Intrinsic Bayesian estimation of VAR impulse responses

Shawn Ni[a], Dongchu Sun[b], and Xiaoqian Sun[c]

[a]Department of Economics, University of Missouri, Columbia, MO 65211, U.S.A.

[b]Department of Statistics, Virginia Tech, Blacksburg, VT 24061, U.S.A.

[c]Department of Statistics, University of Missouri, Columbia, MO 65211, U.S.A.

Abstract

We propose an information-theoretic alternative to the conventional Bayesian posterior mean estimator of impulse responses in Vector Autoregressive (VAR) models. The proposed estimator is based on the intrinsic entropy loss function. The loss function is invariant to nonlinear transformations of parameters, consequently intrinsic estimation of impulse responses is equivalent to that of VAR parameters. The Bayesian estimator under the entropy loss involves a frequentist expectation of regressors. We propose Markov Chain Monte Carlo (MCMC) algorithms to simulate the posterior of the frequentist expectation of regressors and compute the Bayesian estimates. We estimate the VAR impulse responses in two applications.

KEY WORDS: Bayesian VAR, Entropy Loss, Latent Variables, Markov Chain Monte Carlo.

1 Introduction

In applications of Vector Autoregressive (VAR) models, researchers are often more interested in inferences of impulse responses, which are nonlinear functions of VAR parameters, than VAR parameters themselves. The Bayesian approach has been effective for finite sample inferences of impulse responses (for examples, see Sims (1980) and Sims and Zha (1998)). Although posterior distributions of quantities of interest are the ultimate result of Bayesian analysis, they are difficult to report in multiple parameter settings. Bayesian estimation of impulse responses is therefore useful in summarizing the posterior for economic policy making. In Bayesian estimation of impulse responses, the default estimator is the posterior mean. In this paper, we propose an alternative Bayesian estimator based on an information-theoretic argument. We derive the proposed estimator and conduct Markov Chain Monte Carlo (MCMC) simulations for Bayesian estimations in macroeconomic applications.

In the following, we provide an information-theoretic argument for using the intrinsic loss function for Bayesian analysis. For a parameter θ, a Bayesian estimator $\widehat{\theta}$ minimizes the posterior expected loss $L(\widehat{\theta}, \theta)$ ($\widehat{\theta}$ is used to denote an estimator in general and an estimator under a particular loss). The estimator depends on the choice of loss function. For example, the posterior mean estimator is optimal if $L(\theta, \widehat{\theta})$ is a quadratic loss function. On the principle of Bayesian estimation, Bernardo and Juárez (2003) pointed out that

"*In practise, in most situations where point estimation is of interest, an objective point estimation is actually required: objective in the very precise sense of exclusively depending on the assumed probability model (i.e., on the conditional distribution of the data given the parameters) and the available data. Moreover, in purely inferential settings (where interest focuses on the actual mechanism which governs the data) this estimate is typically required to be invariant under one-to-one transformation of either the data or the parameter space.*"

They argue that the loss function should not concern a distance between θ and $\widehat{\theta}$; instead it should be the *intrinsic discrepancy* $L\{f(x \mid \theta), f(x \mid \widehat{\theta})\}$– the distance between the probability models $f(x \mid \theta)$ and $f(x \mid \widehat{\theta})$. Bernardo and Juárez (2003) and Robert (1994, 1996) proposed using the logarithmic divergence (also known as the Kullback-Leibler divergence or the entropy loss) as the intrinsic loss. The intrinsic loss is invariant to transformation of data x or parameter θ. Suppose the subject of interest is a one-to-one transformation of parameter $\alpha = \varphi(\theta)$ instead of θ itself; then the intrinsic loss with respect to α is equivalent to that with respect to θ.

In a VAR setting, intrinsic estimation of impulse responses is advisable. By definition, impulse responses are nonlinear functions of the VAR regression coefficients (represented by $\boldsymbol{B}$ hereafter) and the covariance matrix (represented by $\boldsymbol{\Sigma}$ hereafter) of the VAR error term. The most commonly used VAR impulse responses to orthogonal shocks are based upon Cholesky decomposition of $\boldsymbol{\Sigma}$, which depends on the ordering of variables. Koop et al. (1996) and Pesaran and Shin (1998) proposed generalized impulse responses, which are nonlinear functions of $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$ as well but independent of variable ordering. For many empirical applications, both types of impulse responses may be of interest. When we report the posterior mean of the impulse responses, we in effect make a statement on our best estimate of them with respect to a metric of quadratic distance. However, the VAR parameters $(\boldsymbol{\Sigma},\boldsymbol{B})$ corresponding to the posterior mean of impulse responses are not the posterior mean of $(\boldsymbol{\Sigma},\boldsymbol{B})$, our best estimate based on the quadratic loss. Reporting the posterior mean estimates of generalized impulse responses and those of impulse responses to orthogonal shocks amounts to giving different sets of assessment on the values of the underlying VAR parameters. Intrinsic estimation under the entropy loss avoids this pitfall. Because there is a one-to-one mapping between the VAR parameters $(\boldsymbol{\Sigma},\boldsymbol{B})$ and impulse responses, estimation of $(\boldsymbol{\Sigma},\boldsymbol{B})$ is equivalent to estimation of the VAR impulse responses. In other words, the Bayesian estimators of impulse responses coincide with the impulse response functions of the Bayesian estimators of VAR parameters.

In this paper, we provide analytical as well as empirical results on the Bayesian estimator of VAR impulse responses under the entropy loss. We use $\boldsymbol{\Phi}$ to denote the intercept vector $\boldsymbol{c}$ and coefficient matrix $\boldsymbol{B}$. We show that the entropy loss on $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ is non-separable in $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$, which can be written as the sum of losses pertaining to the covariance matrix $\boldsymbol{\Sigma}$ (a pseudo entry loss) and normalized estimation error of $\boldsymbol{\Phi}$, with the weighting matrix being the frequentist expectation of a nonlinear function of regressors.

Besides the invariance property, the functional form of intrinsic loss function described above lends support to its use. The part of the loss function associated with $\boldsymbol{\Phi}$ can be viewed as quadratic with a weighting matrix that depends on $\boldsymbol{\Sigma}$, $\boldsymbol{\Phi}$ and data. In particular, the $\boldsymbol{\Phi}$ part of the loss is the estimation error in $\boldsymbol{\Phi}$ weighted by their contributions to the normalized estimate errors of residuals. It is in the spirit of Zellner's "precision of the estimation" loss. Under the widely used quadratic loss, estimation errors in elements of $\boldsymbol{\Phi}$ are weighted by constants and the weights do not reflect the contribution of the estimation errors in coefficient to the estimation errors of the residuals (i.e., the estimation errors of the overall model); in contrast, they do under the intrinsic loss function.

To intuitively illustrate how the intrinsic loss function weighs the estimation error, we consider a simple AR(1) model: $y_t = \rho y_{t-1} + \epsilon_t$, for $t = 1, \cdots, T$, where ϵ_t is iid $N(\mathbf{0}, 1)$, $y_0 = 0$, and ρ is the only unknown parameter. Under a flat prior, the posterior of ρ given data $\boldsymbol{y}_D$ is a normal distribution, $N(\widehat{\rho}_M, (\sum_{t=1}^T y_{Dt-1}^2)^{-1})$. The posterior mean $\widehat{\rho}_M$ is the Maximum Likelihood Estimator (MLE), $(\sum_{t=1}^T y_{Dt} y_{Dt-1})(\sum_{t=1}^T y_{Dt-1}^2)^{-1}$. It is well known that $\widehat{\rho}_M$ is biased downward, especially when the true parameter is close to unity (see MacKinnon and Smith, 1998.) For the AR(1) model with normal density f, the entropy loss is $L(\widehat{\rho}, \rho) = \int \log \left\{ \frac{f(\boldsymbol{y}|\rho)}{f(\boldsymbol{y}|\widehat{\rho})} \right\} f(\boldsymbol{y} \mid \rho) d\boldsymbol{y} = \frac{(\rho - \widehat{\rho})^2}{2} \delta(\rho)$, where $\delta(\rho) = E_{\boldsymbol{y}|\rho} \sum_{t=1}^T y_{t-1}^2$. Under the entropy loss, the Bayesian estimator for ρ is $\widehat{\rho}_E = argminE[\delta(\rho)(\rho - \widehat{\rho})^2 \mid \boldsymbol{y}_D] = E\{\delta(\rho)\rho \mid \boldsymbol{y}_D\} / E\{\delta(\rho) \mid \boldsymbol{y}_D\}$. Note that if ρ is positive, ρ and $\delta(\rho)$ are positively correlated. It follows that the Bayes estimator under the entropy loss for positive ρ is always larger than the posterior mean.

The intrinsic estimator has a theoretically appealing interpretation. Note that the entropy loss can be viewed as the weighted squares of estimation error $\rho - \widehat{\rho}$. The weight $\delta(\rho)$ is $E_{\boldsymbol{y}|\rho} \sum_{t=1}^T y_{t-1}^2$, the frequentist expectation of the *posterior precision*. The functional form means that the intrinsic loss function penalizes estimation errors more severely in the region where ρ is likely to generate data that give rise to high posterior precision. The quadratic loss, on the other hand, imposes a constant weight on all regions of ρ.

The rest of the paper is organized as follows. In Section 2, we derive the Bayesian VAR estimator under the entropy loss function. We show that the Bayesian estimator for $\mathbf{\Sigma}$ is larger than the posterior mean. The Bayesian estimator for $\mathbf{\Phi}$, which is also different from the posterior mean, involves frequentist moments of VAR variables. In Section 3, we discuss priors and prove the existence of posterior moments pertaining to the Bayesian estimator. In Section 4, we discuss several MCMC algorithms for posterior simulation. In Section 5, we estimate VAR impulse responses in two applications on the U.S. macroeconomic data. In Section 6, we give some concluding remarks.

2 The Entropy Loss Function and Bayesian Estimation

2.1 VAR Model

A VAR of a p dimensional column variable $\boldsymbol{y}_t$, $(t = 1, .., T,)$ is

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

where the VAR lag L is a known positive integer, the intercept $\boldsymbol{c}$ is a $1 \times p$ unknown vector, $\boldsymbol{B}_j$ is an unknown $p \times p$ matrix, $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are iid $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ errors, and the covariance $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix. We can rewrite (1) in the familiar matrix form,

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

where $\boldsymbol{x}_t' = (1, \boldsymbol{y}_{t-1}', \cdots, \boldsymbol{y}_{t-L}')$,

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

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

Under the normality assumption the likelihood function of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is

$$f(\boldsymbol{\Phi}, \boldsymbol{\Sigma}) \ = \ \frac{1}{|\boldsymbol{\Sigma}|^{T/2}} \mathrm{etr} \left\{ -\frac{1}{2} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}) \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi})' \right\}. \tag{3}$$

Here and hereafter $etr(\mathbf{A})$ is $\exp(trace(\mathbf{A}))$ of matrix $\mathbf{A}$. The MLEs of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ 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

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

We assume that $T \geq Lp + p + 1$ and that the MLEs of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ exist. The MLEs $\widehat{\boldsymbol{\Phi}}_{MLE}$ and $\widehat{\boldsymbol{\Sigma}}_{MLE}$ are known to be unsatisfactory when the sample size T is small relative to the number of parameters in the model. The "over-parameterization" of VARs often leads to erratic behavior of the MLEs. In addition, drawing finite sample inferences of the VAR parameters is a challenge. Frequentist finite sample distribution is not available in analytical form for the VAR model. On the other hand, with a large number of parameters and limited data observations, asymptotic theories may not be good guidance for finite sample properties. In particular, when nonlinear functions of the VAR coefficients (such as impulse responses) are of interest, the asymptotic theory involves approximation of nonlinear functions, and the approximation becomes worse the more nonlinear the functions are (see Kilian (1999)). In practice, the Bayesian approach is widely used for finite sample inferences of VAR model.

