

# Macroeconomic Forecasting with Large Stochastic Volatility in Mean VARs\*

Jamie L. Cross<sup>†</sup>      Chenghan Hou<sup>‡</sup>      Gary Koop<sup>§</sup>      Aubrey Poon<sup>¶</sup>

June 16, 2021

## Abstract

Vector autoregressions with stochastic volatility in both the conditional mean and variance are commonly used to estimate the macroeconomic effects of uncertainty shocks. Despite their popularity, intensive computational demands when estimating such models have made out-of-sample forecasting exercises impractical, particularly when working with large data sets. In this article, we propose an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior and predictive inference in such models that facilitates such exercises. The key insight underlying the algorithm is that the (log-)conditional densities of the log-volatilities possess Hessian matrices that are banded. This enables us to build upon recent advances in band and sparse matrix algorithms for state space models. In a simulation exercise, we evaluate the new algorithm numerically and establish its computational and statistical efficiency over a conventional particle filter based algorithm. Using macroeconomic data for the US we find that such models generally deliver more accurate point and density forecasts over a conventional benchmark in which stochastic volatility only enters the variance of the model.

*Keywords:* Bayesian VARs, Macroeconomic Forecasting, Stochastic Volatility in Mean, State Space Models, Uncertainty.

---

\*The authors gratefully acknowledge Knut Are Aastveit, Joscha Beckmann, Efram Castelnuovo, Joshua Chan, Todd Clark, Yunjong Eo, Dimitris Korobilis, Yong Song, Yiqiao Sun, Shaun Vahey, Benjamin Wong, Tomasz Wozniak, members of the 2021 World Meeting of the International Society for Bayesian Analysis, the 12th RCEA Bayesian Workshop, the 27th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics (SNDE), the Bank of Canada-Norges Bank Macroeconomic Workshop, and seminar participants at the Australian National University and the University of Melbourne for their comments and suggestions in the development of this research. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School. Chenghan Hou would like to acknowledge financial support by the National Natural Science Foundation of China (72003064).

<sup>†</sup>BI Norwegian Business School

<sup>‡</sup>Hunan University. Corresponding author. Email: chenghan.hou@hotmail.com

<sup>§</sup>University of Strathclyde

<sup>¶</sup>University of Strathclyde

# 1 Introduction

Unprecedented levels of macroeconomic uncertainty during the Great Recession and the more recent COVID-19 pandemic have resulted in numerous studies highlighting the adverse effects of uncertainty shocks both in the US and around the world.<sup>1</sup> From an econometric perspective, the workhorse model in measuring uncertainty and its effects on the economy is the stochastic volatility in mean vector autoregression (SVMVAR) model (Mumtaz and Zanetti, 2013; Mumtaz and Theodoridis, 2015, 2017a,b; Mumtaz and Surico, 2018; Carriero et al., 2018, 2020b). These models define uncertainty as common variation in the latent log-volatilities of each equation in the VAR system. To analyze the effects of uncertainty, this common component enters directly into the conditional mean of the model. This critical modeling feature does not come without costs. From a computational perspective, allowing the time-varying second moment of the shocks (i.e., uncertainty) to enter the conditional mean, implies that efficient algorithms for estimating simpler stochastic volatility (SV) models—in which the time-varying second moment does not enter the conditional mean—such as the auxiliary mixture sampler (Kim et al., 1998), can no longer be applied. As a result, scholars have relied on either Metropolis-within-Gibbs (Jacquier et al., 2002) or particle-Gibbs (Andrieu et al., 2010) algorithms to estimate such models; see, e.g., Mumtaz and Zanetti (2013) for the former and Carriero et al. (2018) for the latter. While such methods are perfectly suitable for measuring uncertainty and its effects on the economy, their computational demands have made out-of-sample forecasting exercises practically infeasible, particularly when working with large data sets. Resolving this problem is critical as policy makers, private sector businesses and international organizations all rely on macroeconomic forecasts when formulating their decisions.

In this paper we overcome this hurdle by developing an efficient Markov chain Monte Carlo (MCMC) algorithm that facilitates posterior and predictive inference in SVMVAR models. The key insight underlying the algorithm is that the (log-)conditional densities of the log-volatilities contain Hessian matrices that are banded. Having established this fact,

---

<sup>1</sup>The pre-pandemic literature is vast and we will not attempt to survey it here. Instead, we direct the interested reader to excellent surveys by Bloom (2014) or Castelnuovo et al. (2017) and references therein. For more recent papers on the effects of uncertainty during the COVID-19 pandemic see, e.g. Altig et al. (2020), Baker et al. (2020), Caggiano et al. (2020) and Carriero et al. (2020a).

we propose the use of an independence-chain Metropolis-Hastings algorithm in which the proposal distribution is a carefully selected Gaussian distribution. It is therefore similar to existing algorithms that have been used for the estimation of univariate models with non-Gaussian likelihoods (Shephard and Pitt, 1997), with the key difference that we consider a multivariate model and draw upon recent advances in band and sparse matrix algorithms (Rue et al., 2009; Chan and Jeliazkov, 2009; McCausland, 2012), which have been shown to perform efficiently in the estimation of various state space models (McCausland et al., 2011; Chan, 2017; Poon, 2018; Hou, 2020; Zhang et al., 2020; Leiva-Leon and Uzeda, 2021).

In a simulation exercise, we evaluate the new algorithm numerically and establish its computational and statistical efficiency against the state-of-the-art particle filter algorithm, known as particle Gibbs with ancestor sampling (PGAS) (Lindsten et al., 2014). The proposed algorithm is significantly faster than PGAS. For example, when estimating a small model with three variables in which each variable has its own idiosyncratic stochastic volatility component, as in Mumtaz and Zanetti (2013), our proposed method is about 20 times faster than PGAS. Importantly, this computational gain increases in both number of variables and number of stochastic volatility terms and is therefore well suited for estimating the SVMVAR model with a larger number of variables as in Carriero et al. (2018). We also find that the proposed algorithm exhibits superior mixing properties, thereby providing more numerically efficient estimates. This is especially true in high dimensional models, where we find that PGAS tends to mix poorly. This suggests that our algorithm is scalable and capable of handling large SVMVARs in a way that PGAS is not.

Using our new efficient algorithm we then carry out a macroeconomic forecasting exercise. We focus on three key macroeconomic indicators in the US economy: real GDP growth, personal consumption expenditures (PCE) inflation, and the effective federal funds rate, both in the pre-pandemic period: 1975Q1-2019Q4 and during the recent pandemic period: 2020Q1-2021Q1. We consider large (20 variable) and small (3 variable) variants of the SVMVAR model with different SVM structures, e.g. common and/or idiosyncratic volatilities, as well as different numbers of common volatility terms. As a benchmark model, we use a VAR in which the SV component does not enter the conditional mean (SVVAR), as this model has been shown to perform well against a wide range of alternative time series

models when forecasting macroeconomic variables (e.g. [Clark, 2011](#); [Clark and Ravazzolo, 2015](#); [Carriero et al., 2016](#)). Overall, we find that models with SVM components generally provide superior point and density forecast accuracy relative to those in which the SV does not enter the conditional mean. Focusing on the variables individually, we find that working with large datasets is especially useful when forecasting real GDP growth. The best model is typically one in which we specify idiosyncratic SVM terms for each of the three key macroeconomic indicators and a common SVM term for the remaining variables. Allowing the SVM terms to have a persistent effect on the conditional mean via extra lags in the model is also useful when forecasting real GDP growth. These technical insights are especially notable because it is precisely in these computationally complex models that our scalable algorithm is needed. When forecasting inflation, we find that large models also tend to provide superior point forecasts, however the increased estimation-uncertainty associated with such models results in the smaller, more parsimonious model, providing superior density forecasts. In contrast, the small SVMVAR generally provides the best interest rate forecasts. This suggests that while accounting for macroeconomic uncertainty via the SVMVAR generally improves both point and density forecast accuracy, the exact form of the model specification, i.e. VAR dimension and SVM structure, should be specified in accordance with the users objectives. Finally, when looking at the forecast results over time, we observe the general pattern that while a standard SV specification is sufficient during times of tranquility, such as the Great Moderation, specifying stochastic volatility in the mean is especially important in uncertain, unstable times, such as the Great Recession and recent pandemic period, during which macroeconomic forecasting is known to be especially difficult.

The rest of the paper is structured as follows. In [Section 2](#) we present the SVMVAR model and discuss our novel MCMC algorithm for posterior and predictive inference in such models. [Section 3](#) contains the simulation study which compares the computational and statistical efficiency of our algorithm against PGAS. [Section 4](#) contains the forecasting exercise in which we focus on results for GDP growth, PCE inflation and the Effective Federal Funds Rate. Finally, we conclude in [Section 5](#).

## 2 The Stochastic Volatility in Mean Vector Autoregressive Model

In this section, we define the SVMVAR and develop an efficient MCMC algorithm which allows for computationally-efficient posterior and predictive inference of the SVMVAR.

### 2.1 The model

The SVMVAR is a multivariate time series model for  $n$  variables of interest. We assume a general framework in which these variables can be partitioned into  $g$  groups with each group containing  $n_i$  variables that are driven by a single group-specific volatility for  $i = 1, \dots, g$ ,  $n = \sum_{i=1}^g n_i$ . This framework accommodates everything from the unrestricted SVMVAR ( $g = n$  and  $n_i = 1$  for all  $i$ ) through cases considered in the economic uncertainty literature where a small number of volatility processes, each estimated based on a large number of variables, are interpreted as measures of economic uncertainty. For instance, the unrestricted trivariate SVMVAR in which each variable has idiosyncratic volatility used in [Mumtaz and Zanetti \(2013\)](#) is obtained by specifying  $g = 3$  groups each with  $n_i = 1$  variables,  $i = 1, 2, 3$ . Similarly, the SVMVAR of [Carriero et al. \(2018\)](#) can be viewed as specifying a restricted SVMVAR with  $g = 2$  groups of  $n_1 = 18$  and  $n_2 = 12$  variables. The former are macroeconomic variables and the latter are financial variables and the two volatility processes are interpreted as macroeconomic and financial uncertainty, respectively. This general set-up also facilitates the estimation of more flexible specifications with different combinations of common and idiosyncratic volatilities which we later exploit in the forecasting exercise.

We use the following notation.  $y_{i,t}$  is an  $n_i \times 1$  vector that collects variables in the  $i$ th group at time  $t$  for  $t = 1, \dots, T$ .  $\mathbf{y}_t = (y'_{1,t}, \dots, y'_{g,t})'$  is an  $n \times 1$  vector with  $n = \sum_i n_i$ .  $h_{i,t}$  is the log-volatility of the  $i$ th group at time  $t$ . For later reference, we also define  $\mathbf{h}_{.,t} = (h_{1,t}, \dots, h_{g,t})'$  and  $\mathbf{h}_{i.} = (h_{i,1}, \dots, h_{i,T})'$ .

We write the SVMVAR as

$$\mathbf{B}_0 \mathbf{y}_t = \mathbf{b} + \sum_{i=1}^p \mathbf{B}_i \mathbf{y}_{t-i} + \sum_{j=0}^q \mathbf{A}_j \mathbf{h}_{.,t-j} + \boldsymbol{\epsilon}_t^y, \quad \boldsymbol{\epsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (1)$$

where  $\mathbf{B}_0$  is an  $n \times n$  lower triangular matrix with ones on its diagonal,  $\mathbf{b}$  is an  $n \times 1$  vector of intercepts and  $\mathbf{B}_1, \dots, \mathbf{B}_p$  are  $n \times n$  coefficient matrices. The coefficient matrices  $\mathbf{A}_0, \dots, \mathbf{A}_q$  are of dimension  $n \times g$  and capture the effects of the contemporaneous and lagged group-specific log-volatilities on the levels of the variables. The time-varying error covariance matrix is diagonal<sup>2</sup> and specified as

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \Omega_{1,t} & & \\ & \ddots & \\ & & \Omega_{g,t} \end{pmatrix}, \quad (2)$$

where  $\Omega_{i,t} = \text{diag}(\sigma_{i,1}^2 e^{h_{i,1,t}}, \dots, \sigma_{i,n_i}^2 e^{h_{i,n_i,t}})$ ,  $i = 1, \dots, g$ . The log-volatility associated with the  $i$ th group of variables is assumed to evolve according to a stationary AR(1) process

$$h_{i,t} = \phi_i h_{i,t-1} + \epsilon_{i,t}^h, \quad \epsilon_{i,t}^h \sim \mathcal{N}(0, \sigma_{h,i}^2), \quad (3)$$

where  $|\phi_i| < 1$  and the initial state is initialized with  $h_{i,1} \sim \mathcal{N}(0, \sigma_{h,i}^2 / (1 - \phi_i^2))$ .

## 2.2 Prior

We specify a Normal prior for  $\mathbf{B} = (\mathbf{b}, \mathbf{B}_1, \dots, \mathbf{B}_p)$  based on the popular Minnesota prior which has been shown to be a reliable choice when forecasting macroeconomic variables, particularly when working with large datasets (see, e.g., [Carriero et al., 2009, 2011](#); [Bańbura et al., 2010](#); [Koop and Korobilis, 2010](#); [Koop, 2013](#); [Karlsson, 2013](#); [Carriero et al., 2015](#); [Cross et al., 2020](#)).<sup>3</sup> In recent years, VAR researchers have found forecasts to be improved

---

<sup>2</sup>Note that this assumption does not restrict the contemporaneous relationships between the variables since these are modeled by  $\mathbf{B}_0$ . The diagonality assumption allows for equation by equation estimation of the model which greatly speeds up computation, see [Carriero et al. \(2019\)](#).

<sup>3</sup>It is also worth noting that the MCMC algorithm developed in this paper (with minor modifications) can handle any conditionally Normal prior including a wide range of global-local shrinkage priors such as the Lasso and Horseshoe priors.

by choosing prior hyperparameters optimally (see, e.g., [Giannone et al., 2005](#); [Chan et al., 2019, 2020](#); [Cross et al., 2020](#); [Chan, 2021](#)). In this paper, we therefore develop methods suitable for estimating key prior hyperparameters within the context of SVMVARs.<sup>4</sup> The means in our Normal prior are zero. The prior covariance matrix is diagonal. For the variances, we first define an  $n \times (1 + n^2p)$  variance matrix  $\mathbf{V}_b = (\mathbf{v}_b, \mathbf{V}_{b,1}, \dots, \mathbf{V}_{b,p})$ , where  $\mathbf{v}_b$  and  $\mathbf{V}_{b,i}$  are of dimensions  $n \times 1$  and  $p \times p$  and collect the prior variances of corresponding parameters in  $\mathbf{b}$  and  $\mathbf{B}_i$  respectively. We assume a noninformative prior for the VAR intercepts by setting the variance vector  $\mathbf{v}_b = (100 \dots, 100)'$ . For the VAR coefficients, we use a Minnesota prior form:

$$\mathbf{B}_l^{i,j} \sim \mathcal{N}(0, \mathbf{V}_l^{i,j}),$$

where  $\mathbf{B}_l^{i,j}$  and  $\mathbf{V}_{b,l}^{i,j}$  are the  $(i, j)$ th element of  $\mathbf{B}_l$  and  $\mathbf{V}_{b,l}$  and

$$\mathbf{V}_l^{i,j} = \begin{cases} \frac{\pi_1}{l^2}, & \text{for } i = j, \\ \frac{\pi_1 \pi_2 s_i}{l^2 s_j}, & \text{for } i \neq j. \end{cases}$$

Following a common Minnesota prior strategy, the scale parameters  $s_i^2$  are chosen to be the residual variances of the AR( $p$ ) models for variable  $i$  for  $i = 1, \dots, n$ .<sup>5</sup> Instead of setting the overall shrinkage parameter  $\pi_1$  and the cross-variable shrinkage parameter  $\pi_2$  to fixed values, we treat them as unknown parameters to be estimated using noninformative priors

$$\pi_1 \propto c_1, \quad \pi_2 \propto c_2.$$

We use Normal-gamma priors for the SVM coefficient matrix,  $\mathbf{A} = (\mathbf{A}_0, \dots, \mathbf{A}_q)$  and the contemporaneous matrix  $\mathbf{B}_0$ . Specifically, let  $a_i$ ,  $i = 1, \dots, n \times g \times (q + 1)$  and  $b_{0,j}$ ,

---

<sup>4</sup>Given the need for computational efficiency in the SVMVAR, it is worth highlighting that the methods we develop for doing so integrate out the VAR coefficients which leads to fast mixing of the MCMC algorithm.

