Evaluation of Forecasting Performance Using Bayesian Stochastic Search Variable Selection in a Vector Autoregression

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

Evaluation of Forecasting Performance Using Bayesian Stochastic Search Variable Selection in a Vector Autoregression

Katsuhiro Sugita

Working Paper No.1

August 2018

Evaluation of forecasting performance using Bayesian stochastic search variable selection in a vector autoregression

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 examines forecasting performance of a vector autoregressive (VAR) model by a Bayesian stochastic search variable selection (SSVS) method. We use several artificially generated data sets to evaluate forecasting performance using a direct multiperiod forecasting method with a recursive forecasting exercise. We find that implementing SSVS prior in a VAR improves forecasting performance over unrestricted VAR models for either non-stationary or stationary data. As an illustration of a VAR model with SSVS prior, we investigate US macroeconomic data sets with three variables using a VAR with lag length of ten, and find that the SSVS restrictions on insignificant parameters alleviates over-parameterized problem of VAR and thus offers an appreciable improvement in forecast performance.

1 Introduction

Vector Autoregressive (VAR) models have been widely used to forecast macroeconomic variables and to analyze macroeconomics and policy. However, VAR models tend to have over-parameterization problem, which leads to imprecise inference and thus deteriorates the forecast performance. To remedy this problem, George et al. (2008) apply a Bayesian stochastic search variable selection (SSVS) method to VAR model. 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 SSVS method to various multivariate time series models by Hara and Sillanp (2009), Jochmann et al. (2010), Jochmann et al. (2013) and Koop (2013).

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 1-step ahead mean squared error of forecast error and Kullback-Liebler divergence. In this paper, we examine numerical simulations using several DGPs to evaluate forecasting performances with 1, 4, 8 and 12-step ahead horizons, and compute predictive likelihoods and the mean squared forecast error (MSFE) to compare restricted SSVS VAR model with other unrestricted VAR models. We find that implementing SSVS in VAR improves forecasting performance appreciably, particularly when the data are non-stationary. We then illustrate an application of SSVS VAR model to US macroeconomics model. We choose the lag length of ten for the VAR based on the AIC. With this relatively long lag length and thus a large number of parameters in the model, we find that SSVS can effectively restrict insignificant parameters in the model and thus improve forecasting performance.

The plan of this paper is as follows. Section 2 reviews prior and posterior distributions of SSVS VAR model. We also review methods to evaluate forecasting performances. In Section 3 we describe numerical experiments with artificially generated data, and then examine the results of the numerical simulations. Section 4 illustrates an application to a simple 3-variable VAR of US macroeconomics. Section 5 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 VAR model with SSVS prior

2.1 SSVS prior

In this section, we present a VAR model with SSVS prior, proposed by George et al. (2008). Let y_t be an $n \times 1$ vector of observations at time t, then an unrestricted VAR model with p lag is written as

$$y_t' = C + \sum_{i=1}^{p} y_{t-i}' \Theta_i + \varepsilon_t' \tag{1}$$

for $t = 1, \ldots, T$, where C is a $1 \times n$ vector of an intercept term; Θ_i are $n \times n$ matrices of coefficients for $i = 1, \ldots, p$; 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. 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, we follow George et al. (2008) method to investigate forecasting performances relative to non-SSVS methods such as the MLE, unrestricted Bayesian 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 we can rewrite the VAR model (1) in matrix form as

$$Y = X\Phi + \varepsilon \tag{2}$$

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 unrestricted 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:

$$\phi_j|\gamma_j \sim (1-\gamma_j)N(0, \tau_{0,j}^2) + \gamma_j N(0, \tau_{1,j}^2) \tag{3}$$

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:

$$\begin{aligned} P(\gamma_j = 1) &= p_j \\ P(\gamma_j = 0) &= 1 - p_j \end{aligned} \tag{4}$$

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 (3) can be written as:

$$\phi|\gamma \sim N(0, DD) \tag{5}$$

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

$$d_j = \begin{cases} \tau_{0j} & \text{if } \gamma_j = 0 \\ \tau_{1j} & \text{if } \gamma_j = 1 \end{cases} \tag{6}$$

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, we can decompose the error covariance matrix as:

$$\Sigma^{-1} = \Psi\Psi' \tag{7}$$

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

$$\Psi = \{\psi_{ij}\} = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1n} \\ 0 & \psi_{22} & \cdots & \psi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{nn} \end{bmatrix} \tag{8}$$

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

$$\psi_{ii}^2 \sim G(a_i, b_i) \tag{9}$$

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 $\eta_j = \left(\psi_{1j}, \psi_{2j}, \ldots, \psi_{j-1,j}\right)'$ 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:

$$\eta_j | \omega_j \sim N(0, D_j D_j) \tag{10}$$

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:

$$\begin{aligned} P(\omega_{ij} = 1) &= q_{ij} \\ P(\omega_{ij} = 0) &= 1 - q_{ij} \end{aligned} \tag{11}$$

where q_{ij} is equal to 0.5 for a natural default choice. D_j in (10) is defined as $D_j = diag(\left(h_{1j}, \ldots, h_{j-1,j}\right)$, where

$$h_{ij} = \begin{cases} \kappa_{0ij} & \text{if } \omega_{ij} = 0 \\ \kappa_{1ij} & \text{if } \omega_{ij} = 1 \end{cases}. \tag{12}$$