Bayesian estimators of parameters $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ are determined by the choice of prior, likelihood, and loss function. Many Bayesian researchers may argue that instead of focusing on the estimator of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ under a particular loss function, one should report the posteriors. A related argument is that elicitation of priors is far more important than the consideration of loss functions. We are sympathetic to the arguments and have devoted much of our research efforts on analyzing properties of the posteriors of VAR parameters $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ under various priors. However, we observe that when there are a large number of parameters, the posteriors of the VAR parameters are too complicated to report and point estimates can be a useful summarizing device. Often times researchers are interested in estimating some components in the covariance matrix $\boldsymbol{\Sigma}$ or certain lag coefficients in $\boldsymbol{\Phi}$ or nonlinear functions of VAR parameters. In reporting impulse responses of a VAR, economists often need to summarize the posterior in terms of Bayesian estimates for policy makers, e.g., point estimates of the GDP responses to a monetary policy shock. It is natural to ask whether we should use posterior mean estimator or some alternative estimators. To justify choosing one estimator over another we need to investigate the loss functions based on which the estimators are derived.

2.2 Impulse Responses

A covariance stationary VAR implies $\boldsymbol{y}_t' = \mathbb{E}_0 \boldsymbol{y}_t' + \sum_{j=0}^{t-1} \boldsymbol{\epsilon}_{t-j}' \boldsymbol{H}_j$, where $\boldsymbol{H}_0 = \boldsymbol{I}$ is the p by p identity matrix, and the impulse responses of $\boldsymbol{y}_t$ to a shock $\boldsymbol{\epsilon}_{t-j}$ occurred j periods earlier is $\boldsymbol{H}_j = \sum_{i=1}^{j} \boldsymbol{B}_i \boldsymbol{H}_{j-i}$, where $\boldsymbol{B}_i = 0$ for i larger than L and $\boldsymbol{B}_0 = \boldsymbol{I}$. Note that correlation of the components of the vector of errors $\boldsymbol{\epsilon}_t$ makes the economic interpretation of the $\boldsymbol{H}_j$ matrix ambiguous. Sims (1980) suggests that researchers report impulse responses to orthogonalized (structural) errors. Orthogonalization of the errors can be achieved through the Cholesky decomposition of the covariance matrix,

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi}' \boldsymbol{\Psi}, \tag{6}$$

where $\boldsymbol{\Psi}$ is an upper triangular positive definite matrix. The VAR error vector $\boldsymbol{\epsilon}_t$ is mapped to structural shock vector $\boldsymbol{u}_t$ through $\boldsymbol{u}_t' = \boldsymbol{\epsilon}_t' \boldsymbol{\Psi}^{-1}$. The response of $\boldsymbol{y}_t'$ to a unit shock of the ith element of $\boldsymbol{u}_{t-j}'$ is the ith row of $\boldsymbol{\Psi} \boldsymbol{H}_j$. We call $\boldsymbol{Z}_j = \boldsymbol{\Psi} \boldsymbol{H}_j$ the impulse responses of $\boldsymbol{y}_t$ to structural shocks $\boldsymbol{u}_{t-j}$. By definition, impulse responses are nonlinear functions of $(\boldsymbol{B}, \boldsymbol{\Psi})$. The nonlinearity makes it difficult to derive frequentist inference but does not pose difficulties for Bayesian computations as long as posteriors of $(\boldsymbol{B}, \boldsymbol{\Psi})$ are available.

There is a one-to-one mapping between impulse responses $\boldsymbol{Z}_0, \cdots, \boldsymbol{Z}_L$ to VAR parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{B}_1, \cdots, \boldsymbol{B}_L$. By the definition of impulse responses and the uniqueness of decomposition (6), $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$ uniquely determine $(\boldsymbol{Z}_0, \cdots, \boldsymbol{Z}_L)$. It is straightforward to verify that the reverse is true as well. Note that $\boldsymbol{\Psi} = \boldsymbol{Z}_0$, and $\boldsymbol{B}_1 = \boldsymbol{H}_1 = \boldsymbol{\Psi}^{-1}\boldsymbol{Z}_1 = \boldsymbol{Z}_0^{-1}\boldsymbol{Z}_1$. For $j \leq L$, we can derive $\boldsymbol{B}_j$ from $(\boldsymbol{Z}_0, \cdots, \boldsymbol{Z}_j)$ and $(\boldsymbol{B}_0, \cdots, \boldsymbol{B}_{j-1})$ as follows: From $\boldsymbol{Z}_0^{-1}\boldsymbol{Z}_j = \boldsymbol{H}_j = \sum_{i=1}^{j} \boldsymbol{B}_i \boldsymbol{H}_{j-i}$, we have $\boldsymbol{B}_j = \boldsymbol{Z}_0^{-1}\boldsymbol{Z}_j - \sum_{i=1}^{j-1} \boldsymbol{B}_i \boldsymbol{H}_{j-i}$. Therefore, $\boldsymbol{\Theta} = (\boldsymbol{\Sigma}, \boldsymbol{B})$ is uniquely determined by $(\boldsymbol{Z}_0, \cdots, \boldsymbol{Z}_L)$. It follows that for inferential purposes, the likelihood of VAR model can be written in terms of impulse responses $\boldsymbol{Z} = (\boldsymbol{Z}_0, \cdots, \boldsymbol{Z}_L)$ (in addition to the intercept $\boldsymbol{c}$). For simplicity we denote the one-to-one mapping as $\boldsymbol{Z} = \varphi(\boldsymbol{\Theta})$. In VAR applications, researchers are generally not interested in the intercept vector $\boldsymbol{c}$, which is treated as nuisance parameters.

Note that the quadratic loss is not invariant to parameter transformation. Denote $\widehat{\boldsymbol{Z}}_{Mean}$ as the posterior mean of $\boldsymbol{Z}$ and $\widehat{\boldsymbol{\Theta}}_{Mean}$ the posterior mean of $\boldsymbol{\Theta}$. Then $\widehat{\boldsymbol{Z}}_{Mean}$ is not the same as $\varphi\,(\widehat{\boldsymbol{\Theta}}_{Mean})$. Ni and Sun (2003) found that frequentist averages of quadratic losses with respect to the posterior mean of VAR parameters may not correspond closely with frequentist averages of quadratic losses pertaining to the posterior mean of impulse responses. Under the entropy loss, the estimators have the invariance property so that $\widehat{\boldsymbol{Z}}_E = \varphi(\widehat{\boldsymbol{\Theta}}_E)$. This implies that we can focus on the estimation of $\boldsymbol{\Theta}$ even though the subject of interest is the impulse responses $\boldsymbol{Z}$.

Impulse responses to orthogonalized shocks depend on the ordering of variables. Koop et al. (1996) and Pesaran and Shin (1998) suggested using generalized impulse responses to avoid ordering VAR models. Generalized impulse response depicts the effect of a shock (corresponding to one of the VAR residuals) with the effects of all other shocks integrated out. Using the notation set earlier, the generalized response of $\boldsymbol{y}_t'$ to the ith element shock of $\boldsymbol{\epsilon}_{t-j}'$ of the size of one standard deviation $\sigma_{ii}^{\frac{1}{2}}$ (σ_{ii} is the ith diagonal element of $\boldsymbol{\Sigma}$) is defined as $\sigma_{ii}^{-\frac{1}{2}}$ times the ith row of $\boldsymbol{\Sigma}\boldsymbol{H}_j$, which is equivalent of putting the ith variable at the top of the Cholesky order. The intrinsic estimates of generalized impulse responses and impulse responses to orthogonalized shocks for a VAR model correspond to different nonlinear functions of the same estimates of VAR parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$.

2.3 The Entropy Loss Function

The general form of the entropy loss was defined in Robert (1994, p74). For the VAR model, it is given by

$$\begin{aligned} L_E(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}}; \boldsymbol{\Phi}, \boldsymbol{\Sigma}) &= \int \log\left\{\frac{f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma})}{f(\boldsymbol{Y} \mid \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})}\right\} f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma})\, d\boldsymbol{Y} \\ &= \mathbb{E}_{(\boldsymbol{Y}\mid\boldsymbol{\Phi},\boldsymbol{\Sigma})} \log\left\{\frac{f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma})}{f(\boldsymbol{Y} \mid \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})}\right\}, \end{aligned} \tag{7}$$

where f is the density of VAR variables $\boldsymbol{Y}$. In information theory, $\log(1/f(\boldsymbol{Y}))$ is often used to measure the content of information regarding the VAR parameters when a researcher observes $\boldsymbol{Y}$. Thus, the entropy loss can be interpreted as the expected difference in information gained from data observation when researcher's estimates of the VAR parameters are $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})$ instead of the true parameters $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$.

Note that for computing the frequentist expectation in the loss function, $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})$ are not treated as functions of $\boldsymbol{Y}$. Naturally, the larger the entropy loss the larger the difference between $f(\boldsymbol{Y} \mid \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})$ and the true model $f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma})$. The entropy has been used in a variety of problems in econometrics and statistics as a measure of distance between distributions. For instance, Kitamura and Stutzer (1997) used the Kullback-Leibler distance to derive a frequentist estimator for nonlinear models. Recently, the entropy divergence was used by Robertson, Tallman, and Whiteman (2003) to measure the accuracy of Bayesian VAR forecasting density and by Fernandez-Villaverde and Rubio-Ramirez (2004) to gauge the asymptotic convergence of parameters of models selected based on Bayes factors. The entropy loss function has not been used for estimation of Bayesian VAR parameters, despite the advantages stated in the introduction. The current study will fill the void.

In the following, for the VAR with multivariate normal errors we decompose the entropy loss L_E into two parts. One part measures the loss associated with the covariance matrix $\boldsymbol{\Sigma}$ only; the other part measures the loss of VAR coefficients but is related to the covariance matrix and frequentist expectation $\mathbb{E}_{(\boldsymbol{Y}\mid\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})$. Throughout the paper, the frequentist expectation is always defined with a given initial value $\boldsymbol{Y}_0$ ($\boldsymbol{Y}_0' = (\boldsymbol{y}_0', \boldsymbol{y}_{-1}', \cdots, \boldsymbol{y}_{1-L}')$). In our VAR notations of the previous section, $\boldsymbol{X}$ are the lags of $\boldsymbol{Y}$. We will use both $\boldsymbol{Y}$ and $\boldsymbol{X}$ as symbols of VAR variables. Note that conditional on the initial state $\boldsymbol{Y}_0$, for a finite sample $\mathbb{E}_{(\boldsymbol{Y}\mid\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})$ exists even when the VAR has explosive roots.

