Forecasting with Vector Autoregressions Using Bayesian Variable Selection Methods: Comparison of Direct and Iterated Methods

Ryukyu Economics Working Paper Series, University of the Ryukyus
琉球大学 経済学ワーキングペーパー

Forecasting with Vector Autoregressions Using Bayesian Variable Selection Methods: Comparison of Direct and Iterated Methods

Katsuhiro Sugita

Working Paper No. REWP#02

May 2019

Forecasting with Vector Autoregressions Using Bayesian Variable Selection Methods: Comparison of Direct and Iterated Methods

Katsuhiro Sugita

Faculty of Global and Regional Studies, University of the Ryukyus, Nishihara,

Senbaru, Okinawa, 903-0213, Japan

E-Mail: ksugita@ll.u-ryukyu.ac.jp

Abstract

This paper compares multi-period forecasting performances by direct and iterated method using a Bayesian vector autoregressions with the stochastic search variable selection (SSVS) priors. The forecasting performances are evaluated using the artificially generated data with both nonstationary and stationary process. In theory direct forecasts are more efficient asymptotically and more robust to model misspecification than iterated forecasts, and iterated forecasts tend to bias but more efficient if the one-period ahead model is correctly specified. From the results of the Monte Carlo simulations, iterated forecasts tend to outperform direct forecasts, particularly with longer lag model and with longer forecast horizons. Implementing SSVS prior generally improves forecasting performance over unrestricted VAR model for either nonstationary or stationary data. As an illustration, US macroeconomic data sets with three variables are examined to compare iterated and direct forecasts using the unrestricted VAR model and the SSVS VAR model. Overall, iterated forecasts using model with the SSVS generally best outperform, suggesting that the SSVS restrictions on insignificant parameters alleviates over-parameterized problem of VAR in one-step ahead forecast and thus offers an appreciable improvement in forecast performance of iterated forecasts.

1 Introduction

Vector Autoregressive (VAR) models have been widely used to forecast macroeconomic variables and to analyze macroeconomics and policy. For one-period ahead forecasting, one has to just estimate the model. However, it is often the case that more than one-period forecasting is of interest. In making a multi-period forecast, there are two methods - direct forecast model and iterated forecast model, and there have been several theoretical research in terms of which method is better for multi-period forecasting such as Bhansali (1996, 1997), Clements and Hendry (1996), Kang (2003), Chevillon and Hendry (2005), Ing (2003), and among others. These literature tend to conclude that direct forecasts are more robust to model specification and efficient asymptotically, and thus the direct forecast method is preferable compared with the iterated forecast method, while the iterated forecast method can be more efficient only if the one-period ahead model is correctly specified. However, some empirical research show that iterated forecasts outperform direct forecasts. Ang et al. (2006) find that the iterated forecasts of U.S. GDP growth perform better than the direct forecasts. Marcellino et al. (2006) show that iterated forecasts outperform direct forecasts, especially with longer lag and longer forecast horizon, using 170 U.S. monthly macroeconomic time series for either univariate or multivariate models. Pesaran et al. (2011) state that whether direct or iterated method is better in multi-period forecasting depends upon the sample size, forecast horizon, the underlying DGP, and the methods used to select lag length for the model, and thus it is ultimately an empirical matter.

For multivariate VAR models, there exists over-parameterization problem, which leads to imprecise inference and thus deteriorates the forecast performance. To deal with this dimensionality problem, this paper examines a Bayesian stochastic search variable selection (SSVS) method for VAR models proposed by George et al. (2008), and compares multi-period forecasting performance between the direct and iterated forecast models, instead of a factor-augmented VAR approach that Pesaran et al. (2011) propose. The SSVS method, developed by George and McCulloch (1993) and George and Mcculloch (1997), uses a hierarchical prior where each of the parameters in the model is drawn in a Markov chain Monte Carlo (MCMC) from two different normal distributions - one with a small variance and the other with a large variance. There have been several research that applied the SSVS method to various multivariate time series models by Hara and Sillanp (2009), Jochmann et al. (2010), Jochmann et al. (2013) and Koop (2013). Korobilis (2013) applies the SSVS to not only linear VAR model but also nonlinear VAR models that allow for structural breaks or time-varying parameters to investigates the forecasting performance by using the SSVS with the iterated forecast model, finding that the SSVS does improve forecast performance.

George et al. (2008) investigate numerical simulations and show that implementing SSVS method in VAR can be effective at both selecting a satisfactory model and improving forecast performance based on the one-step ahead mean squared error of forecast error and Kullback-Liebler divergence. In this paper, numerical simulations are conducted using both stationary and nonstationary DGPs to evaluate forecasting performances with 2, 4, 8 and 12-step ahead horizons, and compute the mean squared forecast error (MSFE) to compare direct forecasts with iterated forecasts using restricted SSVS VAR model and unrestricted VAR models. Iterated forecasts are found to outperform direct forecasts for both unrestricted VAR models and the SSVS VAR model, particularly with long-lag model and with long forecasting horizon. Implementing SSVS in VAR is found to be generally improves forecasting performance appreciably. With relatively long lag length and thus a large number of parameters in the model, it seems that SSVS can effectively restrict insignificant parameters in the model and thus improve forecasting performance.

The plan of this paper is as follows. In Section 2 multi-period forecasting using VAR model is described, and method to evaluate forecasting performances. Section 3 reviews prior and posterior distributions of SSVS VAR model. Section 4 illustrates numerical experiments with artificially generated data, and then examines the results of the numerical simulations. Section 5 illustrates an application to a simple 3-variable VAR of US macroeconomics. Section 6 concludes and suggests for future work. All results reported in this paper are generated using Ox version 7.2 for Linux (see Doornik (2013)).

2 Iterated and direct Multi-period Forecasts for VAR Models

This section describes iterated and direct forecasting methods for VAR models. Let y_t be an $n\times 1$ vector of observations at time t, then a VAR model with p lag is written as

$$y_t' = \mu' + \sum_{i=1}^{p} y_{t-i}'\Theta_i + \varepsilon_\tau' \tag{1}$$

for $t = 1,\dots,T$, where μ is a $n\times 1$ vector of an intercept term; Θ_i are $n\times n$ matrices of coefficients for $i = 1,\dots,p$; ε_t are $n\times 1$ independent $N_n(0,\Sigma)$ errors; and the covariance matrix Σ is an $n\times n$ positive definite matrix. The VAR model in (1) can be written as

$$\mathbf{y}_t = \mu + \Theta\mathbf{y}_{t-1} + \varepsilon_t \tag{2}$$

where $\mathbf{y}_t = (y_t',...,y_{t-p+1}')'$, $\mu = (\mu',0,...,0)'$, $\varepsilon_t = (\varepsilon_t',0,..,0)'$ and

$$\Theta = \begin{bmatrix} \Theta_1, ..., \Theta_{p-1} & \Theta_p \\ I_{n(p-1)} & 0_{n(p-1)\times n(p-1)} \end{bmatrix}$$