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

$$\psi_{ij} | \omega_{ij} \sim (1 - \omega_{ij}) N(0, \kappa_{0ij}^2) + \omega_{ij} N(0, \kappa_{1ij}^2) \tag{13}$$

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, we consider stochastic search for both of Ψ and Φ together. Note that, if we treat the unknown indicator parameters to equal to 1, that is $\gamma_j = \omega_{ij} = 1$ for all j and i, then our SSVS VAR is just unrestricted VAR.

2.2 Posteriors

In this section, we show the conditional posterior distribution for each parameter, followed 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

$$\begin{aligned}\mathscr{L}(Y|\Phi,\Sigma) &\propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2}(Y - X\Phi)\Sigma^{-1}(Y - X\Phi)'\right] \\ &\propto \prod_{i=1}^{n} \psi_{ii}^{T} \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n} \psi_{ii}^2 \upsilon_i + \sum_{j=2}^{n} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)' S_{j-1} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)\right\}\right]\end{aligned} \tag{14}$$

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

With the likelihood function (14) and priors of (4), (5), (9), (10), (11) and (13), George et al. (2008) derive the conditional posterior distributions as follows. For the VAR coefficients Φ, the conditional posterior is given as:

$$\Phi|\gamma, \eta, \psi; Y \sim N_m(\mu, \Xi) \tag{15}$$

where

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

For the conditional posterior of γ, let define $\gamma_{(-i)} = (\gamma_1, \ldots, \gamma_{i-1}, \gamma_{i+1}, \ldots, \gamma_m)$, then we have

$$\gamma_j|\phi, \gamma_{j-1}, \eta, \psi; Y \sim Bernoulli\left(u_{j1}/(u_{j1} + u_{j2})\right) \tag{16}$$

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 $\psi_{11}^2, \psi_{22}^2, \ldots, \psi_{nn}^2$ are independent and gamma distributions as:

$$\psi_{jj}^2|\phi,\omega;Y \sim G(a_j + \frac{T}{2}, b_j) \tag{17}$$

where

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

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

$$\eta_j|\phi,\omega,\psi;Y \sim N_{j-1}(\mu_j, \Delta_j) \tag{18}$$

where

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

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

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

$$\omega_{ij}|\phi,\psi,\omega_k, k \neq j;Y \sim Bernoulli\left(u_{ij1}/(u_{ij1} + u_{ij2})\right) \tag{19}$$

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 (15) - (19).

2.3 Forecasting

In this section, we discuss methods for evaluating the forecasting performances among SSVS VAR model and other unrestricted models. For point forecasting, we consider the mean squared forecast error (MSFE) and the one-step ahead average mean squared error (MSE) of forecast error, which measures the forecast error caused by deviation of estimation of coefficients. We employ a direct multiperiod forecasting method using a recursive forecasting exercise.

Let us consider the following VAR model instead of (1),

$$
\begin{aligned}
y'_{\tau+h} &= C + \sum_{i=1}^{p} y'_{\tau-i}\Theta_i + \varepsilon'_\tau \\
&= X_{\tau-1}\Phi + \varepsilon'_\tau
\end{aligned} \tag{20}
$$

where $X_{\tau-1} = \left(1, y'_{\tau-1}, \ldots, y'_{\tau-p}\right)$, $\Phi = (C', \Theta'_1, ..., \Theta'_p)'$, $y'_{\tau+h}$ is a vector of observations at time $\tau+h$ for $\tau = \tau_0, ..., T-h-1$, and $h = 0, 3, 7$, and 11, that are one-, four-, eight- and twelve-step ahead forecasts. Let $Y_{\tau-1} = (X_{\tau-1}, X_{\tau-2}, \ldots, X_1)$, then we estimate $\hat{\Phi}$ 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:

$$
\frac{1}{T-h-\tau_0+1} \sum_{\tau=\tau_0}^{T-h} \left[y_{\tau+h} - \hat{y}_{\tau+h}|\hat{\Phi}, Y_{\tau-1}\right]' \left[y_{\tau+h} - \hat{y}_{\tau+h}|\hat{\Phi}, Y_{\tau-1}\right]. \tag{21}
$$

Note that in this paper we consider the MSFE for all variables by summing up each MSFE for each variable.

