Alternative Bayesian Estimators for Vector-Autoregressive Models

Shawn Ni, Department of Economics, University of Missouri, Columbia, MO 65211, USA

Dongchu Sun, Department of Statistics, University of Missouri, Columbia, MO 65211, USA

Abstract

This paper compares frequentist risks of several Bayesian estimators of the VAR lag parameters and covariance matrix under alternative priors. With the constant prior on the VAR lag parameters, the asymmetric LINEX estimator for the lag parameters does better overall than the posterior mean. The posterior mean of covariance matrix performs well in most cases. The choice of prior has more significant effects on the estimates than the form of estimators. The shrinkage prior on the VAR lag parameters dominates the constant prior, while Yang and Berger's reference prior on the covariance matrix dominates the Jeffreys prior. Estimation of a VAR using the U.S. macroeconomic data reveals significant differences between estimates under the shrinkage and constant priors.

KEY WORDS: Bayesian VAR, pseudo entropy loss, quadratic loss, LINEX loss, Noninformative priors.

JEL CLASSIFICATIONS: C11, C15, C32.

1 Introduction

In this paper we examine Bayesian estimators of Vector-Autoregression (VAR) model under several loss functions and non-informative priors that have been studied in the statistics literature. The Bayesian estimators considered in the study include alternatives to the posterior mean–the default choice by the users of VAR models in practice. Under most of the priors, our numerical analysis shows that commonly used loss functions yield estimators that are quantitatively similar. The prior choice, on the other hand, has significant effects on the Bayesian estimates. We find that for all loss functions considered the diffuse prior commonly employed in the empirical VAR literature is inferior to an alternative.

In the past two decades VAR has become a popular tool for analyzing empirical macroeconomic questions. In addressing policy questions such as the macroeconomic effects of monetary supply shocks, economists often rely on VAR models. A typical VAR of a p dimensional variable, $\mathbf{y}_t$, $(t = 1, .., T)$ can be specified as follows.

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{L} \mathbf{y}_{t-i}\mathbf{B}_i + \boldsymbol{\epsilon}_t,$$

where $(\mathbf{c}', \mathbf{B}'_1, \cdots, \mathbf{B}'_L)$ are the VAR coefficient matrix (denoted by $\boldsymbol{\Phi}'$ hereafter). The covariance matrix of the error term is $\boldsymbol{\Sigma}$. For a VAR with even a modest number of variables and relatively short lag length there are typically hundreds of parameters in matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ to be estimated via Ordinary Least Square or Maximum Likelihood Estimation. Drawing finite sample inferences of the VAR parameters is a challenge. Frequentist finite sample distributions cannot be derived in closed-form, while asymptotic asymptotic theory may not be applicable to a VAR with a large number of parameters and limited data observations, which is usually the case in macroeconomic applications.

In practice, Bayesian procedures are widely used for finite sample inferences of VAR models. Bayesian estimators are derived from minimization of expected posterior loss in the parameter space. Hence the choice of loss function determines the form of Bayesian estimator. In applications of Bayesian procedures, the posterior means of generated VAR lag coefficients and covariance matrix are usually employed as the Bayesian estimators. The posterior means of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are optimal for certain losses on $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$. Bayesian estimators derived from minimizing other commonly used loss functions in the Bayesian statistics literature are rarely studied for VAR models. These loss functions include Yang and Berger's (1994) quadratic and pseudo inverse entropy losses for $\mathbf{\Sigma}$ and Zellner's (1986) LINEX loss for $\mathbf{\Phi}$. The fact that Bayesian estimators derived from these loss functions differ from the posterior mean may be of interest for macroeconomists. For instance, under the most commonly used priors the posterior mean of $\mathbf{\Phi}$ is biased. An asymmetric LINEX estimator may be helpful in correcting the bias.

The properties of Bayesian estimators are also influenced by the choice of prior. Research on the effects of prior choice on VAR posterior distributions is relatively scant. (For some studies on the choice of priors for VARs see Kadiyala and Karlsson 1997 and Ni and Sun 2001). The default prior in the literature for the VAR lag coefficients $\mathbf{\Phi}$ is the constant prior and the default prior for the covariance matrix of the residuals $\mathbf{\Sigma}$ is the Jeffreys prior or a modified version of it used in RATS (the RATS prior hereafter). This combination of priors allows for easy simulation of posterior distributions. However, the Jeffreys prior is known to be deficient in high dimensional settings (see Berger and Bernardo 1992). Bernardo (1979) and Berger and Bernardo (1992) propose an alternative approach of deriving a reference prior. The reference prior for iid covariance matrix is derived by Yang and Berger (1994). The constant prior for $\mathbf{\Phi}$ is known to be inadmissible under quadratic loss for estimation of unknown mean of vector with iid normal distribution (see Berger and Strawderman 1996). An alternative to the constant prior is a 'shrinkage' prior for $\mathbf{\Phi}$, which has been used in estimating the unknown normal mean in iid cases (e.g, Baranchik 1964), and in hierarchical linear mixed models (e.g., Berger and Strawderman, 1996). The shrinkage prior is independent of the sample size and is quite easy to use for computing the posteriors. Ni and Sun (2001) explore the Bayesian posterior mean estimator of $\mathbf{\Phi}$ under the shrinkage prior, but no work has been done to compare the posterior mean estimator with other Bayesian estimators under this prior in the VAR setting. The present study will examine the effect of priors for $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ under alternative loss functions and provide recommendations for VAR users.

In section 2 of the paper we lay out notation for the VAR model. In section 3 we discuss generalized Bayesian estimators under alternative loss functions. In section 4 we examine MCMC simulation results of marginal posteriors of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. In section 5 we compare alternative Bayesian estimates of a VAR using quarterly data of the U.S. economy. In section 6 we offer concluding remarks.

2 VAR Model and the MLE

We consider the VAR model

$$\mathbf{y}_t = \mathbf{c} + \sum_{j=1}^{L} \mathbf{y}_{t-j}\mathbf{B}_j + \boldsymbol{\epsilon}_t, \tag{1}$$

for $t = 1, \cdots, T$, where L is a known positive integer, $\mathbf{c}$ is a $1 \times p$ unknown vector, $\mathbf{B}_j$ is an unknown $p \times p$ matrix, $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are independently identically distributed $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ errors, and $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix. Denote

$$\begin{aligned}
\mathbf{x}_t &= (1, \mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-L}),\\
\mathbf{Y} &= \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_T \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{pmatrix}, \boldsymbol{\Phi} = \begin{pmatrix} \mathbf{c} \\ \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_L \end{pmatrix}.
\end{aligned}$$

Here $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ are $T \times p$ matrices, $\boldsymbol{\Phi}$ is a $(1 + Lp) \times p$ matrix of unknown parameters, $\mathbf{x}_t$ is a $1 \times (1 + Lp)$ row vector, and $\mathbf{X}$ is a $T \times (1 + Lp)$ matrix of observations. We may write the VAR as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Phi} + \boldsymbol{\epsilon}. \tag{2}$$

The likelihood function of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is

$$\begin{aligned}
L(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) &= \frac{1}{|\boldsymbol{\Sigma}|^{T/2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\Phi})' \right\}\\
&= \frac{1}{|\boldsymbol{\Sigma}|^{T/2}} etr\left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi})' \right\}. \qquad (3)
\end{aligned}$$

Here and hereafter $etr(\mathbf{A})$ is $\exp(trace(\mathbf{A}))$ of a matrix $\mathbf{A}$. The Maximum Likelihood Estimator (MLE) of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ are

$$\widehat{\boldsymbol{\Phi}}_M = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \text{ and } \widehat{\boldsymbol{\Sigma}}_M = \mathbf{S}(\widehat{\boldsymbol{\Phi}}_M)/T, \tag{4}$$

respectively, where

$$\mathbf{S}(\boldsymbol{\Phi}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\Phi}). \tag{5}$$

It is well known that when $T \geq Lp + 1$, $(\mathbf{X}'\mathbf{X})^{-1}$ exists with probability one and if $T \geq Lp + p + 1$ the MLEs of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ exist with probability one and $\mathbf{S}(\widehat{\boldsymbol{\Phi}}_M)$ is positive definite. In this paper, we consider only the case that $T \geq Lp + p + 1$ so the MLEs of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ exist.

3 Loss Functions and Bayesian Estimators

A Bayesian estimator of $(\mathbf{\Phi}, \mathbf{\Sigma})$ depends on the sampling distribution, the prior, and the loss function. In this paper we consider loss functions that contain a part measuring the loss associated with the covariance matrix and a part measuring the loss of VAR lag coefficients only. The overall loss with respect to $(\mathbf{\Phi}, \mathbf{\Sigma})$ is in the form of

$$L(\hat{\mathbf{\Phi}}, \hat{\mathbf{\Sigma}}; \mathbf{\Phi}, \mathbf{\Sigma}) \quad = \quad L_{\Phi}(\hat{\mathbf{\Phi}} + L_{\Sigma}(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}). \tag{6}$$

The question we seek to answer is whether alternative loss functions result in Bayesian estimators with significantly different properties.

3.1 Loss Functions for $\mathbf{\Sigma}$

First, we consider a loss function for $\mathbf{\Sigma}$:

$$L_{\Sigma_1}(\widehat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \log|\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}| - p, \tag{7}$$

where p is the number of variables in the VAR. We refer this function as a pseudo entropy loss since it is an entropy loss with respect to $\mathbf{\Sigma}$ only, while the entropy loss is pertaining to both $\mathbf{\Sigma}$ and $\mathbf{\Phi}$.

