Large Hybrid Time-Varying Parameter VARs

Joshua C.C. Chan*
Purdue University

This version: May 2022
First version: September 2019

Abstract

Time-varying parameter VARs with stochastic volatility are routinely used for structural analysis and forecasting in settings involving a few endogenous 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 whether the VAR coefficients and contemporaneous relations among variables are constant or time-varying. Using US datasets of various dimensions, we find evidence that the parameters in some, but not all, equations are time varying. The large hybrid TVP-VAR also forecasts better than many standard benchmarks.

Keywords: large vector autoregression, time-varying parameter, stochastic volatility, macroeconomic forecasting, Bayesian model averaging
JEL classifications: C11, C52, C55, E37, E47

*This paper has benefited from the constructive comments and suggestions from 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, as shown in papers such as Clark (2011), D'Agostino, Gambetti, and Giannone (2013), Koop and Korobilis (2013), Clark and Ravazzolo (2015) and Cross and Poon (2016). 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 increasingly 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). There is now a 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). 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; Kastner and Huber, 2018; Chan, 2020, 2021), the literature on large VARs with time-varying coefficients remains relatively scarce.

We propose a class of models we call hybrid TVP-VARs—VARs in which some equations have time-varying coefficients, whereas the coefficients are constant in others. More precisely, we develop an efficient Bayesian shrinkage and sparsification method that automatically decides, for each equation, (i) whether the VAR coefficients are constant or time-varying, and (ii) whether the parameters of the contemporaneous relations among variables are constant or time-varying. Given the importance of time-varying volatility, all equations feature stochastic volatility. Our framework nests many popular VARs as special cases, ranging from a constant-coefficient VAR with stochastic volatility on one

end of the spectrum to the flexible but highly parameterized TVP-VARs of Cogley and Sargent (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. This structural-form parameterization, however, raises the issue of variable ordering, that is, the assumed order of the variables might affect the model estimates compared to a standard reduced-form TVP-VAR. We investigate this issue empirically and find that the variability of the estimates from this structural-form parameterization is comparable to that of the TVP-VAR of Primiceri (2005).

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 VAR coefficients are time-varying or constant, while the other controls whether the elements of the impact matrix are time-varying or not. Hence, each vector $\boldsymbol{\gamma} \in \{0, 1\}^{2n}$, where n is the number of endogenous variables, characterizes a hybrid TVP-VAR with a particular form of time variation. 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. This data-driven hybrid TVP-VAR can also be interpreted as a Bayesian model average of 2^{2n} hybrid TVP-VARs with different forms of time variation, where the weights are

determined by the posterior model probabilities $p(\boldsymbol{\gamma} \mid \mathbf{y})$. It follows that forecasts from such a model can be viewed as a forecast combination of a wide variety of hybrid TVP-VARs.

The estimation is done using Markov chain Monte Carlo (MCMC) methods. Hence, in contrast to 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 therefore develop algorithms to sample the indicators and the latent states jointly.

Using US datasets of different dimensions, we find evidence that the VAR coefficients and elements of the impact matrix in some, but not all, equations are time varying. In particular, in a formal Bayesian model comparison exercise, we show that there is overwhelming support for the (data-driven) hybrid TVP-VAR relative to a few standard benchmarks, including a constant-coefficient VAR with stochastic volatility and a full-fledged TVP-VAR in which all the VAR coefficients and error covariances are time varying. We further illustrate the usefulness of the hybrid TVP-VAR with a forecasting exercise that involves 20 US quarterly macroeconomic and financial variables. We show that the proposed model forecasts better than many benchmarks. 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.

This paper contributes to the budding literature on developing large TVP-VARs. Earlier papers include Koop and Korobilis (2013, 2018), who propose fast methods to approximate the posterior distributions of large TVP-VARs. Banbura and van Vlodrop (2018) and Götz and Hauzenberger (2018) consider large VARs with only time-varying inter-

cepts. Chan, Eisenstat, and Strachan (2020) model the time-varying coefficients using a factor-like 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. As mentioned above, our estimation approach is exact and fully Bayesian, and the modeling framework is more flexible than many of those in earlier papers. There is also a growing literature on alternative, non-likelihood based approaches. Examples include Giraitis, Kapetanios, and Price (2013) and Petrova (2019) that allow for the estimation of large TVP-VARs without imposing the Cholesky-type stochastic volatility, and hence they avoid the ordering issue. Nevertheless, one main advantage of the likelihood-based approach taken in this paper is that it is flexible and modular. In particular, it is straightforward to incorporate additional useful features into the proposed hybrid model, such as more sophisticated static and dynamic shrinkage priors for VARs (Prüser, 2021; Chan, 2021) or more flexible error distributions to deal with outliers (Carriero, Clark, Marcellino, and Mertens, 2021; Bobeica and Hartwig, 2021).

The rest of the paper is organized as follows. We first introduce the proposed modeling 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 a Monte Carlo study in Section 4 that demonstrates that the proposed methodology works well and can select the correct time-varying or time-invariant structure. 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

We first introduce a class of models we call hybrid time-varying parameter VARs: VARs in which some equations have time-varying coefficients, whereas coefficients 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 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 $\boldsymbol{\Sigma}_t = \text{diag}(\exp(h_{1,t}), \ldots, \exp(h_{n,t}))$. The law of motion of the VAR coefficients and log-volatilites will be specified below. 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.

We note that 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. The main drawback of this representation, however, is that the implied reduced-form estimates depend on how the variables are ordered in the system. We will investigate the extent to which these estimates depend on the ordering in Section 5.2.

2.1 An Equation-by-Equation Representation

It is convenience to introduce some notations. Let $b_{i,t}$ denote the i-th element of $\mathbf{b}_t$ and let $\mathbf{B}_{i,j,t}$ represent 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 of the i-th equation and is of dimension $k_\beta \times 1$ with $k_\beta = np + 1$. Moreover, let $\boldsymbol{\alpha}_{i,t}$ denote the free elements in the i-th row of the contemporaneous impact matrix $\mathbf{A}_t$ for $i = 2, \ldots, n$. That is, $\boldsymbol{\alpha}_{i,t} = (A_{i1,t}, \ldots, A_{i(i-1),t})'$ is of dimension $k_{\alpha_i} \times 1$ with $k_{\alpha_i} = i - 1$. Then, the i-th equation of the system in (1) can be rewritten as:

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

where $\widetilde{\mathbf{x}}_t = (1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}')$ and $\widetilde{\mathbf{w}}_{i,t} = (-y_{1,t}, \ldots, -y_{i-1,t})$. 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.

If we let $\mathbf{x}_{i,t} = (\widetilde{\mathbf{x}}_t, \widetilde{\mathbf{w}}_{i,t})$, 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 $\boldsymbol{\theta}_{i,t} = (\boldsymbol{\beta}'_{i,t}, \boldsymbol{\alpha}'_{i,t})'$ is of dimension $k_{\theta_i} = k_\beta + k_{\alpha_i} = np + i$. 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:

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

$$\boldsymbol{\alpha}_{i,t} = \boldsymbol{\alpha}_{i,t-1} + \boldsymbol{\varepsilon}^\alpha_{i,t}, \qquad \boldsymbol{\varepsilon}^\alpha_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\alpha_i}), \tag{4}$$

$$h_{i,t} = h_{i,t-1} + \varepsilon^h_{i,t}, \qquad \varepsilon^h_{i,t} \sim \mathcal{N}(0, \sigma^2_{h,i}), \tag{5}$$

where the initial conditions $\boldsymbol{\beta}_{i,0}, \boldsymbol{\alpha}_{i,0}$ and $h_{i,0}$ are treated as unknown parameters to be estimated. The system in (2)–(5) specifies a reparameterization of a standard TVP-VAR in which all equations have time-varying parameters and stochastic volatility.

Note that the innovations in (3)-(5) are assumed to be independent across equations. This assumption is partly motivated by the concern of proliferation of correlation parameters, especially when n is large, if the correlations of the innovations are unrestricted. In addition, for $\boldsymbol{\beta}_{i,t}$ and $\boldsymbol{\alpha}_{i,t}$, this independence assumption is important for extending the setup later so that we can turn on and off the time variation in both equations. In contrast, it is feasible to allow the innovations to $h_{i,t}$ to be correlated across equations (with a slight increase of computational cost). In preliminary work we considered such an extension. While the estimation results suggest that the correlation parameters are sizable, this extension leads to only very modest forecast gains (see Appendix D for details). Therefore, in what follows we maintain the independence assumption in (3)-(5) as the baseline.

2.2 The Non-Centered Parameterization

Next, we introduce a framework that allows the model to determine in a data-driven fashion whether the VAR coefficients and the contemporaneous relations among the endogenous variables in each equation are time varying or constant. For that purpose, we adapt the non-centered parameterization of Frühwirth-Schnatter and Wagner (2010) to

our hybrid TVP-VARs. More specifically, for $i = 1, \ldots, n, t = 1, \ldots, T$, we consider the following model:

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\theta}_{i,0} + \gamma_i^{\beta}\widetilde{\mathbf{x}}_t\boldsymbol{\Sigma}_{\beta_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\beta}}_{i,t} + \gamma_i^{\alpha}\widetilde{\mathbf{w}}_{i,t}\boldsymbol{\Sigma}_{\alpha_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\alpha}}_{i,t} + \varepsilon_{i,t}^{y}, \qquad \varepsilon_{i,t}^{y} \sim \mathcal{N}(0, \mathrm{e}^{h_{i,t}}), \tag{6}$$

$$\widetilde{\boldsymbol{\beta}}_{i,t} = \widetilde{\boldsymbol{\beta}}_{i,t-1} + \widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\beta}, \qquad \widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\beta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_\beta}), \tag{7}$$

$$\widetilde{\boldsymbol{\alpha}}_{i,t} = \widetilde{\boldsymbol{\alpha}}_{i,t-1} + \widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\alpha}, \qquad \widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_{\alpha_i}}), \tag{8}$$

$$h_{i,t} = h_{i,t-1} + \varepsilon_{i,t}^{h}, \qquad \varepsilon_{i,t}^{h} \sim \mathcal{N}(0, \sigma_{h,i}^2), \tag{9}$$

where $\widetilde{\boldsymbol{\beta}}_{i,0} = \mathbf{0}$ and $\widetilde{\boldsymbol{\alpha}}_{i,0} = \mathbf{0}$. Here γ_i^{β} and γ_i^{α} are indicator variables that take values of either 0 or 1.

The model in (6)-(9) includes a wide variety of popular VAR specifications. For example, assuming that all indicators take the value of 1, the above model is just a reparameterization of the TVP-VAR in (2)–(5). To see that, define $\boldsymbol{\beta}_{i,t} = \boldsymbol{\beta}_{i,0} + \gamma_i^{\beta}\boldsymbol{\Sigma}_{\beta_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\beta}}_{i,t}$ and $\boldsymbol{\alpha}_{i,t} = \boldsymbol{\alpha}_{i,0} + \gamma_i^{\alpha}\boldsymbol{\Sigma}_{\alpha_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\alpha}}_{i,t}$. Then, when $\gamma_i^{\beta} = \gamma_i^{\alpha} = 1, i = 1, \ldots, n$, it is clear that (6) becomes (2). In addition, we have

$$\boldsymbol{\beta}_{i,t} - \boldsymbol{\beta}_{i,t-1} = \boldsymbol{\Sigma}_{\beta_i}^{\frac{1}{2}}(\widetilde{\boldsymbol{\beta}}_{i,t} - \widetilde{\boldsymbol{\beta}}_{i,t-1}) = \boldsymbol{\Sigma}_{\beta_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\beta},$$
$$\boldsymbol{\alpha}_{i,t} - \boldsymbol{\alpha}_{i,t-1} = \boldsymbol{\Sigma}_{\alpha_i}^{\frac{1}{2}}(\widetilde{\boldsymbol{\alpha}}_{i,t} - \widetilde{\boldsymbol{\alpha}}_{i,t-1}) = \boldsymbol{\Sigma}_{\alpha_i}^{\frac{1}{2}}\widetilde{\boldsymbol{\varepsilon}}_{i,t}^{\alpha}.$$

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

For the intermediate case where $\gamma_i^{\beta} = 1$ and $\gamma_i^{\alpha} = 0, i = 1, \ldots, n$, the proposed model reduces to a structural-form reparameterization of the model in Cogley and Sargent (2005), i.e., a TVP-VAR with stochastic volatility but the contemporaneous relations among the endogenous variables are restricted to be constant. In the extreme case where $\gamma_i^{\beta} = \gamma_i^{\alpha} = 0, i = 1, \ldots, n$, the proposed model then becomes a constant-coefficient VAR with stochastic volatility—a reparameterization of the specification in Carriero, Clark, and Marcellino (2019). More generally, by allowing the indicators γ_i^{β} and γ_i^{α} to take different values, we can have a VAR in which only some equations have time-varying parameters. Note that it is straightforward to include a few additional indicators to allow for more flexible forms of time variation. For example, one can replace γ_i^{β} with two indicators, say, $\gamma_i^{\beta,\text{own}}$ and $\gamma_i^{\beta,\text{other}}$, which control the time variation in the elements of

$\boldsymbol{\beta}_{i,t}$ that correspond to coefficients on own lags and lags of other variables, respectively. The posterior simulator in Section 3.2 can be modified to handle this case, with a slight increase in computation time.

