2010

Bayesian multivariate predictions

Weijie Mao
University of Iowa

BAYESIAN MULTIVARIATE PREDICTIONS

by

Weijie Mao

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Economics
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisor: Professor Emeritus John F. Geweke

ABSTRACT

This work offers two strategies to raise the prediction accuracy of Vector Autoregressive (VAR) Models. The first strategy is to improve the Minnesota prior, which is frequently used for Bayesian VAR models. The improvement is achieved in two ways. First, the variance-covariance matrix of regression disturbances is treated as unknown and random to incorporate parameter uncertainty. Second, the prior variance-covariance matrix of regression coefficients is constructed as a function of the variance-covariance matrix of disturbances, in order to account for dependencies between different equations. Since different prior specifications unavoidably lead to different models, and forecasting capability of any such model is often limited, the second strategy is to build an optimal prediction pool of models by using the conventional log predictive score function. The effectiveness of the proposed strategies is examined for one-step-ahead, multi-4-step-ahead, and single-4-step-ahead predictions through two exercises. One exercise is predicting national output, inflation, and interest rate in the United States, and the other is predicting state tax revenue and personal income in Iowa. The empirical results indicate that a properly selected prior can improve the prediction performance of a BVAR model, and that a real-time optimal prediction pool can outperform a single best constituent model alone.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

BAYESIAN MULTIVARIATE PREDICTIONS

by

Weijie Mao

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Economics
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisor: Professor Emeritus John F. Geweke

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Weijie Mao

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Economics at the December 2010 graduation.

Thesis Committee: John F. Geweke, Thesis Supervisor

Kung-Sik Chan

Beth F. Ingram

Nathan E. Savin

Charles H. Whiteman

To my husband, Jiewu Meng, for his encouragement

ACKNOWLEDGMENTS

I am grateful to my committee members Professors Charles Whiteman, Beth Ingram, Nathan Savin, and Kung-Sik Chan, and especially my advisor, Professor John Geweke, for guidance and advice. My research benefited greatly from Professor John Geweke, who guided me through many technical details in this thesis. As the first reviewer of almost everything I wrote, he continued to offer pertinent comments with insights.

Financial support from the department, Professors Geweke and Whiteman, is kindly acknowledged. Worthy of particular thanks are Professors Geweke and Whiteman, for their very generous financial support during the economic recession between 2008 and 2010. In addition, I owe thanks to my graduate coordinator, Ms. Renea Jay, for her impeccable administrative assistance.

Finally, I would like to thank my parents who respect my choices, and my husband who stands firmly by me.

ABSTRACT

This work offers two strategies to raise the prediction accuracy of Vector Autoregressive (VAR) Models. The first strategy is to improve the Minnesota prior, which is frequently used for Bayesian VAR models. The improvement is achieved in two ways. First, the variance-covariance matrix of regression disturbances is treated as unknown and random to incorporate parameter uncertainty. Second, the prior variance-covariance matrix of regression coefficients is constructed as a function of the variance-covariance matrix of disturbances, in order to account for dependencies between different equations. Since different prior specifications unavoidably lead to different models, and forecasting capability of any such model is often limited, the second strategy is to build an optimal prediction pool of models by using the conventional log predictive score function. The effectiveness of the proposed strategies is examined for one-step-ahead, multi-4-step-ahead, and single-4-step-ahead predictions through two exercises. One exercise is predicting national output, inflation, and interest rate in the United States, and the other is predicting state tax revenue and personal income in Iowa. The empirical results indicate that a properly selected prior can improve the prediction performance of a BVAR model, and that a real-time optimal prediction pool can outperform a single best constituent model alone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Economic forecasting is necessary when decisions have to be made in circumstances, where future values of many economic variables are unknown. Forecasting multivariate time series is virtually ubiquitous in macroeconomics, since policy analysis often considers all levels of economic activities such as output, investment, and consumption simultaneously. A milestone in multivariate prediction was achieved by Sims (1980), who proposed the use of unrestricted vector autoregressive (VAR) models.

A clean and intelligent way of specifying and estimating VAR models is through a Bayesian approach. Bayesian VAR (BVAR) models, which originated from Litterman (1979, 1986) and Doan et al. (1984), overcome the problem of overparameterization and gain remarkable empirical success. Currently BVAR models are combined with Dynamic Stochastic General Equilibrium (DSGE) models to improve macroeconomic forecasting (DeJong et al., 2000; Smets and Wouters, 2005; Del Negro et al., 2007). In any Bayesian inference, a fundamental yet challenging step is prior specification, which influences posterior distributions of unknown parameters and forecasts consequently (Geweke, 2005). Therefore, it is technically important to specify a sensible BVAR prior, in order to improve a model's prediction performance. Moreover, since different modeling techniques unavoidably lead to different models, it is practically desirable to construct an optimal pool of models, in order to achieve better performance over a single model.

This chapter provides a literature review of related topics. Section 1.1 presents the advantages of BVAR models, and discusses the current stage of the integration of BVARs and DSGEs. Since prior specification plays an important role in Bayesian inference, Section 1.2 reviews the widely used Minnesota prior with its extensions. Since there usually exist multiple competing models at the same time due to forecasts' different prior beliefs, Section 1.3 discusses the techniques of combining prediction models. Finally, Section 1.4 lays out the structure of this work.

1.1 Bayesian Prediction

An unrestricted N-variable pth-order VAR model, as suggested by Sims (1980), can be written as

$$\underset{N\times 1}{\mathbf{y}_t} = \sum_{l=1}^{p} \underset{N\times 1}{\mathbf{y}_{t-l}} \underset{N\times N}{\mathbf{A}_l} + \underset{N\times 1}{\boldsymbol{\varepsilon}_t}, \tag{1.1}$$

where $\mathbf{y}_t$ is a vector of variables being forecasted, $\mathbf{A}_l$'s are matrices of unknown coefficients, and $\boldsymbol{\varepsilon}_t$ is a vector of white-noise disturbances.

A VAR model as specified in (1.1) has $N \times (N \times p)$ coefficients in total. Since the number of coefficients to be estimated quickly increases with the number of variables as well as the number of lags in the system, a moderate sized system can be highly overparameterized relative to the number of observations. An overparameterized unrestricted VAR model can explain data "too well". It captures not only important features that are useful for forecasting, but also noisy features that merely reflect accidental or random relationships. Statistically, overparameterization usually causes multicollinearity and loss of degrees of freedom, which lead to inefficient estimates and large out-of-sample forecasting errors. To avoid overparameterization, one approach is to find the maximum lag length in a VAR model and exclude insignificant lags based on statistical tests (e.g., Akaike, 1974;

Schwarz, 1979; Hannan and Quinn, 1979). Another approach is to specify unequal numbers of lags for different equations in a VAR model (e.g., Hsiao, 1981, 1982; Hendry and Mizon, 1993; Clements and Mizon, 1991). However, these approaches rely on classical hypothesis testing.

A Bayesian VAR model, however, offers an intelligent way to overcome overparameterization without relying on classical hypothesis testing. The pioneering work includes Litterman (1979) and Doan, Litterman, and Sims (1984). Theoretically, BVARs have three apparent advantages. First, BVARs avoid overparameterization by imposing general restrictions on coefficients instead of implementing complete exclusions. Excluding long lags from a VAR model amounts to an overly absolute belief that their coefficients are zeros. Such exclusions cannot be revised by any amount of data evidence. By contrast, BVARs impose flexible prior specifications on these coefficients to shrink them more tightly towards zeros than coefficients of short lags. Such prior specifications can be easily overridden by data, if long lags turn out to have strong impact. Second, BVARs enable forecasters to impose prior specifications through probabilistic terms in a fully transparent way. The means of prior distributions reflect forecasters' prior beliefs and best guesses about the true values of unknown parameters. The variances reflect forecasters' confidence on the prior means. Small prior variances indicate that forecasters believe that the true values are not likely to deviate from their guesses (i.e., prior means), and vice versa. This standard specification procedure allows resulting forecasts to be reproduced. Third, BVARs generate complete multivariate density forecasts, by fully incorporating parameter uncertainty instead of simply using point estimates of parameters. A posterior distribution for each parameter can be obtained from its prior distribution and the likelihood function by using the Bayes rule.

Empirically, BVARs achieve remarkable success in both national and regional predictions (e.g., Amirizadeh and Todd, 1984; Litterman, 1986; Dua and Ray, 1995; Gupta and Sichei, 2006).

The empirical success of Bayesian estimation techniques have motivated researchers to develop advanced models by integrating BVARs with other models, such as Dynamic Stochastic General Equilibrium (DSGE) models. DSGE models are micro-founded and optimization-based models that are capable of providing a complete qualitative analysis of the workings of the economy. However, DSGEs were rarely used for quantitative analysis due to unsatisfactory calibration of their structural parameters (e.g., Kydland and Prescott, 1982). On the one side of the integration from DSGEs to BVARs, Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004) showed that information derived from DSGEs can be used to construct priors for BVARs, which can then achieve slightly better prediction performance than BVARs with the Minnesota prior. On the other side of the integration from BVARs to DSGEs, Smets and Wouters (2003) showed that a DSGE model with structural parameters estimated by Bayesian techniques can deliver acceptable forecasts. Del Negro et al. (2007) explicitly used a BVAR with four lags to estimate the state-space representation of a DSGE, and obtained better forecasts than using traditional DSGEs. The current trend in macroeconomic forecasting is clearly to formulate DSGE-BVAR models that can take advantage of economic theories and statistical tools to facilitate policy analysis and produce more accurate forecasts than either DSGE or BVAR alone.

1.2 Bayesian Priors

In any Bayesian application, from calibrating structural parameters of theoretical DSGE models to estimating regression coefficients of atheoretical VAR models, prior specification of unknown parameters is a required and indispensable step. This section begins with a review of the widely used Minnesota prior[1], followed by a discussion of several available modifications in the literature.

The unknown parameters in a VAR model typically include regression coefficients and variances of regression disturbances. In the Minnesota prior proposed by Doan et al. (1984) and Litterman (1986), the variances of disturbances are fixed to estimated residual variances obtained from univariate autoregressions[2], and the coefficients are assumed to have normal distributions with the means and variances representing forecasters' prior beliefs about the coefficients. Certainly it is impractical for forecasters to specify the mean and variance for each coefficient individually, when hundreds of coefficients can be involved. An advantage of the Minnesota prior is to generate a full set of prior means and variances for BVAR coefficients automatically. The prior specification proceeds in two steps. For example, a simple output-inflation VAR model based on (1.1) with two lags of each target variable

[1]The name is due to its development at the University of Minnesota and the Federal Reserve Bank of Minneapolis.

[2]Litterman (1986) specified a full variance-covariance matrix of regression disturbances, but was unable to follow through on its implications for the likelihood function and for forecasting, primarily due to computational difficulty in the early 1970s. A compromise was the equation-by-equation treatment, and consequently the estimated residual variances obtained from univariate regressions were substituted for the posterior distribution.

can be written as

$$\begin{aligned} Output_{t+1} &= a_{11,1}Output_t + a_{11,2}Output_{t-1} \\ &+ a_{12,1}Inflation_t + a_{12,2}Inflation_{t-1} + \varepsilon_{1,t+1}, \end{aligned} \tag{1.2}$$

$$\begin{aligned} Inflation_{t+1} &= a_{21,1}Output_t + a_{21,2}Output_{t-1} \\ &+ a_{22,1}Inflation_t + a_{22,2}Inflation_{t-1} + \varepsilon_{2,t+1}. \end{aligned} \tag{1.3}$$

The first step is to specify priors for the coefficients of the current and past values of the dependent variable in any given equation. These variables are known as *own lags*, which are $Output_t$ and $Output_{t-1}$ in (1.2) and $Inflation_t$ and $Inflation_{t-1}$ in (1.3). The second step is to specify priors for the coefficients of the current and past values of all the other variables except the dependent variable in any given equation. These variables are known as *cross lags*, which are $Inflation_t$ and $Inflation_{t-1}$ in (1.2) and $Output_t$ and $Output_{t-1}$ in (1.3). In the belief that own lags, particularly the first own lags, are important explanatory variables, the prior means of their coefficients are usually set to ones, while the prior means of the coefficients of cross lags are set to zeros. The prior standard deviations then indicate the likely departure of coefficient values from their prior means. Specifically, the standard deviation of the coefficient of lag l of variable j in equation i takes the following form,

$$s(i,j,l) = \begin{cases} \dfrac{\lambda}{l^d} & \text{for coefficients of own lags if } i = j, \\ \dfrac{\lambda}{l^d}\,\theta\,\dfrac{s_i}{s_j} & \text{for coefficients of cross lags if } i \neq j, \end{cases} \tag{1.4}$$

where the hyperparameters d, λ, and θ are assigned numerical values by forecasters, and s_i and s_j are fixed to the estimated standard deviations of the disturbances in univariate autoregressions. The properties of these parameters are summarized as follows.

- The term λ is called the overall tightness parameter. It is the prior standard deviation of the coefficient of the first own lag, and essentially controls the prior standard deviations of all the lag coefficients. In other words, it determines how all the coefficients are concentrated around their prior means. A tighter prior can be produced by decreasing λ.

- The term d is a decay factor, and $1/l^d$ controls the tightness on lag l relative to lag 1. Since coefficients of higher order lags are more likely to be close to zeros than those of lower order lags, prior standard deviations of coefficients decrease as lag length l increases.

- The term θ is called the cross-equation tightness parameter, which represents the tightness of variable j relative to variable i in equation i. Since own lags usually account for most of variation in a dependent variable, the coefficients of cross lags are assigned smaller standard deviations than those of own lags in relative terms by choosing θ between 0 and 1.

- The ratio s_i/s_j is used to account for differences in variability of different variables due to measurement.

For example, based on (1.4), the prior standard deviations of the coefficients of the simple output-inflation VAR model defined in (1.2) and (1.3) are given in the parentheses as follows.

$$\begin{aligned} Output_{t+1} &= \underset{(\lambda)}{a_{11,1}} Output_t + \underset{(\lambda/2^d)}{a_{11,2}}\ Output_{t-1} \\ &+ \underset{(\lambda\theta s_1/s_2)}{a_{12,1}} Inflation_t + \underset{(\lambda/2^d\theta s_1/s_2)}{a_{12,2}} Inflation_{t-1} + \varepsilon_{1,t+1}, \end{aligned} \tag{1.5}$$

$$\begin{aligned} Inflation_{t+1} &= \underset{(\lambda\theta s_2/s_1)}{a_{21,1}} Output_t + \underset{(\lambda/2^d\theta s_2/s_1)}{a_{21,2}} Output_{t-1} \\ &+ \underset{(\lambda)}{a_{22,1}} Inflation_t + \underset{(\lambda/2^d)}{a_{22,2}}\ Inflation_{t-1} + \varepsilon_{2,t+1}. \end{aligned} \tag{1.6}$$

Direct modifications of the Minnesota prior usually focus on alternative specifications for prior standard deviations of coefficients based on (1.4), because prior standard deviations determine how far the true values can deviate from the prior means, which represent forecasters' prior guesses. An immediate alteration is to replace the constant cross-equation tightness parameter θ with a weighting function $f(i,j)$ for variable j in equation i. Doan and Litterman (1986) applied such a weighting function to a BVAR model for predicting both national and state variables. They suggested specifying $f(i,j)$ according to a circle-star structure, in which star (national) variables affect both star and circle (state) variables while circle variables influence primarily only other circle variables. For example, the value of $f(i,j)$ is 0.6 for a national variable in either other national equations or state equations, 0.1 for a state variable in other state equations, and 0.01 for a state variable in national equations. Research along this line of thought includes LeSage and Pan (1995) and LeSage and Krivelyova (1999), both of which emphasized the distinction between variables from neighboring states and variables from non-neighboring states. That is, the value of $f(i,j)$ for a state variable in a neighboring state equation should be different from the value for a state variable in a non-neighboring state equation.

Although the specification of the weighting function $f(i,j)$ improves the forecast accuracy of BVAR models on state variables, two major limitations of BVAR models remain untouched. The first limitation is fixing the unknown variances of disturbances to their estimates obtained from univariate autoregressions. Since the

variances of disturbances are parameters of the likelihood function, such a specification with fixed values is essentially an overly strong restriction on the data likelihood function, and is rarely supported by practice. The second limitation is ignoring potential dependencies between equations. Basically, parameters in any one equation are independent of the parameters in any other equations. This is primarily due to the equation-by-equation prior specification and parameter estimation. Additionally, the prior standard deviations (i.e., $\frac{\lambda}{l^d}\,\theta\,\frac{s_i}{s_j}$ in (1.4)) of the coefficients of cross lags certainly disregard interaction between equations, because the values of s_i and s_j are fixed (see, e.g., Kadiyala and Karlsson, 1993; Geweke and Whiteman, 2006, for a detailed discussion).