<sup>5</sup>As we are using quarterly data in our forecasting exercise, the VAR lag length is set at  $p = 4$ .

$j = 1, \dots, \frac{n(n-1)}{2}$  be the free elements in  $\mathbf{A}$  and  $\mathbf{B}_0$ , then we assume

$$\begin{aligned} a_i &\sim \mathcal{N}(0, V_i^a), & V_i^a &\sim \mathcal{G}(\nu_a, \frac{\nu_a \delta_a}{2}), & i &= 1, \dots, n \times g \times (q+1) \\ b_{0,j} &\sim \mathcal{N}(0, V_j^b), & V_j^b &\sim \mathcal{G}(\nu_b, \frac{\nu_b \delta_b}{2}), & j &= 1, \dots, \frac{n(n-1)}{2}. \end{aligned}$$

Furthermore, we estimate the hyperparameters in the Normal-gamma prior and assume the following priors for them:

$$\begin{aligned} \delta_a &\sim \mathcal{G}(c_1^a, c_2^a), & \nu_a &\sim \mathcal{E}(\mu_a), \\ \delta_b &\sim \mathcal{G}(c_1^b, c_2^b), & \nu_b &\sim \mathcal{E}(\mu_b), \end{aligned}$$

and we set  $c_1^a = 0.001$ ,  $c_2^a = 0.001$ ,  $\mu_a = 1$ ,  $c_1^{b_0} = 0.001$ ,  $c_2^{b_0} = 0.001$  and  $\mu_{b_0} = 1$ .

For the variable-specific variance  $\sigma_{i,j}^2$ , AR coefficients and variances of the group-specific log-volatility, we use the following priors:

$$\begin{aligned} \sigma_{i,j}^2 &\sim \mathcal{IG}(\alpha_{y,ij}, \gamma_{y,ij}) \text{ for } i = 1, \dots, g \text{ and } j = 1, \dots, n_i, \\ \phi_i &\sim \mathcal{N}(\phi_{i,0}, V_{\phi_i}) \text{ for } |\phi_i| < 1, \text{ for } i = 1, \dots, g, \\ \sigma_{h,i}^2 &\sim \mathcal{IG}(\alpha_{h,i}, \gamma_{h,i}) \text{ for } i = 1, \dots, g. \end{aligned}$$

We set the shape parameters as  $\alpha_{y,ij} = 10$  and  $\alpha_{h,i} = 10$ . The scale parameters,  $\gamma_{y,ij}$ , are set so as to match the prior mean of  $\sigma_{i,j}^2$  at the residual variance of the AR( $p$ ) model of the variable  $j$  in the  $i$ th block. For the other scale parameters,  $\gamma_{h,i}$ , we let  $\gamma_{h,i} = 0.1^2(\alpha_{h,i} - 1)$  which implies that the prior mean of  $\sigma_{h,i}^2$  is  $0.1^2$ . For the AR coefficients, we set  $\phi_{i,0} = 0.95$  and  $V_{\phi_i} = 0.2^2$ .

### 2.3 An efficient MCMC algorithm for the SVMVAR

We develop an efficient MCMC algorithm for drawing from the full conditional posterior distributions of the parameters in the SVMVAR. Note that, conditional on  $\mathbf{h}$ , the model becomes a multivariate Normal model. Hence, standard methods exist for drawing from

$\mathbf{B} = (\mathbf{b}, \mathbf{B}_1, \dots, \mathbf{B}_p)$ ,  $\mathbf{A} = (\mathbf{A}_0, \dots, \mathbf{A}_q)$  and  $\mathbf{B}_0$ . The contributions of this paper lie in the methods for drawing  $\mathbf{h}$  and some of the prior hyperparameters. Accordingly, we provide detailed derivations for the latter and describe the former in Appendix A.

### 2.3.1 Drawing $\mathbf{h}$

Since the log-volatilities enter the SVMVAR model in both the conditional mean and the time-varying covariance matrix, the efficient auxiliary mixture sampler developed in Kim et al. (1998) cannot be applied. Early papers by Mumtaz and Zanetti (2013) overcame this difficulty by using a date-by-date independence-chain Metropolis step as described in Jacquier et al. (2002). More recently, the literature has turned to the use of particle MCMC (PMCMC) methods, e.g. Carriero et al. (2018). While perfectly feasible for in-sample analysis, it is well known that these algorithms can be computationally inefficient both in the sense of being computationally costly (i.e. computational cost is proportional to  $T$ ) and, in the case of particle filters, mixing poorly due to the path degeneracy issue. In this paper, we develop an MCMC algorithm that typically is less computationally costly (i.e. computational cost is proportional to  $g$ ) and we demonstrate it mixes better. We do so by first showing that the conditional posteriors of the volatility processes,  $(\mathbf{h}_{1,\cdot}, \dots, \mathbf{h}_{g,\cdot})$ , involves Hessian matrices that are banded. We then draw upon recent developments in band and sparse matrix algorithms, to produce an efficient MCMC algorithm for estimating SVMVARs.

For expository purposes, we work with simplified version of the SVMVAR:

$$\mathbf{y}_t = \sum_{i=0}^q \mathbf{A}_i \mathbf{h}_{\cdot, t-i} + \boldsymbol{\epsilon}_t^y, \quad \boldsymbol{\epsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t). \quad (4)$$

Furthermore, we set the the variable-specific variance  $\sigma_{i,j}^2 = 1$  for all  $i, j$  for expositional simplicity.

The proposed algorithm is designed for sequentially drawing the log-volatilities  $\mathbf{h}_{1,\cdot}, \dots, \mathbf{h}_{g,\cdot}$  from their corresponding full conditional posterior distributions. We first introduce some notation. Let  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ ,  $\tilde{\mathbf{h}}_{i,t} = (h_{i,t}, \dots, h_{i,t-q})'$  and define  $\tilde{\mathbf{A}}_i$  to be an  $n \times (q+1)$  matrix with its  $k$ th column being the  $i$ th column of  $\mathbf{A}_{k-1}$ .

To derive the full conditional distribution of  $\mathbf{h}_{j\cdot} = (h_{j,1}, \dots, h_{j,T})'$  for  $j = 1, \dots, g$ , we write equation (4) as

$$\mathbf{y}_t = \sum_{i=1}^g \tilde{\mathbf{A}}_i \tilde{\mathbf{h}}_{i,t} + \boldsymbol{\epsilon}_t^y, \quad \boldsymbol{\epsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t). \quad (5)$$

Rearranging this equation by putting all terms independent of  $\mathbf{h}_{j\cdot}$  on the left-hand side leads to

$$\tilde{\mathbf{y}}_t^j = \tilde{\mathbf{A}}_j \tilde{\mathbf{h}}_{j,t} + \boldsymbol{\epsilon}_t^y, \quad \boldsymbol{\epsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (6)$$

where

$$\tilde{\mathbf{y}}_t^j = \mathbf{y}_t - \sum_{i \neq j} \tilde{\mathbf{A}}_i \tilde{\mathbf{h}}_{i,t}.$$

Since  $(\tilde{\mathbf{y}}_1^j, \dots, \tilde{\mathbf{y}}_T^j)$  are conditionally independent, the log conditional distribution of  $\mathbf{h}_{j\cdot}$  can be expressed as

$$\begin{aligned} \log p(\mathbf{h}_{j\cdot} | \mathbf{y}) &= \log p(\mathbf{y} | \mathbf{h}_{j\cdot}) + \log p(\mathbf{h}_{j\cdot}) + c_1, \\ &= \sum_{t=1}^T \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t}) + \log p(\mathbf{h}_{j\cdot}) + c_2, \end{aligned} \quad (7)$$

where  $c_1$  and  $c_2$  are normalizing constants. We suppress the other conditioning arguments except for  $\mathbf{h}_{j\cdot}$  for notational convenience.

The conditional posterior in (7) involves a likelihood component and a prior component. The prior component is based on (3). The likelihood component can be obtained by noting that  $\tilde{\mathbf{y}}_t^j$  is Normally distributed (see equation (6)), which gives

$$\log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}_t| - \frac{1}{2} \left( \tilde{\mathbf{y}}_t^j - \tilde{\mathbf{A}}_j \tilde{\mathbf{h}}_{j,t} \right)' \boldsymbol{\Sigma}_t^{-1} \left( \tilde{\mathbf{y}}_t^j - \tilde{\mathbf{A}}_j \tilde{\mathbf{h}}_{j,t} \right) + c_3. \quad (8)$$

To further simplify equation (8), we write  $\tilde{\mathbf{A}}_j = (\tilde{\mathbf{A}}'_{j,1}, \dots, \tilde{\mathbf{A}}'_{j,g})'$  and  $\tilde{\mathbf{y}}_t^j = (\tilde{y}'_{1,t}, \dots, \tilde{y}'_{g,t})'$ , where  $\tilde{\mathbf{A}}_{j,k}$  and  $\tilde{y}'_{k,t}$  are of dimensions  $n_k \times (q+1)$  and  $n_k \times 1$  respectively. Then it can be

shown that equation (8) can be simplified to:

$$\begin{aligned} \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t}) &= -\frac{n_j h_{j,t}}{2} - \frac{1}{2} e^{-h_{j,t}} (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t})' (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t}) \\ &\quad - \frac{1}{2} \sum_{i \neq j} e^{-h_{i,t}} (\tilde{\mathbf{y}}_{i,t}^j - \tilde{\mathbf{A}}_{j,i} \tilde{\mathbf{h}}_{j,t})' (\tilde{\mathbf{y}}_{i,t}^j - \tilde{\mathbf{A}}_{j,i} \tilde{\mathbf{h}}_{j,t}) + c_4, \end{aligned} \quad (9)$$

where both  $c_3$  and  $c_4$  are normalizing constants.

To sample  $\mathbf{h}_{j,\cdot}$ , we propose an acceptance-rejection Metropolis-Hasting step with a Normal proposal with its mean and covariance matrix set to be the mode and the negative inverse of the Hessian of the  $\log p(\mathbf{h}_{j,\cdot} | \mathbf{y})$  evaluated at the mode. To find the mode, the Newton-Raphson method can be applied, which requires us to derive the gradient and Hessian of the log-density  $\log p(\mathbf{h}_{i,\cdot} | \mathbf{y}) = \sum_{t=1}^T \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})$  with respect to  $\mathbf{h}_{i,\cdot}$ . To this end, we will first derive the gradient and the Hessian of  $\log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})$  with respect to  $\tilde{\mathbf{h}}_{j,t}$ .

Let  $\tilde{f}^{j,t}$  be the  $(q+1) \times 1$  gradient vector and  $\tilde{G}^{j,t}$  be the  $(q+1) \times (q+1)$  Hessian of the log density function  $\log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})$ .<sup>6</sup>  $\tilde{f}^{j,t}$  and  $\tilde{G}^{j,t}$  are given by

$$\begin{aligned} \tilde{f}^{j,t} &= -\frac{1}{2} \left( n_j - e^{-h_{j,t}} (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t})' (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t}) \right) \mathbf{d}_j \\ &\quad + \sum_{i=1}^g e^{-h_{i,t}} \tilde{\mathbf{A}}'_{j,i} \left( \tilde{\mathbf{y}}_{i,t}^j - \tilde{\mathbf{A}}_{j,i} \tilde{\mathbf{h}}_{j,t} \right), \end{aligned} \quad (10)$$

$$\begin{aligned} \tilde{G}^{j,t} &= -\frac{1}{2} e^{-h_{j,t}} (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t})' (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t}) \mathbf{d}_j \mathbf{d}'_j \\ &\quad - e^{-h_{j,t}} \tilde{\mathbf{A}}'_{j,j} (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t}) \mathbf{d}'_j - e^{-h_{j,t}} \mathbf{d}_j (\tilde{\mathbf{y}}_{j,t}^j - \tilde{\mathbf{A}}_{j,j} \tilde{\mathbf{h}}_{j,t})' \tilde{\mathbf{A}}_{j,j} \\ &\quad - \sum_{i=1}^g e^{-h_{i,t}} \tilde{\mathbf{A}}'_{j,i} \tilde{\mathbf{A}}_{j,i}, \end{aligned} \quad (11)$$

where  $\mathbf{d}_j$  is the  $j$ th column of the identity matrix with dimension  $q+1$ . Using equations (10) and (11), we can construct the gradient and the Hessian of the log likelihood

---

<sup>6</sup>To be specific, the  $k$ th element of  $\tilde{f}^{j,t}$  is  $\tilde{f}_k^{j,t} = \frac{\partial \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})}{\partial h_{j,t-k-1}}$  and the  $(k,l)$ th element of  $\tilde{G}^{j,t}$  is  $\tilde{G}_{k,l}^{j,t} = \frac{\partial^2 \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})}{\partial h_{j,t-k-1} \partial h_{j,t-l-1}}$  and  $\tilde{f}_k^{j,t} = 0$  and  $\tilde{G}_{k,l}^{j,t} = 0$  for  $t-k-1 < 0$  or  $t-l-1 < 0$ .

$\sum_{t=1}^T \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})$  with respect to  $\mathbf{h}_{j,\cdot}$ . To be specific, we have

$$\mathbf{f}^j = \begin{pmatrix} f_1^j \\ f_2^j \\ \vdots \\ f_T^j \end{pmatrix} := \sum_{t=1}^T \frac{\partial \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})}{\partial \mathbf{h}_{j,\cdot}}, \quad (12)$$

$$\mathbf{G}^j = \begin{pmatrix} G_{1,1}^j & \cdots & G_{q,1}^j & 0 & \cdots & 0 \\ \vdots & G_{2,2}^j & \ddots & \ddots & \ddots & \vdots \\ G_{q+1,1}^j & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & G_{T,T-q}^j \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & G_{T,T-q}^j & \cdots & G_{T,T}^j \end{pmatrix} := \sum_{t=1}^T \frac{\partial^2 \log p(\tilde{\mathbf{y}}_t^j | \tilde{\mathbf{h}}_{j,t})}{\partial \mathbf{h}_{j,\cdot} \partial \mathbf{h}_{j,\cdot}'} \quad (13)$$

where  $f_t^j = \sum_{i=0}^q \tilde{f}_{i+1}^{j,t+i}$  and  $G_{t,t-l}^j = \sum_{i=0}^{q-l} \tilde{G}_{l+i+1,i+1}^{j,t+i}$ .