For Monte Carlo simulation in the next section, since Φ is known a priori for $h = 0$ (1-step ahead forecast), we consider an average MSE of forecast error for point forecasts. The one-step ahead forecast error at period τusing information up to τ-1 can be decomposed into tow parts such as:

$$
y_\tau - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1} = (y_\tau - \hat{y}_\tau|\Phi) + \left(\hat{y}_\tau|\Phi - \hat{y}_\tau|\hat{\Phi}, Y_{\tau-1}\right) \tag{22}
$$

The first term in the right hand side in (22), $y_\tau - \hat{y}_\tau|\Phi$, 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 on $\hat{\Phi}$, 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 average one-step ahead MSE of the forecast error is calculated as;

$$
\frac{1}{N} \sum_{n=1}^{N} \left[\left(\hat{y}_T^{(n)}|\Phi^{(n)} - \hat{y}_T^{(n)}|\hat{\Phi}^{(n)}, Y_{T-1}^{(n)}\right)' \left(\hat{y}_T^{(n)}|\Phi^{(n)} - \hat{y}_T^{(n)}|\hat{\Phi}^{(n)}, Y_{T-1}^{(n)}\right) \right] \tag{23}
$$

where $n = 1, \ldots, N$ is the number of simulation samples. Note that this average one-step ahead MSE of the forecast error is also for all variables by summing up each MSE for each variable.

MSFE or the on-step ahead MSE of forecast error mentioned above is the point forecasts, that is a standard frequentist criterion for forecast evaluation. For Bayesian forecast comparison, predictive likelihoods are more preferable than point forecasts since they provide more comprehensive forecast with the entire predictive density (see Corradi and Swanson (2006) for review). The predictive density $p\left(\hat{y}_{\tau+h}|Y\right)$ can be approximated using MCMC with replications for $k = 1, \ldots, K$ as:

$$\begin{aligned}\hat{p}\left(\hat{y}_{\tau+h}|Y_{\tau-1}\right) &= \frac{1}{K}\sum_{k=1}^{K} p\left(\hat{y}_{\tau+h}|Y_{\tau-1},\Phi^{(k)},\Sigma^{(k)}\right) \\ &= \frac{1}{K}\sum_{k=1}^{K} f_N\left(\hat{y}_{\tau+h}=y_{\tau+h}|X_{\tau-1}\Phi^{(k)},\Sigma^{(k)}\right) \\ &\to p\left(\hat{y}_{\tau+h}|Y\right) \text{ as } K\to\infty\end{aligned} \tag{24}$$

where f_N denotes the normal density such as:

$$f_N\left(\hat{y}_{\tau+h}=y_{\tau+h}|X_{\tau-1}\Phi^{(k)},\Sigma^{(k)}\right)=\frac{|\Sigma^{(k)}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\left(y_{\tau+h}-X_{\tau-1}\Phi^{(k)}\right)'\left(y_{\tau+h}-X_{\tau-1}\Phi^{(k)}\right)\right]. \tag{25}$$

The predictive likelihood is the predictive density in (24) evaluated at the actual value of $y_{\tau+h}$. Then, the sum of the log predictive likelihoods is computed as:

$$\sum_{\tau=\tau_0}^{T-h}\ln\hat{p}\left(\hat{y}_{\tau+h}|Y_{\tau-1}\right)=\sum_{\tau=\tau_0}^{T-h}\ln\left[\frac{1}{K}\sum_{k=1}^{K} f_N\left(\hat{y}_{\tau+h}=y_{\tau+h}|X_{\tau-1}\Phi^{(k)},\Sigma^{(k)}\right)\right] \tag{26}$$

We use the sum of the log predictive likelihoods to evaluate forecasting performance instead of the Kullback-Leibler divergence used by George et al. (2008) to compare predictive density estimates.

3 Monte Carlo Simulation

In this section, we present several Monte Carlo simulations to illustrate forecasting performances of the SSVS for VAR models. We consider five data generating processes (DGPs). For each DGPs, we simulate 100 samples of size $T = 100$ and $T = 200$, and then for each sample we estimate MLE, Bayesian unrestricted VAR, Minnesota prior model and SSVS VAR for comparison. For Bayesian unrestricted VAR, the indicator parameters γ and ω in SSVS VAR are set to equal to one, that is $\gamma_j = \omega_{ij} = 1$ for all j and i. 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:

$$V(\Theta_i)=\frac{\lambda^2}{i^2}\begin{bmatrix} 1 & \theta\hat{\sigma_1}^2/\hat{\sigma_2}^2 & \cdots & \theta\hat{\sigma_1}^2/\hat{\sigma_n}^2 \\ \theta\hat{\sigma_2}^2/\hat{\sigma_1}^2 & 1 & \cdots & \theta\hat{\sigma_2}^2/\hat{\sigma_n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \theta\hat{\sigma_n}^2/\hat{\sigma_1}^2 & \theta\hat{\sigma_n}^2/\hat{\sigma_2}^2 & \cdots & 1 \end{bmatrix}, \tag{27}$$

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.

