat{\alpha}_j \) \( j = 1 \) and residuals \( \hat{(u)}_{t} \) \( t = p+1 \). Using (1.8), residuals from the backward model \( \{\hat{\varepsilon}_t\}_{t=1}^{n-p} \) can be obtained. Obtain the bootstrap sample \( Y^* = \{y^*_{t}\}_{t=1}^{n} \) recursively based on (1.8) by resampling \( \{\hat{\varepsilon}_t\}_{t=1}^{n-p} \) with replacement. That is

(1.9) \[ y^*_{t} = \hat{\alpha}_1 y^*_{t+1} + \ldots + \hat{\alpha}_p y^*_{t+p} + \hat{\varepsilon}^*_t, \]

where \( \hat{\varepsilon}^*_t \sim i.i.d. \) \( \{\hat{\varepsilon}_t\}_{t=1}^{n-p} \) and the starting values of the recursion are the last \( p \) values of the original data. By repeating this process \( b \) times, \( b \) sets of bootstrap sample \( \{Y^*_{1}\}_{1=1}^{b} \) can be obtained. Each bootstrap sample \( Y^*_{1} \) possesses the last \( p \) observations which are same as the last \( p \) observations of the original data. In this way, forecasts generated from each bootstrap sample will be conditional on the last \( p \) observations of the original data. From each bootstrap sample, AR parameters are estimated to yield \( \{\hat{\alpha}_j\}_{j=1}^{p} \) and forecasts made at \( n \) for the period \( n + h \) can be generated as
\[(1.10) \quad y^*_n(h) = \hat{\alpha}_1^* y^*_n(h-1) + \ldots + \hat{\alpha}_p^* y^*_n(h-p) + \hat{u}^*_n,\]

where \(y^*_n(h) = y_{n-j}\) for \(j \leq 0\) and \(\hat{u}^*_t \sim \text{i.i.d.} \{\hat{u}_t^*\}_{t=1}^n\). Repeating this process \(b\) times, one can obtain the bootstrap distribution of forecasts \((y^*_n(h))_{l=1}^b\), which can be used to approximate the sampling distribution of forecasts. Prediction intervals can be constructed by the percentile method described earlier. Thombs & Schucany (1990) proved that the bootstrap procedure given here is asymptotically valid.

Masarotto (1990) adopted a different approach to that of Thombs & Schucany (1990). It involves bootstrapping the distribution of standardised prediction errors. The usual \(100(1-\alpha)\%\) level prediction interval for AR model can be written as

\[(1.11) \quad \left[ \hat{y}_{n}^* (h) \pm Z_{1-\alpha/2} \hat{\sigma}_h^* \right] \]

where \(\hat{y}_{n}^*(h)\) is the forecast made at \(n\) for period \(n+h\) from the estimated model of (1.7), \(Z_{1-\alpha/2}\) is the \(100(1-\alpha/2)\)th percentile of the standard normal distribution, and \(\hat{\sigma}_h^*\) is the estimator of the forecast standard error using the MA(\(\omega\)) representation of (1.7). Masarotto (1990) attempted to find the bootstrap counterpart of \(Z_{1-\alpha/2}\) by bootstrapping standardised prediction errors. Consider

\[(1.12) \quad r_{n,h}^* = (\hat{y}_{n+h}^* - \hat{y}_n^*(h)) / \hat{\sigma}_h^* \]

where \((\hat{y}_t^*)_{t=1}^n\) and \(\hat{y}_n^*(h)\) are generated equivalently to (1.9) and (1.10) based on the forward AR representation (1.7) and \(\hat{\sigma}_h^*\) is the bootstrap counterpart of \(\hat{\sigma}_h\) calculated from \((\hat{y}_t^*)_{t=1}^n\). \(r_{n,h}^*\) can be calculated for each
bootstrap sample to yield the bootstrap distribution of standardised prediction errors \( \{ \hat{r}_{n,h}^{*1} \}_{n,h=1}^b \). The bootstrap prediction interval can be constructed by replacing \( \pm Z_{1-\alpha/2} \) with the corresponding percentiles of \( \{ \hat{r}_{n,h}^{*1} \}_{n,h=1}^b \). That is,

\[
[ \hat{y}_n(h) - \hat{r}_{n,h}^{*1}(\alpha/2) \hat{\sigma}_h, \hat{y}_n(h) + \hat{r}_{n,h}^{*1}(1-\alpha/2) \hat{\sigma}_h ],
\]

where \( \hat{r}_{n,h}^{*1}(\alpha) \) denotes the 100\(\alpha\)th percentile of the distribution \( \{ \hat{r}_{n,h}^{*1} \}_{n,h=1}^b \). Masarotto (1990) proved that this procedure is asymptotically valid.

The simulation findings of Thombs & Schucany (1990) and Masarotto (1990) reveal that bootstrap prediction intervals perform reasonably well in small samples under both normal and a wide range of non-normal innovations. However, there is no evidence that the bootstrap prediction intervals perform substantially better than the conventional ones, or vice versa, in small samples. This suggests that it is not clear whether or when to adopt the bootstrap method over the asymptotic one to construct prediction intervals for AR models. Nevertheless, it appears that the bootstrap in general provides a useful non-parametric alternative to the conventional normal approximation. A practitioner is recommended to adopt the bootstrap prediction intervals as an alternative tool for assessing future uncertainty.

Peters & Freedman (1984) suggested that the bootstrap will provide a useful alternative when the structure of model is complicated or the number of parameters is large relative to the number of data points. Since the VAR model possesses a complicated structure with often large numbers of parameters, the bootstrap might be expected to be more effective than in
the univariate AR case. The evidence so far, however, does not appear to support this view. Runkle (1987) and Griffiths & Lütkepohl (1990) used the bootstrap to construct confidence intervals for variance decompositions and impulse responses in VAR models. Runkle (1987) found empirically that bootstrap confidence intervals are as wide as their asymptotic counterparts providing no additional useful information. Griffiths & Lütkepohl (1990) conducted Monte Carlo simulations to compare small sample performances of asymptotic and bootstrap confidence intervals. As in the univariate case, no strong evidence of bootstrap confidence intervals performing better than their asymptotic counterparts is found. However, it seems too early to conclude that the bootstrap provides only a computationally burdensome alternative to the asymptotic method in VAR modelling and offers little additional information. The bootstrap should be evaluated in the context of other fundamental issues, such as parameter estimation, forecasting and order selection.

It seems that past studies have paid attention mainly to the application of the bootstrap to the provision of variability measures such as standard errors, confidence intervals and prediction intervals. However, as observed by Freedman & Peters (1984b), asymptotic approximations may fail in capturing substantial small sample biases present in parameter estimators. This indicates that the use of the bootstrap as a means of evaluating biases in parameter estimators or forecasts can provide useful insights in econometric and time series modelling. In addition, the application of the bootstrap to a wider range of modelling issues - such as model selection and hypothesis testing - is called for. It should also be mentioned that past studies concentrated on bootstrapping econometric or time series models whose underlying process is stationary. It is because
these past studies are based largely on the work of Freedman (1984) in
which the assumption of stationarity is crucial when establishing
asymptotic validity of the bootstrap. However, statistical analyses of
non-stationary time series are becoming more important; see, for example,
testing and estimation of cointegration of Engle & Granger (1987). Since
these statistical analyses rely heavily on asymptotic theories, the
bootstrap method can provide a useful small sample alternative.

A series of papers by Basawa et al. (1989, 1991a, 1991b) dealt with
the bootstrap when the underlying process follows unstable and explosive
univariate AR(1) processes. Consider an AR(1) process which can be written
as

\[ y_t = \alpha y_{t-1} + u_t, \]

where \( u_t \sim \text{i.i.d. } (0, \sigma_u^2) \) and \( y_0 = 0. \) \( y_t \) is unstable if \( |\alpha| = 1 \), explosive
if \( |\alpha| > 1 \) and stationary if \( |\alpha| < 1 \). The LS estimator for \( \alpha \), denoted \( \hat{\alpha} \), is
given by

\[ \hat{\alpha} = \left( \sum_{t=1}^{n} y_t y_{t-1} \right) \left( \sum_{t=1}^{n} y_t \right)^{-1}. \]

Let \( \hat{\alpha}^* \) be the bootstrap estimator for \( \alpha \) obtained by using the standard
bootstrap procedure given earlier. Basawa et al. (1989) proved that \( \hat{\alpha}^* \) is
asymptotically valid when \( y_t \) is explosive. Basawa et al. (1991a) showed
that \( \hat{\alpha}^* \) is asymptotically invalid when the process is unstable; i.e., when
\( |\alpha| = 1 \), asymptotic behaviour of \( \hat{\alpha} \) is different from that of \( \hat{\alpha}^* \). The
asymptotic behaviour of \( \hat{\alpha}^* \) is different in each of the three cases.
Asymptotic invalidity of \( \hat{\alpha}^* \) when \( |\alpha| = 1 \) arises because this case occurs at
the threshold of the two other cases of $|\alpha| > 1$ and $|\alpha| < 1$, as pointed out by Basawa et al. (1991b).

A modified procedure, which is asymptotically valid, has been proposed by Basawa et al. (1991b). They showed that, when the bootstrap is used for testing $H_0: \alpha = 1$, the standard bootstrap procedure is asymptotically valid if bootstrap samples are generated under $H_0^*$. That is, the bootstrap sample is generated recursively as

$$ y_t^* = y_{t-1}^* + u_t^* = \sum_{l=1}^{t} u_t^*, $$

where $u_t^* \sim i.i.d. (\hat{\epsilon}_t)^n$ and $\hat{u}_t = y_t - \hat{\alpha} y_{t-1}$ with $y_0^* = 0$. The limit distribution of $n(\hat{\alpha}^* - 1)$ in this case is shown to be the same as that of $n(\alpha - 1)$. This provides the basis of testing the significance of $H_0$ by bootstrapping; see, for an empirical application, Harris (1992). From the bootstrap sampling distribution of the test statistic, small sample critical values or p-values of the test suitable for the data set under scrutiny are calculated. For bootstrap parameter estimation of $\alpha$ when $|\alpha| \leq 1$, Basawa et al. (1991b) proposed a sequential procedure. Let $\tilde{\alpha}$ be defined as

$$ \tilde{\alpha} = \hat{\alpha} I(|\hat{\alpha}| \leq 1) + I(\hat{\alpha} > 1) - I(\hat{\alpha} < -1). $$

The bootstrap sample is generated recursively as

$$ \tilde{y}_t^* = \tilde{\alpha} \tilde{y}_{t-1}^* + \epsilon_t^*. $$
where $\epsilon_t^* \sim \text{i.i.d. } (\epsilon_t)^n_{t=1}$; $\tilde{\epsilon}_t = \epsilon_t - \frac{1}{n} \sum_{t=1}^n \epsilon_t' $ the centered residual generated with $\tilde{\alpha}$; and $y_0^* = 0$. Let $\tilde{\alpha}^*$ be the bootstrap estimator for $\alpha$ based on this sequential procedure. Basawa et al. (1991b) proved that asymptotic behaviour of $\tilde{\alpha}^*$ is the standard normal distribution. For application of the bootstrap to a wider range of data generation mechanisms, theoretical investigation into the bootstrap when the underlying process is non-stationary should be extended to higher order and higher dimensional cases.

1.2.3 Bootstrap methods for dependent data

When the standard bootstrap procedure is applied to a process with serial dependence, its dependence structure is assumed to be fully known, as was the case in the previous section. However, the dependence structure is often completely unknown or only partially known. In these situations, the standard bootstrap procedure is unable to capture the unknown structure. To overcome this, the method of the moving block bootstrap (MBB) has been proposed; see, for example, Künsch (1989), Liu and Singh (1992) and Fitzenberger (1995). It is designed to capture the dependence structure without imposing any a priori assumptions on the data generation mechanism. The idea is to form (overlapping) blocks made of sub-series of data and produce bootstrap samples by i.i.d. resampling from the blocks with replacement. A precursor of the MBB, proposed earlier by Carlstein (1986), involves resampling non-overlapping blocks; but the experience of Li & Maddala (1996, p136) suggests inferiority of Carlstein's (1986) method to the MBB.
The bootstrap sample constructed based on the MBB preserves salient features of the dependence structure up to the order of block length. In this respect, the MBB can be applied to a general class of dependent data under no parametric model assumptions. Künnisch (1989) and Liu and Singh (1992) established asymptotic validity of the MBB in bootstrapping sample mean when the data is a realisation of a general dependent and stationary process. Fitzenberger (1995) proved that the MBB is asymptotically valid when it is applied to LS estimation of a regression model whose error terms are realisations of stationary dependent process with heteroscedasticity.

The MBB procedure can be summarised as follows:

i) Consider a realisation of a stationary process \( \{y_t\}_{t=1}^n \). Form blocks of length \( h \) to yield \( B = \{B_t\}_{t=1}^{n-h+1} \) with \( B_t = \{y_t, \ldots, y_{t+h-1}\} \) and \( t = 1, \ldots, n-h+1 \).

ii) Draw \( d \) blocks with replacement from \( B \) to obtain the bootstrap sample \( C = \{G_1, \ldots, G_d\} = \{S_1, \ldots, S_{n^*}\} \) of size \( n^* = d \times h \), where \( G_i \) is i.i.d. from \( B \) with replacement.

iii) Repeat ii) \( b \) times to obtain \( b \) sets of bootstrap sample \( \{C_j\}_{j=1}^b \). The statistic of interest, say \( \tau_j^* \), can be calculated from each bootstrap sample to form a bootstrap distribution of the statistic \( \{\tau_j^*\}_{j=1}^b \).

As in Fitzenberger (1995), the bootstrap sample size \( n^* \) can take any value different from, but usually larger than, \( n \). By drawing extra samples, the
likely performance of the data with a larger number of observations can be evaluated, as pointed out by Freedman (1981). However, to ensure $n = n^*$, draw blocks with $n^* \geq n$ and discard the last $n^* - n$ observations.

Politis & Romano (1994) extended the MBB by allowing block length $h$ to be random and treating the original data as circular. They assume block length to be generated according to a geometric distribution with mean $\bar{h}$; $\bar{h}$ is the average block length in resampling. The circular treatment of data is related to the treatment of end values in resampling. That is, if a block includes the data points beyond the final observation, these data points are filled with initial observations of data. For example, a block of length 5 starting from $y_{n-2}$ is given by $(y_{n-2}, y_{n-1}, y_n, y_1, y_2)$. This procedure is called the stationary bootstrap (SB) because it provides bootstrap samples which are stationary while those generated with the MBB may not; see Proposition 1 of Politis & Romano (1994). Asymptotic validity of the SB in bootstrapping the sample mean is shown by Politis & Romano (1994).

In implementing the MBB or SB, the choice of value of (average) block length $\bar{h}$ and bootstrap sample size $n^*$ is of importance since the outcome of the bootstrap can significantly be affected by these values. Hall et al. (1996) found that the optimal choice of block length depends significantly on context. That is, the optimal choice of block length in the case of bias or variance estimation differs application by application. They also provided an empirical rule by which the optimal choice of block length can be estimated. However, difficulty still remains because this rule requires the choice of another parameter, and the sensitivity of the bootstrap outcomes to this additional choice of parameter is unknown. Further issues
and details regarding the choice of $\tilde{n}$ and $n^*$ are given in Chapter 2.

Applications of the MBB method can be found in Li & Maddala (1993) and Li (1994) who considered bootstrapping cointegrating regressions (see, for example, Engle & Granger, 1987, and Johansen, 1991). The MBB is applied to the residuals of a cointegrating regression, which are estimators of a stationary disturbance process whose dependence structure is unknown. In their Monte Carlo simulations, the usefulness of the MBB, as an alternative to the asymptotic method, in conducting small sample statistical inferences in cointegrating regressions is evaluated. Although the MBB provides favorable results, its asymptotic validity in the context of cointegration regression is not established. This casts doubt on the use of the MBB as a small sample alternative to the asymptotic method in this context. Further theoretical investigations on the applicability of the MBB to econometric and time series modelling is called for.

1.3 Concluding remarks

A substantial decline in computational costs has been experienced in recent years, and, as a result, implementation of the computer-intensive methods such as the bootstrap has become more attractive. As econometric and time series models become more complicated in their structures, it is likely that small sample statistical analyses of these models will become more complicated too. This will inevitably lead to more reliance on asymptotic approximations based on large sample theories whose small sample performance is often unknown and sometimes unsatisfactory. Since the bootstrap method can provide a useful alternative to asymptotic
approximation, its application to increasingly complicated econometric and
time series models is to be expected. This thesis applies the bootstrap
method to VAR, a class of multivariate linear time series models frequently
adopted in practice. Although the bootstrap method has found many
applications in time series econometrics, application of the bootstrap as a
means of investigating fundamental issues in VAR modelling is yet to be
addressed in the literature.

Our attention is restricted to stationary VAR models. The stationarity
assumption is crucial to achieve asymptotic validity, as can be seen from
the work of Freedman (1984). Although Basawa et al. (1991b) suggested an
asymptotically valid bootstrap procedure for the non-stationary AR(1)
model, the asymptotic validity of the bootstrap for higher order or higher
dimensional non-stationary AR models is yet to be established. Moreover,
for non-stationary VAR models, we may need to consider cointegrating
relationships which provide useful information on the long-run
relationships and short-run dynamics of VAR components. However, in this
setting, it is yet to be established whether the bootstrap method is
asymptotically valid. It seems that more theoretical investigation is
required before the bootstrap method can be applied to non-stationary VAR
systems. For this reason, non-stationary VAR models are not examined here.
Chapter 2. VAR Modelling and Bootstrapping

2.1 Introduction

This chapter reviews the methodologies to be employed in this thesis. Issues concerning estimation, forecasting and order selection of VAR models are dealt with and the associated asymptotic theories are introduced. Issues presented in Section 2.2 on VAR modelling are taken in part from Lütkepohl (1991). Stochastic properties and estimation methods for backward VAR models are presented in Section 2.3. The backward representation in the univariate AR case possesses stochastic properties identical to those of the forward representation (Box & Jenkins, 1976; and Thombs & Schucany; 1990). However, as we shall see in Section 2.3, stochastic properties of the backward VAR representation, although closely related, differ from those of the forward VAR representation. In Section 2.4, the bootstrap counterparts of the issues dealt with in Section 2.2 are introduced and their large sample validity is established.

2.2 VAR modelling

Consider the K-dimensional stationary VAR process of order p, denoted by VAR(p), given by

\[(2.1)\quad y_t = \nu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \ldots\]

where $y_t$ is a Kx1 random vector, the $A_i$'s are KxK matrices of coefficients,
and \( v \) is a \( K \times 1 \) vector of intercept terms. The \( K \times 1 \) vector of i.i.d. innovations \( u_t \) is such that \( \text{E}(u_t | I_{t-1}) = 0 \) and \( \text{E}(u_t u'_t | I_{t-1}) = \Sigma_u \), where \( \Sigma_u \) is a \( K \times K \) symmetric positive definite matrix with finite elements. The notation \( I_s \) indicates the set of all available past information at time point \( s \). The process \( y_t \) is assumed stationary, so that the condition

\[
det(I_K - A_1 z - \ldots - A_p z^p) \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1
\]

is satisfied (see, for derivation, p10 of Lütkepohl; 1991), where \( \det(\cdot) \) denotes the determinant. The above condition indicates that all the roots of the reverse characteristic equation \( \det(I_K - A_1 z - \ldots - A_p z^p) = 0 \) lie outside the unit circle (or have modulus greater than one).

It is useful to express the VAR(p) model in (2.1) in a more compact way. A VAR(p) model can be expressed in a \( Kp \)-dimensional VAR(1) form as

\[
(2.2) \quad Y_t = v + \Pi Y_{t-1} + U_t
\]

where \( Y_t, U_t \) and \( v \) are \( Kp \times 1 \) vectors defined as

\[
Y_t = (y'_t, y'_{t-1}, \ldots, y'_{t-p+1})'; \quad U_t = (u'_t, 0', \ldots, 0'); \quad \text{and}
\]

\[
v = (v, 0', \ldots, 0');
\]

while \( \Pi \) and \( \Sigma_u \) are \( Kp \times Kp \) matrices of the form

\[
\Pi = \begin{bmatrix} A_1 & A_2 & \ldots & A_p \\ \vdots & \vdots & \ddots & \vdots \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \end{bmatrix}; \quad \text{and} \quad \Sigma_u = \text{E}(U_t U'_t) = \begin{bmatrix} \Sigma_u & 0' \\ 0 & 0 \end{bmatrix},
\]
where I is a K(p-1) identity matrix, O is a K(p-1)×K null matrix and \( A = [A_t, ..., A_{p-1}] \).

In estimating a VAR model, the most commonly used estimators are the LS and Yule-Walker (YW) estimators (see, for example, Lütkepohl, 1991, Chapter 3); they can easily be computed and share the same well-defined limiting distribution. Since both of the estimators are used throughout this thesis in various aspects of VAR modelling, a brief summary of their derivation and asymptotic properties follow.

Assume that a suitable set of time series data, including p presample values, \( (y_{-p+1}, ..., y_0, y_1, ..., y_n) \) is available. The LS estimator is obtained by applying the standard least squares principle to equation (2.1). Rewrite the model (2.1) as

\[
Y = AZ + U,
\]

where

\[
Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \text{(K×n)}
\]

\[
A = \begin{pmatrix} \nu \\ A_1 \\ \vdots \\ A_p \end{pmatrix} \quad \text{(K×(Kp+1))}
\]

\[
Z = \begin{pmatrix} Z_0 \\ \vdots \\ Z_{n-1} \end{pmatrix} \quad \text{((Kp+1)×n)}
\]

\[
Z_t = \begin{pmatrix} 1 \\ y_t' \\ \vdots \\ y_{t-p+1}' \end{pmatrix}' \quad \text{((Kp+1)×1)}
\]

\[
U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \quad \text{(K×n)}
\]

(the dimensions of the matrices are indicated to the right).

The LS estimators for A and \( \Sigma_u \) are given by

\[
\hat{A} = (\hat{\nu}, \hat{A}_1, ..., \hat{A}_p) = YZ'(ZZ')^{-1}; \text{ and}
\]

\[
\hat{\Sigma}_u(p) = (Y - \hat{A}Z)(Y - \hat{A}Z)' / (n - Kp - 1).
\]
Under suitable regularity conditions, the LS estimator \( \hat{A} \) can be shown to have desirable asymptotic properties, such as consistency and asymptotic normality.

**Theorem 2.1.** (Proposition 3.1 of Lütkepohl, 1991; p66)

Let \( y_t \) be a K-dimensional stationary VAR(p) process as in (2.1). Assume that, for some finite constant \( c \),

\[
E|u_{it} u_{jt} u_{kt} u_{mt}| \leq c, \text{ for } i, j, k, m = 1, \ldots, K \text{ and all } t.
\]

Then it can be shown that:

(i) \( \text{plim}(\hat{A}) = A \); and

(ii) \( n^{1/2} \{ \text{vec}(\hat{A}) - \text{vec}(A) \} \xrightarrow{d} N(0, \Delta^{-1} \otimes \Sigma_u) \),

where \( \Delta = \text{plim}(ZZ'/n) \), and \( \otimes \) and \( \xrightarrow{d} \) respectively indicate Kronecker product and convergence in distribution.


To obtain the YW estimators of VAR parameters, consider the mean-adjusted version of (2.2), which can be written as

\[
(2.3) \quad W_t = \Pi W_{t-1} + U_t,
\]
where \( W_t = (w'_t, \ldots, w'_{t-p+1})' \) with \( w_t = y_t - E(y_t) \). By post-multiplying on both sides of (2.3) by \( W'_{t-1} \) and taking expectations, we obtain

\[
\Gamma_p(1) = \prod_{p} \Gamma(0) \quad \text{or} \quad \Pi = \Gamma_p(1) \Gamma_p(0)^{-1},
\]

where \( \Gamma_p(j) = E(W_t W'_{t-j}) = 
\begin{bmatrix}
\Gamma(j) & \Gamma(j+1) & \cdots & \Gamma(j+p-1) \\
\Gamma(j-1) & \Gamma(j) & \cdots & \Gamma(j+p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma(j-p+1) & \Gamma(j-p+2) & \cdots & \Gamma(j)
\end{bmatrix}_{(K_p \times K_p)}; \quad \text{and} \quad \\
\Gamma(j) = E(w_t w'_{t-j}).
\]

The YW estimator \( \hat{\Pi} \) for \( \Pi \) can be obtained by replacing unknowns with estimators in the above expression. That is,

\[
\hat{\Pi} = \hat{\Gamma}_p(1) \hat{\Gamma}_p(0)^{-1} \quad \text{where} \quad \hat{\Gamma}(h) = 1/n \sum_{t} \tilde{w}_t \tilde{w}'_{t-j},
\]

where \( \tilde{w}_t = y_t - \tilde{y} \) and \( \tilde{y} = 1/n \sum_{t=1}^{n} y_t \). The vector of intercept terms \( \tilde{\nu} \) can be estimated as \( \tilde{\nu} = (I_{K_p} - \hat{\Pi}) \tilde{Y} \) with \( \tilde{Y} = 1/n \sum_{t=1}^{n} Y_t \). The YW estimator for \( \Lambda \) can be written as \( \hat{\Lambda} = (\tilde{\nu}, \tilde{\Lambda}_1', \ldots, \tilde{\Lambda}_p') \); \( \tilde{\nu} \) is obtained as a sub-vector of \( \tilde{\nu} \) and \( \tilde{\Lambda}_i \)'s as sub-matrices of \( \hat{\Pi} \).

By applying the LS principle to (2.3), the LS estimators of \( \Pi \) and \( \nu \) can also be written in an alternative form as

\[
\hat{\Pi} = \left( \sum_{t=1}^{n} W_t W'_{t-1} \right) \left( \sum_{t=1}^{n} W_{t-1} W'_{t-1} \right)^{-1}.
\]
By Theorem 2.1, it is clear that \( \hat{\theta} \) is consistent and asymptotically normal (see Lemma 2.2 for the asymptotic distribution of \( \hat{\theta} \)). The YW estimators \( \hat{\theta} \) and \( \tilde{\theta} \) exhibit small sample properties which may differ from those of the LS estimators. However, the differences are only transient and the LS and YW estimators share the same asymptotic properties.

The linear minimum mean squared error (MSE) point forecast of \( y_{n+h} \) made at \( n \) conditional on the data up to \( n \) for \( h \)-steps ahead is given by

\[
y_n(h) = \nu + A_1 y_{n-1} + \ldots + A_p y_{n-p},
\]

where \( y_n(j) = y_{n+j} \) for \( j \leq 0 \). Obviously, this is the conditional expectation of \( y_{n+h} \) given the past values of \( y \). If \( u_t \) is assumed to be normally distributed, then it can be shown that

\[
y_{n+h} - y_n(h) \sim N(0, \Sigma_y(h)),
\]

where \( \Sigma_y(h) \) is the forecast error covariance matrix or MSE matrix derivable from \( \Sigma_u \) and coefficient matrices of the MA(\( m \)) representation of (2.1) (see p30-33 of Lütkepohl; 1991).

The optimal forecast (2.6) can be estimated by replacing unknown coefficient matrices with estimators, i.e.,

\[
\hat{y}_n(h) = \hat{\nu} + \hat{A}_1 \hat{y}_{n-1} + \ldots + \hat{A}_p \hat{y}_{n-p},
\]

where \( \hat{y}_n(j) = y_{n+j} \) for \( j \leq 0 \). Due to the additional uncertainty brought about by parameter estimation, the exact form of the small sample
distribution of the forecasts is unknown. However, under the assumption of normality, Lütkepohl (1991; p86-7) provides the asymptotic distribution of $\hat{y}_n(h)$, details of which are given in Chapter 4. Construction of interval forecasts, or prediction intervals, for VAR forecasts is also important because they provide a more informative assessment of future uncertainty than point forecasts. Issues concerning interval forecasts from VAR models are to be dealt with in detail in Chapter 5.

In practice, the true AR order $p$ is unknown and has to be estimated. The use of the order selection criteria provides a simple and systematic approach of order selection, in that the selected order is obtained by minimizing an appropriate criterion. It also seems popular in practice because other methods may bring difficulties in their implementations. For instance, the autocorrelation method of Box & Jenkins (1976) often involves the difficulty of making a subjective judgment in order selection. Statistical tests, such as portmanteau, likelihood ratio or score tests (for more details, see Chapter 6 of Choi, 1992; Section 4.2 of Lütkepohl, 1991; and Section 4.3 of Reinsel, 1993), may suffer from serious power problems in small samples and are in fact preferred in practice as diagnostic checks rather than order selection devices.

The general form of order selection criteria is given by

\begin{equation}
(2.8) \quad IC(m) = \ln(\det[\hat{\Sigma}_u(m)]) + mK^2 C(n)/n, \quad m = 0, 1, \ldots, P,
\end{equation}

where $P$ is the maximum AR order to be fitted and $\hat{\Sigma}_u(m) = (n-Km-1)/n \hat{\Sigma}_u(m)$, the maximum likelihood estimator of $\Sigma_u$. The term $C(n)$ is a penalty term for the number of freely estimated parameters ($mK^2$) in the model.
order \( \hat{p} \) is determined in such a way that \( \hat{p} \) minimizes IC(m). Various forms of IC(m) have been proposed, and they differ in the specification of C(n).

By setting C(n) = 2, we obtain the AIC of Akaike (1974). Shibata (1980) has shown that AIC provides asymptotically efficient order selection in the sense that 1-step ahead forecast mean squared error is minimized. The FPE of Akaike (1969) and the CAT of Parzen (1974) can be shown to be asymptotically equivalent to AIC (see Chapter 3 of Choi, 1992; and Section 4.3 of Lütkepohl, 1991). It can be seen from the Monte Carlo results of Mills & Prasad (1992) and Lütkepohl (1985) that the small sample performances of FPE and CAT are also similar to that of AIC. However, these criteria are inconsistent: they possess non-zero probability of over-estimating the true order, when it exists, even in the limit. BIC of Schwarz (1978) and Rissanen (1978), with C(n) = \ln(n), and HQ of Hannan & Quinn (1979), with C(n) = \( c \ln(\ln(n)) \), \( c \geq 2 \), are designed to yield a (strongly) consistent order selection in which the limit probability of choosing the true order is 1. There are other consistent criteria available, such as BEC of Geweke & Meese (1981), but they are found to perform similarly with BIC in small samples by Mills & Prasad (1992). It is also found from their Monte Carlo results that, in small samples, BIC tends to under-estimate and AIC tends to over-estimate the true order; while HQ exhibits a tendency of under-estimation which is milder than that of BIC. This is due entirely to the contribution of C(n) in HQ, which is in between those in AIC and BIC. Since other criteria behave similarly in small and large samples to AIC, BIC and HQ, it seems plausible that they can be used as representatives of the others. On this ground, we restrict our attention to AIC, BIC and HQ in this thesis.
2.3 Backward Representation of VAR models

It is well-known that VAR forecasts are generated conditional on past observations of a time series. To incorporate this conditionality more explicitly in bootstrapping VAR forecasts, we employ an alternative VAR representation in bootstrap procedures, which is called the backward VAR representation. In the next section, bootstrap procedures for VAR forecasts using this backward representation are dealt with in more detail. The backward representation was introduced by Box & Jenkins (1976) as a means of estimating starting values of a univariate time series. It was exploited by Thombs & Schucany (1990) for the univariate AR case and proved useful in generating bootstrap forecasts conditional on past observations. As Thombs & Schucany (1990) pointed out, the forward and backward representations share identical stochastic properties in the univariate AR case. However, as we shall see below, they exhibit slightly different, although closely related, stochastic properties in the VAR case. In this section, parameter estimators for the backward VAR parameters are proposed and their asymptotic properties are discussed.

The backward VAR representation is given by

\[(2.9) \quad y_t = \mu + H_1 y_{t+1} + \ldots + H_p y_{t+p} + v_t, \quad t = 0, 1, 2, \ldots, \]

where \(H_i\)'s are \(K\times K\) matrices of coefficients and \(\mu\) is a \(K\times 1\) vector of intercept terms. The \(K\times 1\) vector random sequence \(v_t\) is such that \(E(v_t | J_{t+1}) = 0, \quad E(v_t v'_t | J_{t+1}) = \Sigma_v, \) where \(\Sigma_v\) is a \(K\times K\) symmetric positive definite with finite elements and \(E(v_t v'_s) = 0 \) for \(s \neq t\). The notation \(J_s\) indicates the
set of all available future information at time point \( s \). By post-multiplying (2.1) and (2.9) by \( y_t' \) (assuming \( \mu \) and \( v \) to be zero vectors without loss of generality) and taking expectations, we have

\[
\Gamma(0) = A_1 \Gamma(1)' + \ldots + A_p \Gamma(p)' + \Sigma_u; \quad \text{and}
\]

\[
\Gamma(0) = H_1 \Gamma(1) + \ldots + H_p \Gamma(p) + \Sigma_v.
\]

Comparing the above two equations, it can be seen that \( \Sigma_u \neq \Sigma_v \) in general. It is easy to see that when the \( u_t \)'s are i.i.d. Gaussian, the \( v_t \)'s are also i.i.d. Gaussian. However, if the \( u_t \)'s are i.i.d. non-Gaussian, the \( v_t \)'s are uncorrelated but may be dependent. This was also observed by Thombs & Schucany (1990) for the univariate case. The condition for stationarity of the backward representation (2.9) will be given later in this section.

The VAR(1) representation of (2.9) can be written as

\[
Y_t = \mu + \Omega Y_{t-1} + V_t'
\]

where \( K \times 1 \) vectors \( V_t \) and \( \mu \) are defined as

\[
V_t = (0', \ldots, 0', v_{t-p+1}')' \quad \text{and} \quad \mu = (0', \ldots, 0', \mu)';
\]

while \( K \times Kp \) matrices \( \Omega \) and \( \Sigma_v \) can be written as

\[
\Omega = \begin{bmatrix}
0 \\
- & I \\
H_p' & H_1'
\end{bmatrix} \quad \text{and} \quad \Sigma_v = E(V_t V_t') = \begin{bmatrix}
0 & 0 \\
0 & \Sigma_v
\end{bmatrix}
\]

with \( H_* = [H_{p-1}', \ldots, H_1'] \).
The backward VAR process (2.10) is ensured to be stationary if the sequence \( \Omega^i \) \( (i = 0, 1, \ldots) \) is absolutely summable (see Propositions C7 and C8 of Lütkepohl, 1991); for the definition of absolute summability, see Lütkepohl (1991; p.488). The sequence \( \Omega^i \) is absolutely summable if all eigenvalues of \( \Omega \) have modulus less than one, or equivalently, if all the roots of the reverse characteristic equation \( \det(I_{kp} - \Omega z) = 0 \) have modulus greater than one (see p.10 of Lütkepohl; 1991). Using the identity (see p.12 of Lütkepohl; 1991)

\[
\det(I_{kp} - \Omega z) = \det(I_k - H_1 z - \ldots - H_p z^p),
\]

the condition for stationarity can be expressed, in a similar form to that of the forward representation, as

\[
\det(I_k - H_1 z - \ldots - H_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1.
\]

Consider the mean-adjusted version of (2.10), which can be written as

\[
(2.11) \quad W_t = \Omega W_{t+1} + V_t.
\]

By post-multiplying \( W'_{t+1} \) on both sides of (2.11) and taking expectation, we obtain

\[
(2.12) \quad \Gamma_p (-1) = \Omega \Gamma_p (0) \text{ or } \Omega = \Gamma_p (-1) \Gamma_p (0)^{-1}.
\]

It can be seen that the coefficient matrices for forward and backward VAR models are related in a special way. By combining (2.4) and (2.12), we
have

\[(2.13) \quad \Pi = \Gamma_p(0)\Omega' \Gamma_p(0)^{-1} \quad \text{or} \quad \Omega' = \Gamma_p(0)^{-1}\Pi \Gamma_p(0).\]

To examine more closely the relationship between the two representations, the definition of the similarity between two matrices is introduced.

**Definition 2.1. Similarity of Two Matrices (Horn & Johnson; 1985)**

An \((n \times n)\) matrix \(B\) is said to be **similar** to an \((n \times n)\) matrix \(A\) if there exists an \((n \times n)\) non-singular matrix \(S\) such that \(B = S^{-1}A S\).