Next, we derive the first and second order derivatives of  $\log p(\mathbf{h}_{j,\cdot})$ . Given the specification in equation (3), we can show that the log conditional density of  $\mathbf{h}_{j,\cdot}$  is given by

$$\log p(\mathbf{h}_{j,\cdot}) = -\frac{1}{2} \mathbf{h}_{j,\cdot}' \mathbf{H}_j' \mathbf{S}_j^{-1} \mathbf{H}_j \mathbf{h}_{j,\cdot} + c_5,$$

where

$$\mathbf{H}_j = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\phi_j & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\phi_j & 1 \end{pmatrix}, \quad \mathbf{S}_j = \begin{pmatrix} \frac{\sigma_{h,j}^2}{1-\phi_j^2} & 0 & \cdots & 0 \\ 0 & \sigma_{h,j}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{h,j}^2 \end{pmatrix}$$

and  $c_5$  is a normalizing constant. Thus, it follows that

$$\frac{\partial \log p(\mathbf{h}_{j,\cdot})}{\partial \mathbf{h}_{j,\cdot}} = -\mathbf{H}_j' \mathbf{S}_j^{-1} \mathbf{H}_j \mathbf{h}_{j,\cdot}, \quad (14)$$

$$\frac{\partial^2 \log p(\mathbf{h}_{j,\cdot})}{\partial \mathbf{h}_{j,\cdot}^2} = -\mathbf{H}_j' \mathbf{S}_j^{-1} \mathbf{H}_j. \quad (15)$$

Combining the results obtained in equations (12) - (15) gives the gradient and the negative Hessian of  $\log p(\mathbf{h}_{j,\cdot}|\mathbf{y})$ :

$$\begin{aligned}\frac{\partial \log p(\mathbf{h}_j|\mathbf{y})}{\partial \mathbf{h}_j} &= \mathbf{f}_j = \mathbf{f}^j - \mathbf{H}'_j \mathbf{S}_j^{-1} \mathbf{H}_j \mathbf{h}_{j,\cdot}, \\ -\frac{\partial^2 \log p(\mathbf{h}_j|\mathbf{y})}{\partial \mathbf{h}_j^2} &= \mathbf{K}_j = \mathbf{H}'_j \mathbf{S}_j^{-1} \mathbf{H}_j - \mathbf{G}^j.\end{aligned}$$

To find the mode  $\hat{\mathbf{h}}_{j,\cdot}$ , we use the Newton-Raphson method and iteratively update

$$\mathbf{h}_j^{(s+1)} = \mathbf{h}_j^{(s)} + \mathbf{K}_j^{-1} \mathbf{f}_j, \quad \text{for } s = 1, 2, \dots, \quad (16)$$

until some convergence criterion is reached, e.g., when  $\|\mathbf{h}_j^{(s+1)} - \mathbf{h}_j^{(s)}\| < \epsilon$ . Given this mode and the negative Hessian evaluated at the mode  $\mathbf{K}_j$ , the Gaussian distribution  $\mathcal{N}(\hat{\mathbf{h}}_{j,\cdot}, \mathbf{K}_j^{-1})$  is then used as our proposal distribution in the acceptance-rejection Metropolis-Hastings step. Note that the resulting  $\mathbf{K}_j$  is not guaranteed to be positive definite. If this is the case, we replace  $\mathbf{K}_j$  by  $\tilde{\mathbf{K}}_j = \mathbf{K}_j + (\delta - \lambda_{\min})\mathbf{I}$ , where  $\lambda_{\min}$  is the smallest eigenvalue of  $\mathbf{K}_j$  and  $\delta$  is a small positive number. It can be shown that  $\tilde{\mathbf{K}}_j$  is always positive definite. Moreover, it can be seen from equations (10) and (15) that the precision matrix  $\mathbf{K}_j$  is a banded matrix. Thus draws from the proposal distribution  $\mathcal{N}(\hat{\mathbf{h}}_{j,\cdot}, \mathbf{K}_j^{-1})$  can be efficiently obtained by applying the precision sampler of [Chan and Jeliazkov \(2009\)](#).

## 2.4 Drawing $\pi$

It is straightforward to derive the posterior of  $\pi = (\pi_1, \pi_2)$  conditional on the VAR coefficients  $(\mathbf{B}_1, \dots, \mathbf{B}_p)$  and the other parameters. However, MCMC efficiency can be improved by integrating out the VAR coefficients. We do this in this sub-section and develop a Metropolis-Hastings algorithm for drawing from the resulting distribution.

Write equation (1) as

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t^y, \quad \boldsymbol{\epsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \quad (17)$$

where  $\tilde{\mathbf{y}}_t = \mathbf{B}_0 \mathbf{y}_t - \sum_{j=0}^q \mathbf{A}_j \mathbf{h}_{\cdot, t-j}$ ,  $\tilde{\mathbf{X}}_t = \mathbf{I}_n \otimes (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$  and  $\boldsymbol{\beta} = \text{vec}((\mathbf{b}, \mathbf{B}_1, \dots, \mathbf{B}_p)')$ .

Stacking equation (17) over  $t = 1, \dots, T$ , gives

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\epsilon}^y, \quad \boldsymbol{\epsilon}^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where  $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}'_1, \dots, \tilde{\mathbf{y}}'_T)'$ ,  $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}'_1, \dots, \tilde{\mathbf{X}}'_T)'$ . We write the Normal prior for the VAR coefficients as  $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta^\pi)$ , where the diagonal covariance matrix has its diagonal elements  $\text{vec}(\mathbf{V}'_b)$ . Note that this depends on  $\pi = (\pi_1, \pi_2)$ . For simplicity, we suppress all other conditioning arguments except  $\boldsymbol{\beta}$  and write the full conditional distribution of  $\pi$  as

$$\begin{aligned} p(\pi|\mathbf{y}, \boldsymbol{\beta}) &\propto p(\tilde{\mathbf{y}}|\boldsymbol{\beta}, \pi)p(\boldsymbol{\beta}) \\ &\propto |\mathbf{V}_\beta^\pi|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}'\mathbf{K}_\beta\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{K}_\beta\hat{\boldsymbol{\beta}}\right)\right) \\ &\propto |\mathbf{V}_\beta^\pi|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'\mathbf{K}_\beta\hat{\boldsymbol{\beta}}\right) \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{K}_\beta(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right), \end{aligned}$$

where  $\mathbf{K}_\beta = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \mathbf{V}_\beta^{\pi-1})$  and  $\hat{\boldsymbol{\beta}} = \mathbf{K}_\beta^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{y}}$ . It can be seen that the VAR coefficients  $\boldsymbol{\beta}$  can be integrated out analytically leading to

$$\begin{aligned} p(\pi|\mathbf{y}) &\propto |\mathbf{V}_\beta^\pi|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'\mathbf{K}_\beta\hat{\boldsymbol{\beta}}\right) \int \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathbf{K}_\beta(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right) d\boldsymbol{\beta} \\ &= |\mathbf{K}_\beta|^{-\frac{1}{2}} |\mathbf{V}_\beta^\pi|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'\mathbf{K}_\beta\hat{\boldsymbol{\beta}}\right). \end{aligned}$$

Since this distribution is nonstandard, we implement a Metropolis-Hastings step to obtain draws of  $\pi$  using a log-Normal proposal distribution. Denote the current draw by  $\pi^o = (\pi_1^o, \pi_2^o)$ . A candidate draw is taken from:

$$\begin{aligned} \pi_1^* &= \exp(x_1^*), \quad \text{with } x_1^* \sim \mathcal{N}(\log \pi_1, v_1), \\ \pi_2^* &= \exp(x_2^*), \quad \text{with } x_2^* \sim \mathcal{N}(\log \pi_2, v_2). \end{aligned}$$

The probability of accepting the candidate draw

$$\alpha(\pi^o, \pi^*) = \min\left\{1, \frac{p(\pi^*|\mathbf{y})}{p(\pi^o|\mathbf{y})} \times \frac{\pi_1^*\pi_2^*}{\pi_1^o\pi_2^o}\right\}.$$

In our forecasting exercise, we set  $v_1 = v_2 = 0.05$ .

The details of the other blocks of the MCMC algorithm are given in Appendix A.

### 3 Simulation Study

In this section, we investigate the performance of our algorithm using artificial data in terms of computation time and MCMC efficiency. We compare our algorithm with the particle Gibbs ancestor sampling (PGAS) algorithm of Lindsten et al. (2014) which has been used in previous stochastic volatility in mean papers such as Carriero et al. (2018). Since the algorithm is for sampling of the latent (log-)volatilities, we work with the simplified SVMVAR in (4) as opposed to the general model in (1). For simplicity, we consider the case where each group has the same number of variables, i.e.,  $\tilde{n} = n_1 = \dots = n_g$ . We then use the model defined in equation (4) to simulate a single data set for each of several different configurations of  $T$ ,  $g$  and  $\tilde{n}$ . The lag order for the log-volatilities in the mean is set at  $q = 2$ . The elements of  $\mathbf{A}$  are drawn from a Normal distribution  $\mathcal{N}(0, 0.5)$ . We set  $(\phi_i, \sigma_{i,h}^2) = (0.95, 0.1)$ ,  $i = 1, \dots, g$ . For the PGAS, we use the bootstrap proposal in the internal particle filter. Computation is done using MATLAB on a desktop with an Intel Core i7-7700 @4.20GHz processor.

#### 3.1 Computational efficiency

Table 1 reports the computation times for obtaining 1000 posterior draws using our proposed method and PGAS. We denote PGAS- $N$  as the PGAS method using  $N$  particles. It is evident that our proposed method is much faster than PGAS for almost every configuration of  $(\tilde{n}, g, T)$ . For example, our proposed method is about 20 times faster than PGAS-50 for estimating a small model  $(\tilde{n}, g, T) = (1, 3, 100)$ . It is worth mentioning that the computational cost of our proposed method is roughly proportional to  $g$ . This is because the proposed method is designed to sequentially sample  $g$  full conditional distributions corresponding to  $\mathbf{h}_{1,\cdot}, \dots, \mathbf{h}_{g,\cdot}$  in each MCMC draw and all  $T$  terms in each of these distributions are drawn jointly which can be quickly done. In contrast the computational cost of the

PGAS is proportional to  $T$  since the particle filtering process in the PGAS needs to sequentially approximate  $T$  filtering distributions for the latent variables  $\mathbf{h}_{1, \dots, \mathbf{h}_T}$ . In practice, most of the time series data used in economics and finance has  $g$  being much smaller than  $T$ . For example, in our forecasting exercise, we consider 20 US quarterly time series for the period 1960Q1-2020Q1. Hence,  $T = 241$  but  $g \leq 20$ . [Carriero et al. \(2018\)](#) has  $g = 2$ . This highlights the practical importance of our proposed method in improving estimation efficiency for real economic problems.

Table 1: Computation times (in minutes) to obtain 1000 posterior draws.

	$T = 100$			$T = 300$			$T = 500$		
	$g = 3$	$g = 10$	$g = 20$	$g = 3$	$g = 10$	$g = 20$	$g = 3$	$g = 10$	$g = 20$
$\tilde{n} = 1$									
proposed method	0.1	0.6	1.7	0.4	1.5	4.3	0.6	2.5	6.4
PGAS-50	1.9	2.0	2.2	5.5	6.5	8.1	9.2	12.0	16.9
PGAS-100	3.5	3.8	4.3	10.8	13.1	16.7	18.7	25.4	33.9
PGAS-300	10.8	12.2	13.3	34.9	41.9	53.0	60.2	80.0	110.5
$\tilde{n} = 3$									
proposed method	0.1	0.7	2.4	0.4	2.0	5.6	0.8	3.1	8.7
PGAS-50	1.8	2.0	2.4	5.5	6.4	8.7	9.3	12.4	17.6
PGAS-100	3.5	4.0	4.7	10.8	13.7	17.7	18.5	25.8	35.9
PGAS-300	11.0	12.0	14.3	34.7	42.5	54.8	59.9	82.0	110.6
$\tilde{n} = 5$									
proposed method	0.2	0.9	3.4	0.5	2.3	7.2	0.8	3.8	11.7
PGAS-50	1.8	2.1	2.5	5.5	6.6	9.1	9.4	12.5	18.3
PGAS-100	3.6	4.0	4.9	11.0	14.0	18.1	18.8	25.8	36.3
PGAS-300	10.8	12.6	15.1	35.2	43.3	54.9	62.5	84.6	112.7

Notes: The proposed method is our Acceptance-Rejection Metropolis-Hastings algorithm presented in Section 2.3. The particle Gibbs ancestor sampling (PGAS) algorithm is developed in [Lindsten et al. \(2014\)](#). We denote PGAS- $N$  as the PGAS method using  $N$  particles, and use  $\tilde{n}, g$  and  $T$  to respectively denote the number of variables in each group, number of groups (equal to the number of stochastic volatility terms) and number of time series observations.

### 3.2 MCMC efficiency

The only cases where PGAS is comparable to our algorithm in terms of the time taken to produce 1,000 draws is when  $T$  is small,  $g$  is large and a small number of particles are taken. See, for instance, results in Table 1 for PGAS-50 when  $T = 100$  and  $g = 20$ . However, this takes no account of MCMC mixing. We investigate this issue by reporting inefficiency

factors for each of the parameters in the model. The inefficiency factor is defined as

$$1 + 2 \sum_{l=1}^L \rho_l,$$

where  $\rho_l$  is the sample autocorrelation at lag length  $l$  obtained from the MCMC draws.  $L$  is chosen to be large enough so that the autocorrelation tapers off. The inefficiency factor is used to measure the relative efficiency loss incurred by using correlated MCMC draws as opposed to independent draws.

The SVM model has many parameters and each will have an inefficiency factor. We therefore summarize them using boxplots. The inefficiency factors for data sets simulated from  $(\tilde{n}, g, T) = (1, 3, 100)$  and  $(\tilde{n}, g, T) = (5, 20, 500)$  are shown in Figure 1 and Figure 2, respectively. These configurations are chosen as the smallest and largest data generating processes in our simulation study. Results for other configurations lie between these two extremes. The center line of each box is the median, the lower and upper edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum. For each approach, the inefficiency factor is computed using 10000 retained MCMC draws after a burnin period of 5000.

It is evident that our proposed method performs as well as PGAS for the smallest case of  $(\tilde{n}, g, T) = (1, 3, 100)$ . It is interesting to note that increasing the particle numbers used in PGAS only slightly reduces the inefficiency factors for drawing the volatilities. However, in the larger case of  $(\tilde{n}, g, T) = (5, 20, 500)$ , our proposed method performs much better than PGAS. As shown in Figure 2, inefficiency factors for  $\mathbf{h}$  for some artificial data sets produced by PGAS-50 and PGAS-100 go as high as 10000. Given that we are only taking 10000 draws from each algorithm this indicates that the posterior draws for some elements  $\mathbf{h}$  are perfectly correlated. Such poor mixing rates are a result of the path degeneracy issue in the PGAS and this seems to be inevitable for high-dimensional problems. Using a larger number of particles only partially alleviates the path degeneracy issue. It is evident that our proposed method mixes much better even than PGAS-300.