We consider the following five DGPs for VARs. All DGPs contain intercept term. DGP 1 is a four-variable VAR with two lags. DGP 1 contains unit roots with parameters

$$\text{DGP1: } \Phi^{(DGP1)} = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \\ 1 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix} \text{ and } \Psi^{(DGP1)} = \begin{bmatrix} 2 & -1 & 1 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Next, DGP 2 is a four-variable VAR with two lags, and is stationary data with parameters

$$\text{DGP 2: } \Phi^{(DGP2)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.5 & 0 & 0 & 0 \\ -0.3 & 0.5 & 0 & 0 \\ 0 & -0.3 & 0.5 & 0 \\ 0 & 0 & -0.3 & 0.5 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix} \text{ and } \Psi^{(DGP2)} = \Psi^{(DGP1)}.$$

DGP 3 is a four-variable VAR with four lags, and also shows stationary series and with parameters

$$\text{DGP 3: } \Phi^{(DGP3)} = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.6 & 0 & 0 & 0 \\ -0.3 & 0.6 & 0 & 0 \\ 0 & -0.3 & 0.6 & 0 \\ 0 & 0 & -0.3 & 0.6 \\ 0 & 0 & 0 & 0 \\ -0.2 & 0 & 0 & 0 \\ 0 & -0.2 & 0 & 0 \\ 0 & 0 & -0.2 & 0 \\ -0.3 & 0 & 0 & -0.2 \\ 0 & -0.3 & 0 & 0 \\ 0 & 0 & -0.3 & 0 \\ 0 & 0 & 0 & -0.3 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.3 \end{bmatrix} \text{ and } \Psi^{(DGP3)} = \Psi^{(DGP1)}.$$

DGP 4 is a six-variable VAR with two lags, and contains unit roots and its parameters are

$$\text{DGP 4: } \Phi^{(DGP4)} = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix}$$

$$\text{and } \Psi^{(DGP4)} = \begin{bmatrix} 2 & -1 & 1 & -1 & 1 & -1 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

DGP 5 is a six-variable VAR with two lags, and is stationary data with parameters

$$\text{DGP 5: } \Phi^{(DGP5)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.4 & 0 & 0 & 0 & 0 & 0 \\ -0.3 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & -0.3 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & -0.3 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0-0.3 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & -0.3 & 0.4 \\ 0.3 & 0 & 0 & 0 & 0 & 0 \\ -0.2 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & -0.2 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & -0.2 & 0.3 & 0 & 0 \\ 0 & 0 & 0 & -0.2 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & -0.2 & 0.3 \end{bmatrix}$$

and $\Psi^{(DGP5)} = \Psi^{(DGP4)}$.

With regard to Ψ of these DGPs, we assume that partial correlation between the i-th component and the j-th component of the error term for $i, j > 1$ in (8) is zero. In all Bayesian models, 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 (4), $q_i = 0.5$ in (11), $a = b = 0.01$ in (9), $c_0 = 0.1$ and $c_1 = 50$ for τ_{kj} and κ_{kij} in (6) and (12). For each sample, MCMC is run with 20,000 draws after 5,000 burn-in. We consider joint restriction search for both Φ and Ψ for SSVS. The lag length is assumed to be known a priori for all models.

Table 1 - 6 present results of numerical simulations with $T = 100$ and 200 by DGP1 - 5. We consider four models labeled by MLE (unrestricted Maximum Likelihood Estimation), BVAR(Bayesian unrestricted VAR with relatively non-informative priors), Minn.prior (Bayesian unrestricted VAR with Minnesota prior) , and SSVS (Bayesian restricted VAR with SSVS priors).

Table 1 presents the one-step ahead average MSE of forecast error over 100 samples, calculated by (23). We find that there is little difference between the MLE and the BVAR for all DGPs, since the BVAR uses relatively non-informative priors, generating similar posterior estimates to those by MLE. The Minnesota prior presents better forecast compared with the MLE or the BVAR for DGP1 and DGP4 that contain unit roots since the Minnesota prior suppose non-stationary process. The SSVS priors show substantially better than the MLE or the BVAR, especially for non-stationary DGP1 and DGP4.

Tables 2 - 6 report the sum of the log predictive likelihoods and the MSFEs for $h =$ 0, 3, 7, and 11, calculated by (26) and (21) respectively. As in the case of the one-step ahead average MSE of forecast error, there is little difference between the MLE and the unrestricted BVAR. From these results, it is often the case that the SSVS performs best and the MLE performs worst in terms of both the predictive likelihood and the MSFE. For any DPG, SSVS leads to higher predictive likelihoods and lower MSFEs, and improvements of forecasting performances by SSVS over MLE are higher with $T = 100$ than with $T = 200$. This is because prior that affects to posterior is having little impact when the size T is large.

In the case of DGP 1 in Table 2 and DGP 4 in Table 5, both of which contain unit roots in the data, the Minnesota prior exceeds to the SSVS for T=200 and $h \geqq 3$. For stationary data such as DGP 2, 3, and 5, the SSVS performs best, while the Minnesota prior performs poorly in some cases due to the assumption of random walk process. Comparing DGP 2 with DGP 3, the SSVS perform much better for the case of DGP 3 than for the case of DGP2. This is because DGP 3 includes longer lags (4 lags) than DGP 2 (2 lags), suggesting restrictions were effectively imposed by the SSVS. Overall, Tables 2 - 6 indicate that implementing SSVS leads to better forecasting performances for either non-stationary or stationary data with small or large size.

