Received 21 April 2022
Revised 28 April 2022
Accepted 28 April 2022

Forecasting with Bayesian vector autoregressive models: comparison of direct and iterated multistep methods

Katsuhiro Sugita
Faculty of Global and Regional Studies, University of the Ryukyus, Okinawa, Japan

Abstract

Purpose – The paper compares multi-period forecasting performances by direct and iterated method using Bayesian vector autoregressive (VAR) models.

Design/methodology/approach – The paper adopts Bayesian VAR models with three different priors – independent Normal-Wishart prior, the Minnesota prior and the stochastic search variable selection (SSVS). Monte Carlo simulations are conducted to compare forecasting performances. An empirical study using US macroeconomic data are shown as an illustration.

Findings – 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.

Originality/value – The paper finds that iterated forecasts using model with the SSVS prior 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.

Keywords Forecasting, Bayesian econometrics, VAR model

Paper type Research paper

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 method and iterated forecast method, and there have been several theoretical research about 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 more efficient asymptotically, and thus the direct forecast method is preferable compared with the iterated forecast method, while the iterated forecast method can

This work was supported by JSPS KAKENHI grant number 20K01591.

Asian Journal of Economics and Banking
Vol. 6 No. 2, 2022
pp. 142-154
Emerald Publishing Limited
e-ISSN: 2633-7991
p-ISSN: 2615-9821
DOI 10.1108/AJEB-04-2022-0044

be more efficient only if the one-period ahead model is correctly specified. However, some empirical research studies show that iterated forecasts outperform direct forecasts. Ang *et al.* (2006) find that the iterated forecasts of the US 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; this paper uses 170 US 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 data generating process (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 an over-parameterization problem, which leads to imprecise inference and thus deteriorates the forecast performance. Some Bayesian approaches to VAR models have been increasingly popular since Bayesian method can shrink VAR models by restricting its prior distributions. In this paper we investigate whether restricted parsimonious VAR models can mitigate misspecification problem and thus improve the forecasting performance of iterated method. Here, an independent Normal-Wishart prior is used for the unrestricted VAR and the Minnesota prior (Minn) and the stochastic search variable selection (SSVS) prior are used for the restricted prior to compare multiperiod forecasting performance between the direct and iterated forecast method.

We conduct numerical simulations using both stationary and nonstationary data generating processes (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 Bayesian VAR models with unrestricted and restricted priors. Iterated forecasts are found to outperform direct forecasts for both unrestricted and restricted VAR models, particularly with long-lag model and with long forecasting horizon. Implementing SSVS in VAR is found to generally improve 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 Bayesian VAR models with three different priors – the independent Normal-Wishart prior, the Minnesota prior and the SSVS prior. 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 three variables VAR of US macroeconomics. Section 6 concludes. This paper is based on preliminary working papers, Sugita (2018), Sugita (2019a) and Sugita (2019b). 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 $t = 1, \ldots, T$, where μ is a $n \times 1$ vector of an intercept term; $\boldsymbol{\Theta}_i$ are $n \times n$ matrices of coefficients for $i = 1, \ldots, p$; $\boldsymbol{\varepsilon}_t$ are $n \times 1$ independent $N_n(0, \Sigma)$ errors; and the covariance matrix $\boldsymbol{\Sigma}$ is an $n \times n$ positive definite matrix.

The one-step ahead forecast $\widehat{y}_{t+1|t}'$ of the VAR model is obtained by estimating the parameters in eq. (1) as $\widehat{y}_{t+1|t}' = \widehat{\mu}_{(I)}' + \sum_{i=1}^{p} y_{t+1-i|t}' \widehat{\Theta}_{(I),i}$, where $\widehat{\mu}_{(I)}$ and $\widehat{\Theta}_{(I),i}$ are the

estimators for μ and Θ_i in eq. (1). 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

$$\widehat{y}'_{t+h} = \widehat{\mu}'_{(I)} + \sum_{i=1}^{p} \widehat{y}'_{t+h-i} \widehat{\Theta}_{(I),i} \tag{2}$$

where $\widehat{y}_{i|t} = y_j$ for $j \leq t$.