The second loss function on $\mathbf{\Sigma}$ is a quadratic loss

$$L_{\Sigma_2}(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - \mathbf{I})^2. \tag{8}$$

The third loss function is a pseudo entropy function on $\mathbf{\Sigma}^{-1}$ consider in Berger and Yang (1994):

$$L_{\Sigma_3}(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p. \tag{9}$$

The following lemma produces Bayesian estimators of $\mathbf{\Sigma}$ with any separate loss with respect to $\mathbf{\Phi}$. In this setting the Bayesian estimators can the be derived independently from minimizing expected posterior loss functions regarding $(\mathbf{\Phi}, \mathbf{\Sigma})$.

Lemma 1 *(a) Under the loss L_{Σ_1}, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_1$, where*

$$\widehat{\mathbf{\Sigma}}_1 \quad = \quad \mathbb{E}(\mathbf{\Sigma} \mid \boldsymbol{Y}). \tag{10}$$

(b) Under the loss L_{Σ_2}, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_2$ given by

$$vec(\widehat{\mathbf{\Sigma}}_2) = \Big[\mathbb{E}\{(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}) \mid \boldsymbol{Y}\}\Big]^{-1} vec(\mathbb{E}(\mathbf{\Sigma}^{-1} \mid \boldsymbol{Y})), \tag{11}$$

where $\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}$ is the Kronecker product of $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Sigma}^{-1}$.

(c) Under the loss L_{Σ_3}, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_3$, where

$$\widehat{\mathbf{\Sigma}}_3 = \Big\{\mathbb{E}(\mathbf{\Sigma}^{-1} \mid \boldsymbol{Y})\Big\}^{-1}. \tag{12}$$

Proof. For part (a), note that

$$\frac{\partial \log(|\widehat{\mathbf{\Sigma}}|)}{\partial \widehat{\mathbf{\Sigma}}} = \widehat{\mathbf{\Sigma}}^{-1}, \quad \frac{\partial tr(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma})}{\partial \widehat{\mathbf{\Sigma}}} = -\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}.$$

The above facts can be used to calculate the derivative of $\mathbb{E}\{L_{\Sigma_1}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \mid \mathbf{Y}\}$ with respect to $\widehat{\mathbf{\Sigma}}$. The desired result follows by letting the derivative be zero. The proof of part (b) is more tedious. For that see Yang and Berger (1994). Part (c) follows from the fact that the derivative of $\mathbb{E}\{L_{\Sigma_3}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \mid \mathbf{Y}\}$ with respect to $\widehat{\mathbf{\Sigma}}$ is $\mathbb{E}(\mathbf{\Sigma}^{-1} \mid \mathbf{Y}) - \widehat{\mathbf{\Sigma}}^{-1}$. □

3.2 Loss Functions for $\mathbf{\Phi}$

The most common loss for $\mathbf{\Phi}$ is the quadratic loss

$$L_{\Phi_1}(\widehat{\mathbf{\Phi}}, \mathbf{\Phi}) = tr\{(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})'\mathbf{W}^{-1}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\}, \tag{13}$$

where $\mathbf{W}^{-1}$ is a constant weighting matrix. If the weighting matrix $\mathbf{W}$ is the identity matrix, then the loss of L_{Φ_1} is simply the sum of squared errors of all elements of $\widehat{\mathbf{\Phi}}$, $\sum_{i=1}^{1+Lp}\sum_{j=1}^{p}(\widehat{\boldsymbol{\phi}}_{i,j} - \boldsymbol{\phi}_{i,j})^2$.

The quadratic function is symmetric, an asymmetric LINEX loss function is explored by Zellner (1986) for estimation of iid normal mean under conjugate prior. A LINEX loss is

$$L_{\Phi_2}(\hat{\boldsymbol{\phi}}, \boldsymbol{\phi}) = \sum_{i=1}^{1+Lp}\Big[\exp\{a(\widehat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i)\} - a(\widehat{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_i) - 1\Big], \tag{14}$$

where a is a hyper-parameter. When a is close to zero, the LINEX loss function is close to be symmetric and not much different from the quadratic loss. When a is a large negative number, the LINEX loss is close to be exponential when $\widehat{\boldsymbol{\phi}}_i < \boldsymbol{\phi}_i$ and close to be linear otherwise. Hence if we suspect that the posterior mean has a downward bias, using the LINEX loss with a negative a parameter should help correcting the bias. The Bayesian estimators of $\mathbf{\Phi}$ under these loss functions are well known.

Lemma 2 *(a) Under the loss L_{Φ_1}, the generalized Bayesian estimator of $\mathbf{\Phi}$ is $\widehat{\mathbf{\Phi}}_1$, where*

$$\widehat{\mathbf{\Phi}}_1 = \mathbb{E}(\mathbf{\Phi} \mid \boldsymbol{Y}). \tag{15}$$

(b) Under the LINEX loss function each elements of the Bayesian estimator $\widehat{\mathbf{\Phi}}_2$ satisfies the condition

$$\widehat{\boldsymbol{\phi}}_i \quad = \quad (-1/a)log[\mathbb{E}\{\exp(-a\boldsymbol{\phi}_i) \mid \boldsymbol{Y}\}] \tag{16}$$

for $i = 1, ..., Lp + 1$.

Proving the Lemma is straightforward. The Lemma shows that the Bayesian estimator on $\mathbf{\Phi}$ under the quadratic loss function is the posterior mean while the estimator under the LINEX loss may be larger or smaller than the posterior mean, depending the sign of the a parameter.

3.3 Priors and Posteriors

Bayesian analysis requires explicit specification of prior on the parameters. The prior may be subjective (informative) or objective (noninformative). Informative priors reflect the beliefs of researchers about the distribution of the parameters of interest. Noninformative priors reflect the vagueness of researchers' knowledge before they observes data. It is not easy to find a subjective prior that is universally justifiable for VAR models in all applications. Noninformative priors are a common choice by VAR users. Note that noninformative priors for $(\mathbf{\Sigma}, \mathbf{\Phi})$ can be derived based on different principles, and are therefore not unique. A recent review of various approaches for deriving noninformative priors is provided by Kass and Wasserman (1996).

The most popular noninformative prior for $\mathbf{\Sigma}$ is the Jeffreys prior (See Geisser 1965, Tiao and Zellner 1964). The Jeffreys prior is derived from the "invariance principle", meaning the prior is invariant to re-parameterization (see Jeffreys 1961 and Zellner 1971). The Jeffreys prior is proportional to the square root of the determinant of the Fisher information matrix. Specifically for the VAR covariance matrix, the Jeffreys prior is

$$\pi_J(\mathbf{\Sigma}) \propto \frac{1}{|\mathbf{\Sigma}|^{(p+1)/2}}. \tag{17}$$

The prior for $\mathbf{\Sigma}$ in RATS is a modified version of the Jeffreys prior

$$\pi_A(\mathbf{\Sigma}) \propto \frac{1}{|\mathbf{\Sigma}|^{(L+1)p/2+1}}. \tag{18}$$

It has been noted, however, that the Jeffreys prior often gives unsatisfactory results for multi-parameter problems. Bernardo (1979) proposes an information-based approach of deriving a reference prior by breaking a single multiparameter problem into a series of problems with fewer numbers of parameters. The reference prior is chosen to maximizes the difference between itself and the posterior, so that maximum information about the parameters of interest is extracted from data. The form of the reference prior depends on the inferential problem at hand and on researchers' ordering of parameters in terms of perceived importance. For examples in which reference priors produce more desirable estimators than Jeffreys priors, see Berger and Bernardo (1992), Sun and Berger (1998), among others. In estimating the variance-covariance matrix $\mathbf{\Sigma}$ based on an iid random sample from a normal population with known mean, Yang and Berger (1994) re-parameterize matrix $\mathbf{\Sigma}$ as $\mathbf{O}'\mathbf{D}\mathbf{O}$ where $\mathbf{D}$ is a diagonal matrix the elements of which are the eigenvalues of $\mathbf{\Sigma}$ (in increasing or decreasing order) and $\mathbf{O}$ is an orthogonal matrix. The following reference prior is derived by placing vectorized $\mathbf{D}$ in front of vectorized $\mathbf{O}$:

$$\pi_R(\mathbf{\Sigma}) \propto \frac{1}{|\mathbf{\Sigma}| \prod_{1 \le i < j \le p} (d_i - d_j)}, \tag{19}$$

where $d_1 > d_2 > \cdots > d_p$ are eigenvalues of $\mathbf{\Sigma}$.

The prior for $(\mathbf{\Phi}, \mathbf{\Sigma})$ can be obtained by putting together priors for $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. In practice, it is often convenient to consider $\boldsymbol{\phi} = vec(\mathbf{\Phi})$, instead of $\mathbf{\Phi}$. A common expression of ignorance about $\boldsymbol{\phi}$ is a (flat) constant prior. A popular noninformative prior for multivariate regression models is called diffuse prior, which consists of a constant prior for $\boldsymbol{\phi}$ and the Jeffreys prior for $\mathbf{\Sigma}$. The joint densities of the constant-Jeffreys prior for $(\mathbf{\Phi}, \mathbf{\Sigma})$ (or $(\boldsymbol{\phi}, \mathbf{\Sigma})$) are of the form $\pi_{CJ}(\boldsymbol{\phi}, \mathbf{\Sigma}) \propto \pi_J(\mathbf{\Sigma})$. A prior similar to π_{CJ} is used by the RATS software package. The constant-RATS prior $\pi_{CA}(\mathbf{\Phi}, \mathbf{\Sigma}) \propto \pi_A(\mathbf{\Sigma})$ is the default choice in RATS and has been used in hundreds of published papers in empirical macroeconomics. For an argument of using constant-RATS instead of constant-Jeffreys prior see Sims and Zha (1999). At last, the constant-reference prior, which has not been commonly used for VAR models, takes the form $\pi_{CR}(\boldsymbol{\phi}, \mathbf{\Sigma}) \propto \pi_R(\mathbf{\Sigma})$.

