

Huber, Florian; Kastner, Gregor; Feldkircher, Martin

Working Paper

Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models

Working Papers in Economics, No. 2018-05

Provided in Cooperation with:

Department of Social Sciences and Economics, University of Salzburg

Suggested Citation: Huber, Florian; Kastner, Gregor; Feldkircher, Martin (2018) : Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models, Working Papers in Economics, No. 2018-05, University of Salzburg, Department of Social Sciences and Economics, Salzburg

This Version is available at:

<https://hdl.handle.net/10419/201673>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Florian Huber, Gregor Kastner, Martin Feldkircher

Should I stay or should I go?
A latent threshold approach to
large-scale mixture innovation models

No. 2018-05

 **WORKING PAPERS**
IN ECONOMICS

Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models

Florian Huber¹, Gregor Kastner^{*2}, and Martin Feldkircher^{†3}

¹Salzburg Centre of European Union Studies (SCEUS), University of Salzburg

²Institute for Statistics and Mathematics, WU Vienna University of Economics and Business

³Foreign Research Division (AUSA), Oesterreichische Nationalbank (OeNB)

November 7, 2018

Abstract

We propose a straightforward algorithm to estimate large Bayesian time-varying parameter vector autoregressions with mixture innovation components for each coefficient in the system. The computational burden becomes manageable by approximating the mixture indicators driving the time-variation in the coefficients with a latent threshold process that depends on the absolute size of the shocks. Two applications illustrate the merits of our approach. First, we forecast the US term structure of interest rates and demonstrate forecast gains relative to benchmark models. Second, we apply our approach to US macroeconomic data and find significant evidence for time-varying effects of a monetary policy tightening.

Keywords: Time-varying parameter vector autoregression with stochastic volatility (TVP-VAR-SV), Change point model, Structural breaks, Term structure of interest rates, Monetary policy, R package `threshtvp`.

Acknowledgments: We sincerely thank the participants of the 7th and 8th European Seminar on Bayesian Econometrics (ESOB 2016 and 2017), the Econometrics Seminar of the Institute for Advanced Studies, Vienna, the WU Brown Bag Seminar of the Institute of Statistics and Mathematics, the 3rd Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance 2017, the NBP Workshop on Forecasting 2017, and in particular Sylvia Frühwirth-Schnatter, Hedibert Freitas Lopes, Herman van Dijk, and Helga Wagner for many helpful comments and suggestions that improved the paper significantly.

Conflict of Interest: The authors declare no conflict of interest.

*Corresponding author. Address: Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, Welthandelsplatz 1, Building D4, 4th Floor, 1020 Vienna, Austria. Phone: +43 1 31336-5593. Fax: +43 1 31336-905593. Email: gregor.kastner@wu.ac.at.

[†]The opinions expressed in this paper are those of the authors and do not necessarily reflect the official viewpoint of the Oesterreichische Nationalbank or the Eurosystem.

1 Introduction

In the last few years, economists in policy institutions and central banks were criticized for not foreseeing the recent financial crisis that engulfed the world economy and led to a sharp drop in economic activity. Critics argued that economists failed to predict the crisis because models commonly utilized at policy institutions back then were too simplistic. For instance, the majority of forecasting models adopted were (and possibly still are) linear and low dimensional. The former implies that the underlying structural mechanisms and the volatility of economic shocks are assumed to remain constant over time – a rather restrictive assumption. The latter implies that only little information is exploited which may be detrimental for obtaining reliable predictions.

In light of this criticism, practitioners started to adopt more complex models that are capable of capturing salient features of time series commonly observed in macroeconomics and finance. These models are based on earlier research that provides considerable evidence, at least for US data, that the influence of certain variables appears to be time-varying ([Stock and Watson, 1996](#); [Cogley and Sargent, 2002, 2005](#); [Primiceri, 2005](#); [Sims and Zha, 2006](#)). This raises additional issues related to model specification and estimation. For instance, do all regression parameters vary over time? Or is time variation just limited to a specific subset of the parameter space? Moreover, as is the case with virtually any modeling problem, the question whether a given variable should be included in the model in the first place naturally arises. Apart from deciding whether parameters are changing over time, the nature of the process that drives the dynamics of the coefficients also proves to be an important modeling decision.