Lemma 1 *Denote the $(1+Lp)\times(1+Lp)$ frequentist expectation matrix as*

$$\boldsymbol{G} = \mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X}). \tag{8}$$

The entropy loss function L_E can be decomposed into two parts,

$$L_E(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Sigma}};\boldsymbol{\Phi},\boldsymbol{\Sigma}) \quad = \quad L_{E1}(\widehat{\boldsymbol{\Sigma}};\boldsymbol{\Sigma}) + L_{E2}(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Sigma}};\boldsymbol{\Phi},\boldsymbol{\Sigma}), \tag{9}$$

where

$$\begin{aligned} L_{E1}(\widehat{\boldsymbol{\Sigma}};\boldsymbol{\Sigma}) &= \frac{T}{2}\Big\{tr(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p\Big\}, & (10)\\ L_{E2}(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Sigma}};\boldsymbol{\Phi},\boldsymbol{\Sigma}) &= \frac{1}{2}tr\Big[\widehat{\boldsymbol{\Sigma}}^{-1}\{(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})'\boldsymbol{G}(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})\}\Big]. & (11)\end{aligned}$$

Proof. Because $\boldsymbol{y}_t - \boldsymbol{X}_t\boldsymbol{\Phi}$, $t=1,\cdots,T$ are iid $N_p(\boldsymbol{0},\boldsymbol{\Sigma})$, we have $\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi}) = \boldsymbol{0}$, $\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}\{(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi})'\boldsymbol{X}\boldsymbol{\Phi}\} = \boldsymbol{0}$, and $\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}\{(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi})\} = T\boldsymbol{\Sigma}$. Then

$$\begin{aligned} &\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}\left\{\log\frac{|\boldsymbol{\Sigma}|^{-T/2}etr\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\Phi})'\}}{|\widehat{\boldsymbol{\Sigma}}|^{-T/2}etr\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\Phi}})\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\Phi}})'\}}\right\}\\ =\ & \frac{T}{2}(\log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p) + \frac{1}{2}tr\left[\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}\left\{(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\Phi}})\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\Phi}})'\right\}\right]\\ =\ & \frac{T}{2}(\log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p) + \frac{T}{2}tr(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) + \frac{1}{2}tr\left[\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}\left\{\boldsymbol{X}(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})'\boldsymbol{X}'\right\}\right]\\ =\ & \frac{T}{2}\left\{tr(\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) - \log|\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p\right\} + \frac{1}{2}tr\left\{\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})'\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})\right\}. \end{aligned}$$

The result follows. □

The part of the intrinsic loss function associated with the regression coefficients turns out to be related to a conventional loss function. For estimation of a matrix parameter such as $\boldsymbol{\Phi}$ in simultaneous equations context, Zellner (1978, 1998) proposed a "precision of the estimation" loss that reflects a reasonable scaling. Denote $\widehat{\epsilon} = \boldsymbol{X}\boldsymbol{\Phi} - \boldsymbol{X}\widehat{\boldsymbol{\Phi}}$, the difference between estimated residuals and the true residuals. With this notation, L_{E2} can be rewritten as $\frac{1}{2}tr\left\{\widehat{\boldsymbol{\Sigma}}^{-1}\mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\widehat{\epsilon}'\widehat{\epsilon})\right\}$, nearly identical to Zellner's "precision of the estimation" loss function in the VAR setting, $tr\left\{\boldsymbol{\Sigma}^{-1}(\widehat{\epsilon}'\widehat{\epsilon})\right\}$, which can also be written as $tr\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\Phi}-\widehat{\boldsymbol{\Phi}})\right\}$. In this loss function, elements of estimation errors in $\boldsymbol{\Phi}$ are weighted according to their contributions to the estimate errors of residuals normalized by the covariance of the residuals. There is, however, an important difference between Zellner's loss function and the intrinsic loss. In Zellner's simultaneous equations model, $\boldsymbol{X}'\boldsymbol{X}$ is taken as given, but in VAR the predetermined variable $\boldsymbol{X}'\boldsymbol{X}$ depends on parameters $(\boldsymbol{\Phi},\boldsymbol{\Sigma})$. Hereafter, we use $\boldsymbol{Y}_D$ to denote the observation of variable $\boldsymbol{Y}$.

Theorem 1 *Under the loss (9), the generalized Bayesian estimator of* $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ *is*

$$\widehat{\boldsymbol{\Phi}}_E = \Big\{\mathbb{E}(\boldsymbol{G} \mid \boldsymbol{Y}_D)\Big\}^{-1}\mathbb{E}(\boldsymbol{G}\boldsymbol{\Phi} \mid \boldsymbol{Y}_D), \tag{12}$$

$$\widehat{\boldsymbol{\Sigma}}_E = \mathbb{E}(\boldsymbol{\Sigma} \mid \boldsymbol{Y}_D) + \frac{1}{T}\mathbb{E}\{(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E)'\boldsymbol{G}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E) \mid \boldsymbol{Y}_D\}, \tag{13}$$

where $\boldsymbol{G}$ *is given by (8).*

Proof. Let $(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}})$ denote an arbitrary estimator of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. For the loss function L_E and posterior $\pi(\boldsymbol{\Phi}, \boldsymbol{\Sigma} \mid \boldsymbol{Y}_D)$, the expected posterior loss is

$$\begin{aligned} R(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}} \mid \boldsymbol{Y}_D) &= \mathbb{E}\Big\{L_E(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}}; \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \mid \boldsymbol{Y}_D\Big\} \\ &= \mathbb{E}\Big\{L_{E1}(\widehat{\boldsymbol{\Sigma}}; \boldsymbol{\Sigma}) \mid \boldsymbol{Y}_D\Big\} + \mathbb{E}\Big\{L_{E2}(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}}; \boldsymbol{\Phi}, \boldsymbol{\Sigma}) \mid \boldsymbol{Y}_D\Big\}. \end{aligned}$$

The Bayesian estimator, which minimizes the expected posterior loss, can be derived through the first order conditions. Note that for any matrices $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$,

$$\frac{\partial tr(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}'\boldsymbol{C})}{\partial \boldsymbol{A}} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{B} + \boldsymbol{C}'\boldsymbol{A}\boldsymbol{B}'.$$

It follows that

$$\frac{\partial R(\widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Sigma}} \mid \boldsymbol{Y}_D)}{\partial \widehat{\boldsymbol{\Phi}}} = \mathbb{E}\{\boldsymbol{G}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}) \mid \boldsymbol{Y}_D\}\widehat{\boldsymbol{\Sigma}}^{-1}. \tag{14}$$

Setting the right hand side of Equation (14) to 0 gives the Bayesian estimator for $\boldsymbol{\Phi}$ in (12). Using the result that for any matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\frac{\partial \log(|\boldsymbol{A}|)}{\partial \boldsymbol{A}} = (\boldsymbol{A}^{-1})', \quad \frac{\partial tr(\boldsymbol{A}^{-1}\boldsymbol{B})}{\partial \boldsymbol{A}} = -(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})',$$

we have

$$\frac{\partial}{\partial \widehat{\boldsymbol{\Sigma}}}R(\widehat{\boldsymbol{\Phi}}_E, \widehat{\boldsymbol{\Sigma}} \mid \boldsymbol{Y}_D) = \mathbb{E}\left\{\frac{T}{2}(-\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1} + \widehat{\boldsymbol{\Sigma}}^{-1}) - \frac{1}{2}\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E)'\boldsymbol{G}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E)\widehat{\boldsymbol{\Sigma}}^{-1} \,\Big|\, \boldsymbol{Y}_D\right\}. \tag{15}$$

Let the right hand side of Equation (15) be 0, we get

$$\mathbb{E}\Big\{\boldsymbol{\Sigma} - \widehat{\boldsymbol{\Sigma}} + \frac{1}{T}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E)'\boldsymbol{G}(\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}_E) \,\Big|\boldsymbol{Y}_D\Big\} = 0.$$

From this equation the estimator $\widehat{\boldsymbol{\Sigma}}_E$ can be derived. $\square$

2.4 Comparing the Intrinsic Bayesian Estimator with the Posterior Mean

In the existing literature, researchers usually consider loss functions that consist of separable parts for $\mathbf{\Phi}$ and $\mathbf{\Sigma}$, with each part taking a given parametric form (e.g., a quadratic function). The overall loss with respect to $(\mathbf{\Phi}, \mathbf{\Sigma})$ is then in the form of

$$L(\widehat{\mathbf{\Phi}}, \widehat{\mathbf{\Sigma}}; \mathbf{\Phi}, \mathbf{\Sigma}) = L_1(\widehat{\mathbf{\Sigma}}; \mathbf{\Sigma}) + L_2(\widehat{\mathbf{\Phi}}; \mathbf{\Phi}). \tag{16}$$

In this setting, a Bayesian estimator $\widehat{\mathbf{\Sigma}}$ is derived separately from the estimator $\widehat{\mathbf{\Phi}}$. Under the normality assumption, the entropy function can be compared to a loss function in the form of (16) that yields the posterior mean estimator.

Consider the following loss functions closely related to the entropy loss: for $\mathbf{\Sigma}$ we consider a pseudo entropy loss function

$$L_1(\widehat{\mathbf{\Sigma}}; \mathbf{\Sigma}) \quad = \quad \frac{T}{2}\{tr(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \log|\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}| - p\}, \tag{17}$$

where p is the number of variables in the VAR; for $\mathbf{\Phi}$ we consider a quadratic function

$$L_2(\widehat{\mathbf{\Phi}}; \mathbf{\Phi}) \quad = \quad \frac{T}{2} tr\{(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})'\mathbf{W}(\widehat{\mathbf{\Phi}} - \mathbf{\Phi})\}, \tag{18}$$

where $\mathbf{W}$ is a constant weighting matrix. Bayesian estimators of $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ can be derived separately from minimizing posterior expected losses regarding $\mathbf{\Sigma}$ and $\mathbf{\Phi}$ respectively. The separable loss function is associated with the posterior mean estimator. The following fact is straightforward.

Fact 1 *(a) Under the loss L_1, the generalized Bayesian estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}}_{Mean} = \mathbb{E}(\mathbf{\Sigma} \mid \boldsymbol{Y}_D)$. (b) Under the loss L_2, the generalized Bayesian estimator of $\mathbf{\Phi}$ is $\widehat{\mathbf{\Phi}}_{Mean} = \mathbb{E}(\mathbf{\Phi} | \boldsymbol{Y}_D)$.*

The above fact shows that using posterior mean as the Bayesian estimator is equivalent to treating the weighting matrix $\mathbb{E}_{(\boldsymbol{Y}|\mathbf{\Phi},\mathbf{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})$ as constant and ignoring the role played by $\widehat{\mathbf{\Sigma}}$ in L_{E2}. The AR(1) example in the introduction shows that placing parameter-dependent weights on the estimation errors is more reasonable. In contrast to the ad hoc separable loss function, the entropy loss involves a more complicated but more plausible weighting scheme. In loss function L_{E2}, elements of estimation errors in $\mathbf{\Phi}$ are weighted according to their expected contributions to the covariance of forecast errors normalized by the inverse of the estimated covariance of the residuals. Therefore the entropy loss is a more natural metric for the fit of estimator in frequentist terms than the quadratic loss. Under the entropy loss, the Bayesian estimator of $\mathbf{\Phi}$ is different from the posterior mean. Since

$\boldsymbol{G} = \mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})$ and $\boldsymbol{\Phi}$ are likely to be positively correlated, the Bayesian estimator of $\boldsymbol{\Phi}$ under the intrinsic loss is likely to be larger than the posterior mean and helpful in correcting the likely downward bias in the posterior mean.

Although L_{E1} is the same as L_1, Theorem 1 shows that the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_E$ under the intrinsic loss is strictly larger than the posterior mean. This result can be explained by the form of the entropy loss. The Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_E$ minimizes the posterior risk by striking an optimal balance between the two parts of the loss, L_{E1} and L_{E2}. The posterior mean $\mathbb{E}(\boldsymbol{\Sigma}|\boldsymbol{Y}_D)$ minimizes L_{E1}-related posterior risk without taking into account the L_{E2}-related risk. The L_{E1}-related posterior risk of Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_E$ derived in Theorem 1 is larger than that of the posterior mean. The larger L_{E1}-related risk of the Bayesian estimator $\widehat{\boldsymbol{\Sigma}}_E$ is more than compensated for by a smaller L_{E2}-related risk.