It can be shown that two similar matrices share the same characteristic polynomial (see Horn & Johnson, 1985; p45). As a result, they share the same eigenvalues. It is also clear from the structure of the similarity that two similar matrices share the same rank. From (2.13), it follows that \(\Pi\) and \(\Omega'\) are similar matrices. This indicates that \(\Pi\) and \(\Omega'\) share the same eigenvalues, trace, rank, characteristic polynomial and determinant because these features are invariant to transpose operation. Hence, the backward model associated with a stationary forward VAR model is guaranteed to be stationary, and vice versa. Moreover, stochastic properties indicated by eigenvalues of a backward model, such as the number of complex roots and proximity to non-stationarity, can also be captured by examining the eigenvalues of the corresponding forward one. It can also be seen that the variance-covariance matrices of innovation processes associated with the forward and backward representations are related in a special way, as the following lemma shows:
Lemma 2.1. Consider the forward and backward representations of VAR(p) model defined in (2.1) and (2.9), and the associated VAR(1) representations (2.2) and (2.10). Then, it can be shown that

\[ \det(\Sigma_u) = \det(\Sigma_v). \]

Proof.

(a) \[
\Sigma_u = \Gamma_p (0) - \Gamma_p (1)\Gamma_p (0)^{-1}\Gamma_p (-1) = \Gamma_p (0) - \Pi_p \Gamma_p (0) \Pi' = (I_{\kappa_p} - \Pi\Pi) \Gamma_p (0); \text{ and}
\]

(b) \[
\Sigma_v = \Gamma_p (0) - \Gamma_p (-1)\Gamma_p (0)^{-1}\Gamma_p (1) = \Gamma_p (0) - \Omega_p \Gamma_p (0) \Omega' = (I_{\kappa_p} - \Omega\Omega) \Gamma_p (0).
\]

Consider

\[
I_{\kappa_p} - \Pi\Pi = \begin{bmatrix} T_1 & T_2 & \ldots & T_p \end{bmatrix} \quad \text{and} \quad (I_{\kappa_p} - \Omega\Omega) = \begin{bmatrix} O_{(K-1)p\times K_p} \\ S_1 \ S_2 \ \ldots \ S_p \end{bmatrix},
\]

where \( T_i = I_{\kappa} -(A_{p+1} + A_{i-1}) \) and \( S_i = I_{\kappa} -(A_{p+1} + H_{i-1}) \) with \( A_0 = H_0 = O_{\kappa \times \kappa} \).

From (a), it can be seen that
\[
\begin{bmatrix}
\Sigma_u & 0 \\
0 & 0 \\
\end{bmatrix}
= 
\begin{bmatrix}
T_1 & T_2 & \ldots & T_p \\
0 & \ldots & \ldots & 0 \\
\end{bmatrix}
\Gamma_p(0).
\]

Define the matrix \( P \) to be a \( K_p \times K_p \) matrix with the first \( K \) rows of \( \Gamma_p(0) \) being replaced by the first \( K \) rows of \( \Sigma_u \). That is,

\[
P = 
\begin{bmatrix}
\Sigma_u & 0 & \ldots & 0 \\
P_1 \\
\end{bmatrix};
\quad P_1 = 
\begin{bmatrix}
\Gamma(-1) \\
\vdots \\
\Gamma_{p-1}(0) \\
\Gamma(-p+1) \\
\end{bmatrix}_{K(p-1) \times K_p}.
\]

Then, \( P \) can be written as

\[
P = 
\begin{bmatrix}
T_1 & T_2 & \ldots & T_p \\
0 & \ldots & \ldots & 0 \\
\end{bmatrix}
\Gamma_p(0).
\]

Similarly, using the relation in (b), define the matrix \( Q \) to be a \( K_p \times K_p \) matrix with the last \( K \) rows of \( \Gamma_p(0) \) being replaced by the last \( K \) rows of \( \Sigma_v \). That is,

\[
Q = 
\begin{bmatrix}
O_{1} & \\
0 & \ldots & \ldots & 0 \\
\end{bmatrix};
\quad Q_1 = 
\begin{bmatrix}
\Gamma(p-1) \\
\vdots \\
\Gamma_{p-1}(0) \\
\Gamma(1) \\
\end{bmatrix}_{K(p-1) \times K_p}
\]

Then, \( Q \) can be written as

\[
Q = 
\begin{bmatrix}
I_{K(p-1)} & O_{K(p-1) \times K} \\
S_p & \ldots & S_1 \\
\end{bmatrix}
\Gamma_p(0).
\]

Evaluating the determinants of \( P \) and \( Q \), we have

\[
det(P) = det(\Sigma_u) det[\Gamma_{p-1}(0)] = det(T_1) det[\Gamma_p(0)]; \quad \text{and}
\]

45
\[
\det(Q) = \det(S) \det(\Gamma_{p-1}(0)) = \det(S_1) \det(\Gamma(0)).
\]

It remains to show that \(\det(T_1) = \det(S_1)\). Consider the matrix \(R\) which can be written as

\[
R = \begin{bmatrix}
I_K & A_p \\
H_p & I_K
\end{bmatrix}.
\]

Then, \(\det(R) = \det(I_K - A_p H_p) = \det(I_K - H_p A_p)\), which concludes the proof.

The above lemma states that, although \(\Sigma_u \neq \Sigma_v\), the magnitudes of variability associated with innovation processes of forward and backward VAR representation are equal when measured in terms of the determinant.

LS estimation for the parameters of the backward representation (2.9) can be conducted in a similar way to that of the forward representation (2.1). Assume that a suitable set of time series data generated by (2.9), including \(p\) post-sample values, \((y_1', \ldots, y_n', y_{n+1}', \ldots, y_{n+p})\) is available. Then (2.9) can be rewritten as

\[
Y = HX + V,
\]

where \(Y = (y_1', \ldots, y_n')\) \((K \times n)\)

\(H = (\mu, H_1', \ldots, H_p')\) \((K \times (K_p + 1))\)

\(X_t = (1, y_t', \ldots, y_{t+p-1}')\) \(((K_p + 1) \times 1)\)

\(X = (X_1', \ldots, X_{n+1}')\) \(((K_p + 1) \times n)\)

\(V = (v_1', \ldots, v_n)\) \((K \times n)\)
The LS estimators for $\hat{H}$ and $\hat{\Sigma}_v$ are provided by

$$
\hat{H} = (\hat{\mu}, \hat{H}_1, \ldots, \hat{H}_p) = YX'(XX')^{-1}; \text{ and }
$$

$$
\hat{\Sigma}_v(p) = (Y - \hat{H} X)(Y - \hat{H} X)' / (n - Kp - 1).
$$

The LS principle can also be applied to (2.11) to obtain the LS estimators for $\hat{\Omega}$ and $\hat{\eta}$, which can be written in an alternative form as

$$(2.14) \quad \hat{\Omega} = \left( \sum_{t=1}^{n} W_t W_t' \right) \left( \sum_{t=1}^{n} W_t W_t' \right)^{-1}.$$ 

It is easy to see that $\hat{H}$ is consistent and asymptotically normal, as it is obtained from the regression of $y_t$ on $y_{t+1}$'s where $y_t$ is a stationary process. However, for completeness, the formal proof is given in Theorem 2.2 below. Note that subsequent theoretical results do not require the use of this theorem; they appeal to the relationship between the forward and backward representations. The concept of martingale (see Definition on p458 of Billingsley, 1995) and the martingale central limit theorem (Billingsley, 1968; p206) are employed in proving asymptotic normality.

**Theorem 2.2.** Suppose $u_t$'s satisfy the assumptions given in Theorem 2.1.

Then it can be shown that

1. plim($\hat{H}$) = $H$; and
2. $n^{1/2}|\text{vec}(\hat{H}) - \text{vec}(H)| \overset{d}{\rightarrow} N(0, \Sigma_v^{-1} \otimes \Sigma_v),$
where $\mathbf{V} = \text{plim}(\mathbf{XX}'/n)$.

Proof.

(i) $\hat{H}$ can be written such that

$$\hat{H} - H = \mathbf{VX}'(\mathbf{XX}')^{-1} = 1/n \mathbf{VX}'(1/n \mathbf{XX}')^{-1}.$$ 

Since components of $\mathbf{VX}'$ are uncorrelated, it is also plausible to assume that $\text{plim}(1/n \mathbf{VX}') = 0$ (see, for example, p259 of Stewart, 1991; and p256 of Davidson & MacKinnon, 1993). It can also be assumed that $\text{plim}(1/n \mathbf{XX}') = \mathbf{V}$ is a finite and non-singular matrix since backward VAR models possess finite autocovariance matrices. Hence, $\text{plim}(\hat{H} - H) = 0$.

(ii) Consider $n^{1/2}(\hat{H} - H) = n^{-1/2}\mathbf{VX}'(1/n \mathbf{XX}')^{-1}$, then

$$n^{1/2}\text{vec}(\hat{H} - H) = [(1/n \mathbf{XX}')^{-1} \otimes \mathbb{I}_{k+1}] n^{-1/2}\text{vec}(\mathbf{VX}')$$

where the rule $\text{vec}(ABC) = (C' \otimes A) \text{vec}(B)$ is used.

Since $\mathbf{VX}' = \sum_{t=1}^{n} v_t \mathbf{X}'_{t+1}$, $\text{vec}(\mathbf{VX}') = \sum_{t=1}^{n} (\mathbf{X}_{t+1} \otimes \mathbb{I}_k) v_t$. Let $F_{t+1}$ be the $\sigma$-field generated by $\{v_{t+1}, v_{t+2}, \ldots\}$. It can be seen that

$$E[n^{-1/2}\text{vec}(\mathbf{VX}')] = n^{-1/2} \sum_{t=1}^{n} E[(\mathbf{X}_{t+1} \otimes \mathbb{I}_k) v_t] = 0,$$

since $\mathbf{X}_{t+1}$ and $v_t$ are uncorrelated.
\[ \text{Cov}(n^{-1/2} \text{vec}(VX')) = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} E[(X_{t+1} \otimes I_k) v_t \cdot v'_{s+1} (X'_{s+1} \otimes I_k)]. \]

**Case.** \( t = s. \)

\[ E[(X_{t+1} \otimes I_k) v_t \cdot v'_{t} (X'_{t+1} \otimes I_k)] = E[(X_{t+1} \otimes I_k) E(v_t v'_{t} | F_{t-1}) (X'_{t+1} \otimes I_k)] \]

\[ = E[(X_{t+1} \otimes I_k) \Sigma_v (X'_{t+1} \otimes I_k)] = E (X_{t+1} X'_{t+1} \otimes \Sigma_v) = (v \otimes \Sigma_v) \]

**Case.** \( t > s. \)

\[ E[(X_{t+1} \otimes I_k) v_t \cdot v'_{s+1} (X'_{s+1} \otimes I_k)] = E[(X_{t+1} \otimes I_k) v_t E(v'_{s+1} | F_{s+1}) (X'_{s+1} \otimes I_k)] \]

\[ = 0. \]

**Case.** \( t < s. \) (similarly to the previous case).

Hence, it follows that \( \text{Cov}(n^{-1/2} \text{vec}(VX')) = v \otimes \Sigma_v. \)

Define \( S_m = \sum_{t=m}^{n} (X_{t+1} \otimes I_k) v_t. \) Then it can be seen that

\[ E(S_m | F_{m+1}) = S_{m+1} \text{ and } F_{m+1} \subset F_m, \]

which indicates that \( S_m \) is a martingale sequence (Billingsley, 1995; p458) in the backward form. The martingale central limit theorem (Billingsley, 1968; p206) is applicable to \( S_m \), and it can be shown that

\[ n^{-1/2} \text{vec}(VX') \xrightarrow{d} N(0, v \otimes \Sigma_v). \]
This indicates that

\[ n^{1/2} \text{vec}(\hat{\Pi} - \Pi) \xrightarrow{d} N(0, \Sigma_{\hat{\Pi}}), \]

where \( \Sigma_{\hat{\Pi}} = [\text{plim } (1/n XX')^{-1} \otimes I_k] (\nabla \otimes \Sigma_v) [\text{plim } (1/n XX')^{-1} \otimes I_k] \)

\[ = (\nabla^{-1} \otimes I_k) (\nabla \otimes \Sigma_v) (\nabla^{-1} \otimes I_k) \]

\[ = (\nabla^{-1} \otimes \Sigma_v). \]

This concludes the proof.

Based on the results obtained in Theorems 2.1 and 2.2, it is trivial to see that \( \hat{\Pi} \) and \( \hat{\Omega} \) are consistent, i.e., \( \text{plim}(\hat{\Pi}) = \Pi \) and \( \text{plim}(\hat{\Omega}) = \Omega \). Asymptotic normality of these estimators can easily be established.

**Lemma 2.2.** Suppose \( u_t \)’s satisfy the assumptions given in Theorem 2.1. Then it can be shown that

(i) \( n^{1/2}[ \text{vec}(\hat{\Pi}) - \text{vec}(\Pi)] \xrightarrow{d} N(0, \Gamma_{\Pi}(0)^{-1} \otimes \Sigma_{\Pi}); \) and

(ii) \( n^{1/2}[ \text{vec}(\hat{\Omega}) - \text{vec}(\Omega)] \xrightarrow{d} N(0, \Phi_p(0)^{-1} \otimes \Sigma_{\Omega}), \)

where \( \Phi_p(0) \) is the submatrix of \( \nabla \) which is obtained by excluding its first row and first column.
Proof.

(i) Let $\Delta = [A_1, \ldots, A_p]$. By Theorem 2.1, it can be seen that

$$n^{1/2} [\text{vec}(\hat{\Delta}) - \text{vec}(\Delta)] \overset{d}{\rightarrow} N(0, \Gamma_p (O)^{-1} \otimes \Sigma_u).$$

Let $J = [I_k, O_{k \times kp}]$, then it can be seen that $(\hat{\Pi} - \Pi) = J (\hat{\Delta} - \Delta)$. Applying the vec operator on both sides gives

$$\text{vec}(\hat{\Pi} - \Pi) = (I_k \otimes J) \text{vec}(\hat{\Delta} - \Delta).$$

It can be shown that

$$n^{1/2} \text{vec}(\hat{\Pi} - \Pi) \overset{d}{\rightarrow} N(0, \Sigma_{\Pi}),$$

where $\Sigma_{\Pi} = (I_k \otimes J) (\Gamma_p (O)^{-1} \otimes \Sigma_u) (I_k \otimes J') = \Gamma_p (O)^{-1} \otimes \Sigma_u$.

(ii) Similarly to the case (i). This concludes the proof.

Alternatively, using (2.13) and (i) of Lemma 2.2, it is easy to see that

$$n^{1/2} [\text{vec}(\hat{\Omega}') - \text{vec}(\Omega')] \overset{d}{\rightarrow} N(0, \Gamma_p (O) \otimes [\Gamma_p (O)^{-1} \otimes (\Gamma_p (O)^{-1})].$$

As a further result, it can be shown that forward and backward innovations are related as...
\[ V_t = -\Omega(I_{Kp} - \Omega^{-1}B)(I_{Kp} - \Pi B)^{-1} U_{t+1} \]

\[ = -\Omega(I_{Kp} + C_1 B + C_2 B^2 + \ldots) U_{t+1}, \]

where $B$ is the backshift operator, and $C_1 = \Pi - \Omega^{-1}$ and $C_k = \Pi C_{k-1}$ for $k \geq 2$. By using the above, stochastic structure between forward and backward innovations can be written as

\[ E(U_{t-k} V_{t}^\prime) = \Sigma_u (\Pi')^k \Gamma (0)^{-1} \Sigma_v \text{ for } k \geq 0; \text{ and} \]

\[ E(U_{t} V_{t-k}^\prime) = -\Sigma_u \Omega'; \ E(U_{t} V_{t-1}^\prime) = 0 \text{ for } k \geq 2. \]

That is, the two innovations are in general correlated except when the backward innovations are lagged more than two periods.

2.4 Bootstrapping in VAR models

In Chapter 1, we have seen two resampling methods for econometric and time series models: one is by resampling the residuals of the estimated equation as appears in Freedman (1984) and Thombs & Schucany (1990); and the other is to resample blocks of observations based on the MBB or SB methods (see, for example, Politis & Romano, 1994). This thesis applies both methods to bootstrapping in various aspects of VAR modelling. In the following two sub-sections, detailed exposition of the aforementioned bootstrap methods in the context of VAR modelling are presented and their large sample validity is provided.
2.4.1 Bootstrapping with resampling residuals

In this subsection, the bootstrap procedures to be employed in parameter estimation and forecasting of VAR models are presented. They are based on parametric resampling and assume the VAR order to be known. To simplify the exposition, the bootstrap procedures for parameter estimation are labelled (E-i) to (E-iii) and those for forecasting are (F-i) to (F-iii).

(E-i) Estimation of the VAR(p) model.

Consider a VAR(p) process of known order p including p presample values, i.e., \( (y_{-p+1}, \ldots, y_0, y_1, \ldots, y_n) \). A VAR(p) model is fitted to an observed process by the LS method to obtain \( \hat{A}_i, i = 1, \ldots, p \), \( \hat{\nu} \) and a vector of residuals

\[
\hat{u}_t = y_t - \hat{\nu} - \hat{A}_1 y_{t-1} - \ldots - \hat{A}_p y_{t-p}, \quad t = 1, \ldots, n.
\]

(E-ii) Resampling to get the bootstrap sample.

An empirical distribution function \( \hat{F}_u \) can be constructed from \( (a_t^*)_{t=1}^n = (a_{t_1}, \ldots, a_{t_n}) \), where \( a_t \) is a centered and rescaled version of \( \hat{u}_t \). As Thombs & Schucany (1990) and others pointed out, the residuals are deflated by fitting. They used the rescaling factor \( [n/(n-p)]^{1/2} \), where \( p \) is the AR order. In this thesis, \( [n/(n-Kp-1)]^{1/2} \) is used analogous to the univariate
case of Thombs & Schucany (1990). That is, the square root of the sample size divided by the degrees of freedom of the model.

The bootstrap sample \( Y^* = (y^*_1, \ldots, y^*_n) \) can then be generated recursively as

\[
y^*_t = \hat{\nu} + \hat{A}_1^* y^*_{t-1} + \ldots + \hat{A}_p^* y^*_{t-p} + a^*_t, \quad t = 1, \ldots, n,
\]

where \( p \) starting values are set equal to \( p \) presample values of the generated series, and \( a^*_t \) is a random draw from \( \hat{F}_u \) with replacement. This resampling procedure can be repeated \( b \) times to obtain a set of \( b \) bootstrap samples \( \{Y^*_i\}_{i=1}^b \).

(E-iii) Bootstrap distributions of parameter estimators

For each bootstrap sample, \( \nu \) and \( A_j (j = 1, \ldots, p) \) are estimated yielding the bootstrap distribution \( \{\hat{\nu}^*_i\}_{i=1}^b \) and \( \{\hat{A}_j^*\}_{i=1}^b \). The distribution of \( \hat{A}_j - A_j \) is approximated by the bootstrap distribution \( \{\hat{A}_j^* - A_j\}_{i=1}^b \), and the distribution of \( \hat{\nu} - \nu \) by \( \{\hat{\nu}^*_i - \hat{\nu}\}_{i=1}^b \). The asymptotic validity of these approximations will be established in Theorem 2.3.

The procedure to obtain the bootstrap distribution of VAR forecasts, based upon the backward representation, can be described as follows (see Thombs & Schucany; 1990, for the univariate equivalent):

(F-i) Estimation of the VAR(p) model.
To bootstrap the forecast error distribution, the backward representation needs to be used so that the final p values at the end of every series and prior to the forecasting period are always the same. The LS residuals based on the backward representation can be generated recursively by using the last p values of the series as starting values as

\[ \hat{v}_t = y_t - \mu - \hat{H}_1 y_{t+1} - \ldots - \hat{H}_p y_{t+p}, \quad t = l-p, \ldots, n-p. \]

(F-i) Resampling to get the bootstrap sample.

An empirical distribution function \( \hat{F}_v \) can be constructed from \( \{\epsilon_t\}_{t=1}^n \), where \( \epsilon_t \) is a centered and rescaled version of \( \hat{v}_t \). The bootstrap sample based on the backward representation \( Y^* = (y_1^*, \ldots, y_n^*) \) can be generated recursively based on this representation as

\[ y_t^* = \mu + \hat{H}_1 y_{t+1}^* + \ldots + \hat{H}_p y_{t+p}^* + \epsilon_t^*, \quad t = 1, \ldots, n, \]

where the p starting values are set equal to the last p values of the original series and \( \epsilon_t^* \) is a random draw from \( \hat{F}_v \) with replacement. This resampling procedure can be repeated b times to obtain a set of b bootstrap samples \( (Y^*)_{i=1}^b \). As indicated above, the last p values of each bootstrap sample are identical to the last p values of the generated series and the forecasts generated, based on a bootstrap sample, will be conditional on the last p values of the original series.
(F-iii) Bootstrap forecast error distribution

A VAR(p) model is then fitted using the bootstrap sample obtained at (F-ii) above. The bootstrap forecasts for the period \(n+h\) made at time \(n\) can then be obtained recursively as

\[
y_n^*(h) = \hat{\nu}^* + \hat{A}_1^* y_{n}(h-1) + \ldots + \hat{A}_p^* y_{n}(h-p) + a_{n+h}^*
\]

where \(y_n^*(j) = y_{n+j}\) for \(j \leq 0\), \(a_{n+h}^*\) is a random draw from \(\hat{F}_u\) with replacement, and the \(\hat{\nu}^*\) and \(\hat{A}_1^*\)'s are the bootstrap parameter estimators based on the bootstrap sample using the LS method. Repeating this for each bootstrap sample, the bootstrap distribution of forecasts \((y_n^*(h))_{i=1}^b\) is obtained. The forecast error distribution \((y_{n+h} - \hat{y}_n(h))\) is approximated by the bootstrap forecast error distribution \((\hat{y}_n(h) - y_n^*(h))_{i=1}^b\). The asymptotic validity of this approximation is established in Theorem 2.4.

The nature of the forecast error distributions generated by bias-corrected parameter estimators is also considered. The bias of parameter estimators is estimated from the bootstrap procedures (E-i) to (E-iii) and a bias-corrected coefficient matrix estimators, denoted \(\hat{A}_j^c\), can be calculated as

\[
\hat{A}_j^c = \hat{A}_j - \text{Bias}(\hat{A}_j), \text{ where Bias}(\hat{A}_j) = \hat{A}_j^* - \hat{A}_j
\]

with \(\hat{A}_j^*\) being the sample mean of \((\hat{A}_j^*)_{i=1}^b\). The correction can be made to \(\hat{\nu}^*\) in the same way to yield \(\hat{\nu}^c\). The forecasts based on bias-corrected parameter estimators are generated as
\[ y_n^c(h) = \overset{\cdot}{\nu} + \overset{\cdot}{A}_1^c y_n^c(h-1) + \ldots + \overset{\cdot}{A}_p^c y_n^c(h-p) + \overset{\cdot}{a}^c_{n+h} \]

where \( y_n^c(j) = y_{n+j} \) for \( j \leq 0 \). This procedure can be repeated \( b \) times to generate the bootstrap forecast error distribution with bias-corrected parameter estimators \( (\overset{\cdot}{y}_n(h) - y_n^c(h))_{i=1}^b \).

If the bootstrap is to be an appealing procedure in the context of VAR modelling, a proof of large sample validity is needed, i.e., one which shows that bootstrap parameter estimators converge to the true population parameters in probability (or in distribution) as sample size increases. Freedman (1984) showed that bootstrapping 2SLS estimators in a general stationary dynamic model is valid in large samples. Since a stable VAR(p) model is a special case of the general model of Freedman (1984), the large sample validity of bootstrapping in VAR(p) model can easily be established, as the following theorem shows.

**Theorem 2.3.** Let \((y_{p+1}, \ldots, y_0, y_1, \ldots, y_{n})\) be a realization from a stationary VAR(p) process (2.1) with its innovations satisfying the condition given in Theorem 2.1. Then, along almost all sample sequences, as \( n \to \infty \), conditionally on the data:

(i) \( \overset{\cdot}{A}_j^c \to A_j \) (\( j = 1, \ldots, p \)); and

(ii) \( \overset{\cdot}{\nu} \to \nu \)

in conditional probability.
Proof.

Based on (2.5), the LS estimator for $\Pi$ can be expressed as

$$\hat{\Pi} = \Pi + \omega_n F_n^{-1},$$

where $F_n = \frac{1}{n} \sum W_{t-1} W'_{t-1}$ and $\omega_n = \frac{1}{n} \sum U_t W_{t-1}'. $ The usual asymptotic theory applies in such a way that $F_n \to E(W W') = \Gamma_p (0)$ and $\omega_n \to 0.$ Resampling from LS residuals $\hat{u}_t'$s, bootstrap sample can be generated as

$$W_t^* = \hat{\Pi} W_{t-1}^* + U_t^*,$$

where $U_t^* = \begin{pmatrix} a_t', \ 0', \ ..., \ 0' \end{pmatrix}'$ and $a_t^*$ is as defined in Subsection 2.3.1.

Then the bootstrap estimator can be expressed as

$$\hat{\Pi}^* = \hat{\Pi} + \omega_n F_n^{-1},$$

with $F_n^* = \frac{1}{n} \sum W_t^* W_{t-1}^*$ and $\omega_n^* = \frac{1}{n} \sum U_t^* W_{t-1}^*.$ Theorem 4 of Freedman (1984) says that, as $n \to \infty$, $F_n^* \to \Gamma_p (0)$ in conditional probability and the conditional law of $\sqrt{n} \omega_n^*$ has the same limit as the unconditional law of $\sqrt{n} \omega_n,$ Hence, $\hat{\Pi}^* \to \Pi$ in conditional probability, and it follows that $\hat{a}_j^* \to \hat{a}_j$ and $\hat{v}^* \to v$ in conditional probability for all $j = 1, ... p,$ thereby proving the theorem.

A companion theorem to deal with the backward representation and bootstrap VAR forecasts is now given. To ensure that innovations of the backward model are independent, we assume innovations of the forward model.
to be normal in establishing the asymptotic validity of bootstrap VAR forecasts.

Theorem 2.4. Let \((y_{-p+1}, \ldots, y_0, y_1, \ldots, y_n)\) be a realization from a stationary VAR(p) process as in (2.1) with its innovations satisfying the condition given in Theorem 2.1. Then, under the assumption of Gaussian innovations, along almost all sample sequences, as \(n \to \infty\), conditionally on the data:

(i) \(\hat{A}_j^* \to A_j \) (\(j = 1, \ldots, p\)) and \(\hat{\nu}^* \to \nu\) in conditional probability; and

(ii) \(\hat{y}_n^*(h) \overset{d}{\to} y_{n+h}^*\).

Proof.

From (2.14), the LS estimator for \(\Omega\) can be expressed as

\[
\hat{\Omega} = \Omega + \frac{\delta}{n} D^{-1}_n,
\]

where \(D_n = \frac{1}{n} \sum W_{t+1} W_{t+1}'\) and \(\delta_n = \frac{1}{n} \sum V_t W_{t+1}'\). The usual asymptotic theory applies in such a way that \(D_n \to E(W_t W_t') = \Gamma_p(0)\) and \(\delta_n \to 0\). Resampling from LS residuals \(\hat{v}_t^*\)'s, bootstrap sample can be generated as

\[
W_{t+1}^* = \hat{\Omega}^* W_{t+1}^* + E_t^*.
\]

where \(V_t^* = (0', \ldots, 0', c_{t-1}')'\) and \(E_t^*\) is as defined in Subsection 2.3.1.
Then the bootstrap estimator for \( \Omega \) based on the LS method can be expressed as

\[
\hat{\Omega}^* = \hat{\Omega} + \delta^* D_n^{-1},
\]

with \( D_n^* = 1/n \sum A_{t+1}^* W_{t+1}^* \) and \( \delta_n^* = 1/n \sum V_t^* W_{t+1}^* \). By the same argument as in Theorem 2.3, it can be seen that \( \hat{\Omega}^* \to \Omega \), \( \hat{H}_j^* \to H_j \) (j = 1, ..., p) and \( \hat{\mu}^* = (I - \hat{H}_1^* - ... - \hat{H}_p^*) \bar{y} \to \mu \), convergence being in conditional probability.

However, bootstrap forecasts should be generated with \( \hat{A}_j^* \)'s conditional on past observations. From (2.13), it can be seen that \( \Pi^* = D_n^* \hat{\Omega}^* D_n^{-1} \).

Since \( D_n^* \to \Gamma(0) \) in conditional probability, it is clear that \( \Pi^* \to \Pi \), \( \hat{A}_j^* \to A_j \) (j = 1, ..., p) and \( \hat{\nu}^* = (I - \hat{A}_1^* - ... - \hat{A}_p^*) \bar{y} \to \nu \) in conditional probability.

The asymptotic validity of VAR forecasts can be shown by induction. For \( h = 1 \),

\[
y_n^*(1) = \hat{\nu}^* + \hat{A}_1^* y_n + ... + \hat{A}_p^* y_{n-p+1} + a_{n+1}^*.
\]

Since \( \hat{\nu}^* \) and \( \hat{A}_1^* \to A_1 \) (i=1, ..., p) in conditional probability and, by Theorem 4 of Freedman (1984), \( a_{n+1}^* \to u_{n+1} \), it follows that \( y_{n}^*(1) \) \( \to \) \( y_{n+1} \) as \( n \to \infty \). For \( h = 2 \),

\[
y_n^*(2) = \hat{\nu}^* + \hat{A}_1^* y_n^*(1) + \hat{A}_2^* y_n + ... + \hat{A}_p^* y_{n-p+1} + a_{n+2}^*.
\]

Since \( \hat{A}_1^* y_n^*(1) \to A_1 y_{n+1} \) and by the same argument for the case of \( h = 1 \), it can be seen that \( y_n^*(2) \) \( \to \) \( y_{n+2} \) as \( n \to \infty \). Since \( y_n^*(h) \) \( \to \) \( y_{n+h} \) as \( n \to \infty \) for
all \( h = 1, \ldots, m-1 \). Then

\[
y_n^*(m) = \hat{\nu}^* + \hat{A}_1^* y_{n}^*(m-1) + \hat{A}_2^* y_{n}^*(m-2) + \ldots + \hat{A}_p^* y_{n}^*(m-p) + a_{n+m}^*.
\]

and \( y_n^*(m) \to y_{n+m}^* \) as \( n \to \infty \) and since \( a_{n+m}^* \to u_{n+m}^* \). This concludes the proof.

In constructing a bootstrap prediction interval based on the percentile method as introduced in Chapter 1, the bootstrap percentiles should be shown to be asymptotically valid so that they converge to the true percentiles as the sample size increases. The convergence of a bootstrap percentile to the true one in distribution can be shown by applying the Glivenko-Cantelli Theorem (see p507 of Mood et al., 1974), as in the univariate case of Thombs & Schucany (1990). In the context of the results obtained in Theorems 2.3 and 2.4, the asymptotic validity of VAR forecasts based on bias-corrected parameter estimators can also be constructed. That is, \( \hat{A}_j^c \to A_j \) (\( j = 1, \ldots, p \)) and \( \hat{\nu}^c \to \nu \) in conditional probability, and \( y_n^{c*}(h) \to y_{n+h}^* \).

2.4.2 Bootstrapping with resampling blocks

The bootstrap method based on the MBB and SB methods, labelled (MSB-i) to (MSB-iii), can be summarized as follows:

(MSB-i) Forming the blocks of sub-series.
Consider a stationary K-dimensional process \( Y = (y_1, \ldots, y_n) \). Let \( B(1, h) = (y_{i}, \ldots, y_{i+h-1}) \) denote a block of length \( h \) starting from \( y_i \), where \( h \) and \( i \) are positive integers.

For \( y_j \in B(1, h) \), if \( j > n \), the corresponding block is not defined in the MBB; whereas, in the SB, \( y_j \) is set to \( y_{j-n} \) (circular treatment of data).

(MSB-ii) Obtaining a bootstrap sample

Let \( h_i \) be an i.i.d. random variable with mean \( \bar{h} \) and \( I_i \) be i.i.d. from the discrete uniform distribution on \( (1, \ldots, n) \). Randomly draw blocks from \( \{B(I_1, h_1), B(I_2, h_2), B(I_3, h_3), \ldots \} \), with replacement, to construct a bootstrap sample of size \( n^* \) denoted as \( G = (S_1, \ldots, S_{n^*}) \). Note that \( S_i \) indicates a member of a randomly drawn block.

(MSB-iii) Obtaining the bootstrap distribution of a statistic

Repeat (MSB-ii) \( b \) times to obtain \( b \) bootstrap samples \( \{G^j\}_{j=1}^b \). The statistic of interest, say \( \tau_j^* \), can be calculated from each bootstrap sample to form a bootstrap distribution \( \{\tau_j^*\}_{j=1}^b \).

The major difference between the MBB and SB methods lies in the treatment of block length generation. In the MBB, \( h_i \)'s are fixed to a constant \( \bar{h} \), while they are randomly generated in the SB. In fact, the MBB can be thought of as a special case of the SB, in which \( h_i = \bar{h} \) for all \( i \) with probability one. As a random structure of block length generation for the SB, Politis and Romano (1994) adopted the geometric distribution which
can be written as

\[(2.15) \quad \text{Prob}(h = i) = (1-r)^{i-1}r, \quad 0 < r \leq 1 \text{ and } i = 1, 2, \ldots ,\]

whose mean and variance respectively are \(1/r\) and \((1-r)/r^2\). Another feature of the SB is that it treats the original data as if it is circular when resampling is conducted, as mentioned in Chapter 1. By treating data as circular, the resampled data is stationary conditional on the original data (see Proposition 1 of Politis and Romano; 1994). It should also be noted that the type of random distribution to be used for block length generation is not crucial in achieving stationarity, as pointed out by Politis and Romano (1994; p1304). Other discrete random distributions whose supports consist of positive integers can be used. In Chapter 6, the zero-truncated Poisson distribution, in addition to the geometric distribution, is also considered.

As mentioned in Chapter 1, to ensure \(n^* = n\) for both MBB and SB procedures, resampling is stopped as soon as \(n^*\) becomes greater or equal to \(n\), and residual observations, if any, can be discarded. The bootstrap sample size \(n^*\) can take any arbitrary number which can differ from the original sample size \(n\) (see Liu and Singh, 1992; Fitzenberger, 1995; and Politis and Romano, 1994). A bootstrap sample with \(n^* > n\) can be obtained simply by drawing additional blocks of data. If the objective of the MBB or SB is to approximate the sampling distribution of a statistic in order to evaluate its properties under a sample size same as the original one, \(n^*\) should be equal to the original sample size \(n\). However, if MBB or SB are conducted to find an accurate point estimator of a parameter without any interest being placed on other distributional properties such as standard
error, \( n^* \) can be any sensible number different from \( n \). Presumably, more accurate bootstrap estimators can be obtained with \( n^* \) larger than \( n \), because, by additional resampling, more information can be incorporated. As Freedman (1981) argued, likely performance of a data set when sample size is larger can be evaluated. However, it seems reasonable to postulate that information content in a data set is finite and there would be a point where the information is exhausted by repeated resampling. In this respect, the choice of \( n^* \) value may be an important factor in improving the accuracy of bootstrap estimators. The choice of \( n^* \) in the context of VAR order selection is discussed in more detail in Chapter 6.

Let \( \bar{h} \) be the average block length. For the MBB, all \( h_i \)'s are fixed at \( \bar{h} \) and \( \bar{h} = 1/r \) for the SB with the geometric distribution given in (2.15). Past studies on the MBB and SB have found that the choice of (average) block length is of great importance; their small sample performances are affected to a great extent by the choice of block length. However, they have failed to provide any sensible guideline for the optimal or reasonable choice of the block length. Recently, Hall et al. (1995) found that optimal choice of block length depends heavily on context. In fact, Künsch (1989) suggested the use of "subjective judgment based on sample correlations". Hence, in choosing the block length, several factors should carefully be considered. It is plausible to assume that a process with longer memory requires longer block length. For example, an AR(1) process would require a longer block than an MA(1) because its autocorrelations gradually die out while those of MA(1) cut off at lag 1. If bootstrapping is conducted for an AR(1) process with a short block length, say 2, whole dependence structure of the original data will not be replicated in bootstrap sample. Hence, the length of memory will play an important role in choosing the block length.
in practice. Another important factor is the sample size; As sample size increases, a high order sample autocorrelation whose population value is non-zero becomes more important. This suggests that a longer block may be considered if the process under investigation has larger sample size. It is also important to consider block length relative to the sample size. If a block length too long relative to the sample size is used, the bootstrap sample generated will exhibit little variation of the original data. For example, the block length 20 for sample size 30 will provide bootstrap samples nearly identical to the original data. This indicates the possible existence of a trade-off between short and long block lengths. That is, a longer block can replicate the higher order dependence structure present in the original data, but will produce a bootstrap sample which exhibits little variation of the original data; and vice versa for a shorter block.