To avoid these limitations of the Minnesota prior, there are alternative prior specifications in literature. One alternative is the "normal-Wishart" prior (e.g., Litterman, 1979; Broemeling, 1985). It assumes that regression coefficients are normally distributed conditional on the variance-covariance matrix of disturbances, which in turn has an inverted Wishart distribution. This prior does not have the two limitations of the Minnesota prior. Additionally it is a natural conjugate prior, which means that the posterior distribution of parameters can be obtained analytically and has the same functional form as the prior distribution. However, this prior has a drawback that it does not treat own lags and cross lags in the same fashion as the Minnesota prior (see, e.g., Geweke and Whiteman, 2006, for a detailed discussion). Another alternative is the "normal-diffuse" prior (e.g., Kadiyala and Karlsson, 1993; Lewis and Whiteman, 2006). It assumes independent Minnesota priors for coefficients in each equation and a diffuse prior for the variance-covariance matrix of disturbances. This prior avoids the disadvantages of the Minnesota prior and the "normal-Wishart" prior. However, because of the use

of a diffuse prior, the posterior distributions of parameters are not in recognizable standard distribution forms. In addition, large differences between the information contained in the prior and the likelihood function can cause the posterior distributions to be bimodal. Therefore, it is of interest to form BVAR priors that overcome the limitations of the Minnesota prior while maintaining its advantages in a fully Bayesian approach.

Another strain of literature takes a theoretical perspective on the prior specification of BVAR models. The earliest work includes DeJong et al. (1993) and Ingram and Whiteman (1994), which used DSGE models to supply additional information to the BVAR prior specification. That is, the prior means and variances of regression coefficients, particularly of the first own lags, are derived theoretically. Most recently, Del Negro and Schorfheide (2004) also used prior information obtained from DSGE models to introduce shrinkage. Specially, they had a tightness parameter to control the weight placed on the DSGE model prior versus the unrestricted VAR prior. Although DSGE-based BVAR models showed comparable forecasting performance to BVAR models by taking advantage of economic theories, they do not technically improve prior specifications from a statistical perspective.

1.3 Combinations of Prediction Models

Forecasters are likely to build different prediction models even for the same target variables due to different prior beliefs or modeling approaches. For example, forecasters can use micro-founded theoretical DSGE and atheoretical time-series BVAR as two different prediction models, and they can even have different BVAR models by simply varying prior specifications. With multiple prediction models available at the same time, there are two options to produce forecasts. The first is

to identify a single best model through a model selection procedure. The second is to pool competing models and then obtain a combined forecast. The second option has become increasingly popular (see, e.g., Timmermann, 2006, for a recent review) because of superior forecast accuracy in the sense of low root mean square errors. The success of combined forecasts primarily owes to the combination of different information sets and the robustness against misspecified models and poor estimations (see, e.g., Hendry and Clements, 2004).

As far as the form of combining forecasts is concerned, there are combinations of point forecasts, combinations of interval forecasts, and combinations of density forecasts. Until recently, most of the work focused on combining point forecasts (see, e.g., Clemen, 1989; Diebold and Lopez, 1996, for reviews). Certainly point forecasts can serve as a starting point. Interval forecasts and event probabilities can be derived from point forecasts, when forecast errors are normally distributed with a constant variance. However, normality and constant variance may not always be supported by data. Moreover, optimal point and interval forecasts can vary under an asymmetric loss function. Therefore, density forecast is more desirable for decision makings, because it provides a complete description of future uncertainty (see, e.g., Tay and Wallis, 2000, for a survey). Consequently, research on combining density forecasts from competing models is of more interest than combining point forecasts or interval forecasts.

Once a strategy of combing density forecasts is selected, its success depends largely on how well combination weights can be determined. A simple and inexpensive approach is to use equal weights. The experience of combining point forecasts suggests that a simple average with equal weights can outperform complicated

weighting schemes (see, e.g., Clemen, 1989). Wallis (2005), a forerunner of combining density forecasts, adopted the approach of equal weights. However, a more sophisticated and sensitive approach is to use optimal weights. Hall and Mitchell (2007) proposed a practical data-driven approach to search the optimal weights, which minimizes the Kullback-Leibler distance between the combined density forecast and the true but unknown density of variables of interest. Geweke and Amisano (2010) recommended using a log scoring function to construct linear weighted pools of prediction models. They provided a detailed theoretical analysis, and focused on competing models that can provide density forecasts (i.e., predictive distributions) of variables under consideration. Most importantly, they discussed at least three advantages of building optimal prediction pools. First, no model is assumed to coincide with the "true" model (i.e., the underlying data generating process). Second, given a data generating process, there exists a unique optimal prediction pool. Third, an optimal prediction pool can substantially outperform any constituent model in the pool.

1.4 Organization

Chapter 2 presents a full Bayesian vector autoregressive (BVAR) model with two major innovations. The first is to treat the unknown variance-covariance matrix of disturbances as a random matrix. The second is to construct the prior variance-covariance matrix of coefficients conditional on the variance-covariance matrix of disturbances. The details regarding the posterior simulation of the unknown parameters are also provided. Chapter 3 describes the construction of an optimal prediction pool by using a conventional log predictive score function. A contribution to the existing research is made by combining multivariate density forecasts in

multi-H-step-ahead and single-H-step-ahead predictions. Chapters 4 and 5 apply the proposed BVAR model and optimal pooling techniques to two forecasting exercises. One exercise is predicting national output, inflation, and interest rate in the United States, and the other is predicting state tax revenue and personal income in Iowa. Finally, Chapter 6 concludes.

CHAPTER 2
A BAYESIAN VECTOR AUTOREGRESSIVE MODEL

We develop a full Bayesian vector autoregressive (BVAR) model, which improves Litterman (1986)'s model in four ways. First, the variance-covariance matrix of regression disturbances is treated as unknown and random to incorporate parameter uncertainty. Second, the prior variance-covariance matrix of regression coefficients is constructed as a function of the variance-covariance matrix of disturbances. Third, the N equations of the model are estimated simultaneously to account for possible contemporaneous correlation of disturbances[1]. Fourth, the model is fully Bayesian. Since the prior distributions lead to posterior distributions that do not have analytical closed-form expressions, a state-of-the-art Metropolis-within-Gibbs sampler is specified to obtain draws from the posterior distributions.

This chapter is organized as follows. Section 2.1 establishes the notation used throughout the chapter. Section 2.2 specifies the data likelihood function of the unknown parameters, followed by Section 2.3 on the prior distributions of the unknown parameters. Finally, Section 2.4 offers a simulation algorithm to obtain draws from the posterior distributions of the unknown parameters.

[1]Litterman (1986, page 31) explicitly admitted the efficiency loss due to the single-equation operation. Geweke and Whiteman (2006, pages 43-45) provided a detailed discussion on this issue.

2.1 Basic Notation

An N-dimensional pth-order Gaussian vector autoregressive [VAR(p)] model can be written as[2]

$$\underset{N\times 1}{\mathbf{y}_t} = \underset{N\times 1}{\mathbf{C}} + \underset{N\times N}{\mathbf{A}_1}\,\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \ldots + \mathbf{A}_p\mathbf{y}_{t-p} + \underset{N\times 1}{\boldsymbol{\varepsilon}_t}, \quad \boldsymbol{\varepsilon}_t \overset{i.i.d}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), \tag{2.1}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is an $N \times 1$ random vector, $\mathbf{C} = (C_1, \ldots, C_N)'$ is an $N \times 1$ vector of intercept terms, $\mathbf{A}_l$ $(l = 1, \ldots, p)$ is an $N \times N$ coefficient matrix of the lth lag, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is an $N \times 1$ vector of disturbances. The disturbances have mean zero $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and the same (nonsingular) variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ for all t. Furthermore, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_s$ are uncorrelated for $t \neq s$.

Assuming that we have T observations (i.e., $\mathbf{y}_1, \ldots, \mathbf{y}_T$) and p presample values (i.e., $\mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0$) for each of the N variables, we define

$$\begin{aligned}
&\mathbf{Y}_T = (\mathbf{y}_1, \ldots, \mathbf{y}_T) && N \times T,\\
&\mathbf{B} = (\mathbf{C},\ \mathbf{A}_1,\ \ldots,\ \mathbf{A}_p) && N \times (1 + Np),\\
&\boldsymbol{\beta} = vec(\mathbf{B}) && (N + N^2p) \times 1,\\
&\mathbf{X}_t = (1,\ \mathbf{y}_{t-1}',\ \ldots,\ \mathbf{y}_{t-p}') && 1 \times (1 + Np),\\
&\mathbf{X}_T = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T \end{bmatrix} && T \times (1 + Np),
\end{aligned} \tag{2.2}$$

where *vec* operation is defined by stacking the rows of a matrix into a column vector, instead of stacking the columns of a matrix into a column vector. Using the above

[2]For specificity, the model includes only the constant as an exogenous variable. However, much of the discussion hereafter can be generalized to allow for the presence of other exogenous variables.

notation, we can write (2.1) more compactly as

$$\begin{aligned} \mathbf{y}_t &= (\mathbf{I}_N \otimes \mathbf{X}_t)\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \\ &= \mathbf{Z}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{2.3}$$

where $\otimes$ denotes the Kronecker product.

To perform Bayesian inference based on (2.3), we follow two principles, *explicit formulation* and *relevant conditioning*, as suggested in Geweke and Whiteman (2006). In the rest of this chapter, we specify three necessary components of Bayesian inference,

- a joint distribution of the observables $\mathbf{Y}_T$ conditional on the unknown parameters $\boldsymbol{\beta}$ and $\mathbf{H} = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$ and model assumptions A, i.e., $p\left(\mathbf{Y}_T \,|\, \boldsymbol{\beta}, \mathbf{H}, A\right)$,
- prior distributions of the unknown parameters, i.e., $p\left(\mathbf{H} \,|\, A\right)$ and $p\left(\boldsymbol{\beta} \,|\, \mathbf{H}, A\right)$, and
- posterior distributions of the unknown parameters, i.e., $p\left(\boldsymbol{\beta} \,|\, \mathbf{H}, \mathbf{Y}_T, A\right)$ and $p\left(\mathbf{H} \,|\, \boldsymbol{\beta}, \mathbf{Y}_T, A\right)$.

2.2 Conditional Distribution of Observables

Since the vector of the disturbances in (2.3) is assumed to have a normal distribution $N\left(0, \mathbf{H}^{-1}\right)$, where $\mathbf{H}$ is the $N \times N$ precision matrix (i.e., $\mathbf{H} = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$), the conditional probability density function (pdf) for the observable $\mathbf{Y}_T$ is

$$\begin{aligned} p\left(\mathbf{Y}_T \,|\, \boldsymbol{\beta}, \mathbf{H}, A\right) &= \prod_{t=1}^{T} p\left(\mathbf{y}_t \,|\, \mathbf{Y}_{t-1}, \boldsymbol{\beta}, \mathbf{H}, A\right) \\ &= (2\pi)^{-\frac{TN}{2}} |\mathbf{H}|^{\frac{T}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\mathbf{SH})\right], \end{aligned} \tag{2.4}$$

where the $N \times N$ matrix $\mathbf{S} = \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$. (The vector $\boldsymbol{\varepsilon}_t$ is defined implicitly in (2.1) and is a function of the parameter vector $\mathbf{C}$ and matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$.)

Throughout the work, the vector of the *observed* values is denoted by $\mathbf{Y}_T^o$, where the superscript o differs $\mathbf{Y}_T^o$ from the *observable random* vector $\mathbf{Y}_T$. Replacing $\mathbf{Y}_T$ with $\mathbf{Y}_T^o$ in (2.4), we have the expression of the likelihood function, $\text{Ł}\,(\boldsymbol{\beta}, \mathbf{H}; \mathbf{Y}_T^o, A) \propto p\,(\mathbf{Y}_T^o \,|\, \boldsymbol{\beta}, \mathbf{H}, A)$.

2.3 Prior Distributions of Unknown Parameters

In the model of (2.3), the unknown parameters are the vector of coefficients $\boldsymbol{\beta}$ and the variance-covariance matrix of disturbances $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. Assumptions about these parameters need to reflect empirical regularities and admit uncertainties.

In the work of Litterman (1986), the assumptions about $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\beta}$ are as follows.

- There was no prior distribution specified for the variance-covariance matrix of disturbances $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. Instead, the unknown variances of disturbances were fixed to the estimated residual variances from unrestricted univariate autoregressions.

- The prior distributions of the unknown regression coefficients (i.e., elements of $\boldsymbol{\beta}$) were independent normal distributions. The prior means of the coefficients of the first own lags were set to ones, while the prior means of the other coefficients were set to zeros. The prior standard deviation of the coefficient of lag l of variable j in equation i, $s(i, j, l)$ was

$$s(i,j,l) = \begin{cases} \dfrac{\lambda}{l} & \text{for coefficients of own lags if } i = j, \\ \dfrac{\lambda}{l}\,\theta\,\dfrac{s_i}{s_j} & \text{for coefficients of cross lags if } i \neq j, \end{cases} \tag{2.5}$$

which was in the same fashion as (1.4) by setting the decay factor d equal to 1. The term s_i was fixed to the estimated standard deviation of the disturbance in an univariate autoregression on the dependent variable i. The hyperparameters λ and θ were called overall tightness and cross-equation tightness, respectively. They were assigned numerical values by forecasters.

A major limitation of Litterman's specification is the fixed-value variances of disturbances, which is an overly strong assumption that ignores parameter uncertainty. Furthermore, given the values of hyperparameters λ and θ, since the values of s_i and s_j in (2.5) were fixed to their estimates from univariate autoregressions, the variance-covariance matrix of the coefficients was also fixed. In particular, the fixed prior standard deviations $\theta\frac{\lambda}{l}\frac{s_i}{s_j}$ did not allow the coefficients of cross lags to be impacted by the unknown disturbances, and did not take account of dependencies between the coefficients in the different equations of the VAR model.

Improving from Litterman's specification and more importantly from a strict Bayesian perspective, we treat $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ as an unknown and random matrix to incorporate parameter uncertainty. The precision matrix (i.e., $\mathbf{H} = \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}$) is assumed to follow a Wishart distribution (see, e.g., Zellner, 1971; Press, 1982, for details),

$$\mathbf{H} \,|\, A \sim W\,(\underline{\mathbf{S}}^{-1}, \underline{\nu}), \tag{2.6}$$

where $\underline{\mathbf{S}}^{-1}$ is a scale matrix and $\underline{\nu}$ is a degrees of freedom parameter. The hyperparameters with underscores in (2.6) will be assigned numerical values as a part of the prior specification. The Wishart density for (2.6) is

$$p\,(\mathbf{H} \,|\, A) \propto |\mathbf{H}|^{(\underline{\nu}-N-1)/2} |\underline{\mathbf{S}}|^{\underline{\nu}/2} \exp\{-\frac{1}{2}\mathrm{tr}(\underline{\mathbf{S}}\mathbf{H})\}, \quad \mathbf{H} > 0,\, \underline{\mathbf{S}} > 0, \tag{2.7}$$

where the matrices $\mathbf{H}$ and $\underline{\mathbf{S}}$ are positive definite.

The vector of coefficients $\boldsymbol{\beta}$ is then assumed to follow a multivariate normal distribution in the spirit of the Minnesota prior. Mathematically,

$$\boldsymbol{\beta} \,|\, (\mathbf{H}, A) \sim N\left(\underline{\boldsymbol{\beta}},\, \underline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}(\mathbf{H})\right), \tag{2.8}$$

where the prior variance-covariance matrix of the coefficients (i.e., $\underline{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \underline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}$) is a function of the variance-covariance matrix of the disturbances (i.e., $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{H}^{-1}$). Specifically, the prior standard deviations of the coefficients in any given equation i $(i = 1, \ldots, N)$ take the following forms,

$$\begin{cases} \kappa & \text{for coefficients of intercept terms,} \\ \dfrac{\lambda}{l} & \text{for coefficients of own lags if } i = j, \\ \dfrac{\lambda}{l}\theta\dfrac{\sigma_i}{\sigma_j} & \text{for coefficients of cross lags if } j \neq i, \end{cases} \tag{2.9}$$

where σ_i is an unknown and random variable. Precisely, it is the square root of the (i, i) element of $\mathbf{H}^{-1}$.

Consequently, the proposed prior specifications in (2.8) and (2.9) improve Litterman's prior by replacing the fixed-value parameters with unknown random variables. A resulting advantage is introducing uncertainty into parameter estimation by admitting the randomness of the unknown parameters.

2.4 Posterior Simulator of Unknown Parameters

Since the proposed priors do not lead to recognizable standard posterior distributions of the unknown parameters, a Metropolis-Hastings algorithm is developed to handle posterior simulations.