For estimation of multivariate unknown normal mean, motivated by Stein's (1956) result on inadmissibility of MLE, some authors (e.g., Baranchik 1964, Berger and Strawderman 1996) advocate the following 'shrinkage' prior as an alternative to the constant prior for $\boldsymbol{\phi}$:

$$\pi_S(\boldsymbol{\phi}) \propto \|\boldsymbol{\phi}\|^{-(J-2)}, \quad \boldsymbol{\phi} \in \mathbb{R}^J, \tag{20}$$

where $J = p(Lp+1)$, the dimension of $\boldsymbol{\phi}$.

We consider three noninformative priors for $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ that have the shrinkage prior on $\boldsymbol{\Phi}$. The shrinkage-Jeffreys prior is given by

$$\pi_{SJ}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) \quad \propto \quad \pi_S(\boldsymbol{\phi})\pi_J(\boldsymbol{\Sigma}). \tag{21}$$

The shrinkage-RATS prior is given by

$$\pi_{SA}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) \quad \propto \quad \pi_S(\boldsymbol{\phi})\pi_A(\boldsymbol{\Sigma}). \tag{22}$$

At last the shrinkage-reference prior for $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ is

$$\pi_{SR}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) \quad \propto \quad \pi_S(\boldsymbol{\phi})\pi_R(\boldsymbol{\Sigma}). \tag{23}$$

The posterior properties of various Bayesian estimators of VAR parameters $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ produced by alternative noninformative priors are a focus of the study. Since the noninformative priors for $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ are improper (i.e., the integrals of which in the parameter space are infinite), it is important to know if the posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ exist under these priors. Ni and Sun (2001) prove that the posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ are indeed proper under all of the prior combinations considered in this paper.

3.4 Conditional Posteriors

Although the posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ exist, for most prior combinations they are not available in closed-form. As in other Bayesian statistics studies we employ Markov Chain Monte Carlo (MCMC) methods for sampling from the posterior. In particular, we use the Gibbs sampling method. More details of the theory can be found in Gelfand and Smith (1990). For MCMC numerical simulations of posterior distributions of $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$, We will make use of the following properties of the conditional posteriors with alternative priors.

Fact 1 *Under the constant-Jeffreys prior for* $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ *the conditional posterior* $\boldsymbol{\phi}$ *given* $\boldsymbol{\Sigma}$ *is multivariate normal* $N(\widehat{\boldsymbol{\phi}}_M, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$ *and the marginal posterior of* $\boldsymbol{\Sigma}$ *is Inverse Wishart* $(\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_M)$, $T - Lp - 1)$. *Here the definition of Inverse Wishart distribution follows that given by Anderson (1984, p268).*

Fact 2 *Under the constant-RATS prior the conditional posterior* $\boldsymbol{\phi}$ *given* $\boldsymbol{\Sigma}$ *is* $N(\widehat{\boldsymbol{\phi}}_M, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$ *and the marginal posterior of* $\boldsymbol{\Sigma}$ *is Inverse Wishart* $(\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_M), T)$.

Fact 3 *Under the constant-reference prior*

(a) the conditional distribution of $\boldsymbol{\phi} = vec(\boldsymbol{\Phi})$ *given* $(\boldsymbol{\Sigma}, \boldsymbol{Y})$ *is*

$$\pi(\boldsymbol{\phi} \mid \boldsymbol{\Sigma}, \boldsymbol{Y}) \sim MVN(\widehat{\boldsymbol{\phi}}_M,\ \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}); \tag{24}$$

(b) the conditional density of $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\phi}, \boldsymbol{Y})$ *is given by*

$$\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\phi}, \boldsymbol{Y}) \propto \frac{etr\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{\phi})\}}{|\boldsymbol{\Sigma}|^{\frac{T}{2}+1}\prod_{1\le i<j\le p}(d_i - d_j)}, \tag{25}$$

where $\boldsymbol{S}(\boldsymbol{\Phi})$ *is defined by (??) and* $d_1, ..., d_p$ *are eigenvalues of* $\boldsymbol{\Sigma}$ *in increasing or decreasing order.*

To simulate the posterior of $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ with the shrinkage prior on $\boldsymbol{\Phi}$, we introduce a latent variable δ. So instead of simulating from the conditional distribution of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ within each Gibbs cycle, we simulate from $\boldsymbol{\Phi}$, $\boldsymbol{\Sigma}$, and δ based on the following fact.

Fact 4 *Consider the shrinkage-reference prior* π_{SR}*.*

(a) The conditional density of $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\phi}, \delta, \boldsymbol{Y})$ *is given by (??).*

(b) The conditional distribution of $\boldsymbol{\phi} = vec(\boldsymbol{\Phi})$ *given* $(\delta, \boldsymbol{\Sigma}, \boldsymbol{Y})$ *is* $N(\boldsymbol{\mu}, \boldsymbol{V})$*, where*

$$\boldsymbol{\mu} \;=\; \delta\Big(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1} \;+\; \delta \boldsymbol{I}_J\Big)^{-1}\widehat{\boldsymbol{\phi}}_M; \tag{26}$$

$$\boldsymbol{V} \;=\; \Big(\boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) \;+\; \delta \boldsymbol{I}_J\Big)^{-1}. \tag{27}$$

(c)The conditional distribution of δ *given* $(\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{Y})$ *is Inverse Gamma* $(\frac{J}{2} - 1, \frac{1}{2}\boldsymbol{\phi}'\boldsymbol{\phi})$.

These conditional posteriors are used for MCMC simulations of posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. Ni and Sun (2001) offer proofs for some of the above facts as well as detailed algorithms of MCMC procedures. The Bayesian estimates of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ under alternative loss functions and priors can be computed when the posterior distributions are simulated.

4 Simulations for Marginal Posteriors of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$

In the following we use numerical examples to evaluate the posteriors of competing estimators. We first generate data samples from a VAR with known parameters, then compute the Bayesian estimates via MCMC simulations. Using the Monte Carlo results, we evaluate the performance of the Bayesian estimators under alternative priors in terms of the frequentist risks, which are the average losses of these estimators at the point of true parameters over the sample distribution.

Example 1 We consider a five variable VAR with one lag (i.e. $p = 5$ and $L = 1$). The size of sample $T = 50$. We choose the weighting matrix $\mathbf{W}$ in the loss function L_{Φ_1} to be the identity matrix. The parameter a in loss L_{Σ_2} is set at -1.50. The true parameters are $\mathbf{\Sigma} = \mathbf{I}_5$, the intercepts are zero, and

$$\mathbf{B}_1 = \begin{pmatrix} .90 & 0 & 0 & 0 & 0 \\ 0 & .92 & 0 & 0 & 0 \\ 0 & 0 & .94 & 0 & 0 \\ 0 & 0 & 0 & .96 & 0 \\ 0 & 0 & 0 & 0 & .98 \end{pmatrix}.$$

We generate $N = 1000$ samples of the five-variable VAR model from the population with the above true parameters. In computation of the Bayesian estimators of $(\mathbf{\Phi}, \mathbf{\Sigma})$, we run $M = 5000$ MCMC cycles after 500 burn-in cycles for each of the 1000 samples. The evaluations of the Bayesian estimators under alternative priors are reported in Tables 1 and 2.

The frequentist risks under loss L, i.e. $E_{\mathbf{Y}|\mathbf{\Phi},\mathbf{\Sigma}} L(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}})$ and $E_{\mathbf{Y}|\mathbf{\Phi},\mathbf{\Sigma}} L(\mathbf{\Phi}, \widehat{\mathbf{\Phi}})$, of MLE and Bayesian estimators using alternative priors on $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ are computed as the average across generated data samples. For instance, $E_{\mathbf{Y}|\mathbf{\Phi},\mathbf{\Sigma}} L(\mathbf{\Phi}, \widehat{\mathbf{\Phi}})$ is computed as $\frac{1}{N}\sum_{j=1}^{N} L(\mathbf{\Phi}, \widehat{\mathbf{\Phi}}_j)$. Here $\widehat{\mathbf{\Phi}}_j$ is Bayesian estimates of sample j $(j = 1, ..., N)$, and $\mathbf{\Phi}$ is the true parameter matrix $\mathbf{B}_1$.

We denote the estimators by the loss function and prior choice. For example, $\widehat{\mathbf{\Sigma}}_{1CA}$ represents the estimator of $\mathbf{\Sigma}$ under loss L_{Σ_1} and the constant-RATS prior combination and $\widehat{\mathbf{\Phi}}_{2SR}$ the estimator of $\mathbf{\Phi}$ under loss L_{Φ_2} and the shrinkage-reference prior combination. Each row of the tables reports the frequentist risks of the corresponding estimator under different loss functions.

Several patterns merge from Tables 1 and 2:

(a) The estimates of $\mathbf{\Sigma}$ are largely independent of the priors on $\mathbf{\Phi}$, and estimates of $\mathbf{\Phi}$ are largely independent of the priors on $\mathbf{\Sigma}$. This is due to the assumption that the loss functions are separable in $\mathbf{\Sigma}$ and $\mathbf{\Phi}$. For example, the $\mathbf{\Sigma}$-related losses of $\widehat{\mathbf{\Sigma}}_{1CA}$ (based on the constant-RATS prior) are quite similar to that of $\widehat{\mathbf{\Sigma}}_{1SA}$ (based on the shrinkage-RATS prior), and the $\mathbf{\Phi}$-related losses of $\widehat{\mathbf{\Phi}}_{1CA}$ (based on the constant-RATS prior) are similar to those of $\widehat{\mathbf{\Phi}}_{1CJ}$ (based on the constant-Jeffreys prior) and $\widehat{\mathbf{\Phi}}_{1CR}$ (based on the constant-reference prior).

