Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Large Hybrid Time-Varying Parameter VARs

CAMA Working Paper 77/2019
October 2019

Joshua C.C. Chan
Purdue University
UTS
Centre for Applied Macroeconomic Analysis, ANU

Abstract

Time-varying parameter VARs with stochastic volatility are routinely used for structural analysis and forecasting in settings involving a few macroeconomic variables. Applying these models to high-dimensional datasets has proved to be challenging due to intensive computations and over-parameterization concerns. We develop an efficient Bayesian sparsification method for a class of models we call hybrid TVP-VARs - VARs with time-varying parameters in some equations but constant coefficients in others. Specifically, for each equation, the new method automatically decides (i) whether the VAR coefficients are constant or time-varying, and (ii) whether the error variance is constant or has a stochastic volatility specification. Using US datasets of various dimensions, we find evidence that the VAR coefficients and error variances in some, but not all, equations are time varying. These large hybrid TVP-VARs also forecast better than standard benchmarks.

Keywords

large vector autoregression, time-varying parameter, stochastic volatility, trend output growth, macroeconomic forecasting

JEL Classification

C11, C52, C55, E37, E47

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Large Hybrid Time-Varying Parameter VARs

Joshua C.C. Chan*
Purdue University, UTS and CAMA

September 2019

Abstract

Time-varying parameter VARs with stochastic volatility are routinely used for structural analysis and forecasting in settings involving a few macroeconomic variables. Applying these models to high-dimensional datasets has proved to be challenging due to intensive computations and over-parameterization concerns. We develop an efficient Bayesian sparsification method for a class of models we call hybrid TVP-VARs—VARs with time-varying parameters in some equations but constant coefficients in others. Specifically, for each equation, the new method automatically decides (i) whether the VAR coefficients are constant or time-varying, and (ii) whether the error variance is constant or has a stochastic volatility specification. Using US datasets of various dimensions, we find evidence that the VAR coefficients and error variances in some, but not all, equations are time varying. These large hybrid TVP-VARs also forecast better than standard benchmarks.

Keywords: large vector autoregression, time-varying parameter, stochastic volatility, trend output growth, macroeconomic forecasting

JEL classifications: C11, C52, C55, E37, E47

*This paper has benefited from the constructive comments and suggestions of many people. I particularly thank Mark Bognanni, Todd Clark, Francesco Corsello, Francis Diebold, Luis Uzeda Garcia, Jaeho Kim, Kurt Lunsford, Michael McCracken, Elmar Mertens, Jouchi Nakajima, Davide Pettenuzzo, Frank Schorfheide, Le Wang, Benjamin Wong and Saeed Zaman, as well as conference and seminar participants at the 27th Annual Symposium of the SNDE, Applied Time Series Econometrics Workshop at the Federal Reserve Bank of St. Louis, the Deutsche Bundesbank, the Federal Reserve Bank of Cleveland, the 5th Hitotsubashi Summer Institute, Oklahoma University and University of Pennsylvania. All remaining errors are, of course, my own.

1 Introduction

Time-varying parameter vector autoregressions (TVP-VARs) developed by Cogley and Sargent (2001, 2005) and Primiceri (2005) have become the workhorse models in empirical macroeconomics. These models are flexible and can capture many different forms of structural instabilities and the evolving nonlinear relationships between the dependent variables. Moreover, they often forecast substantially better than their homoskedastic or constant-coefficient counterparts.[1] In empirical work, however, their applications are mostly limited to modeling small systems involving only a few variables because of the computational burden and over-parameterization concerns.

On the other hand, large VARs that use richer information have become increasing popular due to their better forecast performance and more sensible impulse-response analysis, as demonstrated in the influential paper by Banbura, Giannone, and Reichlin (2010).[2] Since there is a large body of empirical evidence that demonstrates the importance of accommodating time-varying structures in small systems, there has been much interest in recent years to build TVP-VARs for large datasets. While there are a few proposals to build large constant-coefficient VARs with stochastic volatility (see, e.g., Carriero, Clark, and Marcellino, 2016, 2019; Chan, 2018, 2019; Kastner and Huber, 2018), the literature on large VARs with time-varying coefficients remains relatively scarce.[3]

We propose a class of models we call hybrid TVP-VARs—VARs in which some equations have time-varying coefficients and/or stochastic volatility, whereas others have constant coefficients and/or homoscedastic errors. More precisely, we develop an efficient Bayesian shrinkage and sparsification method that automatically decides, for each equation, (i)

[1]The superior forecast performance of time-varying parameter models is demonstrated in many recent papers, such as Clark (2011), D'Agostino, Gambetti, and Giannone (2013), Koop and Korobilis (2013), Clark and Ravazzolo (2014) and Cross and Poon (2016).

[2]There is now a large and rapidly expanding literature that uses large VARs for forecasting and structural analysis. Prominent examples include Carriero, Kapetanios, and Marcellino (2009), Koop (2013), Banbura, Giannone, Modugno, and Reichlin (2013), Carriero, Clark, and Marcellino (2015), Ellahie and Ricco (2017) and Morley and Wong (2019).

[3]There are few attempts in this direction. For example, Koop and Korobilis (2013) and Gefang, Koop, and Poon (2019) propose methods to approximate the posterior distributions of large TVP-VARs. Banbura and van Vlodrop (2018) and Götz and Hauzenberger (2018) consider a large VAR with only time-varying intercepts. Eisenstat, Chan, and Strachan (2018) model the time-varying coefficients using a reduced-rank structure, whereas Huber, Koop, and Onorante (2019) develop a method that first shrinks the time-varying coefficients, followed by setting the small values to zero.

whether the VAR coefficients are constant or time-varying, and (ii) whether the errors are homoscedastic or follow a stochastic volatility specification. Our framework nests many popular VARs as special cases, ranging from a standard homoscedastic, constant-coefficient VAR on one end of the spectrum to the flexible but highly parameterized TVP-VARs of Cogley and Sargent (2001, 2005) and Primiceri (2005) on the other end. More importantly, our framework also includes many hybrid TVP-VARs in between the extremes, allowing for a more nuanced modeling approach of the time-varying structures.

To formulate these large hybrid TVP-VARs, we use a reparameterization of the standard TVP-VAR in Primiceri (2005). Specifically, we rewrite the TVP-VAR in the structural form in which the time-varying error covariance matrices are diagonal. Hence, we can treat the structural TVP-VAR as a system of n unrelated TVP regressions and estimate them one by one. This reduces the dimension of the problem and can substantially speed up computations. This approach is similar to the equation-by-equation estimation approach in Carriero, Clark, and Marcellino (2019) that is designed for the reduced-form parameterization. But since under our parameterization there is no need to obtain the 'orthogonalized' shocks at each iteration as in Carriero, Clark, and Marcellino (2019), the proposed approach is substantially faster. Moreover, under our parameterization the estimation can be parallelized to further speed up computations.

Next, we adapt the non-centered parameterization of the state space model in Frühwirth-Schnatter and Wagner (2010) to our structural TVP-VAR representation. Further, for each equation we introduce two indicator variables, one determines whether the coefficients are time-varying or constant, while the other decides between stochastic volatility and constant variance. By treating these indicators as parameters to be estimated, we allow the data to determine the appropriate time-varying structures, in contrast to typical setups where time variation is assumed. The proposed approach therefore is not only flexible—it includes many state-of-the-art models routinely used in applied work as special cases—it also induces parsimony to ameliorate over-parameterization concerns.

The estimation is done using Markov chain Monte Carlo (MCMC) methods. Hence, in contrast to many earlier attempts to build large TVP-VARs, our approach is fully Bayesian and is exact—it simulates from the exact posterior distribution. There are, however, a few challenges in the estimation. First, the dimension of the model is large and there are thousands of latent state processes—time-varying coefficients and stochastic volatilities—to simulate. To overcome this challenge, in addition to using the equation-by-equation estimation approach described earlier, we also adopt the precision sampler of Chan and Jeliazkov (2009) to draw both the time-invariant and time-varying VAR coefficients, as well as the stochastic volatilities. In our high-dimensional setting the precision sampler substantially reduces the computational cost compared to conventional Kalman filter based smoothers. A second challenge in the estimation is that the indicators and the latent states enter the likelihood multiplicatively. Consequently, it is vital to sample them jointly; otherwise the Markov chain is likely to get stuck. We develop algorithms based on the importance sampling estimators for marginal likelihoods of TVP-VARs in Chan and Eisenstat (2018a,b) to sample the indicators and the latent states jointly.

Using US datasets of different dimensions, we find evidence that the VAR coefficients and error variances in some, but not all, equations are time varying. We further illustrate the usefulness of the proposed hybrid TVP-VARs with a forecasting exercise that involves 20 US quarterly macroeconomic and financial variables. We show that the proposed hybrid TVP-VARs forecast better than many benchmarks, including the conventional homoscedastic, constant-coefficient VAR, the constant-coefficient VAR with stochastic volatility and the full-fledged TVP-VAR where all the VAR coefficients and error variances are time varying. These results suggest that using a data-driven approach to discover the time-varying structures—rather than imposing either constant coefficients or time-varying parameters—is empirically beneficial.

The rest of the paper is organized as follows. We first introduce the proposed framework in Section 2. In particular, we discuss how we combine a reparameterization of the reduced-form TVP-VAR and the non-centered parameterization of the state space model to develop the hybrid TVP-VARs. We then describe the shrinkage priors and the posterior sampler in Section 3. It is followed by an illustration in Section 4 that fits a small dataset to check if the proposed methodology can replicate a model comparison exercise. The empirical application is discussed in detail in Section 5. Lastly, Section 6 concludes and briefly discusses some future research directions.