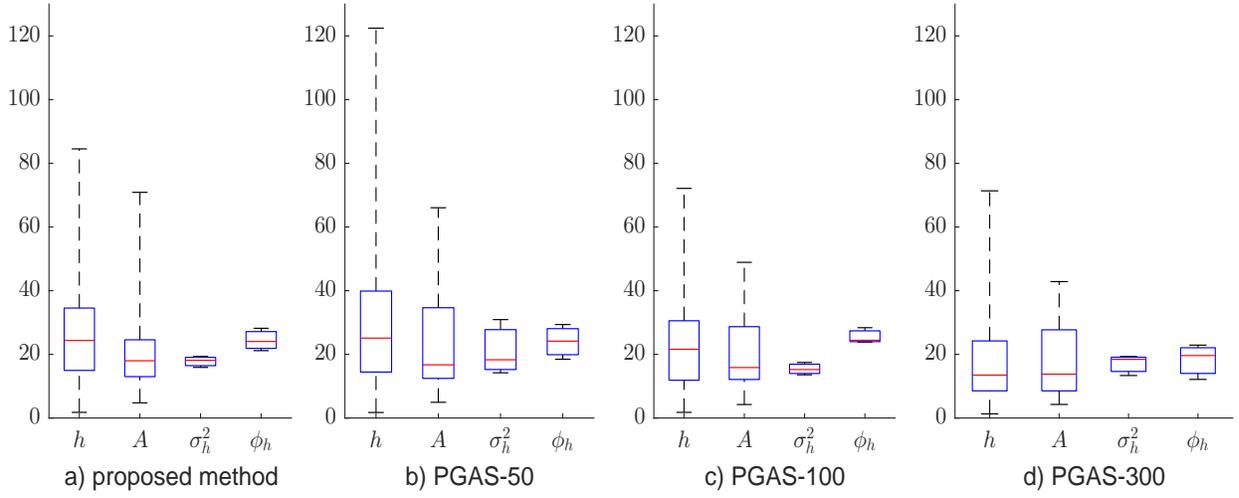


Figure 1: Boxplots of the inefficiency factors for model with  $(\tilde{n}, g, T) = (1, 3, 100)$ .

Notes: The center line of each box is the median, the lower and upper edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum.

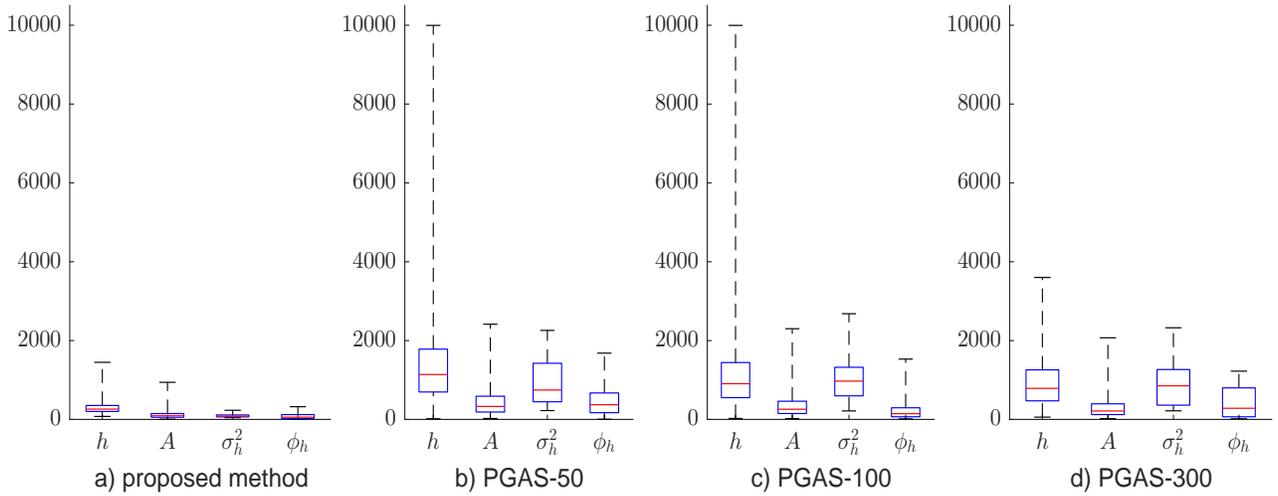


Figure 2: Boxplots of the inefficiency factors for model with  $(\tilde{n}, g, T) = (5, 20, 500)$ .

Notes: The center line of each box is the median, the lower and upper edges of the box represent the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum.

## 4 Forecasting Exercise

We focus on forecasting GDP growth, PCE inflation and the effective federal funds rate over the period from 1960Q1 to 2021Q1. Since numerous studies have documented that larger VARs tend to forecast at least as well as smaller VARs (e.g. [Bańbura et al., 2010](#); [Koop, 2013](#); [Carriero et al., 2019](#)), we consider both a large version which contains the 20 variables listed in [Table 2](#) and a small version that contains only the key macroeconomic indicators being forecast. The set of variables in [Table 2](#) are similar to those used in [Carriero et al. \(2019\)](#). All of the data was obtained from the Federal Reserve Bank of St. Louis' Quarterly Database for Macroeconomic Research (FRED-QD).

Table 2: Description of the data.

FRED-ID	Series Name	Transformation
GDPC1	Real Gross Domestic Production	$\Delta$ log
PCECC96	Real Personal Consumption Expenditures	$\Delta$ log
CMRMTSPLx	Real Manufacturing and Trade Industries Sales	$\Delta$ log
INDPRO	Industrial Production Index	$\Delta$ log
CUMFNS	Capacity Utilization: Manufacturing	no transformation
UNRATE	Civilian Unemployment Rate	no transformation
PAYEMS	All Employees: Total Nonfarm	$\Delta$ log
CES060000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	log
CES060000008	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing	$\Delta$ log
WPSFD49207	Producer Price Index by Commodity for Final Demand: Finished Goods	$\Delta$ log
PPIACO	Producer Price Index for All Commodities	$\Delta$ log
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	$\Delta$ log
FEDFUNDS	Effective Federal Funds Rate	no transformation
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	log
S&P 500	S&P's Common Stock Price Index: Composite	$\Delta$ log
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	$\Delta$ log
TB3SMFFM	3-Month Treasury Constant Maturity Minus Federal Funds Rate	no transformation
T5YFFM	5-Year Treasury Constant Maturity Minus Federal Funds Rate	no transformation
AAAFFM	Moody's Seasoned Aaa Corporate Bond Minus Federal Funds Rate	no transformation
AMDMNOx	Real Manufacturers' New Orders: Durable Goods	$\Delta$ log

Notes: The data was obtained from the Federal Reserve Bank of St. Louis' Quarterly Database for Macroeconomic Research (FRED-QD). The first column contains the identifier of the series at [fred.stlouisfed.org](http://fred.stlouisfed.org). The second column is the series name. The third column specifies any transformations. There are two types of transformations. First, log denotes the natural logarithm. Second,  $\Delta$  log denotes the change (first-difference) of the natural logarithm.

When forecasting we use the first 15 years of data to estimate the models and then consider two forecast evaluation periods 1975Q1 - 2019Q4 (pre-pandemic) and 2020Q1 - 2021Q1 (pandemic). In addition to the unrestricted SVMVAR with  $n_i = 1$  for all  $i$  (so that each equation has its own volatility), we also consider three large SVMVAR model variants in which we restrict the SV process in different ways. First, we assume there is

one common volatility process affecting all variables ( $g = 1$  and  $n_1 = 20$ ). Second, we assume there is one common volatility for the three core variables and a second common volatility for the remaining 17 variables ( $g = 2$ ,  $n_1 = 3$  and  $n_2 = 17$ ). Third, we assume each of the three core variables has its own volatility and the remaining variables share a common volatility ( $g = 4$ ,  $n_1 = n_2 = n_3 = 1$  and  $n_4 = 17$ ). Finally, as is standard in the forecasting literature, we use a simpler VAR with stochastic volatility (SVVAR) as a benchmark. The small SVVAR is used as a benchmark model, while the large SVVAR is used as a competitive model. We highlight that the SVVAR can be viewed as a SVMVAR in which we restrict  $\mathbf{A}_0, \dots, \mathbf{A}_q$  to be zero. Table 3 lists the competing models and their acronyms. In each model we set the VAR lag order at  $p = 4$  and consider forecast horizons of  $m = 1, 4, 8, 12$  quarters. For every SVMVAR, we set  $q = 0$  lags for the SVM component.<sup>7</sup>

Table 3: Competing models

Model	Description
Small SVVAR	3-variable SVVAR model with idiosyncratic SV
Large SVVAR	20-variable SVVAR model with idiosyncratic SV
Small SVMVAR	3-variable SVMVAR model with idiosyncratic SVM
Large SVMVAR	20-variable SVMVAR model with idiosyncratic SVM
Large SVMVAR-1SV	20-variable SVMVAR model with common SV
Large SVMVAR-2SV	20-variable SVMVAR model with two SV terms: One common SV for the 3 core variables and one for the remaining 17 variables
Large SVMVAR-4SV	20-variable SVMVAR model with four SV terms: Idiosyncratic SV for the 3 core variables and one common SV for the remaining 17 variables

Notes: The 3-variable models contain real GDP growth, personal consumption expenditures (PCE) inflation and the effective federal funds rate. The 20-variable models contain these three variables plus the additional 17 variables specified in Table 2.

## 4.1 Pre-pandemic Results

In this section we report the results for the pre-pandemic period: 1975Q1 - 2019Q4. Tables 4 and 5 respectively report the root mean squared forecast errors (RMSFEs) and cumulative log scores (i.e. sums of log-predictive likelihoods) of each variable. Results are benchmarked relative to the small SVVAR. Results of the Diebold-Mariano test of equal forecasting ability

<sup>7</sup>As a robustness check, in Section 4.3 we discuss results in which  $q = 2$  lags to see whether allowing for persistence in the volatility-in-mean process is important.

between the small SVVAR benchmark and an alternative model are indicated using \*\*\*,\*\* and \* to denote the 1%, 5% and 10% levels of significance, respectively.

The overall story told by these tables is that, with some exceptions, adding SVM generally improves forecasts of all three macroeconomic indicators. This statement holds particularly true beyond the one-step-ahead forecast horizon, where SVM models tend to dominate both small and large SVVARs both in terms of RMSFEs and log-predictive likelihoods. We now consider the forecast accuracy of each variable in some more detail.

Consider first GDP growth. With one exception, every SVMVAR provides superior RMSFEs and log-predictive likelihoods relative to the small SVVAR at all forecast horizons. The one exception is for a version of the model with a common volatility process for all 20 variables at the eight-month-ahead forecast horizon. We also observe a clear benefit to working with larger data sets when forecasting real GDP growth. At each horizon, a large VAR provides superior forecast accuracy to a smaller variant. As noted previously, the large SVVAR forecasts best at the one-step-ahead horizon, however the SVM models then improve upon the forecast accuracy at longer horizons. The SVMVAR with four volatility processes forecasts particularly well, although the unrestricted model with 20 distinct volatility processes is also competitive. This reinforces the point made above that larger values of  $g$  are associated with better forecasts. Overall, we conclude that the incorporation of SVM and working with large data sets leads to improved forecasts of GDP growth.

A similar story can be told for inflation, with SVM models generally outperforming the small SVVAR benchmark at all forecast horizons and for both forecast metrics. The unrestricted SVMVAR and restricted SVMVAR-4SV models again forecasts well in terms of point forecasts, however their density forecast results are mixed. In that case, a notable finding in relation to the issue of VAR dimension is that the small SVMVAR generally provides superior forecast accuracy, with large SVMVARs sometimes failing to outperform the benchmark. This is consistent with some findings in the inflation forecasting literature, e.g. [Faust and Wright \(2013\)](#), where simple parsimonious models are hard to beat. Overall, we conclude that the incorporation of SVM is generally useful when forecasting inflation. The VAR dimension, however, should be specified in accordance to the objective, i.e. large

dimensions can provide superior point forecasts, while small dimensions are better suited for density forecasts.

Finally, when forecasting the Effective Federal Funds rate, we again observe that specifying SVM generally improves upon simpler SV specification. There is also strong evidence in favor of using a smaller dataset, with the small SVMVAR providing the best density forecasts beyond the one-step-ahead horizon. At the the one-step-ahead horizon, however, the larger information set allows for greater forecast accuracy, with the large SVMVAR-4SV and large SVVAR models respectively providing the best point and density forecasts at that horizon. Overall, we conclude that the incorporation of SVM leads to improved forecasts of the interest rate. Similar to the case of inflation, the VAR dimension should be specified in accordance to the objective. In this case, large dimensions provide useful information for short-term forecasts, while small dimensions are better suited to longer term forecasts.

Having established that SVM is generally a useful modeling feature when forecasting each of the individual variables, we now determine which model is best across all three variables, on average. To that end, Table 6 contains the joint predictive likelihoods for the three variables. With a few exceptions we observe that models with some form of SVM tend to forecast substantially better than the small SVVAR at all forecast horizons. This result reinforces the story told when investigating the variables on an individual basis. Regarding forecast-horizon, we observe that the large SVVAR provides the best forecasts at the shortest forecast horizon, however adding SVM leads to clear improvements at longer horizons. On average, the unrestricted large SVMVAR and the restricted large SVMVAR with four volatility processes are the best forecasting models. This suggests that allowing for diversity across variables in terms of their volatilities is beneficial.

Table 4: Root Mean Squared Forecast Errors (RMSFEs) relative to a small SVVAR benchmark over the forecast evaluation period: 1975Q1-2019Q4

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	0.97*	0.96**	0.96**	0.95***	0.96
Large SVMVAR	0.89**	0.93**	<b>0.91</b>	<b>0.86*</b>	0.90
Large SVMVAR-1SV	0.89**	0.95**	0.98	0.89*	0.93
Large SVMVAR-2SV	0.90**	0.92**	0.92*	0.90	0.91
Large SVMVAR-4SV	<b>0.87**</b>	<b>0.92**</b>	0.92	0.87**	<b>0.89</b>
Large SVVAR	<b>0.87***</b>	0.95*	0.96	0.96	0.94
Inflation					
Small SVMVAR	0.98	0.96	0.88**	0.85**	0.92
Large SVMVAR	0.94***	<b>0.91**</b>	<b>0.83***</b>	<b>0.81**</b>	<b>0.87</b>
Large SVMVAR-1SV	0.99	1.02	0.95	0.91	0.97
Large SVMVAR-2SV	0.97	0.94	0.88*	0.90	0.92
Large SVMVAR-4SV	<b>0.93***</b>	<b>0.91***</b>	0.89**	0.92	0.91
Large SVVAR	1.01	1.07	1.03	0.97	1.02
Interest Rate					
Small SVMVAR	0.99	0.99	<b>0.93**</b>	<b>0.87***</b>	<b>0.94</b>
Large SVMVAR	0.95	1.02	0.98	0.93	0.97
Large SVMVAR-1SV	0.97	1.01	0.97	0.92	0.97
Large SVMVAR-2SV	0.98	1.11	1.16	1.18	1.11
Large SVMVAR-4SV	<b>0.93</b>	<b>0.98</b>	0.96	0.97	0.96
Large SVVAR	0.97	1.04	0.97	0.90	0.97

Notes: Model acronyms are explained in Table 3. Bold numbers indicate the best forecast performance at each horizon. Diebold-Mariano test are based on the benchmark Small SVVAR, \*\*\*, \*\* and \* denote statistically significant forecast improvements of a given model over the benchmark at the 1%, 5% and 10% level of significance, respectively.