(b) Among the estimators for $\mathbf{\Sigma}$, posterior mean is the best overall for the loss functions considered. The finding provides a justification for the common practice of using the posterior mean as the estimator of the covariance matrix. Note that this does not contradict to the optimality of other Bayesian estimators because a Bayesian estimator minimizes the posterior loss over the parameter space, while the frequentist risk only concerns the loss at the true parameter. Another interesting pattern is that under the RATS prior the Bayesian estimators $\widehat{\mathbf{\Sigma}}_{3CA}$ and $\widehat{\mathbf{\Sigma}}_{3SA}$ are very similar to the MLE $\widehat{\mathbf{\Sigma}}_{MLE}$. Whether it suggests a theoretical result remains to be determined by future research.

(c) Under the constant prior, the estimator $\widehat{\mathbf{\Phi}}_2$ obtained from minimizing the LINEX loss function is slightly worse than estimator $\widehat{\mathbf{\Phi}}_1$ in terms of the quadratic loss but it is much better in terms of the LINEX loss. Hence overall $\widehat{\mathbf{\Phi}}_2$ is better than the posterior mean.

(d) In terms of frequentist risk, the influence of prior exceeds that of the loss functions. Table 1 shows that the reference prior dominates other priors on $\mathbf{\Sigma}$. Under the reference prior, with any loss function the average losses associated with $\mathbf{\Sigma}$ are reduced by more than half for all estimators. Table 2 shows that even with strong asymmetry (a= -1.5) in the loss function L_{Φ_2} there is not much difference between the two estimators of $\mathbf{\Phi}$ under the shrinkage prior. But the prior choice is critically important. Under the constant prior both Bayesian estimators perform poorly, even in comparison with the MLE. The large standard deviations of the frequentist losses under constant prior indicate that these Bayesian estimators are not successful in dealing with outliers. In contrast, under the shrinkage prior both estimators dominate the MLE in terms of the frequentist risk associated with $\mathbf{\Phi}$. It is worthwhile to note that under the shrinkage prior the conditional posterior mean of $\boldsymbol{\phi}$ is $\{\mathbf{\Sigma} \otimes (\delta \mathbf{X}'\mathbf{X})^{-1} + \mathbf{I}_J\}^{-1}\widehat{\boldsymbol{\phi}}_M$ (where $\widehat{\boldsymbol{\phi}}_M$ is MLE of $\boldsymbol{\phi}$), which appears to shrink the estimator of $\boldsymbol{\phi}$ towards zero. However, not all elements of shrinkage-based Bayesian estimator of matrix $\mathbf{\Phi}$ are smaller than their MLE counterparts. Some diagonal elements of Bayesian estimates of the first lag coefficients $\mathbf{B_1}$ can be larger than those of the MLE. In VAR models, by construction the regressors and the lags of error terms are correlated, and the MLEs are biased in finite samples. Even with a downward bias of the MLE, the Bayesian estimator under shrinkage prior improves over the MLE by reducing bias for some elements and at the same time substantially reducing variances.

Example 2 For a robustness check, in the next example we change the true parameters of the $\mathbf{\Sigma}$ matrix and keep $\mathbf{\Phi}$ unchanged. We assume that the residuals of VAR variables have different variances and pairwise correlation of 0.5. Specifically, the upper triangle of the covariance is

$$\mathbf{\Sigma} = \begin{pmatrix} 0.500 & 0.354 & 0.433 & 0.500 & 0.559 \\ & 1.0 & 0.612 & 0.707 & 0.791 \\ & & 1.5 & 0.866 & 0.968 \\ & & & 2.0 & 1.118 \\ & & & & 2.5 \end{pmatrix}.$$

The results are different from that in Example 1 in two aspects.

(a) In Table 3 the dominance of the reference prior is no longer as prominent as in Table 1. This is due to the fact that by construction, the reference prior re-parameterizes $\mathbf{\Sigma}$ as $\mathbf{O}'\mathbf{D}\mathbf{O}$ with diagonal matrix $\mathbf{D}$ being the eigenvalues and $\mathbf{O}$ being an orthogonal matrix. The eigenvalues are placed before the orthogonal matrix in the order of importance. By design the data reveal more information pertaining to the diagonal elements. As a result the diagonal elements are estimated more precisely at the expense of the off-diagonal elements. In Example 1 the true $\mathbf{\Sigma}$ matrix is diagonal hence the reference prior does much better.

(b) In Table 4 the shrinkage prior are much better relative to the constant prior. This is because the pairwise correlations of the VAR variables amplify the variances of MLEs and the variances of Bayesian estimators under constant prior. The shrinkage prior is effective in reducing the variances of the Bayesian estimators and results in smaller losses.

Example 3 In this example we consider a five variable VAR(2) model. The covariance matrix $\mathbf{\Sigma}$ is the same as that in Example 2, the intercepts are set at zero for each equation, and the VAR lag coefficients are

$$\mathbf{B}_1 = \begin{pmatrix} .50 & 0 & 0 & 0 & 0 \\ 0 & .52 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .56 & 0 \\ 0 & 0 & 0 & 0 & .58 \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} 0.30 & 0 & 0 & 0 & 0 \\ 0 & 0.32 & 0 & 0 & 0 \\ 0 & 0 & 0.34 & 0 & 0 \\ 0 & 0 & 0 & 0.36 & 0 \\ 0 & 0 & 0 & 0 & 0.38 \end{pmatrix}.$$

Tables 5 and 6 show that the conclusions drawn from the previous VAR(1) examples still hold qualitatively in Example 3. The Bayesian estimators for $\mathbf{\Sigma}$ under the reference prior still dominate other estimators of $\mathbf{\Sigma}$. As in Tables 2 and 4, under the constant prior the LINEX estimator $\widehat{\mathbf{\Phi}}_2$ does better than the posterior mean $\widehat{\mathbf{\Phi}}_1$ overall.

The significance of the role played by the prior choice depends on the sample size of data. With larger sample sizes the importance of prior should diminish. Now we examine the performance of the Bayesian estimators with a larger sample size. We keep the parameters of Example 3 unchanged while increase the sample size from $T = 50$ to $T = 200$.

The results are in Tables 7 and 8. These tables show that with a larger sample size, frequentist risks of alternative estimators for $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ under various priors are all smaller. This is in part due to the fact that with the enlarged sample size the MLE is less erratic and less biased. While the alternative estimators for $\mathbf{\Sigma}$ yield losses similar in magnitude under all prior combinations, the estimators for $\mathbf{\Phi}$ under the shrinkage prior yield losses only half as much as those under the constant prior. In practice, there is no precise guidance on whether the sample size is large enough to render the choice of prior and the form of estimator unimportant. In this example it appears that 200 observations are adequate. But for a VAR of more variables (larger p), longer lags (larger L), or with different parameters in $\mathbf{B}_1$ or $\mathbf{B}_2$ (e.g., when the VAR is explosive rather than stationary), different choice of prior and loss function may produce vastly different estimates with a data sample of 200 observations.

In the next section we will examine competing estimators under alternative priors in a VAR of the U.S. economy with a data sample of 172 periods.

5 Estimating a VAR of the U.S. Economy

In the past two decades, VAR models have been commonly used for analyzing multivariate time series macroeconomic data and addressing policy questions. Given the fact that macroeconomic data are limited in availability, variable selection for the VARs is always a result of researchers' balancing act between two conflicting considerations. On the one hand, the number of parameters (for $\mathbf{\Sigma}$ and $\mathbf{\Phi}$) in a VAR of p variables and L lags to be estimated is $p(p+1)/2 + p(Lp+1)$, hence including more variables significantly expands the total parameters and the finite sample inferences become less robust; on the other hand, omitting useful variables results in an incorrect model. The Bayesian approach for finite sample inferences may help to mitigate the difficulty a VAR practitioner faces. However, the Bayesian inferences depend on the loss function and prior choice, it is useful to examine whether the commonly adopted Bayesian estimators such as the posterior mean and priors such as the constant-RATS prior give rise to estimates that are similar to alternative estimators under alternative priors. In the following, we compare various Bayesian estimates of a six-variable VAR using quarterly data of the U.S. economy from 1959Q1 to 2001Q4. The lag length of the VAR is two.[1] The variables are real GDP, GDP deflator, world commodity price, M2 money stock, non-borrowed reserves, and the federal funds rate. The commodity price data are obtained from the International Monetary Fund, the rest of data series from the FRED database at the Federal Reserve Bank of St Louis. All variables except the fed funds rate are in logarithms. These variables frequently appear in macroeconomics related VARs (e.g. Sims 1992, Gordon and Leeper 1994, Sims and Zha 1998, and Christiano, Eichenbaum, and Evans 1999). To measure the difference of the alternative Bayesian estimates, we use the posterior mean of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ under the constant-RATS prior combination $(\widehat{\boldsymbol{\Sigma}}_{1CA}, \widehat{\boldsymbol{\Phi}}_{1CA})$ as the benchmark for $(\widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\Phi}})$, and use the loss functions to measure the distance between other estimates from the benchmark estimates. Tables 9 and 10 report the distance between the benchmark Bayesian estimator and alternative estimators. For example, the $\widehat{\boldsymbol{\Sigma}}_{3SR}$ row and L_{Σ_2} column of Table 9 reports the distance $L_{\Sigma_2}(\widehat{\boldsymbol{\Sigma}}_{1CA}, \widehat{\boldsymbol{\Sigma}}_{3SR})$. The main finding in the tables is that the estimates for $\boldsymbol{\Sigma}$ are similar but the estimates are distinctly different for $\boldsymbol{\Phi}$ under constant and shrinkage priors even with a short VAR lag of two.

