1996

Three Essays on VAR Techniques.

Omer Ozcicek
Louisiana State University and Agricultural & Mechanical College

THREE ESSAYS ON VAR TECHNIQUES

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in
The Department of Economics

by
Omer Ozcicek
B.S., Bilkent University, 1990
August 1996

UMI Number: 9706354

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my committee chairman Professor W. Douglas McMillin. I also wish to thank the other members of my committee, Professor R. Carter Hill, Professor Faik Koray and Dr. Ed Watson. I am grateful to Dr. Mark Schafer for his presence on my committee and Janet Daniel for proofreading this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

LIST OF FIGURES viii

ABSTRACT x

CHAPTER

1 INTRODUCTION 1

2 A COMPARISON OF TWO METHODS FOR SPECIFYING VAR MODELS WITH ASYMMETRIC LAG STRUCTURE 5
- 2.1 Introduction 5
- 2.2 Lag Selection Methods 10
- 2.3 Empirical Methodology 18
- 2.4 Simulation Results 23
  - 2.4.1 Lag Specification Performances 25
  - 2.4.2 The IRF Performances 43
  - 2.4.3 The Forecasting Performances 62
  - 2.4.4 The Longer Period Performance of The SIC Criterion . 72
- 2.5 Conclusion 73

3 MULTIVARIATE SHRINKAGE ESTIMATORS FOR IMPROVEMENT OF FORECASTING PERFORMANCE 80
- 3.1 Introduction 80
- 3.2 Literature Review 86
- 3.3 Theoretical Methodology 89
- 3.4 Empirical Methodology and Results 99
  - 3.4.1 Real-Time Data Analysis 99
  - 3.4.2 Monte-Carlo Experiment Results 124
- 3.5 Conclusion 131

4 ESTIMATING A VAR MODEL THAT INCLUDES DISCRETE VARIABLES 134
- 4.1 Introduction 134
- 4.2 General Methodology 140
- 4.3 Empirical Applications 153

4.3.1 Empirical Results 154
4.3.2 Two Alternative Modelling Approaches 163
4.4 Conclusion 168

5 CONCLUSION 170

BIBLIOGRAPHY 175

VITA 180

LIST OF TABLES

2.1 Descriptions of The Simulated Models 24

2.2 Model 1, Percent of Time Lag Length Selected 26

2.3 Model 2, Percent of Time Lag Length Selected 29

2.4 Model 3, Percent of Time Lag Length Selected 31

2.5 Model 4, Percent of Time Lag Length Selected 33

2.6 Model 5, Percent of Time Lag Length Selected 35

2.7 Model 6, Percent of Time Lag Length Selected 37

2.8 Model 7, Percent of Time Lag Length Selected 39

2.9 Model 8, Percent of Time Lag Length Selected 41

2.10 Model 1, IRF mse's 45

2.11 Model 2, IRF mse's 49

2.12 Model 3, IRF mse's 51

2.13 Model 4, IRF mse's 53

2.14 Model 5, IRF mse's 55

2.15 Model 6, IRF mse's 57

2.16 Model 7, IRF mse's 58

2.17 Model 8, IRF mse's 60

2.18 Model 1, Forecast mse's 63

2.19 Model 2, Forecast mse's 65

2.20 Model 3, Forecast mse's 65

2.21 Model 4, Forecast mse's 67

2.22 Model 5, Forecast mse's 67

2.23 Model 6, Forecast mse's 68

2.24 Model 7, Forecast mse's 70

2.25 Model 8, Forecast mse's 70

2.26 Summary Standardized Mean Lag Specification 74

2.27 IRF Response Summary 76

2.28 Forecast Performance Summary 78

3.1 Log Level Model, Theil U and SOFM statistics of BVAR for selected values of θ and λ for 1985:1 - 1989:4 105

3.2 Log Level Model, Theil U and SOFM statistics of BVAR for selected values of d for 1985:1 - 1989:4 107

3.3 Log Level Quarterly Data Theil U Statistics for 1990:1 - 1994:4 109

3.4 Log Level Quarterly Data, BVAR1, BVAR2 and SRVAR Sign Test Values 111

3.5 Log Level Quarterly Data, Overall Fitness Measures 113

3.6 First Difference Quarterly Data Theil U Statistics 114

3.7 First Difference Quarterly Data, BVAR1, BVAR2 and SRVAR Sign Test Values 116

3.8 First Difference Quarterly Data, Overall Fitness Measures 118

3.9 Log Level Monthly Data Theil U Statistics 120

3.10 Log Level Monthly Data, BVAR1, BVAR2 and SRVAR Sign Test Values . 121

3.11 Log Level Monthly Data, Overall Fitness Measures 123

3.12 First Difference Monthly Data Theil U Statistics 124

3.13 First Difference Monthly Data, BVAR1, BVAR2 and SRVAR Sign Test Values 125

3.14 First Difference Monthly Data, Overall Fitness Measures 127

3.15 Model 1, Relative Theil U of BVAR and SRVAR 129

3.16 Model 2, Relative Theil U of BVAR and SRVAR 131

4.1 Variance Decomposition Values From the Probit Model 162

4.2 Variance Decomposition Values From the Regular VAR Model 163

LIST OF FIGURES

2.1 Larger Sample Model 1 IRF_{11} mse's 47

2.2 Smaller Sample Model 1 IRF_{11} mse's 47

2.3 Larger Sample Model 1 IRF_{12} mse's 48

2.4 Larger Sample Model 1 IRF_{21} mse's 48

2.5 Larger Sample Model 2 IRF_{11} mse's 50

2.6 Larger Sample Model 2 IRF_{21} mse's 50

2.7 Larger Sample Model 3 IRF_{11} mse's 52

2.8 Larger Sample Model 3 IRF_{21} mse's 52

2.9 Larger Sample Model 4 IRF_{22} mse's 54

2.10 Larger Sample Model 5 IRF_{12} mse's 56

2.11 Larger Sample Model 6 IRF_{12} mse's 56

2.12 Larger Sample Model 7 IRF_{11} mse's 59

2.13 Smaller Sample Model 7 IRF_{11} mse's 59

2.14 Larger Sample Model 8 IRF_{12} mse's 61

2.15 Larger Sample Model 1 msfe of the First Variable 64

2.16 Smaller Sample Model 1 msfe of the First Variable 64

2.17 Larger Sample Model 1 msfe of the Second Variable 66

2.18 Larger Sample Model 2 msfe of the First Variable 66

2.19 Larger Sample Model 6 msfe of the First Variable 69

2.20 Larger Sample Model 8 msfe of the First Variable 71

2.21 Larger Sample Model 8 msfe of the Second Variable 71

4.1 Impulse Response of M2 From the Probit Model 156

4.2 Impulse Response of Monetary Base From the Probit Model 156

4.3 Impulse Response of NBR From the Probit Model 156

4.4 Impulse Response of FFR From the Probit Model 157

4.5 Impulse Response of 3 Month T-Bill Rate From the Probit Model 157

4.6 Impulse Response of the Spread From the Probit Model 157

4.7 Impulse Response of M2 From the Regular Model 159

4.8 Impulse Response of Monetary Base From the Regular Model 159

4.9 Impulse Response of NBR From the Regular Model 159

4.10 Impulse Response of FFR From the Regular Model 160

4.11 Impulse Response of 3 Month T-Bill Rate From the Regular Model 160

4.12 Impulse Response of the Spread From the Regular Model 160

ABSTRACT

This dissertation consists of three essays on vector autoregressive (VAR) models. The first examines lag selection criteria. Typically, in VAR models all variables have the same lag length in each equation. Hsiao (1981) and Keating (1994) suggest two ways to estimate asymmetric lag structures. In a Monte-Carlo framework, we simulate eight prespecified VAR models and observe how the lag specification methods perform. We also look at the impulse response and forecast performances of these lag selection methods. We find that the AIC criterion performs better in finite samples than the SIC criterion, and Keating's method is superior to Hsiao's method.

In the second part of the dissertation, we employ the Stein-rule estimator to estimate two VAR models, one using quarterly macroeconomic data and another one using monthly macroeconomic data. The forecasts produced by VARs estimated via Stein-rule are contrasted with the forecasts produced by VARs estimated via Bayesian methods and via ordinary least squares (OLS). In general, Bayesian VARs and Stein-rule VARs produce more accurate forecasts than OLS VARs; however, Bayesian forecasts are more expensive and difficult to obtain than either Stein-rule VARs or OLS VARs. We find that usually Stein-rule VARs perform better than Bayesian VARs.

The last part of the dissertation estimates VAR models with a discrete variable. The equation with the discrete variable as the dependent variable is estimated using the probit approach. The impulse response function (IRF) and variance decomposition (VDC) produced from VARs partly estimated via the probit approach are contrasted with

the IRF and VDC values produced from VARs estimated via the OLS technique. The results obtained from these two alternative approaches are different. However, the results from the probit approach show that the IRF and VDC values obtained with the current technique are not plausible in terms of theoretical expectations and that the modeling and estimation of a VAR that includes a discrete variable has to be further investigated.

CHAPTER 1

INTRODUCTION

In economics a common method of policy analysis and forecasting is to construct a large-scale structural economic model. In this approach, different sectors of the economy are described by different sets of equations which are derived from economic theory. However, the identification restrictions imposed by structural models are often claimed to be inappropriate, and therefore, the usage of structural models for policy analysis is criticized.

An alternative approach was introduced by Sims (1980), in which all variables are assumed to be endogenous. In Sims' approach, each variable is explained by the lagged values of all the variables in the model, thus imposing no restrictions on equations. Dynamics are introduced to the model by including higher order lags of the variables in the model. He called this approach vector autoregression (VAR).

VARs are widely used in economics for forecasting and policy analysis. A VAR consists of multiple equations in which the current value of each variable is the dependent variable of one equation and the lagged values of all the variables are the explanatory variables.

This dissertation investigates whether with new techniques we can do better policy analysis with VARs and obtain more accurate forecasts from VARs. Chapter 2 investigates a very fundamental but important issue in VAR modeling: the lag length of the VAR. Typically each variable has the same lag length in each equation.

However, theoretically and practically it does not have to be so. There exist two lag specification methods which allow asymmetry in the lag structure: Hsiao's (1981) method and Keating's (1994) method. In a Monte-Carlo experiment framework, we examine which method is best in selecting the correct lag structure, and which has the best impulse response analysis and forecasting performance. We also investigate the consequences of using a symmetric lag specification if the true lag structure is asymmetric. The symmetric lag specification methods select the lag length which minimizes Akaike's information criterion, Schwarz's information criterion, and a posterior information criterion.

Chapter 3 contrasts the forecasting performances of different methods of imposing restrictions on the coefficients of VAR models. We compare the forecasts of an interest rate, the level of GDP, and the price level. Even though VAR models are parsimonious in the number of variables, often they are not parametrically parsimonious. Overparametrization can be avoided by restricting the number of variables entering the VAR or restricting the lag length. Overparametrization may decrease the bias of the forecast; however, it may also increase the variance of the forecast. On the other hand, putting restrictions on the parameter space will lower the forecast variance; however, if the restrictions are not correct, the forecast will be biased. Thus, if the true nature of the VAR model has an asymmetric lag structure, symmetric VARs may yield poorer forecast performances than asymmetric VARs. Furthermore, typically, asymmetric lag structures require estimation of fewer

coefficients than symmetric VARs. Thus, using Hsiao's method and Keating's method to estimate lag structures may improve forecast performance.

Another possible solution to inefficient forecast estimation was proposed by Litterman (1980), who utilized non-sample information by imposing prior distributions on the model's parameters. The additional information supplied by the prior distribution of the parameters in the system yields more accurate estimates of the parameters, which are expected to improve forecast performance if the imposed prior beliefs adequately describe the true underlying model. This method is called Bayesian vector autoregression (BVAR) in the literature since it imposes a prior distribution on parameters.

In the third chapter, we propose another multivariate time-series forecasting technique, which we will call the Stein-rule VAR (SRVAR). In the Stein-rule VAR, parameter estimates are obtained via a Stein-rule estimator instead of conventional maximum likelihood (ML) or ordinary least squares (OLS) estimators. SRVAR forecasts are easier to compute than BVAR forecasts since BVAR estimation requires specification of a prior distribution and some parameters of that prior distribution. For Stein-rule estimation, the only necessary non-sample information are the restrictions that are imposed on the parameter space.

Chapter 4 examines the estimation of VARs in which there is a qualitative variable in the model. This is important since qualitative variables have been used in the monetary/macro literature to measure the stance of monetary policy, and VARs have been estimated that contain such a variable. In a single equation setting we

know that if the dependent variable is qualitative, we should use a nonlinear estimation technique. However, so far in the VAR literature, qualitative variables are inappropriately treated as ordinary variables; see Carlino and DeFina (1994) and Boschen and Mills (1995). We estimate six VAR models with the appropriate approach and the ordinary approach, and compare their impulse response and variance decomposition results.

Chapter 5 provides a summary and conclusion of the study.

CHAPTER 2

A COMPARISON OF TWO METHODS FOR SPECIFYING VAR MODELS WITH ASYMMETRIC LAG STRUCTURES

2.1 Introduction

In macroeconomics a common method of policy analysis and forecasting is to construct a large-scale structural macroeconomic model. In this approach, different sectors of the economy are described by different sets of equations which are derived from economic theory. The theory identifies which variables are endogenous or exogenous to the system and are included in a particular equation. However, the identification restrictions imposed by structural models are often claimed to be inappropriate, and therefore, the usage of structural models for policy analysis is criticized.

An alternative approach was introduced by Sims (1980), in which all variables are assumed to be endogenous. In Sims' approach, each variable is explained by the lagged values of all the variables in the model, thus imposing no restrictions on equations. Dynamics are introduced to the model by including higher order lags of the variables in the model. He called this approach vector autoregression (VAR).

VARs are widely used in macroeconomics for forecasting and policy analysis. A VAR consists of multiple equations (for each variable in the model there is only one equation), each with exactly the same set of explanatory variables. A VAR model with N variables can be written as:

$$y_t = D+\beta_1 y_{t-1}+\beta_2 y_{t-2}+...+\beta_n y_{t-n}+e_t \tag{2.1}$$

where y_t is the $N\times1$ vector of endogenous variables that are included in the system which are observed at period t, D is a $N\times1$ vector of constant terms, β_i is a $N\times N$ coefficient matrix, n is the maximum lag, and e_t is the $N\times1$ vector of error terms which are distributed as multivariate normal with covariance matrix Σ. Equation (2.1) can be viewed as the reduced form of a structural model,

$$\Phi_0 y_t = C+\Phi_1 y_{t-1}+\Phi_2 y_{t-2}+...+\Phi_n y_{t-n}+\epsilon_t \tag{2.2}$$

where Φ_0 is the contemporaneous effect matrix, ϵ_t is the $N\times1$ vector of white noise error terms which are distributed as multivariate normal with covariance matrix $\sigma^2 I$, C is a $N\times1$ vector of constant terms, and Φ_i is an $N\times N$ coefficient matrix. The VAR model is obtained by multiplying both sides of (2.2) by Φ_0^{-1}. Thus, $D=\Phi_0^{-1}C$, $\beta_i=\Phi_0^{-1}\Phi_i$ and $e_t=\Phi_0^{-1}\epsilon_t$.

VAR models are easy to construct and estimate compared to large structural models. To estimate a VAR model one has to decide which variables to include in the model and the lag length, n. Economic theory determines the variables that are included in the system. However, theory often does not indicate the appropriate lag length of the VAR, and, therefore, statistical methods have been employed to select the appropriate lag length. Sims (1980) originally used a likelihood ratio test to decide whether to use four lags or eight lags. Alternatively, researchers often pick lag lengths which minimize a criterion (discussed in the next section) or pick a lag

length in an ad hoc manner. Typically, the same lag length is used for all variables in all equations, even though there is no theoretical or statistical reason for doing so. Such a VAR will be referred as a symmetric VAR.

This essay is concerned with selection of lag lengths for VAR models. Specifying the correct lag length is important because estimated VAR models with lag lengths different from the true model are misspecified. Braun and Mittnik (1993) show that estimates of a conventional VAR which has fewer lags or more lags than the true model are inconsistent. By inconsistency we mean the estimator does not converge to the true parameter values as the sample size increases. Underspecifying the lag structure imposes incorrect restrictions on the parameter space, which leads to inconsistent VAR coefficient estimates. Overspecifying the lag structure ignores the relevant restrictions that we can impose, which again leads to inconsistent coefficient estimates.[1] Impulse response functions (IRFs) and variance decompositions (VDCs) are two common tools for policy analysis in the VAR technique. IRFs and VDCs are functions of the structural parameters. An IRF is the response of a variable to a specific structural shock. The VDC is the proportion of forecast variance explained by structural shocks to each variable.[2] In an extension of the inconsistent parameter estimation analysis, Braun and Mittnik (1993) also show that a misspecified model has inconsistent IRFs and VDCs.

[1]See Braun and Mittnik (1993, pp. 323-326) for the proof.

[2]A detailed description of IRFs and VDCs can be obtained from Lütkepohl (1993).

Lütkepohl (1993) shows that overfitting (higher order lag length than the true lag length) a VAR causes an increase in the mean-square-forecast-error statistics.[3] This is due to imprecise coefficient estimation. Lütkepohl also demonstrated that the residuals of a underfitted VAR are often autocorrelated. Parameter estimates of a model with autocorrelated innovations which are not adjusted for that autocorrelation are inefficient and forecasts are biased and inefficient.[4]

Until recently, the same lag length has typically been used for all variables in all equations. Neither economic theory nor intuition suggests that all variables should enter all equations with the same lag length. Theoretically it is possible to have different lag lengths for different variables in different equations. Thus, by estimating a symmetric VAR, we face potential misspecification problems due to improper lag structure. If variables in the true underlying model do not all have the same lag length, with a symmetric VAR estimation the estimates will be inconsistent, and it is most likely that redundant parameters will be estimated.

The first method proposed to estimate asymmetric lag structures was by Hsiao (1981). This approach treats each equation separately, i.e. the lag structure of an equation is determined independently from other equations. Keating (1994) criticized Hsiao's method, arguing that the lag structure selected by Hsiao's method is likely to

---

[3]The mean-square-forecast-error of the i^{th} variable is defined as,

$$\frac{1}{H}\sum_{h=1}^{H}\left(y_{i,t+h}-y_{i,t}(h)\right)^2$$

where $y_{i,t}(h)$ is the forecast of $y_{i,t+h}$ made at period t, and H is the total forecast horizon.

[4] Judge et al. (1988).

be biased. Keating (1994) proposed another approach to determine the lag structure of asymmetric VARs. Contrary to Hsiao's method, Keating's method does not treat the equations in the VAR separately. The lag structure estimation is carried out by jointly considering all equations and minimizing a criterion. These two method will be described in detail in the next section.

Symmetric lag selection methods have been intensively studied[5]; however, little is known about the performance of the asymmetric lag estimation methods. Keating uses his lag length estimation method in estimating the long-run structural parameters of a four-variable VAR. He imposes a set of long-run identifying restrictions to recover the structural parameters from the VAR estimates $\beta(L)$ and Σ. The long run structural parameters and structural innovations are functions of $\beta(1)$ and Σ, where $\beta(1)=\beta_1+\beta_2+\ldots+\beta_n$ where β_i's are the coefficient in (2.1). Keating's results show that his method yields smaller asymptotic standard deviations of structural parameters, IRFs and VDCs than a standard symmetric VAR estimate. The source of efficiency are the additional restrictions imposed by the asymmetric VAR. This VAR requires about half the number of parameters to be estimated that a symmetric VAR requires.

Even though Keating argues that his method reduces the standard deviations of the parameter estimates, IRFs, and VDCs, this method's ability to specify the correct lag structure is not known. Moreover, the performance of Hsiao's method in estimating the correct lag structure has not been studied. Basically, it is unknown

[5]See Lütkepohl (1993).

whether Keating's or Hsiao's method is superior in determining the true underlying lag structure. This study aims to shed light on the performance of alternative lag selection methods by employing Monte Carlo experimentation, which generates data from a known model and estimates the lag structure using each method. In this way, we will be able to determine whether Hsiao's method and Keating's method reduce the redundant parameter estimation problem. Furthermore, the IRF and forecast performance of VARs whose lag structures are determined by these methods will be investigated. The VDC performance is skipped to concentrate on the IRF performance only since IRF is more commonly used.

2.2 Lag Selection Methods

Sims (1980) used a likelihood-ratio (LR) test to determine whether to use a VAR with four or eight lags for all variables in each equation. Using LR test statistics[6], a systematic search procedure can be employed to determine the true lag length in a certain range. If a symmetric VAR's upper bound of the order of the lag, say M, is known, but the exact true order is unknown, then, a sequence of tests can be set up to determine the lag length. If $y_t=D+\beta_1 y_{t-1}+\ldots+\beta_M y_{t-M}+e_t$ is a standard VAR, the null hypothesis $H_0{:}\beta_M=0$ can be tested against the alternative $H_A{:}\beta_M\neq 0$ using a LR test with $\alpha=5$ percent significance level. If this null is not rejected, a test for $H_0{:}\beta_{M-1}=0$ against the alternative $H_A{:}\beta_{M-1}\neq 0$ is conducted. This sequence of testing

[6]The description of likelihood-ratio test can be found in Judge et al. (1988) or Lütkepohl (1993).

continues until the null of an insignificant coefficient matrix is rejected. The lag length is set to the length at which the null is rejected.

A second method for finding the lag order of a symmetric VAR is to find the lag length which minimizes a chosen criterion. This can be done by varying the lag length between 1 and M, estimating a VAR, calculating the criterion for each case, and selecting the lag length which minimizes the criterion. There are various proposed and commonly used criteria. If we assume that $e_t \sim N(0,\Sigma)$, then Akaike's information criterion (AIC)[7], Schwarz's information criterion (SIC)[8], Hannan-Quinn's Criterion (HQ)[9], and the forecast prediction error criterion (FPE)[10] are defined respectively as,

$$AIC = \ln|\tilde{\Sigma}| + \frac{2}{T}(\text{number of freely estimated parameters}) \quad ,$$

$$SIC = \ln|\tilde{\Sigma}| + \frac{\ln T}{T}(\text{number of freely estimated parameters}) \quad ,$$

$$HQ = \ln|\tilde{\Sigma}| + 2\frac{\ln \ln T}{T}(\text{number of freely estimated parameters}) \quad ,$$

and,

---

[7]Akaike (1974).

[8]Schwarz (1978).

[9]Hannan and Quinn (1979).

[10]Akaike (1969,1971).

$$FPE = \left[\frac{T+Nn+1}{T-Nn-1}\right]^{N} |\tilde{\Sigma}| \ ,$$

where n is the lag order of the VAR, N is the number of variables in the VAR, T is the sample size, and $\tilde{\Sigma}$ is the estimated covariance matrix Σ. These criteria reward low value of the determinant of residual covariance matrix and penalize increasing number of parameters.

Lütkepohl (1993, Tables 4.6 and 4.7) reports that LR statistics have a poor performance compared to the criteria minimization method in finding the true order in samples equal to or less than 100. Furthermore, his results show that SIC tends to underestimate the lag length more often than do AIC, FPE and HQ.

Recently Phillips (1994) proposed another model specification criterion called posterior information criterion (PIC). Phillips derived PIC from Bayesian analysis in which flat priors are imposed on the parameters to find the posterior distributions. The lag specification performance of PIC is still unexplored. PIC is defined as:

$$PIC = \ln|\tilde{\Sigma}| + \frac{1}{T}\ln|\tilde{\Sigma}^{-1} \otimes X'X| \ ,$$

where $\otimes$ is the kronecker product operator.

Unfortunately, there is not much research on asymmetric VARs. A method to determine asymmetric lag structure was proposed by Hsiao (1981). This method allows the data to determine the lag lengths of each variable in each equation based on the FPE criterion. It does not put any kind of symmetry restriction

into the lag structure. This avoids imposing any potentially spurious or false restrictions on the model.

Consider the bivariate VAR model with stationary variables z and x,

$$z_t=\beta_{10}+\beta_{11}(L)z_t+\beta_{12}(L)x_t+u_t$$
$$x_t=\beta_{20}+\beta_{21}(L)z_t+\beta_{22}(L)x_t+v_t \quad , \tag{2.3}$$

where β_{ij} (i,j=1,2) are the lag polynomials, and β_{i0} (i=1,2) are constant terms. (2.3) is a reduced form model; thus $\beta_{ij}(0)=0$. Hsiao's method chooses a lag structure which yields the smallest FPE. The minimum FPE of z can be obtained by letting the polynomial degree of $\beta_{11}(L)$ and the polynomial degree of $\beta_{12}(L)$ vary between 0 and M, where M is assumed to be the maximum lag length for this variable. This requires $(M+1)^2$ FPE computations. Hsiao argues that this is a lot of computation, and he offers a modification using Granger's causality concept. This method needs 2M FPE calculations per equation which is about M^2 fewer computations than the unmodified procedure.

The modified method determines the lag structure of each equation separately. For instance, consider the first equation in (1.3). If we assume that z is a univariate AR process, i.e. $z_t=\beta_{10}+\beta_{11}(L)z_t+u_t$, we can vary the order of $\beta_{11}(L)$ between 1 and M, and, in each case, estimate the equation and compute the FPE. We would set the order to the value, say m, which minimizes the FPE criterion. In this case the FPE criterion is defined as,

$$\text{FPE} = \frac{\text{T+number of parameters}}{\text{T-number of parameters}}\text{SSE} \;,$$

where SSE is the sum of squares of the residuals from the regression. Next, we set the order of β_{11} to m, include x in the equation, and vary its lag polynomial order between 1 and M. Again, we select the lag length, n, which minimizes the FPE. If the inclusion of x yields a lower FPE statistic than an equation only depending on z, then we say that x Granger-causes z, and leave x in the equation. Otherwise, we exclude x from the first equation since it does not Granger-cause z.

Using the same procedure, the lag structure of the second equation can be obtained. If we assume that x is a univariate AR process, we can set the lag length to a number which yields the minimum FPE criterion. Then, we find the order for β_{21} which minimizes the FPE and check whether z Granger-causes x. Since the explanatory variables can vary from equation to equation, the final model whose lag structure is determined is usually estimated via seemingly unrelated regression (SUR).

Even though Hsiao described this method only for a bivariate VAR, it was extended to an N-variate VAR by Caines, Keng, and Sethi (1981). McMillin and Fackler (1984) have applied this procedure to a three-variable VAR.

A practical problem with Hsiao's method is that the lag structure may depend on the order in which the variables enter each equation. Keating (1994) argues that each final equation tends to have more lags of the dependent variable than in the true model since the own-lags are considered first in lag length determination. This may cause a bias in lag structure specification.

Keating (1994) suggests another method to determine an asymmetric lag structure. Unlike Hsiao's method of equation-by-equation lag length estimation, Keating's method is based on jointly estimating the lag structure of the complete system of equations. This is done by estimating a VAR and using the residuals to find the lag structure which minimizes a particular criterion. In Keating's method the final lag structure specification does not depend on the particular order in which variables are introduced.

If we rewrite the structural model in (2.2) as,

$$\Phi(L)y_t = C + e_t \tag{2.4}$$

where C, ϵ_t and y_t are defined as previously, and $\Phi(L)$ is the N×N lag polynomial matrix in which its element at the i^{th} row and the j^{th} column can be expressed as,

$$\Phi_{ij}(L) = \phi_{0ij} + \phi_{1ij}L + \phi_{2ij}L^2 + \dots + \phi_{n_{ij}ij}L^{n_{ij}} \quad .$$

n_{ij} may be different for each element, allowing asymmetry in the structural model. By premultiplying (2.4) with Φ_0^{-1}, the inverse of the contemporaneous effect matrix, we obtain the reduced form of this model as,

$$\beta(L)y_t = D + e_t \ , \tag{2.5}$$

where $\beta(L) = \Phi_0^{-1}\Phi(L)$ and its k^{th}, j^{th} element equals,

$$\beta_{kj}(L) = \sum_{i=1}^{N} \Phi_0^{ki}\Phi_{ij}(L) \tag{2.6}$$

where Φ_0^{ki} is the k^{th}, j^{th} element of Φ_0^{-1}. Assuming that none of the Φ_0^{-1} elements are

zero, it is clear from (2.6) that $\beta_{1j}(L)$, $\beta_{2j}(L)$,..., and $\beta_{Nj}(L)$ are each linear combinations of the same N lag polynomials $\Phi_{1j}(L)$,...,and $\Phi_{Nj}(L)$ conveying the asymmetry in the structural model into the VAR. The lag length of β_{kj} is the largest of the n_{1j}, n_{2j},...,n_{Nj}. This means that in a VAR, a variable has the same lag length in each equation. Thus, the explanatory variables are the same across equations, which makes equation by equation ordinary least squares estimates efficient. Furthermore, this technique retains the continuous lag structure assumption found in an ordinary VAR; each variable has a lag length between one and some finite maximum lag.

Given this underlying lag structure of the reduced form VAR, Keating suggests computing AIC and SIC statistics for each possible combination of lag lengths. A systematic search can be set up by varying the lag length of each variable between 1 and M, which will require M^N VAR estimations. Out of those M^N lag structures, the one which has the minimum AIC or SIC is selected. The SIC tends to choose lower lag lengths than the AIC. However, Keating argues that the residuals from a regression model whose lag length is chosen by SIC often tend to be autocorrelated. Therefore, Keating suggests using AIC as a criterion.

An important difference between Hsiao's and Keating's methods is that Hsiao considers each equation separately, allowing a variable to have different lag lengths in each equation. In this method, if a variable does not Granger-cause the dependent variable, it is excluded from that equation. However, Keating jointly estimates all the equations in the VAR to find the minimum criterion. The final VAR specification is not completely asymmetric as in Hsiao, because a variable has the same lag length in

all equations. Keating claims that his method is superior in specifying the true lag structure, because in Hsiao's procedure the order which the variables are considered might be important.

Even though Keating shows that every variable has the same lag in each equation in a VAR, this depends on the Φ_0^{-1} matrix having no zero elements. Otherwise, it is possible to have different lag lengths of the same variable across equations in a VAR. For example, consider a two-variable VAR. Assume that the contemporaneous coefficient matrix is lower triangular,

$$\Phi_0 = \begin{bmatrix} \phi_{011} & 0 \\ \phi_{021} & \phi_{022} \end{bmatrix}$$

then,

$$\Phi_0^{-1} = \frac{1}{\phi_{011}\phi_{022}} \begin{bmatrix} \phi_{022} & 0 \\ -\phi_{021} & \phi_{011} \end{bmatrix} .$$

This demonstrates that it is possible to have different lag lengths of a variable in different equations.

However, the alternative to Keating's method is Hsiao's method in which the specified lag structure may depend on the order in which the variables enter each equation. A particular solution to the upward bias of the lag length of the dependent variable in Hsiao's method is to abandon using the Granger causality concept, and consider all combinations of lag lengths in each equation. This was originally considered by Hsiao but abandoned because of the large numbers of lag combinations.

However, we applied this method of lag structure estimation. Since this method did not improve the outcome and to save space, the results from this lag specification method are not reported.

2.3 Empirical Methodology

To find which method is superior for finding the true lag structure, we apply three performance criteria to a Monte Carlo experiment based upon simulation of eight VAR models with known parameters and known lag structure. The standard method of simulating a N-variable VAR is to generate N time series using prespecified model parameters and lag structures and a random number generator.

Assume we have the true underlying data generating process as in (2.2). For convenience we rewrite the reduced form (2.1) as,

$$\begin{aligned} y_t &= D+\beta_1 y_{t-1}+\beta_2 y_{t-2}+\ldots+e_t \\ &= D+\beta_1 y_{t-1}+\beta_2 y_{t-2}+\ldots+\Phi_0^{-1}e_t \ . \end{aligned}$$

From this new representation of the error term, it is easy to see that generating y_t using (2.2) by picking ϵ_t from $N(0,\sigma^2 I)$, or generating y_t using (2.1) by picking e_t from $N(0,\sigma^2\Phi_0^{-1}\Phi_0'^{-1}=\Sigma)$ will yield the same values.[11] Therefore, we can use the reduced form model to generate the data for the Monte-Carlo experiment.

First, we set the initial values $y_0, \ldots, y_{-p}$ to zero for some p, which is usually equal to the largest lag length in which a variable enters the VAR. Then we generate an innovation, e_1, by picking a vector of numbers from a sample of the multivariate

[11]The normal distribution is employed since the normal distribution is generally used in Monte Carlo VAR simulations.

normal distribution with zero mean and covariance matrix Σ. Referring back to equation (2.1), y_1 can be computed given e_1 . Similarly, we can generate e_2, and, given y_1, compute y_2. Any number of observations from the model can be generated using this process. In this experiment 212 observations are generated from each model. The first 74 observations are discarded since the random number generator may not generate truly random numbers at the beginning of the process (typically the first 60 to 80 numbers are discarded in data generating processes). The large sample estimation uses observations from 75 to 200 (126 observations which are 31.5 years of quarterly data) to estimate the VAR, specify the lag structure, and estimate the IRFs and out-of-sample forecasts. For the smaller sample estimation, the first 124 observations are discarded and only 76 observations are used in the estimation process. In this essay, all the default analysis will be the larger sample results.

After generating the data, the methods described in the previous section are applied to specify the lag structure. The symmetric VAR lag length is set to a value which minimizes either the AIC or SIC criteria. Once a lag structure is determined, we estimate the VAR and compute the 24 period horizon IRFs and 12 period-ahead out-of-sample forecasts.

To compute the IRFs, we need to estimate the structural coefficient matrix Φ_0. However, the common problem with VARs is that the structural parameters cannot be recovered from the reduced form estimates without some identification assumptions. For policy implications it is important to know the true coefficient matrix. However, since this study is not concerned with policy analysis, we will restrict the

contemporaneous coefficient matrix Φ_0 to be lower triangular, which was the identifying restriction first used by Sims (1980). If the structural innovations vector, ϵ_t, has a diagonal covariance matrix $\sigma^2 I$, then the covariance matrix of the reduced form innovation vector, e_t, will be $\sigma^2\Phi_0^{-1}\Phi_0'^{-1}$. We will use the Choleski decomposition method to obtain a lower triangular matrix, A, such that $\tilde{\Sigma}=AA'$, where $\tilde{\Sigma}$ is the estimated residual covariance matrix. In this case A^{-1} will be the estimated contemporaneous coefficient matrix.

With Sims' identification assumption, a shock to the first variable in the vector of endogenous variables will have a contemporaneous effect on all variables. A shock to the second variable will have a contemporaneous effect on all variables except the first one. And a shock to the last variable will have a contemporaneous effect on itself only.

A Monte Carlo experiment consists of many simulations (in this study the number of replications is 1000). After each simulation, the lag structure is determined and IRFs and forecasts are computed. Since the data, y_t, is a random variable, the specified lag structures are also random variables, and thus have a distribution. The observations of the specified lag lengths from the simulations enable inference to be drawn about the underlying distribution of each method's lag length determination.

The second performance criterion in this study is the mean-square-error (mse) of the IRFs that each method yields. The mse is the sum of the squared difference between the IRF of the estimated model from the simulated data and the IRF from the

actual prespecified model, divided by the number of repetitions of the Monte Carlo experiment.[12] The IRFs obtained from the constructed model and the $\bar{\Sigma}$ matrix are defined to be the true IRF. A lower mse is desirable. The mse equals to the square of bias plus the variance of estimator. Thus, lower mse can either be obtained from lower bias or lower variance.

Another performance criterion used is the relative out-of-sample forecasting ability of the VARs whose lag structures are estimated by any of the methods (symmetric or asymmetric). The forecasts are obtained using the sample up to and including the 200[th] observation. The 1 to 12 period-ahead forecasts are out-of-sample dynamic forecasts (note that only one h-period-ahead forecast is computed at each simulation). The forecast error is the difference between the forecasts and the samples of the generated data from 201 through 212. The h-period-ahead mean-square-forecast-error (msfe) is the sum of squares of these one thousand h-period-ahead forecast errors, divided by 1000.[13]

---

[12]This can be formulated as,

$$\frac{1}{1000}\sum_{i=1}^{1000}(irf_{kl}^{h,i}-trueirf_{kl}^{h})^2 \quad h=1,...,24.$$

where $irf^{h,i}_{kl}$ is the impulse response at the h^{th} period horizon of the k^{th} variable to a shock to the l^{th} variable at the i^{th} simulation, and likewise $trueirf^{h}_{kl}$ is the impulse response at the h^{th} period horizon of the k^{th} variable to a shock to the l^{th} variable obtained from the defined model.

[13]The forecast mse is,

$$\frac{1}{1000}\sum_{i=1}^{1000}(x1_{h,i}-x1(h)_i)^2 \qquad h=1,...,12$$

where $x1_{h,i}$ is the $200+h^{th}$ observation of the first variable at the i^{th} simulation and $x1(h)_i$

To check the robustness of the methods, we repeat the experiment for different predetermined lag structures and sample sizes. We also simulate a symmetric VAR to see how closely both asymmetric lag specification methods estimate symmetry in the lag structure. Also, we simulate models with asymmetric lag structure as Keating described, and a lag structure in which variables can have different lag lengths across equations.

The only issue remaining is to determine the parameters that will be used in the data-generating process. It is important to have a stationary series in the estimation process. The stationarity condition in terms of the reduced form model is that the roots of $\det(I_{N\times N}-\beta_1 z-\beta_2 z^2-\ldots-\beta_p z^p)=0$ lie outside the unit circle. One way to determine the parameters is to pick numbers from the parameter space that satisfy the stationarity condition. For a large number of variables this is tedious and difficult.

This study estimates a VAR with stationary series and uses the coefficients as our model parameters. Then, we assume that the estimated coefficients are the prespecified coefficients which will generate stationary series. Once the parameters are specified, the roots of the characteristic polynomial are checked to determine whether they are outside the unit circle.

We use the three-month T-bill rate, real GDP, and real M1 money supply in our estimation. Keating (1994) tested and rejected the hypothesis of unit roots in the growth rates of these variables. The parameters of the two-variable VAR models that are to be simulated are obtained by estimating a VAR with the growth rates of the

---

is the forecast of the $200+h^{th}$ observation at the i^{th} simulation.

three-month T-bill rate and real GDP as the variables.[14] For example, one model we simulate is a two-variable VAR whose explanatory variables are three lags of the first variable and one lag of the second variable. The coefficients of this model are obtained from regressing the growth rates of the T-bill rate and the real GDP on a constant, on three lags of the growth rate of the T-bill, and on one lag of the growth rate of real GDP. In the three-variable VAR simulation, the coefficients are obtained from a VAR whose variables are the growth rates of the three-month T-bill rate, real GDP, and real M1 money supply. All the models are checked for stationarity, and the characteristic roots are found to be outside the unit circle.

The covariance matrix of the simulated models, $\tilde{\Sigma}$, are obtained from the same estimated VAR from which the coefficients are obtained. The $\tilde{\Sigma}$ matrix is equated to the covariance matrix of the residuals.

2.4 Simulation Results

Eight models are constructed to be simulated. Each model has a different lag structure. The first six models are two variable VARs, and the last two are three variable VAR models. The descriptions of the models are presented in Table 2.1. The first element in the endogenous variable vector, y (equation 2.1), is referred as the first variable, the second element in y is referred as the second variable and the third element as the third variable.

[14]The growth rate of variable X is defined as, $\ln X_t - \ln X_{t-1}$.

Table 2.1
Descriptions of The Simulated Models

Model 1	2 variable	1. equation	Three lags of the first and the second variables.
		2. equation	same
Model 2	2 variable	1. equation	Three lags of the first, one lag of the second variable.
		2. equation	same
Model 3	2 variable	1. equation	Three lags of the first, one lag of the second variable.
		2. equation	Two lags of the first, one lag of the second variable.
Model 4	2 variable	1. equation	Three lags of the first, one lag of the second variable.
		2. equation	One lag of the first, three lags of the second variable.
Model 5	2 variable	1. equation	Eight lags of the first and the second variables.
		2. equation	same
Model 6	2 variable	1. equation	Eight lags of the first, five lags of the second variable.
		2. equation	same
Model 7	3 variable	1. equation	Three lags of the first, second and the third variables.
		2. equation	same
		3. equation	same
Model 8	3 variable	1. equation	Six lags of the first, four lags of the second and the third variables.
		2. equation	same
		3. equation	same

2.4.1 Lag Specification Performances

The first model is a two-variable symmetric VAR whose variables have three lags. Lags 1 through 8 were searched to find the lag length associated with the minimum value generated by the criterion. Table 2.2 presents the percentages of lag lengths that each method has specified. The first column in this table is the lag length. Panel A presents large sample results. Out of 1000 simulations, the true lag length was specified 81.7% of the time by a method whose criterion is to minimize AIC. Two lags were found 2.5% of the time, and four lags were specified 9.5% of the time. The mean of the specified lag length, which is reported at the last row of Panel A, is 3.2 . The next column presents the percentages specified by the SIC minimizing criterion. 60.8% of the time three lags and 38.9% of the time two lags were specified. The mean in this case is 2.6, and lag lengths are skewed to lower lags (skewness parameter is -0.357). PIC's estimates are close to SIC's; PIC estimates two lags 45.0% of the time and three lags 54.6% of the time. The mean length of estimated lags is 2.5. We see that the SIC and PIC tend to pick a shorter lag than the true lag length.

The next columns present Keating's method using the AIC and SIC criteria (KAIC and KSIC, respectively). The two columns are for the lag lengths of the first and the second variables specified by KAIC. In this method the specifications are more dispersed. For the first variable, 15.2% of the time two lags were specified, 60.1% of the time the true lag length was selected, and 13.9% of the time four lags were specified. For the second variable 5.5% of the time one lag was specified,

Table 2.2
Model 1, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	21.7	0.0	0.0
1	0.0	0.1	0.3	0.0	5.5	1.7	34.5	0.2	3.4	0.5	2.6
2	2.5	38.9	45.0	15.2	7.1	42.6	18.7	5.8	1.3	19.3	1.6
3	81.7	60.8	54.6	60.1	65.0	51.2	44.5	17.9	49.7	51.4	26.8
4	9.5	0.2	0.1	13.9	10.9	4.3	2.1	43.8	9.8	11.0	47.2
5	3.2	0.0	0.0	4.4	5.4	0.2	0.0	13.7	6.1	7.4	10.2
6	1.3	0.0	0.0	2.8	2.7	0.0	0.0	9.4	3.3	4.0	4.2
7	1.1	0.0	0.0	2.2	1.8	0.0	0.0	5.2	2.4	3.7	4.0
8	0.7	0.0	0.0	1.4	1.6	0.0	0.0	4.0	2.3	2.7	3.4
mean	3.2	2.6	2.5	3.3	3.3	2.6	2.1	4.3	2.8	3.5	4.1

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	31.1	0.0	0.0
1	0.0	8.5	11.2	1.6	15.2	11.7	46.5	2.1	8.5	4.2	12.5
2	13.8	56.1	61.6	27.8	12.8	43.5	24.1	16.4	2.5	27.5	5.3
3	71.4	35.4	27.2	44.3	48.0	41.2	27.7	21.8	35.8	42.3	28.1
4	8.5	0.0	0.0	13.1	11.6	3.4	1.6	34.2	7.8	10.3	34.4
5	3.5	0.0	0.0	4.9	4.2	0.2	0.1	8.2	3.5	7.2	8.8
6	1.3	0.0	0.0	3.1	3.5	0.0	0.0	8.0	3.0	3.2	4.4
7	1.0	0.0	0.0	2.8	2.3	0.0	0.0	4.0	3.8	2.5	3.6
8	0.5	0.0	0.0	2.4	2.4	0.0	0.0	5.3	4.0	2.8	2.9
mean	3.1	2.3	2.2	3.2	3.1	2.4	1.8	4.0	2.5	3.2	3.6

7.1% of the time two lags, 65.0% of the time the true lag length, and 10.9% of the time four lags were specified. The next two columns are the results from KSIC. As in the SIC case, the KSIC results tend to have lower values than the true lag. One lag

was found 1.7% of the time, two lags 42.6% of the time, and three lags were specified only 51.2% of the time for the first variable. For the second variable these numbers are 34.5%, 18.7%, and 44.5%. The KAIC has mean lags greater than three, and KSIC has mean lags less than three.

The last section reports the lag specifications of Hsiao's method (hereafter HSIAO). HSIAO allows a variable not to enter an equation. Lag 0 for this part indicates the percentage of the time a variable is not included in an equation. a11 is the lag length of the first variable in the first equation, a12 is the lag length of the second variable in the first equation, a21 is the lag length of the first variable in the second equation, and a22 is the lag length of the second variable in the second equation. HSIAO generates more dispersion than the other methods. 21.7% of the time the second variable is not included in the first equation. The true lag is specified 17.9% of the time for the first variable, and 49.7% of the time for the second variable in the first equation. These numbers are 51.4% and 26.8% for the second equation. Note that the first variable is always included in the second equation even though the second variable is sometimes not included in the first equation. This is due to the structure of the covariance matrix, Σ, of the disturbance term.

Panel B of Table 2.2 presents the results for the same model estimated with the smaller sample. The results show that as the sample size decreases, the frequency of specifying the correct lag length also decreases. Especially for the SIC and KSIC cases, the tendency to pick lower lags is dramatic; only rarely are more than three

lags specified. Even for HSIAO, we can see that lower lags are selected more frequently than in the larger sample.