The posterior conditional distribution of $\boldsymbol{\beta}$ is a multivariate normal (e.g., see Geweke (2005) for a derivation of the results). That is,

$$\boldsymbol{\beta} \,|\, (\mathbf{H}, \mathbf{Y}_T^o, A) \sim N\left(\overline{\boldsymbol{\beta}},\, \overline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}\right), \tag{2.10}$$

where

$$\overline{\mathbf{H}}_{\boldsymbol{\beta}} = \underline{\mathbf{H}}_{\boldsymbol{\beta}}(\mathbf{H}) + \mathcal{Z}_T'(\mathbf{H} \otimes \mathbf{I}_T)\mathcal{Z}_T, \tag{2.11}$$

$$\overline{\boldsymbol{\beta}} = \overline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}\left[\underline{\mathbf{H}}_{\boldsymbol{\beta}}(\mathbf{H})\underline{\boldsymbol{\beta}} + \mathcal{Z}_T'(\mathbf{H} \otimes \mathbf{I}_T)\mathcal{Y}_T^o\right], \tag{2.12}$$

where $\mathcal{Z}_T = \mathbf{I}_N \otimes \mathbf{X}_T$ and $\mathcal{Y}_T^o = vec(\mathbf{Y}_T^o)$.

Since $\underline{\mathbf{H}}_{\boldsymbol{\beta}}$ is a function of $\mathbf{H}$, the posterior distribution of $\mathbf{H}$ does not belong to a recognizable standard distribution. However, the kernel of the posterior conditional distribution of $\mathbf{H}$ is

$$p\left(\mathbf{H} \,|\, \boldsymbol{\beta}, \mathbf{Y}_T^o, A\right) \propto p\left(\boldsymbol{\beta}|\mathbf{H}, A\right)p\left(\mathbf{H} \,|\, A\right)p\left(\mathbf{Y}_T^o \,|\, \boldsymbol{\beta}, \mathbf{H}, A\right). \tag{2.13}$$

Thus a Metropolis-Hastings algorithm can be developed to obtain draws from this posterior distribution. Since a part of (2.13), $p\left(\mathbf{H} \,|\, A\right)p\left(\mathbf{Y}_T^o \,|\, \boldsymbol{\beta}, \mathbf{H}, A\right)$, is proportional to the kernel of a Wishart distribution in $\mathbf{H}$,

$$W\left((\underline{\mathbf{S}} + \mathbf{S})^{-1}, \underline{\nu} + T\right), \tag{2.14}$$

it can be used as a candidate generating density. Consequently, the ratio of the target to the candidate density is proportional to $p\left(\boldsymbol{\beta} \,|\, \mathbf{H}, A\right)$, and the acceptance probability has a relatively simple form

$$\alpha(\mathbf{H}^* \,|\, \mathbf{H}^{(m-1)}) = \min\left[\frac{p\left(\boldsymbol{\beta} \,|\, \mathbf{H}^*, A\right)}{p\left(\boldsymbol{\beta} \,|\, \mathbf{H}^{(m-1)}, A\right)},\, 1\right]. \tag{2.15}$$

Specifically, a Metropolis-within-Gibbs algorithm for the posterior simulation consists of the following steps.

(0) Choose an initial value of $\mathbf{H}^{(0)}$.

(1) Compute $\underline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1^{(m)}}$ conditional on $\mathbf{H}^{-1^{(m-1)}}$ according to (2.9).

(2) Draw $\boldsymbol{\beta}^{(m)}$ from its posterior conditional Gaussian distribution described in (2.10), (2.11), and (2.12).

(3) Take a candidate draw, $\mathbf{H}^*$, from the candidate generating density described in (2.14).

(4) Calculate an acceptance probability α described in (2.15).

(5) Set $\mathbf{H}^{(m)} = \mathbf{H}^*$ with probability α and set $\mathbf{H}^{(m)} = \mathbf{H}^{(m-1)}$ with probability $1 - \alpha$. That is, generate a random number, u, from a Uniform (0,1), if $u < \alpha$, set $\mathbf{H}^{(m)} = \mathbf{H}^*$, otherwise set $\mathbf{H}^{(m)} = \mathbf{H}^{(m-1)}$.

(6) Repeat Steps (1) to (5) M times.

CHAPTER 3
OPTIMAL COMBINATION OF MULTIVARIATE FORECASTS

We offer a practical approach to combining competing prediction models. The optimal weights of the models are obtained by maximizing the conventional log predictive score of the linear prediction pool under consideration. The real-time predictive probability densities of each model in the optimal pool are then evaluated to identify the pivotal observations that account for each model having a positive weight.

This chapter is organized as follows. Section 3.1 presents an analytical closed-form conditional joint distribution of forecasts, which can be used to evaluate predictive probability densities of forecasts. Section 3.2 provides an algorithm for computing conventional log predictive scores of multi-H-step-ahead and single-H-step-ahead predictions. Section 3.3 explains the computation of optimal weights of models for constructing a linear prediction pool. Finally, Section 3.4 outlines a search scheme for pivotal observations that explain why a model is given a positive weight in an optimal prediction pool.

3.1 Predictive Probability Densities of Forecasts

To reach the goal of combining prediction models, we first compute predictive probability densities of forecasts from each competing model. We consider an N-variable and pth-order autoregressive model described in (2.1). A vector of multi-H-step-ahead forecasts from an end-of-sample point t is denoted

$$\underset{NH\times 1}{\boldsymbol{\omega}_{t+1,t+H}} = \begin{bmatrix} \mathbf{y}_{t+1} \\ \mathbf{y}_{t+2} \\ \vdots \\ \mathbf{y}_{t+H} \end{bmatrix}, \tag{3.1}$$

for some $H > 0$.

Given the assumption of normal disturbances, the conditional distribution of $\boldsymbol{\omega}_{t+1,t+H}$ is multivariate normal,

$$\boldsymbol{\omega}_{t+1,t+H} \,|\, (\mathbf{Y}_t^o, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, A) \sim N\,(\underset{NH\times 1}{\boldsymbol{\mu}}, \underset{NH\times NH}{\boldsymbol{\Sigma}_{\boldsymbol{\omega}}}), \tag{3.2}$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is the variance-covariance matrix of regression disturbances, as defined in Chapter 2. The analytical closed-form expressions for the mean vector $\boldsymbol{\mu}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\omega}}$ can be derived as follows.

(a) *Mean Vector.* The conditional expectation of the vector of the hth-step-ahead forecasts conditional on the history $\mathbf{Y}_t^o$ can be denoted

$$\boldsymbol{\mu}_h = \begin{cases} E[\mathbf{y}_{t+h} \,|\, \mathbf{Y}_t^o] & \text{if } h > 0, \\ \mathbf{y}_{t+h}^o & \text{if } h \leq 0. \end{cases} \tag{3.3}$$

Starting with $h = 0$, we have

$$\boldsymbol{\mu}_0 = \mathbf{y}_t^o.$$

Then at $h = 1$, we have

$$\begin{aligned}\boldsymbol{\mu}_1 &= E[\mathbf{y}_{t+1} \,|\, \mathbf{Y}_t^o] \\ &= E[\mathbf{C} + \mathbf{A}_1\mathbf{y}_t + \ldots + \mathbf{A}_p\mathbf{y}_{t+1-p} + \boldsymbol{\varepsilon}_{t+1} \,|\, \mathbf{Y}_t^o] \\ &= \mathbf{C} + \mathbf{A}_1\boldsymbol{\mu}_0 + \ldots + \mathbf{A}_p\boldsymbol{\mu}_{1-p}.\end{aligned}$$

Similarly at $h = 2$, we have

$$\begin{aligned}\boldsymbol{\mu}_2 &= E[\mathbf{y}_{t+2} \,|\, \mathbf{Y}_t^o] \\ &= E[\mathbf{C} + \mathbf{A}_1\mathbf{y}_{t+1} + \ldots + \mathbf{A}_p\mathbf{y}_{t+2-p} + \boldsymbol{\varepsilon}_{t+2} \,|\, \mathbf{Y}_t^o] \\ &= \mathbf{C} + \mathbf{A}_1\boldsymbol{\mu}_1 + \ldots + \mathbf{A}_p\boldsymbol{\mu}_{2-p}.\end{aligned}$$

Recursively, we know

$$\boldsymbol{\mu}_h = \mathbf{C} + \sum_{l=1}^{p} \mathbf{A}_l\boldsymbol{\mu}_{h-l}, \quad \text{for } h = 1, \ldots, H, \tag{3.4}$$

where $\mathbf{A}_l$ is the coefficient matrix of the lth lag in (2.1).

(b) *Variance-Covariance Matrix.* The $(N \times N)$ conditional variance-covariance matrix of the ith-step-ahead and jth-step-ahead forecasts can be denoted

$$\boldsymbol{\Gamma}_{j,i} = cov[\mathbf{y}_{t+j}, \mathbf{y}_{t+i} \,|\, \mathbf{Y}_t^o] = E[(\mathbf{y}_{t+j} - \boldsymbol{\mu}_j)(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)' \,|\, \mathbf{Y}_t^o]. \tag{3.5}$$

By symmetry, we have $\boldsymbol{\Gamma}_{j,i} = \boldsymbol{\Gamma}'_{i,j}$.

By using (3.4), we know

$$\mathbf{y}_{t+j} - \boldsymbol{\mu}_j = \sum_{l=1}^{\min(j-1,p)} \mathbf{A}_l(\mathbf{y}_{t+j-l} - \boldsymbol{\mu}_{j-l}) + \boldsymbol{\varepsilon}_{t+j}. \tag{3.6}$$

(i) For $i < j$, we postmultiply both sides of (3.6) by $(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)'$ and get the

conditional expectation

$$\begin{aligned}\boldsymbol{\Gamma}_{j,i} &= E[(\mathbf{y}_{t+j} - \boldsymbol{\mu}_j)(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)' \,|\, \mathbf{Y}_t^o] \\ &= E\left[\sum_{l=1}^{\min(j-1,p)} \mathbf{A}_l(\mathbf{y}_{t+j-l} - \boldsymbol{\mu}_{j-l})(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)' + \boldsymbol{\varepsilon}_{t+j}(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)' | \mathbf{Y}_t^o\right] \\ &= \sum_{l=1}^{\min(j-1,p)} \mathbf{A}_l \boldsymbol{\Gamma}_{j-l,i}. \end{aligned} \tag{3.7}$$

Notice that $E\left[\boldsymbol{\varepsilon}_{t+j}(\mathbf{y}_{t+i} - \boldsymbol{\mu}_i)' | \mathbf{Y}_t^o\right] = 0$, because $i < j$.

(ii) For $i = j$, we have

$$\boldsymbol{\Gamma}_{j,j} = \sum_{l=1}^{min(j-1,p)} \mathbf{A}_l \boldsymbol{\Gamma}_{j-l,j} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}. \tag{3.8}$$

In summary, the vector of the multi-H-step-ahead forecasts $\boldsymbol{\omega}_{t+1,t+H}$ defined in (3.1) has a conditional multivariate normal distribution as stated in (3.2). The $(NH \times 1)$ mean vector $\boldsymbol{\mu}$ can be expressed

$$\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_H \end{bmatrix},$$

where the conditional expectation of the vector of the single-hth-step-ahead forecasts $\boldsymbol{\mu}_h$ $(h = 1, \ldots, H)$ is defined in (3.4). The $(NH \times NH)$ variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\omega}}$ can be expressed

$$\begin{bmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} & \ldots & \boldsymbol{\Gamma}_{1,H} \\ \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} & \ldots & \boldsymbol{\Gamma}_{2,H} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Gamma}_{H,1} & \boldsymbol{\Gamma}_{H,2} & \ldots & \boldsymbol{\Gamma}_{H,H} \end{bmatrix},$$

where

(a) for $i = j = 1$, $\mathbf{\Gamma}_{1,1} = \mathbf{\Sigma}_{\varepsilon}$, and

(b) for $j = 2, 3, \ldots, H$,

(i) $\mathbf{\Gamma}_{ji}$ and $\mathbf{\Gamma}_{ij} = \mathbf{\Gamma}'_{ji}$ are defined in (3.7) successively for $i = 1, \ldots, j-1$, and

(ii) $\mathbf{\Gamma}_{jj}$ is defined in(3.8).

Furthermore, we can use the property of a multivariate normal distribution,

$$\mathbf{R}\boldsymbol{\omega}_{t+1,t+H} \,|\, (\mathbf{Y}_t^o, \boldsymbol{\beta}, \mathbf{\Sigma}_{\varepsilon}, A) \sim N\left(\mathbf{R}\boldsymbol{\mu},\, \mathbf{R}\mathbf{\Sigma}_{\boldsymbol{\omega}}\mathbf{R}'\right), \tag{3.9}$$

to have conditional distributions of any linear combinations of forecasts by appropriately manipulating the $\mathbf{R}$ matrix in (3.9).

For the purpose of illustration, we consider a vector of multi-4-step-ahead forecasts from a two-variable VAR model, that is,

$$\underset{8\times 1}{\boldsymbol{\omega}_{t+1,t+4}} = \begin{bmatrix} \mathbf{y}_{t+1} \\ \mathbf{y}_{t+2} \\ \vdots \\ \mathbf{y}_{t+4} \end{bmatrix}.$$

We are able to investigate various cases by constructing different $\mathbf{R}$ matrices as follows.

- When multi-4-step-ahead forecasts of the two variables are of interest, we set $\mathbf{R}$ to be an 8×8 identity matrix.

- When single-4-step-ahead forecasts of the two variables are of interest, we set

$$\underset{2\times 8}{\mathbf{R}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

- When multi-4-step-ahead forecasts of the first of the two variables are of interest, we set

$$\underset{4\times 8}{\mathbf{R}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

- When the sum of multi-4-step-ahead forecasts of the first of the two variables is of interest, we set

$$\underset{1\times 8}{\mathbf{R}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

By manipulating the $\mathbf{R}$ matrix as illustrated above, we can have a closed-form expression of the conditional distribution as described in (3.9). Therefore, it is straightforward to compute the predictive probability density of interest[1],

$$M^{-1}\sum_{m=1}^{M} p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \mid \mathbf{Y}^{o}_{t}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Sigma}^{(m)}_{\boldsymbol{\varepsilon}}, A\right) \overset{a.s.}{\to} p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \mid \mathbf{Y}^{o}_{t}, A\right), \tag{3.10}$$

where $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\Sigma}^{(m)}_{\boldsymbol{\varepsilon}}$ are the draws obtained from the posterior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ at the mth iteration of the posterior simulation. The predictive probability density is computed using the observed values $\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H}$, in order to evaluate the performance of a prediction model.

3.2 Log Predictive Scores of Models

After obtaining the predictive probability densities, we proceed to compute the conventional log predictive score, which can be used for evaluating the forecasting performance of competing models. Specifically, for a prediction model A and a

[1]Hereafter, the almost sure convergence is in M.

vector of observed time series $\mathbf{Y}_T^o$, the conventional log predictive score function can be written

$$LS(\mathbf{Y}_T^o, A) = \sum_{t=T_0}^{T-H} \log p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o,\, A\right). \tag{3.11}$$

The first predictive density evaluation starts at T_0+1 conditional on $\mathbf{Y}_{T_0}^o$ (for some $T_0 < T$). The second predictive density evaluation starts at $(T_0+1)+1$ conditional on $\mathbf{Y}_{T_0+1}^o$. There are $T-T_0-H+1$ predictive probability densities to be evaluated. A higher log predictive score generally indicates a better forecasting performance than a lower one.

Two major forecasting scenarios are examined. First, when multi-H-step-ahead forecasts of N variables are of interest, we set $\mathbf{R}$ in (3.11) to be an $(NH \times NH)$ identity matrix. Then (3.11) becomes

$$LS(\mathbf{Y}_T^o, A) = \sum_{t=T_0}^{T-H} \log p\left(\mathbf{y}_{t+1}^o,\, \ldots,\, \mathbf{y}_{t+H}^o \mid \mathbf{Y}_t^o,\, A\right). \tag{3.12}$$

Second, when single-H-step-ahead forecasts of N variables are of interest, we set $\mathbf{R}$ in (3.11) to be an $(N \times NH)$ matrix, i.e., $[\mathbf{0}, \ldots, \mathbf{0}, \mathbf{I}_N]$. Then (3.11) becomes

$$LS(\mathbf{Y}_T^o, A) = \sum_{t=T_0}^{T-H} \log p\left(\mathbf{y}_{t+H}^o \mid \mathbf{Y}_t^o,\, A\right). \tag{3.13}$$

A generalized algorithm for computing the log predictive score of a prediction model consists of two steps.

(1) Compute a numerical approximation of the predictive probability density $p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o,\, A\right)$ at *each* of $t = T_0 : (T-H)$ evaluations, following Steps (a) through (c) below.

   (a) Draw vectors of regression coefficients $\boldsymbol{\beta}^{(m)}$ $(m = 1, \ldots, M)$ and precision matrices of disturbances $\mathbf{H}^{(m)}$ $(m = 1, \ldots, M)$ from their posterior distributions,

- $p\left(\boldsymbol{\beta} \mid \mathbf{H}, \mathbf{Y}_t^o, \mathbf{Z}, A\right) \sim N\left(\overline{\boldsymbol{\beta}}, \overline{\mathbf{H}}_{\boldsymbol{\beta}}^{-1}\right)$ as described in (2.10), and
- $p\left(\mathbf{H} \mid \boldsymbol{\beta}, \mathbf{Y}_t^o, \mathbf{Z}, A\right) \propto p\left(\boldsymbol{\beta} | \mathbf{H}, A\right) p\left(\mathbf{H} \mid A\right) p\left(\mathbf{Y}_t^o \mid \boldsymbol{\beta}, \mathbf{H}, \mathbf{Z}, A\right)$ as described in (2.13).