3 Priors and Properties of the Posterior

3.1 Priors on $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$

An essential part of Bayesian analysis is the choice of prior. A popular informative prior of $\boldsymbol{\phi} = vec(\boldsymbol{\Phi})$ is the normal distribution,

$$\pi_N(\boldsymbol{\phi}) \quad \propto |\boldsymbol{M}_0|^{-1/2} \exp\left\{-\tfrac{1}{2}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\boldsymbol{M}_0^{-1}(\boldsymbol{\phi}-\boldsymbol{\phi}_0)\right\}. \tag{19}$$

A popular informative prior for $\boldsymbol{\Sigma}$ is an inverse Wishart distribution. The most commonly used 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). 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

$$\pi_J(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}. \tag{20}$$

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

$$\pi_A(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(L+1)p/2-1}. \tag{21}$$

Another candidate prior is Zellner's (1997) maximum data informative (MDI) prior,

$$\pi_M(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2}. \tag{22}$$

The joint prior densities of $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ under the normal-Jeffreys, normal-RATS and normal-MDI priors are given by

$$\begin{aligned}
\pi_{NJ}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) &= \pi_N(\boldsymbol{\phi})\pi_J(\boldsymbol{\Sigma}), \\
\pi_{NA}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) &= \pi_N(\boldsymbol{\phi})\pi_A(\boldsymbol{\Sigma}), \\
\pi_{NM}(\boldsymbol{\phi}, \boldsymbol{\Sigma}) &= \pi_N(\boldsymbol{\phi})\pi_M(\boldsymbol{\Sigma}).
\end{aligned}$$

For analysis on prior choice in VAR models, see Kadiyala and Karlsson (1997), Ni and Sun (2003), and Sun and Ni (2004). We consider a class of joint priors,

$$\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma}) = \pi_N(\boldsymbol{\phi})\pi_b^*(\boldsymbol{\Sigma}), \tag{23}$$

where $\pi_N(\boldsymbol{\phi})$ is the normal prior for $\boldsymbol{\phi}$ given by (19), and $\pi_b^*(\boldsymbol{\Sigma})$ $(b \in \mathbb{R})$ is given by

$$\pi_b^*(\boldsymbol{\Sigma}) \propto \frac{1}{|\boldsymbol{\Sigma}|^{b/2}}. \tag{24}$$

Note that π_{NJ}, π_{NA}, and π_{NM} are special cases of (24) when b equals to $p+1$, $(L+1)p+2$ and 1, respectively.

For the choice of prior on impulse responses to orthogonalized shocks, the Jeffreys prior on $\boldsymbol{\Sigma}$ is equivalent to the Jeffreys prior on $\boldsymbol{Z}_0$ (due to the invariance property of the Jeffreys prior), but the normal prior on $\boldsymbol{\phi}$ is not the same as the normal prior on $\boldsymbol{Z}_i$ $(i = 1, \cdots, L)$. The effect of priors of $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ on the posterior mean estimator of impulse responses cannot be easily assessed. The impact of prior choice of $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ on the Bayesian estimator of impulse responses with respect to the entropy loss can be assessed by examining that on the Bayesian estimator of $(\boldsymbol{\Sigma}, \boldsymbol{\Phi})$.

3.2 The Posteriors

The class of commonly used noninformative priors for $\boldsymbol{\Sigma}$ in (24) are improper. Bayesian estimates exist only when the posterior is proper (i.e., integrable) and relevant posterior moments exist under the prior (23). We now give a sufficient condition for the propriety of the posterior. The proof is given in Appendix A.

Theorem 2 *Consider the prior $\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma})$. The posterior of $(\boldsymbol{\phi}, \boldsymbol{\Sigma})$ is proper if $T > 2p - b$.*

3.3 Existence of Posterior Moments

The existence of posterior does not guarantee the existence of posterior moments. In the following, we give a sufficient condition for the existence of the posterior moments of certain functions that are related to the Bayesian estimator under the intrinsic loss. Its proof is given in Appendix B.

Theorem 3 *Consider the prior* $\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma})$*. If* $T > 2h+2p-b$*, the posterior mean of* $\|\boldsymbol{\phi}\|^k\{tr(\boldsymbol{\Sigma}^2)\}^{h/2}$ *is finite, where* h *is a nonnegative integer and* k *is an arbitrary nonnegative integer.*

Based on the theorems above, we can derive the following sufficient condition for the existence of the Bayesian estimators under the intrinsic loss. Its proof is provided in Appendix C.

Theorem 4 *Consider the prior* $\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma})$*. The generalized Bayesian estimator* $(\widehat{\boldsymbol{\Phi}}_E, \widehat{\boldsymbol{\Sigma}}_E)$ *of* $(\boldsymbol{\Phi}_E, \boldsymbol{\Sigma}_E)$ *in Theorem 1 exists if* $T > 2p - b + 2$*.*

The Bayesian estimate of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ can be calculated using the result of Theorem 1 so long as the posterior of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ is simulated. The following well-known result is useful for simulating the posterior of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$. Note that if the prior variance on $\boldsymbol{\phi}$ approaches infinity, the posterior precision of $\boldsymbol{\phi}$ (conditional on $\boldsymbol{\Sigma}$) is proportional to $\boldsymbol{X}'_D\boldsymbol{X}_D$, and the Bayesian estimator of $\boldsymbol{\Phi}$ in (12) is a generalized case of the AR parameter in the introduction.

Fact 2 *(a) Consider the normal prior for* $\boldsymbol{\phi}$ *given in (19). The conditional density of* $\boldsymbol{\phi}$ *given* $(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)$ *is* $N_J(\boldsymbol{\mu}_M, \boldsymbol{V}_M)$*, where*

$$\boldsymbol{\mu}_M = \widehat{\boldsymbol{\phi}}_{MLE} + (\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'_D\boldsymbol{X}_D))^{-1}\boldsymbol{M}_0^{-1}(\boldsymbol{\phi}_0 - \widehat{\boldsymbol{\phi}}_{MLE}); \tag{25}$$

$$\mathbf{V_M} = (\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}'_D\boldsymbol{X}_D))^{-1}. \tag{26}$$

(b) Consider the prior $\pi_b^*(\boldsymbol{\Sigma})$ *in (24). The conditional posterior of* $\boldsymbol{\Sigma}$ *given* $(\boldsymbol{\Phi}; \boldsymbol{Y}_D)$ *is Inverse Wishart* $(\boldsymbol{S}(\boldsymbol{\Phi}), T + b - p - 1)$*, where* $\boldsymbol{S}(\boldsymbol{\Phi})$ *is given by (5).*

There are two definitions of IW distribution (e.g., see Press (1982) and Anderson (1984)). We use the definition given by Anderson (1984).

4 Computing Bayesian Estimate under the Entropy Loss

In most applications of Bayesian VAR models, posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ based on commonly employed priors and likelihood functions are not standard distributions. In these situations, the posterior distributions can be simulated, typically using an MCMC method. Besides the papers cited in the introduction, applications of MCMC are shown to be fruitful for a variety of topics of Bayesian econometrics in the studies of Geweke (1989), Chib and Greenberg (1996), Sims and Zha (1999), and DeJong et al. (2000), among others. Suppose the posteriors are generated in a Markov Chain of M cycles. In the kth MCMC cycle ($k = 1, \cdots, M$), a set of parameter values $(\mathbf{\Phi}_k, \mathbf{\Sigma}_k)$ is simulated. The average of the M simulated parameters is used to approximate the posterior mean.

Theorem 1 shows that to compute the Bayesian estimators of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ we need to compute the quantities $\mathbb{E}\{\boldsymbol{G} \mid \boldsymbol{Y}_D\}$ and $\mathbb{E}\{\boldsymbol{G\Phi} \mid \boldsymbol{Y}_D\}$, where $\boldsymbol{G}$ is a frequentist expectation of sample data. We will discuss the issue in the following.

4.1 Algorithms for Posterior Simulation

To facilitate discussion of computing posterior quantities that involves frequentist moments, we consider a setting more general than VARs. Suppose that the density of a random vector or a matrix variable $\boldsymbol{X}$ for a given unknown parameter (vector) $\boldsymbol{\theta}$ is $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ and the prior for $\boldsymbol{\theta}$ is $\pi(\boldsymbol{\theta})$. Let $\boldsymbol{X}^*$ be a random vector or a matrix with density $f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta})$. The asterisk "*" reflects the fact that the density involves simulated data $\boldsymbol{X}^*$. In the application to VAR models, f^* and f are identical. Let $h(\boldsymbol{\theta})$ be a function of parameter $\boldsymbol{\theta}$. We are interested in the posterior mean of $[\mathbb{E}_{(\boldsymbol{X}^*|\boldsymbol{\theta})}\{g(\boldsymbol{X}^*)\}]h(\boldsymbol{\theta})$ given data $\boldsymbol{X}_D$.

We now show that different expressions of $\mathbb{E}(\mathbb{E}_{(\boldsymbol{X}^*|\boldsymbol{\theta})}\{g(\boldsymbol{X}^*)\}h(\boldsymbol{\theta}) \mid \boldsymbol{X}_D)$ lead to different ways of simulation. First, note that

$$\mathbb{E}(\mathbb{E}_{(\boldsymbol{X}^*|\boldsymbol{\theta})}\{g(\boldsymbol{X}^*)\}h(\boldsymbol{\theta}) \mid \boldsymbol{X}_D) \quad = \quad \int \Big\{ \int g(\boldsymbol{x}^*) f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}) d\boldsymbol{x}^* \Big\} h(\boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{X}_D) d\boldsymbol{\theta}. \qquad (27)$$

In the absence of exogenous variables in VAR, one can derive the formula for frequentist expectation $\boldsymbol{G}$ as a function of $(\mathbf{\Phi}, \mathbf{\Sigma})$. In this more general framework, it means that there is an analytical form of the expectation $\int g(\boldsymbol{x}^*) f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}) d\boldsymbol{x}^*$. The Bayesian estimator can be simulated using the following algorithm.

Algorithm A. In MCMC cycle k, draw $\boldsymbol{\theta}_k$ from $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_D)$ and compute $\int g(\boldsymbol{x}^*) f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}) d\boldsymbol{x}^*$ analytically.

This approach will be difficult to implement when exogenous variables are included in the VAR. Based on (27), we have an algorithm that simulates $\int g(\boldsymbol{x}^*) f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}) d\boldsymbol{x}^*$ via brute force.

Algorithm B. In MCMC cycle k, given parameter $\boldsymbol{\theta}_k$ drawn from $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_D)$ simulate N number of VARs and use the average $\frac{1}{N}\sum_{j=1}^{N} g(\boldsymbol{X}_j^*)$ to approximate $\int g(\boldsymbol{x}^*)f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}_k)d\boldsymbol{x}^*$.

While applicable to more general setting than Algorithm A, Algorithm B is computationally costly. Using Fubini's Theorem, we can change the orders of the terms in the righthand side of (27) and have

$$\mathbb{E}(\mathbb{E}_{(\boldsymbol{X}^*|\boldsymbol{\theta})}\{g(\boldsymbol{X}^*)\}h(\boldsymbol{\theta}) \mid \boldsymbol{X}_D) = \int\int\Big\{g(\boldsymbol{x}^*)h(\boldsymbol{\theta})\Big\}f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_D)d\boldsymbol{x}^*d\boldsymbol{\theta}. \qquad (28)$$

Based on (28), we can simulate using the following algorithm of data augmentation suggested by a referee.

Algorithm C. In MCMC cycle k $(k = 1, \cdots, M)$,

Step 1. Simulate $\boldsymbol{\theta}_k \sim \pi(\boldsymbol{\theta} \mid \boldsymbol{X}_D) \propto f(\boldsymbol{X}_D \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$.

Step 2. Simulate $\boldsymbol{X}_k^* \sim f^*(\boldsymbol{x}^* \mid \boldsymbol{\theta}_k)$.

With simulated random sample $(\boldsymbol{X}_k^*, \boldsymbol{\theta}_k)$, $(k = 1, \cdots, M)$ from the algorithm above, we can approximate the right hand side of (28) by $\frac{1}{M}\sum_{k=1}^{M} g\{\boldsymbol{X}_k^*\}h(\boldsymbol{\theta}_k)$.