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

$$y'_t = \mu' + \sum_{i=1}^{p} y'_{t-h-i} \Theta_i + \boldsymbol{\varepsilon}_t, \tag{3}$$

Then using the estimated coefficients directly to make the forecast of

$$\widehat{y}_{t+h} = \widehat{\mu}_{(D)} + \sum_{i=1}^{p} y_{t-i} \widehat{\Theta}_{(D),i} \tag{4}$$

where $\widehat{\mu}_{(D)}$ and $\widehat{\Theta}_{(D),i}$ are the estimators for μ and Θ_i in eq. (3). Thus, the relative forecast accuracy depends on how accurate $\widehat{\Theta}_{(I),i}$ and $\widehat{\Theta}_{(D),i}$ are estimated. If $\widehat{\Theta}_{(I),i}$ is badly estimated with large errors, then its powered values diverge increasingly from Θ_i. Since the iterated method depends on one-period ahead coefficients $\widehat{\Theta}_{(I),i}$, 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 US macroeconomic time series data, and finds 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 MSFE is the most widely used. Let $y'_{\tau+h}$ is a vector of observations at time $\tau + h$ for $\tau = \tau_0, \ldots, T - h - 1$, and $h = 2$-, 4-, 8- and 12-step ahead forecasts. Then, $\widehat{\Phi} = \left(\widehat{\mu}', \widehat{\Theta}'_1, \ldots, \widehat{\Theta}'_p\right)'$ is estimated for both the direct and iterated method, using information up to $\tau - 1$ to forecast values $\widehat{y}_{\tau+h}$ starting from $\tau = \tau_0$ up to $\tau = T - h - 1$, and calculate the MSFE defined as:

$$\text{MSFE} = \frac{1}{T - h - \tau_0 + 1} \sum_{\tau=\tau_0}^{T-h} \left[y_{\tau+h} - \widehat{y}_{\tau+h} | \widehat{\Phi}, Y_{\tau-1} \right]^2. \tag{5}$$

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

3. Bayesian VARs

This section presents Bayesian VAR models with three different priors – independent Normal-Wishart prior, the Minnesota prior and the SSVS prior. The VAR model in eq. (1) can be written in matrix form as follows:

$$Y = X\Phi + \varepsilon \tag{6}$$

where the $T \times n$ matrix Y is defined as $Y = (y_1, \ldots, y_T)'$; the $T \times (1 + np)$ matrix X is defined as $X = (x_1, \ldots, x_T)'$; the $(1 + np) \times 1$ vector is defined as $x_t = \left(1, y'_{t-1}, \ldots, y'_{t-p}\right)'$, the $(1 + np) \times n$ matrix Φ is defined as $\Phi = (\mu', \Theta'_1, \ldots, \Theta'_p)'$; and the ε is a $T \times n$ matrix with

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$. Based on the VAR model in eq. (1) or eq. (6), we describe briefly the three priors in the following subsections.

3.1 Independent Normal-Wishart prior

The VAR model in eq. (6) with the independent Normal-Wishart prior

$$vec(\Phi) \sim MN(vec(\Phi_0), V_0) \tag{7}$$

$$\Sigma \sim IW(\Sigma_0, \nu_0) \tag{8}$$

where MN refers to a multivariate normal with mean $vec(\Phi_0)$ and covariance-variance matrix V_0: IW refers to an inverted Wishart distribution with parameters Σ_0 and degrees of freedom, ν_0. Unlike the natural conjugate priors, prior for Φ in eq. (7) and Σ in eq. (8) are independently specified. With the joint prior and the likelihood, the conditional posterior densities of $vec(\Phi)$ and Σ are derived as follows:

$$vec(\Phi)|\Sigma, Y \sim MN(vec(\Phi_*), V_*) \tag{9}$$

$$\Sigma|\Phi, Y \sim IW(\Sigma_*, \nu_*) \tag{10}$$

where $V_* = \left[V_0^{-1} + \Sigma \otimes (X'X)\right]^{-1}$ and $vec(B_\star) = V_* \left[V_0^{-1} vec(\Phi_0) + (\Sigma \otimes I_\kappa)^{-1} vec(X'Y)\right]$, $\Sigma_* = (Y - X\Phi)'(Y - X\Phi) + \Sigma_0$, and $\nu_* = T + \nu_0$. Given these conditional posterior specifications above, the Gibbs sampler generates sample draws.

Note that, with zero prior mean $\Phi_0 = 0$ and large prior variance V_0 in eq. (7), the posterior mean for Φ is almost identical to the Maximum likelihood estimator. In this paper the hyperparameters are set at $vec(\Phi_0) = 0$ and $V_0 = 100$ in eq. (7), $\Sigma_0 = 0.1I$, and $\nu = 5$ in eq. (8).