Overall, in the symmetric case, AIC does better than the others. This is to be expected since this is a symmetric lag specification method, and, for this particular case, it performs relatively well. But, for both SIC and KSIC, the performances are not as good as for AIC and KAIC. PIC's lag specification is very close to those of SIC's.

The second model simulated has three lags of the first variable and only one lag of the second variable as explanatory variables. Table 2.3 presents the percentages of lag lengths that are specified. In Panel A, the AIC 70.5% of the time estimates three lags, SIC 69.5% of the time estimates one lag, and PIC 74.1% of the time estimates one lag. The AIC tends to choose the lag length of the variable which has the highest lag length in the model. In a sense, between the choices of one and three, AIC preferred to pick the higher lag. On the contrary, SIC and PIC tends to choose the lowest lag length.

KAIC estimates three lags 72.5% of the time for the first variable, and one lag for the second variable 72.4% of the time. These percentages for KSIC are 63.5% and 94.5%. The means of the specified lags (3.4 and 1.6) are higher than the true values in KAIC. For the first variable this mean value is 2.4 in KSIC.

The correct lag for a11 is specified 53.4% of the time, for a12 30.5%, for a21 54.3%, and for a22 53.1% of the time. The second variable was not included in the first equation 46.8% of the time.

Table 2.3
Model 2, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	46.8	0.0	0.0
1	3.9	69.5	74.1	1.9	72.4	25.4	94.5	0.5	30.5	19.0	57.1
2	9.9	14.7	12.6	3.6	13.3	10.6	5.1	11.0	8.7	2.7	16.1
3	70.5	15.8	13.3	72.5	4.6	63.5	0.3	53.4	4.8	54.3	7.1
4	9.1	0.0	0.0	10.9	4.1	0.4	0.1	16.3	2.9	9.6	8.5
5	3.3	0.0	0.0	5.0	3.1	0.1	0.0	7.7	2.2	6.1	5.5
6	2.0	0.0	0.0	2.6	0.8	0.0	0.0	4.5	1.6	3.1	2.4
7	0.6	0.0	0.0	1.6	0.8	0.0	0.0	3.2	1.5	3.0	1.8
8	0.7	0.0	0.0	1.9	0.9	0.0	0.0	3.4	1.0	2.2	1.5
mean	3.1	1.5	1.4	3.4	1.6	2.4	1.1	3.6	1.3	3.1	2.1

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	54.9	0.0	0.0
1	19.8	84.1	88.0	9.5	70.1	50.9	91.3	2.6	22.5	29.1	64.1
2	18.6	10.4	8.6	8.7	14.0	13.5	7.8	25.1	8.3	4.6	14.1
3	47.5	5.5	3.4	59.3	4.9	34.5	0.5	44.0	3.4	43.0	6.1
4	6.8	0.0	0.0	11.2	4.3	0.8	0.4	11.6	4.5	9.0	6.4
5	3.3	0.0	0.0	4.6	1.7	0.2	0.0	8.5	1.9	4.7	4.2
6	1.8	0.0	0.0	2.3	2.5	0.1	0.0	3.4	2.4	3.1	2.2
7	1.3	0.0	0.0	2.3	1.3	0.0	0.0	2.1	1.6	3.7	1.5
8	0.9	0.0	0.0	2.1	1.2	0.0	0.0	2.7	0.5	2.8	1.4
mean	2.7	1.2	1.2	3.2	1.7	1.9	1.1	3.3	1.1	2.9	1.9

Panel B shows that AIC and SIC tend to pick lags 1 and 2 more often than in Panel A. Still, the most frequently chosen lag length by AIC is three, and one by SIC. KAIC and KSIC also tend to pick lags 1 and 2 more often for the first variable.

Surprisingly, the most frequently chosen lag by KSIC for the first variable is no longer three lags, but one lag (50.9%). In general, HSIAO also specifies the correct lag length less frequently. Based on the three models simulated, we see that a smaller sample causes misspecification more often.

In Model 1 we concluded that AIC performed better. In an asymmetric VAR framework, AIC, SIC and PIC yield misspecified models; KAIC has higher rate of estimating the correct lag structure than either KSIC or HSIAO.

In the next two models, we conduct our experiment on VAR models in which variables have different lags in each equation. Model 3 has three lags of the first variable and one lag of the second variable as explanatory variables in the first equation, and two lags of the first variable and one lag of the second variable as explanatory variables in the second equation. Thus, in this case, lag structures specified by KAIC and KSIC will be misspecified. Only HSIAO has the potential to correctly specify the lag structure.

Table 2.4 presents the results of the lag specification. AIC specifies three lags 51.6% of the time, SIC specifies one lag 79.4% of the time and PIC specifies one lag 82.7% of the time. Based on the results obtained from the previous models, these percentages of estimated lag lengths are not a surprise to us. AIC tends to pick the longer lag in the model, and SIC and PIC tend to select the shorter lag in the model.

KAIC estimates three lags 61.3% of the time for the first variable and KSIC specifies one lag 40.5% of the time. For the second variable one lag is selected most frequently by KAIC and KSIC.

Table 2.4
Model 3, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	42.4	0.0	0.0
1	9.9	79.4	82.7	5.2	73.4	40.5	93.5	0.8	33.4	56.3	39.6
2	26.8	18.4	15.9	14.5	13.3	29.5	6.1	8.7	9.2	23.4	23.1
3	51.6	2.2	1.4	61.3	6.0	29.8	0.4	61.3	5.6	7.7	16.7
4	7.6	0.0	0.0	9.3	3.5	0.2	0.0	12.3	3.0	4.0	6.5
5	2.3	0.0	0.0	4.2	1.7	0.0	0.0	6.9	2.0	2.7	6.7
6	0.9	0.0	0.0	3.3	1.1	0.0	0.0	4.2	1.5	3.0	3.4
7	0.5	0.0	0.0	1.0	0.5	0.0	0.0	3.1	1.8	1.6	1.8
8	0.4	0.0	0.0	1.2	0.5	0.0	0.0	2.7	1.1	1.3	2.2
mean	2.7	1.2	1.2	3.1	1.5	1.9	1.1	3.5	1.2	1.9	2.5

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	51.8	0.0	0.0
1	27.7	86.3	89.8	14.8	69.7	59.2	93.3	3.7	25.5	56.9	50.1
2	32.4	12.4	9.4	26.8	14.2	25.4	6.1	24.3	7.2	20.9	20.6
3	29.3	1.3	0.8	38.8	6.3	15.0	0.4	48.5	4.8	7.8	11.9
4	6.4	0.0	0.0	9.5	3.8	0.3	0.2	10.7	2.9	3.6	5.6
5	1.5	0.0	0.0	2.8	1.2	0.1	0.0	4.1	2.0	3.0	4.6
6	1.0	0.0	0.0	3.2	1.4	0.0	0.0	3.7	1.4	2.8	2.7
7	1.0	0.0	0.0	2.5	1.0	0.0	0.0	2.4	1.7	3.5	2.2
8	0.7	0.0	0.0	1.6	2.4	0.0	0.0	2.6	2.7	1.5	2.3
mean	2.3	1.1	1.1	2.9	1.7	1.6	1.1	3.2	1.2	2.0	2.2

The correct specification in this model is a11=3, a12=1, a21=2 and a22=1.

HSIAO does not correctly specify the lag structure as frequently as we might expect.

For instance, 33.4% of the time one lag is specified for a12, and 23.4% of the time two lags are specified for a21.

Panel B of Table 2.4 reveals the same conclusion. In HSIAO the correct specifications are less frequent than in Panel A. Usually there is a slight increase in the variance of the specified lag lengths, and there is a higher tendency to pick lower lags than in Panel A.

Model 4 is constructed to study the effect of a completely asymmetric model. In this model the explanatory variables of the first equation are three lags for the first variable, and one lag for the second variable. The explanatory variables of the second equation are one lag of the first variable, and three lags of the second variable. As in Model 3, AIC, SIC, KAIC, and KSIC will yield misspecified VARs. HSIAO has the capability to specify the correct lag structure.

Table 2.5 presents the percentages of specified lag lengths. It is not surprising to see that AIC estimated three lags 83.5% of the time. However, this time SIC did not pick the lowest lag length in the model; instead it picked three lags 98.7% of the time. As usual PIC's estimates are close to SIC's with a tendency to pick a little bit lower lags than SIC.

KAIC estimated three lags for both variables most of the time; however, KSIC estimates one lag for the first variable more often. For instance, KSIC selected one lag 70.5% of the time and KAIC 13.4% of the time for the first variable.

HSIAO selected three lags for a11 and a22 11.1% and 70.5% of the time, respectively. One lag is selected for a12 and a21 68.7% and 69.4% of the time. The

Table 2.5
Model 4, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	13.4	0.0	70.5	0.0	0.0	68.7	69.4	0.0
2	0.0	1.1	1.3	18.3	0.0	14.1	0.0	0.1	12.3	12.0	0.0
3	83.5	98.7	98.5	48.4	72.6	15.1	96.6	11.1	6.0	6.5	70.5
4	10.0	0.2	0.2	9.8	15.4	0.3	3.1	13.6	4.2	3.5	12.1
5	2.7	0.0	0.0	4.5	5.0	0.0	0.3	16.2	3.2	3.1	6.7
6	2.0	0.0	0.0	2.5	2.7	0.0	0.0	17.4	2.0	2.5	5.5
7	1.1	0.0	0.0	1.5	2.5	0.0	0.0	20.2	2.0	1.4	2.5
8	0.7	0.0	0.0	1.6	1.8	0.0	0.0	21.4	1.6	1.6	2.7
mean	3.3	3.0	3.0	2.9	3.5	1.4	3.0	5.9	1.8	1.8	3.6

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	0.3	0.0	0.0
1	0.0	0.0	0.1	27.7	0.0	79.1	0.1	0.4	65.7	68.8	0.0
2	0.3	11.8	15.0	25.3	0.4	13.6	1.0	0.4	11.1	11.3	0.1
3	79.8	87.9	84.8	28.2	67.7	7.1	95.0	25.5	6.9	6.2	71.1
4	9.9	0.7	0.1	7.5	14.5	0.2	3.4	19.3	4.2	3.7	14.6
5	4.9	0.0	0.0	4.5	6.5	0.0	0.4	16.1	2.6	3.4	5.1
6	2.2	0.0	0.0	2.6	4.3	0.0	0.0	16.0	2.7	2.2	4.0
7	1.6	0.0	0.0	2.2	2.6	0.0	0.0	12.9	2.7	2.1	2.8
8	1.3	0.0	0.0	2.0	4.0	0.0	0.0	9.4	3.8	2.3	2.3
mean	3.4	2.9	2.8	2.6	3.7	1.3	3.3	5.0	2.0	1.9	3.6

problem here is the overestimation of a11. The overall correct specification of the whole lag structure is 4.6%, which is a very low percentage.

Panel B of Table 2.5 has the small sample results. Generally the estimates are close to their larger sample counterpart. The mean lag lengths are also close to the ones in Panel A.

As shown in the previous results, PIC, SIC and KSIC tend to pick lower lag lengths than the true lag length. However, the lag length that can be specified is bounded from below with one lag, and the true lag length was at most three lags. To reveal the properties of the estimators in VAR models which have variables with high lag lengths (which is often the case in high frequency time series data), the highest true lag length is increased from three to eight in Models 5 and 6. Thus, lag lengths 1 through 16 are searched to find the lag structures which minimize the criteria.

Model 5 has eight lags of both variables in each equation. Table 2.6 presents the percentage of time each lag length is selected. AIC specifies eight lags 67.4% of the time with mean specification of 8.5 lags. The specified lag lengths are skewed a little bit towards high lags, with skewness parameter 0.94. SIC specifies three, four, or five lags most of the time, PIC usually specifies less than six lags, and KSIC usually specifies less than seven lags. KAIC estimates seven or eight lags most of the time. The mean of specified lag lengths with the SIC and PIC criteria are much smaller than the ones specified with the AIC criterion. The usage of the AIC criterion yields lag lengths which are closer to the true one, which in this model is eight. The results clearly shows that the SIC and PIC criteria under-specifies lag structures. Compared to SIC, PIC estimates lower lags more frequently. HSIAO has a dispersed specification which also tends to pick lower lags, but not as often as PIC,

Table 2.6
Model 5, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	1.9	0.0	0.0
1	0.0	8.8	12.8	0.0	0.1	5.1	47.9	0.0	1.2	4.5	1.9
2	0.0	7.7	10.6	0.0	0.1	6.9	6.0	1.2	0.1	0.2	0.2
3	0.5	40.2	45.0	4.6	0.8	49.9	8.8	4.4	1.2	6.9	0.2
4	0.6	12.0	10.0	3.3	3.3	8.5	9.7	1.1	40.7	9.2	1.6
5	3.6	23.7	19.2	2.4	5.3	7.2	15.2	4.9	15.1	3.5	2.4
6	4.2	3.7	1.4	6.3	2.6	15.6	3.4	11.5	6.6	13.7	0.2
7	5.2	0.8	0.3	46.3	1.9	5.8	3.2	31.2	9.4	24.2	21.9
8	67.4	3.1	0.7	18.6	57.5	0.6	5.6	5.1	10.1	15.6	37.7
9	10.9	0.0	0.0	7.6	13.0	0.4	0.2	18.4	4.9	7.0	9.1
10	3.5	0.0	0.0	2.3	5.4	0.0	0.0	4.9	2.1	3.7	11.1
11	1.6	0.0	0.0	2.5	3.9	0.0	0.0	7.3	1.6	3.3	3.9
12	1.2	0.0	0.0	2.1	2.2	0.0	0.0	4.0	2.2	1.7	3.0
13	1.3	0.0	0.0	2.0	1.3	0.0	0.0	1.2	0.9	1.9	2.6
14	0.3	0.0	0.0	0.9	1.0	0.0	0.0	1.3	0.7	1.8	1.2
15	0.0	0.0	0.0	0.8	0.5	0.0	0.0	2.0	0.8	0.9	1.8
16	0.0	0.0	0.0	0.3	1.1	0.0	0.0	1.5	0.5	1.9	1.2
mean	8.5	3.6	3.2	7.5	8.3	3.8	2.9	8.0	5.8	7.1	8.5

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KSIC (1)	KSIC (2)	HSIAO a11	HSIAO a12	HSIAO a21	HSIAO a22
0	-	-	-	-	-	-	-	0.0	12.4	0.0	0.0
1	0.1	41.8	53.4	0.4	5.1	20.7	64.5	0.8	4.3	10.6	14.3
2	0.4	16.1	17.0	1.5	1.1	12.1	10.9	7.2	0.8	1.8	2.3
3	4.2	27.2	23.6	11.1	3.2	49.0	9.5	14.1	3.0	16.2	0.3
4	3.8	7.4	4.0	6.9	8.3	5.2	6.4	3.5	31.3	12.0	4.5
5	10.9	6.0	1.9	4.1	8.3	3.9	5.3	7.8	14.4	2.9	5.6
6	7.9	0.7	0.1	11.5	5.4	6.7	1.2	14.4	8.9	17.1	1.5
7	6.2	0.1	0.0	27.1	3.2	2.1	0.6	24.7	6.4	11.1	21.4
8	40.0	0.0	0.0	10.6	34.7	0.1	1.4	5.7	6.8	8.6	27.3
9	8.8	0.0	0.0	5.6	8.0	0.1	0.2	9.3	2.1	4.1	7.1
10	4.7	0.0	0.0	3.1	5.0	0.0	0.0	3.5	1.6	2.7	5.5
11	3.2	0.0	0.0	5.0	3.3	0.1	0.0	2.8	1.3	4.1	3.0
12	2.3	0.0	0.0	2.2	3.2	0.0	0.0	1.6	1.9	1.9	1.8
13	1.9	0.0	0.0	2.5	2.5	0.0	0.0	1.1	1.1	1.7	1.5
14	1.1	0.0	0.0	1.8	2.7	0.0	0.0	0.8	1.2	1.6	1.6
15	2.2	0.0	0.0	3.2	2.8	0.0	0.0	1.4	1.1	1.0	1.6
16	2.3	0.0	0.0	3.4	3.2	0.0	0.0	1.3	1.4	2.6	0.7
mean	7.9	2.3	1.8	7.5	7.8	2.9	1.9	6.5	5.1	6.1	6.9

SIC and KSIC. The sample means estimated for a11, a12, a21 and a22 are 8.0, 5.8, 7.1 and 8.5, respectively.

In Panel B, AIC specifies the true lag less frequently than in Panel A and the dispersion is larger. PIC, SIC and KSIC picked one, two and three lags most of the time. The lag specifications of KAIC seem to be symmetric around seven lags. The mean values in the smaller sample are lower than the values in larger sample.

In both Panels A and B, KAIC has a higher percentage in estimating the correct lag length than SIC and PIC, even though SIC and PIC are a symmetric lag specification method. However, the same is not true for KSIC.

The sixth model is an asymmetric VAR model, with eight lags of the first variable and five lags of the second variable as explanatory variables in each equation. The simulated lag length results are in Table 2.7. In this case AIC does not usually pick the highest lag length in the true model (8 lags) as it does in models 2 and 3, but picks five, six, seven or eight lags most of the time. SIC picks three, four or five lags most of the time, and so does PIC. Again SIC and PIC pick lower lags than AIC. KAIC does not perform very well in estimating the true lag length, but it gets close. KAIC estimates eight lags 24.6% of the time for the first variable and five lags 42.3% of the time for the second variable. When we look at the mean lag values, we see that the lag structure specified by KAIC is very close to eight for the first variable and close to five for the second variable. KSIC usually estimates fewer than eight lags for the first variable, and less than six lags for the second variable. KSIC tends to pick lower lags than what KAIC picks.

Table 2.7
Model 6, Percent of Time Lag Length Selected

Panel A

	AIC	SIC	PIC	KAIC		KSIC		HSIAO			
Lag				(1)	(2)	(1)	(2)	a11	a12	a21	a22
0	-	-	-	-	-	-	-	0.0	3.1	0.0	0.0
1	0.0	3.7	5.4	0.0	1.3	4.6	33.2	0.1	0.8	1.1	1.9
2	0.0	5.9	7.4	0.0	1.2	2.5	10.3	0.6	0.3	0.2	2.0
3	0.1	32.5	40.4	2.2	1.5	31.1	9.1	2.8	0.9	3.2	0.3
4	0.6	14.2	15.4	1.7	29.8	7.2	22.1	0.2	48.1	4.6	2.4
5	15.8	35.8	27.7	4.2	42.3	8.4	22.9	3.1	24.1	7.3	16.2
6	17.7	5.5	3.2	9.2	8.8	23.1	1.7	13.2	6.4	16.9	3.0
7	41.2	2.4	0.5	40.0	5.0	14.8	0.7	34.0	4.3	21.0	52.6
8	16.2	0.0	0.0	24.6	3.8	7.6	0.0	7.9	3.5	21.6	8.2
9	4.3	0.0	0.0	8.8	2.1	0.6	0.0	22.9	2.2	8.4	4.5
10	1.8	0.0	0.0	3.2	1.1	0.1	0.0	5.7	1.4	4.7	2.1
11	0.9	0.0	0.0	2.5	0.8	0.0	0.0	4.0	0.7	2.9	2.1
12	0.1	0.0	0.0	0.8	0.6	0.0	0.0	1.5	1.3	1.8	2.0
13	0.3	0.0	0.0	1.2	1.0	0.0	0.0	1.2	1.1	1.9	1.0
14	0.6	0.0	0.0	0.7	0.1	0.0	0.0	1.2	0.7	1.6	0.9
15	0.3	0.0	0.0	0.3	0.2	0.0	0.0	0.6	0.5	1.4	0.4
16	0.1	0.0	0.0	0.6	0.4	0.0	0.0	1.0	0.6	1.4	0.4
mean	6.9	4.0	3.6	7.5	5.3	4.8	3.0	7.9	5.1	7.5	7.0

Panel B

	AIC	SIC	PIC	KAIC		KSIC		HSIAO			
				(1)	(2)	(1)	(2)	a11	a12	a21	a22
0	-	-	-	-	-	-	-	0.0	12.4	0.0	0.0
1	0.0	25.4	34.8	0.0	9.4	21.7	50.0	0.6	3.3	5.1	13.7
2	0.2	13.8	16.6	0.7	4.0	8.0	14.8	4.7	1.3	1.0	5.2
3	5.2	37.9	36.9	8.6	5.7	36.7	14.4	7.6	4.8	8.2	0.8
4	5.0	9.6	6.8	5.8	26.0	4.2	9.8	3.2	35.6	7.4	7.8
5	21.7	10.3	4.8	7.2	23.6	5.3	9.4	7.3	17.5	6.0	16.3
6	18.2	2.5	0.1	11.4	9.0	15.2	1.2	15.4	7.6	14.5	4.3
7	24.0	0.4	0.0	24.0	5.3	5.1	0.3	29.5	3.7	13.3	29.6
8	12.0	0.0	0.0	18.2	4.2	3.5	0.1	5.9	3.8	15.8	6.2
9	4.0	0.1	0.0	6.8	3.1	0.3	0.0	11.9	2.2	6.2	5.5
10	1.8	0.0	0.0	5.0	1.5	0.0	0.0	4.7	1.4	4.7	3.1
11	1.5	0.0	0.0	2.6	1.0	0.0	0.0	3.3	0.8	4.7	2.2
12	1.4	0.0	0.0	2.3	1.5	0.0	0.0	1.9	1.2	2.3	1.9
13	1.4	0.0	0.0	1.9	1.5	0.0	0.0	0.6	1.0	3.3	1.2
14	0.6	0.0	0.0	1.6	1.6	0.0	0.0	1.3	1.3	2.5	0.9
15	1.6	0.0	0.0	1.5	1.4	0.0	0.0	0.9	0.8	2.1	0.7
16	1.4	0.0	0.0	2.4	1.2	0.0	0.0	1.2	1.3	2.9	0.6
mean	6.8	2.7	2.3	7.4	5.3	3.5	2.2	7.0	4.7	7.4	6.0

As with KSIC, HSIAO does not do well in specifying the lag structure. HSIAO's estimates of a11 are mostly around eight, a12's around five and four, a21's around seven, and a22's are mostly around seven. Panel B shows that all the methods tend to estimate lower lags more often than in the larger sample.

The smaller sample results presented in Panel B support the evidence that the best method is KAIC. The mean values show that SIC and PIC criteria yield lower lags in the smaller sample than in the larger sample.

So far, we have experimented on two-variable VARs. The results show that the SIC criterion often under-specifies the lag length. The lag estimates of PIC are very close to those of SIC, and more frequently have lower values. The AIC criterion (AIC or KAIC) specifies the correct lag length more often than the SIC and PIC criteria. The equation-by-equation lag specification method (HSIAO) seems to yield very dispersed lag specifications.

The last two models are three-variable VARs. Model 7 has three lags of each variable in each equation. The frequencies of specified lag lengths are in Table 2.8. AIC estimates three lags 82.2% of the time, and SIC estimates two lags 60.6% of the time. Compared to the two-variable case in Model 1, SIC's performance has substantially deteriorated (SIC correctly specifies three lags 60.8% of the time in Model 1). PIC estimates two lags 58.0% of the time and three lags 4.6% of the time. KAIC estimates three lags for the first and third variables more than 50% of the time; however, it only estimates three lags for the second variable 13.7% of the time. Note that, KAIC gets closer to the true values than SIC. KSIC also estimates one or two

Table 2.8
Model 7, Percent of Time Lag Length Selected

Panel A

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KAIC (3)	KSIC (1)	KSIC (2)	KSIC (3)	HSIAO a11	a12	a13	a21	a22	a23	a31	a32	a33
0	-	-	-	-	-	-	-	-	-	0.0	1.5	3.2	3.3	0.0	26.7	0.0	25.1	0.0
1	0.0	30.0	37.4	1.6	16.7	6.0	27.9	64.5	0.0	100.	55.8	15.0	0.9	32.3	12.3	27.1	12.8	21.9
2	10.9	60.6	58.0	18.8	61.5	13.6	29.9	35.1	2.3	0.0	14.8	16.0	53.7	33.9	35.2	7.5	37.3	8.3
3	82.2	9.4	4.6	58.7	13.7	59.7	41.7	0.4	37.0	0.0	11.2	25.6	10.8	15.4	9.8	41.7	9.3	8.0
4	5.6	0.0	0.0	11.2	4.2	12.1	0.5	0.0	31.1	0.0	4.8	27.7	8.0	4.2	6.7	8.6	6.4	24.5
5	1.2	0.0	0.0	4.8	1.4	4.5	0.0	0.0	6.8	0.0	3.3	11.6	9.7	5.6	3.9	5.3	2.5	22.9
6	0.0	0.0	0.0	2.4	0.8	2.4	0.0	0.0	9.3	0.0	2.7	5.8	5.9	3.6	2.4	4.0	2.7	6.3
7	0.0	0.0	0.0	1.2	1.0	1.1	0.0	0.0	8.0	0.0	3.5	5.2	4.0	2.5	1.7	3.4	1.8	5.6
8	0.0	0.0	0.0	1.3	0.7	0.6	0.0	0.0	5.5	0.0	2.4	3.4	3.7	2.5	1.6	2.4	2.1	2.5
mean	3.0	1.8	1.7	3.2	2.2	3.1	2.1	1.4	1.9	1.0	2.1	3.8	3.1	2.5	2.0	2.9	2.0	3.7

Panel B

Lag	AIC	SIC	PIC	KAIC (1)	KAIC (2)	KAIC (3)	KSIC (1)	KSIC (2)	KSIC (3)	HSIAO a11	a12	a13	a21	a22	a23	a31	a32	a33
0	-	-	-	-	-	-	-	-	-	0.0	5.0	13.1	15.1	0.0	37.5	0.0	33.2	0.0
1	2.8	68.4	79.0	6.0	29.8	15.1	52.9	83.5	49.7	99.9	57.3	7.9	3.1	46.1	16.2	35.8	18.4	40.3
2	31.0	29.1	20.6	28.2	45.8	19.3	21.6	15.6	38.4	0.1	10.8	19.6	45.9	27.8	23.6	6.9	23.3	9.4
3	56.7	2.5	0.4	40.0	10.3	42.1	24.1	0.9	11.3	0.0	8.5	23.4	9.3	9.6	8.6	30.0	7.9	8.0
4	5.5	0.0	0.0	10.9	5.0	11.2	1.1	0.0	0.6	0.0	5.0	19.0	7.2	4.1	5.1	8.4	6.0	17.6
5	2.0	0.0	0.0	6.4	3.4	4.6	0.3	0.0	0.0	0.0	4.2	5.1	6.6	4.2	2.6	7.4	3.6	14.6
6	0.7	0.0	0.0	3.7	2.3	3.7	0.0	0.0	0.0	0.0	3.1	4.2	5.0	3.1	1.9	4.3	2.6	4.0
7	0.5	0.0	0.0	2.2	1.8	2.1	0.0	0.0	0.0	0.0	3.2	4.3	3.5	2.1	2.5	3.6	2.1	3.7
8	0.8	0.0	0.0	2.6	1.6	1.9	0.0	0.0	0.0	0.0	2.9	3.4	4.3	3.0	2.0	3.6	2.9	2.4
mean	2.8	1.3	1.2	3.2	2.3	3.0	1.7	1.2	1.6	1.0	2.1	3.0	2.7	2.2	1.7	2.9	1.8	3.0

lags most of the time for the second variable. In the original model, all three variables have three lags; however, in the estimation process the second variable's lag length is constantly underestimated by the methods. The reason for this underestimation might be the structure of the model and the covariance matrix of the disturbance terms.

On average, KSIC under-specifies the lag structure, which makes KAIC preferred over KSIC. HSIAO estimates three lags at most 41.7% of the time and that is for a13. This shows that HSIAO's performance is not very good. The dispersion in HSIAO is larger than the other methods. Note that it is not a coincidence that the third variable is more often excluded from the first and second equations than the other two variables. In Hsiao's method the third variable enters last in the first two equations; therefore, it is excluded more often. The only significant difference in Panel B is that the estimated lags are more often smaller than in Panel A.

Considering the tendency of the SIC and PIC criteria to select lower lags, Model 8 is constructed with six lags of the first variable and four lags of the second and third variables. The Monte-carlo experiment of this model searches up to fourteen lags to estimate the lag length. Table 2.9 has the frequencies of the lag length estimations. AIC mostly estimates four, five and six lags; five lags are estimated 49.3% of the time. SIC and PIC, on the other hand, mostly estimate lags less than five, with two lags having the highest frequency, 52.5% for SIC and 54.2% for PIC. KAIC's estimations are concentrated around six lags for the first variable. For the third variable, four lags are estimated 53.3% of the time. However, KAIC's

Table 2.9

Model 8, Percent of Time Lag Length Selected

Panel A

	AIC	SIC	PIC	KAIC			KSIC			HSIAO								
Lag				(1)	(2)	(3)	(1)	(2)	(3)	a11	a12	a13	a21	a22	a23	a31	a32	a33
0	-	-	-	-	-	-	-	-	-	0.0	1.1	1.7	0.3	0.0	36.5	0.0	32.5	0.0
1	0.0	27.5	35.7	0.0	15.9	3.0	10.4	65.5	45.1	99.8	48.3	2.9	0.2	21.0	4.6	6.5	24.8	15.1
2	0.4	52.5	54.2	0.0	46.8	0.9	8.5	33.3	18.8	0.2	8.0	19.4	3.2	13.6	10.2	1.1	17.3	8.0
3	0.3	6.7	4.6	0.3	6.2	17.8	8.6	0.9	22.4	0.0	8.2	18.6	0.7	15.4	20.9	7.5	4.9	2.3
4	16.9	12.6	5.4	2.9	19.5	53.3	7.7	0.3	13.0	0.0	4.2	16.9	39.2	3.5	11.8	2.1	7.2	8.5
5	49.3	0.7	0.1	25.0	4.1	11.3	36.2	0.0	0.7	0.0	9.8	4.8	15.0	16.4	4.3	4.7	2.9	9.5
6	25.9	0.0	0.0	46.9	2.4	5.1	28.1	0.0	0.0	0.0	4.1	16.6	12.3	6.3	2.5	45.1	1.9	2.1
7	4.7	0.0	0.0	11.0	1.7	2.7	0.4	0.0	0.0	0.0	4.0	5.4	7.7	5.9	2.4	10.3	1.4	31.2
8	0.9	0.0	0.0	5.8	1.0	2.5	0.1	0.0	0.0	0.0	3.0	3.1	8.1	5.7	2.0	8.1	1.0	9.3
9	0.4	0.0	0.0	3.1	0.3	0.9	0.0	0.0	0.0	0.0	2.2	3.1	3.4	3.5	1.1	4.2	1.2	5.1
10	0.3	0.0	0.0	2.1	0.2	0.6	0.0	0.0	0.0	0.0	2.1	1.6	3.4	3.2	0.9	3.4	1.4	3.7
11	0.3	0.0	0.0	0.9	0.4	0.6	0.0	0.0	0.0	0.0	2.0	1.7	2.6	1.9	1.1	1.6	0.6	1.7
12	0.3	0.0	0.0	0.6	0.4	0.5	0.0	0.0	0.0	0.0	0.4	1.3	0.9	1.6	0.8	1.4	0.7	1.6
13	0.1	0.0	0.0	0.6	0.7	0.1	0.0	0.0	0.0	0.0	1.4	1.2	1.6	1.1	0.5	1.6	0.8	1.4
14	0.2	0.0	0.0	0.8	0.4	0.7	0.0	0.0	0.0	0.0	1.2	1.7	1.4	0.9	0.4	2.4	1.4	0.5
mean	5.3	2.1	1.8	6.3	2.9	4.3	4.4	1.4	2.0	1.0	3.3	4.7	5.8	4.5	2.6	6.3	2.2	5.7

(Table con'd.)

Panel B

	AIC	SIC	PIC	KAIC			KSIC			HSIAO								
Lag				(1)	(2)	(3)	(1)	(2)	(3)	a11	a12	a13	a21	a22	a23	a31	a32	a33
0	-	-	-	-	-	-	-	-	-	0.0	3.2	12.6	4.8	0.0	38.6	0.0	38.7	0.0
1	0.3	64.0	77.0	0.0	22.8	11.5	33.7	75.1	60.1	99.3	54.4	8.0	0.8	34.9	9.6	16.1	21.7	33.5
2	5.3	30.0	21.1	1.2	30.0	5.2	16.0	23.4	25.8	0.7	9.3	19.6	8.9	14.5	14.8	2.6	13.7	6.0
3	5.8	4.0	1.2	2.8	8.1	21.8	5.3	1.1	10.7	0.0	7.3	19.7	1.3	13.3	13.5	12.9	4.1	2.7
4	24.8	1.9	0.3	7.6	11.7	25.1	17.0	0.3	3.1	0.0	4.7	12.8	31.6	3.8	6.4	3.1	5.0	12.5
5	28.4	0.1	0.0	23.7	3.6	7.9	12.6	0.1	0.2	0.0	4.6	4.2	13.7	10.5	3.7	3.2	3.2	11.0
6	12.0	0.0	0.0	27.5	2.6	5.3	0.5	0.0	0.1	0.0	3.0	8.2	11.9	4.5	3.0	29.8	2.2	3.0
7	3.5	0.0	0.0	7.6	2.8	4.3	0.0	0.0	0.0	0.0	3.4	2.9	4.6	3.7	2.6	9.2	1.6	16.6
8	1.8	0.0	0.0	6.4	1.8	2.7	0.0	0.0	0.0	0.0	1.5	1.9	7.7	5.1	1.5	5.3	1.6	5.5
9	2.0	0.0	0.0	4.6	1.8	1.5	0.0	0.0	0.0	0.0	2.1	2.7	3.8	3.3	1.6	3.9	1.4	2.8
10	0.5	0.0	0.0	3.0	2.0	2.2	0.0	0.0	0.0	0.0	1.2	1.5	2.6	2.2	1.1	3.1	1.3	1.4
11	1.0	0.0	0.0	2.8	2.1	2.0	0.0	0.0	0.0	0.0	1.2	1.6	2.3	1.3	1.6	3.9	1.7	1.2
12	1.2	0.0	0.0	2.7	2.5	2.3	0.0	0.0	0.0	0.0	1.3	1.1	2.2	1.2	1.0	2.7	0.9	2.3
13	2.8	0.0	0.0	3.4	2.5	2.8	0.0	0.0	0.0	0.0	1.2	1.2	1.7	0.9	0.5	2.7	1.6	1.0
14	10.6	0.0	0.0	6.7	5.7	5.4	0.0	0.0	0.0	0.0	1.6	2.0	2.1	0.8	0.5	1.5	1.3	0.5
mean	6.1	1.4	1.2	7.0	4.2	5.1	3.0	1.3	1.6	1.0	2.9	3.8	5.5	3.7	2.4	5.7	2.3	4.4

estimates for the second variable are underestimated. KSIC estimates six or five lags for the first variable most of the time, one or two lags for the second, and less than five lags for the third variable. KAIC more closely estimates the true lag length than KSIC. With HSIAO too, the second variable's lag length is underestimated. Overall, HSIAO's estimates are poor and dispersed. In Panel B, KAIC and KSIC estimate the correct lag lengths less often than in Panel A. HSIAO displays the same pattern of estimation, with slightly lower percentages of the correct specifications.

The general conclusion of this section can be stated as: In symmetric VARs, AIC is the best method, and KAIC is the second best. In second degree asymmetric VARs (each variable has different lag lengths but they are the same across equations, models 2, 6, 8), KAIC has the best performance and HSIAO has the second best, especially for VAR models with high order lag lengths. In first degree asymmetric VARs (each variable has different lag lengths in each equation, models 3, 4), HSIAO's success ratio is below 30% which makes it relatively undesirable. If we know that the true model is first degree asymmetric, HSIAO should definitely be used to estimate the lag structure; however, since we do not know the true model, KAIC might be preferred because if the true model is symmetric or second order asymmetric, it is the best method; if the model is first degree asymmetric, the loss is not substantial. Overall, PIC's performance is close to SIC's.

2.4.2 The IRF performances

A misspecified model will yield an incorrect IRF of the VAR model. This section investigates the IRF performances of the lag specification methods. The IRF

performance is measured by the mean-square-error where the error is the difference between the IRF from the true model and the IRF from the VAR whose lag structure has been estimated.

The mse of IRFs is defined in footnote 12. We rewrite the formula for convenience:

$$\frac{1}{1000}\sum_{i=1}^{1000}(irf_{kl}^{h,i}-trueirf_{kl}^{h})^2 \qquad h=1,\ldots,24.$$

We compute the IRF for each method and criterion. For a two variable model, there are four different IRF measures which are computed over a 24 period horizon. However, a separate mse is not reported for each horizon in order to conserve space. Since we are interested in the overall IRF performance of each criterion, we report a summary mse measure that incorporates all 24 horizons, i.e.

$$\frac{1}{24{,}000}\sum_{h=1}^{24}\sum_{i=1}^{1000}(irf_{kl}^{h,i}-trueirf_{kl}^{h})^2 \ .$$

In this case the computed mse is for all response horizons and for all replications. (Note that this mse measure can also be written as mse = bias^2 + variance, where bias is the average bias at 1000 replications and variance is the variance of the 1000 IRF responses.)

The IRF mse's of Model 1 are in Table 2.10. The four different IRFs in a two-variable VAR model are: the response of the first variable to a shock to the first variable (IRF_{11}) and to a shock to the second variable (IRF_{12}), and the response of the second variable to a shock to the first (IRF_{21}) and the second variables (IRF_{22}).

Table 2.10
Model 1, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	1.294	1.250**	1.270*	1.496	1.534	1.782
$IRF_{12}(\times 10^{-7})$	1.733**	2.054	2.131	2.033*	2.143	2.353
$IRF_{21}(\times 10^{-3})$	0.879**	1.339	1.451	1.063*	1.478	1.311
$IRF_{22}(\times 10^{-7})$	1.480*	1.470**	1.509	1.743	2.128	2.004
Panel B						
$IRF_{11}(\times 10^{-3})$	2.181	2.094**	2.129*	2.674	2.405	3.107
$IRF_{12}(\times 10^{-7})$	2.933**	3.406*	3.558	3.489	3.418	3.844
$IRF_{21}(\times 10^{-3})$	1.525**	2.115	2.255	1.857*	2.047	2.130
$IRF_{22}(\times 10^{-7})$	2.569**	2.764*	2.926	3.082	3.216	3.406

IRF_{ij} is the response of the i^{th} variable when the j^{th} variable is shocked. The mse values in each row are scaled with the numbers in parentheses. ** identifies the smallest mse and * identifies the second smallest mse in that row.

Both in the larger sample and in the smaller sample AIC and SIC have the smallest IRF mse's overall. The next best performance is either from PIC or KAIC. In almost all cases HSIAO has the largest mse values. Smaller sample IRF mse's show the same pattern as those in the larger samples but with larger values (Panel B). The larger mse values are expected since in the smaller sample the misspecification of the lag structure occurs more often than in the larger sample.

To investigate the IRF performances in different periods, the plots of the IRF mse's are presented. In all IRF mse plots the response of PIC is not included since it is very close to the SIC mse response. Most of the time these two responses are on top of each other making it visually impossible to separate. This is a direct consequence of the similar lag structure specifications of PIC and SIC. By not

including the IRF mse response of PIC, it will be easier to compare the performances of the other methods. We present some of the IRFs graphically. Only a sample of responses will be presented to save space and presented samples will be enough to give an idea about the response pattern. The horizontal axis of these graphs is the IRF horizon and the vertical axis is the IRF mse value. Figures 2.1 and 2.2 are the plots of IRF_{11} from the larger sample and smaller sample, respectively. The plots of the larger sample and the smaller sample closely resemble each other except the smaller sample has larger mse values. Figures 2.3 and 2.4 show the larger sample mse of IRF_{12} and IRF_{21}. These figures shows a typical outcome of the impulse response. There is a huge increase in the short-run mse and after about 4 periods it starts dropping steadily. This tells us that in the short-run, IRF estimates tend to make larger mistakes than in the long-run. This peculiar behavior is due to the stationarity of the model. Since the data series are stationary by construction, no shock has a permanent effect. Therefore, the impulse response has to die out after a certain period. Since the mse's converge to zero, all the estimates have impulse responses which also converge to zero after a certain period.

AIC has the lowest mse in the short-run (periods up to 4). In the long-run (periods starting from 7 and up) SIC often has the lowest mse. HSIAO has the worst long-run performance. Usually KAIC has smaller overall mses than KSIC. The general problem of the SIC criterion is the very high short-run mse values that it yields. Figures 2.3 and 2.4 clearly show that the SIC (and PIC) criterion has the worst short-run performance and the best long-run performance.