4 An empirical analysis

In this section, we apply SSVS VAR model to an empirical study 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. We obtained these data from the Federal Reserve Bank of St. Louise.[1] These data are plotted in Figure 1.

The choice 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. They do not mention reason why the lag length was chosen to be two particularly, however probably by parsimonious reason. Jochmann et al. (2010) use VAR(4) for their SSVS VAR model with reason that the SSVS can find zero restrictions on the parameters of longer lags even if the true lag length is less than 4. However, there is possibility that the true lag length is more than 4. With our data set, the choice of the number of lags is scattered depending on 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 choose VAR (10) as suggested by the AIC. Forecast horizons are 1 quarter, 1 year, 2 year, and 3 year, which corresponds to $h = 0$, 3, 7, and 11 respectively in our VAR in (20). We work with a recursive forecasting exercise using a direct multistep forecasting method, using data up to time $\tau - 1$, where $\tau = \tau_0, \ldots, T - h - 1$, to forecast at time $\tau + h$ for $h = 0$, 3, 7 and 11.

Table 7 presents the sum of the log predictive likelihoods (26) and the MSFEs (21) for the three-variable VAR with lag length of ten for MLE, BVAR, Minn.prior and SSVS models. For any forecast horizon, SSVS improves the forecasting performance with large lag length in terms of the highest predictive likelihood among all models. SSVS reduces the MSFE against the MLE at 38.9% for $h = 0$, 28.8% for $h = 3$, 38.2% for $h = 7$ and 26.5% for $h = 11$. However VAR with the Minnesota prior leads to the lowest MSFEs at longer horizons with $h = 3$ and 11, although there are little differences of the MSFEs between SSVS prior and Minnesota prior.

Since the sum of log predictive likelihoods (26) can be interpreted as the log of the marginal likelihood as suggested by Geweke and Amisano (2011), we can use them for model selection. Thus, we find that SSVS for VAR (10) is preferable model among all models under consideration. Table 8 presents a complete set of posterior means, standard deviations and inclusion probabilities, denoted by $\mathrm{P(inc.)} = \Pr(\gamma_j = 1|Y)$ or $\Pr(\omega_j = 1|Y)$, of the VAR coefficients Φ and the off-diagonal elements of the error covariance Ψ respectively. The posterior inclusion probabilities for the VAR coefficients can be used for model averaging, or choosing single model that includes only coefficients with P(inc.)>0.5. We find that there are only 20 important coefficients with $\Pr(\gamma_j = 1|Y) > 0.5$ out of 90 in Φ, excluding the intercepts, and only 1 important element with $\Pr(\omega_j = 1|Y) > 0.5$ out of the 3 possible off diagonal elements in Ψ. This indicates that SSVS is effective to ensure parsimony in over-parameterized VAR(10) model. If we use SSVS to select lag length of VAR, we find that the number of lag would be ten (or more) but fourth and sixth lag should be excluded since none of the coefficients on the sixth lag variables are important where all inclusion probabilities in the sixth lag variables are less than 0.5..

For the posterior results of Ψ, we find that two out of three off-diagonal elements

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

of Ψ are restricted to be zero as $\Pr(\omega_j = 1|Y) \leq 0.5$. Let define $\varepsilon = \left(\varepsilon_{\text{Unemp}}, \varepsilon_{\text{Inf}}, \varepsilon_{\text{Int}}\right)'$ to denote the VAR residuals and e to denote the structural shock vector, then the point estimates of the Ψ restricts on how regression residuals relate to one another as $1.6413\varepsilon_{\text{Int}} = -1.9905\varepsilon_{\text{Unemp}} + 0.0116\varepsilon_{\text{Inf}} + u_{\text{Int}}$, suggesting unexpected interest rate equals 1.2128 times that of unemployment rate and an idiosyncratic shock with a standard deviation of about 0.6093%.

In this section, we have examined the forecasting performances of SSVS VAR models to compare with other models. We chose VAR (10) which contains relatively large coefficients and found that the SSVS restricts many insignificant coefficients to be zero. We have strong evidence that SSVS improves forecasting performance for VAR model.

5 Conclusion

George et al. (2008) develop method for implementing SSVS in VAR models, and show numerical simulations to illustrate that VAR with SSVS prior improves forecasting performance using one-step ahead forecasting MSE and the KL distance between the unrestricted and SSVS predictive density. In this paper, we investigate VAR model with SSVS prior to examine forecast performances using not only 1-step ahead average MSE of forecast error but also predictive likelihood, MSFE for 1, 4, 8 and 12-step ahead forecasting horizons. We find that SSVS VAR model appreciably improves the forecast performance for all DGPs. This improvement of forecasting is notable particularly when the DGP is random walk process.

We illustrate an application of US macroeconomics to show a benefit of using SSVS prior in a VAR. We choose the lag length of 10 based on the AIC, instead of 2 or 4 lag length, which has been often chosen. With longer lags and thus large number of parameters that may include many insignificant, we find that SSVS alleviates overparameterization problem in VAR model by restricting insignificant parameters of the model, and enables to improve forecasting performance.