3.2 Minnesota prior

Litterman (1986) proposes what we call the Minnesota prior which is shrinkage prior for a Bayesian VAR model with random walk components. For a VAR model with p-the lag in eq. (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:

$$V(\Theta_i) = \frac{\lambda^2}{i^2} \begin{bmatrix} 1 & \theta\hat{\sigma}_1^2/\hat{\sigma}_2^2 & \cdots & \theta\hat{\sigma}_1^2/\hat{\sigma}_n^2 \\ \theta\hat{\sigma}_2^2/\hat{\sigma}_1^2 & 1 & \cdots & \theta\hat{\sigma}_2^2/\hat{\sigma}_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \theta\hat{\sigma}_n^2/\hat{\sigma}_1^2 & \theta\hat{\sigma}_n^2/\hat{\sigma}_2^2 & \cdots & 1 \end{bmatrix}, \tag{11}$$

where $0 < \theta < 1$, and $\Sigma = \text{diag}\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_n^2\right)$. In this paper, the hyperparameters in eq. (11) are set at $\lambda = 0.05$ and $\theta = 0.1$.

3.3 SSVS prior

Without any restriction on the regression coefficients and the covariance matrix in eq. (1), VAR models usually has over-parameterization problem. They contain a very large number of parameters, leading to imprecise inference and deterioration of the forecast performance. To overcome this problem, George *et al.* (2008) apply the Bayesian SSVS method in a VAR. The SSVS method, proposed by George *et al.* (2008) and George and McCulloch (1997), restricts the parameters of the model by using a hierarchical prior on the parameters.

SSVS defines the prior for the VAR coefficient Φ for each element in Φ. Let ϕ_j be each element in Φ, then the prior for ϕ_j is a hierarchical prior with mixture of two normal distributions with different variance conditional on an unknown dummy variable γ_j that takes zero or one:

$$\phi_j|\gamma_j \sim (1-\gamma_j)N\left(0, \tau_{0,j}^2\right) + \gamma_j N\left(0, \tau_{1,j}^2\right) \tag{12}$$

where $\tau_{0,j}^2$ is small and $\tau_{0,j}^2 < \tau_{1,j}^2$. This implies that if $\gamma_j = 0$, that is, the element ϕ_j is restricted to be close to 0 as $\phi_j|\gamma_j \sim N(0, \tau_{0,j}^2)$, the prior for $\phi_j|\gamma_j$ is virtually zero with small variance, on the other hand, if $\gamma_j = 1$, that is, the element ϕ_j is unrestricted as $\phi_j|\gamma_j \sim N(0, \tau_{1,j}^2)$, the prior is relatively non-informative with large variance. The priors on γ_j are assumed to be independent Bernoulli $p_i \in (0, 1)$ random variables as follows:

$$\begin{aligned} P(\gamma_j = 1) &= p_j \\ P(\gamma_j = 0) &= 1 - p_j \end{aligned} \tag{13}$$

where p_j is the prior parameter and $p_j = 0.5$ for a natural default choice.

George and McCulloch (1997) and George *et al.* (2008) use a default semiautomatic approach that sets $\tau_{kj} = c_k \widehat{\sigma}_{\phi_j}$ for $k = 0, 1$, where $\widehat{\sigma}_{\phi_j}$ is the OLS estimates of the standard error of ϕ_j in an unrestricted VAR and pre-selected constants c_0 and c_1 must be $c_0 < c_1$ e.g. $c_0 = 0.1$ and $c_1 = 10$ as used by George *et al.* (2008), Jochmann *et al.* (2010) and Jochmann *et al.* (2013). In this paper, we follow these values for the hyperparameters.

4. Monte Carlo simulations

This section presents Monte Carlo simulations to illustrate forecasting performances for both iterated and direct forecast methods using VAR models. Two 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 priors are compared: (1) inverted Normal-Wishart (INW) prior, (2) Minnesota (Minn) prior and (3) SSVS prior.

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

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

where Ψ is upper-triangular of the Choleski decomposition of $\Sigma^{-1} = \Psi\Psi^{-1}$.

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

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

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 Akaike information criterion (AIC) with $0 \leq p \leq 12$. The first method has the fixed lag length as $p = 4$ is the true lag length. We do not use the Bayesian information criterion (BIC) for the lag length determination since the BIC is generally choosing 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 eq. (4) for the direct forecasts. For each τ in eq. (5), 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 eq. (5) for each estimator by a recursive forecasting exercise of both an iterated and a direct multi-period forecasting method.

