Bayesian Estimates for Vector-Autoregressive Models [1]

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 examines frequentist risks of Bayesian estimates of VAR regression coefficient and error covariance matrices under competing loss functions, under a variety of non-informative priors, and in the normal and Student-t models. Simulation results show that for the regression coefficient matrix an asymmetric LINEX estimator does better overall than the posterior mean. For the error covariance matrix no dominating estimator emerges. We find that the choice of prior has a more significant effect on the estimates than the form of estimator. For the VAR regression coefficients, a shrinkage prior dominates a constant prior. For the error covariance matrix, Yang and Berger's reference prior dominates the Jeffreys prior. Estimation of a VAR using U.S. macroeconomic data yields significantly different estimates under competing priors.

KEY WORDS: Bayesian VAR, pseudo entropy loss, quadratic loss, LINEX loss, Noninformative priors, Student-t distribution.

JEL CLASSIFICATIONS: C11, C15, C32.

[1]The research is supported by a Research Board grant from the University of Missouri-System. Sun's research is also supported by National Science Foundation grants DMS-9972598 and SES-0095919 and a grant from the Missouri Department of Conservation. We thank two referees and an associate editor for their valuable comments. We also thank Professor Arnold Zellner and the referees for suggesting the simulation study on stationarity of the estimated VARs. We are especially grateful to Professor Paul L. Speckman for his careful proof reading and helpful comments.

1 Introduction

The purpose of this paper is to explore properties of Bayesian estimates of Vector-Autoregression (VAR) models under several possible choices of sampling distributions, loss functions and priors. In the past two decades, VAR models have become popular tools for analyzing time series data. Applications of VAR to policy questions such as the macroeconomic effects of monetary supply shocks are too numerous to list (for examples see Sims 1980, 1992 and Christiano et al. 1999). The popularity of VAR models justifies a close look at the technical issues pertaining to VAR estimation and finite sample inference. A typical VAR has hundreds of parameters, which are often estimated through Least Square or Maximum Likelihood by VAR users. There are several problems with the frequentist approach to VAR estimation. First, for some distributions of data (such as Student-t), MLE does not have an analytical form or simply does not exist. Second, drawing finite sample inference for the VAR parameters is a challenge. Frequentist finite sample distributions cannot be derived in closed-form even for normal errors, while asymptotic asymptotic theory may not be applicable to a VAR with a large number of parameters and limited observations of macroeconomic data. Third, in some applications of VAR models, nonlinear functions of VAR parameters are the focus of research. For example, impulse responses are often considered easier to interpret than the regression coefficients. For another example, eigenvalues of the regression coefficient matrix are of interest since they determine the long-term dynamics of the VAR. Finite sample frequentist analysis of these nonlinear functions is more difficult than analysis of the VAR parameters themselves.

The difficulties encountered in the frequentist approach of VAR inference can be circumvented by the Bayesian approach, which combines information from data with the researcher's prior. Many applied econometricians find Bayesian methods appealing not only for philosophical reasons but more importantly for their effectiveness for finite sample inference.

Bayesian inference is influenced by the choice of prior. Elicitation of priors is an important step of Bayesian analysis. When researchers have information on the nature of parameters of interest, they may use informative priors to reflect their beliefs. For example, Doan et al. (1984) and Litterman (1986) observed that many macroeconomic time series approximately follow random walk processes and developed an informative prior known as the "Minnesota prior" that reflects the pattern. In recent Bayesian studies, Pastor (2000) and Pastor and Stambaugh (2000) used finance theory for elicitation of informative priors. Hollifield et al. (2001) developed a method for elicitation of an informative prior for VAR variance decompositions.

In many applications of VAR models, because of the large number of parameters involved, researchers often find it desirable to employ vague or noninformative priors. However, examination of the effects of using different noninformative priors on VAR posterior distributions is relatively rare. For some studies along this line, see Kadiyala and Karlsson (1997), Ni and Sun (2003), and Sun and Ni (2003). The default prior in the literature for the VAR regression coefficients is the constant prior, and the default prior for the error covariance matrix is the Jeffreys prior or a modified version of it used in the RATS software package (the RATS prior hereafter). These combinations of priors allow for easy simulation of posterior distributions and are widely employed in macroeconomic studies. The Jeffreys prior for the covariance matrix has also been widely used for multi-factor asset pricing models in the finance literature.[2] However, the Jeffreys prior is known to be undesirable in high dimensional settings (see Berger and Bernardo 1992). Bernardo (1979) and Berger and Bernardo (1992) proposed an alternative approach of deriving a reference prior. The reference prior for the covariance matrix in an independent and identically distributed (iid) normal model is derived by Yang and Berger (1994). The constant prior on the regression coefficient matrix is known to be inadmissible under quadratic loss for estimation of an unknown mean of a vector with iid normal observations (see Berger and Strawderman 1996). An alternative to the constant prior is a "shrinkage" prior (e.g, Baranchik 1964 and Berger and Strawderman, 1996). The shrinkage prior dominates the constant prior under squared error loss for estimating an unknown normal mean in iid cases, but its applications to the VAR model are scant.

For Bayesian VAR analysis, the subject of interest is the posterior of the VAR parameters. However, reporting the posterior distribution of each parameter is infeasible even for a small VAR. Some summaries of posterior quantities such as the posterior risks of Bayesian estimates would be desirable. Bayesian estimates are derived from minimization of expected posterior loss in the parameter space. However, the choice of loss function determines the form of the Bayesian estimator. In applications of Bayesian procedures, the posterior means of VAR regression coefficients and the error covariance matrix are usually reported as the Bayesian estimates. The posterior mean is optimal for certain loss functions. Bayesian estimates derived from minimizing other commonly used loss functions are rarely studied for VAR models. These loss functions include Yang and Berger's (1994) quadratic and pseudo inverse entropy losses for the error covariance matrix and Zellner's (1986) LINEX loss for the regression coefficient matrix. The fact that Bayesian estimates derived from these loss functions differ from the posterior mean may be of interest for macroeconomists. For instance, the posterior mean of the regression coefficient matrix is often biased. An asymmetric

[2]Some recent examples of include Kandel et al. (1995), Kandel and Stambaugh (1996), Lamoureux and Zhou (1996), Pastor and Stambaugh (1999), and Barberis (2000).

LINEX estimator may be helpful in correcting the bias.

The main objective of the present paper is to examine the effect of using competing priors and estimates for VAR models under various assumptions on the distribution of data. We explore Bayesian estimates under three competing loss functions on the covariance matrix (pseudo entropy loss, quadratic loss, and inverse pseudo entropy loss). We also compare the quadratic and asymmetric LINEX loss functions with respect to the VAR regression coefficient matrix. Each estimator is considered under the following prior combinations: Jeffreys, RATS, or Yang-Berger's reference prior on the covariance matrix, and a constant or shrinkage prior on the VAR coefficient matrix. The noninformative priors for the covariance matrix are derived under the normality assumption. We will evaluate their performance for the normal as well as Student-t VAR models. Our study on the Student-t model extends Geweke's (1993) analysis on univariate time series models to the vector case. We examine the competing VAR estimates in terms of frequentist average losses. We also investigate the posterior of the eigenvalues of the VAR regression coefficient matrix and the posterior of the impulse response functions.

We find that the asymmetric LINEX estimator for the VAR regression coefficient matrix does better overall than the posterior mean. The performance of competing estimates for the error covariance matrix is mixed. The choice of prior has more significant effects on the estimates than the form of the estimates. The shrinkage prior on the VAR regression coefficient matrix dominates the constant prior, while Yang and Berger's reference prior on the covariance matrix dominates the Jeffreys prior. These conclusions are drawn from a numerical example with the normal and Student-t errors. Estimation of a VAR using U.S. macroeconomic data produces significantly different estimates under the shrinkage and constant priors.

The rest of the paper is organized as follows. In Section 2, we define notation for the VAR models with the normal or Student-t errors. In Section 3, we introduce several loss functions for VAR estimation and derive corresponding Bayesian estimates. In Section 4, we present priors, especially noninformative priors. In Section 5, we derive full conditional distributions under both normal and t-errors. Several algorithms under t-errors are given in the appendix. In Section 6, simulation results are given for illustration. We also compare Bayesian estimates of a VAR using quarterly data of the U.S. economy. In Section 7, we offer concluding remarks.

2 VAR Models

A VAR of a p-dimensional column variable, $\boldsymbol{y}_t$, typically has the form

$$\boldsymbol{y}_t' = \boldsymbol{c} + \sum_{i=1}^{L} \boldsymbol{y}_{t-i}' \boldsymbol{B}_i + \boldsymbol{\epsilon}_t' \tag{1}$$

for $t = 1, \cdots, T$, where L is a known positive integer, $\boldsymbol{c}$ is a $1 \times p$ unknown vector, $\boldsymbol{B}_j$ is an unknown $p \times p$ matrix. In the normal VAR, the errors $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are iid $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite error covariance matrix. Define $\boldsymbol{x}_t' = (1, \boldsymbol{y}_{t-1}', \cdots, \boldsymbol{y}_{t-L}')$,

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_T' \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_T' \end{pmatrix}, \ \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{B}_L \end{pmatrix}, \ \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1' \\ \vdots \\ \boldsymbol{\epsilon}_T' \end{pmatrix},$$

where $\boldsymbol{x}_t'$ is a $1 \times (1 + Lp)$ row vector, and $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are $T \times p$ matrices. Here $\boldsymbol{X}$ is a $T \times (1 + Lp)$ matrix of observations, and the regression coefficient matrix $\boldsymbol{\Phi}$ is a $(1 + Lp) \times p$ matrix of unknown parameters. We can rewrite (1) as

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

For the case of normal errors, the likelihood function of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is

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

Here and in the following $etr(\mathbf{A})$ represents $\exp(trace(\mathbf{A}))$ for a matrix $\mathbf{A}$. Under the normality assumption, the MLEs are

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

where

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

It is well known that normal distributions have quite light tails. Alternative heavier tail distributions for the errors $\boldsymbol{\epsilon}_t$ include an independent multivariate-t with parameters $(\mathbf{0}, \boldsymbol{\Sigma})$ and degrees of freedom v, denoted as $t_v(\mathbf{0}, \boldsymbol{\Sigma})$. The density of $t_v(\mathbf{0}, \boldsymbol{\Sigma})$ is given by

$$p(\boldsymbol{s} \mid \boldsymbol{\Sigma}, v) = \frac{\Gamma(\frac{1}{2}(v+p))}{(\pi v)^{p/2}\Gamma(\frac{v}{2})} |\boldsymbol{\Sigma}|^{-1/2} \Big(1 + \frac{1}{v}\boldsymbol{s}'\boldsymbol{\Sigma}^{-1}\boldsymbol{s}\Big)^{-\frac{p+v}{2}}, \quad \boldsymbol{s} \in \mathbb{R}^p. \tag{6}$$

See Johnson and Kotz (1972, p134).