The one-step ahead forecast $\hat{y}'_{t+1|t}$ of the VAR model is obtained by estimating the parameters in (1) as $\hat{y}'_{t+1|t} = \hat{\mu}' + \sum_{i=1}^{p} y'_{t+1-i|t}\hat{\Theta}_i$. To make forecasting further than one-period ahead into the future, there are two methods for making multi-period forecasts - iterated forecasts and direct forecasts methods. Iterated forecasts for the h-period forecasts are obtained recursively as

$$\hat{y}'_{t+h} = \hat{\mu}' + \sum_{i=1}^{p} \hat{y}'_{t+h-i}\hat{\Theta}_i \tag{3}$$

where $\hat{y}_{i|t} = y_j$ for $j \leq t$, thus the iterated forecasts can be computed as

$$\hat{\mathbf{y}}_{t+h} = \sum_{i=0}^{h-1} \hat{\Theta}^i_{(I)}\hat{\mu}_{(I)} + \hat{\Theta}^h_{(I)}\mathbf{y}_{t-1}. \tag{4}$$

Direct forecasts for the multi-period forecasting are obtained by estimating the model

$$y'_t = \mu + \Theta y'_{t-h-1} + \varepsilon_t, \tag{5}$$

then using the estimated coefficients directly to make the forecast of

$$\hat{\mathbf{y}}_{t+h} = \hat{\mu}_{(D)} + \hat{\Theta}_{(D)}\mathbf{y}_{t-1} \tag{6}$$

Thus, the relative forecast accuracy depends on how accurate $\hat{\Theta}_{(I)}$ and $\hat{\Theta}_{(D)}$ are estimated. If $\hat{\Theta}_{(I)}$ is badly estimated with large errors, then its powered values diverge increasingly from $\Theta_{(I)}$. Since the iterated method depends on one-period ahead coefficients $\hat{\Theta}_{(I)}$, the direct method is preferable when the one-period ahead model is not specified correctly. Chevillon and Hendry (2005) evaluate the asymptotic and finite-sample properties of direct forecasting method, and show that, compared with iterated method, the direct method is more efficient asymptotically, more precise in finite samples and more robust against model misspecification. The theoretical advantages of the direct forecasting method over the iterated method are shown by Bhansali (1996, 1997), Clements and Hendry (1996), Kang (2003), and Ing (2003) among others. However, Marcellino et al. (2006) evaluates a large-scale empirical comparison of iterated and direct forecasts using U.S. macroeconomic time series data, and find that iterated forecasts tend to have smaller MSFEs than direct forecasts, contrary to the theoretical preference of direct forecasts.

To evaluate the forecasting performances among several different models, the mean squared forecast error (MSFE) is the most widely used. Let $y'_{\tau+h}$ is a vector of observations at time $\tau+h$ for $\tau = \tau_0, ..., T - h - 1$, and $h = 2$, 4, 8 and 12- step ahead forecasts. Then, $\hat{\Phi} = \left(\hat{\mu}', \hat{\Theta}'_1, \ldots, \hat{\Theta}'_p\right)'$ is estimated for both the direct and iterated method, using information up to $\tau - 1$ to forecast values $\hat{y}_{\tau+h}$ starting from $\tau = \tau_0$ up to $\tau = T - h - 1$, and calculate the MSFE defined as:

$$\text{MSFE} = \frac{1}{T - h - \tau_0 + 1} \sum_{\tau=\tau_0}^{T-h} \left[y_{\tau+h} - \hat{y}_{\tau+h}|\hat{\Phi}, Y_{\tau-1}\right]^2. \tag{7}$$

where $Y_{\tau-1} = (X_{\tau-1}, X_{\tau-2}, \ldots, X_1)$.

MSFE is the point forecasts, and a standard frequentist criterion for forecast evaluation. For Bayesian forecast comparison, predictive likelihoods are also used to evaluate the forecast performance since they provide more comprehensive forecast with the entire predictive density (see Corradi and Swanson (2006) for review). Jochmann et al. (2010), Koop (2013) and others use the predictive likelihood to measure the forecasting performance. However, this paper do not consider the predictive likelihood to measure the forecast performance as Andersson and Karlsson (2007) point out that the predictive likelihood has problem in measuring of forecasting performance. If the forecast error is small, the predictive likelihood prefers the model with small variance, but the model with larger variance is favored if the forecast error is larger than what can be expected from the model with smaller variance.

In comparison of iterated and direct forecasts, it is of interest whether the SSVS VAR model can correctly identify the one-step ahead model since the efficiency of iterated forecasts depend upon the accuracy of $\hat{\Theta}_{(I)}$. Since true value Φ is known a priori for $h = 1$ for Monte Carlo simulation in Section 4, the mean squared error (MSE) of forecast error can be used to examine whether the SSVS prior help to improve the one-step ahead forecast, particularly when the lag is larger than the true lag. The one-step ahead forecast error at period τusing information up to τ-1 can be decomposed into tow parts such as:

$$y_\tau - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1} = (y_\tau - \hat{y}_\tau|\Phi) + \left(\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}\right) \tag{8}$$

The first term in the right hand side in (8), $y_\tau - \hat{y}_\tau|\Phi$, is the sampling error, and the second term, $\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}$, is the forecasting error caused by the deviation of the estimates $\hat{\Phi}$ from the true parameters Φ. For the comparison of forecasting performances among different models, the sampling error, $y_\tau - \hat{y}_\tau|\Phi$, is common to all models under consideration and does not depend upon $\hat{\Phi}$, so that the forecasting error, $\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}$, can be used for evaluation of forecasting performances among different models. The one-step ahead MSE of the forecast error is calculated as;

$$\text{MSE} = \frac{1}{T - \tau_0 + 1} \sum_{\tau=\tau_0}^{T} \left[\left(\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}\right)' \left(\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}\right) \right] \tag{9}$$

Note that this one-step ahead MSE of the forecast error is also for all variables by summing MSE for each variable.

3 VAR model with SSVS prior

3.1 SSVS prior

This section presents a VAR model with SSVS prior, proposed by George et al. (2008). Without any restriction on the regression coefficients and the covariance matrix in (1), VAR models typically contain a very large number of parameters to estimate relative to the number of the observations. This over-parameterization problem leads to imprecision of inference and thus worsens forecasting performance. To overcome this problem, George et al. (2008) implement the SSVS method in a VAR, based on George and McCulloch (1993) and George and Mcculloch (1997). SSVS is a Bayesian MCMC method to take restrictions on the parameters of the model by using a hierarchical prior on the parameters, In this paper, George et al. (2008) method is followed to investigate forecasting performances relative to non-SSVS methods such as the unrestricted VAR and the Minnesota prior model.

Let x_t be a $(1+np) \times 1$ vector with $x_t = (1, y_{t-1}, ..., y_{t-p})'$, then the VAR model (1) can be rewritten in matrix form as

$$Y = X\Phi + \varepsilon \tag{10}$$

where the $T \times n$ matrix Y is defined as $Y = (y_1, ..., y_T)'$; the $T \times (1+np)$ matrix X is defined as $X = (x_1, ..., x_T)'$; the $(1+np) \times n$ matrix Φ is defined as $\Phi = (C', \Theta_1', ..., \Theta_p')'$; and the ε is a $T \times n$ matrix with $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$.