In a recent contribution, [Frühwirth-Schnatter and Wagner \(2010\)](#) focus on model specification issues within the general framework of state space models. Exploiting a non-centered parametrization of the model allows them to rewrite the model in terms of a constant parameter specification, effectively capturing the steady state of the process along with deviations thereof. The non-centered parameterization is subsequently used to search for appropriate model specifications, imposing shrinkage on the steady state part and the corresponding deviations.

Recent research aims to discriminate between inclusion/exclusion of elements of different variables and whether the associated regression coefficients are constant or time-varying ([Koop and Korobilis, 2012, 2013](#); [Kalli and Griffin, 2014](#); [Belmonte, Koop, and Korobilis, 2014](#); [Eisenstat, Chan, and Strachan, 2016](#)). Another strand of the lit-

erature asks whether coefficients are constant or time-varying by assuming that the innovation variance in the state equation is characterized by a change point process (McCulloch and Tsay, 1993; Gerlach, Carter, and Kohn, 2000; Koop, Leon-Gonzalez, and Strachan, 2009; Giordani and Kohn, 2008). The main drawback of this modeling approach is the severe computational burden originating from the need to simulate additional latent states for each parameter. This renders estimation of large dimensional models like vector autoregressions (VARs) unfeasible. To circumvent such problems, Koop, Leon-Gonzalez, and Strachan (2009) estimate a single Bernoulli random variable to discriminate between time constancy and parameter variation for the autoregressive coefficients, the covariances, and the log-volatilities, respectively. This assumption, however, implies that either all autoregressive parameters change over a given time frame, or none of them. Along these lines, Maheu and Song (2018) allow for simultaneous breaks in regression coefficients and volatility parameters. They show that their multivariate approach is inferior to univariate change point models when out-of-sample forecasts are considered and conclude that allowing for independent breaks in each series is important.

In the present paper, we introduce a method that can be applied to a highly parameterized VAR model by combining ideas from the literature on latent threshold models (Neelon and Dunson, 2004; Nakajima and West, 2013a,b; Zhou, Nakajima, and West, 2014; Kimura and Nakajima, 2016) to approximate the latent indicators during Markov chain Monte Carlo (MCMC) sampling. As mentioned above, the main computational hurdle stems from the necessity to apply forward-filtering backward-sampling (FFBS) based algorithms to estimate the indicators that control the time-variation in the regression coefficients.

The key contribution of this paper is to avoid the computationally intensive simulation of the latent indicators by proposing a straightforward approximation to these indicators and thus allow for estimation of large-scale models. In doing so, we mimic the behavior of a standard mixture innovation model by setting the value of an indicator equal to one if the absolute value of the parameter change exceeds a threshold to be estimated. In that case, the corresponding state innovation variance is set to a large value, allowing for large jumps in the regression coefficients. By contrast, if the absolute changes are small (i.e., below the threshold), a state innovation variance close to zero is adopted and the corresponding regression coefficient can be viewed as being constant over that certain stretch in time. Compared to existing algorithms, the

additional costs of estimating the proposed model, henceforth labeled the threshold time varying parameter (TTVP) model, is negligible. To assess systematically, in a data-driven fashion, which predictors should be included in the model, we impose a set of Normal-Gamma priors ([Griffin and Brown, 2010](#)) in the spirit of [Bitto and Frühwirth-Schnatter \(forthcoming\)](#) on the initial state of the system. The TTVP code is bundled in the R package `threshtvp` which is made available from the authors upon request.

We illustrate the empirical merits of our approach by carrying out two empirical exercises. In the first exercise, we predict the US term structure of interest rates. The proposed framework is benchmarked against several constant parameter Bayesian VAR models with stochastic volatility (SV) and hierarchical shrinkage priors, time-varying parameter VARs as well as a multivariate random walk with SV. Moreover, we follow [Diebold and Li \(2006\)](#) and use a model based on the [Nelson and Siegel \(1987\)](#) three factor model. The findings indicate that our proposed TTVP specification outperforms all competing specifications for one-month-ahead as well as three-month-ahead predictions. The forecasting gains appear to be especially pronounced during crises episodes.