We choose a direct multiperiod forecast method for both numerical simulation and an application. Since VAR with SSVS prior greatly improves the 1-step ahead MSE of forecast error, the coefficients are estimated more efficiently and thus an iterated multiperiod forecast method would be preferable. So, it is of interest if we compare a direct method with an iterated multiperiod method using SSVS VAR model.

Acknowledgements

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

References

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.

Geweke, J. and Amisano, G. (2011). "Hierarchical Markov normal mixture models with applications to financial asset returns." *Journal of Applied Econometrics*, 29(January 2010): 1–29.

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.

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`

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.

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

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

Table 1: 1-step ahead average MSE of forecast error

	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5	
T	100	200	100	200	100	200	100	200	100	200
MLE	0.1630	0.0908	0.1324	0.0712	0.3275	0.1281	0.4318	0.1816	0.2889	0.1344
BVAR	0.1611	0.0905	0.1279	0.0699	0.3239	0.1278	0.4245	0.1808	0.2813	0.1327
Minn.prior	0.1251	0.0787	0.1204	0.0691	0.2726	0.1330	0.3159	0.1448	0.2697	0.1416
SSVS	0.0732	0.0359	0.1132	0.0290	0.2560	0.0829	0.1160	0.0499	0.2275	0.0793

Table 2: Log predictive likelihoods and MSFEs by DGP 1

(a) $T = 100$

	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-218.40	1.4524	-590.95	6.3145	-1320.5	17.257	-2276.1	30.941
BVAR	-240.24	1.4432	-558.27	6.1051	-1142.6	16.391	-1856.1	29.225
Minn. prior	-213.41	1.3568	-415.95	4.7059	-842.96	11.937	-1513.4	22.791
SSVS	-203.77	1.2759	-421.09	4.7161	-803.25	11.091	-1305.9	22.500

(b) $T = 200$

	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-314.42	1.3305	-603.27	4.7046	-986.65	11.812	-1427.1	21.343
BVAR	-335.44	1.3296	-617.97	4.6855	-977.86	11.680	-1388.9	21.034
Minn.prior	-345.77	1.3477	-578.29	4.2544	–811.36	9.7949	-1091.4	17.549
SSVS	-310.75	1.2598	-580.56	4.3386	-828.79	10.349	-1107.8	17.854

Table 3: Log predictive likelihoods and MSFEs by DGP 2

(a) $T = 100$

	$h=0$		$h=3$		$h=7$		$h=11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-218.06	1.4216	-361.43	3.6244	-558.68	8.3217	-502.19	12.545
BVAR	-224.80	1.4052	-357.68	3.4838	-538.17	7.8093	-480.89	11.574
Minn.prior	-210.66	1.3577	-325.35	3.1973	-481.97	6.6719	-458.52	10.748
SSVS	-208.20	1.3428	-310.07	3.0091	-448.49	6.4266	-408.39	10.349

(b) $T = 200$

	$h=0$		$h=3$		$h=7$		$h=11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-325.62	1.2930	-454.60	2.7581	-542.98	5.0483	-584.54	7.9008
BVAR	-320.64	1.2913	-457.63	2.7463	-541.13	4.9667	-588.48	7.7129
Minn.prior	-337.84	1.3101	-473.48	2.7969	-577.91	5.0617	-657.20	7.8814
SSVS	-308.00	1.2601	-445.54	2.6517	-527.82	4.7872	-571.32	7.3561

Table 4: Log predictive likelihoods and MSFEs by DGP 3

(a) $T = 100$

	$h=0$		$h=3$		$h=7$		$h=11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-198.52	1.6203	-310.11	4.1628	-402.37	6.5414	-524.01	8.7630
BVAR	-213.33	1.6117	-307.80	4.1029	-367.86	6.3499	-430.44	8.3547
Minn.prior	-196.34	1.5905	-283.81	3.8096	-298.92	4.9912	-342.41	6.2802
SSVS	-176.47	1.4556	-255.60	3.5156	-278.75	5.1400	-281.69	5.8245

(b) $T = 200$

	$h=0$		$h=3$		$h=7$		$h=11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-156.64	1.3513	-222.98	3.1107	-223.16	4.0159	-219.15	5.3372
BVAR	-157.03	1.3509	-222.08	3.1078	-221.43	4.0105	-217.67	5.3230
Minn.prior	-171.55	1.4830	-225.29	3.1879	-220.79	3.9612	-218.68	5.0314
SSVS	-147.36	1.2939	-216.17	3.0265	-217.38	4.0510	-209.02	5.0773

Table 5: Log predictive likelihoods and MSFEs by DGP 4

(a) $T = 100$

	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-314.64	2.4437	-890.84	10.168	-1807.9	23.464	-2780.2	36.746
BVAR	-342.29	2.4221	-821.58	9.9079	-1522.6	22.352	-2238.7	35.367
Minn.prior	-288.46	2.2079	-571.07	6.9395	-1039.1	16.055	-1773.3	27.058
SSVS	-258.20	1.9852	-561.62	6.9047	-972.93	16.022	-1357.9	24.793