The above approaches are equivalent asymptotically. Algorithms A and B calculate or simulate the frequentist expectation first in each MCMC cycle that simulates parameter $\boldsymbol{\theta}$. Algorithms C simulates $\pi(\boldsymbol{X}^*, \boldsymbol{\theta} \mid \boldsymbol{X}_D)$ based on the decomposition $\pi(\boldsymbol{X}^*, \boldsymbol{\theta} \mid \boldsymbol{X}_D) \propto f^*(\boldsymbol{X}^* \mid \boldsymbol{\theta})f(\boldsymbol{\theta} \mid \boldsymbol{X}_D)$. In practice, Algorithm C can often be applied with parameter vector $\boldsymbol{\theta}$ drawn in blocks via Gibbs sampling method illustrated by Gelfand and Smith (1990). Within each MCMC cycle, the posterior of $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ are simulated via Gibbs sampling. There is no need to alter the data-augmented algorithm if deterministic exogenous variables are included in the VAR. If stochastic exogenous variables are included, they simply become part of the simulated data $\boldsymbol{X}^*$. The simulated data are in essence latent variables. They are not the subject of our interests per se, but they are useful for simulation of parameter of interest (i.e., the frequentist expectation $\boldsymbol{G}$). Data augmentation is not uncommon in Bayesian computation, but this data-augmented simulation approach differs from its other uses in the econometrics and statistics literature. Our data-augmented algorithm uses generated regressors as latent parameters. In recent years, data augmentation has been used for various purposes in Bayesian literature. For example, in a seminal paper by Tanner and Wang (1987), data augmentation was used to alter the likelihood function for easier MCMC simulation of the posteriors. In the study by Otrok and Whiteman (1998), the generated latent economic indicator itself is of primary interest. In the paper by Elerian et al. (2001), generated data were used to estimate stochastic differential equations from discrete sample observations. Data augmentation in this study is different in motivation and implementation from that in the literature. Here, the likelihood function of the generated data is the same as that of the sample data, and data augmentation does not make posterior simulation easier. Instead, data augmentation makes it possible to compute frequentist moments used to produce Bayesian estimates under the entropy loss.

5 Applications

5.1 Estimating Responses of CPI to PPI Shocks

VAR models have been widely used in studying how shocks to the Producer Price Index (PPI) are transmitted to the Consumer Price Index (CPI). While some extent of such price pass-through is expected, a number of theories have been proposed on complications in the price transmission. Blomberg and Harris (1995) pointed out that the global commodity markets for crude materials are competitive where prices quickly respond to new information on future demand and supply. In contrast, the prices of final goods and consumer prices respond more slowly due to market rigidities. In addition, shifts in relative demand of commodities and final goods in response to commodity price inflation may weaken the link between the commodity price shock and CPI inflation. The theoretical ambiguity makes the transmission of commodity price to CPI an interesting empirical question.

The existing empirical studies show that VAR estimates on the responses of CPI to shocks in commodity price are generally unstable over sample periods. Blomberg and Harris (1995) and Furlong and Ingenito (1996) found that in the 1970s and early 1980s responses of CPI to shocks to commodity prices are positive and sustained, but for the mid 1980s and 1990s the CPI responses to PPI shocks are much weaker. Clark (1995) and Weinhagen (2002) showed that whether PPI inflation is useful in predicting CPI inflation depends on sample periods.

In these studies, the VAR estimates of impulse responses are obtained through standard software packages, which are based on the RATS prior and posterior mean estimates. In the following, we estimate a bi-variable VAR of monthly PPI of crude materials and CPI of the U.S. economy which was estimated by Furlong and Ingenito (1996). We choose the simplest model because our primary interest is not the model per se but the comparison of the posterior mean of impulse responses with the Bayesian estimate of impulse responses under the entropy loss.

We consider three separate sample periods: January 1970 to December 1979, January 1980 to

December 1989, and January 1990 to May 2005. The series are obtained from the FRED database at the Federal Reserve Bank of St Louis. Based on the Schwarz criterion, for each sample period the lag length of the VAR is two. Jeffreys prior is applied to the covariance matrix $\mathbf{\Sigma}$. The prior for the VAR coefficient $\boldsymbol{\phi}$ is a diffuse Minnesota prior. The prior variance $\boldsymbol{M}_0$ is a diagonal matrix with 20.0 being the diagonal element. For each sample we draw the posterior from $10,000$ MCMC cycles after $1,000$ burn-in runs. MCMC simulation took less than one minute on a 1.7 GHz Pentium 4 PC.

We first examine the posterior average losses. By definition, the intrinsic Bayesian estimator (with subscript E) is optimal under the entropy loss $L_E = L_{E1} + L_{E2}$ in (9). The posterior mean (with subscript $Mean$) is optimal under the sum of L_{E1} and quadratic loss L_2 in (18). Table 1 reports the posterior average losses of both the intrinsic Bayesian estimator and the posterior mean under the two competing loss functions, respectively. As expected, the Bayesian estimate dominates the posterior mean in terms of posterior expected loss under the entropy loss. The difference in the posterior expected loss is mainly due to the difference in the risks of the quadratic term L_{E2}, which is approximately $\frac{1}{2}tr[\widehat{\mathbf{\Sigma}}_E^{-1}\mathbb{E}((\mathbf{\Phi}-\widehat{\mathbf{\Phi}}_E)'\boldsymbol{G}(\mathbf{\Phi}-\widehat{\mathbf{\Phi}}_E)|\boldsymbol{Y}_D)\widehat{\mathbf{\Sigma}}_E^{-1}(\widehat{\mathbf{\Sigma}}_E-\widehat{\mathbf{\Sigma}}_{Mean})]$. The VAR variables are quite persistent in this application. As a result, with a larger $\widehat{\mathbf{\Sigma}}$ the Bayesian estimate substantially reduces the posterior risk, compared with the posterior mean. As we repeat the MCMC, the estimated posterior risks of the posterior mean estimator with respect to the intrinsic loss function vary over a wide range. As a consequence of the large value of $\mathbf{X_D'X_D}$ for this application, estimation errors of the posterior mean in certain dimensions are penalized severely. In contrast, due to the constant weight of quadratic loss in all dimensions of the parameters, the intrinsic Bayesian estimator yields similar posterior risk estimates in the repetitions of the MCMC simulations.

For the sample periods of 1970s and 1990-2005, the posterior average loss of the posterior mean estimate is larger than that of the intrinsic estimate. The lower overall posterior average loss of the Bayesian estimate over the posterior mean is achieved by substantially lowering the risk with respect to L_{E2}. The loss related to $\mathbf{\Sigma}$ under the Bayesian estimate is slightly larger compared to that under the posterior mean estimate. As is noted earlier, the intrinsic estimate improves L_{E2}-related loss with a tradeoff of larger L_{E1}-related loss. The empirical result shows that the intrinsic Bayesian estimate induces lower posterior average L_E loss than the posterior mean by making the L_{E2}-related loss substantially lower and the L_{E1}-related loss only slightly higher. Under the loss function $L_{E1} + L_2$, the posterior mean estimator is optimal. Table 1 shows that in each sub-sample period the intrinsic Bayesian estimator is slightly worse than the posterior mean estimator.

We now turn to the main subject of interest, the estimates of impulse responses. Figure 1 plots the point estimates and the 95% posterior probability bands of responses of CPI (measured in percentage per annum) to a unit shock in PPI of crude materials. We also plug in MLE of $(\mathbf{\Phi}, \mathbf{\Sigma})$ into the impulse responses, they turn out to be similar to the posterior mean of impulse responses. We compare both of these quantities with the intrinsic Bayesian estimates of impulse responses.

The dynamics of the CPI responses in the pre- and post-1980 samples are distinctly different. For the pre-1980 sample, CPI picks up pace after a unit shock of crude material PPI and keeps going up. For the post-1980 sample, the impulse response of CPI levels off after the shock to PPI of crude materials. The change in the dynamics of impulse responses suggests a structural difference in the pre- and post-1980 samples. For each sub-sample period the entropy-loss-based Bayesian estimates of impulse responses (the solid lines) are larger in magnitude than the posterior mean estimates (the dotted lines). This is not surprising, given the result in Section 2 that the Bayesian estimate $\widehat{\mathbf{\Sigma}}_E$ is larger than the posterior mean of $\mathbf{\Sigma}$ and that $\widehat{\mathbf{\Phi}}_E$ is likely to be larger than the posterior mean of $\mathbf{\Phi}$. The difference in the competing estimates of VAR parameters corresponds to a visible difference in point estimates of impulse responses (although the posterior mean of the impulse responses is unequal to the impulse response functions of the posterior mean.) For longer forecasting horizon of the impulse response, the difference between the posterior mean and the intrinsic estimator is more profound. With a relatively small sample size and diffuse priors, the posterior coverage of the data-generating-parameter can be poor; hence being on the tail of the posterior does not imply that the intrinsic estimator is further from the data generating parameter than the posterior mean is.

Note that under the intrinsic loss, a unit shock in PPI of crude materials represent 2.8%, 1.3%, 4.1% for 1970s, 1980s and post-1990s periods, respectively. However, for each unit shock in PPI of crdue materials, the difference in the estimates of impulse responses under the intrinsic estimator for the pre- and post-1980 periods is much smaller than the difference in the posterior mean. In other words, the intrinsic Bayesian estimates suggest that the force of pass-through from PPI to CPI is much persistent than what indicated by the posterior mean estimator.

5.2 Application 2: A VAR of U.S. Macroeconomic Variables

Bayesian VAR models have been commonly used for analyzing multivariate time series macroeconomic data and addressing policy questions. However, not much attention was given to sensitivity of results to researchers' choice of estimator. In the following, we compare various Bayesian estimates of a

VAR using quarterly data of the U.S. economy. The variables are real GDP growth, government expenditure growth, growth of real consumer expenditure on nondurables, CPI inflation, growth of M2 money stock, and the Federal Funds rates (FFR). The data series are obtained from the FRED database at the Federal Reserve Bank of St Louis. These variables have appeared in macroeconomics related VARs (e.g. Sims 1992, Gordon and Leeper 1994, Sims and Zha 1998, and Christiano et al. 1999). The sample period is from 1982Q4 to 2005Q1, during this period the Federal Reserve Bank has been using the Federal Funds rates as the main target of monetary policy. The question of interest to us is how the Federal Funds Rates respond to shocks in macro-variables and how macro-variables respond to shocks in the Federal Funds rates. Based on the BIC, the lag length of the VAR is set at two.

Table 2 reports the posterior average losses. As is expected, the intrinsic Bayesian estimator produces smaller posterior entropy loss than the posterior mean. The latter does better under the loss $L_{E1} + L_2$. The impulse responses in Figure 2 show that FFR rises persistently after a shock in GDP growth, more so based on the intrinsic Bayesian estimator than the posterior mean. There is a hump-shaped response of FFR to a shock in CPI inflation. The FFR response to consumption growth shock is close to zero and not plotted. The slightly positive response of FFR to M2 money stock growth suggests the absence of liquidity effect under the posterior mean. However, the intrinsic Bayesian estimates of FFR response to M2 shocks are larger than the posterior means and exhibit liquidity effect. The responses of macro-variables to shocks in FFR suggest that the latter are not a good indicator of monetary contractions. Both GDP and consumption growth respond positively. Inflation response to FFR shocks is small and not plotted. The impulse responses are consistent with the interpretation that shocks in FFR reflect Federal Reserve's expectation rather than exogenous monetary contraction, an interpretation that agrees with the argument of Gordon and Leeper (1994). The intrinsic estimates offer stronger support to this view than the posterior mean estimates do. The results suggest more sophisticated characterization of monetary policy is needed. The objective here is to show intrinsic Bayesian estimator lead to different estimates than the posterior mean. We will leave the task of developing models for monetary policy analysis to future research.

6 Concluding Remarks

In this paper, we investigate properties of the Bayesian estimates of impulse responses through an information-theoretic approach. We derive Bayesian estimators from an intrinsic entropy loss function and show that they are distinctly different from the posterior mean. We propose an algorithm that uses generated data as latent variables in numerical simulation of Bayesian estimates under the entropy loss. Our data-augmented simulation scheme may be useful for Bayesian analysis in other problems. For example, in the Bayesian VAR literature, the priors on $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are often elicited separately. In this paper, we examine the Bayesian estimate under the joint non-separable loss on $(\mathbf{\Phi}, \mathbf{\Sigma})$ under separate priors. For users of Bayesian VAR, it is of interest to experiment with joint non-informative priors for $(\mathbf{\Phi}, \mathbf{\Sigma})$ under the general principles outlined in Kass and Wasserman (1996). The joint non-informative priors generally involve frequentist moments of VAR variables, where computational issues similar to our study in the paper will arise.