(b) Evaluate the conditional predictive probability density

$$p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o, \boldsymbol{\beta}^{(m)}, \mathbf{H}^{(m)}, A\right),$$

by using each pair of $\boldsymbol{\beta}^{(m)}$ and $\mathbf{H}^{(m)}$ (for $m = 1 : M$) according to the closed-form conditional normal distribution of the forecasts, which is characterized by (3.4), (3.7), and (3.8).

(c) Compute the numerical approximation

$$M^{-1} \sum_{m=1}^{M} p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o, \boldsymbol{\beta}^{(m)}, \mathbf{H}^{(m)}, A\right) \stackrel{a.s.}{\to} p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o, A\right),$$

where M is the total number of iterations of the posterior simulator.

In summary, Steps (a) through (c) are repeated for $T - T_0 - H + 1$ evaluations of predictive probability densities. Each time the posterior simulator is executed to obtain M pairs of $\boldsymbol{\beta}^{(m)}$ and $\mathbf{H}^{(m)}$, then a numerical approximation of the predictive probability density is computed.

(2) Compute a numerical approximation of the log predictive score defined in (3.11). That is,

$$\sum_{t=T_0}^{T-H} \log\left(M^{-1} \sum_{m=1}^{M} p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o, \boldsymbol{\beta}^{(m)}, \mathbf{H}^{(m)}, A\right)\right) \stackrel{a.s.}{\to} \sum_{t=T_0}^{T-H} \log p\left(\mathbf{R}\boldsymbol{\omega}_{t+1,t+H}^o \mid \mathbf{Y}_t^o, A\right). \quad (3.14)$$

3.3 Optimal Weights of Models

After getting the predictive probability densities from J different models, that is,

$$p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right) \quad (j = 1, \ldots, J), \tag{3.15}$$

we consider an optimal linear combination of these predictive probability densities in the form of

$$\sum_{j=1}^{J} w_j p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right); \quad \sum_{j=1}^{J} w_j = 1;\ w_j \geq 0\ (j = 1, \ldots, J). \tag{3.16}$$

To get the optimal weight w_j for the prediction model A_j $(j = 1, \ldots, J)$, we maximize the log predictive score function,

$$f_{T-T_0-H+1}(\mathbf{w}) = \sum_{t=T_0}^{T-H} \log\left[\sum_{j=1}^{J} w_j p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right)\right], \tag{3.17}$$

which is concave in $w_2, \ldots, w_J$. Let $w_1 = 1 - \sum_{j=2}^{J} w_j$, we then have

$$\frac{\partial f_{T-T_0-H+1}(\mathbf{w})}{\partial w_i} = \sum_{t=T_0}^{T-H} \frac{p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_i\right) - p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_1\right)}{\sum_{j=1}^{J} p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right)}, \tag{3.18}$$

where $i = 2, \ldots, J$, and

$$\frac{\partial^2 f_{T-T_0-H+1}(\mathbf{w})}{\partial w_i \partial w_k} = -\sum_{t=T_0}^{T-H} \left[\left(\frac{p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_i\right) - p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_1\right)}{\sum_{j=1}^{J} p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right)}\right)\right.$$

$$\left. \times \left(\frac{p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_k\right) - p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_1\right)}{\sum_{j=1}^{J} p\left(\mathbf{R}\boldsymbol{\omega}^{o}_{t+1,t+H} \,|\, \mathbf{Y}^{o}_{t}, A_j\right)}\right)\right], \tag{3.19}$$

where $i, k = 2, \ldots, J$. The resulting $(J-1) \times (J-1)$ Hessian matrix is negative definite[2], and the computation of the maximum is a conventional convex programming

[2]It is still possible that the Hessian matrix may only be negative semi-definite for some priors and for nested models.

problem. The strategy is using past performance of the pool as a training sample to find the optimal weights that maximize (3.17). That is, the real-time optimal weights are determined recursively at each date t based on information through $t-1$. As indicated in (3.17), the first sample consists of data through T_0 ($T_0 < T$), the second sample consists of data through $T_0 + 1$, and the last sample is through $T - H$.

3.4 Real Time Performance of Models

For each model j that has a positive weight in an optimal linear prediction pool of J models, we are interested in finding a set of time periods such that the optimal weight of model j will become zero, if the predictive probability densities of those time periods are removed from the summation in (3.17). A searching scheme for model j is as follows.

(1) Given a set of $T-T_0-H+1$ predictive probability densities of model j, compute the ratio of its predictive probability density to that of the optimal prediction pool, that is,

$$\frac{p\left(\mathbf{R}\boldsymbol{\omega}^o_{t+1,t+H} \mid \mathbf{Y}^o_t, A_j\right)}{\sum_{j=1}^{J} w_j^* p\left(\mathbf{R}\boldsymbol{\omega}^o_{t+1,t+H} \mid \mathbf{Y}^o_t, A_j\right)} \quad (t = T_0, \ldots, T-H), \tag{3.20}$$

where w_j^*'s (for $j = 1, \ldots, J$) are the optimal weights. Sort the $T - T_0 - H + 1$ ratios in the descending order and identify the time period t of the highest ratio.

(2) Remove the predictive probability density corresponding to the time period identified in Step (1), and re-compute the optimal weights for the J models as well as the optimal log predictive score based on the $(T - T_0 - H + 1) - 1$ predictive probability densities.

(3) Repeat Steps (1) and (2) until the optimal weight of model j is technically zero, i.e., below a certain small value (e.g., 10^{-7}).

The predictive probability densities, which are identified for model j by going through the above search scheme, explain why the model j is given a positive weight in the optimal prediction pool.

CHAPTER 4
PREDICTING U.S. OUTPUT, INFLATION, AND INTEREST RATE

We apply the full Bayesian autoregressive (BVAR) model developed in Chapter 2 and the optimal pooling techniques described in Chapter 3 to predicting output, inflation, and interest rate in the United States. The predictions are made for one-step-ahead, multi-4-step-ahead, and single-4-step-ahead. Throughout this exercise and the one in the next chapter, we are interested in two forecasting questions. First, is it necessary to tailor prior tightness by targets and across forecast horizons for improving prediction accuracy? Second, how much improvement over a single best model can be achieved by a linear prediction pool of models? Specifically, such a linear prediction pool under consideration is constructed in three ways, by using equal weights, optimal weights based on the full sample, and continuously updated optimal weights based on only past realizations. An investigation to these questions helps build better prediction models.

This chapter is organized as follows. Section 4.1 describes the data set. Section 4.2 specifies six BVAR models that differ in their prior tightness. Section 4.3 summarizes and interprets the empirical findings.

4.1 Data

The data are from Smets and Wouters (2007), which consist of U.S. time series of quarterly output, inflation, and interest rate, from 1947.Q1 to 2004.Q4. The dependent variables in the BVAR models are defined as follows.

(a) Output = LN(GDPC96/LNSindex), where GDPC96 is seasonally adjusted real

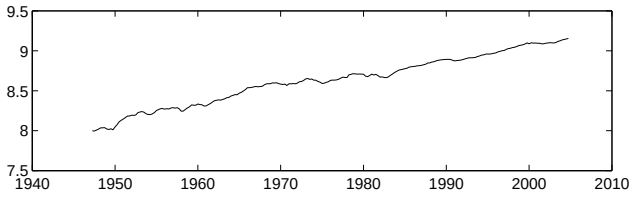
gross domestic product (billions of chained 1996 dollars), and LNSindex is civilian noninstitutional population (age: 16 years and older) index (1992.Q3=1).

(b) Inflation$_t$ = LN(GDPDEF$_t$/GDPDEF$_{t-1}$)×100, where GDPDEF$_t$ is seasonally adjusted implicit price deflator of GDP$_t$, and GDPDEF$_{t-1}$ is GDPDEF of one-quarter ago.

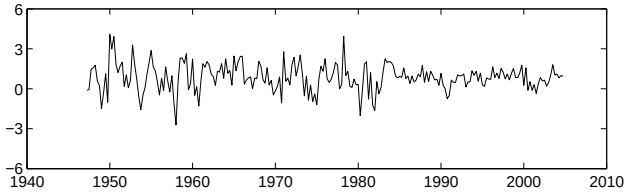
(c) Interest Rate = Federal Funds Rate/4, where Federal Funds Rate is expressed in percentage points.

Figure 4.1 plots four time series. Output displays an almost linear upward trend. Output growth rate, which is computed using the standard percentage change formula, i.e., (GDPC96$_t$/GDPC96$_{t-1}$ − 1)×100, shows fluctuations that appear to be consistent with the U.S. business cycles overall. Inflation is essentially the first difference of natural logarithms of GDP deflators. Although it appears to follow a random walk most of the time, it still shows several substantial jumps around 1951.Q1, 1974.Q4, and 1980.Q4. Interest rate also has several major ups and downs over the five decades. It is usually higher at the peak of a business cycle (e.g., 3.76 at 1980.Q1 and 4.39 at 1981.Q3) and lower at the trough (e.g., 2.46 at 1980.Q3 and 2.32 at 1982.Q4). To explore the data variation in terms of economic expansions and contractions, Table 4.1 lists the peaks and troughs of the U.S. business cycles[1] over the sample period.

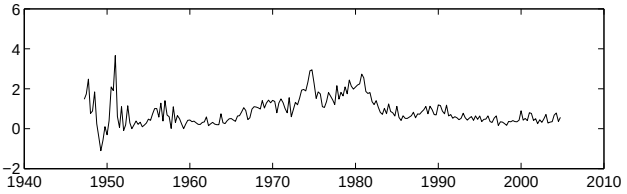
[1]The data are from the National Bureau of Economic Research.



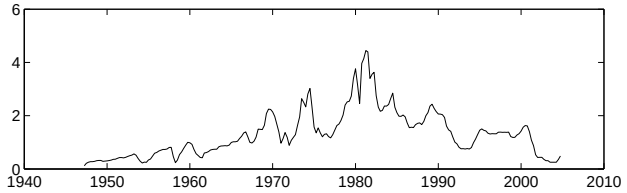
(a) Output



(b) Output Growth Rate (%)



(c) Inflation (%)



(d) Interest Rate (%)

Figure 4.1: U.S. Output, Output Growth Rate, Inflation, and Interest Rate

Table 4.1: U.S. Business Cycles (1947-2004)

Peak	Trough
1948.Q4	1949.Q4
1953.Q2	1954.Q2
1957.Q3	1958.Q2
1960.Q2	1961.Q1
1969.Q4	1970.Q4

Peak	Trough
1973.Q4	1975.Q1
1980.Q1	1980.Q3
1981.Q3	1982.Q4
1990.Q3	1991.Q1
2001.Q1	2001.Q4

4.2 Model Specification

The prior distribution for the precision matrix of regression disturbances (i.e., $\mathbf{H}$) is Wishart as described in (2.6). Since the parameterization of (2.6) implies (see, e.g., Zellner, 1971; Press, 1982, for details)

$$E(\mathbf{H}) = \underline{\nu}\underline{\mathbf{S}}^{-1}, \tag{4.1}$$

$$E(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}) = \frac{\underline{\mathbf{S}}}{\underline{\nu} - N - 1} = \frac{\underline{\mathbf{S}}}{\underline{\nu} - 3 - 1}, \tag{4.2}$$

the scale matrix $\underline{\mathbf{S}}$ is set to an estimated residual variance-covariance matrix from a seemingly unrelated regression using the data through 1965.Q4, and the degrees of freedom parameter $\underline{\nu}$ is set to 5. On the one hand, this combination of $\underline{\mathbf{S}}$ and $\underline{\nu}$ makes the expectation of the variance-covariance matrix of disturbances equal to the estimated residual variance-covariance matrix. On the other hand, the small value of $\underline{\nu}$ allows large prior variation in $\mathbf{H}$, which admits a lack of information and incorporates parameter uncertainty.

The prior distribution for the vector of regression coefficients (i.e., $\boldsymbol{\beta}$) is multivariate normal conditional on $\mathbf{H}$ as described in (2.8) and (2.9). In line with the conventional Minnesota prior, the prior means of the coefficients of the first own lags are set to ones, while the prior means of the other coefficients are set to zeros. The prior standard deviations of the coefficients of the intercept terms (κ) are loosened to 10^2 to allow for large variation. As for the overall tightness (λ) and cross-equation tightness (θ), Doan et al. (1984) and Litterman (1986) recommended to set λ equal to 0.2, and Dua and Ray (1995) recommended to set θ equal to 0.9, when the prediction targets are correlated macroeconomic variables. In this exercise, to investigate the relation between prior tightness and prediction accuracy for different targets across forecast horizons, we consider six combinations of λ and

θ, which represent typical cases of interest, as follows.

(1) $\lambda = 10^2$ and $\theta = 10^{-4}$.

When large prior standard deviations (λ/l) are set for the coefficients of the own lags, and small prior standard deviations ($\frac{\lambda\theta}{l}\frac{\sigma_i}{\sigma_j}$) are set for the coefficients of the cross lags, the equations in the VAR system become similar to univariate autoregressive equations.

(2) $\lambda = 10^2$ and $\theta = 1$.

When large prior standard deviations are set for all the lag coefficients, the VAR system becomes unrestricted.

(3) $\lambda = 0.5$ and $\theta = 0.9$.

(4) $\lambda = 0.2$ and $\theta = 0.9$.

Combinations (3) and (4) fix θ but vary λ. As λ decreases from 0.5 to 0.2, the prior standard deviations for all the lag coefficients are tightened up, consequently the lag coefficients are shrunk towards their prior means. In other words, a smaller λ assumes that the "true" values of the lag coefficients are less likely to deviate far from the prior guesses (i.e., prior means).

(5) $\lambda = 0.2$ and $\theta = 0.6$.

(6) $\lambda = 0.2$ and $\theta = 0.1$.

Combinations (4), (5), and (6) fix λ but vary θ. As θ decreases from 0.9 to 0.6 to 0.1, only the coefficients of the cross lags are shrunk towards their prior zero means. In other words, a smaller θ assumes that the variables are less related.

4.3 Empirical Results

In this exercise, six trivariate BVAR models are constructed by using the combinations of λ and θ described above. Each model is estimated with four lags of each of the three variables. The forecast horizons under consideration are one-step-ahead, multi-4-step-ahead, and single-4-step-ahead. Prediction performance of a model is evaluated through its log predictive score, which is the sum of the logarithms of predictive densities. In each prediction, the first predictive density evaluation always starts from 1966.Q1. There are 156 predictive probability densities in the one-step-ahead prediction, and 153 predictive probability densities in each case of the multi-4-step-ahead and single-4-step-ahead predictions. Moreover, there are two scenarios: the target being the vector of the three variables (i.e., output, inflation, and interest rate); and the target being each of the three variables individually. The analysis focuses on two forecasting concerns as follows.

The first concern relates to the selection of prior tightness for a BVAR model. Intuitively, prior tightness should be tailored to different target variables and forecast horizons. A looser prior probably works better than a tighter one for a more volatile target and a longer forecast horizon, because it allows more variation in parameter estimation. In this exercise, since the data of output, inflation, and interest rate display apparently different patterns over time, and the forecast horizons considered vary from one-step-ahead to multi-4-step-ahead and single-4-step-ahead, it is of interest to examine the effect of prior tightness on prediction accuracy in separate cases.

The second concern relates to the performance of linear prediction pools. A simple way of constructing a linear prediction pool is to assign equal weights to the predictive densities of competing models. However, to optimize the performance of

a linear prediction pool, an effective way is to use the optimal weights that maximize the log predictive score of the pool under consideration[2]. From the mathematics of optimization, the log predictive score of an optimal pool, which is constructed by using the optimal weights computed based on the full sample[3], cannot be lower than that of its best constituent model. However, such a weighting scheme is not applicable in practice, because only past data are available for optimization. A sensible alternative is to continuously re-compute the optimal weights at each time period by using only past realizations[4]. Since computational complexity increases from calculating equal weights to optimal weights, it is of interest to examine how much improvement over a single best model can be achieved by two different linear prediction pools, one with equal weights and the other with real-time optimal weights.

To address the above two concerns, the performance of six individual models and three linear prediction pools are evaluated in terms of log predictive score. At the beginning, one of the six models with the highest log predictive score is identified as the "best" model. Then the contribution of each individual model to an optimal prediction pool is examined, in terms of its optimal weight and supporting

[2]Such an optimization can be conveniently done in a general computing software, e.g., MATLAB.

[3]The full-sample optimal weights are computed only once. For example, computation of the optimal weights for the one-step-ahead prediction uses all the 156 predictive densities at one time.

[4]For the first evaluation, the real-time optimal weights are set to 1/6. For the second evaluation, the real-time optimal weights are computed based on the observations of the past two periods, and so on.

observations[5]. The main purpose is to determine the "best" prior tightness for each target at each forecast horizon. To investigate the potential advantage of constructing an optimal prediction pool, two comparisons are made. One comparison is to see whether a real-time optimal prediction pool can outperform a single best model. The other is to see whether a real-time optimal prediction pool can outperform an equally-weighted[6] prediction pool.