The Monte Carlo simulations for the multi-step forecasting are examined. Table 1 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 $\widehat{y}_{\tau+h}$ are computed for $\tau = 80$ to $\tau = 150 - h - 1$, then we calculate the MSFE defined as eq. (5). Each figure in Table 1 is the average over 100 sample MSFEs. Inspection of Table 1 suggests the following:

(1) Among the three estimators by the INW, the Minn and the SSVS, the SSVS produces the lowest MSFE in most cases, though in a very few cases 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 INW 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.

Model	Method	DGP 1 Forecast horizon 2	4	8	12	DGP 2 Forecast horizon 2	4	8	12
Lag = 4									
INW	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									
INW	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
Lag by AIC									
INW	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 1.
Monte Carlo simulation: average MSFEs

(3) For the INW and the SSVS, the iterated method of forecasts is 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 2 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, $\frac{MSFE(\text{direct})}{MSFE(\text{iterated})}$. 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 INW and the SSVS, the mean values of the relative MSFEs are always greater than 1 (means that the iterated forecasts outperform the direct forecasts), while for the Minn the ratios are either greater or less than 1.

(2) For the INW 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 INW are quite sensitive to the choice of the lag length. As the lag length is longer, the relative MSFEs by the INW are getting larger. However, the relative MSFE by the SSVS is 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.

Model		DGP 1 Forecast horizon 2	4	8	12	DGP 2 Forecast horizon 2	4	8	12
Lag = 4									
INW	Mean	1.043	1.125	1.277	1.335	1.031	1.090	1.221	1.269
	St dev	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
	St dev	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
	St dev	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									
INW	Mean	1.060	1.166	1.357	1.446	1.054	1.132	1.345	1.467
	St dev	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
	St dev	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
	St dev	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
Lag by AIC									
INW	Mean	1.068	1.275	1.729	1.875	1.140	1.253	1.456	1.687
	St dev	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
	St dev	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
	St dev	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

Table 2.
MSFE ratio

(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 lead to better performance than direct forecasts.

These findings show that for given DGPs the SSVS also has almost same properties in forecasting as the INW, 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 1953:1 to 2020:1 with sample size $T = 268$. 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 three-month Treasury bill. These data are obtained from the Federal Reserve Bank of St. Louis [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 larger 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) where 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 3 presents the MSFEs eq. (5) for the three-variable VAR with the lag-length 12 and chosen by the AIC for the INW, 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 12, 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 in ensuring 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 eq. (1), then these models are used to compute the forecast of the level of y_{t+h} such as $\widehat{y}_{t+h} = y_t + \sum_{i=1}^{h} \Delta \widehat{y}_{t+i|t}$ for the

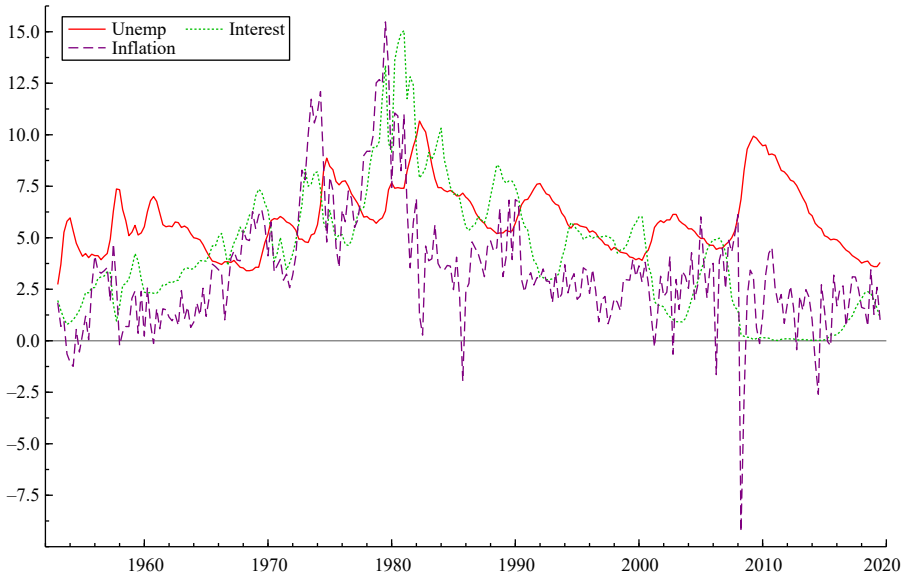

Figure 1.
Data: US unemployment rate, interest rate, inflation rate

Forecast horizon		2		4		8		12	
Model	Variable	Direct	Iterated	Direct	Iterated	Direct	Iterated	Direct	Iterated
Lag = 12									
INW	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
INW	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 3.
MSFEs for US data: level data

iterated forecast, and $\widehat{y}_{t+h} = y_t + \Delta\widehat{y}_{t+h}$ 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 4 presents the MSFEs of the case of the first