The paper can be extended in a number of ways. First, in this paper, we only consider estimators of impulse responses under the intrinsic loss. In some VAR applications, researchers may be interested in estimating nonlinear quantities other than impulse responses, such as variance decomposition. As noted in the introduction, under the intrinsic loss any nonlinear quantities can be estimated using the invariance property of the Bayesian estimator. We will not pursue this subject in details due to constraint in space. Second, the present study is limited to estimation of unrestricted VARs. Some recent examples of Bayesian analysis with restrictions on parameters in simultaneous equation models and VARs with cointegration include Zellner (1998), Gao and Lahiri (2002), Kleibergen and van Dijk (1998), and Kleibergen and Paap (2002). In identified VARs, restrictions are placed on the contemporaneous relationship of VAR variables (e.g., Sims and Zha, 1998, 1999). Furthermore, macroeconomic theories such as DSGE (Dynamic Stochastic General Equilibrium) models are used for prior elicitation of Bayesian VARs (e.g., Ingram and Whiteman, 1994; DeJong, Ingram, and Whiteman, 2000, and Del Negro and Schorfheide, 2005). Suppose a DSGE model has fundamental parameter $\boldsymbol{\theta}$ (on preference, technology, government policies, etc) and its linearized approximation is a VAR model

$$\mathbf{y_t'} = \mathbf{y_{t-1}'}\mathbf{B} + \boldsymbol{\epsilon}_\mathbf{t}', \tag{29}$$

where $\boldsymbol{\epsilon}_t \sim N(0, \mathbf{\Sigma})$. Denote the VAR parameters $\mathbf{\Upsilon} = (\boldsymbol{B}, \mathbf{\Sigma})$. The economic theory corresponds to nonlinear restrictions on $\mathbf{\Upsilon}$ through restrictions of $\boldsymbol{\theta}$. These theoretical restrictions can be reflected in the prior on Υ. In general, there may be unobservable variables in the DSGE model. In such a case, the DSGE model may be approximated by a state space model where Kalman's filter can be used for evaluation of the likelihood. A common practise is to estimate the parameter $\boldsymbol{\theta}$ based on a distance between moments of generated data and observed data, for instance, IRF of restricted and unrestricted VARs. This practice does not take into account the uncertainty in the quantities used for estimation, such as the uncertainty in IRF of either restricted or unrestricted VARs. In a Bayesian analysis, such uncertainties will be fully taken into consideration. Given the economic theory-based-prior and observed data, we can produce Bayesian estimate of $\boldsymbol{\Upsilon}$ under a loss function. The loss function may be with respect to $\boldsymbol{\Upsilon}$ or IRF of VAR in (29). These estimates imply different estimates for $\boldsymbol{\theta}$. A virtue of the entropy loss is the estimating any nonlinear functions of $\boldsymbol{\Upsilon}$, including the IRF of (29), is equivalent to estimating $\boldsymbol{\theta}$. Deriving the functional form of the estimator will be difficult because the likelihood is no longer a standard distribution. Numerical solutions are needed for minimization of the loss function. Bayesian estimation based on intrinsic losses in these settings is challenging and will be left for future research.

Acknowledgments

We thank the editor, an associate editor, and two referees for insightful comments and suggestions that substantially improved the paper. Ni's research was supported by grants from the MU Research Board, MU Research Council, and Hong Kong Research Grant Council grant CUHK4128/03H. D. Sun's research was supported by the National Science Foundation grants DMS-9972598 and SES-0351523, a grant from the MU Research Board, and a grant from the Missouri Department of Conservation. The authors benefited from discussions with Christopher Sims and Edward George.

References

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

Bernardo, J. M. and Jurez, M. A. 2003, Intrinsic estimation, in *Bayesian Statistics* 7 (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, Eds.). Oxford: Oxford University Press, 465-476.

Blomberg B. S. and E. S. Harris, 1995, The commodity-consumer price connection: Fact or fable? Federal Reserve Bank of San Francisco Economic Policy Review, October, 21-38.

Chib, S. and E. Greenberg, 1996, Markov Chain Monte Carlo simulation methods in econometrics, Econometric Theory, 12, 327-335.

Christiano, L.J., M. Eichenbaum, and C. Evans, 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, 65–147.

Clark T. E., 1995, Do producer prices lead consumer prices? Federal Reserve Bank of Kansas City

Economic Review, third quarter, 25-39.

DeJong D.N., B. F. Ingram, and C. H. Whiteman, 2000, A Bayesian approach to dynamic macroeconomics, Journal of Econometrics 98, 203-223.

DelNegro M. and F. Schorfheide, 2005, Priors from general equilibrium models for VARs, International Economic Review, forthcoming.

Elerian O., S. Chib, and N. Shephard, 2001, Likelihood inference for discretely observed nonlinear diffusions, Econometrica 69, 959-993.

Fernandez-Villaverde J. and J. F. Rubio-Ramirez, 2004, Comparing dynamic equilibrium economies to data, Journal of Econometrics 123, 153-187.

Furlong, F. and R. Ingenito, 1996, Commodity prices and inflation, Federal Reserve Bank of San Francisco Economic Review, No. 2, 27-47.

Gao, C. and K. Lahiri, 2002, A comparison of some recent Bayesian and classical procedures for simultaneous equation models with weak instruments, working paper, Department of Economics, State University of New York at Albany.

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

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

Geweke J., 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317-1339.

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

Ingram, B. F. and Whiteman, C. H. 1994, Supplementing the Minnesota prior-forecasting macroeconomic time series using real business cycle model priors, Journal of Monetary Economics 34, 497–510.

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

Kass, R.E. and Wasserman, L. 1996, The selection of prior distributions by formal rules. Journal of American Statistical Association 91, 1343-1370.

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

Kilian, L. (1999). Finite sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses, Review of Economics and Statistics, 81, 652-660.

Kitamura Y. and M. Stutzer, 1997, An information-theoretic alternative to generalized method of moments estimation, Econometrica, 65, 861-874.

Kleibergen, F. and H. K. van Dijk, 1998, Bayesian simultaneous equations analysis using reduced rank structures, Econometric Theory 14, 701-743.

Kleibergen, F. and R. Paap, 2002, Priors, posteriors, and Bayes factors for a Bayesian analysis of cointegration, Journal of Econometrics 111, 223-249.

Koop, G., Pesaran, H., and S. Potter, 1996, Impulse response analysis in nonlinear multivariate models, Journal of Econometrics 74, 119-147.

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

Muirhead, R. 1982, *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York.

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

Otrok C. and C. H. Whiteman, 1998, Bayesian leading indicators: measuring and predicating economic conditions in Iowa. International Economic Review 39, 997-1014.

Pesaran, H. and Y. Shin, 1998, Generalized impulse response analysis in linear multivariate models, Economics Letters 58, 17-29.

Press, J.S., 1982, *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference (2nd edition).* Krieger, Florida.

Robert C.P., 1994, *The Bayesian Choice.* Springer-Verlag, New York.

Robert C.P., 1996, Intrinsic losses. Theory and Decision 40, 191-204.

Robertson J. C., E. W. Tallman, and C. H. Whiteman 2005, Forecasting using relative entropy, Journal of Money, Credit, and Banking 37 383-402.

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

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 T. Zha, 1998, Does monetary policy generate recessions? Federal Reserve Bank of Atlanta working paper 98-12.

Sims, C. A. and T. Zha, 1999, Error bands for impulse responses, Econometrica 67, 1113–1155.

Sun, D. and S. Ni, 2004, Bayesian analysis of vector-autoregressive models with noninformative priors, Journal of Statistical Planning and Inference 121, 291-309.

Tanner M. and Wang W.H., 1987, The calculation of posterior distributions by data augmentation, Journal of American Statistical Association 82, 528-540.

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

Weinhagen, J. 2002, An empirical analysis of price transmission by stage of process, Monthly Labor Review, November, 3-11.

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

Zellner, A., 1978, Estimation of functions of population means and regression coefficients including structural coefficients: a minimum expected loss approach, Journal of Econometrics 8, 127-158.

Zellner, A., 1997, Maximal data information prior distributions, Chapter 8 of Bayesian Analysis in Econometrics and Statistics (Edward Elgar, Lyme, UK).

Zellner, A., 1998, The finite sample properties of simultaneous equations estimates and estimators: Bayesian and non-Bayesian approaches, Journal of Econometrics 83, 185-212.

Appendix A. Proof of Theorem 2

In the following, we let $C, C_1, C_2, \ldots$ be constants depending only on sample size T and observation $\boldsymbol{Y}_D$. By Fact 2, with $J = (LP+1)p$,

$$\int_{\mathbb{R}^J} L(\boldsymbol{\phi}, \boldsymbol{\Sigma})\pi_N(\boldsymbol{\phi})\, d\boldsymbol{\phi} \quad = \quad C|\boldsymbol{\Sigma}|^{T/2}|\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}_D'\boldsymbol{X}_D)|^{-1/2}\, \mathrm{etr}\Big\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE})\Big\},$$

where $\widehat{\boldsymbol{\Phi}}_{MLE}$ and $\boldsymbol{S}(\cdot)$ are defined by (4) and (5), respectively. Since $|\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}_D'\boldsymbol{X}_D)| \geq |\boldsymbol{M}_0|^{-1}$, we have

$$\int\int_{\mathbb{R}^J} L(\boldsymbol{\phi}, \boldsymbol{\Sigma})\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma})\, d\boldsymbol{\phi}\, d\boldsymbol{\Sigma} \quad \leq \quad C|\boldsymbol{M}_0|^{1/2}\int \frac{\mathrm{etr}\Big\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE})\Big\}}{|\boldsymbol{\Sigma}|^{(T+b)/2}}\, d\boldsymbol{\Sigma}.$$

If $T > 2p - b$, the right hand side is equal to

$$C|\boldsymbol{M}_0|^{1/2}\frac{\Gamma_p((T+b-p-1)/2)}{2^{-(T+b-p-1)/2}|\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE})|^{(T+b-p-1)/2}} < \infty,$$

where $\Gamma_p(a)$ is a multivariate gamma function.

Appendix B. Proof of Theorem 3

We first prove a lemma.

Lemma 2 *Suppose that the conditional posterior of $\boldsymbol{\phi}$ given $(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)$ is $N_J(\boldsymbol{\mu}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D), \boldsymbol{V}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D))$. Suppose that for any $\boldsymbol{\Sigma}$, $\|\boldsymbol{\mu}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)\| \leq a_0$ and $\boldsymbol{V}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D) < \boldsymbol{A}_0$, where a_0 is a positive constant and $\boldsymbol{A}_0$ is a positive definite $J \times J$ matrix, both are dependent on data but independent of $\boldsymbol{\Sigma}$. Then for any fixed integer $k \geq 0$, there exists a constant $H(k) > 0$, which is independent of $\boldsymbol{\Sigma}$, such that posterior moment $\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^k) \leq H(k)$.*

Proof. We just need to consider the case $k > 0$ because $\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^k) = 1$ for $k = 0$. Let $\boldsymbol{\mu}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D) = (\mu_1(\boldsymbol{\Sigma}; \boldsymbol{Y}_D), \cdots, \mu_J(\boldsymbol{\Sigma}; \boldsymbol{Y}_D))'$ and $\boldsymbol{V}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D) = (\sigma_{ij}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D))_{J\times J}$. Then by $\|\boldsymbol{\mu}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)\| \leq a_0$ and $\boldsymbol{V}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D) < \boldsymbol{A}_0$, we can easily show that there exists a constant C, such that

$$|\mu_i(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)| \leq C, \ \ 1 \leq i \leq J; \ \ \text{and} \ \ |\sigma_{ij}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)| \leq C, \ \ 1 \leq i, j \leq J.$$

Thus, for even k, there exists $H(k)$ such that $\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^k) \leq H(k)$ because $\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^k) = \mathbb{E}_{\boldsymbol{\Sigma}}(\phi_1^2 + \cdots + \phi_J^2)^{k/2}$ is a polynomial of $(\mu_j(\boldsymbol{\Sigma}; \boldsymbol{Y}_D), j = 1, \cdots, J; \sigma_{ij}(\boldsymbol{\Sigma}; \boldsymbol{Y}_D), 1 \leq i \leq j \leq J)$. For odd k, Jensen's Inequality ensures that $\{\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^k)\}^{1/k} \leq \{\mathbb{E}_{\boldsymbol{\Sigma}}(\|\boldsymbol{\phi}\|^{k+1})\}^{1/(k+1)}$ and thus the result follows. $\square$