Regression models with Student-t errors have been studied by a number of researchers in different contexts. In Zellner's (1976) model the error terms (of different periods) in a univariate regression model form a joint multivariate-t distribution. In his model, given degrees of freedom, the MLEs of the variance and regression coefficients are the same as the MLEs in the normal model. For independent and identically t-distributed vector models, Fernandez et al. (1997) proved propositions on the robustness of both frequentist and Bayesian inferences developed under the normality assumption. Osiewalski and Steel (1993) showed that in a regression model with exogenous regressors and an improper prior the posteriors of the parameters of interest are the same under the normal or t-errors. In Geweke's (1993) univariate time series regression model, the error terms are independent t-distributed and have unknown degrees of freedom. In the present VAR model, the vector of error terms in each period is assumed to follow a multivariate-t distribution with unknown degrees of freedom. The case when the degree of freedom $v = 1$ corresponds to a multivariate Cauchy distribution (cf. Johnson and Kotz, 1972, p134). When $v \to \infty$, $t_v(\mathbf{0}, \mathbf{\Sigma})$ goes to $N_p(\mathbf{0}, \mathbf{\Sigma})$.

The property that the MLE based on the t-distributed errors is the same as the MLE under normal errors no longer holds for the VAR model. For given degrees of freedom parameter v, the MLE of $(\mathbf{\Phi}, \mathbf{\Sigma})$ under the t-distributed errors is not available in closed-form. If v is unknown, MLE for $(\mathbf{\Phi}, \mathbf{\Sigma}, v)$ may not even exist. Fernandez and Steel (1999) discussed the difficulties of conducting likelihood-based inference for t-distributed models when identical data are recorded from independent sampling. However, we do not encounter such a problem in our VAR analysis on a given set of time series observations. To simulate Bayesian estimates, in this paper we will develop MCMC methods based on a hierarchical structure of the multivariate t-distribution.

3 Loss Functions and Bayesian Estimates

A Bayesian estimator of $(\mathbf{\Phi}, \mathbf{\Sigma})$ depends on the distribution model, the prior, and the loss function. We consider loss functions that contain a part measuring the loss associated with the error covariance matrix and a part measuring the loss of the VAR regression coefficients only. The overall loss with respect to estimator $(\hat{\mathbf{\Phi}}, \hat{\mathbf{\Sigma}})$ for $(\mathbf{\Phi}, \mathbf{\Sigma})$ has the form

$$L(\hat{\mathbf{\Sigma}}, \hat{\mathbf{\Phi}}; \mathbf{\Sigma}, \mathbf{\Phi}) \quad = \quad L_\Sigma(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}) + L_\Phi(\hat{\mathbf{\Phi}}; \mathbf{\Phi}). \tag{7}$$

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

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

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

$$L_{\Sigma 1}(\widehat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \log|\hat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}| - p, \tag{8}$$

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 true entropy loss pertains to both $\mathbf{\Sigma}$ and $\mathbf{\Phi}$.

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

$$L_{\Sigma 2}(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1} - \boldsymbol{I})^2. \tag{9}$$

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

$$L_{\Sigma 3}(\hat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad tr(\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}) - \log|\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1}| - p. \tag{10}$$

Note that the loss functions with respect to $\mathbf{\Sigma}$ are separated from loss functions with respect to $\mathbf{\Phi}$. As a result, Bayesian estimates can be derived separately from minimizing expected posterior loss functions regarding $(\mathbf{\Phi}, \mathbf{\Sigma})$ if the minimum is finite. A generalized Bayesian estimator can be defined analogously regardless of whether the Bayes risk is finite.

Lemma 1 *(a) Under the loss $L_{\Sigma 1}$, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_1 = E(\mathbf{\Sigma} \mid \boldsymbol{Y})$.*
(b) Under the loss $L_{\Sigma 2}$, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_2$, given by $vec(\widehat{\mathbf{\Sigma}}_2) = \Big[E\{(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}) \mid \boldsymbol{Y}\}\Big]^{-1} vec\{E(\mathbf{\Sigma}^{-1} \mid \boldsymbol{Y})\}$, where $\otimes$ represents the Kronecker product.
(c) Under the loss $L_{\Sigma 3}$, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_3 = \Big\{E(\mathbf{\Sigma}^{-1} \mid \boldsymbol{Y})\Big\}^{-1}$.

Proof. For part (a), note that

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

The above facts can be used to calculate the derivative of $\mathrm{E}\{L_{\Sigma 1}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \mid \boldsymbol{Y}\}$ with respect to $\widehat{\mathbf{\Sigma}}$. The desired result follows by setting the derivative to zero. The proof of part (b) is more tedious; see Yang and Berger (1994). Part (c) follows from the fact that the derivative of $\mathrm{E}\{L_{\Sigma 3}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) \mid \boldsymbol{Y}\}$ with respect to $\widehat{\mathbf{\Sigma}}$ is $\mathrm{E}(\mathbf{\Sigma}^{-1} \mid \boldsymbol{Y}) - \widehat{\mathbf{\Sigma}}^{-1}$. □

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

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

$$L_{\Phi 1}(\widehat{\mathbf{\Phi}}, \mathbf{\Phi}) \quad = \quad trace\{(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})' \mathbf{W} (\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\}, \tag{11}$$

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

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

$$L_{\Phi 2}(\hat{\mathbf{\Phi}}, \mathbf{\Phi}) = \sum_{i=1}^{1+Lp} \sum_{j=1}^{p} \Big[\exp\{a_{ij}(\widehat{\phi}_{ij} - \phi_{ij})\} - a_{ij}(\widehat{\phi}_{ij} - \phi_{ij}) - 1 \Big], \tag{12}$$

where a_{ij} is a given constant. When a_{ij} is close to zero, the LINEX loss function is nearly symmetric and not much different from quadratic loss. When a_{ij} is a large negative number, the LINEX loss is close to be exponential when $\widehat{\phi}_{ij} < \phi_{ij}$ 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_{ij} parameter should help correct the bias. The Bayesian estimates of $\mathbf{\Phi}$ under these loss functions are well known and are given here for convenience.

Lemma 2 *(a) Under the loss $L_{\Phi 1}$, the generalized Bayesian estimator of $\mathbf{\Phi}$ is*

$$\widehat{\mathbf{\Phi}}_1 = E(\mathbf{\Phi} \mid \boldsymbol{Y}). \tag{13}$$

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

$$\widehat{\boldsymbol{\phi}}_{ij} \quad = \quad -\frac{1}{a_{ij}} \log[E\{\exp(-a_{ij}\boldsymbol{\phi}_{ij} \mid \boldsymbol{Y}\}] \tag{14}$$

for $i = 1, \ldots, Lp+1$ and $j = 1, \ldots, p$.

Proving the lemma is straightforward. The lemma shows that the Bayesian estimate of $\mathbf{\Phi}$ under the quadratic loss function is the posterior mean while the estimate under the LINEX loss may be larger or smaller than the posterior mean, depending on the sign of the constant a_{ij}.

3.3 Impulse Response Functions

A covariance stationary VAR has the moving average representation $\boldsymbol{y}_t' = \mathrm{E}_0 \boldsymbol{y}_t' + \sum_{j=0}^{t-1} \boldsymbol{\epsilon}_{t-j}' \boldsymbol{H}_j$, where $\boldsymbol{H}_0$ is the p by p identity matrix, and the impulse responses of $\boldsymbol{y}_t$ to a shock $\boldsymbol{\epsilon}_{t-j}$ occurring

j periods earlier is $\boldsymbol{H}_j = \sum_{i=1}^{j} \boldsymbol{B}_i \boldsymbol{H}_{j-i}$, with $\boldsymbol{B}_i$=0 for i larger than L. Note that the components of the vector of errors $\boldsymbol{\epsilon}_t$ are correlated since the covariance matrix is unrestricted. For example, the forecasting error of short-term interest rates may be correlated with that of inflation. Suppose a short-term interest rate is the monetary policy indicator. A monetary policy shock should be represented by a shock in the short-term interest rate uncorrelated with other shocks. Thus more economic meaningful impulse responses are ones to orthogonalized (structural) errors. Orthogonalization of the errors can be achieved through the Cholesky decomposition of the covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{\Psi}'\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is an upper triangular positive definite matrix. VAR errors are mapped to structural shocks through $\boldsymbol{u}_t' = \boldsymbol{\epsilon}_t'\boldsymbol{\Psi}^{-1}$. The covariance matrix of the structural error vector $\boldsymbol{u}_t$ is the identity matrix. The impulse response to structural shocks occurring j periods earlier is given by $\boldsymbol{Z}_j = \boldsymbol{\Psi}\boldsymbol{H}_j$. By definition, impulse responses are nonlinear functions of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. The nonlinearity makes it difficult to conduct frequentist inference but does not pose difficulty for Bayesian simulations as long as $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ can be simulated.

The loss function for estimation of the impulse response $\boldsymbol{Z}_j$ (with forecasting horizon j) is assumed to be $L(\boldsymbol{Z}, \widehat{\boldsymbol{Z}}_j) = trace\{(\boldsymbol{Z}_j - \widehat{\boldsymbol{Z}}_j)'\boldsymbol{\Omega}(\boldsymbol{Z}_j - \widehat{\boldsymbol{Z}}_j)\}$. The matrix $\boldsymbol{\Omega}$ weighs the estimation error of each element of the impulse responses. The weighting may be dictated by the economic significance of the element. So long as $\boldsymbol{\Omega}$ is constant, the Bayesian estimate $\widehat{\boldsymbol{Z}}_j$ is the posterior mean $\mathrm{E}(\boldsymbol{Z}_j \mid Y)$. In the numerical example in this paper, we simply set $\boldsymbol{\Omega}$ as the identity matrix.

4 Priors

Bayesian analysis requires explicit specification of a prior on the parameters. Noninformative priors are commonly employed by VAR users since it is difficult to find a universally justifiable subjective prior. Note that noninformative priors for $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ are not unique because they can be derived according to different principles (see Kass and Wasserman 1996).

The most popular noninformative prior for $\boldsymbol{\Sigma}$ is the Jeffreys prior (See Geisser 1965, Tiao and Zellner 1964). The Jeffreys prior is derived from the "invariance principle," according to which the prior is invariant to re-parameterization (see Jeffreys 1961 and Zellner 1971). Specifically, for the VAR covariance matrix, the Jeffreys prior is $\pi_J(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}$. The prior for $\boldsymbol{\Sigma}$ in RATS is a modified version of the Jeffreys prior, $\pi_A(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(L+1)p/2-1}$.

