Econometrics and Statistics 20 (2021) 87–108

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Flexible Mixture Priors for Large Time-varying Parameter Models

Niko Hauzenberger[a,*]

Department of Economics and Salzburg Centre of European Union Studies (SCEUS), University of Salzburg, Mönchsberg 2a, 5020 Salzburg, Austria

ARTICLE INFO

Article history:
Received 15 February 2021
Revised 30 May 2021
Accepted 1 June 2021
Available online 19 June 2021

JEL classification:
C11
C30
C53
E44
E47

Keywords:
Time-varying parameter vector autoregressions
hierarchical modeling
clustering
macroeconomic forecasting

ABSTRACT

Time-varying parameter (TVP) models often assume that the TVPs evolve according to a random walk. This assumption, however, might be questionable since it implies that coefficients change smoothly and in an unbounded manner. This assumption is relaxed by proposing a flexible law of motion for the TVPs in large-scale vector autoregressions (VARs). Instead of imposing a restrictive random walk evolution of the latent states, hierarchical mixture priors on the coefficients in the state equation are carefully designed. These priors effectively allow for discriminating between periods in which coefficients evolve according to a random walk and times where the TVPs are better characterized by a stationary stochastic process. Moreover, this approach is capable of introducing dynamic sparsity by pushing small parameter changes towards zero if necessary. The merits of the model are illustrated by means of two applications. Using synthetic data these flexible modeling techniques yield precise parameter estimates. When applied to US data, the model reveals interesting patterns of low-frequency dynamics in coefficients and forecasts well relative to a wide range of competing models.

1. Introduction

A growing number of papers introduces time-varying parameters (TVP) in econometric models for capturing structural breaks in relations across macroeconomic fundamentals (see, for example, Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006; Korobilis, 2013; Eickmeier et al., 2015; Mumtaz and Theodoridis, 2018; Czudaj, 2019; Paul, 2020) and to achieve more accurate macroeconomic forecasts (see, for instance, Koop and Korobilis, 2012; 2013; D'Agostino et al., 2013; Groen et al., 2013; Bauwens et al., 2015; Hauzenberger et al., 2019; Huber et al., 2020a; 2020b).

In this paper, we focus on estimating TVP vector autoregressive (VAR) models featuring a large number of endogenous variables. Due to severe overfitting issues in large TVP-VARs, special emphasis is paid to important modeling decisions, such as whether coefficients evolve gradually, change abruptly or remain constant for subset of periods. In macroeconomic applications, it is common to assume that coefficients evolve according to a random walk, implying that parameters change

* Corresponding author.
E-mail address: niko.hauzenberger@sbg.ac.at

https://doi.org/10.1016/j.ecosta.2021.06.001

smoothly over time. As noted by the recent literature (see, for example, Lopes et al., 2018; Hauzenberger et al., 2019), however, this assumption may be overly simplistic and lead to model mis-specification.

In large TVP-VARs it is often reasonable to assume that most parameters remain constant over time, while only few vary. To capture this behavior, the Bayesian literature frequently uses shrinkage priors on the state innovation variances to sufficiently push them towards zero (Frühwirth-Schnatter and Wagner, 2010; Belmonte et al., 2014). A severe drawback of this strategy is that it only accounts for the case that a given coefficient is constant for all points in time (labeled static sparsity).

Another common situation faced by researchers is that coefficients change only at certain points in time (this is referred to as dynamic sparsity). Using a mixture distribution on the innovation variances, for example, allows to push small parameter changes towards zero (see, inter alia, McCulloch and Tsay, 1993; Gerlach et al., 2000; Giordani and Kohn, 2008; Koop et al., 2009; Huber et al., 2019). Alternatively, Hauzenberger et al. (2019) introduce a more flexible law of motion by assuming a conjugate hierarchical location mixture prior directly on the time-varying part of the coefficients. This location mixture allows for dynamically adjusting the prior mean on the TVPs to capture situations with a low, moderate or even large number of structural breaks in the coefficients. Both techniques come with drawbacks. For instance, the mixture innovation model of Huber et al. (2019), equipped with a latent threshold mechanism, discriminates between a high and a low innovation variance state. However, the authors do not discard the random walk law of motion, which might be too restrictive. Hauzenberger et al. (2019) use either a conjugate g-prior (Zellner, 1986) or a conjugate Minnesota prior (Doan et al., 1984; Litterman, 1986), potentially lacking flexibility to truly disentangle abrupt from gradual changes.

In this paper, we carefully design suitable mixture priors for the state equation. In a first variant, a mixture prior is not only introduced on the state innovations, but also on the autoregressive coefficients in the state equation to obtain sufficient flexibility. To achieve parsimony in large models, a latent binary indicator determines the law of motion for the TVPs and detect periods in which coefficients evolve according to a random walk and times in which the TVPs are better characterized by a stationary stochastic process. Combined with a mixture on the innovation volatilities and suitable shrinkage priors, this approach is capable of automatically capturing a wide range of typical parameter changes. In a second variant, the sparse finite location mixture model of Hauzenberger et al. (2019) is extended by considering non-conjugate shrinkage priors and by replacing the location with a location-scale mixture. Here, similar to mixture innovation models, an additional mixture on the state variances reflects the notion that structural breaks in coefficients happen infrequently (with potentially large TVP innovations), while most of the time coefficients are constant (with TVP innovations pushed towards zero).

In the previous paragraphs we repeatedly stated that our techniques are well suited to handle overfitting issues in large TVP-VARs. But large TVP models also raise the question of computational feasibility. In this contribution, computational complexity is reduced by using recent advances in estimating large-scale TVP regression (see Chan and Jeliazkov, 2009; McCausland et al., 2011; Hauzenberger et al., 2020). These are based on rewriting the TVP model in its static regression form. In this representation, the TVP model is treated as a very big regression model and the techniques proposed in Bhattacharya et al. (2016) can be used. Since these algorithms are designed for single equation models, we consider the VAR model in its structural representation and estimate a set of unrelated equation-specific TVP regressions.

Based on two applications we investigate the merits of the techniques developed in the paper. First, in an application using synthetic data we illustrate that the proposed methods work well in detecting small and large structural breaks in coefficients. Second, we employ a large US macroeconomic dataset for an empirical application. Our proposed methods reveal interesting patterns in the low-frequency relationship between unemployment and inflation. Moreover, to evaluate predictive performance of our approach, we perform a comprehensive forecasting exercise. This forecasting horse race shows that the proposed framework works well relative to a wide range of competing models. Even for large TVP-VARs introducing flexible mixture priors in the state equation tends to improve forecast accuracy.

The remainder of the paper is structured as follows. Section 2 introduces a TVP regression model with flexible mixture priors and sketches the main contributions of the paper. Section 3 outlines inference in these models, while Section 4 discusses the posterior sampling algorithm of Bhattacharya et al. (2016), when applied to non-centered TVP regressions. Section 5 and Section 6 show the results for artificial data and US data, respectively. Finally, Section 7 summarizes and concludes.

2. Econometric Framework

2.1. A TVP Regression

Let y_t denote a scalar time series and $\boldsymbol{x}_t$ refer to a K-dimensional vector of predictors, then the observation equation for a TVP regression can be written as:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_t^2). \tag{1}$$

Here, $\boldsymbol{\alpha}_t$ is a K-dimensional vector of TVPs that relates $\boldsymbol{x}_t$ to the quantity of interest and ε_t denotes the measurement error with mean zero and time-varying variance σ_t^2. For the state equation of σ_t^2, we assume a stochastic volatility (SV) specification and refer to Appendix A.1 for details.

Typically, $\boldsymbol{\alpha}_t$ is assumed to evolve according to a random walk. In this paper, interest centers on relaxing this assumption. In the following, to achieve both sufficient flexibility and model parsimony, we use two different hierarchical mixture

specifications for $\boldsymbol{\alpha}_t$. In the first variant, we assume that coefficients evolve according to a mixture of a random walk and a white noise process. In the second variant, interest centers on extending the methods proposed in Hauzenberger et al. (2019).

2.2. A Hierarchical Mixture between a Random Walk and a White Noise Process

For a mixture between a random walk and white noise process we assume that the evolution of $\boldsymbol{\alpha}_t$ is given by:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\phi}_t(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_0) + \boldsymbol{\varsigma}_t, \quad \boldsymbol{\varsigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_t), \tag{2}$$

with $\boldsymbol{\alpha}_0$ denoting a K-dimensional intercept vector, $\boldsymbol{\phi}_t$ being a K-dimensional diagonal autoregressive coefficient matrix and $\boldsymbol{\varsigma}_t$ denoting a K-dimensional vector of state innovations, which are centered on zero and feature a $K \times K$-dimensional variance-covariance matrix $\boldsymbol{\Psi}_t$. Moreover, we assume $\boldsymbol{\phi}_t$ and $\boldsymbol{\Psi}_t$ to evolve according to a regime-switching process:

$$\boldsymbol{\phi}_t = \boldsymbol{S}_t, \tag{3}$$

and

$$\boldsymbol{\Psi}_t = \boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0. \tag{4}$$

Here, $\boldsymbol{S}_t = \text{diag}\,(s_{1t}, \dots, s_{Kt})$ denotes a binary indicator matrix with $\{s_{it}\}_{i=1}^K$ being either zero or one, $\boldsymbol{I}_K$ is a K-dimensional identity matrix, and $\bar{\boldsymbol{\Psi}}_1 = \text{diag}\,(\bar{\psi}_{11}, \dots, \bar{\psi}_{K1})$ as well as $\bar{\boldsymbol{\Psi}}_0 = \text{diag}\,(\bar{\psi}_{10}, \dots, \bar{\psi}_{K0})$ denote K-dimensional diagonal matrices. Equation 3 assumes that coefficients evolve according to a mixture of a random walk and a white noise process, while Equation 4 ensures sufficient flexibility of the state innovations. For example, if the covariate-specific indicator $s_{it} = 1$ in the t^{th} period, the i^{th} covariate follows a random walk with state innovation variance $\psi_{it} = \bar{\psi}_{i1}$, while if $s_{it} = 0$ in the t^{th} period, it follows a stochastic process with variance $\psi_{it} = \bar{\psi}_{i0}$.

This specification (henceforth labeled as `TVP-MIX`) nests a wide variety of popular TVP models, such as standard random walk state equations and mixture innovation models. A standard random walk evolution is trivially obtained by setting $\boldsymbol{S}_t = \boldsymbol{I}_K$. A so-called mixture innovation model assumes $\boldsymbol{S}_t = \boldsymbol{I}_K$ and specifies $\boldsymbol{\Psi}_t$ similar to Equation 4 (Gerlach et al., 2000; Giordani and Kohn, 2008; Koop et al., 2009; Huber et al., 2019). Additionally, mixture innovation specifications restrict $\bar{\boldsymbol{\Psi}}_0 = \kappa\hat{\boldsymbol{\Psi}}_0$ with κ being a small value close to zero and $\hat{\boldsymbol{\Psi}}_0$ being a diagonal matrix collecting variable specific scaling parameters. Related to the literature on variable selection (George and McCulloch, 1993; 1997), here $\bar{\boldsymbol{\Psi}}_1$ is commonly referred to as slab component and $\bar{\boldsymbol{\Psi}}_0$ as spike component. Depending on the hierarchical modeling assumptions on $\boldsymbol{S}_t$ and $\boldsymbol{\Psi}_t$ other dynamic sparsification techniques could be recovered as well. These approaches include different forms of dynamic shrinkage processes (see, inter alia, Kalli and Griffin, 2014; Uribe and Lopes, 2017; Kowal et al., 2019; Hauzenberger et al., 2020; Rockova and McAlinn, 2021), latent threshold models (Nakajima and West, 2013) or dynamic model selection techniques (Chan et al., 2012; Koop and Korobilis, 2013).

Apart from discussing the relation to other popular TVP models, it is also worth highlighting additional features of the model proposed in (2) to (4). If a parameter is almost constant, but also features larger abrupt changes for some periods, we would expect that $\bar{\psi}_{i0} > \bar{\psi}_{i1}$. This case is of particular interest, when compared to a standard mixture innovation model with random walk state equation. Conversely, if a coefficient features large, more persistent swings, but also some periods of parameter stability, we would expect $\bar{\psi}_{i0} < \bar{\psi}_{i1}$. Intuitively, the relative proportions of $\bar{\psi}_{i0}$ and $\bar{\psi}_{i1}$ depend mainly on the nature of coefficient changes. If the i^{th} coefficient is constant or negligible (static sparsity), this can be achieved with $\bar{\psi}_{i1}$ and/or $\bar{\psi}_{i0}$ close to zero (Lopes et al., 2018). Note that in the special case of constant coefficients, the proposed specification is not identified. We address this issue in the context of interpreting the state indicators $\boldsymbol{S}_t$.

2.3. A Hierarchical Pooling Specification

For a hierarchical pooling specification, we follow Hauzenberger et al. (2019) and assume that the time-varying part of $\boldsymbol{\alpha}_t$ follows a sparse finite mixture in the spirit of Malsiner-Walli et al. (2016).

The specification of the state equation $\boldsymbol{\alpha}_t$ (labeled as `TVP-POOL`) reads as:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\gamma}_t. \tag{5}$$

Here, $\boldsymbol{\alpha}_0$ denotes a K-dimensional constant coefficient vector and $\boldsymbol{\gamma}_t$ is assumed to be a K-dimensional vector of random coefficients featuring a specific structure. That is, conditional on latent group indicators θ_t that takes an integer value $n \in \{1, \dots, N\}$, $\boldsymbol{\gamma}_t$ follows a multivariate Gaussian distribution:

$$\boldsymbol{\gamma}_t = \boldsymbol{\mu}_n + \boldsymbol{\varsigma}_t, \quad \boldsymbol{\varsigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_t), \quad \text{if} \quad \theta_t = n, \tag{6}$$

where $\boldsymbol{\mu}_n$ refers to the group-specific mean and $\boldsymbol{\Psi}_t$ denotes the variance-covariance matrix. It is also worth noting that θ_t serves as group indicator for $\boldsymbol{\gamma}_t$. The probability that $\boldsymbol{\gamma}_t$ is assigned to cluster n is defined as $P(\theta_t = n) = \omega_n$.

This structure is closely related to the set-up of Hauzenberger et al. (2019). In the following, we extend their location mixture prior to a location-scale mixture prior by introducing a regime-switching specification on $\boldsymbol{\Psi}_t$, similar to Equation 4. That is:

$$\boldsymbol{\Psi}_t = \boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0, \tag{7}$$

with both $\bar{\boldsymbol{\Psi}}_0$ and $\bar{\boldsymbol{\Psi}}_1$ being diagonal matrices and $\boldsymbol{S}_t$ denoting a binary indicator matrix. Similar to standard mixture innovation models one component serves to detect larger breaks, while a second component handles dynamic sparsity. We therefore discard the conjugate prior assumption of Hauzenberger et al. (2019) and instead assume non-conjugate shrinkage priors on both state innovation variances (described in more detail in Sub-section 3.1).

Before proceeding, it is also worth sketching the general idea of this random coefficient specification. This model can be seen as a stochastic variant of multiple break point specifications (Koop and Potter, 2007), which is capable of capturing situations with a low, moderate or even large number of structural breaks. To estimate the number of regimes, we follow Malsiner-Walli et al. (2016) and Hauzenberger et al. (2019) and specify an "overfitting" model by setting N to a large integer (i.e., consider many regimes a priori). To achieve parsimony, we come up with an estimate for the number of clusters $\hat{N}$ (usually $\hat{N} < N$) by specifying a shrinkage prior on both the mixture weights and the component means. Thus, overall shrinkage is determined between two interacting objectives: we aim at eliminating irrelevant clusters, while at the same time avoiding highly overlapping component means.

At this stage one might ask, why we do not assume N different state innovation variances (i.e., using the group indicators θ_t for both $\boldsymbol{\gamma}_t$ and $\boldsymbol{\Psi}_t$)? Here, it is worth discussing two important considerations. First, N denotes a large integer and might lead to overfitting issues without assuming additional hierarchical shrinkage/pooling priors on the state innovation variances. Second, covariate-specific binary indicators ($\boldsymbol{S}_t$) for the scales already render the model highly flexible and it allows to introduce shrinkage on the state innovation variances in a simpler way. Moreover, the two-state mixture on the state variances (see Equation 7) is designed to foster inference about the locations $\boldsymbol{\gamma}_n$, for $n = \{1, \dots, N\}$. We expect that many elements in $\{\boldsymbol{\gamma}_t\}_{t=1}^T$ cluster around zero (i.e., coefficients are constant with $\boldsymbol{\Psi}_t$ close to zero), while occasionally there are structural breaks in some coefficients (requires relatively large values in $\boldsymbol{\Psi}_t$). We aim to correctly detect in particular those two extremes (changes/no changes in $\boldsymbol{\alpha}_t$) with $\boldsymbol{\gamma}_t$.

2.4. The Latent State Indicator Matrix

Sofar we remained silent on the evolution of $\boldsymbol{S}_t$. There are many different possibilities how the binary indicators s_{it}, for $i = \{1, \dots, K\}$, evolve over time. In the following, we assume two laws of motion:

1. **Pooled Markov-switching process:** When assuming a first-order Markov process for each s_{it} independently, sampling the state indicators can be computationally cumbersome, especially if K is large. Since one has to rely on forward filtering backward sampling algorithms, computation time quickly adds up. Therefore, we replace $\boldsymbol{S}_t$ with $s_t \boldsymbol{I}_K$. In the following, s_t is assumed to be common to all K covariates in period t and governed by a joint Markov process. This process is driven by a transition probability matrix given by:

$$P = \begin{pmatrix} p_{00} & 1 - p_{11} \\ 1 - p_{00} & p_{11} \end{pmatrix},$$

   with transition probabilities from state k to l denoted by p_{kl} and following a Beta distribution $p_{kk} \sim \mathcal{B}(c_{0k}, c_{1k})$, for $k = \{0, 1\}$, see Uribe and Lopes (2017).
2. **Independent over time and covariate-specific indicators:** The assumption that a joint indicator governs the evolution of large number of coefficients might be too inflexible in certain cases. For this reason, we also specify covariate-specific indicators, coupled with independent mixture priors (see Lopes et al., 2018). In contrast to covariate-specific Markov processes, mixture priors are assumed to be independent over time and thus do not involve computationally demanding forward filtering backward sampling algorithms. In the following, s_{it} is assumed to be independent Bernoulli distributed with $P(s_{it} = 1) = p_i$ and p_i being Beta distributed: $p_i \sim \mathcal{B}(c_{i,0}, c_{i,1})$.

Moreover, it should be noted that the prior choice on the binary indicators is quite influential. For the random walk/white noise mixture (`TVP-MIX`), the hyperparameters are chosen in such a way that a priori gradual changes have a higher (unconditional) expected duration (with $s_t = 1$) than abrupt changes (with $s_t = 0$). In the empirical application, we therefore set $c_{00} = 0.3, c_{01} = 30, c_{10} = 30, c_{11} = 0.3$ for the Markov-switching process and $c_{i,0} = 0.3$, $c_{i,1} = 30$, $i = \{1, \dots, K\}$, for the independent mixture distribution. For the location-scale mixture (`TVP-POOL`with $\boldsymbol{S}_t$ solely governing the state innovation variances, we take a more agnostic approach by assuming $c_{00} = c_{11} = c_{i,0} = 0.3$ and $c_{01} = c_{10} = c_{i,1} = 3$.

3. Bayesian Inference

To discuss inference for both variants outlined in Section 2, we introduce a very general state equation for $\boldsymbol{\alpha}_t$:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\gamma}_t + \boldsymbol{\phi}_t(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_0) + \boldsymbol{\varsigma}_t, \quad \boldsymbol{\varsigma}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}_t). \tag{8}$$

Equation 8 nests both approaches with the first variant (`TVP-MIX`) being obtained by setting $\boldsymbol{\gamma}_t = \boldsymbol{0}_{K \times 1}$, while the second approach (`TVP-POOL`) is derived by defining $\boldsymbol{\phi}_t = \boldsymbol{0}_{K \times K}$ and $\boldsymbol{\gamma}_t = \boldsymbol{\mu}_n$, if $\theta_t = n$.

3.1. The Non-Centered Parameterization

In this sub-section we exploit the non-centered parameterization to write $\bar{\boldsymbol{\Psi}}_0$ and $\bar{\boldsymbol{\Psi}}_1$ as part of the observation equation, enabling shrinkage on the regime-switching state innovation volatilities (Frühwirth-Schnatter and Wagner, 2010).

We therefore recast the model as follows:

$$\begin{aligned} y_t = \quad & \boldsymbol{x}_t' \underbrace{(\boldsymbol{\alpha}_0 + \sqrt{\boldsymbol{\Psi}_t}\tilde{\boldsymbol{\alpha}}_t)}_{\boldsymbol{\alpha}_t} + \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1), \quad \text{and} \\ \tilde{\boldsymbol{\alpha}}_t = \quad & \tilde{\boldsymbol{\gamma}}_t + \boldsymbol{\phi}_t \tilde{\boldsymbol{\alpha}}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_K), \quad \tilde{\boldsymbol{\alpha}}_0 = \boldsymbol{0}, \boldsymbol{\phi}_1 = \boldsymbol{I}_K. \end{aligned} \tag{9}$$

Here, $\tilde{\boldsymbol{\alpha}}_t$ is a K-dimensional vector of normalized states, defined as $\tilde{\boldsymbol{\alpha}}_t = (\sqrt{\boldsymbol{\Psi}_t})^{-1}(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_0)$ and $\tilde{\boldsymbol{\gamma}}_t = (\sqrt{\boldsymbol{\Psi}_t})^{-1}\boldsymbol{\gamma}_t$ with $\sqrt{\boldsymbol{\Psi}_t} = \text{diag}\,(\sqrt{\psi_{1t}}, \dots, \sqrt{\psi_{Kt}})$ denoting the (matrix) square-root of $\boldsymbol{\Psi}_t$. Using the definition of $\boldsymbol{\Psi}_t$ in Equation 4 (or Equation 7) the observation equation in Equation 9 can be rewritten as:

$$y_t = \boldsymbol{x}_t' \left(\boldsymbol{\alpha}_0 + \boldsymbol{S}_t \sqrt{\bar{\boldsymbol{\Psi}}_1} \tilde{\boldsymbol{\alpha}}_t + (\boldsymbol{I}_K - \boldsymbol{S}_t) \sqrt{\bar{\boldsymbol{\Psi}}_0} \tilde{\boldsymbol{\alpha}}_t \right) + \sigma_t \epsilon_t,$$

and, more compactly, as a standard regression model:

$$y_t = \hat{\boldsymbol{x}}_t' \hat{\boldsymbol{\alpha}} + \sigma_t \epsilon_t.$$

Here, $\hat{\boldsymbol{x}}_t = (\boldsymbol{x}_t', (\boldsymbol{S}_t \boldsymbol{x}_t \odot \tilde{\boldsymbol{\alpha}}_t)', ((\boldsymbol{I}_K - \boldsymbol{S}_t)\boldsymbol{x}_t \odot \tilde{\boldsymbol{\alpha}}_t)')'$ denotes a $3K$-dimensional covariate vector with $\odot$ referring to component-wise multiplication and $\hat{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_0', \sqrt{\bar{\psi}_{11}}, \dots, \sqrt{\bar{\psi}_{K1}}, \sqrt{\bar{\psi}_{10}}, \dots, \sqrt{\bar{\psi}_{K0}})'$ being a $3K$-dimensional coefficient vector.

On the time-invariant $\hat{\boldsymbol{\alpha}}$ we use a hierarchical global-local shrinkage prior (see Polson and Scott, 2010):

$$\hat{\alpha}_j \sim \mathcal{N}(0, \tau_j), \quad \tau_j | \lambda \sim f, \quad \lambda \sim g, \quad \text{for} \quad j = 1, \dots 3K,$$

where $\hat{\alpha}_j$ refers to the j$^{\text{th}}$ element in $\hat{\boldsymbol{\alpha}}$, τ_j induces local shrinkage with mixing density f and λ denotes a global shrinkage parameter with density g. In the empirical application, we focus on the Normal-Gamma (Griffin and Brown, 2010) shrinkage prior and choose f and g accordingly. This shrinkage prior has been proven to be successful in macroeconomic and financial application (see, for example, Huber and Feldkircher, 2019) and is quite common in the literature. The exact prior specification is outlined in Appendix A.2. Still, it is worth noting that any global-local shrinkage prior might be used. Other popular choices are the SSVS prior (George and McCulloch, 1993; 1997), the Horseshoe prior (Carvalho et al., 2010), the Bayesian Lasso (Park and Casella, 2008) or the Triple-Gamma prior (Cadonna et al., 2020). For thorough studies of global-local shrinkage priors in macroeconomic applications, see Follett and Yu (2019), Cross et al. (2020), Huber et al. (2020a), and Kastner and Huber (2020).

3.2. The Static Representation

If interest centers on estimating the latent states $\{\tilde{\boldsymbol{\alpha}}_t\}_{t=1}^T$, we can straightforwardly recast Equation 9 in a static regression form by conditioning on $\boldsymbol{\alpha}_0$, the state innovation volatilities $\{\sqrt{\boldsymbol{\Psi}_t}\}_{t=1}^T$ and the stochastic volatilities in $\boldsymbol{\Sigma} = \text{diag}\,(\sigma_1, \dots, \sigma_T)$. We define $\boldsymbol{y}$ as a T-dimensional vector, $\boldsymbol{X}$ as a $T \times K$-dimensional matrix and $\boldsymbol{\epsilon}$ as a T-dimensional vector with y_t, $\boldsymbol{x}_t'$ and ϵ_t on the t$^{\text{th}}$ position, respectively. Then, the static form of Equation 9 is:

$$\begin{aligned} \boldsymbol{y} = \quad & \boldsymbol{X}\boldsymbol{\alpha}_0 + \boldsymbol{W}\tilde{\boldsymbol{\alpha}} + \boldsymbol{\Sigma}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_T), \quad \text{and} \\ \boldsymbol{\Phi}\tilde{\boldsymbol{\alpha}} = \quad & \tilde{\boldsymbol{\gamma}} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_\nu). \end{aligned}$$

Here, $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_1', \dots, \tilde{\boldsymbol{\alpha}}_T')'$ is a ν $(= TK)$-dimensional latent state vector, $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1', \dots, \tilde{\boldsymbol{\gamma}}_T')'$ is a ν-dimensional intercept vector and $\boldsymbol{\eta}$ is a ν-dimensional shock vector. After defining $\tilde{\boldsymbol{x}}_t' = \boldsymbol{x}_t' \sqrt{\boldsymbol{\Psi}_t}$, the precise structure of $\boldsymbol{W}$ and $\boldsymbol{\Phi}$ is given by:

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{x}_1' & \boldsymbol{0}_{K\times 1}' & \dots & \boldsymbol{0}_{K\times 1}' \\ \boldsymbol{0}_{K\times 1}' & \tilde{\boldsymbol{x}}_2' & \dots & \boldsymbol{0}_{K\times 1}' \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{K\times 1}' & \boldsymbol{0}_{K\times 1}' & \dots & \boldsymbol{x}_T' \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{I}_K & \boldsymbol{0}_{K\times K} & \dots & \boldsymbol{0}_{K\times K} & \boldsymbol{0}_{K\times K} \\ -\boldsymbol{\phi}_2 & \boldsymbol{I}_K & \dots & \boldsymbol{0}_{K\times K} & \boldsymbol{0}_{K\times K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0}_{K\times K} & \boldsymbol{0}_{K\times K} & \dots & -\boldsymbol{\phi}_T & \boldsymbol{I}_K \end{pmatrix},$$

with $\boldsymbol{0}_{K\times K}$ denoting a $K \times K$-dimensional zero matrix. Solving for $\tilde{\boldsymbol{\alpha}}$ yields:

$$\tilde{\boldsymbol{\alpha}} = \boldsymbol{\Phi}^{-1}(\tilde{\boldsymbol{\gamma}} + \boldsymbol{\eta}),$$

implying that $\tilde{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{a}_0, \boldsymbol{\Omega}_0)$ with prior mean $\boldsymbol{a}_0 = \boldsymbol{\Phi}^{-1}\tilde{\boldsymbol{\gamma}}$ and prior variance-covariance matrix $\boldsymbol{\Omega}_0 = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}$ of $\tilde{\boldsymbol{\alpha}}$ (see, for instance, Chan and Jeliazkov, 2009; Chan and Strachan, 2020). In the special case of $\boldsymbol{\phi}_t = \boldsymbol{0}_{K\times K}$, for all t, $\boldsymbol{\Phi}$ (and thus $\boldsymbol{\Omega}_0$) reduces to an identity matrix, while $\boldsymbol{\phi}_t \neq \boldsymbol{0}_{K\times K}$, for any t, induces a (specific) banded lower-triangular (block diagonal) structure of $\boldsymbol{\Phi}$ ($\boldsymbol{\Omega}_0$). Note that $\boldsymbol{\phi}_t = \boldsymbol{0}_{K\times K}$, for $t = \{1, \dots, T\}$, is always true for the `TVP-POOL` model, but not ruled out for the `TVP-MIX` specification. The precise structure of the prior variance-covariance matrix $\boldsymbol{\Omega}_0$ thus solely depends on state indicators $\boldsymbol{S}_t$.

Moreover, the prior mean $\boldsymbol{a}_0$ also depends on the structure of $\boldsymbol{\Phi}$ and $\tilde{\boldsymbol{\gamma}}$. The simplest thing is to set $\boldsymbol{a}_0$ to a zero vector, which we implicitly assume for the `TVP-MIX` variants. For the `TVP-POOL` approach we use a hierarchical mixture prior on $\boldsymbol{a}_0$, described in detail next.

3.3. A Hierarchical Prior Mean

The model outlined in (5) to (7) denotes a sparse finite location-scale mixture. After recasting the model in the non-centered parameterization, we are able to replace the location-scale mixture prior on $\boldsymbol{\alpha}_t$ (outlined in Equation 6) with a location mixture prior on the normalized latent states $\tilde{\boldsymbol{\alpha}}_t$, since the scales of Equation 7 ($\bar{\boldsymbol{\Psi}}_0$ and $\bar{\boldsymbol{\Psi}}_1$) are now part of the observation equation. That is:

$$\tilde{\boldsymbol{\alpha}}_t | \theta_t = n \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_n, \boldsymbol{I}_K), \tag{10}$$

with group-specific mean $\tilde{\boldsymbol{\mu}}_n$, for $n = \{1, \dots, N\}$ and variance-covariance matrix $\boldsymbol{I}_K$. In the following, the prior mean is defined as $\boldsymbol{a}_0 = (\boldsymbol{a}_{01}', \dots, \boldsymbol{a}_{0T}')'$ with $\boldsymbol{a}_{0t} = \tilde{\boldsymbol{\mu}}_n$ if $\theta_t = n$.

The sparse finite location mixture in Equation 10 allows us to use a similar prior set-up as proposed in Malsiner-Walli et al. (2016) and Hauzenberger et al. (2019). To ensure model parsimony we use a Dirichlet prior on $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)'$:

$$\boldsymbol{\omega} | \xi \sim \text{Dir}(\xi, \dots, \xi),$$

with ξ referring to an intensity parameter. The prior on the intensity parameter is specified as:

$$\xi \sim \mathcal{G}(d_0, d_0 N),$$

with $d_0 = 10$ in the empirical application. Here, we closely follow Malsiner-Walli et al. (2016), who show that this prior choice is successful in detecting superfluous components and obtaining a parsimonious mixture representation.

Moreover, on the group means we specify the following shrinkage prior:

$$\tilde{\boldsymbol{\mu}}_n | \tilde{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_0),$$

with $\tilde{\boldsymbol{\mu}}_n$ being centered on zero and prior variance-covariance matrix $\boldsymbol{\Lambda}_0 = \boldsymbol{LRL}$. Here, $\boldsymbol{L} = \text{diag}\ (\sqrt{l_1}, \dots, \sqrt{l_K})$ and $\boldsymbol{R} = \text{diag}(r_1^2, \dots, r_K^2)$ with r_i denoting the range of $\tilde{\boldsymbol{\alpha}}_i = (\tilde{\alpha}_{i1}, \dots, \tilde{\alpha}_{iT})'$. Moreover, we specify a Gamma prior on the elements in $\boldsymbol{L}$:

$$l_i \sim \mathcal{G}(e_0, e_1).$$

Following Yau and Holmes (2011) and Malsiner-Walli et al. (2016), we define $e_0 = e_1 = 0.6$ to push the group-specific prior means towards zero.

4. Posterior computation

In this sub-section, we outline the MCMC sampling step for $\tilde{\boldsymbol{\alpha}}$. We stress that drawing $\hat{\boldsymbol{\alpha}}$ is computationally fast, also for relatively large K, but sampling the ν-dimensional vector $\tilde{\boldsymbol{\alpha}}$ is demanding on time (Hauzenberger et al., 2020). Thus for $\hat{\boldsymbol{\alpha}}$ (and the remaining parameters) we use standard MCMC techniques with sampling steps and conditional posteriors outlined in Appendix A.3.

For $\tilde{\boldsymbol{\alpha}}$, irrespectively of the structure of $\boldsymbol{a}_0$ and $\boldsymbol{\Omega}_0$, we obtain standard conditional Gaussian posterior quantities with $\tilde{\boldsymbol{y}} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}_0)$ and $\tilde{\boldsymbol{W}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{W}$:

$$\tilde{\boldsymbol{\alpha}} | \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}, \boldsymbol{a}_0, \boldsymbol{\Omega}_0 \sim \mathcal{N}(\boldsymbol{a}_1, \boldsymbol{\Omega}_1) \quad \text{with}$$
$$\boldsymbol{\Omega}_1^{-1} = \underbrace{(\tilde{\boldsymbol{W}}'\tilde{\boldsymbol{W}} + \boldsymbol{\Omega}_0^{-1})}_{\nu \times \nu} \quad \text{and} \quad \boldsymbol{a}_1 = \boldsymbol{\Omega}_1(\tilde{\boldsymbol{W}}'\tilde{\boldsymbol{y}} + \boldsymbol{\Omega}_0^{-1}\boldsymbol{a}_0).$$

The main issue, however, is that the inversion of $\boldsymbol{\Omega}_1^{-1}$ is computationally costly, since it is $\nu \times \nu$-dimensional matrix with $\nu = TK$ and $\{T, K\}$ being potentially large integers.

Thus, to avoid high-dimensional full matrix inversions and Cholesky decompositions for drawing the normalized latent states $\tilde{\boldsymbol{\alpha}}$, we rely on the algorithms proposed in Bhattacharya et al. (2016), applied to TVP models in Hauzenberger et al. (2020). This method involves the following steps:

1. Draw a ν-dimensional vector $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_\nu)$.
2. Sample a T-dimensional vector $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_T)$.
3. Define $\boldsymbol{q} = \boldsymbol{a}_0 + \boldsymbol{\Phi}^{-1}\boldsymbol{u}$, with $\boldsymbol{\Phi}^{-1}$ denoting the lower Cholesky factor of $\boldsymbol{\Omega}_0$, and $\boldsymbol{r} = \tilde{\boldsymbol{W}}\boldsymbol{q} + \boldsymbol{v}$.
4. Compute $\tilde{\boldsymbol{\Omega}}_1 = (\boldsymbol{I}_T + \tilde{\boldsymbol{W}}\boldsymbol{\Omega}_0\tilde{\boldsymbol{W}}')^{-1}$.
5. Set $\boldsymbol{f} = \tilde{\boldsymbol{\Omega}}_1(\tilde{\boldsymbol{y}} - \boldsymbol{r})$.
6. Obtain a draw with $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\Omega}_0\tilde{\boldsymbol{W}}'\boldsymbol{f}) + \boldsymbol{q}$.

Using the static representation for a TVP regression the involved matrices are sparse, which can be exploited to achieve additional computational gains (see Chan and Jeliazkov, 2009; Hauzenberger et al., 2019; 2020). Depending on the structure of $\boldsymbol{\Phi}$ there are two extreme cases as briefly discussed in Sub-section 3.2. Computationally, the most expensive case is a random walk state equation ($\boldsymbol{\phi}_t = \boldsymbol{I}_K$, $\forall\ t$), while having no autoregressive structure in the state equation ($\boldsymbol{\phi}_t = \boldsymbol{0}_{K \times K}$, $\forall\ t$) it is computationally less demanding (see Hauzenberger et al. (2020) for a comparison between the two extremes). Recall, the former $\boldsymbol{\Phi}$ has a specific lower triangular structure (rendering $\boldsymbol{\Omega}_0$ block diagonal) and in the latter both $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}_0$ are

diagonal. Thus, even for a random walk state equation (the most dense case), using sparse algorithms pays off in terms of computation. Moreover, if $\boldsymbol{\phi}_t = \boldsymbol{S}_t$, we have to account for an intermediate computational burden lying between the two extremes that eventually depends on the exact structure of $\boldsymbol{\Phi}$ ($\boldsymbol{\Omega}_0$).

4.1. Equation-wise estimation for a TVP-VAR

The methods outlined in the previous sub-section are designed for single equation models. To use these algorithms also for posterior inference in TVP-VARs, we rewrite the multivariate model as a set of unrelated TVP regressions (see, e.g., Carriero et al., 2019; Chan, 2019; Huber et al., 2019).

This can be achieved by using the structural form of the TVP-VAR:

$$\boldsymbol{Y}_t = \boldsymbol{B}_{0t}\boldsymbol{Y}_t + \sum_{\ell=1}^{p} \boldsymbol{B}_{\ell t}\boldsymbol{Y}_{t-\ell} + \boldsymbol{C}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t). \tag{11}$$

Here, $\boldsymbol{Y}_t = (Y_{1t}, \dots, Y_{mt})'$ denotes an m-dimensional vector of endogenous variables with $\boldsymbol{B}_{0t}$ being an $m \times m$-dimensional strictly lower-triangular matrix (with zero main diagonal) defining contemporaneous relationships between the elements of $\boldsymbol{Y}_t$. Moreover, $\boldsymbol{B}_{\ell t}$, for $\ell = 1, \dots, p$, denotes an $m \times m$-dimensional time-varying coefficient matrix, $\boldsymbol{C}_t$ is an m-dimensional intercept vector and $\boldsymbol{\varepsilon}_t$ refers to an m-dimensional Gaussian distributed error vector, centered on zero and with time-varying m-dimensional diagonal variance-covariance matrix $\boldsymbol{\Sigma}_t = \text{diag}\,(\sigma_{1t}^2, \dots, \sigma_{mt}^2)$. Before proceeding, it is convenient to define $\boldsymbol{B}_t = (\boldsymbol{B}_{1t}, \dots, \boldsymbol{B}_{pt})$.

In the following, for $i = 2, \dots, m$, the i$^{\text{th}}$ equation of $\boldsymbol{Y}_t$ is given by:

$$y_{it} = \boldsymbol{x}_{it}' \underbrace{(\boldsymbol{\alpha}_{i0} + \tilde{\boldsymbol{\alpha}}_{it})}_{\boldsymbol{\alpha}_{it}} + \quad \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2),$$

with $\boldsymbol{x}_{it}$ denoting a $K_i(= mp + i)$-dimensional covariate vector with $\boldsymbol{x}_{it} = (\{y_{jt}\}_{j=1}^{i-1}, \boldsymbol{Y}_{t-1}', \dots, \boldsymbol{Y}_{t-p}', 1)'$ and $\boldsymbol{\alpha}_{it} = (\{b_{ij,0t}\}_{j=1}^{i-1}, \boldsymbol{B}_{i\bullet,t}, c_{it})'$ a K_i-dimensional vector of time-varying coefficients. Here $b_{ij,0t}$ refers to the $(i,j)^{th}$ element of $\boldsymbol{B}_{0t}$, $\boldsymbol{B}_{i\bullet,t}$ denotes the i^{th} row of $\boldsymbol{B}_t$ and c_{it} the i^{th} element of $\boldsymbol{C}_t$. Moreover, for the first equation ($i = 1$) we have $\boldsymbol{x}_{1t} = (\boldsymbol{Y}_{t-1}', \dots, \boldsymbol{Y}_{t-p}', 1)'$ and $\boldsymbol{\alpha}_{1t} = (\boldsymbol{B}_{1\bullet,t}, c_{1t})'$.

5. Simulation study

In this section we use synthetic data to illustrate the features of the proposed mixture variants. For the data generating process (DGP) we assume that the number of observations is $T = 100$ and the number of covariates is given by $K = 5$. The covariates are simulated with $\boldsymbol{X}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_T)$ for $j = 1, \dots, (K-1)$ and $\boldsymbol{X}_K = \boldsymbol{\iota}_T$ with $\boldsymbol{\iota}_T$ being a T-dimensional vector of ones. For the error variance σ_t^2, we assume an SV specification with $\log(\sigma_t^2) = h_t$ following a random walk process. That is, $h_t = h_{t-1} + \vartheta_t$ with $\vartheta_t \sim \mathcal{N}(0, 0.1)$ and $h_0 = \log(0.1)$. For the time-varying parameters $\boldsymbol{\alpha}_t$ we introduce quite specific laws of motion. We define $\boldsymbol{\alpha}_0 = (-4, 3, -2, 2, 0)'$ as initial level and assume that both regime-switching autoregressive parameters and regime-switching variances in the state equation are governed by a joint Markov process s_t. Here, we let $\boldsymbol{S}_t = s_t \boldsymbol{I}_K$, $\boldsymbol{\phi}_t = \boldsymbol{S}_t$, $\boldsymbol{\Psi}_t = \boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$ with $\bar{\boldsymbol{\Psi}}_0 = \text{diag}\,(10^{-10}, 0.5, 0.1, 10^{-10}, 1)$ and $\bar{\boldsymbol{\Psi}}_1 = \text{diag}\,(1, 0.1, 0.5, 10^{-10}, 10^{-10})$. The joint indicator s_t is simulated with transition probabilities $p_{00} = 0.6$ and $p_{11} = 0.95$, effectively leading to a higher unconditional probability that $\boldsymbol{\alpha}_t$ follows a random walk evolution.

In particular, the first coefficient features larger, more persistent parameter changes (with $\psi_{11} = 1$), while for a small number of periods $\boldsymbol{\alpha}_{1t}$ is basically constant (achieved through the white noise state equation with ψ_{10} close to zero). The second parameter features small gradual changes over time ($\psi_{21} = 0.1$), but also larger abrupt breaks ($\psi_{20} = 0.5$). The third parameter is similar to the first coefficient, but assumes $\psi_{30} > \psi_{10}$. The fourth coefficient is assumed to be constant over time (with ψ_{40} and ψ_{41} being both close to zero) and, finally, the fifth parameter features a few extremely large breaks (with $\psi_{50} = 1$), while it is otherwise assumed to be constant (here achieved through the random walk state equation with ψ_{51} close to zero).

To assess the flexibility of our approaches we compare them to models assuming a standard random walk evolution of coefficients and to those assuming constant coefficients. Moreover, we consider a typical mixture innovation model as an important benchmark. For each model we use a Normal-Gamma (Griffin and Brown, 2010) prior on the constant part and allow for SV in the measurement error variance. Furthermore, each TVP model features a Normal-Gamma prior on the square root of the state innovation variances, which are potentially regime-switching (see Equation 9).

Panels (a) to (c) in Figure 1 depict the evolution of regression parameters estimated with our proposed methods, while panel (d) shows estimates with a typical mixture innovation model. The red solid lines denote the true coefficients, the blue solid lines the posterior median, the blue shaded areas represent the 68% posterior credible interval, the gray shaded area represent the respective credible set of a standard TVP model with random walk state equation, and the black-dotted lines indicate the 16$^{\text{th}}$, 50$^{\text{th}}$, and 84$^{\text{th}}$ percentiles of a constant coefficient model.

The results with artificial data reveal at least three important features. First, all TVP models yield reasonable estimates for constant coefficients, which is highly crucial for forecasting applications. Focusing on the the fourth parameter, considering a more flexible model pays off to produce less biased and more precise estimates, especially when compared to the constant

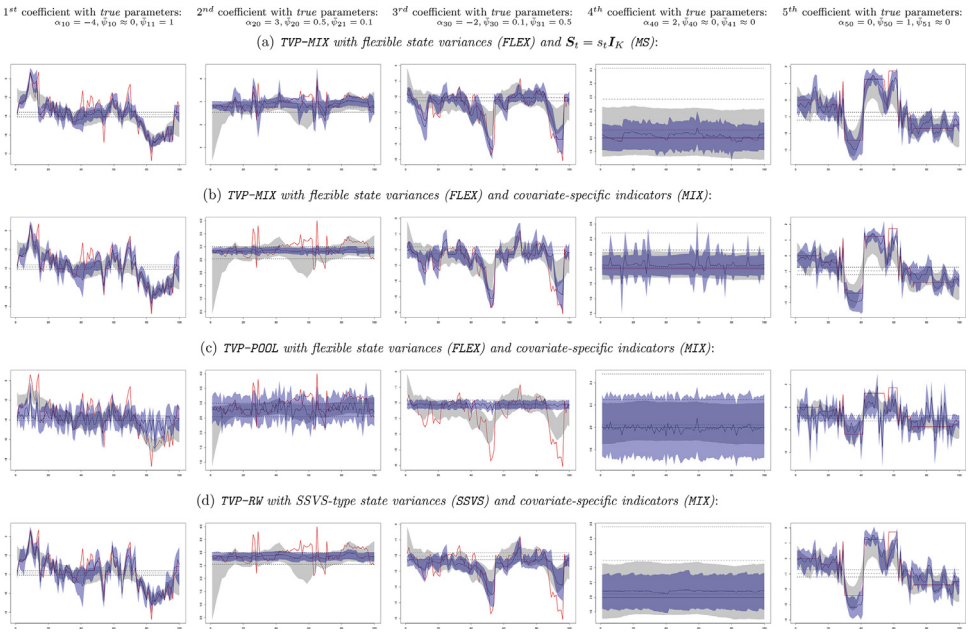


Fig. 1. The blue-shaded areas denote the 68% posterior credible intervals of the proposed methods with the blue solid lines denoting the posterior medians. The gray shaded areas refer to the 68% credible sets of a standard TVP regression with random walk state equation. The black dotted lines indicate the 16th, 50th, and 84th percentiles of a constant coefficient model. Moreover, the red lines denote the true coefficients of $\boldsymbol{\alpha}_t$.

Table 1
Overview of main specifications.

TVP-MIX	$\boldsymbol{\phi}_t = \boldsymbol{S}_t$	$\boldsymbol{\Psi}_t =$	$\boldsymbol{S}_t =$	Related to:
FLEX MS		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	$s_t\boldsymbol{I}_K$	
FLEX MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	
SINGLE		$\bar{\boldsymbol{\Psi}}$		
SSVS MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + \kappa(\boldsymbol{I}_K - \boldsymbol{S}_t)\hat{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	Chan et al. (2012)
TVP-POOL	$\boldsymbol{\phi}_t = \boldsymbol{0}_{K\times K}$			
FLEX MS		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	$s_t\boldsymbol{I}_K$	
FLEX MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	
SINGLE		$\bar{\boldsymbol{\Psi}}$		Hauzenberger et al. (2019)
SSVS MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + \kappa(\boldsymbol{I}_K - \boldsymbol{S}_t)\hat{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	
TVP-RW	$\boldsymbol{\phi}_t = \boldsymbol{I}_K$			
FLEX MS		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	$s_t\boldsymbol{I}_K$	
FLEX MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + (\boldsymbol{I}_K - \boldsymbol{S}_t)\bar{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	
SINGLE		$\bar{\boldsymbol{\Psi}}$		Primiceri (2005)
SSVS MIX		$\boldsymbol{S}_t\bar{\boldsymbol{\Psi}}_1 + \kappa(\boldsymbol{I}_K - \boldsymbol{S}_t)\hat{\boldsymbol{\Psi}}_0$	diag $(s_{1t}, \dots, s_{Kt})$	e.g., Huber et al. (2019)

coefficient model. Second, the TVP-MIX specifications are capable in capturing both rapid shifts and smooth adjustments in the regression coefficients. Our methods in panel (a) and (b) tend to quickly adjust when facing high frequency changes, rendering the methods even more flexible when compared to a typical mixture innovation specification in panel (d). Third, the TVP-POOL model in panel (c) tends to detect sudden changes in the parameters quite well, but is less capable in capturing low frequency movements. This feature differs from a standard random walk evolution assumption on the TVPs. Assuming a standard random walk implies smoothly evolving coefficients renders capturing high frequency changes difficult. Interestingly, the time-varying intercept (the fifth coefficients) tends to soak up movements of other parameters. Models that do not truly detect the large breaks of the third coefficient are particularly prone to this issue (TVP-POOL but also standard random walk, TVP-RW, specifications).

6. Empirical application

Structural analysis and forecasting key macroeconomic indicators is of great relevance for policy makers. In the empirical work, we focus on output growth, inflation, unemployment, and the interest rate. Focusing on these variables we investigate the merits of our approach by using the popular quarterly US data described in McCracken and Ng (2016). The data set includes 165 macroeconomic and financial variables and ranges from 1959:Q1 to 2019:Q4. In the empirical application, we start with 1962:Q1 and use the first observations for transformations.

Table 2
Overview of the best performing models, indicated by bold numbers in Table 3 and Table 4.

Variable	1-quarter-ahead			1-year-ahead			2-years-ahead		
	Size	Class	Subclass	Size	Class	Subclass	Size	Class	Subclass
Point forecasts									
RMSE ratios									
TOT	L-VAR	TVP-POOL	SINGLE	L-VAR	TVP-POOL	SINGLE	FA-VAR	TVP-POOL	SINGLE
GDPC1	L-VAR	TVP-POOL	FLEX MIX	FA-VAR	TVP-POOL	FLEX MS	FA-VAR	TVP-MIX	FLEX MIX
CPIAUCSL	S-VAR	TVP-MIX	FLEX MIX	S-VAR	TVP-RW	Primiceri (2005)	S-VAR	TVP-RW	Primiceri (2005)
UNRATE	L-VAR	TVP-POOL	SSVS MIX	L-VAR	TVP-POOL	SINGLE	FA-VAR	TVP-POOL	SSVS MIX
FEDFUNDS	FA-VAR	TVP-RW	SSVS MIX	FA-VAR	TVP-RW	SSVS MIX	FA-VAR	TVP-POOL	SINGLE
Density forecasts									
LPBFs									
TOT	L-VAR	TVP-POOL	FLEX MIX	L-VAR	TVP-POOL	SSVS MIX	L-VAR	const (NG.)	
GDPC1	L-VAR	TVP-POOL	FLEX MS	FA-VAR	TVP-POOL	SSVS MIX	FA-VAR	TVP-POOL	SINGLE
CPIAUCSL	S-VAR	TVP-MIX	SSVS MIX	L-VAR	const. (NG)		L-VAR	const. (NG)	
UNRATE	L-VAR	TVP-POOL	SSVS MIX	L-VAR	TVP-POOL	SSVS MIX	L-VAR	TVP-MIX	SSVS MIX
FEDFUNDS	L-VAR	TVP-POOL	SSVS MIX	FA-VAR	TVP-RW	FLEX MIX	FA-VAR	const. (Min)	

In Sub-section 6.1 we show some stylized in-sample features of our methods for a small-scale model. By including the four target variables in a small-scale VAR (henceforth S-VAR) we present posterior probabilities of the state indicator matrix $\boldsymbol{S}_t$ and estimate the low-frequency relationship between unemployment and inflation (Phillips Curve). For instance, the recent literature highlights the existence of potential non-linear dynamics in both the inflation persistence and the relationship of the Phillips Curve in the US (for a thorough discussion, see Cogley et al., 2010; Ball and Mazumder, 2011; Watson, 2014; Coibion and Gorodnichenko, 2015; Del Negro et al., 2020).

In Sub-section 6.2, this variable set forms the basis for evaluating the predictive performance of our methods in a comprehensive forecast exercise. For the forecasting exercise, we consider two additional information sets. In our largest specification (L-VAR) we pick 20 macroeconomic indicators, which are commonly considered by the recent literature for forecasting (see, for example, Huber et al., 2020a). In particular, we include financial market indicators that carry important information about the future stance of the economy (see Bańbura et al., 2010). In Appendix Appendix C we provide further details on the specific variable set, included in the largest specification, and the transformation applied. Moreover, we consider a factor-augmented VAR (FA-VAR). Here, we augment the target variables with six principal components compromising information of the remaining variables in the data set, effectively leading to VAR with ten endogenous variables. The number of principal components is motivated by the specification in Stock and Watson (2012), who also consider six factors. In such larger scale-models our methods are capable of handling less frequent (but important) parameter instabilities in a genuine way.

Especially forecasting these important macroeconomic aggregates remains a challenging task, since (at least) two issues arise. First, we have to decide on a set of variables, which we want to include in our econometric model. The recent literature on constant parameter VARs highlights that exploiting large information sets yields forecast gains (see, for example, Bańbura et al., 2010; Koop, 2013). Second, it is well documented that important economic indicators feature instabilities in structural parameters and innovation volatilities. In this context, for example, Stock and Watson (2012), Ng and Wright (2013) and Aastveit et al. (2017) put a special emphasis on the recent financial crisis. In the literature there is strong agreement that SV is important in macroeconomic applications (see Clark, 2011). There is also strong empirical support for shifting parameters in small-scale models (see D'Agostino et al., 2013). However, there is less consensus for time-varying parameters in larger-scale models. With increasing amount of information overall time-variation in parameters tends to reduce. Recent contributions dealing with large-scale TVP-VARs argue that in smaller models the TVP part controls for an omitted variable bias (see Feldkircher et al., 2017; Huber et al., 2020b). Since estimating TVP models with typical MCMC methods remains computationally demanding, several studies take this argument as a reason to opt for approximating the TVP part or rely on dimension reduction techniques, yielding fast inference by accepting a certain risk of mis-specification (see, inter alia, Korobilis, 2019; Chan et al., 2020; Hauzenberger et al., 2020; Huber et al., 2020b; Korobilis and Koop, 2020).

In the following empirical application, note that we consider two lags for every model and allow for SV.

6.1. In-sample evidence

Before proceeding, we briefly elaborate on a potential identification problem when interpreting the state indicators $\boldsymbol{S}_t$ (see Frühwirth-Schnatter, 2001). For the TVP-MIX models, identification is ensured by construction (if coefficients indeed feature time variation). Assuming $\boldsymbol{\phi}_t = \boldsymbol{S}_t$ (see Equation 3) automatically imposes inequality constraints on the autoregressive coefficients in the state equation. However, non-identifiability can occur when coefficients are constant. In such a case, elements in $\boldsymbol{S}_t$ are hard to interpret, since a no change evolution is supported by both a random walk and a white noise process. Interpreting $\boldsymbol{S}_t$ for the TVP-POOL specification is an even more challenging task, since in these models $\boldsymbol{S}_t$ solely controls the evolution of state innovations. Here, inference about the state indicator matrix is only useful in combination with inference about the size of state innovation variances $\bar{\boldsymbol{\Psi}}_0$ and $\bar{\boldsymbol{\Psi}}_1$ and with imposing an inequality restriction ex-post (e.g., $\bar{\psi}_{i0} < \bar{\psi}_{i1}$).

Table 3
Point forecast performance (RMSE ratios) relative to the benchmark (`const (Min.)`). The red shaded row denotes the benchmark (and its RMSE values). Asterisks indicate statistical significance for each model relative to `const (Min.)`at the 1 (***), 5 (**) and 10 (*) percent significance levels.

Specification		1-quarter-ahead					1-year-ahead					2-years-ahead				
Class	Subclass	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS
FA-VAR																
const. (Min.)		0.91*	0.87	0.94	0.79	1.22	0.92**	0.84*	1.03	0.80	1.11	0.93*	1.00	1.04	0.86	0.78***
const. (NG)		0.88**	0.81*	0.95	0.80	1.14	0.90**	0.84**	0.99	0.81	1.02	0.92	1.00	1.01	0.87	0.75**
TVP-MIX	FLEX MIX	0.89**	0.80*	0.97	0.81	0.91	0.94*	0.92	0.98	0.88	0.98	1.01	**0.86**	1.10*	1.08	0.95
	FLEX MS	0.91**	0.83	0.98	0.77	0.89	0.91	0.89	1.00	0.80	0.83*	0.94	0.96	1.00	0.92	0.89
	SINGLE	0.92*	0.82	1.01	0.79	0.85**	0.91*	0.87	1.00	0.81	0.86*	0.97	0.93	1.12	0.94	0.87
	SSVS MIX	0.89**	0.83	0.96	0.79	0.95	0.92*	0.89	0.98	0.83	0.98	0.95	0.93	1.09	0.88	0.92
TVP-POOL	FLEX MIX	0.88**	0.79*	0.95	0.77	1.06	0.88**	0.80**	0.99	0.77	0.96	0.89	0.95	1.01	0.82	0.73*
	FLEX MS	0.88***	0.79*	0.96	0.76	1.05	0.87**	**0.80****	0.98	0.77	0.97	0.89	0.96*	1.00	0.82	0.73*
	SINGLE	0.87***	0.79*	0.95	0.77	1.05	0.87**	0.81**	0.98	0.77	0.95	**0.88**	0.94**	1.00	0.82	**0.73***
	SSVS MIX	0.89**	0.81*	0.96	0.76	1.06	0.88**	0.81**	1.00	0.77	0.98	0.89	0.98	1.01	**0.82**	0.73*
TVP-RW	FLEX MIX	0.89**	0.83	0.95	0.77	0.89**	0.88*	0.86	0.97	0.76	0.81**	0.93	1.00	1.00	0.88	0.79
	FLEX MS	0.89**	0.80	0.96	0.80	0.91	0.90*	0.88	0.99	0.79	0.87	0.95	1.00	1.00	0.93	0.83
	SINGLE	0.91*	0.83	0.99	0.81	0.92	1.04	0.99	1.07	1.02	1.23	1.20	0.90	1.24	1.30*	1.36
	SSVS MIX	0.92*	0.84	0.99	0.79	**0.83****	0.90*	0.90	0.96	0.83	**0.80***	0.96	0.99	1.02	0.94	0.83
L-VAR																
const. (Min.)		1.02	0.99	1.05	0.72	1.52	0.92**	0.93	0.95*	0.79	1.00	0.97**	1.02	0.98	0.95	0.88**
const. (NG)		0.91**	0.85	0.96	0.71	1.10	0.88***	0.89*	0.92**	0.74	0.98	0.92**	1.00	0.97	0.89	0.79**
TVP-MIX	FLEX MIX	1.04	0.90	1.14	0.71	2.19	0.91*	0.82	1.02	0.81	1.09	0.92	0.93	0.99	0.89	0.90
	FLEX MS	1.06	0.92	1.17	0.72	1.51	0.90*	0.86	0.95	0.79	1.08	1.22	1.56	1.30	0.90	1.13
	SINGLE	0.92*	0.86	0.98	0.74	0.90**	0.87*	0.82*	0.94	0.83	0.87*	0.92	0.94	0.97	0.91	0.86
	SSVS MIX	0.96	0.98	0.94	0.72	1.20	0.90	0.86	0.98	0.78	1.00	0.95	0.96	1.06	0.91	0.89
TVP-POOL	FLEX MIX	0.88***	**0.79****	0.96	0.68	0.86***	0.87**	0.87	0.94	0.74	0.89**	0.91	0.99	1.01	0.84	0.79**
	FLEX MS	0.88***	0.82*	0.94	0.67	0.87***	0.86**	0.85*	0.94	0.72	0.90**	0.90	0.99	0.99	0.83	0.77**
	SINGLE	**0.87*****	0.79**	0.95	0.68*	0.88***	**0.86****	0.86*	0.93	**0.71**	0.88**	0.91	0.99	0.99	0.84	0.78**
	SSVS MIX	0.88***	0.81*	0.95	**0.67***	0.89***	0.86**	0.85*	0.94	0.72	0.88**	0.90	0.98	1.01	0.83	0.78**
TVP-RW	FLEX MIX	0.96	0.97	0.97	0.71	1.09	0.88*	0.85	0.94	0.80	0.93*	0.95	1.00	0.97	0.94	0.87
	FLEX MS	0.98	0.97	0.99	0.73	0.93**	0.86*	0.81*	0.92	0.84	0.96	0.96	0.97	0.96	0.96	0.93
	SINGLE	1.24	1.09	1.37	0.75	1.75	0.93	0.88	0.97***	0.84	1.23	0.94	0.91	0.93	0.96	0.95
	SSVS MIX	0.94	0.93	0.95	0.73	1.17	0.88*	0.85	0.93	0.82	0.97	0.96	0.95	0.98	0.95	0.95
S-VAR																
const. (Min.)		0.60	0.83	0.85	0.15	0.11	0.76	1.00	0.87	0.62	0.41	0.91	0.93	0.85	1.10	0.73
const. (NG)		0.99**	0.99*	0.99	1.00	1.01	0.96**	0.94*	0.98	0.98*	0.98	0.97	0.99	0.98	0.97**	0.95
TVP-MIX	FLEX MIX	0.93**	0.94*	**0.92**	0.92	0.91***	0.89	0.84	0.94	0.92	0.93	0.93	0.89	1.01	0.93	0.92
	FLEX MS	0.96	0.96	0.96	0.91	0.87	0.93	0.92	0.95	0.91	0.96	0.96	0.97	1.01	0.92	0.95
	SINGLE	0.97	0.97	0.96	0.97	0.85*	0.93	0.91	0.95	0.95	0.97	0.95	0.96	0.99	0.94	0.93
	SSVS MIX	0.95**	0.96	0.94**	0.98	0.94***	0.91*	0.86	0.94	0.93	0.92	0.95	0.94	1.00	0.94	0.92
TVP-POOL	FLEX MIX	0.98***	0.97**	0.99	0.95	0.98	0.95**	0.93*	0.98	0.94	0.94	0.95	0.99	0.96	0.93	0.90
	FLEX MS	0.98**	0.98*	0.98	0.94**	0.98*	0.95**	0.94*	0.98	0.94	0.94	0.95	0.99	0.97	0.93	0.90*
	SINGLE	0.98**	0.98*	0.99	0.94	0.98	0.95**	0.93*	0.99	0.94	0.94	0.95	0.99	0.97	0.93	0.90
	SSVS MIX	0.98**	0.98*	0.99	0.95	0.97*	0.95**	0.93*	0.98*	0.95	0.94	0.94	0.98	0.97	0.92	0.91
TVP-RW	FLEX MIX	0.96**	0.96	0.95	0.97	0.92**	0.93*	0.94	0.91	0.95	0.94	0.96	0.98	0.96	0.95	0.94
	FLEX MS	0.98	0.99	0.97	0.98	0.88*	0.93	0.91	0.93	0.96	0.98	0.95	0.94	0.99	0.95	0.94
	SINGLE	0.96	0.92	1.00	0.92	0.86***	0.89	0.82	0.95	0.94	0.90	0.93	0.88	1.00	0.96	0.85
	SSVS MIX	0.98	1.00	0.96	0.98	0.89*	0.94	0.92	0.93	0.99	0.97	0.97	0.96	0.97	0.97	0.95
	Primiceri (2005)	0.96	1.00	0.93	0.99	0.96	0.89**	0.88	**0.87**	0.96	0.87	0.93	0.92	**0.92**	0.97	0.85

Table 4
Density forecast performance (LPBFs) relative to the benchmark (`const (Min.)`). The red shaded row denotes the benchmark (and its LPS values). Asterisks indicate statistical significance for each model relative to `const (Min.)`at the 1 (***), 5 (**) and 10 (*) percent significance levels.

Specification		1-quarter-ahead					1-year-ahead					2-years-ahead				
Class	Subclass	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS
FA-VAR																
const. (Min.)		11.35	12.01**	2.23	8.56	-4.19	14.59	8.30**	2.02	14.03	8.69	37.98**	2.12	5.63	12.49	**30.86*** **
const. (NG)		18.88	13.48***	4.19	8.66	0.46	25.90***	9.01***	2.34	14.21	10.93	40.82***	3.08	7.01	12.22	28.66***
TVP-MIX	FLEX MIX	23.23	12.71**	2.78	11.15	12.95***	39.12**	7.60	2.33	26.49	16.00	50.99	3.80	3.97	23.44	13.09
	FLEX MS	29.37	11.74**	5.32	13.84	9.59*	49.54**	7.49	2.32	27.51	17.64	41.96	0.10	3.65	22.67	13.22
	SINGLE	24.87	11.07**	2.24	10.12	17.01***	52.21**	6.25	0.25	25.16	24.39	55.19	0.70	3.17	22.70	18.79
	SSVS MIX	25.93	13.23**	2.59	10.60	6.38*	40.29*	7.92*	1.90	23.53	10.64	40.57	3.38	1.82	25.83	7.45
TVP-POOL	FLEX MIX	26.23	13.49***	3.44	10.51	7.28***	39.69***	10.63**	2.06	19.71	15.86	53.01**	3.61	6.18	22.98	28.67*
	FLEX MS	28.06	14.49***	3.98*	10.24	7.62***	34.75**	10.85***	2.71	17.29	14.71	52.55**	4.07	6.27	21.13	28.34*
	SINGLE	25.63	14.42***	2.98	10.61	6.20***	34.67**	10.54***	2.21	18.32	16.06	46.04**	**4.36**	6.22	19.52	26.47
	SSVS MIX	25.54	14.08***	3.17*	10.06	6.84***	35.28**	**11.02*** **	1.59	18.01	15.29	50.57**	4.04	6.74	23.07	28.09
TVP-RW	FLEX MIX	37.26	11.40**	7.35**	12.30	17.50***	63.22**	7.80	1.93	29.72	**27.66**	69.42*	-1.76	5.94	27.99	29.14
	FLEX MS	26.44	11.88***	5.01	11.44	11.10**	53.18**	7.23	0.70	24.70	20.65	38.45	-0.47	4.42	21.35	18.47
	SINGLE	20.52	11.81**	-2.21	11.01	13.10***	46.68**	4.46	2.24	19.20	17.68	45.86	-1.09	-0.24	13.83	16.93
	SSVS MIX	38.78	11.12**	6.46	9.69	16.94***	47.25***	6.07	0.96	21.05	27.08	47.06*	-1.64	5.28	17.38	25.21
L-VAR																
const. (Min.)		15.05	13.32*	-0.29	12.15	-3.68***	56.73**	2.78	3.22***	32.41	11.78***	61.59***	-3.86	3.18	15.52	18.28**
const. (NG)		28.70	15.97***	1.06	14.87	2.34	70.27***	7.40**	**6.76*** **	36.48	16.42*	**73.50*** **	0.33	**8.80***	17.46**	21.68***
TVP-MIX	FLEX MIX	21.50	13.39**	-3.42	14.97	6.06	57.99*	7.63	-0.49	36.44	6.38	50.44	-3.40	-0.52	27.14	-1.18
	FLEX MS	13.65*	10.32*	-0.35	13.39	7.88	46.02	6.19	-0.16	32.55	7.15	44.48	-6.69	-0.73	28.69	0.75
	SINGLE	28.03	12.75**	1.04	13.65	12.62	60.71*	8.02	2.15	32.82	12.23	60.21	-3.28	1.74	28.66	7.41
	SSVS MIX	16.40*	11.52**	1.00	12.90	11.67	55.25*	8.09	-0.77	34.05	9.97	41.64	-0.87	1.23	**31.10**	1.53
TVP-POOL	FLEX MIX	**54.32**	16.63***	0.86	16.78	22.62***	71.86***	8.08*	3.09	35.86	24.24	63.16***	0.27	2.73	22.08	21.52**
	FLEX MS	50.88	**17.28*** **	1.72	16.95	21.29***	70.43**	8.02**	2.27	36.55	24.52	65.07***	-0.06	2.43	20.10	22.50**
	SINGLE	51.15	16.22***	1.44	16.75	22.24***	72.35***	8.14*	2.47	36.30	24.31	60.97***	-0.50	2.74	24.76	20.51
	SSVS MIX	49.90	16.39***	-0.50	**17.15**	**23.32*** **	**73.08** **	6.80*	2.06	**36.87**	25.34	63.73***	-1.29	2.69	19.82	22.33*
TVP-RW	FLEX MIX	29.38	10.54**	1.22	11.09	12.63	54.24	3.78	1.28	27.88	17.50	59.54	-7.53	-0.86	28.15	15.32
	FLEX MS	15.46	9.99**	-8.43	11.68	13.28	55.15*	5.85	1.40	28.01	13.82	41.09	-5.25	-0.64	25.89	6.06
	SINGLE	12.39	12.04**	-1.38	12.34	7.23	41.96	4.35	0.38	28.51	-2.50	28.63	-9.08*	0.76	20.03	-6.91
	SSVS MIX	18.47	9.93*	0.81	11.50	13.63	55.11	4.67	2.12	27.41	13.44	46.83	-3.83	1.76	23.01	6.32
S-VAR																
const. (Min.)		-22.11	-82.64	-80.74	45.02	86.64	-256.48	-97.26	-83.70	-69.04	-29.94	-383.84	-92.27	-89.84	-125.53	-87.85
const. (NG)		4.55	1.97*	0.71	0.62*	1.39***	4.97	2.64*	1.95*	0.03	2.30	9.15	0.58	3.07	2.99	1.37
TVP-MIX	FLEX MIX	24.41	3.97*	5.52	2.90	12.18***	15.60	9.64**	3.98	8.50	5.03	-3.76	3.00	3.74	-0.02	-7.71
	FLEX MS	18.00	2.66	3.57	6.42	5.86**	9.48	5.25*	-0.44	12.27	-0.22	0.77	-1.60	0.83	10.35	-11.26
	SINGLE	7.18**	2.45	-1.98	0.85	13.26***	12.51	5.42	3.33	6.38	7.12*	-0.80	-2.74	1.19	2.14	-5.03
	SSVS MIX	23.21	3.44*	**8.67** **	2.67	7.84***	22.03	6.18	1.77	10.22	5.15	15.68	1.32	6.23	14.46	-3.63
TVP-POOL	FLEX MIX	16.39	2.47***	2.63**	1.29	6.18***	9.34	3.10*	1.70	3.87*	5.15	14.86	0.07	3.60*	6.37	0.56
	FLEX MS	14.35	2.30**	1.26*	1.66	7.91***	11.74	3.65***	2.04**	7.55*	6.02	14.99	-0.16	4.04**	8.28	1.86
	SINGLE	16.46	2.68***	2.61*	1.84	6.99***	5.61	3.52**	1.73	3.43	5.59	14.77	0.21	3.13	10.10	-0.22
	SSVS MIX	18.75	2.40**	2.84*	1.84	7.03***	11.67	3.66**	2.70***	4.44**	5.74	12.63	0.51	3.68**	7.64	0.59
TVP-RW	FLEX MIX	22.15	2.35	5.20***	2.14	11.56***	31.13*	4.67	6.61*	10.96	9.42	23.66	-2.07	7.42	11.93	2.23
	FLEX MS	17.02	1.76	2.62	1.98	8.41**	10.37	4.80	1.87	8.30	2.86	7.89	-2.64	4.15	3.79	-7.05
	SINGLE	13.68	6.02**	-3.98	1.79	12.17***	6.72	8.60	1.15	1.63	13.13	-11.49	2.34	0.87	-7.78	3.09
	SSVS MIX	20.25	1.26	5.76*	0.31	12.10***	28.93*	5.73	5.24	6.34	9.14	16.00	-2.42	6.40	4.74	-2.84
	Primiceri (2005)	0.56	2.73	2.02	5.58	-10.87***	10.79	4.76	-1.76	10.58	-2.41***	2.45	-0.53	2.58	6.58	2.67

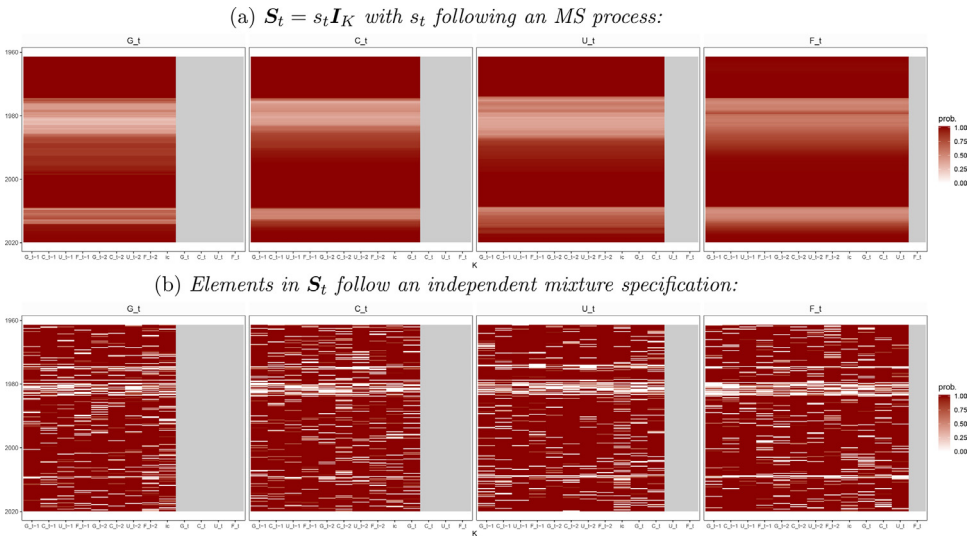


Fig. 2. Posterior distribution of s_{it}, for $i = \{1, \dots, K\}$, for two small-scale TVP−MIX models. Here, G denotes output growth (GDPC1), C the inflation (CPIAUCSL), U the unemployment rate (UNRATE), F refers to the interest rate (FEDFUNDS) and ic to an intercept. Moreover, the structural form in Equation 11 implies that some parameters are not part of the i^{th} equation (denoted by grey shaded areas), due to the strictly lower triangular structure of $\boldsymbol{B}_{0t}$.

Therefore, we solely focus on two variants of a TVP−MIX model to illustrate the switching behavior. Figure 2 depicts the posterior median of the diagonal elements in $\boldsymbol{S}_t$. Panel (a) shows a TVP−MIX model with $\boldsymbol{S}_t = s_t \boldsymbol{I}_K$ and s_t following a first-order Markov process (MS). Panel (b) depicts a specification with elements in $\boldsymbol{S}_t$ following an independent mixture distribution (MIX). A comparison between both approaches highlights that a joint indicator evidently leads to a different posterior median of $\boldsymbol{S}_t$ than covariate-specific indicators. By restricting $\boldsymbol{S}_t = s_t \boldsymbol{I}_K$, all covariates are driven solely by a single indicator that pushes all covariates towards either a random walk or white noise state equation in period t. Conversely, with covariate-specific indicators, we see more dispersion across covariates. However, both approaches agree on a white noise state equation in times of turmoil, suggesting a need for abruptly adjusting parameters in these periods. This model feature is in line with the discussion in Primiceri (2005), who suggests that an economically stable period favours more gradual changes (which are more consistent with a random walk state equation) in the coefficients, while shifts in policy rules require quickly adjusting coefficients (which is better captured by using a white noise state equation).

To further illustrate the proposed methods, we estimate the low-frequency relationship between unemployment and inflation. This low-frequency measure corresponds to a long-run coefficient of distributed-lag regression models (Whiteman, 1984) and disentangles systematic co-movements from short-run fluctuations. Sargent and Surico (2011) and Kliem et al. (2016) suggest that a TVP-VAR framework, additionally, allows to account for changes in the transmission channels (time-varying coefficients) and changes in the error volatilities (SV). Further details can be found in Appendix A.4.

Panel (a) to (c) in Figure 3 depict the obtained low-frequency component with our proposed approaches and panel (d) shows estimates with a standard TVP model assuming a standard random evolution assumption. Starting with a comparison between the random walk/white noise mixture (TVP−MIX) and a classic random walk TVP model, we observe a similar pattern for both approaches during tranquil periods. However, during recessions, the approaches significantly differ. Both TVP−MIX models are capable of detecting a major structural break in the low-frequency relationship after the oil crisis in the 1970s and strongly support a long-lasting stagflation period (i.e., positive relationship between unemployment and inflation). While TVP−MIX methods are designed to quickly capture these large abrupt breaks in parameters, a standard random walk state equation translates into a low-frequency component that gradually adapts over time. However, the TVP−MIX model with covariate-specific indicators MIXis slightly more sensitive with respect to abrupt changes in parameters than the TVP−MIX MS model.

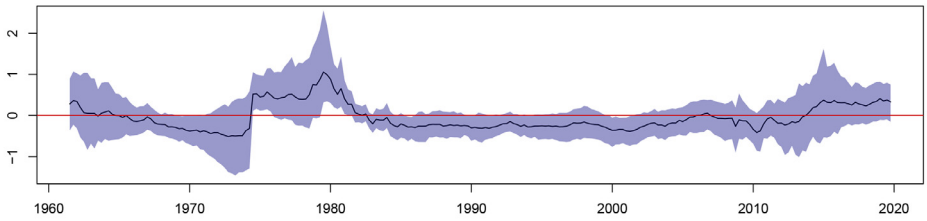
Panel (c) shows the sparse scale-location mixture (TVP−POOL) approach with covariate-specific indicators (MIX). We observe that this method almost resembles a constant coefficient specification with SV. In the mid 1980s and in the financial crisis movement in the low-frequency relationship is slightly more erratic compared to other periods, but it stays mostly constant and significant.

Overall, considering TVP−MIX methods seem to improve the economic interpretability of the low-frequency component, while a TVP−POOL model aggressively pushes coefficients towards a constant evolution, which could pay off for forecasting.

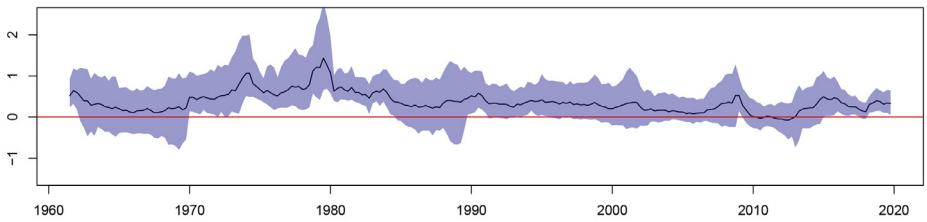
6.2. Forecasting evidence

In the forecast exercise we consider a wide range of models varying along the evolution assumption of parameters and the information set considered.

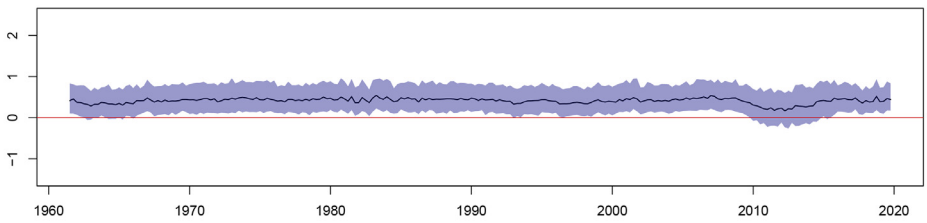
(a) *TVP-MIX with flexible state variances (FLEX) and* $\boldsymbol{S}_t = s_t \boldsymbol{I}_K$ *(MS)*:



(b) *TVP-MIX with flexible state variances (FLEX) and covariate-specific indicators (MIX)*:



(c) *TVP-POOL with flexible state variances (FLEX) and covariate-specific indicators (MIX)*:



(d) *Standard TVP-VAR with random walk state equation:*

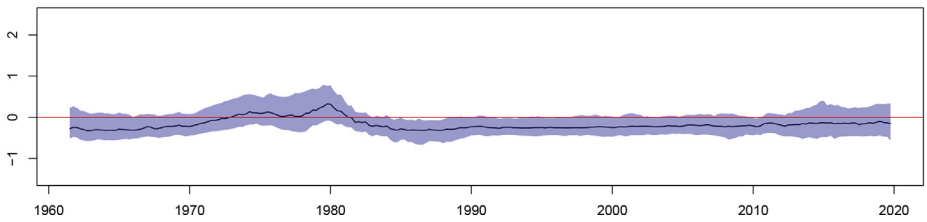


Fig. 3. Low-frequency relationship between the unemployment rate and the inflation. The blue line refers to the posterior median, while the blue-shaded area indicates the 68% posterior credible set. The red line indicates zero.

With respect to the evolution of parameters, it proves convenient to summarize the different specifications (see Table 1). The models differ along three dimensions: the autoregressive parameters $\boldsymbol{\phi}_t$, the innovation variances $\boldsymbol{\Psi}_t$ and the state indicator matrix $\boldsymbol{S}_t$. First, our main specifications vary between a model that assumes a binary indicator matrix on the autoregressive parameter with $\boldsymbol{\phi}_t = \boldsymbol{S}_t$ (labeled as TVP-MIX) and a model that introduces a hierarchical prior on the TVP-part (TVP-POOL). For the latter we implicitly assume that $\boldsymbol{\phi}_t = \boldsymbol{0}_{K \times K}$, for all t, in Equation 5. Regarding the autoregressive parameter, a natural competing model is a standard random walk assumption with $\phi_t = \boldsymbol{I}_K$, for all t (TVP-RW). Second, the models differ in the treatment of the state innovation variances. The most flexible innovation variance specification does not restrict $\bar{\boldsymbol{\Psi}}_0$ and $\bar{\boldsymbol{\Psi}}_1$ (labeled as FLEX), a second specification assumes $\bar{\boldsymbol{\Psi}}_0 = \kappa \bar{\boldsymbol{\Psi}}_0$ (SSVS), while the most restrictive

specification fixes $\boldsymbol{\Psi}_t = \bar{\boldsymbol{\Psi}}$, for all t, to a a single variance state (SINGLE) with $\hat{\boldsymbol{\alpha}}$ collapsing to a $2K$-dimensional vector (see Bitto and Frühwirth-Schnatter, 2019). In the empirical exercise, we set $\kappa = 0.1^6$ and $\hat{\boldsymbol{\Psi}}_0 = \text{diag}\ (\hat{\psi}_1, \dots, \hat{\psi}_K)$ with $\hat{\psi}_i$, for, $i = \{1, \dots, K\}$, denoting ordinary least square (OLS) variances obtained from an AR(p) model (see Huber et al., 2019). Third, with regards to the state indicator matrix $\boldsymbol{S}_t$, we discriminate between a joint Markov-switching indicator (labeled as MS) and covariate-specific indicators following an independent mixture distribution (MIX). Recall, that for the TVP-MIX models $\boldsymbol{S}_t$ adjusts both the autoregressive parameters and the state innovation variances, while for the TVP-POOL and TVP-RW models $\boldsymbol{S}_t$ only controls the state innovations. In the following, we define TVP-MIX, TVP-POOL and TVP-RW as the Class of the TVP model and the combination of the acronyms for the innovation variances and indicator matrix as the Subclass of the specification. A single model is identified by a combination of all three acronyms. For example, a TVP-MIX FLEX MIX specification denotes a model with a random walk/white noise mixture for the state equation, with unrestricted two-state variances and with the elements in $\boldsymbol{S}_t$ following an independent mixture distribution.

All these TVP models feature a Normal-Gamma (Griffin and Brown, 2010) on $\hat{\boldsymbol{\alpha}}$. We compare our methods to two constant parameter specifications and a traditional small-scale TVP-VAR model, where we use exactly the same prior set-up as in Primiceri (2005). A constant coefficient model can be obtained by either offsetting $\tilde{\boldsymbol{\alpha}} = \mathbf{0}_{\nu \times 1}$ or setting $\{\boldsymbol{\Psi}_t\}_{t=1}^T \approx \mathbf{0}_{K \times K}$. One variant features a Normal-Gamma (const.(NG)) prior, while the second variant assumes a Minnesota (const.(MIN)) prior. We consider a non-conjugate Minnesota prior, capturing the notion that own lags are more important than lags from other variables (Doan et al., 1984; Litterman, 1986). We estimate this set of models for three information sets (FA-VAR, L-VAR and S-VAR) with each featuring a different number of endogenous variables. The Primiceri (2005) TVP-VAR is an alternative variant of the TVP-RW SINGLE model. In brief, this specification can be obtained by using pre-specified priors that are calibrated with OLS estimates on an initial training sample of ten years (instead of automatic shrinkage processes), assuming a full state innovation variance-covariance matrix $\bar{\boldsymbol{\Psi}}$ (instead of diagonal $\bar{\boldsymbol{\Psi}}$) and considering random walk SV (instead of AR(1)-SV). This model is considered exclusively for the small information set, since in larger scale specifications the proposed prior set-up would either lead to severe overfitting issues or involve a substantial amount of tuning. Similar to Primiceri (2005) and D'Agostino et al. (2013), every considered specification features two lags and SV.

To assess one-quarter-, one-year- and two-years-ahead predictions, we treat observations ranging from 1962:Q1 to 1999:Q4 as an initial sample and the periods from 2000:Q1 to 2019:Q4 as a hold-out sample. The initial sample is then recursively expanded until the penultimate quarter (2019:Q3) is reached. For each forecast comparison, a small-scale Minnesota VAR with constant parameters (S-VAR const.(MIN)) serves as our benchmark. In the following, Table 2 shows the best performing models for point and density forecasts, effectively being a tractable summary of Table 3 and Table 4. Table 3 depicts root-mean squared error ratios (RMSEs) as point forecast measures and Table 4 the log predictive Bayes factors (LPBFs) as density forecast metrics. In Table B.5 we provide additional results on continuous rank probability score (CRPS) ratios. This alternative density forecast measure is more robust to outliers than log predictive scores (Gneiting and Raftery, 2007). In each table the best performing models within each column are indicated by bold numbers. With three different measures at three different horizons we obtain a comprehensive picture to evaluate our methods jointly and marginally along the four target variables.

Table 2 summarizes the main findings of our forecast exercise. *First*, larger-scale models (FA-VAR, L-VAR) generally outperform the small-scale specifications across horizon-variable combinations, indicating that an increasing amount of information pays off for forecasting (see Bańbura et al., 2010). One exception is inflation. For inflation, S-VARs featuring TVPs yield more accurate forecasts than FA-VARs and L-VARs across all horizons for point forecasts and the one-quarter-ahead horizon for density forecasts. Comparing FA-VARs with L-VARs, the results are mixed. One pattern worth noting is that L-VARs tend to outperform FA-VARs for the one-quarter-ahead horizon while the picture reverses for higher-order forecasts. Second, with respect to parameter changes we see that the TVP-POOL specifications forecast particularly well across all horizons and target variables. These models substantially improve upon a wide range of benchmarks. Overall, Table 2 shows that all TVP classes that provide accurate point predictions generally also perform well in terms of density forecasts.

When examining Table 3 and Table 4 in greater detail, note that a large number of models shown in the tables outperform the Minnesota benchmark in terms of RMSEs (indicated by ratios below one) and in terms of LPBFs (indicated by values above zero). However, the benchmark is a tough competitor when predicting inflation and for higher-order point forecasts. In particular, the result that inflation is hard to predict is a well-known fact in the literature (Stock and Watson, 2007). Commonly, it is found here that more sophisticated models with large information sets are outperformed by rather small-sized and comparatively simpler models (see, for example, Giannone et al., 2015). In particular, the small-scale Primiceri (2005) VAR produces sensible point forecasts for inflation. This finding is much in line with D'Agostino et al. (2013), who provide a thorough study on forecasting with small-scale TVP-VARs.

When focusing on the differences occurring through the varying treatment of parameter evolutions, our proposed methods, the TVP-POOL and TVP-MIX specifications, show that their good performance is mainly driven by improved forecast accuracy for output growth and unemployment. In terms of the innovation variance assumption for these specifications, we observe that additional flexibility tends to improve density forecasts performance and yields accurate point forecasts. For the TVP-POOL models this higher degree of flexibility generally pays off across variables and model sizes. For flexible TVP-MIX specifications, forecast ability tends to improve for S-VARs and FA-VARs and is competitive for the L-VARs. Especially the TVP-MIX SSVS MIX and TVP-MIX FLEX MIX models using a small information set yield quite accurate

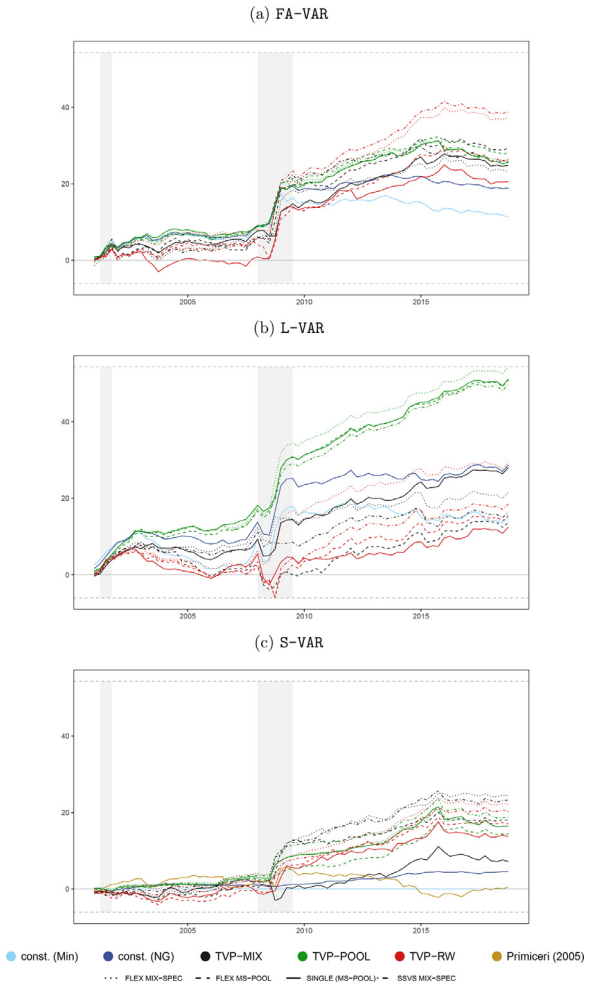


Fig. 4. Evolution of one-quarter-ahead total cumulative LPBFs relative to the benchmark. The gray dashed lines refer to the maximum/minimum Bayes factor over the full hold-out sample. The light gray shaded areas indicate the NBER recessions in the US.

inflation forecasts, being the best performing models for the one-quarter-ahead horizon. Across variables, a notable exceptions is the interest rate for TVP-MIX models. Here, a TVP-MIX SINGLE specification is superior to models assuming a mixture on innovation volatilities.

When assessing random walk state equation (TVP-RW) specifications across the information sets, two things are worth noting. First, a standard TVP model with random walk assumption TVP-RW SINGLE is mostly competitive for one-year- and two-years-ahead forecasts and otherwise forecasts poorly. For the Primiceri (2005) VAR, as an alternative small-scale variant of the TVP-RW SINGLE model, we observe a quite similar pattern, although a few differences in performance are notable. The Primiceri (2005) VAR produces more sensible forecasts for inflation. Apart from that, we find that this model tends to be outperformed for both point and density forecasts by the (main) small-scale TVP-RW SINGLE variant with shrinkage priors (especially at the one-quarter-ahead horizon). We conjecture that these differences in performance may originate from the fact that the Primiceri (2005) model assumes non-adaptive priors and a full state innovation variance-covariance matrix. While for the TVP-RW SINGLE specification the diagonal elements in $\sqrt{\tilde{\mathbf{\Psi}}}$ are pushed towards zero with a Normal-Gamma shrinkage prior, the Primiceri (2005) VAR uses an Inverse-Wishart prior on $\tilde{\mathbf{\Psi}}$, effectively leading to a dense state innovation variance-covariance matrix.

Second, more flexible TVP-RW variants produce quite accurate forecasts for FA-VARs. Constant parameter models with a Normal-Gamma prior show reasonable forecasts for L-VARs (especially for inflation), but lack flexibility in smaller-scale models. This observation is in line with the fact that in larger-scale model, time-variation in coefficients vanishes (see Huber et al., 2020b). However, few parameter instabilities might still be present since, apart from some exceptions, our methods provide improvements when compared to constant coefficient models.

To illustrate the forecast performance over time, Figure 4 depicts the evolution of cumulated joint LPBFs relative to our benchmark. Overall we find that for all four target variable jointly our proposed methods never forecast poorly. Both TVP-MIX (black lines) and TVP-POOL (green lines) methods outperform a standard TVP model with random walk state

equation (TVP−RW SINGLE denoted by the red solid line) across information sets (one exception is the TVP−MIX SINGLE model for the S−VAR). Moreover, allowing for occasional parameter changes during and in the aftermath of financial crisis tends to increase predictive ability.

A few points are worth discussing in greater detail. First, at the beginning of the hold-out sample is characterized by the early 2000s recession. Although this was a quite short crisis, it already leads to a quite diverse model performance across information sets. During this episode and its consecutive three years, L−VARs strictly dominate the other two information sets (FA−VARs and S−VARs). This implies, that for any TVP evolution assumption, the large-scale model outperforms its smaller-scale counterparts. Moreover, during the financial crisis, we observe a substantial increase in LPBFs for a wide range of FA−VARs and L−VARs, while for S−VARs we see similar improvements solely for some TVP-VARs. This feature might indicate that TVPs are capable of mitigating a potential omitted variable bias (Huber et al., 2020b).

Second, within each information set, performance across parameter evolution assumptions is mixed. Evidently, performance of the four L−VARs featuring a TVP−POOLspecification stands out (depicted by green lines). In more tranquil periods they show constant improvements and yield substantial predictive gains during the financial crisis. Especially in the aftermath of the Great Recession, the LPBFs steadily increase compared to other large TVP-VARs. This episode also includes a time characterized by a (sluggish) recovery of the US economy after the financial crisis and the interest rate hitting the zero lower bound. With monetary policy shifting towards unconventional measures it might not only pay off to include financial market variables, such as longer-term yields, but also to allow for occasional changes in transmission channels in these variables. Moreover, it is worth noting that the four TVP−POOL variants tend to perform almost identical until 2010, while afterwards slight performance differences become evident. Hauzenberger et al. (2019) have made a similar observation, when varying the hyperparameters on their conjugate prior of the state equation.

Third, TVP−MIX methods generally forecast well for S−VARs and FA−VARs , while for L−VARs only the TVP−MIX SINGLE specification yields substantial gains. In particular for FA−VARs and S−VARs, a flexible variance modelling (TVP−MIX FLEX MIX and TVP−MIX SSVS MIX) generally pays off. For L−VARs these two models also yield reasonable forecast accuracy, while the TVP−MIX MS forfeits forecast accuracy. Thus, for a large-scale model, the assumption that a joint indicator governs the evolution of large number of coefficients might be less appropriate. Moreover, comparing TVP−MIX with TVP−RW specifications reveals that the random walk/white noise mixture yields gains in tranquil periods for larger-scale models (FA−VARs and L−VARs) and does particularly well in recessions for the small information set (S−VAR). Especially for small-scale VARs the TVP−MIX variants, featuring a mixture distribution on the state innovation volatilities, greatly improve predictive performance relative to the Primiceri (2005) TVP-VAR and all other TVP−RW models during the financial crisis. It is worth noting, nonetheless, that relative to constant parameter specifications the Primiceri (2005) model shows marginal gains in some episodes prior to as well as during the Great Recession, but exhibits a steady deterioration in forecast performance in its aftermath. The TVP−RW SINGLE model forecasts poorly for larger-scale models (FA−VARs and L−VARs) during tranquil periods previous to the financial crisis, while performance slightly recovers in the middle of the Great Recession. A plausible explanation for this pattern might be that spurious movements in coefficients lead to overfitting, widening the predictive density of the TVP−RW SINGLE model. This feature is harming predictive accuracy in stable times, while it is to some extent helpful in times of turmoil (periods characterized by large outliers). In contrast to the TVP−RW SINGLE the other three flexible TVP−RW variants forecast particularly well, suggesting that flexible state innovation volatilities greatly increase precision for TVP coefficients. In particular for FA−VARs these models show improved forecast accuracy after the Great Recession.

7. Closing remarks

It is empirically well documented that macroeconomic time series feature instabilities in the parameters and innovation volatilities. In the literature there is strong agreement that stochastic volatility is important, while, especially in larger-scale models, there is less consensus for time-varying coefficients. With increasing amount of information overall time-variation in parameters tends to reduce, but might be still present at few points in time for some parameters. Detecting such occasional changes is challenging and requires highly flexible modeling techniques. To achieve such flexibility we introduce mixture priors on the time-varying part of the parameters. By additionally using hierarchical shrinkage priors on dynamic state variances, these methods are capable in imposing dynamic sparsity, as well as capturing a wide range of parameter changes. In a simulation study we show that our methods detect both sudden and gradual changes in parameters. In an empirical exercise we find that some coefficients tend to change abruptly in times of turmoil. Moreover, all proposed approaches forecast well. Even for large VARs flexible mixture priors improve forecast accuracy upon a wide range of benchmarks, suggesting that capturing these infrequent instabilities pays off.

Acknowledgments

The author gratefully acknowledges financial support from the Austrian Science Fund (FWF, grant no. ZK 35) and the Oesterreichische Nationalbank (OeNB, Anniversary Fund, project no. 18127). I would like to thank Erricos J. Kontoghiorghes, the associate editor, two anonymous referees, Sylvia Frühwirth-Schnatter, Paul Hofmarcher, Florian Huber, Karin Klieber, Gary Koop, Luca Onorante, Michael Pfarrhofer, Aubrey Poon and Anna Stelzer for valuable comments and suggestions that substantially improved the quality of the paper.

Appendix A. Technical appendix

A1. Stochastic volatility specification

A stochastic volatility specifications assumes that $h_t = \log(\sigma_t^2)$ follows an AR(1)-process:

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \vartheta_t, \quad \vartheta_t \sim \mathcal{N}(0, \psi_h). \tag{A.1}$$

Following Kastner and Frühwirth-Schnatter (2014), we assume Gaussian priors on the initial state $h_0 \sim \mathcal{N}\left(\mu_h, \frac{\psi_h}{1-\phi_h^2}\right)$ and the unconditional mean $\mu_h \sim \mathcal{N}(0, 100)$, a Beta prior on the autoregressive parameter $\frac{\psi_h+1}{2} \sim \mathcal{B}(25, 1.5)$ and a Gamma prior on the state variance $\psi_h \sim \mathcal{G}(1/2, 1/2)$. This quite informative prior on ψ_h pushes the specification towards a random walk.

A2. The Normal-Gamma prior (Griffin and Brown, 2010)

Similar to Bitto and Frühwirth-Schnatter (2019), we introduce class-specific global shrinkage parameters, differentiating between the constant part of the coefficients (labeled λ_a) and regime-switching variances (labeled λ_{ψ_0} and λ_{ψ_1}, respectively). In the following, specify $\tau_j|\lambda_j \sim \mathcal{G}(\varrho_j, \varrho_j\lambda_j/2)$ and $\lambda_j \sim \mathcal{G}(\zeta, \zeta)$ with $\lambda_j = \lambda_k$ and $\varrho_j = \varrho_k$ if $j \in P_k$ for $k = \{a, \psi_0, \psi_1\}$. P_k denotes a classifier (i.e. defines the set of coefficients belonging to the k^{th} group). In the following,

$$P_a = \quad \left\{j : \hat{\alpha}_j \in \boldsymbol{\alpha}_0\right\}, \quad P_{\psi_0} = \quad \left\{j : \hat{\alpha}_j \in \left\{\sqrt{\bar{\psi}_{i0}}\right\}_{i=1}^{K}\right\}, \quad \text{and} \quad P_{\psi_1} = \quad \left\{j : \hat{\alpha}_j \in \left\{\sqrt{\bar{\psi}_{i1}}\right\}_{i=1}^{K}\right\}.$$

Moreover, we learn the hyperparameter ϱ_k in a fully Bayesian fashion and specify $\zeta = 0.01$.

A3. Detailed MCMC algorithm

In this section, we provide details on each sampling step of the MCMC algorithm and on the full conditional posterior distributions. After defining appropriate starting values, we iterate through the following steps 20,000 times and discard the first 10,000 draws as burn-in:

1. The sampling steps (and conditional posteriors) for $\hat{\boldsymbol{\alpha}}_t$, λ_k, τ_j, for $k = \{a, \psi_0, \psi_1\}$ and $j = 1, \dots, 3K$ and ϱ_k are of standard form (Griffin and Brown, 2010):
   (a) Draw $\hat{\boldsymbol{\alpha}}$ from a multivariate Gaussian distribution:
   $$\hat{\boldsymbol{\alpha}}|\boldsymbol{y}, \hat{\boldsymbol{X}}, \boldsymbol{\Sigma}, \{\tau_j\}_{j=1}^{3K} \sim \mathcal{N}\left(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{V}}_1\right).$$
   Here, $\hat{\boldsymbol{X}}$ is a $T \times 3K$-dimensional matrix with $\hat{\boldsymbol{x}}_t'$ on the t^{th} position and:
   $$\hat{\boldsymbol{V}}_1^{-1} = \left((\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{X}})'(\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{X}}) + \text{diag}\ (\tau_1^{-1}, \dots, \tau_{3K}^{-1})\right),$$
   $$\hat{\boldsymbol{\alpha}}_1 = \hat{\boldsymbol{V}}_1\left((\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{X}})'(\boldsymbol{\Sigma}^{-1}\boldsymbol{y})\right).$$
   (b) Sample the local shrinkage scalings $\{\tau_j\}_{j=1}^{3K}$ from a generalized inverse Gaussian (GIG) distribution (Griffin and Brown, 2010):
   $$\tau_j|\hat{\alpha}_j, \lambda_j, \varrho_j \sim \text{GIG}\left(\varrho_j - \frac{1}{2}, \varrho_j\lambda_j, \hat{\alpha}_j^2\right), \text{ for } j = \{1, \dots, 3K\}.$$
   Here, the $\text{GIG}(a, b, c)$ is parameterized as $p(x) \propto x^{a-1}\exp\{-(bx + c/x)/2\}$, $\lambda_j = \lambda_k$ and $\varrho_j = \varrho_k$ if $j \in P_k$ with $k = \{a, \psi_0, \psi_1\}$.
   (c) Sample the associated global shrinkage parameter λ_k, for $k = \{a, \psi_0, \psi_1\}$, from a Gamma distribution distribution:
   $$\lambda_k|\{\tau_j\}_{j\in P_k}, \varrho_k \sim \mathcal{G}\left(\zeta + \varrho_k p_k, \zeta + \frac{\varrho_k}{2}\sum_{j\in P_k}\tau_j\right)$$
   with p_k denoting the cardinality of the set P_k (see Appendix A.2).
   (d) The hyperparameter ϱ_k, for $k = \{a, \psi_0, \psi_1\}$, are updated with a random walk Metropolis Hastings (MH) step. We refer to Bitto and Frühwirth-Schnatter (2019) for details.
2. Draw the normalized latent states $\tilde{\boldsymbol{\alpha}}$ from a ν-dimensional Gaussian distribution by exploiting the static representation (see Section 4).
3. Draw time-varying volatilities $\boldsymbol{\Sigma}$ using the R package `stochvol` (Kastner, 2016).
4. Update binary indicators in $\boldsymbol{S}_t$, depending on its law of motion. We recast state equation back in the centered parameterization and evaluate the following regime-switching specification:
   $$\boldsymbol{\alpha}_t = \begin{cases} \boldsymbol{\alpha}_0 + \boldsymbol{\gamma}_t + \bar{\boldsymbol{\phi}}_0(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_0) + \boldsymbol{\varsigma}_t, & \boldsymbol{\varsigma}_t \sim \mathcal{N}(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}_0) \quad \text{if} \quad s_t = 0, \\ \boldsymbol{\alpha}_0 + \boldsymbol{\gamma}_t + \bar{\boldsymbol{\phi}}_1(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\alpha}_0) + \boldsymbol{\varsigma}_t, & \boldsymbol{\varsigma}_t \sim \mathcal{N}(\boldsymbol{0}, \bar{\boldsymbol{\Psi}}_1) \quad \text{if} \quad s_t = 1, \end{cases} \tag{A.2}$$
   with $\bar{\boldsymbol{\phi}}_0 = \boldsymbol{0}_{K\times K}$, $\bar{\boldsymbol{\phi}}_1 = \boldsymbol{I}_K$ and $\boldsymbol{\gamma}_t = \boldsymbol{0}_{K\times 1}$ for the `TVP-MIX` model. For the `TVP-POOL` model we set $\bar{\boldsymbol{\phi}}_0 = \bar{\boldsymbol{\phi}}_1 = \boldsymbol{0}_{K\times K}$.

- s_t follows a first-order MS process (MS):
  (a) Conditional on the other parameters in Equation A.2, we follow Kim and Nelson (1999) and sample $\{s_t\}_{t=1}^T$ using standard algorithms.
  (b) Conditional on $\{s_t\}_{t=1}^T$ we update transition probabilities by sampling $p_{00} \sim \mathcal{B}(T_{00}+c_{00}, T_{01}+c_{10})$ and $p_{11} \sim \mathcal{B}(T_{11}+c_{01}, T_{10}+c_{11})$ both from a Beta distribution with T_{kl}, denoting the number of transitions from the k^{th} to the l^{th} regime.
- Covariate-specific indicators with $\{s_{it}\}_{i=1}^K$ independent over time (MIX):
  (a) Conditional on the other parameters in Equation A.2 we evaluate both regimes in Equation A.2 and sample s_{it} for each period and covariate independently from a Bernoulli distribution.
  (b) Conditional on $\{s_{it}\}_{t=1}^T$ we are able to update the success probability for each covariate by sampling from a Beta distribution $p_i \sim \mathcal{B}(T_{i,0}+c_{i,0}, T_{i,1}+c_{i,1})$, for $i=\{1,\dots,K\}$, with $T_{i,k}$ denoting the number of periods in the k^{th} regime.

5. For the specification with a hierarchical prior on $\tilde{\boldsymbol{\gamma}}_t$ and $\boldsymbol{\phi}_1 = \boldsymbol{\phi}_0 = \mathbf{0}_{K\times K}$, we need five additional sampling steps (details can be found in Malsiner-Walli et al. (2016) and Hauzenberger et al. (2019)):
   (a) Draw mixture weights $\boldsymbol{\omega}$ from a Dirichlet distribution:

$$\boldsymbol{\omega}|\boldsymbol{\theta}, \xi \sim \mathcal{D}\text{ir}(\xi_1,\dots,\xi_N),$$

   with $\boldsymbol{\theta} = (\theta_1,\dots,\theta_T)'$ and $\xi_n = \xi + T_n$, where T_n denotes the number of periods assigned to group n.
   (b) Update the hyperparemter ξ with random walk Metropolis-Hastings step (for details, see Malsiner-Walli et al., 2016).
   (c) Sample group indicators $\theta_t \in \{1,\dots,N\}$ for each $\tilde{\boldsymbol{\alpha}}_t$ from a Multinomial distribution:

$$\text{P}(\theta_t = n|\omega_n, \tilde{\boldsymbol{\mu}}_n) \propto \omega_n f_{\mathcal{N}}(\tilde{\boldsymbol{\alpha}}_t|\tilde{\boldsymbol{\mu}}_n, \boldsymbol{I}_K), \quad \text{for} n = \{1,\dots,N\}.$$

   (d) The full conditional posterior of $\tilde{\boldsymbol{\mu}} = \text{vec}(\tilde{\boldsymbol{\mu}}_1,\dots,\tilde{\boldsymbol{\mu}}_N)$ follows a multivariate Gaussian distribution:

$$\tilde{\boldsymbol{\mu}}|\boldsymbol{\Lambda}_0, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{c}_1, \boldsymbol{\Lambda}_1),$$

   with $\boldsymbol{\theta} = (\theta_1,\dots,\theta_T)'$ and:

$$\begin{aligned} \boldsymbol{\Lambda}_1 &= \left(\boldsymbol{I}_K \otimes \boldsymbol{\Xi}'\boldsymbol{\Xi} + \boldsymbol{I}_N \otimes \boldsymbol{\Lambda}_0^{-1}\right)^{-1}, \\ \boldsymbol{c}_1 &= \boldsymbol{\Lambda}_1\left(\text{vec}(\boldsymbol{\Xi}'\tilde{\boldsymbol{A}})\right). \end{aligned}$$

   Here, $\boldsymbol{\Xi}$ denotes a $T\times N$ matrix with $(t,n)^{th}$ element given by $\mathcal{I}(\theta_t = n)$, where $\mathcal{I}(\bullet)$ refers to the indicator function and $\tilde{\boldsymbol{A}}$ collects $\tilde{\boldsymbol{\alpha}}$ in a $T\times K$ matrix.
   (e) Sample shrinkage parameters $\{l_j\}_{j=1}^K$ from a GIG distribution:

$$l_j|\boldsymbol{R}, \{\tilde{\boldsymbol{\mu}}_n\}_{n=1}^N \sim \text{GIG}\left(e_0 - \frac{N}{2}, 2e_1, \frac{\sum_{n=1}^N \tilde{\mu}_{jn}^2}{r_j}\right).$$

   with $\tilde{\mu}_{jn}$, for $n=\{1,\dots,N\}$, denoting the j^{th} element of $\tilde{\boldsymbol{\mu}}_n$.

A4. The spectral decomposition

To obtain a time-varying low-frequency measure between two endogenous variable, we follow Sargent and Surico (2011) and Kliem et al. (2016). We therfore recast a TVP-VAR model in its companion form:

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{J}\boldsymbol{Z}_t \\ \boldsymbol{Z}_t &= \boldsymbol{F}_t\boldsymbol{Z}_{t-1} + \boldsymbol{E}_t, \quad \boldsymbol{E}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Upsilon}_t) \end{aligned}$$

In the following, the spectral density of $\boldsymbol{Y}_t$ at the very low frequency $\rho = 0$ is given by:

$$\boldsymbol{\Pi}_t(\rho=0) = \boldsymbol{J}\left(\boldsymbol{I}_{mp+1} - \boldsymbol{F}_t\right)\boldsymbol{\Upsilon}_t\left(\boldsymbol{I}_{mp+1} - \boldsymbol{F}_t'\right)^{-1}\boldsymbol{J}'.$$

For $\rho = 0$ the low-frequency relationship $\pi_{ij,t}$ between two variables $(Y_{it}, Y_{jt}) \in \boldsymbol{Y}_t$ can be derived with:

$$\pi_{ij,t} = \frac{\Pi_{ij,t}(\rho=0)}{\Pi_{jj,t}(\rho=0)}$$

with $\Pi_{ij,t}$ denoting the $(i,j)^{th}$ element in $\boldsymbol{\Pi}_t$.

Appendix B. Additional forecasting results

Table B1
Density forecast performance (CRPS ratios) relative to the benchmark (`const (Min.)`). The red shaded row denotes the benchmark (and its CRPS values). Asterisks indicate statistical significance for each model relative to `const (Min.)` at the 1 (***), 5 (**) and 10 (*) percent significance levels.

Specification		1-quarter-ahead					1-year-ahead					2-years-ahead				
Class	Subclass	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS	TOT	GDPC1	CPIAUCSL	UNRATE	FEDFUNDS
FA-VAR																
const. (Min.)		0.92	0.85	0.97	0.89	1.10	0.90**	0.85**	1.02	0.81	0.94	0.88**	0.98	0.97	0.87	0.71***
const. (NG)		0.90**	0.82**	0.97	0.89	1.04	0.89**	0.84**	1.00	0.82	0.89	0.88**	0.97	0.97	0.87	0.72**
TVP-MIX	FLEX MIX	0.89***	0.82**	0.96	0.89	0.85*	0.91**	0.89	0.98	0.86	0.88	0.97	0.95	1.03	0.99	0.89
	FLEX MS	0.90**	0.84*	0.97	0.85	0.88	0.91*	0.89*	1.02	0.82	0.85	0.95	1.00	1.00	0.94	0.87
	SINGLE	0.90**	0.84*	0.98	0.88	0.80**	0.90**	0.88**	1.01	0.83	0.79**	0.96	0.98	1.04	0.96	0.82
	SSVS MIX	0.89**	0.83*	0.97	0.88	0.92	0.91**	0.88**	1.01	0.83	0.88	0.95	0.97	1.05	0.90	0.89
TVP-POOL	FLEX MIX	0.88**	0.80**	0.97	0.86	0.95	0.87**	0.81**	1.00	0.79	0.85	0.86*	0.94*	0.97	0.82	0.70**
	FLEX MS	0.88***	0.80**	0.96	0.86	0.95	0.87**	**0.81*** **	0.99	0.79	0.86	0.85*	0.94*	0.95	0.82	0.71**
	SINGLE	0.88***	0.80**	0.95*	0.86	0.95	0.87**	0.81***	1.00	0.79	0.85	**0.85***	0.93**	0.96	**0.81**	0.70**
	SSVS MIX	0.89**	0.82**	0.97	0.86	0.95	0.87**	0.81**	0.99	0.79	0.87	0.86*	0.95*	0.96	0.82	**0.70****
TVP-RW	FLEX MIX	0.89**	0.85*	0.95	0.85	0.82**	0.87*	0.87**	0.99	0.78	**0.75****	0.92	1.04	0.97	0.90	0.74
	FLEX MS	0.89**	0.83**	0.96	0.88	0.85	0.90*	0.88*	1.01	0.82	0.81**	0.95	1.03	0.99	0.95	0.81
	SINGLE	0.91**	0.84*	1.00	0.89	0.87	0.97	0.94	1.02	0.95	0.98	1.06	1.00	1.09	1.16*	0.98
	SSVS MIX	0.90**	0.85*	0.97	0.88	0.79**	0.89*	0.90	0.98	0.85	0.75**	0.95	1.03	0.98	0.97	0.78
L-VAR																
const. (Min.)		0.94	0.87	1.03	0.83	1.10	0.90**	0.95*	0.95***	0.79	0.87	0.94**	1.03	0.95	0.94	0.83***
const. (NG)		0.89**	0.82***	0.97	0.81	0.99	0.85***	0.89**	0.90***	0.74	0.84	0.88**	0.99	0.91	0.87	0.76**
TVP-MIX	FLEX MIX	0.97	0.87	1.08	0.81	1.17	0.92*	0.89	1.04	0.79	0.95	1.00	1.06	1.09	0.91	0.95
	FLEX MS	0.97	0.88	1.07	0.84	1.10	0.96	0.95	1.06	0.81	1.00	1.44	1.75	1.79	1.05	1.24
	SINGLE	0.92	0.85**	1.02	0.83	0.88	0.90*	0.88*	1.00	0.81	0.86*	0.96	1.04	1.02	0.91	0.88
	SSVS MIX	0.91	0.88	0.97	0.82	0.92	0.90*	0.89	1.01	0.78	0.88	0.97	1.03	1.05	0.90	0.91
TVP-POOL	FLEX MIX	0.87***	**0.80*****	0.97	0.78*	**0.79*****	0.85**	0.88*	0.95	0.73	0.77**	0.89	1.00	0.98	0.83	0.75**
	FLEX MS	**0.86*****	0.80***	0.95	0.77*	0.83*	0.85**	0.87**	0.94	**0.72**	0.78**	0.88	0.99	0.98	0.82	0.74**
	SINGLE	0.87***	0.80***	0.96	0.77*	0.85	**0.84****	0.86**	0.94	0.72	0.77***	0.88*	0.99	0.97	0.82	0.75**
	SSVS MIX	0.87***	0.81***	0.96	**0.77***	0.85	0.84**	0.87**	0.94	0.72	0.76**	0.88	1.00	0.97	0.82	0.74**
TVP-RW	FLEX MIX	0.93	0.89	0.98	0.83	0.93	0.90*	0.91	0.97	0.82	0.85*	0.97	1.09	1.02	0.93	0.86
	FLEX MS	0.93	0.89	1.00	0.83	0.89	0.92*	0.90	0.99	0.84	0.92	0.99	1.07	1.02	0.95	0.94
	SINGLE	1.03	0.92	1.17	0.85	1.14	0.97	0.93	1.03	0.85	1.10	1.05	1.12	1.04	1.01	1.04
	SSVS MIX	0.92	0.89	0.97	0.84	0.95	0.91*	0.90	0.97	0.84	0.92	0.99	1.04	1.00	0.96	0.94
S-VAR																
const. (Min.)		0.24	0.44	0.41	0.07	0.05	0.38	0.54	0.44	0.31	0.22	0.49	0.48	0.46	0.58	0.44
const. (NG)		0.98**	0.98*	0.99	1.00	1.00	0.96**	0.94**	0.97**	0.96**	0.96	0.97	0.99	0.96	0.96**	0.95
TVP-MIX	FLEX MIX	0.92***	0.92**	**0.91***	0.96	0.88***	0.90	0.84*	0.94	0.93	0.90	0.95	0.93	0.97	0.98	0.91
	FLEX MS	0.95*	0.96	0.96	0.94	0.92*	0.95	0.92*	0.98	0.93	0.98	0.98	1.01	1.00	0.95	0.97
	SINGLE	0.96*	0.96	0.96	0.99	0.85**	0.93	0.90*	0.95	0.97	0.93	0.98	1.00	0.99	0.99	0.92
	SSVS MIX	0.94***	0.95	0.92**	0.99	0.90***	0.91*	0.87*	0.95	0.93	0.90	0.96	0.98	0.98	0.96	0.91
TVP-POOL	FLEX MIX	0.97***	0.97**	0.98	0.98	0.95***	0.95**	0.94**	0.98*	0.94	0.92*	0.94	0.99	0.95	0.92	0.91
	FLEX MS	0.98***	0.97**	0.99	0.97*	0.95**	0.94**	0.94**	0.98**	0.93	0.92*	0.94	1.00	0.96	0.92	0.90
	SINGLE	0.97***	0.97**	0.98	0.97	0.95**	0.95**	0.93**	0.99	0.94	0.92	0.94	0.99	0.95	0.92	0.90
	SSVS MIX	0.97***	0.97*	0.98	0.97*	0.95***	0.95**	0.93**	0.98*	0.94	0.93	0.94	0.99	0.95	0.92	0.91
TVP-RW	FLEX MIX	0.96***	0.96	0.95	1.00	0.90***	0.93*	0.93	0.92*	0.97	0.91	0.96	1.01	0.93	0.98	0.92
	FLEX MS	0.97	0.98	0.97	0.99	0.89**	0.94	0.91*	0.95	0.98	0.96	2.26	3.86	2.30	1.62	1.31
	SINGLE	0.95*	0.91*	0.99	0.97	0.84***	0.91	0.82*	0.96	1.00	0.88	0.96	**0.92**	0.99	1.07	0.84
	SSVS MIX	0.97	0.99	0.95	1.01	0.88**	0.94	0.91	0.93	1.01	0.93	0.98	1.01	0.95	1.03	0.93
	Primiceri (2005)	0.96	0.98	0.94	0.97	1.04	0.90*	0.88*	**0.89***	0.96	0.93	0.94	0.96	**0.90**	1.00	0.91

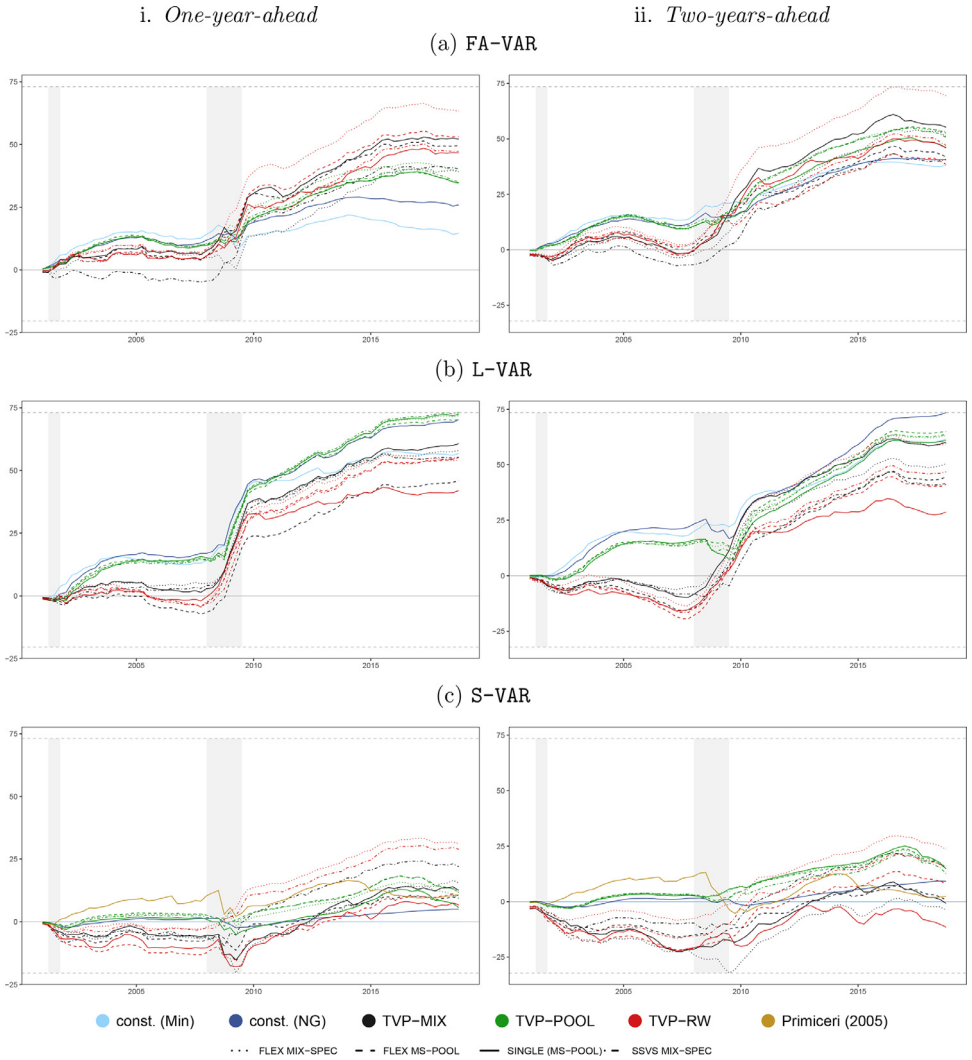


Fig. B.1. Evolution of one- and two-years-ahead total cumulative LPBFs relative to the benchmark. The gray dashed lines refer to the maximum/minimum Bayes factor over the full hold-out sample. The light gray shaded areas indicate the NBER recessions in the US.

Appendix C. Data

In this section we provide further details on the variable used for the large-scale VAR (L-VAR). Table C.1 lists the exact description and provides further information on the transformation of the indicators. The gray shaded rows denote our target variables.

For the factor-augmented VAR (FA-VAR) we consider the full data set, compromising 165 variables. For brevity we refer to McCracken and Ng (2016) for a detailed description and transformation codes. All variables, serving as a basis for the principal components, are transformed to stationarity as suggested in McCracken and Ng (2016). Finally, we standardise the data by demeaning each variable and dividing through the standard deviation. Especially for principal components standardising is important due to the scale variance of the components.

Table C1
Data for the US is obtained from the FRED data base of the Federal Reserve of St. Louis. The column `Transformation`shows the transformation applied to each variable. Following McCracken and Ng (2016), (1) implies no transformation, (5) denotes growth rates, defined as log first differences $\log\left(\frac{x_t}{x_{t-1}}\right)$ and (7) denotes differences in percentage changes with $\Delta\left(\frac{x_t - x_{t-1}}{x_{t-1}}\right)$. All variables are standardized by subtracting the mean and dividing by the standard deviation.

FRED.Mnemonic	Description	Transformation
GDPC1	Real Gross Domestic Product	5
PCECC96	Real Personal Consumption Expenditures	5
FPIx	Real private fixed investment	5
GCEC1	Real Government Consumption Expenditures and Gross Investment	5
INDPRO	IP:Total index Industrial Production Index (Index 2012=100)	5
CE16OV	Civilian Employment (Thousands of Persons)	5
UNRATE	Civilian Unemployment Rate (Percent)	1
CES0600000007	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	1
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started	5
PERMIT	New Private Housing Units Authorized by Building Permits	5
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	5
GDPCTPI	Gross Domestic Product: Chain-type Price Index	5
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	5
CES0600000008	Average Hourly Earnings of Production and Nonsupervisory Employees	5
FEDFUNDS	Effective Federal Funds Rate (Percent)	1
GS1	1-Year Treasury Constant Maturity Rate (Percent)	1
GS10	10-Year Treasury Constant Maturity Rate (Percent)	1
TOTRESNS	Total Reserves of Depository Institutions	5
NONBORRES	Reserves Of Depository Institutions, Nonborrowed	7
S.P.500	S & P's Common Stock Price Index: Composite	5

References

Aastveit, K.A., Carriero, A., Clark, T.E., Marcellino, M., 2017. Have standard VARs remained stable since the crisis? Journal of Applied Econometrics 32 (5), 931–951.

Ball, L., Mazumder, S., 2011. Inflation dynamics and the Great Recession. Brookings Papers on Economic Activity 42 (1 (Spring), 337–405.

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

Bauwens, L., Koop, G., Korobilis, D., Rombouts, J.V., 2015. The contribution of structural break models to forecasting macroeconomic series. Journal of Applied Econometrics 30 (4), 596–620.

Belmonte, M.A., Koop, G., Korobilis, D., 2014. Hierarchical shrinkage in time-varying parameter models. Journal of Forecasting 33 (1), 80–94.

Bhattacharya, A., Chakraborty, A., Mallick, B.K., 2016. Fast sampling with Gaussian scale mixture priors in high-dimensional regression. Biometrika 103 (4), 985–991.

Bitto, A., Frühwirth-Schnatter, S., 2019. Achieving shrinkage in a time-varying parameter model framework. Journal of Econometrics 210 (1), 75–97.

Cadonna, A., Frühwirth-Schnatter, S., Knaus, P., 2020. Triple the gamma–a unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. Econometrics 8 (2), 20.

Carriero, A., Clark, T.E., Marcellino, M., 2019. Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. Journal of Econometrics 212 (1), 137–154.

Carvalho, C.M., Polson, N.G., Scott, J.G., 2010. The horseshoe estimator for sparse signals. Biometrika 97 (2), 465–480.

Chan, J., Eisenstat, E., Strachan, R., 2020. Reducing the state space dimension in a large TVP-VAR. Journal of Econometrics 218 (1), 105–118.

Chan, J.C., 2019. Large hybrid time-varying parameter VARs. CAMA Working Paper 77/2019.

Chan, J.C., Jeliazkov, I., 2009. Efficient simulation and integrated likelihood estimation in state space models. International Journal of Mathematical Modelling and Numerical Optimisation 1 (1-2), 101–120.

Chan, J.C., Koop, G., Leon-Gonzalez, R., Strachan, R.W., 2012. Time-varying dimension models. Journal of Business & Economic Statistics 30 (3), 358–367.

Chan, J.C., Strachan, R.W., 2020. Bayesian state space models in macroeconometrics. Journal of Economic Surveys 0 (ja).

Clark, T., 2011. Real-time density forecasts from BVARs with stochastic volatility. Journal of Business & Economic Statistics 29, 327–341.

Cogley, T., Primiceri, G.E., Sargent, T.J., 2010. Inflation-gap persistence in the us. American Economic Journal: Macroeconomics 2 (1), 43–69.

Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics 8 (2), 262–302.

Coibion, O., Gorodnichenko, Y., 2015. Is the phillips curve alive and well after all? inflation expectations and the missing disinflation. American Economic Journal: Macroeconomics 7 (1), 197–232.

Cross, J.L., Hou, C., Poon, A., 2020. Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. International Journal of Forecasting 36 (3), 899–915.

Czudaj, R.L., 2019. Dynamics between trading volume, volatility and open interest in agricultural futures markets: A bayesian time-varying coefficient approach. Econometrics and Statistics 12, 78–145.

D'Agostino, A., Gambetti, L., Giannone, D., 2013. Macroeconomic forecasting and structural change. Journal of Applied Econometrics 28 (1), 82–101.

Del Negro, M., Lenza, M., Primiceri, G.E., Tambalotti, A., 2020. What's up with the Phillips Curve? NBER Working Papers. National Bureau of Economic Research.

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

Eickmeier, S., Lemke, W., Marcellino, M., 2015. Classical time varying factor-augmented vector auto-regressive models-estimation, forecasting and structural analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178 (3), 493–533.

Feldkircher, M., Huber, F., Kastner, G., 2017. Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs? arXiv:1711.00564.

Follett, L., Yu, C., 2019. Achieving parsimony in bayesian vector autoregressions with the horseshoe prior. Econometrics and Statistics 11, 130–144.

Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association 96 (453), 194–209.

Frühwirth-Schnatter, S., Wagner, H., 2010. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. Journal of Econometrics 154 (1), 85–100.

George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. Journal of the American Statistical Association 88 (423), 881–889.

George, E.I., McCulloch, R.E., 1997. Approaches for Bayesian variable selection. Statistica Sinica 339–373.

Gerlach, R., Carter, C., Kohn, R., 2000. Efficient Bayesian inference for dynamic mixture models. Journal of the American Statistical Association 95 (451), 819–828.
Giannone, D., Lenza, M., Primiceri, G.E., 2015. Prior selection for vector autoregressions. The Review of Economics and Statistics 97 (2), 436–451.
Giordani, P., Kohn, R., 2008. Efficient Bayesian inference for multiple change-point and mixture innovation models. Journal of Business & Economic Statistics 26 (1), 66–77.
Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.
Griffin, J., Brown, P., 2010. Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis 5 (1), 171–188.
Groen, J.J., Paap, R., Ravazzolo, F., 2013. Real-time inflation forecasting in a changing world. Journal of Business & Economic Statistics 31 (1), 29–44.
Hauzenberger, N., Huber, F., Koop, G., 2020. Dynamic shrinkage priors for large time-varying parameter regressions using scalable Markov Chain Monte Carlo methods. arXiv:2005.03906.
Hauzenberger, N., Huber, F., Koop, G., Onorante, L., 2019. Fast and flexible Bayesian inference in time-varying parameter regression models. arXiv:1910.10779.
Huber, F., Feldkircher, M., 2019. Adaptive shrinkage in Bayesian vector autoregressive models. Journal of Business & Economic Statistics 37 (1), 27–39.
Huber, F., Kastner, G., Feldkircher, M., 2019. Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models. Journal of Applied Econometrics 34 (5), 621–640.
Huber, F., Koop, G., Onorante, L., 2020. Inducing sparsity and shrinkage in time-varying parameter models. Journal of Business & Economic Statistics 0 (ja).
Huber, F., Koop, G., Pfarrhofer, M., 2020. Bayesian inference in high-dimensional time-varying parameter models using integrated rotated Gaussian approximations. arXiv:2002.10274.
Kalli, M., Griffin, J., 2014. Time-varying sparsity in dynamic regression models. Journal of Econometrics 178 (2), 779–793.
Kastner, G., 2016. Dealing with stochastic volatility in time series using the r package stochvol. Journal of Statistical Software 69 (5), 1–30.
Kastner, G., Frühwirth-Schnatter, S., 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. Computational Statistics & Data Analysis 76, 408–423.
Kastner, G., Huber, F., 2020. Sparse bayesian vector autoregressions in huge dimensions. Journal of Forecasting 39 (7), 1142–1165.
Kim, C.-J., Nelson, C.R., 1999. Has the US economy become more stable? A Bayesian approach based on a markov-switching model of the business cycle. The Review of Economics and Statistics 81 (4), 608–616.
Kliem, M., Kriwoluzky, A., Sarferaz, S., 2016. On the low-frequency relationship between public deficits and inflation. Journal of Applied Econometrics 31 (3), 566–583.
Koop, G., 2013. Forecasting with medium and large Bayesian VARs. Journal of Applied Econometrics 28 (2), 177–203.
Koop, G., Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review 53 (3), 867–886.
Koop, G., Korobilis, D., 2013. Large time-varying parameter VARs. Journal of Econometrics 177 (2), 185–198.
Koop, G., Leon-Gonzalez, R., Strachan, R.W., 2009. On the evolution of the monetary policy transmission mechanism. Journal of Economic Dynamics and Control 33 (4), 997–1017.
Koop, G., Potter, S.M., 2007. Estimation and forecasting in models with multiple breaks. The Review of Economic Studies 74 (3), 763–789.
Korobilis, D., 2013. Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. Oxford Bulletin of Economics and Statistics 75 (2), 157–179.
Korobilis, D., 2019. High-dimensional macroeconomic forecasting using message passing algorithms. Journal of Business & Economic Statistics 0 (ja).
Korobilis, D., Koop, G., 2020. Bayesian dynamic variable selection in high dimensions. MPRA:100164.
Kowal, D.R., Matteson, D.S., Ruppert, D., 2019. Dynamic shrinkage processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81 (4), 781–804.
Litterman, R.B., 1986. Forecasting with Bayesian vector autoregressions – five years of experience. Journal of Business & Economic Statistics 4 (1), 25–38.
Lopes, H.F., McCulloch, R.E., Tsay, R.S., 2018. Parsimony inducing priors for large scale state-space models. hedibert.org/wp-content/uploads/2018/09/lopes-mcculloch-tsay-2018.pdf.
Malsiner-Walli, G., Frühwirth-Schnatter, S., Grün, B., 2016. Model-based clustering based on sparse finite Gaussian mixtures. Statistics and Computing 26 (1-2), 303–324.
McCausland, W.J., Miller, S., Pelletier, D., 2011. Simulation smoothing for state–space models: A computational efficiency analysis. Computational Statistics & Data Analysis 55 (1), 199–212.
McCracken, M.W., Ng, S., 2016. Fred-md: A monthly database for macroeconomic research. Journal of Business & Economic Statistics 34 (4), 574–589.
McCulloch, R.E., Tsay, R.S., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. Journal of the American Statistical Association 88 (423), 968–978.
Mumtaz, H., Theodoridis, K., 2018. The changing transmission of uncertainty shocks in the US. Journal of Business & Economic Statistics 36 (2), 239–252.
Nakajima, J., West, M., 2013. Bayesian analysis of latent threshold dynamic models. Journal of Business & Economic Statistics 31 (2), 151–164.
Ng, S., Wright, J.H., 2013. Facts and challenges from the great recession for forecasting and macroeconomic modeling. Journal of Economic Literature 51 (4), 1120–1154.
Park, T., Casella, G., 2008. The Bayesian lasso. Journal of the American Statistical Association 103 (482), 681–686.
Paul, P., 2020. The time-varying effect of monetary policy on asset prices. Review of Economics and Statistics 102 (4), 690–704.
Polson, N.G., Scott, J.G., 2010. Shrink globally, act locally: Sparse Bayesian regularization and prediction. Bayesian Statistics 9, 501–538.
Primiceri, G., 2005. Time varying structural autoregressions and monetary policy. The Review of Economic Studies 72 (3), 821–852.
Rockova, V., McAlinn, K., 2021. Dynamic variable selection with spike-and-slab process priors. Bayesian Analysis 16 (1), 233–269.
Sargent, T.J., Surico, P., 2011. Two illustrations of the quantity theory of money: Breakdowns and revivals. American Economic Review 101 (1), 109–128.
Sims, C.A., Zha, T., 2006. Were there regime switches in US monetary policy? American Economic Review 96 (1), 54–81.
Stock, J.H., Watson, M.W., 2007. Why has us inflation become harder to forecast? Journal of Money, Credit and banking 39, 3–33.
Stock, J.H., Watson, M.W., 2012. Disentangling the Channels of the 2007-2009 Recession. NBER Working Papers. National Bureau of Economic Research.
Uribe, P.V., Lopes, H.F., 2017. Dynamic sparsity on dynamic regression models. hedibert.org/wp-content/uploads/2018/06/uribe-lopes-Sep2017.pdf.
Watson, M.W., 2014. Inflation persistence, the nairu, and the great recession. American Economic Review 104 (5), 31–36.
Whiteman, C.H., 1984. Lucas on the quantity theory: Hypothesis testing without theory. The American Economic Review 74 (4), 742–749.
Yau, C., Holmes, C., 2011. Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination. Bayesian Analysis 6 (2), 329–351.
Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. Studies in Bayesian Econometrics and Statistics 6, Edited by: Goel, P. Zellner, A. 233–243.