Proof of Theorem 3. Because the posterior is proper from the assumptions, it is enough to show that

$$\int\int_{\mathbb{R}^J} \|\boldsymbol{\phi}\|^k \{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^{h/2} L(\boldsymbol{\phi},\boldsymbol{\Sigma})\pi_b(\boldsymbol{\phi},\boldsymbol{\Sigma})d\boldsymbol{\phi}\, d\boldsymbol{\Sigma} < \infty. \tag{30}$$

By Fact 2, $\boldsymbol{\phi} \sim N(\boldsymbol{\mu}_M, \boldsymbol{V}_M)$, where $\boldsymbol{\mu}_M = \boldsymbol{\mu}_M(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)$ and $\boldsymbol{V}_M = \boldsymbol{V}_M(\boldsymbol{\Sigma}; \boldsymbol{Y}_D)$ are given by (25) and (26), respectively. It follows that $\|\boldsymbol{\mu}_M(\boldsymbol{\Sigma})\|^k < \|\widehat{\boldsymbol{\phi}}_{MLE}\|^k + \|\boldsymbol{\phi}_0\|^k$. It is also obvious that $\boldsymbol{V}_M < \boldsymbol{M}_0$. Note that $\boldsymbol{\phi}_0$ and $\boldsymbol{M}_0$ are constants, $\widehat{\boldsymbol{\phi}}_{MLE}$ depends on data but is independent of $\boldsymbol{\Sigma}$. It follows from Lemma 2 that for any given integer $k \geq 0$, there is a positive constant C_1 such that

$$\int_{\mathbb{R}^J} \|\boldsymbol{\phi}\|^k \pi(\boldsymbol{\phi} \mid \boldsymbol{\Sigma}, \boldsymbol{Y}_D) d\boldsymbol{\phi} \ < C_1.$$

The marginal posterior of $\boldsymbol{\Sigma}$ given $\boldsymbol{Y}_D$ has the form

$$\begin{aligned} m(\boldsymbol{\Sigma} \mid \boldsymbol{Y}_D) &= C_2 \pi_b(\boldsymbol{\Sigma}) \int L(\boldsymbol{\phi},\boldsymbol{\Sigma})\pi(\boldsymbol{\phi})\, d\boldsymbol{\phi} \\ &= C_3 \frac{|\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}_D'\boldsymbol{X}_D)|^{-1/2}}{|\boldsymbol{\Sigma}|^{(T+b)/2}} \,\mathrm{etr}\Big\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE})\Big\}. \end{aligned}$$

Therefore, the left hand side of (30) is bounded by

$$C_4 \int \frac{\{\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^{h/2}}{|\boldsymbol{\Sigma}|^{(T+b)/2}} \,\mathrm{etr}\Big\{-\frac{\eta_p}{2}\boldsymbol{\Sigma}^{-1}\Big\}\, d\boldsymbol{\Sigma} \tag{31}$$

because $|\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{X}_D'\boldsymbol{X}_D)| \geq |\boldsymbol{M}_0|^{-1}$ and $\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE}) > \eta_p \boldsymbol{I}_p$, where $\eta_1 > \eta_2 \cdots > \eta_p > 0$ are the eigenvalues of $\boldsymbol{S}(\widehat{\boldsymbol{\Phi}}_{MLE})$. Making the orthogonal decomposition $\boldsymbol{\Sigma} = \boldsymbol{O}'\boldsymbol{\Lambda}\boldsymbol{O}$, where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$, and using Theorem 3.2.17 in Muirhead (1982), (31) becomes

$$C_4 \frac{\pi^{p^2/2}}{\Gamma_p(p/2)} \int \prod_{1 \leq i < j \leq p} (\lambda_i - \lambda_j) \frac{(\sum_{i=1}^p \lambda_i^2)^{h/2}}{\prod_{i=1}^p \lambda_i^{(T+b)/2}} \exp\Big\{-\frac{\eta_p}{2\lambda_i}\Big\}\, d\boldsymbol{\lambda} \tag{32}$$

because the integrated function in (31) is just a function of $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)$. Note that $\{\sum_{i=1}^p \lambda_i^2\}^{h/2} \leq (p\lambda_1)^h$ and $\prod_{1 \leq i < j \leq p}(\lambda_i - \lambda_j) \leq \prod_{i=1}^p \lambda_i^{p-i}$. So (32) is bounded by

$$C_5 \int \frac{1}{\lambda_1^{(T+b-2h-2p+2)/2}} \exp\Big(-\frac{\eta_p}{2\lambda_1}\Big) d\lambda_1 \prod_{i=2}^p \int \frac{1}{\lambda_i^{(T+b-2p+2i)/2}} \exp\Big(-\frac{\eta_p}{2\lambda_i}\Big) d\lambda_i. \tag{33}$$

Note that $\int_0^\infty x^{-(\alpha+1)} e^{-\beta/x} dx$ is finite if and only if $\alpha > 0$ and $\beta > 0$. Thus (33) is finite if $T > 2h + 2p - b$. The theorem then follows.

Appendix C. Proof of Theorem 4

By (12) and (13), $(\hat{\boldsymbol{\Phi}}_E, \hat{\boldsymbol{\Sigma}}_E)$ exists if and only if all of $\mathbb{E}(\boldsymbol{G} \mid \boldsymbol{Y}_D)$, $\mathbb{E}(\boldsymbol{G\Phi} \mid \boldsymbol{Y}_D)$, $\mathbb{E}(\boldsymbol{\Phi}'\boldsymbol{G\Phi} \mid \boldsymbol{Y}_D)$ and $\mathbb{E}(\boldsymbol{\Sigma} \mid \boldsymbol{Y}_D)$ exist. By Theorem 3, it is easy to see that if $T > 2p - b + 2$, $\mathbb{E}(\boldsymbol{\Sigma} \mid \boldsymbol{Y}_D) < \infty$ because $|\sigma_{st}| \leq \{tr(\boldsymbol{\Sigma}^2)\}^{1/2}$, $1 \leq s, t \leq p$. In order to prove that $\mathbb{E}(\boldsymbol{G} \mid \boldsymbol{Y}_D)$, $\mathbb{E}(\boldsymbol{G\Phi} \mid \boldsymbol{Y}_D)$ and $\mathbb{E}(\boldsymbol{\Phi}'\boldsymbol{G\Phi} \mid \boldsymbol{Y}_D)$ exist, we first consider the structure of $\boldsymbol{G} = \mathbb{E}_{(\boldsymbol{Y}|\boldsymbol{\Phi},\boldsymbol{\Sigma})}(\boldsymbol{X}'\boldsymbol{X})$. To simplify notation, the asterisk "*" that signifies simulated data is omitted here.

By (1), we have

$$\boldsymbol{y}_i' = \boldsymbol{g}_i(\boldsymbol{\phi}) + \sum_{j=1}^{i} \boldsymbol{\epsilon}_j' \boldsymbol{R}_{ij}(\boldsymbol{\phi}), \;\; i = 1, \cdots, T,$$

where $\boldsymbol{g}_i(\boldsymbol{\phi})$ is a p-dimensional row vector and each element of $\boldsymbol{g}_i(\boldsymbol{\phi})$ is a polynomial of the elements of $\boldsymbol{\phi}$ with degree less than $i+1$, $\boldsymbol{R}_{ij}(\boldsymbol{\phi})$ is a $p \times p$ matrix and each element of $\boldsymbol{R}_{ij}(\boldsymbol{\phi})$ is a polynomial of the elements of $\boldsymbol{\phi}$ with degree less than i. In addition, since

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{X} &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{y}_0 & \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_{T-1} \\ \cdot & \cdot & \cdots & \cdot \\ \boldsymbol{y}_{1-L} & \boldsymbol{y}_{2-L} & \cdots & \boldsymbol{y}_{T-L} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{y}_0' & \boldsymbol{y}_{-1}' & \cdots & \boldsymbol{y}_{1-L}' \\ 1 & \boldsymbol{y}_1' & \boldsymbol{y}_0' & \cdots & \boldsymbol{y}_{2-L}' \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \boldsymbol{y}_{T-1}' & \boldsymbol{y}_{T-2}' & \cdots & \boldsymbol{y}_{T-L}' \end{pmatrix} \\
&= \begin{pmatrix} T & \sum_{i=1}^{T-1} \boldsymbol{y}_i' & \sum_{i=1}^{T-1} \boldsymbol{y}_{i-1}' & \cdots & \sum_{i=1}^{T-1} \boldsymbol{y}_{i-L+1}' \\ \sum_{i=1}^{T-1} \boldsymbol{y}_i & \sum_{i=1}^{T-1} \boldsymbol{y}_i\boldsymbol{y}_i' & \sum_{i=1}^{T-1} \boldsymbol{y}_i\boldsymbol{y}_{i-1}' & \cdots & \sum_{i=1}^{T-1} \boldsymbol{y}_i\boldsymbol{y}_{i+1-L}' \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \sum_{i=1}^{T-1} \boldsymbol{y}_{i-L+1} & \sum_{i=1}^{T-1} \boldsymbol{y}_{i+1-L}\boldsymbol{y}_i' & \sum_{i=1}^{T-1} \boldsymbol{y}_{i+1-L}\boldsymbol{y}_{i-1}' & \cdots & \sum_{i=1}^{T-1} \boldsymbol{y}_{i+1-L}\boldsymbol{y}_{i+1-L}' \end{pmatrix},
\end{aligned}$$

we can easily show that $\mathbb{E}(\boldsymbol{G} \mid \boldsymbol{Y}_D)$ exists if

$$\mathbb{E}\left(\prod_{i=1}^{p(Lp+1)} \phi_i^{k_i} \mid \boldsymbol{Y}_D\right) < \infty, \quad k_i \geq 0, \sum_{i=1}^{p(Lp+1)} k_i \leq 2(T-1)$$

and

$$\mathbb{E}\left(\sigma_{st} \prod_{i=1}^{p(Lp+1)} \phi_i^{k_i} \mid \boldsymbol{Y}_D\right) < \infty, \quad 1 \leq s, t \leq p; \quad k_i \geq 0, \sum_{i=1}^{p(Lp+1)} k_i \leq 2(T-2).$$

Because $|\prod_{i=1}^{p(Lp+1)} \phi_i^{k_i}| \leq \|\boldsymbol{\phi}\|^{\sum_{i=1}^{p(Lp+1)} k_i}$, we know that $\mathbb{E}(\boldsymbol{G} \mid \boldsymbol{Y}_D)$ exists if

$$\mathbb{E}\left(\|\boldsymbol{\phi}\|^t \mid \boldsymbol{Y}_D\right) < \infty, \quad 0 \leq t \leq 2(T-1) \tag{34}$$

and

$$\mathbb{E}\left[\{tr(\boldsymbol{\Sigma}^2)\}^{1/2}\|\boldsymbol{\phi}\|^t \mid \boldsymbol{Y}_D\right] < \infty, \quad 0 \le t \le 2(T-2). \tag{35}$$

Theorem 3 ensures that (34) and (35) hold under the condition $T > 2p - b + 2$. Similarly, we can show that

$$\mathbb{E}(\boldsymbol{G}\boldsymbol{\Phi} \mid \boldsymbol{Y}_D) < \infty \quad \text{and} \quad \mathbb{E}(\boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi} \mid \boldsymbol{Y}_D) < \infty \tag{36}$$

if $T > 2p - b + 2$ and thus the result follows.