It has been noted, however, that the Jeffreys prior often gives unsatisfactory results for multiparameter problems. Bernardo (1979) proposed an information-based approach of deriving a reference prior by breaking a single multiparameter problem into consecutive problems with fewer numbers of parameters. 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 where the reference priors produce more desirable estimates than the Jeffreys priors, see Berger and Bernardo (1992) and 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-parameterized matrix the $\mathbf{\Sigma}$ as $\mathbf{O}'\mathbf{D}\mathbf{O}$, where $\mathbf{D}$ is a diagonal matrix whose elements 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 giving priority to vectorized $\mathbf{D}$ over vectorized $\mathbf{O}$: $\pi_R(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-1} \prod_{1 \le i < j \le p} (d_i - d_j)^{-1}$, here $d_1 > d_2 > \cdots > d_p$ 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 the vectorized $\mathbf{\Phi}$, $vec(\mathbf{\Phi})$, which we denote as $\boldsymbol{\phi}$. A common expression of ignorance about $\boldsymbol{\phi}$ is a (flat) constant prior. A popular noninformative prior for multivariate regression models is called the diffuse prior, which consists of a constant prior for $\boldsymbol{\phi}$ and the Jeffreys prior for $\mathbf{\Sigma}$. The joint density of the constant-Jeffreys prior for $(\mathbf{\Phi}, \mathbf{\Sigma})$ (or $(\boldsymbol{\phi}, \mathbf{\Sigma})$) is in the form $\pi_{CJ}(\boldsymbol{\phi}, \mathbf{\Sigma}) \propto \pi_J(\mathbf{\Sigma})$. 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 for the constant-RATS instead of the constant-Jeffreys prior, see Sims and Zha (1999). Finally, 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 an unknown multivariate normal mean, motivated by Stein's (1956) result on inadmissibility of the MLE, some authors (e.g., Baranchik 1964, Berger and Strawderman 1996) have advocated the following "shrinkage" prior as an alternative to the constant prior for $\boldsymbol{\phi}$:

$$\pi_S(\boldsymbol{\phi}) \propto \|\boldsymbol{\phi}\|^{-(J-2)}, \quad \boldsymbol{\phi} \in \mathbb{R}^J, \tag{15}$$

where the dimension of $\boldsymbol{\phi}$, J, equals $p(Lp+1)$. Note that the prior (15) has a two-stage hierarchical structure. Let $\pi_S(\boldsymbol{\phi} \mid \delta)$ be the normal density $N_J(\mathbf{0}, \delta \boldsymbol{I}_J)$. We can write

$$(\boldsymbol{\phi} \mid \delta) \;\sim\; N_J(\mathbf{0}, \delta \boldsymbol{I}_J), \;\text{ and }\; \pi(\delta) \;\propto\; 1, \tag{16}$$

since

$$\int_0^\infty \pi_S(\boldsymbol{\phi} \mid \delta)\pi(\delta) d\delta \;\propto\; \int_0^\infty \frac{1}{(2\pi\delta)^{J/2}} \exp\{-\frac{1}{2\delta}\boldsymbol{\phi}'\boldsymbol{\phi}\} d\delta \;=\; \frac{\Gamma(J/2)}{(2\pi)^{J/2}(\boldsymbol{\phi}'\boldsymbol{\phi})^{J/2-1}},$$

which is proportional to (15).

We consider three noninformative priors for $(\mathbf{\Phi}, \mathbf{\Sigma})$ with the shrinkage prior on $\mathbf{\Phi}$, the shrinkage-Jeffreys prior $\pi_{SJ}(\boldsymbol{\phi}, \mathbf{\Sigma}) = \pi_S(\boldsymbol{\phi})\pi_J(\mathbf{\Sigma})$, the shrinkage-RATS prior $\pi_{SA}(\boldsymbol{\phi}, \mathbf{\Sigma}) = \pi_S(\boldsymbol{\phi})\pi_A(\mathbf{\Sigma})$, and the shrinkage-reference prior $\pi_{SR}(\boldsymbol{\phi}, \mathbf{\Sigma}) = \pi_S(\boldsymbol{\phi})\pi_R(\mathbf{\Sigma})$.

The noninformative priors for $(\mathbf{\Phi}, \mathbf{\Sigma})$ are improper (i.e., the integrals of which in the parameter space are infinite). Ni and Sun (2003) and Sun and Ni (2003) provided conditions under which the posteriors of $(\mathbf{\Phi}, \mathbf{\Sigma})$ are proper with the prior combinations considered in this paper.

We now turn to the prior for the degrees of freedom parameter v in the Student-t errors. The parameter v in the Student-t likelihood function is commonly regarded as a positive integer but can be any positive number. We will apply a gamma (a, b) prior for $w = v/2$, the density of which is given by

$$[w] \quad = \quad \frac{b^a}{\Gamma(a)} w^{a-1} \exp(-bw), \quad w > 0, \tag{17}$$

where (a, b) are positive known constants. When $a = 1$, it is the exponential prior given in Geweke (1993).

5 Bayesian Computation and Conditional Posteriors

The posterior quantities of $(\boldsymbol{\phi}, \mathbf{\Sigma})$ are not available in closed-form for most of the priors we considered. In this study, we use Gibbs sampling MCMC methods to sample from the posteriors (see Gelfand and Smith, 1990). The first step of the MCMC computation is to find the full conditional distributions of $(\boldsymbol{\phi}, \mathbf{\Sigma})$.

5.1 Conditional Posteriors under the Normal Errors

We will make use of the following results for the normal model.

Fact 1 *Consider the constant-Jeffreys prior for* $(\boldsymbol{\phi}, \mathbf{\Sigma})$.
(a) The conditional posterior of $\boldsymbol{\phi}$ *given* $(\mathbf{\Sigma}, \boldsymbol{Y})$ *is* $N_J(\widehat{\boldsymbol{\phi}}_{MLE}, \mathbf{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$
(b) The marginal posterior of $\mathbf{\Sigma}$ *given* $\boldsymbol{Y}$ *is inverse Wishart* $(\boldsymbol{S}(\widehat{\mathbf{\Phi}}_{MLE}), T - Lp - 1)$.

Fact 2 *Consider the constant-RATS prior* π_{CA}.
(a)The conditional posterior of $\boldsymbol{\phi}$ *given* $(\mathbf{\Sigma}; \boldsymbol{Y})$ *is* $N_J(\widehat{\boldsymbol{\phi}}_{MLE}, \mathbf{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$.
(b) The marginal posterior of $\mathbf{\Sigma}$ *given* $\boldsymbol{Y}$ *is Inverse Wishart* $(\boldsymbol{S}(\widehat{\mathbf{\Phi}}_{MLE}), T)$.

Fact 3 *Consider the constant-reference prior* π_{CR}.

(a) The conditional distribution of $\boldsymbol{\phi}$ *given* $(\boldsymbol{\Sigma}; \boldsymbol{Y})$ *is* $N_J(\widehat{\boldsymbol{\phi}}_{MLE}, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$.

(b) The conditional density of $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\phi}; \boldsymbol{Y})$ *is*

$$\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\phi}; \boldsymbol{Y}) \propto \frac{etr\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{\Phi})\}}{|\boldsymbol{\Sigma}|^{\frac{T}{2}+1} \prod_{1 \le i < j \le p} (\lambda_i - \lambda_j)}, \tag{18}$$

where $\boldsymbol{S}(\boldsymbol{\Phi})$ *is defined by (5).*

For the shrinkage prior, the hierarchical structure (16) suggests a nice computational formula. Instead of simulating from the conditional distribution of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ within each Gibbs cycle, we use δ as a latent variable and simulate from $\boldsymbol{\Phi}$, $\boldsymbol{\Sigma}$, and δ based on the following fact.

Fact 4 *Consider the shrinkage-reference prior* π_{SR}. *The full conditional posteriors of* $(\boldsymbol{\Sigma}, \boldsymbol{\phi}, \delta)$ *given data* $\boldsymbol{Y}$ *are as follows.*

(a) The conditional density of $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\phi}, \delta; \boldsymbol{Y})$ *is in (18).*

(b) The conditional distribution of $\boldsymbol{\phi}$ *given* $(\delta, \boldsymbol{\Sigma}; \boldsymbol{Y})$ *is* $N_J(\boldsymbol{\mu}_S, \boldsymbol{V}_S)$, *where*

$$\boldsymbol{\mu}_S = \delta\Big(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1} + \delta \boldsymbol{I}_J\Big)^{-1} \widehat{\boldsymbol{\phi}}_{MLE}; \tag{19}$$

$$\boldsymbol{V}_S = \Big(\boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) + \frac{1}{\delta}\boldsymbol{I}_J\Big)^{-1}. \tag{20}$$

(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 the posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. Ni and Sun (2003) offered 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.

5.2 Conditional Posteriors under Student-t Errors

The conditional posterior distributions of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ under the t-errors are quite complicated. We can, however, use the hierarchical structure of the t-distribution to introduce latent variables. In particular, let $(\boldsymbol{s} \mid q) \sim N_p(0, q^{-1}\boldsymbol{\Sigma})$, where $q \sim gamma(v/2, v/2)$ has the density given by

$$p(q \mid v) = \frac{(v/2)^{v/2} q^{\frac{v}{2}-1}}{\Gamma(v/2)} \exp(-\frac{v\,q}{2}), \quad \text{for} \quad q > 0. \tag{21}$$

It follows that the marginal density of $\boldsymbol{s}$ is

$$p(\boldsymbol{s} \mid \boldsymbol{\Sigma}, v) = \int_0^\infty p(\boldsymbol{s} \mid q, \boldsymbol{\Sigma}, v) p(q \mid v) dq \propto |\boldsymbol{\Sigma}|^{-1/2} (v + \boldsymbol{s}'\boldsymbol{\Sigma}^{-1}\boldsymbol{s})^{-\frac{v+p}{2}},$$

which is the same as the density of $t_v(\mathbf{0}, \boldsymbol{\Sigma})$ given by (6).