Table 5: Cumulative Log predictive likelihoods relative to a small SVVAR benchmark over the forecast evaluation period: 1975Q1-2019Q4

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	2.99	6.12**	3.03	1.59	3.43
Large SVMVAR	13.78***	9.93**	7.71	7.21	9.66
Large SVMVAR-1SV	15.68**	9.22*	-1.62	5.92	7.30
Large SVMVAR-2SV	10.65*	9.86*	<b>10.63*</b>	8.01	9.79
Large SVMVAR-4SV	15.64**	<b>11.39**</b>	10.29*	<b>14.72***</b>	<b>13.01</b>
Large SVVAR	<b>18.78***</b>	7.08**	3.82	4.12	8.45
Inflation					
Small SVMVAR	0.27	<b>1.81</b>	8.21**	<b>9.13**</b>	4.86
Large SVMVAR	0.60	-1.27	<b>12.57**</b>	8.90	<b>5.20</b>
Large SVMVAR-1SV	-0.58	-14.34	5.76	2.28***	-1.72
Large SVMVAR-2SV	-19.59	1.09	4.43	3.94	-2.53
Large SVMVAR-4SV	-10.40	0.32	2.67	-0.19	-1.90
Large SVVAR	<b>1.65</b>	-4.77	0.76	-6.39	-2.19
Interest Rate					
Small SVMVAR	1.45	<b>9.74**</b>	<b>15.65***</b>	<b>24.58***</b>	<b>12.85</b>
Large SVMVAR	4.34	-7.06	-11.17	3.00	-2.72
Large SVMVAR-1SV	-63.84	1.27	17.51	9.28	-8.94
Large SVMVAR-2SV	-45.05	-3.56	-5.16	-8.61	-15.59
Large SVMVAR-4SV	-10.84	-6.29	-0.35	3.32	-3.54
Large SVVAR	<b>7.78</b>	-4.77	-4.64	5.94	1.08

Notes: Model acronyms are explained in Table 3. Bold numbers indicate the best forecast performance at each horizon. Diebold-Mariano test are based on the benchmark Small SVVAR, \*\*\*, \*\* and \* denote statistically significant forecast improvements of a given model over the benchmark at the 1%, 5% and 10% level of significance, respectively.

Table 6: Cumulative Joint Log Predictive Likelihoods (Three variables)

Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	2.46	2.19**	8.77***	5.00***	4.61
Large SVMVAR	13.29	<b>5.09</b>	<b>18.04</b>	<b>11.58</b>	<b>12.00</b>
Large SVMVAR-1SV	19.38**	-0.29	4.52***	7.83**	7.86
Large SVMVAR-2SV	-4.00	-5.72	4.87***	10.41***	1.39
Large SVMVAR-4SV	3.72	7.57***	11.33*	8.98	7.90
Large SVVAR	<b>22.80**</b>	2.44**	6.55**	-4.82	6.74

Notes: Model acronyms are explained in Table 3. Bold numbers indicate the best forecast performance at each horizon. Diebold-Mariano test are based on the benchmark Small SVVAR, \*\*\*,\*\* and \* denote statistically significant forecast improvements of a given model over the benchmark at the 1%, 5% and 10% level of significance, respectively.

Before proceeding to the results for the recent pandemic period, it is also interesting to investigate when the forecast improvements obtained by specifying SVM have occurred. To that end, Figure 3 presents the cumulative joint and individual variable log-predictive likelihoods across all four forecast horizons, benchmarked against the small SVVAR over the forecast evaluation period.

From a historical perspective, the forecast evaluation period can be divided into four epochs: i) the volatile period until the early 1980s, ii) the Great Moderation from the early 1980s to the Great Recession, iii) the Great Recession and subsequent recovery and iv) the pandemic of 2020, in addition to containing five NBER recessions. These episodes are highlighted in Figure 3 by shading the NBER recession dates in dark gray and the Great Moderation in light gray. With some exceptions, a general pattern is that most of the large SVMVARs handled the earliest period best, but then showed relatively little further improvement throughout the Great Moderation. When the financial crisis hit, the large SVMVARs also tended to forecast better than the other models, however, with the exception of real GDP, the relative forecast performance of most models then drops towards the end of the Great Recession. This suggests that specifying stochastic volatility in the mean is generally important in uncertain, unstable times, and less important during times of tranquility, where a standard SV specification is sufficient.

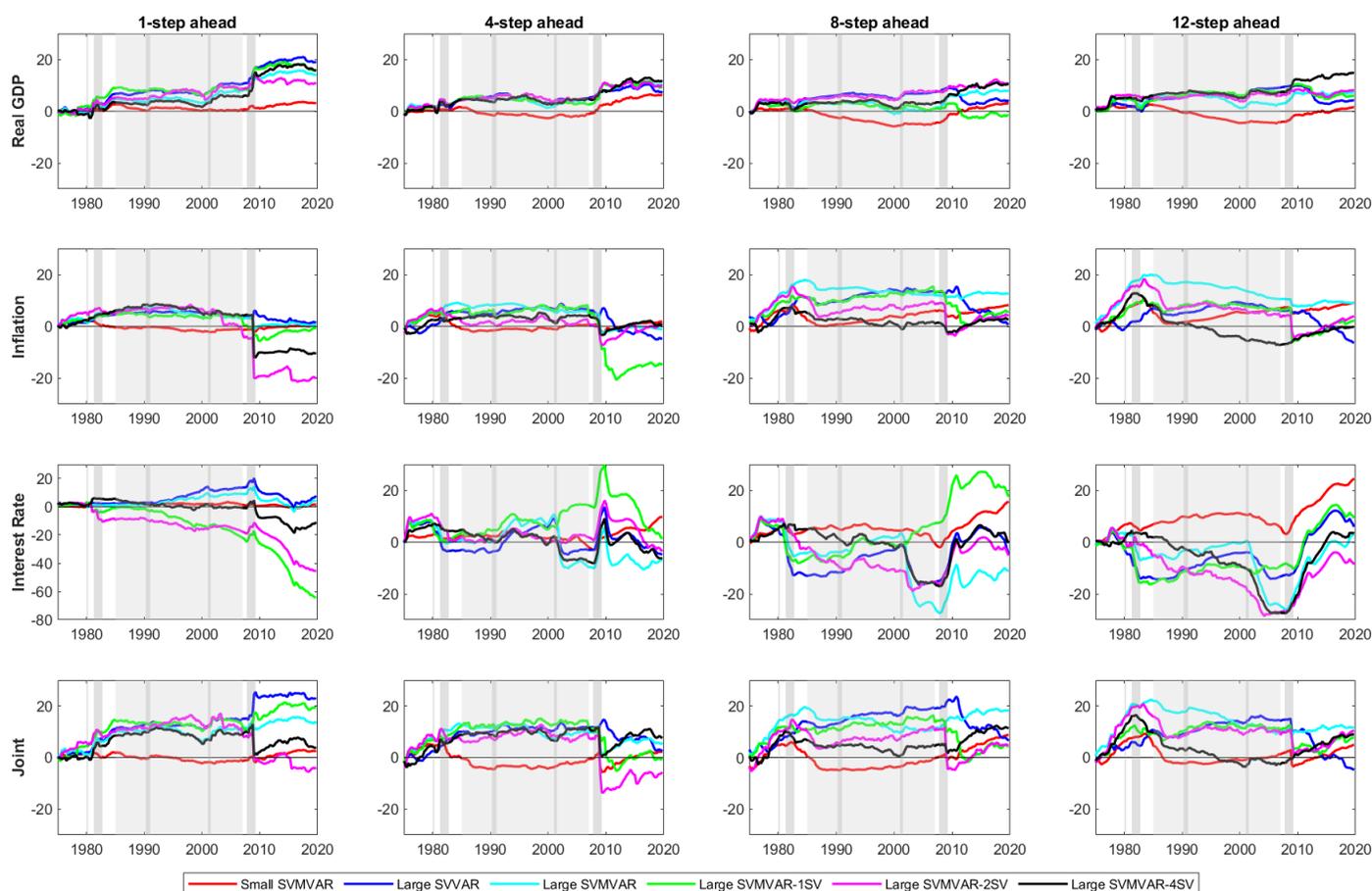


Figure 3: Cumulative joint and individual predictive likelihoods over the forecast evaluation period: 1975-2020

Notes: The panels show cumulative predictive likelihoods, relative to a small SVVAR benchmark over the forecast evaluation period: 1975-2020, at 1-, 4-, 8- and 12-step-ahead forecast horizons. The first, second and third rows contain the results for real GDP, inflation and the interest rate, respectively. The fourth row contains the joint cumulative predictive likelihood. The dark shaded regions represent the five NBER recession dates during this period: 1980Q1-Q2, 1981Q2-1982Q4, 1990Q2-1991Q1, 2001Q1-Q4, 2007Q4-2009Q2, and the light shaded region represents the Great Moderation period: 1985Q1-2006Q4. Model acronyms are explained in Table 3.

As discussed previously, the large SVVAR is often one of the better performing models when forecasting one-step-ahead. At longer horizons, however, this model is usually beaten by a comparable model which includes stochastic volatility in mean. A key insight from Figure 3 is that the large SVVAR has generally not forecast well since the financial crisis, while the large SVMVAR has done relatively well. This is especially notable for inflation.

We also previously found that, with some exceptions, large SVMVARs with restricted volatility processes, such as the large SVMVAR-4SV model, generally tend to forecast well. Focusing on the joint predictive likelihoods, we observe that these exceptions are almost always due to very brief periods of bad forecast performance at the time of the financial crisis. That is, there are several cases where forecast performance (relative to the small VAR benchmark) deteriorates abruptly for only a single quarter or two right at the time of the financial crisis. A detailed examination of results for the individual variables shows that inflation is the variable for which these deterioration are occurring. But even in these cases, this brief deterioration is subsequently replaced by period of relative improvement in forecast performance towards the end of the sample.

## 4.2 Pandemic Results

Point and density forecast results during the pandemic period are provided in Tables 7 and 8, respectively.<sup>8</sup> Overall we find that the SVMVAR models generally forecast better than those without SVM. This is particularly notable for the point forecasts, where the RMSFEs for SVMVARs are always lower than the SVVARs. In terms of individual variable performance, on average, the best SVM model improves upon the point forecast accuracy of the small SVVAR by 12% in the case of real GDP, 5% for inflation and 5% for the interest rate. The real GDP improvements are especially impressive given the relatively large fluctuations observed during this period. Turning to density forecasts, we again observe that the SVMVARs are generally superior over the pandemic period, however the large SVVAR is more competitive. It is especially interesting that the large SVMVARs with common volatility, i.e. SVMVAR-1SV and SVMVAR-2SV, are delivering relatively good forecast performance for during the pandemic. These common volatility models tended to forecast relatively poorly prior to 2020 but are among the best performing models during the pandemic. A likely explanation for this result is that there was a larger degree of co-movement in macroeconomic volatility during the pandemic. This would lead to a larger

---

<sup>8</sup>The Diebold-Mariano test statistic converges asymptotically to a normal distribution and theory predicts an over-rejection of the null hypothesis in small samples. Given the small number of observations during the COVID-19 pandemic, we have therefore omitted the Diebold-Mariano test statistics from Tables 7 and 8.

increase in predictive variances during the pandemic, and the common volatility model is designed to capture this exact phenomenon. Thus, an extreme realization would look less unlikely for the common volatility models leading to better values for the log predictive likelihoods. In summary, we are finding that specifying SVM generally improves accuracy when forecasting of macroeconomic variables and these gains are often substantial. During the pandemic period, we are finding that working with large SVMVARs is important and it is precisely in these large models that our scalable MCMC algorithm are needed.

Table 7: Root Mean Square Forecast Errors during the pandemic period: 2020-2021

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	0.95	1.00	1.00	0.99	0.98
Large SVMVAR	<b>0.88</b>	<b>0.99</b>	<b>0.99</b>	0.99	<b>0.96</b>
Large SVMVAR-1SV	0.97	0.99	0.98	0.99	0.98
Large SVMVAR-2SV	1.03	0.99	0.98	<b>0.98</b>	1.00
Large SVMVAR-4SV	1.03	1.00	0.99	0.99	1.00
Large SVVAR	0.95	0.99	0.98	0.99	0.98
Inflation					
Small SVMVAR	<b>0.95</b>	0.95	0.92	0.98	0.95
Large SVMVAR	0.98	0.98	1.06	0.97	1.00
Large SVMVAR-1SV	0.96	0.96	0.78	0.96	<b>0.91</b>
Large SVMVAR-2SV	1.01	0.96	0.91	0.98	0.97
Large SVMVAR-4SV	1.14	<b>0.92</b>	<b>0.76</b>	<b>0.88</b>	0.93
Large SVVAR	1.16	0.94	0.89	1.12	1.03
Interest Rate					
Small SVMVAR	1.03	1.07	1.16	1.31	1.14
Large SVMVAR	<b>0.95</b>	0.89	0.96	<b>0.79</b>	0.90
Large SVMVAR-1SV	1.48	0.76	0.58	1.33	1.04
Large SVMVAR-2SV	1.50	<b>0.75</b>	<b>0.53</b>	1.01	0.95
Large SVMVAR-4SV	1.20	<b>0.75</b>	0.71	0.69	<b>0.84</b>
Large SVVAR	1.19	0.77	0.55	1.30	0.95

Notes: Model acronyms are explained in Table 3. Bold numbers indicate the best forecast performance at each horizon.

Table 8: Cumulative Log predictive likelihoods during the pandemic period: 2020-2021

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	-7.29	-22.56	-24.47	4.28	-12.51
Large SVMVAR	-30.82	-8.34	-28.17	-14.41	-20.43
Large SVMVAR-1SV	<b>-0.64</b>	19.74	9.76	26.52	13.84
Large SVMVAR-2SV	-4.54	<b>29.20</b>	<b>14.09</b>	<b>27.84</b>	<b>16.65</b>
Large SVMVAR-4SV	-128.82	-87.49	-37.00	-20.67	-68.50
Large SVVAR	-7.38	-8.59	-21.90	4.20	-8.42
Inflation					
Small SVMVAR	2.13	0.02	0.27	0.31	0.68
Large SVMVAR	3.11	-0.06	-0.27	-0.05	0.68
Large SVMVAR-1SV	<b>4.48</b>	-1.40	0.70	0.03	<b>0.95</b>
Large SVMVAR-2SV	3.45	-2.02	-0.76	0.18	0.21
Large SVMVAR-4SV	1.35	-0.01	<b>1.15</b>	<b>0.53</b>	0.75
Large SVVAR	1.77	<b>0.18</b>	0.44	-0.49	0.48
Interest Rate					
Small SVMVAR	-2.43	-5.77	-4.20	-1.60	-3.50
Large SVMVAR	-5.84	-2.06	0.65	1.14	-1.53
Large SVMVAR-1SV	<b>3.17</b>	<b>7.03</b>	3.71	-0.52	3.35
Large SVMVAR-2SV	2.31	6.51	4.52	0.27	<b>3.40</b>
Large SVMVAR-4SV	-4.17	2.88	3.19	<b>1.83</b>	0.93
Large SVVAR	-2.85	1.64	<b>4.65</b>	-0.40	0.76

Notes: Model acronyms are explained in Table 3. Bold numbers indicate the best forecast performance at each horizon.