Our benchmark estimates are as follows. Under the constant-RATS prior, the posterior means are

$$\widehat{\boldsymbol{\Sigma}}_{1CA} = \begin{pmatrix} 0.0037 & 0.0015 & 0.0013 & 0.0028 & 0.0006 & 0.0001 \\ 0.0015 & 0.0006 & 0.0006 & 0.0011 & 0.0003 & 0.0000 \\ 0.0013 & 0.0006 & 0.0020 & 0.0010 & 0.0002 & 0.0001 \\ 0.0028 & 0.0011 & 0.0010 & 0.0021 & 0.0005 & 0.0000 \\ 0.0006 & 0.0003 & 0.0002 & 0.0005 & 0.0013 & -0.0001 \\ 0.0001 & 0.0000 & 0.0001 & 0.0000 & -0.0001 & 0.0001 \end{pmatrix},$$

[1] A four-lag VAR produces similar results.

$$\widehat{\mathbf{\Phi}}_{1CA} = \begin{pmatrix} 7.6508 & 3.0508 & 3.489 & 5.6024 & 2.3428 & 0.0246 \\ -1.4406 & -1.1026 & -1.2361 & -1.9825 & -0.7285 & 0.1141 \\ 3.0555 & 2.6087 & 1.0809 & 2.0534 & 2.3192 & -0.4009 \\ -0.2109 & -0.0771 & 1.2914 & -0.1748 & -0.1646 & 0.0629 \\ 0.6838 & 0.2295 & 0.4314 & 1.8888 & -0.5640 & 0.0238 \\ -0.5571 & -0.2194 & -0.3025 & -0.4265 & 0.9542 & -0.0092 \\ -0.2794 & 0.0048 & -0.2142 & -0.3078 & -0.4271 & 1.0502 \\ 0.9843 & 0.5174 & 0.5669 & 0.9140 & 0.2875 & -0.1266 \\ -3.7121 & -1.9043 & -.5284 & -2.5485 & -2.4442 & 0.3636 \\ 0.0551 & 0.0248 & -0.3735 & 0.066 & 0.0833 & -0.0432 \\ 0.3835 & 0.2135 & 0.1359 & -0.0945 & 0.8896 & -0.0012 \\ 0.6352 & 0.2513 & 0.366 & 0.4722 & 0.012 & 0.0055 \\ -0.1588 & -0.1062 & 0.0776 & 0.0439 & 0.2301 & -0.1291 \end{pmatrix}.$$

Tables 9 and 10 show that overall the estimates under the shrinkage-reference prior are most different from the benchmark. To explore the difference in the estimates we report the posterior means under the shrinkage-reference prior.

$$\widehat{\mathbf{\Sigma}}_{1SR} = \begin{pmatrix} 0.0039 & 0.0016 & 0.0014 & 0.0029 & 0.0006 & 0.0001 \\ 0.0016 & 0.0007 & 0.0006 & 0.0012 & 0.0003 & 0.0000 \\ 0.0014 & 0.0006 & 0.0020 & 0.0011 & 0.0003 & 0.0001 \\ 0.0029 & 0.0012 & 0.0011 & 0.0022 & 0.0005 & 0.0000 \\ 0.0006 & 0.0003 & 0.0003 & 0.0005 & 0.0014 & -0.0001 \\ 0.0001 & 0.0000 & 0.0001 & 0.0000 & -0.0001 & 0.0001 \end{pmatrix},$$

$$\widehat{\mathbf{\Phi}}_{1SR} = \begin{pmatrix} 7.6150 & 3.0361 & 3.4727 & 5.5752 & 2.3341 & 0.0244 \\ -0.3538 & -0.6538 & -0.7502 & -1.1513 & -0.4358 & 0.1169 \\ 0.3872 & 1.5022 & -0.0621 & 0.0076 & 1.1720 & -0.3611 \\ -0.1996 & -0.0725 & 1.2890 & -0.1660 & -0.1581 & 0.0621 \\ 0.6114 & 0.2021 & 0.3767 & 1.8362 & -0.3544 & -0.0012 \\ -0.4598 & -0.1792 & -0.2597 & -0.3524 & 0.9843 & -0.0094 \\ -0.0991 & 0.0797 & -0.1261 & -0.1686 & -0.3043 & 1.0398 \\ -0.1113 & 0.0649 & 0.0774 & 0.0760 & -0.0094 & -0.1293 \\ -1.1606 & -0.8463 & -0.4363 & -0.5924 & -1.3471 & 0.3255 \\ 0.0866 & 0.0379 & -0.3531 & 0.0900 & 0.0929 & -0.0428 \\ 0.5045 & 0.2610 & 0.2123 & -0.0045 & 0.6985 & 0.0234 \\ 0.5474 & 0.2152 & 0.3268 & 0.4054 & -0.0074 & 0.0048 \\ 0.0185 & -0.0324 & 0.1429 & 0.1788 & 0.2674 & -0.1256 \end{pmatrix}.$$

Under alternative priors the difference in estimates for $\mathbf{\Sigma}$ is negligible. On the other hand the difference in the estimates for $\mathbf{\Phi}$ is quite large:

$$\widehat{\mathbf{\Phi}}_{1CA} - \widehat{\mathbf{\Phi}}_{1SR} = \begin{pmatrix} 0.0358 & 0.0147 & 0.0163 & 0.0272 & 0.0087 & 0.0002 \\ -1.0868 & -0.4488 & -0.4859 & -0.8312 & -0.2927 & -0.0028 \\ 2.6683 & 1.1065 & 1.1430 & 2.0458 & 1.1472 & -0.0398 \\ -0.0113 & -0.0046 & 0.0024 & -0.0088 & -0.0065 & 0.0008 \\ 0.0724 & 0.0274 & 0.0547 & 0.0526 & -0.2096 & 0.0250 \\ -0.0973 & -0.0402 & -0.0428 & -0.0741 & -0.0301 & 0.0002 \\ -0.1803 & -0.0749 & -0.0881 & -0.1392 & -0.1228 & 0.0104 \\ 1.0956 & 0.4525 & 0.4895 & 0.8380 & 0.2969 & 0.0027 \\ -2.5515 & -1.0580 & -1.0921 & -1.9561 & -1.0971 & 0.0381 \\ -0.0315 & -0.0131 & -0.0204 & -0.0240 & -0.0096 & -0.0004 \\ -0.1210 & -0.0475 & -0.0764 & -0.0900 & 0.1911 & -0.0246 \\ 0.0878 & 0.0361 & 0.0392 & 0.0668 & 0.0194 & 0.0007 \\ -0.1773 & -0.0738 & -0.0653 & -0.1349 & -0.0373 & -0.0035 \end{pmatrix}.$$

The most prominent differences of the two estimates under the alternative priors are reflected in the third and the ninth rows of $\mathbf{\Phi}$, which correspond to the first and second lag parameters of the GDP deflator equation. As in our numerical examples, under the constant-RATS prior the posterior mean of $\mathbf{\Phi}$ is very close to the MLE. Most elements of the third row of $\widehat{\mathbf{\Phi}}_{1CA}$ are in similar magnitude with the elements in the ninth row of the same column, but have the opposite signs. The pattern of estimates also emerge in the GDP equation (of the second and eighth rows) and the non-borrowed reserves equation (of the sixth and the twelfth rows). The pattern of estimates of this VAR is not uncommon for macroeconomic applications of VARs because macroeconomic time series data often exhibit strong serial and pairwise correlations and the VAR models are 'over-parameterized' with no restrictions on the matrix $\mathbf{\Phi}$. In empirical applications such as the present one, the MLE estimates of the first and second lag coefficients are not only in similar magnitude and opposite signs, they are also often found to be very sensitive to model specification and sample period. The fact the estimates of $\mathbf{\Phi}$ under the constant-RATS prior combination are similar to MLE suggests a possibility of improvement by using alternative priors in place of the constant prior. As is discussed earlier, the shrinkage estimator of James and Stein (1961), which motivates the the shrinkage prior, reduces quadratic frequentist loss in estimation of multivariate normal mean. The James-Stein estimator is known for improving efficiency in the presence of multicollinearity in regression models. Assessing the improvement of shrinkage-prior-based estimators requires computation of losses. The frequentist risks cannot be calculated since the true parameters are unknown in the application and the theoretical results on admissibility of estimators under alternative priors have not been established in the VAR framework, but most likely the use of shrinkage prior for $\mathbf{\Phi}$ improves upon the MLE and the Bayesian estimators based on the constant prior.