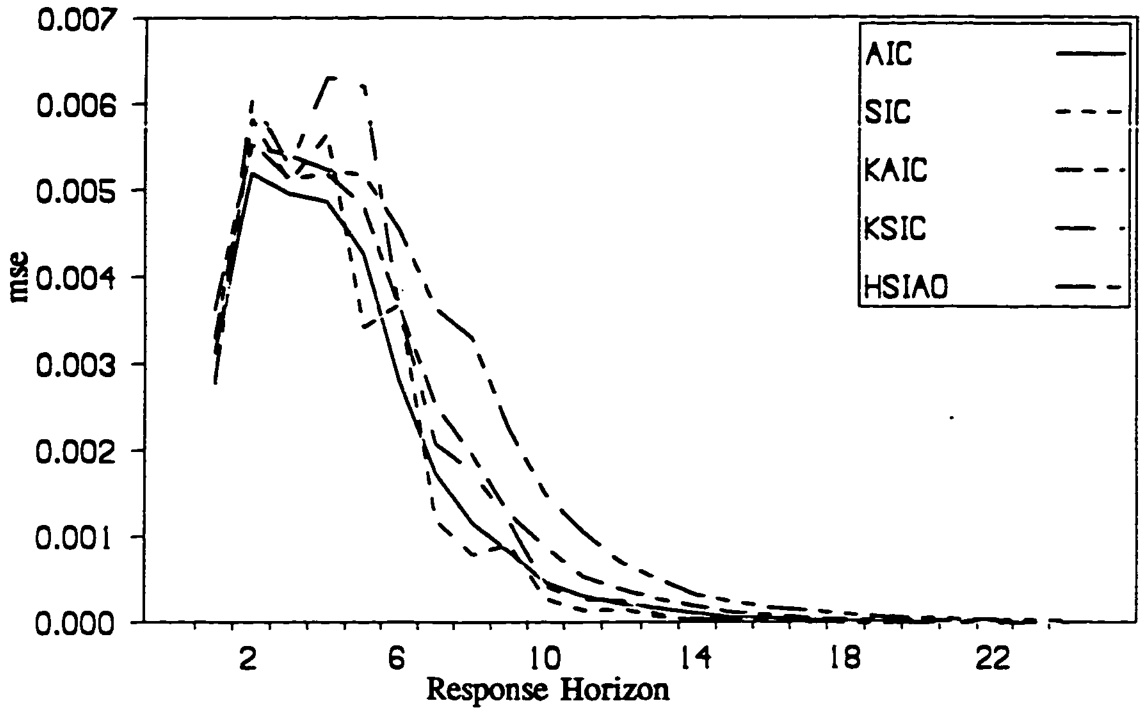

Figure 2.1
Larger Sample Model 1 IRF_{11} mse's

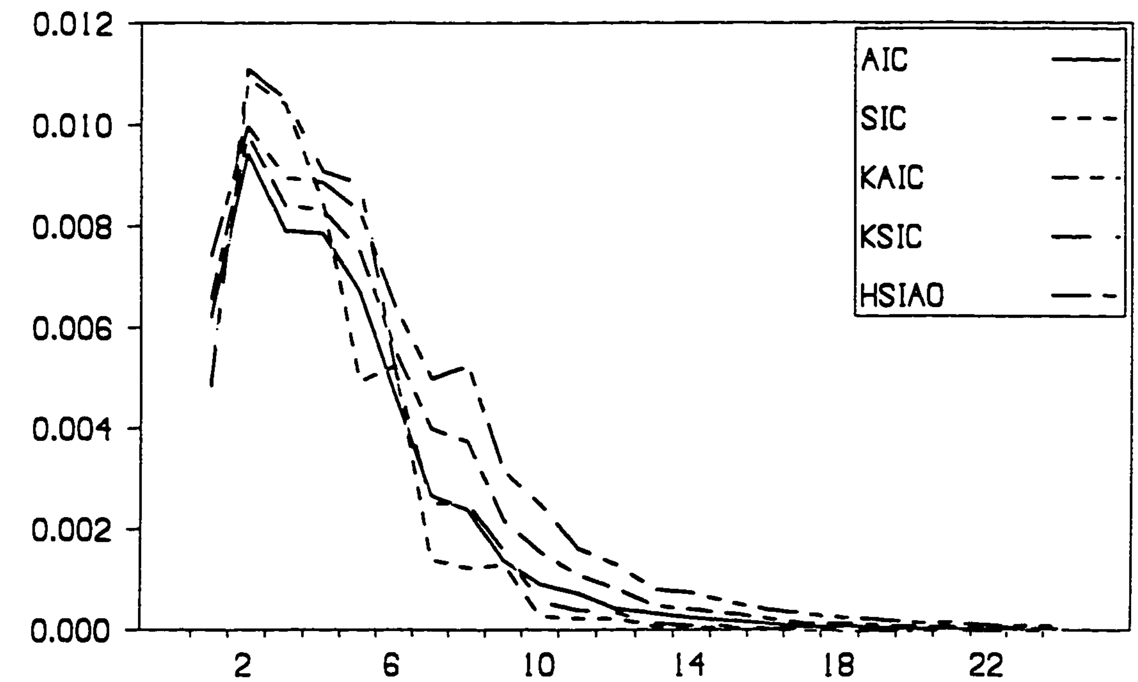

Figure 2.2
Smaller Sample Model 1 IRF_{11} mse's

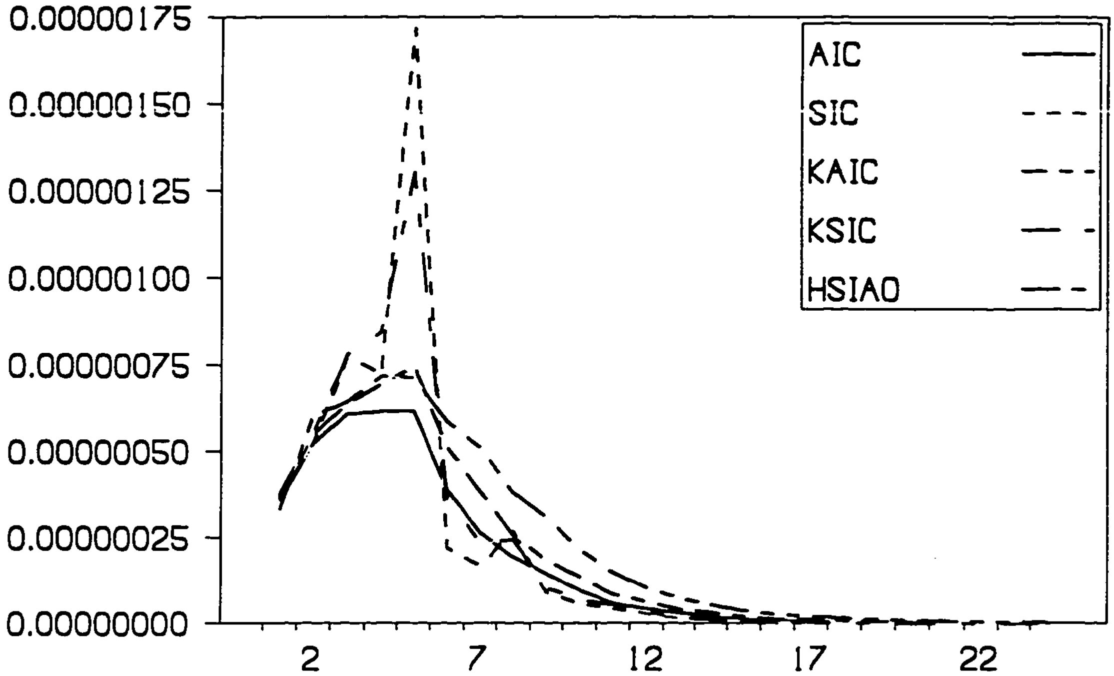

Figure 2.3
Larger Sample Model 1 IRF_{12} mse's

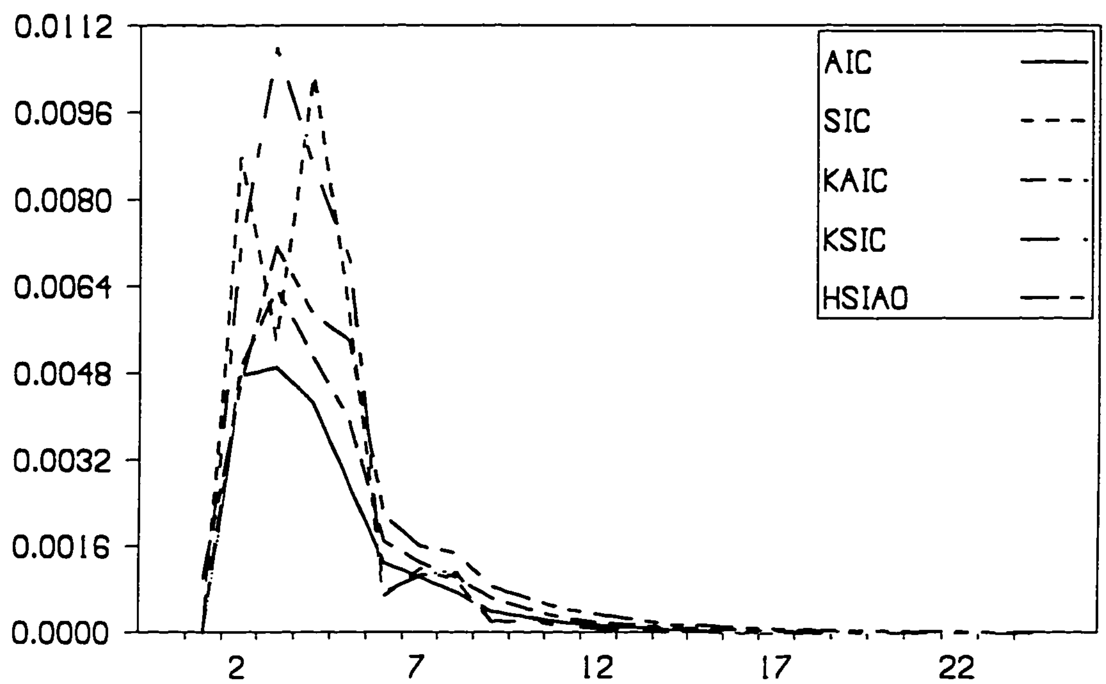

Figure 2.4
Larger Sample Model 1 IRF_{21} mse's

Table 2.11
Model 2, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{3})$	1.477*	2.976	3.056	1.429**	1.748	1.703
$IRF_{12}(\times 10^{7})$	2.451	2.979	2.993	2.331*	2.315**	2.786
$IRF_{21}(\times 10^{3})$	0.781	0.479	0.463*	0.550	0.323**	0.703
$IRF_{22}(\times 10^{7})$	1.607	1.149	1.095*	1.280	0.888**	1.440
Panel B						
$IRF_{11}(\times 10^{3})$	2.639*	3.570	3.616	2.414**	2.851	2.709
$IRF_{12}(\times 10^{7})$	4.050	3.920	3.921	3.895*	3.665**	4.363
$IRF_{21}(\times 10^{3})$	1.248	0.598	0.563*	0.977	0.590**	1.201
$IRF_{22}(\times 10^{7})$	2.587	1.410*	1.325**	2.218	1.496	2.468

For explanations see the footnote of Table 2.10.

The performance of AIC deteriorates in Model 2. Table 2.11 presents the IRF performance results. For this particular model, Keating's method with AIC and SIC criteria usually has the lowest IRF mse. In the smaller sample case, presented in Panel B, KAIC has the same performance level. PIC and KAIC each have the lowest mse values in only one case. AIC and HSIAO do not have any superior performances. Figure 2.5 shows IRF_{11}. SIC and KSIC have larger short-run IRF mse's, which drop substantially in the long-run. KAIC usually has the lowest mse in the short-run. AIC, KAIC and HSIAO have smooth responses than SIC and KSIC which have high peaks in the short-run. For the responses of the second variable, KAIC does not perform well. Figure 2.6 shows IRF mse of the response of the second variable shock to the first variable. KSIC has lower mse in almost all

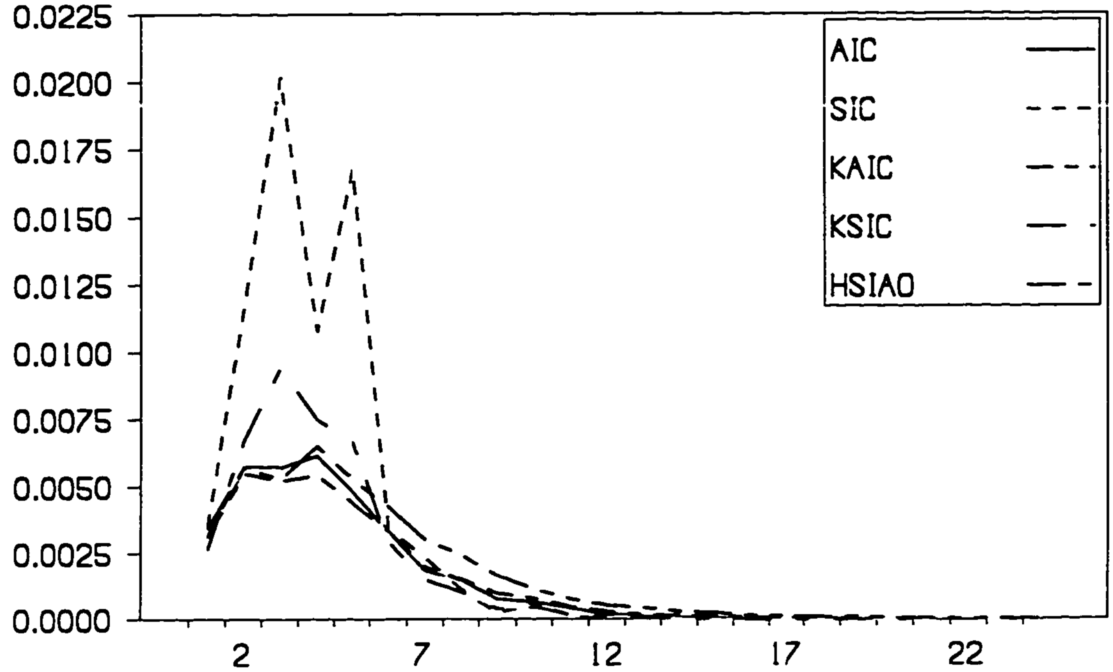

Figure 2.5
Larger Sample Model 2 IRF_{11} mse's

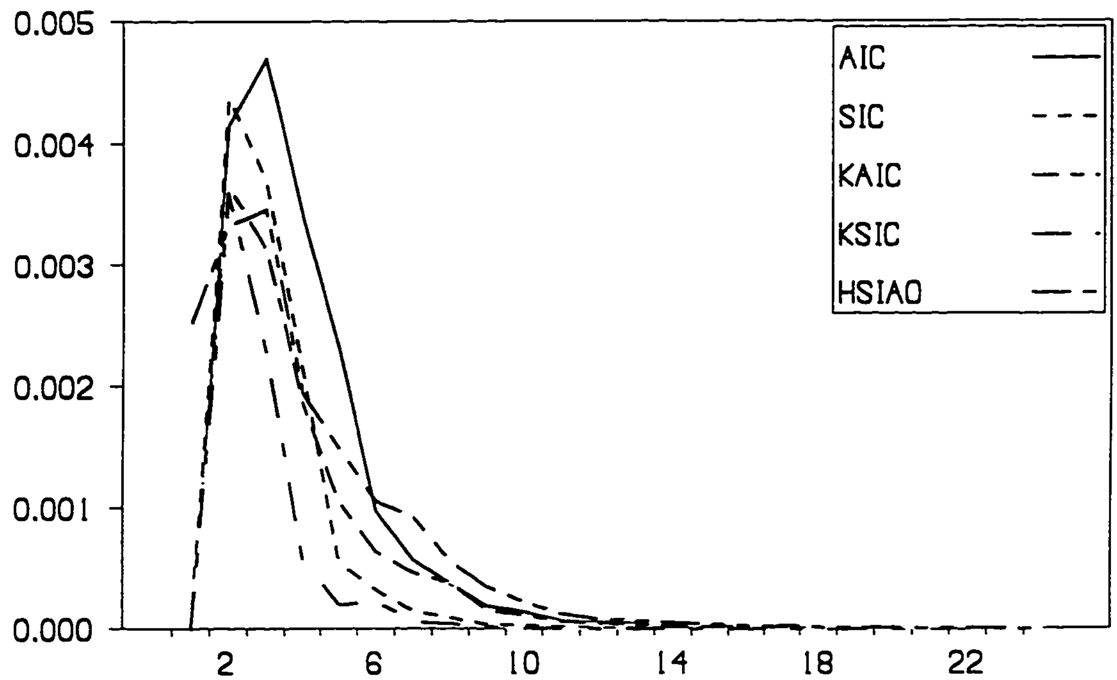

Figure 2.6
Larger Sample Model 2 IRF_{21} mse's

Table 2.12
Model 3, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	1.655	3.216	3.274	1.488**	2.368	1.599*
$IRF_{12}(\times 10^{-7})$	2.148*	2.149	2.150	2.166*	2.044**	2.184
$IRF_{21}(\times 10^{-3})$	0.848	0.514	0.490*	0.619	0.462**	0.778
$IRF_{22}(\times 10^{-7})$	1.640	1.316	1.312*	1.362	1.144**	1.649
Panel B						
$IRF_{11}(\times 10^{-3})$	2.770	3.643	3.692	2.596**	3.211	2.686*
$IRF_{12}(\times 10^{-7})$	3.287	2.787*	2.732**	3.550	2.946	3.517
$IRF_{21}(\times 10^{-3})$	1.274	0.671*	0.637**	1.165	0.738	1.359
$IRF_{22}(\times 10^{-7})$	2.647	1.918	1.902*	2.572	1.791**	2.877

For explanations see the footnote of Table 2.10.

periods. The shape of IRF_{12} mse is similar to IRF_{11} mse, and the shape of IRF_{22} is similar to IRF_{21} (not shown here).

For Model 1, AIC produced better results than the other methods. The results for Model 2 show that in an asymmetric VAR framework, AIC does not perform well anymore. Instead, KAIC and KSIC perform well, generating low mse values.

In Model 3, variables have different lag lengths in each equation; thus, AIC, SIC, KAIC, and KSIC will yield misspecified models. The IRF mses are presented in Table 2.12. In the larger sample KSIC and KAIC have low mse values, but in the smaller sample PIC, KAIC and KSIC have low mse values. The mse values of SIC are very close to those of PIC. Figures 2.7 and 2.8 presents the mse plots of IRF_{11} and IRF_{21}. The response of IRF_{12} is similar to IRF_{11} and response of IRF_{22} is very similar to IRF_{21}. In Figure 2.7 HSIAO has the lowest and SIC and KSIC have the

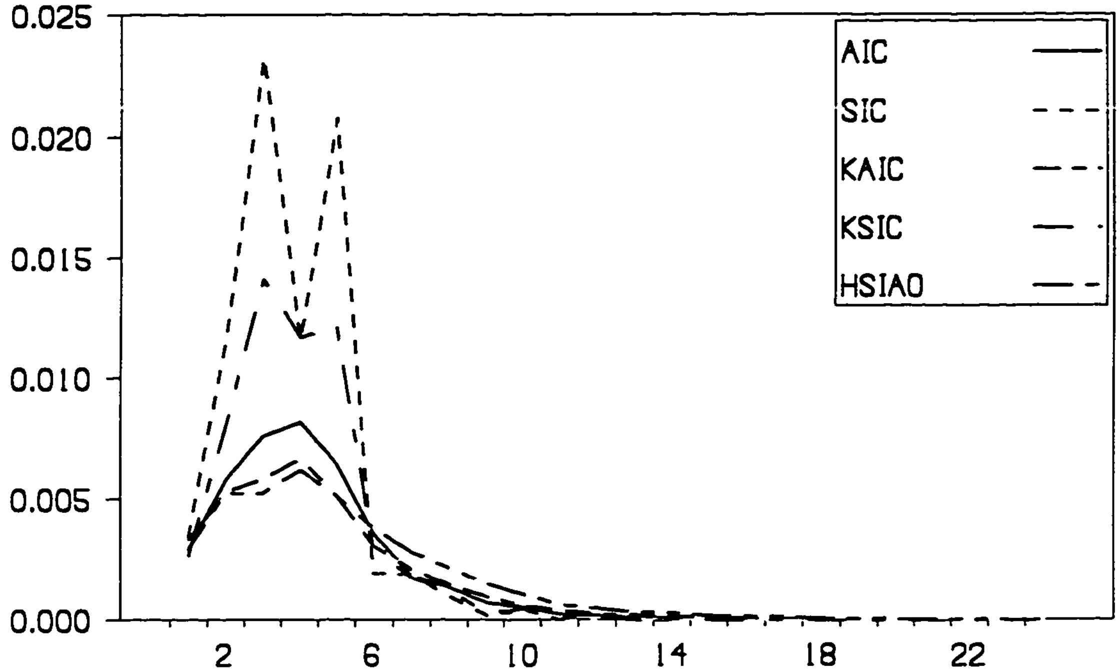

Figure 2.7
Larger Sample Model 3 IRF_{11} mse's

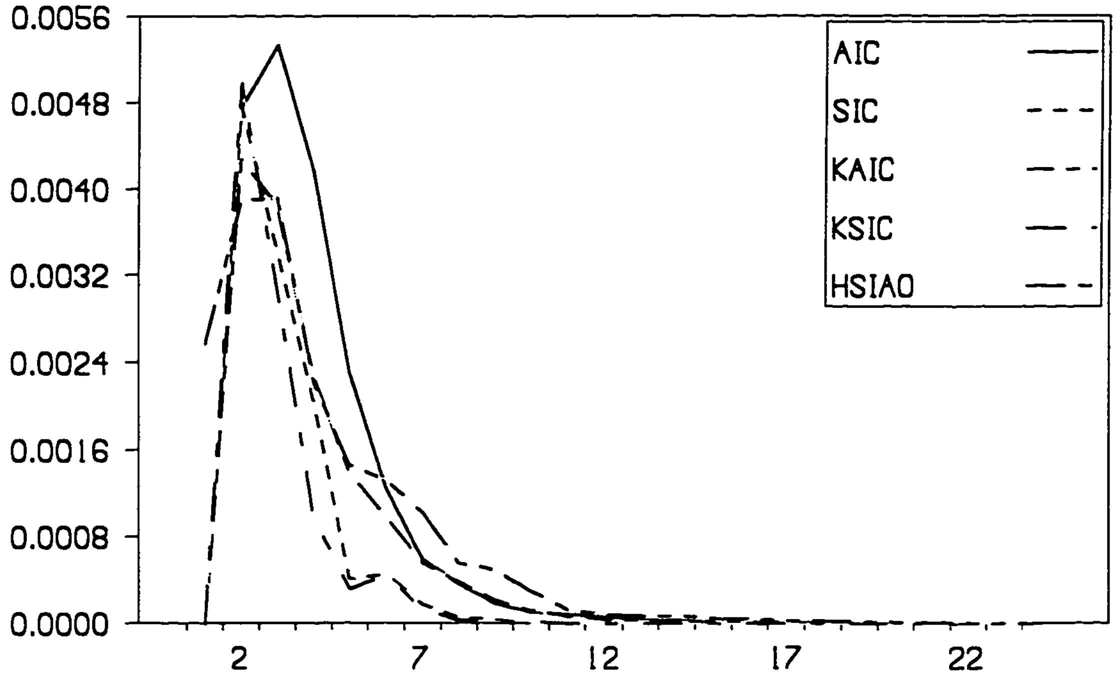

Figure 2.8
Larger Sample Model 3 IRF_{21} mse's

Table 2.13
Model 4, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	1.489	1.316**	1.317*	1.883	2.201	2.041
$IRF_{12}(\times 10^{-7})$	16.19	14.73*	14.74	17.72	13.67**	17.32
$IRF_{21}(\times 10^{-3})$	0.661	0.483**	0.484*	0.898	0.949	0.641
$IRF_{22}(\times 10^{-7})$	6.560	5.604**	5.664*	8.122	6.680	7.793
Panel B						
$IRF_{11}(\times 10^{-3})$	2.566	2.339**	2.369*	3.161	2.835	3.217
$IRF_{12}(\times 10^{-7})$	26.89	26.69*	27.38	29.25	20.85**	28.99
$IRF_{21}(\times 10^{-3})$	1.201	0.862**	0.868*	1.613	1.345	1.186
$IRF_{22}(\times 10^{-7})$	12.12	11.82*	12.47	14.25	10.53**	13.71

For explanations see the footnote of Table 2.10.

highest mse in the short-run. AIC has high mse values, especially for the second variables' responses. In both figures, HSIAO has the highest long-run IRF mse.

Table 2.13 presents the mse's of the IRFs of Model 4. In both the larger and the smaller samples SIC performs best. KSIC has the second best performance. Over here too PIC closely follows SIC. HSIAO, which is the only method capable of correctly specifying the lag structure, does not have the lowest mse in any case. AIC and KAIC are the worst performing methods for this model.

The impulse response of both variables to a shock to the first variable shows a typical hump-shaped response (not shown here), high in the short-run and converging to zero in the long-run. The responses of both variables to a shock to the second variable are explosive. Unlike the previous cases, in the long-run, the responses do not approach zero; instead they continuously increase. Figure 2.9 presents the mse

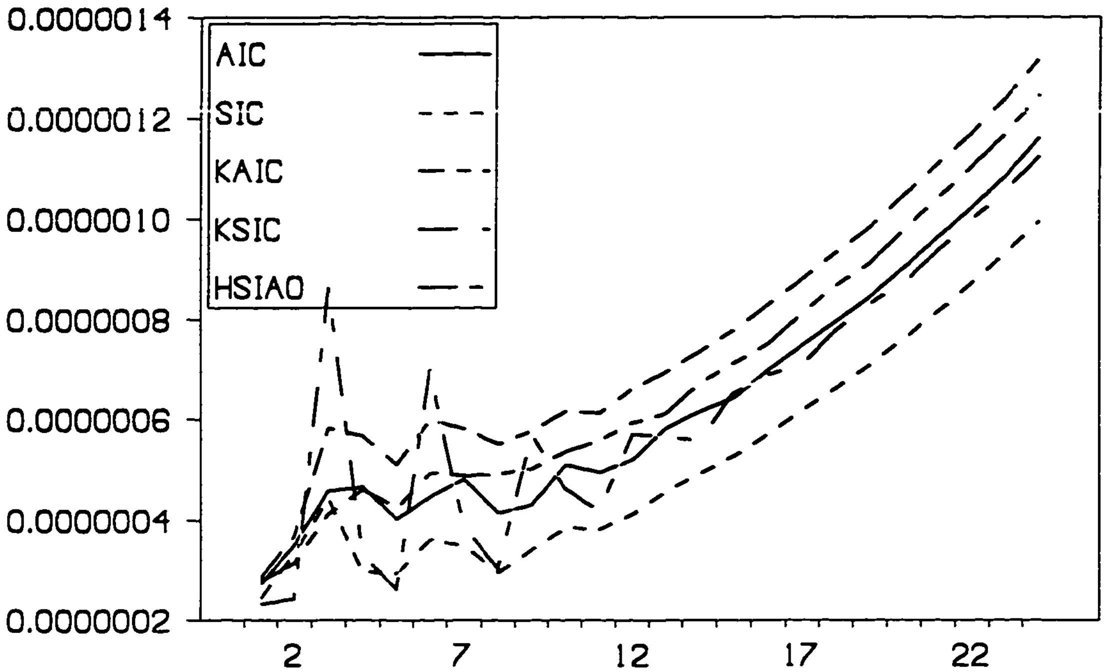

Figure 2.9
Larger Sample Model 4 IRF_{22} mse's

plot of IRF_{22} as a sample response shape. SIC has the best performance, and KAIC has the worst.

The explosive behavior is due to the misspecification of the model. To test this argument, we did a Monte Carlo experiment using Model 4, with 1000 drawings. In each simulation, a VAR model with the lag structure of Model 4 was estimated and thereafter, the IRF was computed. The IRF mse of this experiment was not explosive as in Figure 2.9; rather, it was a flat line. This result shows that a completely misspecified VAR can generate unreliable outcomes.

Models 5 and 6 have longer lag lengths than the previous models. The IRF mse results of Model 5 are presented in Table 2.14. AIC has the lowest mse in most cases and at most horizons. The second best performance is from KAIC in Panel A

Table 2.14
Model 5, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	3.079**	4.369	4.703	3.249*	4.256	3.496
$IRF_{12}(\times 10^{-7})$	5.359**	9.437	9.975	5.614*	8.904	6.171
$IRF_{21}(\times 10^{-3})$	1.942**	3.122	3.318	2.086	3.187	2.078*
$IRF_{22}(\times 10^{-7})$	4.253**	7.371	7.617	4.688*	7.387	4.818
Panel B						
$IRF_{11}(\times 10^{-3})$	5.327**	6.612	7.094	5.595	5.957	5.554*
$IRF_{12}(\times 10^{-7})$	9.179**	12.72	13.31	9.558*	11.76	9.846
$IRF_{21}(\times 10^{-3})$	3.375**	3.574	3.586	3.640	3.618	3.391*
$IRF_{22}(\times 10^{-7})$	7.341*	8.037	7.968	7.970	7.909	7.320**

For explanations see the footnote of Table 2.10.

and from HSIAO in Panel B. PIC usually has the highest mse values. The fluctuations of IRF mse's in the short-run that SIC and KSIC displayed previously are more severe in this model. Figure 2.10 displays a sample response pattern. The mse's of IRF_{12} start decreasing after the 8^{th} period. In general, in this model, SIC and KSIC do not do well in the short-run, but they improve their performance in the long-run.

Model 6 is an asymmetric model. Table 2.15 presents mse's of IRFs for Model 6, and the mse of the IRF of the response of the first variable to a shock to the second variable is plotted in Figure 2.11. It is surprising to see that AIC has a lower mse more often than KAIC in the smaller sample, since AIC leads to a misspecified model in each draw. In the larger sample, KAIC has the lowest mse.

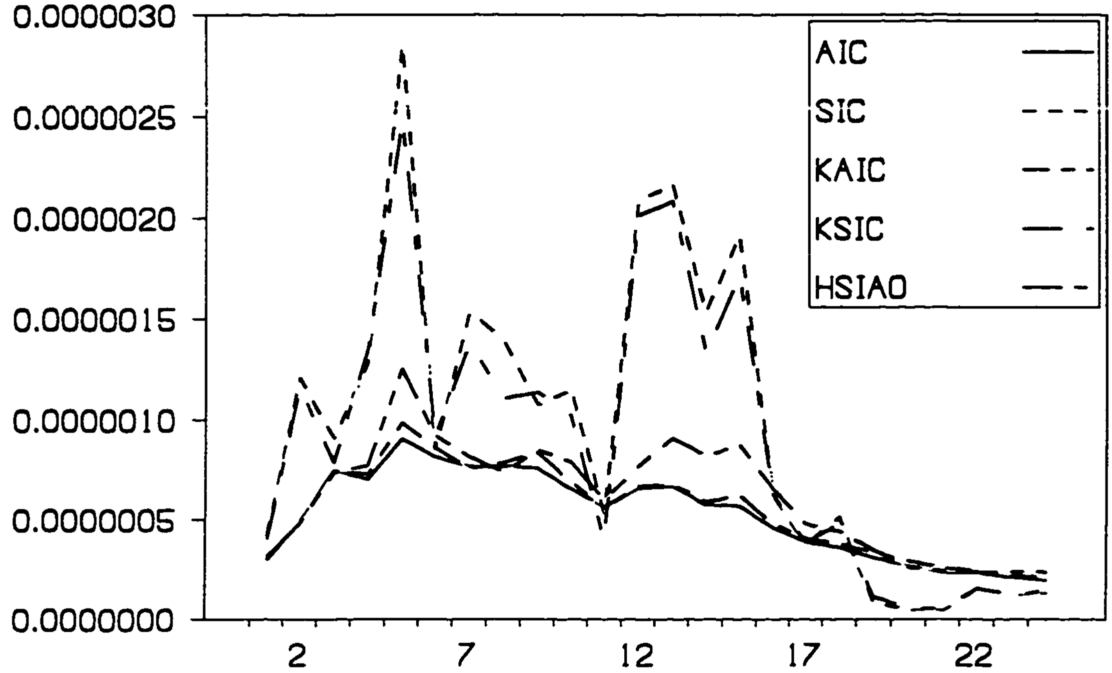

Figure 2.10
Larger Sample Model 5 IRF_{12} mse's

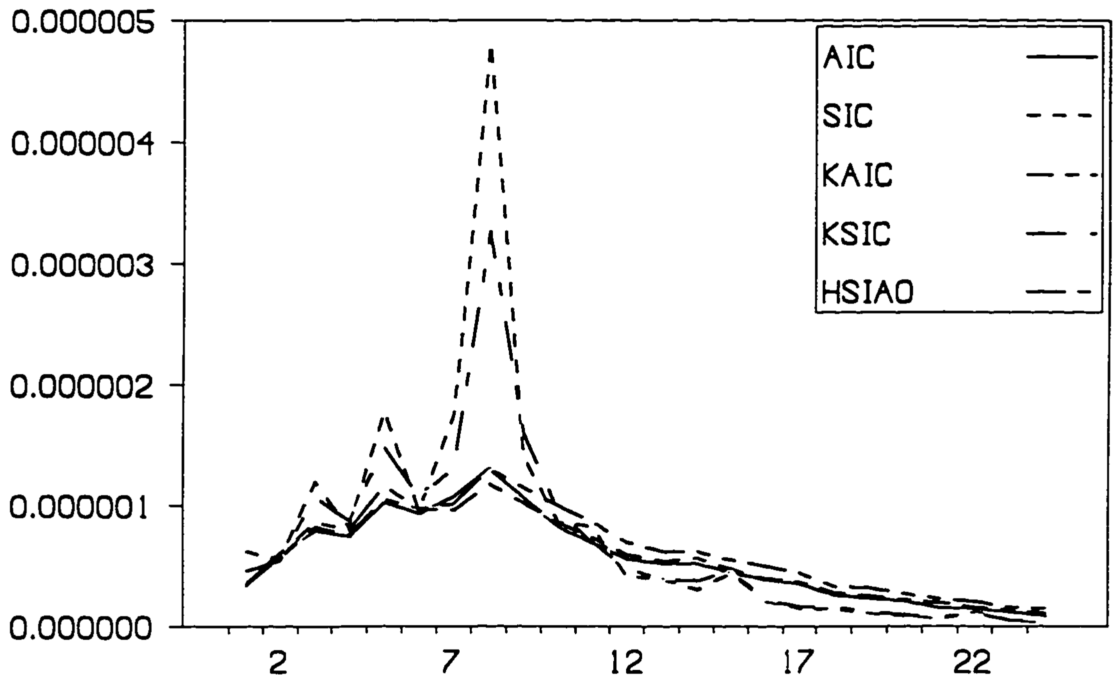

Figure 2.11
Larger Sample Model 6 IRF_{12} mse's

Table 2.15
Model 6, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	3.032*	3.814	4.045	3.008**	3.594	3.257
$IRF_{12}(\times 10^{-7})$	5.714*	7.472	8.017	5.685**	6.709	6.203
$IRF_{21}(\times 10^{-3})$	1.865*	2.682	2.958	1.759**	2.807	1.831
$IRF_{22}(\times 10^{-7})$	3.243*	3.854	4.210	3.159**	3.713	3.420
Panel B						
$IRF_{11}(\times 10^{-3})$	5.085**	5.975	6.348	5.286*	5.622	5.461
$IRF_{12}(\times 10^{-7})$	9.900**	12.38	13.26	9.992*	10.95	11.01
$IRF_{21}(\times 10^{-3})$	3.269**	3.796	3.937	3.310	3.840	3.277*
$IRF_{22}(\times 10^{-7})$	5.712	5.415*	5.544	5.917	5.085**	5.823

For explanations see the footnote of Table 2.10.

The IRF plots of Model 6, which are not presented here, are similar to the shape of IRF_{12} mse in Figure 2.11. The mse's of SIC and KSIC fluctuate around a path with occasional large peaks. SIC and KSIC have the lowest long-run mse's, and, generally, AIC and KAIC have the lowest short-run mse's.

The last two models are three variable VARs. Model 7 is a symmetric VAR model with three lags of each variable. The IRF mse, presented in Table 2.16, show that AIC has the best performance; SIC has the second best performance. However, in the smaller sample, PIC and KSIC do better than in the larger sample. Most of the time HSIAO has the largest IRF mse. Figure 2.12 presents the mse plot of the response of the first variable when the first variable is shocked, and Figure 2.13 is the analogous smaller sample plot. In general, SIC and KSIC have the largest mses in the short-run and the lowest mses in the long-run. In the short-run AIC and KAIC

Table 2.16
Model 7, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	1.394^{**}	2.660	2.868	1.581^{*}	1.910	3.098
$IRF_{12}(\times 10^{-7})$	1.437	1.392^{**}	1.408^{*}	1.645	1.576	2.396
$IRF_{13}(\times 10^{-7})$	2.536^{**}	3.694	3.796	2.857^{*}	2.868	3.842
$IRF_{21}(\times 10^{-3})$	0.834	0.797^{*}	0.776^{**}	0.852	0.907	1.474
$IRF_{22}(\times 10^{-7})$	1.205^{**}	1.231^{*}	1.283	1.313	1.584	1.814
$IRF_{23}(\times 10^{-7})$	1.755	1.700^{*}	1.717	1.787	1.637^{**}	2.484
$IRF_{31}(\times 10^{-3})$	0.875^{**}	1.823	1.937	1.086^{*}	1.807	2.247
$IRF_{32}(\times 10^{-7})$	0.856^{**}	1.014^{*}	1.060	1.043	1.072	1.475
$IRF_{33}(\times 10^{-7})$	1.646^{**}	2.658	2.799	2.002^{*}	2.349	2.964
Panel B						
$IRF_{11}(\times 10^{-3})$	2.552^{**}	4.108	4.352	2.878^{*}	3.217	3.866
$IRF_{12}(\times 10^{-7})$	2.400	2.114^{*}	2.107^{**}	2.855	2.437	3.903
$IRF_{13}(\times 10^{-7})$	4.514^{**}	5.161	5.232	5.065	4.555^{*}	5.670
$IRF_{21}(\times 10^{-3})$	1.402	1.054^{*}	1.032^{**}	1.541	1.196	2.353
$IRF_{22}(\times 10^{-7})$	2.072	1.975^{**}	2.025^{*}	2.385	2.201	2.971
$IRF_{23}(\times 10^{-7})$	2.969	2.449	2.416^{*}	3.122	2.361^{**}	3.769
$IRF_{31}(\times 10^{-3})$	1.538^{**}	2.420	2.524	1.803^{*}	2.296	2.885
$IRF_{32}(\times 10^{-7})$	1.489^{*}	1.501	1.539	1.814	1.465^{**}	2.378
$IRF_{33}(\times 10^{-7})$	2.812^{**}	3.554	3.656	3.371	2.910^{*}	4.370

For explanations see the footnote of Table 2.10.

usually do well. AIC is always better than KAIC in the long-run. HSIAO has the largest long-run mse, and its short-run performance is not better than AIC or KAIC. SIC and KSIC seem to converge to their long-run path faster than AIC and KAIC in the smaller sample. The IRF mses that are not shown here have similar characteristics to the IRF_{11} mse. In the short-run response, SIC and KSIC have

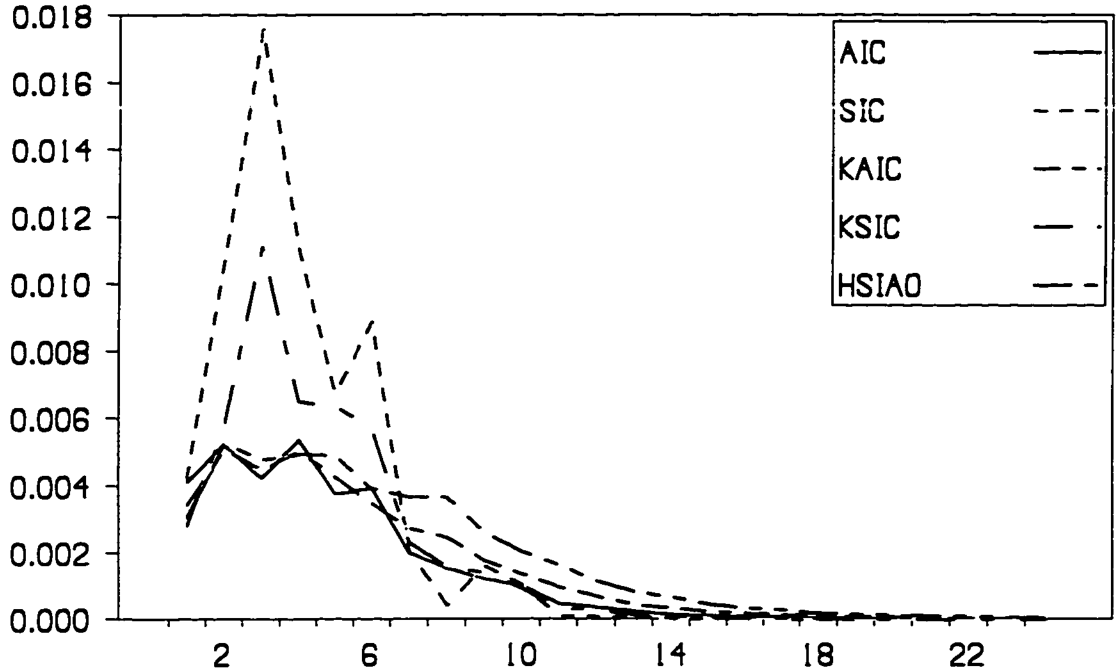

Figure 2.12
Larger Sample Model 7 IRF_{11} mse's

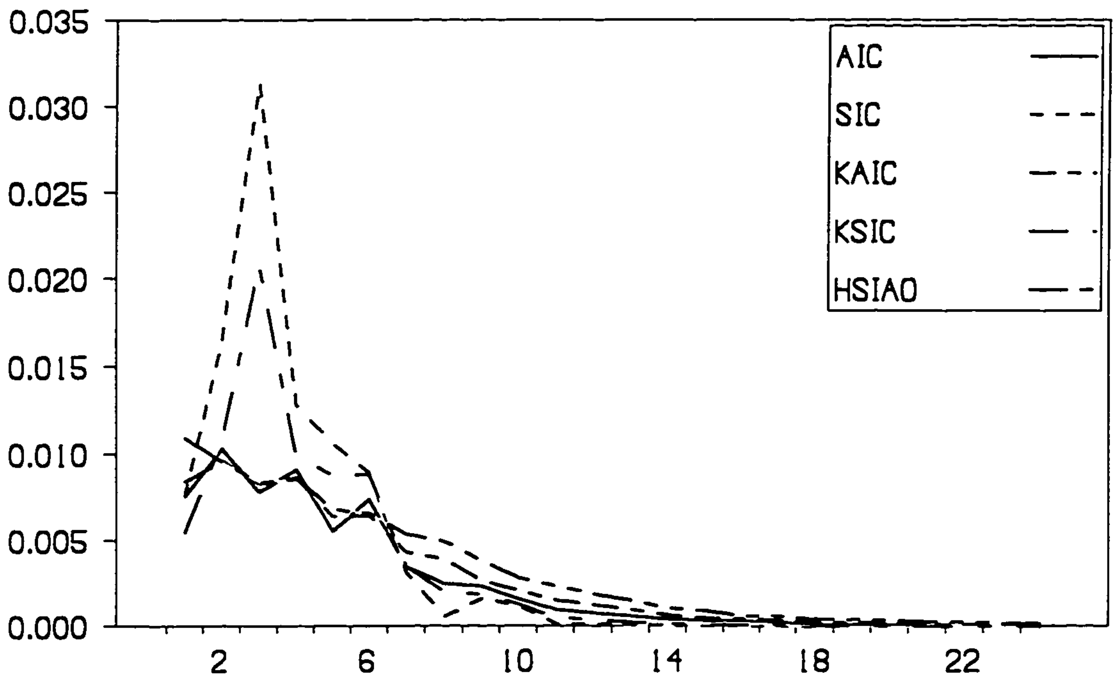

Figure 2.13
Smaller Sample Model 7 IRF_{11} mse's

Table 2.17
Model 8, IRF mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
$IRF_{11}(\times 10^{-3})$	2.292**	3.993	4.295	2.325*	2.596	5.578
$IRF_{12}(\times 10^{-7})$	2.296**	3.241	3.366	2.469*	2.554	5.007
$IRF_{13}(\times 10^{-7})$	3.889*	6.278	6.601	3.804**	4.365	6.010
$IRF_{21}(\times 10^{-3})$	1.505	1.116*	1.112**	1.303	1.475	2.431
$IRF_{22}(\times 10^{-7})$	1.912**	2.006*	2.123	1.954*	2.073	2.415
$IRF_{23}(\times 10^{-7})$	2.519	2.730	2.810	2.236*	2.224**	3.308
$IRF_{31}(\times 10^{-3})$	1.504**	3.548	3.742	1.638*	2.879	3.299
$IRF_{32}(\times 10^{-7})$	1.771**	4.156	4.465	1.777*	2.454	2.438
$IRF_{33}(\times 10^{-7})$	2.700*	3.469	3.601	2.720	2.559**	3.037
Panel B						
$IRF_{11}(\times 10^{-3})$	4.843*	5.530	5.897	4.870	4.261**	6.673
$IRF_{12}(\times 10^{-7})$	4.601	4.239	4.236*	4.863	4.014**	8.182
$IRF_{13}(\times 10^{-7})$	7.697*	8.610	8.821	7.751	7.151**	9.311
$IRF_{21}(\times 10^{-3})$	2.930	1.541**	1.560*	2.689	1.916	3.238
$IRF_{22}(\times 10^{-7})$	4.284	2.682**	2.757	4.131	2.752*	3.874
$IRF_{23}(\times 10^{-7})$	4.799	3.718*	3.806	4.224	3.292**	4.746
$IRF_{31}(\times 10^{-3})$	3.117**	3.585	3.506	3.154*	3.364	3.928
$IRF_{32}(\times 10^{-7})$	3.527	4.367	4.301	3.374**	3.489	3.454*
$IRF_{33}(\times 10^{-7})$	5.690	3.302	3.101	5.498	3.039**	4.458

For explanations see the footnote of Table 2.10.

relatively large mse values; however, in the long-run, SIC and KSIC have relatively smaller mse values. AIC and KAIC have small short-run mse values.

Model 8 has six lags of the first variable and four lags of the second and the third variables in all equations. Table 2.17 presents the IRF mse's. In the larger sample, AIC and KAIC usually have lower mse values than the other methods. AIC

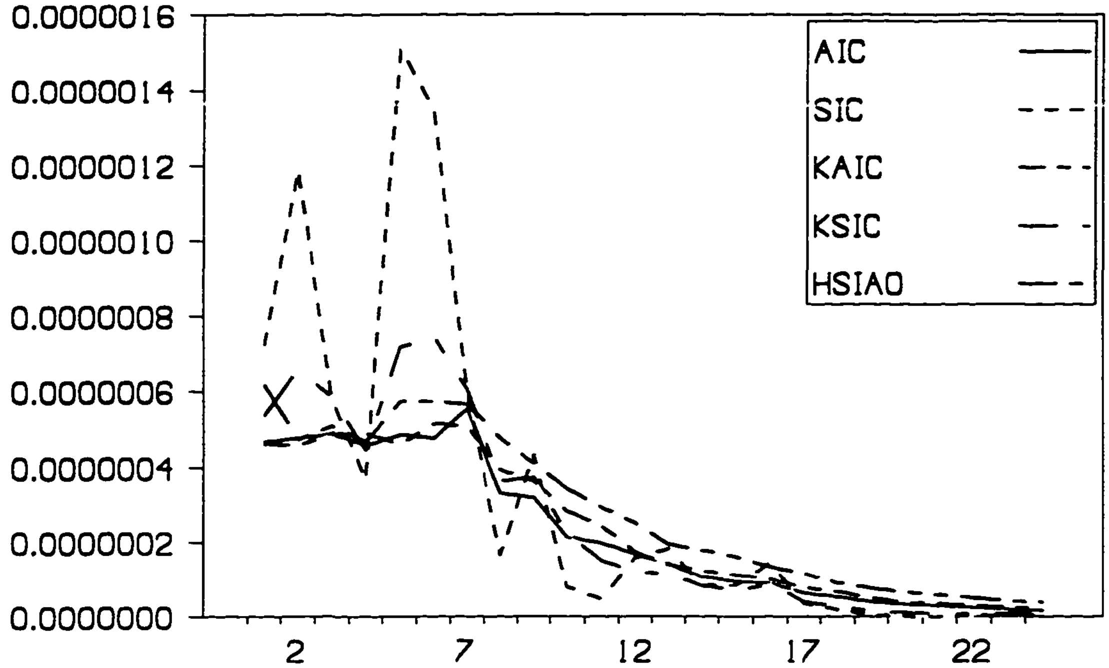

Figure 2.14
Larger Sample Model 8 IRF_{12} mse's

seems to perform a little better than KAIC. However, in the smaller sample, KSIC performs the best, and KAIC performs better than AIC.

In Figure 2.14, the mse of IRF_{12} shows a typical response pattern in which SIC and KSIC have larger mse's in the short-run, and in the long-run they have smaller mse values. In the short-run, AIC and KAIC perform equally well. In the smaller sample, KAIC and AIC are not superior; however, they also do not have large swings as SIC and KSIC have. (The response figures are not shown.)