Asymptotic validity of the MBB procedure has been discussed by several authors: that of bootstrap sample mean is established by Künsch (1989) for general stationary processes; and by Liu and Singh (1992) for $q$-dependent processes. Fitzenberger (1995) showed that the bootstrap LS estimators of a regression model, when observations are generated as strong mixing sequences, are asymptotically valid. Politis and Romano (1994) established the asymptotic validity of the sample mean obtained with the SB method. In establishing the asymptotic validity here, we assume the underlying process to be strictly stationary and ergodic. An ergodic process can be described as an identically distributed stationary process whose observations are asymptotically independent. For example, consider the case of univariate AR(1) process where covariance between two of its members $k$-period apart can be written as $\sigma^2 \alpha^k / (1-\alpha^2)$, where $\sigma^2$ is the innovation variance and $\alpha$ is the AR parameter with $-1 < \alpha < 1$. It can be seen that observations in AR(1)
processes are asymptotically independent because the covariance tends to zero as the time difference between two observations $k$ tends to infinity.

A formal and rigorous treatment of ergodic processes can be found in White (1984; Sec. III.3). To make the exposition simple, however, we adopt the definition of an ergodic process given by Davidson & MacKinnon (1993). Since the definition of ergodic process accompany the concept of strict stationarity, the definition of strict stationarity is given first as below:

**Definition 2.2. Strict Stationarity**

Let $G_1$ be the joint distribution function of the sequence $(Z_1, Z_2, \ldots,)$ and let $G_{1+\tau}$ be the joint distribution function of the sequence $(Z_{1+\tau}, Z_{2+\tau}, \ldots)$. The sequence $(Z_t)$ is strictly stationary if $G_1 = G_{1+\tau}$ for each $\tau \geq 1$.

The above definition states that a process is said to be strictly stationary if two sets of consecutive observations $\tau$ period apart have the identical joint distribution function. Now the definition of ergodic process is stated below:

**Definition 2.3. Ergodicity (Davidson & MacKinnon, 1993; p132)**

The strictly stationary sequence $(y_t)$ is said to be **ergodic** if, for any two bounded mappings $Y: \mathbb{R}^{k+1} \to \mathbb{R}$ and $Z: \mathbb{R}^{l+1} \to \mathbb{R}$,
\[
\lim_{n \to \infty} |E(Y_{y_1}, \ldots, y_{1+k})Z(y_{1+n}, \ldots, y_{1+n+1})| \\
= |E(Y_{y_1}, \ldots, y_{1+k})| |Z(y_{1+n}, \ldots, y_{1+n+1})|.
\]

An ergodic process has the property that two sets of observations are asymptotically independent. That is, two sets of observations are closer to independence if they are further apart from each other. A stationary VAR process is an ergodic process, since its autocorrelations decline geometrically as the time difference between two observations increases (see, for the univariate AR case, Sec. 3.3 of Brockwell & Davis, 1991). Now we introduce the Ergodic theorem, which is the law of large numbers applicable to ergodic processes.

**Theorem 2.5.** Ergodic Theorem (White, 1984; p42)

Let \( Z_t \) be a strictly stationary ergodic sequence with \( E|Z_t| < \infty \). Then \( \bar{Z}_n = 1/n \sum_{t=1}^{n} Z_t \to E(Z_t) \) almost surely.

Note that convergence occurs almost surely in the above theorem. See White (1984; Definition 2.9) for the definition of almost sure convergence or strong consistency, which in turn implies convergence in probability (plim) or weak convergence (see White, 1984; Theorem 2.24).

It is mentioned earlier that the magnitude of sample size can play an important role in choosing the (average) block length for MBB and SB. For ergodic processes such as VAR, high order sample autocorrelations, whose
population values are non-zero, become important as sample size increases. Hence, in establishing the asymptotic validity, it is reasonable to assume that the block lengths increase with the sample size. For example, Liu and Singh (1992) assumed that $\tilde{h} = O(n)$ so that $\tilde{h}/n$ is bounded below and above; Fitzenberger (1995) assumed that $n^* = O(n)$, $n = O(n^*)$ and $\tilde{h} = o(n^{1/2})$; Künsch (1989) has found that the choice of $\tilde{h} = O(n^{1/3})$ minimizes the mean squared error of the MBB estimator of variance. Politis and Romano (1994) assumed that $1/\tilde{h} \to 0$ as $n \to \infty$, which indicates that $\tilde{h} = O(n^k)$ with $k > 0$. Hence, with the exception of Politis & Romano (1994), past studies assume that block length increases with sample size at a lower or the same rate. However, if $\tilde{h} = O(n^k)$ with $k > 1$, $\tilde{h}$ may soon exceed $n$ as $n$ gets larger, in which case the bootstrap sample is identical to the original one. Hence, it seems reasonable to consider $0 < k \leq 1$. If we allow $n^* > n$, it is also reasonable to assume that $n^* = O(n)$. Hence, in establishing the asymptotic validity in the context of VAR modelling in this thesis, it is assumed that $\tilde{h} = O(n^{1/2})$ and $d = O(n^{1/2})$ so that $n^* = O(n)$, where $d$ is the mean number of blocks to be drawn.

It can be shown that the MBB and SB procedures produces a bootstrap sample whose autocorrelation structure is asymptotically the same as that of the original process as the sample size goes to infinity. For simplicity in proof, assume the process to have zero mean without loss of generality.

Theorem 2.6. Let $\{y_t\}$ be a K-dimensional strictly stationary ergodic sequence with $E|y_t| < \infty$. Suppose that $\{y_t\}$ is uniformly bounded in probability and that $E(y_t y_{t-k}') = \Gamma(k)$. Suppose that the bootstrap is conducted as in (MSB-i) to (MSB-iii) and that $\tilde{h} = O(n^{1/2})$ and $d = O(n^{1/2})$.
Then, for a given set of bootstrap sample \( \mathcal{G}^j \), it can be shown that

\[
C(k) = \frac{1}{n^*} \sum_{k+1}^{n^*} S'_t \gamma'_{t-k} \to \Gamma(k) \text{ almost surely (a.s.) as } n \to \infty,
\]

where \( \mathcal{G}^j \) and \( S_i \)'s are as defined in (MSB-i) to (MSB-iii).

Proof.

For a given set of bootstrap sample \( \mathcal{G}^j \), we divide \( S'_t \gamma'_{t-k} \) into two groups. The first includes those which consist of \( y'_t \)'s k period apart and the second includes those otherwise. That is,

(a) \( S'_t \gamma'_{t-k} = y_s y'_{s-k} = g_k(t) \) if \( S_t, S_{t-k} \in B(I,h) \);

(b) \( S'_t \gamma'_{t-k} = y_s y'_{s-w} = g(w) \) if \( S_t \in B(I,h) \) but \( S_{t-k} \notin B(I,h) \),

for some I and h, where \( k \neq w \) in general and \( t = 1, \ldots, n^*, s = 1, \ldots, n \), and \( w \) is any integer in between 1 and \( n \). Note that, on average, there are at least \( (n^* - k \bar{d}) \) terms in (a) and at most \( k(\bar{d} - 1) \) terms in (b). Since \( \bar{d} = O(n^{1/2}) \) and \( n^* = O(n) \), the proportion of the members in (a) will increase as \( n \) increases, while that of (b) will decrease.

Let \( C(k) = \frac{1}{n^*} \sum_{k+1}^{n^*} S'_t \gamma'_{t-k} \) be the estimator of autocovariance of order \( k \) for the bootstrap sample. Then it can be seen that

\[
C(k) = \frac{1}{n^*} \sum_{t \in \{a\}} g_k(t) + \frac{1}{n^*} \sum_{t \in \{b\}} g_w(t) = \frac{1}{n^*} \sum_{t \in \{a\}} g_k(t) + O(n^{-1}).
\]

Hence \( C(k) \to \Gamma(k) \) as \( n \to \infty \) a.s. by the Ergodic theorem (Theorem 2.5). This
concludes the proof.

Since a strictly stationary VAR process is an ergodic sequence, the autocorrelation structure of a bootstrap sample based on a VAR model is valid in large samples by Theorem 2.6. This result provides the basis for establishing the asymptotic validity of bootstrap parameter estimators when the MBB is conducted on VAR models. In Theorem 2.7, we consider the case of fitting a VAR(m) process to a VAR(p) model when m ≥ p.

**Theorem 2.7.** Let \( \{y_t\} \) be a strictly stationary process which follows an VAR(p) model defined in (2.1). Suppose that the bootstrap is conducted as in (MSB-i) to (MSB-iii) and that the assumptions in Theorem 2.6 are satisfied. Then it can be shown that

\[
\left\{ A_i^* \right\}_{i=1}^m \to \left\{ A_i \right\}_{i=1}^m \ a.s.; \text{ and}
\]

\[
\text{plim} [\hat{\Sigma}_u^*(m)] = \Sigma_u \text{ for } p \leq m,
\]

where \( A_{p+1} = 0 \) for \( i > 0 \) and \( \hat{\Sigma}_u^*(m) \) is the bootstrap counterpart of \( \Sigma_u \).

**Proof.**

The YW equations can be written as

\[
\Gamma(j) = A_1 \Gamma(j-1) + \ldots + A_q \Gamma(j-q), \ j = 1, \ldots, m.
\]

70
The YW estimators $\{\widehat{A}_i^*\}_{i=1}^m$ satisfy

$$C(j) = \widehat{A}_1^* C(j-1) + \ldots + \widehat{A}_m^* C(j-m), \ j = 1, \ldots, m.$$ 

By Theorem 2.6, $C(k) \to \Gamma(k)$ a.s. as $n$ tends to infinity.

Hence, $\{\widehat{A}_1^*\}_{i=1}^m \to \{A_1\}_{i=1}^m$ a.s. for $p \leq m$. By resorting to the well-known result that the YW estimators are asymptotically equivalent to the LS estimators, we see that $\{\widehat{A}_1^*\}_{i=1}^m \to \{A_1\}_{i=1}^m$ a.s. By Slutsky's theorem, $\operatorname{plim}[\Sigma^*_u(m)] = \Sigma_u$ for $p \leq m$. This concludes the proof.

In Chapter 6, the MBB and SB procedures introduced here are applied to the problem of small sample (V)AR order selection.

2.5 Concluding Remarks

This chapter is designed to review theoretical aspects of the issues to be dealt with in this thesis. Parameter estimators and their asymptotic properties are given for both forward and backward VAR models. Useful results are found by examining the stochastic structure and estimation methods of the backward representation of VAR models. These results will prove to be valuable not only in bootstrapping VAR forecasts conditionally on past observations, but also in other cases where VAR modelling in the backward form may be required. For example, in backcasting stationary VAR or VARMA models, parameter estimators for backward VAR models and their
asymptotic properties discussed in this chapter will be useful.

This chapter also presents bootstrap procedures to be used in various facets of VAR modelling. Bootstrap procedures based on parametric resampling, which has been conventionally used in the literature, are applied to bootstrap VAR parameter estimators and forecasts; and those based on newly emerging non-parametric resampling are also discussed. These bootstrap procedures are proved here to be asymptotically valid in the context of VAR models, thus satisfying a requirement that sensible bootstrap procedures should possess. On this basis, they can be used to evaluate unknown small sample properties in VAR modelling, as an alternative to conventionally employed asymptotic approximations.
Chapter 3. Biases of VAR Parameter Estimators

3.1 Introduction

The widespread use of AR models in econometric and time series analyses has prompted many theoreticians to examine the small sample properties of AR parameter estimators. The LS and YW estimators are popularly adopted in practice when estimating AR parameters, and past studies found these estimators to be, in general, biased in small samples (see, Sawa, 1978; Tjostheim & Paulsen, 1983; and Shaman & Stein, 1988, among others). This is basically because the current value of an AR model is generated as a linear function of its past values together with an innovation, which can be regarded as a case of lagged dependent variables in the context of linear regression. The presence of lagged dependent variables renders AR parameter estimators biased in small samples. As a simple example, consider a stationary univariate AR(1) process of the form

\[ y_t = \alpha y_{t-1} + u_t, \]

with \( u_t \) being i.i.d. with zero mean and \(-1 < \alpha < 1\). Given a series of \( n \) observations, the LS estimator, \( \hat{\alpha} \), for \( \alpha \) can be written as

\[ \hat{\alpha} = \alpha + \sum_{t=2}^{n} \frac{y_{t-1} u_t}{\sum_{t=2}^{n} y_{t-1}^2}. \]

Even though \( E(y_{t-1} u_t) = 0 \) for all \( t \), the expectation of the second term in the above expression is not, in general, zero as the numerator is not independent of the denominator. This example can be extended to the VAR
case to show that the LS parameter estimators are, in general, biased.

Since biases are present in VAR parameter estimators, it is of interest to examine how serious they are in small samples and how, if at all, they are related to the parameters of the VAR model. Past studies have paid attention mainly to univariate AR models. Sawa (1978) numerically evaluated the exact bias of the LS coefficient estimator of $\alpha$ in (3.1). He concluded that bias can be serious for sample sizes of less than 100, and that bias worsens as the model tends towards non-stationarity, i.e., $|\alpha|$ approaches unity. Nankervis & Savin (1988) re-calculated and extended Sawa's (1978) numerical evaluations and found similar results. Tjostheim & Paulsen (1983) simulated biases of parameter estimators in univariate AR and VAR models of orders 1 and 2, and drew similar conclusions to Sawa (1978). One important result of this work was their conclusion that YW estimators generate substantially larger biases than LS. For this reason, LS estimators are preferred over YW, and we examine biases of LS estimators only. Shaman & Stein (1988) considered univariate AR models up to order six and provided explicit expressions for the biases of parameter estimators to $O(n^{-1})$. They express biases as simple linear functions of the unknown parameters; for example, the bias to $O(n^{-1})$ of $\hat{\alpha}$ is $-2\alpha$.

The purpose of this chapter is to examine the behaviour of small sample biases of VAR parameter estimators by Monte Carlo bootstrapping when the LS method is used. Although Tjostheim & Paulsen (1983) examined the VAR case before, their primary concern was to compare magnitudes of biases generated from LS and YW estimators; and their experimental design features were restrictive. The procedures carried out to obtain bootstrap distributions of LS parameter estimators, $\{\hat{A}_j^* - \hat{A}_j^b\}_{j=1}^b$ and $\{\hat{\nu}_i^* - \hat{\nu}_i^b\}_{i=1}^b$,
are summarized in Section 2.2.1 of Chapter 2. By Theorem 2.1, these bootstrap procedures are asymptotically valid, which provides the basis for using the bootstrap distributions to estimate the unknown properties of sampling distributions \( \{ \hat{A}_j A_j \} \) and \( \{ \hat{\nu} \nu \} \). Let \( \alpha_{j,kl} \) be the \((k,l)\) element of \( A_j \). Then the bootstrap bias of \( \hat{\alpha}_{j,kl} \), the \((k,l)\) element of \( \hat{A}_j \), can be estimated as

\[
\text{Bias}(\hat{\alpha}_{j,kl}) = \hat{\alpha}_{j,kl}^* - \hat{\alpha}_{j,kl}^1
\]

where \( \hat{\alpha}_{j,kl}^* \) is the sample mean of the distribution \( \{ \hat{\alpha}_{j,kl}^{*1} \} \), the \((k,l)\) element of \( \{ \hat{A}_j^{*1} \} \). The bootstrap bias of an element of \( \hat{\nu} \) is estimated in the same manner.

In evaluating small sample biases of VAR parameter estimators, the main concern is their sensitivity to model structure, such as proximity to non-stationarity. In particular, we attempt to find those parts of the parameter space in which small sample biases change rapidly. We also seek to provide a summary of how biases are related to the unknown parameters of the VAR model. For this purpose, response surfaces (see Hendry & Harrison, 1974; and Hendry, 1984) for biases are constructed. A response surface is a regression model which relates the outcomes from a Monte Carlo experiment to the parameter values of the model chosen by the analyst. With a single equation, a satisfactorily specified response surface provides a simple and compact summary of the results of the experiment. It can also be used to interpolate and to infer likely experimental outcomes corresponding to parameter values not chosen by the experimenter. A review of the method of constructing response surfaces is given in the next section.
Initially, to make the Monte Carlo experiments for constructing response surfaces more manageable, the vector of VAR intercepts is assumed to be known. In practice, however, the vector of intercepts will typically not be known and will have to be estimated, so it is of interest to examine the consequences of relaxing the assumption of known intercepts. The small sample behaviour of biases under known intercepts is compared to that of unknown intercepts in Section 3.5. The case of the univariate AR(1) model has been dealt with by Sawa (1978) who found that biases are larger when the intercept is unknown. We examine the case of VAR models of orders 1 and 2 here. Finally, biases used in constructing response surfaces are obtained under the assumption of normal innovations. The sensitivity of small sample biases to departures from innovation normality is investigated. Distributions displaying fatter tails than the normal, as well as asymmetric distributions are considered to represent the various types of innovation processes that may be encountered in practice.

The organization of this chapter is as follows: Section 3.2 reviews the method of constructing response surfaces; Section 3.3 summarizes the experimental designs to be used in constructing response surfaces. The designs here will provide the basis of those for other simulations to be conducted in later sections; Sections 3.4 to 3.5 provides results on small sample biases and response surfaces; the consequence of assuming the vector of intercepts to be unknown and the influence of non-normal innovations on small sample biases are respectively dealt with in Sections 3.6 and 3.7; and conclusions appear in Section 3.8. Throughout the thesis, all tables and figures are presented in the appendices at the end of each chapter.
3.2 Response Surfaces

When designing Monte Carlo experiments, it may sometimes be the case that the parameter values chosen can represent only a limited set of points in the parameter space. In such circumstances, it is not reasonable to claim generality for the results obtained. To try to attenuate such difficulties, the experimental design should be framed in such a way that the selected parameter values effectively span the parameter space. Consider, for example, a stationary univariate AR(1) process in which an open interval \((-1, 1)\) constitutes the parameter space for the AR coefficient. Since stochastic properties associated AR models with positive and negative AR coefficients differ, it is plausible to divide the parameter space into two representative parts based on the sign of the autocorrelation. AR coefficients close to \(-1\) and \(1\) can constitute an additional representative part of the parameter space. This is because stochastic properties of near non-stationary models may differ from those of the models whose autoregressive coefficients are far from \(-1\) and \(1\). Hence, a reasonable choice in this case can be \([-0.95, -0.9, -0.7, ..., 0.7, 0.9, 0.95]\).

The presentation of Monte Carlo results is also an important issue. One way of approaching this task is to construct a response surface (see, for example, Hendry & Harrison, 1974; and Hendry, 1984). According to Hendry (1984), a response surface is a regression model whose dependent variable is the value of the statistic of interest obtained in the experiment and whose independent variables are appropriate functions of the parameter values of the model chosen by the experimenter. Hence, a satisfactorily specified response surface provides a summary of results in
a compact and easily assimilated way. A satisfactorily specified response surface also indicates how the statistic would react when parameter values not chosen in the experiment are considered. In other words, it can reduce what Hendry (1984) termed "the problem of specificity"; namely, that results obtained from a Monte Carlo experiment are associated only with the parameter values chosen by the experimenter, whereas, to be useful, Monte Carlo results which cover the entire parameter space are desired. Hence, it is important to be able to interpolate effectively the experimental outcomes associated with the parameter values not chosen, by using those which have been chosen. Tables and graphs are frequently used to present the results of the experiment, but they may not be useful as a means of interpolation. Response surface can provide an effective way of interpolation, as the interpolated outcomes can be obtained by substituting the parameter values into the response surface; but, as Maasoumi & Phillips (1982) stressed, it is essential to have satisfactorily specified response surfaces for this purpose.

Suppose a Monte Carlo experiment is conducted to evaluate properties of a statistic, such as the bias or standard error. Let \( \theta \) denote the vector of parameters. Then it can be expected that the property, denote it by \( \tau \), is expressible as a function of \( \theta \), \( \tau = \Phi(\theta) \), where the explicit form of \( \Phi \) is unknown. Let \( i \) denote an experiment in which a set of parameter values \( \theta_i \) is chosen and a Monte Carlo estimate of \( \tau_i \), denoted \( \hat{\tau}_i \), is obtained. Then a response surface can be specified as

\[
\hat{\tau}_i = F(\theta_i) + u_i, \quad i = 1, \ldots, M,
\]

where \( F \) is a functional form chosen by the experimenter, \( u_i \) is an error
term and M is the total number of experiments. Note that $u_i$ comprises two components: the measurement error $\tau_i - \hat{\tau}_i$; and the approximation error $\hat{\phi} - F$. The choice of $F$ depends on the nature of the problem to be investigated. But it is usually the case that the fitted linear response surface possesses a linear functional form with its variables being non-linear in the parameters of the model.

The general form of a response surface for the first moment of an estimator is given by Hendry (1984). Let $\phi_j$ ($j = 1, ..., k$) denote appropriate functions of the parameter values chosen in the experiment. This general form can be written as

\[(3.2) \quad n(\hat{\tau} - \tau) = \gamma_0 + \sum_{j=1}^{k} \gamma_j \phi_j + \gamma_k I,\]

where $n$ is the sample size and $I$ represents the inconsistency which is the difference between $\tau$ and the probability limit of $\hat{\tau}$. If $\hat{\tau}$ is a consistent estimator of $\tau$, the value of $I$ is zero. In a simple case where there is only one parameter $\theta$ in the model, the variables $\phi_j$ include functions such as $\theta^2$ or $\theta/n$, etc. In the case of more than two parameters, $\phi_j$'s can include interaction terms between parameters (possibly) scaled by appropriate functions of sample size. From the Monte Carlo experiment given as an example in Hendry (1984), the estimated response surface constructed for the bias of autoregressive coefficient estimator for univariate AR(1) processes of the form (3.1) is

\[(3.3) \quad E(\hat{\alpha} - \alpha) = -1.8 \alpha/n + 43 \alpha^2/n^3.\]

This result compares closely to the theoretical derivation by Shaman &
Stein (1988), which gives \( \lim_{n \to \infty} nE(\hat{\alpha} - \alpha) = -2\alpha \). From this example, it is seen that a correctly specified response surface can provide a good approximation to the true relationship between the statistic of interest and parameter values. For more examples of response surface analysis, see Hendry & Harrison (1974) and Hendry (1984).

According to Hendry (1984; pp949-950), a correctly specified response surface should possess the following properties:

(i) the error term \( u_i \) should be of small magnitude, so that the approximation and measurement errors are small;

(ii) the experimental design must adequately cover the parameter space and the region in which \( \Phi \) changes most rapidly should be highlighted;

(iii) the function \( F() \) must be estimated in an efficient way; and

(iv) the estimated response surface should accurately predict the outcomes not used in its fitting as a function of the corresponding parameter values.

The LS method is used here to estimate response surfaces. Diagnostic checks have to be provided to ensure that the response surface is statistically adequate and correctly specified. The presence of autocorrelation or heteroscedasticity may indicate a mis-specified response surface. It also renders LS estimation inefficient and provides biased standard errors of response parameter estimators. This may, obviously, lead to incorrect decisions on the significance of parameters in the regression

80
equation. The diagnostic checks employed are summarized below:

1. The predictive failure (PF) test, proposed by Hendry (1979), is employed to check whether experimental outcomes for parameter points not used in the fitting of the response surface, but which were included in the Monte Carlo experiments, can be satisfactorily predicted. This test examines the constancy of estimated parameters in a regression equation when the data is divided into two subsets. The first subset is used for estimation and the second for the purpose of evaluating prediction performance. It is based on the idea that econometric relations formulated in the first subset will provide accurate predictions for the data points in the second if the econometric formulations are constant over the two subsets of data. "Predictive failure" therefore indicates that, according to Hendry (1979), the model formulation can be rejected against an unspecified alternative using the subset of data for prediction. Let \( y_1 = X_1 \beta_1 + u_1 \) and \( y_2 = X_2 \beta_2 + u_2 \) be models for estimation and prediction periods, of length \( T_1 \) and \( T_2 \) respectively, with \( u_i \sim \text{IN}(0, \sigma_i^2) \) (\( i = 1, 2 \)). The null hypothesis of parameter constancy can be written as

\[
H_0 : \beta_1 = \beta_2, \quad \sigma_1^2 = \sigma_2^2.
\]

It was shown by Chow (1960) that

\[
PF = f_2^T V^{-1} f_2 / T_2 s_2^2 \sim F_{T_2 \tau T_1 - \omega} \quad \text{under } H_0,
\]

where \( f_2 = y_2 - X_2 b_1 \); \( V = I + X_2 (X_2' X_2)^{-1} X_2' \) with \( I \) being an identity matrix; \( \omega \) is the dimension of \( \beta_1 \); and \( b_1 \) and \( s_1^2 \) are the LS estimators of \( \beta_1 \) and \( \sigma_1^2 \). Hence, the statistic \( PF \) can serve as a tool of testing for the predictive
failure of a response surface by checking the parameter constancy over two subsets of experimental outcomes.

2. The Durbin-Watson statistic (DW) (Durbin & Watson; 1950, 1951) can be used to test for mis-specification in response surfaces. DW values significantly different from 2, signal the presence of autocorrelation and indicate a strong possibility of mis-specification. In obtaining the DW statistic, the experimental outcomes should be ordered according to the changes in a single parameter, as recommended by Maasoumi & Phillips (1982). In constructing response surfaces for biases of VAR parameter estimator, we follow Maasoumi & Phillips (1982) and Davidson & MacKinnon (1993, p761) and order experimental outcomes according to increasing sample size, unless stated otherwise.

3. Mis-specification of response surfaces can also be detected by checking for the presence of heteroscedasticity. Two alternative LS covariance matrices are calculated: one is the conventional one and the other is based on the heteroscedasticity-consistent covariance matrix estimator of White (1980). The latter provides the covariance matrix estimator which is asymptotically valid even when there exists heteroscedasticity of unknown form. According to Hendry (1984), substantial discrepancies between the standard errors obtained from the conventional and White's (1980) covariance estimators can indicate the presence of heteroscedasticity or mis-specification.

Since we are dealing only with stationary VAR models, the parameter space is the collection of the parameter values in the coefficient matrices whose eigenvalues all lie outside the unit circle, when those in the
Intercept vector and variance-covariance matrix of innovations are fixed. In fact, the eigenvalues here refer to the roots of the characteristic equation for the forward VAR model defined in Chapter 2. The term "eigenvalues" are used throughout the thesis only for simplicity of expression. The parameter space can be divided into several different parts based on the properties of eigenvalues. Parameter values associated with eigenvalues that are all real; all complex; and a mixture of real and complex represent one plausible partition of the parameter space. This is basically because stochastic properties of VAR processes associated with each of these parts of the parameter space may differ. For example, a univariate AR(2) model exhibits different autocorrelation properties depending on the properties of its eigenvalues (see Figure 3.2 of Box et al., 1994). Parameters which give eigenvalues close to the unit circle may form an additional region for special scrutiny because VAR processes associated with this region exhibit different stochastic properties from those with eigenvalues well outside the unit circle.

In the bivariate VAR(1) case, the characteristic equation is quadratic and yields two eigenvalues. In this case, the arguments of the last paragraph suggest that the parameter space be divided into three parts: parameters which give eigenvalues: i) both real; ii) both complex; and iii) close to the unit circle. In the VAR(2) case where four eigenvalues are involved, the parameter space is then divided into four parts; namely, parameters which give eigenvalues: i) all real; ii) all complex; iii) 2 real and 2 complex; and iv) close to the unit circle. In designing the Monte Carlo experiments, the parameter values should be chosen so that the associated VAR models effectively span the parameter space, paying attention to each important part of the parameter space. Response surfaces
are constructed for each parameter in the coefficient matrix as a function of its own and other parameters of the model. That is, four response surfaces corresponding to each element of the coefficient matrix would be constructed for the bivariate VAR(1) case.

3.3 Experimental Designs for Response Surfaces

As mentioned in Chapter 1, VAR models can involve large numbers of parameters. A large number of parameters implies a high dimensional parameter space, and may cause difficulties in designing appropriate Monte Carlo experiments for VAR models. Firstly, a substantially large number of parameter choices may be required to form a set of parameters which span the entire parameter space of interest. Secondly, there are many potential regressors in constructing response surfaces. This is because the response surface can take any (linear) functional form which may be non-linear in variables. That is, the right-hand side variables in the response surface can be any function of the parameter values chosen in the experiment. Hence, the number of candidate regressors in the response surface regression can rapidly proliferate, making the specification search for a satisfactory response surface rather difficult. To alleviate this difficulty, a limited structure of VAR model is considered in constructing response surfaces. Attention is restricted to zero-mean bivariate VAR models of orders 1 and 2. Initially, we assume that the vector of intercept terms \( \nu \) is known and equal to a zero vector. The innovation variance-covariance structure in the bivariate VAR(1) case can be written as \( \text{vech}(\Sigma_u) = (\sigma^2_1, \sigma^2_{12}, \sigma^2_2) \), where \( \sigma^2_i \) represents the variance of the ith (\( i = 1, 2 \)) innovation and \( \sigma_{12} \) the covariance; define the correlation
coefficient by \( \rho = \sigma_{12} / \sigma_1 \sigma_2 \). It is well-known that \( \sigma_1^2 \) can be restricted to unity without loss of generality. In an attempt to keep a simple experimental structure for constructing response surfaces, \( \sigma_2^2 \) is also restricted to unity. The consequences of this restriction will be examined using the PF test. As a result of these restrictions, the innovation variance-covariance structure is \( \text{vech}(\Sigma_u) = (1, \rho, 1) \). Only normal innovations are considered below because biases do not seem sensitive to departures from normality (see Section 3.6).

Table 3.1 and 3.2, respectively, summarize the bivariate VAR models of orders 1 and 2 considered in this study. They are chosen according to properties of their characteristic roots. Eleven VAR(1) models (labelled M1a to M1k) and seventeen VAR(2) models (labelled M2a to M2q) are considered. Sample sizes are 30, 60 and 100 with \( \rho = (0.3, 0.8) \). Taking into account variation in \( \rho \) and \( n \), Table 3.1 yields 66 VAR(1) models and Table 3.2 102 VAR(2) models. To conduct the PF test, the experimental outcomes associated with M1c are not included in estimation in the VAR(1) case; and those with M2b and M2j in the VAR(2) case. This means that, in the VAR(1) case, response surfaces are constructed with sixty experimental outcomes, with the six associated with M1c used for the PF test. In the VAR(2) case, ninety outcomes are used, leaving out the twelve associated with M2b and M2j for the PF test. The number of bootstrap iterations \( b \) is set to 300 and Monte Carlo iterations \( N \) to 100.

3.4 Constructing Response Surfaces

We begin by plotting biases of parameter estimators of an element of
the coefficient matrices, scaled by sample size. Figure 3.1 plots absolute scaled biases of the (1,1) element of $A_1$ for VAR(1) models and Figure 3.2 plots those of the (1,1) element of $A_2$ for VAR(2) models, sorted into ascending order. At first glance, it can be seen that absolute scaled biases are less than 2 in most cases. Those greater than 2 can be regarded as biases unusually large in magnitude and hereafter are termed as "large biases". The incidence of large biases will be examined in more detail below. This will help identify the part of the parameter space in which biases change most rapidly, and it will prove to yield useful information when constructing response surfaces.

From Figure 3.1, the groups of points represented by A and B are associated with large biases. They correspond, respectively, to the near non-stationary models M1a and M1h. Similar observations can also be made by plotting the absolute scaled biases of the other elements of $A_1$. In Figure 3.2, the groups of points represented by A to D are associated with large biases. The points at A and B correspond respectively to M2e and M2q, both of which are near non-stationary models. However, the points at C and D (each appears in two locations, associated with $\rho = 0.3$ and 0.8; larger biases occurring when $\rho = 0.8$) correspond to the models M2d and M2k, which are stable models with roots well outside the unit circle. In fact, these models possess a common property that their component series are highly correlated: the correlations are -0.97 and -0.85 for models M2d and M2k respectively. This is evidence that models with highly correlated component series can generate large biases in parameter estimators. Hence, it seems that there are reasons other than near non-stationarity which can generate large biases in parameter estimates. Similar features are evident when plotting the absolute scaled biases of the other elements of $A_1$ and $A_2$.  

86
It is of interest to examine more closely the sensitivity of biases of VAR parameter estimators to the structure of the model. For this purpose, two additional sets of simulations were conducted. Firstly, in model Mia, the $(2, 2)$ element of the coefficient matrix is allowed to vary with increment 0.1 in the range $-0.6$ to $1.3$, yielding 20 cases of VAR(1) models. That is,

$$A_1 = \begin{bmatrix} 0.5 & 0.3 \\ -0.6 & \alpha \end{bmatrix}; \text{ where } \alpha = -0.6 [0.1] 1.3.$$ 

The variation of $\alpha$ from $-0.6$ to $1.3$ changes the eigenvalues of the model monotonically from $(3.33, 2.50)$ to $(1.09, 1.09)$. The values of $n$ considered are 30 and 60 to yield a total of 40 cases, with the value of $\rho$ fixed at 0.8. The values of $b$ and $N$ are as before. This experiment is designed to examine how biases react as the model tends towards non-stationarity. Figure 3.3 plots the mean absolute biases corresponding to each value of $\alpha$. It can be seen that biases decrease initially, but soon steadily increase as the model tends to non-stationarity. A response surface is constructed for this experiment. To conduct the DW test, the experimental outcomes are ordered according to increasing values of $\alpha$; and those associated with $\alpha = 0.5$ are not included in fitting for the purpose of conducting the PF test. Let $\hat{m}_{\text{ab}}$ denote the mean absolute bias. The preferred response surface is

$$n \hat{m}_{\text{ab}} = 0.41 - 0.15 \alpha + 0.51 \alpha^2 + 0.62 \alpha^3 - 0.52 \alpha^4 - 0.69 \alpha^5 + 0.64 \alpha^6$$

$$(0.03) \quad (0.09) \quad (0.20) \quad (0.19) \quad (0.40) \quad (0.18)$$

$$[0.02] \quad [0.05] \quad [0.19] \quad [0.08] \quad [0.36] \quad [0.21]$$

$$= 0.41 - 0.15 \alpha + 0.51 \alpha^2 + 0.60 \alpha^3 (1 - \alpha)(1 - \alpha^2)$$
\[ \hat{R}^2 = 0.9952, \text{ PF: } F_{2,31} = 0.10 \text{ (p-value = 0.90)}, \]
\[ s = 0.02, \text{ DW} = 1.95, \]

where \( s \) denotes standard error of regression. Other specifications which involve lower powers of \( \alpha \) were also tested, but they were all found to be statistically inadequate due mainly to low DW values. The numbers in parentheses ( ) are the conventional LS estimated standard errors and those in parentheses [ ] are estimated standard errors based on White's (1980) heteroscedastic-consistent adjustment. The diagnostics reported suggest that the response surface is statistically adequate and exhibits no evidence of mis-specification. It can be seen that biases of VAR parameter estimators are significantly related to the parameter values of the model in a highly non-linear fashion.