Now we introduce independent latent variables $q_1, \cdots, q_T$ that are iid $\sim gamma(v/2, v/2)$. For given $q_1, \cdots, q_T$, and $\boldsymbol{\Sigma}$, VAR error vectors $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are independent, with $\boldsymbol{\epsilon}_t \sim N_p(0, q_t^{-1}\boldsymbol{\Sigma})$. Let $\boldsymbol{Q} = diag(q_1, \cdots, q_T)$. The joint likelihood of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}, v)$ is then

$$\begin{aligned} l^*(\boldsymbol{Q}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}, v) &\propto \frac{(v/2)^{\frac{vT}{2}}}{[\Gamma(v/2)]^T} \frac{\prod_{t=1}^T q_t^{\frac{p+v}{2}-1} \exp(-\frac{v\,q_t}{2})}{|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \exp\Big\{-\sum_{t=1}^T \frac{q_t(\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})}{2}\Big\} \\ &= \frac{(v/2)^{\frac{vT}{2}}}{[\Gamma(v/2)]^T} \frac{\prod_{t=1}^T q_t^{\frac{p+v}{2}-1} \exp(-\frac{v\,q_t}{2})}{|\boldsymbol{\Sigma}|^{\frac{T}{2}}} etr\Big\{-\frac{1}{2}\boldsymbol{S}_Q(\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}\Big\}, \end{aligned} \quad (22)$$

where

$$\boldsymbol{S}_Q(\boldsymbol{\Phi}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})'\boldsymbol{Q}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}). \quad (23)$$

The following facts are useful for Bayesian simulation in the Student-t model.

Fact 5 *Consider the constant-Jeffreys prior for* $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ *and gamma* (a, b) *prior for* $w = v/2$.
(a) The conditional posterior of $\boldsymbol{\phi}$ *given* $(\boldsymbol{\Sigma}, \boldsymbol{Q}, v; \boldsymbol{Y})$ *is* $N(\widehat{\boldsymbol{\phi}}_Q, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})^{-1})$. *Here* $\widehat{\boldsymbol{\phi}}_Q$ *is defined as vectorized* $\widehat{\boldsymbol{\Phi}}_Q$, *and*

$$\widehat{\boldsymbol{\Phi}}_Q = (\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{Y}. \quad (24)$$

(b) The conditional posterior of $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\phi}, \boldsymbol{Q}, v; \boldsymbol{Y})$ *is Inverse Wishart* $(\boldsymbol{S}_Q(\boldsymbol{\Phi}), T)$.
(c) Given $(\boldsymbol{\phi}, \boldsymbol{\Sigma}, v; \boldsymbol{Y})$, $q_1, \cdots, q_T$ *are independent and*

$$(q_t \mid \boldsymbol{\phi}, \boldsymbol{\Sigma}, v; \boldsymbol{Y}) \sim gamma\Big(\tfrac{1}{2}(v+p),\ \tfrac{1}{2}\{v + (\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})\}\Big). \quad (25)$$

(d) Given $(\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}; \boldsymbol{Y})$, *the conditional density of* $w = v/2$ *has the form*

$$[w \mid \boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}; \boldsymbol{Y}] = \frac{w^{Tw+a-1}}{[\Gamma(w)]^T} \left\{\prod_{t=1}^T q_t\right\}^w \exp\Big\{-\Big(b + \sum_{t=1}^T q_t\Big)w\Big\}, \ \textit{for } w > 0. \quad (26)$$

(e) If $T/2 + a - 1 > 0$, *the conditional posterior of* w *in part (d) is log-concave.*

Proof. The proof for parts (a)-(d) is easy. For part (e), denote $g(w) = \log[w \mid \boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}; \boldsymbol{Y}]$. It is enough to show that $\frac{\partial^2}{\partial w}g(w) < 0$. Clearly,

$$g(w) = (Tw + a - 1)\log(w) - T\log\{\Gamma(w)\} + w\log\Big(\prod_{t=1}^T q_t\Big) - \Big(b + \sum_{t=1}^T q_t\Big)w.$$

It follows from Formula (1.46) of Bowman and Shenton (1988) that

$$\frac{d^2}{dw^2}\log[\Gamma(w)] = \frac{1}{w} + \frac{1}{2w^2} + \frac{2\pi}{w}\int_0^\infty \frac{y\sqrt{s}}{(w^2+s)(y-1)^2}ds,$$

where $y = \exp(2\pi\sqrt{s})$. Consequently,

$$\frac{d^2}{dw^2} g(w) \quad = \quad -\frac{\frac{T}{2}+a-1}{w^2} - \frac{2\pi T}{w}\int_0^\infty \frac{y\sqrt{s}}{(w^2+s)(y-1)^2} ds,$$

which is negative if $a + T/2 - 1 > 0$. □

For MCMC simulations, one can sample successively from the full conditional distribution, given by Fact 5(a)-(d), treating $(\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}, v)$ in four blocks in updating. Alternatively, we can divide $(\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}, v)$ into three blocks, $\boldsymbol{Q}$, $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ and v. We first sample $(\boldsymbol{Q} \mid \boldsymbol{\phi}, \boldsymbol{\Sigma}, v; \boldsymbol{Y})$ according to Fact 5(c); within the block $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$, we sample $(\boldsymbol{\phi} \mid \boldsymbol{\Sigma}, \boldsymbol{Q}, v; \boldsymbol{Y})$ according to Fact 5(a), then instead of using property Fact 5(b), we sample $(\boldsymbol{\Sigma} \mid \boldsymbol{Q}; \boldsymbol{Y})$ according to $(\boldsymbol{\Sigma} \mid \boldsymbol{Q}, v; \boldsymbol{Y}) \sim$ Inverse Wishart $(\boldsymbol{S}_Q(\widehat{\boldsymbol{\Phi}}_Q), T-Lp-1)$. Chib et al. (1991) offered a theoretical discussion on simulation of the degrees of freedom parameter v. Gilks and Wild (1992) proposed an adaptive rejection sampling scheme for simulating from a log-concave density. We will adopt Gilks and Wild's method for simulating the conditional posterior of w based on Fact 5(e).

Fact 6 *Consider π_{CA}, the constant-RATS prior for $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$, and gamma (a, b) for $w = v/2$. The conditional posteriors of $\boldsymbol{\phi}$, $\boldsymbol{Q}$, and v are the same as parts (a), (c) and (d) in Fact 5, respectively, while the conditional posterior of $\boldsymbol{\Sigma}$ given $(\boldsymbol{Q}, v; \boldsymbol{Y})$ is Inverse Wishart $(\boldsymbol{S}_Q(\widehat{\boldsymbol{\Phi}}_Q), T)$.*

Fact 7 *Consider π_{CR}, the constant-reference prior for $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$, and gamma (a, b) prior for $w = v/2$. The conditional posteriors of $\boldsymbol{\phi}$, $\boldsymbol{Q}$, and v are the same as parts (a), (c) and (d) in Fact 5, respectively, while the conditional posterior density of $\boldsymbol{\Sigma}$ given $(\boldsymbol{Q}, v; \boldsymbol{Y})$ is*

$$\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\phi}, \boldsymbol{Q}, v; \boldsymbol{Y}) \propto \frac{etr\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_Q(\boldsymbol{\phi})\}}{|\boldsymbol{\Sigma}|^{\frac{T}{2}+1}\prod_{1\le i<j\le p}(\lambda_i - \lambda_j)}. \tag{27}$$

Fact 8 *Consider π_{SR}, the shrinkage-reference prior and gamma (a, b) prior for $w = v/2$.*

(a) The conditional distribution of $\boldsymbol{\phi}$ given $(\delta, \boldsymbol{\Sigma}, \boldsymbol{Q}, v; \boldsymbol{Y})$ is $N_J(\boldsymbol{\mu}_Q, \boldsymbol{V}_Q)$, where

$$\boldsymbol{\mu}_Q \quad = \quad \delta\Big(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})^{-1} \ + \ \delta \boldsymbol{I}_J\Big)^{-1}\widehat{\boldsymbol{\phi}}_Q; \tag{28}$$

$$\boldsymbol{V}_Q \quad = \quad \Big(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X} \ + \ \frac{1}{\delta}\boldsymbol{I}_J\Big)^{-1}. \tag{29}$$

(b) The conditional distribution of δ given $(\boldsymbol{\phi}, \boldsymbol{\Sigma}, \boldsymbol{Q}, v; \boldsymbol{Y})$ is Inverse Gamma $(\frac{J}{2} - 1, \frac{1}{2}\boldsymbol{\phi}'\boldsymbol{\phi})$.

(c) The conditional density of $\boldsymbol{\Sigma}$ given $(\boldsymbol{\phi}, \delta, \boldsymbol{Q}; \boldsymbol{Y})$ is given by (27).

(d) For $t = 1, \cdots, T$, $(q_t \mid \boldsymbol{\phi}, \delta, \boldsymbol{\Sigma}, v; \boldsymbol{Y})$ has the distribution (25).

(e) Given $(\boldsymbol{\phi}, \delta, \boldsymbol{\Sigma}, \boldsymbol{Q}; \boldsymbol{Y})$, the conditional posterior of $w = v/2$ is the same as part (d) in Fact 5.

6 Numerical Examples

6.1 Simulations Results

In the following we use a numerical example to evaluate the properties of competing Bayesian estimates. We first generate data samples from a VAR with known parameters and then compute the Bayesian estimates via MCMC simulations. Using the Monte Carlo results, we evaluate the performance of the Bayesian estimates under competing priors in terms of the frequentist risks given the true parameters. The frequentist risks under a loss L, i.e. $E_{\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma}} L(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}})$ and $E_{\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma}} L(\boldsymbol{\Phi}, \widehat{\boldsymbol{\Phi}})$, of the MLE and Bayesian estimates using alternative priors on $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are estimated as the average loss pertaining to estimates $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Phi}}$ across generated data samples. For instance, $E_{\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma}} L(\boldsymbol{\Phi}, \widehat{\boldsymbol{\Phi}})$ is estimated as $\frac{1}{N}\sum_{n=1}^{N} L(\boldsymbol{\Phi}, \widehat{\boldsymbol{\Phi}}_n)$. Here $\widehat{\boldsymbol{\Phi}}_n$ is the Bayesian estimate from sample n ($n = 1, ..., N$), and $\boldsymbol{\Phi}$ is the true parameter matrix $\boldsymbol{\Phi}$.

We denote the estimates according to a loss function and a prior. For example, $\widehat{\boldsymbol{\Sigma}}_{1CA}$ represents the estimator of $\boldsymbol{\Sigma}$ under loss $L_{\Sigma 1}$ and the constant-RATS prior combination, and $\widehat{\boldsymbol{\Phi}}_{2SR}$ represents the estimator of $\boldsymbol{\Phi}$ under loss $L_{\Phi 2}$ and the shrinkage-reference prior combination. Each row of Tables 1 and 2 reports the frequentist average and standard deviations of losses of the corresponding estimator under different loss functions.