Unlike the frequentist risk, the posterior risk for the given data sample can be computed using the MCMC simulation output. For some loss functions this can be done at little added cost. For example, with the posterior mean estimator $\widehat{\mathbf{\Sigma}} = \mathbb{E}(\mathbf{\Sigma} \mid \mathbf{Y})$ the posterior loss of $L_{\Sigma_1}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})$ is

$$\mathbb{E}\Big[\{tr(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \log|\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}| - p\} \mid \mathbf{Y}\Big] = \log|\widehat{\mathbf{\Sigma}}| - \mathbb{E}\{(\log|\mathbf{\Sigma}|) \mid \mathbf{Y}\}. \tag{28}$$

To compute $\mathbb{E}\{(\log|\mathbf{\Sigma}|) \mid \mathbf{Y}\}$, we decompose the $\mathbf{\Sigma}_k$ matrix in the kth MCMC cycle as $\mathbf{\Sigma}_k = \mathbf{O}_k\mathbf{D}_k\mathbf{O}_k'$, where $\mathbf{D}_k$ is the diagonal matrix that consists of eigenvalues of $\mathbf{\Sigma}_k$, i.e., $D_k = diag(d_{k1}, d_{k2}, .., d_{kp})$, and $\mathbf{O}_k$ is an orthogonal matrix with $\mathbf{O}_k\mathbf{O}_k' = \mathbf{I}$. It follows that $\widehat{\mathbb{E}}\{(\log|\mathbf{\Sigma}|) \mid \mathbf{Y}\} = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{p} log|d_{ki}|$, which can be computed in MCMC runs without restoring the simulated $\mathbf{\Sigma}_k$ matrices for $k = 1, ..., M$. The posterior risk of the estimator $\widehat{\mathbf{\Phi}} = \mathbb{E}(\mathbf{\Phi} \mid \mathbf{Y})$ associated with the loss $L_{\Phi_1}(\widehat{\mathbf{\Phi}}, \mathbf{\Phi})$ is

$$\mathbb{E}\Big[tr\{(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})'(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\} \mid \mathbf{Y}\Big] \quad = \quad tr\Big[\mathbb{E}(\mathbf{\Phi}'\mathbf{\Phi} \mid \mathbf{Y}) - \{\mathbb{E}(\mathbf{\Phi} \mid \mathbf{Y})\}'\{\mathbb{E}(\mathbf{\Phi} \mid \mathbf{Y})\}\Big]. \tag{29}$$

Both $\mathbb{E}(\mathbf{\Phi} \mid \mathbf{Y})$ and $\mathbb{E}(\mathbf{\Phi}'\mathbf{\Phi} \mid \mathbf{Y})$ can be computed in the process of MCMC simulations without the storage of the entire MCMC outputs.

The posterior risks of the posterior mean estimator under alternative priors turn out to be quite different. For instance, under the constant-RATS prior the posterior risks (**??**) and (**??**) are 0.126 and 13.703, respectively. Under the shrinkage-reference prior combination, the corresponding posterior losses are 0.123 and 6.470, respectively. The posterior losses under the shrinkage-reference prior are smaller because the posterior distributions are tighter for the distribution of the VAR lag coefficients $\mathbf{\Phi}$.

6 Concluding Remarks

The paper compares frequentist risks of several Bayesian estimators of the VAR lag parameters $\mathbf{\Phi}$ and covariance matrix $\mathbf{\Sigma}$ under alternative priors. With the constant prior on $\mathbf{\Phi}$, the asymmetric LINEX estimator $\mathbf{\Phi}$ does better overall than the posterior mean. With the shrinkage prior on $\mathbf{\Phi}$ the LINEX estimator and the posterior mean yield similar losses. The posterior mean of $\mathbf{\Sigma}$ performs well in most cases. The choice of prior has more significant effects on the Bayesian estimates than the choice of loss function. The shrinkage prior on $\mathbf{\Phi}$ dominates the constant prior, while Yang and Berger's reference prior on $\mathbf{\Sigma}$ dominates the Jeffreys prior and the RATS prior. Estimation of a VAR using the U.S. macroeconomic data reveals significant difference between estimates under the shrinkage and constant priors.

The study may be extended in several directions. First, note that the list of noninformative priors examined in the present paper is by no means exhaustive. Other noninformative priors applicable to the VAR framework need to be explored. Second, in this paper Bayesian estimators are derived from loss functions separable in $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. As a result, the inferences of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are largely independent. Our future research concerns joint Bayesian inferences of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ based on an intrinsic loss function such as the entropy loss. Derivation and computation of Bayesian estimators are more challenging under the entropy loss on $(\mathbf{\Phi}, \mathbf{\Sigma})$ because the joint loss is non-separable in $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ and it involves computing moments of sample distribution for which closed-form expressions are unavailable.

References

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd edition, Wiley, New York.

Baranchik, A. J. (1964). Multiple regression and estimation of the mean of multivariate normal distribution. Technical Report 51, Dept. Statistics, Stanford University.

Berger, J.O. and Bernardo, J.M. (1992). On the development of reference priors. In *Bayesian Analysis IV*, J.M. Bernardo, et. al., (Eds.). Oxford University Press, Oxford.

Berger, J.O. and Strawderman, W.E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Annals of Statistics,* **24,** 931-951.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B*, **41**, 113-147.

Christiano, L. J., Eichenbaum, M. and Evans C. (1999), Monetary policy shocks: What have we learned and to what end? in: J.B. Taylor and M. Woodford eds., *Handbook of Macroeconomics*, Volume **1**, pp. 65–147.

Geisser, S. (1965). Bayesian estimation in multivariate analysis. *Annals of Mathematical Statistics* **36**, 150-159.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

Gordon, D. B. and Leeper, E. M. (1994), The dynamic impacts of monetary policy: an exercise in tentative identification, *Journal of Political Economy* **102**, 1228–1247.

James, W. and Stein, C. (1961) Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematics, statistics, and probability.* **1**, 361-380. University of California Press, Berkeley.

Jeffreys, H. (1961) *Probability Theory.* Oxford University Press, New York.

Kadiyala, K.R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics.* **12,** 99-132.

Kass, R.E. and Wasserman, L. (1996). The selection of prior distributions by formal rules, *J. Amer. Statist. Assoc.*, **91**, 1343-1370.

Ni, S. and D. Sun (2001). A Monte Carlo study on frequentist risks of Bayesian estimators of vector-autoregressive models based on noninformative priors. Submitted.

Sims, Christopher A. (1992), Interpreting the macroeconomic time series facts: The effects of monetary policy, *European Economic Review* **38**, pp. 975 – 1000.

Sims, C.A. and Zha T. (1998). Does monetary policy generate recessions? *Federal Reserve Bank of Atlanta working paper* **98-12**.

Sims, C. A. and Zha T. (1999). Error Bands for Impulse Responses, *Econometrica*, **67**, 1113–1155.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the third Berkeley symposium*, **Vol. 1**, 197-206, Berkeley: University of California Press.

Sun, D. and Berger, J.O. (1998). Reference priors under partial information. *Biometrika*, **85**, 55-71.

Tiao, G.C. and Zellner, A. (1964). On the Bayesian estimation analysis of multivariate regression, *Journal of Royal Statistical Society, B,* **26,** 389-399.

Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *Annals of Statistics,* **22,** 1195-1211.

Zellner, A (1971). *An Introduction to Bayesian Inference in Econometrics.* John Wiley & Sons, New York.

Zellner, A (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association,* **81**. 446-451.

Table 1: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Sigma}$ in Example 1. (Standard deviations are in parentheses.)

	L_{Σ_1}	L_{Σ_2}	L_{Σ_3}
$\widehat{\mathbf{\Sigma}}_{MLE}$	.7690 (.3723)	.6570 (.1950)	.4764 (.1706)
$\widehat{\mathbf{\Sigma}}_{1CA}$	.5512 (.2847)	.6635 (.2505)	.3968 (.1508)
$\widehat{\mathbf{\Sigma}}_{2CA}$	1.0595 (.4645)	.7528 (.1888)	.6035 (.1951)
$\widehat{\mathbf{\Sigma}}_{3CA}$	.7691 (.3724)	.6571 (.1953)	.4765 (.1706)
$\widehat{\mathbf{\Sigma}}_{1CJ}$	.4273 (.2066)	.8914 (.4053)	.4002 (.1518)
$\widehat{\mathbf{\Sigma}}_{2CJ}$	.7692 (.3724)	.6572 (.1953)	.4766 (.1706)
$\widehat{\mathbf{\Sigma}}_{3CJ}$	.5513 (.2847)	.6638 (.2508)	.3969 (.1509)
$\widehat{\mathbf{\Sigma}}_{1CR}$	.1660 (.1279)	.2720 (.1835)	.1441 (.0942)
$\widehat{\mathbf{\Sigma}}_{2CR}$	.3300 (.2529)	.3375 (.1781)	.2294 (.1431)
$\widehat{\mathbf{\Sigma}}_{3CR}$	.2284 (.1825)	.2794 (.1620)	.1729 (.1130)
$\widehat{\mathbf{\Sigma}}_{1SA}$	.5515 (.2847)	.6636 (.2499)	.3970 (.1508)
$\widehat{\mathbf{\Sigma}}_{2SA}$	1.0598 (.4647)	.7527 (.1887)	.6035 (.1951)
$\widehat{\mathbf{\Sigma}}_{3SA}$	.7694 (.3726)	.6570 (.1949)	.4765 (.1706)
$\widehat{\mathbf{\Sigma}}_{1SJ}$	.4271 (.2062)	.8908 (.4040)	.4000 (.1513)
$\widehat{\mathbf{\Sigma}}_{2SJ}$	.7693 (.3722)	.6572 (.1952)	.4766 (.1706)
$\widehat{\mathbf{\Sigma}}_{3SJ}$	.5512 (.2844)	.6636 (.2502)	.3969 (.1507)
$\widehat{\mathbf{\Sigma}}_{1SR}$	.1582 (.1248)	.2589 (.1721)	.1375 (.0903)
$\widehat{\mathbf{\Sigma}}_{2SR}$	.3169 (.2436)	.3271 (.1707)	.2214 (.1374)
$\widehat{\mathbf{\Sigma}}_{3SR}$	.2185 (.1764)	.2688 (.1538)	.1660 (.1084)

Table 2: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Phi}$ in Example 1. (Standard deviations are in parentheses.)