Table 1: Posterior Average Loss of three sample periods in Application 1

1970.1-1979.12	L_{E1}	L_{E2}	$L_E = L_{E1} + L_{E2}$	L_2	$L_{E1} + L_2$
$\widehat{\mathbf{\Sigma}}_{Mean}, \widehat{\mathbf{\Phi}}_{Mean}$	1.597	570.594	572.191	203.800	205.397
	(0.018)	(5246.690)	(5246.691)	(3.928)	(3.929)
$\widehat{\mathbf{\Sigma}}_E, \widehat{\mathbf{\Phi}}_E$	2.576	13.201	15.777	265.353	267.929
	(0.999)	(3.975)	(4.915)	(135.621)	(136.389)
1980.1-1989.12	L_{E1}	L_{E2}	$L_E = L_{E1} + L_{E2}$	L_2	$L_{E1} + L_2$
$\widehat{\mathbf{\Sigma}}_{Mean}, \widehat{\mathbf{\Phi}}_{Mean}$	1.597	3680.203	3681.800	32.124	33.722
	(0.018)	(53806.513)	(53806.514)	(0.590)	(0.592)
$\widehat{\mathbf{\Sigma}}_E, \widehat{\mathbf{\Phi}}_E$	2.409	12.123	14.532	42.472	44.881
	(0.816)	(3.951)	(4.723)	(18.151)	(18.785)
1990.1-2005.5	L_{E1}	L_{E2}	$L_E = L_{E1} + L_{E2}$	L_2	$L_{E1} + L_2$
$\widehat{\mathbf{\Sigma}}_{Mean}, \widehat{\mathbf{\Phi}}_{Mean}$	1.564	173936.005	173937.569	747.090	748.654
	(0.073)	(5177054.074)	(5177054.145)	(13.323)	(13.330)
$\widehat{\mathbf{\Sigma}}_E, \widehat{\mathbf{\Phi}}_E$	3.427	20.397	23.824	884.158	887.585
	(2.584)	(8.395)	(10.843)	(443.842)	(445.246)

Note: The subscript '$Mean$' represents posterior mean estimator and 'E' represents Bayesian estimator under the entropy loss. The losses L_{E1} and L_{E2} are given in (9). The entropy loss is $L_E = L_{E1} + L_{E2}$. The quadratic loss L_2 is given in (18). The intrinsic Bayesian estimator is optimal under the entropy loss, the posterior mean estimator is optimal under the loss $L_{E1} + L_2$. Posterior average loss is computed for each MCMC simulation (each with $10,000$ cycles after $1,000$ burn-in runs). We conduct $1,000$ replications of the simulations and report the mean of the posterior average loss (with standard deviations in the brackets) of these replications.

Table 2: Posterior Average Loss in Application 2: 1983.Q1-2005.Q1

	L_{E1}	L_{E2}	L_E	L_2	$L_{E1}+L_2$
$\widehat{\mathbf{\Sigma}}_{Mean}, \widehat{\mathbf{\Phi}}_{Mean}$	13.214	18254.000	18267.214	87.322	100.536
	(0.072)	(240313.650)	(240313.654)	(0.552)	(0.577)
$\widehat{\mathbf{\Sigma}}_E, \widehat{\mathbf{\Phi}}_E$	25.431	69.229	94.660	119.859	145.290
	(4.299)	(7.945)	(12.169)	(26.018)	(30.113)

Note: The subscript '$Mean$' represents posterior mean estimator and 'E' represents Bayesian estimator under the entropy loss. The losses L_{E1} and L_{E2} are given in (9). The entropy loss is $L_E = L_{E1} + L_{E2}$. The quadratic loss L_2 is given in (18). The intrinsic Bayesian estimator is optimal under the entropy loss, the posterior mean estimator is optimal under the loss $L_{E1} + L_2$. Posterior average loss is computed for each MCMC simulation (each with $10,000$ cycles after $1,000$ burn-in runs). We conduct $1,000$ replications of the simulations and report the mean of the posterior average loss (with standard deviations in the brackets) of these replications.

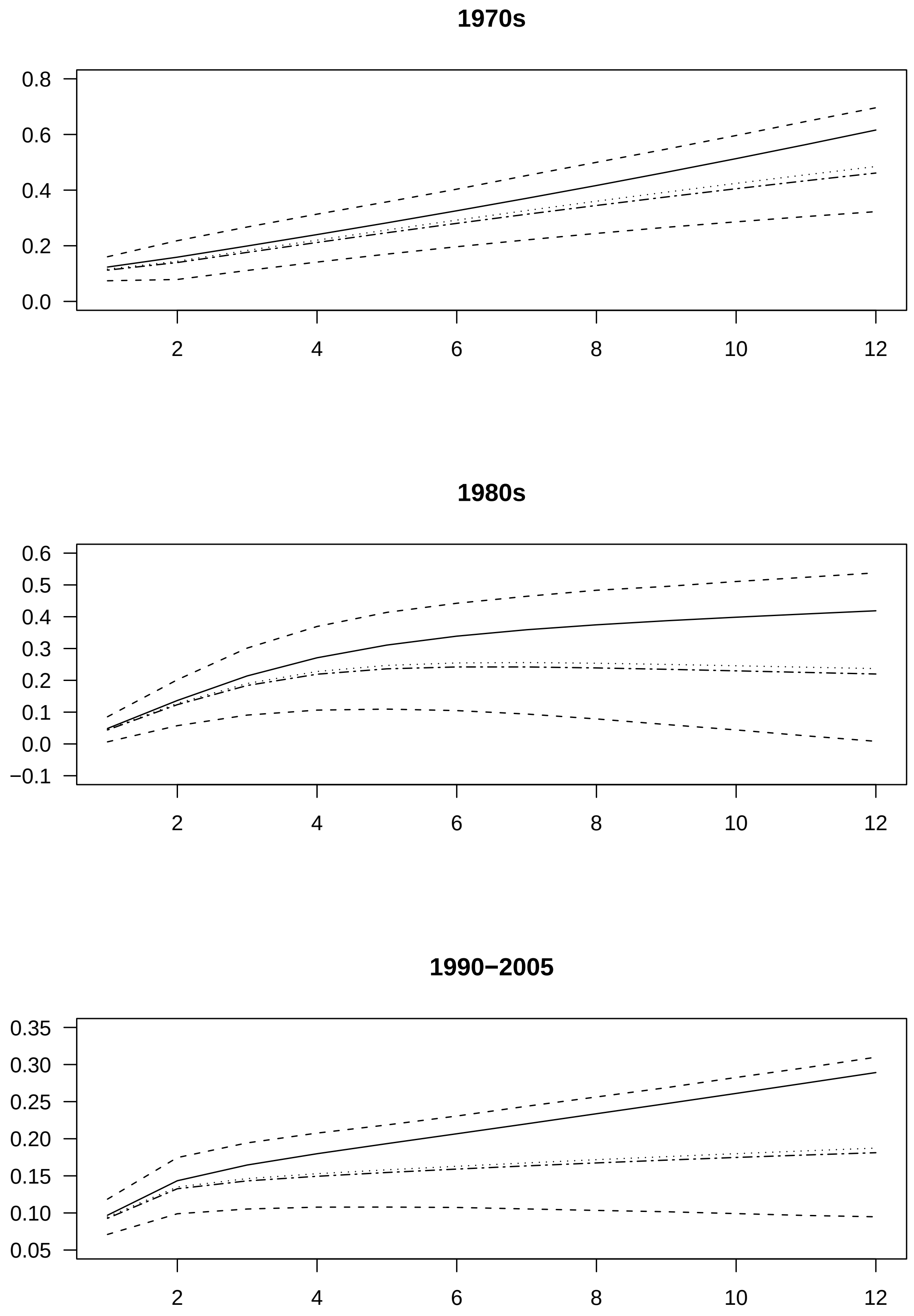

Figure 1. Estimates of Impulse Responses of Application 1: Response of CPI to PPI Raw Material Shocks.

Solid line=Bayesian estimates of impulse responses under the entropy loss; dotted line=posterior mean of impulse responses; dash-dotted lines= plugging MLE into functions of impulse response; dashed line= 95% posterior probability bands.

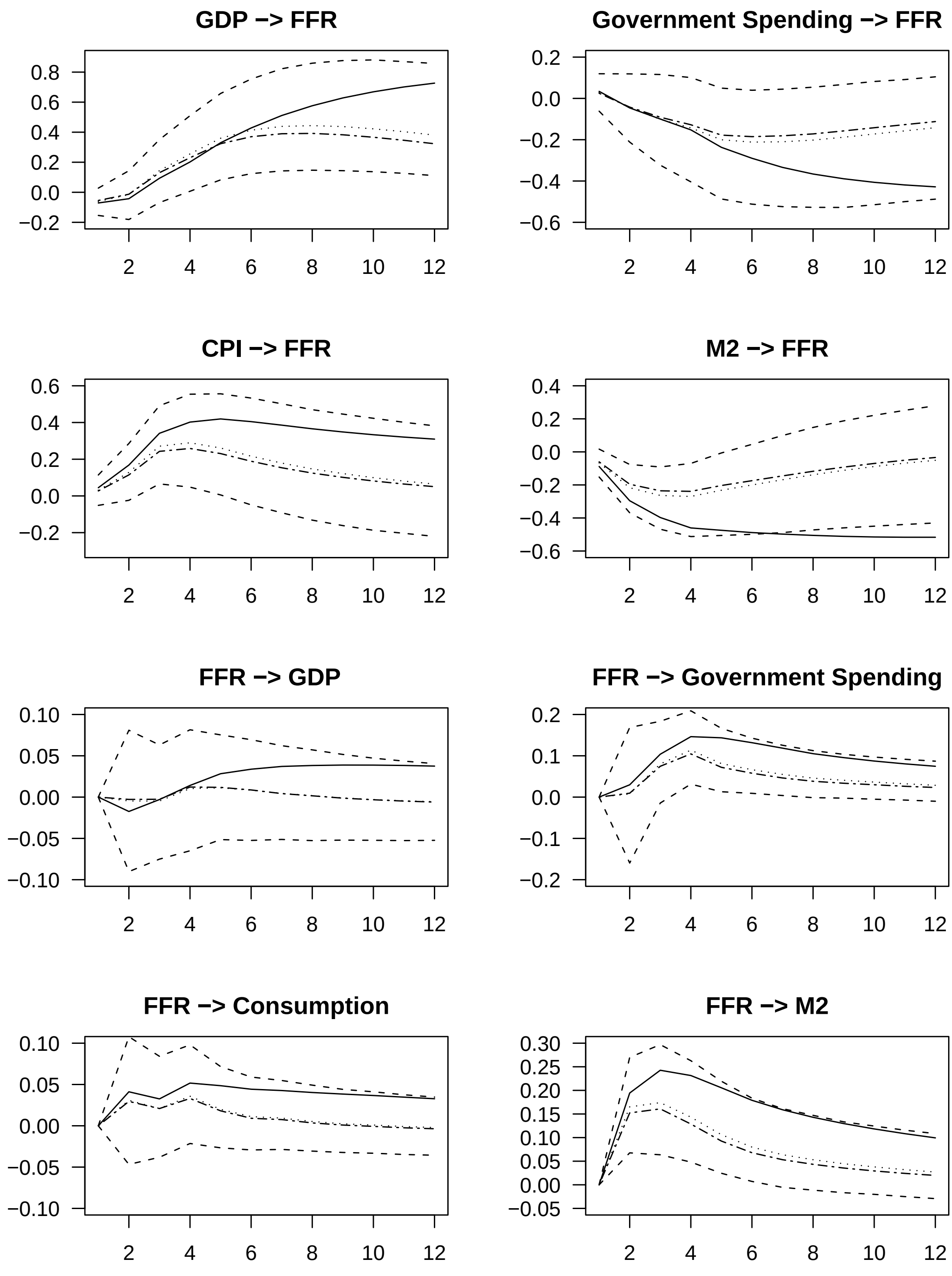

Figure 2. Estimates of Impulse Responses of Application 2: VAR of U.S. Macroeconomic Variables. Solid line=Bayesian estimates of impulse responses under the entropy loss; dotted line=posterior mean of impulse responses; dash-dotted lines= plugging MLE into functions of impulse response; dashed line= 95% posterior probability bands.