As mentioned in the introduction, besides frequentist risks of estimates of parameters of the VAR, we are interested in estimates of certain nonlinear functions of these parameters such as impulse responses. Losses pertaining to impulse responses are computed as follows. For the nth data set generated in the experiment, we denote the impulse response matrix for the ith step after the shock as $\widehat{\boldsymbol{Z}}_i^{(n)}$. The accuracy in estimation of the impulse responses (with forecasting horizon i) is measured by the frequentist average of sum of squared errors

$$R_{Imp,i} = \frac{1}{N}\sum_{n=1}^{N} trace\{(\boldsymbol{Z}_i - \widehat{\boldsymbol{Z}}_i^{(n)})'(\boldsymbol{Z}_i - \widehat{\boldsymbol{Z}}_i^{(n)})\}. \tag{30}$$

In addition to evaluating the overall performance of VAR estimates, we also investigate stationarity of the estimated VARs. VARs with one lag are stationary if the absolute values of the real eigenvalues (and amplitudes of the complex eigenvalues) of the regression coefficient matrix $\boldsymbol{B}_1$ are less than unity. VARs of more than one lag can be rewritten as a VAR with one lag through transformation of variables. For simplicity, simulations in this study were based on a VAR with one lag. For each prior, we compared the eigenvalues of the posterior mean of the $\boldsymbol{B}_1$ matrix with those used for generating the data.

We generated $N = 1000$ samples from a five-variable VAR model with one lag (i.e., $p = 5$ and $L = 1$) with true parameters given in the following. The elements of $\mathbf{\Sigma}$ are $\sigma_{i,j} = 1.0$ if $i = j$ and 0.5 otherwise, which implies that the VAR errors have a pairwise correlation of 0.5. The elements of $\mathbf{\Phi}$ are $\boldsymbol{c}$=0, $\boldsymbol{B}_1 = \boldsymbol{I}_5$. By assumption, the VAR regression coefficients contains unit roots. The number of sample observations, T, is 50. In computation of the Bayesian estimates of $(\mathbf{\Phi}, \mathbf{\Sigma})$, we ran $M = 10000$ MCMC cycles with 500 burn-in cycles for each of the 1000 samples. We choose the weighting matrix $\mathbf{W}$ in the loss function $L_{\Phi 1}$ to be the identity matrix. The parameter a in loss $L_{\Sigma 2}$ is set as follows. We let the parameter corresponding to the intercept term ϕ_{1j} be 0.001, implying a near symmetric loss, and that corresponding to non-intercept term be -4.0. The value -4 was used by Zellner (1986) for estimating the normal means. We first consider the normal model and then a Student-t model.

Simulations for the six prior combinations took less than 2 minutes for each of the 1000 samples on a Pentium 4 1.7 GHz PC. The simulations converged quite quickly. There is little difference when the Markov Chain is shortened to 5000 from 10000. The acceptance rates for the Metropolis step for simulation of $\mathbf{\Sigma}$ under the reference prior are about thirty percent.

Several conclusions can be drawn from the numerical example.

(a) The estimates of $\mathbf{\Sigma}$ are largely unrelated to the priors on $\mathbf{\Phi}$, and estimates of $\mathbf{\Phi}$ are not affected much by the priors on $\mathbf{\Sigma}$. Tables 1 and 2 report the frequentist risks of the estimates of $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ under competing priors for the VAR with normal errors. In Table 1, the $\mathbf{\Sigma}$-related losses of $\widehat{\mathbf{\Sigma}}_{1CA}$ (based on the constant-RATS prior) are quite similar to those of $\widehat{\mathbf{\Sigma}}_{1SA}$ (based on the shrinkage-RATS prior), and in Table 2 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). This is due to the assumption that the loss functions are separable in $\mathbf{\Sigma}$ and $\mathbf{\Phi}$.

(b) 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 quadratic loss, but it is moderately better in terms of LINEX loss. Hence overall $\widehat{\mathbf{\Phi}}_2$ is better than the posterior mean. The Bayesian estimates of $\mathbf{\Phi}$ under the shrinkage prior yield smaller frequentist average losses than their counterparts under the constant prior.

(c) 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 for all estimates. Table 2 shows that the Bayesian estimates of $\mathbf{\Phi}$ based on the shrinkage prior yield smaller frequentist risks than their counterparts based on a constant prior. Even with strong asymmetry ($a = -4.0$) in the loss function $L_{\Phi 2}$, there is not much difference between the two estimates of $\mathbf{\Phi}$ under the shrinkage prior. In contrast, prior choice is critically important. Under the constant prior, both Bayesian estimates perform poorly, even in comparison with the MLE. The large standard deviations of the frequentist losses under the constant prior indicate that these Bayesian estimates are not successful in dealing with outliers. In comparison, under the shrinkage prior both estimates dominate the MLE in terms of the frequentist risk associated with $\mathbf{\Phi}$. The Bayesian estimator under the shrinkage prior improves over the MLE by reducing bias for some elements and at the same time substantially reducing variances. Table 3 reports the frequentist risks of impulse responses for the normal model. Under the shrinkage prior, the frequentist risks of impulse responses are smaller than those under the constant prior.

(d) A question of interest for applied econometricians is whether the conclusions on comparison of priors are robust when the VAR errors are not normal. We find that they are. We simulate 1000 samples using the Student-t model with degrees of freedom $v = 8$ and the parameter values for $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ unchanged. We find similar performance of the estimates and priors as in the normal model. To save space, the counterparts of Tables 1 and 2 are not reported. We use gamma(1,0.5) as the prior for the nuisance parameter v. Experiments with alternative priors on v show that our general conclusion is robust. Tables 4 reports frequentist losses of impulse responses, with the normal VAR errors replaced by Student-t distributed errors and the degrees of freedom is treated as an unknown parameter. The posteriors are simulated under the algorithm for t-errors given in the appendix. Consistent with our finding on the estimates of VAR parameters under competing priors and data generating models, Table 4 is not qualitatively different from Table 3.

(e) Economists are often concerned with inference on stationarity of time series models. Much research has been done on the properties of MLE of autoregressive (AR) models. It is known that the MLE of the AR(1) coefficient with unit root has a downward bias. MacKinnon and Smith (1998) showed that the downward bias of ML estimates for an AR(1) coefficient is nonlinear in the true AR parameter. When the true parameter is near unity, the downward bias is particularly severe. There is a rich Bayesian literature on the unit root model. Phillips (1991) and Sims (1991) debated the merit of using a constant prior for the AR coefficient. Stambaugh (1999) conducted Bayesian analysis of a model with two equations, one a regression of asset return on lagged dividend yield and the other an AR(1) equation of the dividend yield. He showed that the posterior mean of the slope parameter in the asset return equation is smaller than the bias-corrected OLS, while that of the dividend yield equation is larger. Zellner and Hong (1989) estimated an AR(3) model for GDP growth and found that eighty percent of the posterior eigenvalues have a complex pair.

We now examine the posterior mean of eigenvalues of the VAR regression coefficient matrix $\boldsymbol{B}_1$. The true coefficient matrix of the VAR in the numerical example has five eigenvalues of unity. We found that all eigenvalues of the MLE and posterior mean of Bayesian estimates of the matrix were smaller in magnitude than the true value. About forty percent of the eigenvalues in the 1000 samples turned out to be complex numbers. The imaginary parts of these complex numbers were much smaller than the real parts. In the following, we focus on the eigenvalues with the largest and smallest amplitude. For the normal model, the average of the largest eigenvalues of the MLE over the 1000 samples is 0.960 and the standard deviation is 0.036. The average of the smallest eigenvalues is 0.510, with a standard deviation of 0.145. There is clearly a downward bias even for the largest eigenvalues. For the 1000 samples, the statistics of the eigenvalues for posterior means of $\boldsymbol{B}_1$ under constant prior were almost identical to those of the MLE. The well known downward bias of the MLE for AR(1) coefficients exists for the eigenvalues of the posterior mean of $\boldsymbol{B}_1$ as well under the constant prior. Under the shrinkage prior, the downward bias in eigenvalues of $\boldsymbol{B}_1$ improves slightly. For example, the average of largest eigenvalues under the shrinkage-reference prior is 0.965 and the standard deviation is 0.030. The average of the smallest eigenvalues is 0.534 and the standard deviation is 0.133. A closer examination of the impulse responses reveals from a different angle a pattern of the downward bias of the MLE and the Bayesian estimates under the constant prior. For instance, the true impulse responses of the fifth variable to a shock of the fifth equation at the first lag (denoted as $z_{1,(5,5)}$) is 0.775, and at the 12th lag, the element $z_{12,(5,5)}$ is 0.775 as well. The frequentist averages of the same elements corresponding to the MLE are $\hat{z}_{1,(5,5)} = 0.397$, and $\hat{z}_{12,(5,5)} = 0.004$. Similar numbers are obtained for Bayesian estimates under the constant priors. The same elements corresponding to the Bayesian estimates under the shrinkage-reference prior are $\hat{z}_{1,(5,5)} = 0.605$, and $\hat{z}_{12,(5,5)} = 0.216$. The downward bias under the shrinkage prior is still quite sizeable but is much smaller than that of the MLE and Bayesian estimates under the constant prior.

We conducted more simulations with several parameter settings. To save space, we will not report the details of the simulations but will briefly describe the results instead. When the sample size T is increased, the role of the prior diminishes. When the size of the VAR, p, or the number of lags, L, increases, the opposite is true. When the degrees of freedom parameter of the Student-t distribution v increases, the results become more similar to the case with normal errors. The shrinkage prior tends to shrink the intercept terms more than the non-intercept terms. When the true value of the intercept term is set at a positive value, the dominance of the shrinkage prior over the constant prior becomes less prominent with respect to the symmetric quadratic loss. Stronger correlations of the VAR variables amplify the variances of MLEs and the variances of Bayesian estimates under the constant prior. The shrinkage prior is effective in reducing the variances of the Bayesian estimates and results in smaller losses compared to the constant prior when the VAR errors have strong pairwise correlations. Lastly, regarding the loss function, when the parameter in the asymmetric loss function a is set at -8 instead of -4 the Bayesian estimator $\widehat{\mathbf{\Phi}}_2$ becomes more different from the posterior mean $\widehat{\mathbf{\Phi}}_1$. The downward bias of the posterior mean is better corrected by the estimator based on the loss function with stronger asymmetry.

The above exercises demonstrate that the frequentist properties of alternative estimates can be substantially different under competing priors. In the next section, we will compare Bayesian VAR estimates of the U.S. economy with a data sample of 172 quarters. We will show that the estimates are quite different, and it is important to be aware of the effect of using alternative priors.