A second additional experiment allows the value of \( \rho \) to vary while fixing all other parameters. A VAR(1) model with coefficient matrix

\[
A_1 = \begin{bmatrix}
0.1 & 0.6 \\
0.8 & 0.1 
\end{bmatrix}
\]

is considered as this model yields extreme values of correlation between the component series when the disturbances are highly correlated. Note that the model yields the eigenvalues 16.63 and 3.36. The variation of \( \rho \) from -0.95 to 0.95 (with increment 0.1) changes the correlation between component series from -0.90 to 0.96. This experiment is designed to elicit the sensitivity of biases to the correlation between component series. In all, there are 40 VAR models here. Figure 3.4 plots the mean absolute biases for each value of \( \rho \). Large increases in mean absolute biases at the extreme values of \( \rho \) are clearly present. The outcomes are ordered according
to increasing values of $\rho$ in constructing a response surface. For the purpose of the PF test, four additional experimental outcomes corresponding to $\rho = \pm 0.9$ and sample sizes of 30 and 60 are used. The preferred response surface is

\[
\hat{\eta} = 0.58 + 0.95 \rho^2 - 11.26 \rho^4 + 45.02 \rho^6 - 70.98 \rho^8 + 39.33 \rho^{10}
\]

\[
(0.70) \quad (5.64) \quad (17.15) \quad (21.78) \quad (9.76)
\]

\[
[0.75] \quad [5.92] \quad [17.71] \quad [22.31] \quad [9.97]
\]

\[
\hat{R}^2 = 0.9681, \quad PF: F_{4,34} = 1.40 \quad (p\text{-value} = 0.25),
\]

\[
s = 0.05, \quad DW = 2.47.
\]

These results agree with the previous finding that biases are significantly related to the parameters of the model in a highly non-linear fashion. It can also be observed from Figure 3.4 that biases change rapidly when absolute values of $\rho$ are greater than 0.75 but they are virtually unchanged when otherwise. An alternative, but more parsimonious, response surface can be obtained by exploiting this feature. Let $\pi = (|\rho| - 0.75)I(\rho > 0.75)$, where $I()$ is an indicator function which takes the value 1 if the condition in the parenthesis is true and is zero otherwise. It can be written as

\[
\hat{\eta} = 0.61 + 9.01 \pi - 127.97 \pi^2 + 567.87 \pi^3
\]

\[
(2.57) \quad (35.84) \quad (117.06)
\]

\[
[1.81] \quad [26.57] \quad [89.76]
\]

\[
\hat{R}^2 = 0.9665, \quad PF: F_{4,36} = 0.85 \quad (p\text{-value} = 0.50),
\]

\[
s = 0.05, \quad DW = 2.35.
\]
The findings suggest that near non-stationarity is not the only cause of large biases in VAR parameter estimators. VAR models whose component series are highly correlated can also generate large biases. Moreover, the biases seem to be related to the parameters of the model in a highly non-linear manner.

Response surfaces for biases of all bivariate VAR(1) parameter estimators are constructed based on (3.2). Attention is restricted to capture biases to O(n^{-1}), as in Shaman & Stein (1988). Suppressing the coefficient matrix subscript, the dependent variable of response surfaces are the values of n(\hat{\alpha}_{kl} - \alpha_{kl}) (k, l = 1, 2) obtained by Monte Carlo bootstrapping. Functional forms of response surfaces are limited to be linear in their parameters, but the variables are allowed to be non-linear combinations of model parameters. Hence, the vector $\gamma_i$ in (3.2), the independent variables of the response surfaces, may include $\alpha_{kl}$, $\rho$, $n$ and their interactions and transformations in non-linear form. The preferred response surfaces are reported in Appendix 3.1. It can be seen that they are all highly non-linear in the parameters of the VAR model. These equations exhibit desirable statistical properties: good fit reflected by high $R^2$s; standard errors of regression are small in magnitude; individual significance of independent variables in most cases; and effective forecastability. The DW values reported suggest no evidence of autocorrelation or mis-specification in each equation. The standard errors of parameter estimators calculated using both the conventional and White's (1980) robust form appear very close for each equation and rarely affect the result on the individual significance of parameters. This suggests no indication of heteroscedasticity or mis-specification.
The plots of the preferred response surfaces are presented in Figures 3.5 to 3.8. The preferred response surface of \( n(\hat{\alpha}_{kl} - \alpha_{kl}) \) is plotted against \( \alpha_{kl} \) for \( \rho = 0.3 \) and 0.8 when other \( \alpha_{ij} \) values are set to 0. It should be noted that the plots are used to describe the shapes of response surfaces, not to perform out-of-sample forecasting. As Maasoumi & Phillips (1982) pointed out, extreme care should be taken to conduct out-of-sample forecasting based upon estimated response surfaces. For all four cases, it can be seen that \( n(\hat{\alpha}_{kl} - \alpha_{kl}) \) is sensitive to changes in \( \rho \) and \( \alpha_{k1} \) in a non-linear fashion. The structure of inter-dependence observed in the preferred response surfaces should also be noted. That is, the biases of \( \hat{\alpha}_{11} \) and \( \hat{\alpha}_{22} \) share the same functional form and depend on \( \alpha_{11} \), \( \alpha_{22} \) and \( \rho \) and their interactions. Similarly, the biases of \( \hat{\alpha}_{12} \) and \( \hat{\alpha}_{21} \) share the same functional form and depend on \( \alpha_{12} \), \( \alpha_{21} \) and \( \rho \), and their interactions.

It was mentioned earlier that attention is restricted to VAR processes with a limited structure in an attempt to reduce the number of parameters. This limited structure allows the disturbance variance-covariance matrix to be of the form \( \text{vech}(\Sigma_u) = (1, \rho, 1)' \). One may argue that the structure of disturbances dealt with here is too restrictive and that the response surfaces constructed are not suitable for explaining small sample behaviour of biases which arise with a more general structure. To defend this point, three VAR(1) models (M1b, M1c and M1k) are examined with two variance-covariance structures of disturbances: \( \text{vech}(\Sigma_u^1) = (1, 0.424, 2)' \) and \( \text{vech}(\Sigma_u^2) = (1, 1.6, 4)' \), implying that the correlation coefficient is, respectively, 0.3 and 0.8. The PF test is used to check whether the biases generated with these structures are adequately predicted by using the response surfaces constructed above. The results are presented in Table 3.3 where F statistics and p-values are reported for each response surface. In
each case, it can be seen that the null hypothesis of predictive adequacy is not rejected for any significance level typically used in statistical inference. Hence, it seems likely that little generality is lost by employing the limited disturbance variance-covariance structure used here.

We now turn to estimating response surfaces for biases of VAR(2) parameter estimators. Although various specifications of response surfaces were considered, none of them were found to be statistically adequate. In Appendix 3.2, examples of response surfaces for the VAR(2) case are given. They possess the structures of non-linearity and symmetry which were observed in the case of VAR(1). However, in most cases, very high values of $s$ and low values of $\bar{R}^2$ are reported. This means that these response surfaces do not fit very well to the associated experimental outcomes, suggesting a strong possibility of the presence of substantially large approximation and measurement errors [see (i) in Section 3.2]. Many other complicated structures were attempted but without any noticeable improvement. This may be attributable to the fact that VAR(2) models involve a larger number of parameters. Since construction of response surfaces for biases of VAR(2) parameter estimators should be feasible in principle, statistically adequate response surfaces for the VAR(2) case can be obtained if a large number of additional experimental outcomes are attached to the existing ones. This can be done by employing additional choices of parameter values and conducting additional experiments. Moreover, there are eight response surfaces to be constructed for the VAR(2) case, which are highly likely to be non-linear functions of the nine parameters and their interactions. This suggests that, even with the additional experimental outcomes, a specification search for a correctly specified response surface in the VAR(2) case would be very difficult. It
seems that alternative techniques to linear regression may have to be employed to cope with complexities in estimating response surfaces for the VAR(2) case; the method based on neural networks (see, among others, Kuan & White, 1994) is a candidate. However, this method is not pursued here and is a subject of future research. Nevertheless, the findings of this section suggest that biases of VAR(2) parameter estimators behave similarly to those of the VAR(1) parameter estimators. As we have seen in Figure 3.2, biases in the VAR(2) case are found to be especially sensitive when the model is nearly non-stationary or when the component series are highly correlated. It is also likely that response surfaces of VAR(2) parameter estimators are highly non-linear and complicated functions of the parameter values of the model.

3.5 Some out-of-sample forecasting results

As we have seen earlier, Shaman & Stein (1988) found that the theoretical bias to $O(n^{-1})$ for the LS estimator of the AR(1) coefficient, $\alpha$, is $-2\alpha$. In this chapter, response surfaces for biases of the VAR(1) parameter estimators are constructed and found to be statistically adequate. An interesting comparison can be made by imposing some parameter restrictions on the estimated VAR(1) response surfaces. By setting appropriate parameter values to zero, we can obtain expressions for univariate AR(1) parameter estimator biases. They can be compared with the theoretical expression $-2\alpha$. We pay attention to response surfaces of $n(\hat{\alpha}_{11} - \alpha_{11})$ and $n(\hat{\alpha}_{22} - \alpha_{22})$. These response surfaces are re-estimated by adopting the same specification and simulation results as before, except for the intercept term being restricted to zero. This is to be compatible
with the expression \(-2\alpha\) in the univariate AR(1) case which indicates that the true response surface cannot involve the intercept term. Although not reported, it is found that parameter estimates and results of diagnostic tests in VAR(1) response surfaces are not sensitive to the exclusion of the intercept term. By setting \(a_{22} = a_{12} = a_{21} = \rho = 0\) for the re-estimated response surface of \(\hat{\sigma}_{11} - \sigma_{11}\) and \(a_{11} = a_{12} = a_{21} = \rho = 0\) for that of \(\hat{\sigma}_{22} - \sigma_{22}\), we obtain the univariate AR(1) case as

\[
\begin{align*}
\hat{\sigma}_{11} - \sigma_{11} &= -2.01 \sigma_{11} - 0.12 \sigma_{11}^2; \text{ and} \\
\hat{\sigma}_{22} - \sigma_{22} &= -1.58 \sigma_{22} + 0.16 \sigma_{22}^2.
\end{align*}
\]

It can be seen that the coefficient of \(\sigma_{11}\) is nearly \(-2\) and that of \(\sigma_{22}\) is reasonably close to \(-2\). Using the standard error estimate based on White's (1980) heteroscedasticity-consistent covariance matrix estimator, obtained from the response surface fitted without the intercept term, we cannot reject the null hypothesis that the coefficient of \(\sigma_{22}\) is equal to \(-2\) at the 1% significance level (p-value = 4.9%). This indicates that the use of VAR(1) response surfaces as a means of forecasting the univariate AR(1) outcomes can provide fairly accurate results; this may be a desirable property that VAR(1) response surfaces should possess.

3.6 Effect of unknown intercepts

So far in this chapter, the vector of intercepts has been assumed known, in order to reduce the number of parameters involved in constructing
response surfaces. In practical applications of VAR models, however, the vector of intercepts is unknown and has to be estimated. Inclusion of intercept terms in the model adds sampling variability. Consequently, parameter estimators may suffer from larger biases in small samples than when intercept terms are treated as known parameters; as in the univariate case of Sawa (1978). All VAR(1) models except M1f and M1g in Table 3.1 are chosen to examine the effect of unknown intercepts. The VAR(2) models selected are M2a to M2e in Table 3.2. Note that those models which were found to generate large biases in parameter estimators are included, along with the near non-stationary models M1a, M1h and M2e, and the high correlation model M2d. In every experiment, the vector of intercepts is set to (0, 0)', but is assumed to be unknown and estimated. With other designs being as before, there are 54 cases of VAR(1) and 30 cases of VAR(2). Only normal innovations are considered because biases are found to be robust to departures from normality of innovations, as will be seen in the next section.

Figures 3.9 and 3.10 depict biases of VAR parameter estimators when the vector of intercepts is known and unknown; the former for the VAR(1) case and the latter VAR(2). What is plotted are the mean absolute biases across all VAR parameter estimators, excluding intercepts, multiplied by sample size. For VAR(1) models, biases in parameter estimators with unknown intercepts are always larger than those with known intercepts. This is also true in the VAR(2) case. It is noticeable that biases exhibit similar patterns regardless of the presence of intercept terms; a feature also evident in the univariate AR case examined by Sawa (1978). This means that a model which generates large biases with known intercepts also generates large biases when intercepts are unknown. Consider, for example, the points
in Figure 3.10 associated with the near non-stationary model M2e which are labelled 25 to 30, and the points associated with the high correlation model M2f labelled 19 to 24. They exhibit larger biases than the others for both known and unknown intercepts. Hence, for all cases considered, biases are found to be larger when the vector of intercepts is estimated. However, the basic properties of small sample biases, such as the sensitivity of biases to near non-stationarity and high correlation in component series, seem to remain unaffected.

3.7 Biases under non-normal innovations

The investigation has so far assumed the innovations of VAR processes to be normally distributed. However, the normality assumption may be too restrictive, especially in small samples. It may often be the case that the observed residuals exhibit a disproportionate number of occasions on which they would be considered outliers under a normal innovation model so that it is more plausible to assume a distribution with fatter tails than the normal for innovations. Residuals may also be observed in such a way that an asymmetric distribution is more appropriate than a symmetric one. When the assumption of normality is no longer tenable, it is of interest to examine the consequence of non-normality on biases of VAR parameter estimators in small samples. This section examines sensitivity of biases to departures from normality of innovations. The univariate Student's $t$ distribution is used as a basis for generating a bivariate random variable which is a representative of distributions with fatter tails than the normal, and centred exponential and chi-squared distributions are used as representatives of asymmetric distributions. The non-normal distributions
to be used here and elsewhere in this thesis are introduced in more detail below.

A variant of the multivariate Student t distribution (henceforth called the approximate Student t) is employed here. Breusch, Robertson & Welsh (1993) showed that empirical implications of multivariate t models are the same as those of the normal model, and that they are indistinguishable for the purpose of statistical inference. Hence, it is preferable that any K-dimensional error process be drawn from K independent univariate t distributions rather than a K-dimensional multivariate t distribution. The imposition of a correlation structure on the K independent univariate distributions involves a linear combination of t distributions, which will not be distributed exactly as a t variate but it is still fat-tailed with moments and quantiles close to those of t.

An illustration of the procedure suggested in the last paragraph in the bivariate case is now given. Let \( u_1 \) and \( u_2 \) be two independent t variates, each with \( \tau \) degrees of freedom; i.e., \( u_1 \sim t_\tau \) and \( u_2 \sim t_\tau \). Consider a transformation

\[
v_1 = u_1; \text{ and } \quad v_2 = \rho \ u_1 + (k - \rho^2)^{1/2} \ u_2
\]

where \( \rho/\sqrt{k} \) is the correlation coefficient between \( v_1 \) and \( v_2 \). It follows that \( v_1 \sim t_\tau \), but clearly \( v_2 \) is not t distributed. However, preliminary simulations conducted with \( \tau \geq 3 \) reveal that \( v_2 \) (for \( k = 1 \)) is a fat-tailed distribution with moments and quantiles close to those of \( t_\tau \). The value of
k plays the role of rescaling the random variables in a way that $v_2$ has larger variance than $v_1$ if $k > 1$. When $k = 1$, the variances of $v_1$ and $v_2$ are equal.

The bivariate exponential distribution employed here is a version of a distribution given in Moran (1967). It is designed so that each marginal distributions is univariate exponential. Let $x = (x_1, x_2)$, where

$$x_1 = 1/2 (u_1^2 + u_2^2) \text{ and } x_2 = 1/2 (u_3^2 + u_4^2).$$

Here all u's are standard normal variables with $(u_1, u_3)$ and $(u_2, u_4)$ being mutually independent. Moreover, each pair has a bivariate normal distribution with correlation $\rho$. Then it follows that

$$E(x_i) = 1, \text{ VAR}(x_i) = 1 \text{ for } i = 1, 2, \text{ and } \text{Corr}(x_1, x_2) = \rho^2,$$

where Corr($\cdot$) indicates the correlation between two random variables.

Consider a transformation $y_1 = x_1$ and $y_2 = \sqrt{k} \cdot x_2$, where the value of $k$ will play the role of changing the variance of $x_2$ without affecting the correlation structure. The distribution is centred so that it has zero mean for use as a bivariate innovation process.

A bivariate $\chi^2$ distribution is also employed to reflect asymmetry not as extreme as that of the exponential distribution. It can be constructed by taking the sum of squares of standard normal distributions as the following theorem shows:
Theorem 3.1. Let \( X_{i1} \sim \text{NID}(0,1), \ X_{2i} \sim \text{NID}(0,1) \) and \( E(X_{11}X_{2i}) = \rho \ (i = 1, \ldots, m) \). Consider the random variables \( S_1 = \sum_{i=1}^{m} X_{11}^2 \) and \( S_2 = \sum_{i=1}^{m} X_{2i}^2 \), both of which follow \( \chi_m^2 \) distributions. Then, \( \text{Corr}(S_1, S_2) = \rho^2 \).

Proof.

It is clear that \( S_1 \) and \( S_2 \) follow \( \chi_m^2 \) distributions as they are a sum of squares of \( m \) independent standard normal variates. Consider the standardized versions of \( S_1 \) and \( S_2 \)

\[
t_1 = (S_1 - m) / \sqrt{2m} \quad \text{and} \quad t_2 = (S_2 - m) / \sqrt{2m}.
\]

To show that the correlation between \( S_1 \) and \( S_2 \) is \( \rho^2 \), evaluate \( E(t_1 t_2) \).

Note that summations run from 1 to \( m \) in all cases.

\[
E(t_1 t_2) = E\left( \left( \sum_{i=1}^{m} X_{11}^2 - m \right) \left( \sum_{i=1}^{m} X_{2i}^2 - m \right) \right) / 2m
\]

\[
= E\left( \sum_{i=1}^{m} X_{11}^2 \sum_{i=1}^{m} X_{2i}^2 - m \sum_{i=1}^{m} X_{11}^2 - m \sum_{i=1}^{m} X_{2i}^2 + m^2 \right) / 2m
\]

\[
= E\left( \sum_{i=1}^{m} X_{11}^2 \sum_{i=1}^{m} X_{2i}^2 - m^2 \right) / 2m
\]

\[
= E\left[ \sum_{i \neq j} X_{11}^2 X_{2j}^2 + \sum_{i \neq j} \sum_{j \neq l} X_{11}^2 X_{2j}^2 - m^2 \right] / 2m
\]

\[
= \left[ \sum_{i=1}^{m} E(X_{11}^2 X_{2i}^2) + m(m-1) - m^2 \right] / 2m
\]

since \( E(X_{11}^2 X_{2j}^2) = E(X_{11}^2)E(X_{2j}^2) = 1 \) for \( i \neq j \). Now, the bivariate standard normal distribution with correlation coefficient \( \rho \) has moment generating
function

\[ M(q_1, q_2) = \exp\left( q_1^2 + q_2^2 + 2\rho q_1 q_2 \right)/2 \].

Then it can be shown that \( E(x_{11}^2 x_{21}^2) = 2\rho^2 + 1 \) by evaluating \( \frac{\partial^2 M(q_1, q_2)}{\partial q_1 \partial q_2} \) at \( q_1 = q_2 = 0 \). Hence,

\[ E(t_1 t_2) = [m(2\rho^2 + 1) + m(m-1) - m^2]/2m = \rho^2. \]

The above theorem could be extended to the multivariate case, as described in Johnson and Kotz (1972; p225-226). The distribution is centred and standardized so that it has a zero mean vector and unit variances for its components for use as a bivariate innovation process. The transformation \( y_1 = s_1 \) and \( y_2 = \sqrt{k} \ s_2 \) again changes the variance of the second component without affecting the correlation structure.

The VAR(1) models selected are M1a, M1b, M1c and M1e and the VAR(2) models are M2a, M2b, M2d and M2e. Note that models M1a and M2e are nearly non-stationary and M2d has highly correlated component series. These models were found to generate the largest biases in parameter estimators earlier. The variance-covariance structure used for normal innovations is \( \text{vech}(\Sigma_\nu) = (1, 0.3, 1)' \), and equivalent structures, in the sense of preserving the correlation structure between \( u_1 \) and \( u_2 \), are given to each of the non-normal innovations. That is, the value of \( \rho \) given for the approximate Student's t, exponential and chi-squared distributions are, 0.3, \( \sqrt{0.3} \) and \( \sqrt{0.3} \), respectively, with \( k \) set to unity in every case. The vector of
intercepts is fixed at the null vector but free to be estimated. Other experimental features are as before.

Figure 3.11 plots the mean absolute biases of VAR parameter estimators, scaled by sample size, for each model employed and for all types of innovations considered. Three adjacent points correspond to one VAR model, each point representing standardized mean absolute biases associated with sample sizes in the order of 30, 60 and 100. For VAR(1) models, it can be seen that the magnitudes of biases are nearly identical for all types of innovations. The same holds true for the case of VAR(2), with the exception being for models M2d and M2e when approximate Student's t innovations are used. It seems that fat-tail behaviour, combined with a model structure that generates large biases, contributes to instability which causes biases under approximate Student's t innovations to be somewhat different from those under other types of innovations. The differences, however, are not substantial. It can be seen from Figure 3.11 that the maximum difference, which arises when model M2d is employed with sample size 60, is around 1.2. This means that, on average, additional biases of approximately 0.02 in absolute terms are associated for each VAR parameter estimators when approximate Student's t innovations are used. However, in general, we observe that biases of VAR parameter estimators are not sensitive to departures from normality of the innovation processes.

3.8 Concluding Remarks

In this chapter, bivariate VAR models of orders 1 and 2 have been employed to investigate small sample behaviour of biases of VAR parameter
estimators. It is found that near non-stationarity is not the only cause of large biases of VAR parameter estimators. A VAR model with its eigenvalues well outside the unit circle can generate large biases when its component series are highly correlated. Response surfaces for the biases of VAR(1) parameter estimators to $O(n^{-1})$ are found to be highly non-linear and complicated functions of the unknown parameters and their interactions. This contrasts with the findings of Shaman & Stein (1988), where biases to $O(n^{-1})$ in the AR(1) case are found to be simple linear functions of the parameters of the model. This may be due to the complicated structure of VAR models, involving factors such as interactions among the variables in the system. Despite much effort to construct response surfaces for biases of VAR(2) parameter estimators, none of those attempted were found to be statistically adequate. However, the findings of this chapter suggest that response surfaces for biases of VAR(2) parameter estimators may also possess highly non-linear and complicated structures, as those for VAR(1) parameter estimators. It seems that an alternative method of estimating response surfaces to linear regression should be considered to obtain statistically adequate response surfaces. Estimation of response surfaces for the VAR(2) case using neural networks is the subject of future research.

As in the univariate case, VAR models with unknown intercepts are found to generate larger biases in parameter estimators than those with known intercepts. Departures from normality of innovations, however, seem to make little difference to the magnitude of biases of parameter estimators. It appears that small sample performances of biases of the univariate AR parameter estimators observed in past studies also hold in the VAR case. However, biases in the VAR case are found to exhibit more
complicated structure, which seems to reflect the complex structure of VAR models.
Appendix 3.1. Estimated Response Surfaces for VAR(1) Models

Let $a_{ij} = n(\alpha_{ij} - \alpha_{ij})$, then the response surfaces can be expressed as

$$
\begin{align*}
\alpha_{11} &= -0.62 - 1.84 \alpha_{11} + 0.50 \alpha_{11}^2 + 0.62 \alpha_{11}\rho^2 + 19.16 \alpha_{11}^3 \rho^2 \\
&\quad - 22.26 \alpha_{11}^3 \rho^3 - 4.94 \alpha_{22}^3 \rho^2 + 4.41 \alpha_{22}^3 \rho^3 \\
&\quad (3.64) \quad (1.14) \quad (1.37) \quad (2.95) \quad (0.89) \quad (1.06) \\
\alpha &= -0.62 - 1.84 \alpha_{11} + 0.50 \alpha_{11}^2 + 0.62 \alpha_{11}\rho^2 \\
&\quad + 19.16 \alpha_{11}^3 \rho^2 (1-\rho) - 4.94 \alpha_{22}^3 \rho^2 (1-\rho) \\
\bar{R}^2 &= 0.917, \text{ PF: } F_{6,52} = 0.147 (0.989), s = 0.281, \text{ DW} = 1.57.
\end{align*}
$$

$$
\begin{align*}
\alpha_{22} &= 0.07 - 1.61 \alpha_{22} + 0.11 \alpha_{22}^2 - 1.18 \alpha_{11}\alpha_{22} \rho^2 \\
&\quad - 8.48 \alpha_{11}^3 \rho^2 + 10.04 \alpha_{11}^3 \rho^3 + 11.37 \alpha_{22}^3 \rho^2 - 11.89 \alpha_{22}^3 \rho^3 \\
&\quad (2.03) \quad (2.51) \quad (2.33) \quad (2.79) \quad (1.40) \quad (1.73) \quad (2.78) \quad (3.17) \\
\alpha &= 0.07 - 1.61 \alpha_{22} + 0.11 \alpha_{22}^2 - 1.18 \alpha_{11}\alpha_{22} \rho^2 \\
&\quad - 8.48 \alpha_{11}^3 \rho^2 (1-\rho) + 11.37 \alpha_{22}^3 \rho^2 (1-\rho) \\
\bar{R}^2 &= 0.766, \text{ PF: } F_{6,52} = 0.640 (0.697), s = 0.468, \text{ DW} = 2.06.
\end{align*}
$$

$$
\begin{align*}
\alpha_{12} &= -0.28 - 0.83 \alpha_{12} + 0.42 \alpha_{21} - 1.72 \alpha_{12}^2 + 3.53 \alpha_{12}^3 - 0.57 \alpha_{12}^3 \\
&\quad - 3.68 \alpha_{21}^3 - 1.61 \alpha_{12} \rho^2 - 1.94 \alpha_{21} \rho^2 + 3.66 \alpha_{12}^2 \rho^2 + 6.67 \alpha_{21}^2 \rho^2 \\
&\quad (0.46) \quad (0.47) \quad (0.52) \quad (0.50) \quad (0.45) \quad (0.22) \quad (0.29) \quad (0.14) \quad (0.49) \quad (0.31) \quad (0.41) \quad (0.48) \quad (0.82) \\
&\quad + 10.66 \alpha_{12} \alpha_{21} \rho^2 \\
&\quad (1.35) \quad (1.01)
\end{align*}
$$
\[ \bar{R}^2 = 0.842, \text{ PF: } F_{6, 48} = 1.134 \text{ (0.357)}, \ s = 0.262, \ DW = 2.40. \]

\[
\begin{align*}
\alpha_{21} &= -0.18 - 1.32 \alpha_{12} + 0.02 \alpha_{21} + 1.25 \alpha_{12}^2 + 2.36 \alpha_{21}^2 - 0.27 \alpha_{12}^3 \\
&\quad (0.09) \quad (0.25) \quad (0.23) \quad (0.28) \quad (0.45) \quad (0.21) \quad (0.12) \\
&\quad [0.10] \quad [0.26] \quad [0.19] \quad [0.13] \quad [0.18] \quad [0.12] \\
&- 2.32 \alpha_{21}^3 + 3.26 \alpha_{12} \rho^2 + 4.48 \alpha_{21} \rho^2 - 3.41 \alpha_{12}^2 \rho^2 - 10.01 \alpha_{21}^2 \rho^2 \\
&\quad (0.48) \quad (0.46) \quad (0.52) \quad (0.49) \quad (0.93) \quad (0.93) \\
&\quad [0.46] \quad [0.28] \quad [0.40] \quad [0.52] \quad [0.97] \\
\end{align*}
\]

\[ \bar{R}^2 = 0.875, \text{ PF: } F_{6, 48} = 1.566 \text{ (0.178)}, \ s = 0.257, \ DW = 2.32 \]
Appendix 3.2. Estimated Response Surfaces for VAR(2) Models

Let \( a_{k,ij} = n (\hat{\alpha}_{k,ij} - \alpha_{k,ij}) \) (\( k = 1, 2 \)), then the response surfaces can be expressed as

\[
a_{1,11} = 0.09 - 0.58 \alpha_{1,11} - 0.29 \alpha_{1,11}^2 - 0.65 \alpha_{1,22} - 0.09 \alpha_{1,22} \rho - 0.25 \rho + 0.68 \rho^2 \\
(0.15) (0.07) (0.12) (0.05) (0.08) (0.10) \\
= [0.10] [0.10] [0.16] [0.07] [0.10]
\]

\( \hat{R}^2 = 0.707 \), PF: \( F_{12,83} = 0.039 \) (1.000), \( s = 0.634 \), DW = 1.88

\[
a_{1,22} = -0.13 - 0.87 \alpha_{1,11} - 0.27 \alpha_{1,11}^2 - 0.65 \alpha_{1,22} - 0.11 \alpha_{1,22} \rho - 0.25 \rho + 0.06 \rho^2 \\
(0.15) (0.07) (0.12) (0.05) (0.08) (0.11) \\
= [0.09] [0.11] [0.17] [0.08] [0.11]
\]

\( \hat{R}^2 = 0.699 \), PF: \( F_{12,83} = 1.037 \) (0.424), \( s = 0.632 \), DW = 1.98

\[
a_{1,21} = 0.05 + 0.10 \alpha_{1,12} + 0.02 \alpha_{1,12}^2 - 0.08 \alpha_{1,12} \rho + 0.14 \alpha_{1,21} \\
(0.14) (0.06) (0.02) (0.07) (0.06) (0.04) \\
= [0.13] [0.05] [0.02] [0.06] [0.04]
\]

\( \hat{R}^2 = 0.111 \), PF: \( F_{12,81} = 0.306 \) (0.987), \( s = 0.554 \), DW = 1.43

\[
a_{1,12} = 0.30 + 0.06 \alpha_{1,12} - 0.04 \alpha_{1,12}^2 - 0.16 \alpha_{1,12} \rho - 0.13 \alpha_{1,21} \\
(0.17) (0.07) (0.02) (0.09) (0.07) (0.07) \\
= [0.15] [0.07] [0.01] [0.08] [0.07]
\]

\( \hat{R}^2 = 0.106 \), PF: \( F_{12,81} = 0.306 \) (0.987), \( s = 0.554 \), DW = 1.43
\[ R^2 = 0.299, \text{ PF: } F_{12, 81} = 0.166 \text{ (1.000), } s = 0.673, \text{ DW = 1.91} \]

\[ a_{2,11} = 0.06 + 0.47 \alpha_{2,11}^2 - 1.51 \alpha_{2,22}^2 - 0.62 \alpha_{2,11} \rho - 0.43 \alpha_{2,22} \rho \]
\[
\begin{array}{cccc}
(0.22) & (0.09) & (0.21) & (0.21) & (0.27) & (0.21) & (0.33) \\
(0.19) & (0.11) & (0.25) & (0.21) & & & \\
\end{array}
\]

\[ R^2 = 0.524, \text{ PF: } F_{12, 85} = 2.166 \text{ (0.021), } s = 1.282, \text{ DW = 1.87} \]

\[ a_{2,22} = -0.09 + 0.47 \alpha_{2,11}^2 - 0.59 \alpha_{2,22}^2 + 0.31 \alpha_{2,11} \rho - 1.20 \alpha_{2,22} \rho \]
\[
\begin{array}{cccc}
(0.29) & (0.13) & (0.28) & (0.27) & (0.36) & (0.27) & (0.44) \\
(0.28) & (0.13) & (0.35) & (0.27) & & & \\
\end{array}
\]

\[ R^2 = 0.219, \text{ PF: } F_{12, 85} = 0.929 \text{ (0.522), } s = 1.679, \text{ DW = 2.16} \]

\[ a_{2,21} = 0.58 - 1.19 \alpha_{2,12} + 0.60 \alpha_{2,12}^2 - 0.51 \alpha_{2,21} - 0.40 \alpha_{2,21}^2 \]
\[
\begin{array}{cccc}
(0.33) & (0.27) & (0.29) & (0.16) & (0.13) & (0.30) & (0.18) & (0.11) \\
(0.24) & (0.30) & & & & & & \\
(0.41) & (0.86) & (0.33) & (0.63) & & & & \\
\end{array}
\]

\[ R^2 = 0.284, \text{ PF: } F_{12, 83} = 1.378 \text{ (0.193), } s = 1.307, \text{ DW = 2.31} \]

\[ a_{2,12} = 0.05 - 0.08 \alpha_{2,12} + 0.10 \alpha_{2,12}^2 + 0.02 \alpha_{2,21} + 0.09 \alpha_{2,21}^2 \]
\[
\begin{array}{cccc}
(0.46) & (0.38) & (0.39) & (0.23) & (0.17) & (0.37) & (0.19) & (0.13) \\
(0.24) & (0.33) & (0.39) & (0.23) & & & & \\
(0.48) & (0.78) & & & & & & \\
\end{array}
\]

\[ R^2 = 0.077, \text{ PF: } F_{12, 83} = 0.023 \text{ (1.000), } s = 1.799, \text{ DW = 2.16} \]
### Table 3.1. Design of VAR(1) models

<table>
<thead>
<tr>
<th>Model</th>
<th>$A_1$</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1a</td>
<td>$\begin{bmatrix} 0.5 &amp; 0.3 \ -0.6 &amp; 1.35 \end{bmatrix}$</td>
<td>1.05, 1.11</td>
</tr>
<tr>
<td>M1b</td>
<td>$\begin{bmatrix} 1 &amp; -1.5 \ 0.75 &amp; -1.25 \end{bmatrix}$</td>
<td>-2, 4</td>
</tr>
<tr>
<td>M1c</td>
<td>$\begin{bmatrix} -0.5 &amp; -0.5 \ 1 &amp; 0.3 \end{bmatrix}$</td>
<td>-0.28 ± 1.66i</td>
</tr>
<tr>
<td>M1d</td>
<td>$\begin{bmatrix} 0.7 &amp; 0.4 \ 0.4 &amp; -0.4 \end{bmatrix}$</td>
<td>1.2, -1.88</td>
</tr>
<tr>
<td>M1e</td>
<td>$\begin{bmatrix} 0.8 &amp; 0.4 \ 0.4 &amp; -0.5 \end{bmatrix}$</td>
<td>1.09, -1.63</td>
</tr>
<tr>
<td>M1f</td>
<td>$\begin{bmatrix} 0.95 &amp; 0.4 \ -0.4 &amp; 0.65 \end{bmatrix}$</td>
<td>1.02 ± 0.47i</td>
</tr>
<tr>
<td>M1g</td>
<td>$\begin{bmatrix} 0.85 &amp; 0.4 \ -0.4 &amp; 0.55 \end{bmatrix}$</td>
<td>1.11 ± 0.59i</td>
</tr>
<tr>
<td>M1h</td>
<td>$\begin{bmatrix} 0.95 &amp; 0 \ 0 &amp; 0.9 \end{bmatrix}$</td>
<td>1.05, 1.11</td>
</tr>
<tr>
<td>M1i</td>
<td>$\begin{bmatrix} 0.9 &amp; -0.1 \ 0.3 &amp; 0.55 \end{bmatrix}$</td>
<td>1.33, 1.42</td>
</tr>
<tr>
<td>M1j</td>
<td>$\begin{bmatrix} -1.2 &amp; -0.8 \ 0.8 &amp; 0.1 \end{bmatrix}$</td>
<td>1.05 ± 0.89i</td>
</tr>
<tr>
<td>M1k</td>
<td>$\begin{bmatrix} -0.9 &amp; 1 \ -0.8 &amp; 0.9 \end{bmatrix}$</td>
<td>± 10</td>
</tr>
</tbody>
</table>
Table 3.2. Design of VAR(2) models

<table>
<thead>
<tr>
<th>Model</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2a</td>
<td>$\begin{bmatrix} 0.8 &amp; -2 \ 0.9 &amp; 0.5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.4 &amp; 4/3 \ 0 &amp; 0.5 \end{bmatrix}$</td>
<td>$0.5 \pm 1.321, 1 \pm 1.21$</td>
</tr>
<tr>
<td>M2b</td>
<td>$\begin{bmatrix} 1.2 &amp; 0.6 \ -1.55 &amp; -0.4 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -0.2 &amp; 0 \ 0.8 &amp; 0.5 \end{bmatrix}$</td>
<td>$1 \pm 1.221, \pm 2$</td>
</tr>
<tr>
<td>M2c</td>
<td>$\begin{bmatrix} -2 &amp; -3.67 \ 3 &amp; 3.65 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; -0.72 \ -2.19 &amp; 0.8 \end{bmatrix}$</td>
<td>$1.26 \pm 0.391, 2.55 \pm 4.241$</td>
</tr>
<tr>
<td>M2d</td>
<td>$\begin{bmatrix} 2 &amp; 3.031 \ -2.424 &amp; -3 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0 &amp; 0.075 \ 1 &amp; 1.3 \end{bmatrix}$</td>
<td>$-1.91, 1.99, -1.70 \pm 0.761$</td>
</tr>
<tr>
<td>M2e</td>
<td>$\begin{bmatrix} 1.75 &amp; 1 \ 1 &amp; 1.85 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -1.765 &amp; -0.95 \ -0.9 &amp; -0.855 \end{bmatrix}$</td>
<td>$1.05, 1.11, 1.11, 1.17$</td>
</tr>
<tr>
<td>M2f</td>
<td>$\begin{bmatrix} 0.9 &amp; -1.2 \ 0.4 &amp; -0.5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.4 &amp; -0.3 \ 0 &amp; 0.2 \end{bmatrix}$</td>
<td>$-2.5, -2, 2, 1.25$</td>
</tr>
<tr>
<td>M2g</td>
<td>$\begin{bmatrix} 1.55 &amp; -0.3 \ 0 &amp; 0.4 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -0.57 &amp; 0.1 \ 0 &amp; 0.45 \end{bmatrix}$</td>
<td>$-2, 1.05, 1.11, 1.66$</td>
</tr>
<tr>
<td>M2h</td>
<td>$\begin{bmatrix} -0.928 &amp; 0.573 \ -2.872 &amp; 1.928 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.439 &amp; -0.520 \ 0.992 &amp; -1.142 \end{bmatrix}$</td>
<td>$0.98 \pm 1.731, 3.02, 5.74$</td>
</tr>
<tr>
<td>M2i</td>
<td>$\begin{bmatrix} -0.55 &amp; -2.993 \ 0.119 &amp; 1.002 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -0.2 &amp; 0.8 \ -0.1 &amp; -0.55 \end{bmatrix}$</td>
<td>$-0.38 \pm 1.341, 1.15 \pm 1.191$</td>
</tr>
<tr>
<td>M2j</td>
<td>$\begin{bmatrix} 4.75 &amp; 3.9 \ -4.95 &amp; -4 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -1.275 &amp; -1.425 \ 1.425 &amp; 1.625 \end{bmatrix}$</td>
<td>$-2, 4, 1.51 \pm 0.851$</td>
</tr>
<tr>
<td>M2k</td>
<td>$\begin{bmatrix} 2.15 &amp; 1 \ -1.26 &amp; 0.25 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -0.85 &amp; -0.61 \ 0.76 &amp; 0.31 \end{bmatrix}$</td>
<td>$1.24 \pm 0.641, 1.44 \pm 0.681$</td>
</tr>
<tr>
<td>M2l</td>
<td>$\begin{bmatrix} -1.1 &amp; -2.08 \ 1.25 &amp; 2.05 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.938 &amp; 0.85 \ -0.613 &amp; -0.615 \end{bmatrix}$</td>
<td>$1.34 \pm 0.651, -2, 4$</td>
</tr>
<tr>
<td>M2m</td>
<td>$\begin{bmatrix} -1.2 &amp; -1.84 \ 1 &amp; 1.55 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.438 &amp; 0.4 \ -0.288 &amp; -0.295 \end{bmatrix}$</td>
<td>$2.68 \pm 1.321, -2, 3.98$</td>
</tr>
<tr>
<td>M2n</td>
<td>$\begin{bmatrix} 1 &amp; -1.33 \ 2.5 &amp; -1.1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1.2 &amp; 0.73 \ 1.6 &amp; 1 \end{bmatrix}$</td>
<td>$-1.49 \pm 1.571, 1.76 \pm 1.871$</td>
</tr>
<tr>
<td>M2o</td>
<td>$\begin{bmatrix} -0.2 &amp; 2 \ 0.5 &amp; -1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 0.48 &amp; 1.6 \ -0.3 &amp; -1.5 \end{bmatrix}$</td>
<td>$-1.25, 1.67, 1 \pm 1$</td>
</tr>
<tr>
<td>M2p</td>
<td>$\begin{bmatrix} -0.6 &amp; -0.5 \ -0.8 &amp; -1.05 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -0.08 &amp; -0.2 \ -0.16 &amp; -0.6 \end{bmatrix}$</td>
<td>$-5, -4, -2.5, -1.25$</td>
</tr>
<tr>
<td>M2q</td>
<td>$\begin{bmatrix} 0.4 &amp; -0.02 \ 4 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} -2 &amp; -0.513 \ 1.245 &amp; 0 \end{bmatrix}$</td>
<td>$1 \pm 0.51, 0.5 \pm 1$</td>
</tr>
</tbody>
</table>
Table 3.3. The results of the PF tests.