SSVS defines the prior for the VAR coefficient Φ not as a whole but as all of the elements in Φ. Let $\phi = vec(\Phi)$ and m be the number of restricted elements in Φ, then the prior for each element, ϕ_j,

$j = 1, 2, \ldots, m$, is a hierarchical prior with mixture of two normal distributions conditional on an unknown dummy variable γ_j that takes zero or one:

$$\phi_j|\gamma_j \sim (1-\gamma_j)N(0, \tau_{0,j}^2) + \gamma_j N(0, \tau_{1,j}^2) \tag{11}$$

where $\tau_{0,j}^2$ is small when $\gamma_j = 0$ and $\tau_{1,j}^2$ is large when $\gamma_j = 1$. This implies that if $\gamma_j = 0$, that is, the element ϕ_j is restricted to be zero, the prior for ψ_j is virtually zero, while if $\gamma_j = 1$, that is, the element ϕ_j is unrestricted, the prior is relatively noninformative. With regard to priors on γ_j, SSVS assumes independent Bernoulli $p_i \in (0,1)$ random variables:

$$\begin{aligned} P(\gamma_j = 1) &= p_j \\ P(\gamma_j = 0) &= 1 - p_j \end{aligned} \tag{12}$$

where p_j is the prior hyperparameter that is set to be 0.5 for a natural default choice. Let $\gamma = (\gamma_1, \ldots, \gamma_m)$, then the prior for ϕ in (11) can be written as:

$$\phi|\gamma \sim N(0, DD) \tag{13}$$

where D is a diagonal matrix as $D = diag(d_1, \ldots, d_m)$; where

$$d_j = \begin{cases} \tau_{0j} & \text{if } \gamma_j = 0 \\ \tau_{1j} & \text{if } \gamma_j = 1 \end{cases} \tag{14}$$

George and Mcculloch (1997) and George et al. (2008) recommend to use a default semiautomatic approach that sets $\tau_{kj} = c_k \hat{\sigma}_{\phi_j}$ for $k = 0, 1$, where $\hat{\sigma}_{\phi_j}$ is the least squares estimates of the standard error of ϕ_j, which is the coefficients in an unrestricted VAR. The pre-selected constants c_0 and c_1 must be $c_0 < c_1$. George et al. (2008), Jochmann et al. (2010) and Jochmann et al. (2013) set $c_0 = 0.1$ and $c_1 = 10$, however these numbers should be adjusted by researcher to obtain an optimal forecasting performance. George and McCulloch (1993) recommend these numbers to be such that $\tau_{k1}^2/\tau_{k0}^2 \leq 10000$, otherwise the MCMC would be very slow to converge if τ_{k1}/τ_{k0} is chosen extremely large.

SSVS also considers the restrictions on the covariances in Σ. Let Ψ be a $n \times n$ upper-triangular matrix, the error covariance matrix can be decomposed as:

$$\Sigma^{-1} = \Psi\Psi' \tag{15}$$

where the upper-triangular matrix Ψ can be obtained by the Choleski decomposition of Σ and expressed as:

$$\Psi = \{\psi_{ij}\} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1n} \\ 0 & \psi_{22} & \cdots & \psi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{nn} \end{bmatrix} \tag{16}$$

For the diagonal elements of Ψ, let define $\psi = (\psi_{11}, \psi_{22}, \ldots, \psi_{nn})'$, then prior for each element of ψ is assumed as:

$$\psi_{ii}^2 \sim G(a_i, b_i) \tag{17}$$

where $G(a_i, b_i)$ denotes the gamma distribution with mean a_i/b_i and variance a_i/b_i^2. For the elements above the diagonal, let define $\eta_j = (\psi_{1j}, \psi_{2j}, \ldots, \psi_{j-1,j})'$ for $j = 2, \ldots, n$, and $\eta = (\psi_{12}, \psi_{13}, \psi_{23}, \ldots, \psi_{n-1,n})' = (\eta_2', \ldots, \eta_n')'$, then priors on η is assumed as:

$$\eta_j|\omega_j \sim N(0, D_j D_j) \tag{18}$$

where $\omega_j = \left(\omega_{1j}, \ldots, \omega_{j-1,j}\right)'$ is a vector of dummy variables which are assumed to be independent Bernoulli as:

$$\begin{aligned} P(\omega_{ij} = 1) &= q_{ij} \\ P(\omega_{ij} = 0) &= 1 - q_{ij} \end{aligned} \tag{19}$$

where q_{ij} is equal to 0.5 for a natural default choice. D_j in (18) is defined as $D_j = diag(\left(h_{1j}, \ldots, h_{j-1,j}\right)$, where

$$h_{ij} = \begin{cases} \kappa_{0ij} & \text{if } \omega_{ij} = 0 \\ \kappa_{1ij} & \text{if } \omega_{ij} = 1 \end{cases}. \tag{20}$$

The choice of κ_{kij} for $k = 0, 1$ can be determined by using a semiautomatic default approach that is similar considerations for setting τ_{kj}, that is, we set $\kappa_{kij} = c_k \hat{\sigma}_{\psi_{ij}}$ with values of $c_0 < c_1$, where $\hat{\sigma}_{\psi_{ij}}$ is an estimate of the standard error associated with off-diagonal element of Ψ. With this prior, SSVS also considers restrictions on the off-diagonal elements of Ψ, and thus each element of η_j is a mixture of two normal distributions so that

$$\psi_{ij}|\omega_{ij} \sim (1 - \omega_{ij}) N(0, \kappa_{0ij}^2) + \omega_{ij} N(0, \kappa_{1ij}^2) \tag{21}$$

where κ_{0ij}^2 is small when $\omega_{ij} = 0$ and κ_{1ij}^2 is large when $\omega_{ij} = 1$.

This summarizes the SSVS hierarchical prior for VAR model. George et al. (2008) and Jochmann et al. (2010) consider three patterns of SSVS - restrictions only for Ψ and for Φ separately, and then for both. In this paper, stochastic search for both of Ψ and Φ together is considered. Note that, if the unknown indicator parameters is set to equal to 1, that is $\gamma_j = \omega_{ij} = 1$ for all j and i, then the SSVS VAR is just unrestricted VAR.

3.2 Posteriors

With the SSVS priors that described above, the conditional posterior distribution for each parameter is obtained. Following by George et al. (2008), let s_{ij} be the elements of $S = (Y - X\Phi)'(Y - X\Phi)$, $s_j = \left(s_{1j}, \ldots, s_{j-1,j}\right)'$, and S_j be the upper left $j \times j$ block of S, then the likelihood function is

$$\begin{aligned} \mathscr{L}(Y|\Phi, \Sigma) &\propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}(Y - X\Phi)\Sigma^{-1}(Y - X\Phi)'\right] \\ &\propto \prod_{i=1}^{n} \psi_{ii}^T \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n} \psi_{ii}^2 \upsilon_i + \sum_{j=2}^{n} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)' S_{j-1} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)\right\}\right] \end{aligned} \tag{22}$$

where $\upsilon_1 = s_{11}$ and $\upsilon_i = |S_i|/|S_{i-1}|$ for $i = 2, \ldots, n$.