6.2 Estimating a VAR of the U.S. Economy

In the past two decades, Bayesian VAR models have been commonly used for analyzing multivariate time series macroeconomic data and addressing policy questions. However, not much attention has been given to the sensitivity of results with respect to researchers' choice of estimator and prior. 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. We set the lag length of the VAR at two, in accordance with the Schwarz criterion. The variables are real GDP, GDP deflator, world commodity price, the Federal Funds rates, non-borrowed reserves and M2 money stock. The commodity price data were obtained from the International Monetary Fund and the rest of data series from the FRED database at the Federal Reserve Bank of St Louis. All variables except the Federal Funds rates 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 et al. 1999). We focus on the role of priors by limiting our attention to the normal model. The Gibbs sampling algorithms for the posteriors under various priors in this application are based on the conditional posteriors given in Section 5.1. The number of MCMC cycles is set at 10000.

6.2.1 Bayesian Estimates and Posterior Risks

As in our numerical example, under the constant prior the posterior mean of $\mathbf{\Phi}$ is very close to the MLE but quite different from that under the shrinkage prior. To be specific, the most prominent differences of the two estimates under alternative priors on $\mathbf{\Phi}$ are in the third and the ninth rows of $\mathbf{\Phi}$, which correspond to the first and second lag parameters of the GDP deflator variable. Most elements of the third row of $\widehat{\mathbf{\Phi}}$ under the constant prior are in similar magnitude to the elements in the ninth row of the same column but have the opposite signs. The pattern of estimates also emerge in the GDP variable (of the second and eighth rows), the non-borrowed reserves variable (of the sixth and the twelfth rows), and the M2 money stock variable (of the seventh and thirteenth rows). The estimates of this VAR suggest some degree of collinearity, which is not uncommon for macroeconomic applications of VARs because macroeconomic time series data often exhibit strong serial and pairwise correlations. Moreover, 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 of similar magnitude and opposite signs, they are also often found to be very sensitive to model specification and sample period. The fact that the estimates of $\mathbf{\Phi}$ under the constant prior are similar to MLEs suggests a possibility of improvement by using alternative priors in place of the constant prior.

As discussed earlier, the shrinkage estimator of James and Stein (1961), which motivates the shrinkage prior, reduces quadratic frequentist loss in estimating a multivariate normal mean. The James-Stein estimator is also known for improving efficiency in the presence of multicollinearity in regression models. Assessing the improvement of shrinkage-prior-based estimates requires computation of losses. In real applications, frequentist risks cannot be calculated since the true parameters are unknown. Theoretical results on admissibility of estimates under alternative priors have not been established in the VAR framework, but most likely the use of a shrinkage prior for $\mathbf{\Phi}$ improves upon the MLE and Bayesian estimates based on the constant prior.

Unlike the frequentist risk, the posterior risk for a given data set can be computed using the MCMC simulation output. For some loss functions this can be done at little additional cost. For example, with the posterior mean estimator $\widehat{\mathbf{\Sigma}} = \mathrm{E}(\mathbf{\Sigma} \mid \boldsymbol{Y})$, the posterior loss of $L_{\Sigma 1}(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})$ is

$$\mathrm{E}\Big[\{trace(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \log|\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}| - p\} \mid \boldsymbol{Y}\Big] = \log|\widehat{\mathbf{\Sigma}}| - \mathrm{E}\{(\log|\mathbf{\Sigma}|) \mid \boldsymbol{Y}\}. \tag{31}$$

To compute $\mathrm{E}\{(\log|\mathbf{\Sigma}|) \mid \boldsymbol{Y}\}$, we decompose the $\mathbf{\Sigma}_k$ matrix in the kth MCMC cycle as $\mathbf{\Sigma}_k = \boldsymbol{O}_k \boldsymbol{D}_k \boldsymbol{O}_k'$, where $\boldsymbol{D}_k$ is the diagonal matrix that consists of eigenvalues of $\mathbf{\Sigma}_k$, i.e., $D_k = diag(d_{k1},$

$d_{k2}, .., d_{kp})$, and $\boldsymbol{O}_k$ is an orthogonal matrix with $\boldsymbol{O}_k \boldsymbol{O}_k' = \boldsymbol{I}$. It follows that $\widehat{\mathrm{E}}\{(\log|\boldsymbol{\Sigma}|) \mid \boldsymbol{Y}\} = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{p} log|d_{ki}|$, which can be computed in MCMC runs. The posterior risk of the estimator $\widehat{\boldsymbol{\Phi}} = \mathrm{E}(\boldsymbol{\Phi} \mid \boldsymbol{Y})$ associated with the loss $L_{\Phi 1}(\widehat{\boldsymbol{\Phi}}, \boldsymbol{\Phi})$ is

$$\mathrm{E}\Big[trace\{(\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi})'(\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi})\} \mid \boldsymbol{Y}\Big] = trace\Big[\mathrm{E}(\boldsymbol{\Phi}'\boldsymbol{\Phi} \mid \boldsymbol{Y}) - \{\mathrm{E}(\boldsymbol{\Phi} \mid \boldsymbol{Y})\}'\{\mathrm{E}(\boldsymbol{\Phi} \mid \boldsymbol{Y})\}\Big]. \quad (32)$$

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

The posterior risks of the posterior mean estimator under alternative priors turned out to be quite different. For instance, under the constant-RATS prior the posterior risks (31) and (32) were 0.126 and 13.703, respectively. Under the shrinkage-reference prior combination, the corresponding posterior losses were 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 regression coefficient matrix $\boldsymbol{\Phi}$.

6.2.2 Impulse Responses

We now compare impulse responses under the constant-RATS and shrinkage-reference priors. To save space, we only plot responses of GDP to shocks in the GDP deflator. We compare the GDP response to an inflation shock under competing priors. As noted earlier, under the constant prior the third and ninth rows of $\widehat{\boldsymbol{\Phi}}_{1CA}$ show similar magnitude but opposite signs. For example, the posterior means of elements $\phi_{(3,1)}$ and $\phi_{(9,1)}$ are 3.057 and -3.714, respectively. In contrast, under the shrinkage prior, the posterior means of elements $\phi_{(3,1)}$ and and $\phi_{(9,1)}$ are 0.358 and -1.134. The difference in the Bayesian estimates $\widehat{\boldsymbol{\Phi}}_{1CA}$ and $\widehat{\boldsymbol{\Phi}}_{1SR}$ corresponds to quite different impulse responses. Figure 1 shows that under the constant prior, after a one-unit inflation shock, GDP initially surges by as much as one percent and then quickly drops. The long-run effect is about -0.2 percent. Given the fact that price level increases after an inflation shock, one may interpret the shock as a positive shock in demand. Under the shrinkage prior, a distinctly different pattern emerges. GDP does not change by much immediately after the inflation shock and drifts slowly below zero. Under the shrinkage prior, an inflation shock plays the role of a negative supply shock (with the consequence of increases in price and decreases in output). Obviously, using different priors on the same data set leads to quite different conclusions.

7 Concluding Remarks

This paper compares frequentist risks of several Bayesian estimates of the VAR regression coefficients $\mathbf{\Phi}$ and covariance matrix $\mathbf{\Sigma}$ under competing priors and data distributions. The asymmetric LINEX estimator $\mathbf{\Phi}$ does better overall than the posterior mean. We do not find an estimator for $\mathbf{\Sigma}$ dominating in all cases. It is therefore well advised to compare all estimates of $\mathbf{\Sigma}$ in applications. The choice of prior has more effect on the Bayesian estimates than the choice of loss function. The shrinkage prior on $\mathbf{\Phi}$ dominates the constant prior in the numerical example. Yang and Berger's reference prior on $\mathbf{\Sigma}$ dominates the Jeffreys prior and the RATS prior. These conclusions hold for Student-t model as well as the normal model. Estimation of a VAR using U.S. macroeconomic data reveals significant difference between estimates of VAR regression coefficients under the shrinkage and constant priors. Similarly, impulse responses of GDP to an inflation shock are shown to be distinctly different under the competing priors.

The study may be extended in several directions. First, 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 estimates are derived from loss functions separable in $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. As a result, inferences for $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are largely independent. Our future research concerns joint Bayesian inference of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ based on an intrinsic joint loss function such as the entropy loss. Third, the present paper only considers linear VAR models. Recently, applications of nonlinear VAR models (e.g., Altissimo and Giovanni 2001) have become more common. Extending our analysis to nonlinear Bayesian VARs will be an interesting future research topic.

Appendix: Sampling from the Posterior of $(\mathbf{\Phi}, \mathbf{\Sigma})$ under Student-t Errors

The algorithms for MCMC computations of posterior distributions of $(\boldsymbol{\phi}, \mathbf{\Sigma})$ depend on the priors. For brevity we only outline the algorithms with the constant prior on $\boldsymbol{\phi}$ and the Jeffreys and reference priors on $\mathbf{\Sigma}$. We only outline the algorithms with Student-t errors, which have one more step than the algorithms with the normal errors. We take the degrees of freedom of the Student-t distribution, v, as unknown. Following the discussion after Fact 5, we use an MCMC algorithm to sample from the joint posterior distribution $(\boldsymbol{\phi}, \mathbf{\Sigma}, \boldsymbol{Q}, v)$. The algorithm used for simulating the posterior under the constant-Jeffreys prior can be found in the following.

Algorithm CJT: Suppose at cycle k we have $(\mathbf{\Phi}_{k-1}, \mathbf{\Sigma}_{k-1}, \boldsymbol{Q}_{k-1}, v_{k-1})$ sampled from cycle $k-1$.

Step 1: Simulate $q_{tk} \sim gamma\Big(\frac{1}{2}(v_{k-1}+p),\ \frac{1}{2}\{v_{k-1} + (\boldsymbol{y}_t - \boldsymbol{x}_t\mathbf{\Phi}_{k-1})'\mathbf{\Sigma}_{k-1}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}_t\mathbf{\Phi}_{k-1})\}\Big)$.

Step 2: Calculate $\widehat{\mathbf{\Phi}}_k = (\boldsymbol{X}'\boldsymbol{Q}_k\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}_k\boldsymbol{Y}$, where $\boldsymbol{Q}_k = Diag(q_{1k}, \cdots, q_{Tk})$.

Step 3: Simulate $\mathbf{\Sigma}_k \sim IW(\boldsymbol{S}_k(\widehat{\mathbf{\Phi}}_k), T - Lp - 1)$, where $\boldsymbol{S}_k(\mathbf{\Phi}) = (\boldsymbol{Y} - \boldsymbol{X}\mathbf{\Phi})'\boldsymbol{Q}_k(\boldsymbol{Y} - \boldsymbol{X}\mathbf{\Phi})$.

Step 4: Simulate $\boldsymbol{\phi}_k \sim N(\widehat{\boldsymbol{\phi}}_k,\ \mathbf{\Sigma}_k \otimes (\boldsymbol{X}'\boldsymbol{Q}_k\boldsymbol{X})^{-1})$.