<table>
<thead>
<tr>
<th>$\Sigma_u$</th>
<th>M1c</th>
<th>M1b</th>
<th>M1k</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$</td>
<td>0.126 (.944)</td>
<td>0.579 (.630)</td>
<td>0.424 (.737)</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.623 (.599)</td>
<td>0.081 (.970)</td>
<td>0.221 (.882)</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.034 (.991)</td>
<td>0.367 (.777)</td>
<td>0.151 (.929)</td>
</tr>
<tr>
<td>$a_{21}$</td>
<td>0.171 (.915)</td>
<td>0.110 (.954)</td>
<td>0.718 (.546)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Sigma_u^2$</th>
<th>M1c</th>
<th>M1b</th>
<th>M1k</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{11}$</td>
<td>0.321 (.810)</td>
<td>0.143 (.934)</td>
<td>0.243 (.867)</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>1.211 (.315)</td>
<td>0.924 (.436)</td>
<td>1.387 (.257)</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>0.021 (.996)</td>
<td>0.772 (.515)</td>
<td>1.272 (.295)</td>
</tr>
<tr>
<td>$a_{21}$</td>
<td>0.432 (.731)</td>
<td>0.945 (.426)</td>
<td>0.386 (.764)</td>
</tr>
</tbody>
</table>

Note. $a_{1j} = n (\hat{\alpha}_{1j} - \alpha_{1j})$

The entries are F values with p-values in the parentheses.

F statistics for $a_{11}$ and $a_{22}$ are distributed as $F_{3,52}$; and those for $a_{12}$ and $a_{21}$ are $F_{3,48}$.
Appendix 3.4. Figures

Figure 3.1. Biases of VAR(1) Models: (1,1) Element of A1 Matrix

Figure 3.2. Biases of VAR(2) Models: (1,1) Element of A2 Matrix
Figure 3.3. Sensitivity of Biases to the Value of Alpha

Figure 3.4. Sensitivity of Biases to the Value of Rho
Figure 3.5. Response Surface for alpha11: alpha22=0

Figure 3.6. Response Surface for alpha21: alpha12=0
Figure 3.7. Response Surface for alpha12: alpha21=0

Figure 3.8. Response Surface for alpha22: alpha11=0
Figure 3.9. Biases with Known and Unknown Intercepts:
VAR(1) Case

Figure 3.10. Biases with Known and Unknown Intercepts:
VAR(2) Case
Figure 3.11. Biases under Various Types of Innovations
Chapter 4. VAR Forecasting and Bootstrapping

4.1 Introduction

Forecasting is a natural application of the VAR model as it specifies the relationship between present and past realisations of the variables in the system. There are many studies which have used the VAR as a means of forecasting; see Sims (1982), Litterman (1986b), Fackler & Krieger (1986), McNees (1986) and Liu et al. (1994) among others. VAR forecasts are generated with estimated parameters, which bring additional sampling variability to forecasting. This renders knowledge on small sample properties of forecasts incomplete - such as the bias, variability and normality of forecast error distribution - even with an innovation process whose stochastic properties are perfectly known.

Unconditional forecasts generated from estimated univariate AR or VAR models can be shown to be unbiased under weak conditions (see, Dufour, 1984 & 1985). However, VAR forecasts are generated conditional on past history and are in general biased, as pointed out by Dufour (1985; p399). The bias of conditional forecasts in the univariate case is investigated by Maekawa (1987). He considered the univariate AR(1) model of the form (3.1), generated with a Gaussian error process, and provided an expression for forecast bias to O(n^{-1}) when the LS estimator is used for parameter estimation. The expression for forecast bias is given as

\begin{equation}
E(\hat{y}_{n+h}|y_n) = y_n \frac{1}{n} \left[ h\omega_n \frac{1}{n} \sigma^2 (\tau^2 - 3) + 0.5h(h-1)(1-\omega_n^2)\alpha^{h-2} \right],
\end{equation}
where \( h \) represents the forecasting period and \( \gamma^2 = \sigma_{u}^2/(1 - \alpha^2) \) with \( \sigma_{u}^2 \) and \( \alpha \) respectively being the variance of the error process and AR(1) coefficient. It can be seen from (4.1) that forecast biases generated from univariate AR(1) models are determined by the values of \( \alpha, h, \sigma_{u}^2 \) and the final observation on which the forecasts are conditional. Detailed numerical evaluation of (4.1) will be carried out later in this chapter. In the VAR case, the model structure indicated by VAR coefficient matrices (such as near non-stationarity and correlations among VAR components) may affect the magnitudes of small sample forecasts biases, analogous to \( \alpha \) values in the univariate AR(1) case. One of the major tasks of this chapter is to investigate small sample properties of VAR forecast biases and their determinants, paying attention to the conditionality and the VAR model structure. The conditionality and its effect on forecast biases are explicitly considered by employing the backward VAR representation (2.9).

The variability of forecasts in small samples is estimated in practice by asymptotic formulae based on large sample results, a brief summary of which is given below. For the optimal forecast \( \hat{y}_n(h) \) generated as in (2.7), the forecast error can be written as

\[
y_{n+h} - \hat{y}_n(h) = [y_{n+h} - y_n(h)] + [y_n(h) - \hat{y}_n(h)]
\]

\[
= \sum_{i=0}^{h-1} \phi_i u_{n+h-i} + [y_n(h) - \hat{y}_n(h)],
\]

where \( \phi_i \) are the coefficients of the MA(\( \omega \)) representation of (2.1). The variability of forecasts can be estimated by the MSE matrix, which is of the form
\[ \Sigma^*_{y}(h) = \text{MSE}(\hat{y}_n(h)) = E \{ y_{n+h} - \hat{y}_n(h) \} \{ y_{n+h} - \hat{y}_n(h) \}' \]

\[ = \Sigma^*_{y}(h) + E \{ y_n(h) - \hat{y}_n(h) \} \{ y_n(h) - \hat{y}_n(h) \}', \]

with \( \Sigma^*_{y}(h) = \sum_{i=0}^{h-1} \phi_i \Sigma \phi_i' \). The term with the expectation operator in the above expression is unknown in small samples. However, it can be approximated by resorting to the following limiting distribution given in Lütkepohl (1991; p86-7). That is, for a Gaussian process \( y_t \), conditional on past observations,

\[ n^{1/2} \{ \hat{y}_n(h) - y_n(h) \} \xrightarrow{d} N(0, \Omega(h)), \]

where \( \Omega(h) = E \left[ \frac{\partial y_n(h)}{\partial \beta'} \left( \Delta^{-1} \otimes \Sigma_u \right) \frac{\partial y_n(h)'}{\partial \beta} \right] \)

\[ = \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \text{tr} \{ (A')^{h-1-i} \Delta^{-1} A^{h-1-j} \Delta \} \phi_i \Sigma_u \phi_j', \]

with \( \beta = \text{vec}(\Lambda) \); and \( \Lambda \) and \( \Delta \) as defined in Chapter 2 (see Theorem 2.1). Hence, the following asymptotic approximation can be made to the MSE matrix to have

\[ \Sigma^*_{y}(h) = \Sigma^*_{y}(h) + 1/n \tilde{\Omega}(h). \]

Samaranayake & Hasza (1988) provided a formal proof that the above asymptotic approximation for the MSE matrix is valid. The estimated MSE matrix can be obtained by replacing unknowns with estimators to give

\[ \hat{\Sigma}^*_{y}(h) = \hat{\Sigma}^*_{y}(h) + 1/n \tilde{\Omega}(h). \]
The estimated asymptotic standard errors of forecasts are obtained by taking the square root of the diagonal elements of (4.3).

Note that the above asymptotic standard errors are derived under the assumption that the underlying VAR process is generated with normally distributed innovations. However, if the normality assumption is untenable, which is often the case in practice, the asymptotic standard errors may provide misleading assessment of future uncertainty. It is, therefore, of interest to examine whether the asymptotic standard errors are robust to a wide range of non-normal innovations. Moreover, if the asymptotic standard errors approximate the true variability poorly in small samples, statistical inferences on future uncertainty based upon the asymptotic standard errors can be misleading. This motivates us to examine reliability of the asymptotic standard errors in approximating the true variability in small samples.

The normality of VAR forecast error distributions in small samples is also under scrutiny in this chapter. Asymptotic approximations based on the limit distribution (4.2) assume forecast error distributions to be normally distributed. However, when the assumption of normality is not tenable, statistical inferences based on the normality assumption may be misleading. For example, a prediction interval constructed with the normality assumption when the underlying forecast error distribution is non-normal may lead to misleading assessments of future uncertainty. In the zero-mean univariate AR(1) case, Phillips (1979) investigated small sample properties of the forecast error distribution when the parameter is estimated by the LS method. He found that forecast error distributions generated with estimated univariate AR(1) models are in general skewed, and the direction
of skewness depends on the sign of the final observation on which forecasts are conditional. That is, for a positive AR(1) parameter, a positive (negative) final observation generates a negatively (positively) skewed forecast error distribution. The extent of skewness is determined by the strength of statistical dependence between the final observation and parameter estimator. If the final observation is an outlier, the skewness of forecast error distribution is less marked. Since VAR forecasts are also generated conditional on past observations, it is expected that final observations also play an important role in determining the shape of VAR forecast error distributions. The normality of VAR forecast error distribution will be examined in this chapter, paying special attention to the effect of final observations.

Since analytical treatments of the small sample issues discussed above would be intractable or may even be impossible, the bootstrap method is adopted to examine small sample properties. Its details in the VAR forecasting context are given in (F-i) to (F-iii) of Chapter 2; Theorem 2.4 proves that this bootstrap procedure is asymptotically valid. This provides the basis of using the bootstrap forecast error distribution \( \{ \hat{y}_n(h) - y_n^*(h) \}_{i=1}^b \) as a means of approximating the unknown sampling distribution of forecast error \( y_{n+h} - \hat{y}_n(h) \) in small samples. The bias and variability of forecasts can be estimated as follows:

\[
\text{Bias}[y_n^*(h)] = \hat{y}_n(h) - y_n^*(h) \quad \text{with} \quad \hat{y}_n^*(h) = 1/b \sum_{i=1}^b y_n^*(h)_i; \quad \text{and}
\]

\[
\hat{\Sigma}_y^*(h) = \text{MSE}[y_n^*(h)] = \sum_{i=1}^b \{ y_n^*(h)_i - \hat{y}_n^*(h) \} \{ y_n^*(h)_i - \hat{y}_n^*(h) \}' / (b-1).
\]

121
Bootstrap standard errors of forecast are obtained by taking the square
roots of the diagonal elements of the bootstrap MSE matrix. Bootstrap
forecast error distributions generated with bias-corrected parameter
estimators \( \{\hat{y}_n(h) - \gamma^*_n(h)^b\}_{i=1}^b \) are also introduced in Section 2.3.1 of
Chapter 2, from which we can obtain alternative standard errors of
forecasts to be called bias-corrected standard errors. Their major
attraction is that they measure the variability of forecasts generated with
parameter estimators whose small sample biases are removed.

The purpose of this chapter is to investigate the small sample issues
mentioned above by Monte Carlo simulations. Bootstrap estimators of VAR
forecast bias and variability are obtained from the bootstrap procedures
nested under each Monte Carlo iteration. Bivariate VAR models of orders 1
and 2 are considered as data generation mechanisms; six VAR(1) (M1a to M1f
and M1 in Table 3.1) and four VAR(2) (M2a, M2b, M2d and M2f from Table
3.2) are chosen. Note that model M1a and M2d are, respectively, a near
non-stationary and a model with highly correlated VAR components which are
found to generate large biases in VAR parameter estimators. Unless stated
otherwise, the following experimental design features are used for Monte
Carlo simulations for the VAR case: normal innovations with unit variances
and correlation coefficient 0.5; the vector of intercepts is set to \((0, 0)^T\)
but assumed to be unknown; the numbers of bootstrap and Monte Carlo
iterations are set respectively to 300 and 100; and the sample size used
are 30, 60 and 100. More general variance-covariance structures of
innovations are examined in preliminary simulations; but the results are
found to be qualitatively similar to those of the simpler structure. The
non-normal innovations used are as described in Chapter 3. Forecasting
periods considered are \( h = 1, 2, \ldots, 10 \) for numerical evaluations in the
univariate AR(1) case; and h = 1, 4 and 8 for Monte Carlo simulations in the VAR case.

The organisation of the chapter is as follows: Section 4.2 summarises the test statistics for skewness and kurtosis of a distribution and their small sample properties; Section 4.3 deals with the issue of resampling backward VAR residuals; Sections 4.4 to 4.7 present the results and discussions; and conclusions are drawn in Section 4.8.

4.2 Tests for skewness and kurtosis

The moments of a random variable often carry useful information on its distributional properties. For example, the first two moments, the mean and variance, summarize the information concerning the location and spread of the distribution. Information regarding symmetry and kurtosis of a distribution is carried by the third and fourth moments. Consider a standard normal variate $u$, then it can be shown that

$$E(u^3) = 0; \text{ and } E(u^4) = 3.$$ 

The third moment is a measure of symmetry, and a distribution with the positive (negative) third centred moment is skewed to the left (right) with 0 indicative of symmetry. The fourth centred moment measures kurtosis of distribution with a value larger (smaller) than three indicative of fatter (shorter) tails than the standard normal. Based on these properties, tests for skewness and kurtosis can be constructed. They can be used to test whether a distribution is asymmetric or flatter than the normal. The joint
test for skewness and kurtosis can also be used to test for the normality of a distribution, since the property of the third centred moment being equal to zero and the fourth moment to three is a distinct feature of the standard normal distribution.

The properties presented above are extended below to obtain the test statistics for skewness and kurtosis for the multivariate case. Let $z_t$ be a $K$-dimensional Gaussian process with $z_t \sim N(\mu, \Sigma)$ and $P$ be a lower triangular matrix such that $PP' = \Sigma$. Then,

$$w_t = (w_{1t}', \ldots, w_{Kt}')' = P^{-1}(z_t - \mu) \sim N(0, I_K),$$

(4.4) \hspace{1cm} E(w_{1t}^3, \ldots, w_{Kt}^3)' = 0_K; \text{ and}$

(4.5) \hspace{1cm} E(w_{1t}^4, \ldots, w_{Kt}^4)' = 3_K,$

where $0_K$ and $3_K$ are, respectively, $K$ element vectors of 0's and 3's. These properties can be used as a basis for obtaining test statistics to detect non-normality of a multivariate distribution.

Suppose a $K \times N$ data matrix $(z_{1t}, \ldots, z_{nt})$ with unknown mean and covariance matrix is available. Then it can be shown that

$$n^{1/2} \left[ \begin{array}{c} b_1 \\ b_z - 3_K \end{array} \right] \overset{d}{\rightarrow} N \left[ 0, \left( \begin{array}{cc} 6I_K & 0 \\ 0 & 24I_K \end{array} \right) \right],$$

where $b_i = (b_{1i}, \ldots, b_{K_i})'$ with $b_{ki} = 1/n \sum v_{ki}^3.$

124
\( b_2 = (b_{12}, \ldots, b_{k2})' \) with \( b_{k2} = 1/n \sum v_{kt}' \).

\( v_t = (v_{1t}', \ldots, v_{kt}')' = P_s^{-1}(z_t - \bar{z}); \quad \bar{z} = 1/n \sum z_t' \).

\( P_s P_s' = S \); and

\( S = 1/(n-1) \sum (z_t - \bar{z})(z_t - \bar{z})' \).

From the above results, the following statistics can be obtained:

\[ \lambda_1 = nb_1 b_1'/6 \chi^2(K); \]

\[ \lambda_2 = n(b_2 - 3\lambda_k)'(b_2 - 3\lambda_k)/24 \chi^2(K); \text{ and} \]

\[ \lambda_3 = \lambda_1 + \lambda_2 \chi^2(2K). \]

The statistic \( \lambda_1 \) tests for the null hypothesis of (4.4), \( \lambda_2 \) for the null of (4.5) and \( \lambda_3 \) for the null of (4.4) and (4.5) jointly. The ideas presented here are taken from Lütkepohl (1991; Sec. 4.5).

Since these tests are asymptotic tests whose small sample properties are unknown, it is important to examine their small sample reliability before they are implemented. As we shall see later, these tests are applied to bootstrap VAR forecast error distributions generated with the number of bootstrap iterations equal to 1000. That is, these test statistics are applied to a distribution with the number of data points equal to 1000. Although this number seems to be large enough to justify the use of asymptotic theories, a Monte Carlo experiment was conducted to evaluate the
performance of these tests when the sample size is 1000. Bivariate normal distributions with zero means and the following variance-covariance structures are considered: \( \text{vech}(\Sigma) = (1, 0.5, 1)' \); \( (1, -1, 4)' \); and \( (1, 0.7, 2)' \). That is, variances of each component are unequal but correlations are fixed at 0.5 in absolute terms. Non-normal distributions considered are as mentioned before; their variance-covariance structures are set to those above (see Chapter 3 for details of assigning a variance-covariance structure to these non-normal distributions). The empirical size and power properties are evaluated with significance levels of 5% and 10%. The number of Monte Carlo iterations is set to 1000.

The rejection frequencies are tabulated in Table 4.1. The entries for the normal distributions represent the empirical sizes and those for the others represent the empirical powers. The results are reported only for the case with \( \text{vech}(\Sigma) = (1, 0.7, 2)' \); the results obtained with the other \( \Sigma \) matrices were similar. It can be seen that the tests exhibit satisfactory performance in terms of size and power properties: empirical sizes are quite close to the nominal ones and powers are fairly high. Especially, these tests perform exceptionally well in detecting asymmetry of the distribution, judging from the empirical powers reported for the exponential and chi-squared distributions. Hence, these results suggest that the above tests statistics are powerful to detect departures from skewness, kurtosis and normality of the distribution, when the sample size is 1000.

4.3 Resampling backward VAR residuals
In Chapter 2, it is mentioned that innovation process \( \{v_t\} \) of the backward VAR model given by (2.9), corresponding to a forward model with non-normal innovation process, is uncorrelated but not independent. This means that, when bootstrap VAR forecasts are generated using the backward VAR representation (see (F-i) to (F-iii) in Chapter 2), i.i.d. resampling from the backward residuals \( \{\hat{v}_t\} \) is mis-specified as the underlying disturbances are not i.i.d. One way of overcoming this problem is to resample forward residuals, whose underlying disturbances are i.i.d., and obtain backward residuals by exploiting the relationship between the two. As a simple example, consider the VAR(1) case where forward and backward innovations, respectively from (2.1) and (2.9), are related as

\[
v_t = - H_1^{-1} \{ I_k - H_1^{-1} B \} \{ I_k - A B \}^{-1} u_{t+1} \\
= - H_1^{-1} \{ I_k + C_1 B + C_2 B^2 + \ldots \} u_{t+1},
\]

where \( B \) is the backshift operator, and \( C_1 = A_i - H_1^{-1} \) and \( C_k = A_i C_{k-1} \) for \( k \geq 2 \). Backward residuals can be obtained from forward residuals by using the above relationship. The infinite polynomial in \( B \) above can be truncated at some point \( j \) where the contribution of \( H_1 C_j \) becomes negligible. The above relationship can easily be extended to the higher order VAR case using the VAR(1) representations of forward and backward VAR(p) models. In the univariate case, Thombs & Schucany (1990) pointed out that the backward residuals obtained as above exhibit consistently greater variability than predicted by the model. Although not reported in detail, simulation results in the VAR case drew similar conclusion. Hence, despite the failure of backward innovations being independent when forward innovations are non-normal, direct i.i.d. resampling of backward residuals is adopted here,
as recommended by Thombs & Schucany (1990).

A simulation experiment is conducted to examine whether the direct i.i.d. resampling of backward residuals is acceptable. Bootstrap forecast biases and standard errors under normal and non-normal distributions are compared when sample sizes are large enough to justify the use of asymptotic theories. Under normal forward innovations where backward innovations are i.i.d., the bootstrap is well-specified and bootstrap quantities are shown to be well-behaved so that they converge to the true values as sample size increases (see Theorem 2.4). Under non-normal forward innovations where backward innovations can be dependent, there is no such assurance as the bootstrap is mis-specified. However, if bootstrap quantities under non-normal innovations are found to be fairly close to those under the normal with increasingly large sample sizes, the mis-specified bootstrap can be seen to be robust to non-i.i.d. innovations.

The forecast biases and bootstrap standard errors of forecasts are compared when sample sizes are 500 and 1000. The number of Monte Carlo iterations and bootstrap iterations respectively are set to 10 and 100 to lighten the burden of computation. Other design features are the same as before. It is found that the bootstrap quantities under non-normal innovations are very close to those under normality, when forecast biases and bootstrap standard errors of forecasts are compared. An example is given in Table 4.2 where the standard errors of the forecast for model M2a under normal and exponential innovations are given. It can be seen that the bootstrap standard errors under exponential innovations are quite close to those obtained under normal based on the asymptotic and bootstrap methods. Hence, it would seem that, when innovations are non-normal, bootstrapping
of VAR forecast error distribution by means of direct resampling of backward residuals is acceptable.

4.4 VAR forecast biases

We begin by numerically evaluating forecast biases (4.1) generated from a univariate AR(1) model. The parameter values used are

\[ n = 30, 60, 100; \alpha = 0.1 [0.2] 0.9; \text{ and } y_n = 1 [1] 5. \]

The value of \( \sigma_u^2 \) is fixed at unity without loss of generality. Only positive values of \( \alpha \) and \( y_n \) are considered because forecast biases associated with negative values are mirror images to those associated with positive ones, as apparent from (4.1).

Figure 4.1 plots forecast biases for each value of \( y_n \) when the \( \alpha \) and \( n \) values are, respectively, fixed at 0.5 and 60. It can be seen that large final observations generate substantial biases especially when forecast horizons are short. However, forecast biases die out quickly as \( h \) increases. The plots for other \( \alpha \) values appear similar except when \( \alpha = 0.9 \); the case of \( \alpha = 0.9 \) appears in Figure 4.2. In Figure 4.2, the cases of \( y_n = 0 \) and 1 are not plotted because the associated biases are negligible compared to those of the other cases. It can be seen that substantially large biases can be generated as the value of final observation increases. A feature different from that observed in Figure 4.1 is that biases continue to increase as \( h \) increases, instead of dying out quickly. Although not reported in detail, biases are calculated for \( h \) values larger than 10
when \( \alpha = 0.9 \). It is found that biases continue to increase until \( h \) reaches approximately 25 and start decreasing afterwards. Hence, when the model is near non-stationary, it appears that biases die out far more slowly than the models with the \( \alpha \) value well inside the stationarity region. Figure 4.3 plots biases for each \( \alpha \) value when the values of \( y_n \) and \( n \) are respectively fixed at 3 and 60. It is again evident that biases increase as \( \alpha \) approaches unity. When the model is near non-stationary with \( \alpha = 0.9 \), biases continue to increase as \( h \) increases, exhibiting no sign of reduction for the \( h \) values considered.

Two important factors which determine the magnitudes of univariate AR forecast biases are identified: the value of final observation and near non-stationarity. Since a VAR is a generalisation of the univariate AR(1), these two factors may also play important roles in determining the magnitude of VAR forecast biases. In this section, Monte Carlo simulations are conducted to examine the factors which determine the magnitudes of VAR forecast biases in small samples. Special attention is paid to the influence of final observations on the magnitude of forecasts biases. The models which generate large biases in VAR parameter estimators — near non-stationary models and models with highly correlated components as found in Chapter 3 — are also examined.

In the VAR(1) case, simulations are conducted with three vectors of final observations \((0, 0), (2, 2)\) and \((5, 5)\), respectively labelled FOV1 to FOV3. That is, final observations deviate further from the mean vector of the process with the same values in each component. Although these choices seem ad hoc, they will serve our purpose of examining the reaction of VAR forecast biases to the change of final observations. The case of different
values in each VAR components - e.g., (0, 5) and (5, 0) - was also examined, but not reported here as it provided similar results. We pay attention to three VAR(1) models M1d, M1f and M1a, as similar results have been obtained from the other models. It can be seen from Table 3.1 that, among these three, M1a has eigenvalues closest to the unit circle, while M1d furthest from the unit circle. The results are presented in Figures 4.4 to 4.6 for models M1d, M1f and M1a respectively, in terms of average absolute forecast biases (AAFB; average of absolute forecast biases arising from K VAR components). A common feature observed from these figures is that forecast biases increase as the values of final observations increase. This indicates that large VAR forecast biases can be generated when final observations are far larger than the mean of the process. It can also be observed from Figures 4.4 to 4.6 that forecast biases are peaked respectively at $h = 1, 4, 8$. That is, as models tends to non-stationarity, the speed of VAR forecast biases decreasing as $h$ increases gets slower, which agrees with the observation made in the univariate AR(1) case.

In the VAR(2) case, forecasts are conditional on the final two observations $y_{n-1}$ and $y_n$. The matrices of final observations, labelled FOM1 to FOM7, considered for the VAR(2) case are

$$
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix},
\begin{bmatrix}
2 & 2 \\
2 & 2
\end{bmatrix},
\begin{bmatrix}
5 & 5 \\
5 & 5
\end{bmatrix},
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix},
\begin{bmatrix}
2 & 2 \\
2 & 2
\end{bmatrix},
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix},
\text{and }
\begin{bmatrix}
5 & 5 \\
5 & 5
\end{bmatrix},
$$

where the first row entry represents $y_{n-1}$ and the second $y_n$. The first three matrices indicate the case where final observations deviate further from the mean with each entry and each component possessing identical values. The last four matrices of the above are introduced to evaluate the
forecast biases when only one of $y_{n-1}$ and $y_n$ deviates from the mean. Although, as in the VAR(1) case, the matrices of final observations are chosen in an ad hoc way, they will allow us to examine the effect of final observations on the magnitude of VAR forecast biases. As in the VAR(1) case, the case of different values in each VAR component exhibited similar results.

The results obtained in the VAR(2) case is qualitatively similar to the VAR(1) case. The general finding is that forecast biases increase as final observations deviate further from the mean and decrease as $h$ increases. The matrices of final observation with only one component being outliers are also found to generate biases as large as or sometimes even larger than the case where both components are outliers. Examples can be found in Figures 4.7 and 4.8, where AAFB for models M2a and M2b are plotted when sample size is 60. Note that, in Figure 4.7, the points associated with FOM1 and FOM2 are much smaller than the others and not included in the figure for clearer exposition. It is evident from both figures that biases are in general larger as final observations deviate further from the mean. The tendency that forecast biases decrease as $h$ increases is also evident in Figures 4.7 and 4.8.

When models which generate large biases in parameter estimators are considered, the magnitudes of forecast biases are found to be much larger than those of the others. Figure 4.9 plots the case of M2d which is a model with highly correlated VAR components. It can be seen from Figures 4.7 and 4.8 that biases, measured in terms of AAFS, are in the range of 0 to 0.2. However, if model M2d is considered as in Figure 4.9, biases range from 0 to 1.2. An additional near non-stationary model, M2q in Table 3.2, was also
experimented with, and the associated VAR forecasts are found to be biased as severely as those from model M2d. The results reported so far in this section are associated only with the case when n = 60. However, the results when n = 30 or 100 are found to be similar. It is also found, plausibly, that VAR forecast biases decrease as sample size increases.

The finding of this section suggests that forecast biases in VAR models depend heavily on the values of final observations on which forecasts are conditional. Forecast biases are seen to be negligible only when the values of final observations are identical or fairly close to zero (or the mean of the process), which agrees with the finding of Maekawa (1987) in the univariate AR(1) case. When forecasts are generated with the models which generate large biases in parameter estimators, one should also be prepared for substantially large biases.

4.5 Variability of VAR forecasts

In this section, small sample properties of asymptotic standard errors are examined and compared to those of their bootstrap counterparts. The final p observations are not fixed but free to vary in repeated sampling. In the next section, the effect of conditionality is examined by fixing final observations. Other experimental design features are as before. Asymptotic and bootstrap standard errors are also compared to the bias-corrected standard errors. The bias-corrected standard errors are calculated only under normal innovations because they are found to be insensitive to departures from normality in preliminary simulations.
It is observed, for all cases considered, that asymptotic standard errors are fairly close to bias-corrected standard errors, and these standard errors are larger than their bootstrap counterparts. This suggests the possibility that bootstrap standard errors under-estimate the true variability in small samples. The differences among the three measures decrease as sample size increases, as might be anticipated. Of course, the possibility of asymptotic standard errors over-estimating the true variability cannot not be ruled out. However, based on the findings to be reported in Chapter 5, we maintain that bootstrap standard errors under-estimate future uncertainty. In Chapter 5, small sample properties of asymptotic and bootstrap prediction regions are compared. It is found that asymptotic prediction regions outperform their bootstrap counterparts with a tendency of the latter under-estimating the true variability.

An example can be seen in Table 4.4 where the results from model M2a under normal innovations are reported. When models which generate large small sample biases in parameter estimators are considered, the asymptotic standard errors of forecasts are still found to be reasonably close to the bias-corrected standard errors. The bootstrap standard errors are again found to be smaller than their asymptotic counterparts exhibiting the tendency of under-estimation. The results for model M2d, a model with highly correlated VAR components, are reported in Table 4.5 under normal innovations. Hence, it seems that the asymptotic standard errors perform well even for those models which generate large biases in parameter estimators. It can also be observed that the models with large biases in parameter estimators exhibit variability much larger than models with moderate biases (compare Tables 4.4 and 4.5). This tendency becomes more obvious as the forecasting horizon increases.
Small sample performances of asymptotic standard errors of VAR forecasts under non-normal innovations are evaluated. It is found that asymptotic standard errors perform reasonably well and that features observed under normal innovations, such as the possibility of under-estimation by bootstrap standard errors, are also present in the case of non-normal innovations. An example is given in Table 4.6 where asymptotic and bootstrap forecast standard errors are compared for model M2b. Similar results are also obtained for all other cases of non-normal innovations. Hence, the results obtained in this section suggest that asymptotic standard errors provide a reliable and robust assessment of future uncertainty for a wide range of non-normal innovations which are fat-tailed or asymmetric.

4.6 Effect of final observations on forecast variability

In this section, variability of VAR forecasts is evaluated by explicitly incorporating the property of conditionality. In Section 4.4, we have seen that VAR forecast biases are sensitive to the final observations on which forecasts are conditional. A natural question is how the variability of VAR forecasts reacts to the values of final observation. As in Section 4.4, the final p observations of VAR processes are fixed in repeated sampling. Maekawa (1987) provided an expression for the conditional MSE to $O(n^{-1})$ for the univariate AR(1) case, which can be written as [see, for the univariate AR(p) case, Fuller & Hasza, 1981]

$$
E[(\hat{y}_{n+h} - y_{n+h})^2 | y_n] \sim (1-\alpha^2)\sigma_u^2 / (1-\alpha^2) + y_n^2 / n \sigma^2 \alpha^{2h-1}(1-\alpha^2),
$$

135
Numerical evaluations of (4.6) are performed using the same parameter values as those used in the evaluation of (4.1) in Section 4.4. Figure 4.10 plots MSE of forecasts for the univariate AR(1) model with $\alpha = 0.5$ for each value of the final observations considered when sample size is 60. The plots with the other $\alpha$ values appear similar except when $\alpha = 0.9$. It can be seen that MSE's are sensitive to changes in the final observation when forecasting horizons are short, but the variations of MSE's disappear quickly as the forecasting horizon increases. This figure can be compared to Figure 4.1, in which forecast biases exhibit similar patterns. The case when $\alpha = 0.9$ is plotted in Figure 4.11. The variation of MSE's now increase as the forecasting horizon increases. Note that similar patterns were observed in Figure 4.2 for forecast biases.