2 Hybrid TVP-VARs

In this section we introduce a class of models we call hybrid time-varying parameter VARs, under which some equations have time-varying coefficients and stochastic volatility while coefficients and variances in other equations remain constant. To that end, let $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ be an $n \times 1$ vector of endogenous variables at time t. The TVP-VAR of Primiceri (2005) can be reparameterized and written in the following structural form:

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{b}_t + \mathbf{B}_{1,t}\mathbf{y}_{t-1} + \cdots + \mathbf{B}_{p,t}\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t^y, \quad \boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{1}$$

where $\mathbf{b}_t$ is an $n \times 1$ vector of time-varying intercepts, $\mathbf{B}_{1,t}, \ldots, \mathbf{B}_{p,t}$ are $n \times n$ VAR coefficient matrices, $\mathbf{A}_t$ is an $n \times n$ lower triangular matrix with ones on the diagonal and $\boldsymbol{\Sigma}_t = \mathrm{diag}(\exp(h_{1,t}), \ldots, \exp(h_{n,t}))$. The law of motion of the VAR coefficients and log-volatilites will be specified below. Here we note that since the system in (1) is written in the structural form, the covariance matrix $\boldsymbol{\Sigma}_t$ is diagonal by construction. Consequently, we can estimate this recursive system equation by equation without loss of efficiency.[4]

For later reference, we introduce some notations. Let $b_{i,t}$ denote the i-th element of $\mathbf{b}_t$ and $\mathbf{B}_{i,j,t}$ be the i-th row of $\mathbf{B}_{j,t}$. Then, $\boldsymbol{\beta}_{i,t} = (b_{i,t}, \mathbf{B}_{i,1,t}, \ldots, \mathbf{B}_{i,p,t})'$ is the intercept and VAR coefficients for the i-th equation. Moreover, let $\boldsymbol{\alpha}_{i,t}$ denote the free elements in the i-th row of the impact matrix $\mathbf{A}_t$, i.e., $\boldsymbol{\alpha}_{i,t} = (A_{i1,t}, \ldots, A_{i(i-1),t})'$. Then, the i-th equation of the system in (1) can be rewritten as:

$$y_{i,t} = \widetilde{\mathbf{w}}_{i,t}\boldsymbol{\alpha}_{i,t} + \widetilde{\mathbf{x}}_t\boldsymbol{\beta}_{i,t} + \varepsilon_{i,t}^y, \quad \varepsilon_{i,t}^y \sim \mathcal{N}(0, \mathrm{e}^{h_{i,t}}),$$

where $\widetilde{\mathbf{w}}_{i,t} = (-y_{1,t}, \ldots, -y_{i-1,t})$ and $\widetilde{\mathbf{x}}_t = (1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')$. Note that $y_{i,t}$ depends on the contemporaneous variables $y_{1,t}, \ldots, y_{i-1,t}$. But since the system is triangular, when we perform the change of variables from $\boldsymbol{\varepsilon}_t^y$ to $\mathbf{y}_t$ to obtain the likelihood function, the density function remains Gaussian.

[4]Carriero, Clark, and Marcellino (2019) pioneer a similar equation-by-equation estimation approach for a large reduced-form constant-coefficient VAR with stochastic volatility. The main advantage of the structural-form representation is that it allows us to rewrite the VAR as n unrelated regressions, and it leads to a more efficient sampling scheme.

If we let $\mathbf{x}_{i,t} = (\widetilde{\mathbf{w}}_{i,t}, \widetilde{\mathbf{x}}_t)$, we can further simplify the i-th equation as:

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\theta}_{i,t} + \varepsilon^y_{i,t}, \quad \varepsilon^y_{i,t} \sim \mathcal{N}(0, \mathrm{e}^{h_{i,t}}), \tag{2}$$

where $\boldsymbol{\theta}_{i,t} = (\boldsymbol{\alpha}'_{i,t}, \boldsymbol{\beta}'_{i,t})'$ is of dimension $k_{\theta_i} = np + i$. Hence, we have rewritten the TVP-VAR in (1) as n unrelated regressions. Finally, the coefficients and log-volatilities are assumed to evolve as independent random walks:

$$\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,t-1} + \boldsymbol{\varepsilon}^\theta_{i,t}, \qquad \boldsymbol{\varepsilon}^\theta_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\theta_i}), \tag{3}$$

$$h_{i,t} = h_{i,t-1} + \varepsilon^h_{i,t}, \qquad \varepsilon^h_{i,t} \sim \mathcal{N}(0, \sigma^2_{h,i}), \tag{4}$$

where the initial conditions $\boldsymbol{\theta}_{i,0}$ and $h_{i,0}$ are treated as unknown parameters to be estimated. The system in (2)–(4) specifies a reparameterization of a standard TVP-VAR in which all equations have time-varying parameters and stochastic volatility.

Next, we introduce a framework that allows the model to determine—in a data-driven fashion—whether the coefficients in each equation are time varying or constant and whether the errors are homoskedastic or they follow a stochastic volatility process. For that purpose, we adapt the non-centered parameterization of Frühwirth-Schnatter and Wagner (2010) to our hybrid TVP-VARs. More specifically, for $i = 1, \dots, n, t = 1, \dots, T$, we consider the following model:

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\theta}_{i,0} + \gamma^\theta_i \mathbf{x}_{i,t}\boldsymbol{\Sigma}^{\frac{1}{2}}_{\theta_i}\widetilde{\boldsymbol{\theta}}_{i,t} + \varepsilon^y_{i,t}, \qquad \varepsilon^y_{i,t} \sim \mathcal{N}(0, \mathrm{e}^{h_{i,0} + \gamma^h_i \sigma_{h,i}\widetilde{h}_{i,t}}), \tag{5}$$

$$\widetilde{\boldsymbol{\theta}}_{i,t} = \widetilde{\boldsymbol{\theta}}_{i,t-1} + \widetilde{\boldsymbol{\varepsilon}}^\theta_{i,t}, \qquad \widetilde{\boldsymbol{\varepsilon}}^\theta_{i,t} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\theta_i}}), \tag{6}$$

$$\widetilde{h}_{i,t} = \widetilde{h}_{i,t-1} + \widetilde{\varepsilon}^h_{i,t}, \qquad \widetilde{\varepsilon}^h_{i,t} \sim \mathcal{N}(0, 1), \tag{7}$$

where $\widetilde{\boldsymbol{\theta}}_{i,0} = \mathbf{0}$ and $\widetilde{h}_{i,0} = 0$, and γ^θ_i and γ^h_i are indicator variables that take values of either 0 or 1.[5]

The model in (5)-(7) includes a wide variety of popular VAR specifications. For example, assuming that the indicators all take the value of 1, the above model is just a reparameterization of the TVP-VAR in (2)–(4). To see that, define $\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,0} + \gamma^\theta_i \boldsymbol{\Sigma}^{\frac{1}{2}}_{\theta_i}\widetilde{\boldsymbol{\theta}}_{i,t}$ and

[5]It is straightforward to include a few additional indicators to allow for more sophisticated forms of time variation in coefficients. For example, one can replace γ^θ_i with two indicators, say, γ^α_i and γ^β_i, which control the time variation in $\boldsymbol{\alpha}_{i,t}$ and $\boldsymbol{\beta}_{i,t}$ respectively. The posterior simulator in Section 3.2 can be modified to handle this case, at the cost of a slight increase of computation time.

$h_{i,t} = h_{i,0} + \gamma_i^h \sigma_{h,i} \widetilde{h}_{i,t}$. Then, (5) reduces to (2). In addition, we have

$$\boldsymbol{\theta}_{i,t} - \boldsymbol{\theta}_{i,t-1} = \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}}(\widetilde{\boldsymbol{\theta}}_{i,t} - \widetilde{\boldsymbol{\theta}}_{i,t-1}) = \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\theta},$$
$$h_{i,t} - h_{i,t-1} = \sigma_{h,i}(\widetilde{h}_{i,t} - \widetilde{h}_{i,t-1}) = \sigma_{h,i}\widetilde{\varepsilon}_{i,t}^{h}.$$

Hence, $\boldsymbol{\theta}_{i,t}$ and $h_{i,t}$ follow the same random walk processes as in (3) and (4), respectively. We have therefore showed that when $\gamma_i^\theta = \gamma_i^h = 1, i = 1, \ldots, n$, the proposed model reduces to a TVP-VAR with stochastic volatility.

In the other extreme, when all the indicators are zero, the proposed model becomes a homoscedastic constant-coefficient VAR in structural form. For the intermediate case when $\gamma_i^\theta = 0$ and $\gamma_i^h = 1, i = 1, \ldots, n$, then the proposed model reduces to a constant-coefficient VAR with stochastic volatility—a structural-form reparameterization of the specification in Carriero, Clark, and Marcellino (2019). More generally, by allowing the indicators γ_i^θ and γ_i^h to take different values, we could have a VAR in which only some equations have time-varying coefficients or stochastic volatility.

These indicators are not fixed but are estimated from the data. More precisely, we specify that each γ_i^θ follows an independent Bernoulli distribution with success probability $\mathbb{P}(\gamma_i^\theta = 1) = p_i^\theta, i = 1, \ldots, n$. Similarly for γ_i^h: $\mathbb{P}(\gamma_i^h = 1) = p_i^h$. These success probabilities p_i^θ and $p_i^h, i = 1, \ldots, n$, are in turn treated as parameters to be estimated.

In contrast to typical setups where time variation in parameters is assumed (e.g. Cogley and Sargent, 2001, 2005; Primiceri, 2005), here the proposed model puts positive probabilities in simpler models in which the VAR coefficients are constant or the errors are homoscedastic. The values of the indicators are determined by the data, and these time-varying features are turned on only when they are warranted. The proposed model therefore is not only flexible in the sense that it includes a wide variety of specifications popular in applied work as special cases, it also induces parsimony to combat over-parameterization concerns.

3 Priors and Bayesian Estimation

In this section we first describe in detail the priors for the time-invariant parameters. We then outline the posterior simulator to estimate the model in (5)–(7).

3.1 Priors

For notational convenience, stack $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{\theta}_i = (\boldsymbol{\theta}'_{i,1}, \ldots, \boldsymbol{\theta}'_{i,T})'$ and $\mathbf{h}_i = (h_{i,1}, \ldots, h_{i,T})'$ over $t = 1, \ldots, T$, and collect $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^n$, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}_{i=1}^n$ and $\mathbf{h} = \{\mathbf{h}_i\}_{i=1}^n$ over $i = 1, \ldots, n$. Similarly define $\widetilde{\boldsymbol{\theta}}_i$ and $\widetilde{\mathbf{h}}_i$. In our model, the time-invariant parameters are $\boldsymbol{\gamma}^\theta = (\gamma_1^\theta, \ldots, \gamma_n^\theta)'$, $\boldsymbol{\gamma}^h = (\gamma_1^h, \ldots, \gamma_n^h)'$, $\boldsymbol{\Sigma}_\theta = \{\boldsymbol{\Sigma}_{\theta_i}\}_{i=1}^n$, $\boldsymbol{\Sigma}_h = \{\sigma_{i,h}^2\}_{i=1}^n$, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{1,0}, \ldots, \boldsymbol{\theta}'_{n,0})'$, $\mathbf{h}_0 = (h_{1,0}, \ldots, h_{n,0})'$, $\mathbf{p}^\theta = (p_1^\theta, \ldots, p_n^\theta)'$ and $\mathbf{p}^h = (p_1^h, \ldots, p_n^h)'$. Below we give the details of the priors on these time-invariant parameters.