The results in this section support the conclusion that AIC and KAIC are superior to SIC, PIC and KSIC, especially in the larger sample. AIC is better than KAIC for the symmetric VAR models, and KAIC is usually better than AIC for the asymmetric cases. The superiority of KAIC over AIC also depends on the degree of

asymmetry. For instance, when the lag length difference between two variables is only two, such as in Models 2 and 8, the performance of AIC and KAIC are close; however, when this difference increases, such as in Model 6, KAIC dominates. HSIAO does not perform better than the others even for asymmetric models where each equation has a different lag structure. In the smaller sample (sample size 76), SIC and KSIC more often have lower mse values than AIC and KAIC, especially in Models 7 and 8.

If we separate the bias and variance parts of the IRF mse measure, we see that most of the time SIC, PIC and KSIC have lower variance, and AIC and KAIC have lower bias. This fact is a result of lag specification performances. The results show that AIC criterion yields lag estimates closer to their true value than the SIC and the PIC criteria. Therefore, it can be expected that AIC and KAIC should have lower IRF bias. SIC, PIC and KSIC impose restrictions on the lag length towards lower lags, thus yielding lower variance.

2.4.3 The Forecasting Performances

In Table 2.18 Panel A, we present the forecast mse (msfe) of the first model. AIC has the smallest msfe's for the first variable, and SIC has the best performance for the second variable. In the smaller sample (Panel B), SIC has the smallest msfe's, and AIC has the second best forecasting performance. Figure 2.15 shows the msfe of the first variable, and Figure 2.16 shows the smaller sample msfe of the first variable. In the larger sample AIC usually has the best short-run (periods up to 4) performance, and in the smaller sample, SIC and KSIC usually have the best short-run performance.

Table 2.18
Model 1, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.756**	0.757*	0.760	0.762	0.765	0.758
V2($\times 10^{-4}$)	0.967	0.963**	0.966	0.964*	0.969	0.978
Panel B						
V1	0.769**	0.769**	0.781	0.772	0.793	0.770
V2($\times 10^{-4}$)	0.960	0.956**	0.958*	0.972	0.959	0.986

Vj is the j^{th} variable. The mse values in each row are scaled with the numbers in parentheses. ** identifies the smallest mse and * identifies the second smallest mse for that particular variable.

In both cases SIC has good long-run (periods 7 and up) performance. HSIAO does not do well, especially in the smaller sample. Both in the larger and smaller sample HSIAO usually has the largest mse value in all periods. The plot of the msfe of the second variable is shown in Figure 2.17. Visually it is hard to differentiate the mse values from each other. We can see that in the short-run AIC has the lowest msfe.

The msfe of the second model is presented in Table 2.19. KAIC has the lowest mse of the first variable, and KSIC has the lowest msfe of the second variable. With the asymmetric VAR, Keating's method yields lower mse values. Figure 2.18 presents the msfe of the first variable. KAIC has better short-run performance than KSIC, and KSIC has better long-run performance than KAIC. SIC has the largest short-run msfe. HSIAO usually has the largest long-run msfe. For the second variable the short-run and long run performances (not shown here) are similar to the

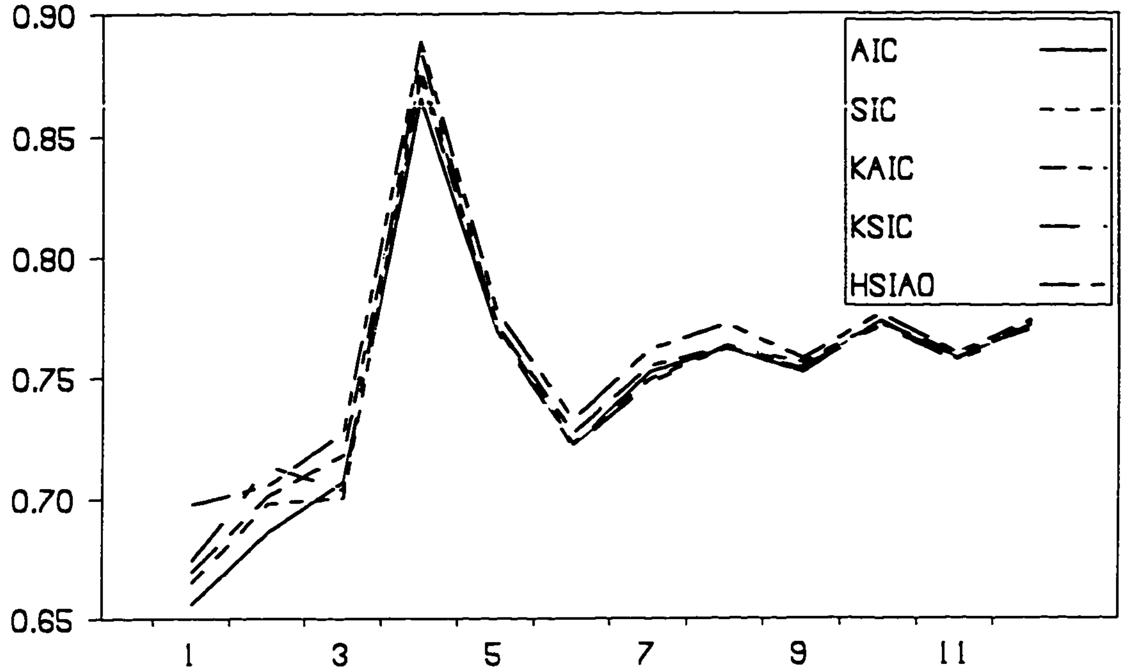

Figure 2.15
Larger Sample Model 1 msfe of the First Variable

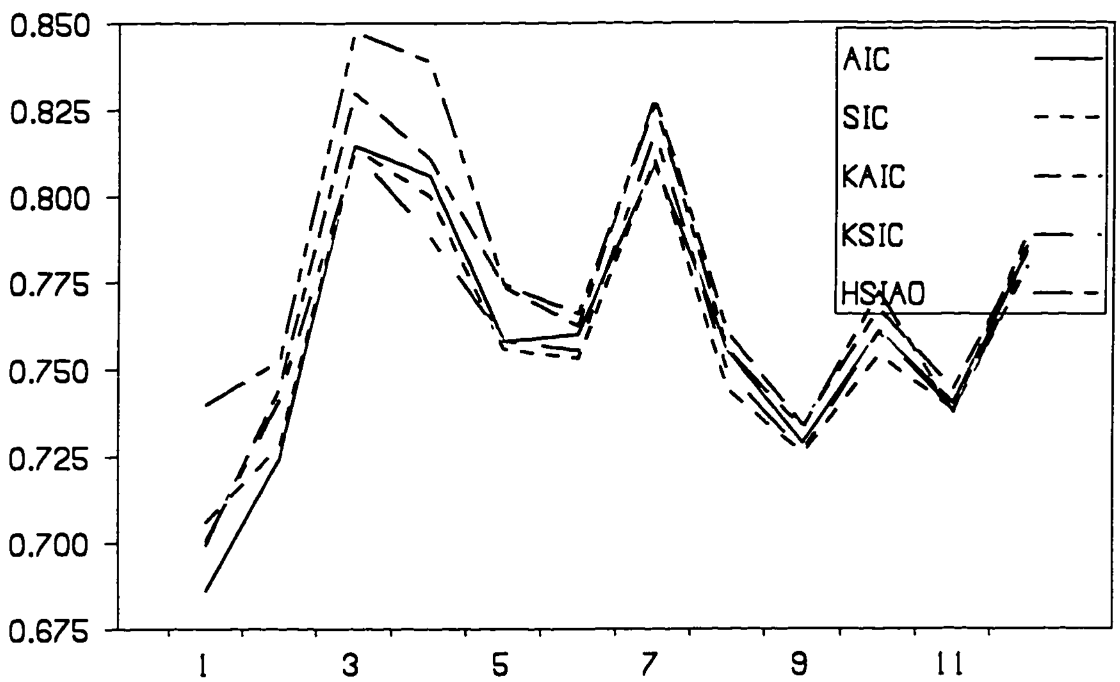

Figure 2.16
Smaller Sample Model 1 msfe of the First Variable

Table 2.19
Model 2 Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.738	0.745	0.737*	0.736**	0.738	0.745
V2($\times 10^{4}$)	1.114	1.110	1.108*	1.110	1.105**	1.119
Panel B						
V1	0.747	0.742*	0.743	0.741**	0.749	0.743
V2($\times 10^{4}$)	1.207	1.186	1.784**	1.203	1.184**	1.208

For explanations see the footnote of Table 2.18.

Table 2.20
Model 3, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.736	0.740	0.735**	0.735**	0.737	0.740
V2($\times 10^{4}$)	1.101	1.093**	1.093**	1.104	1.094	1.107
Panel B						
V1	0.737	0.733*	0.736	0.730**	0.740	0.733*
V2($\times 10^{4}$)	1.144	1.128	1.126*	1.157	1.125**	1.162

For explanations see the footnote of Table 2.18.

performances of the first variable. Smaller sample forecast performances are similar to the larger sample forecast performances.

The forecast mse of Model 3 is presented in Table 2.20. For this model PIC has the overall lowest msfe. In the smaller sample, KAIC and KSIC have the lowest values. Since only HSIAO is the method which can potentially correctly specify the asymmetric lag structure in this model, it is surprising to see that HSIAO usually has the largest msfes.

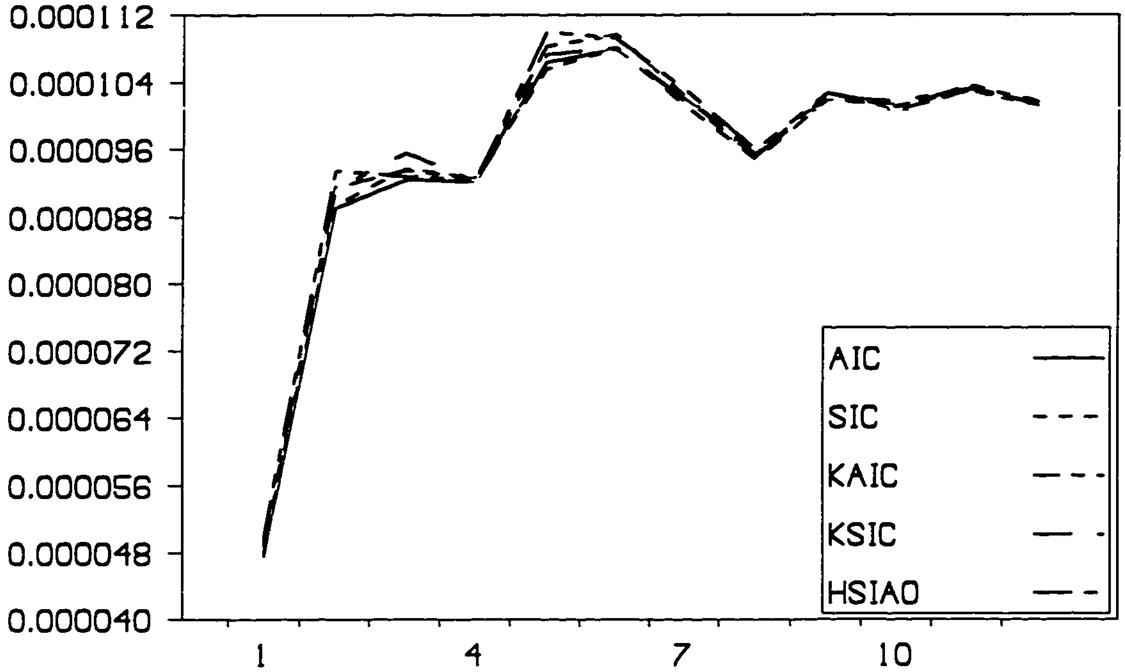

Figure 2.17
Larger Sample Model 1 msfe of the Second Variable

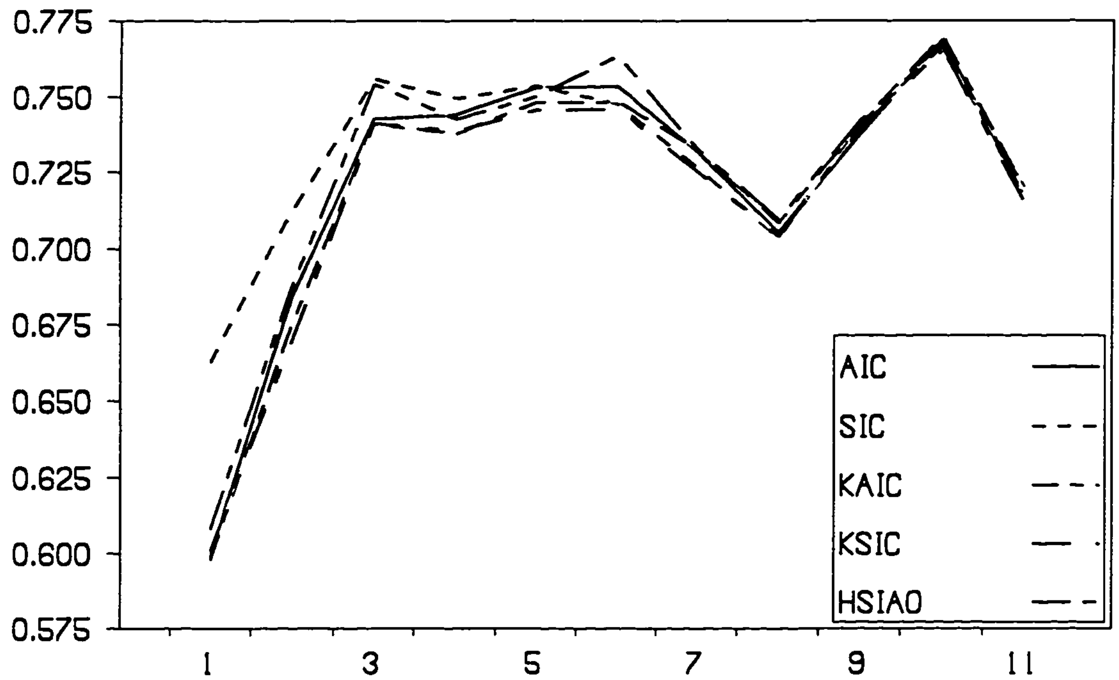

Figure 2.18
Larger Sample Model 2 msfe of the First Variable

Table 2.21
Model 4, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.796	0.793*	0.802	0.801	0.808	0.763**
V2($\times 10^{-4}$)	2.815	2.786	2.785*	2.839	2.780**	2.810
Panel B						
V1	0.825	0.817*	0.838	0.820	0.839	0.816**
V2($\times 10^{-4}$)	3.333	3.315	3.314*	3.378	3.246**	3.409

For explanations see the footnote of Table 2.18.

Table 2.22
Model 5, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.775**	0.790	0.781*	0.790	0.786	0.795
V2($\times 10^{-4}$)	1.194**	1.285	1.301	1.208*	1.290	1.216
Panel B						
V1	0.839	0.811**	0.859	0.813	0.837	0.812*
V2($\times 10^{-4}$)	1.312*	1.330	1.330	1.350	1.309**	1.334

For explanations see the footnote of Table 2.18.

Model 4 is the asymmetric VAR for which we observed the explosive IRFs. The msfe of this model is presented in Table 2.21. HSIAO has the lowest msfe for the first variable and KSIC has the lowest msfe for the second variable. Similarly, as in the IRF mse, the msfe of the second variable increases as the forecast horizon increases.

The forecast mse of Model 5 is reported in Table 2.22. This is a symmetric VAR model, and, as in the previous cases for symmetric VARs, AIC has the best

Table 2.23
Model 6, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.772**	0.784	0.772**	0.779	0.775	0.782
V2($\times 10^{-4}$)	1.341	1.333*	1.334	1.340	1.326**	1.340
Panel B						
V1	0.824	0.794	0.828	0.793*	0.822	0.791**
V2($\times 10^{-4}$)	1.425	1.367*	1.375	1.444	1.351**	1.460

For explanations see the footnote of Table 2.18.

performance. However, in the smaller sample, AIC does not perform as well. Usually, AIC has the smallest short-run msfe followed by KAIC (not reported here), and, in the long-run SIC (and PIC) and KSIC have the smallest msfe.

Model 6 is an asymmetric model. Its msfe is presented in Table 2.23. Surprisingly, KAIC does not show superiority, especially in the larger sample. KSIC is the best forecaster of the second variable; however, no method dominates the forecast performance for the first variable. Figure 2.19 shows the msfe of the first variable. KAIC has the lowest msfe in the short-run. In the long-run, SIC has the lowest msfe, and HSIAO has the largest. The msfe plot of the second variable (not shown here) is similar to Figure 2.17. KAIC and AIC have the smallest mse in the short-run.

The last two models have three variables. Model 7 is a symmetric VAR model. Table 2.24 presents the msfe. AIC has the lowest msfe for variables one and three followed by KAIC. SIC has the lowest msfe for the second variable. HSIAO

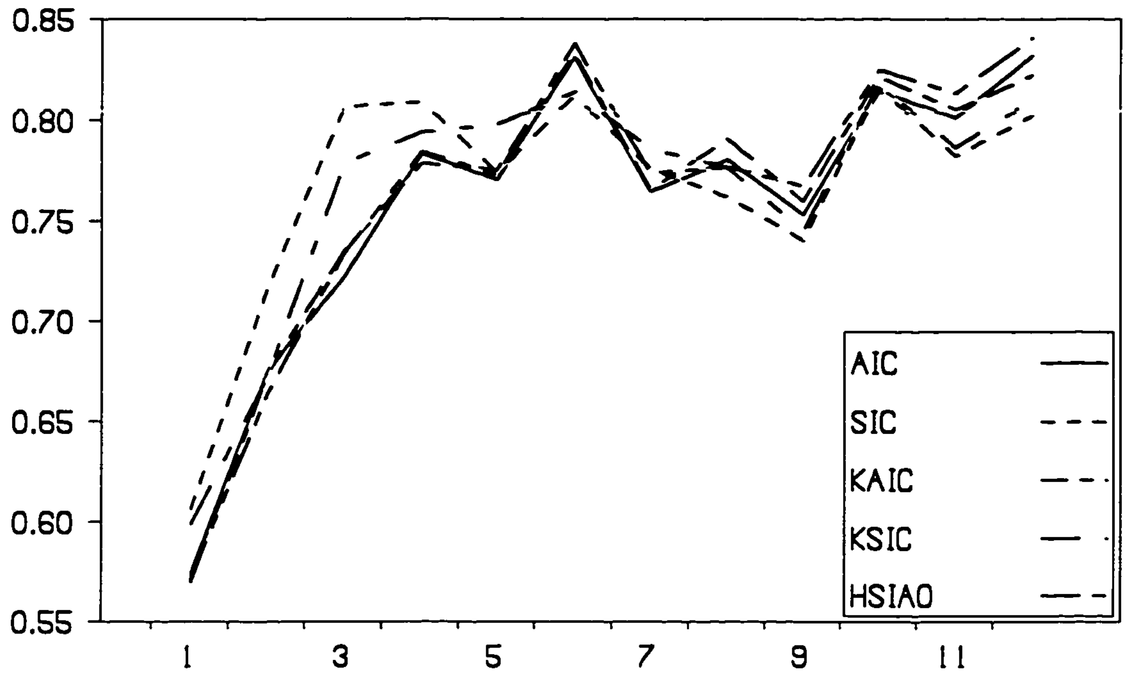

Figure 2.19
Larger Sample Model 6 msfe of the First Variable

has the worst performance. In Panel B, KSIC has the best performance followed by SIC.

Finally, forecast mse's of Model 8 are presented in Table 2.25. AIC has the best forecasting performance in the larger sample, and, in the smaller sample, KSIC and SIC perform well. KAIC has the lowest short-run msfe for the second variable, and SIC and KSIC have the lowest long-run msfe's of these variables as shown in Figures 2.20 and 2.21.

In this section, AIC has the best short-run forecasting performance for symmetric VAR models. However, in small samples, SIC and KSIC more often have a smaller mse than AIC. For asymmetric VAR models, KSIC has the best performance; KAIC usually has the lowest msfe only in the short-run. Even though

Table 2.24
Model 7, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.757**	0.770	0.715	0.763*	0.763*	0.777
V2($\times 10^{-4}$)	0.848	0.842**	0.843*	0.852	0.846	0.860
V3($\times 10^{-4}$)	1.209*	1.225	1.231	1.218	1.208**	1.229
Panel B						
V1	0.784**	0.793	0.793	0.796	0.786*	0.809
V2($\times 10^{-4}$)	0.892	0.885**	0.887	0.906	0.885**	0.913
V2($\times 10^{-4}$)	1.255	1.248*	1.250	1.267	1.245**	1.279

For explanations see the footnote of Table 2.18.

Table 2.25
Model 8, Forecast mse's

Panel A	AIC	SIC	PIC	KAIC	KSIC	HSIAO
V1	0.732**	0.741	0.744	0.738	0.734*	0.775
V2($\times 10^{-4}$)	0.834*	0.844	0.849	0.838	0.819**	0.866
V3($\times 10^{-4}$)	1.128*	1.137	1.145	1.136	1.121**	1.172
Panel B						
V1	0.904	0.787*	0.789	0.886	0.783**	0.834
V2($\times 10^{-4}$)	1.001	0.881*	0.885	0.977	0.873**	0.925
V2($\times 10^{-4}$)	1.296	1.098*	1.100	1.256	1.095**	1.196

For explanations see the footnote of Table 2.18.

SIC and KSIC underspecify the lag length, they outperform AIC and KAIC which more closely specify the true lag length. This peculiar result has also been observed by Hafer and Sheehan (1989).

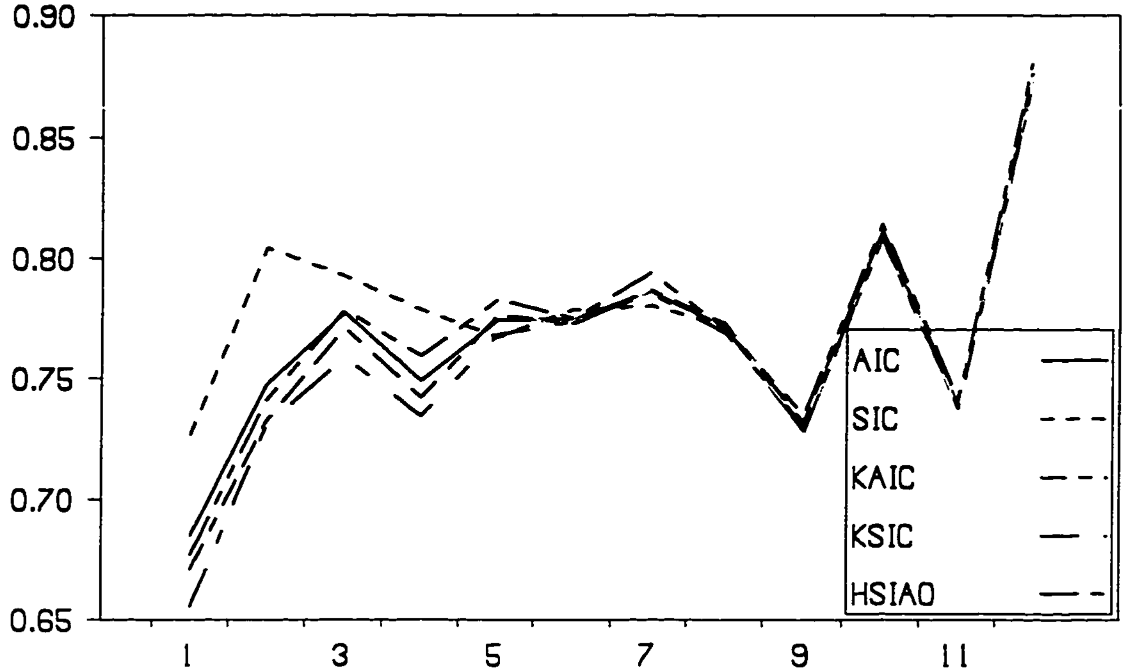

Figure 1.20
Larger Sample Model 8 msfe of the First Variable

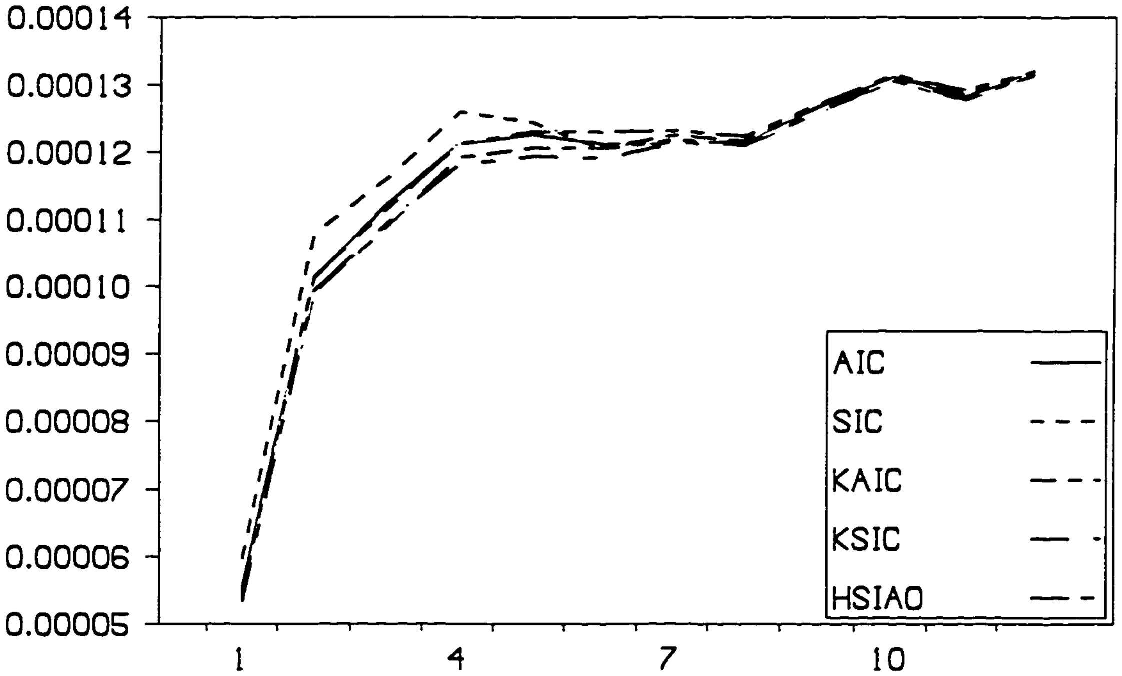

Figure 1.21
Larger Sample Model 8 msfe of the Second Variable

2.4.4 The Longer Period Performance of The SIC Criterion

A counter-intuitive property of the SIC and KSIC is that they usually have long-run mse's of IRFs and forecast errors lower than those of AIC and KAIC, even though the former methods tend to underspecify the lag length. To shed light on this puzzle, we simulated Model 1 1000 times and generated 200 observations each time and discarded the first 74. We estimated two symmetric VARs, one correctly specified with three lags, and the other misspecified with two lags. Then we computed the mse of IRFs, VDCs, and the forecasts.

The misspecified VAR yielded long-run IRF mse's lower than those of the correctly specified VAR and yielded high peaks in the short-run mse, except at the response of the second variable to a shock to the second variable IRF. In the response of the second variable to a shock to the second variable case, the misspecified model had lower mse in the short-run and a high mse in the long run. The msfe of the misspecified model is only smaller than the correct model in the long-run mse of the first variable.

Christiano and Eichenbaum (1989) provide evidence on the effect of underspecification on unit root tests. They show that the fit at low frequency is sacrificed when an ARMA(2,2) model is estimated from a data set which is created from an ARMA(3,3) model. Christiano and Eichenbaum argued that maximum likelihood seeks to minimize the average percentage error of the discrepancy between the theoretical spectral density matrix of the fitted model and the true spectral density matrix. Thus, an underspecified model cannot completely fit the data at all

frequencies. Based on Christiano and Eichenbaum's conclusion, our results can be interpreted as showing that the misspecified model does not completely fit the data, and it usually sacrifices the short-run fit to have a better long-run fit.

2.5 Conclusion

The eight experiments that have been conducted used different VAR models to control for certain factors such as the number of variables and the lag structure.

Table 2.26 is a summary table where the mean lag specifications in the larger sample are reported. The standardized mean lag value is the ratio of the estimated mean lag length presented previously (Tables 2.2 - 2.9), and the true lag length. Thus, standardized mean lag value closer to one indicates estimated mean lag value closer to the true lag length. The numbers in parenthesis are the corresponding standard deviations scaled by 10^{-2}. The results show that the SIC and PIC criteria tend to choose lower lag lengths than the true ones (for a sample size of 124 or lower). The AIC criterion more often estimates the correct lag length than the SIC and PIC criteria. These results are consistent with Lütkepohl (1993, Tables 4.6 and 4.7), Hafer and Sheehan (1989), and Geweke and Meese (1981). For asymmetric VAR cases, Keating's method with the AIC criterion does relatively well in estimating the lag structure. For symmetric VAR cases, the symmetric lag specification method with AIC criterion does better than Keating's method with AIC criterion. The results show that Keating's method performs better than the equation-by-equation lag specification method, Hsiao's method, in estimating the lag structure, except in

Table 2.26
Summary Standardized Mean Lag Specification

Model 1	a11	a12	a21	a22	Model 2	a11	a12	a21	a22
AIC	1.07	1.07	1.07	1.07		1.03	3.10	1.03	3.10
	(0.85)	(0.85)	(0.85)	(0.85)		(1.01)	(0.03)	(1.01)	(0.03)
SIC	0.87	0.87	0.87	0.87		0.49	1.46	0.49	1.46
	(0.51)	(0.51)	(0.51)	(0.51)		(0.79)	(2.28)	(0.79)	(2.28)
PIC	0.85	0.85	0.85	0.85		0.46	1.39	0.46	1.39
	(0.53)	(0.53)	(0.53)	(0.53)		(0.75)	(2.25)	(0.75)	(2.25)
KAIC	1.11	1.09	1.11	1.09		1.12	1.62	1.12	1.62
	(1.21)	(1.28)	(1.21)	(1.28)		(1.18)	(4.15)	(1.18)	(4.15)
KSIC	0.86	0.72	0.86	0.72		0.80	1.06	0.80	1.06
	(0.64)	(0.98)	(0.64)	(0.98)		(0.92)	(0.83)	(0.92)	(0.83)
HSIAO	1.45	0.93	1.15	1.35		1.21	1.13	1.04	2.11
	(1.15)	(2.02)	(1.50)	(1.39)		(1.48)	(5.27)	(1.64)	(5.33)
Model 3					Model 4				
AIC	0.91	2.72	1.36	2.72		1.10	3.29	3.29	1.10
	(1.06)	(3.19)	(1.59)	(3.19)		(0.85)	(2.56)	(2.56)	(0.85)
SIC	0.41	1.20	0.61	1.20		1.00	2.99	2.99	1.00
	(0.49)	(1.48)	(0.74)	(1.48)		(0.12)	(0.36)	(0.36)	(0.12)
PIC	0.39	1.19	0.59	1.19		1.00	2.99	2.99	1.00
	(0.45)	(1.34)	(0.67)	(1.34)		(0.13)	(0.39)	(0.39)	(0.13)
KAIC	1.04	1.55	1.56	1.55		0.98	3.52	2.95	1.17
	(1.23)	(3.69)	(1.84)	(3.69)		(1.44)	(3.42)	(0.43)	(1.11)
KSIC	0.63	1.10	0.95	1.10		0.48	3.04	1.40	1.01
	(0.88)	(0.85)	(1.32)	(0.85)		(0.79)	(0.64)	(2.38)	(0.21)
HSIAO	1.18	1.21	0.98	2.46		1.95	1.83	1.80	1.22
	(1.40)	(5.34)	(2.44)	(5.48)		(1.75)	(5.07)	(4.97)	(1.30)
Model 5					Model 6				
AIC	1.01	1.01	1.01	1.01		0.86	1.38	0.86	1.38
	(0.52)	(0.52)	(0.52)	(0.52)		(0.59)	(0.94)	(0.59)	(0.94)
SIC	0.45	0.45	0.45	0.45		0.50	0.80	0.50	0.80
	(0.60)	(0.60)	(0.60)	(0.60)		(0.51)	(0.82)	(0.51)	(0.82)
PIC	0.40	0.40	0.40	0.40		0.45	0.73	0.45	0.73
	(0.53)	(0.53)	(0.53)	(0.53)		(0.48)	(0.78)	(0.48)	(0.78)
KAIC	0.94	1.03	0.94	1.03		0.94	1.05	0.94	1.05
	(0.84)	(0.79)	(0.84)	(0.79)		(0.73)	(1.24)	(0.73)	(1.24)
KSIC	0.48	0.36	0.48	0.36		0.60	0.60	0.60	0.60
	(0.64)	(0.88)	(0.64)	(0.88)		(0.78)	(1.06)	(0.78)	(1.06)
HSIAO	1.00	0.72	0.88	1.06		0.99	1.02	0.94	1.39
	(1.07)	(1.07)	(1.20)	(0.94)		(0.87)	(1.56)	(1.04)	(1.38)

(Table con'd.)

Model 7	a11	a12	a13	a21	a22	a23	a31	a32	a33
AIC	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	(0.51)	(0.51)	(0.51)	(0.51)	(0.51)	(0.51)	(0.51)	(0.51)	(0.51)
SIC	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
	(0.62)	(0.62)	(0.62)	(0.62)	(0.62)	(0.62)	(0.62)	(0.62)	(0.62)
PIC	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)	(0.40)
KAIC	1.06	0.74	1.03	1.06	0.74	1.03	1.06	0.74	1.03
	(1.18)	(1.14)	(1.16)	(1.18)	(1.14)	(1.16)	(1.18)	(1.14)	(1.16)
KSIC	0.72	0.45	0.64	0.72	0.45	0.64	0.72	0.45	0.64
	(0.88)	(0.51)	(0.84)	(0.88)	(0.51)	(0.84)	(0.88)	(0.51)	(0.84)
HSIAO	0.33	0.71	1.26	1.05	0.83	0.65	0.98	0.66	1.24
	(0.00)	(1.94)	(1.78)	(1.95)	(1.83)	(1.90)	(1.81)	(1.93)	(2.03)
Model 8									
AIC	0.88	1.32	1.32	0.88	1.32	1.32	0.88	1.32	1.32
	(0.60)	(0.89)	(0.89)	(0.60)	(0.89)	(0.89)	(0.60)	(0.89)	(0.89)
SIC	0.34	0.52	0.52	0.34	0.52	0.52	0.34	0.52	0.52
	(0.50)	(0.73)	(0.73)	(0.50)	(0.73)	(0.73)	(0.50)	(0.73)	(0.73)
PIC	0.30	0.45	0.45	0.30	0.45	0.45	0.30	0.45	0.45
	(0.40)	(0.61)	(0.61)	(0.40)	(0.61)	(0.61)	(0.40)	(0.61)	(0.61)
KAIC	1.04	0.72	1.09	1.04	0.72	1.09	1.04	0.72	1.09
	(0.84)	(1.61)	(1.40)	(0.84)	(1.61)	(1.40)	(0.84)	(1.61)	(1.40)
KSIC	0.73	0.34	0.51	0.73	0.34	0.51	0.73	0.34	0.51
	(0.88)	(0.41)	(0.89)	(0.88)	(0.41)	(0.89)	(0.88)	(0.41)	(0.89)
HSIAO	0.17	0.83	1.17	0.96	1.11	0.64	1.05	0.54	1.43
	(0.24)	(2.56)	(2.29)	(1.32)	(2.50)	(2.23)	(1.40)	(2.34)	(2.42)

a_{ij} is the lag length of the j^{th} variable in the i^{th} equation. The numbers in the parenthesis are the standard deviation of the standardized mean selected lag length scaled by 10^{-2}.

asymmetric cases where each equation has different lag lengths in each equation (first degree asymmetry).

Besides estimating the frequencies of specified lag lengths, IRF and mean square forecast errors of the different methods are compared. The summary of IRF performance is presented in Table 2.27 which identifies the best three IRF performances (the average of the 24 periods) for each case and model in the larger

Table 2.27
IRF Response Summary

	Model 1						Model 2					
	AIC	SIC	PIC	KAIC	KSIC	HSIAO	AIC	SIC	PIC	KAIC	KSIC	HSIAO
IRF_{11}	(3)	(1)	(2)				(2)			(1)		(3)
IRF_{12}	(1)	(3)		(2)			(3)			(2)	(1)	
IRF_{21}	(1)			(2)		(3)		(3)	(2)		(1)	
IRF_{22}	(2)	(1)	(3)					(3)	(2)		(1)	

	Model 3						Model 4					
	AIC	SIC	PIC	KAIC	KSIC	HSIAO	AIC	SIC	PIC	KAIC	KSIC	HSIAO
IRF_{11}	(1)			(1)		(2)	(3)	(1)	(2)			
IRF_{12}	(2)			(3)	(1)			(2)	(3)		(1)	
IRF_{21}		(3)	(2)		(1)			(1)	(2)			(3)
IRF_{22}		(3)	(2)		(1)		(3)	(1)	(2)			

	Model 5						Model 6					
	AIC	SIC	PIC	KAIC	KSIC	HSIAO	AIC	SIC	PIC	KAIC	KSIC	HSIAO
IRF_{11}	(1)			(2)		(3)	(2)			(1)		(3)
IRF_{12}	(1)			(2)		(3)	(2)			(1)		(3)
IRF_{21}	(1)			(2)		(3)	(2)			(1)		(3)
IRF_{22}	(1)			(2)		(3)	(2)			(1)		(3)
IRF_{13}												
IRF_{23}												
IRF_{31}												
IRF_{32}												
IRF_{33}												

	Model 7						Model 8					
	AIC	SIC	PIC	KAIC	KSIC	HSIAO	AIC	SIC	PIC	KAIC	KSIC	HSIAO
IRF_{11}	(1)			(2)	(3)		(1)			(2)	(3)	
IRF_{12}	(3)	(1)	(2)				(1)			(2)	(3)	
IRF_{21}	(3)	(2)	(1)					(2)	(1)	(3)		
IRF_{22}	(1)	(2)	(3)				(1)	(3)		(2)		
IRF_{13}	(1)			(2)	(3)		(2)			(1)	(3)	
IRF_{23}		(2)	(3)		(1)		(3)			(2)	(1)	
IRF_{31}	(1)			(2)	(3)		(1)			(2)	(3)	
IRF_{32}	(1)	(2)		(3)			(1)			(3)		(3)
IRF_{33}	(1)			(2)	(3)		(2)			(3)	(1)	

(1) marks the method which has the lowest IRF mse, (2) marks the second lowest and (3) marks the third lowest.

sample simulation. In symmetric VARs (Models 1, 5 and 7), AIC has the best IRF mse performance. KAIC performs well in Models 6 and 8, and KSIC performs well in Models 2 and 3. The performances of SIC and PIC are very close to each other. SIC, PIC and KSIC have large swings in IRF mse's in the shorter periods, but have lower mse's in the longer periods than the other methods. Since the SIC and PIC criteria underestimate the lag length most of the time, the econometric model will not fit the sample information well. The evidence indicates that in underspecified models the high frequency fit of the data is sacrificed to have better lower frequency fit. For the symmetric VAR models, AIC most often has the lowest IRF mse, and, for the asymmetric VARs, KAIC most often has the lowest short-run IRF mse. HSIAO usually has the largest mse's in the long run even in first degree asymmetric models.

The forecast performance is summarized in Table 2.28 which identifies the lowest three forecast mses for each variable and model in the larger sample simulation. For symmetric VARs usually AIC has the lowest msfe. In asymmetric VAR cases, KSIC performs better. KAIC has low msfe's only in the short-run forecast. However, overall, AIC and KSIC usually have the lowest average msfe. In the forecast performance, too, the equation-by-equation specification method does not do well.

In conclusion, if we look at the overall performance it is preferred to use a method which employs the AIC criterion rather than SIC or PIC. However, if is the longer periods that we are interested in, SIC or PIC will yield lower long-run IRF mse. The results show that, in asymmetric models, the difference between the

Table 2.28
Forecast Performance Summary

	Variable	AIC	SIC	PIC	KAIC	KSIC	HSIAO
Model 1	V1	(1)	(2)				(3)
	V2	(3)	(1)		(2)		
Model 2	V1	(3)		(2)	(1)	(3)	
	V2		(3)	(2)		(1)	
Model 3	V1	(3)		(1)	(1)		
	V2		(1)	(1)		(3)	
Model 4	V1	(3)	(2)				(1)
	V2		(3)	(2)		(1)	
Model 5	V1	(1)		(2)		(3)	
	V2	(1)			(2)		(3)
Model 6	V1	(1)		(1)		(3)	
	V2		(2)	(3)		(1)	
Model 7	V1	(1)			(2)	(2)	
	V2		(1)	(2)		(3)	
	V3	(2)			(3)	(1)	
Model 8	V1	(1)			(3)	(2)	
	V2	(2)			(3)	(1)	
	V3	(2)			(3)	(1)	

(1) marks the method which has the lowest IRF mse, (2) marks the second lowest and (3) marks the third lowest

symmetric lag specification model with the AIC criterion and Keating's method with the AIC criterion is not substantial. However, even though there is not a big loss using the symmetric lag specification method with AIC criterion, Keating's method has some advantage, especially in IRF computations. When the VAR model becomes larger and more asymmetric, this advantage becomes substantial.

CHAPTER 3

MULTIVARIATE SHRINKAGE ESTIMATION FOR IMPROVEMENT OF FORECASTING PERFORMANCE

3.1 Introduction

Time series techniques have been a popular method in forecasting and model building since the introduction of autoregressive-integrated-moving average (ARIMA) models by Box and Jenkins (1970). ARIMA models are usually very parsimonious in the number of parameters since the dependent variable is explained only by its lags and by the current innovation and its lags.

ARIMA models are a univariate time series technique which do not incorporate the information that may be contained in other relevant variables. Sims (1980) suggested an extension of ARIMA models which is a multivariate time series modeling technique called a vector autoregressive (VAR) model. In a VAR model no restrictions are imposed on the way the variables interact. All the variables in the VAR are considered to be endogenous and are explained by their own lags and the lags of all other variables in the model. Economic theory is used to specify which variables enter the model. VAR models are parsimonious in the number of variables since only a few variables are included in the model. In contrast, structural models often have hundreds of variables. A VAR model can be viewed as the reduced form of a structural model of the same dimension; however, knowledge of the reduced form parameters is enough in forecasting. Therefore, forecasting with VARs is relatively easy and inexpensive compared to structural models.

Even though VAR models are parsimonious in the number of variables, often they are not parametrically parsimonious. The more variables and the longer the lags on the variables in the model, the more parameters that have to be estimated. After a certain number of variables, we face the danger of an overparametrized model. Lütkepohl (1993), in a Monte-Carlo experiment framework, shows that VARs with higher lag orders than the true underlying model generate out-of-sample forecasts with higher mean-square-forecast-error (msfe).[1] Thus, for forecasting purposes, lags longer than the true lags are not advantageous.

Overparametrization can be avoided by restricting the number of variables entering the VAR or restricting the lag length. However, Braun and Mittnik (1993) prove that VAR model estimators will be inconsistent if VAR models exclude relevant variables or have higher or lower lag orders than the true model. With inconsistent estimators, coefficient estimates do not converge to their true values. Moreover, Lütkepohl (1993) shows that the residuals from an estimated VAR which has a lag length less than the true underlying model are autocorrelated. The consequence of autocorrelation is biased and inefficient predictors.[2] Overparametrization may decrease the bias of the forecast; however, it may also increase the variance of the

---

[1]Mean-square-forecast-error is defined for the i^{th} variable in the model as,

$$\frac{1}{H}\sum_{h=1}^{H}(y_{i,t+h}-y_{i,t}(h))^2$$

where $y_{i,t}(h)$ is the h-period ahead forecast of $y_{i,t+h}$ made at period t, and H is the total forecast horizon. The total number of forecasts is H.

[2]Judge et al. (1988), pp 402-405.

forecast. On the other hand, putting restrictions on the parameter space will lower the forecast variance; however, if the restrictions are not correct, the forecast will be biased. Thus, unless correct restrictions are imposed, there usually exists a tradeoff between lower bias and higher variance (msfe equals the sum of the variance of the forecast and the square of the bias).

A possible solution to inefficient forecast estimation was proposed by Litterman (1980) and extended by Doan, Litterman, and Sims (1984), who utilized non-sample information by imposing prior distributions on the model's parameters. The additional information supplied by the prior distribution of the parameters in the system yields more accurate estimates of the parameters, which are expected to improve the forecast performance if the imposed prior beliefs adequately describe the true underlying model. This method is called Bayesian vector autoregression (BVAR) in the literature since it imposes a prior distribution on parameters. The prior can be thought of as restrictions imposed on the parameters. These restrictions are not derived from economic theory. For example, the "Minnesota prior" of Litterman assumes that the variables follow a random walk. The restrictions of the prior distribution are different from restrictions imposed on least squares estimates, because if the prior does not adequately represent the true model, the data have the opportunity to override the restrictions. However, in least squares estimation, the sample information does not have such an opportunity.

Another popular shrinkage estimator is the Stein-rule estimator. Like its Bayesian counterpart, the Stein-rule estimator shrinks the estimates toward a known

prior vector by imposing restrictions on the estimator and allowing the sample information to override the restrictions if the restrictions are not supported by the data. Hill et al. (1991) and Knight et al. (1992) show that forecasts from Stein-rule estimators usually have lower msfe than ordinary least squares (OLS) forecasts.