Step 5: Simulate w conditional on $\boldsymbol{\phi}_k$, $\mathbf{\Sigma}_k$, and $\boldsymbol{Q}_k$ from distribution (26) using the Gilks-Wild (1992) adaptive rejection sampling scheme, and let $v_k = 2w$.

The algorithm using the constant-RATS prior is similar to the one above, with the exception that in Step 3 the distribution of the Inverse Wishart has different degrees of freedom: $\mathbf{\Sigma}_k \sim IW(\boldsymbol{S}_k(\widehat{\mathbf{\Phi}}_k), T)$.

It is much more difficult to simulate from the conditional distribution of $\mathbf{\Sigma}$ under the reference prior. We adopt the hit-and-run algorithm used in Yang and Berger (1994). In implementing the algorithm, we use a one-to-one transformation of $\mathbf{\Sigma}$, namely $\mathbf{\Sigma}^* = \log(\mathbf{\Sigma})$ or $\mathbf{\Sigma} = \exp(\mathbf{\Sigma}^*)$ in the sense that $\mathbf{\Sigma} = \sum_{j=0}^{\infty} \mathbf{\Sigma}^{*j}/j!$. Then the conditional posterior density of $\mathbf{\Sigma}^*$ given $(\boldsymbol{Q}, \boldsymbol{\phi}, \boldsymbol{Y}, v)$ is

$$\pi(\mathbf{\Sigma}^* \mid \boldsymbol{Q}, \boldsymbol{\phi}, v, \boldsymbol{Y}) = \pi(\mathbf{\Sigma}^* \mid \boldsymbol{S}_Q(\boldsymbol{\phi})) \quad \propto \quad \frac{etr\{-\frac{T}{2}\mathbf{\Lambda}^* - \frac{1}{2}(\exp \mathbf{\Sigma}^*)^{-1}\boldsymbol{S}_Q(\boldsymbol{\phi})\}}{\prod_{i<j}(\lambda_i^* - \lambda_j^*)}, \tag{33}$$

where $\mathbf{\Sigma}^* = \boldsymbol{O}'\mathbf{\Lambda}^*\boldsymbol{O}$, $\boldsymbol{O}$ is an orthogonal matrix, and $\mathbf{\Lambda}^* = diag(\lambda_1^*, \cdots, \lambda_p^*)$ with $\lambda_1^* > \ldots > \lambda_p^*$. Note that $\exp(\mathbf{\Sigma}^*) = \boldsymbol{O}'\exp(\mathbf{\Lambda}^*)\boldsymbol{O}$. To simulate $\mathbf{\Sigma}^*$ from (33), we use the following algorithm.

Algorithm CRT: Suppose at cycle k we have $(\mathbf{\Phi}_{k-1}, \mathbf{\Sigma}_{k-1}, \boldsymbol{Q}_{k-1}, v_{k-1})$ sampled from cycle $k-1$.

Step 1: Same as Step 1 of Algorithm CJT.

Step 2: Same as Step 2 of Algorithm CJT.

Step 3: Simulate $\boldsymbol{\phi}_k \sim N(\widehat{\boldsymbol{\phi}}_k,\ \mathbf{\Sigma}_{k-1} \otimes (\boldsymbol{X}'\boldsymbol{Q}_k\boldsymbol{X})^{-1})$.

Step 4: Decompose $\mathbf{\Sigma}_{k-1} = \boldsymbol{O}'\mathbf{\Lambda}\boldsymbol{O}$, where $\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_p)$, $\lambda_1 > \lambda_2 \cdots > \lambda_p$, and $\boldsymbol{O}'\boldsymbol{O} = \boldsymbol{I}$. Let $\lambda_i^* = \log(\lambda_i)$, $\mathbf{\Lambda}^* = diag(\lambda_1^*, \cdots, \lambda_p^*)$, and $\mathbf{\Sigma}_{k-1}^* = \boldsymbol{O}\mathbf{\Lambda}^*\boldsymbol{O}'$.

Step 5: Select a random symmetric $p \times p$ matrix $\boldsymbol{V}$, with elements $v_{ij} = z_{ij}/\sqrt{\sum_{l \le m} z_{lm}^2}$, where $z_{ij} \sim N(0,1)$, $1 \le i \le j \le p$. The other elements of $\boldsymbol{V}$ are defined by symmetry.

Step 6: Generate $\rho \sim N(0,1)$ and set $\boldsymbol{W} = \mathbf{\Sigma}_{k-1}^* + \rho\boldsymbol{V}$. Decompose $\boldsymbol{W} = \boldsymbol{H}'\boldsymbol{C}^*\boldsymbol{H}$, where $\boldsymbol{C}^* = diag(c_1^*, \cdots, c_p^*)$, $c_1^* > c_2^* > \cdots > c_p^*$, and $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}$. Calculate $\boldsymbol{S}_k(\mathbf{\Phi}_k) = \boldsymbol{S}_{kQ} =$

$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}_k)'\boldsymbol{Q}_k(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}_k)$. Compute

$$\begin{aligned}\alpha_k &= \log(\pi(\exp(\boldsymbol{W}) \mid \boldsymbol{S}_{kQ})) - \log(\pi(\exp(\boldsymbol{\Sigma}^*_{k-1})) \mid \boldsymbol{S}_{kQ}) \\ &= \frac{T}{2}\sum_{i=1}^{p}(\lambda_i^* - c_i^*) + \frac{1}{2}trace\{((\exp\boldsymbol{\Sigma}^*_{k-1})^{-1} - (\exp\boldsymbol{W})^{-1})\boldsymbol{S}_{kQ}\} \\ &\quad + \sum_{i<j}\log(\lambda_i^* - \lambda_j^*) - \sum_{i<j}\log(c_i^* - c_j^*).\end{aligned}$$

Step 7: Generate $u \sim Unif(0,1)$.

If $u \leq min(1, \exp(\alpha_k))$, let $\boldsymbol{\Sigma}^*_k = \boldsymbol{W}$ and $\boldsymbol{\Sigma}_k = \boldsymbol{HCH}'$, where $\boldsymbol{C} = diag(e^{c_1}, \cdots, e^{c_p})$; otherwise, let $\boldsymbol{\Sigma}^*_k = \boldsymbol{\Sigma}^*_{k-1}$ and $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{k-1}$.

Step 8: Simulate w conditional on $\boldsymbol{\phi}_k$, $\boldsymbol{\Sigma}_k$, and $\boldsymbol{Q}_k$ from distribution (26) using the Gilks-Wild (1992) adaptive rejection sampling scheme, and let $v_k = 2w$.

When the shrinkage prior is used to replace the constant prior for $\boldsymbol{\phi}$, the algorithms for Bayesian computation need to be modified by adding one step for drawing $\boldsymbol{\phi}$ using Fact 8. In cycle k, $\boldsymbol{\phi}_k$ is drawn in two-steps. First, parameter δ_k is drawn from an inverse gamma distribution, which depends on $\boldsymbol{\phi}_{k-1}$. Then $\boldsymbol{\phi}_k$ is drawn from a multivariate normal distribution that depends on δ_k, $\boldsymbol{\Sigma}_k$, and the data sample.

Algorithms for simulating the posterior of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ with the normal errors are similar to the algorithms presented above, except that in the normal model the steps for simulating $\boldsymbol{Q}$ and v are omitted and $\boldsymbol{Q}$ is fixed at $\boldsymbol{I}_T$.

References

Altissimo, F. and Giovanni, L. V. (2001). The non-linear dynamics of output and unemployment in the U.S. *Journal of Applied Econometrics*, **16**, 461–86.

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.

Barberis, N. (2000). Investing for the long run when the returns are predictable. *Journal of Finance*, **55,** 225–64.

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

Bowman, K.O. and Shenton, L.R. (1988). *Properties of Estimators for the Gamma Distribution.* Marcel Dekker Inc., New York.

Chib, S., Osiewalski, J. and Steel, M. F. J. (1991) Posterior inference on the degrees of freedom parameter in multivariate-t regression models. *Economics Letters*, **37**, 391–397.

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.

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

Fernandez, C. and Steel, M. F. J. (1999). Multivariate Student-t regression models: Pitfalls and inference. *Biometrika*, **86**, 153–167.

Fernandez, C., Osiewalski, J. and Steel, M. F. J. (1997). Classical and Bayesian inference robustness in multivariate regression models. *Journal of American Statistical Association*, **92**, 1434–1444.

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.

Geweke J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, **8**, s19–40.

Gilks, W.R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs sampling. *Applied Statistics,* **41**, 337-348.

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.

Hollifield, B., Koop, G., and Li, K. (2001). A Bayesian analysis of a variance decomposition of stock returns. *Working paper*, Carnegie Mellon University.

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.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley, 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.

Kandel, S., McCulloch R., and Stambaugh, R. F. (1995). Bayesian inference and portfolio efficiency. *Review of Financial Studies*, **8**, 1–53.

Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns. *Journal of Finance*, **51**, 385–424.

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

Lamoureux, C. G. and Zhou, G. (1996). Temporary Components of Stock Returns: What Do the Data Tells Us?, *Review of Financial Studies*, **9**, 1033–1059.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregression-five years of experience. *Journal of Business and Economic Statistics*, **4**, 25-38.

MacKinnon, J. G. and Smith A. A. Jr. (1998). Approximate bias correction in econometrics, *Journal of Econometrics*, **85**, 205–230.

Ni, S. and D. Sun (2003). Noninformative priors and frequentist risks of Bayesian estimators of vector-autoregressive models. *Journal of Econometrics,* **115** 159–197.

Pastor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective *Journal of Financial Economics*, **56**, 335–81.

Pastor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance*, **55**, 179–223.

Osiewalski J. and Steel, M. F. J. (1993). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**, 345–363.

Phillips, P.C.B. (1991). To criticize the critics: an objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, **6**, 423-434.

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

Sims, C.A. (1991). Comment by Christopher A. Sims on 'To criticize the critics'. by Peter C.B. Phillips. *Journal of Applied Econometrics*, **6**, 423-434.

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

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, **54**, 375–421.

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.

Sun, D. and Ni, S. (2003). Bayesian Analysis of Vector-Autoregressive Models with Noninformative Priors. *Journal of Statistical Planning and Inference*, in press.

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. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association,* **71**, 400–405.

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

Zellner, A. and Hong C. (1989). Forecasting international growth rates using Bayesian shrinkage and other procedures *Journal of Econometrics,* **40**, 183–202.

Table 1: Frequentist Average Losses (standard deviations in parentheses) of Competing Bayes Estimates of $\mathbf{\Sigma}$ in the Numerical Example with Normal Errors.