Since $\boldsymbol{\theta}_0$, the initial conditions for the VAR coefficients, is high-dimensional when n is large, appropriate shrinkage is crucial. We assume a Minnesota-type prior on $\boldsymbol{\theta}_0$ along the lines in Sims and Zha (1998).[6] More specifically, consider $\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{a}_{\boldsymbol{\theta}_0}, \mathbf{V}_{\boldsymbol{\theta}_0})$, where the prior mean $\boldsymbol{\theta}_0$ is set to be zero to induce shrinkage and the prior covariance matrix $\mathbf{V}_{\boldsymbol{\theta}_0}$ is block-diagonal with $\mathbf{V}_{\boldsymbol{\theta}_0} = \text{diag}(\mathbf{V}_{\boldsymbol{\theta}_{1,0}}, \ldots, \mathbf{V}_{\boldsymbol{\theta}_{n,0}})$—here $\mathbf{V}_{\boldsymbol{\theta}_{i,0}}$ is the prior covariance matrix for $\boldsymbol{\theta}_{i,0}, i = 1, \ldots, n$. For each $\mathbf{V}_{\boldsymbol{\theta}_{i,0}}$, we in turn assume it to be diagonal with the k-th diagonal element $(V_{\boldsymbol{\theta}_{i,0}})_{kk}$ set to be:

$$(V_{\boldsymbol{\theta}_{i,0}})_{kk} = \begin{cases} \frac{\kappa_1}{l^2}, & \text{for the coefficient on the } l\text{-th lag of variable } i, \\ \frac{\kappa_2 s_i^2}{l^2 s_j^2}, & \text{for the coefficient on the } l\text{-th lag of variable } j, j \neq i, \\ \frac{\kappa_3 s_i^2}{s_j^2}, & \text{for the } j\text{-th element of } \boldsymbol{\alpha}_i, \\ \kappa_4 s_i^2, & \text{for the intercept,} \end{cases}$$

where s_r^2 denotes the sample variance of the residuals from regressing $y_{r,t}$ on $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-4}$, $r = 1, \ldots, n$. Here the prior covariance matrix $\mathbf{V}_{\boldsymbol{\theta}_0}$ depends on four hyperparameters—$\kappa_1, \kappa_2, \kappa_3$ and κ_4—that control the degree of shrinkage for different types of coefficients. For simplicity, we set $\kappa_3 = 1$ and $\kappa_4 = 100$. These values imply moderate shrinkage for

[6] See also Doan, Litterman, and Sims (1984), Litterman (1986) and Kadiyala and Karlsson (1997). For a textbook discussion of the Minnesota prior, see, e.g., Koop and Korobilis (2010), Del Negro and Schorfheide (2012) and Karlsson (2013).

the coefficients on the contemporaneous variables and no shrinkage for the intercepts.

The remaining two hyperparameters are κ_1 and κ_2, which control the overall shrinkage strength for coefficients on own lags and those on lags of other variables, respectively. Departing from Sims and Zha (1998), here we allow κ_1 and κ_2 to be different, as one might expect that coefficients on lags of other variables would be on average smaller than those on own lags. Chan (2019) in fact finds empirical evidence in support of this so-called cross-variable shrinkage. In addition, we treat κ_1 and κ_2 as unknown parameters to be estimated rather than fixing them to some subjective values. This is motivated by a few recent studies, such as Carriero, Clark, and Marcellino (2015) and Giannone, Lenza, and Primiceri (2015), which show that by selecting this type of overall shrinkage hyperparameters in a data-based fashion, one can substantially improve the forecast performance of the resulting VAR.[7]

We assume gamma priors for the hyperparameters κ_1 and κ_2: $\kappa_j \sim \mathcal{G}(c_{1,j}, c_{2,j}), j = 1, 2$. We set $c_{1,1} = c_{1,2} = 1$, $c_{2,1} = 1/0.04$ and $c_{2,2} = 1/0.04^2$. These values imply that the prior modes are at zero, which provides global shrinkage. The prior means of κ_1 and κ_2 are 0.04 and 0.04^2 respectively, which are the fixed values used in Carriero, Clark, and Marcellino (2015).

Next, following Frühwirth-Schnatter and Wagner (2010), the square roots of the diagonal elements of $\mathbf{\Sigma}_{\theta_i} = \text{diag}(\sigma^2_{\theta_i,1}, \ldots, \sigma^2_{\theta_i,k_{\theta_i}}), i = 1, \ldots, n$ and $\mathbf{\Sigma}_h = \text{diag}(\sigma^2_{h,1}, \ldots, \sigma^2_{h,n})$ are independently distributed as mean 0 normal random variables: $\sigma_{\theta_i,j} \sim \mathcal{N}(0, S_{\theta_i,j})$ and $\sigma_{h,i} \sim \mathcal{N}(0, S_{h,i})$, with $j = 1, \ldots, k_{\theta_i}, i = 1, \ldots, n$. The success probabilities p^θ_i and p^h_i are assumed to have beta distributions: $p^\theta_i \sim \mathcal{B}(a_{p^\theta}, b_{p^\theta})$ and $p^h_i \sim \mathcal{B}(a_{p^h}, b_{p^h}), i = 1, \ldots, n$. Finally, the initial conditions $\mathbf{h}_0$ are assumed to be jointly Gaussian: $\mathbf{h}_0 \sim \mathcal{N}(\mathbf{a}_{\mathbf{h}_0}, \mathbf{V}_{\mathbf{h}_0})$.

3.2 Posterior Simulator

We now turn to the estimation of the model in (5)–(7) given the prior described in the previous section. There are a few challenges in the estimation. First, since $\boldsymbol{\theta}_i$ becomes

[7]Moreover, this data-based Minnesota prior is also found to forecast better than many recently introduced adaptive shrinkage priors such as the normal-gamma prior, the Dirichlet-Laplace prior and the horseshoe prior. For example, this is demonstrated in a comprehensive forecasting exercise in Cross, Hou, and Poon (2019).

degenerate when $\gamma_i^\theta = 0$, making its sampling nonstandard (similarly for $\mathbf{h}_i$). To sidestep this problem, we will use the parameterization in terms of $\widetilde{\boldsymbol{\theta}}_i$ and $\widetilde{\mathbf{h}}_i$. Then, given the posterior draws of $\widetilde{\boldsymbol{\theta}}_i$, $\widetilde{\mathbf{h}}_i$ and other parameters, we can recover the posterior draws of $\boldsymbol{\theta}_i$ and $\mathbf{h}_i$ using the definitions $\boldsymbol{\theta}_{i,t} = \boldsymbol{\theta}_{i,0} + \gamma_i^\theta \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} \widetilde{\boldsymbol{\theta}}_{i,t}$ and $h_{i,t} = h_{i,0} + \gamma_i^h \sigma_{h,i} \widetilde{h}_{i,t}$.

Second, since $\widetilde{\boldsymbol{\theta}}_i$ and the indicator γ_i^θ enter the likelihood multiplicatively, it is vital to sample them jointly (similarly for $\widetilde{\mathbf{h}}_i$ and γ_i^h); otherwise the Markov chain might get stuck. To see this, consider an alternative sampling scheme in which we simulate $\widetilde{\boldsymbol{\theta}}_i$ given γ_i^θ, followed by sampling γ_i^θ given $\widetilde{\boldsymbol{\theta}}_i$. Suppose $\gamma_i^\theta = 0$ in the last iteration. Given $\gamma_i^\theta = 0$, $\widetilde{\boldsymbol{\theta}}_i$ does not enter the likelihood and we simply sample it from its state equation. Since the sampled $\widetilde{\boldsymbol{\theta}}_i$ has no relation to the data, the implied time variation in the VAR coefficients would not match the data. Consequently, it is highly likely that the model would prefer no time variation, i.e., $\gamma_i^\theta = 0$. Hence, it is unlikely for the Markov chain to move away from $\gamma_i^\theta = 0$ once it is there. Therefore, in the following posterior sampler we jointly draw $\widetilde{\boldsymbol{\theta}}_i$ and γ_i^θ, as well as $\widetilde{\mathbf{h}}_i$ and γ_i^h.

Given the model in (5)–(7) and the above the prior described in the previous section, we can simulate from the joint posterior distribution using the following posterior sampler that sequentially samples from:

1. $p(\gamma_i^\theta, \widetilde{\boldsymbol{\theta}}_i \,|\, \mathbf{y}, \widetilde{\mathbf{h}}, \boldsymbol{\theta}_0, \mathbf{h}_0, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}^h, \mathbf{p}^\theta, \mathbf{p}^h, \boldsymbol{\kappa})$, $i = 1, \ldots, n$;
2. $p(\gamma_i^h, \widetilde{\mathbf{h}}_i \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_0, \mathbf{h}_0, \boldsymbol{\Sigma}_\theta, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}^\theta, \mathbf{p}^\theta, \mathbf{p}^h, \boldsymbol{\kappa})$, $i = 1, \ldots, n$;
3. $p(\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}}, \boldsymbol{\theta}_{i,0} \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{h}}, \mathbf{h}_0, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}^\theta, \boldsymbol{\gamma}^h, \mathbf{p}^\theta, \mathbf{p}^h, \boldsymbol{\kappa})$, $i = 1, \ldots, n$;
4. $p(\sigma_{h,i}, h_{i,0} \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{h}}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\theta}_0, \boldsymbol{\gamma}^\theta, \boldsymbol{\gamma}^h, \mathbf{p}^\theta, \mathbf{p}^h, \boldsymbol{\kappa})$, $i = 1, \ldots, n$;
5. $p(p_i^\theta, p_i^h \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{h}}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\theta}_0, \boldsymbol{\gamma}^\theta, \boldsymbol{\gamma}^h, \boldsymbol{\Sigma}_h, \mathbf{h}_0, \boldsymbol{\kappa})$, $i = 1, \ldots, n$;
6. $p(\boldsymbol{\kappa} \,|\, \mathbf{y}, \widetilde{\boldsymbol{\theta}}, \widetilde{\mathbf{h}}, \boldsymbol{\Sigma}_\theta, \boldsymbol{\theta}_0, \boldsymbol{\gamma}^\theta, \boldsymbol{\gamma}^h, \boldsymbol{\Sigma}_h, \mathbf{h}_0, \mathbf{p}^\theta, \mathbf{p}^h)$.