In this study, we propose another multivariate forecasting time-series technique, which we will call the Stein-rule VAR (SRVAR). In the Stein-rule VAR, parameter estimates are obtained via a Stein-rule estimator instead of conventional maximum likelihood (ML) or OLS estimators. SRVAR forecasts are easier to compute than BVAR forecasts since BVAR estimation requires specification of a prior distribution and some parameters of that prior distribution. For Stein-rule estimation, the only necessary non-sample information is the restrictions that are imposed on the parameter space. However, contrary to BVARs, a SRVAR has not been used before; thus, its forecasting ability is unknown.

Typically a VAR estimate assumes that all variables have the same lag length in each equation. However, Keating (1994) argues that it is possible to have VAR models in which each variable has a different lag length. He proposes a method to estimate VAR models with asymmetric lag structure. Keating assumes that the variables of a VAR can have different lag lengths in an equation; however, a variable will have the same lag length in different equations. A second method devised by Hsiao (1981) estimates the lag structures of an asymmetric VAR in which each variable can have different lag lengths in different equations.

In this essay, we compare the forecasting performance of unrestricted VAR to four shrinkage VAR estimator forecasts. Bayesian VARs and Stein-rule VARs shrink the parameter values while asymmetric VARs often decrease the number of parameters in the VAR model. The h-period-ahead forecasts of the variables of interest will be compared based on the Theil U statistic. The Theil U statistic is the ratio of the rmsfe of the forecasting model to the rmsfe of a naive model such as a random walk model (rmsfe is the square root of the msfe statistic). The appealing part of the Theil U statistic is that it gives a unit free measure which makes comparisons across variables possible. Further discussion of forecast performance criteria can be found in Fildes (1992).

The forecasting performance of an estimator may change from variable to variable. Thus, if an estimation method has the best forecasting performance for one variable, this does not necessarily mean that it will have the best forecasting performance for all variables. The Theil U statistic can only measure the performance for one variable. However, since the Theil U statistic is a unit-free measure, overall forecast performance can be examined by comparing the sum of the Theil U statistics of the variables that are forecasted. Each Theil U statistic is a ratio of two rmsfes. Thus, it is a measure of relative forecast error. Then, the sum of the Theil U statistics can be interpreted as a measure of total relative forecast error of a forecasting method. A second method of evaluating overall forecast performance was suggested by Sims (1982). Sims' overall forecast performance of an estimation method for variables in the system that are forecasted is evaluated by examining the

log of the determinant of the sample covariance matrix of the forecast errors of the forecasted variables.

We also compare statistically the forecast accuracies of the different forecasting models with a null hypothesis of equal forecast accuracy. Within a Monte-Carlo experiment framework, Diebold and Mariano (1995) show that the sign-test has desirable finite-sample performance. The sign-test compares the forecast errors from two different forecast methods. Let d_t be the difference of the two forecast errors and T the sample size of the forecast; then the test statistic is as follows:

$$S_{2a} = \frac{S_2 - 0.5T}{\sqrt{0.25T}} \sim N(0,1) \quad ,$$

where

$$S_2 = \sum_{t=1}^{T} I(d_t) \quad \text{and} \quad I(d_t) = \begin{cases} 1 & \text{if } d_t > 0 \\ 0 & \text{otherwise} \end{cases} \quad .$$

Thus, if a forecasting method consistently yields lower or higher forecast errors than another method, the sign test will differentiate between the forecast accuracies of these two methods. However, if, in the pairwise comparison, a method does not have lower or higher forecast errors sufficiently more often than another forecasting model, the two forecast performances will not be distinguishable from each other. In this case, the value of the sign-test statistic will be low, and the test will not reject equal forecast accuracies.

The empirical study of the forecasting performance of multivariate shrinkage estimators consists of two parts. The first part uses real-time macroeconomic data to

investigate the relative forecasting performance of the estimators. The second part reports the results of Monte-Carlo experiments comparing the forecast performances of BVARs and SRVARs.

3.2 Literature Review

There are various studies comparing VAR forecasts with forecasts of professional forecasting services. Litterman (1979) compared a VAR forecast with seven major forecasters from 1970-1975, and found that VAR forecasts performed better in many cases, especially at longer horizons. Webb (1984) compared forecasts from a major consulting service, a survey of professional forecasters, and a VAR model from 1977-1983. Average values from the survey are used as the forecast. He found that VAR forecasts of inflation had the worst accuracy among the forecasting methods. In a real GNP growth forecast, the VAR does not perform as well as the survey at the one-quarter-ahead forecast but performs as well at the four-quarter-ahead. VAR forecasts of real GNP growth and the commercial paper rate were better than the professional service forecasts. Lupoletti and Webb (1986) show that short-run forecasts (one and two quarters ahead) of inflation, the growth rate of real GNP, and the 90-day T-bill rate obtained from a VAR model which also contains the capacity utilization rate and monetary base are substantially less accurate than the forecasts produced by major forecasting services. At longer horizons (four and six quarters ahead), however, the VAR forecasts were competitive with those of the forecasting services. Thus, the evidence shows that VAR forecasts are competitive with expensive professional forecasts, especially at longer horizons.

BVAR forecasts are shown to be superior to univariate ARIMA methods by Doan, Litterman and Sims (1984). Furthermore, Litterman (1986) shows that a BVAR's forecasts are better in most cases compared to univariate autoregressive and ARIMA forecasts, especially for long-run forecasts. Litterman also compares the performance of BVARs to various commercial firms' forecasts. Based on the root-mean-forecast-error criterion, he concludes that the BVAR performs as well as structural and sophisticated forecasting models for the period 1976-1979. However, for 1980-1984, BVAR forecasts of the GNP deflator do not perform as well as the professional forecasters'. McNees (1986) computes and compares BVAR forecasts with professional service forecasts. He shows that BVAR forecasts are generally the most accurate for real GNP, the unemployment rate, and real residential investment, and among the least accurate for the implicit price deflator and the 90-day T-Bill rate. Other studies which show that BVAR models are at least as accurate as forecasts from structural economic model are Dua and Smyth (1995) and Sarantis and Steward (1995).

Lupoletti and Webb (1986) compare an unrestricted VAR forecast with a BVAR forecast for 1980-1983 and find that BVAR forecasts of real GNP and 90-day Treasury-bill rate are more accurate than the unrestricted VAR. However, unrestricted VAR forecasts of the GNP deflator are more accurate than BVAR forecasts. Fackler, McMillin and Silver (1990) forecast nominal GNP and personal income in a system where different measures of monetary aggregates are included. Their results show that, in almost all cases, BVAR forecasts with Litterman's prior

distribution performed better than the unrestricted VAR forecast in one-year-ahead out-of-sample forecasting. Liu et al. (1994) show that BVAR models have much better exchange rate forecasts than unrestricted VARs.

Kadiyala and Karlsson (1993) compare BVAR models with an unrestricted OLS forecast. The priors of the models are the Minnesota prior, Normal-Wishart prior, diffuse prior, normal-diffuse prior, and extended natural conjugate prior. The OLS forecast performs significantly worse than BVAR forecasts. Among the BVARs, the diffuse and Normal-Wishart priors do especially well. This result shows that different prior distributions may display different forecast performances.

In summary, the results show that the forecasting accuracy of unrestricted VAR models is comparable to the forecasting accuracy of expensive structural models, especially in long-run forecasts. VARs often suffer from overparametrization which tends to lower the forecasting accuracy. BVAR models are designed to overcome this problem by imposing restrictions on the parameter space. The forecasting evidence indicates that, generally, BVARs have a better forecasting performance than VAR and structural models. However, BVARs do not always have the most accurate forecasts.

Stein-rule estimators can be derived from empirical Bayes estimators and are shown to have lower risk functions than maximum-likelihood (ML) and OLS estimators under the square-forecast-error loss criterion.[3,4] A lower risk function

---

[3]See Judge and Bock (1978).

generally indicates a lower average loss incurred using that particular estimator which is expected to yield lower forecast loss. As in the Bayesian case, Stein-rule estimators also incorporate non-sample information. Thus, it is expected that forecasts obtained from this estimator will be improved relative to unrestricted VAR forecasts. Hill et al. (1991) and Knight et al. (1992) show that single equation Stein-rule estimators have lower msfe than the single equation OLS estimator forecasts. The Stein-rule VAR will be the first of its kind in the VAR literature to our knowledge.

Not much is known about the forecasting performances of asymmetric VARs. Based on a Monte-Carlo experiment, the previous essay of this dissertation shows that Keating's method generally has better forecasting performance than symmetric VARs if the true model is asymmetric. Hsiao's method of lag specification does not yield a more accurate forecast. This essay will further investigate the forecasting performance of these asymmetric VARs using actual macroeconomic data.

3.3 Theoretical Metholology

Bayesian econometricians estimate model parameters by first imposing a prior distribution, and then finding the posterior distribution of the parameters conditioned on the sample data. Bayesians express their uncertainty about the value of an unknown parameter in terms of a probability distribution, which is called the prior distribution. Bayesians obtain the posterior distribution of the parameters by

[4]The risk function is defined as $R(b,\beta)=E(b-\beta)'X'X(b-\beta)$ where E is the expectation operator, and b is the estimator of β.

minimizing an average of a loss function, where the average is calculated from the losses that would be obtained from repeated samples.

To obtain more accurate forecasts, Litterman (1979) advocated using a Bayesian estimator to estimate the coefficients of a VAR. Litterman's analysis is based on a VAR written as:

$$y_t = D + \sum_{k=1}^{p} \beta^k y_{t-k} + e_t \qquad t=1,\ldots,T$$

$$E[e_t e_s'] = \begin{cases} \Sigma & \text{if } s=t \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

y_t is the $N \times 1$ variable vector, D is the $N \times 1$ deterministic component vector, β^k is a $N \times N$ coefficient matrix, e_t is the $N \times 1$ error vector, and p is the lag length of the VAR. Each equation of the VAR can be written as:

$$y_{i,t} = D_i + \sum_{k=1}^{p} \beta_i^k y_{t-k} + e_{i,t} \qquad t=1,\ldots,T. \tag{3.2}$$

β_i^k is the i^{th} row of the β^k matrix. It can be seen from (3.2) that the explanatory variables are the same in all equations. Litterman (1986) assumes that the coefficients have a prior normal distribution,

$\beta_{i,j}^k \sim N(1,(\delta_{ij}^k)^2)$ for $k=1$, and $i=j$

$\beta_{i,j}^k \sim N(0,(\delta_{ij}^k)^2)$ otherwise.

Statistically, it is assumed that all variables are random walks. The first own lag of a variable has a unit prior mean, and the rest of the coefficients have a zero prior mean. The prior variances decrease as the lag increases; this can be formulated as,

$$\delta_{ij}^{k} = \begin{cases} \dfrac{\lambda}{k^{d}} & \text{if i=j} \\ \dfrac{\theta\lambda\hat{\sigma}_{i}}{k^{d}\hat{\sigma}_{j}} & \text{if i}\neq\text{j} \end{cases} \tag{3.3}$$

where δ_{ij}^{k}= the standard deviation of the prior distribution of the j^{th} variable's k^{th} lag in equation i, d is the decay parameter defining the speed of decaying towards zero, θ= weight of cross variable lags, and λ= prior variance of first own lag. $\hat{\sigma}_i$ and $\hat{\sigma}_j$ are the estimated standard error of the unrestricted OLS residuals from the i^{th} and the j^{th} equations, respectively. This prior distribution is called the Minnesota prior (or sometimes the Litterman prior) in the BVAR literature. Note that we impose no prior distribution on the deterministic term D.

Since the explanatory variables are the same in each equation, we can estimate each equation's coefficients separately using OLS. Therefore, the discussion to follow will be based on the i^{th} equation of the VAR. The result can be applied to any equation in the VAR without any difficulty. If we define the i^{th} equation as,

$$Y = X\beta + e\ , \tag{3.4}$$

where X contains all lagged variables and the deterministic components, and $e \sim N(0,\sigma^2 I)$, then the imposed prior can be represented as,

$$\beta \sim N(\bar{\beta},\bar{\Sigma}_{\beta})\ , \tag{3.5}$$

Where $\bar{\beta}$ is the prior imposed coefficient values and $\bar{\Sigma}_{\beta}$ is the prior imposed covariance matrix of the coefficients. Then, it can be shown that the posterior expected value of β is

$$\bar{\bar{\beta}} = (A+X'X)^{-1}(A\bar{\beta}+X'Y) \,, \tag{3.6}$$

where for convenience we set $A=\sigma^2\bar{\Sigma}_\beta^{-1}$, and σ^2 is the variance of the disturbance term e.[5] Note that to find the posterior coefficient values, we need to specify σ^2. A Bayesian solution that takes a diffuse prior distribution for σ leads to a normal-t posterior density for β, which would require numerical integration to calculate the posterior mean. Instead, Litterman suggests using an approximation $\hat{\sigma}_i$, the estimated standard error of the unrestricted OLS residuals.

Until now it was assumed that the scale factors d, θ and λ in the prior distribution were known. However, they have to be determined. There is no commonly accepted method or criterion to determine these values. Litterman reports results for different values of λ and argues that the one with $\lambda=0.2$ yields the minimum forecast error. Furthermore, he arbitrarily picks $\theta=0.2$ and $d=1$. Fackler, McMillin, and Silver (1990) and Spencer (1993) use a systematic grid search to find the optimum scale parameters. The options are to make a systematic grid search or pick some numbers. In this study we do a systematic grid search to select values for d, λ and θ.

A similar estimator to (3.6) is obtained by Theil and Goldberger (see Theil (1971)) employing a different approach. The Theil-Goldberger estimator combines sample and non-sample (prior) information by formulating the prior as a regression equation. Assume that we have the regression model in (3.4) with $e \sim N(0,\Sigma)$.

---

[5]Details can be found in Judge et al. (1988), chapter 7.

Further assume that the researcher has some prior knowledge or belief about the coefficients which can be represented as,

$$r = R\beta + v \tag{3.7}$$

where R is a known JxK matrix of rank J, E(v)=0, E(vv')=V_0, and V_0 is a known nonsingular covariance matrix. The elements of r can be interpreted as the best guesses of the corresponding elements of $R\beta$ and the diagonal elements of V_0 as measures of uncertainty with respect to these guesses. In (3.4) Bayesians impose restrictions on the coefficients by specifying the prior means and prior covariance matrix. However, (3.7) imposes restrictions on the linear combinations of the coefficients where the uncertainty is defined by the covariance matrix V_0.

The Theil-Goldberger estimator is derived by combing (3.4) and (3.7) under the assumption that sample and prior information are independent. Combining (3.4) and (3.7), we obtain

$$\begin{bmatrix} Y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} e \\ v \end{bmatrix} \quad .$$

The Theil-Goldberger estimator of β can be obtained as,

$$b_{TG} = (X'\Sigma^{-1}X + R'V_0^{-1}R)^{-1}(X'\Sigma^{-1}Y + R'V_0^{-1}r) \quad . \tag{3.8}$$

The details of this derivation can be found in Theil (1971, pp 247-249). We see that (3.6) and (3.8) are similar estimators where the Bayesian estimator is a special case of b_{TG}. The two approaches would be the same if R in (3.7) is the identity matrix, Σ

equals $\sigma^2 I$, and the elements of V_0 are formulated such that the off-diagonal elements of V_0 are zero and the variances are given by (3.3).

The non-sample information may vary from equation to equation. In a VAR model there is more than one equation. Different restrictions can be imposed on different equations, so that for each equation the matrix R and the vector r can be different. In this essay, we will impose Litterman's restrictions, which restrict the coefficient of the first own-lag to one, and the rest of the coefficients to zero. R and r will be different for each equation in the VAR since the position of the own lags will change in the explanatory variable matrix which is the same for all equations. For instance, in the first equation, the own lags will be positioned right after the intercept; however, in the second equation, the own lags will be positioned after the lag values of the first variable.

The Stein-rule is another shrinkage estimator. Like its Bayesian counterpart, the Stein-rule estimator can shrink the ML estimator toward a known prior vector by imposing restrictions on the estimator. Assume we have the single equation model described by equation (3.4). The unrestricted OLS and ML estimator is $b=(X'X)^{-1}X'Y$. Suppose we have some non-sample information about the parameters this equation, which we characterize as $R\beta=r$. These restrictions represent the researchers beliefs or are restrictions suggested by economic theory. Note that these restrictions are exact; that is, the restricted parameters will obey the constraints.

The explanation of the Stein-rule estimator will be given in a single equation context. We will use the same estimator for all equations except with different R and

r matrixes. If we impose the restriction $R\beta=r$ on the coefficients, the restricted ML becomes:

$$b^{*}=b-(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb-r) \tag{3.9}$$

A Stein-rule estimator is a linear combination of the restricted ML and unrestricted ML estimators given as,

$$b^{*}=\left(1-\frac{(J-2)s}{(T-K+2)u}\right)b+\left(\frac{(J-2)s}{(T-K+2)u}\right)b^{*} \tag{3.10}$$

where u is,

$$u=(b-b^{*})'(X'X)(b-b^{*}) \tag{3.11}$$

and s is the sum-of-the-squares of the residuals from the unrestricted model, K is the number of parameters in the unrestricted model, and J is the number of restrictions imposed on the unrestricted model to obtain the restricted model.[6] Note that, since s and u will be different for each equation, the weight in (3.10) will have different values for each equation. Equation (3.10) can also be presented as,

$$b^{*}=\left[1-\frac{c}{f}\right]b+\left[\frac{c}{f}\right]b^{*} \tag{3.12}$$

where c is a constant and f is the F test statistics of the restrictions which is written as,

[6]See Judge et al. (1985) pp. 876-81.

$$u = \frac{(b-b^*)'(X'X)(b-b^*)/(J-2)}{s/(T-K+2)} \; .$$

If the data support the non-sample information, then u will be small and the weight attached to b^* will be relatively large. Conversely, if the data do not support the information, then the weight of b^* will be small in the estimator equation. See Knight et al. (1992) for derivation and further discussion.

The main idea of Stein-rule estimation is to obtain an estimate whose value will be between the unrestricted and restricted estimates. Note that to have a convex combination of b and b^*, c has to be less than f. However, if c is larger than f, the weight of b in (3.12) will be a negative number, which is not desirable. To overcome this problem we can use the positive-part Stein-rule estimator.

$$b^{s+} = I_{(c,\infty)}(f)\left[1-\frac{c}{f}\right](b-b^*)+b^* \tag{3.13}$$

where I(f) is an indicator function defined as

$$I_{(c,\infty)}(f) = \begin{cases} 1 & \text{if } \; c<f<\infty \\ 0 & \text{otherwise} \end{cases}$$

where c and f are as defined before. If we wish to consider how much shrinkage toward the restricted estimator is optimal from the viewpoint of minimizing out-of sample prediction risk, we can modify (3.13) as,

$$b^{s+}(m) = I_{(mc,\infty)}(f)\left[1-\frac{mc}{f}\right](b-b^*)+b^* \tag{3.14}$$

where m controls the degree of shrinkage. As $m \rightarrow \infty$ then $b^{s+}(m) \rightarrow b^*$.

In this essay, the restrictions imposed in Stein-rule estimation will be the same as the Minnesota prior; the deterministic coefficients are not restricted, the coefficient of the own lag is restricted to one, and the rest of the coefficients are restricted to zero.

Unlike forecasting with a BVAR, forecasting with a SRVAR does not require the researcher to determine some parameters before forecasting. The forecaster has to provide only the lag structure of the model, the restrictions that are imposed, and the multiplier m. The restricted and unrestricted estimates are combined by using weights estimated from the sample.

The three approaches to obtaining shrinkage estimators (Bayesian, Theil-Goldberger, and Stein-rule) differ from each other in the way they measure the uncertainty about the restrictions. Bayesian methods define a prior distribution where the variance of this distribution represents the uncertainty; Theil-Goldberger do not define a prior distribution, but the restrictions are allowed to have a covariance matrix; a Stein-rule measures the uncertainty by the test statistic of the restrictions. A large test statistic value defined by (3.11), which has a chi-square distribution, means that the data are not compatible with the restrictions and therefore the data determine the parameter estimation.

BVAR and Stein-rule estimators use non-sample information to shrink the estimated parameters towards prior values. As noted earlier, besides these shrinkage estimators, other have been methods proposed which shrink the number of estimated parameters. Keating (1994) and Hsiao (1981) offer procedures to determine the

number of lags of each variable that are included in a VAR. Often these methods suggest estimating VARs with asymmetric lag structures which have fewer parameters than the conventional VARs. Once the lag structure is identified, the parameters can be estimated via OLS or seemingly unrelated regressions.

In Keating's method, it is assumed that each variable has the same lag length in all equations, but different variables are allowed to have different lag lengths. Hsiao uses an equation-by-equation method in determining the lag structure. The details of lag structure specification via Keating's method and Hsiao's method are provided in the first essay of this dissertation.

Hereafter we will call the VAR model in which the lag structure is specified by Keating's method as KVAR, and the VAR model in which the lag structure is specified by Hsiao's method as HVAR. An additional forecasting technique considered is a Stein-rule asymmetric VAR in which the lag structure is determined by Keating's method and the coefficients are estimated via the Stein-Rule (KSRVAR). An asymmetric lag structure can be used in BVAR estimation. However, since our purpose in using an asymmetric lag structure is to investigate whether an improvement can be achieved, we consider only the asymmetric lag structure SRVAR and compare it with a symmetric SRVAR.

Once we obtain the coefficients using any of the methods described above, the h-period-ahead forecast of the variable of interest can be obtained by chain computations. After the one-period-ahead forecast is computed, this forecast and

the predetermined variables are used to compute the two-period-ahead forecast. We proceed this way until we get the h-period-ahead forecast.

3.4 Empirical Methodology and Results

3.4.1 Real-Time Data Analysis

The real-time data (actual real data) performances of the estimators are investigated for quarterly data and monthly data. In this essay we focus on forecasting the six-month commercial paper rate (RATE), real GDP, and the GDP deflator. The focus is on these variables since Litterman (1979), Doan, Litterman and Sims (1984), Webb (1984), Lupoletti and Webb (1986), and Litterman (1986) report the performance of forecasters of real GNP, the inflation rate, and different measures of short-term asset interest rates.[7] A monetary aggregate, M2, is included in the VAR system because it is argued that changes in the money supply have an effect on real output, at least in the short-run (see Friedman and Schwartz (1963), Christiano and Ljungqvist (1988), Stock and Watson (1989) and Fackler, McMillin and Silver (1990)). Even though a money supply measure is included in the system, we are not interested in forecasting the money supply. All forecasts are out-of-sample forecasts, at 1, 2, 3, 4, 6 and 8 periods ahead for quarterly data, and 1, 2, 3, 6, 12 and 24 periods ahead for monthly data. In monthly forecasting, real GDP is replaced by industrial production (IP), and the price deflator by the consumer price index.

---

[7]Recently the U.S. government started concentrating on GDP rather than GNP. Therefore, in the literature the focus has shifted from GNP to GDP. That is why GDP is used in this study instead of GNP.

The data are from CITIBASE. Since data for M2 are only available beginning in 1959, the data cover the 1959 to 1994 period. Monthly observations of the interest rate and M2 are converted to quarterly data by arithmetic averaging. Usually, the time-series of real GDP, the price level and the money supply are modeled as first-differences in their logarithms (GGDP, INF and GM2, respectively) or in their logarithm levels (LGDP, LPIND and LM2, respectively), and the interest rate is in its level form (see Webb (1984), Lupoletti and Webb (1986), Litterman (1986) McNees (1986), and Kadiyala and Karlsson(1993)). To investigate the effect of time-series modeling on the relative forecast performances, this essay compares the forecast performances of the estimators when real GDP, IP, the price level and M2 are modeled both in log levels and log differences. In both models, the interest rate is in level form.

To construct the forecasting system, certain parameters of the BVAR model and the lag length of the autoregressive process have to be predetermined. Unfortunately, a set of parameter values which is optimum for the forecasting of one variable usually will not be the optimum for another variable. Spencer (1993) chooses the values for the forecast of one particular variable; he indicates, "It will be presumed that the primary interest is in forecasting of a single variable. If there is an interest in forecasting more than one variable, the forecaster will want to select separate BVAR models for each." (Spencer 1993, p.414). We are interested in the forecasts of three variables in this essay; thus we should have three sets of parameters. In this essay, we also pick a set of parameters which yields the best

Sims' overall forecast performance measure (hereafter SOFM). This will be used as a criterion to select one set of parameter values to forecast all three variables. The forecasting performance outcome of this practice will be referred to as BVAR1, and BVAR2 will be the forecast outcome obtained by specifying different sets of parameters to forecast each variable.

Various criteria have been proposed to select the lag length of a VAR. Our previous essay shows that Akaike's Information Criterion (AIC) most consistently specifies the true underlying lag structure. Therefore, in specifying the symmetric VAR, and in Keating's method, the AIC criterion is used. The lag structures of BVAR1, BVAR2 and SRVAR will be the same one as the unrestricted VAR's lag structure, so that the lag structures will have the same effect on the outcomes of all estimators. This will allow comparison of the forecast accuracies of these three estimators based solely on the estimation technique.

The first forecast exercise is done using quarterly data and modeling the variables in log levels. Because 8 lags of quarterly data go back 2 years, the maximum lag length with quarterly data will be 8. In most cases, two years of lags are sufficient to provide the necessary information for the most accurate forecast without overparametrization of the model. Therefore, lags 1 through 8 are considered in searching for the optimum lag length. The lag structures are determined using only the 1959:1-1989:4 sample. By using a sample period through 1989, we are pretending that the sample information after 1989 is not available to the forecaster, who has to determine the hyperparameters and the lag structure. A lag length of 6 is

chosen for the symmetric lag VARs, and lags 6, 1, 4, and 5 are estimated for the lag lengths of RATE, LGDP, LM2 and LPIND, respectively, by Keating's method. Hsiao's method specifies these lag lengths for the first equation as 6, 3, 4 and 1; for the second equation as 5, 2, 2 and 0;[8] for the third equation as 6, 0, 2 and 3; and for the fourth equation as 2, 1, 0 and 4.

For the BVAR, the grid search method is used to find the optimum parameter values. A grid search is not based on a theoretic derivation, such as the AIC or the likelihood-ratio test statistic. A grid search does not yield a global optimum; rather, it finds the optimum parameters from a limited set of parameters. Grid search is a simple way to find acceptable parameters to be used in forecasting which will yield low forecast error. To conduct a grid search, a pilot forecasting period is selected, and the forecaster uses the forecast in this period to determine the optimum parameters. In this essay the pilot period is 1985:1-1989:4. The parameter values that have to be determined are systematically altered, and one-by-one the corresponding Theil U and Sims' overall forecast measure (SOFM) are calculated. The forecaster then picks the set of parameters that yields the lowest Theil U or SOFM. The search starts using the sample through 1984:4 to estimate the coefficients and to compute the h-period ahead out-of-sample forecasts. Then the sample is updated one period, and the coefficients are reestimated and h-period-ahead forecasts are computed. This process is stopped after the values of 1989:4 are forecasted, yielding (20 - h) forecasts for each variable. Again, we are pretending that the

[8]0 lag indicates that the price level is not included in the GDP equation.

sample after 1989 is not available to determine the hyperparameters. A set of parameter values which yields the lowest Theil U or SOFM statistic at one horizon does not necessarily yield the lowest Theil U or SOFM statistic at other horizons. Selecting the appropriate hyperparameters is difficult if forecast performance for different sets of hyperparameters yields different results at different forecast horizons. Consequently, the researcher cannot mechanically pick the best set, but must use judgement in selecting the set of hyperparameters. For instance, if the researcher has a preference for lower forecast error in the short-run, parameters which tend to yield lower Theil U or SOFM statistics in the short-run will be chosen.

In a BVAR, three parameters that have to be obtained before estimating the forecasting model are d (the decay parameter), θ (the weight of cross variable lags), and λ (prior variance of first own lag). To decide which values of these hyperparameters to use in the grid search, we used the values in previous studies as prior information. Previous studies have used a decay parameter around 1, and θ and λ values typically around 0.2. Furthermore, in the grid search, the number of different combinations of the parameters should be a manageable amount, since the forecaster will use his/her judgement in picking the best parameter set. Therefore, in this essay, to find the most suitable hyperparameters, we obtain the Theil U and SOFM statistics corresponding to the d values of 0.5, 1 and thereafter one unit increments up to 7. Values of θ and λ are obtained by searching between 0.05 and 0.50 with 0.05 increments. Thus, the total number of different combinations of the parameters is 8x10x10=800. Instead of computing Theil U and SOFM statistics for

each combination, the grid search is divided in two parts. The first part alters only the values of θ and λ, keeping d=1, which yields 10x10=100 different combinations. Once the parameter values of θ and λ which yield low Theil U or SOFM statistics are selected, the search for the best decay parameter is conducted. Fixing θ and λ at the values selected in the previous step, the accuracy statistics for different values of d are calculated. A detailed description of the grid search can be found in Spencer (1993). It is not feasible to report all the grid search results; therefore, the Theil U and SOFM statistics of selected lags and hyperparameters are reported.

To demonstrate how the hyperparameters are selected from a grid search, Table 3.1 presents selected θ and λ values, and the Theil U and the SOFM statistic they yield. The third through the eighth columns are the forecasting horizons. Clearly it is seen that a parameter set which yields the lowest Theil U or SOFM statistics for some forecast horizons may not yield the lowest statistics for other horizons. That is why the forecaster has to pick the parameters which he/she thinks is the most suitable. The first three sections of this table presents the Theil U statistics of each three forecasted variables, which are used for selecting the hyperparameters of BVAR2. The fourth section presents Sims' overall forecast accuracy statistic, which is used to select the hyperparameters of BVAR1. BVAR1 uses only one set of parameters to forecast all three variables whereas BVAR2 uses different parameters to forecast each variable. For the interest rate variable, θ=0.50 and λ=0.05 yield the lowest values at all forecast horizons. Even though θ=0.25 and λ=0.05 do not yield the lowest 2 period-ahead LGDP Theil U statistic, these values are selected to be used

Table 3.1
Log Level Model, Theil U and SOFM statistics of BVAR for selected values of θ and λ for 1985:1 - 1989:4

RATE		Forecast Horizon					
θ	λ	1	2	3	4	6	8
0.35	0.05	1.082	1.111	1.132	1.154	1.182	1.170
0.40	0.05	1.078	1.107	1.128	1.150	1.180	1.169
0.45	0.05	1.074	1.102	1.123	1.146	1.178	1.168
0.50	0.05	1.069	1.097	1.119	1.143	1.177	1.168
0.50	0.10	1.110	1.165	1.209	1.267	1.360	1.372
LGDP							
0.05	0.20	0.442	0.310	0.288	0.236	0.164	0.096
0.05	0.25	0.438	0.302	0.282	0.232	0.174	0.116
0.15	0.10	0.449	0.322	0.295	0.251	0.202	0.148
0.25	0.05	0.442	0.312	0.279	0.225	0.148	0.075
0.30	0.05	0.448	0.319	0.287	0.237	0.173	0.107
LPIND							
0.05	0.40	0.243	0.260	0.249	0.283	0.346	0.403
0.05	0.45	0.242	0.259	0.248	0.282	0.346	0.404
0.05	0.50	0.242	0.259	0.247	0.282	0.346	0.405
0.20	0.45	0.241	0.266	0.251	0.292	0.363	0.428
0.20	0.50	0.240	0.265	0.250	0.290	0.361	0.427
SOFM							
0.05	0.25	-12.36	-11.07	-10.26	-9.96	-9.58	-9.24
0.05	0.30	-12.37	-11.09	-10.26	-9.99	-9.64	-9.03
0.05	0.35	-12.35	-11.04	-10.21	-9.94	-9.58	-8.68
0.10	0.30	-12.17	-10.70	-9.88	-9.58	-9.29	-8.12
0.10	0.35	-12.09	-10.59	-9.79	-9.48	-9.17	-8.04

in BVAR2 since they perform well in 1, 3, 4, 6 and 8 period ahead LGDP forecasts. For price index forecasting, the optimum parameter values we pick are θ=0.05 and λ=0.50. These parameters yield the lowest 1, 2, 3, 4, and 6 period-ahead forecast Theil U statistics. The selected θ and λ parameter values of BVAR1 are 0.05 and

0.30, respectively. Again, these parameter values do not have the best forecast performance in all 6 forecast horizons. However, $\theta=0.05$ and $\lambda=0.30$ yields the lowest 1, 2, 3, 4 and 6 periods-ahead SOFM and the second lowest 8 periods-ahead SOFM.

The second step in the grid search is to find the optimal decay parameter. Table 3.2 presents the Theil U and SOFM statistics for different values of d when θ and λ are set at the values determined in the first step. For RATE, d=7 is the most suitable choice. For LGDP, LPIND we pick d=1, since it yields the lowest Theil U statistics in all periods. The last section of Table 3.2 is for the BVAR1 decay parameter. For BVAR1 we pick d=1, since it yields the lowest SOFM values in all periods.

As stated in (3.14), we have the option to determine what value to use for the Stein-rule estimator multiplier (m). Two different values of the multiplier, 1 and 2, are used to forecast the three variables. Setting m=2 shrinks the estimator further towards the unrestricted coefficient estimates, and setting m=1 coincides to the original Stein-rule estimator derived in (3.13). The forecast results for 1985-1989 showed that m=1 yields the best results in terms of low Theil U statistics for all variables.

After determining the lag lengths and the necessary parameters, we are ready to compute the out of sample forecasts between 1990:1 and 1994:4. We start with the data through 1989:4 to estimate the coefficients and to compute the h-period-ahead forecasts. Then, the data set is updated one period and the coefficients are

Table 3.2
Log Level Model, Theil U and SOFM statistics of BVAR for selected values of d for 1985:1 - 1989:4

RATE	Forecast Horizon					
d	1	2	3	4	6	8
1	1.069	1.097	1.119	1.143	1.177	1.168
2	1.069	1.082	1.087	1.092	1.095	1.071
3	1.067	1.076	1.078	1.079	1.078	1.053
4	1.067	1.075	1.076	1.076	1.074	1.049
5	1.067	1.074	1.075	1.075	1.073	1.048
6	1.067	1.074	1.075	1.075	1.073	1.048
7	1.067	1.074	1.075	1.075	1.073	1.048
LGDP						
0.5	0.447	0.322	0.293	0.248	0.190	0.130
1	0.442	0.312	0.279	0.225	0.148	0.075
2	0.448	0.322	0.286	0.232	0.154	0.099
3	0.451	0.326	0.290	0.236	0.159	0.108
4	0.451	0.327	0.291	0.237	0.161	0.110
5	0.452	0.327	0.291	0.238	0.161	0.111
6	0.452	0.327	0.291	0.238	0.161	0.111
LPIND						
0.5	0.249	0.256	0.271	0.295	0.364	0.427
1	0.242	0.247	0.259	0.282	0.347	0.406
2	0.311	0.349	0.385	0.426	0.522	0.607
3	0.566	0.659	0.720	0.769	0.865	0.940
4	0.766	0.835	0.871	0.898	0.965	1.022
5	0.837	0.882	0.905	0.925	0.985	1.038
6	0.857	0.893	0.913	0.931	0.989	1.041
SOFM						
0.5	-12.27	-10.93	-10.13	-9.86	-9.39	-8.67
1	-12.37	-11.09	-10.26	-9.99	-9.64	-9.03
2	-12.06	-10.68	-9.76	-9.38	-8.91	-8.69
3	-11.37	-9.99	-9.11	-8.70	-8.23	-8.23
4	-11.16	-9.83	-8.98	-8.58	-8.13	-8.15
5	-11.11	-9.79	-8.96	-8.56	-8.11	-8.14
6	-11.09	-9.79	-8.95	-8.56	-8.11	-8.14

reestimated and another set of h-period-ahead forecasts are computed. This procedure is stopped after the forecasts for 1994:4 are computed. The Theil U statistics are presented in Table 3.3. The first column defines the forecasted variable and the next six columns are the Theil U statistics for 1, 2, 3, 4, 6 and 8 period ahead forecasts. The numbers in parenthesis show the order of the Theil U statistics from the smallest to the largest. Forecast accuracies which are identified to be different from the no-change forecast accuracy by the sign test are marked by "*".

SRVAR and VAR have the best Theil U statistics obtained from forecasting the interest rate; however, the performances of VAR and SRVAR are not statistically different from the no-change forecast. The two BVAR methods yield Theil U statistics larger than one which means lower accuracy than the no-change forecast. The forecast performances of BVAR are statistically inferior to the random walk forecasts. For this variable, the Theil U statistics for BVAR2 are lower than for BVAR1.

The lowest Theil U statistics from forecasting the log level of GDP are obtained from HVAR. However, the sign test does not show any statistical difference from the no-change forecast accuracy. The BVARs have statistically better forecasts than the random walk in 1, 2 and 3 periods ahead, and SRVAR has statistically better forecasts in periods 2 and 3.

All forecasting methods yield statistically more accurate forecasts of the price level than the no-change forecast. In the 1, 2, 3 and 4 periods ahead forecast, SRVAR has the lowest Theil U statistic; in periods 6 and 8, KSRVAR has the lowest

Table 3.3
Log Level Quarterly Data Theil U Statistics for 1990:1 - 1994:4

RATE	1	2	3	4	6	8
VAR	0.870(2)	0.669(1)	0.652(1)	*0.818(1)	1.176(3)	*1.354(3)
BVAR1	*1.131(7)	*1.227(6)	*1.246(7)	*1.285(6)	*1.383(6)	*1.433(6)
BVAR2	*1.062(5)	*1.096(5)	*1.120(5)	*1.140(5)	*1.169(2)	*1.176(1)
SRVAR	0.829(1)	0.716(2)	0.727(2)	0.871(2)	1.148(1)	1.309(2)
KVAR	0.985(4)	0.942(4)	0.837(3)	0.962(3)	*1.301(5)	*1.422(4)
HVAR	1.129(6)	*1.242(7)	*1.242(6)	*1.466(7)	*1.942(7)	*2.170(7)
KSRVAR	0.976(3)	0.935(3)	0.839(4)	0.966(4)	*1.300(4)	*1.424(5)
LGDP						
VAR	0.902(5)	0.930(7)	1.002(7)	1.043(7)	1.066(7)	0.904(7)
BVAR1	*0.787(2)	*0.807(3)	*0.817(3)	0.811(3)	0.745(3)	0.628(3)
BVAR2	*0.809(3)	*0.799(1)	*0.788(1)	0.769(1)	0.686(2)	0.584(2)
SRVAR	0.833(4)	*0.845(4)	*0.885(4)	0.910(4)	0.913(4)	0.787(4)
KVAR	0.911(7)	0.901(6)	0.920(6)	0.942(6)	0.959(6)	0.820(6)
HVAR	0.766(1)	0.805(2)	0.800(2)	0.777(2)	0.651(1)	0.482(1)
KSRVAR	0.904(6)	0.894(5)	0.912(5)	0.932(5)	0.946(5)	0.808(5)
LPIND						
VAR	*0.294(2)	*0.279(4)	*0.222(2)	*0.202(4)	*0.299(5)	*0.405(5)
BVAR1	*0.295(4)	*0.275(3)	*0.245(6)	*0.239(6)	*0.231(3)	*0.371(4)
BVAR2	*0.290(3)	*0.271(2)	*0.238(5)	*0.226(5)	*0.277(4)	*0.339(3)
SRVAR	*0.283(1)	*0.261(1)	*0.202(1)	*0.188(1)	*0.303(6)	*0.412(6)
KVAR	*0.303(6)	*0.284(6)	*0.230(3)	*0.196(3)	*0.204(2)	*0.246(2)
HVAR	*0.330(7)	*0.314(7)	*0.281(7)	*8.299(7)	*0.452(7)	*0.565(7)
KSRVAR	*0.301(5)	*0.282(5)	*0.227(4)	*0.192(2)	*0.201(1)	*0.243(1)

Theil U statistics with a "*" indicate forecast accuracies which are significantly different than those produced by the no-change forecast. The numbers in parentheses are the order from the smallest to the largest Theil U statistics.

Theil U statistics. Comparing the performances of BVAR1 and BVAR2, we see that BVAR2 is usually more accurate.

Table 3.4 reports the sign-test statistics comparing BVAR1, BVAR2 and SRVAR forecast accuracy to the other forecasting methods' accuracies. The test statistic has a standard normal distribution. A claim of equal forecast accuracy will

be rejected at 5% level if the test statistic is larger than 1.645. BVAR2 is statistically more accurate than BVAR1 in forecasting RATE, but does not have statistically more accurate LGDP forecasts. In forecasting LPIND, BVAR2 has statistically better forecasts than BVAR1 at horizons 4, 6 and 8. BVAR1 is never statistically more accurate than BVAR2. SRVAR is statistically more accurate than BVAR1 and BVAR2 in forecasting 3 and 4 periods-ahead for RATE, and it less accurate than the BVARs in 3, 4, 6 and 8 periods-ahead LGDP forecasts. Furthermore, the unrestricted VAR is more accurate than BVAR1 in 2, 3 and 4 period-ahead forecasting of RATE. The performance of SRVAR in forecasting RATE is not statistically different from that of the VAR, and, at some horizons, is better than the BVARs, KVAR, HVAR and KSRVAR. Thus, SRVAR is preferred in forecasting the interest rate and the price index, while BVAR2 is preferred in forecasting GDP.

Table 3.5 presents Sims' overall forecast measure SOFM, and the sum of Theil U statistics of the three forecasted variables (SUM) for each forecast-horizon. SOFM and SUM agree that SRVAR has the lowest values in 1 and 2 period-ahead forecasts. SOFM and SUM identify BVAR2 as the best long-run forecaster. However, overall, in all periods, SRVAR performs well. The price level forecast performance of SRVAR substantially deteriorates at forecast horizons 6 and 8, causing the overall performance also to deteriorate in those horizons. On the other hand, the interest rate forecast of BVAR2 improves at forecast horizons 6 and 8, raising the long-run overall forecast performances. The overall forecast measures show that BVAR2 performs better than BVAR1.

Table 3.4
Log Level Quarterly Data, BVAR1, BVAR2 and SRVAR Sign Test Values

	BVAR1					
RATE	1	2	3	4	6	8
VAR	0.89	3.44	3.30	2.18	1.29	1.39
BVAR2	2.24	2.52	2.36	2.18	2.32	2.50
SRVAR	1.34	3.44	3.77	2.67	1.81	1.39
KVAR	1.34	1.61	1.89	2.67	1.29	-0.28
HVAR	0.45	-0.23	-0.47	-1.70	-3.87	-3.61
KSRVAR	1.34	1.61	1.89	2.67	1.29	0.28
LGDP						
VAR	0.89	-2.06	-2.36	-2.18	-1.29	-1.39
BVAR2	-0.45	1.61	1.41	1.21	0.77	1.39
SRVAR	-0.89	-0.23	-2.36	-3.15	-2.84	-2.50
KVAR	-0.89	-2.06	-2.36	-2.67	0.77	-0.83
HVAR	0.89	-0.23	0.00	-0.24	0.26	-0.28
KSRVAR	-0.89	-1.61	-2.36	-2.18	-0.77	-1.39
LPIND						
VAR	0.45	1.61	0.94	2.67	0.26	-0.83
BVAR2	1.34	0.69	1.89	2.18	3.36	3.05
SRVAR	0.89	1.15	1.41	1.70	-0.26	-0.28
KVAR	0.00	-0.23	0.47	1.21	1.81	2.50
HVAR	-1.34	0.69	-1.41	-2.18	-3.87	-3.61
KSRVAR	0.00	-0.23	0.47	1.21	2.32	2.50

	BVAR2					
RATE	1	2	3	4	6	8
VAR	-0.89	1.15	1.89	-2.18	0.77	-2.50
SRVAR	0.89	1.61	2.36	2.18	0.26	-1.39
KVAR	-0.45	-0.23	0.94	0.24	-1.81	-2.50
HVAR	-0.89	-1.15	-0.94	-2.67	-2.84	-3.61
KSRVAR	-0.45	-0.23	0.94	0.24	-1.81	-2.50
LGDP						
VAR	-0.45	-1.61	-1.89	-1.70	-1.81	-1.94
SRVAR	0.00	0.69	-2.36	-3.15	-2.32	-2.50
KVAR	0.00	-1.61	-2.83	-2.67	-1.81	-0.83
HVAR	0.89	-0.23	0.00	-0.24	-0.26	0.28
KSRVAR	0.00	-1.15	-2.83	-2.67	-1.81	-0.83
LPIND						
VAR	0.45	1.15	0.94	-1.70	-0.26	-0.83
SRVAR	0.45	1.15	0.47	1.70	-0.77	-1.94
KVAR	-0.89	0.23	0.47	0.73	1.29	2.50
HVAR	-1.34	-0.69	-1.41	-3.15	-3.87	-3.61
KSRVAR	-0.89	0.23	0.94	0.73	1.29	2.50

(Table con'd.)