Four major findings are summarized as follows.

- A larger overall tightness (λ) works better than a smaller one for predicting interest rate. This can be seen through three perspectives. First, it is through the λ value of the best individual model. For example, in the one-step-ahead prediction, the best single model for predicting interest rate has its λ value equal to 10^2 (Table 4.5a). Second, the relative competitiveness of a model with a larger λ can be seen through a two-model optimal prediction pool. For example, two models with ($\lambda = 10^2$) and ($\lambda = 0.2$) are paired in the multi-4-step-ahead prediction of interest rate. Figure 4.3 portrays the log predictive score of this two-model pool as a function of model weights. It can be seen that the log predictive score keeps increasing, as the weight of the model with ($\lambda = 10^2$) increases, and eventually reaches the maximum value when the weight of the model with ($\lambda = 10^2$) reaches about 0.7. This indicates that the model with ($\lambda = 10^2$) is more competitive than the model with ($\lambda = 0.2$) for predicting interest rate. Third, the relative competitiveness of a

[5]The supporting observations account for each model's admittance to the full-sample optimal pool. That is, if such supporting observations were removed, then the model would not have a positive weight in the full-sample optimal pool.

[6]Note that equal weights are placed on the predictive densities of the competing models, not on their log predictive scores.

model with a larger λ can also be seen through a real-time optimal prediction pool of multiple models. For example, in the multi-4-step-ahead prediction of interest rate (Table 4.9b), the sum of the average real-time optimal weights of the two models with ($\lambda = 10^2$) is about 0.8, which is higher than 0.2, the sum of the average real-time optimal weights of the three models with ($\lambda = 0.2$). By contrast, a smaller λ works better than a larger one for predicting output and inflation, which can be seen through the same perspectives.

- The advantage of using a larger λ for predicting interest rate becomes more noticeable in the single-4-step-ahead prediction than in the one-step-ahead prediction. In the one-step-ahead prediction (Table 4.5b), the sum of the average real-time optimal weights of the two models with ($\lambda = 10^2$) is about 0.8. In the single-4-step-ahead prediction (Table 4.13b), the sum of the average real-time optimal weights of these two models is close to 1.0. Correspondingly, the advantage of using a smaller λ for predicting output and inflation also becomes more noticeable at a longer forecast horizon. For example, in the single-4-step-ahead prediction of output (Table 4.11b), the optimal weights of the models with ($\lambda = 10^2$) and ($\lambda = 0.5$) technically become zeros in both full-sample and real-time optimal pools.

- When the target is the vector of the three variables, a linear prediction pool of models with different levels of prior tightness outperforms a single best model, even the prediction pool is simply constructed by using equal weights instead of optimal weights. This can be seen from Tables 4.2, 4.6, and 4.10 for the predictions at the three forecast horizons. However, when the target is one of the three variables, an equally-weighted prediction pool does not necessarily

outperform a single best model. This can be seen from the one-step-ahead prediction of inflation (Table 4.4), multi-4-step-ahead predictions of output and inflation (Tables 4.7 and 4.8), and single-4-step-ahead predictions of output, inflation, and interest rate (Tables 4.11, 4.12, and 4.13). Moreover, the disadvantage of an equally-weighted prediction pool becomes more noticeable at a longer forecast horizon. For example, in the one-step-ahead prediction of output, the equally-weighted prediction pool outperforms the best single model by about 1 point (Table 4.3). However, in the single-4-step-ahead prediction of output, the equally-weighted prediction pool underperforms the best single model by about 7 points (Table 4.11).

- A real-time optimal prediction pool outperforms its single best constituent model alone and the equally-weighted prediction pool in all the predictions in this exercise. Moreover, the log predictive score of a real-time optimal pool is much higher than that of a single best model, especially when the real-time optimal weights converge more quickly. For example, in the one-step-ahead prediction, the real-time optimal weights in the interest rate prediction appear to converge more quickly than those weights in the inflation prediction (Figure 4.2). In the interest rate prediction, the log predictive score of the real-time optimal pool is about 30% higher than that of the single best model (Table 4.5). By contrast, in the inflation prediction, the log predictive score of the real-time optimal pool is only 3% higher than that of the single best model (Table 4.4).

The above findings have two implications on prior specification of BVAR models and performance of linear prediction pools.

- When future values of target variables are likely to be volatile and deviate from their most recent historical values, a larger λ works better than a smaller one, and vice versa. This is because a larger λ definitely loosens up the prior standard deviations of all the lag coefficients, which in turn allows the lag coefficients to deviate far from their prior means.

- When the target is a vector of several variables and a multi-H-step-ahead prediction is of interest, a linear prediction pool of models with different λ values can outperform a single best model. This is because a combination of prior tightness can accommodate various situations more easily than a single level of prior tightness.

These implications are supported by the data characteristics that also explain why the findings arise from this exercise.

- Interest rate appears more volatile than the other two variables. For example, the standard deviation of output is about 0.2 for the predictive density evaluation period from 1966.Q1 to 2004.Q4, while the standard deviation of interest rate is about 0.8 over the same period. For another example, the values of output at 1981.Q2 and 1982.Q2 are 8.697 and 8.673 respectively, which are apparently very close; consequently the predictive density of output at 1982.Q2 (i.e., the 4-step-ahead prediction from 1981.Q2) is identified as a supporting observation for a model with ($\lambda = 0.2$), as seen in Table 4.11c. During the same period, the actual interest rate at 1981.Q2 is 4.45, which is actually the highest value over the five decades (1947-2004), and then decreases by 18.4% to 3.63 four quarters later at 1982.Q2; consequently the predictive density of interest rate at 1982.Q2 (i.e., the 4-step-ahead prediction from 1981.Q2) is

identified as a supporting observation for a model with ($\lambda = 10^2$), as indicated in Table 4.13c.

- The advantage of using a combination of prior tightness is so salient for predicting a vector of variables that even a prediction model with equal weights can outperform a single best model. For example, in the single-4-step-ahead prediction of the vector, four models with their λ values ranging from 10^2 to 0.5 to 0.2 have their average real-time optimal weights greater than 0.01, and the equally-weighted prediction pool outperforms the single best model (Table 4.10). By contrast, in the single-4-step-ahead prediction of output, since only the two models with ($\lambda = 0.2$), out of the six models, contribute more substantively to the prediction than the other models, the equally-weighted prediction pool does worse than the single best model (Table 4.11).

Table 4.2: Log Scores for One-step-ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	374.0
$\lambda = 10^2 \quad \theta = 1$	370.4
$\lambda = 0.5 \quad \theta = 0.9$	375.1
$\lambda = 0.2 \quad \theta = 0.9$	383.0
$\lambda = 0.2 \quad \theta = 0.6$	383.6
$\lambda = 0.2 \quad \theta = 0.1$	380.7

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	401.5
Full-sample optimal weights	0.332	0.316	0.000	0.000	0.224	0.128	404.7
Average real-time optimal weights	0.446	0.213	0.002	0.093	0.134	0.113	408.8

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1969.2 1971.3 1972.1 1975.1 1980.1
$\lambda = 10^2 \quad \theta = 1$	1971.1 1972.2 1973.1 1974.4 1980.2
$\lambda = 0.2 \quad \theta = 0.6$	1968.4 1971.2 1973.4 1981.2
$\lambda = 0.2 \quad \theta = 0.1$	1981.1

Table 4.3: Log Scores for One-step-ahead Predictions of Output

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	517.8
$\lambda = 10^2 \quad \theta = 1$	520.5
$\lambda = 0.5 \quad \theta = 0.9$	521.9
$\lambda = 0.2 \quad \theta = 0.9$	520.9
$\lambda = 0.2 \quad \theta = 0.6$	520.8
$\lambda = 0.2 \quad \theta = 0.1$	521.1

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	523.0
Full-sample optimal weights	0.000	0.000	0.620	0.000	0.000	0.380	523.8
Average real-time optimal weights	0.102	0.001	0.275	0.050	0.248	0.324	525.6

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.5 \quad \theta = 0.9$	1969.4 1972.2 1974.1 1974.3 1975.1 1977.2 1980.2 1981.2 1981.4
$\lambda = 0.2 \quad \theta = 0.1$	1971.2 1978.2 1981.3

Table 4.4: Log Scores for One-step-ahead Predictions of Inflation

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-68.8
$\lambda = 10^2 \quad \theta = 1$	-78.6
$\lambda = 0.5 \quad \theta = 0.9$	-72.5
$\lambda = 0.9 \quad \theta = 0.2$	-66.4
$\lambda = 0.2 \quad \theta = 0.6$	-65.1
$\lambda = 0.2 \quad \theta = 0.1$	-65.3

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-67.7
Full-sample optimal weights	0.000	0.000	0.000	0.000	0.572	0.428	-64.6
Average real-time optimal weights	0.001	0.001	0.018	0.134	0.311	0.535	-63.4

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.2 \quad \theta = 0.6$	1974.3 1974.4 1979.2 1991.2
$\lambda = 0.2 \quad \theta = 0.1$	1981.2

Table 4.5: Log Scores for One-step-ahead Predictions of Interest Rate

(a) Performance of Single Model

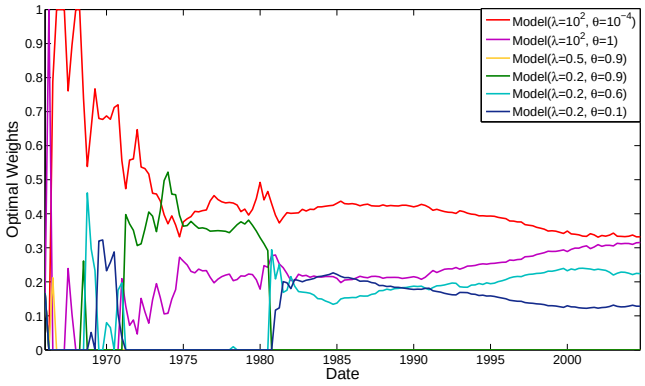
Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-72.2
$\lambda = 10^2 \quad \theta = 1$	-58.1
$\lambda = 0.5 \quad \theta = 0.9$	-68.5
$\lambda = 0.2 \quad \theta = 0.9$	-69.1
$\lambda = 0.2 \quad \theta = 0.6$	-69.4
$\lambda = 0.2 \quad \theta = 0.1$	-72.7

(b) Performance of Linear Pools

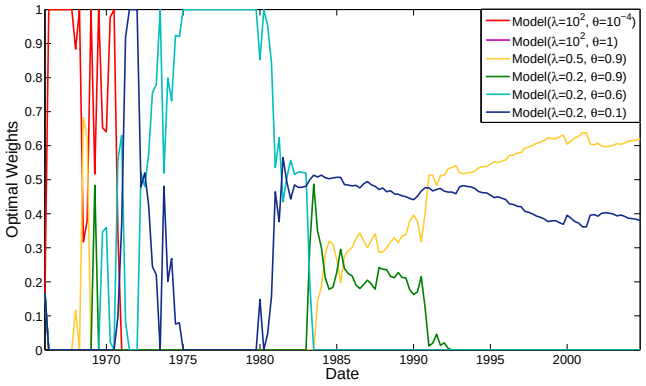
	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-49.6
Full-sample optimal weights	0.228	0.518	0.000	0.254	0.000	0.000	-43.5
Average real-time optimal weights	0.310	0.449	0.001	0.235	0.004	0.001	-40.6

(c) Reasons of Models Being in the Full-sample Optimal Pool

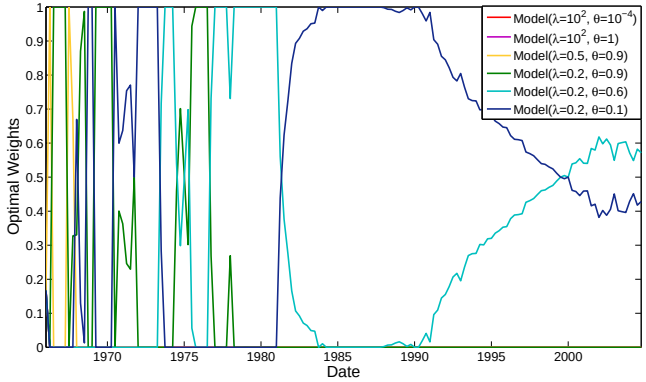
Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1969.2 1972.1
$\lambda = 10^2 \quad \theta = 1$	1971.1 1971.4 1972.2 1973.1 1974.4 1980.4 1982.3
$\lambda = 0.2 \quad \theta = 0.9$	1971.2 1973.4 1981.1 1981.2



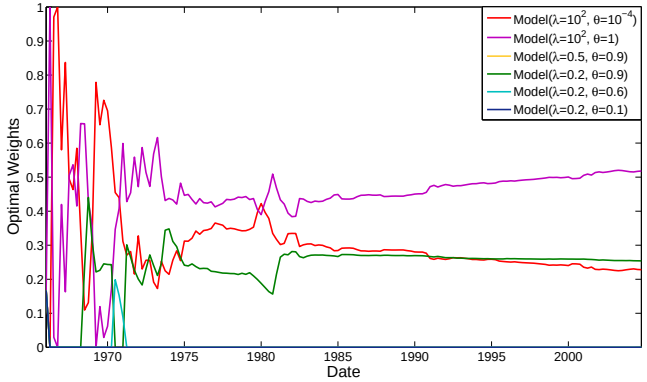
(a) Vector



(b) Output



(c) Output



(d) Output

Figure 4.2: Evolution of Real-time Optimal Weights: One-step-ahead Predictions

Table 4.6: Log Scores for Multi-4-step-ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	1396.2
$\lambda = 10^2 \quad \theta = 1$	1350.0
$\lambda = 0.5 \quad \theta = 0.9$	1439.4
$\lambda = 0.2 \quad \theta = 0.9$	1472.9
$\lambda = 0.2 \quad \theta = 0.6$	1480.6
$\lambda = 0.2 \quad \theta = 0.1$	1470.2

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	1541.3
Full-sample optimal weights	0.378	0.185	0.000	0.176	0.136	0.125	1545.7
Average real-time optimal weights	0.536	0.077	0.001	0.165	0.116	0.106	1550.8

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1966.2 1966.3 1968.4 1969.1 1969.2 1971.3 1974.4 1975.1 1979.1 1979.2 1979.3 1979.4
$\lambda = 10^2 \quad \theta = 1$	1970.2 1974.1 1974.2
$\lambda = 0.2 \quad \theta = 0.9$	1970.4 1971.1 1973.1
$\lambda = 0.2 \quad \theta = 0.6$	1980.2
$\lambda = 0.2 \quad \theta = 0.1$	1980.3 1980.4 1981.1 1981.2

Table 4.7: Log Scores for Multi-4-step-ahead Predictions of Output

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	2029.3
$\lambda = 10^2 \quad \theta = 1$	2022.9
$\lambda = 0.5 \quad \theta = 0.9$	2034.6
$\lambda = 0.2 \quad \theta = 0.9$	2037.1
$\lambda = 0.2 \quad \theta = 0.6$	2036.4
$\lambda = 0.2 \quad \theta = 0.1$	2039.4

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	2038.1
Full-sample optimal weights	0.000	0.000	0.197	0.249	0.000	0.554	2041.2
Average real-time optimal weights	0.008	0.001	0.072	0.106	0.330	0.484	2042.7

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.5 \quad \theta = 0.9$	1983.1 1983.2
$\lambda = 0.2 \quad \theta = 0.9$	1972.2 1974.1 1981.4
$\lambda = 0.2 \quad \theta = 0.1$	1970.2 1971.2 1974.4 1980.3 1981.1 1981.3

Table 4.8: Log Scores for Multi-4-step-ahead Predictions of Inflation

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-271.1
$\lambda = 10^2 \quad \theta = 1$	-302.8
$\lambda = 0.5 \quad \theta = 0.9$	-285.0
$\lambda = 0.2 \quad \theta = 0.9$	-263.9
$\lambda = 0.2 \quad \theta = 0.6$	-259.7
$\lambda = 0.2 \quad \theta = 0.1$	-258.4

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-267.3
Full-sample optimal weights	0.000	0.000	0.000	0.000	0.484	0.516	-257.2
Average real-time optimal weights	0.010	0.001	0.008	0.005	0.313	0.663	-255.6

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.2 \quad \theta = 0.6$	1973.3 1973.4 1974.1 1974.2 1974.3 1979.2
$\lambda = 0.2 \quad \theta = 0.1$	1981.1 1981.2 1981.3

Table 4.9: Log Scores for Multi-4-step-ahead Predictions of Interest Rate

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-331.5
$\lambda = 10^2 \quad \theta = 1$	-307.9
$\lambda = 0.5 \quad \theta = 0.9$	-297.1
$\lambda = 0.2 \quad \theta = 0.9$	-292.4
$\lambda = 0.2 \quad \theta = 0.6$	-296.6
$\lambda = 0.2 \quad \theta = 0.1$	-303.0

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-218.8
Full-sample optimal weights	0.164	0.617	0.000	0.005	0.148	0.067	-205.9
Average real-time optimal weights	0.122	0.683	0.005	0.032	0.098	0.059	-202.1