In the second application, we use a medium-scale US macroeconomic dataset to investigate the degree of time-variation in the macroeconomic responses to a contractionary monetary policy shock. We find evidence for abrupt changes of inflation responses. These show a considerable price puzzle during the 1960s which suddenly disappears in the early 1980s. Effects on variables reflecting economic activity also vary over time but do so more gradually. They are especially pronounced during the aftermath of the global financial crisis in 2008/09 indicating evidence for increased effectiveness of monetary policy.

The paper is structured as follows. Section [2](#) introduces the modeling approach, the prior setup and the corresponding MCMC algorithm for posterior simulation. Section [3](#) illustrates the behavior of the model by showcasing scenarios with no, few, and many jumps in the state equation, alongside a standard TVP specification with sustained movement. In Section [4](#), we predict the US term structure of interest rates. In Section [5](#), we apply the model to a medium-scale US macroeconomic dataset and investigate during which periods VAR coefficients display the largest amount of time-variation; furthermore, we scrutinize the associated implications on dynamic responses with respect to a monetary policy shock. Finally, Section [6](#) concludes.

2 Econometric framework

In this section, we first introduce a univariate dynamic regression model that is capable of discriminating between constant and time-varying parameters at each point in time. This stylized framework is used to discuss the main ideas of the paper. We then subsequently generalize this model framework to the VAR case that is used in the empirical applications.

2.1 A mixture innovation model

Consider the following dynamic regression model,

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_t + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_t^2), \quad (2.1)$$

where \mathbf{x}_t is a K -dimensional vector of explanatory variables and $\boldsymbol{\beta}_t = (\beta_{1t}, \dots, \beta_{Kt})'$ a vector of regression coefficients. The error term u_t is assumed to be independently normally distributed with (potentially) time-varying variance σ_t^2 . This model assumes that the relationship between elements of \mathbf{x}_t and y_t is not necessarily constant over time, but changes subject to some law of motion for $\boldsymbol{\beta}_t$. Typically, researchers assume that the j th element of $\boldsymbol{\beta}_t, \beta_{jt}$ ($j = 1, \dots, K$), follows a random walk process,¹

$$\beta_{jt} = \beta_{j,t-1} + e_{jt}, \quad e_{jt} \sim \mathcal{N}(0, \vartheta_j), \quad (2.2)$$

with ϑ_j denoting the innovation variance of the latent states. [Eq. \(2.2\)](#) implies that parameters evolve gradually over time, ruling out abrupt changes. While being conceptually flexible, in the presence of only a few breaks in the parameters, this model generates spurious movements in the coefficients that could be detrimental for the empirical performance of the model ([D'Agostino, Gambetti, and Giannone, 2013](#)).

Thus, we deviate from [Eq. \(2.2\)](#) by specifying the innovations of the state equation e_{jt} to be a spike-and-slab mixture distribution. More concretely, let

$$e_{jt} \sim \mathcal{N}(0, \theta_{jt}), \quad (2.3)$$

$$\theta_{jt} = s_{jt} \vartheta_{j1} + (1 - s_{jt}) \vartheta_{j0}, \quad (2.4)$$

¹A recent exception to this is [Ročková and McAlinn \(2018\)](#) who assume a stationary thresholded AR(1) process to model the evolution of the coefficients.

where ϑ_{j1} (the slab variance) and ϑ_{j0} (the spike variance) are state innovation variances with $\vartheta_{j1} \gg \vartheta_{j0}$ and ϑ_{j0} set close to zero. Furthermore, s_{jt} is an indicator variable that follows a Bernoulli distribution, i.e.,

$$s_{jt} = \begin{cases} 1 & \text{with probability } p_j, \\ 0 & \text{with probability } 1 - p_j. \end{cases} \quad (2.5)$$

This model is a relatively standard mixture innovation model (McCulloch and Tsay, 1993; Gerlach, Carter, and Kohn, 2000; Giordani and Kohn, 2008).² Eq. (2.3) and Eq. (2.4) state that if s_{jt} equals one, we assume that the change in β_{jt} is normally distributed with zero mean and variance ϑ_{j1} . On the contrary, if s_{jt} equals zero, the innovation variance is set close to zero, effectively implying that $\beta_{jt} \approx \beta_{j,t-1}$, i.e., almost no change from period $(t - 1)$ to t .