	SRVAR					
RATE	1	2	3	4	6	8
VAR	-1.34	0.23	1.89	1.70	-0.26	-1.39
KVAR	-1.79	-1.61	-1.89	-2.18	-2.32	-2.50
HVAR	-2.24	-3.44	-3.30	-3.15	-2.84	-3.61
KSRVAR	-1.34	-1.61	-1.89	-2.18	-2.32	-2.50
LGDP						
VAR	-1.34	-1.61	-0.94	-0.73	-1.29	-1.39
KVAR	-0.45	-1.61	-0.94	-1.21	0.26	-0.28
HVAR	0.89	-0.23	0.00	-0.24	0.26	-0.28
KSRVAR	-0.45	-1.61	-0.94	-1.21	0.26	0.83
LPIND						
VAR	0.00	-0.69	-0.47	0.73	1.29	1.39
KVAR	0.45	-1.61	-0.47	-0.24	1.81	1.39
HVAR	-1.79	-2.06	-1.89	-2.18	-2.84	-3.05
KSRVAR	0.45	-1.61	0.47	-0.24	1.81	1.94

The log level quarterly data forecast results show that SRVAR performs at least as well as any BVAR estimator. We also compare the forecasting performance of the estimators when the models are estimated using log differences for GDP, PIND and M2. However, ultimately, the log level forecasts of GDP and price level are computed by transforming the growth rate forecasts to log level forecasts. We compute the log level forecast so that we can compare the results of the first difference specification explained with the results of the log level model. The same procedure is applied as demonstrated above. The AIC criterion for lag lengths 1 through 8 are computed, and 5 lags are found to be optimal for the symmetric lag VARs. As before, the BVARs and SRVAR have the same lag structure as the symmetric VAR. KVAR has 7 lags for RATE, 2 lags for the growth rate of GDP (GGDP), 1 lag for the growth rate of M2 (GM2), and 5 lags for inflation (INF) in each equation. Hsiao's method specifies the lag lengths for the first equation as 6, 2,

Table 3.5
Log Level Quarterly Data, Overall Fitness Measures

SOFM	1	2	3	4	6	8
VAR	$-11.63^{(3)}$	$-10.29^{(2)}$	$-9.68^{(3)}$	$-9.06^{(5)}$	$-7.54^{(5)}$	$-7.22^{(5)}$
BVAR1	$-11.60^{(4)}$	$-10.05^{(5)}$	$-9.37^{(6)}$	$-9.12^{(4)}$	$-7.98^{(2)}$	$-7.90^{(2)}$
BVAR2	$-11.72^{(2)}$	$-10.27^{(3)}$	$-9.92^{(1)}$	$-9.39^{(1)}$	$-8.24^{(1)}$	$-8.11^{(1)}$
SRVAR	$-11.82^{(1)}$	$-10.44^{(1)}$	$-9.80^{(2)}$	$-9.23^{(3)}$	$-7.81^{(4)}$	$-7.52^{(4)}$
KVAR	$-11.48^{(7)}$	$-9.96^{(7)}$	$-9.31^{(7)}$	$-8.73^{(7)}$	$-7.49^{(7)}$	$-7.06^{(7)}$
HVAR	$-11.61^{(5)}$	$-10.16^{(4)}$	$-9.66^{(4)}$	$-9.33^{(2)}$	$-7.98^{(2)}$	$-7.86^{(3)}$
KSRVAR	$-11.50^{(6)}$	$-9.99^{(6)}$	$-9.43^{(5)}$	$-8.74^{(6)}$	$-7.52^{(6)}$	$-7.09^{(6)}$
SUM						
VAR	$2.067^{(2)}$	$1.877^{(2)}$	$1.876^{(2)}$	$2.064^{(2)}$	$2.540^{(6)}$	$2.663^{(6)}$
BVAR1	$2.212^{(6)}$	$2.309^{(6)}$	$2.308^{(6)}$	$2.335^{(6)}$	$2.435^{(3)}$	$2.431^{(2)}$
BVAR2	$2.161^{(3)}$	$2.166^{(5)}$	$2.145^{(5)}$	$2.135^{(5)}$	$2.133^{(1)}$	$2.099^{(1)}$
SRVAR	$1.945^{(1)}$	$1.822^{(1)}$	$1.815^{(1)}$	$1.969^{(1)}$	$2.365^{(2)}$	$2.509^{(5)}$
KVAR	$2.199^{(5)}$	$2.128^{(4)}$	$1.987^{(4)}$	$2.101^{(4)}$	$2.464^{(5)}$	$2.488^{(4)}$
HVAR	$2.224^{(7)}$	$2.362^{(7)}$	$2.323^{(7)}$	$2.542^{(7)}$	$3.045^{(7)}$	$3.218^{(7)}$
KSRVAR	$2.181^{(4)}$	$2.111^{(3)}$	$1.977^{(3)}$	$2.090^{(3)}$	$2.446^{(4)}$	$2.475^{(3)}$

3, and 2; for the second equation as 5, 2, 0, and 5; for the third equation as 6, 8, 1, and 8; and for the fourth equation as 2, 0, 0, and 3.

The grid search of the hyperparameters of the BVAR estimators uses the same intervals as before. For convenience, only the parameters values that have been selected are reported without going into the details of how they were chosen. The hyperparameters are chosen such that the log level Theil U and overall fitness statistics are minimized. BVAR1 values are $d=1$, $\theta=0.15$ and $\lambda=0.40$. BVAR2 hyperparameters to forecast RATE are $d=2$, $\theta=0.05$, and $\lambda=0.40$. These numbers for forecasting LGDP are 0.5, 0.05, and 0.50, and for forecasting LPIND are 7, 0.40, and 0.10, respectively. For Stein-rule estimation the multiplier is set to one, since $m=1$ had better forecasting performance than $m=2$ for periods 1985-1989.

Table 3.6
First Difference Quarterly Data Theil U Statistics

RATE	1	2	3	4	6	8
VAR	$1.012^{(6)}$	$0.927^{(4)}$	$0.903^{(5)}$	$0.876^{(6)}$	$0.820^{(5)}$	$0.778^{(5)}$
BVAR1	$0.871^{(1)}$	$^{*}0.842^{(3)}$	$^{*}0.831^{(3)}$	$^{*}0.832^{(4)}$	$^{*}0.840^{(6)}$	$0.857^{(6)}$
BVAR2	$0.960^{(3)}$	$0.980^{(6)}$	$0.982^{(7)}$	$0.990^{(7)}$	$0.993^{(7)}$	$0.989^{(7)}$
SRVAR	$0.887^{(2)}$	$0.788^{(1)}$	$0.765^{(2)}$	$0.764^{(2)}$	$0.755^{(2)}$	$0.775^{(4)}$
KVAR	$1.045^{(7)}$	$0.997^{(7)}$	$0.944^{(6)}$	$0.847^{(5)}$	$0.766^{(3)}$	$0.726^{(1)}$
HVAR	$0.960^{(3)}$	$0.808^{(2)}$	$0.701^{(1)}$	$0.662^{(1)}$	$^{*}0.650^{(1)}$	$^{*}0.734^{(2)}$
KSRVAR	$0.993^{(5)}$	$0.936^{(5)}$	$0.894^{(4)}$	$0.824^{(3)}$	$0.768^{(4)}$	$0.744^{(3)}$
LGDP						
VAR	$1.175^{(6)}$	$1.017^{(6)}$	$0.926^{(6)}$	$0.904^{(6)}$	$0.850^{(4)}$	$0.773^{(4)}$
BVAR1	$0.906^{(1)}$	$0.825^{(1)}$	$0.773^{(1)}$	$0.749^{(1)}$	$0.662^{(1)}$	$0.556^{(1)}$
BVAR2	$0.971^{(2)}$	$0.896^{(2)}$	$0.847^{(3)}$	$0.812^{(3)}$	$0.707^{(2)}$	$0.584^{(2)}$
SRVAR	$1.098^{(4)}$	$0.936^{(3)}$	$0.842^{(2)}$	$0.796^{(2)}$	$0.719^{(3)}$	$0.597^{(3)}$
KVAR	$1.108^{(5)}$	$0.948^{(5)}$	$0.883^{(5)}$	$0.892^{(5)}$	$0.918^{(6)}$	$0.864^{(6)}$
HVAR	$1.281^{(7)}$	$1.163^{(7)}$	$1.161^{(7)}$	$1.783^{(7)}$	$1.150^{(7)}$	$0.991^{(7)}$
KSRVAR	$1.088^{(3)}$	$0.937^{(4)}$	$0.867^{(4)}$	$0.866^{(4)}$	$0.872^{(5)}$	$0.816^{(5)}$
LPIND						
VAR	$0.902^{(4)}$	$0.777^{(4)}$	$0.677^{(4)}$	$0.659^{(4)}$	$0.629^{(4)}$	$0.573^{(2)}$
BVAR1	$0.935^{(5)}$	$^{*}0.819^{(5)}$	$0.786^{(5)}$	$0.887^{(5)}$	$1.076^{(5)}$	$1.171^{(5)}$
BVAR2	$0.977^{(6)}$	$0.960^{(7)}$	$0.977^{(7)}$	$1.079^{(7)}$	$^{*}1.291^{(7)}$	$^{*}1.412^{(7)}$
SRVAR	$0.854^{(1)}$	$0.726^{(3)}$	$0.580^{(3)}$	$^{*}0.505^{(3)}$	$0.575^{(3)}$	$^{*}0.597^{(4)}$
KVAR	$0.871^{(3)}$	$0.700^{(1)}$	$0.542^{(1)}$	$^{*}0.483^{(1)}$	$0.543^{(1)}$	$0.553^{(1)}$
HVAR	$1.037^{(7)}$	$0.887^{(6)}$	$0.822^{(6)}$	$0.937^{(6)}$	$1.159^{(6)}$	$1.247^{(6)}$
KSRVAR	$0.862^{(2)}$	$^{*}0.704^{(2)}$	$0.553^{(2)}$	$0.495^{(2)}$	$0.562^{(2)}$	$0.577^{(3)}$

Theil U statistics with a "*" indicate forecast accuracies which are significantly different than those produced by the no-change forecast. The numbers in parentheses are the order from the smallest to the largest Theil U statistics.

The Theil U statistics are reported in Table 3.6. With the first difference model, HVAR has the best overall interest rate forecast, since it has the lowest Theil U statistics at horizons 3, 4 and 6, and the second lowest values at horizons 2 and 8. SRVAR usually has lower Theil U values than do the BVARs for this variable. The most accurate LGDP forecasts are obtained from the BVAR1 estimator. KVAR yields

the most accurate price level forecast. In this model BVAR1 has a lower Theil U statistic than BVAR2 at all horizons. Almost all forecast performances are not statistically different from the no-change forecast performance. Compared to the log level model, the Theil U statistics of LPIND are much higher.

The sign test results of the BVARs and SRVAR are reported in Table 3.7. There are few cases in which forecast accuracies of BVARs and SRVAR are statistically different from each other. BVAR1 has a statistically more accurate 8 period-ahead forecast of LGDP than SRVAR, and SRVAR has statistically more accurate 4, 6 and 8 periods-ahead forecasts of LPIND than either BVAR1 or BVAR2.

SRVAR is never statistically less accurate than VAR, KVAR, HVAR and KSRVAR. BVAR1 is less accurate than KVAR in forecasting 4 periods-ahead for LGDP, and for 6 and 8 periods-ahead for LPIND. VAR is statistically better than BVAR1 in forecasting 6 and 8 periods-ahead LPIND; BVAR is more accurate than VAR in forecasting 2, 6 and 8 periods-ahead for LGDP.

The overall forecast accuracy measures of the first difference model are in Table 3.8. In 1 period-ahead, BVAR1 has the lowest overall forecast accuracy measures, and in 2, 3, 4, 6 and 8 periods-ahead SRVAR has the best values. With this model SRVAR usually does not have the lowest Theil U statistics, but, SRVAR performs well in forecasting all the three variables and therefore, it has the best overall forecast performance. RATE and LGDP forecasts of BVAR1 produce low Theil U statistics; however, LPIND forecasts of BVAR1 are substantially worst than SRVAR. Unexpectedly, BVAR2 usually has inferior values compared to the other

Table 3.7
First Difference Quarterly Data, BVAR1, BVAR2 and SRVAR Sign Test Values

	BVAR1					
RATE	1	2	3	4	6	8
VAR	-0.89	-1.15	0.00	-0.24	1.29	1.94
BVAR2	-0.89	-2.06	-1.89	-2.18	-2.32	-2.50
SRVAR	-0.45	0.69	1.41	1.21	1.81	1.94
KVAR	-0.45	-0.69	-0.94	-0.24	1.29	1.94
HVAR	-0.45	0.23	0.47	1.21	1.81	3.05
KSRVAR	0.00	-0.69	-0.47	-0.24	1.29	1.94
LGDP						
VAR	-0.89	-2.06	-1.89	-1.21	-2.32	-2.50
BVAR2	0.00	-2.52	-1.89	-2.67	-1.81	-1.94
SRVAR	-0.45	-0.69	0.00	0.73	0.77	-2.50
KVAR	-1.34	-1.61	-1.89	2.18	-3.87	-3.61
HVAR	-2.24	-2.52	-3.30	-4.12	-3.87	-3.61
KSRVAR	-1.34	-1.15	-2.36	-1.21	-3.87	-3.61
LPIND						
VAR	0.89	-0.69	0.00	1.21	2.32	2.50
BVAR2	0.45	-1.15	-1.89	-0.73	-1.81	-3.05
SRVAR	1.79	0.69	1.41	2.18	2.84	3.05
KVAR	0.00	1.15	1.89	1.70	3.36	3.61
HVAR	-0.89	-1.15	-0.47	-0.24	-1.29	-0.83
KSRVAR	0.89	1.15	1.41	1.70	3.36	3.61
	BVAR2					
RATE	1	2	3	4	6	8
VAR	0.45	0.23	0.47	0.73	1.81	1.94
SRVAR	0.45	0.69	0.94	1.21	1.81	1.94
KVAR	0.45	-0.23	0.00	0.73	1.81	1.94
HVAR	0.00	1.15	0.47	1.70	2.32	2.50
KSRVAR	0.45	0.23	0.00	0.73	1.81	1.94
LGDP						
VAR	-0.89	-1.61	-0.94	-0.73	-2.32	-2.50
SRVAR	-0.45	-0.23	0.47	1.21	1.29	-0.83
KVAR	-0.45	-0.23	-1.41	-1.21	-3.87	-3.61
HVAR	-3.13	-2.96	-3.30	-4.12	-3.87	-3.61
KSRVAR	-0.45	-0.23	-0.47	-1.70	-3.36	-3.61
LPIND						
VAR	-0.45	-0.23	0.47	1.70	2.84	3.05
SRVAR	0.45	1.61	1.89	2.18	2.84	3.05
KVAR	0.00	0.69	1.41	2.18	3.36	3.61
HVAR	-1.79	-0.23	1.41	1.21	1.81	1.94
KSRVAR	0.00	1.15	1.89	2.18	3.36	3.61

(Table con'd.)

	SRVAR					
RATE	1	2	3	4	6	8
VAR	-1.34	-1.15	-1.89	-0.73	0.26	0.83
KVAR	-1.79	-0.69	-1.89	-0.73	0.26	0.83
HVAR	-1.79	-0.69	0.94	1.70	1.29	-0.83
KSRVAR	-1.34	-0.69	-1.41	-0.73	0.26	0.83
LGDP						
VAR	-0.89	-2.06	-1.41	-0.73	-1.29	-1.94
KVAR	0.89	0.23	0.00	-1.21	-2.32	-3.05
HVAR	-0.45	-2.06	-1.89	-2.18	-2.84	-3.61
KSRVAR	0.89	1.15	0.47	-1.21	-1.81	-3.05
LPIND						
VAR	-1.34	-1.61	-1.89	-1.70	0.26	1.39
KVAR	-0.45	-1.15	-0.94	-1.21	-0.26	0.83
HVAR	-2.24	-1.15	-0.47	-1.21	-3.36	-3.61
KSRVAR	0.00	0.23	0.47	-0.24	-0.77	0.83

imposing better restrictions. We only use the Minnesota prior since this is the common practice in VAR analysis, and we are only interested in comparing the forecast performances of the shrinkage estimators.

We also forecast monthly macroeconomic variables. To estimate the lag structure, a search over lag lengths from 1 to 18 is performed. The AIC criterion forecasts. VAR has an average performance, and the asymmetric VARs usually are not better than VAR. This result shows that shrinkage estimators can have better forecasting performance than unrestricted VAR forecasts.

The reader should note that these particular results of BVAR and SRVAR forecasts depend on the restrictions imposed. For instance, we found that for the first difference model, imposing zero restrictions on all coefficients except the intercept yields better forecast results. Thus, a forecaster can improve his/her forecasts by picked 13 lags for the symmetric VAR. As before, the BVARs and SRVAR have the

Table 3.8
First Difference Quarterly Data, Overall Fitness Measures

SOFM	1	2	3	4	6	8
VAR	-11.71$^{(5)}$	-10.31$^{(4)}$	-9.59$^{(5)}$	-9.17$^{(4)}$	-8.12$^{(4)}$	-7.61$^{(4)}$
BVAR1	-11.97$^{(1)}$	-10.46$^{(2)}$	-9.71$^{(2)}$	-9.22$^{(2)}$	-8.20$^{(2)}$	-7.72$^{(2)}$
BVAR2	-11.64$^{(6)}$	-9.87$^{(7)}$	-8.92$^{(7)}$	-8.47$^{(7)}$	-7.33$^{(7)}$	-6.70$^{(6)}$
SRVAR	-11.96$^{(2)}$	-10.62$^{(1)}$	-10.03$^{(1)}$	-9.69$^{(1)}$	-8.78$^{(1)}$	-8.24$^{(1)}$
KVAR	-11.73$^{(4)}$	-10.29$^{(5)}$	-9.60$^{(4)}$	-9.14$^{(5)}$	-8.10$^{(5)}$	-6.60$^{(7)}$
HVAR	-11.45$^{(7)}$	-10.02$^{(6)}$	-9.32$^{(6)}$	-8.80$^{(6)}$	-7.90$^{(6)}$	-7.48$^{(5)}$
KSRVAR	-11.81$^{(3)}$	-10.36$^{(3)}$	-9.65$^{(3)}$	-9.19$^{(3)}$	-8.15$^{(3)}$	-7.62$^{(3)}$
SUM						
VAR	3.089$^{(6)}$	2.721$^{(5)}$	2.507$^{(5)}$	2.439$^{(4)}$	2.300$^{(4)}$	2.124$^{(2)}$
BVAR1	2.712$^{(1)}$	2.487$^{(2)}$	2.390$^{(4)}$	2.467$^{(5)}$	2.579$^{(5)}$	2.584$^{(5)}$
BVAR2	2.909$^{(3)}$	2.836$^{(6)}$	2.806$^{(7)}$	2.881$^{(7)}$	2.991$^{(7)}$	2.984$^{(6)}$
SRVAR	2.894$^{(2)}$	2.451$^{(1)}$	2.187$^{(1)}$	2.065$^{(1)}$	2.049$^{(1)}$	1.969$^{(1)}$
KVAR	3.025$^{(5)}$	2.645$^{(4)}$	2.368$^{(3)}$	2.222$^{(3)}$	2.227$^{(3)}$	2.143$^{(4)}$
HVAR	3.279$^{(7)}$	2.857$^{(7)}$	2.684$^{(6)}$	2.782$^{(6)}$	2.959$^{(6)}$	2.972$^{(7)}$
KSRVAR	2.944$^{(4)}$	2.577$^{(3)}$	2.314$^{(2)}$	2.185$^{(2)}$	2.203$^{(2)}$	2.136$^{(3)}$

same lag structure as the symmetric VAR. KVAR has 14 lags for RATE, 2 lags for the log level of industrial production (LIP), 7 lags for the log level of M2 (LM2), and 10 lags for consumer price index (LPIND) in each equation. Hsiao's method specifies the lag lengths for the first equation as 17, 3, 6, and 15; for the second equation as 4, 6, 7, and 1; for the third equation as 15, 0, 11, and 9; and for the fourth equation as 5, 1, 4, and 10.

The grid search for the values of the hyperparameters of the BVAR estimators uses the same intervals as before. For convenience, only the parameter values that have been selected are reported. The hyperparameters are chosen such that the Theil U and overall fitness statistics are minimized. BVAR1 values are $d=1$, $\theta=0.05$ and $\lambda=0.05$. BVAR2 hyperparameters to forecast RATE are $d=1$, $\theta=0.05$, and

$\lambda=0.25$. These numbers for forecasting LIP are 7.0, 0.05, and 0.05, and for forecasting LPIND are 1, 0.05, and 0.35, respectively. For the Stein-rule estimation, the multiplier is set to one, since $m=1$ has better forecasting performance than $m=2$ for the 1985-1989 period.

The Theil U statistics are reported in Table 3.9, and the sign-test statistics are in Table 3.10. Almost all interest rate forecasts of the no-change method are statistically the most accurate, and all price level forecasts of the no-change method are statistically the worst. After the no-change method, BVAR1 has the lowest RATE forecast errors in horizons 1, 2, 3 and 6; BVAR1 has the lowest LIP forecast error in the 8-period ahead horizon, and the lowest LPIND forecast error in horizons 1, 2 and 3. SRVAR is only the most accurate forecaster for LIP in horizons 2, 6 and 12. BVAR1 has statistically better forecasts than SRVAR in 2, 3, 6, 12 and 24 period-ahead RATE forecasting; 8 period-ahead LIP forecast; 3, 6 and 12 period-ahead LPIND forecasts. On the other hand, SRVAR only has a statistically better forecast than BVAR1 in the 3 period-ahead LIP forecasting. The unrestricted VAR forecasts are never the most accurate. HVAR is statistically superior in the 12 and 24 period-ahead forecasts for RATE and LPIND; KSRVAR is superior in forecasting 1 and 2 periods-ahead for LIP.

The overall forecast measures SOFM and SUM are reported in Table 3.11. BVAR has the best overall performance. The second best forecaster is HVAR. SRVAR has the third best performance. Even though BVAR2 picks the hyperparameters for each forecasted variable separately, it's performance is lower

Table 3.9
Log Level Monthly Data Theil U Statistics

RATE	1	2	3	6	12	24
VAR	*1.400$^{(6)}$	*1.753$^{(6)}$	*1.864$^{(6)}$	*1.866$^{(5)}$	*1.461$^{(3)}$	*1.365$^{(4)}$
BVAR1	*1.068$^{(1)}$	1.127$^{(1)}$	*1.169$^{(1)}$	*1.251$^{(1)}$	*1.279$^{(2)}$	*1.279$^{(3)}$
BVAR2	*1.420$^{(7)}$	*1.723$^{(5)}$	*1.850$^{(5)}$	*1.853$^{(4)}$	*1.466$^{(4)}$	*1.262$^{(2)}$
SRVAR	*1.283$^{(3)}$	*1.592$^{(3)}$	*1.705$^{(3)}$	*1.790$^{(3)}$	*1.596$^{(5)}$	*1.597$^{(6)}$
KVAR	*1.340$^{(5)}$	*1.777$^{(7)}$	*1.922$^{(7)}$	*1.962$^{(7)}$	*1.612$^{(6)}$	*1.381$^{(5)}$
HVAR	*1.212$^{(2)}$	*1.447$^{(2)}$	*1.492$^{(2)}$	*1.410$^{(2)}$	*1.195$^{(1)}$	*1.224$^{(1)}$
KSRVAR	*1.309$^{(4)}$	*1.654$^{(4)}$	*1.790$^{(4)}$	*1.880$^{(6)}$	*1.690$^{(7)}$	*1.569$^{(7)}$
LIP						
VAR	0.949$^{(5)}$	*0.841$^{(4)}$	*0.767$^{(4)}$	*0.731$^{(4)}$	0.982$^{(4)}$	*1.535$^{(7)}$
BVAR1	*0.957$^{(6)}$	0.954$^{(6)}$	0.958$^{(6)}$	0.990$^{(7)}$	0.982$^{(4)}$	*0.696$^{(1)}$
BVAR2	1.002$^{(7)}$	0.993$^{(7)}$	0.988$^{(7)}$	*0.975$^{(6)}$	*0.936$^{(3)}$	*0.705$^{(2)}$
SRVAR	0.900$^{(4)}$	0.802$^{(1)}$	*0.745$^{(2)}$	*0.710$^{(1)}$	*0.776$^{(1)}$	1.058$^{(3)}$
KVAR	0.895$^{(2)}$	*0.817$^{(2)}$	*0.746$^{(3)}$	*0.721$^{(3)}$	1.025$^{(6)}$	*1.391$^{(6)}$
HVAR	0.903$^{(3)}$	*0.865$^{(5)}$	0.836$^{(5)}$	*0.900$^{(5)}$	*1.076$^{(7)}$	1.117$^{(5)}$
KSRVAR	*0.877$^{(1)}$	*0.802$^{(1)}$	*0.743$^{(1)}$	0.715$^{(2)}$	*0.869$^{(2)}$	1.102$^{(4)}$
LPIND						
VAR	*0.505$^{(4)}$	*0.440$^{(6)}$	*0.407$^{(5)}$	*0.374$^{(5)}$	*0.401$^{(5)}$	*0.458$^{(5)}$
BVAR1	*0.479$^{(1)}$	*0.404$^{(1)}$	*0.378$^{(1)}$	*0.332$^{(2)}$	*0.338$^{(2)}$	*0.440$^{(4)}$
BVAR2	*0.664$^{(7)}$	*0.643$^{(7)}$	*0.637$^{(7)}$	*0.628$^{(7)}$	*0.697$^{(7)}$	*0.812$^{(7)}$
SRVAR	*0.499$^{(3)}$	*0.433$^{(4)}$	*0.402$^{(4)}$	*0.374$^{(5)}$	*0.431$^{(6)}$	*0.539$^{(6)}$
KVAR	*0.512$^{(6)}$	*0.438$^{(5)}$	*0.401$^{(3)}$	*0.351$^{(4)}$	*0.343$^{(3)}$	*0.352$^{(2)}$
HVAR	*0.495$^{(2)}$	*0.414$^{(2)}$	*0.472$^{(6)}$	*0.320$^{(1)}$	*0.305$^{(1)}$	*0.327$^{(1)}$
KSRVAR	*0.505$^{(4)}$	*0.430$^{(3)}$	*0.393$^{(2)}$	*0.345$^{(3)}$	*0.345$^{(4)}$	*0.373$^{(3)}$

Theil U statistics with a "*" indicate forecast accuracies which are significantly different than those produced by the no-change forecast. The numbers in parentheses are the order from the smallest to the largest Theil U statistics.

than that of BVAR1 which uses one set of hyperparameters to forecast all three variables.

The variables in the growth rate model are the interest rate, growth rate of industrial production (GIP), growth rate of M2 (GM2), and the growth rate of the consumer price index (GPIND). With this model we forecast RATE, LIP and LPIND

Table 3.10
Log Level Monthly Data, BVAR1, BVAR2 and SRVAR Sign Test Values

	BVAR1					
RATE	1	2	3	6	12	24
VAR	-1.81	-5.08	-5.51	-4.72	-2.71	-0.82
BVAR2	-3.87	-4.56	-4.73	-4.72	-2.71	3.12
SRVAR	-1.55	-4.04	-5.25	-5.26	-3.86	-5.43
KVAR	-4.13	-4.56	-4.99	-4.72	-3.57	-0.82
HVAR	-1.55	-1.69	-3.41	-2.56	1.86	3.45
KSRVAR	-3.36	-4.69	-4.99	-4.72	-4.14	-4.77
LIP						
VAR	-0.26	1.69	1.84	1.75	-1.57	-3.45
BVAR2	-1.03	-0.65	-0.79	0.13	0.43	-0.16
SRVAR	-0.26	1.69	2.36	1.48	1.57	-2.79
KVAR	1.03	2.21	2.10	2.02	-1.57	-2.79
HVAR	0.77	1.43	0.79	0.40	-1.29	-2.79
KSRVAR	1.03	2.47	2.10	3.10	1.00	-2.79
LPIND						
VAR	-1.55	-1.69	-2.36	-3.91	-2.43	1.48
BVAR2	-4.65	-4.82	-5.25	-5.26	-6.14	-6.08
SRVAR	-1.81	-1.17	-2.63	-3.64	-2.14	-1.48
KVAR	-1.29	-1.17	-1.58	-2.56	-0.71	3.78
HVAR	-0.52	0.39	0.00	-0.67	1.00	2.79
KSRVAR	-1.29	-2.73	-2.36	-4.18	-4.43	-3.78

	BVAR2					
RATE	1	2	3	6	12	24
VAR	0.77	0.39	0.53	0.94	0.71	-4.77
SRVAR	2.58	1.95	2.10	1.75	-4.43	-5.75
KVAR	-0.52	-0.13	0.00	-1.75	-3.00	-5.43
HVAR	2.58	3.25	3.94	5.26	5.86	2.14
KSRVAR	0.77	1.17	1.05	0.67	-5.00	-6.08
LIP						
VAR	0.26	1.69	2.89	1.75	-1.86	-3.45
SRVAR	0.00	1.69	3.41	1.21	1.57	-2.79
KVAR	1.81	1.69	3.15	2.56	-1.00	-2.79
HVAR	0.52	1.69	2.36	-0.13	-2.43	-2.79
KSRVAR	1.81	2.47	3.41	2.56	0.71	-2.47
LPIND						
VAR	-4.39	5.34	5.78	5.53	6.14	5.75
SRVAR	4.39	5.34	6.30	5.53	6.14	4.77
KVAR	4.39	4.82	5.78	5.53	6.43	6.08
HVAR	4.65	5.08	5.78	5.80	6.71	6.08
KSRVAR	2.07	2.21	3.15	2.02	1.86	-0.82

(Table con'd.)

RATE	SRVAR 1	2	3	6	12	24
VAR	-4.39	-4.56	-4.73	-2.83	4.71	5.75
KVAR	-2.32	-2.47	-2.63	-0.94	1.29	4.44
HVAR	0.52	2.73	3.15	5.26	6.71	6.08
KSRVAR	-1.29	-2.21	-1.58	-0.67	-0.14	1.81
LIP						
VAR	-1.55	-1.69	-1.05	0.13	-3.86	-3.78
KVAR	0.77	-0.65	0.26	-0.67	-3.86	-2.79
HVAR	0.52	-1.43	-2.10	-3.37	-4.14	-1.48
KSRVAR	0.77	-0.39	-1.05	-0.40	-2.14	-1.15
LPIND						
VAR	1.55	0.91	0.79	0.67	3.29	4.77
KVAR	0.26	0.91	0.79	0.40	3.57	4.77
HVAR	1.03	-0.39	1.58	0.67	1.86	3.45
KSRVAR	-0.77	-1.17	-1.31	-2.29	-2.71	-2.47

and compare the performance across different models. The symmetric VAR has 13 lags. KVAR has 7 lags for RATE, 2 lags for GIP, 7 lags for GM2, and 12 lags forGPIND in each equation. Hsiao's method specifies the lag lengths for the first equation as 9, 3, 5, and 6; for the second equation as 4, 2, 6, and 5; for the third equation as 8, 2, 10, and 10; and for the fourth equation as 5, 0, 3, and 9.

The grid search of the hyperparameters of the BVAR estimators uses the same intervals as before. BVAR1 values are $d=1$, $\theta=0.10$ and $\lambda=0.30$. BVAR2 hyperparameters to forecast RATE are $d=1$, $\theta=0.25$, and $\lambda=0.05$. These numbers for forecasting LIP are 1.0, 0.10, and 0.45, and for forecasting LPIND are 1, 0.50, and 0.30, respectively. For SRVAR, the multiplier is set to one since $m=1$ has better forecasting performance than $m=2$ for periods 1985-1989.

The Theil U statistics are reported in Table 3.12 and the sign-test results are reported in Table 3.13. BVAR1 is most accurate in forecasting 1, 2, 3 and 6 period-

Table 3.11
Log Level Monthly Data, Overall Fitness Measures

SOFM	1	2	3	6	12	24
VAR	-12.45$^{(6)}$	-11.00$^{(6)}$	-10.21$^{(6)}$	-8.83$^{(6)}$	-7.36$^{(6)}$	-5.61$^{(7)}$
BVAR1	-12.74$^{(1)}$	-11.42$^{(1)}$	-10.60$^{(1)}$	-9.17$^{(1)}$	-7.82$^{(1)}$	-6.84$^{(2)}$
BVAR2	-12.25$^{(7)}$	-10.80$^{(7)}$	-10.00$^{(7)}$	-8.77$^{(7)}$	-7.66$^{(2)}$	-6.96$^{(1)}$
SRVAR	-12.57$^{(3)}$	-11.13$^{(3)}$	-10.32$^{(3)}$	-8.89$^{(4)}$	-7.44$^{(4)}$	-5.95$^{(5)}$
KVAR	-12.48$^{(5)}$	-11.02$^{(5)}$	-10.23$^{(5)}$	-8.85$^{(5)}$	-7.36$^{(6)}$	-5.78$^{(6)}$
HVAR	-12.62$^{(2)}$	-11.19$^{(2)}$	-10.41$^{(2)}$	-9.01$^{(2)}$	-7.56$^{(3)}$	-6.08$^{(3)}$
KSRVAR	-12.56$^{(4)}$	-11.11$^{(4)}$	-10.30$^{(4)}$	-8.90$^{(3)}$	-7.44$^{(4)}$	-6.01$^{(4)}$
SUM						
VAR	2.854$^{(6)}$	3.033$^{(5)}$	3.038$^{(5)}$	2.970$^{(5)}$	2.844$^{(4)}$	3.358$^{(7)}$
BVAR1	2.504$^{(1)}$	2.485$^{(1)}$	2.505$^{(1)}$	2.573$^{(1)}$	2.600$^{(2)}$	2.459$^{(1)}$
BVAR2	3.086$^{(7)}$	3.360$^{(7)}$	3.474$^{(7)}$	3.456$^{(7)}$	3.100$^{(7)}$	2.780$^{(3)}$
SRVAR	2.682$^{(3)}$	2.827$^{(3)}$	2.851$^{(3)}$	2.875$^{(3)}$	2.803$^{(3)}$	3.194$^{(6)}$
KVAR	2.804$^{(5)}$	3.033$^{(5)}$	3.070$^{(6)}$	3.035$^{(6)}$	2.980$^{(6)}$	3.124$^{(5)}$
HVAR	2.609$^{(2)}$	2.726$^{(2)}$	2.700$^{(2)}$	2.629$^{(2)}$	2.576$^{(1)}$	2.669$^{(2)}$
KSRVAR	2.691$^{(4)}$	2.885$^{(4)}$	2.926$^{(4)}$	2.940$^{(4)}$	2.904$^{(5)}$	3.044$^{(4)}$

ahead for RATE, and 1, 2 and 3 period-ahead for LPIND. BVAR2 has the lowest Theil U statistic in forecasting LIP, and 6, 12 and 24 periods-ahead of LPIND. SRVAR does not perform well.

The overall forecast performance measures are in Table 3.14. SOFM identifies BVAR1 as the overall best forecaster and BVAR2 as the second best. SUM identifies BVAR1 and BVAR2 as the best and second best forecasters at horizons 1, 2, 3 and 6. HVAR has the third best performance. SRVAR and VAR are the worst two forecasters.

3.4.2 Monte-Carlo Experiment Results

Two Monte-Carlo experiments are conducted to further investigate the relative forecasting performances of Bayesian and Stein-rule VARs. We concentrate on the

Table 3.12
First Difference Monthly Data Theil U Statistics

RATE	1	2	3	6	12	24
VAR	1.100 (7)	1.276 (7)	*1.291 (7)	*1.121 (7)	0.776 (1)	*0.743 (1)
BVAR1	*0.890 (1)	0.947 (1)	*0.962 (1)	0.965 (1)	0.915 (6)	0.944 (6)
BVAR2	0.962 (2)	0.979 (2)	0.987 (2)	1.000 (4)	0.989 (7)	0.987 (7)
SRVAR	1.050 (5)	1.198 (6)	*1.211 (6)	1.094 (6)	0.811 (2)	*0.760 (2)
KVAR	1.031 (4)	1.168 (5)	*1.147 (5)	0.999 (3)	*0.828 (3)	*0.850 (3)
HVAR	1.051 (6)	*1.128 (3)	1.133 (4)	1.004 (5)	*0.856 (5)	*0.859 (5)
KSRVAR	1.019 (3)	1.151 (4)	*1.131 (3)	0.995 (2)	*0.832 (4)	*0.852 (4)
LIP						
VAR	0.913 (5)	*0.777 (5)	*0.728 (5)	*0.582 (4)	*0.463 (6)	*0.248 (6)
BVAR1	*0.772 (2)	*0.641 (2)	*0.566 (2)	*0.449 (2)	*0.324 (3)	*0.174 (2)
BVAR2	*0.769 (1)	*0.633 (1)	*0.551 (1)	*0.428 (1)	*0.303 (1)	*0.166 (1)
SRVAR	0.897 (4)	*0.770 (4)	*0.719 (4)	*0.585 (7)	*0.469 (7)	*0.257 (7)
KVAR	0.923 (7)	0.806 (7)	*0.753 (7)	*0.584 (5)	*0.416 (4)	*0.231 (4)
HVAR	0.839 (3)	*0.709 (3)	*0.630 (3)	*0.461 (3)	*0.323 (2)	*0.177 (3)
KSRVAR	0.917 (6)	0.803 (6)	*0.750 (6)	*0.584 (5)	*0.419 (5)	*0.231 (4)
LPIND						
VAR	0.880 (7)	0.811 (7)	0.732 (7)	0.613 (7)	0.499 (4)	*0.390 (2)
BVAR1	0.817 (1)	*0.717 (1)	*0.647 (1)	*0.546 (2)	*0.504 (6)	0.525 (7)
BVAR2	0.835 (2)	0.743 (3)	0.670 (3)	0.544 (1)	*0.438 (1)	*0.374 (1)
SRVAR	0.856 (6)	0.789 (6)	0.717 (6)	0.607 (6)	0.496 (3)	*0.393 (3)
KVAR	0.847 (5)	0.759 (5)	0.682 (5)	0.570 (5)	0.504 (6)	*0.455 (5)
HVAR	0.837 (3)	0.739 (2)	0.653 (2)	0.549 (3)	0.472 (2)	*0.417 (4)
KSRVAR	0.842 (4)	0.754 (4)	0.678 (4)	0.569 (4)	0.503 (5)	*0.455 (5)

Theil U statistics with a "*" indicate forecast accuracies which are significantly different than those produced by the no-change forecast. The numbers in parentheses are the order from the smallest to the largest Theil U statistics.

forecasting performances of these two estimators since, in this essay, the Stein-rule VAR is presented as an alternative to the Bayesian VAR. The experiments are designed to explore the effect of using correct or incorrect restrictions on forecast accuracy.

Table 3.13
First Difference Monthly Data, BVAR1, BVAR2 and SRVAR Sign Test Values

	BVAR1					
RATE	1	2	3	6	12	24
VAR	-1.29	-3.52	-3.68	-3.37	3.57	3.45
BVAR2	-1.81	-1.43	0.00	-2.83	-4.14	-3.45
SRVAR	-0.77	-2.47	-3.41	-3.10	3.57	3.45
KVAR	-1.29	-1.69	-2.10	-0.40	3.57	2.79
HVAR	-1.29	-3.25	-3.15	-0.13	3.00	2.14
KSRVAR	-1.29	-1.69	-2.36	-0.94	3.57	2.79
LIP						
VAR	-1.81	-2.47	-1.84	-2.83	-4.43	-2.79
BVAR2	0.77	0.39	0.00	0.13	1.29	2.79
SRVAR	-1.55	-1.43	-2.10	-3.10	-4.43	-3.12
KVAR	-2.58	-1.95	-2.63	-2.83	-3.29	-1.81
HVAR	-1.55	-1.43	-1.84	-1.75	-1.00	1.81
KSRVAR	-2.32	-1.69	-2.89	-3.10	-3.57	-2.47
LPIND						
VAR	-2.07	-0.13	-0.79	-0.67	-1.00	2.47
BVAR2	0.52	0.13	-0.79	-0.13	1.29	5.10
SRVAR	-1.29	-0.65	-0.26	-1.21	-1.00	2.47
KVAR	-1.03	0.65	0.26	-0.13	0.43	2.79
HVAR	0.00	0.39	0.00	-0.40	1.00	4.11
KSRVAR	-1.03	0.91	0.00	-0.40	0.43	3.12

	BVAR2					
RATE	1	2	3	6	12	24
VAR	1.03	-1.43	-3.68	-0.40	3.57	3.45
SRVAR	1.55	-1.43	-3.68	-0.40	3.86	3.45
KVAR	0.26	-0.39	-2.63	0.94	3.86	2.79
HVAR	0.26	-0.91	-1.84	2.29	3.86	2.79
KSRVAR	0.26	-0.39	-2.36	0.94	3.86	2.79
LIP						
VAR	-1.55	-1.95	-3.41	-3.10	-4.43	-3.45
SRVAR	-1.81	-2.47	-3.15	-3.37	-4.43	-2.79
KVAR	-2.58	-2.47	-3.94	-3.10	-4.14	-2.47
HVAR	-2.07	-1.69	-1.05	-2.56	-2.43	-0.49
KSRVAR	-2.58	-2.21	-3.94	-3.10	-4.14	-2.47
LPIND						
VAR	-2.58	-0.39	-0.79	-2.56	-3.00	0.49
SRVAR	-1.81	-0.91	-0.79	-2.56	-3.00	-0.16
KVAR	-0.77	0.65	0.79	-0.40	-1.57	-2.14
HVAR	0.00	1.17	2.63	2.02	-1.00	-1.15
KSRVAR	-1.03	0.91	0.79	-0.40	-1.86	-2.47

(Table con'd.)

	SRVAR					
RATE	1	2	3	6	12	24
VAR	-2.84	-2.99	-3.94	-2.29	2.43	3.12
KVAR	-0.77	0.91	1.05	2.02	-0.71	-2.79
HVAR	-1.03	0.91	1.05	2.29	-0.71	-3.12
KSRVAR	-0.52	0.91	1.58	2.02	-1.29	-3.12
LIP						
VAR	-0.26	-0.91	-1.05	0.40	2.14	2.79
KVAR	-1.81	-1.95	-3.41	0.13	6.14	3.78
HVAR	2.58	1.95	2.89	5.53	4.43	3.12
KSRVAR	-1.03	-0.65	-1.84	-0.40	6.14	3.78
LPIND						
VAR	-2.32	-1.17	-1.05	-1.75	-1.86	0.49
KVAR	0.00	0.91	1.05	1.21	1.29	-1.81
HVAR	0.52	1.43	2.36	2.83	2.43	0.49
KSRVAR	0.26	0.39	0.53	1.48	1.29	-1.81

Two VAR models are simulated. The first experiment is designed to find out which estimator has the most accurate forecast when the imposed zero restrictions are correct. We use the same restrictions that are employed in the previous section which assumes that the underlying data generating process is a random walk process. The second experiment is designed to investigate the outcome when the restrictions are not correct.

The number of repetitions in the experiments is set to 1000. In each simulation, 204 observations are created for each variable. The first 60 observations are discarded, leaving 144 observations (note that this is the number of observations we have used in the quarterly data). Both estimators employ the same symmetric lag structure. As in the real-time data, 124 observations are used to determine the optimum lag length using the AIC criterion. Since we are concerned with the relative forecast accuracies, we report the differences in Theil U statistics.

Table 3.14
First Difference Monthly Data, Overall Fitness Measures

SOFM	1	2	3	6	12	24
VAR	-12.60$^{(7)}$	-11.20$^{(7)}$	-10.40$^{(7)}$	-9.14$^{(7)}$	-8.16$^{(6)}$	-7.20$^{(5)}$
BVAR1	-12.98$^{(1)}$	-11.70$^{(1)}$	-11.05$^{(1)}$	-9.73$^{(1)}$	-8.55$^{(1)}$	-7.58$^{(1)}$
BVAR2	-12.90$^{(2)}$	-11.72$^{(2)}$	-11.01$^{(2)}$	-9.67$^{(2)}$	-8.49$^{(2)}$	-7.56$^{(2)}$
SRVAR	-12.67$^{(6)}$	-11.29$^{(6)}$	-10.49$^{(6)}$	-9.17$^{(6)}$	-8.11$^{(7)}$	-7.11$^{(7)}$
KVAR	-12.69$^{(5)}$	-11.34$^{(5)}$	-10.56$^{(5)}$	-9.33$^{(4)}$	-8.22$^{(4)}$	-7.21$^{(4)}$
HVAR	-12.75$^{(3)}$	-11.52$^{(3)}$	-10.79$^{(3)}$	-9.59$^{(3)}$	-8.48$^{(3)}$	-7.56$^{(2)}$
KSRVAR	-12.71$^{(4)}$	-11.36$^{(4)}$	-10.58$^{(4)}$	-9.33$^{(4)}$	-8.21$^{(5)}$	-7.20$^{(6)}$
SUM						
VAR	2.893$^{(7)}$	2.866$^{(7)}$	2.751$^{(7)}$	2.316$^{(7)}$	1.738$^{(3)}$	1.381$^{(1)}$
BVAR1	2.479$^{(1)}$	2.305$^{(1)}$	2.174$^{(1)}$	1.960$^{(1)}$	1.743$^{(4)}$	1.643$^{(7)}$
BVAR2	2.567$^{(2)}$	2.355$^{(2)}$	2.208$^{(2)}$	1.973$^{(2)}$	1.730$^{(2)}$	1.527$^{(4)}$
SRVAR	2.803$^{(6)}$	2.757$^{(6)}$	2.647$^{(6)}$	2.286$^{(6)}$	1.776$^{(7)}$	1.410$^{(2)}$
KVAR	2.801$^{(5)}$	2.733$^{(5)}$	2.582$^{(5)}$	2.153$^{(5)}$	1.748$^{(5)}$	1.536$^{(5)}$
HVAR	2.727$^{(3)}$	2.576$^{(3)}$	2.416$^{(3)}$	2.014$^{(3)}$	1.650$^{(1)}$	1.454$^{(3)}$
KSRVAR	2.779$^{(4)}$	2.708$^{(4)}$	2.559$^{(4)}$	2.148$^{(4)}$	1.755$^{(6)}$	1.540$^{(6)}$

The first model consists of three first order autoregressive processes, two of which are random walk processes:

$$y_{1,t} = 0.05 + 0.9\, y_{1,t-1}$$

$$y_{2,t} = 0.05 + 1.0\, y_{2,t-1}$$

$$y_{3,t} = 0.00 + 1.0\, y_{3,t-1} \ .$$

In a VAR model framework, the first own-lag is different from zero, and the other slope coefficients are equal to zero.