	$L_{\Sigma 1}$	$L_{\Sigma 2}$	$L_{\Sigma 3}$
$\widehat{\mathbf{\Sigma}}_{MLE}$	0.861 (0.403)	0.681 (0.189)	0.516 (0.178)
$\widehat{\mathbf{\Sigma}}_{1CA}$	0.608 (0.308)	0.646 (0.222)	0.415 (0.153)
$\widehat{\mathbf{\Sigma}}_{2CA}$	1.187 (0.501)	0.803 (0.192)	0.660 (0.205)
$\widehat{\mathbf{\Sigma}}_{3CA}$	0.861 (0.403)	0.681 (0.189)	0.516 (0.178)
$\widehat{\mathbf{\Sigma}}_{1CJ}$	0.450 (0.222)	0.800 (0.344)	0.389 (0.142)
$\widehat{\mathbf{\Sigma}}_{2CJ}$	0.862 (0.403)	0.681 (0.189)	0.516 (0.178)
$\widehat{\mathbf{\Sigma}}_{3CJ}$	0.609 (0.309)	0.645 (0.222)	0.415 (0.153)
$\widehat{\mathbf{\Sigma}}_{1CR}$	0.281 (0.172)	0.434 (0.219)	0.234 (0.113)
$\widehat{\mathbf{\Sigma}}_{2CR}$	0.546 (0.314)	0.489 (0.184)	0.353 (0.161)
$\widehat{\mathbf{\Sigma}}_{3CR}$	0.386 (0.238)	0.419 (0.178)	0.273 (0.133)
$\widehat{\mathbf{\Sigma}}_{1SA}$	0.609 (0.308)	0.646 (0.222)	0.415 (0.153)
$\widehat{\mathbf{\Sigma}}_{2SA}$	1.187 (0.501)	0.803 (0.192)	0.661 (0.205)
$\widehat{\mathbf{\Sigma}}_{3SA}$	0.862 (0.403)	0.682 (0.189)	0.516 (0.178)
$\widehat{\mathbf{\Sigma}}_{1SJ}$	0.449 (0.221)	0.801 (0.345)	0.389 (0.143)
$\widehat{\mathbf{\Sigma}}_{2SJ}$	0.862 (0.403)	0.682 (0.189)	0.516 (0.178)
$\widehat{\mathbf{\Sigma}}_{3SJ}$	0.609 (0.308)	0.646 (0.222)	0.415 (0.153)
$\widehat{\mathbf{\Sigma}}_{1SR}$	0.261 (0.163)	0.418 (0.208)	0.221 (0.107)
$\widehat{\mathbf{\Sigma}}_{2SR}$	0.505 (0.300)	0.464 (0.177)	0.332 (0.154)
$\widehat{\mathbf{\Sigma}}_{3SR}$	0.356 (0.226)	0.399 (0.169)	0.256 (0.126)

Table 2: Frequentist Average Losses (standard deviations in parentheses) of Competing Bayes Estimates of $\mathbf{\Phi}$ in the Numerical Example with Normal Errors.

	$L_{\Phi 1}$	$L_{\Phi 2}$
$\widehat{\mathbf{\Phi}}_{MLE}$	11.183 (14.070)	11.614 (6.898)
$\widehat{\mathbf{\Phi}}_{1CA}$	11.184 (14.068)	11.611 (6.893)
$\widehat{\mathbf{\Phi}}_{2CA}$	11.135 (14.055)	9.416 (5.154)
$\widehat{\mathbf{\Phi}}_{1CJ}$	11.184 (14.094)	11.613 (6.905)
$\widehat{\mathbf{\Phi}}_{2CJ}$	11.134 (14.079)	9.156 (4.947)
$\widehat{\mathbf{\Phi}}_{1CR}$	11.185 (14.070)	11.615 (6.894)
$\widehat{\mathbf{\Phi}}_{2CR}$	11.135 (14.057)	9.319 (5.066)
$\widehat{\mathbf{\Phi}}_{1SA}$	1.552 (0.597)	7.254 (3.349)
$\widehat{\mathbf{\Phi}}_{2SA}$	1.523 (0.596)	6.313 (2.786)
$\widehat{\mathbf{\Phi}}_{1SJ}$	1.419 (0.501)	7.174 (3.250)
$\widehat{\mathbf{\Phi}}_{2SJ}$	1.387 (0.499)	6.155 (2.653)
$\widehat{\mathbf{\Phi}}_{1SR}$	1.261 (0.359)	7.176 (3.332)
$\widehat{\mathbf{\Phi}}_{2SR}$	1.231 (0.355)	6.198 (2.739)

Table 3: Frequentist Average Losses of Impulse Responses in the Numerical Example with Normal Errors.

Horizon	MLE	CA	CJ	CR	SA	SJ	SR
1	0.963	0.911	0.848	0.854	0.734	0.662	0.640
2	1.745	1.657	1.603	1.640	1.331	1.249	1.250
3	2.389	2.276	2.238	2.290	1.851	1.772	1.781
4	2.898	2.767	2.747	2.804	2.282	2.208	2.220
5	3.302	3.158	3.152	3.210	2.638	2.568	2.580
6	3.625	3.470	3.476	3.533	2.931	2.867	2.877
7	3.886	3.723	3.738	3.793	3.176	3.116	3.125
8	4.101	3.929	3.953	4.005	3.382	3.325	3.332
9	4.278	4.101	4.131	4.180	3.556	3.502	3.508
10	4.427	4.246	4.283	4.330	3.706	3.653	3.659
11	4.554	4.371	4.417	4.460	3.835	3.785	3.789
12	4.662	4.483	4.540	4.579	3.947	3.900	3.903

Table 4: Frequentist Average Losses of Impulse Responses in the Numerical Example with Student-t Errors (the true degrees of freedom parameter $v = 8$.)

Horizon	CA	CJ	CR	SA	SJ	SR
1	0.930	0.867	0.875	0.747	0.673	0.654
2	1.684	1.630	1.670	1.334	1.252	1.255
3	2.307	2.272	2.327	1.843	1.764	1.776
4	2.798	2.782	2.842	2.264	2.190	2.206
5	3.186	3.184	3.246	2.610	2.541	2.558
6	3.492	3.502	3.564	2.895	2.831	2.849
7	3.737	3.756	3.816	3.133	3.073	3.090
8	3.934	3.960	4.019	3.333	3.276	3.293
9	4.095	4.127	4.184	3.502	3.448	3.465
10	4.229	4.266	4.321	3.646	3.595	3.612
11	4.342	4.386	4.438	3.771	3.721	3.738
12	4.441	4.493	4.542	3.879	3.832	3.848

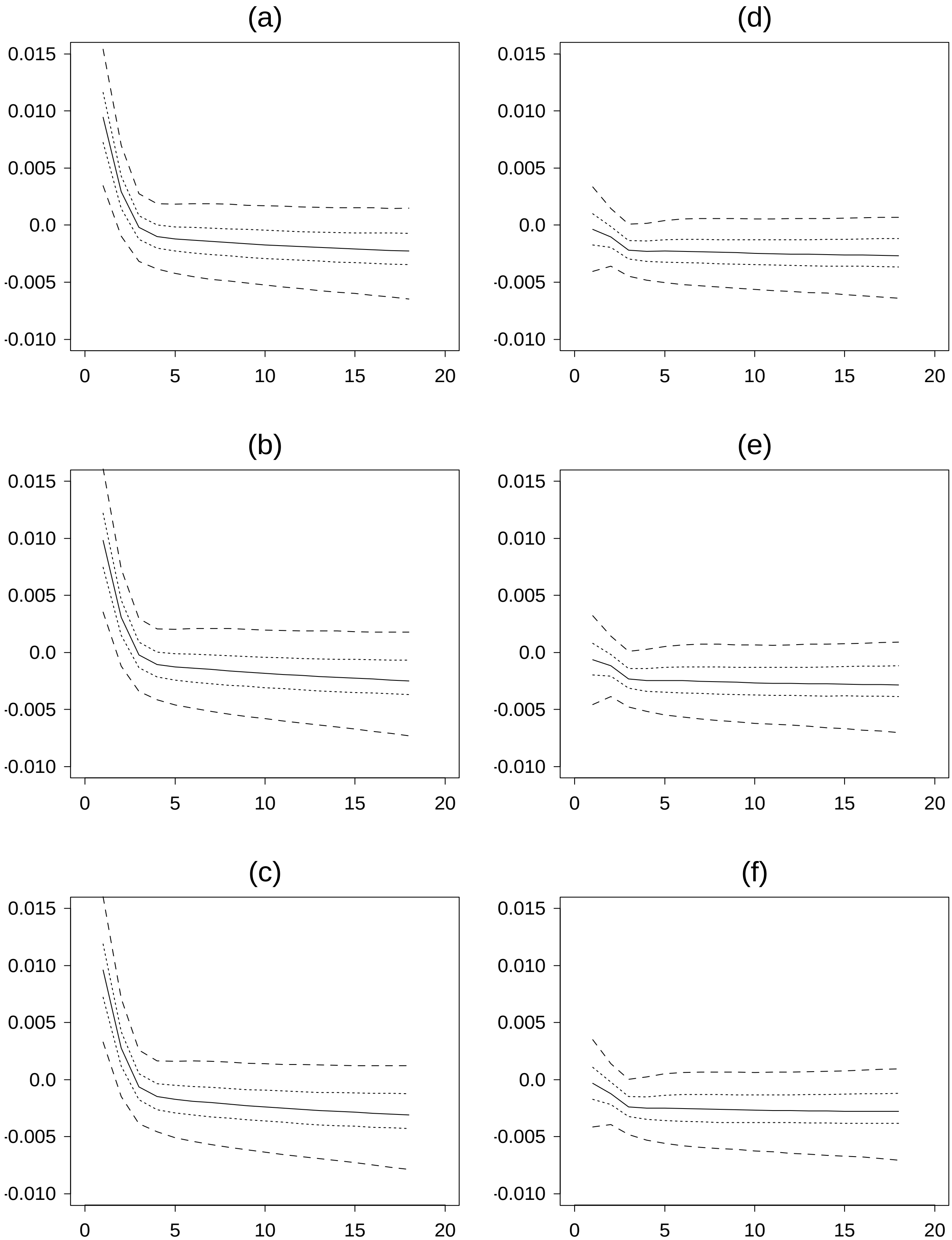

Figure 1: Responses of GDP to an Inflation Shock.

Dashed line = 10 or 90 percentile, dotted line =32 or 68 percentile, solid line = posterior mean of impulse responses. The panels are GDP responses (a) under constant-RATS prior; (b) under constant-Jeffreys prior; (c) under constant-reference prior; (d) under shrinkage-RATS prior; (e) under shrinkage-Jeffreys prior; (f) under shrinkage-reference prior.