To investigate the effect of final observations on the variability of VAR forecasts, Monte Carlo simulations are conducted with experimental design features as before. For simplicity, only bootstrap standard errors of forecasts are reported, although asymptotic standard errors provide similar results. The results obtained are qualitatively similar to those obtained in the univariate AR(1) case. Examples can be found in Figures 4.12 and 4.13, where bootstrap standard errors are reported respectively for model M1a and M2b. The variability is measured in terms of the average forecast standard errors (AFSE), which is the average of forecast standard errors arising from K VAR components. As in the univariate case, standard errors are sensitive to changes of final observations when $h = 1$. As $h$ increases, however, variations in standard error become less sensitive. These features are evident for both cases and they can be compared to those observed earlier in Figures 4.4 and 4.5, where VAR forecast biases exhibit
similar patterns. The case of VAR models which generate large biases in parameter estimators are of interest. Figure 4.14 reports bootstrap standard errors for model M2d, which is a model with highly correlated VAR components. It is observed that the variations of forecasts standard errors increase as final observations deviate further from the mean vector. This figure can be compared to Figure 4.11, where variability of forecasts in the univariate AR(1) case exhibit similar patterns. A similar observation can be made for the near non-stationary model M1a. Hence, in the VAR case, we observe that substantially large variability of forecasts can be generated when models with large biases in parameter estimators are considered or final observations are outliers.

4.7 Normality of VAR forecast error distributions

In this section, the tests for skewness and kurtosis, introduced in Section 4.2, are applied to bootstrap VAR forecast error distributions in order to examine the normality of the underlying distributions in small samples. Only normal innovations with variance-covariance structures \( \text{vech}(\Sigma_u) = (1, 0.5, 1)'; (1, 0.7, 2)'; (1, -1, 4)'; \) and \( (1, -1, 4)' \) are used. The models chosen in Section 4.4 are again used here with all other design features as before, except for the number of the bootstrap iterations which is now set to 1000 instead of 300. Although not reported here in detail, the simulation results become stable when the number of bootstrap iterations is as large as 1000. This also agrees with the number of bootstrap iterations generally required in constructing bootstrap confidence intervals (see, for example, Section 9 of Efron, 1987). To construct a sensible bootstrap confidence interval, any skewness and kurtosis in the underlying
distribution should be well reflected in the bootstrap distributions. Since the normality tests examine the skewness and kurtosis of a distribution, the decision to use 1000 bootstrap iterations seems reasonable.

Although all three statistics introduced in Section 4.2 are employed, only the results associated with the joint test are reported here. It is found that the normality assumption cannot be justified for VAR forecast error distributions even when the sample size is 100. The departure from normality is particularly severe when the forecasting horizon is 1. It is also observed that the frequency of rejecting normality decreases as sample size and forecasting horizon increase. Examples are given in Table 4.7 where frequencies of rejecting normality of forecast error distributions are given for models M1c and M2b when $\text{vech}(\Sigma_u) = (1, 0.5, 1)'$. It can be seen that the frequency of rejecting normality is substantially high, although the frequency decreases as the sample size or forecasting horizon increase. When models with large biases in parameter estimators are considered, the frequencies of rejecting normality are substantially higher than when the models with moderate biases are considered.

Now the final $p$ observations of the VAR processes are fixed in repeated sampling to examine more explicitly the influence of conditionality on the normality of forecast error distribution. The vectors or matrices of final observations used are the same as those used previously. Other design features are as before. The finding suggests that final observations affect the forecast error distributions so that the frequency of rejecting normality increases as the final observations deviate further from the mean of the series. Examples are given for models M1c and M2b in Tables 4.8 and 4.9. It can be seen from Table 4.8 that the
frequency of rejecting normality increases as the final observations deviate further from the mean. For the VAR(2) model M2b, final observations with only one component being outliers can also exhibit a high frequency of rejecting normality. Although not reported in detail, these features are accentuated when models with large biases in parameter estimators are considered. Hence, it can be concluded that the normality of VAR forecast error distribution becomes less likely if final observations are further away from the mean of the series.

It is found in this section that normality of VAR forecast error distribution is not acceptable in small samples. The non-normality is found to be particularly severe when final observations are outliers or models with large biases in parameter estimators are used. There is a tendency, although not substantial, that the frequency of non-normality decreases as forecasting horizon increases.

In this chapter, bivariate VAR models of orders 1 and 2 were employed to investigate small sample properties associated with VAR forecasts. The small sample properties examined are bias, variability and normality of VAR forecasts. As in the univariate AR(1) case, biases are found to be large when the forecasting horizon is short, but diminish as the forecasting horizon increases. The speed of decline is slower when near non-stationary models are considered. Two factors are identified as important determinants of biases, variability and normality of VAR forecasts in small samples. The first is the model structures which generate large biases in VAR parameter estimators. The second is the values of final observations on which VAR forecasts are conditional. When VAR forecasts are generated with final observations some distance away from the mean, forecasts are found to generate large biases.

Three variability measures of VAR forecasts are compared, and asymptotic standard errors of forecasts, popularly employed in practice, are found to perform reliably in small samples under a wide range of innovation processes including the normal. The assumption of normality of VAR forecasts error distribution is found not to be tenable in small samples. This indicates that care should be taken when inferences on VAR forecasts are made based on the asymptotic method. An inferential procedure based on the bootstrap, to be investigated in the next chapter, may provide a useful alternative in small samples. An interesting feature found is that, when the forecasting horizon is shorter, VAR forecast biases are found to be larger; and normality is less likely. As the forecasting horizon increases, biases diminish; and the possibility of normality
increases. Now it is generally accepted that forecasts are more reliable when the forecasting horizon is shorter, however, our results in the VAR case show otherwise.

As discussed in Section 4.3, resampling backward residuals causes mis-specification when disturbances are non-normal. A simulation experiment with a limited design is used to examine the robustness of the mis-specified bootstrap. A more formal analytical treatment of this issue remains as a subject for future research.
Appendix 4.1. Tables.

Table 4.1. Size and Power Properties of Normality Testing (n=1000)

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Student's t</th>
<th>Expo</th>
<th>Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>5.6</td>
<td>82.3</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>4.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>5.1</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Student's t</th>
<th>Expo</th>
<th>Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>10.0</td>
<td>86.5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>10.0</td>
<td>86.5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>10.1</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note. Entries in the first panel are associated with 5% significance level and those in the second panel are 10% significance level.

Table 4.2. Mean Standard Errors of Forecasts for Model M2a

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>Normal (Boot)</th>
<th>Expo (Boot)</th>
<th>Normal (Asy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>1</td>
<td>1.0307</td>
<td>1.0166</td>
<td>1.0413</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.9505</td>
<td>2.2188</td>
<td>2.9339</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.4947</td>
<td>2.5892</td>
<td>3.4719</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>0.9786</td>
<td>0.9980</td>
<td>0.9770</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.8969</td>
<td>2.2649</td>
<td>2.8893</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.4961</td>
<td>2.5532</td>
<td>3.4549</td>
</tr>
</tbody>
</table>

Note. Boot and Asy respectively represent bootstrap and asymptotic quantities. The first entries are standard errors of forecasts for the first component of VAR system. The second entries are those for the second component.
Table 4.3. Bootstrap forecast Errors and standard errors from the Model M2b when innovations are exponential.

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>var1</th>
<th></th>
<th>var2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>fe</td>
<td>ste</td>
<td>fe</td>
<td>ste</td>
</tr>
<tr>
<td>-----</td>
<td>---</td>
<td>-------</td>
<td>---</td>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>0.060</td>
<td>1.017</td>
<td>-0.132</td>
<td>1.029</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>0.033</td>
<td>2.082</td>
<td>-0.005</td>
<td>2.182</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.025</td>
<td>2.157</td>
<td>-0.013</td>
<td>2.268</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.072</td>
<td>1.047</td>
<td>-0.045</td>
<td>1.023</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>-0.008</td>
<td>2.039</td>
<td>-0.028</td>
<td>2.166</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.016</td>
<td>2.106</td>
<td>0.020</td>
<td>2.225</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>0.025</td>
<td>1.021</td>
<td>-0.061</td>
<td>1.016</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.008</td>
<td>2.140</td>
<td>0.019</td>
<td>2.275</td>
</tr>
</tbody>
</table>

Note. fe: bootstrap forecast error
ste: bootstrap standard error of forecast
var1: the first component of the process
var2: the second component of the process innovation is exponential with k = 2
ν = (0, 0)′
Table 4.4. Forecast standard errors from Model M2a when innovations are normal

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>var1</th>
<th></th>
<th>var2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>asy</td>
<td>boot</td>
<td>bc</td>
<td>asy</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.052</td>
<td>1.064</td>
<td>1.010</td>
<td>1.066</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>3.238</td>
<td>2.618</td>
<td>3.085</td>
<td>2.528</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.955</td>
<td>2.918</td>
<td>3.897</td>
<td>2.904</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.032</td>
<td>1.028</td>
<td>1.002</td>
<td>1.032</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>3.141</td>
<td>2.721</td>
<td>3.064</td>
<td>2.458</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.770</td>
<td>3.120</td>
<td>3.744</td>
<td>2.773</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.025</td>
<td>1.022</td>
<td>1.008</td>
<td>1.022</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>3.072</td>
<td>2.810</td>
<td>3.038</td>
<td>2.410</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.701</td>
<td>3.215</td>
<td>3.647</td>
<td>2.728</td>
</tr>
</tbody>
</table>

Note. asy: asymptotic standard errors  
boot: bootstrap standard errors  
bc: standard errors generated with bias-correction  
var1: the first component of the process  
var2: the second component of the process
Table 4.5. Forecast standard errors from Model M2d when innovations are normal

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>var1</th>
<th>var2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>asy</td>
<td>boot</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.055</td>
<td>1.142</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.040</td>
<td>1.040</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>9.111</td>
<td>8.333</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.024</td>
<td>1.027</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>9.090</td>
<td>8.506</td>
</tr>
</tbody>
</table>

Note. asy: asymptotic standard errors
      boot: bootstrap standard errors
      bc: standard errors generated with bias-correction
      var1: the first component of the process
      var2: the second component of the process

145
Table 4.6. Bootstrap and asymptotic forecast standard errors from the Model M2b when the innovations are exponential

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>var1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>asy</td>
<td>boot</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.054</td>
<td>1.017</td>
<td>1.064</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>2.140</td>
<td>2.082</td>
<td>2.343</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.159</td>
<td>2.157</td>
<td>2.365</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.044</td>
<td>1.047</td>
<td>1.014</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>2.122</td>
<td>2.039</td>
<td>2.282</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.139</td>
<td>2.106</td>
<td>2.301</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.030</td>
<td>1.021</td>
<td>1.016</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>2.101</td>
<td>2.074</td>
<td>2.253</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.122</td>
<td>2.140</td>
<td>2.276</td>
</tr>
</tbody>
</table>

Note. asy: asymptotic standard errors
      boot: bootstrap standard errors
      var1: the first component of the process
      var2: the second component of the process
Table 4.7. Frequency of Rejecting Normality for the Distribution of Forecasts (Joint test for skewness and kurtosis using $\lambda_3$ with 5% significance level)

<table>
<thead>
<tr>
<th></th>
<th>Model M1c</th>
<th></th>
<th>Model M2b</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 30$</td>
<td>$n = 60$</td>
<td>$n = 100$</td>
<td>$n = 30$</td>
</tr>
<tr>
<td>1</td>
<td>99</td>
<td>96</td>
<td>89</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>83</td>
<td>77</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>8</td>
<td>86</td>
<td>80</td>
<td>70</td>
<td>74</td>
</tr>
</tbody>
</table>

Note. Entries are frequencies out of 100.
Table 4.8. Frequency of Rejecting Normality for the Distribution of Forecasts (Joint test for skewness and kurtosis using $\lambda_3$ with 5% significance level): Model Mlc.

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>FOV1</th>
<th>FOV2</th>
<th>FOV3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>100</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>54</td>
<td>62</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>53</td>
<td>55</td>
<td>97</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>95</td>
<td>90</td>
<td>98</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>33</td>
<td>36</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>30</td>
<td>34</td>
<td>98</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>88</td>
<td>86</td>
<td>97</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>29</td>
<td>27</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>26</td>
<td>27</td>
<td>97</td>
</tr>
</tbody>
</table>
Table 4.9. Frequency of Rejecting Normality for the Distribution of Forecasts (Joint test for skewness and kurtosis using $\lambda_3$ with 5% significance level): Model M2b.

<table>
<thead>
<tr>
<th>n</th>
<th>h</th>
<th>FOM1</th>
<th>FOM2</th>
<th>FOM3</th>
<th>FOM4</th>
<th>FOM5</th>
<th>FOM6</th>
<th>FOM7</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>67</td>
<td>70</td>
<td>97</td>
<td>69</td>
<td>66</td>
<td>71</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>61</td>
<td>69</td>
<td>97</td>
<td>59</td>
<td>58</td>
<td>73</td>
<td>97</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
<td>97</td>
<td>93</td>
<td>100</td>
<td>96</td>
<td>95</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>62</td>
<td>72</td>
<td>100</td>
<td>62</td>
<td>62</td>
<td>73</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>64</td>
<td>59</td>
<td>100</td>
<td>63</td>
<td>62</td>
<td>63</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>88</td>
<td>92</td>
<td>100</td>
<td>89</td>
<td>89</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>55</td>
<td>57</td>
<td>100</td>
<td>54</td>
<td>66</td>
<td>66</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>52</td>
<td>48</td>
<td>100</td>
<td>50</td>
<td>49</td>
<td>48</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. Entries are frequencies out of 100.
Appendix 4.2. Figures

Figure 4.1. Biases of Forecasts: Univariate AR(1); $n = 60$; alpha = 0.5.

Figure 4.2. Biases of Forecasts: Univariate AR(1); $n = 60$; alpha = 0.9.
Figure 4.3. Biases of Forecasts: Univariate AR(1); \( n = 60, \gamma_n = 3 \).

Figure 4.4. Biases of Forecasts: Model M1d; \( n = 60 \).
Figure 4.5. Biases of Forecasts: Model M1f; \( n = 60 \).

Figure 4.6. Biases of Forecasts: Model M1a; \( n = 60 \).
Figure 4.7. Biases of Forecasts: Model M2a; n = 60.

Figure 4.8. Biases of Forecasts: Model M2b; n=60.
Figure 4.9. Biases of Forecasts: Model M2d; n = 60.

Figure 4.10. MSE’s of Forecasts: Univariate AR(1); alpha = 0.5; n = 60.
Figure 4.11. MSE's of Forecasts: Univariate AR(1); alpha = 0.9; n = 60.

Figure 4.12. Standard Errors of Forecasts: Model M1c; n = 60
Figure 4.13. Standard Errors of Forecasts: Model M2b; n = 60.

Figure 4.14. Standard Errors of Forecasts: Model M2d; n=60.
Chapter 5. Bootstrapping Prediction Regions of VAR Models

5.1 Introduction

Prediction intervals, along with point forecasts, provide a useful tool in predicting the future course of economic variables. However, as found by Dalrymple (1987), point forecasts are more popularly adopted in practice mainly because they are simpler to generate. A drawback of point forecasts is that a single number as an estimate of the future event carries virtually no information as to how much uncertainty is associated. Prediction intervals, on the other hand, provide the upper and lower limits of forecasts with a prescribed probability content and evaluate future uncertainty in a more informative manner. That is, a future scenario indicated by prediction intervals can provide information on the extent of uncertainty associated as well as the direction of future values. In a survey paper on prediction intervals by Chatfield (1993), the importance and benefit of using prediction intervals (together with or instead of point forecasts) are discussed in detail in the context of general econometric and time series models.

In VAR forecasting in practice, the major focus has been generation and evaluation of point forecasts, while little attention has been paid to prediction intervals or regions (see, amongst others, Fackler & Krieger, 1986, Litterman, 1986b, McNees, 1986, and Liu et al., 1994). It is mainly because of the computational burden involved in constructing a VAR prediction region (denoted PR hereafter) which takes the form of an ellipse or ellipsoid. When VAR forecasting is conducted with two or three variables
in the system, graphical representation of a PR in the form of an ellipse or ellipsoid (called the prediction ellipse or ellipsoid, and denoted PE) is possible but can be quite demanding. With four or more variables, construction of a PE which involves all variables in the system requires sophisticated computational techniques. An easier way of constructing a PR is to use Bonferroni’s method (see Lütkepohl, 1991, pp34-5), which gives a cubical approximation to the PE to yield what is called the prediction cube (PC). From this point on, the term PR includes the PE and PC. However, as we shall see later in this chapter, the use of PC as an approximation to PE can lead to a serious over-estimation of future uncertainty. For this reason, the use of PR’s for VAR forecasting may be less appealing to practitioners.

Issues related to small sample properties of PR’s for VAR models are also of importance. The PR’s for VAR forecasts can be constructed using asymptotic formulae based upon the limit distribution of VAR forecasts as given in (4.2) of Chapter 4. The reliability of the asymptotic approximation based on (4.2) is, however, unknown and as too is that of the small sample performance of the PR’s for VAR. Moreover, the limit distribution (4.2) is obtained under the assumption that the VAR innovation process follows a normal distribution. When normality is not acceptable, a PR based on the asymptotic approximation may also be misleading. Hence, practitioners may choose not to employ a PR for VAR forecasts because they are unsure about its reliability in small samples as well as its robustness to non-normality.

The bootstrap provides a useful alternative to the asymptotic PR in small samples. In the univariate AR case, as mentioned in Chapter 1, Thombs
& Schucany (1990) and Masarotto (1990) considered the use of bootstrap prediction intervals (PI’s) as an alternative to their asymptotic counterparts. Small sample performances of the bootstrap and asymptotic PI’s were compared under normal and a wide range of non-normal innovations to conclude that the former provides a useful tool in assessing future uncertainty in small samples. Chatfield (1993) also recommended the use of the bootstrap PI’s as a general-purpose alternative, especially when asymptotic formulae are not available or there are doubts about model assumptions. However, no attempt has so far been made to consider PR’s based on the bootstrap as an alternative to their asymptotic counterparts in multivariate time series models such as VAR. In particular, it was found in Chapter 4 that, even with a normal innovation process, the assumption of normality of VAR forecast error distribution is not tenable in small samples. This result provides a basis for the idea that the bootstrap PR’s may be preferred to their asymptotic counterparts because they are free from the assumption of normality. Hence, it is also of particular interest to examine small sample performances of bootstrap PR’s in the context of VAR models. The basic procedures to obtain bootstrap VAR forecast distributions were introduced in Chapter 2 [procedures (F-i) to (F-iii)]. The method of constructing a PR from a bootstrap forecast distribution will be given later.

The purpose of this chapter is to address the two issues outlined above in relation to why PR’s are not popular for VAR forecasting in practice. Firstly, we evaluate small sample performances of asymptotic PR’s (APR) and examine the reliability of bootstrap PR’s (BPR) as an alternative to their asymptotic counterparts. We adopt measures such as empirical coverages and volumes, to be introduced later in this chapter, for
evaluations and comparisons. Secondly, the extent of over-estimation by PC is evaluated and the use of what may be called the improved PC (IPC) is proposed in an attempt to overcome the problem of over-estimation of future uncertainty. It is hoped that the findings here stimulate interest in using PR's in VAR modelling.

The organisation of this chapter is as follows: in Section 5.2, the methods of constructing asymptotic and bootstrap PR's for VAR models are given; Section 5.3 presents experimental designs for simulations; Section 5.4 examines small sample properties of asymptotic PR's along with their bootstrap counterparts; Section 5.5 proposes the use of IPC in an attempt to alleviate the problem of over-estimation by PC; and conclusions are drawn in Section 5.6.

5.2 Prediction Regions for VAR forecasts

This section provides a review of PR's for VAR forecasts. The APR's and their bootstrap counterparts are introduced. Under the assumption of a normally distributed innovation process, it was stated in Chapter 2 that

\[ y_{n+h} - y_n(h) \sim \mathcal{N}[0, \Sigma_y(h)]. \]

That is, VAR forecast error follows a K-dimensional multivariate normal with variance-covariance matrix \( \Sigma_y(h) \) as defined in Chapter 4. It follows that

\[(5.1) \quad Q[y_n(h)] = [y_{n+h} - y_n(h)]' \Sigma_y(h)^{-1} [y_{n+h} - y_n(h)] \sim \chi^2(K). \]
This result provides the basis of constructing a PR for VAR forecasts. Consider the truncation of (5.1) at a point, typically $\chi^2_{\alpha}(K)$ where $\chi^2_{\alpha}(K)$ denotes the 100(1-$\alpha$)th percentile of the $\chi^2(K)$ distribution with $0 \leq \alpha \leq 1$. That is,

$$(5.2) \quad PE[y_n(h), 1-\alpha] = \{y_{n+h} : y_{n+h} \text{ satisfies } Q[y_n(h)] \leq \chi^2_{\alpha}(K)\}.$$ 

The above expression defines the interior of an ellipsoid on the $K$-dimensional plane of forecasts. It has probability content $(1-\alpha)$ and is called a 100(1-$\alpha$)% PE. The probability $(1-\alpha)$ is called the nominal coverage, and it can formally be written as (see Stein, 1985)

$$(5.3) \quad C(PE[y_n(h), 1-\alpha]) = \text{Prob}(y_{n+h} : y_{n+h} \in PE[y_n(h), 1-\alpha]).$$ 

The shape of a PE, its orientation and relative lengths of its semi-axes, depends entirely on the structure of $\Sigma_y(h)$. If $\Sigma_y(h)$ is a diagonal matrix, VAR forecasts are linearly uncorrelated and the semi-axes of the ellipsoid are parallel to the axes of forecasts. If it is a diagonal matrix whose diagonal elements are equal, the PE is a spheroid. Unequal diagonal elements of $\Sigma_y(h)$ yield unequal lengths of semi-axes of the ellipsoid. If $\Sigma_y(h)$ is not diagonal, i.e., if components of VAR forecasts are linearly correlated, then semi-axes of the ellipsoid are not parallel to the axes of forecasts.

As mentioned earlier, construction of a PE can be quite demanding, especially when more than three variables are involved in the VAR system. An easier way is to give a cubical approximation to PE, which yields a PC.
To construct a $100(1-\alpha)\%$ PC, we use Bonferroni's method (for more details, see Lütkepohl, 1991, pp34-5). The idea is that if we choose $100(1-\alpha/K)\%$ PI's for each of the $K$ components of $y_t$, then the resulting region will have probability at least $(1-\alpha)$ of containing the $K$ variables jointly. This region is called the PC, although, more precisely, it is a rectangular prism. The PC for $K$ components with nominal coverage of at least $1-\alpha$ can be written as

\begin{equation}
\text{PC}(y_{n+h}, 1-\alpha) = \{ y_{n+h} : y_{n+h} \in \bigcup_{k=1}^{K} \{ y_{k,n} \pm z_{\tau} \sigma_k(h) \} \}.
\end{equation}

In (5.4), $y_{k,n}(h)$ is the $k$th component of $y_{n}(h)$, $z_{\tau}$ is the upper $100\tau\%$ percentile of the standard normal distribution with $\tau = 0.5(\alpha/K)$ and $\sigma_k(h)$ is the standard error of the forecast of the $k$th component of $y_t$, given by the square root of the $k$th diagonal element of $\Sigma(h)$. A graphical illustration of a PC, as an approximation to the PE, will be given later in Section 5.5.

In practice, VAR parameters are unknown and have to be estimated. As we have seen in the previous chapter, parameter estimation introduces additional uncertainty to VAR forecasting, causing departure of VAR forecast error distribution from normality in small samples. Hence, with parameter estimation, the exact result in (5.1) will fail to hold in small samples. However, based upon the limit distribution given in (4.2) in Chapter 4, asymptotic approximations can be made in small samples. The unknown parameters in (5.1) can be replaced by estimators, and it can be shown that

\[ Q(\hat{y}_{n+h}) = [y_{n+h} - \hat{y}_{n}(h)]' \hat{\Sigma}_y^{-1} [y_{n+h} - \hat{y}_{n}(h)] \sim \chi^2(K), \]
where the symbol \( \hat{\alpha} \) indicates "asymptotically distributed" and \( \hat{\Sigma}_y(h) \) is as defined in Chapter 4. The asymptotic PE and PC (APE and APC), denoted \( \text{PE}(\hat{y}_n(h), 1-\alpha) \) and \( \text{PC}(\hat{y}_n(h), 1-\alpha) \), can be constructed by replacing unknowns in (5.2) and (5.4) with estimators. It is clear that these APR’s converge to the true PR’s in large samples.

The bootstrap counterparts of the APE and APC are now introduced. To construct the BPR’s, we refer to procedures (F-i) to (F-iii), presented in Chapter 2, to obtain a bootstrap forecast distribution \( \{y_n(h)\}_{i=1}^{b} \). The bootstrap forecasts have been shown to be asymptotically valid by Theorem 2.2. That is, bootstrap forecasts converge to the true future values in large samples, which provides the basis of using a bootstrap PR for VAR forecasts as an alternative to its asymptotic counterpart in small samples.

Consider the quadratic form

\[
Q[\hat{y}_n(h)] = \{y_{n+h} - \hat{y}_n(h)\}' \hat{\Sigma}_y(h)^{-1} \{y_{n+h} - \hat{y}_n(h)\},
\]

where \( \hat{\Sigma}_y(h) \) is as given in Chapter 4 and \( \hat{y}_n(h) \) is the sample mean of \( \{y_n(h)\}_{i=1}^{b} \). Then the bootstrap PE (BPE) with nominal coverage 1-\( \alpha \), denoted \( \text{PE}(\hat{y}_n(h), 1-\alpha) \), is defined as

\[
\text{PE}(\hat{y}_n(h), 1-\alpha) = \{y_{n+h} : y_{n+h} \text{ satisfies } Q[\hat{y}_n(h)] \leq \hat{Q}_{1-\alpha}\},
\]

where \( \hat{Q}_{1-\alpha} \) is the 100(1-\( \alpha \))th percentile of the distribution

\[
\{\{y_n(h)^1 - \hat{y}_n(h)\}' \hat{\Sigma}_y(h)^{-1} \{y_n(h)^1 - \hat{y}_n(h)\}\}_{i=1}^{b}.
\]
To obtain the bootstrap PC, we consider the same cubical approximation as was used in the asymptotic case, based on Bonferroni's method, to yield the bootstrap PC (BPC). The BPC with nominal coverage of at least $1-\alpha$, denoted $\text{PC}(\hat{y}_n^*)$, is defined as

$$\text{PC}(\hat{y}_n^*, 1-\alpha) = \{y_{n,n}; y_{n,n} \in \bigcup_{k=1}^{K} [y_{k,n}^*(h,\tau), y_{k,n}^*(h,1-\tau)]\},$$

where $y_{k,n}^*(h,\tau)$ is the $100\tau$th percentile of the distribution $\{y_{k,n}^*(h)\}_{1=1}^{b}$ and $\tau = 0.5\alpha/K$. From this point on, the term PI, when it is used in the context of the PR's for VAR models, indicates a single interval which constitutes a component interval of the PC. That is, in the bivariate case, a PC consists of two PI's; and an API (BPI) indicates a component interval generated from an APC (BPC).

5.3 Experimental Design

The data generating mechanisms used here are bivariate VAR models of order one and two. The VAR(1) models employed are M1a, M1c and M1k in as given Table 3.1; and the VAR(2) models are M2a, M2b, M2d and M2f as given in Table 3.2. Model M1a is a near non-stationary model and M2d is a model with highly correlated VAR components, both of which have been found to generate large biases in parameter estimators and forecasts. The sample sizes considered are 30, 60 and 100. The number of Monte Carlo iterations $N$ is 100 and the number of bootstrap iterations $b$ is set to 1000. The number of bootstrap iterations is found to be sufficient from preliminary simulations.
The intercept term \( \nu \) is assumed to be unknown and fixed for all VAR processes to \((0, 0)\)'. In evaluating and comparing small sample properties of asymptotic and bootstrap PR's, normal innovations with zero mean and the following variance-covariance matrices are considered:

\[
\Sigma_u^1 = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 2 \end{bmatrix} \quad \text{and} \quad \Sigma_u^2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}.
\]

These specifications imply that components of innovations are allowed to possess different variances and they are moderately correlated with correlation coefficients being equal to 0.5 and -0.5 respectively. Unequal variances are employed in this chapter in an attempt to give a more general stochastic structure for innovation processes. As in Chapter 4, forecasting horizons, \( h \), chosen are 1, 4 and 8. The nominal coverages of PR's used are 80% and 90%. By Bonferroni's method, the corresponding PC's possess nominal coverages of at least 80% and 90%, and the individual PI's at least 90% and 95% respectively.

The non-normal distributions introduced in Chapter 3 are again employed to evaluate the robustness of asymptotic PR's to non-normal innovations. Approximate Student's \( t \) with \( \tau = 4 \) is used as a representative of distributions with fatter tails than the normal; and the chi-squared with the degrees freedom 5, and exponential distributions are used as representatives of asymmetric distributions with the asymmetry of the latter being more extreme than that of the former. The variance-covariance structure equivalent to \( \Sigma_u^1 \) is assigned to each non-normal distributions: for approximate Student's \( t \), \( \rho = 0.707 \) and \( k = 2 \); and for chi-squared and exponential, \( \rho = 0.7 \) and \( k = 2 \).
As measures of evaluating and comparing small sample performances of asymptotic or bootstrap PR's, empirical coverage and volume are employed. The empirical coverage is defined as the proportion of the true future values which belong to the interior of the APR or BPR constructed. It is likely that a PR, with its empirical coverage being substantially different from the nominal one, provides a misleading interpretation of future uncertainty. The volume of a PR, on the other hand, represents the extent of uncertainty associated with the forecasts. A PR with larger volume indicates that larger uncertainty is associated with forecasts than the one with smaller volume. Hence, given that its empirical coverage is reasonably close to the nominal one, a PR with smaller volume should be preferred since the larger volume implies an over-estimation of future uncertainty. Previous Monte Carlo studies in the univariate case (see Stein, 1987; Thombs & Schucany, 1990; and Masarotto, 1990) also have adopted these measures.

In evaluating the empirical coverages, the true future values need to be generated. Let \( \{y_{n+h}^i\}_{i=1}^F \) be the set of \( F \) true future values generated as

\[
y_{n+h} = \nu + A_1 y_{n+h-1} + \ldots + A_p y_{n+h-p} + u_{n+h}
\]

with starting values equal to the last \( p \) values of the original series and \( u_{n+h} \) is i.i.d. innovations. The number of true future values \( F \) is set to 100, which was found to be adequate in preliminary simulations. The empirical coverage of a PE is calculated as the proportion of future values that belong to the PE and can be expressed as

\[
\hat{C}(\text{PE}(.)) = \frac{\# \{ y_{n+h}^i : y_{n+h}^i \in \text{PE}(.) \}}{F}, \quad i = 1, \ldots, F.
\]
where # represents "the number of" and \( \{y^1_{n+h i=1}\}^F \) are the true future values generated from the true model. The empirical coverage of a PC, denoted \( \hat{C}(\text{PC}[.]) \), can be defined in a similar way. The volume of a PE can be calculated by the formula (see Adkins & Carter Hill; 1990)

\[
V(\text{PE}[.]) = \left[ \pi^{K/2} \Gamma(0.5 \, K + 1) \right] \, C^{K/2} \, (\text{det}[\Sigma_y(h)^{-1}])^{-0.5},
\]

where \( C = \chi^2_\alpha(K) \). Note that \( C^{K/2} (\text{det}[\Sigma_y(h)^{-1}])^{-0.5} = \prod_{i=1}^{K} a_i \), where \( a_i \) is the length of ith semi-axis of the PE. The volume of a PC, denoted \( V(\text{PC}[.]) \), can be calculated obviously by multiplying the lengths of each PI which is a component of the PC.

5.4 Performances of asymptotic prediction regions

All comparisons below are made with normal innovations with \( \Sigma_{u,1} \), nominal coverage 90\% and sample size 60 unless stated otherwise. The results obtained in the other cases are found to be qualitatively similar. The case of VAR processes with non-normal innovations is dealt with later in this section.

The coverages of asymptotic PR's are found to be always larger, and in general to be closer to the nominal coverage than those of their bootstrap counterparts. The volumes of APR's are found to be slightly larger than those of their bootstrap counterparts. This suggests that APR's generally provide conservative but more accurate assessment of future uncertainty than their bootstrap counterparts; and that BPR's possess a tendency of
under-estimating the future uncertainty in terms of coverage properties. This feature is illustrated in Figures 5.1 and 5.2 where mean coverages of APE and BPE are compared respectively for model M1c and M2a. Their volume properties are summarised in Table 5.1. The under-estimation by BPE is more evident for model M2a. Similar conclusions can be drawn when the comparisons are made in terms of PC, a cubical approximation to the PE. That is, APC's provide more conservative and accurate assessment of future uncertainty than their bootstrap counterparts. This result suggests that APE’s or APC’s should be preferred to their bootstrap counterparts in small samples when comparisons are made in terms of coverage and volume properties. The property of bootstrap PR's being liberal agrees with the results obtained earlier by Stein (1985) and Thombs & Schucany (1990), who respectively examined the cases of linear regression and univariate AR models.

It is again of particular interest to consider VAR models which generate large biases in parameter estimators. It is found in Chapter 3 that near non-stationary models and models with highly correlated VAR components generate large biases in parameter estimators. In Chapter 4, these models are found to generate large forecast biases; large variability; and high frequency of non-normality of VAR forecast error distributions. Here we attempt to evaluate small sample performances of APR and BPR of these models. It is found that APR’s still outperform their bootstrap counterparts. Table 5.2 compares, when the sample size is 100, mean coverages and volumes of APE and BPE for models M1a and M2d, which are, respectively, near non-stationary and have highly correlated VAR components. For model M2d, it can be seen that coverages of APE’s are much closer to the nominals than those of BPE’s, while their volumes are

168
reasonably close to each other. In the case of model M1a, APE's exhibit coverages closer to the nominals with smaller volumes than those of the BPE's. Although APE appears to perform better, extreme care should be taken in this case as the performance of APE is still very poor. This contrasts to the univariate finding of Thombs & Schucany (1990) where near non-stationarity makes no substantial difference in small sample performances of bootstrap PI's. As for the BPR's, the extent of under-estimation is found to be greater when near non-stationary models and models with highly correlated VAR components are employed. A larger number of parameters and a more complicated structure of VAR models may be attributed to this difference.

The robustness of APR’s to non-normal innovations are also evaluated. We allow innovations of VAR processes to follow non-normal distributions as specified in Section 5.3 and evaluate performance of the APR’s in terms of their coverage and volume properties. Table 5.3 compares mean coverages and volumes of APE generated with normal and non-normal innovations when models M2a and M2b are used with the sample size 30. It can be seen that coverages and volumes are not sensitive to the departure of innovations from normality. Coverages from non-normal innovations are very close to those from normal, although the latter are found to be closer to the nominal coverage than the former. It can also be seen from Table 5.3 that volumes of APE’s are not sensitive to departure of innovations from normality. The results obtained here suggest that APE’s perform reasonably well under a wide range of non-normal innovations in small samples. Similar results can be found when comparisons are made in terms of APC when innovations are non-normal. The robustness of APR’s to non-normal innovations also agrees with the result obtained in the univariate case of Thombs & Schucany
(1990), where asymptotic PI's are found to perform well with innovations generated with skewed distributions such as the exponential.

5.5 Improved prediction cubes for VAR forecasts

In an attempt to overcome the problem of PC over-estimating the future uncertainty, an improved PC is proposed in this section. The extent of over-estimation, in fact, depends on the strength of correlation among VAR forecasts, as we shall see below graphically and numerically for a simple case. When the components of forecasts are moderately or strongly correlated, the volume of the PC can be far larger than that of the PE. Figure 5.3 illustrates the case where components of forecasts are moderately correlated in the bivariate case. Note that the semi-axes of the PE are not parallel to the axes of forecasts because the forecasts are linearly correlated. Over-estimation of future uncertainty by the PC appears clear as the area of the PC is far larger than that of the PE. As correlation gets stronger given the fixed variances, the shape of PE will get thinner, while that of the PC is unchanged resulting in a more serious over-estimation. When the correlation is zero, the semi-axes of the PE become parallel to the axes of forecasts, in which case a close approximation can be made by using the PC. Note that the intervals consisting of the PC should be shorter than the intervals whose limits are extreme points of the PE. The underlying reasons are given in detail by Davidson & MacKinnon (1993; p75).