### 4.3 Robustness

In our main analysis, we set  $q = 0$  lags in the SVM component of the measurement equation in the VAR model. As a robustness check, we re-did the analysis with  $q = 2$  lags as in [Mumtaz and Zanetti \(2013\)](#) and [Carriero et al. \(2018\)](#), to determine whether accounting

for possible persistence in the volatility-in-mean process impacts the forecast performance. The results, which are reported in Appendix B, are broadly consistent with those presented in the main analysis, however some notable exceptions exist. When forecasting real GDP, we find that the large SVMVAR-4SV model systematically provides the best point and density forecast at each horizon. This model also generally provides superior forecast accuracy over the best SVMVAR model with  $q = 0$  lags. This suggests that allowing persistence in the volatility-in-mean process is important for real GDP, however the same conclusion does not always hold for inflation and the interest rate. In the case of point forecasts, we find that the RMSFEs for the best SVMVAR with  $q = 2$  lags are competitive with, but are generally no better than, those with  $q = 0$ . In the case of density forecasts, we also find that specifying  $q = 2$  lags generates more accurate longer run inflation forecasts, but provides no improvements for the interest rate over the best SVMVAR with  $q = 0$ . Taken together, this suggests that specifying persistence in the volatility-in-mean process is useful when forecasting real GDP, however, the more parsimonious model without any such persistence is sufficient when forecasting inflation and the interest rate. Moreover, the qualitative conclusion that SVMVARs can improve upon the conventional SVVAR benchmark is generally robust to lag selection in the volatility-in-mean process.

Our discussion in the main analysis relates to forecast performance relative to the small SVVAR benchmark. Probability integral transforms (PITs) can be used to shed light on absolute forecast performance. We have calculated the PITs for each variable, model and forecast horizon. The interested reader can find them in Appendix B. We note here only that the SVMVARs tend to produce predictive densities which are well calibrated. Furthermore, where there are small departures from Uniformity in the PITs, they are less than for the comparable SVVAR models. Inflation and the interest rate forecasts are particularly well-calibrated. For GDP growth there is some evidence of poor calibration in the small models indicating that their predictive variances are too large. In light of the discussion relating to the pandemic, this makes sense. The larger predictive variances produced by the benchmark model, which were problematic in pre-pandemic times, were of benefit during the pandemic, in the sense that the extreme pandemic realizations did not lie so far out in the tails of the predictive distribution.

## 5 Conclusion

Macroeconomic VARs with stochastic volatility in both the conditional mean and variance have been widely used to document the adverse economic consequences of uncertainty in recent years, however intensive computational demand have made out-of-sample forecasting exercises practically infeasible, particularly when working with large data sets. In this paper, we develop an efficient MCMC algorithm for posterior and predictive inference in such models. The key insight underlying the algorithm is that the (log-)conditional densities of the (log-)volatilities possess Hessian matrices that are banded. This enables us to build upon recent advances in band and sparse matrix algorithms to efficiently sample the log-volatilities in SVMVAR models.

In a simulation exercise, we evaluate the new algorithm numerically and establish its computational and statistical efficiency over the state-of-the-art particle filter algorithm, particle Gibbs with ancestor sampling (PGAS). The proposed MCMC algorithm is shown to be significantly faster than PGAS. It also possesses superior mixing properties, thereby providing more numerically efficient estimates. This is especially true in high dimensional models, where we find that PGAS tends to mix poorly. This suggests that our algorithm is scalable and capable of handling large SVMVARs in a way that PGAS is not.

Using macroeconomic data for the US we find that such models generally provide superior point and density forecasts relative to a commonly used benchmark in which stochastic volatility only enters the variance of the model. One exception is for the one-step-ahead density forecasts in the pre-pandemic period, in which the large SVVAR model is tough to beat. When looking at the forecast results over time, however, we observe the general pattern that while a standard SV specification is sufficient during times of tranquility, such as the Great Moderation, specifying stochastic volatility in the mean is especially important in uncertain, unstable times, such as the Great Recession and recent pandemic period, during which macroeconomic forecasting is a notoriously difficult task. The SVMVAR might therefore be of interest to policy makers, private sector businesses and international organizations who all rely on macroeconomic forecasts when formulating their decisions.

## References

- D. Altig, S. Baker, J. M. Barrero, N. Bloom, P. Bunn, S. Chen, S. J. Davis, J. Leather, B. Meyer, E. Mihaylov, et al. Economic uncertainty before and during the COVID-19 pandemic. *Journal of Public Economics*, 191:104274, 2020.
- C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342, 2010.
- S. R. Baker, N. Bloom, S. J. Davis, and S. J. Terry. Covid-induced economic uncertainty. Technical report, National Bureau of Economic Research, 2020.
- M. Bańbura, D. Giannone, and L. Reichlin. Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92, 2010.
- N. Bloom. Fluctuations in uncertainty. *Journal of Economic Perspectives*, 28(2):153–76, 2014.
- G. Caggiano, E. Castelnuovo, and R. Kima. The global effects of Covid-19-induced uncertainty. *Economics Letters*, 194:109392, 2020.
- A. Carriero, G. Kapetanios, and M. Marcellino. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25(2):400–417, 2009.
- A. Carriero, G. Kapetanios, and M. Marcellino. Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761, 2011.
- A. Carriero, T. E. Clark, and M. Marcellino. Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73, 2015.
- A. Carriero, T. E. Clark, and M. Marcellino. Common drifting volatility in large Bayesian VARs. *Journal of Business and Economic Statistics*, 34(3):375–390, 2016.
- A. Carriero, T. E. Clark, and M. Marcellino. Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, 100(5):799–815, 2018.
- A. Carriero, T. E. Clark, and M. Marcellino. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1):137–154, 2019.
- A. Carriero, T. Clark, M. Marcellino, and E. Mertens. Measuring uncertainty and its effects in the COVID-19 era. Technical report, Federal Reserve Bank of Cleveland, 2020a.
- A. Carriero, T. E. Clark, and M. Marcellino. Assessing international commonality in macroeconomic uncertainty and its effects. *Journal of Applied Econometrics*, 35(3):273–293, 2020b.
- E. Castelnuovo, G. Lim, and G. Pellegrino. A short review of the recent literature on uncertainty. *Australian Economic Review*, 50(1):68–78, 2017.

- J. C. Chan. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business and Economic Statistics*, 35(1): 17–28, 2017.
- J. C. Chan. Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting*, 37(3):1212–1226, 2021. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2021.01.002>. URL <https://www.sciencedirect.com/science/article/pii/S0169207021000029>.
- J. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120, 2009.
- J. C. Chan, L. Jacobi, D. Zhu, et al. How sensitive are VAR forecasts to prior hyperparameters? An automated sensitivity analysis. *Advances in Econometrics*, 40:229–248, 2019.
- J. C. Chan, L. Jacobi, and D. Zhu. Efficient selection of hyperparameters in large Bayesian VARs using automatic differentiation. *Journal of Forecasting*, 39(6):934–943, 2020.
- T. E. Clark. Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341, 2011.
- T. E. Clark and F. Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2015.
- J. L. Cross, C. Hou, and A. Poon. Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915, 2020.
- J. Faust and J. Wright. Forecasting inflation. *Handbook of Economic Forecasting*, 2:2–56, 2013.
- D. Giannone, M. Lenza, and L. Reichlin. Prior selection for vector autoregressions. *Review of Economics and Statistics*, 27:436–451, 2005.
- C. Hou. Time-varying relationship between inflation and inflation uncertainty. *Oxford Bulletin of Economics and Statistics*, 82(1):83–124, 2020.
- E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 20(1):69–87, 2002.
- S. Karlsson. Forecasting with Bayesian vector autoregression. *Handbook of Economic Forecasting*, 2:791–897, 2013.
- S. Kim, N. Shephard, and S. Chib. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393, 1998.

- G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358, 2010.
- G. M. Koop. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203, 2013.
- D. Leiva-Leon and L. Uzeda. Endogenous time variation in vector autoregressions. *Review of Economic and Statistics*, 2021.
- F. Lindsten, M. Jordan, and T. Schon. Partical gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15:2145–2184, 2014.
- W. J. McCausland. The HESSIAN method: Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics*, 168(2):189–206, 2012.
- W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state–space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55(1):199–212, 2011.
- H. Mumtaz and P. Surico. Policy uncertainty and aggregate fluctuations. *Journal of Applied Econometrics*, 33(3):319–331, 2018.
- H. Mumtaz and K. Theodoridis. The international transmission of volatility shocks: An empirical analysis. *Journal of the European Economic Association*, 13(3):512–533, 2015.
- H. Mumtaz and K. Theodoridis. The changing transmission of uncertainty shocks in the US. *Journal of Business and Economic Statistics*, pages 1–14, 2017a.
- H. Mumtaz and K. Theodoridis. Common and country specific economic uncertainty. *Journal of International Economics*, 105:205–216, 2017b.
- H. Mumtaz and F. Zanetti. The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking*, 45(4):535–558, 2013.
- A. Poon. Assessing the synchronicity and nature of Australian state business cycles. *Economic Record*, 94(307):372–390, 2018.
- H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, 71(2):319–392, 2009.
- N. Shephard and M. K. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997.
- B. Zhang, J. C. Chan, and J. L. Cross. Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts. *International Journal of Forecasting*, 36(4):1318–1328, 2020.

## A Appendix: Additional Details of MCMC Algorithm

Drawing  $\mathbf{B} = (\mathbf{b}, \mathbf{B}_1, \dots, \mathbf{B}_p)$ ,  $\mathbf{A} = (\mathbf{A}_0, \dots, \mathbf{A}_q)$  and  $\mathbf{B}_0$

Note that the parameters in  $\Theta = (\mathbf{B}, \mathbf{A}, \mathbf{B}_0)$  are conditionally independent across equations and thus we can sample them equation by equation. To be specific, we first let  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$  and

$$\mathbf{U} = \begin{pmatrix} \mathbf{h}'_{,1} & \mathbf{0}_{1 \times g} & \cdots & \cdots & \mathbf{0}_{1 \times g} \\ \mathbf{h}'_{,2} & \mathbf{h}'_{,1} & \mathbf{0}_{1 \times g} & \cdots & \mathbf{0}_{1 \times g} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{h}'_{,T} & \mathbf{h}'_{,T-1} & \cdots & \cdots & \mathbf{h}'_{,T-q} \end{pmatrix}$$

then we can rewrite equation (1) more compactly as:

$$\mathbf{Y}\mathbf{B}'_0 = \mathbf{X}\mathbf{B}' + \mathbf{U}\mathbf{A}' + \mathbf{E}, \quad \text{vec}(\mathbf{E}') \sim \mathcal{N}(\mathbf{0}, \Sigma),$$

where  $\mathbf{X}$  is the  $T \times k$  matrix with its  $t$ th row being  $(1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})$  and  $\Sigma = \text{diag}(\Sigma_1, \dots, \Sigma_T)$ . Let  $\mathbf{Y}_{1:i}$  be the matrix collecting the first  $i$  columns of  $\mathbf{Y}$  and let  $\mathbf{y}_{i,\cdot}$  to be the  $i$ th column of  $\mathbf{Y}$ , then we can write the  $i$ th equation as

$$\mathbf{y}_{i,\cdot} = \mathbf{X}\mathbf{b}_i + \mathbf{U}\mathbf{a}_i - \mathbf{Y}_{1:i-1}\mathbf{b}_{0,i} + \mathbf{e}_i, \quad \mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_i),$$

where  $\mathbf{b}_i$ ,  $\mathbf{a}_i$  and  $\mathbf{e}_i$  are the  $i$ th columns of  $\mathbf{B}'$ ,  $\mathbf{A}'$  and  $\mathbf{E}$  respectively. The diagonal covariance matrix  $\mathbf{D}_i = \text{diag}(\Sigma_{1,i,i}, \dots, \Sigma_{T,i,i})$  has its  $t$ th diagonal element  $\Sigma_{t,i,i}$  being the  $i$ th diagonal element of  $\Sigma_t$ . The parameter vector associated with  $\mathbf{Y}_{1:i-1}$  is  $\mathbf{b}_{0,i} = (b_{0,i,1}, \dots, b_{0,i,i-1})'$ , where  $b_{0,i,j}$  is the  $(i, j)$ th element of  $\mathbf{B}_0$ . Thus, we have

$$\mathbf{y}_{i,\cdot} = \mathbf{W}_i\boldsymbol{\theta}_i + \mathbf{e}_i, \quad \mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_i),$$

where  $\boldsymbol{\theta}_i = (\mathbf{b}'_i, \mathbf{a}'_i, \mathbf{b}'_{0,i})'$  and  $\mathbf{W}_i = (\mathbf{X}, \mathbf{U}, \mathbf{Y}_{1:i-1})$ . As the elements in  $\Theta$  have independent Normal priors  $\boldsymbol{\theta}_i \sim \mathcal{N}(\mathbf{0}, \Phi_i)$ , where the covariance matrix  $\Phi_i$  can be constructed based on the the prior variances of  $\Theta$  defined in the body of the paper. Standard linear regression results can be applied to derive the full conditional distribution of  $\boldsymbol{\theta}_i$ :

$$(\boldsymbol{\theta}_i | \bullet) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_i, \mathbf{Q}_i^{-1})$$

with  $\bullet$  indicating the data and the remaining parameters, and the precision matrix and the mean are given by  $\mathbf{Q}_i = \mathbf{W}'\mathbf{D}_i^{-1}\mathbf{W} + \Phi_i^{-1}$  and  $\hat{\boldsymbol{\theta}}_i = \mathbf{Q}_i^{-1}\mathbf{W}'\mathbf{D}_i^{-1}\mathbf{y}_i$ .

### Drawing $(V_{i,j}^a, \delta_a, \nu_a)$ and $(V_{i,j}^b, \delta_b, \nu_b)$

Note that  $V_{i,j}^a$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, g(q+1)$ , are conditionally independent. The full conditional distribution of  $V_{i,j}^a$  is given by

$$\begin{aligned} p(V_{i,j}^a | a_{i,j}, \nu_a, \delta_a) &\propto p(a_{i,j} | V_{i,j}^a) p(V_{i,j}^a | \nu_a, \delta_a) \\ &\propto (V_{i,j}^a)^{-\frac{1}{2}} \exp\left(-\frac{a_{i,j}^2}{2V_{i,j}^a}\right) \times (V_{i,j}^a)^{\nu_a-1} \exp\left(-\frac{\nu_a \delta_a}{2} V_{i,j}^a\right) \\ &\propto (V_{i,j}^a)^{\nu_a-\frac{1}{2}-1} \exp\left(-\frac{1}{2}\left(\nu_a \delta_a V_{i,j}^a + \frac{a_{i,j}^2}{V_{i,j}^a}\right)\right), \end{aligned}$$

which is the kernel of a generalized inverse Gaussian distribution and thus we obtain

$$(V_{i,j}^a | a_{i,j}, \nu_a, \delta_a) \sim \mathcal{GIG}\left(\nu_a - \frac{1}{2}, \nu_a \delta_a, a_{i,j}^2\right).$$

To sample  $\delta_a$ , we know that the full conditional distribution of  $\delta_a$  is given by

$$\begin{aligned} p(\delta_a | \{V_{i,j}^a\}_{1 \leq i \leq n, 1 \leq j \leq g(q+1)}, c_1^a, c_2^a) &\propto \prod_{i,j} p(V_{i,j}^a | \nu_a, \delta_a) p(\delta_a | c_1^a, c_2^a) \\ &\propto \delta_a^{k_a \nu_a} \exp\left(-\frac{\nu_a \delta_a}{2} \sum_{i,j} V_{i,j}^a\right) \times \delta_a^{c_1^a-1} \exp(-c_2^a \delta_a) \\ &\propto \delta_a^{c_1^a + k_a \nu_a - 1} \exp\left(-\left(c_2^a + \frac{\nu_a}{2} \sum_{i,j} V_{i,j}^a\right) \delta_a\right), \end{aligned}$$

where  $k_a = ng(q+1)$ . This implies that

$$(\delta_a | \{V_{i,j}^a\}_{1 \leq i \leq n, 1 \leq j \leq g(q+1)}, c_1^a, c_2^a) \sim \mathcal{G}\left(c_1^a + k_a \nu_a, c_2^a + \frac{\nu_a}{2} \sum_{i,j} V_{i,j}^a\right).$$

Finally, the full conditional distribution of  $\nu_a$  is nonstandard and its log conditional density is given by

$$\begin{aligned} \log p(\nu_a | \{V_{i,j}^a\}_{1 \leq i \leq n, 1 \leq j \leq g(q+1)}, \delta_a, \mu_a) = & k_a \left( \nu_a \log \frac{\nu_a \delta_a}{2} - \log \Gamma(\nu_a) \right) + (\nu_a - 1) \sum_{i,j} \log V_{i,j}^a \\ & - \frac{\delta_a \nu_a}{2} \sum_{i,j} V_{i,j}^a - \log \mu_a - \frac{\nu_a}{\mu_a} + c_1, \end{aligned}$$

where  $\Gamma(\cdot)$  is the Gamma function and  $c_1$  is a normalizing constant. To obtain a proposal distribution, we first derive the first and second order derivatives of the log-density:

$$\begin{aligned} \frac{d}{d\nu_a} \log p(\nu_a | \{V_{i,j}^a\}_{1 \leq i \leq n, 1 \leq j \leq g(q+1)}, \delta_a, \mu_a) &= k_a \left( \log \frac{\nu_a \delta_a}{2} - \Psi(\nu_a) + 1 \right) + \sum_{i,j} \log V_{i,j}^a - \frac{\delta_a}{2} \sum_{i,j} V_{i,j}^a - \frac{1}{\mu_a}, \\ \frac{d^2}{d\nu_a^2} \log p(\nu_a | \{V_{i,j}^a\}_{1 \leq i \leq n, 1 \leq j \leq g(q+1)}, \delta_a, \mu_a) &= k_a \left( \frac{1}{\nu_a} - \Psi'(\nu_a) \right), \end{aligned}$$

where  $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$  and  $\Psi'(x) = \frac{d^2}{dx^2} \log \Gamma(x)$  are respectively the digamma and trigamma functions. Using the first and second derivatives derived above, we can use the Newton-Raphon method to obtain the mode  $\hat{\nu}_a$  and the negative Hessian evaluated at the mode,  $K_{\nu_a}$  of this log full conditional density. We then implement an independence chain Metropolis-Hastings step to draw  $\nu_a$  using the Normal proposal  $\mathcal{N}(\hat{\nu}_a, K_{\nu_a}^{-1})$ . Similar steps can be used for sampling  $(V_{i,j}^b, \delta_b, \nu_b)$ .

## Drawing $(\sigma_{i,j}^2, \phi_i, \sigma_{h,i}^2)$

To sample  $\sigma_{i,j}^2$ , we let

$$\widehat{\mathbf{E}} = (\mathbf{YB}'_0 - \mathbf{XB}' - \mathbf{UA}') \odot \begin{pmatrix} e^{-\frac{1}{2}h_{1,1}} \mathbf{1}_{1 \times n_1} & e^{-\frac{1}{2}h_{2,1}} \mathbf{1}_{1 \times n_2} & \dots & e^{-\frac{1}{2}h_{g,1}} \mathbf{1}_{1 \times n_g} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-\frac{1}{2}h_{1,T}} \mathbf{1}_{1 \times n_1} & e^{-\frac{1}{2}h_{2,T}} \mathbf{1}_{1 \times n_2} & \dots & e^{-\frac{1}{2}h_{g,T}} \mathbf{1}_{1 \times n_g} \end{pmatrix}$$

where  $\odot$  is the Hadamard product. Let  $\mathbf{e}_{i,j}$  be the  $(j + \sum_{k=1}^{i-1} n_k)$ th column of  $\widehat{\mathbf{E}}$ , then it can be shown that the full conditional distribution can be expressed as follows:

$$(\sigma_{i,j}^2 | \mathbf{e}_{i,j}) \sim \mathcal{IG} \left( \alpha_{y,ij} + \frac{T}{2}, \gamma_{y,ij} + \frac{1}{2} \mathbf{e}'_{i,j} \mathbf{e}_{i,j} \right).$$

Next, an independence chain Metropolis-Hastings step is used to obtain draws of  $\phi_i$  using a truncated Normal proposal  $\mathcal{N}(\widehat{\phi}_i, D_{\phi}^{-1}) \mathbf{1}(|\phi_i| < 1)$ , where  $D_{\phi,i} = \sum_{t=2}^T h_{i,t}^2 / \sigma_{h,i}^2 + V_{\phi_i}^{-1}$  and

$\hat{\phi}_i = D_\phi^{-1} \left( \sum_{t=2}^T h_{i,t} h_{i,t-1} / \sigma_{i,h}^2 + V_{\phi_i} \phi_{i,0} \right)$ . Given a draw from the proposal  $\phi_i^*$ , the acceptance probability is given by  $\min \left\{ 1, \left( \frac{1-\phi_i^{*2}}{1-\phi_i^2} \right)^{\frac{1}{2}} \times \exp \left( -\frac{h_{i,1}^2 (\phi_i^2 - \phi_i^{*2})}{2\sigma_{h,i}^2} \right) \right\}$ .

The full conditional distribution of  $\sigma_{h,i}^2$  is inverse-Gamma:

$$(\sigma_{h,i}^2 | \phi_i) \sim \mathcal{IG} \left( \alpha_{h,i} + \frac{T}{2}, \gamma_{h,i} + \frac{1}{2} \left( (1 - \phi_i^2) h_{i,1}^2 + \sum_{t=2}^T (h_{i,t} - \phi_i h_{i,t-1})^2 \right) \right).$$

## B Appendix: Additional Results

### B.1 Results setting $q = 2$

Table 9: RMSFEs for SVMVARs with  $q = 2$

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	0.97*	0.95***	0.95***	0.94*	0.95
Large SVMVAR	0.89**	0.95	0.91	0.86*	0.90
Large SVMVAR-1SV	1.34	5.90	9.63	2.00	4.72
Large SVMVAR-2SV	0.96**	0.97**	0.92*	0.87	0.93
Large SVMVAR-4SV	<b>0.88***</b>	<b>0.91***</b>	<b>0.90**</b>	<b>0.85**</b>	<b>0.89</b>
Inflation					
Small SVMVAR	0.98	0.95	0.89**	0.86**	0.92
Large SVMVAR	0.94**	0.94	0.84**	<b>0.80***</b>	0.88
Large SVMVAR-1SV	2.10	6.12	5.60	3.43	4.31
Large SVMVAR-2SV	0.99	<b>0.90**</b>	0.83**	0.84*	0.89
Large SVMVAR-4SV	<b>0.92***</b>	0.91**	<b>0.82***</b>	0.81**	<b>0.86</b>
Interest Rate					
Small SVMVAR	0.98	<b>0.99</b>	<b>0.93**</b>	<b>0.87***</b>	<b>0.94</b>
Large SVMVAR	0.96	1.05	1.00	0.92	0.98
Large SVMVAR-1SV	6.14	25.39	5.77	10.16	11.87
Large SVMVAR-2SV	1.00	1.03	0.97	0.94	0.98
Large SVMVAR-4SV	<b>0.94</b>	1.01	0.94	0.88*	<b>0.94</b>

Notes: Model acronyms are explained in Table 3 of the paper. Bold numbers indicate the best forecast performance at each horizon. Diebold-Mariano test are based on the benchmark Small SVMVAR, \*\*\*, \*\* and \* denote statistically significant forecast improvements of a given model over the benchmark at the 1%, 5% and 10% level of significance, respectively.

Table 10: Cumulative Log predictive likelihoods for SVMVARs with  $q = 2$ 

Real GDP					
Models	$h = 1$	$h = 4$	$h = 8$	$h = 12$	Average
Small SVMVAR	1.17	6.17**	3.14	1.32	2.95
Large SVMVAR	12.62**	2.28	3.19	2.20	5.07
Large SVMVAR-1SV	-11.45	-11.16	-3.90	-5.92	-8.11
Large SVMVAR-2SV	13.65**	7.40*	5.73	9.23	9.00
Large SVMVAR-4SV	<b>16.12***</b>	<b>13.48***</b>	<b>11.98**</b>	<b>13.90**</b>	<b>13.87</b>
Inflation					
Small SVMVAR	<b>-0.37</b>	<b>-1.68</b>	9.31**	8.30**	3.89
Large SVMVAR	-16.73	-7.95	9.16**	5.65	-2.47
Large SVMVAR-1SV	-26.12	-37.76	-11.94	-2.66	-19.62
Large SVMVAR-2SV	-44.58	-7.36	<b>16.67**</b>	<b>11.37</b>	-5.98
Large SVMVAR-4SV	-3.55	-2.57	14.33**	11.11	<b>4.83</b>
Interest Rate					
Small SVMVAR	-0.94	8.05**	<b>15.87***</b>	<b>26.48***</b>	<b>12.37</b>
Large SVMVAR	<b>-0.82</b>	-9.36	-2.13	11.06	-0.31
Large SVMVAR-1SV	-86.04	-32.75	-26.19	-27.29	-43.07
Large SVMVAR-2SV	-32.19	<b>4.52</b>	12.03	11.06	-1.14
Large SVMVAR-4SV	-3.93	-1.97	-1.26	5.07	-0.52
Joint					
Small SVMVAR	1.89	<b>2.97</b>	9.53	3.90	4.57
Large SVMVAR	-3.32	-9.43	12.22	1.69	0.29
Large SVMVAR-1SV	-30.56	-49.72	-9.67	-21.07	-27.75
Large SVMVAR-2SV	-20.59	-16.47	15.84***	<b>15.39***</b>	-1.46
Large SVMVAR-4SV	<b>10.65**</b>	2.74***	<b>21.03***</b>	13.54***	<b>11.99</b>

Notes: Model acronyms are explained in Table 3 of the paper. Bold numbers indicate the best forecast performance at each horizon. Diebold-Mariano test are based on the benchmark Small SVMVAR, \*\*\*,\*\* and \* denote statistically significant forecast improvements of a given model over the benchmark at the 1%, 5% and 10% level of significance, respectively.

## B.2 Probability Integral Transformations (Full Sample)

To distinguish between models with  $q = 0$  and  $q = 2$  lags in the measurement equation of the SVMVAR models, we augment the model names using the notation 1 and 2, respectively. For instance, the SVMVAR with  $q = 0$  and  $q = 2$  are respectively labeled as SVMVAR1 and SVMVAR2.