With the likelihood function (22) and priors of (12), (13), (17), (18), (19) and (21), George et al. (2008) derive the conditional posterior distributions as follows. For the VAR coefficients Φ, the conditional posterior is given as:

$$\Phi|\gamma, \eta, \psi; Y \sim N_m(\mu, \Xi) \tag{23}$$

where

$$\Xi = \left[\left(\Psi\Psi'\right) \otimes \left(X'X\right) + (DD)^{-1}\right],^{-1}$$
$$\mu = \Xi\left[\left\{(\Psi\Psi)' \otimes \left(X'X\right)\right\}\hat{\psi}_M\right],$$
$$\hat{\psi}_M = vec(\hat{\Psi}_M) = vec\left[\left(X'X\right)^{-1}X'Y\right].$$

Let define $\gamma_{(-i)} = (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, , \gamma_m)$, then the conditional posterior of γ can be obtained as,

$$\gamma_j|\phi, \gamma_{j-1}, \eta, \psi; Y \sim Bernoulli\left(u_{j1}/(u_{j1}+u_{j2})\right) \tag{24}$$

where

$$u_{j1} = \frac{1}{\tau_{0j}}\exp\left(-\frac{\phi_j^2}{2\tau_{0j}^2}\right)p_i,$$
$$u_{j2} = \frac{1}{\tau_{1j}}\exp\left(-\frac{\phi_j^2}{2\tau_{1j}^2}\right)(1-p_i).$$

The conditional posterior distributions of $\psi_{11}^2, \psi_{22}^2, \dots, \psi_{nn}^2$ are independent and gamma distributions as:

$$\psi_{jj}^2|\phi, \omega; Y \sim G(a_j + \frac{T}{2}, b_j) \tag{25}$$

where

$$b_j = \begin{cases} b_1 + \frac{s_{11}}{2} & \text{if } j = 1 \\ b_j + \frac{1}{2}\left\{s_{jj} - s_j'\left[V_{j-1} + (D_jD_j)^{-1}\right]^{-1}s_j\right\} & \text{if } j = 2, \dots, n \end{cases}$$

The conditional posterior distributions of $\eta_2, \dots, \eta_n$ are independent and given as;

$$\eta_j|\phi, \omega, \psi; Y \sim N_{j-1}(\mu_j, \Delta_j) \tag{26}$$

where

$$\Delta_j = \left[S_{j-1} + (D_jD_j)^{-1}\right]^{-1},$$

$$\mu_j = -\psi_{jj}\Delta_j s_j.$$

Finally, the conditional posterior distribution of ω_{ij} for $j = 2, \dots, n$ and $i = 1, \dots, j-1$ is derived as:

$$\omega_{ij}|\phi, \psi, \omega_k, k \neq j; Y \sim Bernoulli\left(u_{ij1}/(u_{ij1}+u_{ij2})\right) \tag{27}$$

where

$$u_{ij1} = \frac{1}{\kappa_{1ij}}\exp\left(-\frac{\psi_{ij}^2}{2\kappa_{1ij}^2}\right)q_{ij},$$
$$u_{ij2} = \frac{1}{\kappa_{0ij}}\exp\left(-\frac{\psi_{ij}^2}{2\kappa_{0ij}^2}\right)(1-q_{ij}).$$

The MCMC stochastic search algorithm is obtained by drawing sequentially the above conditional distributions (23) - (27).

4 Monte Carlo Simulation

This section presents Monte Carlo simulations to illustrate forecasting performances for both iterated and direct forecast methods using VAR models. Two data generating processes (DGPs) are considered: one follows non-stationary process and the other follows stationary process. For each DGPs, 100 samples of size $T = 150$ were simulated, and then for each sample, three types of estimators are compared: (1) MLE; (2) Minn, which is based on Bayesian VAR model with Minnesota prior; (3) SSVS. With regard to the Minnesota prior, Litterman (1986) proposes shrinkage prior for a Bayesian VAR model with random walk components. For a VAR model with p-the lag in (1), the Minnesota prior for the coefficients assumes that the importance of the lagged variables is shrinking with the lag length, so that the prior is tighter around zero with lag length such that $\Theta_i \sim N(\bar{\Theta}_i, V(\Theta_i))$ where the expected values of Θ_i is defined as $\bar{\Theta}_1 = I_n$ and $\bar{\Theta}_2 = \cdots = \bar{\Theta}_p = 0_n$, and the variance of Θ_i is given as:

$$V(\Theta_i) = \frac{\lambda^2}{i^2} \begin{bmatrix} 1 & \theta \hat{\sigma_1}^2 / \hat{\sigma_2}^2 & \cdots & \theta \hat{\sigma_1}^2 / \hat{\sigma_n}^2 \\ \theta \hat{\sigma_2}^2 / \hat{\sigma_1}^2 & 1 & \cdots & \theta \hat{\sigma_2}^2 / \hat{\sigma_n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \theta \hat{\sigma_n}^2 / \hat{\sigma_1}^2 & \theta \hat{\sigma_n}^2 / \hat{\sigma_2}^2 & \cdots & 1 \end{bmatrix}, \tag{28}$$

where $0 < \theta < 1$, and $\Sigma = diag\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right)$. In our simulations, we set $\lambda = 0.05$ and $\theta = 0.1$. Note that these parameters should be adjusted for optimal results.

The following two DGPs for VARs are considered for this experiment. Both DGPs contain intercept term. DGP 1 is a four-variable VAR with four lags, containing unit roots with parameters

$$\text{DGP1: } \Phi^{(DGP1)} = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \\ 0.8 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.3 \\ 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix}, \text{ and } \Psi^{(DGP1)} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Next, DGP 2 is also a four-variable VAR with four lags, but stationary data with parameters

$$
\text{DGP 2: } \Phi^{(DGP2)} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.6 & 0 & 0 & 0 \\ -0.3 & 0.6 & 0 & 0 \\ 0 & -0.3 & 0.6 & 0 \\ 0 & 0 & -0.3 & 0.6 \\ 0 & 0 & 0 & 0 \\ -0.2 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & 0 \\ 0 & 0 & -0.2 & 0 \\ -0.3 & 0 & 0 & -0.2 \\ 0 & -0.3 & 0 & 0 \\ 0 & 0 & -0.3 & 0 \\ 0 & 0 & 0 & -0.3 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix}, \text{ and } \Psi^{(DGP2)} = \Psi^{(DGP1)}.
$$

With regard to Ψ of these DGPs, partial correlation between the i-th component and the j-th component of the error term for $i, j > 1$ in (16) is assumed to be zero. The prior for the intercept term is not restricted and assigned a normal with a zero mean and a variance of 50 for relatively non-informative prior. For the SSVS priors, the hyperparameters are set at $p_i = 0.5$ in (12), $q_i = 0.5$ in (19), $a = b = 0.01$ in (17), $c_0 = 0.1$ and $c_1 = 50$ for τ_{kj} and κ_{kij} in (14) and (20). Each DGP is repeated 100 times to obtain 100 samples.