Suppose we want to construct a PE in the bivariate case with the nominal coverage of 95% (α = 0.05). Consider the simple case with
vech($\Sigma_y(h)$) = (1, \rho, 1)'. Since $\alpha = 0.05$, $C = \chi^2_\alpha(K) \approx 5.99$. Then it is easy to see that

$$V(PE[.]) \approx \pi 5.99 (1-\rho^2)^{1/2}.$$  

Consider the case where $\rho = 0$. The PE is a circle whose area is $5.99\pi$. From (5.4), the length of an interval generated from the corresponding PC is $2z_{\tau} \approx 4.48$ since $\sigma_k(h) = 1$, where $\tau = 0.5(\alpha/K) = 0.0125$ and $z_{\tau} \approx 2.24$. The area of the PC is $(2z_{\tau})^2 \approx 20.07$, while the area of the PE is $5.99\pi \approx 18.89$. Hence, even when components of forecasts are uncorrelated, the PC still over-estimates the area of the PE. Since $\rho = 0$ maximises $(1-\rho^2)^{1/2}$, as the absolute value of $\rho$ increases, the volume of the PE declines. Geometrically, the shape of the PE gets thinner, with its semi-axes becoming unparallel to the axes of forecasts, reducing its volume. The area of the PC, however, is unchanged because the variances of each component of forecasts are fixed at unity. Hence, we see that the extent of over-estimation of future uncertainty by the PC gets larger as the correlation between components of forecasts increases. This result can also be extended to deal with the higher dimensional case or the case with $\Sigma_y(h)$ possessing more general structure.

A VAR system involves a set of variables linearly inter-related with an innovation process whose components can be correlated. It is therefore highly likely that components of a VAR system are correlated and so are the associated VAR forecasts. Hence, the use of a PC for VAR forecasts can lead to a serious over-estimation of future uncertainty especially when components of forecasts are more than moderately correlated, as we have seen above. In this section, we propose the use of an alternative PC which
can attenuate the problem of over-estimation. This alternative, called the improved PC (IPC), can be obtained by considering a transformation which removes correlation among components of VAR forecasts.

Since $\Sigma_y(h)$ is a $K \times K$ symmetric positive definite matrix, it has $K$ distinct eigenvectors and $K$ positive real eigenvalues. Let $S_y(h)$ be a diagonal matrix with its diagonal elements being the eigenvalues of $\Sigma_y(h)$. As $\Sigma_y(h)$ is a positive definite matrix, $\Sigma_y(h) = P_1 P_1'$ by the Choleski decomposition, where $P_1$ is a lower triangular matrix whose diagonals are all positive. Let $P_2$ be a diagonal matrix such that $S_y(h) = P_2 P_2'$. Consider the transformation $g_n(h) = P_2^{-1} y_{n+h}$. Then, it can easily be seen that

$$E[g_n(h)g_n(h)'] = S_y(h).$$

Components of the transformed forecasts $g_n(h)$ are uncorrelated as $S_y(h)$ is diagonal. Hence, this transformation removes any correlation present in the forecasts, and rotates the PE so that its semi-axes are parallel to those of the transformed forecasts. Note that the direction of rotation depends on the ordering of eigenvalues in $S_y(h)$. The component assigned with smaller eigenvalue will exhibit shorter semi-axis. Again, the ordering of eigenvectors in $M$ will determine the direction of rotation.

Consider the quadratic form

$$Q[g_n(h)] = [g_{n+h} - g_n(h)]' S_y(h)^{-1} [g_{n+h} - g_n(h)],$$

where $g_{n+h} = P_2^{-1} y_{n+h}$. The transformed PE with nominal coverage $1-\alpha$ can be defined as
\[ \text{PE}[g_n(h), 1-\alpha] = \{g_{n,h} : g_{n,h} \text{ satisfies } Q[g_n(h)] \leq \chi^2_\alpha(K) \}. \]

Note that the volume of \( \text{PE}[g_n(h), 1-\alpha] \) is identical to that of \( \text{PE}[y_n(h), 1-\alpha] \). The corresponding PC, based on Bonferroni's method, with nominal coverage at least \( 1-\alpha \), defined as the IPC, can be written as

\[ \text{PC}[g_n(h), 1-\alpha] = \{g_{n,h} : g_{n,h} \in \bigcup_{k=1}^{K} [g_{k,n}(h) \pm z_\alpha s_k(h)] \}. \]

where \( g_{k,n}(h) \) is the \( k \)th component of \( g_n(h) \) and \( s_k(h) \) is square root of the \( k \)th diagonal element of \( S_y(h) \). Again, its coverage can be defined in the same way as in (5.3) and the calculation of its volume can be done in an obvious way.

The asymptotic IPC (AIPC) \( \text{PC}[\hat{g}_n(h)] \) can be constructed from \( \hat{g}_n(h) = \hat{P}_2 \hat{P}_1^{-1} y_n(h) \) and \( \hat{S}_y(h) = \hat{P}_2 \hat{P}_2' \), where \( \hat{P}_1 \) and \( \hat{P}_2 \) are the estimators of \( P_1 \) and \( P_2 \). The bootstrap IPC (BIPC) with nominal coverage at least \( 1-\alpha \), denoted \( \text{PC}[\hat{g}_n(h)] \), can be obtained in a similar way from the transformed bootstrap forecasts \( \hat{g}_n(h)^\prime = \hat{P}_2^* \hat{P}_1^{-1} y_n(h)^\prime \) and \( \hat{S}_y(h) = \hat{P}_2^* \hat{P}_2^* \), where \( \hat{P}_1^* \) and \( \hat{P}_2^* \) are the bootstrap counterparts of \( \hat{P}_1 \) and \( \hat{P}_2 \).

Geometrical interpretation of the IPC in the bivariate case can be seen by comparing Figures 5.3 and 5.4 where the PC and IPC are illustrated with the corresponding PE's. As mentioned earlier, when forecasts are moderately or strongly correlated, the PC over-estimates the future uncertainty in terms of larger area (or volume) than that of the PE, as appeared in Figure 5.3. Figure 5.4 exhibits the transformed PE and IPC in the plane of the transformed forecasts. The IPC approximates the
transformed PE whose semi-axes are parallel to the axes of the transformed forecasts, and provides a reasonably accurate assessment through the improved volume property. The improvement in volume property can be expected to be greater in higher dimensional cases where the forecasts are moderately or strongly correlated. Now suppose we want to evaluate whether the region F in Figure 5.3 is likely to be realized in the future. The use of the PC is misleading as the region F is located outside the PE but inside the PC. However, if the IPC is used as in Figure 5.4, a correct judgment on the future uncertainty can be made as the region F is transformed to the region G which is located outside both of the IPC and PE.

Note that, in the bivariate case, one of the PI's obtained from the IPC will be shorter than the corresponding PI obtained from the PC, while the length of the other PI will virtually be unchanged. The reduction of the volume of the cube caused by the shorter interval represents the extent of over-estimated uncertainty or gain from using the IPC. As mentioned earlier, the question of which one of them will become shorter depends on the ordering of eigenvalues in $\Sigma_y(h)$. The component assigned with smaller eigenvalue will exhibit shorter PI in the IPC. This feature is also true in higher dimensions. The method of assigning eigenvalues here is to order them in the same order as the magnitudes of the variances of component series.

We begin by comparing performances of PC and IPC in assessing future uncertainty. The experimental designs are as given in Section 5.3. As before, comparisons are made with normal innovations with $\Sigma_u^1$, nominal coverage 90% and sample size 60. But the results obtained in the other
cases are qualitatively similar. In Figures 5.5 and 5.6, coverage and volume properties of asymptotic PE, PC and IPC are compared for model M2b. It can be seen from Figure 5.5 that coverages of PE are close to those of IPC, but coverages of PC are higher than those of the others. The extent of over-estimation by PC is more pronounced when volume properties are compared, as appeared in Figure 5.6. For each \( h \) considered, it can be seen that the volumes of PC are larger than those of the others, while the volumes of IPC are reasonably close to those of PE. In fact, the volumes of PC are about 30\% more than those of the PE. Hence, we observe that the use of IPC can provide reasonably accurate assessment of future uncertainty in terms of coverage and volume properties. The case of the bootstrap PE, PC and IPC appears similar.

The model M2b discussed above in fact generates forecasts with moderate correlations since \((\rho_1, \rho_4, \rho_8) = (0.5, -0.6, -0.6)\), where \( \rho_h \) indicates the true correlation of forecast when forecasting period is \( h \). The gain of using IPC would be much higher if the model under consideration generates higher correlations among forecasts. Consider the model Mlk with two different structures of variance-covariance of disturbances \( \Sigma^1_u \) and \( \Sigma^2_u \). The model Mlk yields \((\rho_1, \rho_4, \rho_8) = (0.5, 0.73, 0.73)\) when combined with \( \Sigma^1_u \); and \((\rho_1, \rho_4, \rho_8) = (-0.5, 0.58, 0.58)\) when combined with \( \Sigma^2_u \). The volume properties of each case are presented in Figures 5.7 and 5.8. It can be seen that IPC again provides volumes reasonably close to those of PE in both cases and that the extent of over-estimation by PC appears to be larger in Figure 5.7 where VAR forecasts possess higher correlations. More precisely, the volumes of PC are approximately 25\%, 60\% and 60\% larger, respectively for \( h = 1, 4, 8 \), than those of PE when \( \Sigma^1_u \) is used; and approximately 24\%, 32\% and 32\% larger when \( \Sigma^2_u \) is used. Hence, we observe
that the gain of using IPC over PC is evident, and it gets higher as the correlation of forecasts increases when h is more than 4.

An extreme case of highly correlated VAR forecasts can be seen from model M2d, a model with highly correlated VAR components. It can be seen that \((\rho_4, \rho_8) = (0.5, -0.98, -0.98)\) when \(\Sigma^1\) is employed. A substantially large over-estimation of future uncertainty is expected when \(h = 4\) or \(8\). Figures 5.9 and 5.10 respectively compare empirical coverages and volumes of asymptotic PE, PC and IPC. Although only moderate over-estimation in terms of coverage properties is evident, over-estimation to a greater extent is made by PC when volume properties are considered. The volumes of PC are about 5 times bigger than those of PE and IPC when h is 4 or 8. The results obtained from this and the last two paragraphs suggest that a substantial over-estimation of future uncertainty can be made by employing PC, especially in terms of volume properties. The use of IPC is strongly recommended as its empirical coverages and volumes are reasonably close to those of PE, especially when forecasts are more than moderately correlated.

An interesting point can be observed by comparing the mean and median coverages. It is found that asymptotic and bootstrap PR's generate coverages whose medians are larger than their means for all cases considered. An example can be seen in Figure 5.11 where mean and median coverages for BIPC from model M2b are given. Earlier work by Thombs & Schucany (1990) for the univariate AR case found that the distribution of coverage is skewed with its mean smaller than the median. Stein (1985) also found that the coverage of a bootstrap PI of a purely random process asymptotically follows a beta distribution with parameters \(\beta(n+1)\) and
(1-\(\beta\))(n+1), where \(\beta\) is the value of nominal coverage. When \(\beta\) takes the values usually employed in practice such as 0.95 or 0.9, the associated beta distribution is skewed to the left, with its median being greater than the mean. Hence, the feature of coverage distributions being skewed to the left found in the simpler cases seems to also hold in the VAR case.

Other general findings worth mentioning are summarised below. It can be seen that, as sample size increases, empirical coverages increase and get closer to the nominal coverage and volumes decrease. An example can be found for model M2a in Tables 5.1, 5.3 and Figure 5.2 where coverage and volumes are presented with sample size 30 and 60. This feature agrees with the finding of Thombs & Schucany (1990) in the univariate case. It is also observed that, as \(h\) increases, the empirical coverages increase and get closer to the nominal coverage. It is also plausible to observe that volumes increase with \(h\), which indicates that uncertainty associated with forecasts increases as the rate of extrapolation increases. Again plausibly, as sample size increases, empirical coverages of asymptotic and bootstrap PR's exhibit a tendency of converging to the nominal one. Volumes of asymptotic and bootstrap PR's get closer each other as sample size increases.

5.6 Concluding Remarks

At the outset of this chapter, the two possible reasons as to why practitioners may choose not to employ PR's for VAR forecasts were discussed. One is the possibility that the PC can substantially over-estimate future uncertainty; and the other is unknown small sample
properties of the asymptotic approximations involved. It was found in this chapter that the use of the PC as an approximation to the PE can cause serious over-estimation of future uncertainty, especially when forecasts are more than moderately correlated. However, the IPC, proposed here as an alternative to the PC, is found to provide reasonably accurate assessment of future uncertainty, especially in terms of volume properties. The gain of using IPC was found to be greater when forecasts are correlated to a higher degree. Hence, the use of the IPC is strongly recommended when PR’s are employed for multivariate models such as VAR because forecasts are likely to be more than moderately correlated. A drawback of using the IPC is that it provides a prediction for the transformed true future value. However, the transformation is only linear and the forecasts for the true future value can easily be recovered. The idea of the IPC can easily be applied to statistical inference of jointly inter-related parameters as an alternative to the confidence ellipsoid.

Small sample properties of asymptotic and bootstrap PR’s were compared. It was found that asymptotic PR’s provide more conservative and accurate assessment of future uncertainty than their bootstrap counterparts, while the latter was found to exhibit a tendency of under-estimating future uncertainty and sometimes provide too liberal assessments. Asymptotic PR’s were also found to be robust to a wide range of non-normal innovations in small samples, in terms of coverage and volume properties. These findings agree with that of the univariate case of Thombs & Schucany (1990) and Masarotto (1990). Practitioners are recommended to employ asymptotic PR’s in evaluating uncertainty associated with VAR forecasts because they are found to perform reasonably well in small samples and also computationally cheaper to construct than their bootstrap
counterparts. A subject of future research is to refine the bootstrap PR's so that they can be a more effective alternative to their asymptotic counterparts. In fact, Efron's (1981, 1987) BC and $B_{C_{a}}$ methods have been applied to VAR forecasts. The bootstrap PR's based on these methods, however, were found to make little improvement. Disappointing performance of $B_{C_{a}}$ confidence intervals, in the context of SUR models, were also reported in a recent Monte Carlo study of Rilstone & Veall (1996).
Appendix 5.1. Tables.

Table 5.1. Mean volumes of PR's (n=60)

<table>
<thead>
<tr>
<th>h</th>
<th>APE</th>
<th>BPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.65</td>
<td>12.54</td>
</tr>
<tr>
<td>M1c</td>
<td>4</td>
<td>32.65</td>
</tr>
<tr>
<td>8</td>
<td>33.38</td>
<td>33.34</td>
</tr>
<tr>
<td>1</td>
<td>18.63</td>
<td>18.80</td>
</tr>
<tr>
<td>M2a</td>
<td>4</td>
<td>185.74</td>
</tr>
<tr>
<td>8</td>
<td>254.58</td>
<td>176.21</td>
</tr>
</tbody>
</table>
Table 5.2. Mean coverages and volumes of PR's (n=100)

<table>
<thead>
<tr>
<th></th>
<th>APE</th>
<th></th>
<th>BPE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coverage</td>
<td>volume</td>
<td>coverage</td>
<td>volume</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>88.93</td>
<td>17.78</td>
<td>83.42</td>
</tr>
<tr>
<td>Mia</td>
<td>4</td>
<td>87.45</td>
<td>68.15</td>
<td>75.26</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>84.61</td>
<td>136.83</td>
<td>70.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>89.12</td>
<td>18.27</td>
<td>71.78</td>
</tr>
<tr>
<td>M2d</td>
<td>4</td>
<td>89.25</td>
<td>354.66</td>
<td>86.92</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>88.79</td>
<td>408.81</td>
<td>87.32</td>
</tr>
</tbody>
</table>

Note. Coverages are in percentages.
Nominal Coverage is 90%.
Table 5.3. Coverages and volumes of APE’s (n=30)

Coverages (in percentage)

<table>
<thead>
<tr>
<th>h</th>
<th>Normal</th>
<th>Student's t</th>
<th>Exponential</th>
<th>Chi-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85.90</td>
<td>84.71</td>
<td>85.23</td>
<td>85.20</td>
</tr>
<tr>
<td>8</td>
<td>88.65</td>
<td>85.24</td>
<td>86.88</td>
<td>85.01</td>
</tr>
<tr>
<td>M2a</td>
<td>87.63</td>
<td>85.28</td>
<td>85.29</td>
<td>84.35</td>
</tr>
<tr>
<td>1</td>
<td>86.32</td>
<td>82.94</td>
<td>84.67</td>
<td>85.42</td>
</tr>
<tr>
<td>8</td>
<td>88.88</td>
<td>89.06</td>
<td>89.95</td>
<td>89.93</td>
</tr>
<tr>
<td>M2b</td>
<td>88.92</td>
<td>87.75</td>
<td>88.25</td>
<td>89.17</td>
</tr>
</tbody>
</table>

Volumes

<table>
<thead>
<tr>
<th>h</th>
<th>Normal</th>
<th>Student's t</th>
<th>Exponential</th>
<th>Chi-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.48</td>
<td>24.48</td>
<td>20.33</td>
<td>20.24</td>
</tr>
<tr>
<td>8</td>
<td>285.98</td>
<td>312.21</td>
<td>300.27</td>
<td>336.24</td>
</tr>
<tr>
<td>M2a</td>
<td>200.94</td>
<td>225.14</td>
<td>210.27</td>
<td>232.17</td>
</tr>
<tr>
<td>1</td>
<td>19.46</td>
<td>24.07</td>
<td>13.21</td>
<td>12.95</td>
</tr>
<tr>
<td>8</td>
<td>78.24</td>
<td>74.03</td>
<td>77.27</td>
<td>76.29</td>
</tr>
<tr>
<td>M2b</td>
<td>77.64</td>
<td>72.25</td>
<td>70.24</td>
<td>71.14</td>
</tr>
</tbody>
</table>

Note. Coverages are in percentages.
Nominal Coverage is 90%.
Appendix 5.2. Figures.

Figure 5.1. Coverages of PE: n=60; 90% nominal; M1c

Figure 5.2. Coverages of PE's: n=60; 90% nominal; M2a.
Figure 5.3. Prediction Ellipse and Prediction Cube

Figure 5.4. Improved Prediction Cube
Figure 5.5. Coverages of APR's: n=60; 90% Nominal; M2b; SIG1.

Figure 5.6. Volumes of APR's: n=60; 90% Nominal; M2b; SIG1.
Figure 5.7. Volumes of APR's: n=60; 90% Nominal; M1k; SIG1.

Figure 5.8. Volumes of APR's: n=60; 90% Nominal; M2d; SIG1.
Figure 5.9. Coverages of APR's: n=60; 90% Nominal; M2d; SIG1.

Figure 5.10. Volumes of APR's: n = 60; 90% Nominal; M1k; SIG2.
Figure 5.11. Coverages of BPE's: n=60; 90% Nominal; M1e; SIG1.
Chapter 6. VAR Order Selection and Bootstrapping

6.1 Introduction

Order selection is a crucial statistical task in VAR modelling. The reliability of statistical and structural analyses with VAR models - such as forecasting, variance decompositions and impulse response analysis - depends in large part on the integrity of order selection. Under-estimation of the true order yields inconsistent parameter estimators which may lead to misleading interpretations of further statistical analyses. Although over-estimation of the true order preserves consistent parameter estimators, a large number of additional, but unnecessary, parameters has to be estimated which can cause excessive sampling variability. It is more often the case in practice that the true VAR order is non-existent. For example, consider the case of a VAR approximation to a more general class of time series models such as VARMA. Although the true VAR order in this case is infinite, there exists a finite approximate VAR order which ensures variability of the associated error structure, measured by the determinant of the innovation variance-covariance matrix, sufficiently close to that of the original process (see, for the univariate case, Poskitt & Tremayne, 1987). Under-estimation or otherwise of this approximate VAR order will bring similar consequences to those of the case where the true order is existent.

As a means of VAR order selection, the use of order selection criteria (see Chapter 4 of Lütkepohl, 1991) is considered in this chapter. It is purely an objective method in which the VAR order is automatically selected.
by minimising an appropriate criterion. Statistical tests, such as the portmanteau test (Sec. 4.4.3 of Lütkepohl, 1991), likelihood ratio test (Sec. 4.4.2 Lütkepohl, 1991) and score test (see, for example, Poskitt & Tremayne, 1982), may be used for the same purpose. However, these tests may often exhibit poor power and size properties in small samples, casting doubt on their suitability as VAR order selection devices. Moreover, in practice, they are more popular as diagnostic checks rather than order selection devices. Attention is restricted to the AIC, BIC and HQ order selection criteria; details of these are given in Chapter 2. As mentioned in Chapter 2, other order selection criteria are considered here to be superfluous due to their similar small sample properties to the AIC, BIC and HQ trio; see Mills & Prasad (1992) and Lütkepohl (1985).

In this chapter, the bootstrap VAR order selection method is proposed. The MBB and SB methods introduced in Chapter 2 are adopted; they generate bootstrap samples in which salient features of dependence structure present in the original data are replicated. Theorems 2.6 and 2.7 provide asymptotic validity of these methods in the sense that a bootstrap sample generated with the MBB or SB methods yields autocorrelations and VAR parameter estimators converging asymptotically to the true values in probability. Through repeated resampling, a large number of bootstrap samples can be generated. From each bootstrap sample, the VAR order is selected by using a order selection criterion to obtain the bootstrap distribution of VAR orders. The bootstrap order can be determined by taking the mode of the bootstrap distribution. This estimator provides an alternative to the VAR order selected by applying an order selection criterion directly to the observed data, hereafter called the asymptotic order and denoted \( \hat{p} \). To be more precise, we reintroduce the bootstrap order
selection procedure, labelled (OS-i) to (OS-iii), based on the MBB and SB. The notations used are the same as those employed in Chapter 2. The bootstrap order, denoted $\hat{p}^*$, is determined as follows:

(OS-i) Forming the blocks of sub-series

Consider a stationary K-dimensional VAR(p) process $Y = (y_1, \ldots, y_n)$ as given in (2.1). Let $B(I,h) = (y_{i}, \ldots, y_{i+h-1})$ denote a block of length $h$ starting from $y_i$, where $h$ and $I$ are positive integers.

For $y_j \in B(I,h)$, if $j > n$, the corresponding block is not defined in the MBB; whereas, in the SB, $y_j$ is set to $y_{j-n}$ (circular treatment of data).

(OS-ii) Obtaining a bootstrap sample

Let $h_1$ be an i.i.d. random variable with mean $\bar{h}$ and $I_i$ be i.i.d. from the discrete uniform distribution on {1, ..., n}. Randomly draw blocks from \{B(I_1,h_1), B(I_2,h_2), B(I_3,h_3), ... \}, with replacement, to construct a bootstrap sample of size $n^*$ denoted as $G = (S_{1}, \ldots, S_{n^*})$. Note that $S_1$ indicates a member of a randomly drawn block.

(OS-iii) Obtaining the bootstrap distribution of selected orders

Repeat (OS-ii) $b$ times to obtain $b$ sets of bootstrap samples $\{G_j^b\}_{j=1}^b$. For each bootstrap sample, VAR order can be selected using an order selection criterion to form the bootstrap distribution of selected orders $\{\hat{p}^*_j\}_{j=1}^b$. That is, $\hat{p}^*_j$ minimizes
\[ IC^*_j(m) = \ln(\det[\tilde{\Sigma}^*_u(m)]) + mK^2 C(n^*)/n^*, \quad m = 0, 1, \ldots, P, \]

where \( \tilde{\Sigma}^*_u(m) \) is the bootstrap counterpart of \( \tilde{\Sigma}_u(m) \) obtained from \( G^J \). The bootstrap order \( \hat{p}^* \) is determined by taking the mode of \( \{\hat{p}^*_j\}_{j=1}^b \).

As mentioned in Chapter 2, the SB procedure of Politis & Romano (1994) adopts the geometric distribution as given in (2.15) as a probability generation mechanism of \( h_i \)'s. For other details of the MBB and SB with respect to the choices of \( n^* \) and \( \tilde{n} \), see Chapter 2. In short, \( n^* \) is allowed to be different from, but usually greater than, \( n \); and \( \tilde{n} \) should be chosen by carefully considering the sample size and the length of memory of the process. The empirical choices of these values in the context of VAR order selection are examined later in this chapter.

Lemma 6.1 below provides an asymptotic result when the VAR order is under-estimated. It will be useful in establishing asymptotic validity of bootstrap order selection.

**Lemma 6.1.** Let \( \{y_t\} \) be a K-dimensional strictly stationary VAR(p) process as defined in (2.1). Suppose that the bootstrap is conducted as in (OS-I) to (OS-iii) and that the assumptions in Theorems 2.6 and 2.7 are satisfied. Then,

\[ \text{plim}[\tilde{\Sigma}^*_u(m)] = \text{plim}[\tilde{\Sigma}_u(m)] = \Sigma_u + \Delta_m \quad \text{for} \quad m < p. \]
where $\Lambda_m$ is a positive definite matrix.

Proof.

The VAR(p) model can be expressed in the multivariate regression form as

(a) \[ Y = Z_1B_1 + Z_2B_2 + U, \]

where

\[
\begin{align*}
Y &= (y_{1}', ..., y_{n}')' \\
B_1 &= (v', A_1', ..., A_m')' \\
B_2 &= (A_{m+1}', ..., A_p')' \\
Z_1 &= (Z_1^1, ..., Z_{n-1}^1) \\
Z_2 &= (Z_0^2, ..., Z_{n-1}^2) \\
Z_t^1 &= (1, y_{t}', ..., y_{t-m+1}') \\
Z_t^2 &= (y_{t-m}', ..., y_{t-p+1}') \\
U &= (u_{1}', ..., u_{n}')'
\end{align*}
\]

(nxK) \quad ((Km+1)xK) \quad [Kx(Km+1)] \quad [n\times(Kp+1)] \quad [n\times(Kp+1)] \quad ((Km+1)x1) \quad ((Km+1)x1) \quad (nxK)

(the dimensions of the matrices are indicated to the right).

The VAR(m) model can be written as

(b) \[ Y = Z_1B_1 + E. \]

Consider LS estimation for (a) and (b) to yield

\[ Y = Z_1\hat{B}_1 + Z_2\hat{B}_2 + \hat{U}; \text{ and } Y = Z_1\hat{B}_1 + \hat{E}, \]

where $\hat{B}_1$ and $\hat{B}_2$ are the LS estimators for $B_1$ and $B_2$ in (a), $\hat{B}_1$ is the LS
estimator for $B_1$ in (b), while $\hat{U}$ and $\hat{E}$ are corresponding LS residuals. It can be shown that

$$\hat{E}'\hat{E} = \hat{U}'\hat{U} + \hat{B}_2' Z_2' M_1 Z_1 \hat{B}_2'$$

where $M_1 = I_n - Z_1 (Z_1' Z_1)^{-1} Z_1'$. Note that $Z_2' M_1 Z_2$ is a matrix whose elements are sample autocovariances so that plim($Z_2' M_1 Z_2/n$) exists and positive definite. Since plim($\hat{U}'\hat{U}/n$) = $\Sigma_u$ by Corollary 3.2.1. of Lütkepohl (1991) and plim($\hat{B}_2$) = $B_2$ by Theorem 2.1, it can be seen that

$$\text{plim}(\hat{E}'\hat{E}/n) = \text{plim}[\hat{\Sigma}_u(m)] = \Sigma_u + \Delta_m,$$

where $\Delta_m = B_2' \text{plim}(Z_2' M_1 Z_2/n) B_2$.

For a bootstrap sample, we have

$$\hat{E}^*\hat{E}^* = \hat{U}^*\hat{U}^* + \hat{B}_2^* Z_2^* M_1^* Z_1^* \hat{B}_2^*$$

where $\hat{E}^*$, $\hat{U}^*$, $\hat{B}_2^*$, $Z_2^*$ and $M_1^*$ are the bootstrap counterparts of $\hat{E}$, $\hat{U}$, $\hat{B}_2$, $Z_2$ and $M_1$. Note that $Z_2^* M_1^* Z_2^*$ consists of bootstrap autocovariance matrices so that plim($Z_2^* M_1^* Z_2^*/n$) = plim($Z_2' M_1 Z_2/n$) by Theorem 2.6. Since, by Theorem 2.7, plim($\hat{U}^*\hat{U}^*/n$) = $\Sigma_u$ and plim($\hat{B}_2^*$) = $B_2$, the proof concludes as

$$\text{plim}(\hat{E}^*\hat{E}^*/n) = \text{plim}[\hat{\Sigma}_u^*(m)] = \Sigma_u + \Delta_m.$$

In Theorem 6.1 below, it is shown that the bootstrap and asymptotic
orders behave equivalently in large samples in the sense that the probability limit of the order selection criterion they minimise is identical.

**Theorem 6.1.** Let \( \{y_t\} \) be a K-dimensional strictly stationary VAR(p) process as defined in (2.1). Suppose that the bootstrap is conducted as in (OS-i) to (OS-iii) and that the assumptions in Theorems 2.6 and 2.7 are satisfied with \( P = O(n^{1/2}) \). Then it can be shown that

\[
\text{plim}[IC(m)] = \text{plim}[IC^*_j(m)].
\]

**Proof.**

Note that, for \( m = 0, 1, \ldots, P \),

\[
\hat{p} = \arg\min \left( IC(m) = \ln(\det[\widehat{\Sigma}_u(m)]) + mK^2 C(n)/n \right); \quad \text{and}
\]

\[
\hat{p}^*_j = \arg\min \left( IC^*_j(m) = \ln(\det[\widehat{\Sigma}^*_j(m)]) + mK^2 C(n^*)/n^* \right).
\]

\( C(n^*)/n^* \) is asymptotically equivalent to \( C(n)/n \) since \( n^* = O(n) \).

**Case.** \( m \geq p \).

It is well-known that \( \text{plim}[\widehat{\Sigma}_u(m)] = \Sigma_u \). By Theorem 2.7, \( \text{plim}[\widehat{\Sigma}^*_j(m)] = \Sigma_u \).

This indicates that \( \text{plim}[IC(m)] = \text{plim}[IC^*_j(m)] \).

195
Case. $m < p$.

By Lemma 6.1, $\text{plim}([\Sigma_u(m)]) = \text{plim}([\Sigma^*(m)]) = \Sigma_u + \Delta_m$. The proof concludes as $\text{plim}([IC(m)]) = \text{plim}([IC^*_j(m)])$.

**Remark 1.** By Theorem 6.1, the use of a consistent criterion leads to

$$p = \text{argmin} \{ \text{plim}([IC(m)]) \} = \text{argmin} \{ \text{plim}([IC^*_j(m)]) \}$$

$$= \text{plim}(\hat{p}) = \text{plim}(\hat{p}^*) = \text{plim}(\hat{p}^*),$$

since $\hat{p}^* = \text{mode}(\hat{p}_j)_{j=1}^b$.

**Remark 2.** If an inconsistent criterion is used,

$$\text{argmin} \{ \text{plim}([IC(m)]) \} = \text{argmin} \{ \text{plim}([IC^*_j(m)]) \} \geq p,$$

since $\lim_{n \to \infty} \text{Pr}(\hat{p} \geq p) = 1$.

Having established the asymptotic validity of bootstrap VAR order selection, its small sample performances are of particular interest. As mentioned in Chapter 2, Mills & Prasad (1992) and Lütkepohl (1985) found that small sample performances of order selection criteria, especially in
terms of selecting the true order, can often be poor. It is hoped that bootstrap VAR order selection improves small sample performances of asymptotic orders in selecting the true order. As $n^*$ is allowed to be larger than $n$, the bootstrap order provides the likely VAR order selected when a sample size larger than the original one is used, making more use of the asymptotic theories involved. In this chapter, Monte Carlo simulation is conducted to investigate small sample performances of bootstrap order selection. The major concern is whether the bootstrap method can provide a superior alternative to conventional VAR order selection. When the true order is non-existent, such as the case of VAR approximation to VARMA, the ability of selecting the approximate VAR order is evaluated. Comparisons are made by examining empirical frequencies of the asymptotic and bootstrap orders. The empirical frequencies here are referred to as the distribution of selected orders from Monte Carlo simulations. Each Monte Carlo iteration generates asymptotic and bootstrap VAR orders, with the latter being obtained from the bootstrap procedures nested under each Monte Carlo iteration. Basic descriptive statistics for empirical frequencies are used where necessary. Univariate AR models are also considered in this chapter because bootstrap order selection has never been applied to the univariate case. The bootstrap order selection procedures, based on MBB and SB, and their asymptotic validity in the context of univariate AR models can be established without any difficulty, as a special case of VAR.

The organisation of this chapter is as follows: in Section 6.2, experimental designs are summarised; results and discussions are presented in Section 6.3; Section 6.4 applies the bootstrap order selection method to economic data; and conclusions are drawn in Section 6.5.
6.2 Experimental Design

Tables 6.1 and 6.2 respectively report the univariate and bivariate AR models used in Monte Carlo simulations along with their eigenvalues. These models are stationary so that all of their eigenvalues lie outside the unit circle. Univariate models used are AR models of order one to three, one MA(1) and one ARMA(1,1); while, for the bivariate case, VAR models of orders 1 and 2 and one VMA(1) model are considered. Note that models M5 and M13 have some of their eigenvalues fairly close to the unit circle, indicating the presence of longer memory than the others. These models are used to examine how bootstrap order selection performs in small samples when the underlying process has a long memory. The sample sizes $n$ considered are 30, 60 and 100, and normal innovations are used throughout. For the univariate case, error variances are fixed at unity for all cases, while the variance-covariance matrix of innovations used for the bivariate case is $\text{vech}(\Sigma_u) = (1, 0.5, 1)'$. All elements of the intercept vector $\nu$ are fixed at unity for both the univariate and bivariate cases. Preliminary simulations found that the bootstrap order selection results are not sensitive to the changes in the structures of $\Sigma_u$ and $\nu$. The number of bootstrap iterations, $b$, is set to 100, and likewise the number of Monte Carlo iterations, $N$. The size of the bootstrap sample $n^*$ is set to 200, as this was found to be adequate in preliminary simulations. Detailed simulations result on the choice of $n^*$ is reported in the next section. A wide range of block lengths are simulated. Let $\bar{h}_n$ be the average block length chosen when sample size is $n$. The values of $\bar{h}_n$ considered here are $\bar{h}_{30} = (5, 10, 15)$; $\bar{h}_{60} = (5, 10, 15, 20)$; and $\bar{h}_{100} = (5, 10, 15, 20, 30)$ for both the univariate and bivariate cases.
To find the approximate VAR order for a VARMA class process, the multivariate extension of the method used by Poskitt & Tremayne (1987) is adopted. It involves non-linear minimization of the determinant of innovation variance-covariance matrix with respect to approximating VAR parameters, and provides a guideline of the approximate VAR order to be selected by order selection criteria over a range of small and moderate sample sizes. As a simple example, we consider the case of VAR(1) approximation to a VMA(1) model, which can be written as

\[ x_t = M_1 u_{t-1} + u_t = (I + M_1 B) u_t \approx (I - A_1 B)^{-1} \epsilon_t, \]

where \( M_1 \) and \( A_1 \) are coefficient matrices, \( u_t \sim i.i.d.(0, \Sigma_u) \) and \( \epsilon_t \sim i.i.d.(0, \Sigma_\epsilon) \). It can easily be seen that

\[ \Sigma_\epsilon \approx \Sigma_u + (M_1 - A_1) \Sigma_u (M_1 - A_1)' + MA_1 \Sigma (MA_1)'. \]

Given that \( M_1 \) is known, the VAR coefficient matrix \( A_1 \), and the parameters therein, are chosen so that \( \text{det}(\Sigma_\epsilon) \) is minimised. Let the minimised value be \( \text{det}(\tilde{\Sigma}_\epsilon(1)) \). This procedure can be repeated for the case of higher order VAR approximations to yield \( \text{det}(\tilde{\Sigma}_\epsilon(m)), m = 0, 1, 2, \ldots, M \), with \( M \) being a reasonably large integer. Analogous to the univariate AR case of Poskitt & Tremayne (1987), a potential approximate VAR order, when sample size is small or moderate, can be detected by examining the values \( \text{det}(\tilde{\Sigma}_\epsilon(m)) \). The VAR order which yields \( \text{det}(\tilde{\Sigma}_\epsilon(m)) \) reduced substantially from \( \text{det}(\tilde{\Sigma}_\epsilon(m-1)) \) and sufficiently close to \( \text{det}(\Sigma_u) \) can be a good candidate for the approximate VAR order in small and moderate samples.
Table 6.3 reports the determinants of error variance-covariance matrix (error variances) of (V)AR approximating models to the (V)ARMA class models M6, M7 and M14. It can be seen that, for models M6 and M7, reductions in error variances are relatively large at AR orders of one and two. However, for both models, variances at AR order 1 seem still too large relative to the innovation variance which is set to 1. This suggests that the AR(2) model is most likely to be selected by an order selection criterion for both models over a wide range of small and moderate samples. By similar reasoning, a VAR(2) model seems to be a close candidate for VMA(1) model M14.