	L_{Φ_1}	L_{Φ_2}
$\widehat{\mathbf{\Phi}}_{MLE}$	3.9199 (3.8931)	8.7284 (32.8418)
$\widehat{\mathbf{\Phi}}_{1CA}$	3.9192(3.8866)	8.7075 (32.3707)
$\widehat{\mathbf{\Phi}}_{2CA}$	4.2583 (4.8835)	5.5381 (12.7549)
$\widehat{\mathbf{\Phi}}_{1CJ}$	3.9202 (3.8936)	8.7546 (33.3711)
$\widehat{\mathbf{\Phi}}_{2CJ}$	4.3768 (5.2561)	5.2834 (11.1703)
$\widehat{\mathbf{\Phi}}_{1CR}$	3.9197 (3.8896)	8.7037 (32.5073)
$\widehat{\mathbf{\Phi}}_{2CR}$	4.2689 (4.9135)	5.2939 (10.8249)
$\widehat{\mathbf{\Phi}}_{1SA}$	1.2523 (.5009)	1.5511 (.7845)
$\widehat{\mathbf{\Phi}}_{2SA}$	1.2725 (.5422)	1.4603 (.6438)
$\widehat{\mathbf{\Phi}}_{1SJ}$	1.1673 (.4264)	1.4364 (.6442)
$\widehat{\mathbf{\Phi}}_{2SJ}$	1.1870 (.4648)	1.3570 (.5347)
$\widehat{\mathbf{\Phi}}_{1SR}$	1.0743 (.3071)	1.3079 (.4348)
$\widehat{\mathbf{\Phi}}_{2SR}$	1.0909 (.3340)	1.2465 (.3756)

Table 3: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Sigma}$ in Example 2. (Standard deviations are in parentheses.)

	L_{Σ_1}	L_{Σ_2}	L_{Σ_3}
$\widehat{\mathbf{\Sigma}}_{MLE}$	.7683 (.3626)	.6560 (.1937)	.4753 (.1651)
$\widehat{\mathbf{\Sigma}}_{1CA}$	.5518 (.2751)	.6649 (.2556)	.3969 (.1446)
$\widehat{\mathbf{\Sigma}}_{2CA}$	1.0578 (.4548)	.7505 (.1863)	.6015 (.1912)
$\widehat{\mathbf{\Sigma}}_{3CA}$	.7685 (.3627)	.6561 (.1940)	.4754 (.1653)
$\widehat{\mathbf{\Sigma}}_{1CJ}$	.4290 (.1977)	.8964 (.4211)	.4017 (.1470)
$\widehat{\mathbf{\Sigma}}_{2CJ}$	.7685 (.3627)	.6561 (.1939)	.4754 (.1652)
$\widehat{\mathbf{\Sigma}}_{3CJ}$	.5519 (.2750)	.6649 (.2554)	.3969 (.1446)
$\widehat{\mathbf{\Sigma}}_{1CR}$	.3859 (.1727)	.7716 (.3442)	.3563 (.1232)
$\widehat{\mathbf{\Sigma}}_{2CR}$	.6556 (.3117)	.6031 (.1803)	.4223 (.1481)
$\widehat{\mathbf{\Sigma}}_{3CR}$	.4842 (.2367)	.6072 (.2237)	.3577 (.1261)
$\widehat{\mathbf{\Sigma}}_{1SA}$	.5518 (.2751)	.6650 (.2557)	.3969 (.1446)
$\widehat{\mathbf{\Sigma}}_{2SA}$	1.0578 (.4551)	.7504 (.1863)	.6015 (.1912)
$\widehat{\mathbf{\Sigma}}_{3SA}$	.7685 (.3629)	.6561 (.1938)	.4754 (.1652)
$\widehat{\mathbf{\Sigma}}_{1SJ}$	.4290 (.1979)	.8959 (.4216)	.4015 (.1472)
$\widehat{\mathbf{\Sigma}}_{2SJ}$	.7687 (.3628)	.6562 (.1938)	.4755 (.1652)
$\widehat{\mathbf{\Sigma}}_{3SJ}$	.5519 (.2751)	.6648 (.2554)	.3969 (.1446)
$\widehat{\mathbf{\Sigma}}_{1SR}$	.3708 (.1666)	.7502 (.3534)	.3452 (.1253)
$\widehat{\mathbf{\Sigma}}_{2SR}$	.6343 (.3011)	.5925 (.1827)	.4122 (.1459)
$\widehat{\mathbf{\Sigma}}_{3SR}$	.4666 (.2285)	.5938 (.2303)	.3478 (.1260)

Table 4: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Phi}$ in Example 2. (Standard deviations are in parentheses.)

	L_{Φ_1}	L_{Φ_2}
$\widehat{\mathbf{\Phi}}_{MLE}$	6.3680 (7.3344)	37.5860 (379.1343)
$\widehat{\mathbf{\Phi}}_{1CA}$	6.3704(7.3302)	37.3729 (373.5502)
$\widehat{\mathbf{\Phi}}_{2CA}$	7.0974 (10.0639)	12.0093 (59.9915)
$\widehat{\mathbf{\Phi}}_{1CJ}$	6.3652 (7.3279)	37.5890 (380.1670)
$\widehat{\mathbf{\Phi}}_{2CJ}$	7.4269 (11.4723)	10.4756 (46.1726)
$\widehat{\mathbf{\Phi}}_{1CR}$	6.3682 (7.3333)	37.4291 (376.6031)
$\widehat{\mathbf{\Phi}}_{2CR}$	7.1591 (10.4047)	11.6565 (55.5603)
$\widehat{\mathbf{\Phi}}_{1SA}$	1.3037 (.4934)	1.5886 (.6739)
$\widehat{\mathbf{\Phi}}_{2SA}$	1.3180 (.5226)	1.4953 (.5833)
$\widehat{\mathbf{\Phi}}_{1SJ}$	1.2151 (.4327)	1.4748 (.5785)
$\widehat{\mathbf{\Phi}}_{2SJ}$	1.2302 (.4595)	1.3915 (.5078)
$\widehat{\mathbf{\Phi}}_{1SR}$	1.1849 (.3824)	1.4325 (.5034)
$\widehat{\mathbf{\Phi}}_{2SR}$	1.1989 (.4049)	1.3572 (.4497)

Table 5: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Sigma}$ in Example 3, with sample size T=50. (Standard deviations are in parentheses.)

	L_{Σ_1}	L_{Σ_2}	L_{Σ_3}
$\widehat{\mathbf{\Sigma}}_{MLE}$	1.1540 (.5251)	.7887 (.1939)	.6396 (.2051)
$\widehat{\mathbf{\Sigma}}_{1CA}$	.8220 (.4141)	.7085 (.2173)	.5050 (.1779)
$\widehat{\mathbf{\Sigma}}_{2CA}$	1.5588 (.6398)	.9351 (.2008)	.8100 (.2325)
$\widehat{\mathbf{\Sigma}}_{3CA}$	1.1541 (.5254)	.7888 (.1940)	.6397 (.2052)
$\widehat{\mathbf{\Sigma}}_{1CJ}$	.4797 (.2333)	1.1070 (.5325)	.4678 (.1745)
$\widehat{\mathbf{\Sigma}}_{2CJ}$	.8716 (.4323)	.7144 (.2092)	.5240 (.1822)
$\widehat{\mathbf{\Sigma}}_{3CJ}$	.6158 (.3261)	.7535 (.2987)	.4408 (.1623)
$\widehat{\mathbf{\Sigma}}_{1CR}$	.4192 (.1936)	.9160 (.4295)	.4020 (.1460)
$\widehat{\mathbf{\Sigma}}_{2CR}$	.7184 (.3594)	.6467 (.1905)	.4549 (.1614)
$\widehat{\mathbf{\Sigma}}_{3CR}$	.5238 (.2686)	.6740 (.2547)	.3885 (.1394)
$\widehat{\mathbf{\Sigma}}_{1SA}$	.8221 (.4136)	.7086 (.2174)	.5051 (.1778)
$\widehat{\mathbf{\Sigma}}_{2SA}$	1.5590 (.6393)	.9351 (.2006)	.8100 (.2323)
$\widehat{\mathbf{\Sigma}}_{3SA}$	1.1543 (.5249)	.7888 (.1939)	.6398 (.2050)
$\widehat{\mathbf{\Sigma}}_{1SJ}$	.4794 (.2329)	1.1066 (.5322)	.4676 (.1745)
$\widehat{\mathbf{\Sigma}}_{2SJ}$	.8715 (.4321)	.7144 (.2091)	.5239 (.1821)
$\widehat{\mathbf{\Sigma}}_{3SJ}$	.6156 (.3258)	.7534 (.2986)	.4407 (.1623)
$\widehat{\mathbf{\Sigma}}_{1SR}$	.3950 (.1895)	.8011 (.3607)	.3665 (.1307)
$\widehat{\mathbf{\Sigma}}_{2SR}$	.6881 (.3465)	.6197 (.1837)	.4379 (.1575)
$\widehat{\mathbf{\Sigma}}_{3SR}$	.5010 (.2616)	.6231 (.2283)	.3678 (.1335)

Table 6: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Phi}$ in Example 3, with sample size T=50. (Standard deviations are in parentheses.)