This modeling approach provides a great deal of flexibility, nesting a plethora of simpler model specifications. The interesting cases are characterized by situations where $s_{jt} = 1$ only for some t . For instance, it could be the case that parameters tend to exhibit strong movements at given points in time but stay constant for the majority of the time. An unrestricted time-varying parameter model would imply that the parameters are gradually changing over time, depending on the innovation variance in Eq. (2.2). Another prominent case would be a structural break model with an unknown number of breaks (for a Bayesian exposition, see e.g. Koop and Potter, 2007). Recently, this framework has been extended by Uribe and Lopes (2017) who model the indicator as a first-order two-state Markov process.

2.2 Mitigating the computational burden through thresholding

Unfortunately, estimation of the model described in the previous section is computationally cumbersome if K is large as in multivariate systems like VARs, even though there exist several estimation strategies. One strand of the literature (see McCulloch and Tsay, 1993) estimates the indicators conditional on the states using single-step Gibbs updating within a larger MCMC algorithm. This, however, often results in poor mixing properties of the algorithm since the states and the indicators are typically highly correlated. The more recent literature (Gerlach, Carter, and Kohn, 2000) sim-

²The main difference is that the literature typically assumes that $\vartheta_{j0} \equiv 0$ for all j (for an exception, see Carter and Kohn, 1994).

ulates the indicators after integrating out the latent states using Kalman-filter-based algorithms. Unfortunately, this procedure has to be repeated for each coefficient during MCMC sampling, turning computationally prohibitive even for moderate K . Thus, researchers often resort to models where only a small number of indicators is introduced that determines the amount of time-variation for certain parts of the parameter space in the system (for a VAR application, see, [Koop, Leon-Gonzalez, and Strachan, 2009](#)).

The key innovation of the present paper is to circumvent this issue by proposing a relatively simple approximation that makes immediate use of the fact that Gibbs sampling generates draws from the joint posterior by sampling from the full conditionals. Similarly to the early literature on mixture innovation models mentioned above ([McCulloch and Tsay, 1993](#)), we also condition on the states to simulate the indicators s_{jt} during MCMC simulation. However, instead of directly sampling from this full conditional distribution, we introduce one additional auxiliary parameter per coefficient, the threshold d_j , which in turn renders the indicators conditionally deterministic. More concretely, in the l th iteration of our MCMC algorithm, after obtaining draws $\{\beta_{jt}^{(l)}\}_{t=1,\dots,T}$ conditional on draws of the indicators $\{s_{jt}^{(l-1)}\}_{t=1,\dots,T}$ and the remaining parameters, the corresponding realization of s_{jt} is given through

$$s_{jt}^{(l)} = \begin{cases} 1 & \text{if } |\Delta\beta_{jt}^{(l)}| > d_j^{(l-1)}, \\ 0 & \text{if } |\Delta\beta_{jt}^{(l)}| \leq d_j^{(l-1)}, \end{cases} \quad (2.6)$$

where $d_j^{(l-1)}$ denotes the $(l-1)$ th draw of a coefficient-specific threshold d_j to be estimated and $\Delta\beta_{jt}^{(l)} := \beta_{jt}^{(l)} - \beta_{j,t-1}^{(l)}$. [Eq. \(2.6\)](#) indicates that if the absolute period-on-period change of the l th draw of β_{jt} exceeds the $(l-1)$ th draw of the threshold d_j , we set $s_{jt}^{(l)} = 1$ and thus use a large variance. By contrast, if the change in the current draws of the parameter is too small, the innovation variance is set close to zero, effectively implying that $\beta_{jt} \approx \beta_{j,t-1}$. The detailed description of the MCMC sampler, along all required full conditionals, can be found in [Section 2.5](#).

Compared to a standard mixture innovation model that postulates s_{jt} as a sequence of independent Bernoulli variables, our approach, labeled the threshold mixture innovation model, mimics this behavior by assuming that regime shifts are governed by a deterministic law of motion, conditionally on the current draw of $\{\beta_{jt}\}_{t=1,\dots,T}$ and d_j . The main advantage of our approach relative to standard mixture innovation models is

that instead of having to estimate a full sequence of s_{jt} for all j , the proposed framework only relies on a single additional parameter per coefficient. This renders estimation of high dimensional models such as vector autoregressions (VARs) feasible. The additional computational burden turns out to be negligible relative to an unrestricted TVP-VAR, see again Section 2.5 for more information.

Our model is also related to the latent thresholding approach put forward in Nakajima and West (2013a) within the time series context. However, while in their model latent thresholding discriminates between the inclusion or exclusion of a given covariate at time t , our model uses information on the changes in a given regression coefficient to mimic the behavior of a mixture innovation model. In addition, while the model proposed in Nakajima and West (2013a) assumes that the thresholded process enters Eq. (2.1) directly, our approach is based on estimating a non-linear model for the state equation. Nevertheless, notice that if the indicators are treated as augmented data, our model is a conditionally linear Gaussian state space model and thus standard algorithms can be used to estimate the latent states.

2.3 A multivariate extension with stochastic volatility

The model proposed in the previous subsection can be straightforwardly generalized to the VAR case with multivariate SV by letting \mathbf{y}_t be an m -dimensional response vector. In this case, Eq. (2.1) becomes

$$\mathbf{y}_t = \mathbf{x}_t' \boldsymbol{\beta}_t + \mathbf{u}_t, \quad (2.7)$$

with $\mathbf{x}_t' = \{\mathbf{I}_m \otimes \mathbf{z}_t'\}$, where $\mathbf{z}_t = (\mathbf{y}_{t-1}', \dots, \mathbf{y}_{t-P}')'$ includes the P lags of the endogenous variables.³ The vector $\boldsymbol{\beta}_t$ now contains the dynamic autoregressive coefficients with dimension $K = m^2 P$ where each element follows the state evolution given by Eqs. (2.2) to (2.6). The vector of white noise shocks \mathbf{u}_t is distributed as

$$\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_m, \boldsymbol{\Sigma}_t). \quad (2.8)$$

Hereby, $\mathbf{0}_m$ denotes an m -variate zero vector and $\boldsymbol{\Sigma}_t = \mathbf{V}_t \mathbf{H}_t \mathbf{V}_t'$ is a time-varying variance-covariance matrix. The matrix \mathbf{V}_t is a lower triangular matrix with unit diagonal and $\mathbf{H}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{mt}})$. We assume that the logarithm of the variances

³In the empirical application, we also include an intercept term which we omit here for simplicity.

evolves according to

$$h_{it} = \mu_i + \rho_i(h_{i,t-1} + \mu_i) + \nu_{it}, \quad i = 1, \dots, m, \quad (2.9)$$

where μ_i and ρ_i are equation-specific mean and persistence parameters and $\nu_{it} \sim \mathcal{N}(0, \zeta_i)$ is an equation-specific white noise error with variance ζ_i . For the covariances in \mathbf{V}_t we impose a random walk state equation with spike-and-slab error variances in analogy to Eq. (2.3) and Eq. (2.4).

Conditional on the ordering of the variables, it is straightforward to estimate the model on an equation-by-equation basis, augmenting the i th equation with the contemporaneous values of the preceding $(i - 1)$ equations, leading to a Cholesky-type decomposition of the variance-covariance matrix. Thus, the i th equation (for $i = 2, \dots, m$) is given by

$$y_{it} = \tilde{\mathbf{z}}_{it}' \tilde{\boldsymbol{\beta}}_{it} + u_{it}. \quad (2.10)$$

Here, $\tilde{\mathbf{z}}_{it} = (\mathbf{z}'_t, y_{1t}, \dots, y_{i-1,t})'$ denotes the augmented vector of regressors, while $\tilde{\boldsymbol{\beta}}_{it} = (\boldsymbol{\beta}'_{it}, \tilde{v}_{i1,t}, \dots, \tilde{v}_{i,i-1,t})'$ is a vector of latent states with dimension $K_i = mP + i - 1$ where $\boldsymbol{\beta}'_{it}$ refers to the coefficients associated with \mathbf{z}'_t in the i th equation and $\tilde{v}_{ij,t}$ denotes the dynamic regression coefficients on the j th contemporaneous value in the i th equation. Note that for the first equation we have $\tilde{\mathbf{z}}_{1t} = \mathbf{z}_t$ and $\tilde{\boldsymbol{\beta}}_{1t} = \boldsymbol{\beta}_{1t}$. The law of motion of the j th element of $\tilde{\boldsymbol{\beta}}_{it}$ reads

$$\tilde{\beta}_{ij,t} = \tilde{\beta}_{ij,t-1} + e_{ij,t}, \quad e_{ij,t} \sim \mathcal{N}(0, \theta_{ij,t}). \quad (2.11)$$

Hereby, $\theta_{ij,t}$ is defined analogously to Eq. (2.4).

While not being order-invariant, this specific way of stating the model yields two significant computational gains. First, the matrix operations involved in estimating the latent state vector become computationally less cumbersome. Second, we can exploit parallel computing and estimate each equation simultaneously on a grid.

2.4 Prior specification

We impose a Normal-Gamma prior (Griffin and Brown, 2010) on each element of $\tilde{\boldsymbol{\beta}}_{i0}$, the initial state of the i th equation,

$$\tilde{\beta}_{ij,0} | \tau_{ij} \sim \mathcal{N}\left(0, \frac{2}{\lambda_i^2} \tau_{ij}^2\right), \quad \tau_{ij}^2 \sim \mathcal{G}(a_i, a_i), \quad (2.12)$$

for $i = 1, \dots, m$ and $j = 1, \dots, K_i$. Hereby, λ_i^2 and a_i are hyperparameters and τ_{ij}^2 denotes an idiosyncratic scaling parameter that applies an individual degree of shrinkage on each element of $\tilde{\beta}_{i0}$. The hyperparameter λ_i^2 serves as an equation-specific shrinkage parameter that shrinks all elements of $\tilde{\beta}_{i0}$ that belong to the i th equation towards zero while the local shrinkage parameters τ_{ij} provide enough flexibility to also allow for non-zero values of $\tilde{\beta}_{ij,0}$ in the presence of a tight equation-specific prior. For the equation-specific scaling parameter λ_i^2 we impose a Gamma prior, $\lambda_i^2 \sim \mathcal{G}(b_0, b_1)$, with b_0 and b_1 being hyperparameters chosen by the researcher. In typical applications we specify b_0 and b_1 to render this prior effectively non-influential.

If the innovation variances of the observation equation are assumed to be constant over time, we impose a Gamma prior on σ_i^{-2} with hyperparameters c_0 and c_1 , i.e., $\sigma_i^{-2} \sim \mathcal{G}(c_0, c_1)$. By contrast, if stochastic volatility is introduced, we follow [Kastner and Frühwirth-Schnatter \(2014\)](#) and impose a normally distributed prior on μ_i with mean zero and variance 100, a Beta prior on ρ_i with $(\rho_i + 1)/2 \sim \mathcal{B}(a_\rho, b_\rho)$, and a Gamma distributed prior on $\zeta_i \sim \mathcal{G}(1/2, 1/(2B_\zeta))$.

In the paper at hand, we only estimate the slab variance $\vartheta_{ij,1}$ from the data and set $\vartheta_{ij,0} = \xi \times \hat{\vartheta}_{ij}$, where $\hat{\vartheta}_{ij}$ denotes the variance of the OLS estimate for automatic scaling which we treat as a constant specified a priori. The multiplier ξ is set to a fixed constant close to zero, effectively turning off any time-variation in the parameters. As long as $\vartheta_{ij,0}$ is not chosen too large, the specific value of the spike variance proves to be rather non-influential in the empirical applications that follow. Note that in principle, also the spike variance $\vartheta_{ij,0}$ could be estimated from the data and a suitable shrinkage prior could be employed to push $\vartheta_{ij,0}$ towards zero.

We use an Inverse-Gamma prior on the slab innovation variances in the state specification, i.e., $\vartheta_{ij,1}^{-1} \sim \mathcal{G}(r_{ij,0}, r_{ij,1})$ for $i = 1, \dots, m$ and $j = 1, \dots, K_i$.⁴ Again, $r_{ij,0}$ and $r_{ij,1}$ denote scalar hyperparameters. This choice implies that we artificially bound $\vartheta_{ij,1}$ away from zero, implying that in the upper regime we do not exert strong shrinkage. This is in contrast to a standard time-varying parameter model, where this prior is usually set rather tight to control the degree of time variation in the parameters (see, e.g., [Primeri, 2005](#)). It is noteworthy that in our model the degree of time variation is governed

⁴Of course, it would also be possible to use a (restricted) Gamma prior on $\vartheta_{ij,1}$ in the spirit of [Frühwirth-Schnatter and Wagner \(2010\)](#). However, we have encountered some issues with such a prior if the number of observations in the regime associated with $s_{ij,t} = 1$ is small. This stems from the fact that the corresponding conditional posterior distribution is generalized inverse Gaussian, a distribution that can be heavy tailed and under certain conditions leads to excessively large draws of $\vartheta_{ij,1}$.

by the thresholding mechanism instead.