Step 3 to Step 6 are relatively straightforward and we leave the details to Appendix A. Here we focus on the first two steps.

Step 1. Since γ_i^θ and $\widetilde{\boldsymbol{\theta}}_i$ enter the likelihood multiplicatively, we sample them jointly to improve efficiency. This is done by first drawing the indicator γ_i^θ marginally of $\widetilde{\boldsymbol{\theta}}_i$—but conditional on other parameters—and then sample $\widetilde{\boldsymbol{\theta}}_i$ from its full conditional distribution. The latter of these two steps is straightforward because given γ_i^θ, we have a linear Gaussian state space model in $\widetilde{\boldsymbol{\theta}}_i$. Specifically, we stack the observation equation (5) over $t = 1, \ldots, T$:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\theta}_{i,0} + \gamma_i^\theta \mathbf{Z}_i \widetilde{\boldsymbol{\theta}}_i + \boldsymbol{\varepsilon}_i^y, \qquad \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{h}_i}),$$

where $\boldsymbol{\Sigma}_{\mathbf{h}_i} = \text{diag}(\text{e}^{h_{i,1}}, \ldots, \text{e}^{h_{i,T}})$ with $h_{i,t} = h_{i,0} + \gamma_i^h \sigma_{h,i} \widetilde{h}_{i,t}$,

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i,1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix}, \quad \mathbf{Z}_i = \begin{pmatrix} \mathbf{x}_{i,1}\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i,2}\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{i,T}\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} \end{pmatrix}.$$

Next, stack the state equation (7) over $t = 1, \ldots, T$:

$$\mathbf{H}_{k_{\theta_i}} \widetilde{\boldsymbol{\theta}}_i = \boldsymbol{\varepsilon}_i^{\widetilde{\theta}}, \quad \boldsymbol{\varepsilon}_i^{\widetilde{\theta}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{Tk_{\theta_i}}),$$

where $\mathbf{H}_{k_{\theta_i}}$ is the first difference matrix of dimension k_{θ_i}. Since $\mathbf{H}_{k_{\theta_i}}$ is a square matrix with unit determinant, it is invertible. It then follows that $\widetilde{\boldsymbol{\theta}}_i \sim \mathcal{N}(\mathbf{0}, (\mathbf{H}'_{k_{\theta_i}} \mathbf{H}_{k_{\theta_i}})^{-1})$. Finally, using standard linear regression results, we have

$$(\widetilde{\boldsymbol{\theta}}_i \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \gamma_i^\theta) \sim \mathcal{N}\left(\widehat{\widetilde{\boldsymbol{\theta}}}_i, \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}^{-1}\right), \tag{8}$$

where

$$\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i} = \mathbf{H}'_{k_{\theta_i}} \mathbf{H}_{k_{\theta_i}} + \gamma_i^\theta \mathbf{Z}_i' \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} \mathbf{Z}_i, \qquad \widehat{\widetilde{\boldsymbol{\theta}}}_i = \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}^{-1} \left(\gamma_i^\theta \mathbf{Z}_i' \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0})\right). \tag{9}$$

Since the precision matrix $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ is a band matrix, one can sample from $(\widetilde{\boldsymbol{\theta}}_i \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i, \gamma_i^\theta}, \boldsymbol{\theta}_{i,0})$ efficiently using the algorithm in Chan and Jeliazkov (2009).

To sample γ_i^θ marginal of $\widetilde{\boldsymbol{\theta}}_i$, first note that

$$p(\gamma_i^\theta \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) \propto \left[\int_{\mathbb{R}^{Tk_{\theta_i}}} p(\mathbf{y}_i \,|\, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \gamma_i^\theta) p(\widetilde{\boldsymbol{\theta}}_i) \text{d}\widetilde{\boldsymbol{\theta}}_i\right] p(\gamma_i^\theta),$$

where both the conditional likelihood $p(\mathbf{y}_i \,|\, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \gamma_i^\theta)$ and the prior density $p(\widetilde{\boldsymbol{\theta}}_i)$ are Gaussian. It turns out that the above integral admits an analytical expression. In fact, using a similar derivation in Chan and Grant (2016), one can show that

$$\int_{\mathbb{R}^{Tk_{\theta_i}}} p(\mathbf{y}_i \,|\, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \gamma_i^\theta) p(\widetilde{\boldsymbol{\theta}}_i) \mathrm{d}\widetilde{\boldsymbol{\theta}}_i \\ = (2\pi)^{-\frac{T}{2}} |\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^T h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0})'\boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0}) - \widehat{\widetilde{\boldsymbol{\theta}}}_i' \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i} \widehat{\widetilde{\boldsymbol{\theta}}}_i\right)}, \tag{10}$$

where $\widehat{\widetilde{\boldsymbol{\theta}}}_i$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ are defined in (9). When $\gamma_i^\theta = 0$, $|\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}| = 1$ and $\widehat{\widetilde{\boldsymbol{\theta}}}_i = \mathbf{0}$. It follows that

$$\mathbb{P}(\gamma_i^\theta = 0 \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) \propto (1 - p_\theta)(2\pi)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^T h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0})'\boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0})\right)}. \tag{11}$$

Similarly,

$$\mathbb{P}(\gamma_i^\theta = 1 \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) \\ \propto p_\theta(2\pi)^{-\frac{T}{2}} |\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1)|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^T h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0})'\boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_{i,0}) - \widehat{\widetilde{\boldsymbol{\theta}}}_i(1)' \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1) \widehat{\widetilde{\boldsymbol{\theta}}}_i(1)\right)}, \tag{12}$$

where $\widehat{\widetilde{\boldsymbol{\theta}}}_i(1)$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1)$ denote respectively $\widehat{\widetilde{\boldsymbol{\theta}}}_i$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ evaluated at $\gamma_i^\theta = 1$. A draw from this Bernoulli distribution is standard once we have normalized the probabilities in (11) and (12).

Step 2. Again we sample γ_i^h and $\widetilde{\mathbf{h}}_i$ jointly to improve efficiency. Unlike Step 1, however, we cannot first sample γ_i^h marginally of $\widetilde{\mathbf{h}}_i$ followed by drawing $\widetilde{\mathbf{h}}_i$ from its full conditional distribution, because the latter distribution is nonstandard. Instead, we implement an independence-chain Metropolis-Hastings step using the joint proposal density $q(\gamma_i^h, \widetilde{\mathbf{h}}_i) = q(\gamma_i^h)q(\widetilde{\mathbf{h}}_i \,|\, \gamma_i^h)$.

Ideally, we would like to use a joint proposal that is "close" to the joint full posterior distribution of γ_i^h and $\widetilde{\mathbf{h}}_i$. To that end, we first use the marginal posterior $p(\gamma_i^h \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \sigma_{h,i}^2, h_{i,0})$—unconditional on $\widetilde{\mathbf{h}}_i$—as the marginal proposal of γ_i^h. More precisely, it is a Bernoulli distribution with success probability $\mathbb{P}(\gamma_i^h = 1 \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \sigma_{h,i}^2, h_{i,0})$. To compute this success probability, we need to evaluate the marginal density

$$p(\gamma_i^h \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \sigma_{h,i}^2, h_{i,0}) \propto \left[\int_{\mathbb{R}^T} p(\mathbf{y}_i \,|\, \widetilde{\mathbf{h}}_i, \boldsymbol{\theta}_i, \sigma_{h,i}^2, h_{i,0}, \gamma_i^h) p(\widetilde{\mathbf{h}}_i) \mathrm{d}\widetilde{\mathbf{h}}_i\right] p(\gamma_i^h),$$

where both the conditional likelihood $p(\mathbf{y}_i \,|\, \widetilde{\mathbf{h}}_i, \boldsymbol{\theta}_i, \sigma^2_{h,i}, h_{i,0}, \gamma_i^h)$ and the prior $p(\widetilde{\mathbf{h}}_i)$ are Gaussian. Unfortunately, the above integral is not available in closed-form. However, it can be evaluated via the importance sampling approach in Chan and Eisenstat (2018a,b) that uses a Gaussian approximation of the theoretical zero-variance importance sampling density. After normalization, we can therefore calculate $\mathbb{P}(\gamma_i^h = 1 \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \sigma^2_{h,i}, h_{i,0})$ and obtain a candidate draw $\gamma_i^{h*} \sim q(\gamma_i^h)$. If $\gamma_i^{h*} = 0$, we obtain a candidate $\widetilde{\mathbf{h}}_i^*$ from the prior distribution $p(\widetilde{\mathbf{h}}_i)$; if $\gamma_i^{h*} = 1$, $\widetilde{\mathbf{h}}_i^*$ is drawn from a Gaussian approximation of $p(\widetilde{\mathbf{h}}_i \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \sigma^2_{h,i}, h_{i,0}, \gamma_i^h = 1)$ using the methods in Chan (2017). Given the current draw $(\gamma_i^h, \widetilde{\mathbf{h}}_i)$, the candidate draw $(\gamma_i^{h*}, \widetilde{\mathbf{h}}_i^*)$ is accepted with probability

$$\min\left\{1, \frac{p(\mathbf{y}_i \,|\, \gamma_i^{h*}, \widetilde{\mathbf{h}}_i^*, \boldsymbol{\theta}_i, \sigma^2_{h,i}, \mathbf{h}_{i,0}) p(\gamma_i^{h*}) p(\widetilde{\mathbf{h}}_i^*)}{p(\mathbf{y}_i \,|\, \gamma_i^{h}, \widetilde{\mathbf{h}}_i, \boldsymbol{\theta}_i, \sigma^2_{h,i}, \mathbf{h}_{i,0}) p(\gamma_i^{h}) p(\widetilde{\mathbf{h}}_i)} \times \frac{q(\gamma_i^h, \widetilde{\mathbf{h}}_i)}{q(\gamma_i^{h*}, \widetilde{\mathbf{h}}_i^*)}\right\}.$$

Finally, we use the sampled value of $\widetilde{\mathbf{h}}_i$ to get a draw of $\mathbf{h}_i$ by setting $\mathbf{h}_i = h_{i,0}\mathbf{1}_T + \gamma_i^h \sigma_{h,i} \widetilde{\mathbf{h}}_i$. The details of the remaining steps are provided in Appendix A.

4 Illustration

In this section we apply the proposed model in a simple setting to demonstrate that it can recover salient patterns of the data. To that end, we aim to replicate the the model comparison exercise in Chan and Eisenstat (2018b). More specifically, Chan and Eisenstat (2018b) compare a number of small hybrid TVP-VARs with different combinations of time-varying parameters and constant coefficients using a dataset that consists of US quarterly GDP deflator inflation, real GDP growth, and Fed funds rate. Since there are three variables in the VAR and in each equation the coefficients can either be constant or time varying—all the models include stochastic volatility—there are in total eight hybrid TVP-VARs. The model comparison there is done by computing the marginal likelihood of each model. For each hybrid TVP-VAR, the model first needs to be estimated. Then, an adaptive importance sampling estimator is used to compute the marginal likelihood. Obviously this approach is too computationally intensive to be applicable to large systems.

Here we fit the same dataset using the proposed hybrid TVP-VAR.[8] Of particular interest are the estimates of the indicators, γ_i^θ and γ_i^h, as each can be interpreted as the posterior model probability comparing an equation with time-vary parameters versus one with constant parameters. The estimates are reported in Table 1.

Table 1: Posterior estimates of γ_i^θ and γ_i^h for the hybrid TVP-VAR with $n = 3$ variables.

Equation	Chan and Eisenstat (2018b)	γ_i^θ	γ_i^h
GDP deflator	time varying	1	1
Real GDP	time varying	0.97	1
Fed funds rate	constant	0	1

The best model in Chan and Eisenstat (2018b) is found to be the hybrid TVP-VAR that has time-varying parameters in the GDP deflator and real GDP equations, whereas the interest rate equation has constant coefficients. Consistent with the model comparison result there, our estimates of γ_i^θ for the three equations are, respectively, 1, 0.97 and 0, indicating that one needs time-varying parameters in only the first two equations but not the third. Moreover, since all the indicators γ_i^h are estimated to be one, there is strong evidence for stochastic volatility in all equations, supporting the modeling choice of including stochastic volatility in all models in Chan and Eisenstat (2018b).

Overall, our results confirm that the proposed hybrid model can recover salient patterns—such as time-varying conditional means and variances—in the data. Comparing to performing a formal Bayesian model comparison, our approach is computationally efficient as it only requires fitting the proposed hybrid TVP-VAR once. Moreover, it is also applicable to fitting large systems, as our empirical application in the following section demonstrates.

5 Application: Trend Estimation and Forecasting

In this section we fit a large US macroeconomic dataset set to demonstrate the usefulness of the proposed model. After describing the dataset in Section 5.1, we first present the

[8]To make our results comparable to those in Chan and Eisenstat (2018b), we use the same priors there whenever it is possible.

full sample results in Section 5.2, particularly, the estimates of the trend output growth and trend inflation from the model. We then consider a pseudo out-of-sample forecasting exercise in Section 5.3. We show that the forecast performance of the proposed model compares favorably to a range of standard benchmarks.

5.1 Data and Prior Hyperparameters

The US dataset for our empirical application consists of 20 quarterly variables with a sample period from 1959Q1 to 2018Q4. It is sourced from the FRED-QD database at the Federal Reserve Bank of St. Louis as described in McCracken and Ng (2016). Our dataset contains a variety of standard macroeconomic and financial variables, such as Real GDP, industrial production, inflation rates, labor market variables, money supply and interest rates. They are transformed to stationarity, typically to annualized growth rates. The complete list of variables and how they are transformed is given in Appendix B.

We use the priors described in Section 3.1. In particular, since the data are transformed to growth rates, we set the prior mean of $\boldsymbol{\theta}_0$ to be zero, i.e., $\mathbf{a}_{\boldsymbol{\theta}_0} = \mathbf{0}$. For the prior hyperparameters on κ_1 and κ_2, we set $c_{1,1} = c_{1,2} = 1$, $c_{2,1} = 1/0.04$ and $c_{2,2} = 1/0.04^2$. These values imply that the prior means of κ_1 and κ_2 are respectively 0.04 and 0.04^2. For the hyperparameters of the initial conditions $\mathbf{h}_0$, we set $\mathbf{a}_h = \mathbf{0}$ and $\mathbf{V}_h = 10 \times \mathbf{I}_n$. Next, the hyperparameter of $\sigma_{h,i}$ is set so that the prior mean is 0.1. In other words, the difference between consecutive log-volatilities is within 0.2 with probability of about 0.95. Similarly, the implied prior mean of $\sigma_{\theta_i,j}$ is 0.005^2 if it is associated with a VAR coefficient and 0.01^2 for an intercept. Finally, we set $a_{p^\theta} = b_{p^\theta} = a_{p^h} = b_{p^h} = 0.1$. These values imply prior modes at 0 and 1, whereas the prior mean is 0.5.

5.2 Full Sample Results

In this section we report the full sample results of the hybrid TVP-VAR fitted using all $n = 20$ variables. Of particular interest are the posterior estimates of γ_i^θ and γ_i^h, the indicators that control time variation in the VAR coefficients and the variance, respectively. The estimates are reported in Table 2.

As the table shows, the estimates of γ_i^h for most equations are close to one, suggesting that the stochastic volatility specification is needed for most equations. Our results thus confirm the importance of allowing for time-varying variance in modeling macroeconomic and financial data. However, there seems to be much less time variation in the VAR coefficients—for the majority of the equations, the estimates of γ_i^θ are close to zero, implying that time variation in those equations is essentially turned off. Overall, while we find evidence for time-varying VAR coefficients and stochastic volatility in some equations, not all equations need both forms of time variation. Our results therefore highlight the empirical relevance of the proposed hybrid TVP-VAR.

Table 2: Posterior estimates of γ_i^θ and γ_i^h for the hybrid TVP-VAR with $n = 20$ variables.

Equation	γ_i^θ	γ_i^h
Real GDP	0.96	1
PCE inflation	0.90	1
Unemployment	0	1
Fed funds rate	0	1
Industrial production index	0.99	1
Real average hourly earnings in manufacturing	1	0.15
M1	1	1
PCE	0.90	1
Real disposable personal income	0.06	1
Industrial production: final products	0	1
All employees: total nonfarm	0	0.95
Civilian employment	0	0.68
Nonfarm business section: hours of all persons	0	0.85
GDP deflator	0	1
CPI	0	1
PPI	1	1
Nonfarm business sector: real compensation per hour	1	1
Nonfarm business section: real output per hour	0	1
10-year treasury constant maturity rate	0	1
M2	1	1

Next, we report the model-implied long-run trends of GDP growth and PCE inflation. Given the current slow growth and low inflation macroeconomic environment, there is much interest to understand whether this is transitory as part of the slow recovery from the Great Recession, or they reflect deeper structural changes in the economy. To that end, we compute the potentially time-varying unconditional means of the dependent variables implied by the hybrid TVP-VARs. More precisely, given a set of posterior draws of the structural-form VAR coefficients $\mathbf{A}_t, \mathbf{b}_t, \mathbf{B}_{1,t}, \ldots, \mathbf{B}_{p,t}$, we first compute the implied reduced-form VAR coefficients $\widetilde{\mathbf{b}}_t = \mathbf{A}_t^{-1}\mathbf{b}_t$, $\widetilde{\mathbf{B}}_{j,t} = \mathbf{A}_t^{-1}\mathbf{B}_{j,t}, j = 1, \ldots, p$. Then, set $\mathbf{c}_t = (\widetilde{\mathbf{b}}_t', \mathbf{0}, \ldots, \mathbf{0})'$ and let $\mathbf{P}_t$ denote the companion matrix associated with the reduced-form VAR parameters, i.e.,

$$\mathbf{P}_t = \begin{pmatrix} \widetilde{\mathbf{B}}_{1,t} & \widetilde{\mathbf{B}}_{2,t} & \cdots & \cdots & \widetilde{\mathbf{B}}_{p,t} \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_n & \mathbf{0} \end{pmatrix}.$$

Finally, we use the first n elements of $\boldsymbol{\mu}_t = (\mathbf{I}_{np} - \mathbf{P}_t)^{-1}\mathbf{c}_t$ as the unconditional means of the dependent variables and interpret them as the long-run trend estimates.

Figure 1 reports the long-run trend estimates of real GDP growth and PCE inflation rate. As a comparison, we also report the corresponding long-run trend estimates for a small system ($n = 3$) with real GDP, PCE inflation and unemployment, as well as a medium system ($n = 7$) with four additional variables: Fed funds rate, industrial production, real average hourly earnings in manufacturing and M1 money stock. These variables are similar to those used in Morley and Wong (2019) for estimating the output gap.

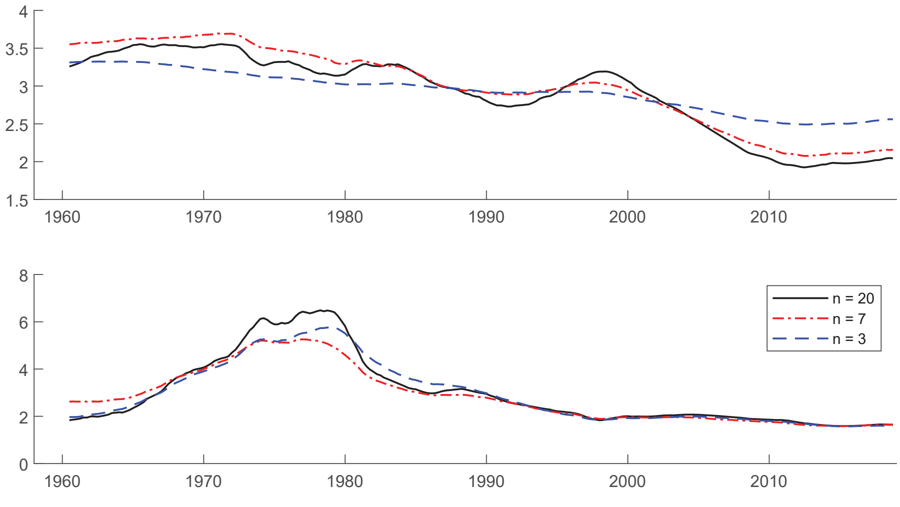

Figure 1: Long-run trend estimates (posterior medians) of real GDP growth (top) and PCE inflation rate (bottom).

It is evident from the figure that that there is substantial time variation in the trend output growth under the large hybrid TVP-VAR over the past six decades. Specifically, the trend growth rate fluctuates around 3.5% from the beginning of the sample until early 1970s. It then begins a steady decline and reaches a trough of about 3% in early 1990s. There is a slight pickup in trend growth in late 1990s, followed by a long, gradual decline that begins around 2000 to about 1.9% in 2012. At the end of the sample, the trend output growth is 2%. The general evolution of the trend output growth is similar to those obtained previously in the literature (e.g. Berger, Everaert, and Vierke, 2016; Grant and Chan, 2017) using unobserved components models. Moreover, the timing of the apparent slowdown in trend output is consistent with the breakdates identified in Perron and Wada (2009) and Morley and Panovska (2019). In particular, our results suggest that the decline of trend output started before the onset of the Great Recession.

Compared to the large VAR, the trend output estimates under the small and medium VARs show less time variation—e.g., the pickup in trend output in 2000 is less apparent in both cases. This highlights the fact that estimates from VARs of different sizes could be quite different, and using a richer set of variables might give better estimates.

Figure 1 also depicts the long-run trend estimates of PCE inflation. Here the three VARs give similar estimates, except for the Great Inflation period. In particular, the results show a gradual decline in trend inflation from early 1980s to late 1990s. Trend inflation has mostly stayed at around 2% since early 2000s, although there is a slight decline after 2010. At the end of the sample, the trend inflation is about 1.7%.

5.3 Forecasting Results

In this section we evaluate the forecast performance of the proposed hybrid TVP-VARs relative to a few standard benchmarks. In particular, we consider the conventional homoscedastic, constant-coefficient VAR (by setting all the indicators to 0), the constant-coefficient VAR with stochastic volatility (by setting all γ_i^h to 0 and all γ_i^θ to 1), and the full-fledged TVP-VAR (by setting all the indicators to 1). All models use all $n = 20$ variables. The sample period is from 1959Q1 to 2018Q4, and the evaluation period starts at 1985Q1 and runs till the end of the sample.

We perform a recursive forecasting exercise using an expanding window. More specifically, in each forecasting iteration t, we use only data up to time t, denoted as $\mathbf{y}_{1:t}$, to estimate the models. We then evaluate both point and density forecasts. We use the conditional expectation $\mathbb{E}(y_{i,t+m} \,|\, \mathbf{y}_{1:t})$ as the m-step-ahead point forecast for variable i and the predictive density $p(y_{i,t+m} \,|\, \mathbf{y}_{1:t})$ as the corresponding density forecast.

The metric used to evaluate the point forecasts from model M is the root mean squared forecast error (RMSFE) defined as

$$\text{RMSFE}_{i,m}^{M} = \sqrt{\frac{\sum_{t=t_0}^{T-m}(y_{i,t+m}^{\text{o}} - \mathbb{E}(y_{i,t+m} \,|\, \mathbf{y}_{1:t}))^2}{T-m-t_0+1}},$$

where $y_{i,t+m}^{\text{o}}$ is the actual observed value of $y_{i,t+m}$. For RMSFE, a smaller value indicates better forecast performance. To evaluate the density forecasts, the metric we use is the average of log predictive likelihoods (ALPL):

$$\text{ALPL}_{i,m}^{M} = \frac{1}{T-m-t_0+1} \sum_{t=t_0}^{T-m} \log p(y_{i,t+m} = y_{i,t+m}^{\text{o}} \,|\, \mathbf{y}_{1:t}),$$

where $p(y_{i,t+m} = y_{i,t+m}^{\text{o}} \,|\, \mathbf{y}_{1:t})$ is the predictive likelihood. For this metric, a larger value indicates better forecast performance.

To compare the forecast performance of model M against the benchmark B, we follow Carriero, Clark, and Marcellino (2015) to report the percentage gains in terms of RMSFE, defined as

$$100 \times (1 - \text{RMSFE}_{i,m}^{M}/\text{RMSFE}_{i,m}^{B}),$$

and the percentage gains in terms of ALPL:

$$100 \times (\text{ALPL}_{i,m}^{M} - \text{ALPL}_{i,m}^{B}).$$

Figure 2 reports the forecasting results of the hybrid TVP-VAR, where we use the conventional homoscedastic, constant-coefficient VAR as the benchmark. The top panel shows the percentage gains in RMSFE for all 20 variables, and the bottom panel presents the corresponding results in ALPL.

For both 1- and 4-step-ahead point forecasts, the hybrid TVP-VAR outperforms the benchmark for almost all variables (all but one for 1-step-ahead and all but two for 4-

step ahead). For a few variables, such as Fed funds rate, CPI inflation and industrial production, the former outperforms the benchmark by more than 10% for 1-step-ahead forecasts. Overall, the median percentage gains in RMSFE for 1- and 4-step-ahead forecasts are, respectively, 4.7% and 6.1%.

For density forecasts, the hybrid TVP-VAR performs even better relative to the benchmark—it outperforms the benchmark for all variables in both forecast horizons. The median percentage gains in ALPL for 1- and 4-step-ahead forecasts are 13% and 11%, respectively. Moreover, for many variables the percentage gains are more than 20%. These results are consistent with the small VAR literature that shows allowing for time-varying structures substantially improves forecast performance compared to VARs with constant parameters, especially for density forecasts.

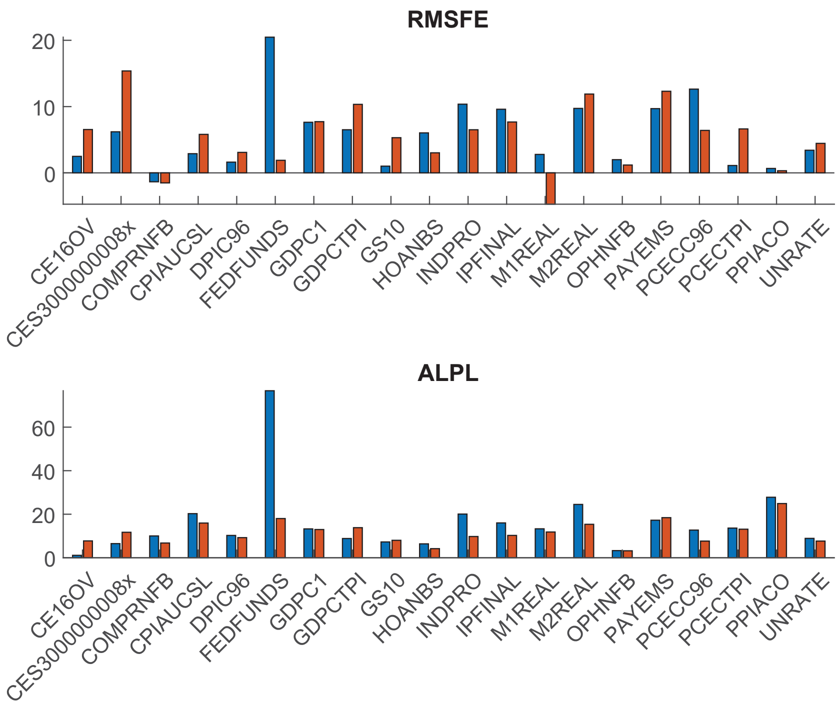

Figure 2: Forecasting results of the hybrid TVP-VAR compared to the benchmark: a standard homoscedastic, constant-coefficient VAR. The top panel shows the percentage gains in root mean squared forecast error of the asymmetric conjugate prior. The bottom panel presents the percentage gains in the average of log predictive likelihoods.

Next, we compare the forecast performance of the hybrid TVP-VAR with that of the constant-coefficient VAR with stochastic volatility (VAR-SV), and the results are reported in Figure 3. For both 1- and 4-step-ahead density forecasts, the hybrid TVP-VAR substantially outperforms the VAR-SV for most variables. The median percentage gains in RMSFE and ALPL are 7.3% and 11.5%, respectively. For point forecasts, the results are similar, though the gains are more modest. The median percentage gains in RMSFE for 1- and 4-step-ahead forecasts are, respectively, 1.5% and 2.7%. Overall, these results suggest that allowing for time variation in VAR coefficients—with appropriate shrinkage and sparsification—can further enhance the forecast performance of a VAR with stochastic volatility.

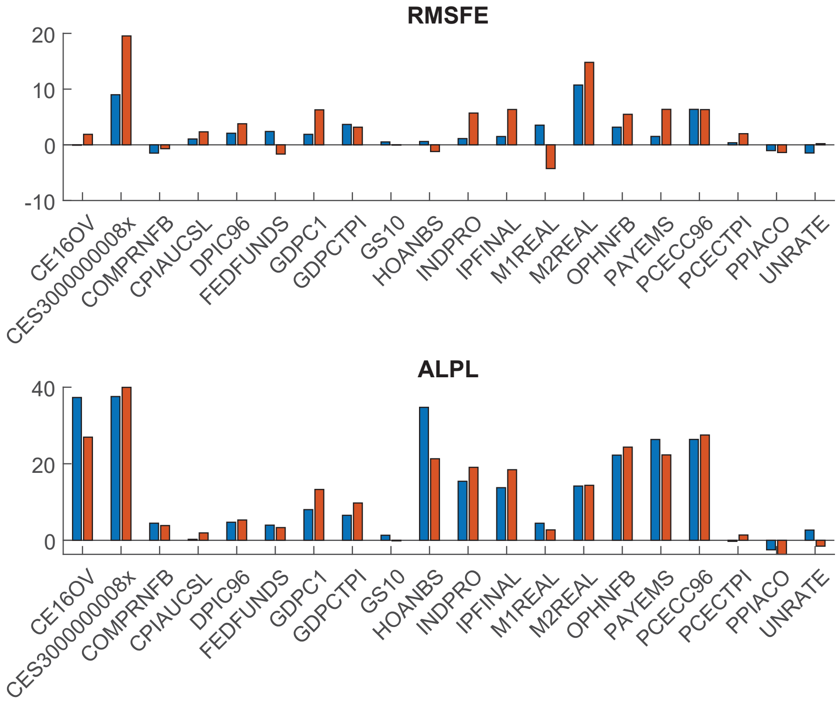

Figure 3: Forecasting results of the hybrid TVP-VAR compared to the benchmark VAR-SV. The top panel shows the percentage gains in root mean squared forecast error of the asymmetric conjugate prior. The bottom panel presents the percentage gains in the average of log predictive likelihoods.

Finally, Figure 4 compares the forecast performance of the the hybrid TVP-VAR with that of the full-fledged TVP-VAR where all the VAR coefficients and error variances are time varying. For point forecasts, the hybrid TVP-VAR performs slightly worse than the benchmark. In particular, the median percentage gains in RMSFE for 1- and 4-step-ahead forecasts are −1.8% and −1.0%, respectively. However, for density forecasts, the hybrid TVP-VAR does substantially better for most variables. The median percentage gains in ALPL are 8.6% and 9.9%, respectively. These results suggest that the hybrid TVP-VAR constructs better predictive distributions, possibly due to the shrinkage and sparsification built into the model.

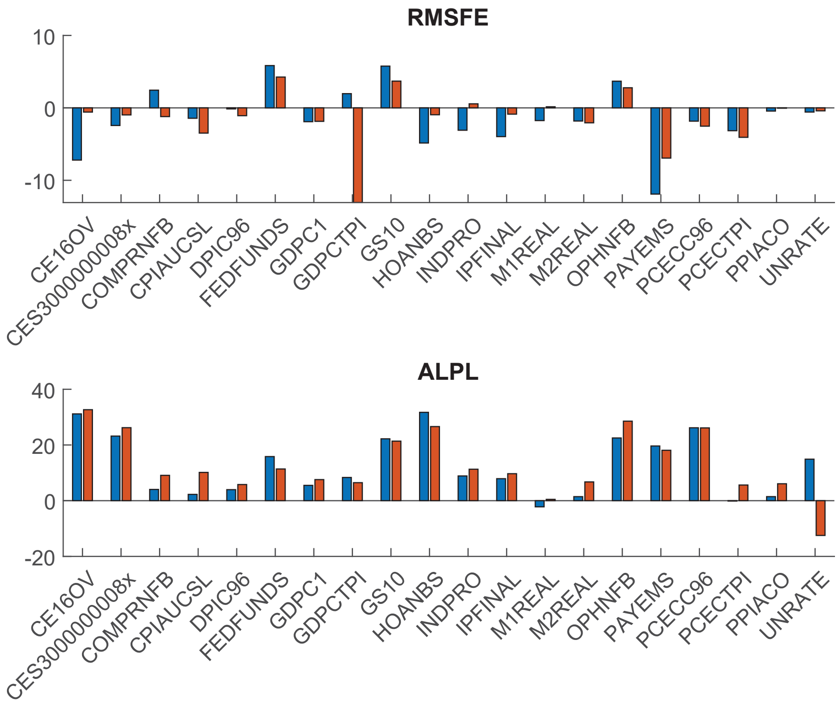

Figure 4: Forecasting results of the hybrid TVP-VAR compared to the benchmark TVP-VAR. The top panel shows the percentage gains in root mean squared forecast error of the asymmetric conjugate prior. The bottom panel presents the percentage gains in the average of log predictive likelihoods.

Overall, these forecasting results show that the proposed hybrid TVP-VAR forecasts better than many state-of-the-art time-varying models. These results highlight the advantages of using a data-driven approach to discover the time-varying structures—rather than imposing either constant coefficients or time variation in parameters.

6 Concluding Remarks and Future Research

We have developed a new class of models we call hybrid TVP-VARs, i.e., VARs with time-varying parameters in some equations but not in others. Using US data, we found evidence that while VAR coefficients and error variances in some equations are time varying, the data prefer constant coefficients or homoscedastic errors in others. In a forecasting exercise that involves 20 macroeconomic and financial variables, we demonstrated the superior forecast performance of the proposed hybrid TVP-VARs compared to standard benchmarks.

In future work, it would be interesting to develop methods to allow for more sophisticated forms of time variation in coefficients. For example, while it is straightforward to introduce multiple indicators in each equation to control the time variation in different groups of VAR coefficients, it is more difficult to handle a group of coefficients that spans across equations within the current equation-by-equation estimation framework. Moreover, dynamic sparsification—e.g., restricting a coefficient to be constant in some periods but allowing it to be time-varying in others—would be an interesting and important extension.

Appendix A: Estimation Details

In this appendix we provide estimation details of the hybrid TVP-VAR given in (2)-(4). In particular, we describe the details of the remaining steps of the posterior sampler.

Step 3. The parameters $\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}}$ and $\boldsymbol{\theta}_{i,0}$ can be sampled easily as their joint distribution is Gaussian. To see that, let $\boldsymbol{\mu}_i^\theta = (\boldsymbol{\theta}_{i,0}', \sigma_{\theta_i,1}, \ldots, \sigma_{\theta_i,k_{\theta_i}})$ and define $\mathbf{w}_{i,t}^\theta = (\mathbf{x}_{i,t}, \gamma_i^\theta \mathbf{x}_{i,t} \odot \widetilde{\boldsymbol{\theta}}_{i,t})$, where $\odot$ denotes the component-wise product. Then, we can rewrite (5) as a linear regression:

$$\mathbf{y}_{i,t} = \mathbf{w}_{i,t}^\theta \boldsymbol{\mu}_i^\theta + \varepsilon_{i,t}^y.$$

Since both $\boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}}$ and $\boldsymbol{\theta}_{i,0}$ have Gaussian priors, the implied prior of $\boldsymbol{\mu}_i^\theta$ is also Gaussian: $\boldsymbol{\mu}_i^\theta \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\boldsymbol{\mu}_i^\theta})$, where $\mathbf{V}_{\boldsymbol{\mu}_i^\theta} = \text{diag}(\mathbf{V}_{\boldsymbol{\theta}_{i,0}}, S_{\theta_i,1}, \ldots, S_{\theta_i,k_{\theta_i}})$. Define $\mathbf{W}_i^\theta$ by stacking $\mathbf{w}_{i,t}^\theta$ over $t = 1, \ldots, T$. It follows that the full conditional distribution of $\boldsymbol{\mu}_i^\theta$ is given by

$$(\boldsymbol{\mu}_i^\theta \,|\, \mathbf{y}_i, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \gamma_i^\theta, \gamma_i^h) \sim \mathcal{N}(\widehat{\boldsymbol{\mu}}_i^\theta, \mathbf{K}_{\boldsymbol{\mu}_i^\theta}^{-1}),$$

where $\mathbf{K}_{\boldsymbol{\mu}_i^\theta} = \mathbf{V}_{\boldsymbol{\mu}_i^\theta}^{-1} + (\mathbf{W}_i^\theta)' \boldsymbol{\Sigma}_{h_i}^{-1} \mathbf{W}_i^\theta$ and $\widehat{\boldsymbol{\mu}}_i^\theta = \mathbf{K}_{\boldsymbol{\mu}_i^\theta}^{-1} (\mathbf{W}_i^\theta)' \boldsymbol{\Sigma}_{h_i}^{-1} \mathbf{y}_i$.

Step 4. The full conditional distribution of $\sigma_{h,i}$ and $h_{i,0}$ is nonstandard, and we simulate $\sigma_{h,i}$ and $h_{i,0}$ using an independence-chain Metropolis-Hastings step. Specifically, the full conditional density of $(\sigma_{h,i}, h_{i,0})$ is given by

$$p(\sigma_{h,i}, h_{i,0} \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \widetilde{\mathbf{h}}_i, \gamma_i^h) \propto p(\mathbf{y}_i \,|\, \sigma_{h,i}, h_{i,0}, \boldsymbol{\theta}_i, \widetilde{\mathbf{h}}_i, \gamma_i^h) p(\sigma_{h,i}) p(h_{i,0}),$$

where both priors $p(\sigma_{h,i})$ and $p(h_{i,0})$ are Gaussian and

$$\log p(\mathbf{y}_i \,|\, \sigma_{h,i}, h_{i,0}, \boldsymbol{\theta}_i, \widetilde{\mathbf{h}}_i, \gamma_i^h) = -\frac{T}{2} \log(2\pi \mathrm{e}^{h_{i,0}}) - \frac{1}{2} \sigma_{h,i} \gamma_i^h \sum_{t=1}^T \widetilde{h}_{i,t} - \frac{1}{2} \sum_{t=1}^T \mathrm{e}^{-h_{i,0} - \sigma_{h,i} \gamma_i^h \widetilde{h}_{i,t}} (\varepsilon_{i,t}^y)^2.$$

Hence, we can readily compute the gradient and Hessian of $\log p(\sigma_{h,i}, h_{i,0} \,|\, \mathbf{y}_i, \boldsymbol{\theta}_i, \widetilde{\mathbf{h}}_i, \gamma_i^h)$ with respect to $(\sigma_{h,i}, h_{i,0})$. Then, the mode of this full conditional density can be obtained by, e.g., Newton-Raphson method. Finally, we construct a Gaussian proposal with mean and precision matrix (inverse of the covariance matrix) set to be, respectively, the mode and the negative Hessian of $\log p(\sigma_{h,i}, h_{i,0} \,|\, \mathbf{y}_i, \widetilde{\mathbf{h}}_i, \gamma_i^h)$.