Forecast horizon		2		4		8		12	
Model	Variable	Direct	Iterated	Direct	Iterated	Direct	Iterated	Direct	Iterated
Lag = 11									
INW	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
INW	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

Table 4.
MSFEs for US data: first difference data

difference data. The MSFEs of the first difference data tend to have lower MSFEs than the case of the level data, particularly for the inflation rates. These results 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. Conclusions

This paper examines comparison of direct and iterated multistep forecasting performance using three estimators for VAR model – the inverted Normal-Wishart (INW) prior, the Minnesota prior and the SSVS prior. Theoretically, direct method is 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. 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 produced 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 INW 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 INW 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 smaller 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 INW, particularly with longer lag and longer forecast horizon.

Note

1. https://fred.stlouisfed.org

References

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

Bhansali, R.J. (1996), "Asymptotically efficient autoregressive model selection for multistep prediction", *Annals of the Institute of Statistical Mathematics*, Vol. 48 No. 3, pp. 577-602.

Bhansali, R. (1997), "Direct autoregressive predictors for multistep prediction: order selection and performance relative to the plug in predictors", *Statistica Sinica*, Vol. 7, pp. 425-449.

Chevillon, G. and Hendry, D.F. (2005), "Non-parametric direct multi-step estimation for forecasting economic processes", *International Journal of Forecasting*, Vol. 21 No. 2, pp. 201-218.

Clements, M.P. and Hendry, D.F. (1996), "Multi-step estimation for forecasting", *Oxford Bulletin of Economics and Statistics*, Vol. 58 No. 4, pp. 657-684.

Cogley, T. and Sargent, T.J. (2005), "Drifts and volatilities: monetary policies and outcomes in the post WWII US", *Review of Economic Dynamics*, Vol. 8 No. 2, pp. 262-302.

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

George, E.I. and McCulloch, R.E. (1997), "Approaches for Bayesian variable selection", *Statistica Sinica*, Vol. 7, pp. 339-373.

George, E.I., Sun, D. and Ni, S. (2008), "Bayesian stochastic search for VAR model restrictions", *Journal of Econometrics*, Vol. 142 No. 1, pp. 553-580.

Ing, C.-k. (2003), "Multistep prediction in autoregressive processes", *Econometric Theory*, Vol. 19 No. 2, pp. 254-279.

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*, Vol. 26 No. 2, pp. 326-347, doi: 10.1016/j.ijforecast.2009.11.002.

Jochmann, M., Koop, G., León-González, 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*, Vol. 28 No. 4, pp. 62-81.

Kang, I.B. (2003), "Multi-period forecasting using different models for different horizons: an application to US economic time series data", *International Journal of Forecasting*, Vol. 19 No. 3, pp. 387-400.

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*, Vol. 33 No. 4, pp. 997-1017.

Litterman, R.B. (1986), "Forecasting with Bayesian vector autoregressions: five year experience", *Journal of Business Economic Statistics*, Vol. 4 No. 1, pp. 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*, Vol. 135 No. 1, pp. 499-526.

Pesaran, M.H., Pick, A. and Timmermann, A. (2011), "Variable selection, estimation and inference for multi-period forecasting problems", *Journal of Econometrics*, Vol. 164 No. 1, pp. 173-187.

Primiceri, G.E. (2005), "Time varying structural vector autoregressions and monetary policy", *Review of Economic Studies*, Vol. 72 No. 3, pp. 821-852.

Sugita, K. (2018), *Evaluation of forecasting performance using Bayesian stochastic search variable selection in a vector autoregression*, Ryukyu Economics Working Paper #1, University of the Ryukyus.

Sugita, K. (2019a), *Forecasting with vector autoregressions by Bayesian model averaging*, Ryukyu Economics Working Paper #3, University of the Ryukyus.

Sugita, K. (2019b), *Forecasting with vector autoregressions using Bayesian variable selection methods: comparison of direct and iterated methods*, Ryukyu Economics Working Paper #2, University of the Ryukyus.

Corresponding author
Katsuhiro Sugita can be contacted at: ksugita@grs.u-ryukyu.ac.jp