### Real GDP

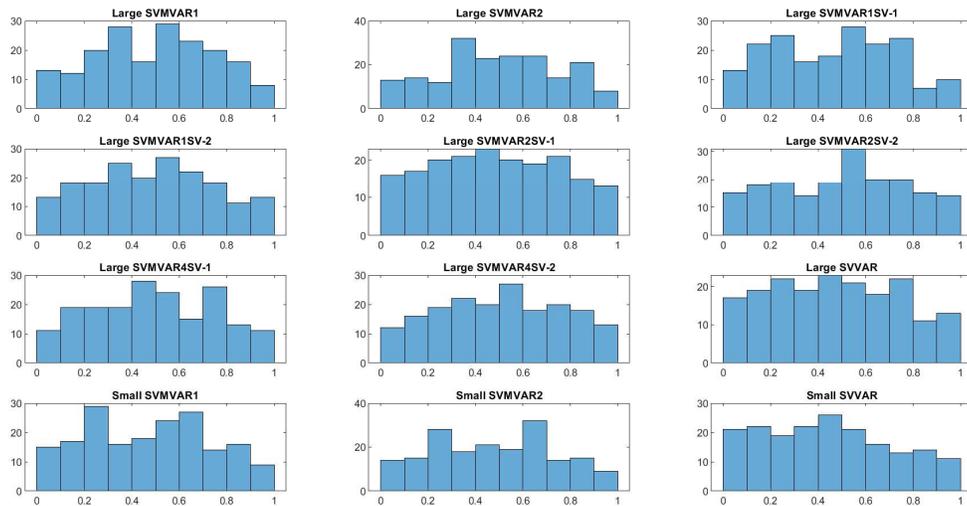


Figure 4: Probability Integral Transformations: One-step-ahead forecasts

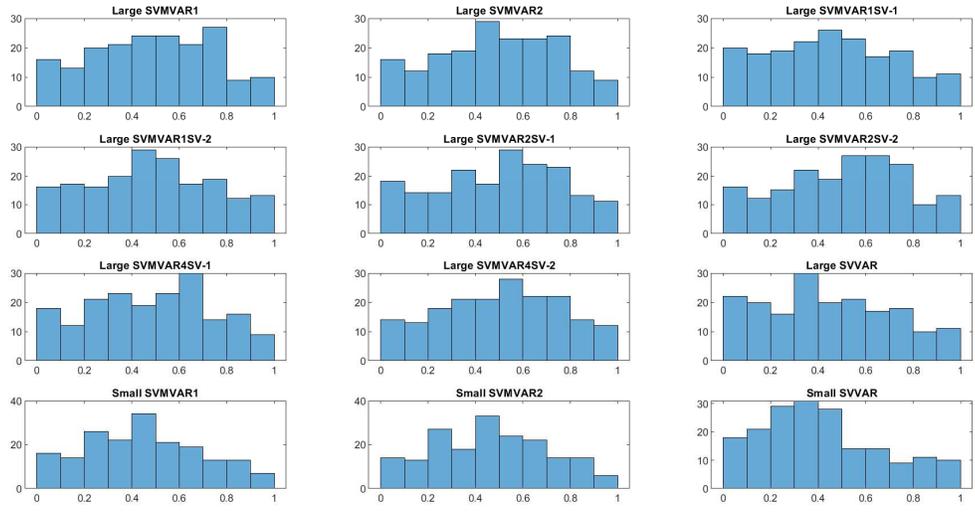


Figure 5: Probability Integral Transformations: Four-step-ahead forecasts

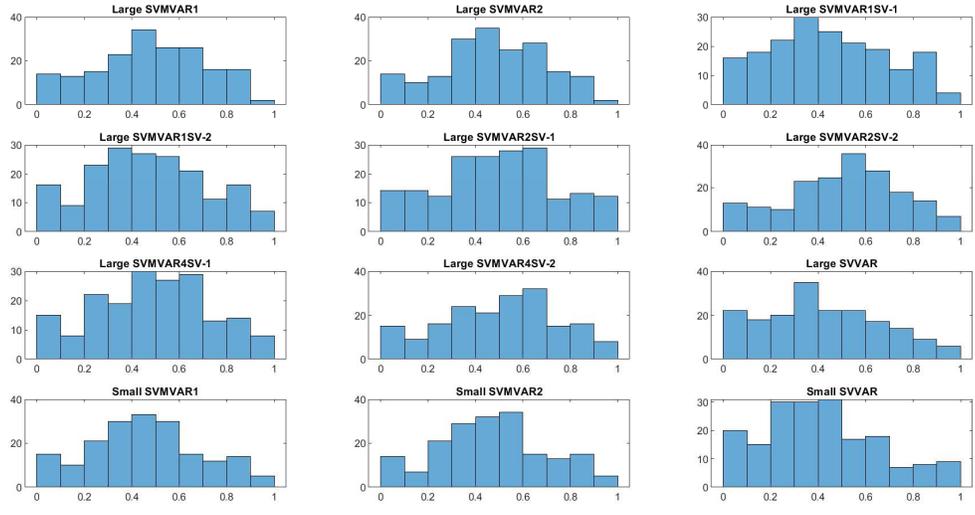


Figure 6: Probability Integral Transformations: Eight-step-ahead forecasts

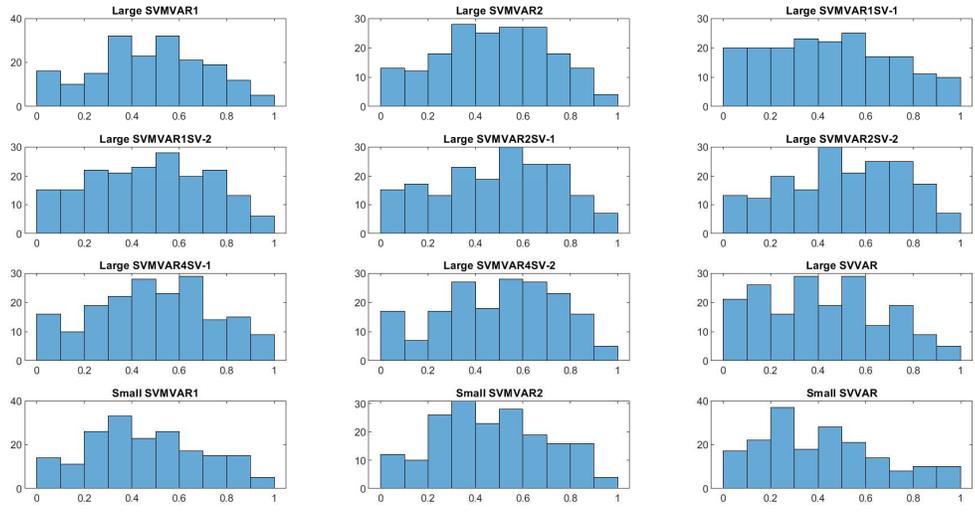


Figure 7: Probability Integral Transformations: Twelve-step-ahead forecasts

**Inflation**

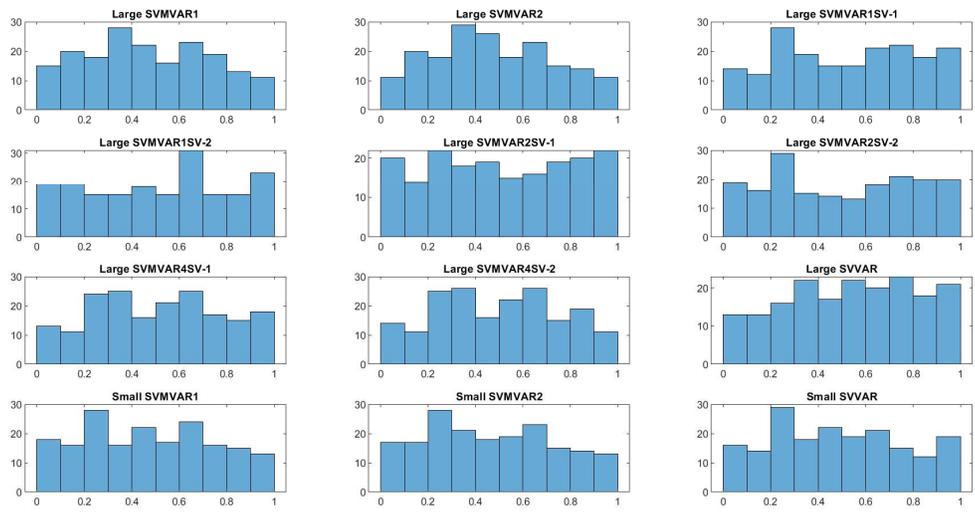


Figure 8: Probability Integral Transformations: One-step-ahead forecasts

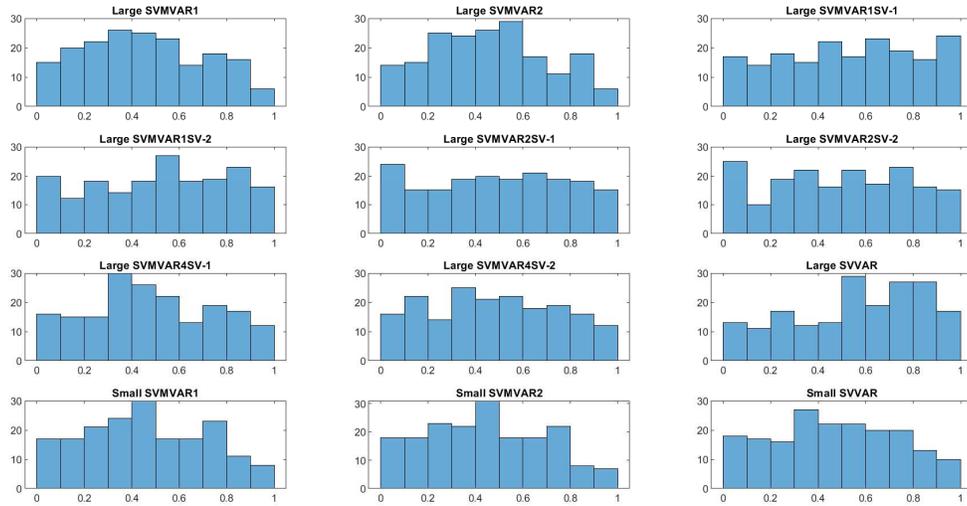


Figure 9: Probability Integral Transformations: Four-step-ahead forecasts

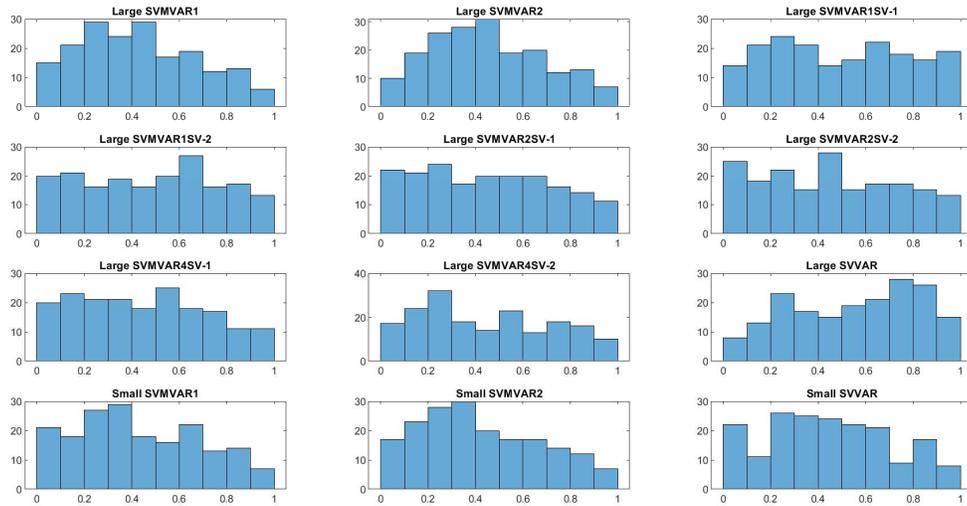


Figure 10: Probability Integral Transformations: Eight-step-ahead forecasts

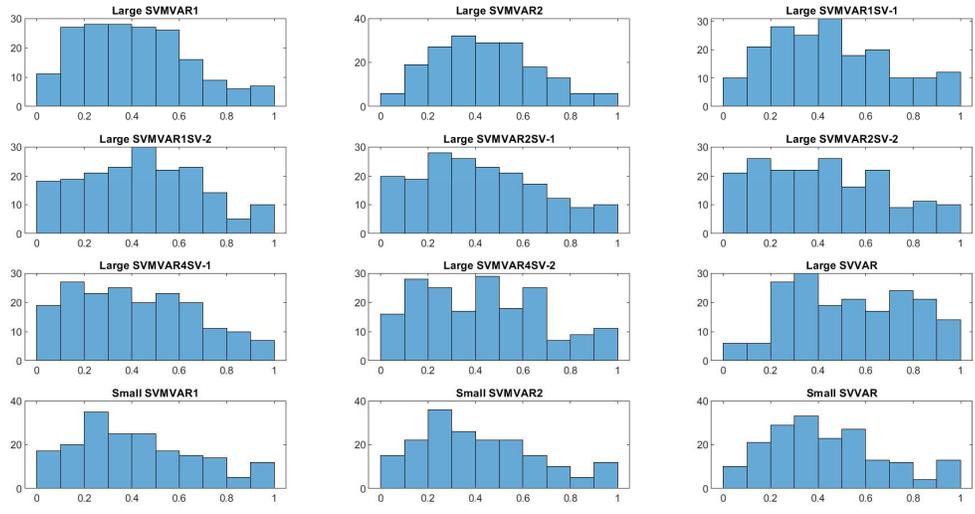


Figure 11: Probability Integral Transformations: Twelve-step-ahead forecasts

Interest Rate

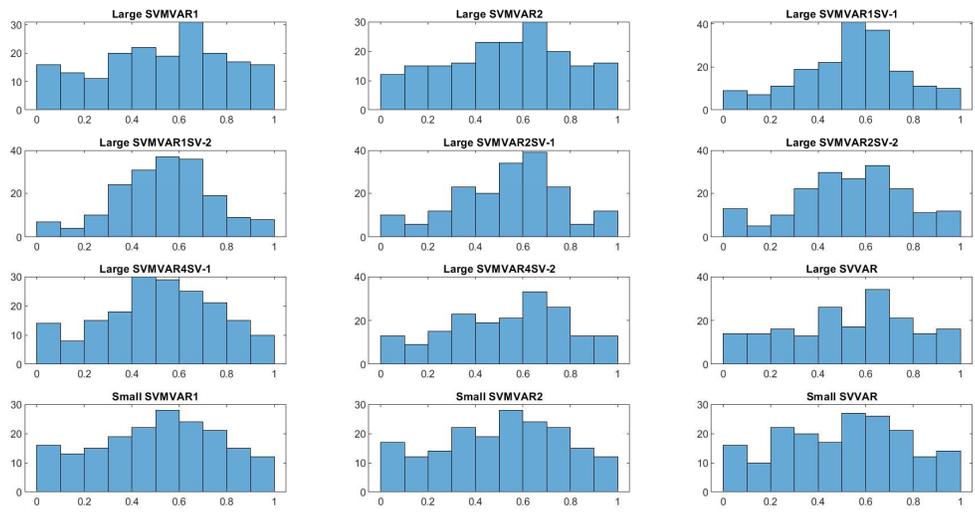


Figure 12: Probability Integral Transformations: One-step-ahead forecasts

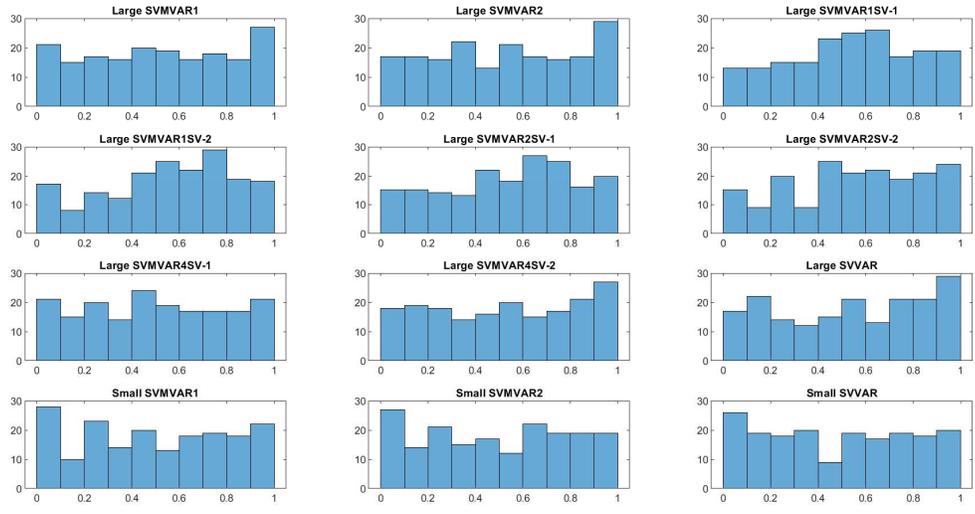


Figure 13: Probability Integral Transformations: Four-step-ahead forecasts

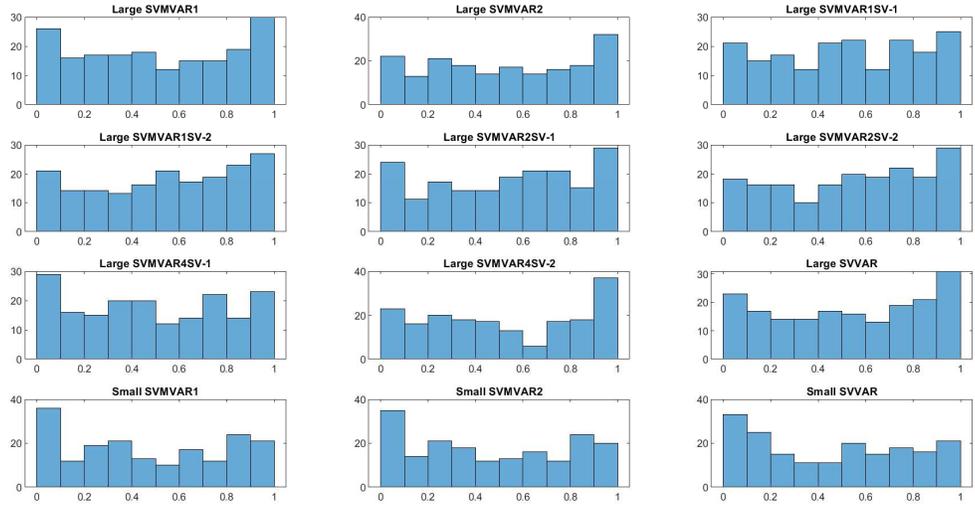


Figure 14: Probability Integral Transformations: Eight-step-ahead forecasts

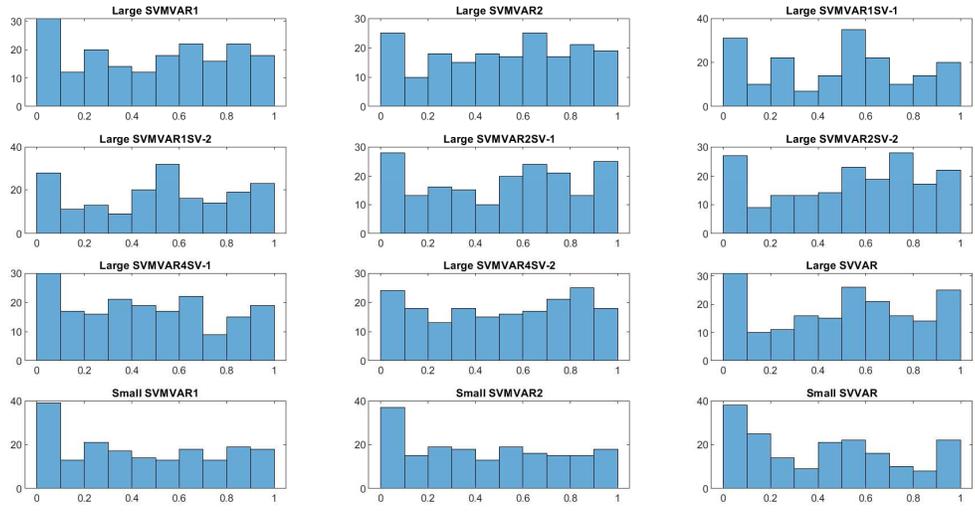


Figure 15: Probability Integral Transformations: Twelve-step-ahead forecasts