To determine optimum hyperparameter values of the Bayesian VAR, Sims' overall forecast measure is employed. Using the same methodology as in the previous section, one-period-ahead forecasts for sample observations 105-124 are computed. The parameters which yield the smallest one period-ahead SOFM are picked to

forecast observations 125-144 of the sample. By looking only to the one period-ahead SOFM, the forecaster's involvement in picking the optimum hyperparameters is avoided. The intervals for the grid search are selected after a pilot Monte-Carlo study with larger intervals (0.05 to 0.95 for θ and λ, and 0.5 to 7.5 for the decay parameter) but with only 100 repetitions. The grid search interval for θ is 0.60 - 0.80 with 0.05 increments. The value of λ is picked from 0.50, 0.55, 0.60, 0.65 and 0.70, and d is picked from 5, 6 and 7. A search over lag lengths from 1 to 3 is performed.

Starting with the first 124 observations, the h-period-ahead dynamic forecasts of the variables are computed and compared to the generated data to find the forecast error. The forecasting continues until the forecast of the 144^{th} observation is computed.

Since we are interested in relative forecast performance, Table 3.15 reports the difference between the Theil U statistic of the BVAR and the SRVAR (Theil U of BVAR minus Theil U statistic of SRVAR). The numbers in parenthesis are the standard deviations of the difference. The standard deviations are computed from the differences in the Theil U statistics. Assume that $U\_BVAR_{i,h,r}$ is the Theil U statistic of the BVAR from the i^{th} variable's h-period-ahead forecast at the r^{th} repetition, and $U\_SRVAR_{i,h,r}$ is similarly defined for SRVAR. Then the numbers in parenthesis are the square roots of,

Table 3.15
Model 1, Relative Theil U of BVAR and SRVAR

	Forecast Horizon		
Variable	1	3	6
1	0.002 (0.015)	0.003 (0.021)	0.003 (0.033)
2	0.010 (0.018)	0.023 (0.068)	0.034 (0.078)
3	0.004 (0.016)	0.010 (0.027)	0.016 (0.052)

The numbers in parenthesis are the standard deviations.

$$\frac{1}{1000}\sum_{r=1}^{1000}(U_{BVAR_{i,h,r}} - U_{SRVAR_{i,h,r}})^2 - \text{mean}^2 \quad ,$$

where

$$\text{mean} = \frac{1}{1000}\sum_{r=1}^{1000}(U_{BVAR_{i,h,r}} - U_{SRVAR_{i,h,r}}) \quad .$$

The statistics are all positive, indicating that SRVAR has always lower Theil U statistic for all three variables. However, the standard deviations show that none of the differences are statistically different than zero.

The second model also consists of three equations, in which the first variable is affected by its lag only, and the second and the third variables are affected by the lags of all three variables. Thus, the restrictions on the first equation will be near correct (the slope coefficient is 0.9 not 1), and the restrictions on the last two equations will be wrong. With this simulation, we aim to find the consequence of wrong restrictions on the relative forecast performance. The model can be described as:

$$y_{1,t} = 0.5 + 0.9\, y_{1,t} + e_{1,t}$$

$$y_{2,t} = 4.8 - 4.3\, y_{1,t} + 1.2\, y_{2,t-1} - 0.2\, y_{2,t-2} - 0.01\, y_{3,t-1} + e_{2,t}$$

$$y_{3,t} = -3.0 + 1.1\, y_{1,t} + 0.001\, y_{2,t-1} + 1.6\, y_{3,t-1} - 0.6\, y_{3,t-2} + e_{3,t}$$

where,

$$e \sim N\left[0, \begin{bmatrix} 1.03 & 0 & 0 \\ 0 & 758 & 23.7 \\ 0 & 23.7 & 130 \end{bmatrix}\right] .$$

In this case, the zero restrictions of Minnesota prior are not correct. There are slope coefficients other than the own-lag coefficients which are different than zero. The coefficients of the first equation is chosen so that it will be a stationary data generating process. The coefficients of the second and third equations are obtained by replacing y_1 with 6-month commercial paper rate, y_2 with M2, and y_3 with the GDP deflator, and estimating the coefficients of the model presented above.

To select the optimum lag length, a search over lag lengths from 1 to 4 is performed. To find a small range of hyperparameter values to be used in the grid search of the actual experiment, a pilot experiment with 100 repetitions was conducted. In this pilot study, the values of θ and λ are altered between 0.05 and 0.95 with 0.1 intervals. On average, 0.77 was selected for θ and 0.84 was selected for λ. The decay parameter was altered between 0.5 and 7.5 with intervals of one, and it was found that 5.4 is the average selected value. Thus, in the actual experiment, the grid search values of θ are 0.65, 0.70, 0.75, 0.80 and 0.85; these numbers are 0.75, 0.80, 0.85, 0.90 and 0.95 for λ, and 4, 5 and 6 for the decay

Table 3.16
Model 2, Relative Theil U of BVAR and SRVAR

	Forecast Horizon		
Variable	1	3	6
1	0.001 (0.036)	0.012 (0.065)	0.027 (0.109)
2	0.007 (0.017)	0.007 (0.024)	0.007 (0.030)
3	0.135 (0.142)	0.076 (0.176)	0.047 (0.257)

The numbers in parenthesis are the standard deviations.

parameter. The results are in Table 3.16. In this simulation experiment too, SRVAR yields lower Theil U statistics than BVAR; however, SRVAR and BVAR have close performances. The standard deviations of the differences between the Theil U statistics of these two methods are large enough to make the differences statistically not different than zero.

3.5 Conclusion

In this essay we have investigated the forecasting performance of various estimators, but mainly concentrating on Bayesian VAR and Stein-rule VAR forecasting performances. With quarterly data, the best forecasters of the 6-month-commercial-paper rate, the log level of GDP and the log level of GDP deflator are obtained from BVAR and SRVAR. With monthly data, the most accurate forecasts of the 6-month-commercial-paper rate, log level of IP and log level of CPI are obtained from BVAR. However, statistically in most cases BVAR and SRVAR have the same forecast accuracy.

SRVAR's forecast is easier to obtain than the BVAR forecast. In Stein-rule estimation, only the lag length has to be determined before estimation. There is no need for defining a prior distribution as in BVAR. Furthermore, BVAR forecast requires determination of at least two hyperparameters, which is a time consuming process.

The results show that one forecasting method does not always yield the lowest forecast error for all variables at all forecast horizons. In general BVAR and SRVAR forecasts are more accurate than the unrestricted VAR forecasts. SRVAR has the worst performance only with the growth rate model forecasting monthly variables. If we look at the overall forecast performance measures (SOFM and SUM) for quarterly data, then SRVAR is superior to BVAR; however, for monthly data, the BVAR has the most accurate forecasts. The asymmetric VARs (KVAR and HVAR) usually have high forecast errors. Using the Stein-rule method to estimate an asymmetric VAR does not improve the forecast; KSRVAR and KVAR almost always perform in a very similar manner.

The Monte-carlo experiment results, show that forecast performance of SRVAR is superior to BVAR; however, statistically the forecast accuracies of BVAR and SRVAR are not different.

Overall, the results show that SRVAR can be considered as an alternative to BVAR, especially when it is costly to estimate BVAR forecasts which might be the case in large VAR models. To forecast different variables, different methods may be

used. However, for most cases the overall forecast performance of SRVAR is at least as accurate as the other forecasters.

CHAPTER 4

ESTIMATING A VAR MODEL THAT INCLUDES DISCRETE VARIABLES

4.1 Introduction

Vector autoregressive (VAR) analysis is a popular technique used by economists. In general, the number of equations in a VAR model is the same as the number of variables included in the model. The current value of each variable is the dependent variable of one equation and the lagged values of all the variables are explanatory variables. Ordinary least squares (OLS) is used to estimate the coefficients of the VAR model, and, therefore, it is assumed that every variable is continuous. This essay focuses upon the appropriate way to estimate VAR models that contain a discrete variable(s) as a model variable(s). An appropriate technique is developed and is applied to estimate the effects of monetary policy actions. These results are contrasted with results derived from estimating this type of VAR model with OLS, the standard VAR estimation technique. The technique is applied to monetary economics since discrete measures of monetary policy have recently been used to estimate the effects of monetary policy.

Monetary economics is mainly concerned with the effect of monetary policies on the economy. Therefore, in monetary economics, it is crucial to appropriately measure the performance of the economy and the stance of monetary policy (contractionary, neutral, expansionary), so that the relationship between the economic activity and monetary policy can be analyzed. There are abundant and fairly accurate

data available measuring most economic variables. However, it is often debatable how the current position of monetary policy should be measured.

Traditionally, the stance of monetary policy was measured by broad monetary aggregates such as the monetary base, M1 or M2 (Friedman and Schwartz (1963), Andersen and Jordan (1968), Sims (1972), Sargent (1976), King and Plosser (1984), Eichenbaum and Singleton (1986), Stock and Watson (1989)). However, it is generally accepted that monetary aggregates are also influenced by non-policy effects such as aggregate demand and aggregate supply changes. Therefore, it can be argued that the effects of changes in broad monetary aggregates on the economy are not only attributable to monetary policy changes, but also to money demand and aggregate demand and supply changes. Policy analysis in a VAR framework is done by estimating the effects of pure policy shocks on economic variables. Based on the argument that broad monetary aggregates are not a good measure of monetary policy changes, we can argue that shocks to monetary aggregates should not be used as monetary policy shocks. Thus, studies which use these aggregates as proxies for monetary policy measures yield misleading conclusions about the effects of policy changes.

Recently, in a VAR framework, Bernanke and Blinder (1992) and Bernanke and Mihov (1995) advocate using the federal funds rate (FFR) as a measure of monetary policy, whereas Christiano, Eichenbaum and Evans (1994) advocate using non-borrowed reserves (NBR). Bernanke and Blinder (1992) use the Federal Reserve's

(Fed) operating procedure[1] as a guide to select a monetary policy variable. They argue that in the last three decades the Fed generally used the FFR as an operating target (a measure that can be closely controlled); therefore, FFR changes primarily reflect policy changes. Christiano et al. argue that the level of NBR is directly controlled by the Fed and the Fed varies NBR in order to achieve its FFR target; thus it is a better measure of the Fed's policy stance than are broader monetary aggregates.

Bernanke and Blinder (1992) and Christiano et al. (1994) estimate structural VARs. The orthogonalized residuals of the VAR equation with the monetary policy variable as the dependent variable are treated as unexpected shocks to monetary policy. The orthogonalized residuals are used in impulse response function (IRF) and variance decomposition (VDC) analysis. However, orthogonalization does not necessarily separate changes in monetary aggregates due to policy changes from changes due to non-policy changes. Christiano et al.'s and Bernanke and Blinder's approach of identifying monetary policy relies on the assumption of constant operating policy. In general, changes in operating procedure such as occurred in 1979-1982 will alter not only the coefficients in the reaction function of the Fed but also how the Fed responses to contemporaneous information. The reaction function is an equation which explains how the Fed changes the monetary policy variable as a result of a change in the other variables in the system. Thus, treating the reaction function as constant over a sample that includes changes in operating procedure may generate

[1]A procedure designed by the Fed to closely follow certain economic variables and try to attain certain targets.

misleading estimates of monetary policy shocks. These measures of monetary policy are contaminated and the orthogonalized residuals from structural VARs can not be interpreted as unexpected shocks only to monetary policy. Therefore, VAR studies using NBR and FFR may yield misleading conclusions as was the case with the broader monetary aggregates.

An alternative way of determining the stance of monetary policy is the narrative approach pioneered by Romer and Romer (1989). Romer and Romer read all the minutes of the Federal Reserve Board meetings to determine the periods in which there was an intentional decision to conduct contractionary monetary policy specifically to fight inflation. The dates of these intentional contractionary monetary policies are October 1947, September 1955, December 1968, April 1974, August 1978 and October 1979. Then they created an index, marking the contractionary monetary policy periods with 1 and the other periods with 0. Boschen and Mills (1991) generated a similar index also by reading the minutes of the Federal Reserve Board meetings. The Boschen and Mills index (BMI) is scaled from -2 to +2 with unit intervals. In their index, -2 indicates a severe tightening in monetary policy and +2 an expansionary monetary policy; 0 is a neutral policy. These indices concentrate only on the intentions and the statements of the policy makers; thus, they are robust to operating procedure changes. However, a policy index is determined by reading and interpreting the minutes of the Federal Open Market Committee meetings; therefore, these indices contain a subjective element. That is, a statement which is perceived as signalling a contractionary monetary policy action by one reader may be considered as

neutral monetary policy action by another reader. Furthermore, a change in the Fed's action may be a reaction to a change in economic conditions. Thus, policy indices are not completely reflections of exogenous policy changes, but also capture endogenous changes as well.

As can be seen, there is no single monetary variable which is commonly accepted as a measure of exogenous monetary policy changes. Some researchers may prefer to use standard measures like NBR and the FFR as proxies for policy changes, while other researchers may prefer to use the indices.

Despite the limitations of the policy indices, these indices have been used in a number of empirical studies of the effect of monetary policy actions on the economy. Romer and Romer (1989) and Morgan (1993) have employed either the Romer and Romer index or the BMI as right-hand side variables in a regression to estimate the effects of monetary policy on economic activity. The estimation of these equations by OLS poses no conceptual econometric problem. However, several studies including Carlino and DeFina (1994) and Boschen and Mills (1995) have recently estimated VARs which include the BMI as a model variable.

Carlino and DeFina (1994) estimate three VAR models using OLS in order to examine whether monetary policy has symmetric effects across regions in the U.S.. Each estimated VAR model has either BMI, FFR or NBR as the monetary policy variable. The other variables in the VAR model are the regional growth rates of real personal incomes. In their VAR models monetary policy does not contemporaneously

affect regional incomes. Based on IRF and VDC analysis, they conclude that monetary policy actions have similar effects on different regions.

Boschen and Mills (1995) investigate the relationship between narrative-based measures of monetary policy[2] and money market indicators of policy[3] in a bivariate VAR framework. They treat the discrete variables as an ordinary variable and estimate the VARs using OLS. They conclude that different monetary policy indices have similar effects on money market variables.

The consequence of estimating an equation with a discrete dependent variable via ordinary least squares (OLS) is a heteroscedastic error term and inefficient coefficient estimates. The forecasts produced by OLS are not discrete, and, therefore, are not the forecasts of the discrete dependent variable. Furthermore, it is not clear what the forecasts produced by OLS represent. OLS forecasts have continuous values and these forecasts may take values larger (smaller) than the largest (smallest) value of the discrete variable. Thus, the forecasts will have no bound whereas the discrete variable is bounded with its largest and smallest values. Appropriate nonlinear optimization techniques do not generate heteroscedastic errors in equations with a dependent discrete variable, and the predicted values are discrete values as they supposed to be. Therefore, it is more appropriate to estimate an econometric equation with a discrete dependent variable using a non-linear optimization procedure (Madalla (p. 16, 1983)).

---

[2]Boschen and Mills, Romer and Romer, Potts and Luckett and Poole indices.

[3]M2, monetary base, NBR, FFR, T-bill, and the spread between FFR and T-bill rate.

Consequently, an alternative procedure needs to be used to consistently estimate IRFs and VDCs from a VAR model with a discrete variable. That is the subject of this essay. Section II describes the appropriate procedure. Section III applies this new procedure to a macroeconomic VAR model with the Boschen and Mills index as one of the variables. Section IV concludes.

4.2 General Methodology

Often the main purpose of estimating a VAR model is to conduct policy analysis via computation of the IRFs and VDCs. Therefore the statistical significance of estimated coefficient values are of no concern to the researcher. The definitions of IRFs and VDCs are given in Lütkepohl and Reimers (1992). If we go over the computation from an ordinary VAR, the IRF and VDC determination from a VAR with a discrete variable will clearer to the reader. Consider a K-dimensional VAR model written as,

$$y_t = \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + e_t \quad , \tag{4.1}$$

where $y_t = (y_{1t}, \dots, y_{Kt})'$, the β_i are (KxK) coefficient matrices, p is the lag length, and e_t is the error term such that $e_t \sim N(0,\Omega)$ for all t. Note that Ω is not necessarily a diagonal matrix. The vector moving average (VMA) representation of (4.1) can be obtained by repeated substitution of the lagged values of y_t into (4.1). The VMA representation will be

$$y_t = e_t + \Phi_1 e_{t-1} + \Phi_2 e_{t-2} + \dots \quad , \tag{4.2}$$

where

$$\Phi_n = \sum_{j=1}^{n} \Phi_{n-j}\beta_j \quad , \quad n=1,2..., \tag{4.3}$$

in which $\Phi_0=I_K$ and $\beta_j=0$ for $j>p$. The $(i,k)^{th}$ element of Φ_n represents the response of variable y_i to a unit shock in variable k, n periods ago. However, since the error terms $(e_{1t},...,e_{Kt})$ in (4.1) are generally correlated, a shock to the k^{th} variable cannot generally be interpreted as a pure shock solely to the k^{th} variable. A shock to any other variable will contemporaneously affect the k^{th} variable too.

Orthogonalized impulses can be obtained from orthogonal residuals (residuals that are uncorrelated with each other). Consider a system of simultaneous equations:

$$A_0y_t = A_1y_{t-1} + \ldots + A_py_{t-p} + \epsilon_t \quad , \tag{4.4}$$

where A_i are KxK coefficient matrices and ϵ_t is distributed as $N(0,\Omega^D)$. Ω^D is a diagonal matrix so that the disturbances are orthogonal to each other. The model defined in (4.1) is the reduced form of (4.4) where $\beta_i=A_0^{-1}A_i$ and $e_t=A_0^{-1}\epsilon_t$. The relationship between the reduced form shocks and orthogonal shocks is $e=A_0^{-1}\epsilon$, and the VMA representation in (4.2) with orthogonal shocks can be written as:

$$y_t = A_0^{-1}\epsilon_t + \Phi_1A_0^{-1}\epsilon_{t-1} + \Phi_2A_0^{-1}\epsilon_{t-1} + \ldots \quad .$$

The A_0 matrix can be uniquely estimated once the system is identified. To fully identify the system (in order to fulfill the rank and order conditions of identification), we need to impose additional restrictions on the coefficients.[4] The essence of VAR modeling is to impose the fewest number of restrictions possible. Originally Sims

[4]See Fomby et al. (1984) Ch. 20 and Green (1993) Ch 20.

(1980) estimated the VMA coefficients by imposing a lower triangular structure on the contemporaneous coefficient matrix. In this case, the system is fully recursive and is just identified.[5] A shock to the second variable has a contemporaneous effect on all variables except the first one. A shock to the last variable affects contemporaneously only the last variable. However, the theoretical and practical validity of restricting A_0 to be lower triangular is often debated.

Restrictions other than full recursivenes can also be imposed. There are K^2 excess parameters in a set of K simultaneous equations. If we attain the orthogonal errors property which imposes K(K-1)/2 restrictions on the covariance matrix, the number of additional restrictions necessary to identify the system in (4.4) is K(K+1)/2. Possible restrictions on A_0 are normalization of the coefficient of the dependent variable to one, exclusion of variables (which imposes zero restrictions on the contemporaneous effect of these variables), and linear restrictions. The diagonal elements of A_0 are the coefficients of the K dependent variables of the K equations. Thus, with the normalization restrictions, the diagonal elements of A_0 will be set to one. Exclusions impose zero restrictions on coefficients. For instance, if we want to impose that in the i^{th} equation the j^{th} variable does not contemporaneously affect the dependent variable, then the $(i,j)^{th}$ element of A_0 will be set to zero. Linear restrictions are restrictions imposed on linear combinations of the elements of A_0. Another possible way to achieve identification is to impose restrictions on the VMA parameters in (4.2). A typical restriction on the VMA coefficients sets the sum of the

[5]Fomby et al. (1984, p.467) , Green (1993, p.596).

$(i,j)^{th}$ elements of all Φ_n's to zero. This restriction suggests that a shock to the j^{th} element in the system does not have any long-run effect on the i^{th} variable. The estimation of A_0 can be carried out such that the imposed restrictions on the VMA coefficients will be fulfilled.

The restrictions do not have to only just-identify the system; overidentified systems can also be estimated. Bernanke and Mihov (1995) and Doan (1992) demonstrate how to estimate the contemporaneous coefficient matrix of (3.4) if the system is overidentified.[6]

Since in (4.4) the error terms are typically assumed to be orthogonal, a shock to the i^{th} variable will be independent from a shock to the k^{th} variable. In the literature, the orthogonal shocks are interpreted as pure shocks to the corresponding variable. For instance, if one of the variables is the money supply, the orthogonalized residuals of the money supply equation will be interpreted as the structural money supply shock.

In fully recursive systems, the estimation of A_0 and Ω^D is fairly simple. Let P be the Cholesky decomposition of Ω, such that $PP'=\Omega$. Note that the covariance matrix of $P^{-1}e_t$ is the identity matrix. If we multiply both sides of (4.1) with P^{-1} we will obtain a scaled version of (4.4) where Ω^D equals the identity matrix; however, the

---

[6]Consider the relationship between structural error term and reduced form error term, $A_0e_t=\epsilon_t$, where e_t is the error vector in (4.1) and ϵ_t is the error vector in (4.4). Then the A_0 matrix can be estimated by minimizing,

$$-2\log|A_0| + \log|A_0 S A_0'| \quad ,$$

where S is the sample covariance matrix of e.

scaling does not affect the policy analysis. A one standard deviation shock has an orthogonalized impulse response defined as,

$$\Theta_n = \Phi_n P = \Phi_n A_0^{-1} \quad , \tag{4.5}$$

where the $(i,k)^{th}$ element of Θ_n represents the response of variable y_i to a unit shock in variable k, n periods ago.

Once the VMA coefficients are estimated, computing the VDC values is easy. Specifically,

$$w_{ik,h} = \sum_{n=0}^{h-1} \Theta_{ik,n}^2 / MSE_k(h) \qquad h=1,2...., \tag{4.6}$$

where $w_{ik,h}$ is the h-period-ahead-forecast variance of y_i explained by variations in variable k, $\Theta_{ik,n}$ is the ik^{th} element of Θ_n, and $MSE_k(h)$ is the k^{th} diagonal element of

$$MSE(h) = \Omega + \sum_{n=1}^{h-1} \Phi_n \Omega \Phi_n' = PP' + \sum_{n=1}^{h-1} \Phi_n PP' \Phi_n' = \Theta_0 \Theta_0' + \sum_{n=1}^{h-1} \Theta_n \Theta_n' \quad ,$$

the mean-square-error matrix of the optimal h-step-ahead forecast of the y_t process.

We have examined a typical VAR estimation and IRF and VDC computation. Thus, a typical approach to IRF and VDC analysis is to estimate the coefficient of the VAR model via OLS and the contemporaneous coefficient matrix using the residual covariance matrix. If one of the variables in the VAR model is a discrete variable, we can not use the same approach previously described. An equation with discrete dependent variable should not be estimated via OLS. Qualitative dependent variable models are estimated by maximizing a nonlinear function, and with this estimation method, the residuals are not observable. When the residuals are not observed, the

residual covariance matrix can not be computed; thus, we have to use an alternative approach to estimate the contemporaneous coefficient matrix..

To be able to estimate the parameters of a model when the dependent variable is the Boschen and Mills index[7], we will assume that the underlying response model can be described as,

$$y^{*}=X\beta+e,$$

where y^{*} is the underlying response variable and X the set of explanatory variables. Even though y^{*} cannot be directly observed, it is assumed to exist and is only known by the Federal Reserve itself. In the econometric literature, the underlying response variable is called the latent variable. A latent variable is a hidden variable which is assumed to exist but cannot be observed by the researcher. Outsiders can only observe the choice which the Federal Reserve makes. The choice is made according to

$$y=\begin{cases} -2 & \text{if } y^{*}<\mu_1 \\ -1 & \text{if } \mu_1<y^{*}\le\mu_2 \\ 0 & \text{if } \mu_2<y^{*}\le\mu_3 \\ 1 & \text{if } \mu_3<y^{*}\le\mu_4 \\ 2 & \text{if } \mu_4<y^{*} \end{cases}.$$

The threshold values, μ, are also unobservable to outsiders and have to be estimated, together with the coefficients. The alternatives are ordered (-2=strongly contractionary monetary policy, -1=contractionary monetary policy, 0=neutral, 1=expansionary, and 2=strongly expansionary policy). A common assumption in

[7]The index takes the values of -2, -1, 0, 1, 2.

models with an ordered discrete variable as the dependent variable is that the error term is normally distributed. Therefore, the appropriate estimation technique would be ordered probit estimation. Since y^* is not observable, neither is the residual vector e. Predicted choice variables are obtained by estimating the coefficient vector in $y^*=X\beta+e$. Given a sample of the explanatory variables X_0, the forecasted latent variable will be $\hat{y}^*_0=X_0\hat{\beta}$. The forecasted choice variable $\hat{y}_0$ can be obtained by comparing $\hat{y}^*_0$ with the estimated threshold values. See Madalla (1983) for details.

This essay discusses several approaches to modeling a VAR when one of the variables is discrete and describes methods to estimate the IRF and VDC values. Assume that the i^{th} variable of the VAR, i.e. y_i, is a discrete variable. The VAR model can be outlined as,

$$\begin{aligned} y_{1t} &= X\beta^1 + e_{1t} \\ &\vdots \\ y_{it}^* &= X\beta^i + e_{it} \\ &\vdots \\ y_{Kt} &= X\beta^K + e_{Kt} \quad , \end{aligned} \tag{4.7}$$

where $X=(y_{t-1},\ldots,y_{t-p})$ and $y_t=(y_{1t},\ldots,y_{it},\ldots,y_{Kt})'$.[8] Note that in this model the i^{th} dependent variable is the unobservable latent variable, y_{it}^*, and the i^{th} variable in y_t is the corresponding observable discrete variable y_{it}. This particular modeling approach assumes that the latent variable, y^*, is unobservable by economic agents, but only the choice made is observable. Therefore, the lagged values of the choice, not the lagged

---

[8]If we represent the model in (4.7) as $y^*_t=X\beta+e_t$, where y^*_t is a $1\times K$ vector and β is a $Kp\times K$ coefficient matrix, then β^k is the coefficient vector of the k^{th} equation which is the k^{th} column vector of β.

values of the latent variable can influence the economic variables. However, since the value of the latent variable is observed by the policy maker (Fed), the lagged values of y^* can affect the response. But the lagged values of the latent variable are also not included on the right-hand side of the i^{th} equation for one reason: we wanted to use the same explanatory variables in all equations

The i^{th} equation of (4.7) can be estimated using optimization procedures, and the rest of the equations can be estimated using OLS. However, we cannot use the repeated substitution method to find the VMA coefficients of the model presented in (4.7). To show why, rewrite (4.7) as,

$$y_t^* = \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + e_t \quad , \tag{4.8}$$

where $y_t^* = (y_{1t}, \ldots, y_{it}^*, \ldots, y_{Kt})'$ and $y_t = (y_{1t}, \ldots, y_{it}, \ldots, y_{Kt})'$. y_t^* and y_t are $1 \times K$ vectors.[9] We can represent y^*_{t-1} as $\beta_1 y_{t-2} + \ldots + \beta_p y_{t-p} + e_t$. However, this model does not generate y_{t-1} as a function of its lagged values. In this case, we can not repeatedly substitute the lagged values to estimate the VMA coefficients. An alternative procedure to obtain the VMA coefficients is to simulate the model.

One standard deviation IRFs can be obtained by simulating a model defined in (4.1) and (4.7). Thus, we have to first estimate the system. The equation with the latent dependent variable (the i^{th} equation in (4.7)) is estimated via non-linear methods, and the other equations are estimated via OLS. The presence of a

[9]The general form of the model is $y_t^* = X\beta + e_t$, where X is defined as before. In this case β_j is a $K \times K$ matrix which collects all the coefficients of the j^{th} lagged values in all K equations. Since the model includes p lags, there are p of β_j coefficient matrices.

deterministic component in (4.7) does not change the result. The IRF is the response relative to the deterministic part. Therefore, for the sake of simplicity, we will ignore the intercept term.

The simulation estimates the VMA coefficients of (4.1) which were defined as Φ_i in (4.2). To implement the simulation, first set $y_{t-1}=\ldots=y_{t-p}=0$ in (4.8). Set $e_{1t}=1$, and all other elements of e_t to zero. Thus, $\hat{y}_t^*=e_t$. To find the value of y at t+1, place $\hat{y}_t$ into model (4.8), $\hat{y}^*_{t+1}=\beta_1\hat{y}_t$. $\hat{y}^*_{it+1}$ will be the estimated latent variable; however, we need the discrete variable, $\hat{y}_{it+1}$ to substitute into the right hand side to simulate the model. To do this, we look at the estimated threshold value, and determine the appropriate choices. The forecasted latent value is converted to the forecasted choice based on the following relationship,

$$y=\begin{cases} -2 & \text{if } \hat{y}^*_{it+1}<\hat{\mu}_1 \\ -1 & \text{if } \hat{\mu}_1<\hat{y}^*_{it+1}\le\hat{\mu}_2 \\ 0 & \text{if } \hat{\mu}_2<\hat{y}^*_{it+1}\le\hat{\mu}_3 \\ 1 & \text{if } \hat{\mu}_3<\hat{y}^*_{it+1}\le\hat{\mu}_4 \\ 2 & \text{if } \hat{\mu}_4<\hat{y}^*_{it+1} \end{cases}.$$

Thus, on the right hand side of the equations, we will always have the discrete variable, not the forecasted latent variable. We estimate $\hat{y}^*_{t+2}$ as $\beta_1\hat{y}_{t+1}+\beta_2\hat{y}_t$. $\hat{y}^*_{t+1}=(\hat{y}_{1t+1},\ldots,\hat{y}^*_{it+1},\ldots,\hat{y}_{Kt+1})'$ is the first column of Φ_1 in (4.2), the orthogonal VMA representation; $\hat{y}^*_{t+2}$ is the first column of Φ_2, and so on. To find the second columns of the orthogonal VMA coefficient matrices, initially set $y_{t-1}=\ldots=y_{t-p}=0$, $e_{2t}=1$ and

all other elements of e_t to zero, and repeat the simulation described above. Note that the IRF described above traces the response of the discrete variable.

The relationship between the reduced form shocks and orthogonal shocks is $e=A_0^{-1}\epsilon$, and the VMA representation in (4.2) with orthogonal shocks can be written as:

$$y_t = A_0^{-1}\epsilon_t + \Phi_1 A_0^{-1}\epsilon_{t-1} + \Phi_2 A_0^{-1}\epsilon_{t-1} + \dots \quad .$$

One standard deviation shocks can be obtained by writing ϵ as Cv, where $v_t \sim N(0,I_K)$ and C is a constant such that $CC' = \Omega^D$. Since Ω^D is a diagonal matrix, C is also a diagonal matrix. With this further decomposition the VMA representation becomes:

$$\begin{aligned} y_t &= A_0^{-1}CC^{-1}\epsilon_t + \Phi_1 A_0^{-1}CC^{-1}\epsilon_{t-1} + \Phi_2 A_0^{-1}CC^{-1}\epsilon_{t-2} + \dots \\ &= \Theta_0 v_t + \Theta_1 v_{t-1} + \Theta_2 v_{t-2} + \dots \end{aligned}$$

where $\Theta_n = \Phi_n A_0^{-1}C$ and $v_n = C^{-1}\epsilon_n$ Thus, if we can estimate A_0 and Ω^D, we can obtain the orthogonal IRF as in (4.5) where $P = A_0^{-1}C$.

A fully recursive system of simultaneous equations (which imposes K normalization restrictions and K(K-1)/2 zero restrictions on the contemporaneous coefficient matrix) looks like:

$$\begin{aligned} y_{1t} &= XA^1 + \epsilon_{1t} \\ y_{2t} &= XA^2 - a_{21,0}y_{1t} + \epsilon_{2t} \\ &\vdots \\ y_{it}^* &= XA^i - a_{i1,0}y_{1t} - \dots - a_{i(i-1),0}y_{i-1t} + \epsilon_{it} \\ &\vdots \\ y_{Kt} &= XA^K - a_{K1,0}y_{1t} - \dots - a_{KK-1,0}y_{K-1t} + \epsilon_{Kt} \quad , \end{aligned} \tag{4.9}$$

where A^j is the coefficient matrix of the predetermined variables of the j^{th} equation,

and $a_{jk,0}$ is the contemporaneous effect coefficient of the k^{th} variable in the j^{th} equation. A general system of simultaneous equation can be written as,

$$
\begin{aligned}
y_{1t} &= XA^{1} - a_{12,0}y_{2t} - \cdots - a_{1K,0}y_{Kt} + \epsilon_{1t} \\
y_{2t} &= XA^{2} - a_{21,0}y_{1t} - \cdots - a_{2K,0}y_{Kt} + \epsilon_{2t} \\
&\vdots \\
y_{it}^{*} &= XA^{i} - a_{i1,0}y_{1t} - \cdots - a_{iK,0}y_{Kt} + \epsilon_{it} \\
&\vdots \\
y_{Kt} &= XA^{K} - a_{K1,0}y_{1t} - \cdots - a_{KK-1,0}y_{K-1t} + \epsilon_{Kt} \quad .
\end{aligned}
$$

Systems other than fully recursive can be also estimated by imposing appropriate K(K-1)/2 restrictions on this general system (we are already imposing normalization restrictions). Note that y_{it} is an endogenous discrete variable which may appear on the right hand side of the equations other than the i^{th} equation. Since y_{it} is a discrete variable and is endogenous, a transformed version of this variable has to be used to consistently estimate its coefficient. Heckman (1978) suggests replacing the endogenous discrete variable, y_{it}, with $F(X\hat{\beta}^{i})$ where F(.) is the standard normal distribution function and $\hat{\beta}^{i}$ is the estimated coefficient vector of the i^{th} equation of the reduced form model presented in (4.7)(see also Amemiya (1978)). Heckman originally describes his procedure for a probit model (a model with two alternatives only); we adapted his procedure by estimating an ordered probit model with five alternatives. Even though y_{it} is replaced with $F(X\hat{\beta}^{i})$, the estimated coefficient is the coefficient of y_{it}. The equations with continuous dependent variables are estimated using the OLS method. The i^{th} equation is estimated via nonlinear optimization.

The VDC is the proportion of forecast variance explained by the structural shocks of each variable. Once the orthogonal VMA coefficients Θ_n are estimated, we

can estimate VDC values by substituting the values in (4.6). As shown before $MSE(h)=\Theta_0\Theta_0'+\Theta_1\Theta_1'+\ldots+\Theta_{h-1}\Theta_{h-1}'$.

Often the standard deviations of IRFs and VDCs are also reported. Typically these standard deviations are obtained from a Monte Carlo simulation. The simulation generates observations from the model defined in (4.1). The essence of the simulation is to generate K random series from a normal distribution with zero mean and covariance matrix Ω. If ϵ_t in (4.4) is distributed as $N(0,\Omega^D)$, then $\Omega=A_0^{-1}\Omega^D A_0^{-1\prime}$ which is the covariance matrix of the error term in (4.8). Thus, an estimate of Ω can be obtained from estimating A_0 and Ω^D in (4.4). The generated series are the simulated error terms in (4.7), $e_1,\ldots e_K$. Observations from (4.7) are generated by initially setting $y_0=\ldots=y_{-p}=0$. The first observation of the first variable will be $y_{11}=e_{11}$. The first observation of the second variable equals to e_{21}, and this iterative process generates the first observations of the K variables. The t^{th} observation of the first variable is generated as

$$y_{1t}=\beta_1^1 y_{t-1}+\beta_2^1 y_{t-2}+\ldots+\beta_p^1 y_{t-p}+e_{1t}.$$

The t^{th} observation of the second variable is generated as,

$$y_{2t}=\beta_1^2 y_{t-1}+\beta_2^2 y_{t-2}+\ldots+\beta_p^2 y_{t-p}+e_{2t}.$$

The generated i^{th} variable will be transformed to its equivalent discrete choice representation. The coefficient and threshold values used in the Monte Carlo simulation are the estimated values of the system defined in (4.7).

The sample size of the generated data will be the same as the sample size of the original VAR estimation, say T. Since initially the random number generator may not

generate a sequence of random numbers, usually the first 30 to 60 generated observations are discarded. Taking this fact into account, the simulated sample size will be larger that T. However, we will discard a part of the generated sample, so that we will end up with a sample size of T. This whole process generates a sample of data obtained from the same population as the sample for which we estimated the VAR. In this way we can generate many samples from the same population from which we estimated the VAR. The generated samples are used to compute new IRFs and VDCs. Since the data are generated from random samples, each estimation will yield different coefficients and thus different IRFs and VDCs. The standard deviations of the IRFs and VDCs are the standard deviations of these samples of IRFs and VDCs obtained from the Monte Carlo simulation. The IRF variances are defined as,

$$\frac{1}{N}\sum_{i=1}^{N}(IRF_{jk}^{i}-\bar{IRF}_{jk})^2 \quad ,$$

and

$$\bar{IFR}_{jk}=\frac{1}{N}\sum_{i=1}^{N}IRF_{jk}^{i} \quad ,$$

where N is the number of repetitions in the Monte Carlo experiment and IRF^{i}_{jk} is the impulse response of the k^{th} variable when j^{th} variable is shocked at the i^{th} repetition. The variances of the VDCs are defined similarly. See Chapter One of this dissertation for a detailed explanation of Monte Carlo simulation.

4.3 Empirical Applications

In this section we demonstrate the difference in IRF and VDC results when the monetary policy index is treated as an ordinary variable as done by Boschen and Mills (1995) and when it is treated as a discrete variable. Boschen and Mills (1995) use various monetary policy indices to study the relationship between narrative-based measures of monetary policy and money market indicators. They estimate several bivariate VARs, each consisting of one policy index and one money market variable. The money market variables are the first-difference of the log M2 (ΔM2), first-difference of the St. Louis Federal Reserve Bank monetary base (ΔBASE), first-difference of the log of nonborrowed reserves (ΔNBR), the federal funds rate (FFR), three-month treasury bill rate (TBILL), and the six-month commercial paper rate less three-month treasury bill rate (SPREAD). The policy indices they consider are Boschen and Mills, Romer and Romer (1992), Poole (1971), Uselton (1974), Potts and Luckett (1978), and Kimelman (1981). Each variable has eighteen lags in the VAR. Boschen and Mills ordered the index variables first in the VAR and use a Cholesky decomposition, assuming that the money market variables do not contemporaneously affect monetary policy, but monetary policy contemporaneously affects the money market variables. With this ordering they assume that, due to information delay, the Fed does not respond to changes in money market variables within a month.

4.3.1 Empirical Results

This section describes the results obtained by applying the method described in the methodology part. We adopt the same framework used by Boschen and Mills (1995); however, we concentrate only on the relationship between money market variables and BMI. All variables are monthly observations, starting from 1953:1 and ending 1994:12. The FFR series are available staring from 1955:1. The BMI series is from Boschen and Mills (1995).[10] The nonborrowed reserves variable is constructed as total reserves adjusted for reserve requirement changes minus borrowed reserves; both series are obtained from Federal Reserve Bank of St. Louis Data Bank. The resulting series is seasonally adjusted using SAS's X11 procedure. The rest of the series are obtained from CITIBASE.[11]

The standard errors of the IRF and VDC estimates are obtained from a Monte-Carlo simulation. First, two samples of random error terms are generated, each from a normal distribution with zero mean and a variance equal to one. We put these two series together to obtain a $N \times 2$ matrix. N is the sample size of the generated random errors and will be defined more precisely later. Because the numbers generated in the early stage of the process may not be truly random, the first 30 observations of this

---

[10]An updated series was obtained from Dr. Boschen.

[11]The M2 series provided by CITIBASE starts from January, 1959. The M2 series from 1953:1 to 1958:12 was obtained in the same way as described in Boschen and Mills (1995), footnote number 9. The pre 1959 M2 data is obtained as the sum of currency, demand deposits, time deposits (reported in Banking and Monetary Statistics 1941-1970), mutual saving bank deposits and savings and loan shares data (reported in Friedman and Schwartz (1970)).

sample are discarded. If we define C as the Cholesky decomposition of $\hat{\Omega}=\hat{A}_0^{-1}\hat{\Omega}^D\hat{A}_0^{-1\prime}$, the created error terms are multiplied by C to convert them to $N(0,\hat{\Omega})$. After obtaining the error term with desired properties, the data from the VAR model are generated as explained in the previous section. To eliminate the effect of setting the values of the variables initially to zero in the data generating process, an additional 30 observations are discarded. Thus, we end up with N-60 observations of the generated data, which is the same sample size of the actual data we have used to estimate the point estimates. The Monte-Carlo experiment has 500 replications.

The IRFs of the monetary variables to a one standard deviation shock to BMI are now presented. An increase in the index is an expansionary policy, so the responses are results from an expansionary monetary policy. The IRFs obtained from the probit model with the associated one standard deviation bands are presented in Figures 4.1 to 4.6. The impulse responses of the monetary aggregates (M2, BASE, NBR) are displayed in terms of cumulative effects.[12] The cumulative responses of the growth rates are obtained to convert the responses to their levels form. Sine the interest rates and the spread are already estimated in the levels, their IRFs are not accumulated. After an expansionary monetary shock, M2 and BASE increase for about 7 periods and then start decreasing. NBR increases for about 5 periods and

---

[12]The cumulative effect at the i^{th} impulse period is calculated as,

$$\sum_{j=1}^{i} IRF_j \qquad i=1\ldots36,$$

where IRF_j is the impulse response of the growth rate of monetary aggregate at the j^{th} impulse period.