These indicators are not fixed but are estimated from the data. More precisely, we specify that each γ_i^β follows an independent Bernoulli distribution with success probability $\mathbb{P}(\gamma_i^\beta = 1) = p_i^\beta, i = 1, \ldots, n$. Similarly for γ_i^α: $\mathbb{P}(\gamma_i^\alpha = 1) = p_i^\alpha$. These success probabilities p_i^β and $p_i^\alpha, i = 1, \ldots, n$, 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 and the contemporaneous relations among the variables are constant. 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.

2.3 An Exploration of the Model Space

The proposed hybrid TVP-VAR can also be viewed as a Bayesian model average of a wide variety TVP-VARs with different forms of time variation. To see that, let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n)'$ denote the vector of indicator variables with $\boldsymbol{\gamma}_i = (\gamma_i^\beta, \gamma_i^\alpha)$. Note that each value of $\boldsymbol{\gamma} \in \{0, 1\}^{2n}$ corresponds to a particular TVP-VAR in which the time variation of the i-th equation is characterized by $\boldsymbol{\gamma}_i$. For example, $\boldsymbol{\gamma} = \mathbf{0}$ corresponds to a constant-coefficient VAR with stochastic volatility. Then, the posterior distribution of any model parameters under the proposed model can be represented as the posterior average with respect to $p(\boldsymbol{\gamma} \,|\, \mathbf{y})$, i.e., the posterior model probabilities of the collection of 2^{2n} TVP-VARs with different forms of time variation, where $\mathbf{y}$ denotes the data. For example, the joint distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the time-varying VAR coefficients and free elements of the contemporaneous impact matrix, can be represented as

$$p(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, \mathbf{y}) = \sum_{\mathbf{c} \in \{0,1\}^{2n}} p(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, \mathbf{y}, \boldsymbol{\gamma} = \mathbf{c}) p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y}).$$

For a small VAR with $n = 3$ variables (and the additional assumption that $\gamma_i^\beta = \gamma_i^\alpha$), Chan and Eisenstat (2018b) estimate all $2^3 = 8$ TVP-VARs and the corresponding posterior model probabilities. For larger n, this approach of computing $p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y})$ and sampling from $p(\boldsymbol{\beta}, \boldsymbol{\alpha} \,|\, \mathbf{y}, \boldsymbol{\gamma} = \mathbf{c})$ for all 2^{2n} possible models is clearly infeasible. In contrast, by including the model indicator $\boldsymbol{\gamma}$ in the estimation, we simultaneously explore the parameter space and the model space. This latter approach is convenient and computationally feasible for large systems.

It is also instructive to investigate how the value of the model indicator $\boldsymbol{\gamma}$ is determined by the data. To fix ideas, suppose we wish to compare two TVP-VARs, represented as $\boldsymbol{\gamma} = \mathbf{c}_1$ and $\boldsymbol{\gamma} = \mathbf{c}_2$. Let $p(\mathbf{y} \,|\, \boldsymbol{\gamma} = \mathbf{c}_j)$ denote the marginal likelihood under model $\boldsymbol{\gamma} = \mathbf{c}_j, j = 1, 2$, i.e.,

$$p(\mathbf{y} \,|\, \boldsymbol{\gamma} = \mathbf{c}_j) = \int p(\mathbf{y} \,|\, \boldsymbol{\psi}_j, \boldsymbol{\gamma} = \mathbf{c}_j) p(\boldsymbol{\psi}_j \,|\, \boldsymbol{\gamma} = \mathbf{c}_j) \mathrm{d}\boldsymbol{\psi}_j, \tag{10}$$

where $\boldsymbol{\psi}_j$ is the collection of model-specific time-invariant parameters and time-varying states (in our setting these parameters and states are common across models and $\boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$), $p(\mathbf{y} \,|\, \boldsymbol{\psi}_j, \boldsymbol{\gamma} = \mathbf{c}_j)$ is the (complete-data) likelihood and $p(\boldsymbol{\psi}_j \,|\, \boldsymbol{\gamma} = \mathbf{c}_j)$ is the prior density. Then, the posterior odds ratio in favor of model $\boldsymbol{\gamma} = \mathbf{c}_1$ against model $\boldsymbol{\gamma} = \mathbf{c}_2$ is given by:

$$\frac{p(\boldsymbol{\gamma} = \mathbf{c}_1 \,|\, \mathbf{y})}{p(\boldsymbol{\gamma} = \mathbf{c}_2 \,|\, \mathbf{y})} = \frac{p(\boldsymbol{\gamma} = \mathbf{c}_1)}{p(\boldsymbol{\gamma} = \mathbf{c}_2)} \times \frac{p(\mathbf{y} \,|\, \boldsymbol{\gamma} = \mathbf{c}_1)}{p(\mathbf{y} \,|\, \boldsymbol{\gamma} = \mathbf{c}_2)},$$

where $p(\boldsymbol{\gamma} = \mathbf{c}_1)/p(\boldsymbol{\gamma} = \mathbf{c}_2)$ is the prior odds ratio. It follows that if both models are equally probable *a priori*, i.e., $p(\boldsymbol{\gamma} = \mathbf{c}_1) = p(\boldsymbol{\gamma} = \mathbf{c}_2)$, the posterior odds ratio between the two models is then equal to the ratio of the two marginal likelihoods, or the Bayes factor. More generally, under the assumption that each TVP-VAR has the same prior probability, the value of the model indicator $\boldsymbol{\gamma}$ is determined by the marginal likelihood $p(\mathbf{y} \,|\, \boldsymbol{\gamma})$. That is, if the TVP-VAR represented by $\boldsymbol{\gamma} = \mathbf{c}$ forecasts the data better (as one-step-ahead density forecasts), the value $\mathbf{c}$ will have a higher weight.

3 Priors and Bayesian Estimation

In this section we first describe in detail the priors on the time-invariant parameters. We then outline the posterior simulator to estimate the model described in (6)–(9)

3.1 Priors

For notational convenience, stack $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i,1}', \ldots, \boldsymbol{\beta}_{i,T}')'$, $\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_{i,1}', \ldots, \boldsymbol{\alpha}_{i,T}')'$ and $\mathbf{h}_i = (h_{i,1}, \ldots, h_{i,T})'$ over $t = 1, \ldots, T$, and collect $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^n$, $\boldsymbol{\beta} = \{\boldsymbol{\beta}_i\}_{i=1}^n$, $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_i\}_{i=1}^n$ and $\mathbf{h} = \{\mathbf{h}_i\}_{i=1}^n$ over $i = 1, \ldots, n$, and similarly define $\widetilde{\boldsymbol{\beta}}_i$ and $\widetilde{\boldsymbol{\alpha}}_i$. Furthermore, let $\boldsymbol{\Sigma}_{\theta_i} = \text{diag}(\boldsymbol{\Sigma}_{\beta_i}, \boldsymbol{\Sigma}_{\alpha_i})$ and $\boldsymbol{\gamma}_i = (\gamma_i^\beta, \gamma_i^\alpha)$. In our model, the time-invariant parameters are $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n)'$, $\boldsymbol{\Sigma}_\theta = \{\boldsymbol{\Sigma}_{\theta_i}\}_{i=1}^n$, $\boldsymbol{\Sigma}_h = \{\sigma_{i,h}^2\}_{i=1}^n$, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{1,0}', \ldots, \boldsymbol{\theta}_{n,0}')'$, $\mathbf{h}_0 = (h_{1,0}, \ldots, h_{n,0})'$, $\mathbf{p}^\beta = (p_1^\beta, \ldots, p_n^\beta)'$ and $\mathbf{p}^\alpha = (p_1^\alpha, \ldots, p_n^\alpha)'$. Below we give the details of the priors on these time-invariant parameters.

Since $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'$, the initial conditions of 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); see also Doan, Litterman, and Sims (1984), Litterman (1986) and Kadiyala and Karlsson (1997). We refer the readers to Koop and Korobilis (2010), Del Negro and Schorfheide (2012) and Karlsson (2013) for a textbook discussion of the Minnesota prior. 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 when the variables are in growth rate 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 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. In fact, Carriero, Clark, and Marcellino (2015) and Chan (2021) find 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 papers, 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. In addition, 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 (2020).

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 $\mathbf{\Sigma}_{\theta_i} = \text{diag}(\sigma^2_{\theta_i,1}, \ldots, \sigma^2_{\theta_i,k_{\theta_i}}), i = 1, \ldots, n$, are independently distributed as mean 0 normal random variables: $\sigma_{\theta_i,j} \sim \mathcal{N}(0, S_{\theta_i,j}), i = 1, \ldots, n, j = 1, \ldots, k_{\theta_i}$. We assume each $\sigma^2_{h,i}$ follow a conventional inverse-gamma priors: $\sigma^2_{h,i} \sim \mathcal{IG}(\nu_{h,i}, S_{h,i}), i = 1, \ldots, n$. The success probabilities p^β_i and p^α_i are assumed to have beta distributions: $p^\beta_i \sim \mathcal{B}(a_{p^\beta}, b_{p^\beta})$ and $p^\alpha_i \sim \mathcal{B}(a_{p^\alpha}, b_{p^\alpha}), i = 1, \ldots, n$. Finally, the elements of the initial condition $\mathbf{h}_0$ are assumed to be Gaussian: $h_{i,0} \sim \mathcal{N}(a_{h_{i,0}}, V_{h_{i,0}})$.

3.2 The Posterior Simulator

We now turn to the estimation of the model in (6)–(9) given the prior described in the previous section. There are a few challenges in the estimation. First, since $\boldsymbol{\beta}_i$ becomes degenerate when $\gamma^\beta_i = 0$, making its sampling nonstandard (similarly for $\boldsymbol{\alpha}_i$). To sidestep this problem, we will use the parameterization in terms of $\widetilde{\boldsymbol{\beta}}_i$ and $\widetilde{\boldsymbol{\alpha}}_i$. Then, given the posterior draws of $\widetilde{\boldsymbol{\beta}}_i$, $\widetilde{\boldsymbol{\alpha}}_i$ and other parameters, we can recover the posterior draws of $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ using the definitions $\boldsymbol{\beta}_{i,t} = \boldsymbol{\beta}_{i,0} + \gamma^\beta_i \mathbf{\Sigma}^{\frac{1}{2}}_{\beta_i} \widetilde{\boldsymbol{\beta}}_{i,t}$ and $\boldsymbol{\alpha}_{i,t} = \boldsymbol{\alpha}_{i,0} + \gamma^\alpha_i \mathbf{\Sigma}^{\frac{1}{2}}_{\alpha_i} \widetilde{\boldsymbol{\alpha}}_{i,t}$.

Second, since $\widetilde{\boldsymbol{\beta}}_i$ and the indicator γ_i^β enter the likelihood in (6) multiplicatively, it is vital to sample them jointly (similarly for $\widetilde{\boldsymbol{\alpha}}_i$ and γ_i^α); otherwise the Markov chain might get stuck. To see this, consider a simpler sampling scheme in which we simulate $\widetilde{\boldsymbol{\beta}}_i$ given γ_i^β, followed by sampling γ_i^β given $\widetilde{\boldsymbol{\beta}}_i$. Suppose $\gamma_i^\beta = 0$ in the last iteration. Given $\gamma_i^\beta = 0$, $\widetilde{\boldsymbol{\beta}}_i$ does not enter the likelihood and we simply sample it from its state equation. Since the sampled $\widetilde{\boldsymbol{\beta}}_i$ 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., $\gamma_i^\beta = 0$. Hence, it is unlikely for the Markov chain to move away from $\gamma_i^\beta = 0$ once it is there. It is therefore necessary to sample both $\widetilde{\boldsymbol{\beta}}_i$ and γ_i^β in the same step. In addition, since the pair $(\widetilde{\boldsymbol{\beta}}_i, \gamma_i^\beta)$ and $(\widetilde{\boldsymbol{\alpha}}_i, \gamma_i^\alpha)$ enters the likelihood additively, we sample them jointly to further improve efficiency.

Next, define $\widetilde{\boldsymbol{\theta}}_i = (\widetilde{\boldsymbol{\theta}}_{i,1}', \ldots, \widetilde{\boldsymbol{\theta}}_{i,T}')'$ with $\widetilde{\boldsymbol{\theta}}_{i,t} = (\widetilde{\boldsymbol{\beta}}_{i,t}', \widetilde{\boldsymbol{\alpha}}_{i,t}')'$. Then, one can simulate from the joint posterior distribution using the following posterior sampler that sequentially samples from:

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

Step 2 to Step 7 mainly involve standard sampling techniques and we leave the details to Appendix A. Here we focus on the first step.

Step 1. We sample the four blocks of parameters $\widetilde{\boldsymbol{\beta}}_i, \gamma_i^\beta, \widetilde{\boldsymbol{\alpha}}_i$ and γ_i^α jointly to improve efficiency. This is done by first drawing the indicators $\boldsymbol{\gamma}_i = (\gamma_i^\beta, \gamma_i^\alpha)$ marginally of $\widetilde{\boldsymbol{\theta}}_{i,t} = (\widetilde{\boldsymbol{\beta}}_{i,t}', \widetilde{\boldsymbol{\alpha}}_{i,t}')'$—but conditional on other parameters—and then sample $\widetilde{\boldsymbol{\beta}}_i$ and $\widetilde{\boldsymbol{\alpha}}_i$ from their joint conditional distribution. The latter of these two steps is straightforward because

given γ_i^β and γ_i^α, we have a linear Gaussian state space model in $\boldsymbol{\theta}_{i,t}$. Specifically, we stack the observation equation (6) over $t = 1, \ldots, T$:

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

where $\boldsymbol{\Sigma}_{\mathbf{h}_i} = \mathrm{diag}(\mathrm{e}^{h_{i,1}}, \ldots, \mathrm{e}^{h_{i,T}})$,

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i,1} \\ \vdots \\ \mathbf{x}_{i,T} \end{pmatrix}, \quad \mathbf{Z}_{\boldsymbol{\gamma}_i} = \begin{pmatrix} (\gamma_i^\beta \widetilde{\mathbf{x}}_1, \gamma_i^\alpha \widetilde{\mathbf{w}}_{i,1}) \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\gamma_i^\beta \widetilde{\mathbf{x}}_2, \gamma_i^\alpha \widetilde{\mathbf{w}}_{i,2}) \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\gamma_i^\beta \widetilde{\mathbf{x}}_T, \gamma_i^\alpha \widetilde{\mathbf{w}}_{i,T}) \boldsymbol{\Sigma}_{\theta_i}^{\frac{1}{2}} \end{pmatrix}.$$

Here note that the matrix $\mathbf{Z}_{\boldsymbol{\gamma}_i}$ depends on the indicators $\boldsymbol{\gamma}_i = (\gamma_i^\beta, \gamma_i^\alpha)$. Next, stack the state equations (7)-(8) 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 $k_{\theta_i} = k_\beta + k_{\alpha_i}$. 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

$$(\widetilde{\boldsymbol{\theta}}_i \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \boldsymbol{\gamma}_i) \sim \mathcal{N}\left(\widehat{\widetilde{\boldsymbol{\theta}}}_i, \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}^{-1}\right), \tag{11}$$

where

$$\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i} = \mathbf{H}'_{k_{\theta_i}} \mathbf{H}_{k_{\theta_i}} + \mathbf{Z}'_{\boldsymbol{\gamma}_i} \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} \mathbf{Z}_{\boldsymbol{\gamma}_i}, \qquad \widehat{\widetilde{\boldsymbol{\theta}}}_i = \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}^{-1} \left(\mathbf{Z}'_{\boldsymbol{\gamma}_i} \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0}) \right). \tag{12}$$

Since the precision matrix $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ is a band matrix, one can sample $(\widetilde{\boldsymbol{\theta}}_i \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \boldsymbol{\gamma}_i)$ efficiently using the algorithm in Chan and Jeliazkov (2009). It is worth noting that one could include an additional step to accept or reject the draw $\widetilde{\boldsymbol{\theta}}_i$ to ensure stationarity by checking the roots of the characteristic polynomial associated with the implied reduced-form VAR coefficients along the lines suggested in Cogley and Sargent (2005).

To sample $\boldsymbol{\gamma}_i = (\gamma_i^\beta, \gamma_i^\alpha)$ marginal of $\widetilde{\boldsymbol{\theta}}_i$, it suffices to compute the four probabilities that

$\boldsymbol{\gamma}_i = (0,0), \boldsymbol{\gamma}_i = (0,1), \boldsymbol{\gamma}_i = (1,0)$, and $\boldsymbol{\gamma}_i = (1,1)$. To that end, note that

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

where both the conditional likelihood $p(\mathbf{y}_i \,|\, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \boldsymbol{\gamma}_i)$ 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

$$\begin{aligned} &\int_{\mathbb{R}^{Tk_{\theta_i}}} p(\mathbf{y}_i \,|\, \widetilde{\boldsymbol{\theta}}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}, \boldsymbol{\gamma}_i) p(\widetilde{\boldsymbol{\theta}}_i) \mathrm{d}\widetilde{\boldsymbol{\theta}}_i \\ &\qquad = (2\pi)^{-\frac{T}{2}} |\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^{T} h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0})' \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0}) - \widehat{\widetilde{\boldsymbol{\theta}}}_i' \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i} \widehat{\widetilde{\boldsymbol{\theta}}}_i \right)}, \end{aligned} \tag{13}$$

where $\widehat{\widetilde{\boldsymbol{\theta}}}_i$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ are defined in (12). Then, one can compute the relevant probabilities using the expression in (13). For example, when $\boldsymbol{\gamma}_i = (0,0)$, $|\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}| = 1$ and $\widehat{\widetilde{\boldsymbol{\theta}}}_i = \mathbf{0}$. It follows that

$$\begin{aligned} &\mathbb{P}(\boldsymbol{\gamma}_i = (0,0) \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) \\ &\qquad \propto (1 - p_i^{\beta})(1 - p_i^{\alpha})(2\pi)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^{T} h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0})' \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0}) \right)}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} &\mathbb{P}(\boldsymbol{\gamma}_i = (1,1) \,|\, \mathbf{y}_i, \mathbf{h}_i, \boldsymbol{\Sigma}_{\theta_i}, \boldsymbol{\theta}_{i,0}) \\ &\quad \propto p_i^{\beta} p_i^{\alpha} (2\pi)^{-\frac{T}{2}} |\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1,1)|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{t=1}^{T} h_{i,t} + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0})' \boldsymbol{\Sigma}_{\mathbf{h}_i}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_{i,0}) - \widehat{\widetilde{\boldsymbol{\theta}}}_i(1,1)' \mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1,1) \widehat{\widetilde{\boldsymbol{\theta}}}_i(1,1) \right)}, \end{aligned}$$

where $\widehat{\widetilde{\boldsymbol{\theta}}}_i(1,1)$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}(1,1)$ denote respectively $\widehat{\widetilde{\boldsymbol{\theta}}}_i$ and $\mathbf{K}_{\widetilde{\boldsymbol{\theta}}_i}$ evaluated at $\boldsymbol{\gamma}_i = (1,1)$. The probabilities that $\boldsymbol{\gamma}_i = (0,1)$ and $\boldsymbol{\gamma}_i = (1,0)$ can be computed similarly. A draw from this 4-point distribution is standard once we normalize the probabilities. The details of the remaining steps are provided in Appendix A.

4 A Monte Carlo Study

In this section we first conduct a series of simulated experiments to assess how well the posterior sampler works in recovering the time-varying structure in the data generating process. We then document the runtimes of estimating the hybrid TVP-VARs of different dimensions to assess how well the posterior sampler scales to larger systems.

First, we generate 300 datasets from the hybrid VAR in (6)–(9) with $n = 12$ variables and sample size $T = 200$, $T = 400$ or $T = 800$. We set the vector of indicators $\boldsymbol{\gamma}$ by repeating the four combinations $(0,0), (0,1), (1,0), (1,1)$ three times — that allows us to study the effect of different combinations of time-varying pattens as well as their positions in the system. We generate $\boldsymbol{\beta}_0$, the initial conditions of the VAR coefficients, stochastically as follows. The intercepts are drawn independently from the uniform distribution on the interval $(-10, 10)$, i.e., $\mathcal{U}(-10, 10)$. For the VAR coefficients, the diagonal elements of the first VAR coefficient matrix are iid $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are from $\mathcal{U}(-0.2, 0.2)$. All other elements of the j-th ($j > 1$) VAR coefficient matrices are iid $\mathcal{N}(0, 0.1^2/j^2)$. Finally, the elements of $\boldsymbol{\alpha}_0$ are drawn independently from $\mathcal{U}(-0.5, 0.5)$.

If the coefficient $\theta_{ij,t}$ is time-varying (i.e., the associated indicator γ_i^{α} or γ_i^{β} is 1), it is generated from the state equation (3) or (4) with $\sigma^2_{\theta_i,j} = 0.01^2$ if $\theta_{ij,t}$ is a VAR coefficient and $\sigma^2_{\theta_i,j} = 0.1^2$ if it is an intercept for $i = 1, \ldots, n, j = 1, \ldots, k_{\theta_i}$. Finally, for the log-volatility processes, we draw $h_{0,i} \sim \mathcal{U}(-2, 2)$ and set $\sigma_h^2 = 0.1, i = 1, \ldots, n$.

In the Monte Carlo study we use the priors described in Section 3.1 with the following hyperparameters. The prior means of the initial conditions $\boldsymbol{\theta}_0$ and $\mathbf{h}_0$ are set to be zero $\mathbf{a}_{\boldsymbol{\theta}_0} = \mathbf{0}$ and $\mathbf{a}_h = \mathbf{0}$, and the prior covariance matrix of $\mathbf{h}_0$ is $\mathbf{V}_h = 10 \times \mathbf{I}_n$. The hyperparameter of $\sigma_{\theta_i,j}$ is set so that the implied prior mean of $\sigma_{\theta_i,j}$ is 0.01^2 if it is associated with a VAR coefficient and 0.1^2 for an intercept. Finally, we set the hyperparameters of γ_i^{β} and γ_i^{α} to be $a_{p^{\beta}} = b_{p^{\beta}} = a_{p^{\alpha}} = b_{p^{\alpha}} = 0.5$. These values imply that the prior modes are at 0 and 1, whereas the prior mean is 0.5.

Given a dataset and the priors described above, we estimate the hybrid VAR using the posterior sampler in Section 3.2 and obtain the posterior mode of $\boldsymbol{\gamma}$. We repeat this procedure for all the datasets and compute the frequencies of γ_i^{β} and γ_i^{α} being one, $i = 1, \ldots, n$. The results are reported in Table 1.

Overall, the posterior sampler works well and is able to recover the true time-varying structure in the simulated data on average. While it is harder to pin down the correct value of γ_i^α compared to γ_i^β, the frequencies of identifying the true value of γ_i^α are still reasonably good, even for a small sample of $T = 200$. In addition, these results substantially improve when the sample size increases from $T = 200$ to $T = 800$. All in all, these Monte Carlo results confirm that the proposed hybrid model can recover salient patterns—such as time-varying conditional means and covariances—in the data.

To further investigate the effect of the beta prior on γ_i^β and γ_i^α, we repeat the Monte Carlo experiments but assume a uniform prior on the unit interval $(0, 1)$, i.e., $\gamma_i^\beta, \gamma_i^\alpha \sim \mathcal{B}(1, 1) = \mathcal{U}(0, 1)$. Hence, both the prior means and modes are 0.5. The Monte Carlo results are similar to the baseline case and they are reported in Appendix D.

Table 1: Frequencies (%) of the posterior modes of γ_i^β and γ_i^α being one in 300 datasets.

Equation	True γ_i^β	True γ_i^α	$T = 200$		$T = 400$		$T = 800$	
			γ_i^β	γ_i^α	γ_i^β	γ_i^α	γ_i^β	γ_i^α
1	0	0	0.05	–	0.06	–	0.05	–
2	0	1	0.02	0.73	0.04	0.88	0.05	0.93
3	1	0	0.92	0.40	0.98	0.25	1.00	0.12
4	1	1	0.94	0.58	0.98	0.64	1.00	0.75
5	0	0	0.01	0.08	0.02	0.02	0.02	0.00
6	0	1	0.01	0.93	0.03	0.96	0.04	0.98
7	1	0	0.92	0.34	0.97	0.13	1.00	0.03
8	1	1	0.91	0.68	0.95	0.80	1.00	0.94
9	0	0	0.01	0.03	0.03	0.00	0.02	0.01
10	0	1	0.06	0.90	0.04	0.94	0.10	0.99
11	1	0	0.89	0.29	0.94	0.11	1.00	0.02
12	1	1	0.92	0.77	0.93	0.88	0.99	0.96

Next, we document the runtimes of estimating the hybrid TVP-VARs of different sizes to assess how well the posterior sampler scales to higher dimensions. More specifically, Table 2 reports the runtimes (in minutes) to obtain 1,000 posterior draws from the hybrid models with $n = 10, 20, 30$ variables and $T = 400, 800$ time periods. The posterior sampler is implemented in MATLAB on a standard desktop with an Intel Core i7-7700 @3.60 GHz processor and 64 GB memory. As a comparison, we also include the corresponding runtimes of fitting the TVP-VAR of Primiceri (2005) using the algorithm in

Del Negro and Primiceri (2015). Note that the algorithm in Del Negro and Primiceri (2015) samples all the time-varying VAR coefficients $\boldsymbol{\beta}$ in one block and it tends to be very computationally intensive for larger systems. One potential solution is to develop an equation-by-equation estimation procedure similar to that in Carriero, Chan, Clark, and Marcellino (2021). Since the algorithm is designed for models with a constant contemporaneous impact matrix, extending it to handle the TVP-VAR of Primiceri (2005)—which features a time-varying contemporaneous impact matrix—would be an interesting future research direction.

Table 2: The runtimes (in minutes) to obtain 1,000 posterior draws from the hybrid TVP-VAR with n variables and T time periods. All VARs have $p = 2$ lags.

	$T = 400$			$T = 800$		
	$n = 10$	$n = 20$	$n = 30$	$n = 10$	$n = 20$	$n = 30$
Hybrid TVP-VAR	4	29	94	8	59	188
Primiceri (2005)	12	209	–	25	415	–

It is evident from the table that for typical applications with 15-30 variables, the proposed model can be estimated reasonably quickly. In addition, using the recursive representation that admits straightforward equation-by-equation estimation, fitting the proposed model is much faster than estimating the TVP-VAR of Primiceri (2005), even though the former is more flexible.

5 Application: Model Comparison 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 investigate how different variable orderings affect the estimates from the proposed hybrid TVP-VAR relative to the TVP-VAR of Primiceri (2005) in Section 5.2. We then present the full sample results in Section 5.3. In particular, we conduct a formal Bayesian model comparison exercise to shed light on the time-varying patterns of the model parameters. We then consider a pseudo out-of-sample forecasting exercise in Section 5.4. 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 (2021). 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 C.

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 hyperparameters of $\sigma^2_{h,i}$ are set so that the prior mean is 0.1. Similarly, the hyperparameters of $\sigma_{\theta_i,j}$ are chosen so that the implied prior mean is 0.01^2 if it is associated with a VAR coefficient and 0.1^2 for an intercept. Finally, we set $a_{p^\theta} = b_{p^\theta} = a_{p^h} = b_{p^h} = 0.5$. These values imply prior modes at 0 and 1, whereas the prior mean is 0.5.

5.2 The Role of Variable Ordering

Since the proposed hybrid TVP-VAR is written in the recursive structural form—which is used as a computational device and not as an identification scheme—one naturally wonders how the assumed order of the variables affects the model estimates compared to standard TVP-VARs such as the model in Primiceri (2005). Conceptually, the proposed hybrid TVP-VAR is not order invariant due to two components. First, the VAR coefficients are in structural form, and their priors induce priors on the reduced-form parameters that depend on the order of the variables. Second, the multivariate stochastic volatility specification is constructed based on the lower triangular impact matrix $\mathbf{A}_t$. And since priors are independently elicited for $\mathbf{A}_t$ and the stochastic volatility, the implied prior on the covariance matrix is not order invariant.

Popular TVP-VARs such as Cogley and Sargent (2005) and Primiceri (2005) share the second component but not the first (since they are formulated as reduced-form VARs); see, e.g., the discussion in Primiceri (2005) and Carriero, Clark, and Marcellino (2019). Hence, one might expect that estimates from the proposed hybrid TVP-VAR would be more sensitive to how the variables are ordered compared to those of Primiceri (2005). On the other hand, as discussed in Section 2.3, the proposed hybrid TVP-VAR can be viewed as a Bayesian model average of a wide variety of VARs with many different forms of time variation. Since some of these VARs are more parsimonious and have restricted time variation—including the model of Carriero, Clark, and Marcellino (2019) that is less prone to the ordering issue—the resulting Bayesian model average estimates could in principle be more robust to different orderings compared to Primiceri (2005). Hence, whether the ordering issue is more severe in the proposed hybrid TVP-VAR relative to Primiceri (2005) is an empirical question, and we investigate this issue below.

It is worth noting that we choose the TVP-VAR of Primiceri (2005) to be our benchmark, even though it is not order invariant, because it is generally viewed as the state-of-the-art and it is widely used as a reduced-form VAR for both forecasting (e.g., D'Agostino, Gambetti, and Giannone, 2013) and structural analysis using non-recursive identification schemes (e.g., Benati, 2008; Baumeister and Peersman, 2013). There are a few recent papers that aim to develop order-invariant VARs with multivariate stochastic volatility, such as Bognanni (2018), Shin and Zhong (2020), Arias, Rubio-Ramirez, and Shin (2021) and Chan, Koop, and Yu (2021). These models, however, are either designed for small TVP-VARs or they do not feature time-varying VAR coefficients.

Now, we investigate how different variable orderings impact the estimates from the proposed hybrid TVP-VAR relative to the TVP-VAR of Primiceri (2005). To that end, we use 6 variables—real GDP, PCE inflation, unemployment, Fed funds rate and industrial production and real average hourly earnings in manufacturing—and we consider all $6! = 720$ possible orderings. For each ordering of these 6 variables, we fit the proposed model and obtain the fitted values (i.e., the time-varying conditional means). We then compute the mean squared errors (against the observed values) of the 6 variables. We repeat this exercise for the TVP-VAR of Primiceri (2005). Since there are 720 MSEs for each variable and each model, to better summarize the results we report the boxplots of the MSEs in Figure 1. The middle line of each box denotes the median, while the lower and upper lines represent, respectively, the 25- and the 75-percentiles. The whiskers

extend to the maximum and minimum.

Since the goal of this exercise is to compare the variability of the fitted means, we normalize the MSEs by the medians (so that the red line of each boxplot is one). Overall, the variability of the estimates from the proposed hybrid TVP-VAR is comparable to that of the model of Primiceri (2005). It is also interesting to note that for the majority of the variables the variability is relatively small. One exception is the unemployment rate—this is partly due to the very small base rate (e.g., the MSE of the unemployment is less than 1% of the MSE of the real GDP).

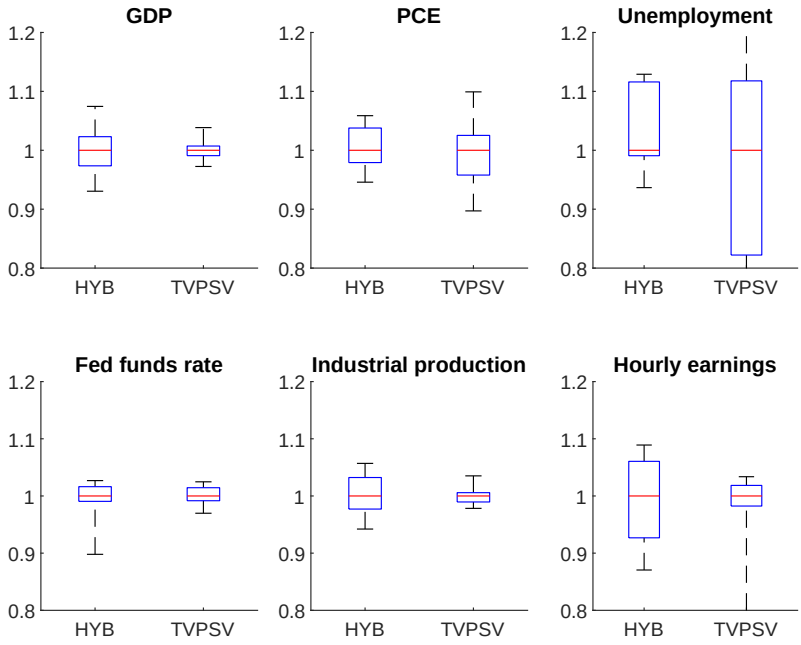

Figure 1: Boxplots of the relative mean squared errors of the fitted means for the proposed hybrid TVP-VAR (HYB) and the model of Primiceri (2005) (TVPSV).

Next, we investigate the variability of the variance estimates due to different orderings. For that purpose, we obtain the fitted variance of each variable in each ordering, and compute the mean squared error against the 'realized volatility' values (defined as $\text{RV}_{i,t} = (y_{i,t} - \widehat{y}_{i,t})^2$, where $\widehat{y}_{i,t}$ is the fitted value from the regression of $y_{i,t}$ on an intercept and $y_{i,t-1}, \ldots, y_{i,t-4}$). The results are reported in Figure 2. Again, the results show that the variability of the variance estimates from the proposed model is comparable to that of the

TVP-VAR of Primiceri (2005), even though the former uses a recursive structural-form representation.

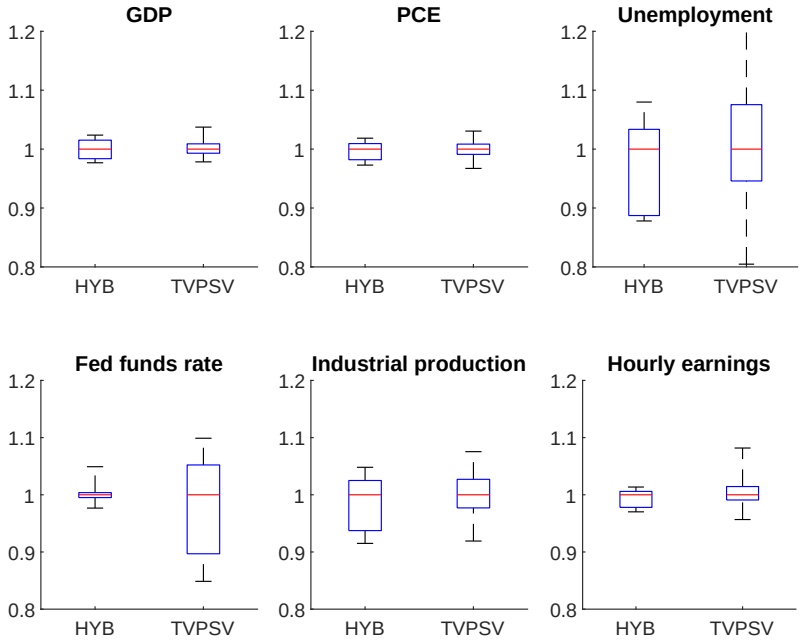

Figure 2: Boxplots of the relative mean squared errors of the estimated variances for the proposed hybrid TVP-VAR (HYB) and the model of Primiceri (2005) (TVPSV).

We also investigate the variability of point and density forecasts similar to the exercise in Arias, Rubio-Ramirez, and Shin (2021). More specifically, for each variable ordering, we compute the root mean squared forecast error and the average of log predictive likelihoods (see Section 5.4 for more details) for each of the 6 variables from the proposed model. Consistent with the results in Arias, Rubio-Ramirez, and Shin (2021), the variability of point forecasts is relatively small. In addition, the variability of the forecasts from the proposed model is similar to a full-fledged TVP-VAR where all the VAR coefficients and elements of the impact matrix are time varying. The details are reported in Appendix D.

5.3 Full Sample Results

In this section we report the full sample results of the hybrid TVP-VAR fitted using all $n = 20$ variables, which are ordered as listed in Table 3. Of particular interest are the posterior means of γ_i^β and γ_i^α, the indicators that control the time variation in the VAR coefficients and the elements of the impact matrix, respectively. These estimates are reported in Table 3. The results clearly show that while many estimates are essentially 0, others are close to 1. In other words, while the time variation in many equations is essentially turned off, there is strong evidence for time-varying parameters in some equations. Since there is substantial heterogeneity in the time-variation pattern across equations, conventional approaches of either assuming time variation in all equations or restricting all parameters to be constant are unlikely to fit the time-variation pattern well.

Table 3: Posterior means of γ_i^β and γ_i^α for the hybrid TVP-VAR with $n = 20$ variables.

Equation	Mnemonic	γ_i^β	γ_i^α
Real GDP	GDPC1	0.97	–
PCE inflation	PCECTPI	0.72	0.41
Unemployment	UNRATE	0	0.33
Fed funds rate	FEDFUNDS	0	1
Industrial production index	INDPRO	1	0.46
Real average hourly earnings in manufacturing	CES3000000008x	0.02	1
M1	M1REAL	0.99	0.89
Real PCE	PCECC96	0.01	0.97
Real disposable personal income	DPIC96	0	0.49
Industrial production: final products	IPFINAL	0	0.24
All employees: total nonfarm	PAYEMS	0	0.07
Civilian employment	CE16OV	0	0.95
Nonfarm business section: hours of all persons	HOANBS	0	0.35
GDP deflator	GDPCTPI	0	0
CPI	CPIAUCSL	0	1
PPI	PPIACO	0.47	1
Nonfarm business sector: real compensation per hour	COMPRNFB	0	1
Nonfarm business section: real output per hour	OPHNFB	0	0
10-year treasury constant maturity rate	GS10	0	0
M2	M2REAL	1	0.94

In addition, the estimates of γ_i^β and γ_i^α for the same equation are often of very different

magnitudes, suggesting that time variation in one group of parameters does not necessarily imply time variation in the other. These results thus confirm the usefulness of having two separate indicators for each equation. Overall, while we find evidence for time variation in VAR coefficients and elements of the impact matrix in some equations, not all equations need both forms of time variation. These results therefore highlight the empirical relevance of the proposed hybrid TVP-VAR.

To assess the efficiency of the proposed posterior sampler, we compute the inefficiency factors of the posterior draws, which are reported in Appendix D. The values of the inefficiency factors are comparable to those of conventional TVP-VARs. The results thus show that the posterior sampler is efficient in terms of producing posterior draws that are not highly autocorrelated.

Next, we consider a Bayesian model comparison exercise to compare the proposed hybrid TVP-VAR to various VARs with different time-variation patterns. In general, to compare models using the Bayes factor, one would require the computation of the marginal likelihood given in (10). Despite recent advances, obtaining the marginal likelihood for high-dimensional models with multiple latent states remains a nontrivial task. Fortunately, one much simpler approach is available when one wishes to compare nested models. More specifically, for nested models, the Bayes factor can be calculated using the Savage-Dickey density ratio (Verdinelli and Wasserman, 1995), which requires only the estimation of the unrestricted model. More importantly, no explicit computation of the marginal likelihood is needed. This approach has been used to compute the Bayes factor in many empirical applications, including Koop and Potter (1999), Deborah and Strachan (2009) and Koop, Leon-Gonzalez, and Strachan (2010).

More specifically, suppose we wish to compare the proposed hybrid TVP-VAR against a TVP-VAR characterized by the vector of indicators $\boldsymbol{\gamma} = \mathbf{c} \in \{0, 1\}^{2n}$ (e.g., $\mathbf{c} = \mathbf{0}$ represents a constant-coefficient VAR with stochastic volatility.) Since the latter is a restricted version of the former, the Bayes factor in favor of the unrestricted model can be obtained using the Savage-Dickey density ratio via

$$\mathrm{BF}_{\mathrm{u},\mathbf{c}} = \frac{p(\boldsymbol{\gamma} = \mathbf{c})}{p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y})},$$

where the numerator and the denominator are, respectively, the marginal prior and pos-

terior densities of $\boldsymbol{\gamma}$ evaluated at $\mathbf{c}$. Hence, to compute the relevant Bayes factor, one only needs to evaluate two densities at a point. Intuitively, if $\boldsymbol{\gamma} = \mathbf{c}$ is less likely under the posterior density relative to the prior density, i.e., $p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y}) < p(\boldsymbol{\gamma} = \mathbf{c})$, then it is viewed as evidence against the restriction $\boldsymbol{\gamma} = \mathbf{c}$ and the unrestricted model is favored with $\log \mathrm{BF}_{\mathrm{u},\mathbf{c}} > 0$. Any two nested models can be compared similarly. We provide the technical details on evaluating the two densities at $\boldsymbol{\gamma} = \mathbf{c}$ in Appendix B.

We compare the proposed hybrid TVP-VAR to four VARs with stochastic volatility. For each of these VARs, $\gamma_1^\beta = \cdots = \gamma_n^\beta$ and $\gamma_1^\alpha = \cdots = \gamma_n^\alpha$, and they are denoted as HYB-$(\gamma_i^\beta, \gamma_i^\alpha)$. The first is a full-fledged TVP-VAR where all the VAR coefficients and elements of the impact matrix are time varying with $\gamma_i^\beta = \gamma_i^\alpha = 1, i = 1, \ldots, n$; this is a structural-form version of the TVP-VAR in Primiceri (2005) and we denote this model as HYB-$(1, 1)$. The second is a VAR with time-varying VAR coefficients but a constant impact matrix, i.e., $\gamma_i^\beta = 1, \gamma_i^\alpha = 0, i = 1, \ldots, n$, which we denote as HYB-$(1, 0)$; this is a structural-form parameterization of the TVP-VAR in Cogley and Sargent (2005). The third is a constant-coefficient VAR with $\gamma_i^\beta = \gamma_i^\alpha = 0, i = 1, \ldots, n$, which we denote as HYB-$(0, 0)$; this is a structural-form variant of the VAR in Carriero, Clark, and Marcellino (2019). Lastly, we also consider a version in which the VAR coefficients are constant but the elements of the impact matrix are time varying, i.e., $\gamma_i^\beta = 0, \gamma_i^\alpha = 1, i = 1, \ldots, n$, which we denote as HYB-$(0, 1)$.

Table 4 reports the log Bayes factors of the proposed hybrid TVP-VAR against these four VARs with stochastic volatility. It is clear that there is overwhelming support for the hybrid TVP-VAR relative to the alternatives. For instance, for $n = 20$ the Bayes factor in favor of the hybrid TVP-VAR against the best alternative HYB-$(0, 1)$ is $\mathrm{e}^{401} \approx 1.42 \times 10^{174}$, suggesting the posterior model probability of the hybrid TVP-VAR is practically 1. These model comparison results are consistent with the estimates of $\boldsymbol{\gamma}$ reported in Table 3, which indicate substantial heterogeneity in the time-variation pattern across equations—hence, fixing these indicators to either 0 or 1 for all equations is likely to be too restrictive.

Table 4: Log Bayes factors of the proposed hybrid TVP-VAR against four VARs with stochastic volatility: structural-form versions of Primiceri (2005) (HYB-$(1,1)$), Cogley and Sargent (2005) (HYB-$(1,0)$) and Carriero, Clark, and Marcellino (2019) (HYB-$(0,0)$), as well as a VAR with constant VAR coefficients but a time-varying impact matrix (HYB-$(0,1)$). A positive log Bayes factor indicates that the hybrid TVP-VAR is favored.

	$n = 3$	$n = 6$	$n = 20$
HYB-(1,1)	18	60	1003
HYB-(1,0)	12	89	1035
HYB-(0,1)	10	54	401
HYB-(0,0)	8	181	2540

To investigate the support for the hybrid TVP-VAR across model dimensions, we also consider a small system ($n = 3$) with only real GDP, PCE inflation and unemployment, as well as a medium system ($n = 6$) with three additional variables: Fed funds rate, industrial production and real average hourly earnings in manufacturing. For both model dimensions, the hybrid TVP-VAR is strongly favored by the data compared to the alternatives.

It is interesting to note that the best model among the four benchmarks changes across model dimensions. More specifically, for the small system with $n = 3$ variables, the best model is HYB-$(0,0)$, the constant-coefficient VAR with stochastic volatility, even though it is the most restrictive among the four VARs. This suggests that the additional flexibility in allowing time-varying coefficients does not sufficiently fit the data better to justify the added model complexity. This is inline with the results in Chan and Eisenstat (2018a), who find that, for fitting a 3-variable US dataset, a constant-coefficient VAR with stochastic volatility performs well relative to various VARs with different forms of time variation according to the marginal likelihood.

While HYB-$(0,0)$ is the preferred model for the small system, when more variables are included HYB-$(0,1)$ is strongly favored. For instance, for $n = 20$, allowing for time variation in the impact matrix increases the log marginal likelihood by 2,139—comparing HYB-$(0,0)$ and HYB-$(0,1)$. But further allowing for time variation in the VAR coefficients reduces the log marginal likelihood by 602—comparing HYB-$(0,1)$ and HYB-$(1,1)$. Taken together, these results suggest that it is generally useful to allow for time variation in the impact matrix. In contrast, adding time variation in the VAR coefficients should be done judiciously in large systems: while simply allowing all VAR coefficients to be

time-varying can be detrimental, using the proposed data-based approach to add time variation equation-wise can substantially improve model-fit.

5.4 Forecasting Results

Next, we evaluate the forecast performance of the proposed hybrid TVP-VAR relative to a few standard benchmarks. In particular, we consider 1) a conventional homoscedastic and constant-coefficient VAR; 2) a constant-coefficient VAR with stochastic volatility and a constant impact matrix (by setting all γ_i^β and γ_i^α to 0; denoted as HYB-$(0,0)$); and 3) a full-fledged TVP-VAR (by setting all the indicators to 1; denoted as HYB-$(1,1)$). 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

$$100 \times (1 - \text{RMSFE}^M_{i,m}/\text{RMSFE}^B_{i,m}),$$

and the percentage gains in terms of ALPL:

$$100 \times (\text{ALPL}^M_{i,m} - \text{ALPL}^B_{i,m}).$$

Figure 3 reports the forecasting results of the hybrid TVP-VAR, where we use a 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-quarter-ahead point forecasts, the hybrid TVP-VAR outperforms the benchmark for almost all variables (all but two for 1-quarter-ahead and one for 4-quarter ahead). For a few variables, such as the federal funds rate, real personal consumption expenditure and industrial production, the hybrid TVP-VAR outperforms the benchmark by more than 10% for 1-quarter-ahead forecasts (the differences in forecast performance are also statistically significant at 0.05 level according to the test of Diebold and Mariano (1995)). Overall, the median percentage gains in RMSFE for 1- and 4-quarter-ahead forecasts are, respectively, 5.1% and 6.0%.

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-quarter-ahead forecasts are 14% and 13%, respectively. Moreover, for many variables the percentage gains are more than 20%. These results are consistent with numerous studies in the small VAR literature, such as Clark (2011), D'Agostino, Gambetti, and Giannone (2013) and Clark and Ravazzolo (2015), that show allowing for time-varying structures substantially improves forecast performance compared to VARs with constant parameters, especially for density forecasts.

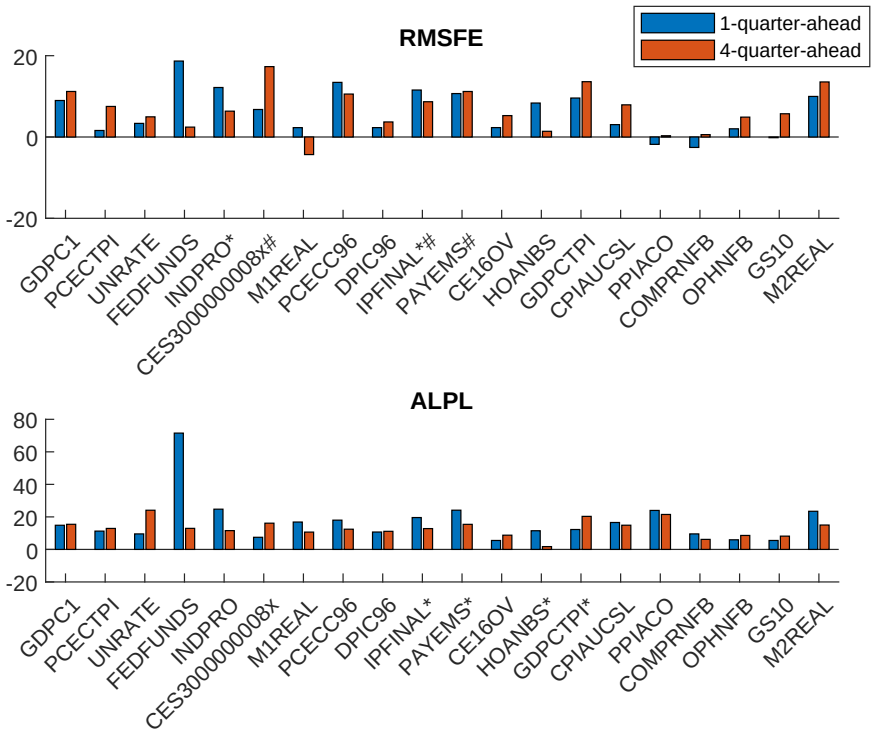

Figure 3: Forecasting results of the hybrid TVP-VAR compared to the benchmark: a standard homoscedastic, constant-coefficient VAR. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic indicate rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Next, we compare the forecast performance of the hybrid TVP-VAR with that of HYB-$(0,0)$, the constant-coefficient VAR with stochastic volatility. The results are reported in Figure 4. For both 1- and 4-quarter-ahead point forecasts, the hybrid TVP-VAR outperforms the HYB-$(0,0)$ for most variables. The median percentage gains in RMSFE are 1.7% and 4.1%, respectively. For density forecasts, the results are similar: the median percentage gains in ALPL for 1- and 4-quarter-ahead forecasts are, respectively, 0.8% and 3.1%. 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.

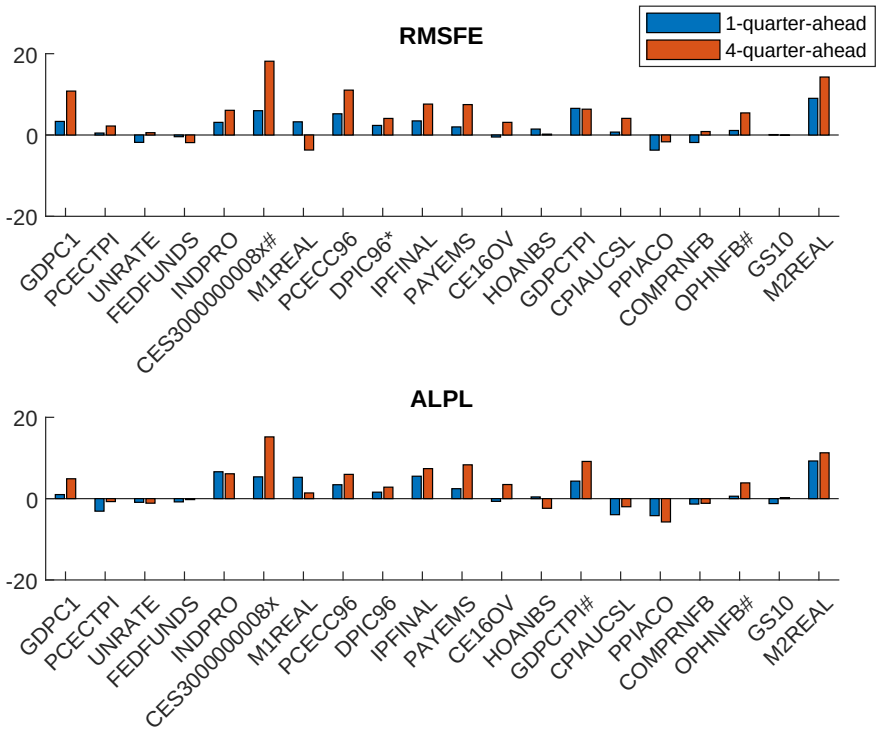

Figure 4: Forecasting results of the hybrid TVP-VAR compared to the benchmark HYB-$(0, 0)$. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic indicate rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Finally, Figure 5 compares the forecast performance of the the hybrid TVP-VAR with that of HYB-$(1, 1)$, the full-fledged TVP-VAR where all the VAR coefficients and error variances are time varying. Again, for both point and density forecasts, the hybrid TVP-VAR performs better than the benchmark for most variables. In particular, the median percentage gains in RMSFE for 1- and 4-quarter-ahead forecasts are 1.0% and 1.4%, respectively; the median percentage gains in ALPL are 2.4% and 2.8%, respectively. These results suggest that imposing time variation in all equations is not necessary and would adversely impact the forecast performance.

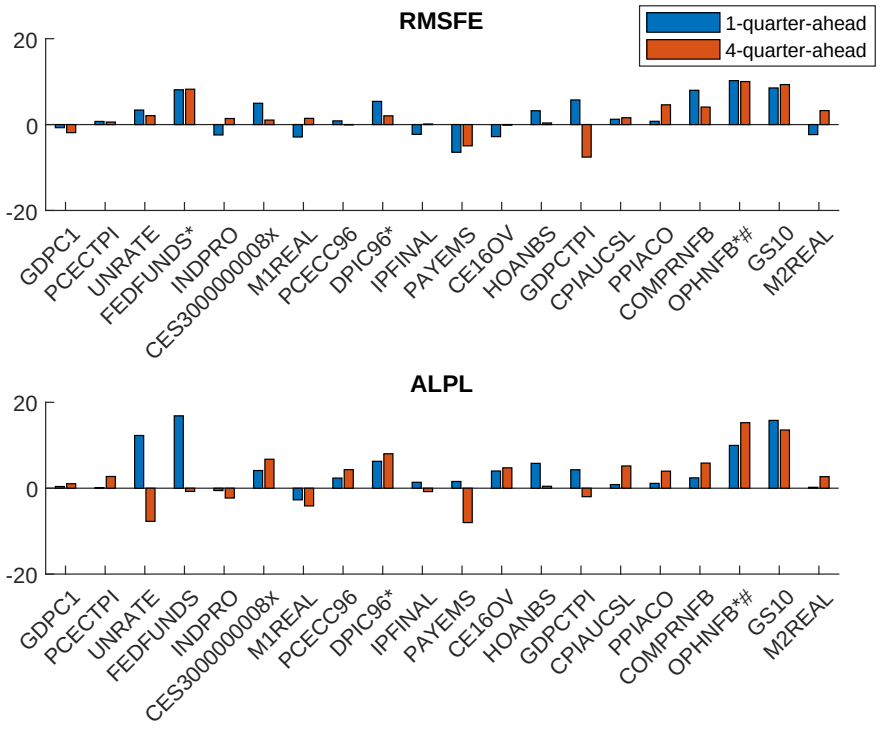

Figure 5: Forecasting results of the hybrid TVP-VAR compared to the benchmark HYB-$(1,1)$. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic indicate rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Overall, these forecasting results show that the proposed hybrid TVP-VAR forecasts better than many state-of-the-art time-varying models. These forecasting 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.

To better understand the sources of forecast gains, Figure 6 compares the forecast performance of HYB-$(1,1)$, the full-fledged TVP-VAR, to HYB-$(0,0)$, the constant-coefficient VAR with stochastic volatility. The results are mixed: while HYB-$(1,1)$ does slightly better in terms of point forecasts for the majority of the variables, it performs worse in terms of density forecasts for many variables. These results suggest that allowing for time-varying VAR coefficients in all equations does not necessarily improve forecast per-

formance. (The forecast performance of HYB-$(1,1)$ and HYB-$(1,0)$ are very similar, as reported in Appendix D). This finding is also consistent with the full-sample estimation results presented in Table 3: while the data clearly favors time-varying VAR coefficients in a few equations, for the majority of the equations time variation is not needed.

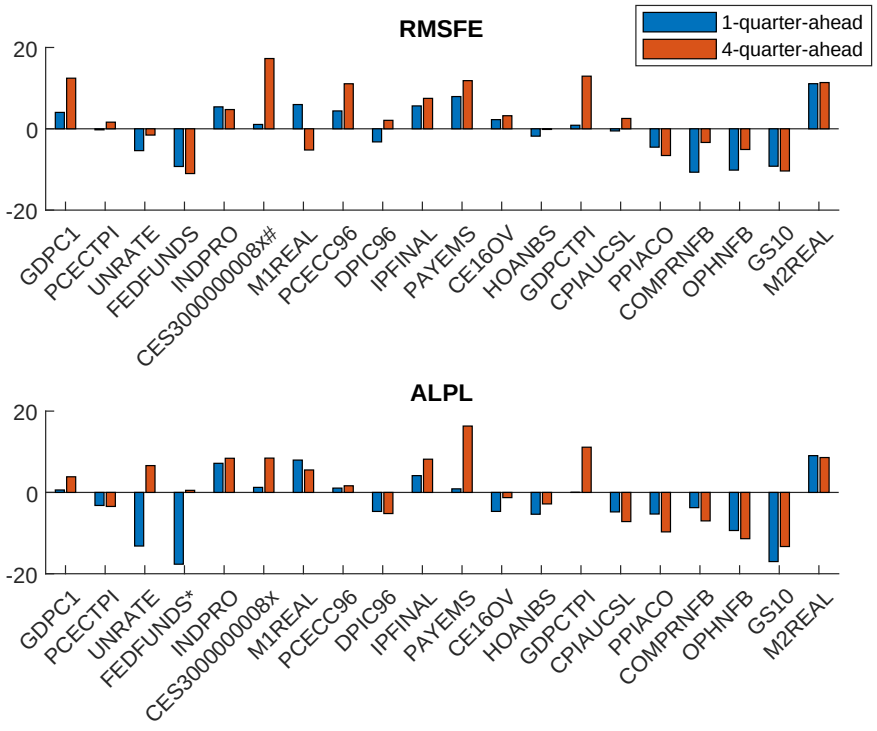

Figure 6: Forecasting results of HYB-$(1,1)$ compared to the benchmark HYB-$(0,0)$. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic indicate rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Finally, one can also interpret the superior forecast performance of the hybrid TVP-VAR through the lens of Bayesian forecast combinations (Min and Zellner, 1993; Andersson and Karlsson, 2008). As discussed in Section 2.3, the proposed hybrid TVP-VAR can be viewed as a Bayesian model average of a wide variety TVP-VARs with different forms of time variation, where each component model is characterized by the vector of indicators $\boldsymbol{\gamma}$. Then, the forecasts from the hybrid TVP-VAR can be interpreted as a forecast combination weighted by the posterior model probabilities $p(\boldsymbol{\gamma} \,|\, \mathbf{y})$. Consistent with the

large literature on forecast combinations, here we find that the forecast combination of the hybrid TVP-VAR performs better than many individual component models, including HYB-$(1,1)$ and HYB-$(0,0)$.

6 Concluding Remarks and Future Research

This paper has developed what 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 covariances in some equations are time varying, the data prefers constant coefficients 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 use hybrid TVP-VARs for structural analysis. Since they are formulated in the recursive structural-form, structural analysis using large VARs identified by recursive zero restrictions, such as the application in Ellahie and Ricco (2017), can directly use the proposed models.

In addition, developing an order-invariant version of these hybrid TVP-VARs would be an interesting and important extension. This would involve changing two components. First, one requires the use of the reduced-form VAR representation, but model indicators $\boldsymbol{\gamma}$ can be introduced similarly. An equation-by-equation estimation procedure can be developed along the lines in Carriero, Chan, Clark, and Marcellino (2021), though computation would be more intensive. Second, one would need to replace the Cholesky stochastic volatility model with an order-invariant model suitable for large VARs, such as Carriero, Clark, and Marcellino (2016) or Chan, Koop, and Yu (2021). The trade-off between the two stochastic volatility models is between computational speed and model flexibility: the former can be estimated quickly but the latter is more flexible. Finding a good modeling approach among all these choices—with an eye on the additional computational costs—would be an interesting research direction.

Appendix A: Estimation Details

In this appendix we provide estimation details of the hybrid TVP-VAR given in (6)-(9). In particular, we describe the details of Step 2 - Step 7 of the posterior sampler outlined in Section 3.2 of the main text.

Step 2. To sample $\mathbf{h}_i$, the log-volatility vector of the i-th equation, $i = 1, \ldots, n$, we can use the auxiliary mixture sampler of Kim, Shephard, and Chib (1998). More specifically, we first compute the residuals $\varepsilon^y_{i,t}, t = 1, \ldots, T$, using (6). Then, we transform these residuals as $\mathbf{y}^*_i = (\log(\varepsilon^y_{i,1})^2, \ldots, \log(\varepsilon^y_{i,T})^2)'$. Finally, we implement the auxiliary mixture sampler in conjunction with the precision sampler of Chan and Jeliazkov (2009) to sample $\mathbf{h}_i$ using $\mathbf{y}^*_i$ as data.

Step 3. The parameters $\boldsymbol{\Sigma}^{\frac{1}{2}}_{\theta_i} = \text{diag}(\boldsymbol{\Sigma}^{\frac{1}{2}}_{\beta_i}, \boldsymbol{\Sigma}^{\frac{1}{2}}_{\alpha_i})$ and $\boldsymbol{\theta}_{i,0}$ can be sampled easily as their joint distribution is Gaussian. To see that, let $\boldsymbol{\mu}^\theta_i = (\boldsymbol{\theta}'_{i,0}, \sigma_{\beta_i,1}, \ldots, \sigma_{\beta_i,k_\beta}, \sigma_{\alpha_i,1}, \ldots, \sigma_{\alpha_i,k_{\alpha_i}})'$ and define $\mathbf{w}^\theta_{i,t} = (\mathbf{x}_{i,t}, \gamma^\beta_i \widetilde{\mathbf{x}}_t \odot \widetilde{\boldsymbol{\beta}}_{i,t}, \gamma^\alpha_i \widetilde{\mathbf{w}}_{i,t} \odot \widetilde{\boldsymbol{\alpha}}_{i,t})$, where $\odot$ denotes the component-wise product. Then, we can rewrite (6) as a linear regression:

$$y_{i,t} = \mathbf{w}^\theta_{i,t} \boldsymbol{\mu}^\theta_i + \varepsilon^y_{i,t}.$$

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

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

where $\mathbf{K}_{\boldsymbol{\mu}^\theta_i} = \mathbf{V}^{-1}_{\boldsymbol{\mu}^\theta_i} + (\mathbf{W}^\theta_i)' \boldsymbol{\Omega}^{-1}_{\mathbf{h}_i} \mathbf{W}^\theta_i$ and $\widehat{\boldsymbol{\mu}}^\theta_i = \mathbf{K}^{-1}_{\boldsymbol{\mu}^\theta_i} (\mathbf{W}^\theta_i)' \boldsymbol{\Omega}^{-1}_{\mathbf{h}_i} \mathbf{y}_i$ with $\boldsymbol{\Omega}_{\mathbf{h}_i} = \text{diag}(\mathrm{e}^{h_{i,1}}, \ldots, \mathrm{e}^{h_{i,T}})$.

Steps 4-5. The full conditional distributions of $\sigma^2_{h,i}$ and $h_{i,0}$ are standard and they can be sampled easily. In particular, their full conditional distributions are

$$\begin{aligned}
(\sigma^2_{h,i} \,|\, \mathbf{h}_i, h_{i,0}) &\sim \mathcal{IG}\left(\nu_{h,i} + \frac{T}{2}, S_{h,i} + \frac{1}{2}\sum_{t=1}^{T} (h_{i,t} - h_{i,t-1})^2\right), \\
(h_{i,0} \,|\, \mathbf{h}_i, \sigma^2_{h,i}) &\sim \mathcal{N}(\widehat{h}_{i,0}, K^{-1}_{h_{i,0}}),
\end{aligned}$$

where $K_{h_{i,0}} = 1/V_{h_{i,0}} + 1/\sigma^2_{h,i}$ and $\widehat{h}_{i,0} = K^{-1}_{h_{i,0}} (a_{h_{i,0}}/V_{h_{i,0}} + h_{i,1}/\sigma^2_{h,i})$.

Step 6. Next, given the independent beta priors on p_i^β and $p_i^\alpha, i = 1, \ldots, n$ their full conditional posterior distributions are also beta distributions. In fact, we have:

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

Step 7. To implement Step 7, we follow the sampling approach in Chan (2021). 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 $(n-1)np$. Furthermore, 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 generalized inverse Gaussian distribution:

$$(\kappa_1 \,|\, \boldsymbol{\theta}_0) \sim \mathcal{GIG}\left(c_{1,1} - \frac{np}{2}, 2c_{2,1}, \sum_{(i,j)\in S_{\kappa_1}} \frac{\theta_{ij,0}^2}{C_{ij}}\right).$$

Similarly, κ_2 also has a generalized inverse Gaussian distribution:

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

Appendix B: Technical Details on Model Comparison

This appendix outlines the technical details on comparing hybrid TVP-VARs using the Savage-Dickey density ratio. First, suppose we want to compare the proposed hybrid TVP-VAR with $\boldsymbol{\gamma}$ unrestricted against a TVP-VAR characterized by $\boldsymbol{\gamma} = \mathbf{c} \in \{0,1\}^{2n}$. Then, the Bayes factor in favor of the unrestricted model can be expressed as the Savage-Dickey density ratio

$$\mathrm{BF}_{\mathrm{u},\mathbf{c}} = \frac{p(\boldsymbol{\gamma} = \mathbf{c})}{p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y})},$$

provided that the priors under the restricted and unrestricted models satisfy a compatibility condition (see, e.g., Verdinelli and Wasserman, 1995, for details). The priors described in Section 3.1 of the main text assume that the restricted (i.e., $\boldsymbol{\gamma}$) and unrestricted parameters are independent *a priori*, which implies the compatibility condition.

Next, we outline how one can evaluate the marginal prior and posterior densities at $\boldsymbol{\gamma} = \mathbf{c}$. First, to evaluate $p(\boldsymbol{\gamma} = \mathbf{c})$, recall that under our priors, γ_i^β and γ_i^α follow independent Bernoulli distributions with success probabilities p_i^β and p_i^α, respectively, for $i = 1, \ldots, n$. These success probabilities, in turn, are assumed to have beta distributions: $p_i^\beta \sim \mathcal{B}(a_{p^\beta}, b_{p^\beta})$ and $p_i^\alpha \sim \mathcal{B}(a_{p^\alpha}, b_{p^\alpha})$. Hence, the marginal prior of γ_i^β (unconditional on p_i^β) can be computed as

$$\begin{aligned} p(\gamma_i^\beta) &= \int_0^1 (p_i^\beta)^{\gamma_i^\beta} (1 - p_i^\beta)^{1-\gamma_i^\beta} \times \frac{\Gamma(a_{p^\beta} + b_{p^\beta})}{\Gamma(a_{p^\beta})\Gamma(b_{p^\beta})} (p_i^\beta)^{a_{p^\beta}-1} (1 - p_i^\beta)^{a_{p^\beta}-1} \mathrm{d}p_i^\beta \\ &= \frac{\Gamma(a_{p^\beta} + b_{p^\beta})\Gamma(\gamma_i^\beta + a_{p^\beta})\Gamma(1 - \gamma_i^\beta + b_{p^\beta})}{\Gamma(a_{p^\beta})\Gamma(b_{p^\beta})\Gamma(a_{p^\beta} + b_{p^\beta} + 1)}, \end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function. A similar expression can be derived for $p(\gamma_i^\alpha)$. It follows that the marginal prior for $\boldsymbol{\gamma}$ has the following analytical expression

$$\begin{aligned} p(\boldsymbol{\gamma}) &= \prod_{i=1}^n p(\gamma_i^\beta) p(\gamma_i^\alpha) \\ &= \prod_{i=1}^n \frac{\Gamma(a_{p^\beta} + b_{p^\beta})\Gamma(\gamma_i^\beta + a_{p^\beta})\Gamma(1 - \gamma_i^\beta + b_{p^\beta})}{\Gamma(a_{p^\beta})\Gamma(b_{p^\beta})\Gamma(a_{p^\beta} + b_{p^\beta} + 1)} \frac{\Gamma(a_{p^\alpha} + b_{p^\alpha})\Gamma(\gamma_i^\alpha + a_{p^\alpha})\Gamma(1 - \gamma_i^\alpha + b_{p^\alpha})}{\Gamma(a_{p^\alpha})\Gamma(b_{p^\alpha})\Gamma(a_{p^\alpha} + b_{p^\alpha} + 1)}, \end{aligned}$$

which can be easily evaluated at any point $\mathbf{c}$.

Next, we can evaluate $p(\boldsymbol{\gamma} = \mathbf{c} \,|\, \mathbf{y})$ with $\mathbf{c} = (c_{1,1}, c_{1,2}, \ldots, c_{n,1}, c_{n,2})'$ using the Monte Carlo average:

$$\frac{1}{R} \sum_{r=1}^{R} \prod_{i=1}^{n} \mathbb{P}\left(\boldsymbol{\gamma}_i = (c_{i,1}, c_{i,2}) \,|\, \mathbf{y}_i, \mathbf{h}_i^{(r)}, \boldsymbol{\Sigma}_{\theta_i}^{(r)}, \boldsymbol{\theta}_{i,0}^{(r)}\right),$$

where $\mathbf{h}_i^{(r)}, \boldsymbol{\Sigma}_{\theta_i}^{(r)}, \boldsymbol{\theta}_{i,0}^{(r)}, i = 1, \ldots, n, r = 1, \ldots, R$ are posterior draws from the unrestricted model. Analytical expressions of the above posterior probabilities are given in Section 3.2 of the main text. For numerical stability, both the prior and posterior densities are computed in log scale.

Thus, we have shown how one can compare the proposed hybrid TVP-VAR with $\boldsymbol{\gamma}$ unrestricted against any TVP-VAR characterized by $\boldsymbol{\gamma} = \mathbf{c} \in \{0, 1\}^{2n}$. To compare two TVP-VARs characterized by $\boldsymbol{\gamma} = \mathbf{c}_1$ and $\boldsymbol{\gamma} = \mathbf{c}_2$, note that one can express the Bayes factor in favor of $\boldsymbol{\gamma} = \mathbf{c}_1$ as

$$\mathrm{BF}_{\mathbf{c}_1, \mathbf{c}_2} = \frac{\mathrm{BF}_{\mathrm{u}, \mathbf{c}_2}}{\mathrm{BF}_{\mathrm{u}, \mathbf{c}_1}} = \frac{p(\boldsymbol{\gamma} = \mathbf{c}_1 \,|\, \mathbf{y})}{p(\boldsymbol{\gamma} = \mathbf{c}_2 \,|\, \mathbf{y})}.$$

This expression can be evaluated using posterior draws from the unrestricted model as before.

Appendix C: Data

The dataset covers 20 quarterly variables sourced from the FRED-QD database at the Federal Reserve Bank of St. Louis (McCracken and Ng, 2021). The sample period is from 1959Q1 to 2018Q4. Table 5 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 5: 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	$400\Delta \log$
Real 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	$400\Delta \log$

Appendix D: Additional Results

In this appendix we provide additional simulation and empirical results. First, we investigate the effect of the prior on γ_i^β and γ_i^α by replacing the beta prior $\mathcal{B}(0.5, 0.5)$ by the uniform prior $\mathcal{U}(0, 1)$ in the Monte Carlo experiments. Table 6 reports the frequencies of the posterior modes of γ_i^β and γ_i^α being one in 300 datasets. All in all, the simulation results are similar to the baseline case, suggesting that the posterior estimates are not sensitive to the prior on γ_i^β and γ_i^α.

Table 6: Frequencies (%) of the posterior modes of γ_i^β and γ_i^α being one in 300 datasets with $T = 400$. The two sets of priors for γ_i^β and γ_i^α are $\mathcal{B}(0.5, 0.5)$ and $\mathcal{U}(0, 1)$ priors.

Equation	True γ_i^β	True γ_i^α	$\mathcal{B}(0.5, 0.5)$ prior		$\mathcal{U}(0, 1)$ prior	
			γ_i^β	γ_i^α	γ_i^β	γ_i^α
1	0	0	0.06	–	0.06	–
2	0	1	0.04	0.88	0.03	0.87
3	1	0	0.98	0.25	0.99	0.24
4	1	1	0.98	0.64	0.98	0.62
5	0	0	0.02	0.02	0.00	0.02
6	0	1	0.03	0.96	0.02	0.96
7	1	0	0.97	0.13	0.98	0.11
8	1	1	0.95	0.80	0.95	0.81
9	0	0	0.03	0.00	0.01	0.00
10	0	1	0.04	0.94	0.02	0.96
11	1	0	0.94	0.11	0.97	0.10
12	1	1	0.93	0.88	0.93	0.90

Next, we compute the inefficiency factors of the posterior draws from the 20-variable hybrid TVP-VAR in Section 5 of the main text, defined as

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

where ρ_l is the sample autocorrelation at lag length l and L is chosen to be large enough so that the autocorrelation tapers off. In the ideal case where the posterior draws are independent, the corresponding inefficiency factor is 1. Figure 7 reports the inefficiency factors, obtained using 10,000 posterior draws after a burn-in period of 1,000. The results

show that the proposed posterior sampler is efficient in terms of producing posterior draws that are not highly autocorrelated.

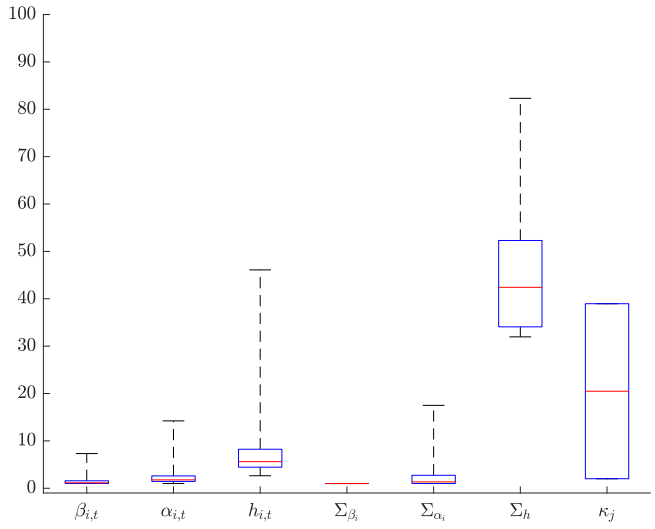

Figure 7: Boxplots of inefficiency factors of posterior draws from the 20-variable hybrid TVP-VAR.

Finally, we report additional forecasting results. First, Figure 8 reports forecasting results of the full-fledged TVP-VAR HYB-$(1,1)$ relative to the benchmark HYB-$(1,0)$, a TVP-VAR with stochastic volatility and a constant impact matrix (by setting all γ_i^β to 1 and all γ_i^α to 0). The results show that the forecast performance of the two models are mostly similar.

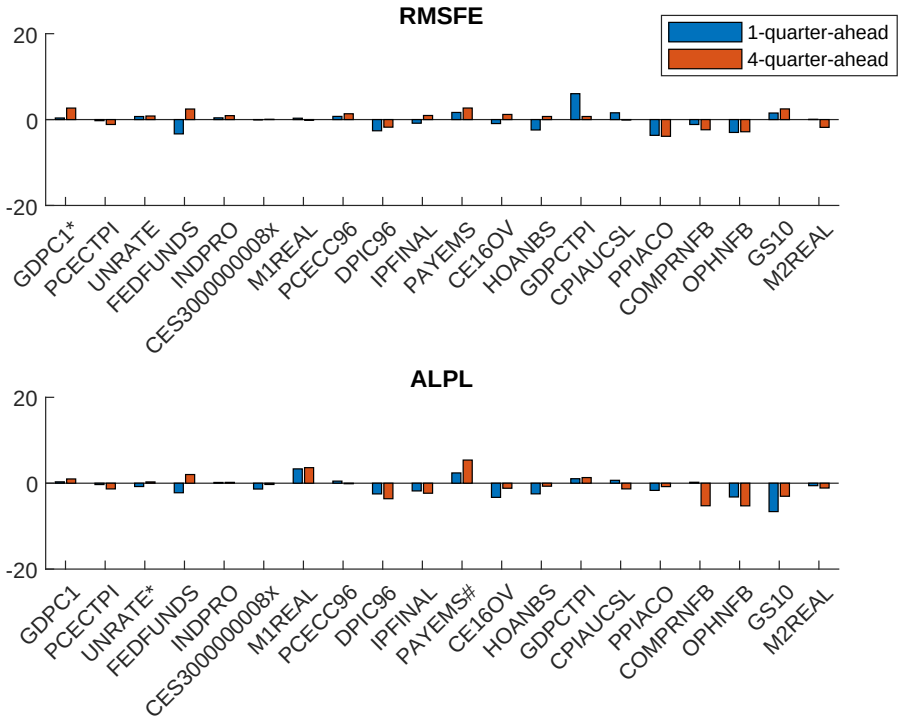

Figure 8: Forecasting results of HYB-$(1,1)$ compared to the benchmark HYB-$(1,0)$. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic indicate rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Next, Figure 9 reports the forecasting results of an extension of the hybrid TVP-VAR where the innovations to $h_{i,t}$ are correlated across equations relative to the benchmark HYB where the innovations are independent. The results show that the forecast performance of the extension is on average better in both point and density forecasts, although the forecast gains are modest. For example, the median percentage gains in RMSFE for 1- and 4-quarter-ahead forecasts are 0.6% and 0.4%, respectively; the median percentage gains in ALPL are 1.5% and 0.6%, respectively.

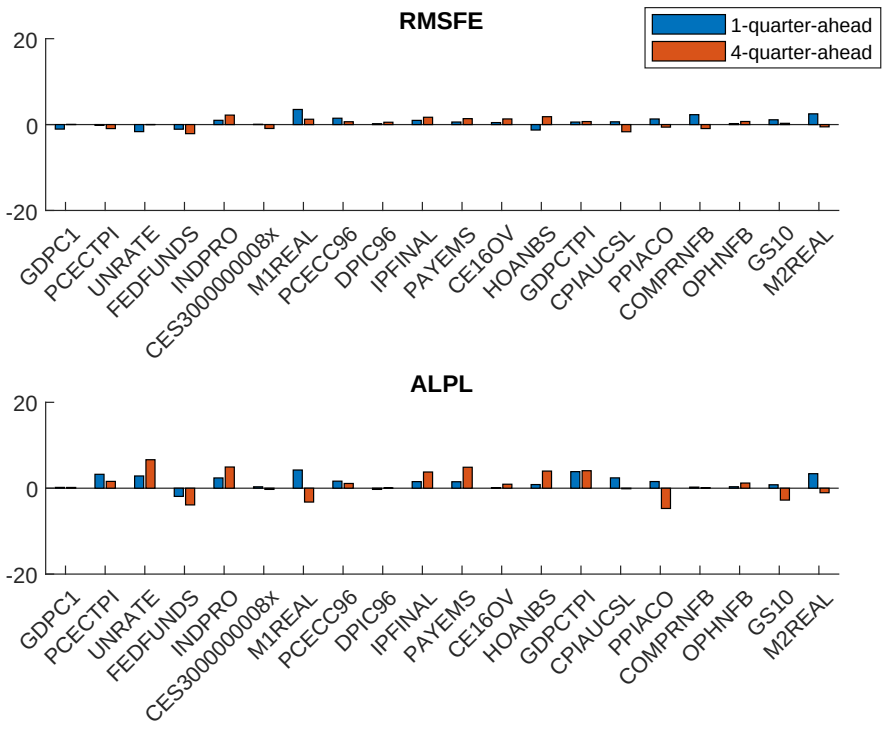

Figure 9: Forecasting results of an extension of HYB where the innovations to $h_{i,t}$ are correlated across equations compared to the benchmark HYB. The top and bottom panels show, respectively, the percentage gains in the root mean squared forecast error and the average of log predictive likelihoods of the proposed hybrid TVP-VAR. The symbols * and # after the mnemonic rejection of equal forecast accuracy at significance level 0.05 using the test in Diebold and Mariano (1995) for 1- and 4-quarter-ahead point forecasts, respectively.

Next, we investigate the variability of point and density forecasts of 6 variables—real GDP, PCE inflation, unemployment, Fed funds rate and industrial production and real average hourly earnings in manufacturing—from the proposed model. We consider all 720 possible variable orderings and the setup of the forecasting exercise is described in Section 5.4 in the main text. For each variable ordering, we compute the root mean squared forecast error and the average of log predictive likelihoods from the proposed model as well as the HYB-$(1, 1)$. We normalize the two metrics by those of the benchmark ordering, and the results are reported in Figures 10-11.

Consistent with the results in Arias, Rubio-Ramirez, and Shin (2021), the variability of point forecasts is relatively small. Moreover, the variability of both point and density

forecasts from the proposed model is similar to those in HYB-$(1,1)$, a structural-form parameterization of the model in Primiceri (2005).

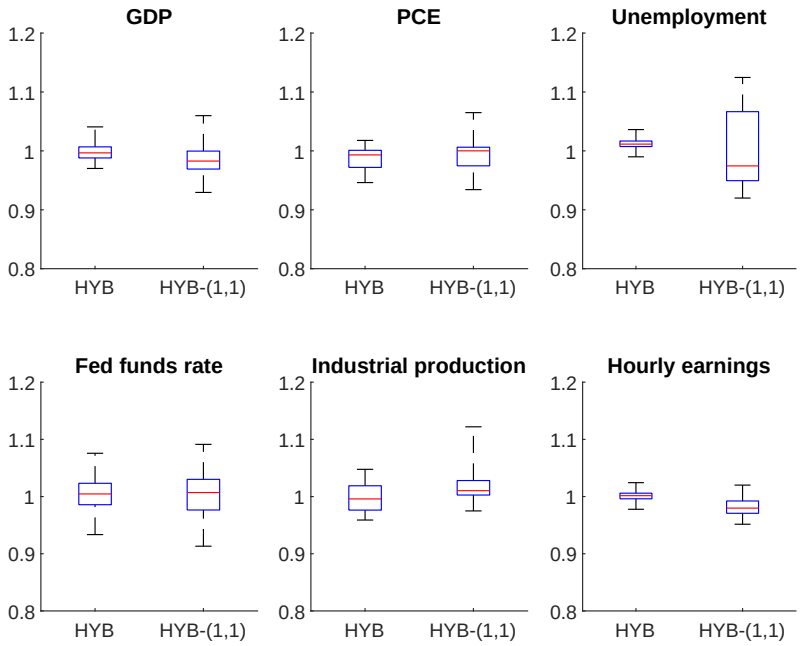

Figure 10: Boxplots of the relative mean squared forecast errors of 1-quarter-ahead point forecasts from the proposed hybrid TVP-VAR (HYB) and the full-fledge TVP-VAR (HYB-$(1,1)$).

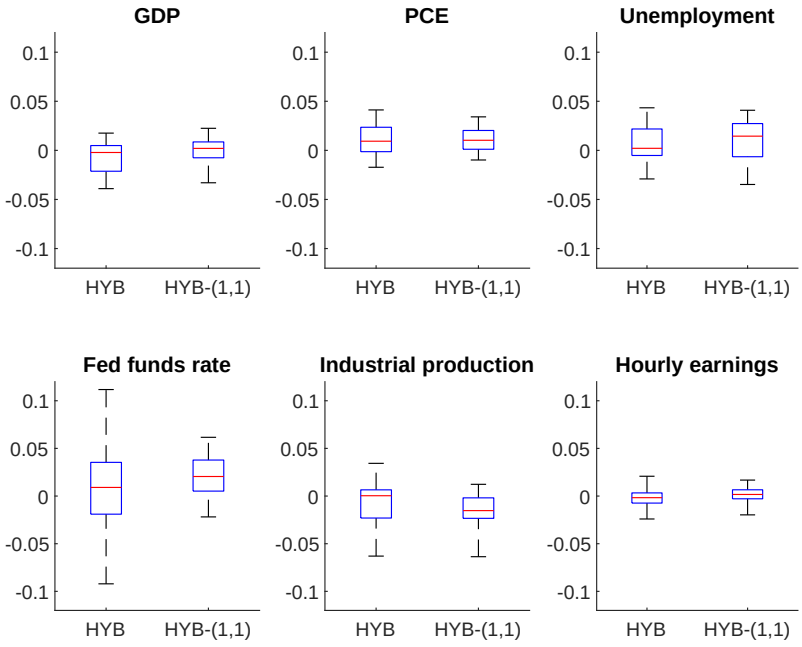

Figure 11: Boxplots of the relative average log predictive likelihoods of 1-quarter-ahead density forecasts from the proposed hybrid TVP-VAR (HYB) and the full-fledge TVP-VAR (HYB-$(1,1)$).

References

Andersson, M. K., and S. Karlsson (2008): "Bayesian forecast combination for VAR models," *Advances in Econometrics*, 23, 501–524.

Arias, J. E., J. F. Rubio-Ramirez, and M. Shin (2021): "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Federal Reserve Bank of Philadelphia Working Papers*.

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*.

Baumeister, C., and G. Peersman (2013): "Time-varying effects of oil supply shocks on the US economy," *American Economic Journal: Macroeconomics*, 4(5), 1–28.

Benati, L. (2008): "The "great moderation" in the United Kingdom," *Journal of Money, Credit and Banking*, 40(1), 121–147.

Bobeica, E., and B. Hartwig (2021): "The COVID-19 shock and challenges for time series models," *ECB Working Paper*.

Bognanni, M. (2018): "A class of time-varying parameter structural VARs for inference under exact or set identification," *FRB of Cleveland Working Paper*.

Carriero, A., J. C. C. Chan, T. E. Clark, and M. G. Marcellino (2021): "Corrigendum to: Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, forthcoming.

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*, 212(1), 137–154.

Carriero, A., T. E. Clark, M. G. Marcellino, and E. Mertens (2021): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *CEPR Discussion Paper No. DP15964*.

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. (2020): "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business and Economic Statistics*, 38, 68–79.

——— (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37(3), 1212–1226.

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., E. Eisenstat, and R. W. Strachan (2020): "Reducing the State Space Dimension in a Large TVP-VAR," *Journal of Econometrics*, 218(1), 105–118.

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.

Chan, J. C. C., G. Koop, and X. Yu (2021): "Large order-invariant Bayesian VARs with stochastic volatility," *Working Paper*.

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 (2015): "Macroeconomic Forecasting Performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30(4), 551–575.

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 (2020): "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 36(3), 899–915.

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.

Deborah, G., and R. W. Strachan (2009): "Nonlinear impacts of International business cycles on the U.K.—A Bayesian smooth transition VAR approach," *Studies in Nonlinear Dynamics and Econometrics*, 14(1), 1–33.

Del Negro, M., and G. E. Primiceri (2015): "Time-varying structural vector autoregressions and monetary policy: a corrigendum," *Review of Economic Studies*, 82(4), 1342–1345.

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

Diebold, F. X., and R. S. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–265.

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

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.

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

Giraitis, L., G. Kapetanios, and S. Price (2013): "Adaptive forecasting in the presence of recent and ongoing structural change," *Journal of Econometrics*, 177(2), 153–170.

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

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*.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65(3), 361–393.

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.

——— (2018): "Variational Bayes inference in high-dimensional time-varying parameter models," *Available at SSRN 3246472*.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2010): "Dynamic probabilities of restrictions in state space models: an application to the Phillips curve," *Journal of Business and Economic Statistics*, 28(3), 370–379.

Koop, G., and S. M. Potter (1999): "Bayes factors and nonlinearity: evidence from economic time series," *Journal of Econometrics*, 88(2), 251–281.

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 (2021): "FRED-QD: A quarterly database for macroeconomic research," *Federal Reserve Bank of St. Louis Review*, 103(1), 1–44.

Min, C., and A. Zellner (1993): "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics*, 56(1-2), 89–118.

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

Petrova, K. (2019): "A quasi-Bayesian local likelihood approach to time varying parameter VAR models," *Journal of Econometrics*, 212(1), 286–306.

Primiceri, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72(3), 821–852.

Prüser, J. (2021): "The horseshoe prior for time-varying parameter VARs and monetary policy," *Journal of Economic Dynamics and Control*, p. 104188.

Shin, M., and M. Zhong (2020): "A new approach to identifying the real effects of uncertainty shocks," *Journal of Business and Economic Statistics*, 38(2), 367–379.

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

Verdinelli, I., and L. Wasserman (1995): "Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio," *Journal of the American Statistical Association*, 90(430), 614–618.