As for determination of the lag length p, three different methods are used as (1) $p = 4$ (fixed); (2) $p = 8$ (fixed); and (3) p chosen by the AIC with $0 \leq p \leq 12$. The first method of the fixed lag length as $p = 4$ is the true lag length. We do not use the BIC for the lag length determination since the BIC is generally choose short lag length, and the use of SSVS means that short lag model is not required to consider. For the selection of lag by the AIC, the AIC is computed at each date τ, where $\tau_0 \leq \tau \leq T - h$, based on the one-step ahead regression for the iterated forecasts, and on the h-step ahead regression (6) for the direct forecasts. For each τ in (7), MCMC is run with 20,000 draws after 5,000 burn-in from $\tau = \tau_0$ up to $\tau = T - h - 1$ to compute the MSFEs in (7) for each estimators by a recursive forecasting exercise of both an iterated and a direct multi-period forecasting method.

First, the average one-step ahead forecasting MSE (9) over 100 samples are calculated to see whether the SSVS can correctly identify the one-step ahead model compared with other estimators. Table 1 shows the results of the Monte Carlo simulation using both DGP 1 and 2 with $T = 150$ for three models labeled by MLE, Minn (VAR with Minnesota prior) and SSVS by computing pseudo-out-of sample forecasts $\hat{y}_{\tau+1}$ for $\tau = 80$ to $\tau = 150 - 1$. The table shows that SSVS estimator performs substantially better than the MLE and the Minn, suggesting that the SSVS can effectively restrict the insignificant elements in the models. The MLE is seemed to be quite sensitive in the lag length. The average MSE by the MLE deteriorates more than twice when the lag length increases to $p = 8$ from $p = 4$. Compared with the MLE, both the Minn and the SSVS seem to restrict the insignificant parameters to be zero in longer lag terms.

Next, the Monte Carlo simulations for the multi-step forecasting are examined. Table 2 summarizes the MSFEs of both iterated and direct forecasts methods with forecast horizon 2-, 4-, 8-, and 12-steps ahead. The MSFE in the table are the sum of the MSFE for each variable. For all series, pseudo-out-of sample forecasts $\hat{y}_{\tau+h}$ are computed for $\tau = 80$ to $\tau = 150 - h - 1$, and calculate the MSFE defined as (7). Each figure in Table 2 is the average over 100 sample MSFEs. Inspection of Table 2 suggests

1. Among the three estimators by the MLE, the Minn and the SSVS, the SSVS produces the lowest MSFE in most cases, though in a very few case of direct forecasts the Minn shows barely better

performances than the SSVS.

2. The forecast performances by SSVS prior tends to be insensitive to the choice of the lag length, while the MLE estimator considerably deteriorates the performances as the lag length is longer. That is, even if the lag length is more than 4 (that is the true lag length), the SSVS treats the coefficients on longer lags to be zero, while the forecast performances of other two models are largely depend upon the selection of the lag length. The Minnesota prior effectively provides shrinkage in parameters of the longer lags.

3. For the MLE and the SSVS, the iterated method of forecasts are better than the direct method, though for the Minn the results by iterated method are better for DGP2 than those by the direct method, but in some cases worse for DGP1.

4. For these DGPs, the SSVS model with iterated forecast performs best for any forecast horizon.

Table 3 illustrates the distributions of the relative MSFE, that is the ratios of the MSFE of the direct forecast to the MSFE of the iterated forecast for different forecast horizons, MSFE(direct)/MSFE(iterated). The table shows the mean, standard deviations, 95% highest posterior density intervals (HPDI) of the relative MSFE, and pr.(<1), which is probability that the ratio is less than 1 (the direct forecasts performs better than the iterated forecasts). The following results are found:

1. For the MLE and the SSVS, the mean values of the relative MSFEs are always more than 1 (means that the iterated forecasts outperform the direct forecasts), while for the Minn the ratios are either more or less than 1.

2. For the MLE and the SSVS, the mean values of the relative MSFEs are getting large as the forecast horizons are longer, meaning that the relative performance of the iterated forecasts improves with the forecast horizon.

3. The MSFE ratios by the MLE are quite sensitive to the choice of the lag length. As the lag length is longer, the relative MSFEs by the MLE are getting larger. However, the relative MSFE by the SSVS are not affected by the choice of the lag length due to the insensitivities of the SSVS to the lag length.

4. For all three estimators, the standard deviations of the relative MSFE are larger as the forecasts horizon is longer.

5. Except for the Minn, the probability that the ratio is less than 1 is generally smaller with long-lag length.

6. In the case of the Minn with non-stationary data, direct forecasts tend to have lower MSFEs than iterated forecasts, though with stationary data, the results are opposite as iterated forecasts are better performance than direct forecasts.

These findings show that for given DGPs the SSVS also has almost same properties in forecasting as the MLE, as suggested by Marcellino et al. (2006). That is, forecasting performance by the SSVS also shows that the iterated forecast tends to outperform the direct forecast, especially with long-lag and longer forecast horizon. For the Minn, the results are ambiguous. Since the Minnesota prior set its prior mean for the coefficients on the first own lag to be 1 and other coefficients to be zero, the Minnesota prior prone to produce misspecified parameter estimates.

5 An empirical analysis

For an empirical example of comparison of the direct and iterated forecasts using Bayesian VAR models, this section considers multivariate model of US macroeconomics that uses three variables - unemployment rate, inflation rate and interest rate. A VAR model that uses these variables has been analyzed by Cogley and Sargent (2005), Primiceri (2005), Koop et al. (2009), and Jochmann et al. (2010), among many others. Our US data are quarterly, from 1953:I to 2017:III with sample size $T = 259$. Unemployment rate is measured by the civilian unemployment rate, inflation by the 400 times the difference of the log of CPI, which is the GDP chain-type price index, and interest rate by the 3-month treasury bill. These data are obtained from the Federal Reserve Bank of St. Louise.[1], and are plotted in Figure 1.

The selection of the number of lags in a VAR affects efficiency in estimation and thus forecasting performances. Cogley and Sargent (2005) and Primiceri (2005) work with VAR(2) to analyze US macroeconomy with the three variables without mentioning any particular reason how the lag length is chosen. Jochmann et al. (2010) use VAR(4) for their SSVS VAR model because the SSVS can find zero restrictions on the parameters of longer lags even if the true lag length is less than 4. However, the true lag length might be more than 4. With our data set, the number of lags is scattered depending upon which criterion we use - VAR(10) by the AIC, VAR(4) by the Hannan-Quinn criterion, and VAR(2) by the BIC. Even if the true lag length is less than 10, the SSVS can set zero restrictions on the longer lags, thus we consider VAR (12) and VAR(AIC) that the lag length is chosen by the AIC. Forecast horizons are 2, 4, 8 and 12-period ahead. We work with a recursive forecasting exercise using both direct and iterated multistep forecasting method, with data up to time $\tau - 1$, where $\tau = \tau_0, \ldots, T-h-1$, and $\tau_0 = 80$.