(b) $T = 200$

	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-240.46	2.0387	-453.10	7.2985	-697.53	17.325	-940.70	30.330
BVAR	-248.42	2.0371	-464.21	7.2685	-690.68	17.118	-932.92	29.980
Minn.prior	-237.08	2.0914	-405.72	6.1949	-522.95	13.191	-679.50	23.575
SSVS	-221.52	1.8677	-410.26	6.2075	-544.47	14.526	-697.09	24.710

Table 6: Log predictive likelihoods and MSFEs by DGP 5

(a) $T = 100$

	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-285.39	2.3226	-462.69	5.9286	-730.46	17.769	-1069.2	37,441
BVAR	-307.20	2.2952	-460.03	5.7424	-645.79	15.907	-799.50	29.742
Minn.prior	-279.30	2.1960	-441.97	5.6980	-679.37	15.834	-867.30	30.304
SSVS	-265.53	2.1140	-383.56	4.9508	-536.04	15.008	-621.14	27.899

(b) $T = 200$

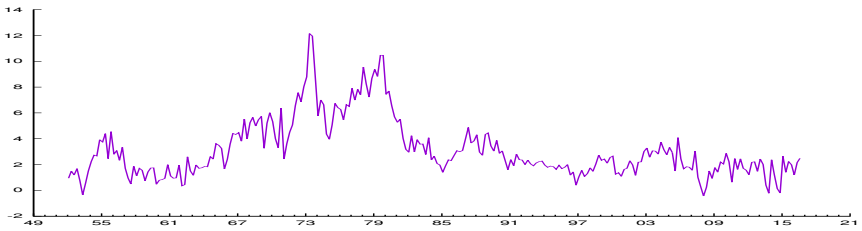
	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-232.25	1.9728	-320.60	4.2765	-352.85	9.8018	-363.33	21.914
BVAR	-235.05	1.9710	-321.80	4.2709	-352.09	9.6982	-360.18	21.427
Minn.prior	-249.87	2.0258	-351.63	4.7073	-421.35	10.782	-484.01	25.150
SSVS	-221.98	1.8000	-309.36	4.0449	-344.56	9.1659	-349.68	20.521

Table 7: Predictive Likelihood and MSFE for US Macroeconomic analysis

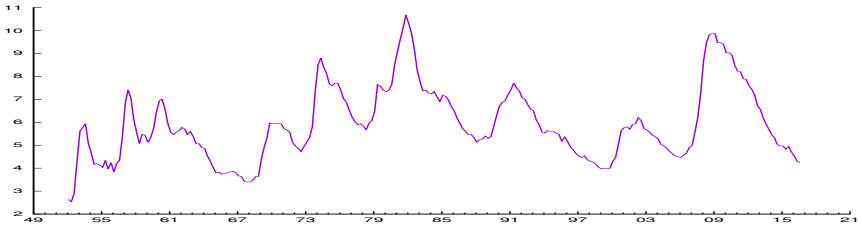
	$h = 0$		$h = 3$		$h = 7$		$h = 11$	
	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE	Pr.L	MSFE
MLE	-1178.8	3.3154	-2540.1	11.799	-5058.3	27.786	-7617.2	41.608
BVAR	-1036.7	3.1341	-2025.9	10.951	-3133.0	24.017	-5432.5	38.678
Minn.prior	-906.96	2.1542	-1613.8	8.3161	-3162.9	18.874	-3725.4	28.748
SSVS	-673.16	2.0245	-1476.1	8.3962	-2361.1	18.181	-3017.2	30.578

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

(a) Inflation Rate



(b) Unemployment Rate



(c) Interest Rate

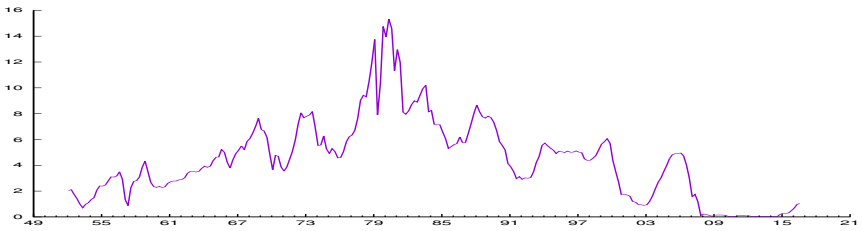


Table 8: Results for SSVS-VAR (10)