6.3 Result and discussions

It is found in this chapter that the bootstrap method can lead to substantial improvements in VAR order selection in small samples. Empirical frequencies of bootstrap orders are in general more concentrated at the true order than those of the asymptotic orders. An example can be found in Figure 6.1 where AIC is used for AR(2) model M3 with sample size 60. It can be seen that asymptotic orders are massed at the true order with 42% frequency and exhibit a tendency to over-estimate. If bootstrap order selection is used, this tendency disappears markedly, and empirical frequencies become more concentrated at the true order. The order selection based on the MBB and SB respectively provide about 63% and 52% frequencies at the true order. Note that the average block lengths used for both of MBB and SB procedures are 5. Figure 6.2 exhibits the case of BIC when VAR(2) model M8 is simulated with sample size 60. Asymptotic orders are located at the true order with 58% frequency, but now there is a tendency to
under-estimate. The MBB and SB procedures, with average block lengths 10 employed for both, alleviate this tendency and provide empirical frequencies more concentrated at the true order. Note that order selections based on the MBB and SB achieve more than 80% at the true order. Similar results can be obtained if HQ is considered. However, it should be noted that the choice of block length is crucial in obtaining superior performance of bootstrap order selection. Bootstrap order selection is found to perform better than the conventional one only when a range of reasonable choices of block length is made.

Although we have seen that bootstrap orders can outperform asymptotic orders in small samples, there are two important issues to be examined. Firstly, it is of interest to compare small sample performances of the MBB and SB order selection methods. The two methods are closely related to each other and provide asymptotically equivalent order selection. It is, therefore, of interest to find out which method performs better in small samples in the context of VAR order selection. By doing this, we can assess the contribution of circular treatment of data and random block lengths adopted in SB order selection. Secondly, the choice of block length should carefully be considered because, as we shall see later, it affects the outcomes of bootstrap order selection to a great extent. An analytical approach to the optimal block lengths, as in recent work by Hall et al. (1995), would be difficult in the context of VAR order selection. It is hoped that simulation results obtained in this paper can provide a guideline as to what range of block lengths is reasonable in practical applications of bootstrap order selection.

We begin by comparing small sample performances of MBB and SB order
selections. It is found that MBB and SB order selections exhibit similar performances throughout, but the former in general outperforms the latter in selecting the true order to a degree. Examples can be found in Figures 6.1 and 6.2 where the MBB method performs better than the SB with higher frequencies of choosing the true order. It may be the case that blocks with random length are less effective in replicating the dependence structure of the underlying process because they add more uncertainty to bootstrap procedures. The circular treatment of data is related only to the end effect in obtaining a bootstrap sample and may have little impact on the overall performance of the bootstrap. The inferior performance of SB order selection may also be attributed to the nature of the geometric distribution. It is an asymmetric distribution skewed severely to the right. This means that blocks shorter than the mean are realized more often than those longer. This property may not be suitable for AR order selection, because features of high order dependency may not effectively be captured when artificial data sets are produced by resampling. It seems that a symmetric discrete random variable may be more suitable in dealing with VAR models as it gives equal weights in assigning randomness to each side of the mean. Consider the zero-truncated Poisson distribution written as

\[
\text{Prob}(h = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 1, 2, \ldots,
\]

where \( \lambda \) is a positive parameter. For reasonably large value of \( \lambda \), this distribution is nearly symmetric around the mean \( \lambda \), although slightly skewed to the right. Moreover, its variance \( \lambda \) is substantially smaller than that of the geometric distribution with the same mean. As mentioned in Chapter 2, the mean and variance of the geometric distribution given in
(2.15) are respectively $1/r$ and $(1-r)/r^2$; and it can be seen that the geometric distribution with the mean $\lambda$ has the variance $\lambda(\lambda-1)$. Another variant of the SB method can be obtained by allowing only for the circular treatment of data in the MBB. This case is also of interest as it enables us to evaluate the contribution of the circular treatment of data in small samples.

Simulations are conducted to compare small sample performances of the MBB, SB, MBB with circularity (MBBC) and SB with the zero-truncated Poisson distribution (SBP). These alternatives are found to perform similarly in small samples. However, the tendency for MBB order selection to perform better than the others is evident in most cases. An example can be found in Figure 6.3 where these procedures are compared when HQ is used for model M9 with sample size 60. The above bootstrap alternatives exhibit similar performances although they all exhibit empirical frequencies which perform far better than those of asymptotic orders. It can also be seen that MBB order selection outperforms all other alternatives. This indicates that the circular treatment of data and random generation of block length seem to exert no beneficial impact on small sample performances of bootstrap order selection. On this basis, preferences may be given to MBB order selection, and all comparisons below are made by using the results obtained with MBB order selection.

From this point on, special attention is paid to the choice of block length and its impact on small sample performances of bootstrap order selection. In general, the result of bootstrap order selection is found to react sensitively to the choice of block length. From Figure 6.4, it can be seen that, when the sample size is 30 for AR(2) model M4, the bootstrap
with AIC improves asymptotic orders with block length 5, but its performance worsens if a higher block length 10 is used. In Figure 6.5 where BIC is employed for VAR(1) model M11 with sample size 100, the bootstrap requires block length 10 for the most improved result, but its performance again worsens if a higher block length 15 is used. The case of HQ appears similar. This result indicates that the choice of block length is crucial to obtain bootstrap order selection which improves performance of the conventional order selection. A close examination will be given below to find reasonable choices of block length in the context of (V)AR order selection in small samples.

Before we look at specific examples, a summary of the findings is given. To improve small sample performances of the asymptotic VAR order selection, the bootstrap with AIC requires block lengths 5 to 10 in most cases, regardless of sample size. On the other hand, the bootstraps with BIC and HQ generally require longer blocks than AIC. Moreover, the block lengths required for BIC and HQ increase with sample size, with a tendency of those of BIC to be slightly longer than those of HQ. That is, when the sample size is 30, they require block length 5 or 10, but with sample size 100, block lengths as long as 20 or 30 are often required. The length of memory of the underlying process is also found to be an important factor, as expected, especially for BIC and HQ. These criteria require longer blocks when the underlying process possesses longer memory. However, it is found that the block lengths required for AIC are insensitive to the length of the memory: even with a longer memory process, AIC still performs well with block length of 10 in most cases.

Table 6.4 presents a summary of the results obtained for AR(2) model
M3. The bootstrap results associated with the block lengths which provide the most improved performances are reported throughout. In all cases, it is evident that substantial improvements have been made in the ability to select the true order by employing the bootstrap order selection. The descriptive statistics reported indicate that bootstrap orders are more accurate (mean, median and mode closer to the true order); and less variable (smaller standard errors). Moreover, bootstrap orders exhibit empirical frequencies at the true order substantially higher than those of asymptotic orders. For AIC, it can be seen that block length 5 is required for all sample sizes considered; evidence that the block lengths required for AIC are insensitive to changes in sample size. For BIC and HQ, block lengths 10 to 20 are required in most cases, with those for BIC being more sensitive to the changes in sample size than those of the HQ. Similar results are obtained in the VAR case. An example can be found in Table 6.5 where descriptive statistics for empirical frequencies of asymptotic and bootstrap orders for VAR(2) model M9 are presented. As in the univariate case, superior performance of the bootstrap is evident. It can be seen that, corresponding to sample sizes (30, 60, 100), AIC requires block lengths (5, 5, 10); while HQ and BIC respectively require (5, 10, 15) and (10, 20, 30).

For AR models with longer memory structure, similar results as above are obtained. It is found that the block lengths required for BIC and HQ are slightly longer than those above and that those for AIC are found to be insensitive to the length of memory. An example can be found in Table 6.6 where descriptive statistics for the case of AR(3) model M5 are given. It can be seen that the bootstrap substantially improves small sample performances of AR order selection. Corresponding to sample sizes (30, 60,
100), AIC requires block lengths (5, 10, 10); while HQ and BIC respectively (10, 15, 20) and (15, 20, 30). For longer memory VAR process M13, similar results to those found in the univariate case are obtained. The case of AIC is reported in Figure 6.6 for sample sizes 60 and 100. It can be seen that AIC requires block length 10 regardless of sample size for improved small sample performances of VAR order selection. Although not reported in the figure, AIC requires block length 5 for sample size 30. As in the univariate case, BIC and HQ require longer block lengths than AIC, which react more sensitively to changes in sample sizes when a long memory is present in the underlying process. Hence, it is observed that block length should be chosen by carefully considering the sample size and length of memory, especially when BIC and HQ are used. In practice, AIC may be preferred as it entails the choice of a small range of block lengths which are virtually insensitive to changes in sample size and length of memory.

Similar findings are obtained in the case of non-existent (V)AR order. That is, bootstrap orders are more concentrated at the approximate (V)AR order than asymptotic orders. An example can be found in Figure 6.7 where results obtained by using AIC and BIC are reported for VMA(1) model M14 when the sample size is 100. It can be seen that bootstrap orders provide higher frequencies at the approximate VAR order 2 than the asymptotic orders. The case of HQ appears similar. In Table 6.7, more detailed results for the MA(1) model M6 are presented. As in the case of the existent true VAR order, AIC requires block length 5 for all sample sizes considered while BIC and HQ sometimes require longer block lengths as sample size increases. Since MA processes possess short memories, it seems that longer blocks required for BIC and HQ here attribute only to increasing sample size. The case of ARMA(1,1) model M7 exhibits similar results. Since an MA
process has shorter memory than an AR, it seems that longer blocks required for HQ and BIC for the MA(1) case attribute only to increasing sample size.

An interesting feature found is that block lengths required for AIC are least sensitive and those for BIC are most sensitive to changes in sample size and memory of the underlying process. The sensitivity of block length required for HQ stands in between those of AIC and BIC. This may have a close relationship with the penalty terms for the number parameters to be fitted which are different for each criterion. In their simulation study on the order selection of AR models, Bhansali & Downham (1977) found the following results: if dependence of \( y_t \) on \( y_{t-p} \) is strong, the frequency of choosing the true order increases by employing the BIC or HQ type criteria; but if the dependence is weak, the frequency may not necessarily increase by employing BIC or HQ. Note that a longer block length can yield a bootstrap sample whose pth order sample autocorrelation is higher if its population value is non-zero. Hence, the findings of Bhansali and Downham (1977) in part explain the sensitivity of bootstrap order selection to block length found here. For AR(p) process, by employing longer blocks, high order dependencies including the pth order can get stronger in bootstrap samples and the bootstrap results based on BIC and HQ will improve asymptotic order selection. However, further investigation is called for to provide a more concrete theoretical basis for the insensitiveness of bootstrap order selection results based on AIC to sample size and length of memory of the underlying process.

Now attention is paid to the choice of bootstrap sample size \( n^* \) and its effect on performance of bootstrap order selection. The value so far used is 200 as it was found to be reasonable in preliminary simulations.
Simulations are conducted with a number of \( n^* \) values ranging from 200 to 1000 with the addition of 100. It is found that performances of bootstrap order selection improve with increasing value of \( n^* \) until it reaches a turning point, after which its performance slightly deteriorates. Examples for AR(2) model M3 are presented in Figures 6.8 and 6.9 when sample size is 60. It can be seen from Figure 6.8 that, when AIC is used, performance of bootstrap order selection slightly deteriorates when \( n^* \) is greater than 700. Similar observation for the case of HQ can be made from Figure 6.9 where deterioration begins when \( n^* \) is 1000. This observation is in accordance with the conjecture made in chapter 2 that the information content in the observed data is finite. The turning point observed may be thought of as the point where the information contained in the original data is exhausted by repeated resampling and no further information can be extracted. The deterioration may occur as a result of increased sampling variability brought about by excessive resampling which adds no extra information. The general finding is that improvement in order selection performance can be made when \( n^* \) is in the range of 200 to 500. Hence, the choice of \( n^* \) within the range of 200 to 500 seems to be reasonable in practical applications of the bootstrap order selection methods.

6.4 Applications to economic data

In this section, the bootstrap order selection method is applied to two sets of economic data. The first data set consists of 70 observations of yields from batch chemical process analyzed by Box et al. (1994). Judging from the time plots and autocorrelations reported in Box et al. (1994), the process is stationary and possesses a memory structure which

208
cannot be regarded as being long. They examined the sample autocorrelations and other diagnostics to conclude that the preferred AR order is 2. Hence, it is highly likely that the true AR order of the process is 2. Let \( \hat{p}_A \), \( \hat{p}_B \), and \( \hat{p}_H \) denote the asymptotic orders from the data when AIC, BIC and HQ are used respectively. It is found here that \( (\hat{p}_A, \hat{p}_B, \hat{p}_H) = (3, 1, 3) \). These orders make no clear agreement among themselves and also differ from the order preferred by Box et al. (1994).

The second is a bivariate data set with 98 observations consisting of weekly production schedule and billing figures, analyzed by Reinsel (1993). By employing the likelihood ratio test, Reinsel (1993, p. 52) concluded the VAR(3) as a preferred model. The asymptotic orders selected by order selection criteria here are \( (\hat{p}_A, \hat{p}_B, \hat{p}_H) = (3, 3, 3) \), agreeing with the order selected by Reinsel (1993). This indicates that the true VAR order is highly likely to be 3. Sample autocorrelations and cross-correlations for the data set reported in Reinsel (1993; p. 51) suggest that the process possesses memory structure which may be regarded as long. In fact, the time plot of billing figures exhibits a mild upward trend, suggesting that it may be a near non-stationary model. The moduli of eigenvalues calculated from VAR(3) parameter estimates reported in Reinsel (1993) are \( (1.125, 1.23, 1.72, 1.99) \), one of which is fairly close to the unit circle. Hence, in light of the simulation findings obtained in the previous section, care should be taken in specifying block lengths for this data set, especially when BIC and HQ are used.

In conducting bootstrap order selection in this section, block lengths employed are \( (5, 10, 15, 20, 30) \) with the bootstrap sample size and number of bootstrap iterations being set respectively to 200 and 100. Other
bootstrap sample sizes ranging from 300 to 500 are also simulated to obtain similar results. The bootstrap results for the first data set are reported in Figure 6.10. It can be seen that, for all criteria considered, the bootstrap orders are 2, which is the order preferred by Box et al. (1994). In fact, bootstrap orders are found to be 2 for a wide range of block lengths including those reported in Figure 6.10. They achieve mode at 2 with substantially high frequencies when AIC is used with block lengths 5 or 10 and when BIC and HQ are used with block length 15 or 20. These block lengths are compatible with those found to be reasonable in simulation experiments reported in the previous section. Note that the asymptotic orders for this data set mentioned above are not equal to 2. Hence, we have seen that bootstrap order selection provides a valuable alternative to the asymptotic orders which may exhibit unsatisfactory performance.

The results for the second data set are reported in Figure 6.11. It can be seen that, for all criteria considered, the bootstrap orders are 3, which agree with the order preferred by Reinsel (1993) and the asymptotic orders mentioned earlier. Although Figure 6.11 reports the results associated only with the selected block lengths, the bootstrap order 3 is obtained with a wide range of block lengths. In particular, the mode at 3 is achieved with substantially high frequencies when AIC is used with block length 10 and when BIC and HQ are used with block length 15 or 20. These block lengths are also compatible with those found to be reasonable for longer memory processes in the previous section. Hence, it is seen that bootstrap order selection provides an additional piece of information which agrees with the AR order found to be adequate by an alternative statistical method.
6.5 Concluding remarks

The bootstrap method of AR order selection is proposed and its small sample performances are evaluated. Simulation results suggest that the bootstrap method can improve small sample performances of the asymptotic method to a great extent. It is found that empirical frequencies of bootstrap orders are in general more concentrated at the true order than those of asymptotic orders. The case of the non-existent true order is also examined and a similar conclusion can be drawn. The bootstrap order selection method is also applied to economic data and found to provide a valuable objective alternative to other conventional order selection methods.

The implementation of bootstrap order selection requires the choice of block length which is found to be crucial in its small sample performances. The sample size and length of memory of the underlying process are identified as important factors that should be considered in choosing the block length. BIC and HQ are found to require block lengths which react sensitively to these factors, while the block lengths required for AIC are found to be virtually insensitive. In general, AIC requires block lengths 5 to 10. As for the bootstrap sample size, the choice in the range of 200 to 500 was found to be reasonable when sample size is less than 100. A practitioner may find it useful to employ AIC because it requires a smaller range of block lengths for improved order selection, which are insensitive to sample size and length of memory of the underlying process.
### Appendix 6.1. Tables

#### Table 6.1. Design of Univariate Models

<table>
<thead>
<tr>
<th>Models</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 [ y_t = 1 + 0.4 y_{t-1} + u_t ]</td>
<td>2.5</td>
</tr>
<tr>
<td>M2 [ y_t = 1 - 0.8 y_{t-1} + u_t ]</td>
<td>-1.25</td>
</tr>
<tr>
<td>M3 [ y_t = 1 + 0.9 y_{t-1} - 0.2 y_{t-2} + u_t ]</td>
<td>2.5, 2</td>
</tr>
<tr>
<td>M4 [ y_t = 1 - 0.5 y_{t-1} - 0.125 y_{t-2} + u_t ]</td>
<td>-2 ± 2i</td>
</tr>
<tr>
<td>M5 [ y_t = 1 + 0.5 y_{t-1} + 0.56 y_{t-2} - 0.18 y_{t-3} + u_t ]</td>
<td>1.11, -1.45, 3.45</td>
</tr>
<tr>
<td>M6 [ y_t = 1 + 0.6 u_{t-1} + u_t ]</td>
<td>1.67</td>
</tr>
<tr>
<td>M7 [ y_t = 1 + 0.5 y_{t-1} + 0.4 u_{t-1} + u_t ]</td>
<td>2(AR), -2.5(MA)</td>
</tr>
</tbody>
</table>

#### Table 6.2. Design of Bivariate Models

<table>
<thead>
<tr>
<th>VAR(2)</th>
<th>A₁</th>
<th>A₂</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>[ 0.5 -0.1 ]</td>
<td>[ -0.3 0 ]</td>
<td>-2.72±1.95i, 0.72±1.56i</td>
</tr>
<tr>
<td></td>
<td>0.4 -0.5</td>
<td>0.2 -0.1</td>
<td></td>
</tr>
<tr>
<td>M9</td>
<td>[ 0.9 0.6 ]</td>
<td>[ -0.2 0.1 ]</td>
<td>-2 ± 2i, 2, 2.25</td>
</tr>
<tr>
<td></td>
<td>0 -0.5</td>
<td>0 -0.125</td>
<td></td>
</tr>
<tr>
<td>M10</td>
<td>[ 0.6 -0.5 ]</td>
<td>[ 0 0.2 ]</td>
<td>-2.87, -2.06, 1.37, 2.45</td>
</tr>
<tr>
<td></td>
<td>0 -0.3</td>
<td>-0.25 0.3</td>
<td></td>
</tr>
<tr>
<td>M11</td>
<td>[ 1.2 0.6 ]</td>
<td>[ -0.2 0 ]</td>
<td>1±1.22i, ±2</td>
</tr>
<tr>
<td></td>
<td>-1.55 -0.4</td>
<td>0.8 0.5</td>
<td></td>
</tr>
<tr>
<td>M12</td>
<td>[ 0.9 -1.2 ]</td>
<td>[ 0.4 -0.3 ]</td>
<td>-2.5, ±2. 1.25</td>
</tr>
<tr>
<td></td>
<td>0.4 -0.5</td>
<td>0 0.2</td>
<td></td>
</tr>
<tr>
<td>M13</td>
<td>[ 1.55 -0.3 ]</td>
<td>[ -0.57 0.1 ]</td>
<td>-2, 1.05, 1.11, 1.67</td>
</tr>
<tr>
<td></td>
<td>0 0.4</td>
<td>0 0.45</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VMA(1)</th>
<th>Coefficient Matrix</th>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>M14</td>
<td>[ 0.6 0 ]</td>
<td>2.5, 1.67</td>
</tr>
<tr>
<td></td>
<td>1.2 0.4</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.3. Error variances of AR approximating models

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>M6</td>
<td>1.36000</td>
<td>1.09500</td>
<td>1.03132</td>
<td>1.01093</td>
<td>1.00389</td>
</tr>
<tr>
<td>M7</td>
<td>2.08000</td>
<td>1.10461</td>
<td>1.01983</td>
<td>1.01660</td>
<td>1.01169</td>
</tr>
<tr>
<td>M14</td>
<td>2.39320</td>
<td>1.19534</td>
<td>0.91100</td>
<td>0.81272</td>
<td>0.77462</td>
</tr>
</tbody>
</table>

Note. For models M6 and M7, $\sigma^2 = 1$; for model M14, $\text{vech}(\Sigma_u) = (1, 0.5, 1)'$. 
<table>
<thead>
<tr>
<th>Mode</th>
<th>n = 30</th>
<th></th>
<th>n = 60</th>
<th></th>
<th>n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mode</td>
<td>median</td>
<td>mean</td>
<td>std</td>
<td>f</td>
</tr>
<tr>
<td>AIC</td>
<td>1</td>
<td>2</td>
<td>1.90</td>
<td>1.23</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>B(h=5)</td>
<td>2</td>
<td>1.78</td>
<td>0.52</td>
<td>62</td>
</tr>
<tr>
<td>BIC</td>
<td>1</td>
<td>1</td>
<td>1.26</td>
<td>0.61</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>B(h=10)</td>
<td>2</td>
<td>1.86</td>
<td>0.89</td>
<td>41</td>
</tr>
<tr>
<td>HQ</td>
<td>1</td>
<td>1</td>
<td>1.60</td>
<td>0.94</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>B(h=5)</td>
<td>2</td>
<td>1.58</td>
<td>0.53</td>
<td>54</td>
</tr>
<tr>
<td>AIC</td>
<td>2</td>
<td>2</td>
<td>2.31</td>
<td>1.35</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>B(h=5)</td>
<td>2</td>
<td>1.66</td>
<td>0.49</td>
<td>64</td>
</tr>
<tr>
<td>BIC</td>
<td>1</td>
<td>1</td>
<td>1.49</td>
<td>0.74</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>B(h=15)</td>
<td>2</td>
<td>1.76</td>
<td>0.81</td>
<td>49</td>
</tr>
<tr>
<td>HQ</td>
<td>1</td>
<td>2</td>
<td>1.80</td>
<td>1.03</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>B(h=10)</td>
<td>2</td>
<td>1.78</td>
<td>0.72</td>
<td>50</td>
</tr>
</tbody>
</table>

Note. A: Asymptotic Orders; B: Bootstrap Orders with block length h; f: frequencies at the true order.
Table 6.5. Empirical Frequencies of Asymptotic and Bootstrap Orders  
(Model M9, p = 2)

<table>
<thead>
<tr>
<th></th>
<th>n = 30</th>
<th></th>
<th>n = 60</th>
<th></th>
<th>n = 100</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mode</td>
<td>median</td>
<td>mean</td>
<td>std</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>2.42</td>
<td>1.48</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>B(h=5)</td>
<td>2</td>
<td>2</td>
<td>1.90</td>
<td>0.41</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1.29</td>
<td>0.45</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>B(h=10)</td>
<td>2</td>
<td>2</td>
<td>1.79</td>
<td>0.55</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>HQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>1.71</td>
<td>0.83</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>B(h=5)</td>
<td>2</td>
<td>2</td>
<td>1.60</td>
<td>0.49</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

|        | n = 60 |        | n = 100 |        |        |
|        | mode | median | mean  | std   | f      |
| AIC    |      |        |       |       |        |
| A      | 2    | 2      | 2.17  | 0.83  | 76     |
| B(h=5) | 2    | 2      | 1.87  | 0.36  | 85     |
| BIC    |      |        |       |       |        |
| A      | 2    | 2      | 1.53  | 0.50  | 53     |
| B(h=20)| 2    | 2      | 1.93  | 0.65  | 74     |
| HQ     |      |        |       |       |        |
| A      | 2    | 2      | 1.73  | 0.49  | 69     |
| B(h=10)| 2    | 2      | 1.89  | 0.44  | 82     |

|        | n = 100 |        |        |        |
|        | mode | median | mean  | std   | f      |
| AIC    |      |        |       |       |        |
| A      | 2    | 2      | 2.22  | 0.73  | 85     |
| B(h=10)| 2    | 2      | 2.10  | 0.36  | 92     |
| BIC    |      |        |       |       |        |
| A      | 2    | 2      | 1.69  | 0.46  | 69     |
| B(h=30)| 2    | 2      | 1.87  | 0.42  | 81     |
| HQ     |      |        |       |       |        |
| A      | 2    | 2      | 1.88  | 0.41  | 85     |
| B(h=15)| 2    | 2      | 1.95  | 0.30  | 91     |

Note. A: Asymptotic Orders; B: Bootstrap Orders with block length h;  
f: frequencies at the true order
<table>
<thead>
<tr>
<th></th>
<th>mode</th>
<th>median</th>
<th>mean</th>
<th>std</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>2</td>
<td>2</td>
<td>2.68</td>
<td>1.31</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.54</td>
<td>1.16</td>
<td>49</td>
</tr>
<tr>
<td>BIC</td>
<td>2</td>
<td>2</td>
<td>1.78</td>
<td>0.91</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.95</td>
<td>1.22</td>
<td>39</td>
</tr>
<tr>
<td>HQ</td>
<td>2</td>
<td>2</td>
<td>2.22</td>
<td>1.18</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.92</td>
<td>1.11</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 60</th>
<th>mode</th>
<th>median</th>
<th>mean</th>
<th>std</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>2</td>
<td>3</td>
<td>3.12</td>
<td>1.22</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.96</td>
<td>0.75</td>
<td>67</td>
</tr>
<tr>
<td>BIC</td>
<td>2</td>
<td>2</td>
<td>2.22</td>
<td>0.67</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.69</td>
<td>1.00</td>
<td>45</td>
</tr>
<tr>
<td>HQ</td>
<td>2</td>
<td>2</td>
<td>2.70</td>
<td>0.98</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.98</td>
<td>0.91</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 100</th>
<th>mode</th>
<th>median</th>
<th>mean</th>
<th>std</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>3</td>
<td>3</td>
<td>3.22</td>
<td>1.06</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.90</td>
<td>0.46</td>
<td>84</td>
</tr>
<tr>
<td>BIC</td>
<td>2</td>
<td>2</td>
<td>2.50</td>
<td>0.64</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.59</td>
<td>0.62</td>
<td>57</td>
</tr>
<tr>
<td>HQ</td>
<td>3</td>
<td>3</td>
<td>2.73</td>
<td>0.76</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2.82</td>
<td>0.54</td>
<td>71</td>
</tr>
</tbody>
</table>

Note. A: Asymptotic Orders. B: Bootstrap Orders with block length h.

f: frequencies at the true order
Table 6.7. Empirical Frequencies of Asymptotic and Bootstrap Orders.  
(Model M6, $p^* = 2$)

|      | AIC |  
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|      | mode| median | mean  | std  | f   |
| n = 30 |     |       |       |      |     |
| AIC | 1   | 2     | 2.20  | 1.55 | 29  |
|     | 2   | 2     | 2.01  | 0.69 | 65  |
| B(h=5) |     |       |       |      |     |
| BIC | 1   | 1     | 1.16  | 0.92 | 27  |
|     | 2   | 2     | 1.46  | 0.67 | 53  |
| HQ  | 1   | 2     | 1.74  | 1.24 | 33  |
|     | 2   | 2     | 1.75  | 0.57 | 64  |
| n = 60 |     |       |       |      |     |
| AIC | 2   | 2     | 2.81  | 1.35 | 43  |
|     | 2   | 2     | 1.87  | 0.56 | 73  |
| B(h=5) |     |       |       |      |     |
| BIC | 2   | 2     | 1.77  | 0.82 | 51  |
|     | 2   | 2     | 1.90  | 0.77 | 61  |
| HQ  | 2   | 2     | 2.08  | 1.03 | 51  |
|     | 2   | 2     | 1.71  | 0.48 | 69  |
| n = 100 |     |       |       |      |     |
| AIC | 2   | 3     | 3.18  | 1.35 | 38  |
|     | 2   | 2     | 1.88  | 0.45 | 78  |
| B(h=5) |     |       |       |      |     |
| BIC | 2   | 2     | 1.88  | 0.71 | 60  |
|     | 2   | 2     | 1.97  | 0.71 | 67  |
| HQ  | 2   | 2     | 2.50  | 1.12 | 52  |
|     | 2   | 2     | 2.18  | 0.75 | 70  |

Note. A: Asymptotic Orders; B: Bootstrap Orders with block length h;  
f: frequencies at the true order. $p^*$: approximate AR order
Figure 6.1. Frequencies of AR Orders: AIC; M3(p=2); n=60.

Figure 6.2. Frequencies of AR Orders: BIC; M8(p=2); n=60.
Figure 6.3. Frequencies of VAR Orders: HQ; M8(p=2); n=60.

Figure 6.4. Frequencies of AR Orders: AIC; M4(p=2); n=30.
Figure 6.5. Frequencies of VAR Orders: BIC; M11(p=1); n=100.

Figure 6.6. Frequencies of VAR Orders: AIC; M13(p=2).
Figure 6.7. Frequencies of VAR Orders: M14(p=2); n=100.

Figure 6.8. Frequencies of AR Orders: AIC; M3(p=2); n=60.
Figure 6.9. Frequencies of AR Orders: HQ; M3(p=2); n=60.

Figure 6.10. Frequencies of Bootstrap AR Orders: Data Set 1.
Figure 6.11. Frequencies of Bootstrap VAR Orders: Data Set 2.
Chapter 7. Conclusion

This thesis extended the use of the bootstrap method to VAR modelling in small samples. The comparison of small sample performances of asymptotic and bootstrap methods in the context of parameter estimation, forecasting and order selection has been the major focus. In all, the bootstrap has proven to be a useful tool for estimating small sample properties of VAR models. In most cases, the bootstrap performed more reliably than conventional asymptotic methodology; however, our analysis did uncover cases where this proved otherwise. The attraction of the bootstrap lies, of course, in the former case, where its use is strongly recommended to practitioners. In the latter case, the usefulness of the bootstrap is confined to a benchmark role in evaluating reliability of asymptotic formulae in small samples.

Bias is a small sample property of a statistic that cannot be estimated by conventional first-order asymptotic approximation. In practice, where first order asymptotic approximation is widely used, the presence of biases in parameter estimators and forecasts may be totally overlooked. In Chapter 3, the bootstrap method was used to examine the major determinants of small sample biases of VAR parameter estimators. Response surfaces for biases of VAR(1) parameter estimators were constructed and it was shown how they provided a useful summary of how biases of VAR(1) parameter estimators are related to the parameters of the model. Simulations were conducted in Chapter 4 to evaluate small sample properties of VAR forecast biases. The major finding here was that VAR parameter estimators and forecasts can be severely biased in small samples.
The model structure, as characterised by VAR coefficient matrices, was found to be the major determinant of biases. Notably, as the model tended towards non-stationarity or as the correlation among VAR components increased, biases in VAR parameter estimators and forecasts also increased. It is strongly recommended in practice to adopt the bootstrap method as a means of estimating small sample biases. Ignoring the presence of substantial biases may lead to misleading forecasts and statistical inferences. This applies not only to the VAR case, but also to other cases where parameter estimators or forecasts are known or suspected to be biased in small samples.

The variability of VAR forecasts are typically measured by using asymptotic formulae. Their small sample reliability and robustness to non-normality are compared to those of their bootstrap counterparts. In Chapter 4, it was found that asymptotic standard errors perform reasonably well under normal and a range of non-normal innovations. Similar results were found in Chapter 5 where the asymptotic PR's provide a more accurate assessment of future uncertainty in small samples than their bootstrap counterparts. Hence, it is suggested for the case of VAR forecasting that asymptotic formulae be preferred by practitioners. Efficient resampling methods, see Hall (1992, Appendix II), may improve the performance of the bootstrap in the context of VAR forecasting, and is the subject of future research. Construction of the bootstrap PR's based on Masarotto's (1990) method may also be of interest for future research, as those in this thesis were constructed based on Thombs & Schucany's (1990).

In Chapter 6, the MBB and SB methods were applied to VAR order selection using AIC, BIC and HQ. Attention was focused upon whether and
when the bootstrap method improves small sample performances of VAR order selection in choosing the true order, when it exists, or approximate VAR order, otherwise. Simulation results revealed that the bootstrap method improved small sample performances of AIC, BIC and HQ to a great extent. However, care should be taken as these improvements are contingent upon reasonable choices of (average) block length and bootstrap sample sizes. Although guidelines for such choice, obtained from simulation experiments, were provided, future research will need to be directed towards a theoretical derivation of the optimal block length in the context of VAR order selection. The MBB and SB methods can be used in other areas where asymptotic methods exhibit unsatisfactory small sample performances. For example, bootstrap order selection based on the MBB and SB can be applied to univariate ARMA models as an alternative to the order selection method proposed by Hannan and Rissanen (1982).

To be a useful alternative to asymptotic methods in small samples, the bootstrap should be asymptotically valid. In this thesis, the asymptotic validity of the bootstrap in the context of VAR parameter estimation, forecasting and order selection was established. This thesis also provided details of stochastic properties of backward stationary VAR models. The LS parameter estimators of backward stationary VAR and their asymptotic properties were also derived. These findings can be used in other applications where the conditionality of a multivariate stationary process on past observations plays an important role.

It should be noted that the range of econometric and time series models to which the bootstrap method can be applied is somewhat restricted. As we have seen in Chapter 1, from the work of Freedman (1984), the model
should satisfy (or be a special case of) equation (1.6) and several regularity conditions to ensure asymptotic validity of the bootstrap. Equation (1.6) is a linear simultaneous equations model with a dynamic structure whose parameters can be estimated by, for example, the 2SLS method. However, an econometric model can sometimes possess a special structure in which an iterative algorithm is required for parameter estimation. These special structures include non-linearity in parameters (see, for example, Chapter 2 of Davidson & MacKinnon, 1993) and limited dependent variables (Maddala, 1983) among others. Under these circumstances, to the best of my knowledge, there is no formal and unified proof - such as that of Freedman (1984) - which shows that bootstrap statistics converge to their true values in large samples. This raises doubts about the usefulness of the bootstrap in the context of these models. Future efforts should be directed towards investigating the large sample behavior of the bootstrap in non-linear models.

In establishing the asymptotic validity of the bootstrap in equation (1.6), it is crucial that the underlying process be stationary. Since the work by Dickey & Fuller (1979) on unit roots testing and Engle & Granger (1987) on cointegration, statistical analysis of non-stationary time series data has attracted much attention. Since statistical inferences in these contexts rely heavily on asymptotic approximations, the bootstrap can play an important role in determining small sample properties. Recent work by Basawa et al. (1989, 1991a, 1991b) provide asymptotic validity of the bootstrap in both univariate non-stationary and the explosive AR(1) case. However, more investigation is needed to determine whether the bootstrap can be applied to non-stationary data with a higher order and/or higher dimension; for example, a non-stationary VAR model whose components are
cointegrated.

As we have seen in Chapter 1, many studies have found the bootstrap method to be a useful tool in investigating small sample properties for a wide range of econometric and time series models. In this thesis, we have seen that the bootstrap is also useful in the VAR case. However, it seems that the bootstrap method is yet to be widely adopted by applied economists in their empirical work. Two reasons for this may be suggested: firstly, the method is not well-known to applied economists; and secondly, there is virtually no easy-to-use computer package designed to provide bootstrap statistics for econometric and time series models. It seems that the second reason is more fundamental than the first and that the first occurs mainly as a result of the second. Hence, development of user-friendly computer packages capable of conducting the bootstrap is required. Incorporation of bootstrap routines into widely-used computer packages is strongly encouraged.
References


Chatfield, C (1993), "Calculating Interval Forecasts", *Journal of Business*


Davies, N., C.M. Triggs & P. Newbold (1977), "Significance Levels of the Box-Peirce Portmanteau Statistic in Finite Samples", *Biometrika*, 64, 517-522.


232


Johns, M.V. Jr. (1988), "Importance Sampling for Bootstrap Confidence


Liu, R.Y. & K. Singh (1992), "Moving Block Jackknife and Bootstrap Capture