Table 4 presents the MSFEs (7) for the three-variable VAR with the lag-length 12 and chosen by the AIC for the MLE, the Minn and the SSVS estimators. For any forecast horizon, iterated forecasts have lower MSFE than direct forecasts. With enough long lag length of 12, the SSVS improves the forecast performance among other methods. However, with the lag selected by the AIC, almost half of the MSFEs by VAR with the Minnesota prior have the lowest MSFEs. Compared with the fixed lag length of twelve, the lag length chosen by the AIC is shorter than 12, and the MSFEs are smaller than the MSFEs by the models with lag length 12. This indicates that the lag length 12 may be too long, containing unnecessary lags or parameters, though the SSVS is supposed to restrict insignificant coefficients to be zero. This indicates that SSVS is effective to ensure parsimony in over-parameterized VAR(12) model.

The three variables used in this empirical analysis appear to be nonstationary, and thus transformation to stationary data by taking their first difference is also considered. For this case, the forecasting models are estimated using Δy_t instead of y_t in (1), then these models are used to compute the forecast of the level of y_{t+h}. such as $\hat{y}_{t+h} = y_t + \sum_{i=1}^{h} \Delta\hat{y}_{t+i|t}$ for the iterated forecast, and $\hat{y}_{t+h} = y_t + \Delta\hat{y}_{t+h}$ for the direct forecasts. All elements of the prior mean for the Minnesota prior are set to be zero as $\bar{\Theta}_1 = 0_n$ since all series are transformed to be stationary by the first differencing. Table 5 presents the MSFEs of the case of the first differenced data. The MSFEs of the first differenced data tend to have lower MSFEs than the case of the level data, particularly for the inflation rates. These result also indicate that iterated forecasts outperform direct forecasts and the SSVS improves forecast performances than other models though the Minn produces better in some cases.

6 Conclusion

This paper examines comparison of direct and iterated multistep forecasting performance using three estimators for VAR model - the MLE, the Minnesota prior and the SSVS prior. Theoretically direct method is more preferable since the direct forecasts are prone to be efficient and more robust to model misspecification. Iterated forecasts are more efficient if the one-step ahead model is not misspecified.

[1] https://fred.stlouisfed.org

Since George et al. (2008) show VAR with SSVS prior greatly improves the one-step ahead forecast, the coefficients are estimated more efficiently and thus an iterated multi-period forecast method would be more efficient than the direct method. So, it is of interest if direct forecasts are compared with iterated forecasts using SSVS VAR model. Although Pesaran et al. (2011) noted that whether the direct or iterated method produce better forecasts is ultimately an empirical question, this paper considers the case of three estimators of VAR for comparison of direct and iterated method using two DGPs and US macroeconomics data. The results are exactly same as Marcellino et al. (2006), that is, iterated forecasts for the MLE and SSVS estimators have lower MSFEs than direct forecasts, particularly if the models are with long-lag and longer forecast horizon, while it is ambiguous for the case of the Minnesota prior. The SSVS estimator tends to appreciably improve the forecast performance against other estimators by the MLE and the Minnesota prior in most cases.

As an empirical example an application of US macroeconomics is studied to show a benefit of using SSVS prior in a VAR. With longer lags and thus large number of parameters that may include many insignificant, it seems that SSVS alleviates over-parameterization problem in VAR model by restricting insignificant parameters of the model, and enables to improve forecasting performance, although the Minnesota prior also produces shorter MSFEs in some case than SSVS since the Minnesota prior provides shrinkages on the longer lags.

With these results, iterated forecasts are found to produce better forecast performances than direct forecasts, and Bayesian method such as the Minnesota prior model and the SSVS model outperform the MLE, particularly with longer lag and longer forecast horizon.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K03661.

References

Andersson, M. K. and Karlsson, S. (2007). "Bayesian forecast combination for VAR models." In *Advances in Econometrics*, 501–524.

Ang, A., Piazzesi, M., and Wei, M. (2006). "What does the yield curve tell us about GDP growth?" *Journal of Econometrics*, 131(1-2): 359–403.

Bhansali, R. (1997). "Direct autoregressive predictors for multistep prediction: Order selection and performance relative to the plug in predictors." *Statistica Sinica*, 7: 425–449.
URL `http://www.jstor.org/stable/24306087`

Bhansali, R. J. (1996). "Asymptotically efficient autoregressive model selection for multistep prediction." *Annals of the Institute of Statistical Mathematics*, 48(3): 577–602.

Chevillon, G. and Hendry, D. F. (2005). "Non-parametric direct multi-step estimation for forecasting economic processes." *International Journal of Forecasting*, 21(2): 201–218.

Clements, M. P. and Hendry, D. F. (1996). "Multi-Step Estimation for Forecasting." *Oxford Bulletin of Economics and Statistics*, 58(4): 657–684.
URL `http://doi.wiley.com/10.1111/j.1468-0084.1996.mp58004005.x`

Cogley, T. and Sargent, T. J. (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII US." *Review of Economic dynamics*, 8(2): 262–302.

Corradi, V. and Swanson, N. R. (2006). "Predictive density evaluation." *Handbook of economic forecasting*, 1: 197–284.

Doornik, J. A. (2013). *Object-Oriented Matrix Programming using Ox*. London: Timberlake Consultants Press.

George, E. I. and McCulloch, R. E. (1993). "Variable selection via Gibbs sampling." *Journal of the American Statistical Association*, 88(423): 881–889.

George, E. I. and Mcculloch, R. E. (1997). "Approaches for Bayesian Variable Selection." *Statistica Sinica*, 7: 339–373.

George, E. I., Sun, D., and Ni, S. (2008). "Bayesian stochastic search for VAR model restrictions." *Journal of Econometrics*, 142(1): 553–580.

Hara, R. B. O. and Sillanp, M. J. (2009). "A Review of Bayesian Variable Selection Methods : What , How and Which." *Bayesian Analysis*, (1): 85–118.

Ing, C.-k. (2003). "Multistep prediction in autoregressive process." *Econometric Theory*, 19: 254–279.

Jochmann, M., Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2013). "Stochastic search variable selection in vector error correction models with an application to a model of the UK macroeconomy." *Journal of Applied Econometrics*, 28(1): 62–81.

Jochmann, M., Koop, G., and Strachan, R. W. (2010). "Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks." *International Journal of Forecasting*, 26(2): 326–347. URL `http://dx.doi.org/10.1016/j.ijforecast.2009.11.002`

Kang, I. B. (2003). "Multi-period forecasting using different models for different horizons: An application to U.S. economic time series data." *International Journal of Forecasting*, 19(3): 387–400.

Koop, G. (2013). "Forecasting with Medium and Large Bayesian VARs." *Journal of Applied Econometrics*, 28: 177–203.

Koop, G., Leon-Gonzalez, R., and Strachan, R. W. (2009). "On the evolution of the monetary policy transmission mechanism." *Journal of Economic Dynamics and Control*, 33(4): 997–1017.