	unemp		inflation		interest	
	mean (s.d.)	P (inc.)	mean (s.d.)	P (inc.)	mean (s.d.)	P (inc.)
constant	0.1848 (0.0797)	——	0.5433 (0.3125)	——	0.0841 (0.2141)	——
unemp (-1)	1.6074 (0.0744)	1.0000	-0.1192 (0.2217)	0.2996	-0.9058 (0.2191)	0.9865
inflation (-1)	0.0044 (0.0107)	0.2028	0.6411 (0.0725)	1.0000	0.0517 (0.0554)	0.5393
interest (-1)	0.0047 (0.0103)	0.1704	0.0089 (0.0304)	0.0875	0.8322 (0.0644)	0.9838
unemp (-2)	-0.5701 (0.1584)	0.9898	0.0781 (0.3244)	0.1377	1.1198 (0.3886)	0.9838
inflation (-2)	0.0098 (0.0165)	0.3209	0.0922 (0.1008)	0.5385	0.0153 (0.0380)	0.2010
interest (-2)	0.0011 (0.0066)	0.0622	0.0030 (0.0237)	0.0543	-0.0403 (0.0771)	0.2660
unemp (-3)	-0.1816 (0.1437)	0.7002	-0.0635 (0.2126)	0.0947	-0.1848 (0.2756)	0.3686
inflation (-3)	-0.0049 (0.0136)	0.1823	0.1058 (0.1099)	0.5623	0.1701 (0.0570)	0.9741
interest (-3)	0.0011 (0.0061)	0.0587	0.0074 (0.0320)	0.0753	0.2425 (0.0737)	0.9910
unemp (-4)	-0.0045 (0.0527)	0.1178	-0.0157 (0.1035)	0.0520	0.0011 (0.0929)	0.0800
inflation (-4)	-0.0005 (0.0060)	0.0880	0.1105 (0.1047)	0.6104	-0.0180 (0.0419)	0.2127
interest (-4)	0.0010 (0.0059)	0.0532	-0.0012 (0.0254)	0.0571	-0.0174 (0.0479)	0.1518
unemp (-5)	0.1933 (0.0972)	0.8616	0.0071 (0.0609)	0.0388	-0.0287 (0.0758)	0.0993
inflation (-5)	-0.0030 (0.0108)	0.1336	-0.0134 (0.0427)	0.1439	-0.1435 (0.0619)	0.9242
interest (-5)	0.0033 (0.0096)	0.0974	-0.0085 (0.0328)	0.0893	0.0035 (0.0245)	0.0720
unemp (-6)	0.0131 (0.0510)	0.1344	0.0102 (0.0545)	0.0344	-0.0277 (0.0649)	0.0842
inflation (-6)	0.0119 (0.0188)	0.3442	0.0025 (0.0244)	0.0831	0.0101 (0.0305)	0.1456
interest (-6)	0.0004 (0.0055)	0.0452	-0.0566 (0.0856)	0.3263	-0.0004 (0.0214)	0.0680
unemp (-7)	0.0194 (0.0608)	0.1422	0.0102 (0.0512)	0.0304	-0.0071 (0.0451)	0.0387
inflation (-7)	-0.0014 (0.0075)	0.0980	-0.0124 (0.0398)	0.1371	-0.0029 (0.0172)	0.0856
interest (-7)	0.0001 (0.0061)	0.0597	-0.0282 (0.0662)	0.1844	-0.3272 (0.0663)	0.9987
unemp (-8)	-0.4494 (0.0857)	1.0000	0.0014 (0.0463)	0.0260	-0.0071 (0.0435)	0.0362
inflation (-8)	0.0032 (0.0098)	0.1395	-0.0039 (0.0267)	0.0938	0.0022 (0.0149)	0.0779
interest (-8)	-0.0021 (0.0079)	0.0724	-0.0080 (0.0380)	0.0908	0.2441 (0.0680)	0.9814
unemp (-9)	0.5278 (0.1059)	1.0000	0.0015 (0.0432)	0.0219	0.0030 (0.0524)	0.0369
inflation (-9)	0.0044 (0.0124)	0.1674	0.0049 (0.0287)	0.1065	0.0133 (0.0321)	0.1884
interest (-9)	-0.0011 (0.0052)	0.0596	-0.0043 (0.0415)	0.0907	0.0097 (0.0372)	0.1010
unemp (-10)	-0.1974 (0.0572)	0.9911	-0.0021 (0.0290)	0.0372	0.0159 (0.0468)	0.0867
inflation (-10)	-0.0131 (0.0191)	0.3867	0.0606 (0.0733)	0.4755	-0.0033 (0.0189)	0.0985
interest (-10)	-0.0011 (0.0050)	0.0639	0.0556 (0.0799)	0.3772	-0.0054 (0.0231)	0.0970

Note: P(inc.) denotes posterior probability of inclusion of parameter, $\Pr\left(\gamma_j = 1|Y\right)$.

$$\hat{\Psi} = \begin{bmatrix} 4.0458 & 0.1463 & 1.9905 \\ (0.1908) & (0.3000) & (0.2955) \\ & 1.0412 & -0.0116 \\ & (0.0495) & (0.0406) \\ & & 1.6413 \\ & & (0.0785) \end{bmatrix}, \hat{\omega} = \begin{bmatrix} 1.0000 & 0.2155 & 1.0000 \\ & 1.0000 & 0.0844 \\ & & 1.0000 \end{bmatrix}$$

Note: The posterior standard deviation is in parentheses.