	L_{Φ_1}	L_{Φ_2}
$\widehat{\mathbf{\Phi}}_{MLE}$	7.5518 (6.4369)	23.5908 (219.3303)
$\widehat{\mathbf{\Phi}}_{1CA}$	7.5513(6.4312)	28.6443 (219.6075)
$\widehat{\mathbf{\Phi}}_{2CA}$	8.2089 (9.1060)	11.7866 (36.0243)
$\widehat{\mathbf{\Phi}}_{1CJ}$	7.5583 (6.4506)	28.4572 (211.7202)
$\widehat{\mathbf{\Phi}}_{2CJ}$	8.7972 (11.4382)	9.9813 (21.3180)
$\widehat{\mathbf{\Phi}}_{1CR}$	7.5573 (6.4484)	28.9039 (227.2678)
$\widehat{\mathbf{\Phi}}_{2CR}$	8.4998 (10.3763)	11.1306 (34.2694)
$\widehat{\mathbf{\Phi}}_{1SA}$	1.0140 (.2642)	1.1986 (.3224)
$\widehat{\mathbf{\Phi}}_{2SA}$	1.0041 (.2632)	1.1591 (.3063)
$\widehat{\mathbf{\Phi}}_{1SJ}$	.8833 (.2073)	1.0484 (.2534)
$\widehat{\mathbf{\Phi}}_{2SJ}$	.8709 (.2064)	1.0091 (.2414)
$\widehat{\mathbf{\Phi}}_{1SR}$	.8608 (.1876)	1.0210 (.2306)
$\widehat{\mathbf{\Phi}}_{2SR}$	.8485 (.1864)	.9837 (.2200)

Table 7: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Sigma}$ in Example 3, with sample size T=200. (Standard deviations are in parentheses.)

	L_{Σ_1}	L_{Σ_2}	L_{Σ_3}
$\widehat{\mathbf{\Sigma}}_{MLE}$	.1042 (.0444)	.1584 (.0536)	.0896 (.0332)
$\widehat{\mathbf{\Sigma}}_{1CA}$	.0917 (.0388)	.1539 (.0547)	.0826 (.0307)
$\widehat{\mathbf{\Sigma}}_{2CA}$	.1213 (.0506)	.1706 (.0554)	.0896 (.0366)
$\widehat{\mathbf{\Sigma}}_{3CA}$	.1043 (.0444)	.1584 (.0536)	.0896 (.0332)
$\widehat{\mathbf{\Sigma}}_{1CJ}$	.0813 (.0315)	.1711 (.0671)	.0819 (.0301)
$\widehat{\mathbf{\Sigma}}_{2CJ}$	.0935 (.0396)	.1541 (.0543)	.0835 (.0310)
$\widehat{\mathbf{\Sigma}}_{3CJ}$	.0849 (.0349)	.1572 (.0586)	.0802 (.0296)
$\widehat{\mathbf{\Sigma}}_{1CR}$	.0832 (.0310)	.1759 (.0680)	.0840 (.0300)
$\widehat{\mathbf{\Sigma}}_{2CR}$	.0943 (.0386)	.1579 (.0550)	.0849 (.0306)
$\widehat{\mathbf{\Sigma}}_{3CR}$	.0864 (.0341)	.1619 (.0596)	.0821 (.0295)
$\widehat{\mathbf{\Sigma}}_{1SA}$	.0917 (.0388)	.1539 (.0547)	.0826 (.0307)
$\widehat{\mathbf{\Sigma}}_{2SA}$	.1213 (.0506)	.1706 (.0554)	.1006 (.0366)
$\widehat{\mathbf{\Sigma}}_{3SA}$	.1043 (.0444)	.1584 (.0536)	.0896 (.0332)
$\widehat{\mathbf{\Sigma}}_{1SJ}$	.0814 (.0315)	.1712 (.0672)	.0820 (.0301)
$\widehat{\mathbf{\Sigma}}_{2SJ}$	.0935 (.0397)	.1541 (.0543)	.0835 (.0310)
$\widehat{\mathbf{\Sigma}}_{3SJ}$	.0849 (.0349)	.1573 (.0586)	.0803 (.0297)
$\widehat{\mathbf{\Sigma}}_{1SR}$	.0832 (.0318)	.1728 (.0671)	.0833 (.0301)
$\widehat{\mathbf{\Sigma}}_{2SR}$	.0956 (.0396)	.1577 (.0548)	.0854 (.0311)
$\widehat{\mathbf{\Sigma}}_{3SR}$	.0871 (.0351)	.1604 (.0590)	.0821 (.0298)

Table 8: Frequentist Risks of Alternative Bayesian Estimators of $\mathbf{\Phi}$ in Example 3, with sample size T=200. (Standard deviations are in parentheses.)

	L_{Φ_1}	L_{Φ_2}
$\widehat{\mathbf{\Phi}}_{MLE}$	.7602 (.3847)	.8748 (.5133)
$\widehat{\mathbf{\Phi}}_{1CA}$	.7601 (.3841)	.8746 (.5105)
$\widehat{\mathbf{\Phi}}_{2CA}$	.7661 (.3871)	.8591 (.4708)
$\widehat{\mathbf{\Phi}}_{1CJ}$	.7603 (.3844)	.8747 (.5112)
$\widehat{\mathbf{\Phi}}_{2CJ}$	.7670 (.3877)	.8589 (.4695)
$\widehat{\mathbf{\Phi}}_{1CAR}$	.7602 (.3848)	.8749 (.5129)
$\widehat{\mathbf{\Phi}}_{2CR}$	.7665 (.3876)	.8591 (.4717)
$\widehat{\mathbf{\Phi}}_{1SA}$	.3000 (.0816)	.3416 (.0936)
$\widehat{\mathbf{\Phi}}_{2SA}$	.2995 (.0821)	.3376 (.0918)
$\widehat{\mathbf{\Phi}}_{1SJ}$	.2925 (.0786)	.3330 (.0901)
$\widehat{\mathbf{\Phi}}_{2SJ}$	.2919 (.0791)	.3290 (.0883)
$\widehat{\mathbf{\Phi}}_{1SR}$	.2959 (.0793)	.3369 (.0909)
$\widehat{\mathbf{\Phi}}_{2SR}$	.2953 (.0799)	.3329 (.0892)

Table 9: Distance Between the Benchmark Bayesian Estimator $\widehat{\mathbf{\Sigma}}_{1CA}$ and Alternative Bayesian Estimators of $\mathbf{\Sigma}$, U.S. Quarterly Data 1959Q1-2001Q4.

	L_{Σ_1}	L_{Σ_2}	L_{Σ_3}
$\widehat{\mathbf{\Sigma}}_{MLE}$	.0053	.0100	.1603
$\widehat{\mathbf{\Sigma}}_{1CA}$	0	0	0
$\widehat{\mathbf{\Sigma}}_{2CA}$	.0053	.0100	.0051
$\widehat{\mathbf{\Sigma}}_{3CA}$	.0053	.0100	.0051
$\widehat{\mathbf{\Sigma}}_{1CJ}$	.0193	.0431	.0204
$\widehat{\mathbf{\Sigma}}_{2CJ}$	.0001	.0003	.0001
$\widehat{\mathbf{\Sigma}}_{3CJ}$	.0040	.0084	.0041
$\widehat{\mathbf{\Sigma}}_{1CR}$	.0322	.0773	.0353
$\widehat{\mathbf{\Sigma}}_{2CR}$	.0054	.0115	.0056
$\widehat{\mathbf{\Sigma}}_{3CR}$	.0138	.0317	.0148
$\widehat{\mathbf{\Sigma}}_{1SA}$	.0000	.0001	.0000
$\widehat{\mathbf{\Sigma}}_{2SA}$	.0203	.0365	.0192
$\widehat{\mathbf{\Sigma}}_{3SA}$	.0052	.0099	.0051
$\widehat{\mathbf{\Sigma}}_{1SJ}$	.0195	.0436	.0206
$\widehat{\mathbf{\Sigma}}_{2SJ}$	.0001	.0003	.0001
$\widehat{\mathbf{\Sigma}}_{3SJ}$	.0040	.0085	.0041
$\widehat{\mathbf{\Sigma}}_{1SR}$	.0342	.0829	.0376
$\widehat{\mathbf{\Sigma}}_{2SR}$	.0110	.0232	.0113
$\widehat{\mathbf{\Sigma}}_{3SR}$	.0177	.0409	.0190

Table 10: Distance Between the Benchmark Bayesian Estimator $\widehat{\mathbf{\Phi}}_{1CA}$ and Alternative Bayesian Estimators of $\mathbf{\Phi}$, U.S. Quarterly Data 1959Q1-2001Q4.

	L_{Φ_1}	L_{Φ_2}
$\widehat{\mathbf{\Phi}}_{MLE}$	.0008	.0009
$\widehat{\mathbf{\Phi}}_{1CA}$	0	0
$\widehat{\mathbf{\Phi}}_{2CA}$	4.6800	3.8900
$\widehat{\mathbf{\Phi}}_{1CJ}$	.0013	.0015
$\widehat{\mathbf{\Phi}}_{2CJ}$	5.6773	4.5720
$\widehat{\mathbf{\Phi}}_{1CR}$	.0025	.0029
$\widehat{\mathbf{\Phi}}_{2CR}$	5.6664	4.5472
$\widehat{\mathbf{\Phi}}_{1SA}$	33.2468	86.0668
$\widehat{\mathbf{\Phi}}_{2SA}$	33.7148	61.0389
$\widehat{\mathbf{\Phi}}_{1SJ}$	34.8034	93.0472
$\widehat{\mathbf{\Phi}}_{2SJ}$	35.3489	66.0322
$\widehat{\mathbf{\Phi}}_{1SR}$	34.1431	89.4955
$\widehat{\mathbf{\Phi}}_{2SR}$	34.6330	63.0639