Korobilis, D. (2013). "VAR forecasting using Bayesian variable selection." *Journal of Applied Econometrics*, 28(2): 204–230.

Litterman, R. B. (1986). "Forecasting with Bayesian Vector Autoregressions: Five Years Experience." *Journal of Business & Economic Statistics*, 4(1): 25–38.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series." *Journal of econometrics*, 135(1): 499–526.

Pesaran, M. H., Pick, A., and Timmermann, A. (2011). "Variable selection, estimation and inference for multi-period forecasting problems." *Journal of Econometrics*, 164(1): 173–187. URL `http://dx.doi.org/10.1016/j.jeconom.2011.02.018`

Primiceri, G. E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies*, 72: 821–852.

7 Tables and Figures

Table 1: Average one-step ahead forecasting MSE

	DGP1			DGP2		
	MLE	Minn	SSVS	MLE	Minn	SSVS
Lag = 4	1.168	0.965	0.575	1.042	1.353	0.785
Lag = 8	2.753	1.405	0.689	2.667	2.000	0.909
AIC	2.747	1.197	0.665	2.469	1.556	0.852

Note: The average lag length by the AIC is 5.22 and 5.20 for DGP 1 and DGP2 respectively.

Table 2: Monte Carlo Simulation: Average MSFE

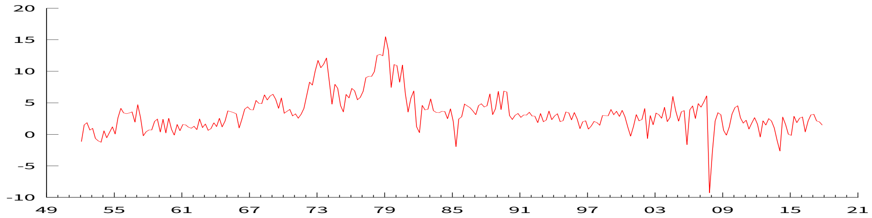
		DGP 1 forecast horizon				DGP 2 forecast horizon			
Model	method	2	4	8	12	2	4	8	12
Lag = 4									
MLE	direct	7.352	11.62	25.33	40.50	8.536	13.48	19.40	24.76
	iterated	7.047	10.29	19.46	29.39	8.273	12.34	15.80	19.39
Minn	direct	7.103	10.50	21.77	34.75	8.858	13.58	18.40	23.15
	iterated	7.114	10.52	21.31	35.10	8.349	12.27	16.16	20.37
SSVS	direct	6.517	10.14	21.92	34.28	8.091	12.47	17.74	22.72
	iterated	6.381	9.405	17.43	26.03	7.729	11.72	15.38	18.87
Lag = 8									
MLE	direct	9.579	15.49	34.15	55.85	11.11	17.67	25.50	32.35
	iterated	9.035	13.18	24.59	36.95	10.54	15.61	19.02	22.29
Minn	direct	7.768	11.49	24.26	39.23	9.825	14.83	20.17	24.58
	iterated	8.013	12.43	27.12	40.94	9.452	13.77	18.43	23.84
SSVS	direct	6.788	10.80	23.87	37.53	8.407	13.05	18.50	23.81
	iterated	6.621	9.838	18.23	27.15	7.962	12.04	15.75	19.26
AIC									
MLE	direct	9.766	16.74	44.74	79.95	11.86	18.98	27.88	40.70
	iterated	9.183	13.31	25.55	42.84	10.43	15.31	19.46	24.07
Minn	direct	7.340	11.25	25.19	41.81	9.280	14.00	18.48	23.99
	iterated	7.550	11.50	24.15	45.29	8.841	13.01	17.25	21.99
SSVS	direct	6.681	10.78	24.16	37.88	8.437	13.15	18.65	24.22
	iterated	6.595	9.750	18.23	27.36	7.937	11.97	15.71	19.28

Table 3: MSFE ratio

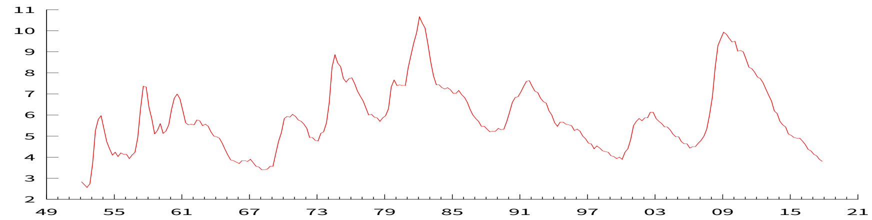
		DGP 1				DGP 2			
		forecast horizon				forecast horizon			
Model		2	4	8	12	2	4	8	12
Lag = 4									
MLE	mean	1.043	1.125	1.277	1.335	1.031	1.090	1.221	1.269
	stdev	0.027	0.079	0.206	0.320	0.029	0.063	0.117	0.172
	HPDI (L)	0.991	0.938	0.912	0.821	0.974	0.982	1.031	0.915
	HPDI (H)	1.100	1.272	1.790	2.162	1.090	1.242	1.486	1.676
	pr.(<1)	0.06	0.05	0.07	0.12	0.07	0.09	0.02	0.05
Minn	mean	0.999	0.997	1.016	0.999	1.062	1.107	1.137	1.132
	stdev	0.031	0.071	0.193	0.294	0.035	0.063	0.097	0.154
	HPDI (L)	0.944	0.850	0.660	0.507	0.993	0.987	0.980	0.895
	HPDI (H)	1.052	1.123	1.462	1.607	1.141	1.249	1.363	1.410
	pr.(<1)	0.50	0.49	0.49	0.51	0.03	0.04	0.04	0.22
SSVS	mean	1.021	1.074	1.241	1.293	1.047	1.061	1.136	1.173
	stdev	0.026	0.073	0.208	0.289	0.041	0.063	0.122	0.180
	HPDI (L)	0.964	0.951	0.862	0.803	0.971	0.938	0.961	0.892
	HPDI (H)	1.073	1.248	1.722	2.066	1.135	1.203	1.469	1.583
	pr.(<1)	0.23	0.12	0.08	0.13	0.11	0.17	0.06	0.12
Lag = 8									
MLE	mean	1.060	1.166	1.357	1.446	1.054	1.132	1.345	1.467
	stdev	0.040	0.096	0.253	0.421	0.034	0.072	0.180	0.312
	HPDI (L)	0.982	0.988	0.942	0.803	0.986	1.021	1.022	1.097
	HPDI (H)	1.146	1.396	1.976	2.871	1.119	1.304	1.772	2.059
	pr.(<1)	0.07	0.04	0.08	0.10	0.06	0.01	0.01	0.01
Minn	mean	0.970	1.166	0.905	0.834	1.040	1.079	1.092	1.030
	stdev	0.034	0.096	0.224	0.328	0.035	0.061	0.108	0.176
	HPDI (L)	0.899	0.750	0.537	0.307	0.966	0.975	0.915	0.705
	HPDI (H)	1.044	1.126	1.324	1.570	1.105	1.206	1.323	1.381
	pr.(<1)	0.81	0.84	0.69	0.75	0.15	0.10	0.20	0.41
SSVS	mean	1.024	1.091	1.289	1.349	1.055	1.081	1.159	1.206
	stdev	0.027	0.073	0.211	0.299	0.043	0.062	0.122	0.207
	HPDI (L)	0.974	0.965	0.942	0.821	0.981	0.986	0.961	0.876
	HPDI (H)	1.080	1.242	1.722	2.054	1.137	1.235	1.461	1.774
	pr.(<1)	0.14	0.09	0.05	0.09	0.05	0.05	0.05	0.12
AIC									
MLE	mean	1.068	1.275	1.729	1.875	1.140	1.253	1.456	1.687
	stdev	0.163	0.313	0.596	0.861	0.155	0.290	0.519	0.695
	HPDI (L)	0.760	0.791	0.905	0.845	0.871	0.794	0.690	0.787
	HPDI (H)	1.466	1.974	3.136	4.556	1.459	1.957	2.626	3.614
	pr.(<1)	0.35	0.20	0.04	0.09	0.22	0.20	0.22	0.12
Minn	mean	0.975	0.990	1.073	1.066	1.052	1.080	1.076	1.091
	stdev	0.055	0.130	0.305	0.454	0.061	0.104	0.128	0.190
	HPDI (L)	0.871	0.753	0.607	0.373	0.928	0.926	0.839	0.750
	HPDI (H)	1.075	1.235	1.718	2.218	1.211	1.373	1.363	1.500
	pr.(<1)	0.72	0.54	0.47	0.50	0.16	0.17	0.29	0.34
SSVS	mean	1.014	1.106	1.313	1.367	1.062	1.097	1.171	1.224
	stdev	0.041	0.110	0.237	0.353	0.051	0.077	0.139	0.237
	HPDI (L)	0.946	0.927	0.979	0.795	0.965	0.973	0.982	0.914
	HPDI (H)	1.087	1.348	1.877	2.251	1.187	1.273	1.460	1.883
	pr.(<1)	0.37	0.15	0.07	0.12	0.10	0.12	0.04	0.12

Figure 1: Data: Inflation Rate, Unemployment Rate, Interest Rate

(a) Inflation Rate



(b) Unemployment Rate



(c) Interest Rate

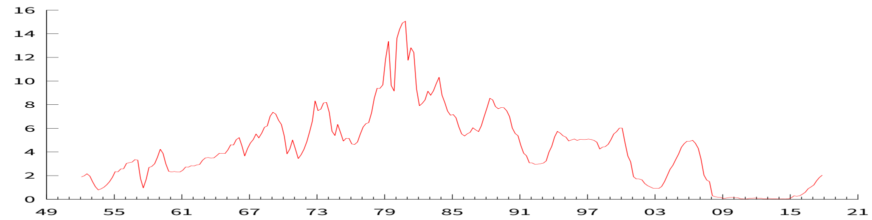


Table 4: MSFEs

		forecast horizon							
		2		4		8		12	
Model	Variable	direct	iterated	direct	iterated	direct	iterated	direct	iterated
Lag = 12									
MLE	Unemp.	0.514	0.133	2.299	0.988	3.513	3.117	5.585	4.771
	Inflation	3.600	2.063	5.829	3.791	10.93	8.507	20.49	11.32
	Interest	2.758	1.414	6.026	3.308	12.60	10.72	17.23	14.87
Minn	Unemp.	0.505	0.139	2.234	0.994	3.386	3.072	5.536	5.274
	Inflation	2.379	1.184	4.372	2.688	9.092	6.161	16.99	9.681
	Interest	1.954	0.885	4.343	2.549	9.315	7.908	13.56	11.48
SSVS	Unemp.	0.359	0.073	1.346	0.634	3.839	2.166	4.243	2.808
	Inflation	2.252	1.136	4.085	2.514	9.443	6.098	15.57	8.937
	Interest	2.133	1.001	4.112	2.530	10.10	7.563	17.94	11.23
Lag by AIC									
	Average Lag	9.220	9.958	7.446	9.958	9.768	9.958	5.863	9.958
MLE	Unemp.	0.449	0.113	1.910	0.801	3.319	2.507	5.571	3.670
	Inflation	3.292	1.954	4.634	3.092	10.74	6.009	18.53	6.859
	Interest	2.516	1.260	5.333	2.897	11.66	8.718	13.81	11.51
Minn	Unemp.	0.435	0.120	1.889	0.818	3.223	2.656	5.295	4.354
	Inflation	2.034	1.138	3.500	2.297	9.104	5.103	13.84	7.159
	Interest	1.903	0.838	3.802	2.311	8.967	6.947	10.40	10.30
SSVS	Unemp.	0.351	0.072	1.389	0.619	3.827	2.142	4.129	2.817
	Inflation	2.038	1.079	3.182	2.315	9.553	5.525	15.26	7.755
	Interest	2.020	0.966	3.790	2.580	9.799	7.691	14.48	11.60

Table 5: MSFE first differenced data

Model	Variable	2 direct	2 iterated	4 direct	4 iterated	8 direct	8 iterated	12 direct	12 iterated
L = 11									
MLE	Unemp.	0.352	0.114	1.206	0.787	3.394	2.613	5.361	4.104
	Inflation	3.427	1.723	3.612	2.582	4.829	4.268	8.325	5.246
	Interest	2.035	1.439	4.256	3.143	8.903	9.314	13.69	12.37
Minn	Unemp.	0.343	0.112	1.184	0.782	3.375	2.642	5.264	3.898
	Inflation	1.739	0.965	2.579	1.469	3.920	2.336	5.421	2.889
	Interest	1.777	0.879	3.479	2.344	7.146	6.484	10.06	8.915
SSVS	Unemp.	0.365	0.069	1.271	0.565	3.204	2.361	4.669	3.913
	Inflation	1.611	0.906	2.217	1.368	3.651	2.385	4.684	2.975
	Interest	1.777	1.063	3.610	2.430	6.930	6.877	9.787	10.08
Lag by AIC									
	Average Lag	7.851	10.54	6.042	10.54	3.494	10.54	1.000	10.54
MLE	Unemp.	0.355	0.114	1.150	0.786	3.124	2.607	4.798	4.431
	Inflation	2.099	1.916	2.671	2.533	3.846	3.967	4.757	5.200
	Interest	1.827	1.622	3.930	3.276	7.278	9.547	9.858	12.08
Minn	Unemp.	0.348	0.112	1.145	0.790	3.147	2.665	4.797	4.421
	Inflation	1.695	0.941	2.471	1.367	3.453	2,508	4.552	3.249
	Interest	1.843	0.869	3.642	2.445	7.181	7.141	9.834	10.23
SSVS	Unemp.	0.360	0.067	1.287	0.574	3.196	2.317	4.634	3.901
	Inflation	1.616	0.904	2.265	1.335	3.647	2.413	4.406	3.253
	Interest	1.644	1.022	3.494	2.493	6.979	6.971	9.704	10.67