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1968.4 1969.1 1969.2
$\lambda = 10^2 \quad \theta = 1$	1966.4 1967.3 1970.1 1970.2 1971.2 1971.4 1974.1 1974.2 1974.3 1980.1 1982.2 1982.3
$\lambda = 0.2 \quad \theta = 0.9$	1981.1
$\lambda = 0.2 \quad \theta = 0.6$	1980.2 1980.4
$\lambda = 0.2 \quad \theta = 0.1$	1973.3 1980.3

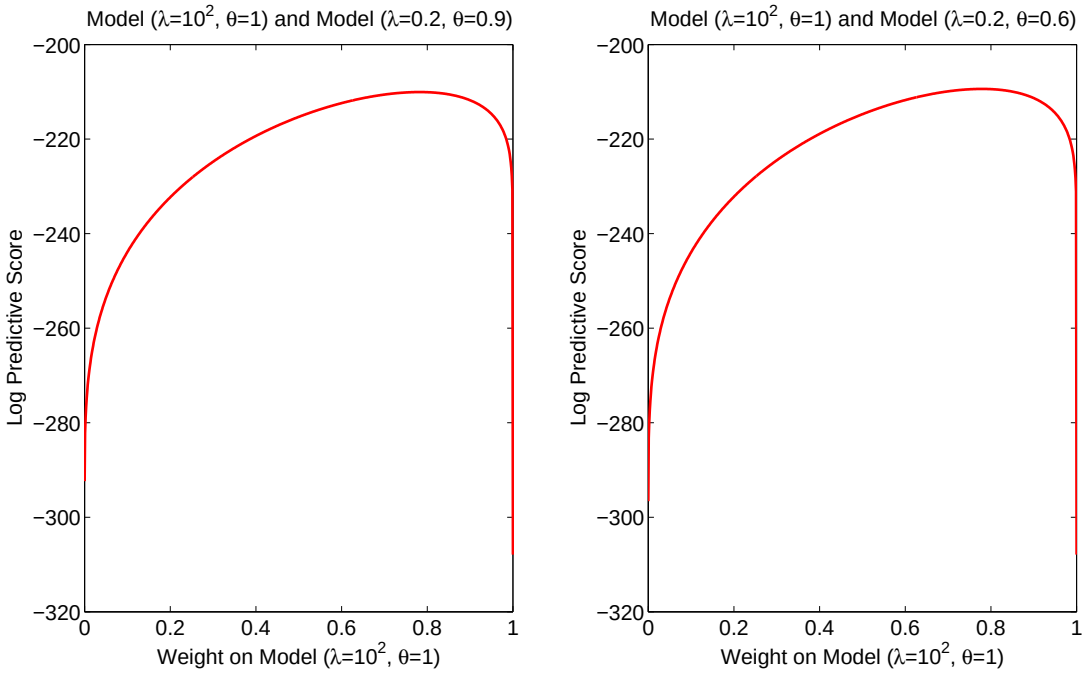
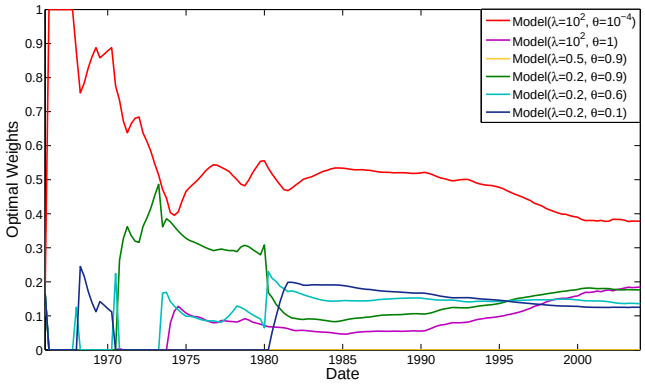
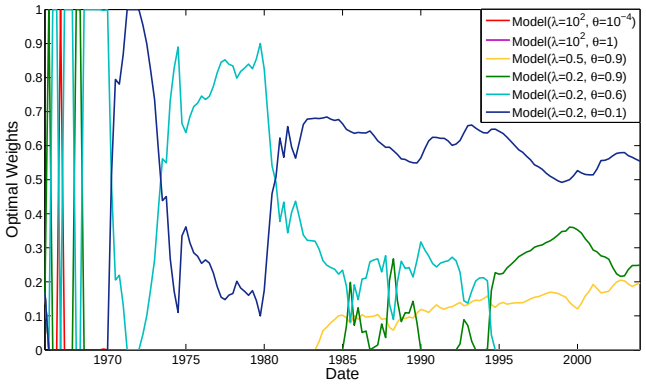


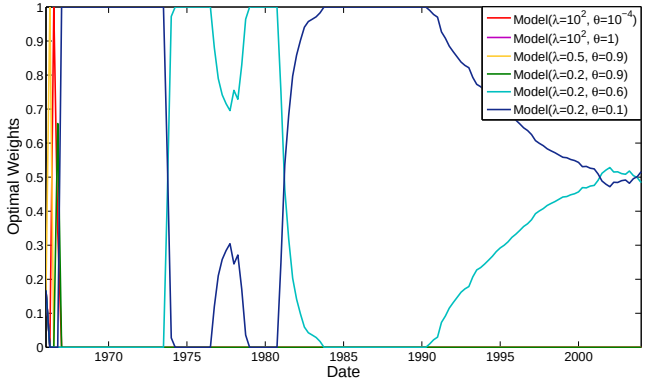
Figure 4.3: Log Predictive Score Function of Model Weight:
Multi-4-step-ahead Prediction of Interest Rate



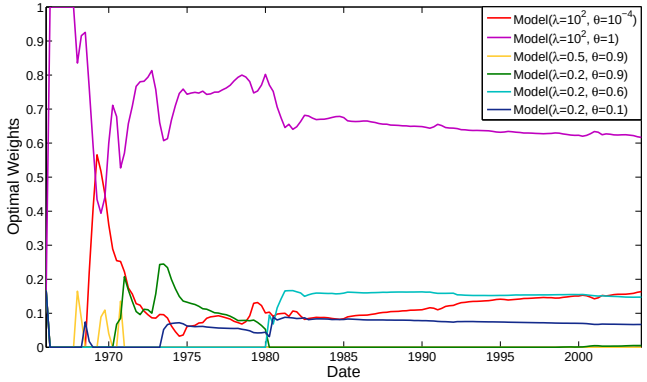
(a) Vector



(b) Output



(c) Output



(d) Output

Figure 4.4: Evolution of Real-time Optimal Weights: Multi-4-step-ahead Predictions

Table 4.10: Log Scores for Single-4-step-ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	41.2
$\lambda = 10^2 \quad \theta = 1$	25.3
$\lambda = 0.5 \quad \theta = 0.9$	22.1
$\lambda = 0.2 \quad \theta = 0.9$	17.3
$\lambda = 0.2 \quad \theta = 0.6$	20.8
$\lambda = 0.2 \quad \theta = 0.1$	32.3

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	47.5
Full-sample optimal weights	0.378	0.456	0.000	0.000	0.000	0.166	52.4
Average real-time optimal weights	0.531	0.342	0.052	0.008	0.001	0.066	54.5

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1968.3 1969.1 1969.2 1979.1 1979.2 1980.4
$\lambda = 10^2 \quad \theta = 1$	1970.1 1970.2 1972.2 1972.3 1973.1 1973.2 1973.4 1974.1 1974.2 1974.3 1978.1 1978.3
$\lambda = 0.2 \quad \theta = 0.1$	1971.2 1981.1 1981.2

Table 4.11: Log Scores for Single-4-step-ahead Predictions of Output

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	356.2
$\lambda = 10^2 \quad \theta = 1$	351.2
$\lambda = 0.5 \quad \theta = 0.9$	356.3
$\lambda = 0.2 \quad \theta = 0.9$	361.0
$\lambda = 0.2 \quad \theta = 0.6$	361.7
$\lambda = 0.2 \quad \theta = 0.1$	369.2

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	362.6
Full-sample optimal weights	0.000	0.000	0.000	0.000	0.119	0.881	369.3
Average real-time optimal weights	0.001	0.001	0.001	0.001	0.295	0.700	370.8

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.2 \quad \theta = 0.6$	1974.1 1981.2
$\lambda = 0.2 \quad \theta = 0.1$	1970.2 1971.2 1974.4 1980.1 1980.2 1980.3 1981.1 1981.3 1982.2 1982.3

Table 4.12: Log Scores for Single-4-step-ahead Predictions of Inflation

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-134.2
$\lambda = 10^2 \quad \theta = 1$	-141.2
$\lambda = 0.5 \quad \theta = 0.9$	-138.7
$\lambda = 0.2 \quad \theta = 0.9$	-131.7
$\lambda = 0.2 \quad \theta = 0.6$	-129.3
$\lambda = 0.2 \quad \theta = 0.1$	-128.2

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-131.0
Full-sample optimal weights	0.000	0.000	0.000	0.000	0.427	0.573	-127.4
Average real-time optimal weights	0.021	0.001	0.001	0.001	0.276	0.700	-126.0

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 0.2 \quad \theta = 0.6$	1973.4 1974.1 1980.1
$\lambda = 0.2 \quad \theta = 0.1$	1981.1 1981.2 1981.3

Table 4.13: Log Scores for Single-4-step-ahead Predictions of Interest Rate

(a) Performance of Single Model

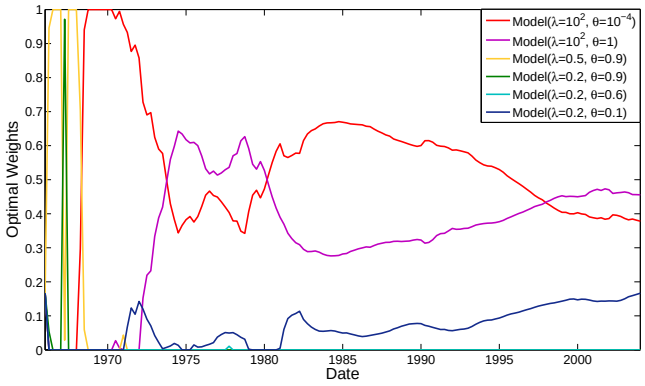
Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	-183.6
$\lambda = 10^2 \quad \theta = 1$	-184.1
$\lambda = 0.5 \quad \theta = 0.9$	-198.3
$\lambda = 0.2 \quad \theta = 0.9$	-211.5
$\lambda = 0.2 \quad \theta = 0.6$	-214.9
$\lambda = 0.2 \quad \theta = 0.1$	-209.9

(b) Performance of Linear Pools

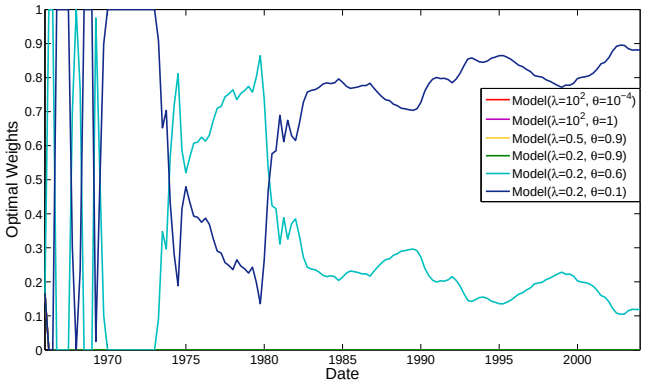
	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	-187.8
Full-sample optimal weights	0.447	0.554	0.000	0.000	0.000	0.000	-177.2
Average real-time optimal weights	0.412	0.545	0.040	0.001	0.001	0.001	-175.6

(c) Reasons of Models Being in the Full-sample Optimal Pool

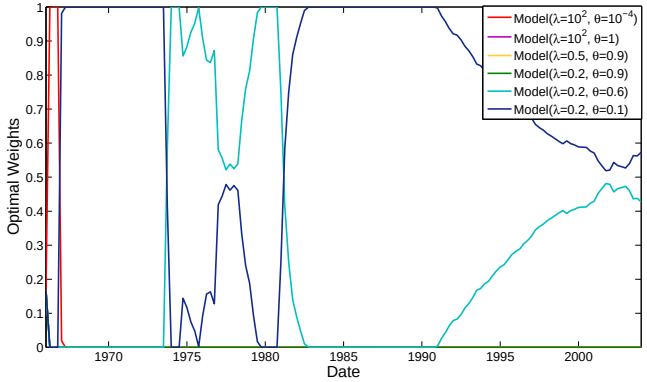
Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1968.3 1968.4 1969.1 1969.2 1979.2 1980.3
$\lambda = 10^2 \quad \theta = 1$	1966.4 1969.3 1969.4 1970.1 1970.2 1972.2 1973.1 1974.3 1974.4 1978.3 1982.2



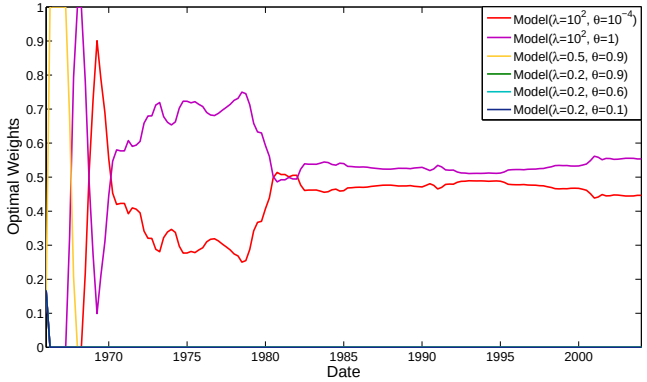
(a) Vector



(b) Output



(c) Output



(d) Output

Figure 4.5: Evolution of Real-time Optimal Weights: Single-4-step-ahead Predictions

CHAPTER 5
PREDICTING IOWA TAX REVENUE AND PERSONAL INCOME

We provide a forecasting exercise on tax revenue and personal income in the state of Iowa, for demonstrating the effectiveness of the proposed full Bayesian autoregressive (BVAR) model and the optimal pooling techniques described in Chapters 2 and 3. We are interested in the selection of prior tightness and the performance of real-time optimal prediction pools.

This chapter is organized as follows. Section 5.1 describes the data set. Section 5.2 specifies six BVAR models. Compared with the models used in Chapter 4, the models in this chapter have extra deterministic predictors including trends and seasonal dummies in addition to the constants, however, the same levels of prior tightness in terms of the six combinations of λ and θ are still applied. Section 5.3 summarizes and interprets the empirical results of one-step-ahead predictions as well as multi-4-step-ahead and single-4-step-ahead predictions.

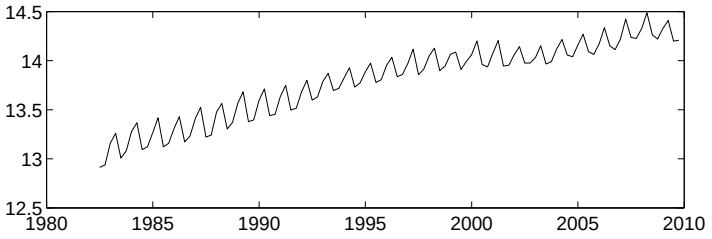
5.1 Data

The data set consists of quarterly time series of tax revenue and personal income in the state of Iowa from 1982.Q3 to 2009.Q4. There are 110 observations for each variable. The definitions of the dependent variables in the BVAR models are as follows.

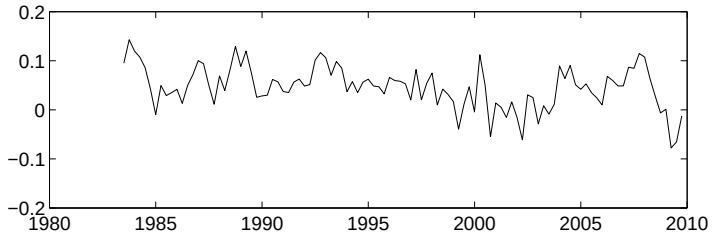
(a) Tax Revenue = LN(general tax receipts), where the data of general tax receipts (in thousands) are not seasonally adjusted. The monthly data are available from the state, and then converted into quarterly data by taking a three-month

average[1].

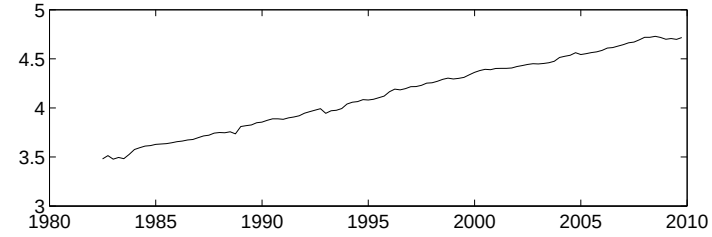
(b) Personal Income= LN(SQ5N), where SQ5N is the Iowa personal income (in billions), which is seasonally adjusted and available quarterly from the Bureau of Economic Analysis.