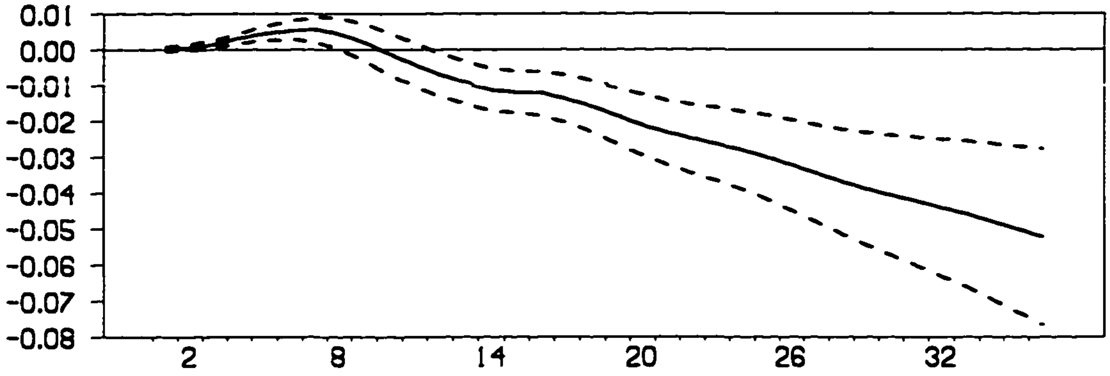

Figure 4.1
Impulse Response of M2 From the Probit Model

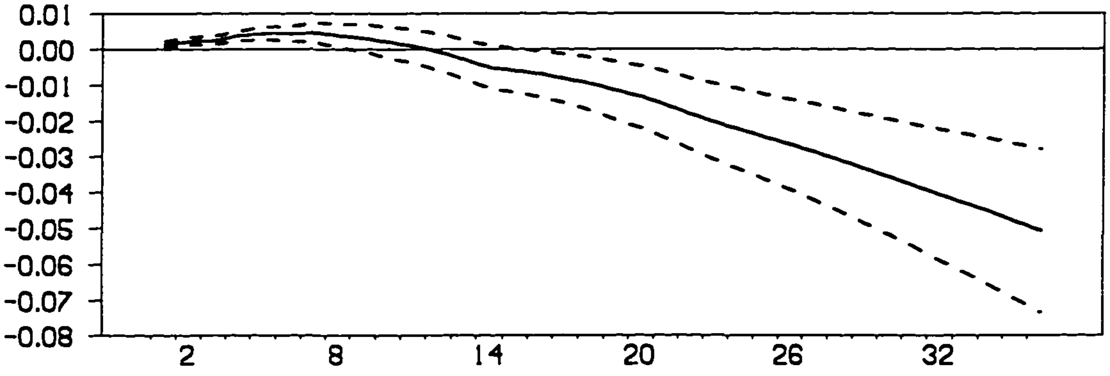

Figure 4.2
Impulse Response of Monetary Base From the Probit Model

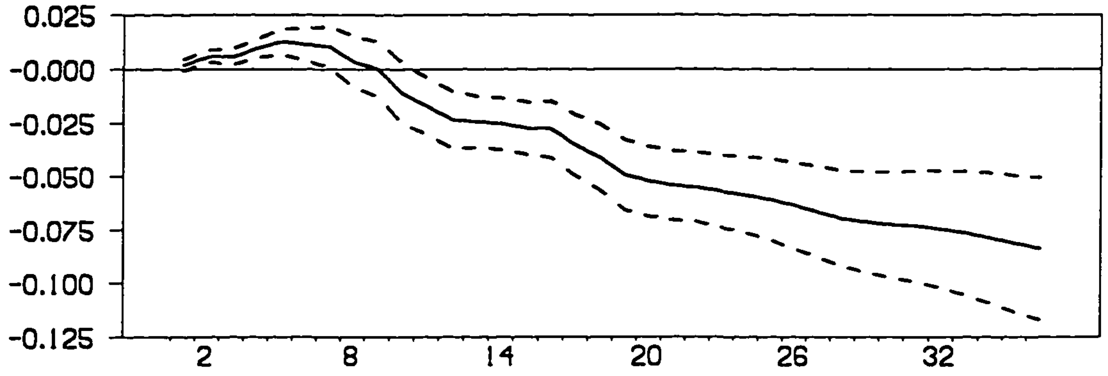

Figure 4.3
Impulse Response of NBR From the Probit Model

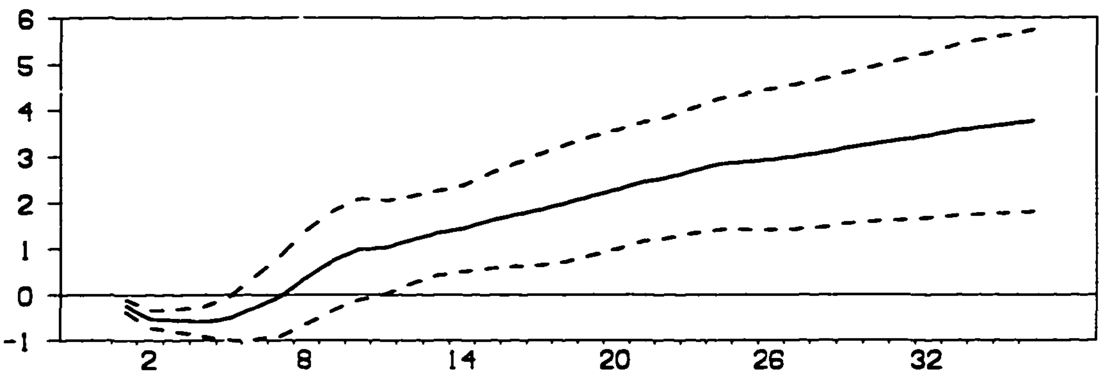

Figure 4.4
Impulse Response of FFR From the Probit Model

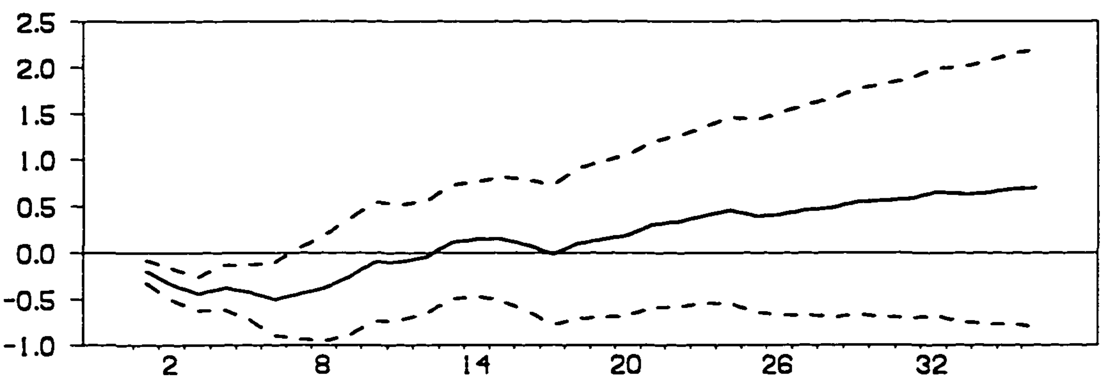

Figure 4.5
Impulse Response of 3 Month T-Bill Rate From the Probit Model

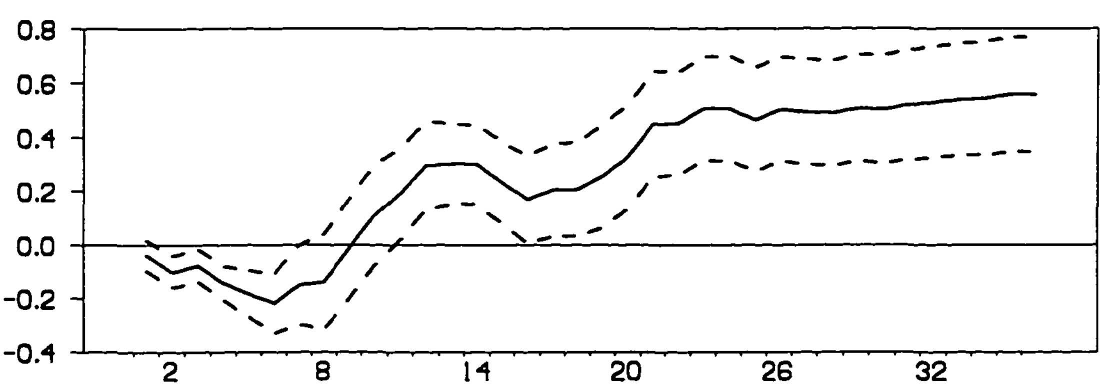

Figure 4.6
Impulse Response of the Spread From the Probit Model

then starts decreasing. The interval estimates fall below zero, indicating a long-run decrease. These results are unexpected. An expansionary policy shock should have resulted in an increase in the monetary aggregates and a return to the initial values over longer horizons. Interest rates display the opposite reaction. FFR and TBILL initially decrease for about 5 periods (which can be interpreted as evidence of a liquidity effect), and then start increasing. SPREAD decreases for about 6 periods and starts increasing. SPREAD starts increasing at a slower pace around the 21st period. The interval estimates for FFR and SPREAD rise above zero at the longer horizons while the interval estimate for TBILL is centered around zero at longer horizons.

The same IRFs are shown in Figures 4.7 - 4.12 for the case when BMI is treated as an ordinary continuous variable. The only difference in computation is that the discrete nature of the index is not taken into account. Therefore, the equation with the BMI as the dependent variables is also estimated via OLS. Again, the IRFs are responses to an expansionary monetary policy. In this case there is an increase in M2 and BASE. NBR increases until the 13th period and then starts decreasing. The interval estimate includes zero over longer horizons, as one would expect from a transitory shock to monetary policy. Compared with the results obtained from the probit model, FFR and TBILL display larger drops and the interval estimates return to their original values. SPREAD has a smaller drop compared to the probit model response, and increases back to its original value after 32 periods.

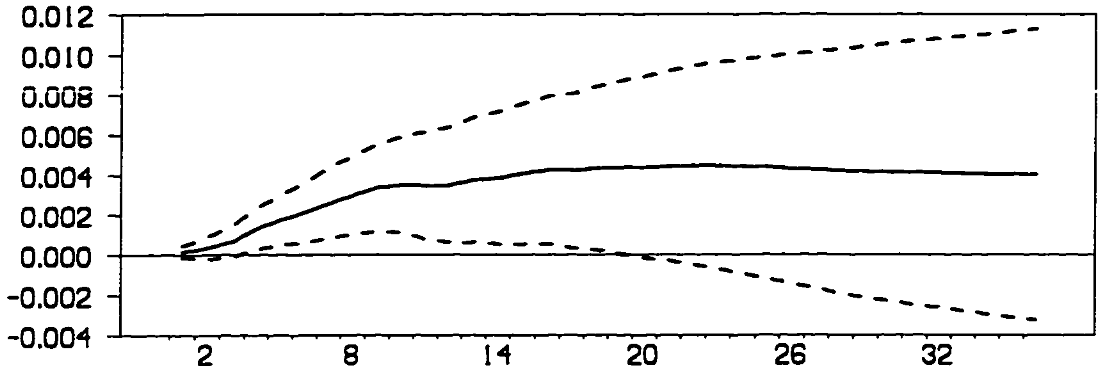

Figure 4.7
Impulse Response of M2 From the Regular Model

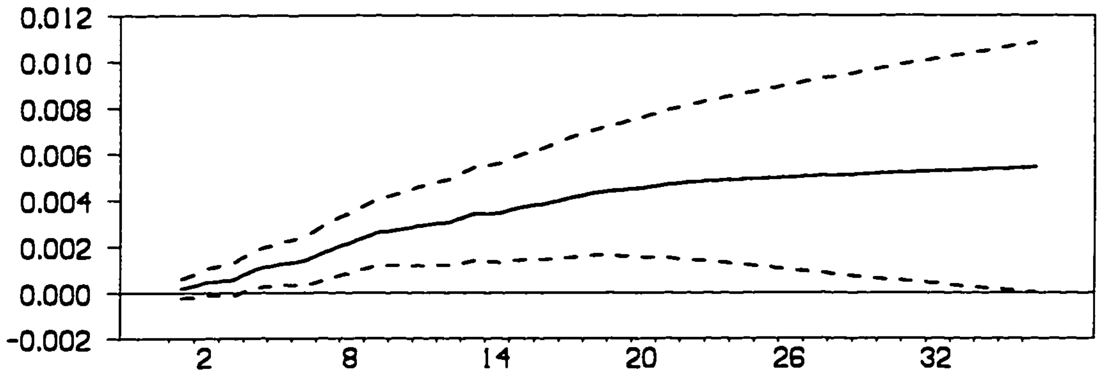

Figure 4.8
Impulse Response of Monetary Base From the Regular Model

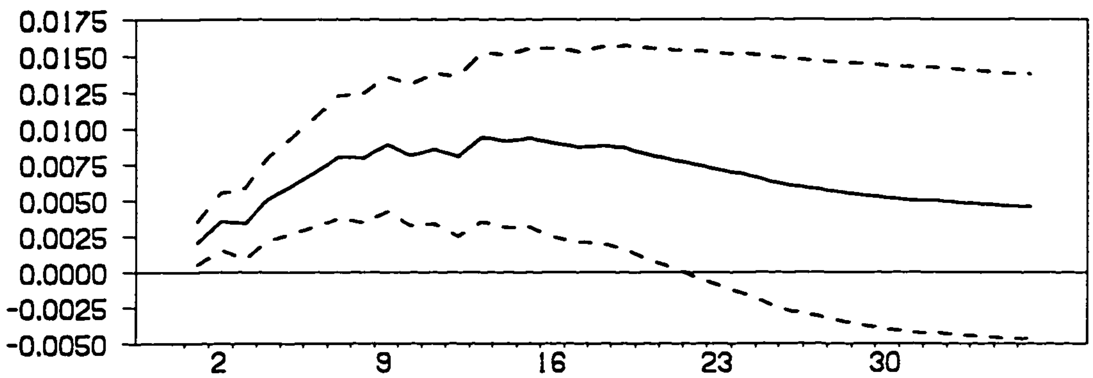

Figure 4.9
Impulse Response of NBR From the Regular Model

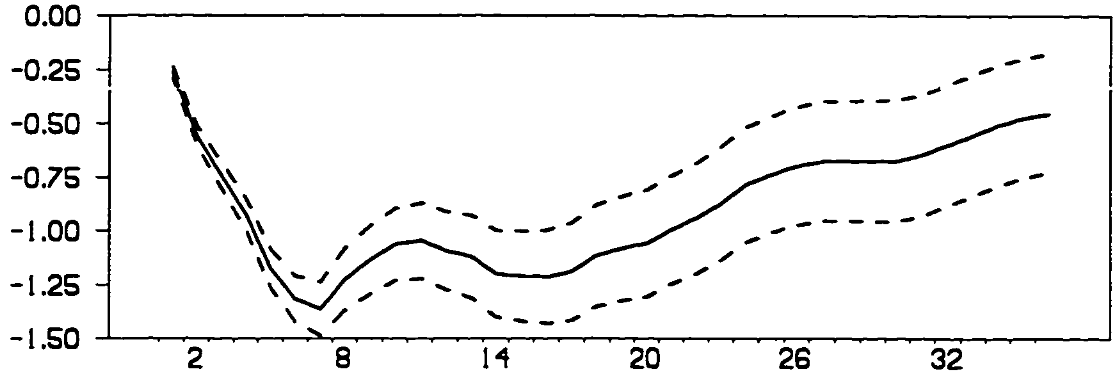

Figure 4.10
Impulse Response of FFR From the Regular Model

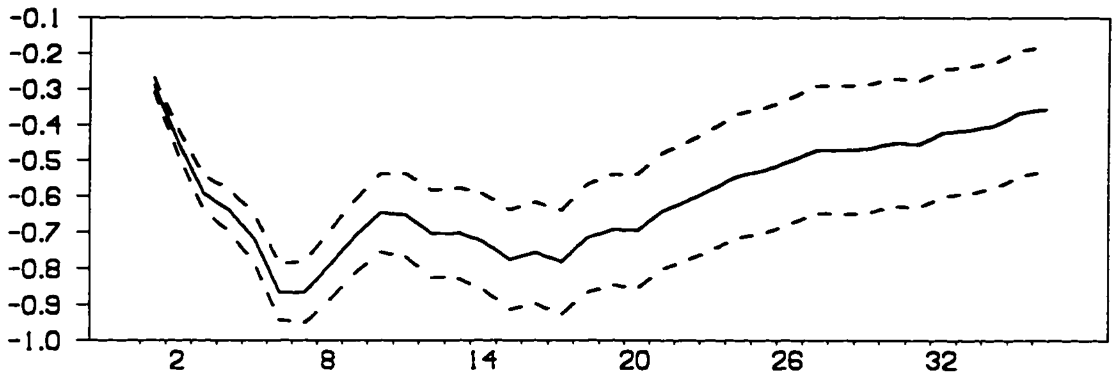

Figure 4.11
Impulse Response of 3 Month T-Bill Rate From the Regular Model

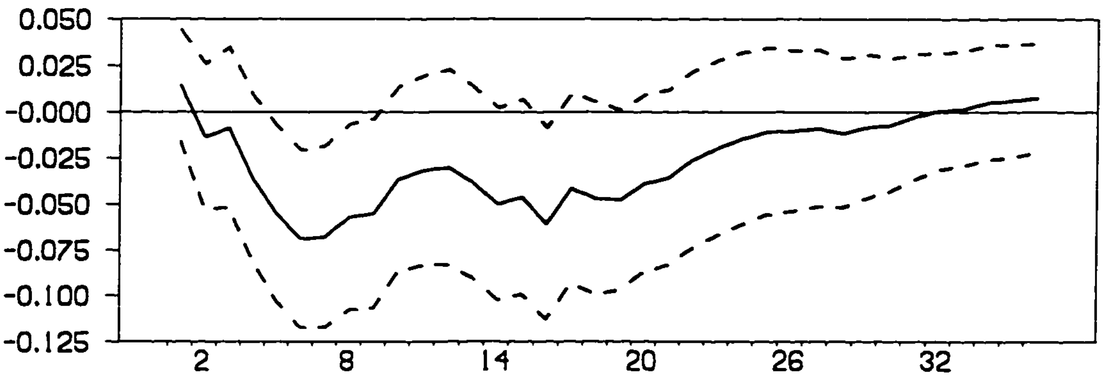

Figure 4.12
Impulse Response of the Spread From the Regular Model

Thus, the main difference resulting from the two estimation approaches is that the money aggregates move in the expected manner following an expansionary monetary policy, using the ordinary technique, while using the probit technique produces a negative long-run response of the monetary variable after an expansionary monetary policy. The largest increase of M2 produced by the probit estimation is about 0.007 where this value is about 0.005 for the OLS estimator. For the monetary base, these numbers, are 0.005 and 0.006, respectively; for NBR, they are 0.013 and 0.01, respectively. Even though the highest values of the IRFs produced by the two different approaches are similar, the IRFs of the probit approach reach their highest values in a much shorter period (about 6-9 periods) than the IRFs of the OLS approach (about 13-36 periods). After reaching the largest values, the point estimates of the monetary aggregates start decreasing in the probit case while they stay at their new level in the OLS case.

The impulse response of the FFR obtained from the probit estimation goes down by about 0.5 percent while the FFR response of the OLS approach goes down by about 1.3 percent. These numbers for the TBILL are 0.5% for the probit case and 0.85% for the OLS case. For SPREAD they are about 0.2% and 0.07%, respectively. After reaching the lowest values, in both approaches the interest rates and the spread increase; however, in the probit approach the responses quickly reach zero (about in 7-9 periods) while in the OLS approach it takes a much longer time for the interest rates and the spread to reach or get close to zero.

Table 4.1
Variance Decomposition Values From the Probit Model

	Percentage of Variance due to BMI					
Months	ΔM2	ΔBASE	ΔNBR	FFR	TBILL	SPREAD
3	12.6 (0.08)	27.1 (0.08)	14.2 (0.11)	30.8 (0.11)	27.6 (0.10)	11.9 (0.09)
12	81.3 (0.11)	54.8 (0.13)	68.8 (0.11)	23.4 (0.17)	15.8 (0.18)	52.5 (0.14)
36	91.6 (0.08)	91.0 (0.09)	79.4 (0.06)	51.7 (0.18)	9.5 (0.21)	88.8 (0.08)

The standard deviations are given in the parentheses.

The difference in these two approaches is the way the first equation which has BMI as the dependent variable is estimated. We obtain different IRFs because the coefficient estimates of the first equation using the different estimation techniques are completely different from each other. The coefficient estimates of the second equations are exactly the same since in both cases we use OLS to estimate those coefficients.

The VDC values of the probit model for selected horizons are presented in Table 4.1. The numbers in parentheses are the standard deviations. Almost all 36 period-ahead forecast errors of ΔM2, ΔBASE, ΔNBR and SPREAD can be attributed to BMI shocks. BMI shocks explain more of the forecast errors of these variables as the forecast horizon increases. Overall, at all three forecast horizons most of the forecast errors of FFR and TBILL are due to their own shocks. BMI similarly explains the forecast variance of SPREAD and the money aggregates. A larger portion of SPREAD is explained by BMI than FFR and TBILL. The reason for this

Table 4.2
Variance Decomposition Values From the Regular VAR Model

	Percentage of Variance due to BMI					
Months	ΔM2	ΔBASE	ΔNBR	FFR	TBILL	SPREAD
3	2.4 (0.09)	1.4 (0.07)	4.8 (0.08)	43.6 (0.04)	44.0 (0.05)	0.4 (0.04)
12	13.5 (0.10)	8.7 (0.08)	9.4 (0.07)	79.9 (0.06)	72.2 (0.07)	12.9 (0.12)
36	12.8 (0.09)	10.8 (0.08)	11.3 (0.06)	82.5 (0.10)	72.3 (0.12)	21.1 (0.10)

The standard deviations are given in the parentheses.

might be that the interest rate spread is more directly related to changes in the stance of the monetary policy.

Table 4.2 presents the VDC values from the model which treats the monetary index as an ordinary continuous variable. In this case BMI shocks explain a large portion of FFR and TBILL forecast errors and a small portion of ΔM2, ΔBASE, ΔNBR and SPREAD forecast errors. For all variables the portion explained by BMI increases as the forecast horizon increases. Thus, in terms VDC results, the two estimation techniques yield different results. Overall, the probit estimation finds a larger effect of monetary policy on monetary variables than the ordinary Cholesky approach.

4.3.2 Two Alternative Modeling Approaches

There are several other approaches that can be considered in modeling the system considered here. In this section we discuss two alternative modeling approaches. We do not report the results of the IRF and VDC analysis, since they do

not yield plausible results. Therefore in this section we merely describe the models and the estimation procedures.

The VAR model in the previous section assumes that each variable is explained by the lagged values of the discrete policy index and the monetary variable. We have two equations and the structural system is defined as,

$$
\begin{aligned}
y_{1t}^{*} &= A_{11}^{1} y_{1t-1} + \ldots + A_{1p}^{1} y_{1t-p} + A_{21}^{1} y_{2t-1} + \ldots + \epsilon_{1t} \\
y_{2t} &= A_{11}^{2} y_{1t-1} + \ldots + A_{1p}^{2} y_{1t-p} + A_{21}^{2} y_{2t-1} + \ldots - a_{21,0} y_{1t} + \epsilon_{2t} \quad ,
\end{aligned}
$$

where p is the lag length which is 18, y^{*}_{1} is the underlying value of the policy action and y_1 is the corresponding discrete choice, and y_2 is the second variable in the system. A^{m}_{kj} is the coefficient of the k^{th} variable's j^{th} lag in the m^{th} equation. $a_{21,0}$ is an element of the contemporaneous coefficient matrix A_0. An alternative modeling approach would be to put the latent variable on the right hand side rather than the discrete variable. As before, we assume that monetary policy does not respond contemporaneously to the other variable. In the first alternative approach , it is the latent variable that directly affects the monetary system,

$$
\begin{aligned}
y_{1t}^{*} &= A_{11}^{1} y_{1t-1}^{*} + \ldots + A_{1p}^{1} y_{1t-p}^{*} + A_{21}^{1} y_{2t-1} + \ldots + \epsilon_{1t} \\
y_{2t} &= A_{11}^{2} y_{1t-1}^{*} + \ldots + A_{1p}^{2} y_{1t-p}^{*} + A_{21}^{2} y_{2t-1} + \ldots - a_{21,0} y_{1t} + \epsilon_{2t} \quad .
\end{aligned}
\tag{4.10}
$$

The policy index y_i is a proxy for the latent variable. Thus, by including the lagged latent variables on the right hand side, we are assuming that we are already capturing the effect of monetary policy on the other variables.

The problem with this modeling strategy is that the latent variable is not observed and therefore the corresponding coefficients cannot directly estimated. A

possible solution to this problem is to estimate an ordered probit equation using a set of observed explanatory variables, and then obtain the forecasted value of the latent variable and replace the forecasted values in (4.10). A logical choice of explanatory variables are the lagged values of the continuous variable in the VAR model, y_2, which is the second variable in (4.10). Basically we can estimate the following equation using nonlinear maximization methods (ordered probit):

$$y_{1t}^{*} = \Pi_{11} y_{2t-1} + \ldots + \Pi_{1p} y_{2t-p} + v_t \tag{4.11}$$

There are p coefficients in this equation. After estimating the coefficients we can obtain the forecasted values of the latent variable y^{*}_{1t}, and replace the lagged latent variables in (4.10) with the lagged forecasted latent variables $\hat{y}^{*}_{1t-j}$.

Originally, in (4.10) the lagged values of the latent variable are also on the right hand side of the first equation. There is an inconsistency between (4.10) and (4.11), because in (4.11) the first variable which is the monetary policy variable depends on the lagged observed variables, whereas in (4.10) the first variable depends on the explanatory variables that are in (4.11) and also on its own lagged values. The dependent variables in these two equations are the same; however, they are explained by different variables, which causes an economic inconsistency. To overcome this inconsistency, we can assume that the first equation in (4.10) only depends on the lagged values of the observed continuous variable as in (4.11). This alteration violates one of the modeling assumption of VAR, which states that the right hand side variables in each equation are the same. We prefer to solve the inconsistency problem and model the system a little bit differently than the VAR approach. Thus,

the structural VAR model that we estimate is

$$\begin{aligned} y^*_{1t} &= A^1_{21} y_{2t-1} + A^1_{22} y_{2t-2} + \ldots + \epsilon_{1t} \\ y_{2t} &= A_{11} \hat{y}^*_{1t-1} + A^2_{12} \hat{y}^*_{1t-2} + \ldots + A^2_{21} y_{2t-1} + \ldots + A^2_{2p} y_{2t-p} - a^2_{10} \hat{y}^*_{1t} + \epsilon_{2t} \quad , \end{aligned} \tag{4.12}$$

where the values in hats are the forecasted values obtained from the first equation of (4.12).

We still have a problem with (4.12). (4.12) is not estimable, because the current forecasted latent variable, $\hat{y}^*_{1t}$, is perfectly correlated with the lagged values of the monetary variable y_2. When we estimate the first equation we specify that the latent variable depends on 18 lags of the monetary variable. In this case the forecasted value $\hat{y}^*_{1t}$ is a function of the 18 lags of y_{2t} and therefore, when we put $\hat{y}^*_{1t}$ and the explanatory variables of y^*_{1t} into the same equation's right hand side we face the problem of perfect multicollinearity. The multicollinearity can be eliminated if we include other variables into the first equation or include more lags. We did not want to include more variables in the system, therefore we are left with only one alternative, including more lags in the first equation. In this system of simultaneous equations the first variable depends on p^* lags of the second variable, and the second variable depends on p lags of each variable and the current value of the first variable ($p^*>p$). In this application $p^*=24$ and $p=18$.

The VMA coefficients of the VAR can be estimated from the reduced form coefficients estimates. The reduced form of the structural model can be written as

$$\begin{bmatrix} y^*_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0 & \beta^1_{21} \\ \beta^2_{11} & \beta^2_{21} \end{bmatrix} \begin{bmatrix} y^*_{1t-1} \\ y_{2t-1} \end{bmatrix} + \ldots + \begin{bmatrix} 0 & \beta^1_{2p} \\ \beta_{1p2} & \beta^2_{2p} \end{bmatrix} \begin{bmatrix} y^*_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} 0 & \beta^1_{2p+1} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y^*_{1t-p-1} \\ y_{2t-p-1} \end{bmatrix} + \ldots + \begin{bmatrix} 0 & \beta^1_{2p^*} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y^*_{1t-p^*} \\ y_{2t-p^*} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad .$$

The maximum lag of the VAR in this case is 24 which is the lag length in the first equation. The VMA coefficients can be estimated either by the repeated substitution method or by the simulation method which are described in section (4.2). The orthogonal IRF can be obtained by multiplying the VMA coefficient matrices Φ_i by the inverse of the contemporaneous coefficient matrix which is estimated in (4.12).

In the second alternative model, we want to include a monetary policy variable on the right hand side of the reaction function (the first equation). Therefore, we model the reaction function as a function of the lagged values of the discrete variable and the monetary variable. The monetary variable is affected by the lagged latent variable and lagged values of itself, and the contemporaneous value of the monetary policy variable y^{*}_{it}. Basically,

$$\begin{aligned} y_{1t}^{*} &= A_{11}^{1} y_{1t-1} + A_{12}^{1} y_{1t-2} + \ldots + A_{21}^{1} y_{2t-1} + \ldots + A_{2p}^{1} y_{2t-p} + \ldots + \epsilon_{1t} \\ y_{2t} &= A_{11} \hat{y}_{1t-1}^{*} + A_{12}^{2} \hat{y}_{1t-2}^{*} + \ldots + A_{21}^{2} y_{2t-1} + \ldots + A_{2p}^{2} y_{2t-p} - a_{10}^{2} \hat{y}_{1t}^{*} + \epsilon_{2t} \quad . \end{aligned}$$

Since the explanatory variables of the first equation and the second equation are different, the forecasted value of the latent variable is not completely correlated with the lagged values of the monetary variable. Thus, we do not have the multicollinearity problem encountered in the previous case. The lag length of each variable is 18. Since the explanatory variable of the first equation is not the latent variable but the discrete variable, we have to use the simulation technique to obtain the VMA coefficients. The technique of obtaining the VMA is the same as the one used in section 4.3.1 which is explained in detail in section 4.2. Even though the technique is the same, there is a slight procedural adjustment. When a forecasted

value of the latent variable is obtained, it has to be converted to the corresponding discrete choice to be able to feed it back into the right hand side of the first equation. Since the second equation depends on the latent variable, we feed the forecasted latent variable into the second equation. After estimating the VMA coefficients the orthogonal IRF can be obtained by multiplying the VMA coefficients with the inverse of the contemporaneous coefficient matrix as described in (4.5).

As previously indicated we estimated these two different models and obtained their IRFs and VDCs. However, the responses of monetary variables to an expansionary monetary shock were not plausible values, thus we do not report those results here.

4.4 Conclusion

In this essay the appropriate approach to estimating a VAR model that includes a discrete variable and then computing IRFs and VDCs is examined. Several VAR models were estimated using an appropriate method, and the results were contrasted with those obtained from VARs estimated using OLS.

Even though the highest and the lowest IRF values produced by these two separate methods are comparable with each other, in general, the results are quite different. Unexpectedly, after an expansionary monetary shock, the probit approach yields decreasing money aggregate (M2, monetary base and NBR) responses. We would expect that a one-time expansionary policy will, in general, increase the money supply in the short-run but have no lasting effects in the long-run. However, the IRFs of the probit model are not completely unacceptable, since for the first 6-9 periods,

money aggregates actually increase after an expansionary policy. The unexpected part is in the periods after the responses reach their peak value, in which money aggregate values continuously decrease.

The probit model produces unexpected results for the two interest rates and the spread, too. With both the probit approach and the OLS approach, the impulse responses of these three variables decrease after an expansionary monetary policy. However, with the OLS approach, the responses of FFR, TBILL, and SPREAD more close to zero at the 36^{th} period, whereas, in the probit approach, these values are much higher than zero. The higher interest rates and lower money aggregates are consistent, because a decrease in money supply causes excess demand for money which causes the interest rates to increase. However, one would not expect those movements in the long-run following a one-time expansionary monetary shock.

The VDC values of these two approaches are different, too. The probit approach attributes most of the forecast errors of M2, NBR, BASE and SPREAD to the innovations in the monetary policy, while the OLS approach attributes most of the forecast errors of the FFR and TBILL to the policy innovations. With the OLS estimation, most of the forecast errors of the monetary aggregates and the SPREAD are explained by innovations to themselves.

The most serious problems of the method we have investigated is the decrease in the money supply and the increase in the interest rate after an expansionary monetary policy. These results suggest that the probit technique in IRF and VDC analysis has to be further developed and investigated.

CHAPTER 5

CONCLUSION

In this dissertation we investigated policy analysis and forecasting within VAR models. Chapter 2 focuses upon specification of the lag length in VAR models while Chapter 3 focuses upon the forecasting performances of different type of VARs. Chapter 4 examines the estimation of a VAR with a qualitative variable.

The results in Chapter 2 show that generally Schwarz's information criterion (SIC) and Phillips' posterior information criterion (PIC) tend to choose lower lag lengths than the true ones (for a sample size of 124 or lower). Both in symmetric and asymmetric models, Akaike's information criterion (AIC) more often estimates the correct lag length than the SIC and PIC criteria. Keating's method with the AIC criterion does relatively well in estimating the lag structure. For symmetric VAR cases, the symmetric lag specification method with AIC criterion does better than Keating's method with AIC criterion. The results show that Keating's method performs better than the equation-by-equation lag specification methods, Hsiao's method, in estimating the lag structure, except in asymmetric cases where each equation has different lag lengths in each equation (first degree asymmetry).

In symmetric VARs, VARs specified using the AIC criterion generate the most accurate IRFs. Keating's method yields low IRF mse values in asymmetric models with long lags. For symmetric VARs usually picking the lag length with AIC yields better forecast performances. In asymmetric VAR cases, Keating's method with SIC

performs better in forecasting. Even though using SIC may yield lower forecast mse, the reason for that is lower forecast variance. In general, using SIC yields higher forecast bias.

In summary, a method which employs the AIC criterion rather than SIC or PIC is generally preferred. The results show that, in asymmetric models, the difference between the symmetric lag specification model with the AIC criterion and Keating's method with the AIC criterion is not substantial. However, even though there is not a big loss using the symmetric lag specification method with AIC criterion, Keating's method has some advantage, especially in IRF computations. When the VAR model becomes larger and more asymmetric, this advantage becomes substantial. Keating's method is more reliable than Hsiao's method in picking the correct lag structure. The forecasts, impulse responses, and variance decompositions produced by a VAR in which the lag structure is specified by Keating's method are more accurate than in a VAR in which the lag structure is specified by Hsiao's method.

In Chapter 3 the forecasting performance of various estimators was investigated. With quarterly data, the best forecasters of the 6-month-commercial-paper rate, the log level of GDP and the log level of GDP deflator are obtained from Bayesian VAR (BVAR) and Stein-rule VAR (SRVAR). With monthly data, the most accurate forecasts of the 6-month-commercial-paper rate, log level of IP and log level of CPI are also obtained from BVAR.

The SRVAR's forecast is easier to obtain than the BVAR forecast. In Stein-rule estimation, only the lag length has to be determined before estimation. There is no

need for defining a prior distribution as in estimating a BVAR. Furthermore, generating a BVAR forecast requires determination of at least two hyperparameters which can be time consuming.

The results show that one particular forecasting method does not always yield the lowest forecast error for all variables at all forecast horizons. In general, BVAR and SRVAR forecasts are more accurate than the unrestricted VAR forecasts. SRVAR has the worst performance only with the monthly growth rate forecasting model. If we look at the overall forecast performance measures (SOFM and SUM) for quarterly data, then SRVAR is superior to BVAR; however, for monthly data, BVAR has the most accurate forecasts. The VARs with asymmetric lag structures usually have high forecast errors.

We also did two Monte-Carlo experiments to investigate the effects of correct restrictions and incorrect restrictions on the forecast performances of BVAR and SRVAR. The Monte-carlo experiment results show that SRVAR produces lower mean-square-forecast-error statistics than BVAR; however, statistically the forecast accuracies of BVAR and SRVAR are not different.

Overall, the results show that SRVAR can be considered as an alternative to BVAR, especially when it is costly to estimate BVAR forecasts which might be the case for large VAR models. To forecast different variables, different methods may be used. However, for most cases the overall forecast performance of SRVAR is more accurate than the other forecast methods.

In Chapter 4 the appropriate approach to estimating a VAR model that includes a discrete variable and computing IRFs and VDCs is examined. Several VAR models were estimated using an appropriate method and the results were contrasted with those obtained from VARs estimated using OLS.

Even though the highest and the lowest IRF values produced by these two separate methods are compatible with each other, in general the results are quite different. Unexpectedly, after an expansionary monetary shock, the probit approach yields decreasing money aggregate (M2, monetary base and NBR) responses. We would expect that an expansionary policy will increase the money supply in the short-run. However, the IRFs of the probit model are not completely unacceptable, since the first 6-9 periods money aggregates actually increase after an expansionary policy. The unexpected part is the following periods after the responses reach their peak value, in which money aggregate values continuously decrease.

The probit model produces unexpected results for the two interest rates and the spread, too. With both the probit approach and the OLS approach, the impulse responses of these three variables decrease after an expansionary monetary policy. However, with the OLS approach, the responses of FFR, TBILL, and SPREAD take values close to zero 36 periods after the shock, whereas, in the probit approach these values are much higher than zero.

The VDC values of these two approaches are different, too. The probit approach attributes most of the forecast errors of M2, NBR, the monetary base and the spread between 6 month commercial paper rate and 3 month T-bill rate to the

innovations in the monetary policy, while the OLS approach attributes most of the forecast errors of the FFR and 3 month T-bill rate to the policy innovations. With the OLS estimation most of the forecast errors of the monetary aggregates and the SPREAD are explained by innovations to themselves.

The most serious problems of the method we have investigated is the decrease in the money supply and increase in the interest rate after an expansionary monetary policy. These results suggest that the probit technique in IRF and VDC analysis has to be further developed and investigated.

BIBLIOGRAPHY

Akaike, Hirotugu. "Fitting Autoregressive Models for Prediction." *Annals of the Institute of Statistical Mathematics* 21, (2, 1969): 243-247.

_____"Autoregressive Model Fitting for Control." *Annals of the Institute of Statistical Mathematics* 23, (2, 1971): 163-180.

_____"A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* AC-19, (December,1974): 716-723.

Amemiya, Takeshi. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46, (September, 1978): 1193-1205.

Andersen, Leonall C. and Jerry L. Jordan. "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization." *Federal Reserve Bank of St. Louis Review* 50, (November, 1968): 11-24.

Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics 1941-1970.* Washington D.C.: Board of Governors of the Federal Reserve System, 1976.

Boschen, John F. and Leonard O. Mills. "The Relation Between Narrative and Money Market Indicators of Monetary Policy." *Economic Inquiry* 33, (January, 1995): 24-44.

_____. "The Effects of Countercylical Monetary Policy On Money and Interest Rates: An Evaluation of Evidence From FOMC Documents." *Federal Reserve Bank of Philadelphia Working Paper* No. 91-20, (1991).

Box, George E.P. and Gwilym M. Jenkins. 1970. *Time Series Analysis Forecasting and Control*, San Francisco: Holden-Day.

Braun, Phillip A. and Stefan Mittnik. "Misspecifications in Vector Autoregressions and Their Effects on Impulse Responses and Variance Decompositions." *Journal of Econometrics* 59, (October,1993): 319-341.

Caines, R. E., C.W. Keng, and S.P. Sethi. "Causality Analysis and Multivariate Autoregressive Modelling With an Application to Supermarket Sales Analysis." *Journal of Economic Dynamics and Control* 5, (August, 1981): 267-98.

Carlino, Gerald and Robert DeFina. "Does Monetary Policy Have Different Regional Effects?" *Federal Reserve Bank of Philadelphia Working Paper* No.94-23, (1994).

Christiano, Lawrence J. and Martin Eichenbaum. "Unit Roots In Real GDP: Do We Know, and Do We Care?" *NBER working Paper* No. 3130, October 1989.

Christiano, Lawrence J. and Lars Ljungquist. "Money Does Granger-Cause Output in the Bivariate Money Output Relation." *Journal of Monetary Economics* 22, (September, 1988): 217-235.

Dua, Pami and David J. Smyth. "Forecasting US Home Sales Using BVAR Models and Survey Data On Households Buying Attitudes for Homes." *Journal of Forecasting* 14, (May, 1995): 217-227.

Doan, Thomas A., Robert B. Litterman, and Christopher Sims. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3, (1, 1984): 1-100.

Eichenbaum, Martin and Kenneth Singleton. "Do Equilibrium Real Business Cycle Theories Explain Post-War Business Cycles?" edited by Stanly Fischer *NBER Macroeconomics Annual*, Vol 1, Cambridge Mass.: MIT Press, (1986): 91-134.

Fackler, James S., W. Douglas McMillin, and J. Lew Silver. "Are Monetary Services Indexes Superior to Conventional Monetary Aggregate as Intermediate Targets?" *Applied Economics* 22, (12, 1990): 1751-1759.

Fildes, Robert. "The Evaluation of Extrapolative Forecasting Methods." *International Journal of Forecasting* 8, (1, 1992): 81-98.

Friedman, Milton and Anna J. Schwartz. 1963. *A Monetary History of the United States 1867-1960*, Princeton: Princeton University Press.

______. 1970. *Monetary Statistics of the United States*, New York: Columbia University Press.

Geweke, John and Richard Meese. "Estimating Regression Models of Finite But Unknown Order." *International Economic Review* 22, (February, 1981):55-70.

Hafer, Rik W. and Richard G. Sheehan. "The Sensitivity of VAR Forecasts to Alternative Lag Structures." *International Journal of Forecasting* 5, (3, 1989):399-408.

Hannan, Edward J. and Barry G. Quinn. "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society* B41, (2, 1979): 190-195.

Greene, William, 1992. *Econometric Analysis*, New York:MacMillan.

Heckman, James J. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46, (July, 1978):931-953.

Hsiao, Cheng. "Autoregressive Modelling and Money-Income Causality Detection." *Journal of Monetary Economics* 7, (January, 1981): 85-106.

Judge, George G., and Marry E. Bock. 1978. *Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Amsterdam: North-Holland.

Judge, George G., William E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*, Second Edition, New York: Wiley.

Judge, Gearge G., R.Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee. 1988. *Introduction to The Theory and Practice of Econometrics*, Second Edition, New York: Wiley.

Kadiyala, K. Rao and Sune Karlsson. "Forecasting With Generalized Bayesian Vector Autoregressions." *Journal of Forecasting* 12, (April, 1993): 365-378.

Keating, John W. "Vector Autoregressive Models with Asymmetric Lag Structure." *Working Paper #184, Washington University*, March 1994.

Kimelman, Nancy Jo. "Post-War Monetary Policymaking in the U.S." *Ph.D. Dissertation, Brown University*, 1981.

King, Robert and Charles Plosser, "Money, Credit and Prices in Real Business Cycles," *American Economic Review* 74, (June, 1984):363-380.

Knight, John D., R. Carter Hill and C. F. Sirmans. "Biased Prediction of Housing Values." *Journal of American Real Estate and Urban Economics Association* 20, (3, 1992): 427-456.

Litterman, Robert B. "Techniques of Forecasting Using Vector Autoregressions," *Working Paper No. 115, Federal Reserve Bank of Minneapolis*, 1979.

_____. "Forecasting With Bayesian Vector Autoregression - Five Years of Experience." *Journal of Business and Economic Statistics* 4, (January, 1986): 25-38.

Liu, Te-Ru, Mary E. Gerlow and Scott H. Irwin. "The Performance of Alternative VAR Models in Forecasting Exchange Rates." *International Journal of Forecasting* 10, (November 1994): 419-433.

Lupoletti, William M. and Roy H.Webb. "Defining and Improving the Accuracy of Macroeconomic Forecast: Contributions from a VAR Model." *The Journal of Business* 59, (April, 1986): 263-285.

Lütkepohl, Helmut, 1993. *Introduction to Multiple Time Series Analysis*, Second Edition, Berlin: Springer-Verlag.

Lütkepohl, Helmut and Hans-Eggert Reimers. "Impulse Response Analysis of Cointegrated Systems." *Journal of Economic Dynamics and Control* 16, (January, 1992):53-78.

Maddala, G. S. 1983. *Limited-Dependent Qualitative Variables In Econometrics*, Cambridge, Cambridge University Press.

McMillin, W. Douglas and James S. Fackler. "Monetary vs. Credit Aggregates: An Evaluation of Monetary Policy Targets." *Southern Economic Journal* 50, (January, 1984):711-723.

McNees, Stephen K. "Forecasting Accuracy of Alternative Techniques: A Comparison of Macroeconomic Forecasts." *Journal of Business and Economic Statistics* 4, (January, 1986): 5-15.

Morgan, Donald P. "Asymmetric Effects of Monetary Policy." *Economic Review, Federal Reserve Bank of Kansas City* 78, (2,1993):21-34.

Phillips, Peter C.B. "Model Determination and Macroeconomic Activity." *Cowles Foundation*, Working Paper, (September, 1994).

Poole, William. "Rules-of-Thumb for Guiding Monetary Policy." in *Open Market Policies and Operating Procedures-Staff Studies*, Board of Governors of the Federal Reserve System, 1971, 135-189.

Potts, Glen T. and Dudley G. Luckett. "Policy Objectives of the Federal Reserve System." *The Quarterly Journal of Economics* 43, (August, 1978): 525-534.

Romer, Christina D. and David H. Romer. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." edited by Olivier J. Blanchard and Stanly Fischer editors, *NBER Macroeconomic Annual*, (1989):121-169.

_____. "Monetary Policy Matters." *Working Paper, University of California at Berkeley*, 1992.

Sarantis, Nicholas and Chris Stewart. "Structural, VAR and BVAR Models of Exchange Rate Determination: a Comparison of Their Forecasting Performance." *Journal of Forecasting* 14, (May 1995): 201-215.

Sargent, Thomas. "A Classical Macroeconomic Model For The United States." *Journal of Political Economy* 87, (January, 1976): 207-237.

Schwarz, Gideon. "Estimating the Dimension of a Model." *Annals of Statistics* 6, (March, 1978): 461-464.

Sims, Christopher. "Money, Income and Causality." *American Economic Review* 62, (September, 1972):540-542.

_____. "Macroeconomics and Reality." *Econometrica* 48, (January,1980): 1-48.

_____. "Policy analysis with econometric models." *Brookings Papers on Economic Activity* 1, (1, 1982): 107-154.

Stock, James H. and Mark W. Watson. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40, (January, 1989): 161-181.

Uselton, Gene C., 1974. *Lags in the Effects of Monetary Policy: A Nonparametric Analysis*, New York: Marcel Decker.

Webb, Roy H., "Vector Autoregressions as a Tool For Forecasting Evaluation," *Economic Review Federal Reserve Bank of Richmond* 70, (1, 1984): 3-11.

VITA

Omer Ozcicek received a bachelor of science degree in Electrical Engineering from Bilkent University, Ankara, Turkey in 1990. In 1991 he came to Baton Rouge and entered the graduate program in the Department of Economics in Louisiana State University. He is expected to earn his doctor of philosophy degree in August, 1996.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Omer Ozcicek

Major Field: Economics

Title of Dissertation: Three Essays On VAR Techniques

Approved:

W. Douglas McMillin

Major Professor and Chairman

John W. Larkin

Dean of the Graduate School

EXAMINING COMMITTEE:

R. Carter Hill

Date of Examination: