Intrinsic Bayesian estimation of linear time series models

Shawn Ni & Dongchu Sun

Taylor & Francis
Taylor & Francis Group

Check for updates

Intrinsic Bayesian estimation of linear time series models

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

[a]Department of Economics, University of Missouri, Columbia, MO, USA; [b]School of Statistics, East China Normal University, Shanghai, People's Republic of China; [c]Department of Statistics, University of Missouri, Columbia, MO, USA

ABSTRACT

Intrinsic loss functions (such as the Kullback–Leibler divergence, i.e. the entropy loss) have been used extensively in place of conventional loss functions for independent samples. But applications in serially correlated samples are scant. In the present study, we examine Bayes estimator of Linear Time Series (LTS) model under the entropy loss. We derive the Bayes estimator and show that it involves a frequentist expectation of regressors. We propose a Markov Chain Monte Carlo procedure that jointly simulates the posteriors of the LTS parameters with frequentist expectation of regressors. We conduct Bayesian estimation of an LTS model for seasonal effects in some U.S. macroeconomic variables.

1. Introduction

To analyse dynamics of multi-variate economic systems, researchers frequently employ Linear Time Series (LTS) models (see for example, Sims, 1980 and the ensuing literature). Bayesian inference of such models often requires point estimate of parameters because reporting the entire posterior distribution is made difficult by a prohibitively large number of parameters. A critical aspect in Bayesian estimation is the choice of loss function.

In this study, we derive Bayes estimator of LTS models based on the intrinsic loss. We illustrate a computational problem arising from serial correlation in the models when applying the intrinsic loss and our solution to the problem.

A loss function $L(\theta, \hat{\theta})$ measures the distance between the parameter θ and its estimate $\hat{\theta}$. Such a metric is often specified for convenience given the problem at hand instead of grounding on a general principle. Bernardo and Juárez (2003) noted that for inferential purposes, what matters most is not the distance between θ and $\hat{\theta}$, instead it is the *intrinsic loss* – the distance between the probability model $f(x \mid \hat{\theta})$ (corresponding to the estimate $\hat{\theta}$) and $f(x \mid \theta)$ (corresponding to the actual parameter θ). 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 has a number of desirable properties not generally possessed by conventional loss functions. For example, it is invariant

to transformation of data x or parameter θ and has the additive property that the loss of the sum of two independent data sets is the sum over the two losses corresponding to each data set.

The intrinsic loss has been used for independent samples for Bayesian estimation. It has also be used in various contexts for time series data. For instance, Kitamura and Stutzer (1997) used the Kullback–Leibler distance to derive a frequentist estimator for nonlinear models. Solo et al. (2001) used the Kullback–Leibler distance for evaluation of signal processing model. Robertson et al. (2005) used the entropy divergence for evaluation of forecasting density. Fernandez-Villaverde and Rubio-Ramirez (2004) used the Kullback–Leibler distance to evaluate dynamic equilibrium models in economics. However, employing the intrinsic loss for Bayesian estimation of time series models leads to technical challenges.

To illustrate the difference of the intrinsic loss in independent models and serially correlated models, consider the following examples. First, suppose $\boldsymbol{y} = \{y_1, \ldots, y_T\}$, where y_t $(t = 1, \ldots, T)$ are independently identically distributed (iid) $N(\rho, 1)$ and we are interested in estimating the mean parameter ρ under the entropy loss

$$\kappa(\hat{\rho} \mid \rho) = \int \log \left\{ \frac{f(\boldsymbol{y} \mid \rho)}{f(\boldsymbol{y} \mid \hat{\rho})} \right\} f(\boldsymbol{y} \mid \rho) \mathrm{d}\boldsymbol{y}$$
$$= \mathrm{E}_{\boldsymbol{y}\mid\rho} \log \left\{ \frac{f(\boldsymbol{y} \mid \rho)}{f(\boldsymbol{y} \mid \hat{\rho})} \right\}.$$

By the assumption on the model $f(\boldsymbol{y} \mid \rho) \propto \exp\{-\frac{1}{2} \times \sum_{t=1}^{T}(y_t - \rho)^2\}$. It is easy to verify that $\kappa(\hat{\rho} \mid \rho) = (T/2)(\rho - \hat{\rho})^2$. In this case, the intrinsic loss coincides with the commonly used quadratic loss, which implies that the Bayes estimator of ρ is the posterior mean. Now consider an AR(1) model: $y_t = \rho y_{t-1} + \epsilon_t$, for $t = 1, \ldots, T$, where ϵ_t is iid $N(0, 1)$, and ρ is the only unknown parameter.

The entropy loss is still

$$\kappa(\hat{\rho} \mid \rho) = \int \log\left\{\frac{f(\boldsymbol{y} \mid \rho)}{f(\boldsymbol{y} \mid \hat{\rho})}\right\} f(\boldsymbol{y} \mid \rho)\mathrm{d}\boldsymbol{y}$$

$$= \mathrm{E}_{\boldsymbol{y}\mid\rho} \log\left\{\frac{f(\boldsymbol{y} \mid \rho)}{f(\boldsymbol{y} \mid \hat{\rho})}\right\},$$

but now f is the density of the AR variable. Substitute in the distribution of data

$$\kappa(\hat{\rho} \mid \rho) = \frac{(\rho - \hat{\rho})^2}{2} \mathrm{E}_{\boldsymbol{y}\mid\rho} \sum_{t=1}^{T} y_{t-1}^2 = (\rho - \hat{\rho})^2 \delta(y_0, \rho),$$

where

$$\delta(y_0, \rho) = \frac{1}{2}\mathrm{E}_{\boldsymbol{y}\mid\rho} \sum_{t=1}^{T} y_{t-1}^2$$

$$= \frac{1}{2} \sum_{t=1}^{T} \left\{ \rho^{2(t-1)} y_0^2 + 1 + \rho^2 + \cdots + \rho^{2(t-2)} \right\}$$

$$= \frac{1}{2} \left\{ \frac{1 - \rho^{2T}}{1 - \rho^2} y_0^2 + \frac{1}{1 - \rho^2} \left(T - \frac{1 - \rho^{2T}}{1 - \rho^2} \right) \right\}.$$

It is obvious that $\delta(y_0, \rho)$ is an increasing function of ρ^2 and is nonnegative for any ρ. A Bayes estimator (which is called a generalised Bayes estimator if the prior is improper) minimises the Bayesian posterior expected loss. If the entropy loss is employed, the Bayes estimator for ρ with a given initial condition y_0 is $\hat{\rho}_E = \mathrm{argmin}\ \mathrm{E}_{\rho\mid\boldsymbol{y}}[\delta(y_0, \rho)(\rho - \hat{\rho})^2] = \mathrm{E}_{\rho\mid\boldsymbol{y}}\{\delta(y_0, \rho)\rho\}/\mathrm{E}_{\rho\mid\boldsymbol{y}}\{\delta(y_0, \rho)\}$. Note that if ρ is positive, ρ and $\delta(y_0, \rho)$ are positively correlated. It follows that the Bayes estimator under the entropy loss for a positive ρ is larger than the posterior mean. It is well known that the MLE (and the posterior mean under constant prior) of $\hat{\rho}_M$ is biased downward, especially when the true parameter is close to unity (see MacKinnon & Smith, 1998) Note that under the constant prior $\sum_{t=1}^{T} y_{t-1}^2$ is the posterior precision (i.e. the inverse of posterior variance) for ρ. Hence the weight for the square of estimation error in the intrinsic loss function, $\delta(y_0, \rho) = \mathrm{E}_{\boldsymbol{y}\mid\rho} \sum_{t=1}^{T} y_{t-1}^2$, is larger in the region of ρ where the posterior precision is high. It is in the spirit of Zellner's (1978, 1998) 'precision of the estimation' loss. This is in contrast of quadratic loss that imposes the same weight on all regions of ρ.

Now we turn to the model of interest. The LTS of a p dimensional column endogenous variable $\boldsymbol{y}_t$ and a q dimensional column exogenous (predetermined) variable $\boldsymbol{x}_{0t}$ ($t = 1, \ldots, T$) has the form:

$$\boldsymbol{y}_t' = \boldsymbol{x}_{0t}'\boldsymbol{B}_0 + \sum_{j=1}^{L} \boldsymbol{y}_{t-j}'\boldsymbol{B}_j + \boldsymbol{\epsilon}_t, \qquad (1)$$

where L is a known positive integer, $\boldsymbol{B}_0$ is a $q \times p$ unknown matrix, $\boldsymbol{B}_j$ is an unknown $p \times p$ matrix, $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_T$ are iid $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ errors, and $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix.

A special case of the above LTS model is all of the lag coefficients $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_L$ are zero (i.e. all regressors (with $q > p$) are exogenous variables.) The exogenous variables may be functions of time. For example, in modelling of climate temperature or holiday consumer spending, seasonal dummies may be introduced in the model. In economic applications, these exogenous variables may also be variables of government policies. Another special case is $\boldsymbol{x}_{0t}$ is a $1 \times p$ constant vector with elements of unity. The regressors only include lags of the variable $\boldsymbol{y}_t$. The LTS model becomes a Vector AutoRegression (VAR), which is commonly used for modelling of macroeconomic time series.

We can rewrite Equation (1) in the familiar matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Phi} + \boldsymbol{\epsilon}, \qquad (2)$$

where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_T' \end{pmatrix} = (X_0, X_1); \quad X_0 = \begin{pmatrix} \boldsymbol{x}_{01}' \\ \vdots \\ \boldsymbol{x}_{0T}' \end{pmatrix},$$

$$X_1 = \begin{pmatrix} \boldsymbol{y}_0' & \cdots & \boldsymbol{y}_{1-L}' \\ \vdots & \vdots & \vdots \\ \boldsymbol{y}_{T-1}' & \cdots & \boldsymbol{y}_{T-L}' \end{pmatrix};$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{B}_0 \\ \boldsymbol{\Phi}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_0 \\ \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{B}_L \end{pmatrix}; \quad \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_T' \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1' \\ \vdots \\ \boldsymbol{\epsilon}_T' \end{pmatrix}.$$

Here X_0 and X_1 are $T \times q$ and $T \times Lp$; the former does not depend on parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$, but the latter does. $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are $T \times p$ matrices, $\boldsymbol{\Phi}$ is a $(q + Lp) \times p$ matrix of unknown parameters, $\boldsymbol{x}_t'$ is a $1 \times (q + Lp)$ row vector, and $\boldsymbol{X}$ is a $T \times (q + Lp)$ matrix of observations. The likelihood function of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ based on $\boldsymbol{Y}$ is then

$$f(\boldsymbol{Y} \mid \boldsymbol{\Phi}, \boldsymbol{\Sigma})$$

$$\propto \frac{1}{|\boldsymbol{\Sigma}|^{T/2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{x}_t\boldsymbol{\Phi})' \right\}$$

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

$$(3)$$

Here and hereafter $\mathrm{etr}(\boldsymbol{A})$ is $\exp(\mathrm{tr}(\boldsymbol{A}))$ of a matrix $\boldsymbol{A}$.

The present paper achieves two goals. The first one is derivation of the Bayes estimator of LTS model under the entropy loss. We show that the entropy loss on $(\mathbf{\Phi}, \mathbf{\Sigma})$ is non-separable in $\mathbf{\Sigma}$ and $\mathbf{\Phi}$, which can be written as the sum of losses pertaining to the covariance matrix $\mathbf{\Sigma}$ and normalised estimation error of $\mathbf{\Phi}$. The form of the $\mathbf{\Phi}$-part of the entropy loss for LTS is $\mathrm{tr}\{\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{\Phi} - \widehat{\mathbf{\Phi}})'\mathrm{E}_{(X|\mathbf{\Phi},\mathbf{\Sigma})}(X'X)(\mathbf{\Phi} - \widehat{\mathbf{\Phi}})\}$, where $\widehat{\mathbf{\Sigma}}$ is the Bayes estimator of $\mathbf{\Sigma}$. Under the entropy loss, the Bayes estimator distinctly differs from the posterior mean and differs from that of the iid multivariate normal model. The part of the intrinsic loss function associated with the regression coefficients turns out to be related with a conventional loss function. For estimation of a matrix parameter such as $\mathbf{\Phi}$ in the simultaneous equations context, Zellner (1978, 1998) proposed a 'precision of estimation' loss that can also be written as $\mathrm{tr}\{\mathbf{\Sigma}^{-1}(\mathbf{\Phi} - \widehat{\mathbf{\Phi}})'(X'X)(\mathbf{\Phi} - \widehat{\mathbf{\Phi}})\}$. However, in Zellner's simultaneous equations model, $X'X$ is taken as given, but in LTS the predetermined variable $X'X$ depends on parameters $(\mathbf{\Phi}, \mathbf{\Sigma})$.

The second goal concerns numerical estimation of the intrinsic Bayes estimator via Markov Chain Monte Carlo (MCMC). We propose a general algorithm that generate regressors as latent parameters in simulation of posteriors of parameters of LTS models. Data augmentation in this study differs from that in Tanner and Wang's 1987 seminal paper in motivation and implementation. Tanner and Wang use data augmentation to alter the likelihood function for easier MCMC simulation of the posteriors. In this study, the likelihood function of the generated data is the same as the likelihood of the sample data. Here, data augmentation does not make it easier for posterior simulation. Instead, it makes it possible to compute frequentist moment $\mathrm{E}_{(X|\mathbf{\Phi},\mathbf{\Sigma})}(X'X)$ of the LTS variables. The frequentist moment, simulated jointly with parameters, is used to produce Bayes estimates under the entropy loss.

Besides the choice of loss function, the choice of prior also plays a pivotal role in Bayesian estimation. Jeffreys prior on $\mathbf{\Sigma}$ (see Zellner, 1971) is a noninformative prior for $\mathbf{\Sigma}$ that gives rise to conditional posteriors in well known distributions. Ni et al. (2007) conducted Bayesian estimation of VAR model under the entropy loss, using the Jefferys prior for $\mathbf{\Sigma}$. However, despite its popularity the Jeffreys prior is known for producing unsatisfactory results in multi-parameter settings. In this study we simulate the LTS model under a combination of normal prior on regression parameters and Yang and Berger (1994) reference prior on $\mathbf{\Sigma}$. The conditional posteriors of $\mathbf{\Sigma}$ are simulated using a Metropolis-Hastings algorithm. Our empirical application shows that despite the fact that LTS models involve a large number of parameters and a large number of latent variables, the data-augmentation algorithm is quite efficient.

In Section 2 of the paper, we derive the Bayes estimator of LTS models under the entropy loss function and discuss computation of the weighting matrix in the Bayes estimator. In Section 3, we present a general algorithm using generated data as latent parameters. In Section 4, we lay out the MCMC algorithm for computing $(\mathbf{\Phi}, \mathbf{\Sigma})$ in the LTS model. In Section 5, we first compare intrinsic Bayes estimator with other estimators in a numerical example and then estimate a LTS model using seasonally unadjusted macroeconomic data. In Section 6 we offer concluding remarks.

2. Entropy loss function for the iid multivariate and LTS models

2.1. Entropy loss function for the iid model

We first consider the entropy loss function (Robert, 1994, p. 74) for a multivariate normal distribution. Let $\mathbf{Y} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ be a random sample from $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. One can compute the entropy loss function as

$$
\begin{aligned}
L(\tilde{\boldsymbol{\mu}}, &\widetilde{\mathbf{\Sigma}}; \boldsymbol{\mu}, \mathbf{\Sigma}) \\
&= \int \log \left\{ \frac{p(\mathbf{x} \mid \boldsymbol{\mu}, \mathbf{\Sigma})}{p(\mathbf{x} \mid \tilde{\boldsymbol{\mu}}, \widetilde{\mathbf{\Sigma}})} \right\} p(\mathbf{x} \mid \boldsymbol{\mu}, \mathbf{\Sigma}) \, \mathrm{d}\mathbf{x} \\
&= \frac{T}{2} \Big\{ \mathrm{tr}(\mathbf{\Sigma} \widetilde{\mathbf{\Sigma}}^{-1}) - \log |\mathbf{\Sigma} \widetilde{\mathbf{\Sigma}}^{-1}| - p \\
&\quad + (\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})' \widetilde{\mathbf{\Sigma}}^{-1} (\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) \Big\}.
\end{aligned}
$$

Here $p(\mathbf{x} \mid \boldsymbol{\mu}, \mathbf{\Sigma})$ is the density of $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. Clearly, the loss function has two parts. One part is related to the means $\tilde{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ (with $\widetilde{\mathbf{\Sigma}}$ as the weighting matrix), and the other part is related to $\widetilde{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$. The following fact states that Bayes estimator for $\boldsymbol{\mu}$ is the posterior mean of $\boldsymbol{\mu}$ but that of $\mathbf{\Sigma}$ is larger than the posterior mean of $\mathbf{\Sigma}$.

Fact 2.1: *Under the entropy loss L, the generalised Bayes estimator of $(\boldsymbol{\mu}, \mathbf{\Sigma})$ is*

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_{\mathrm{iid}} &= \mathrm{E}\{\boldsymbol{\mu} \mid \mathbf{Y}\}, \\
\widehat{\mathbf{\Sigma}}_{\mathrm{iid}} &= \mathrm{E}(\mathbf{\Sigma} \mid \mathbf{Y}) + \mathrm{var}(\boldsymbol{\mu} \mid \mathbf{Y}).
\end{aligned}
$$

Note that $\mathbf{Y}$ represents data, expectation $\mathrm{E}(.)$ and variance $\mathrm{var}(.)$ are with respect to the posterior distribution. The proof is in the appendix.

2.2. Entropy loss functions for LTS models

Recall that for the LTS model (2), the likelihood function of $(\mathbf{\Phi}, \mathbf{\Sigma})$ is of the form (3). The entropy loss for the LTS model is

$$
L(\widetilde{\mathbf{\Phi}}, \widetilde{\mathbf{\Sigma}}; \mathbf{\Phi}, \mathbf{\Sigma}) = \mathrm{E}_{(Y|\mathbf{\Phi},\mathbf{\Sigma})} \log \left\{ \frac{f(\mathbf{Y} \mid \mathbf{\Phi}, \mathbf{\Sigma})}{f(\mathbf{Y} \mid \widetilde{\mathbf{\Phi}}, \widetilde{\mathbf{\Sigma}})} \right\}, \quad (4)
$$

where for computing the expectation on the right-hand side, $(\widetilde{\Phi}, \widetilde{\Sigma})$ is not a function of Y. The entropy loss $L(\widetilde{\Phi}, \widetilde{\Sigma}; \Phi, \Sigma)$ can be decomposed into two parts. One part measures the loss associated with the covariance matrix only, while the second part measures the loss of coefficients Φ, but related to the covariance matrix Σ as well. Because $y_t - X_t\Phi$ $(t = 1, \ldots, T)$ are iid $N_p(0, \Sigma)$, we have

$$E_{(X|\Phi,\Sigma)}(Y - X\Phi) = 0,$$
$$E_{(X|\Phi,\Sigma)}\{(Y - X\Phi)'X\Phi\} = 0,$$
$$E_{(X|\Phi,\Sigma)}\{(Y - X\Phi)'(Y - X\Phi)\} = T\Sigma.$$

Then

$$E_{(X|\Phi,\Sigma)}\left\{\log\frac{|\Sigma|^{-T/2}\mathrm{etr}\{-\frac{1}{2}(Y-X\Phi)\Sigma^{-1}(Y-X\Phi)'\}}{|\widehat{\Sigma}|^{-T/2}\mathrm{etr}\{-\frac{1}{2}(Y-X\widehat{\Phi})\widehat{\Sigma}^{-1}(Y-X\hat{\Phi})'\}}\right\}$$

$$= \frac{T}{2}(\log|\widehat{\Sigma}\Sigma^{-1}| - p)$$
$$+ \frac{1}{2}\mathrm{tr}\left[E_{(X|\Phi,\Sigma)}\left\{(Y - X\widehat{\Phi})\widehat{\Sigma}^{-1}(Y - X\hat{\Phi})'\right\}\right]$$

$$= \frac{T}{2}(\log|\widehat{\Sigma}\Sigma^{-1}| - p) + \frac{T}{2}\mathrm{tr}(\widehat{\Sigma}\Sigma^{-1})$$
$$+ \frac{1}{2}\mathrm{tr}\left[E_{(X|\Phi,\Sigma)}\left\{X(\Phi - \widehat{\Phi})\widehat{\Sigma}^{-1}(\Phi - \widehat{\Phi})'X'\right\}\right]$$

$$= \frac{T}{2}\left\{\mathrm{tr}(\widehat{\Sigma}^{-1}\Sigma) - \log|\widehat{\Sigma}^{-1}\Sigma| - p\right\}$$
$$+ \frac{1}{2}\mathrm{tr}\left\{\widehat{\Sigma}^{-1}(\Phi - \widehat{\Phi})'E_{(X|\Phi,\Sigma)}(X'X)(\Phi - \widehat{\Phi})\right\}.$$

The result of this derivation can be summarised by the following lemma.

Lemma 2.1: *The entropy loss function for the LTS model is*

$$\frac{T}{2}\left\{\mathrm{tr}(\widetilde{\Sigma}^{-1}\Sigma) - \log|\widetilde{\Sigma}^{-1}\Sigma| - p\right\}$$
$$+ \frac{1}{2}\mathrm{tr}\left\{\widetilde{\Sigma}^{-1}(\Phi - \widetilde{\Phi})'W(\Phi - \widetilde{\Phi})\right\}, \quad (5)$$

where

$$W = \frac{1}{T}E_{(X|\Phi,\Sigma)}(X'X). \quad (6)$$

This Lemma can be proved using algebra similar to that in the iid case. However, there is an important difference. For the LTS model, Bayes estimators involve matrix W, a frequentist expectation of $X'X$ for given parameters (Φ, Σ). For the iid case, no such term is present. The next theorem gives the form of the Bayes estimators under the entropy loss.

Theorem 2.1: *The generalised Bayes estimator of (Φ, Σ) under the entropy loss is*

$$\widehat{\Phi}_E = \left\{E(W \mid Y)\right\}^{-1}E(W\Phi \mid Y), \quad (7)$$

$$\widehat{\Sigma}_E = E(\Sigma \mid Y) + E\left\{(\Phi - \widehat{\Phi}_E)'W(\Phi - \widehat{\Phi}_E) \mid Y\right\}. \quad (8)$$

The above theorem can be proved similarly as Fact 2.1.

Under the special case with no lag coefficients in the regression, we have $W = E_{(X|\Phi,\Sigma)}(X'X) = X_0'X_0$, which is not a function of Σ and Φ. It follows that the Bayes estimator of $\widehat{\Phi}_E$ is the posterior mean, as it is for the iid model. This observation is stated in the following remark.

Remark 1: If $B_j = 0$ for $j > 0$, then

$$\widehat{\Phi}_E = E\{\Phi \mid Y\},$$

$$\widehat{\Sigma}_E = E(\Sigma \mid Y) + \frac{1}{T}X_0'X_0\mathrm{Var}(\Phi \mid Y).$$

However, the Bayes estimator for the LTS model is generally different from the iid case. The Bayes estimator $\widehat{\Phi}_E$ for the LTS model is not the posterior mean. To compare the estimator $\widehat{\Phi}_E$ with the posterior mean, note that in general

$$\widehat{\Phi}_E = E(\Phi \mid Y) + \left\{E(W \mid Y)\right\}^{-1}\mathrm{Cov}(W, \Phi \mid Y).$$

Because $W = E_{(X|\Phi,\Sigma)}(X'X)$ and Φ are likely to be positively correlated, the Bayes estimator of Φ under the intrinsic loss is likely to be larger than the posterior mean. It is known that MLE and the posterior mean of Φ under a diffuse prior is likely to have a downward bias when the true parameters are closed to random walk, a typical pattern of macroeconomic data. The form of Bayes estimator of Φ based on the intrinsic loss is helpful in correcting the bias in the posterior mean.

The estimator in LTS model involves the frequentist expectation W. The W matrix depends on specifications of the regressors X. If the regressors are specified as functions of lags of Y, the computation for W matrix becomes nontrivial.

Using notation in Equation (1), the frequentist expectation matrix W can be written as

$$E_{(X|\Phi,\Sigma)}(X'X) = \begin{pmatrix} X_0'X_0 & E_{(X|\Phi,\Sigma)}(X_0'X_1) \\ E_{(X|\Phi,\Sigma)}(X_1'X_0) & E_{(X|\Phi,\Sigma)}(X_1'X_1) \end{pmatrix}.$$

For exogenous variables X_0, there is no need deriving general closed-form expression for the terms in the above matrix as functions of parameter Φ and Σ. On the other hand, due to the serial correlations of y_t, computation of $E_{(X|\Phi,\Sigma)}(X_1'X_1)$ is not straightforward. In the presence of exogenous variables no analytical expression for W is available. In the following, we discuss approaches to Bayesian estimation under the entropy loss for the general LTS model.

3. Approaches of computing the expectation $E_{(X|\Phi,\Sigma)}(X'X)$

Theorem 2.1 shows that under the entropy loss the Bayes estimator for (Φ, Σ) involves the frequentist expectation $W = E_{(X|\Phi,\Sigma)}(X'X)$, and we need to compute the posterior moments $E(W \mid Y)$, $E(W\Phi \mid Y)$, and $E(\Phi'W\Phi \mid Y)$.

The frequentist expectation $E_{(X|\Phi,\Sigma)}(X'X)$ depends on Φ and Σ. For the LTS model $E_{(X|\Phi,\Sigma)}(X'X)$ does not have an analytical form and needs to be computed numerically for a given (Φ, Σ). We use Y and X to denote observed data in the LTS model (2). We generate Y^* and X^* from the same model in (2) given parameters Φ and Σ, in order to compute $E_{(X|\Phi,\Sigma)}(X'X)$. There is only one observed data set Y and X but there are many sets of generated Y^* and X^*. Suppose Φ and Σ need to be simulated by an MCMC algorithm, then Y^* and X^* need to be generated for each draw of Φ and Σ.

One approach to computing $E_{(X|\Phi,\Sigma)}(X'X)$ is straight forward but time-consuming: for each Φ_k and Σ_k drawn in the kth MCMC cycle we generate many sets of X^* and use the average of $X^{*\prime}X^*$ to approximate $E_{(X|\Phi,\Sigma)}(X'X)$. While this approach is possible in theory its high computational cost renders it infeasible in practice. For practical purposes, we must take an alternative approach to compute Bayes estimates.

Fortunately, we have an alternative approach that does not require much additional computational cost beyond simulating (Φ, Σ). Suppose we simulate one set of data X_k^* from the LTS model in each MCMC cycle with simulated parameters of $(\Phi_{k-1}, \Sigma_{k-1})$, and then simulate the parameters of the next MCMC cycle (Φ_k, Σ_k) conditional on both the sample data X and the simulated data X_k^*. We will demonstrate that the posterior moments such as $E(W \mid Y)$, $E(W\Phi \mid Y)$, and $E(\Phi'W\Phi \mid Y)$ can be computed through simulated parameters (Φ_k, Σ_k) and the jointly simulated data X_k^* (for $k = 1, \ldots, M$). The simulated data are in essence latent parameters. They are not the subject of our interests per se but are useful for simulation of parameter of interest (i.e. the frequentist expectation W). Data augmentation is not uncommon in Bayesian simulations, but as we noted in the introduction, this data-augmented simulation approach differs from its other uses in the econometrics and statistics literature. One question of practical importance remains though: The number of elements in simulated matrix X_1^* has the dimension of $T \times Lp$, which can be quite large. Do we have to simulate very long Markov chains to assure the averages are good approximates of the posterior mean? Fortunately, our numerical results show that the answer to the question is "no".

In the following we propose a general algorithm that formalises the data-augmentation idea discussed above.

3.1. A general algorithm using data as latent parameters

Suppose that observed data X has the density $f(x \mid \theta)$, where parameter vector θ is unknown. A prior $\pi(\theta)$ can be informative or noninformative. Let X^* be a random vector (or a matrix) with the density $f(x^* \mid \theta)$. Let $h(\theta)$ be a function of the parameters θ. We are interested in the posterior mean of the quantity $E_{(X^*|\theta)}g(X^*, h(\theta))$ given the data X.

Our algorithm is based on the following fact:

$$E\{E_{(X^*|\theta)}g(X^*, h(\theta)) \mid X\}$$

$$= \frac{\int \left\{\int g(x^*, h(\theta))f(x^*|\theta)dx^*\right\}f(X \mid \theta)\pi(\theta)d\theta}{\int f(X \mid \theta)\pi(\theta)d\theta}$$

$$= \int \int g(x^*, h(\theta))\,\pi(x^*, \theta \mid X)\,dx^*d\theta,$$

where

$$\pi(x^*, \theta \mid X) = \frac{f(x^* \mid \theta)f(X \mid \theta)\pi(\theta)}{\int \int f(\tilde{x}^* \mid \tilde{\theta})f(X \mid \tilde{\theta})\pi(\tilde{\theta})d\tilde{x}^*d\tilde{\theta}}.$$

$$(9)$$

If we have a random sample $(X_k^*, \theta_k), k = 1, \ldots, M$, from the joint distribution of (9), we can estimate $E(E_{(X^*|\theta)}\{g(X^*, \theta)\}h(\theta) \mid X)$ by using the result

$$\hat{E}[E\{g(X^*, h(\theta)) \mid \theta\} \mid X] = \hat{E}_{(X^*,\theta|X)}\{g(X^*, h(\theta))\}$$

$$= \frac{1}{M}\sum_{k=1}^{M} g(X_k^* h(\theta_k)).$$

The problem becomes to generate observations from the joint distribution of (X^*, θ) given the data X. For this task the following MCMC method can be used.

Suppose that at the beginning of cycle k we have $(X_{k-1}^*, \theta_{k-1})$.

Simulating full conditional posterior: We sample from

$$\pi(\theta \mid X^*, X) \propto f(X^* \mid \theta)f(X \mid \theta)\pi(\theta).$$

Step 1. Simulate $X_k^* \sim f(x^* \mid \theta_{k-1})$.
Step 2. Simulate $\theta_k \sim \pi(\theta \mid X_k^*, X) \propto f(X_k^* \mid \theta)f(X \mid \theta)\pi(\theta)$.

4. Bayesian estimation of (Φ, Σ) in LTS models

4.1. Priors

The Bayes estimator of LTS depends on the prior of (Φ, Σ) We assume prior independence so the prior for $\pi(\Phi, \Sigma)$ is $\pi(\phi)\pi(\Sigma)$, the product of priors for Φ and Σ.

For estimation of regression coefficient Φ, a popular informative prior of $\phi = \text{vec}(\Phi)$ is the normal

distribution, $N(\boldsymbol{\phi}_0, \boldsymbol{M}_0)$, with hyperparameters $\boldsymbol{\phi}_0$ and $\boldsymbol{M}_0$:

$$\pi_N(\boldsymbol{\phi}) \propto |\boldsymbol{M}_0|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)' \boldsymbol{M}_0^{-1}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right\}. \tag{10}$$

A popular class of non-informative prior on $\boldsymbol{\Sigma}$ is $\pi_b(\boldsymbol{\Sigma}) \propto 1/|\boldsymbol{\Sigma}|^{b/2}$. If $b = p + 1$, $\pi_b(\boldsymbol{\Sigma})$ becomes the Jeffreys prior (see Zellner, 1971)

Ni et al. (2007) examined intrinsic Bayes estimator under prior $\pi_b(\boldsymbol{\phi}, \boldsymbol{\Sigma}) = \pi_N(\boldsymbol{\phi})\pi_b(\boldsymbol{\Sigma})$.

In the appendix we show posteriors $\pi(\boldsymbol{\Sigma} \mid \boldsymbol{X}^*, \boldsymbol{X})$ and $\pi(\boldsymbol{\Phi} \mid \boldsymbol{X}^*, \boldsymbol{X})$ can be obtained in analytical form.

As mentioned in the introduction, in multiple-parameter settings the Jeffreys prior often has undesirable properties. Bernardo (1979) proposed an approach of deriving a reference prior by breaking a single multi-parameter problem into a consecutive series of problems with fewer numbers of parameters. 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 $\boldsymbol{\Sigma}$ based on an iid random sample from a normal population with known mean, Yang and Berger (1994) re-parameterised matrix $\boldsymbol{\Sigma}$ as $\boldsymbol{O}'\boldsymbol{D}\boldsymbol{O}$, where $\boldsymbol{D}$ is a diagonal matrix the elements of which are the eigenvalues of $\boldsymbol{\Sigma}$ (in increasing or decreasing order) and $\boldsymbol{O}$ is an orthogonal matrix. The following reference prior is derived by giving vectorised $\boldsymbol{D}$ higher priority over vectorised $\boldsymbol{O}$:

$$\pi_R(\boldsymbol{\Sigma}) \propto \frac{1}{|\boldsymbol{\Sigma}| \prod_{1 \leq i < j \leq p}(d_i - d_j)}, \tag{11}$$

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

For numerical and empirical exercise in the is study we use the normal-reference prior $\pi_N(\boldsymbol{\phi})\pi_R(\boldsymbol{\Sigma})$. The conditional densities under the normal-reference prior are in the appendix. Ni and Sun (2003) proved that the posteriors of $(\boldsymbol{\Phi}, \boldsymbol{\Sigma})$ are proper under the normal-reference prior. But the conditional posterior $\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\Phi}, \boldsymbol{X}^*, \boldsymbol{X})$ does not have an analytical form and must be sampled numerically.

4.2. A simulation algorithm for LTS models under the normal-reference prior

We employ an MCMC method to sample from the posterior. In particular, we use the Gibbs sampling method (cf. Gelfand & Smith, 1990). The following algorithm simulates the posteriors of LTS parameters conditional on both the sample and generated data.

Suppose that at cycle k, we have $(\boldsymbol{\Phi}_{k-1}, \boldsymbol{\Sigma}_{k-1})$ (with an initial draw of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$, e.g. the MLE.)

Algorithm MCMC:
Step 1. Generate $\boldsymbol{Y}_k^* | \boldsymbol{\Sigma}_{k-1}, \boldsymbol{\Phi}_{k-1}$.

Simulate $\boldsymbol{y}_{k,t}^{*'} \sim N(\boldsymbol{x}_{0t}'\boldsymbol{B}_0 + \sum_{i=1}^{L} \boldsymbol{y}_{k,t-i}^{*'}\boldsymbol{B}_{k-1,i}, \boldsymbol{\Sigma}_{k-1})$, for $t = 1, \ldots, T$. Define

$$\boldsymbol{Y}_k^* = \begin{pmatrix} \boldsymbol{y}_{k,1}^{*'} \\ \vdots \\ \boldsymbol{y}_{k,T}^{*'} \end{pmatrix} \text{ and } \boldsymbol{X}_k^* = \begin{pmatrix} \boldsymbol{x}_{01}' & \boldsymbol{y}_{k,-1}^{*'} & \cdots & \boldsymbol{y}_{k,-L}^{*'} \\ \boldsymbol{x}_{02}' & \boldsymbol{y}_{k,0}^{*'} & \cdots & \boldsymbol{y}_{k,1-L}^{*'} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_{0T}' & \boldsymbol{y}_{k,T-1}^{*'} & \cdots & \boldsymbol{y}_{k,T-L}^{*'} \end{pmatrix}.$$

Step 2. Generate $\boldsymbol{\Phi}_k | \boldsymbol{\Sigma}_{k-1}, \boldsymbol{Y}_k^*, \boldsymbol{Y}$.
Simulate $\boldsymbol{\phi}_k = \text{vec}(\boldsymbol{\Phi}_k) \sim N(\boldsymbol{\mu}_k, \boldsymbol{V}_k)$, where

$$\boldsymbol{\mu}_k = \boldsymbol{V}_k^*[\{\boldsymbol{\Sigma}_{k-1}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}$$
$$+ \boldsymbol{X}_k^{*'}\boldsymbol{X}_k^*)\}\hat{\boldsymbol{\phi}}_{\text{MLE}k}^* + \boldsymbol{M}_0^{-1}\boldsymbol{\phi}_0]; \tag{12}$$

$$\hat{\boldsymbol{\phi}}_{\text{MLE}k}^* = \text{vec}\left\{\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_k^{*'}\boldsymbol{X}_k^*\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{X}_k^{*'}\boldsymbol{Y}_k^*\right)\right\}. \tag{13}$$

$$\boldsymbol{V}_k^* = \{\boldsymbol{M}_0^{-1} + \boldsymbol{\Sigma}_{k-1}^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_k^{*'}\boldsymbol{X}_k^*)\}^{-1}. \tag{14}$$

Steps 3 to 6 generate $\boldsymbol{\Sigma}_k | \boldsymbol{\Sigma}_{k-1}, \boldsymbol{\Phi}_k, \boldsymbol{Y}_k^*, \boldsymbol{Y}$.
Step 3: Calculate $\boldsymbol{S}_k = S(\boldsymbol{\Phi}_k) + S^*(\boldsymbol{\Phi}_k) = (\boldsymbol{Y} - \boldsymbol{X} \times \boldsymbol{\Phi}_k)'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\Phi}_k) + (\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\Phi}_k)'(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\Phi}_k)$. Decompose $\boldsymbol{\Sigma}_{k-1} = \boldsymbol{O}\boldsymbol{D}\boldsymbol{O}'$, where $\boldsymbol{O}$ is an orthogonal matrix, $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_p)$ and $d_1 > d_2 \cdots > d_p$. Let $d_i^\# = \log(d_i)$,

$$\boldsymbol{D}^\# = \text{diag}(d_1^\#, \ldots, d_p^\#) \quad \text{and} \quad \boldsymbol{\Sigma}_{k-1}^\# = \boldsymbol{O}\boldsymbol{D}^\#\boldsymbol{O}'.$$

Step 4: Select a random symmetric $p \times p$ matrix $\boldsymbol{V}$, with elements, $v_{ij} = z_{ij}/\sqrt{\sum_{l \leq m} z_{lm}^2}$, where $z_{ij} \sim N(0, 1)$ $(1 \leq i \leq j \leq p$, the other elements of $\boldsymbol{V}$ are defined by symmetry).

Step 5: Generate $\lambda \sim N(0, 1)$ and set $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{k-1}^\# + \lambda\boldsymbol{V}$. Decompose $\boldsymbol{\Psi} = \boldsymbol{Q}\boldsymbol{C}^\#\boldsymbol{Q}'$, where $\boldsymbol{Q}$ is an orthogonal matrix, $\boldsymbol{C}^\# = \text{diag}(c_1^\#, \ldots, c_p^\#)$ and $c_1^\# > c_2^\# \cdots > c_p^\#$. Compute

$$\beta_k = T\sum_{i=1}^{p}(d_i^\# - c_i^\#)$$
$$+ \frac{1}{2}\text{tr}[\{(\exp\boldsymbol{\Sigma}_{k-1}^\#)^{-1} - (\exp\boldsymbol{\Psi})^{-1}\}\boldsymbol{S}_k]$$
$$+ \sum_{i<j}\log(d_i^\# - d_j^\#) - \sum_{i<j}\log(c_i^\# - c_j^\#).$$

Step 6: Define $\boldsymbol{C} = \text{diag}(\exp(c_1^\#), \ldots, \exp(c_p^\#))$ and $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q}'$. Simulate $u \sim \text{uniform}(0, 1)$ and let

$$\boldsymbol{\Sigma}_k = \begin{cases} \widetilde{\boldsymbol{\Sigma}}, & \text{if } u \leq \min\{1, \exp(\beta_k)\}, \\ \boldsymbol{\Sigma}_{k-1}, & \text{otherwise.} \end{cases}$$

Note the acceptance probability $\exp(\beta_k) = \pi(\boldsymbol{\Psi}|\boldsymbol{\Phi}_k, \boldsymbol{Y}_k^*, \boldsymbol{Y})/\pi(\boldsymbol{\Sigma}_{k-1}|\boldsymbol{\Phi}_k, \boldsymbol{Y}_k^*, \boldsymbol{Y})$, where the conditional posterior $\pi(\boldsymbol{\Sigma}|\boldsymbol{\Phi}, \boldsymbol{Y}^*, \boldsymbol{Y})$ is given in (A14). To accelerate the convergence, we repeat Step 6 up to five times until a new candidate is accepted.

4.3. Computing the posterior average loss

From Lemma 2.1, given the estimate $(\widehat{\Phi}, \widehat{\Sigma})$, which is computed for a given data sample Y, we write the posterior average loss $\mathrm{E}_{((\Phi,\Sigma)|Y)}L(\widehat{\Phi}, \widehat{\Sigma}, \Phi, \Sigma)$ as $\mathrm{E}_{(\Phi,\Sigma)|Y}L_1(\widehat{\Sigma};\Sigma) + \mathrm{E}_{(\Phi,\Sigma)|Y}L_2(\widehat{\Sigma}, \widehat{\Phi}, \Sigma, \Phi)$. We decompose the Σ_k in the kth MCMC cycle as $\Sigma_k = QD_kQ'$, where D_k is the diagonal matrix that consists of eigenvalues of Σ_k: $D_k = \mathrm{diag}(d_{k1}, d_{k2}, \ldots, d_{kp})$, and Q is an orthogonal matrix with $QQ' = I$.

The posterior average loss under the intrinsic loss can be computed using the posterior draws of $(\Phi_k, \Sigma_k, X_k^*)$ generated by the MCMC procedure ($k = 1, 2, \ldots, M$), with

$$
\widehat{\mathrm{E}}_{((\Phi,\Sigma)|Y)}L_1(\widehat{\Sigma};\Sigma)
$$

$$
= \widehat{\mathrm{E}}_{((\Phi,\Sigma)|Y)}\left\{\frac{T}{2}\{\mathrm{tr}(\widehat{\Sigma}^{-1}\Sigma) - \log|\widehat{\Sigma}^{-1}\Sigma| - p\}\right\}
$$

$$
= \frac{T}{2}\left\{\mathrm{tr}\left(\widehat{\Sigma}^{-1}\frac{1}{M}\sum_{k=1}^{M}\Sigma_k\right)\right.
$$

$$
\left. + \log|\widehat{\Sigma}| - p - \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{p}\log|d_{ki}|\right\}; \quad (15)
$$

$$
\widehat{\mathrm{E}}_{((\Phi,\Sigma)|Y)}L_2(\widehat{\Phi}, \widehat{\Sigma}; \Phi, \Sigma)
$$

$$
= \widehat{\mathrm{E}}_{((\Phi,\Sigma)|Y)}\frac{1}{2}\mathrm{tr}[\widehat{\Sigma}^{-1}\{(\Phi - \widehat{\Phi})'W(\Phi - \hat{\Phi})\}]
$$

$$
= \frac{1}{2}\left\{\mathrm{tr}\left(\widehat{\Sigma}^{-1}\frac{1}{M}\sum_{k=1}^{M}\Phi_k'X_k^{*'}X_k^*\Phi_k\right)\right.
$$

$$
\left. + \mathrm{tr}\left(\widehat{\Phi}\widehat{\Sigma}^{-1}\widehat{\Phi}'\frac{1}{M}\sum_{k=1}^{M}X_k^{*'}X_k^*\right)\right\}
$$

$$
- \mathrm{tr}\left(\widehat{\Sigma}^{-1}\widehat{\Phi}'\frac{1}{M}\sum_{k=1}^{M}X_k^{*'}X_k^*\Phi_k\right), \quad (16)
$$

where $W = \mathrm{E}_{(X|\Phi,\Sigma)}(X'X)$.

Note that all terms in the posterior entropy loss are functions of simulated Σ, Φ, and X^* over the MCMC cycles. The moments of the simulated parameters can be computed in the MCMC cycles, just as the posterior mean, without the need of storage of all of the simulated parameters.

5. A numerical example and an empirical study

5.1. A numerical example

In this section we first simulate data from an LTS model

$$
y_t = c + (y_{t-1}, x_t)B + \epsilon_t, \quad (17)
$$

for $t = 1, \ldots, T$. The dimension of the VAR variable y_t is 5. The exogenous variable x_t is a scalar representing seasonal cycles, with $x_1 = 0$, $x_2 = 1$, $x_3 = 0$, $x_4 = -1$, and $x_t = x_{t-4}$ for $t > 4$.

Now we let the true parameters be

$$
\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix},
$$

$$
B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \end{pmatrix}.
$$

The last row of matrix B are the parameters of the seasonal dummies. The discussion in Section 2 shows with this parameter setting there is no closed-form expression of the frequentist expectation $\mathrm{E}(X'X|\Sigma, \Phi)$ and we need to simulate $\mathrm{E}_{(X|\Phi,\Sigma)}(X'X)$ using the data-augmentation algorithms proposed in Section 4.

We generate one data sample (T) of 100 observations from the LTS model with the above parameters for Σ and Φ. The MLE of the parameters are

$$
\widehat{\Sigma}_{\mathrm{MLE}} = \begin{pmatrix} 0.93201 & 0.12358 & -0.08462 & 0.27246 & -0.11782 \\ 0.12358 & 1.76018 & 0.02217 & -0.16508 & 0.26475 \\ -0.08462 & 0.02217 & 2.44980 & -0.07874 & -0.01940 \\ 0.27246 & -0.16508 & -0.07874 & 3.44815 & -0.80056 \\ -0.11782 & 0.26475 & -0.01940 & -0.80056 & 4.03612 \end{pmatrix},
$$

$$
\widehat{\Phi}_{\mathrm{MLE}} = \begin{pmatrix} -0.05633 & 0.14700 & 0.01158 & 0.16497 & 0.26224 \\ 0.51936 & 0.24397 & 0.21705 & -0.1209 & -0.25149 \\ 0.0214 & 0.41804 & -0.03495 & 0.06221 & 0.28700 \\ -0.04826 & -0.03265 & 0.50279 & -0.03351 & -0.10127 \\ -0.06176 & 0.01439 & 0.14174 & 0.37051 & -0.01758 \\ -0.0194 & -0.0014 & -0.08389 & 0.01734 & 0.39826 \\ 0.15900 & 0.22462 & -0.21547 & -0.01671 & 1.06891 \end{pmatrix}.
$$

We conduct the simulation with a diffuse prior on Φ and the Yang–Berger reference prior on Σ. The length of MCMC cycles set at 100,000. The intrinsic Bayes estimates are

$$
\widehat{\Sigma}_E = \begin{pmatrix} 1.25871 & 0.17644 & -0.08826 & 0.26711 & -0.13247 \\ 0.17644 & 2.33691 & 0.00497 & -0.15278 & 0.28849 \\ -0.08826 & 0.00497 & 2.97993 & -0.03929 & -0.07036 \\ 0.26711 & -0.15278 & -0.03929 & 3.76133 & -0.74181 \\ -0.13247 & 0.28849 & -0.07036 & -0.74181 & 4.57260 \end{pmatrix},
$$

$$
(18)
$$

$$
\widehat{\Phi}_E = \begin{pmatrix} -0.05655 & 0.09316 & -0.00092 & 0.11812 & 0.16943 \\ 0.46112 & 0.18679 & 0.13921 & -0.09552 & -0.13973 \\ 0.02891 & 0.37594 & -0.01426 & 0.05658 & 0.18949 \\ -0.04417 & -0.03156 & 0.45267 & -0.0236 & -0.10665 \\ -0.05809 & 0.02705 & 0.11909 & 0.33560 & 0.02602 \\ -0.02400 & 0.00105 & -0.08313 & 0.02921 & 0.34976 \\ 0.11188 & 0.08832 & -0.08980 & 0.04434 & 0.34897 \end{pmatrix}.
$$

$$
(19)
$$

The acceptance rate for the Metropolis step employed for sampling of Σ from the posterior conditional on other parameters and data is 24%.

Table 1. Posterior average loss of the estimates in the Example.

	EL_1	EL_2	$EL = E(L_1 + L_2)$
MLE (Φ_{MLE}, Σ_{MLE})	11.2940	22.6213	33.9154
Posterior mean (Φ_{Mean}, Σ_{Mean})	7.5414	17.7350	25.2764
Intrinsic Bayes estimator (Φ_E, Σ_E)	7.9295	13.6053	21.5348

Notes: The loss functions are $L_1(\widehat{\Sigma}, \Sigma) = (T/2)\{tr(\Sigma \widetilde{\Sigma}^{-1}) - \log |\Sigma \widetilde{\Sigma}^{-1}|$ $- p\}$, and $L_2(\widehat{\Sigma}, \widehat{\Phi}, \Sigma, \Phi) = \frac{1}{2}tr[\widehat{\Sigma}^{-1}\{(\Phi - \widehat{\Phi})'E_{(X|\Phi,\Sigma)}(X'X)(\Phi - \widehat{\Phi})\}]$. The posterior average loss for L_1 and L_2 are computed by (15) and (16).

Now we compare the Bayes estimator with the MLE and posterior mean. The posterior mean obtained using Option 1 is

$$\widehat{\Sigma}_{Mean} = \begin{pmatrix} 1.17992 & 0.16657 & -0.08449 & 0.25618 & -0.12700 \\ 0.16657 & 2.20489 & 0.00476 & -0.14659 & 0.27587 \\ -0.08449 & 0.00476 & 2.81882 & -0.03801 & -0.06557 \\ 0.25618 & -0.14659 & -0.03801 & 3.57382 & -0.71561 \\ -0.12700 & 0.27587 & -0.06557 & -0.71561 & 4.35069 \end{pmatrix},$$
(20)

$$\widehat{\Phi}_{Mean} = \begin{pmatrix} -0.0599 & 0.09731 & -0.00193 & 0.12276 & 0.17558 \\ 0.43287 & 0.18741 & 0.14185 & -0.09236 & -0.13907 \\ 0.02772 & 0.35030 & -0.01270 & 0.05628 & 0.19118 \\ -0.04650 & -0.03116 & 0.42790 & -0.02733 & -0.10462 \\ -0.05718 & 0.02602 & 0.12006 & 0.31275 & 0.02437 \\ -0.02373 & -0.00041 & -0.08285 & 0.02851 & 0.32792 \\ 0.11167 & 0.08778 & -0.08838 & 0.04414 & 0.34923 \end{pmatrix}.$$
(21)

It is known that the posterior mean of Σ minimises the expected posterior loss of L_1. But it does not minimise the intrinsic loss because the estimator of Σ also influences the weight of Φ related loss in L_2. Table 1 reports the average posterior loss of the MLE, posterior mean, and the Bayes estimator for the data sample generated in the example. The Bayes estimator improves L_2-related risk with a tradeoff of larger L_1-related risk. Table 1 shows that the Bayes estimator induces lower posterior risk than the posterior mean by making the L_2-related risk substantially lower and the L_1-related risk only slightly higher. Both posterior mean and the Bayes estimator dominate the MLE.

5.2. An empirical study: seasonal effects in a macroeconomic model

We now turn to an empirical application of the Bayesian estimation under the entropy loss. We estimate an LTS consisting of seasonal dummies and four macroeconomic variables: the return of Standard and Poor 500 stock price index (which represents weighted stock prices of large companies), the 3-month Treasury Bill rate, the growth rate of payroll (including government jobs as well as private sector jobs), and the growth rate of industrial production (in that order). All series are measured in percentage terms. The series are monthly data from 1970:1 to 2002:12 and are not seasonally-adjusted. The payroll data are obtained from the Bureau

of Labor Statistics, the rest series are obtained from the Federal Reserve Board. There are 12 column dummy variables, representing January to December.

The role of seasonal fluctuations in business cycles has been noted by a number of economists. Barky and Miron (1989) argued that for the U.S. economy the characteristics of seasonal fluctuations are similar to the conventional characterisations of business cycles. Cecchetti et al. (1997) estimated production function of various industries based on how their responses to seasonal shocks vary with the state of business cycle. Ghysels (1988) showed that univariate seasonal adjustment of endogenous variables is not harmless because information on the interactions among endogenous variables will be lost. Miron and Beaulieu (1996) provided a survey on econometric and economic issues on understanding business cycles through seasonal fluctuations. In the finance literature, numerous studies argue that stock returns appear to have a seasonal components. Rozeff and Kinney (1976) documented a celebrated "turn of the year" effect, which refers to the seemingly abnormally high returns in January and July, especially for stocks with small-market capitalisations. A number of theories have been developed to explain the phenomenon. Reinganum (1983) attributed the high stock return in January to the end of year tax-loss selling in December. Chang and Pinegar (1989) found that industrial production trails the seasonal movement of stock returns by one month. The reported point estimates of the seasonal effects in the literature are model dependent and based on OLS or MLE. We will estimate the seasonal effects. Our primary interest lies in comparison of posterior mean with Bayes estimate under the entropy loss.

In this section, we employ an LTS model

$$y_t' = (x_t', y_{t-1}', \ldots, y_{t-L}')\Phi + \epsilon_t,$$
(22)

for $t = 1, \ldots, T$. The dimension of the VAR variable y_t is four. The exogenous variable $x_{0t} = (x_{0t,1}, \ldots, x_{0t,12})$ is a 12-dimensional vector representing seasonal cycles, with $x_{0i,1}$ equals 1 if period i is January, and 0 otherwise; $x_{0i,12} = 1$ if period i is December and 0 otherwise.

Based on the Schwarz criterion, for each sample period the lag length L of the LTS is 2. The Yang–Berger reference prior is applied to the covariance matrix Σ. The prior for the LTS coefficient ϕ is a rather diffuse $N(0, M_0)$. Here M_0 is a diagonal matrix with 10.0 being the diagonal element for parameters corresponding to the dummy variables and 2.0 being the diagonal element for parameters corresponding to the lag coefficients. We draw the posterior from M MCMC cycles after $0.1 \times M$ burn-in runs. The MCMC length M is set at 50,000, 100,000, and 1,000,000.

Under algorithm MCMC, reducing the length of MCMC cycles to 100,000 or 50,000 from 1,000,000 makes little difference. The MLE, posterior mean, and

Bayes estimate of the covariance matrix Σ are as follows. As dictated by the theoretical result, the Bayes estimate under the entropy loss, Σ_E is larger than the posterior mean Σ_{Mean}.

$$\widehat{\Sigma}_{\text{MLE}} = \begin{pmatrix} 19.775 & -0.335 & -0.094 & 0.059 \\ -0.335 & 0.254 & 0.055 & 0.011 \\ -0.094 & 0.055 & 0.937 & 0.070 \\ 0.059 & 0.011 & 0.070 & 0.044 \end{pmatrix}.$$

With $M = 1,000,000$, the posterior mean and entropy-based Bayes estimates are

$$\widehat{\Sigma}_{\text{Mean}} = \begin{pmatrix} 20.908 & -0.354 & -0.100 & 0.063 \\ -0.354 & 0.271 & 0.058 & 0.012 \\ -0.100 & 0.058 & 0.996 & 0.075 \\ 0.063 & 0.012 & 0.075 & 0.048 \end{pmatrix}.$$

$$\widehat{\Sigma}_E = \begin{pmatrix} 22.729 & -0.384 & -0.101 & 0.078 \\ -0.384 & 0.295 & 0.057 & 0.012 \\ -0.101 & 0.057 & 1.074 & 0.081 \\ 0.078 & 0.012 & 0.081 & 0.052 \end{pmatrix}.$$

The posterior standard deviations of the elements of the covariance matrix are

$$\begin{pmatrix} 1.538 & 0.124 & 0.235 & 0.052 \\ 0.124 & 0.020 & 0.027 & 0.006 \\ 0.235 & 0.027 & 0.074 & 0.012 \\ 0.052 & 0.006 & 0.012 & 0.003 \end{pmatrix}.$$

The difference between estimates $\widehat{\Sigma}_{\text{Mean}}$ and $\widehat{\Sigma}_E$ is large relative to the posterior standard deviations.

The above point estimates and standard deviations are similar to those with $M = 50,000$. The MCMC algorithm yields posteriors with few outliers. This is because in Step 2 the MCMC algorithm, ϕ is generated from an average of sample data and generated data, instead of sample data alone. A few outliers in the posterior may affect the posterior mean slightly but can change the posterior risk and the Bayesian estimate substantially because the few explosive parameters carry disproportionately large weights in the posterior average loss.

The intrinsic Bayes estimate dominates the posterior mean by a large margin in terms of posterior expected loss. The large difference in the posterior expected loss is mainly due to the difference in the Φ-related risk, i.e. the quadratic term in (5). This difference in risk is approximately

$$\frac{1}{2}\text{tr}\Big[\widehat{\Sigma}_E^{-1}\text{E}\{(\Phi - \widehat{\Phi}_E)'W$$
$$\times (\Phi - \widehat{\Phi}_E) \mid Y\}\widehat{\Sigma}_E^{-1}(\widehat{\Sigma}_E - \widehat{\Sigma}_{\text{Mean}})\Big],$$

which is proportional to the frequentist expectation W, and the latter is comparable to $X'X$. $X'X$ is quite large in this application, largely due to the strong serial correlation of the 3-month Tbill rates. As a result, with a larger $\widehat{\Sigma}$ the Bayesian estimate substantially reduces the posterior risk, compared with the posterior mean. Simulation with $M = 1,000,000$ shows that the posterior average loss of the posterior mean estimate (3230.6) is larger than that of the intrinsic Bayes estimate (69.7). The lower overall posterior average loss of the intrinsic Bayes estimate is achieved by substantially lowering the risk of the quadratic part, from 3225.3 for the posterior mean to 61.8. The first term of the loss related to Σ of the intrinsic Bayes estimate is slightly larger (7.9) compared to that under the posterior mean estimate (5.3). As is noted earlier, the intrinsic Bayes estimate improves Φ-related loss with a tradeoff of larger Σ-related loss. The empirical result shows that the Bayesian estimate induces lower posterior average loss than the posterior mean by making the Φ-related loss substantially lower and the Σ-related loss only slightly higher.

We now turn to compare the estimates of the regression coefficient Φ. Table 2 reports the MLE, posterior mean, the intrinsic Bayes estimate, and posterior standard deviations. As a consequence of applying a rather diffuse prior, the posterior mean Φ_{Mean} is quite similar to the MLE. For stock returns of Standard and Poor 500 index, MLE and the posterior mean estimate indicate a moderate positive seasonal factor in January, which is smaller than the seasonal factor of March, April, October, November, and December. Most surprisingly, October registers the largest seasonal gain, despite the fact that the sample included the 1987 October sell off. The data of recent years suggest that the estimates of seasonality in large capitalisation stock returns are quite sensitive to the sample period and regression model. In comparison to the MLE and the posterior mean, the intrinsic Bayesian estimate shows a smaller January effect and much smaller end-of-the-year positive seasonal returns. The sum of the seasonal coefficients of the MLE is above 11% while that of the intrinsic Bayesian estimates is about half as much. The large discrepancy between the posterior mean and the intrinsic Bayesian estimate s casts doubt on the robustness of the seasonality of returns of Standard and Poor 500 stock.

Compared to the stock return, the seasonality of industrial production growth rate is much more robust. The most distinct pattern is a steep decline in July followed by a surge in August; then weakness in the end of the year precedes a strong rebound in February. The strong showing of industrial production in February and August is consistent with the pattern reported in Chang and Pinegar (1989) while the predicted industrial production by Standard and Poor stock returns is quite small. The magnitude of the seasonal effects by the entropy-loss-based Bayesian estimates is on average slightly larger than that of the posterior mean.

Lastly, we examine the estimates of the employment growth rate equation. The most prominent seasonal patterns are the decline in January followed by a rebound in February and March and the weakness in

Table 2. Estimates of three equations.

	Stock return equation			IP growth equation			Payroll growth equation		
	MLE	PMean (SD)	BE	MLE	PMean(SD)	BE	MLE	PMean (SD)	BE
Jan	1.507	1.469 (1.093)	0.986	0.701	0.701 (0.243)	0.680	−2.019	−2.019 (0.053)	−2.054
Feb	0.737	0.743 (2.344)	0.784	3.980	3.961 (0.532)	3.977	0.842	0.840 (0.117)	0.892
Mar	1.780	1.749 (2.355)	1.662	0.179	0.186 (0.537)	0.123	0.912	0.912 (0.118)	0.955
Apr	1.534	1.461 (1.170)	0.905	−1.286	−1.281 (0.262)	−1.416	0.518	0.518 (0.057)	0.466
May	0.370	0.304 (1.381)	−0.356	−0.138	−0.134 (0.312)	−0.226	0.485	0.485 (0.068)	0.414
June	0.242	0.195 (1.412)	−0.444	2.092	2.095 (0.318)	2.003	0.455	0.455 (0.070)	0.381
July	0.637	0.585 (1.338)	−0.019	−4.589	−4.584 (0.298)	−4.759	−1.284	−1.284 (0.066)	−1.352
Aug	0.822	0.753 (1.830)	0.485	5.270	5.257 (0.414)	5.251	0.262	0.261 (0.091)	0.255
Sept	−1.881	−1.846 (1.798)	−1.964	1.021	1.025 (0.405)	0.998	1.040	1.041 (0.089)	1.049
Oct	2.162	2.072 (1.377)	1.583	−0.409	−0.403 (0.308)	−0.614	0.155	0.155 (0.067)	0.106
Nov	1.891	1.815 (1.240)	1.271	−1.332	−1.33 (0.278)	−1.464	−0.011	−0.011 (0.061)	−0.075
Dec	1.473	1.419 (1.149)	0.884	−0.792	−0.792 (0.257)	−0.843	0.017	0.017 (0.056)	−0.031
stock1	−0.012	−0.011 (0.052)	0.025	0.013	0.013 (0.011)	0.012	0.000	0.000 (0.002)	0.001
int1	−0.581	−0.580 (0.384)	−0.518	0.152	0.153 (0.084)	0.132	0.030	0.030 (0.018)	0.025
ip1	−0.198	−0.201 (0.243)	−0.208	0.052	0.052 (0.054)	0.076	0.021	0.021 (0.012)	0.021
emp1	0.292	0.310 (1.017)	0.543	1.072	1.063 (0.231)	1.100	0.218	0.217 (0.051)	0.255
stock2	−0.040	−0.039 (0.052)	−0.002	0.009	0.009 (0.011)	0.009	0.000	0.000 (0.002)	0.000
int2	0.589	0.594 (0.386)	0.587	−0.207	−0.207 (0.085)	−0.176	−0.036	−0.036 (0.019)	−0.028
ip2	−0.382	−0.373 (0.241)	−0.365	−0.087	−0.087 (0.053)	−0.064	0.046	0.046 (0.012)	0.045
emp2	0.008	0.019 (0.973)	0.198	−0.038	−0.036 (0.222)	−0.006	0.183	0.183 (0.049)	0.219

Notes: stock lag 1, int lag 1, ip lag 1, and emp lag 1 represent the first lag of stock return, the first lag 3-month Tbill rate, the first lag of industrial production growth, and the first lag of private sector payroll growth. The lag 2 variables are the second lag of the corresponding variables. Estimates are obtained with MCMC cycle size $M = 1,000,000$.

July followed by a recovery in September. The estimated seasonality in payroll growth is somewhat different from that of the industrial production growth. Note that in year 2002 about 83% of the payroll consists of service sector jobs while 85% of industrial production concerns the manufacture sector. The subject of interest is the point estimates. Similar to the industrial production equation, the entropy-loss-based Bayes estimates for the payroll growth rate equation are similar to the posterior mean.

In summary, the Bayesian estimates based on the entropy loss show a qualitatively similar seasonal pattern to that of the posterior mean estimates for industrial production and employment growth but a distinctly different one for stock returns. The posterior average loss of the Bayes estimates with respect to the entropy loss is substantially smaller than that of the posterior mean.

6. Concluding remarks

In this paper we investigate properties of Bayes estimators of LTS model $(\mathbf{\Phi}, \mathbf{\Sigma})$ derived from the entropy loss function. These estimators are distinctly different from the multivariate iid model because of the serial correlation of the time series variables. Bayesian computation under the entropy loss requires simulating a frequentist moment of the regressors. We propose a data-augmenting algorithm for simulation of posteriors and computation of Bayes estimators under the entropy loss and a normal-reference prior. The algorithm that draws from the full conditional posterior is shown to be quite efficient. A novel approach taken in this paper concerns generating data in an MCMC as latent parameters. This idea may be useful for other contexts for simulating complicated posterior moments. Our empirical application to a macroeconomic problem shows that the Bayes estimates under the entropy loss can differ substantially from the posterior mean.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr. Shawn Ni holds a PhD in Economics from University of Minnesota. He is currently Middlebush Professor of Economics and Adjunct Professor of Statistics at University of Missouri-Columbia. He conducts research on a wide range of empirical economics topics and Bayesian statistics.

Dr. Dongchu Sun holds a PhD in Statistics from Purdue University. He is a research professor of statistics at the University of Nebraska-Lincoln and East China Normal University. His research interests includes Bayesian analysis, small area estimation, decision theory, business and econometrics, space-time and longitudinal models, and smoothing splines.

References

Barky, R. B., & Miron, J. A. (1989). The seasonal cycle and the business cycle. *The Journal of Political Economy*, *97*(3), 503–534. https://doi.org/10.1086/261614

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

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of Royal Statistical Society Ser. B*, *41*, 113–147.

Bernardo, J. M., & Juárez, M. A. (2003). Intrinsic estimation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith & M. West (Eds.), *Bayesian statistics* (Vol. 7, pp. 465–476). Oxford University Press.

Cecchetti, S. G., Kashyap, A. K., & Wilcox, D. W. (1997). Interactions between the seasonal and business cycles in

production and inventories. *American Economic Review*, *87*, 884–892.

Chang, E. C., & Pinegar, M. J. (1989). Seasonal fluctuations in industrial production and stock market seasonals. *The Journal of Financial and Quantitative Analysis*, *24*(1), 59–74. https://doi.org/10.2307/2330748

Fernandez-Villaverde, J., & Rubio-Ramirez, J. F. (2004). Comparing dynamic equilibrium economies to data. *Journal of Econometrics*, *123*(1), 153–187. https://doi.org/10.1016/j.jeconom.2003.10.031

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409. https://doi.org/10.1080/01621459.1990.10476213

Ghysels, E. (1988). A study toward a dynamic theory of seasonality for economic time series. *Journal of the American Statistical Association*, *83*(401), 168–172. https://doi.org/10.1080/01621459.1988.10478583

Harville, D. A. (1998). *Matrix algebra from a statistician's perspective*. Taylor & Francis Group.

Kitamura, Y., & Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, *65*(4), 861–874. https://doi.org/10.2307/2171942

MacKinnon, J. G., & Smith, A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, *85*(2), 205–230. https://doi.org/10.1016/S0304-4076(97)00099-7

Miron, J. A., & Beaulieu, J. J. (1996). What have macroeconomists learned about business cycles from the study of seasonal cycles?. *The Review of Economics and Statistics*, *78*(1), 54–66. https://doi.org/10.2307/2109847

Ni, S., & Sun, D. (2003). Noninformative priors and frequentist risks of Bayesian estimators of vector-autoregressive models. *Journal of Econometrics*, *115*(1), 159–197. https://doi.org/10.1016/S0304-4076(03)00099-X

Ni, S., Sun, D., & Sun, X. (2007). Intrinsic Bayesian estimation of vector autoregression impulse responses. *Journal of Business and Economic Statistics*, *25*(2), 163–176.https://doi.org/10.1198/073500106000000378

Reinganum, M. R. (1983). The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects. *Journal of Financial Economics*, *12*(1), 89–104.https://doi.org/10.1016/0304-405X(83)90029-6

Robert, C. P. (1994). *The Bayesian choice*. Springer-Verlag.

Robert, C. P. (1996). Intrinsic losses. *Theory and Decision*, *40*(2), 191–214. https://doi.org/10.1007/BF00133173

Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking*, *37*(3), 383–401. https://doi.org/10.1353/mcb.2005.0034

Rozeff, M. S., & Kinney, W. R. (1976). Capital market seasonality: The case of stock returns. *Journal of Financial Economics*, *3*(4), 379–402. https://doi.org/10.1016/0304-405X(76)90028-3

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48. https://doi.org/10.2307/1912017

Solo, V., Purdon, P., Weisskoff, R., & Brown, E. (2001). A signal estimation approach to functional MRI. *IEEE Transactions on Medical Imaging*, *20*(1), 26–35. https://doi.org/10.1109/42.906422

Sun, D., & Berger, J. O. (1998). Reference priors under partial information. *Biometrika*, *85*, 55–71. https://doi.org/10.1093/biomet/85.1.55

Sun, D., & Ni, S. (2004). Bayesian analysis of VAR models with noninformative priors. *Journal of Statistical Planning and Inference*, *121*(2), 291–309. https://doi.org/10.1016/S0378-3758(03)00116-2

Tanner, M., & Wang, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458

Yang, R., & Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, *22*(3), 1195–1211. https://doi.org/10.1214/aos/1176325625

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. John Wiley & Sons.

Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss approach. *Journal of Econometrics*, *8*(2), 127–158. https://doi.org/10.1016/0304-4076(78)90024-6

Zellner, A. (1998). The finite sample properties of simultaneous equations estimates and estimators: Bayesian and non-Bayesian approaches. *Journal of Econometrics*, *83*(1–2), 185–212. https://doi.org/10.1016/S0304-4076(97)00069-9

Appendix. Proof of Fact 2.1 and posterior properties

A1 Proof of Fact 2.1

Proof: Let $(\tilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$ denote an arbitrary estimator of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For the entropy loss function L and posterior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{Y})$, the expected posterior loss is

$$R(\tilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}} \mid \boldsymbol{Y}) = \frac{T}{2}\mathrm{E}\Big\{(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \mathrm{tr}(\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}^{-1}) - \log|\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}^{-1}| - p\Big\}.$$

The Bayes estimator, which minimises the expected posterior loss, can be derived through conditions on first-order derivatives. Note that because $\widetilde{\boldsymbol{\Sigma}}$ is symmetric,

$$\frac{\partial R(\tilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})}{\partial \tilde{\boldsymbol{\mu}}} = 2\mathrm{E}\left(\widetilde{\boldsymbol{\Sigma}}(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu})\right).$$

Let the derivative be 0 yields $\hat{\boldsymbol{\mu}}_{\mathrm{iid}} = \mathrm{E}(\boldsymbol{\mu} \mid \boldsymbol{Y})$.

The following identities are known (e.g. Harville, 1998, p. 327) for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\frac{\partial \log(|\boldsymbol{A}|)}{\partial \boldsymbol{A}} = 2\boldsymbol{A}^{-1} - \mathrm{diag}(\boldsymbol{A}^{-1}),$$

$$\frac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^{-1})}{\partial \boldsymbol{B}} = -2(\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}) + \mathrm{diag}(\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}^{-1}).$$

Here $\mathrm{diag}(\boldsymbol{A})$ is a diagonal matrix, whose diagonal elements are these from $\boldsymbol{A}$. Using the conclusion that the estimator for $\boldsymbol{\mu}$ is the posterior mean, we have $\mathrm{E}\{(\hat{\boldsymbol{\mu}}_{\mathrm{iid}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}}_{\mathrm{iid}} - \boldsymbol{\mu})' \mid \boldsymbol{Y}\} = \mathrm{var}(\boldsymbol{\mu} \mid \boldsymbol{Y})$. Taking this result to the derivative with respect to $\widetilde{\boldsymbol{\Sigma}}$, we have

$$\frac{\partial R(\hat{\boldsymbol{\mu}}_{\mathrm{iid}}, \widetilde{\boldsymbol{\Sigma}})}{\partial \widetilde{\boldsymbol{\Sigma}}}$$

$$= \mathrm{E}\frac{T}{2}\Big\{-2\widetilde{\boldsymbol{\Sigma}}^{-1}\mathrm{var}(\boldsymbol{\mu}|\boldsymbol{Y})\widetilde{\boldsymbol{\Sigma}}^{-1} + \mathrm{diag}(\widetilde{\boldsymbol{\Sigma}}^{-1}\mathrm{var}(\boldsymbol{\mu}|\boldsymbol{Y})\widetilde{\boldsymbol{\Sigma}}^{-1})$$

$$-2\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}^{-1} + \mathrm{diag}(\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}^{-1}) + 2\widetilde{\boldsymbol{\Sigma}}^{-1} - \mathrm{diag}(\widetilde{\boldsymbol{\Sigma}}^{-1})\Big\}.$$

$$= \frac{T}{2}\Big\{2\widetilde{\mathbf{\Sigma}}^{-1}(\mathbf{I} - [\mathrm{E}(\mathbf{\Sigma}\mid \mathbf{Y}) + \mathrm{var}(\boldsymbol{\mu}\mid \mathbf{Y})]\widetilde{\mathbf{\Sigma}}^{-1}) +$$

$$- \mathrm{diag}(\widetilde{\mathbf{\Sigma}}^{-1}(\mathbf{I} - [\mathrm{E}(\mathbf{\Sigma}\mid \mathbf{Y}) + \mathrm{var}(\boldsymbol{\mu}\mid \mathbf{Y})]\widetilde{\mathbf{\Sigma}}^{-1}))\Big\}.$$

The derivative is 0 when

$$\mathbf{I} = [\mathrm{E}(\mathbf{\Sigma}\mid \mathbf{Y}) + \mathrm{var}(\boldsymbol{\mu}\mid \mathbf{Y})]\widetilde{\mathbf{\Sigma}}^{-1},$$

which yields $\widehat{\mathbf{\Sigma}}_{\mathrm{iid}} = \mathrm{E}(\mathbf{\Sigma}\mid \mathbf{Y}) + \mathrm{var}(\boldsymbol{\mu}\mid \mathbf{Y})$. ∎

A2 Posterior properties

A2.1 $(\theta\mid X^*, X)$ under normal prior for Φ and Jeffreys type prior for Σ

A commonly used noninformative prior for $\mathbf{\Sigma}$ is the Jeffreys prior $\pi_J(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(p+1)/2}$. The prior for $\mathbf{\Sigma}$ in the RATS statistical package is a modified version of the Jeffreys prior, $\pi_A(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-(L+1)p/2-1}$. Zellner's maximum data informative (MDI) prior, $\pi_M(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-1/2}$. For analysis on prior choice in VAR models see Ni and Sun (2003), and Sun and Ni (2004). We consider a class of joint priors,

$$\pi_b(\boldsymbol{\phi}, \mathbf{\Sigma}) = \pi_N(\boldsymbol{\phi})\pi_b^*(\mathbf{\Sigma}), \tag{A1}$$

where $\pi_N(\boldsymbol{\phi})$ is the normal prior for $\boldsymbol{\phi}$ given by (10), and $\pi_b^*(\mathbf{\Sigma})$ ($b \in \mathbb{R}$) is given by

$$\pi_b^*(\mathbf{\Sigma}) \propto \frac{1}{|\mathbf{\Sigma}|^{b/2}}. \tag{A2}$$

Note that π_{NJ}, π_{NA}, and π_{NM} are special cases of (A2) when b equals to $p+1$, $(L+1)p+2$ and 1, respectively.

We propose using the posterior quantities conditional on the simulated data X^*, $h(\theta \mid X^*, X)$, instead of the marginal posterior, $h(\theta \mid X)$, as the estimator of $h(\theta)$. Note that the posterior

$$f(\mathbf{\Phi}, \mathbf{\Sigma}\mid X^*, X) \propto \frac{1}{|\mathbf{\Sigma}|^{T+b/2}}$$

$$\times \mathrm{etr}\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\mathbf{\Phi})\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{\Phi})'\right\}$$

$$\times \mathrm{etr}\left\{-\frac{1}{2}(\mathbf{Y}^* - \mathbf{X}^*\mathbf{\Phi})\mathbf{\Sigma}^{-1}(\mathbf{Y}^* - \mathbf{X}^*\mathbf{\Phi})'\right\}$$

$$\times \exp\left\{-\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)'\mathbf{M}_0^{-1}(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\right\}.$$

We would like to express the posterior of $\mathbf{\Sigma}$ in terms of X^*. Integrating out $\mathbf{\Phi}$ results in

$$f(\mathbf{\Sigma}\mid X^*, X)$$

$$= \frac{|\mathbf{M}_0 + \mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}|^{1/2}}{|\mathbf{\Sigma}|^{T+b/2}}$$

$$\times \mathrm{etr}\left\{-\frac{1}{2}\mathbf{\Sigma}^{-1}[(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{\Phi}}_M)'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{\Phi}}_M)\right.$$

$$+ (\mathbf{Y}^* - \mathbf{X}^*\widehat{\mathbf{\Phi}}_M^*)'(\mathbf{Y}^* - \mathbf{X}^*\widehat{\mathbf{\Phi}}_M^*)]\Big\}$$

$$\times \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)'[\mathbf{M}_0 + \mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X}\right.$$

$$+ \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}]^{-1}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_0)\Big\}$$

$$\times \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\phi}}_M - \hat{\boldsymbol{\phi}}_M^*)'\mathbf{\Sigma}^{-1} \otimes [(\mathbf{X}'\mathbf{X})^{-1}\right.$$

$$+ (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}]^{-1}(\hat{\boldsymbol{\phi}}_M - \hat{\boldsymbol{\phi}}_M^*)\Big\},$$

where the lower case defines the vec operator: $\boldsymbol{\phi} = \mathrm{vec}(\mathbf{\Phi})$, $\widetilde{\mathbf{\Phi}} = (\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}(\mathbf{X}'\mathbf{Y} + \mathbf{X}^{*\prime}\mathbf{Y}^*)$, $\widehat{\mathbf{\Phi}}_M = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and $\widehat{\mathbf{\Phi}}_M^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^*$.

The posterior $f(\mathbf{\Phi}, \mathbf{\Sigma}\mid X^*, X)$ has a closed form when the prior for $\boldsymbol{\phi}$ can be written as $N(\boldsymbol{\phi}_0, \mathbf{M}_0)$ with

$$\mathbf{M}_0 = \mathbf{\Sigma} \otimes \mathbf{\Omega}_0, \tag{A3}$$

where $\mathbf{\Omega}_0$ is a $Lp+1$ by $Lp+1$ known covariance matrix. In the extreme case of $\mathbf{\Omega}_0^{-1} \to 0$, $\mathbf{M}_0^{-1} \to 0$ and the prior approaches a constant prior. Under the assumption (A3), we have

$$f(\mathbf{\Sigma}\mid X^*, X) \propto |\mathbf{\Omega}_0 + (\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}|^{p/2}$$

$$|\mathbf{\Sigma}|^{-T - \frac{b-Lp-1}{2}}\mathrm{etr}\left(-\frac{1}{2}\mathbf{V}\mathbf{\Sigma}^{-1}\right),$$

where

$$\mathbf{V}(X, X^*)$$

$$= (\widehat{\mathbf{\Phi}}_M - \widehat{\mathbf{\Phi}}_M^*)'[(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}](\widehat{\mathbf{\Phi}}_M - \widehat{\mathbf{\Phi}}_M^*)$$

$$+ (\widehat{\mathbf{\Phi}} - \mathbf{\Phi}_0)'[\mathbf{\Omega}_0 + (\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}]^{-1}(\widetilde{\mathbf{\Phi}} - \mathbf{\Phi}_0)$$

$$+ (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{\Phi}}_M)'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{\Phi}}_M)$$

$$+ (\mathbf{Y}^* - \mathbf{X}^*\widehat{\mathbf{\Phi}}_M^*)'(\mathbf{Y}^* - \mathbf{X}^*\widehat{\mathbf{\Phi}}_M^*).$$

$f(\mathbf{\Sigma}\mid X^*, X) \sim IW\{\mathbf{V}, 2T + b - (L+1)p - 2\}$. The mean of $\mathbf{\Sigma}\mid X^*, X$ is $\mathbf{V}(X, X^*)/(2T + b - (L+1)p - 2 - p - 1) = \mathbf{V}(X, X^*)/(2T + b - (L+2)p - 3)$. The posterior mean $\mathbf{\Sigma}\mid X$ is estimated by $\sum_{j=1}^M(\mathbf{\Sigma}\mid X_j^*, X)/M$.

The marginal posterior of $f(\mathbf{\Phi}\mid X^*, X)$ can be obtained by integrating out $\mathbf{\Sigma}$ in $f(\mathbf{\Phi}, \mathbf{\Sigma}\mid X^*, X)$. It is easy to verify that

$$f(\mathbf{\Phi}, \mathbf{\Sigma}\mid X^*, X) \propto |\mathbf{\Sigma}|^{-T-b/2}\mathrm{etr}\left(-\frac{1}{2}\mathbf{U}\mathbf{\Sigma}^{-1}\right)$$

$$= |\mathbf{\Sigma}|^{-(\nu+p+1)/2}\mathrm{etr}\left(-\frac{1}{2}\mathbf{U}\mathbf{\Sigma}^{-1}\right), \tag{A4}$$

where the degree of freedom $\nu = 2T + b - p - 1$ and

$$\mathbf{U} = (\mathbf{\Phi} - \mathbf{\Phi}^u)'(\mathbf{\Omega}_0^{-1} + \mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)(\mathbf{\Phi} - \mathbf{\Phi}^u) + \mathbf{V}(X, X^*), \tag{A5}$$

$$\mathbf{\Phi}^u = (\mathbf{\Omega}_0^{-1} + \mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}(\mathbf{\Omega}_0^{-1}\mathbf{\Phi}_0 + \mathbf{X}'\mathbf{Y} + \mathbf{X}^{*\prime}\mathbf{Y}^*). \tag{A6}$$

It follows from (A4) that

$$f(\mathbf{\Phi}\mid X^*, X) \propto |\mathbf{U}|^{-\nu/2}$$

$$= \big|(\mathbf{\Phi} - \mathbf{\Phi}^u)'(\mathbf{\Omega}_0^{-1}$$

$$+ \mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^*)(\mathbf{\Phi} - \mathbf{\Phi}^u) + \mathbf{V}(X, X^*)\big|^{-\nu/2}, \tag{A7}$$

which follows a matrix version of the Student-t distribution. The mean of $\mathbf{\Phi}\mid X^*, X$ is $\mathbf{\Phi}^u$. To calculate the intrinsic Bayes estimator, note that the frequentist expectation $\widehat{W}$ can be estimated by $\sum_{j=1}^M X_j^{*\prime}X_j^*/M$. The posterior mean $\mathrm{E}(W\mathbf{\Phi}\mid X)$ is estimated by $\sum_{j=1}^M X_j^{*\prime}X_j^*\mathbf{\Phi}(X_j^*, X)/M$.

A2.2 Conditional posteriors under normal prior for Φ and the Yang-Berger reference prior for Σ

Fact A.1: *Consider the normal prior for ϕ given in* (10). *The conditional density of ϕ given $(\Sigma; Y)$ is $N_J(\mu_M, V_M)$, where*

$$\mu_M = \hat{\phi}_{\text{MLE}} + \{M_0^{-1} + \Sigma^{-1} \otimes (X'X)\}^{-1} M_0^{-1}(\phi_0 - \hat{\phi}_{\text{MLE}}); \tag{A8}$$

$$V_M = \{M_0^{-1} + \Sigma^{-1} \otimes (X'X)\}^{-1}, \tag{A9}$$

where

$$\hat{\phi}_{\text{MLE}} = \text{vec}\{(X'X)^{-1}X'Y\}. \tag{A10}$$

Fact A.2: *Consider the normal prior for ϕ given in* (10). *The conditional density of ϕ given $(\Sigma; Y, Y^*)$ is $\sim N(\phi_M^*, V_M^*)$, where*

$$V_M^* = \{M_0^{-1} + \Sigma^{-1} \otimes (X'X + X^{*'}X^*)\}^{-1},$$

$$\phi_M^* = V_M^*[\{\Sigma^{-1} \otimes (X'X + X^{*'}X^*)\}\hat{\phi}_{\text{MLE}}^* + M_0^{-1}\phi_0];$$

$$\hat{\phi}_{\text{MLE}}^* = \text{vec}\left\{\left(X'X + X^{*'}X^*\right)^{-1}\left(X'Y + X^{*'}Y^*\right)\right\}.$$

Fact A.3: *The conditional density of Σ given (ϕ, Y) is*

$$\pi(\Sigma|\Phi, Y) \propto \frac{\text{etr}\{-\frac{1}{2}\Sigma^{-1}S(\Phi)\}}{|\Sigma|^{T/2+1} \prod\limits_{1 \leq i < j \leq p}(d_i - d_j)}, \tag{A11}$$

where $S(\Phi) = (Y - X\Phi)'(Y - X\Phi)$.

Fact A.4: *The conditional density of Σ given (ϕ, Y, Y^*) is*

$$\pi(\Sigma|\Phi, Y, Y^*) \propto \frac{\text{etr}\{-\frac{1}{2}\Sigma^{-1}(S(\Phi) + S^*(\Phi))\}}{|\Sigma|^{T+1} \prod\limits_{1 \leq i < j \leq p}(d_i - d_j)}, \tag{A12}$$

where $S^*(\Phi) = (Y^* - X^*\Phi)'(Y^* - X^*\Phi)$.

We have shown how simulated data facilitate computation of a frequentist moment in the Bayes estimator. In the appendix, we will show that the simulated data can also be used to reduced the variance of MCMC, making the simulation more efficient.

For simulation of Σ, we adopt a hit-and-run algorithm used in Yang and Berger (1994). In implementing the algorithm, we consider a one-to-one transformation $\Sigma^\# = \log(\Sigma)$, or $\Sigma = \exp(\Sigma^\#)$ in the sense that

$$\Sigma = \sum_{j=0}^{\infty} \frac{(\Sigma^\#)^j}{j!}.$$

The reason for simulating Σ as $\exp(\Sigma^\#)$ is to ensure the generated Σ matrices are positive definite. It can be shown that the conditional posterior density of $\Sigma^\#$ given (Φ, Y) is

$$\pi(\Sigma^\#|\Phi, Y)$$

$$= \pi(\Sigma^\#|S(\Phi)) \propto \frac{\text{etr}\{-\frac{T}{2}|\Sigma^\#| - \frac{1}{2}(\exp\Sigma^*)^{-1}S(\Phi)\}}{\prod_{i<j}(d_i^\# - d_j^\#)}, \tag{A13}$$

and that the conditional posterior density of Σ^* given (ϕ, Y, Y^*) is then

$$\pi(\Sigma^\#|\Phi, Y, Y^*) = \pi(\Sigma^\#|S(\Phi), S^*(\Phi))$$

$$\propto \frac{\text{etr}[-T|\Sigma^\#| - \frac{1}{2}(\exp\Sigma^*)^{-1}\{S(\Phi) + S^*(\Phi)\}]}{\prod_{i<j}(d_i^\# - d_j^\#)}, \tag{A14}$$

where $\Sigma^\# = O'D^\#O$, O is an orthogonal matrix, $D^\# = \text{diag}(d_1^\#, \ldots, d_p^\#)$, with $d_1^\# \geq \ldots \geq d_p^\#$. Note that $\exp(\Sigma^\#) = O'\exp(D^\#)O$.