(a) Tax Revenue



(b) Tax Revenue Growth



(c) Personal Income

Figure 5.1: Tax Revenue, Tax Revenue Growth, and Personal Income in Iowa

[1]Before a three-month average is computed, the total racing and gaming receipts and the one-sixth of sales tax receipts (starting from August 2008) are subtracted from the general tax receipts in the same way as Otrok and Whiteman (1997).

Figure 5.1 plots three time series. Since tax revenue is not seasonally adjusted, it shows a seasonal pattern of wavy fluctuations. Tax revenue growth, which is computed through the fourth-order differencing (i.e., tax revenue$_t$ $-$tax revenue$_{t-4}$), displays several major fluctuations. By contrast, personal income shows an almost linear upward trend over the entire sample period from 1982 to 2009.

5.2 Model Specification

The two-variable BVAR model uses four lags of each variable and a set of deterministic predictors, including constants, trends, and seasonal dummies[2]. The prior distribution for the precision matrix of regression disturbances (i.e., $\mathbf{H}$) is Wishart, as described in (2.6). The scale matrix $\underline{\mathbf{S}}$ is set to an estimated residual variance-covariance matrix from a seemingly unrelated regression using the data through 1990.Q4, and the degrees of freedom parameter $\underline{\nu}$ is set to 4. The prior distribution for the vector of regression coefficients (i.e., $\boldsymbol{\beta}$) is multivariate normal conditional on $\mathbf{H}$, as described in (2.8) and (2.9). The prior means of the coefficients of the first own lags are set to ones, while the prior means of the other coefficients are set to zeros. The prior standard deviations of the coefficients of the deterministic predictors are set to 10^2, while the prior standard deviations of the lag coefficients are set as suggested by the Minnesota prior. As for the overall tightness (λ) and cross-equation tightness (θ), the six combinations, which are introduced in Chapter 4, are applied. They are:

[2]Slightly different from the work of Otrok and Whiteman (1997), the dummy variables of tax policy changes are not included in the models. This is because the most recent tax policy change started in 2000.Q3, if policy dummies were included, the first predictive density evaluation would be only after 2001.Q1, which would result in a smaller sample of predictive densities.

(1) $\lambda = 10^2$ and $\theta = 10^{-4}$;

(2) $\lambda = 10^2$ and $\theta = 1$;

(3) $\lambda = 0.5$ and $\theta = 0.9$;

(4) $\lambda = 0.2$ and $\theta = 0.9$;

(5) $\lambda = 0.2$ and $\theta = 0.6$; and

(6) $\lambda = 0.2$ and $\theta = 0.1$.

The purpose of using such six combinations is to present typical cases of interest. Combinations (1) and (2) represent two extreme cases. One makes the equations in the VAR system close to univariate autoregressions, and the other makes the VAR system unrestricted. Combinations (4), (5), and (6) fix λ but vary θ, where the choice of setting λ equal to 0.2 is recommended by Litterman (1986). Combinations (3) and (4) then fix θ but vary λ. The choice of setting θ equal to 0.9 is recommended by Dua and Ray (1995).

5.3 Empirical Results

In this exercise, prediction performance is evaluated through the conventional log predictive score, which is the sum of the logarithms of predictive densities. In each prediction, the first predictive density evaluation always starts from 1991.Q1. There are 76 predictive probability densities in the one-step-ahead prediction, and 73 predictive probability densities in each case of the multi-4-step-ahead and single-4-step-ahead predictions. In addition to the six BVAR models, linear prediction pools of models are constructed by using equal weights and optimal weights. Specifically, the full-sample optimal weights are computed only once based on all the predictive

densities from 1991.Q1 to 2009.Q4, while the real-time optimal weights are updated at each time period based on only past predictive densities. Weights are then placed on predictive densities, and log predictive scores are computed.

Four major findings can be summarized from the results.

- A larger overall tightness parameter (λ) works better for predicting tax revenue, while a smaller λ works better for predicting personal income at all the three forecast horizons. For example, in the one-step-ahead prediction of tax revenue, the two models with ($\lambda = 10^2$) have their full-sample optimal weights summed up to 0.868 (Table 5.2b), while in the one-step-ahead prediction of personal income, the two models with ($\lambda = 0.2$) have their full-sample optimal weights summed up to 0.952 (Table 5.3b).

- The advantage of using a "best" λ becomes more salient at a longer forecast horizon. For example, in the single-4-step-ahead prediction of tax revenue, the full-sample optimal weight of the model with ($\lambda = 10^2$) reaches 0.896 (Table 5.8b), while in the single-4-step-ahead prediction of personal income, the full-sample optimal weights of the two models with ($\lambda = 0.2$) reach 1.000 (Table 5.9b).

- For the prediction of the vector of the two variables, the equally-weighted prediction pool[3] outperforms the single best model at all the three forecast horizons. However, for the prediction of each variable, the equally-weighted prediction pool does worse than the single best model. For example, in the

[3]An equally-weighted prediction pool is constructed by placing equal weights on the predictive densities, not on log predictive scores, of the models under consideration.

one-step-ahead prediction of the vector, the log predictive score of the equally-weighted pool is 349.4, which is about 2 points higher than 347.1, the score of the single best model (Table 5.1). In the one-step-ahead prediction of tax revenue, the log predictive score of the equally-weighted pool is 141.1, which is about 2 points lower than 143.0, the score of the single best model (Table 5.2). Moreover, the difference between the equally-weighted pool and the single best model also becomes more salient at the multi-step-ahead predictions. For example, in the multi-4-step-ahead prediction of the vector, the difference is about 4 points (Table 5.4). In the multi-4-step-ahead prediction of tax revenue, the difference is about 18 points (Table 5.5).

- The real-time optimal prediction pool does outperform the single best model and the equally-weighted prediction pool in all the predictions in this exercise. For example, in the one-step-ahead prediction of tax revenue (Table 5.2), the log predictive score of the real-time optimal pool is 148.1, which is 7 points (5.0%) higher than that of the equally-weighted prediction pool, and 5 points (3.5%) higher than that of the single best model.

The above findings deliver two messages, which are supported by the data characteristics.

- A larger overall tightness parameter (λ) works better for a volatile target. In this exercise, a model with ($\lambda = 10^2$) predicts tax revenue better than a model with ($\lambda = 0.2$), because a larger λ can account for extraordinary volatility of tax revenue. For example, Iowa experienced a historical devastating flood in 1993.Q3, and tax revenue dropped by 16% from 1993.Q2 to 1993.Q3, while personal income dropped by 0.5% over the same period. Consequently, in

the one-step-ahead prediction of tax revenue, the model with ($\lambda = 10^2$, $\theta = 1$) is weighted more heavily than the other models in both full-sample and real-time optimal prediction pools (Table 5.2b). Moreover, the predictive density at 1993.Q3 is also identified as a supporting observation for the model with ($\lambda = 10^2$, $\theta = 1$) (Table 5.2c). Since the coming of a major flood is almost unpredictable a quarter ago, it is encouraging to see that a model can account for the impact of such a sudden event. Basically, a larger λ allows regression coefficients to deviate far from their prior means[4], which in turn incorporates larger uncertainty into parameter estimation and therefore delivers better forecasts than a smaller λ.

- A linear prediction pool of models with different levels of prior tightness can outperform its single best constituent model alone. Basically, a combination of different prior tightness is more capable of adapting to various situations over time than a single level of prior tightness. For example, in the prediction of tax revenue, although the two models with ($\lambda = 10^2$) have total optimal weights higher than 0.85, the optimal pools also include a model with ($\lambda = 0.2$) at all the three forecast horizons (Tables 5.2b, 5.5b, and 5.8b). In addition, the observation that the equally-weighted pool can easily outperform the single best model for predicting the vector also confirms the advantage of using different levels of prior tightness.

[4]The posterior means of coefficients of longer lags (i.e., the third and fourth lags) are actually greater in absolute value for tax revenue than for personal income.

Table 5.1: Log Scores for One-Step-Ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	347.1
$\lambda = 10^2 \quad \theta = 1$	342.6
$\lambda = 0.5 \quad \theta = 0.9$	345.2
$\lambda = 0.2 \quad \theta = 0.9$	339.3
$\lambda = 0.2 \quad \theta = 0.6$	340.2
$\lambda = 0.2 \quad \theta = 0.1$	340.5

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	349.4
Full-sample optimal weights	0.335	0.390	0.000	0.000	0.000	0.274	351.6
Average real-time optimal weights	0.406	0.414	0.002	0.002	0.011	0.164	354.4

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1994.1 1994.3 2000.1
$\lambda = 10^2 \quad \theta = 1$	1991.3 1993.1 1993.3 1995.3 2003.1 2005.1
$\lambda = 0.2 \quad \theta = 0.1$	2000.2 2007.4 2009.2

Table 5.2: Log Scores for One-Step-Ahead Predictions of Tax Revenue

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	140.1
$\lambda = 10^2 \quad \theta = 1$	143.0
$\lambda = 0.5 \quad \theta = 0.9$	139.0
$\lambda = 0.2 \quad \theta = 0.9$	128.6
$\lambda = 0.2 \quad \theta = 0.6$	128.6
$\lambda = 0.2 \quad \theta = 0.1$	129.1

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	141.1
Full-sample optimal weights	0.308	0.560	0.000	0.000	0.000	0.133	145.6
Average real-time optimal weights	0.350	0.518	0.055	0.015	0.002	0.060	148.1

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1994.1 1996.3
$\lambda = 10^2 \quad \theta = 1$	1991.3 1993.3 1997.1 1997.3 1998.2 1999.1
$\lambda = 0.2 \quad \theta = 0.1$	2000.2 2009.2

Table 5.3: Log Scores for One-Step-Ahead Predictions of Personal Income

(a) Performance of Single Model

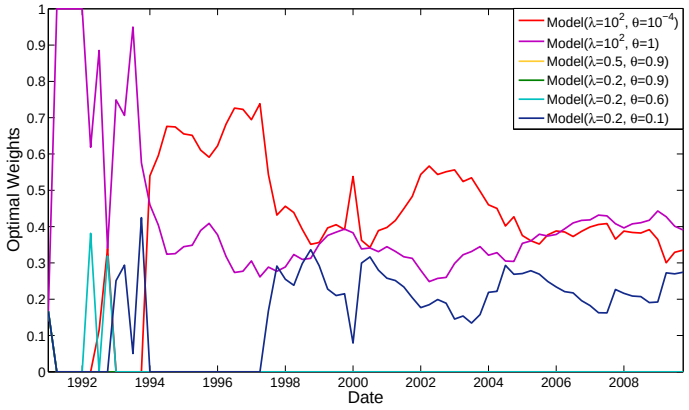
Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	206.6
$\lambda = 10^2 \quad \theta = 1$	200.1
$\lambda = 0.5 \quad \theta = 0.9$	206.2
$\lambda = 0.9 \quad \theta = 0.2$	210.7
$\lambda = 0.2 \quad \theta = 0.6$	211.7
$\lambda = 0.2 \quad \theta = 0.1$	211.5

(b) Performance of Linear Pools

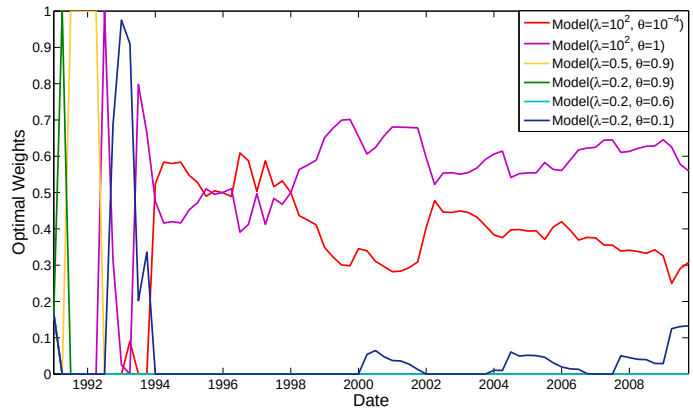
	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	209.9
Full-sample optimal weights	0.000	0.048	0.000	0.000	0.798	0.154	211.8
Average real-time optimal weights	0.098	0.167	0.002	0.002	0.152	0.579	213.1

(c) Reasons of Models Being in the Full-sample Optimal Pool

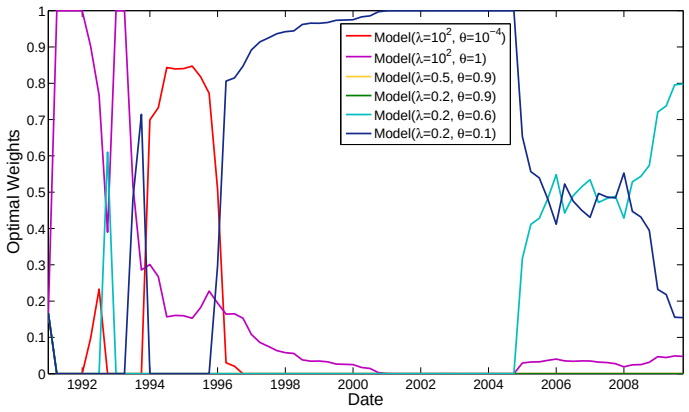
Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 1$	1993.1
$\lambda = 0.2 \quad \theta = 0.6$	2003.1 2005.1
$\lambda = 0.2 \quad \theta = 0.1$	2004.1



(a) Vector



(b) Tax Revenue



(c) Personal Income

Figure 5.2: Evolution of Real-time Optimal Weights: One-step-ahead Predictions

Table 5.4: Log Scores for Multi-4-Step-Ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	1329.9
$\lambda = 10^2 \quad \theta = 1$	1310.3
$\lambda = 0.5 \quad \theta = 0.9$	1320.7
$\lambda = 0.2 \quad \theta = 0.9$	1295.8
$\lambda = 0.2 \quad \theta = 0.6$	1298.8
$\lambda = 0.2 \quad \theta = 0.1$	1302.1

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	1334.2
Full-sample optimal weights	0.417	0.369	0.000	0.000	0.000	0.214	1341.3
Average real-time optimal weights	0.540	0.326	0.003	0.002	0.025	0.104	1343.4

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1993.3 1993.4 1994.1 1994.2 1994.3 2000.1 2001.2 2001.3 2001.4
$\lambda = 10^2 \quad \theta = 1$	1991.1 1992.4 1994.4 1995.1 1998.4 1999.1 2002.3 2002.4 2003.1 2004.4 2005.1 2008.2
$\lambda = 0.2 \quad \theta = 0.6$	1996.4
$\lambda = 0.2 \quad \theta = 0.1$	1997.1 1997.2 2003.4 2009.1

Table 5.5: Log Scores for Multi-4-Step-Ahead Predictions of Tax Revenue

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	534.5
$\lambda = 10^2 \quad \theta = 1$	551.5
$\lambda = 0.5 \quad \theta = 0.9$	531.3
$\lambda = 0.2 \quad \theta = 0.9$	487.1
$\lambda = 0.2 \quad \theta = 0.6$	487.1
$\lambda = 0.2 \quad \theta = 0.1$	487.8

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	533.6
Full-sample optimal weights	0.172	0.805	0.000	0.000	0.000	0.022	552.9
Average real-time optimal weights	0.125	0.852	0.016	0.002	0.002	0.003	553.9

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	1993.4 1994.1
$\lambda = 10^2 \quad \theta = 1$	1991.4 1992.1 1992.4 1993.1 1993.2 1993.3 1998.1 1998.2 1998.4 1999.1 2000.3 2000.4 2003.4 2004.1 2006.1 2006.2 2006.4 2008.2
$\lambda = 0.2 \quad \theta = 0.1$	2009.1

Table 5.6: Log Scores for Multi-4-Step-Ahead Predictions of Personal Income

(a) Performance of Single Model

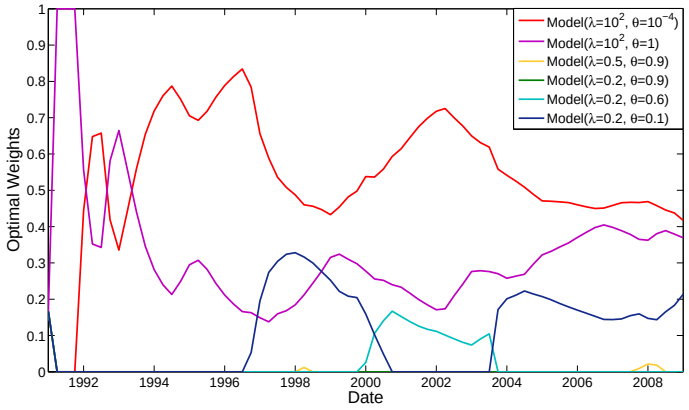
Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	795.2
$\lambda = 10^2 \quad \theta = 1$	774.7
$\lambda = 0.5 \quad \theta = 0.9$	799.7
$\lambda = 0.2 \quad \theta = 0.9$	813.5
$\lambda = 0.2 \quad \theta = 0.6$	815.0
$\lambda = 0.2 \quad \theta = 0.1$	814.1