Step 5. Next, given the independent beta priors on p_i^θ and $p_i^h, i = 1, \ldots, n$ their full conditional posterior distributions are also beta distributions. In fact, we have:

$$(p_i^\theta \,|\, \gamma_i^\theta) \sim \mathcal{B}(a_{p^\theta} + \gamma_i^\theta, b_{p^\theta} + 1 - \gamma_i^\theta),$$
$$(p_i^h \,|\, \gamma_i^h) \sim \mathcal{B}(a_{p_h} + \gamma_i^h, b_{p_h} + 1 - \gamma_i^h).$$

Step 6. To implement Step 6, we follow the sampling approach in Chan (2019). First note that κ_1 and κ_2 only appear in their priors $\kappa_j \sim \mathcal{G}(c_{1,j}, c_{2,j}), j = 1, 2$, and in the prior covariance matrices $\mathbf{V}_{\boldsymbol{\theta}_{i,0}}, i = 1, \ldots, n$. Letting $\theta_{ij,0}$ denote the j-th element of $\boldsymbol{\theta}_{i,0}$, we define the index set S_{κ_1} to be the collection of indexes (i, j) such that $\theta_{ij,0}$ is a coefficient associated with an own lag. That is, $S_{\kappa_1} = \{(i, j) : \theta_{ij,0}$ is a coefficient associated with an own lag$\}$. Similarly, define S_{κ_2} as the set that collects all the indexes (i, j) such that $\theta_{ij,0}$ is a coefficient associated with a lag of other variables. It is easy to check that the numbers of elements in S_{κ_1} and S_{κ_2} are respectively np and $(n-1)np$. Further, for $(i, j) \in S_{\kappa_1} \cup S_{\kappa_2}$, let

$$C_{ij} = \begin{cases} \frac{1}{l^2}, & \text{for the coefficient on the } l\text{-th lag of variable } i, \\ \frac{s_i^2}{l^2 s_j^2}, & \text{for the coefficient on the } l\text{-th lag of variable } j, j \neq i. \end{cases}$$

Then, we have

$$\begin{aligned} p(\kappa_1 \,|\, \boldsymbol{\theta}_0) &\propto \prod_{(i,j)\in S_{\kappa_1}} \kappa_1^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\kappa_1 C_{ij}}\theta_{ij,0}^2} \times \kappa_1^{c_{1,1}-1} \mathrm{e}^{-\kappa_1 c_{2,1}} \\ &= \kappa_1^{c_{1,1}-\frac{np}{2}-1} \mathrm{e}^{-\frac{1}{2}\left(2c_{2,1}\kappa_1 + \kappa_1^{-1} \sum_{(i,j)\in S_{\kappa_1}} \frac{\theta_{ij,0}^2}{C_{ij}}\right)}, \end{aligned}$$

which is the kernel of the $\mathcal{GIG}\left(c_{1,1} - \frac{np}{2}, 2c_{2,1}, \sum_{(i,j)\in S_{\kappa_1}} \frac{\theta_{ij,0}^2}{C_{ij}}\right)$ distribution. Similarly, we have

$$(\kappa_2 \,|\, \boldsymbol{\theta}_0) \sim \mathcal{GIG}\left(c_{1,2} - \frac{(n-1)np}{2}, 2c_{2,2}, \sum_{(i,j)\in S_{\kappa_2}} \frac{\theta_{ij,0}^2}{C_{ij}}\right).$$

Appendix B: Data

The dataset covers 20 quarterly variables sourced from the FRED-QD database at the Federal Reserve Bank of St. Louis (McCracken and Ng, 2016). The sample period is from 1959Q1 to 2018Q4. Table 3 lists all the variables and describes how they are transformed. For example, $\Delta \log$ is used to denote the first difference in the logs, i.e., $\Delta \log x = \log x_t - \log x_{t-1}$.

Table 3: Description of variables used in empirical application.

Variable	Mnemonic	Transformation
Real Gross Domestic Product	GDPC1	$400\Delta \log$
Personal Consumption Expenditures: Chain-type Price index	PCECTPI	$400\Delta \log$
Civilian Unemployment Rate	UNRATE	no transformation
Effective Federal Funds Rate	FEDFUNDS	no transformation
Industrial Production Index	INDPRO	$400\Delta \log$
Real Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	CES3000000008x	$400\Delta \log$
Real M1 Money Stock	M1REAL	$400\Delta \log$
Personal Consumption Expenditures	PCECC96	$400\Delta \log$
Real Disposable Personal Income	DPIC96	$400\Delta \log$
Industrial Production: Final Products	IPFINAL	$400\Delta \log$
All Employees: Total nonfarm	PAYEMS	$400\Delta \log$
Civilian Employment	CE16OV	$400\Delta \log$
Nonfarm Business Section: Hours of All Persons	HOANBS	$400\Delta \log$
Gross Domestic Product: Chain-type Price index	GDPCTPI	$400\Delta \log$
Consumer Price Index for All Urban Consumers: All Items	CPIAUCSL	$400\Delta \log$
Producer Price Index for All commodities	PPIACO	$400\Delta \log$
Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	$400\Delta \log$
Nonfarm Business Section: Real Output Per Hour of All Persons	OPHNFB	$400\Delta \log$
10-Year Treasury Constant Maturity Rate	GS10	no transformation
Real M2 Money Stock	M2REAL	$400\Delta \log$

References

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013): "Now-casting and the real-time data flow," in *Handbook of Economic Forecasting*, vol. 2, pp. 195–237. Elsevier.

Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92.

Banbura, M., and A. van Vlodrop (2018): "Forecasting with Bayesian Vector Autoregressions with Time Variation in the Mean," *Tinbergen Institute Discussion Paper 2018-025/IV*.

Berger, T., G. Everaert, and H. Vierke (2016): "Testing for time variation in an unobserved components model for the U.S. economy," *Journal of Economic Dynamics and Control*, 69, 179–208.

Carriero, A., T. E. Clark, and M. G. Marcellino (2015): "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, 30(1), 46–73.

——— (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business and Economic Statistics*, 34(3), 375–390.

——— (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, Forthcoming.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417.

Chan, J. C. C. (2017): "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.

——— (2018): "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business and Economic Statistics*, Forthcoming.

——— (2019): "Minnesota-Type Adaptive Hierarchical Priors for Large Bayesian VARs," *CAMA Working Paper 61/2019*.

Chan, J. C. C., and E. Eisenstat (2018a): "Bayesian Model Comparison for Time-Varying Parameter VARs with Stochastic Volatility," *Journal of Applied Econometrics*, 33(4), 509–532.

——— (2018b): "Comparing Hybrid Time-Varying Parameter VARs," *Economics Letters*, 171, 1–5.

Chan, J. C. C., and A. L. Grant (2016): "Fast Computation of the Deviance Information Criterion for Latent Variable Models," *Computational Statistics and Data Analysis*, 100, 847–859.

Chan, J. C. C., and I. Jeliazkov (2009): "Efficient Simulation and Integrated Likelihood Estimation in State Space Models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1), 101–120.

Clark, T. E. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business and Economic Statistics*, 29(3), 327–341.

Clark, T. E., and F. Ravazzolo (2014): "Macroeconomic Forecasting Performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, Forthcoming.

Cogley, T., and T. J. Sargent (2001): "Evolving post-world war II US inflation dynamics," *NBER Macroeconomics Annual*, 16, 331–388.

——— (2005): "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8(2), 262–302.

Cross, J., C. Hou, and A. Poon (2019): "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *Working Paper*.

Cross, J., and A. Poon (2016): "Forecasting structural change and fat-tailed events in Australian macroeconomic variables," *Economic Modelling*, 58, 34–51.

D'Agostino, A., L. Gambetti, and D. Giannone (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101.

Del Negro, M., and F. Schorfheide (2012): "Bayesian Macroeconometrics," in *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3(1), 1–100.

Eisenstat, E., J. C. C. Chan, and R. W. Strachan (2018): "Reducing Dimensions in a Large TVP-VAR," *Working Paper series 18-37, Rimini Centre for Economic Analysis*.

Ellahie, A., and G. Ricco (2017): "Government purchases reloaded: Informational insufficiency and heterogeneity in fiscal VARs," *Journal of Monetary Economics*, 90, 13–27.

Frühwirth-Schnatter, S., and H. Wagner (2010): "Stochastic model specification search for Gaussian and partial non-Gaussian state space models," *Journal of Econometrics*, 154, 85–100.

Gefang, D., G. Koop, and A. Poon (2019): "Variational Bayesian inference in large Vector Autoregressions with hierarchical shrinkage," *CAMA Working Paper*.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97(2), 436–451.

Götz, T., and K. Hauzenberger (2018): "Large mixed-frequency VARs with a parsimonious time-varying parameter structure," *Deutsche Bundesbank Discussion Paper*.

Grant, A. L., and J. C. C. Chan (2017): "Reconciling Output Gaps: Unobserved Components Model and Hodrick-Prescott Filter," *Journal of Economic Dynamics and Control*, 75, 114–121.

Huber, F., G. Koop, and L. Onorante (2019): "Inducing Sparsity and Shrinkage in Time-Varying Parameter Models," *arXiv preprint arXiv:1905.10787*.

Kadiyala, K., and S. Karlsson (1997): "Numerical Methods for Estimation and inference in Bayesian VAR-models," *Journal of Applied Econometrics*, 12(2), 99–132.

Karlsson, S. (2013): "Forecasting with Bayesian vector autoregressions," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann, vol. 2 of *Handbook of Economic Forecasting*, pp. 791–897. Elsevier.

Kastner, G., and F. Huber (2018): "Sparse Bayesian vector autoregressions in huge dimensions," *arXiv preprint arXiv:1704.03239*.

Koop, G. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28(2), 177–203.

Koop, G., and D. Korobilis (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends in Econometrics*, 3(4), 267–358.

——— (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198.

Litterman, R. (1986): "Forecasting With Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

McCracken, M. W., and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business and Economic Statistics*, 34(4), 574–589.

Morley, J., and I. B. Panovska (2019): "Is Business Cycle Asymmetry Intrinsic in Industrialized Economies?," *Macroeconomic Dynamics*, pp. 1–34.

Morley, J., and B. Wong (2019): "Estimating and accounting for the output gap with large Bayesian vector autoregressions," *Journal of Applied Econometrics*, forthcoming.

Perron, P., and T. Wada (2009): "Let's take a break: Trends and cycles in US real GDP," *Journal of Monetary Economics*, 56(6), 749–765.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Sims, C. A., and T. Zha (1998): "Bayesian methods for dynamic multivariate models," *International Economic Review*, 39(4), 949–968.