(b) Performance of Linear Pools

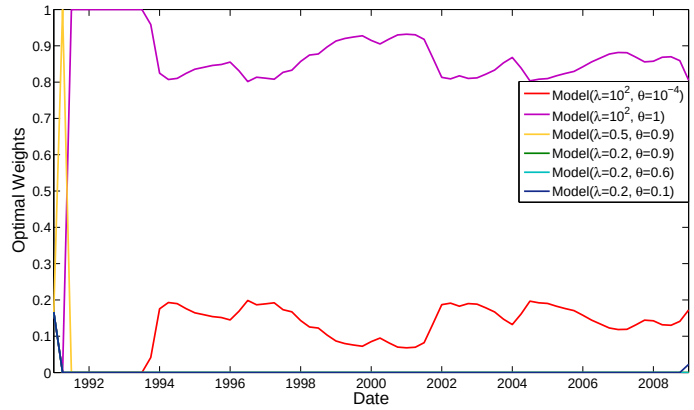
	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	810.2
Full-sample optimal weights	0.000	0.039	0.000	0.000	0.654	0.307	815.4
Average real-time optimal weights	0.106	0.194	0.013	0.002	0.219	0.466	817.6

(c) Reasons of Models Being in the Full-sample Optimal Pool

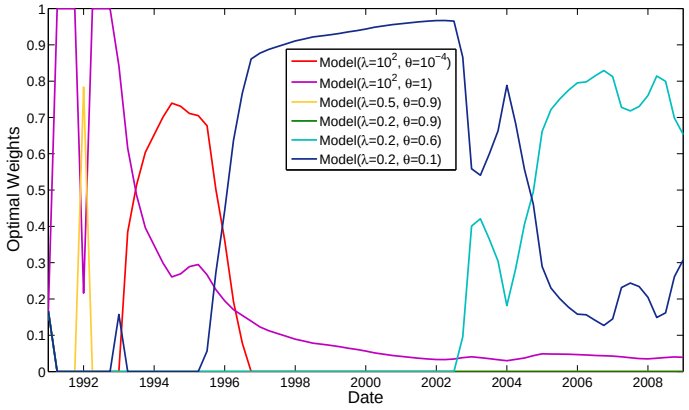
Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 1$	1992.3
$\lambda = 0.2 \quad \theta = 0.6$	2002.3 2002.4 2003.1 2005.1
$\lambda = 0.2 \quad \theta = 0.1$	1992.4 1993.1



(a) Vector



(b) Tax Revenue



(c) Personal Income

Figure 5.3: Evolution of Real-time Optimal Weights: Multi-4-step-ahead Predictions

Table 5.7: Log Scores for Single-4-Step-Ahead Predictions of the Vector

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	246.5
$\lambda = 10^2 \quad \theta = 1$	250.2
$\lambda = 0.5 \quad \theta = 0.9$	249.4
$\lambda = 0.2 \quad \theta = 0.9$	256.7
$\lambda = 0.2 \quad \theta = 0.6$	258.1
$\lambda = 0.2 \quad \theta = 0.1$	257.5

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	259.2
Full-sample optimal weights	0.210	0.084	0.000	0.000	0.706	0.000	260.2
Average real-time optimal weights	0.178	0.258	0.057	0.006	0.264	0.238	262.5

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 10^{-4}$	2001.2 2001.3
$\lambda = 10^2 \quad \theta = 1$	1992.3 2000.2
$\lambda = 0.2 \quad \theta = 0.6$	1995.3 1996.2 1996.3 1996.4 1997.1 2003.2 2003.4 2004.1 2006.4 2007.1 2008.3 2008.4

Table 5.8: Log Scores for Single-4-Step-Ahead Predictions of Tax Revenue

(a) Performance of Single Model

Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	107.3
$\lambda = 10^2 \quad \theta = 1$	112.8
$\lambda = 0.5 \quad \theta = 0.9$	105.1
$\lambda = 0.2 \quad \theta = 0.9$	102.4
$\lambda = 0.2 \quad \theta = 0.6$	103.1
$\lambda = 0.2 \quad \theta = 0.1$	103.8

(b) Performance of Linear Pools

	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	109.4
Full-sample optimal weights	0.000	0.896	0.000	0.000	0.104	0.000	113.0
Average real-time optimal weights	0.076	0.834	0.002	0.002	0.009	0.077	114.1

(c) Reasons of Models Being in the Full-sample Optimal Pool

Model	Supporting Observations
$\lambda = 10^2 \quad \theta = 1$	1993.1 1998.4 1999.4 2000.1 2000.2 2000.3 2000.4 2001.1 2001.2 2001.3 2001.4 2002.1 2002.2 2002.4
$\lambda = 0.2 \quad \theta = 0.6$	2003.4

Table 5.9: Log Scores for Single-4-Step-Ahead Predictions of Personal Income

(a) Performance of Single Model

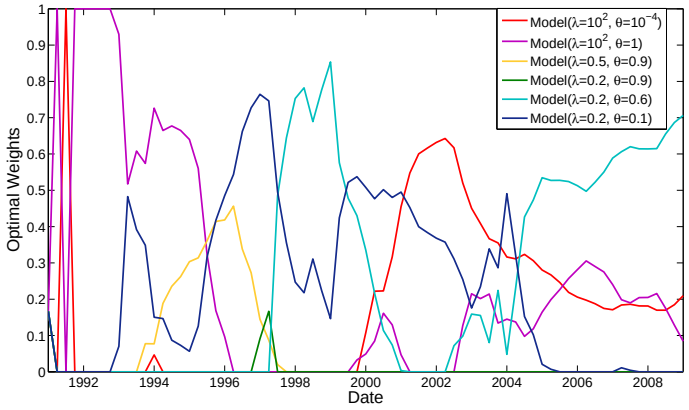
Parameters of Prior	Log Scores
$\lambda = 10^2 \quad \theta = 10^{-4}$	138.15
$\lambda = 10^2 \quad \theta = 1$	139.97
$\lambda = 0.5 \quad \theta = 0.9$	147.43
$\lambda = 0.2 \quad \theta = 0.9$	153.90
$\lambda = 0.2 \quad \theta = 0.6$	153.95
$\lambda = 0.2 \quad \theta = 0.1$	151.65

(b) Performance of Linear Pools

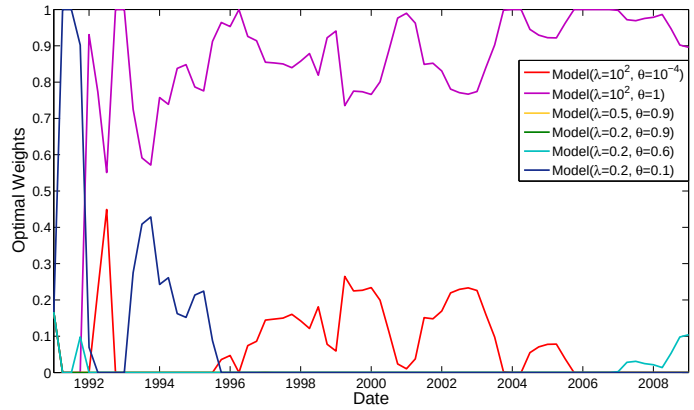
	$\lambda = 10^2$ $\theta = 10^{-4}$	$\lambda = 10^2$ $\theta = 1$	$\lambda = 0.5$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.9$	$\lambda = 0.2$ $\theta = 0.6$	$\lambda = 0.2$ $\theta = 0.1$	Log Score
Equal weights	0.167	0.167	0.167	0.167	0.167	0.167	150.02
Full-sample optimal weights	0.000	0.000	0.000	0.267	0.733	0.000	153.96
Average real-time optimal weights	0.042	0.105	0.144	0.038	0.350	0.321	155.13

(c) Reasons of Models Being in the Full-sample Optimal Pool

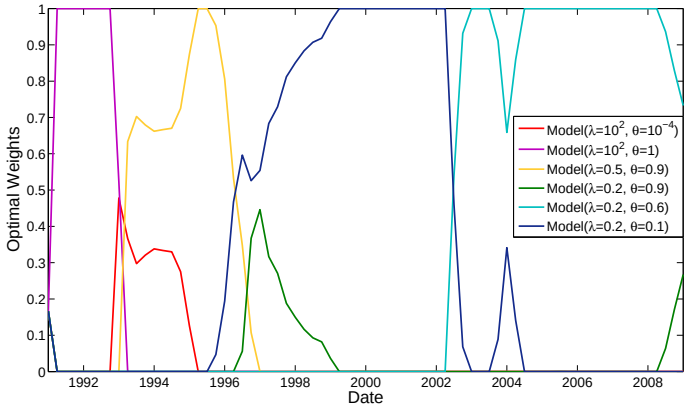
Model	Supporting Observations
$\lambda = 0.2 \quad \theta = 0.9$	1992.2
$\lambda = 0.2 \quad \theta = 0.6$	1993.1



(a) Vector



(b) Tax Revenue



(c) Personal Income

Figure 5.4: Evolution of Real-time Optimal Weights: Single-4-step-ahead Predictions

CHAPTER 6
CONCLUSION

This work improves the prediction performance of Vector Autoregressive (VAR) models in two ways. First, a full Bayesian VAR (BVAR) model is developed to overcome several major limitations of previous BVAR models. Second, an optimal prediction pool of competing models is constructed to achieve higher prediction accuracy over a single best model.

The proposed BVAR model has four desirable features compared with previous BVAR models. First, the variance-covariance matrix of regression disturbances is treated as unknown and random to incorporate parameter uncertainty. Second, the prior variance-covariance matrix of regression coefficients is constructed as a function of the variance-covariance matrix of the disturbances. Therefore, realizations of these coefficients are governed by the prior specification of the disturbances, and dependencies between the coefficients of different equations are effectively taken into account, since the off-diagonal elements of the variance-covariance matrix of the disturbances are not necessarily zeros. Third, all the equations of the model are estimated simultaneously, in order to account for possible contemporaneous correlations of the disturbances. Fourth, the model is fully Bayesian. A state-of-the-art Metropolis-within-Gibbs algorithm is specified to obtain draws from the posterior distributions, which are not in standard forms. Consequently, the proposed model does not have previous BVAR models' limitations on parameter specification and estimation.

The proposed optimal prediction pool offers a sensible alternative to model selection, when there exist multiple competing models. Since the forecasting capability of any single best model can be limited, it is desirable to combine competing models so that they can complement each other. Specifically, it is more desirable to combine density forecasts from competing models to obtain a complete description of future uncertainty. To simplify the implementation, a practical algorithm is offered for computing the optimal weights of competing models by maximizing a conventional log predictive score function. In addition, to further understand why a certain constituent model is given a positive weight in an optimal prediction pool, a search scheme is provided for identifying supporting observations of each model.

The effectiveness of the proposed techniques is examined through two forecasting exercises. One exercise is predicting national output, inflation, and interest rate in the United States. The other is predicting state tax revenue and personal income in Iowa. For practical purposes, the investigation goes beyond one-step-ahead predictions to multi-H-step-ahead and single-H-step-ahead predictions, and focuses on the selection of BVAR prior tightness and the performance of linear prediction pools.

The results indicate that a properly selected prior tightness can improve prediction accuracy. Specifically, in the setting of the Minnesota prior, a larger overall tightness (λ) works better than a smaller one, when future values of the target variables are likely to deviate far from their most recent historical values, and vice versa. By definition, a large λ loosens up the prior standard deviations of all the lag coefficients, which in turn allows the coefficients to deviate far from their prior means. Since a prior specification with a large λ truly admits large variation in the target variables and fully incorporates parameter uncertainty, it delivers better

forecasts than a prior specification with a smaller λ. The advantage of using a larger λ for a volatile target becomes more salient at a longer forecast horizon (e.g., single-4-step-ahead predictions). However, when the target is a vector of different variables and a multi-H-step-ahead prediction is of interest, it becomes desirable to consider a prediction pool of models with different levels of prior tightness, in order to accommodate various situations and achieve higher prediction accuracy over a single best model.

Since the benefit of using a combination of different prior tightness is so salient for predicting a vector, even a prediction pool that is simply constructed by using equal weights can outperform a single best model. However, when the target is a certain variable of the vector, a prediction pool with equal weights does not necessarily outperform a single best model. In such a case, it becomes necessary to construct an optimal prediction pool. Since a prediction pool with the optimal weights computed based on the entire sample cannot do worse than any best constituent model by the mathematics of optimization, it would be arbitrary to assert that such an optimal pool can surely deliver better forecasts over time in practice. A sensible approach to evaluating an optimal prediction pool is through its real-time performance. That is, real-time optimal weights are computed using only past data at each time period. The empirical results of this work are encouraging, as they indicate that such a real-time optimal prediction pool works better than any best constituent model alone in all the predictions under consideration. In addition, the improvement over a single best constituent model is more noticeable in the circumstances, when the real-time optimal weights converge more quickly.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*, 163–180.

Amirizadeh, H. and R. M. Todd (1984). More growth ahead for ninth district states. *Federal Reserve Bank of Minneapolis Quarterly Review 8*(4), 8 – 17.

Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models*. New York: Marcel Dekker.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*(4), 559–583.

Clements, M. P. and G. E. Mizon (1991). Empirical analysis of macroeconomic time series: Var and structural models. *European Economic Review 35*(4), 887–917.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (1993). Analyzing vars with monetary business cycle model priors. *Proceedings of the American Statistical Association, Bayesian Statistics Section*, 160–169.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (2000). A bayesian approach to dynamic macroeconomics. *Journal of Econometrics 98*(2), 203 – 223.

Del Negro, M. and F. Schorfheide (2004). Priors from general equilibrium models for VARs. *International Economic Review 45*(2), 643 – 673.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit of new keynesian models. *Journal of Business and Economic Statistics 25*(2), 123 – 143.

Diebold, F. X. and J. A. Lopez (1996). Forecast evaluation and combination. NBER Technical Working Papers 0192, National Bureau of Economic Research, Inc.

Doan, T. and R. Litterman (1986). *RATS Version 2.0; Users' Manual*, Chapter VAR Econometrics. Estima, Minneapolis, MN.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*(1), 1–100.

Dua, P. and S. C. Ray (1995). A bvar model for the connecticut economy. *Journal of Forecasting 14*(3), 167 – 180.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Hoboken: Wiley.

Geweke, J. and G. Amisano (2010). Optimal prediction pools. *Journal of Econometrics*. forthcoming.

Geweke, J. and C. Whiteman (2006). *Bayesian Forecasting*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 1, pp. 3–80. Elsevier.

Gupta, R. and M. M. Sichei (2006). A bvar model for the south african economy. *South African Journal of Economics 74*(3), 391 – 409.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological) 41*(2), 190–195.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *Econometrics Journal 7*(1), 1–31.

Hendry, D. F. and G. E. Mizon (1993). *Evaluating Dynamic Econometric Models by Encompassing the VAR.*, pp. 272–300. Cambridge, Mass. and Oxford:.

Hsiao, C. (1981). Autoregressive modelling and money-income causality detection. *Journal of Monetary Economics 7*(1), 85–106.

Hsiao, C. (1982). Autoregressive modeling and causal ordering of economic variables. *Journal of Economic Dynamics and Control 4*(3), 243–259.

Ingram, B. F. and C. H. Whiteman (1994). Supplanting the 'minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors. *Journal of Monetary Economics 34*(3), 497 – 510.

Kadiyala, K. R. and S. Karlsson (1993). Forecasting with generalized bayesian vector auto regressions. *Journal of Applied Econometrics 12*(3-4), 365–378.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica 50*(6), 1345 – 1370.

LeSage, J. P. and A. Krivelyova (1999). A spatial prior for bayesian vector autoregressive models. *Journal of Regional Science 39*(2), 297 – 317.

LeSage, J. P. and Z. Pan (1995). Using spatial contiguity as bayesian prior information in regional forecasting models. *International Regional Science Review 18*(1), 33 – 53.

Lewis, K. F. and C. H. Whiteman (2006). Empirical bayesian density forecasting in iowa and shrinkage for the monte carlo era. Discussion Paper Series 1: Economic Studies 2006,28, Deutsche Bundesbank, Research Centre.

Litterman, R. B. (1979). *Techniques Of Forecasting Using Vector Autoregressions*. Ph. D. thesis, University of Minnesota.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics 4*(1), 25–38.

Otrok, C. and C. Whiteman (1997). What to do when the crystal ball is cloudy: Conditional and unconditional forecasting in iowa. *Proceedings of the National Tax Association*, 326 334.

Press, S. J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* (Second ed.). Dover Publications.

Schwarz, G. (1979). Estimating the dimension of a model. *Annals of Statistics 6*(2), 461–464.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Smets, F. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association 1*(5), 1123 – 1175.

Smets, F. and R. Wouters (2005). Comparing shocks and frictions in us and euro area business cycles: A bayesian dsge approach. *Journal of Applied Econometrics 20*(2), 161 – 183.

Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review 97*(3), 586–606.

Tay, A. S. and K. F. Wallis (2000). Density forecasting: A survey. *Journal of Forecasting 19*(4), 235–254.

Timmermann, A. G. (2006). *Forecast Combinations*, Chapter 4, pp. 135–196. Handbook of Economic Forecasting. Amsterdam: North-Holland.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics 67*, 983–994.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. J. Wiley, New York.