Finally, the prior specification of the baseline model is completed by imposing a uniform distributed prior on the thresholds,

$$d_{ij} \sim \mathcal{U}(\pi_{ij,0}, \pi_{ij,1}), \quad j = 1, \dots, K_i. \quad (2.13)$$

Here, $\pi_{ij,0}$ and $\pi_{ij,1}$ denote the boundaries of the prior that have to be specified carefully. In our examples, we use $\pi_{ij,0} = 0.1 \times \sqrt{\vartheta_{ij,1}}$ and $\pi_{ij,1} = 1.5 \times \sqrt{\vartheta_{ij,1}}$. This prior bounds the thresholds away from zero, implying that a certain amount of shrinkage is always imposed on the autoregressive coefficients. Setting $\pi_{ij,0} = 0$ for all i, j would also be a feasible option but we found in simulations that being slightly informative on the presence of a threshold improves the empirical performance of the proposed model markedly. It is worth noting that even under the assumption that $\pi_{0j} > 0$, our framework performs well in simulations where the data is obtained from a non-thresholded version of our model. This stems from the fact that in a situation where parameters are expected to evolve smoothly over time, the average period-on-period change of $\beta_{ij,t}$ is small, implying that $0.1 \times \sqrt{\vartheta_{ij,1}}$ is close to zero and the model effectively shrinks small parameter movements to zero.

2.5 Posterior simulation

We sample from the joint posterior distribution of the model parameters by utilizing an MCMC algorithm. Conditional on the thresholds d_{ij} , the remaining parameters can be simulated in a straightforward fashion. After initializing the parameters using suitable starting values we iterate between the following six steps.

1. We start with equation-by-equation simulation of the full history $\{\tilde{\beta}_{it}\}_{t=0,1,\dots,T}$ for each i by means of a standard forward filtering backward sampling algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) while conditioning on the remaining parameters of the model.
2. The reciprocals of the slab innovation variances, $\vartheta_{ij,1}^{-1}$, $i = 1, \dots, m$, $j = 1, \dots, K_i$, have conditional density

$$p(\vartheta_{ij,1}^{-1}|\bullet) = p(\vartheta_{ij,1}^{-1}|d_{ij}, \tilde{\beta}_{ij,0:T}) \propto p(\tilde{\beta}_{ij,0:T}|\vartheta_{ij,1}^{-1}, d_{ij})p(d_{ij}|\vartheta_{ij,1}^{-1})p(\vartheta_{ij,1}^{-1}), \quad (2.14)$$

where $\tilde{\beta}_{ij,0:T} = (\tilde{\beta}_{ij,0}, \dots, \tilde{\beta}_{ij,T})'$. This is a Gamma distribution, i.e.,

$$\vartheta_{ij,1}^{-1} | \bullet \sim \mathcal{G} \left(r_{ij,0} + \frac{T_{ij,1}}{2} + \frac{1}{2}, r_{ij,1} + \frac{\sum_{t=1}^T s_{ij,t} (\tilde{\beta}_{ij,t} - \tilde{\beta}_{ij,t-1})^2}{2} \right), \quad (2.15)$$

with $T_{ij,1} = \sum_{t=1}^T s_{ij,t}$ denoting the number of time periods that feature time variation in the j th parameter and the i th equation.

3. Combining the Gamma prior on τ_{ij}^2 with the Gaussian likelihood yields a Generalized Inverted Gaussian (GIG) distribution

$$\tau_{ij}^2 | \bullet \sim \mathcal{GIG} \left(a_i - \frac{1}{2}, \tilde{\beta}_{ij,0}^2, a_i \lambda_i^2 \right), \quad (2.16)$$

where the density of $\mathcal{GIG}(\kappa, \chi, \psi)$ is proportional to $z^{\kappa-1} \exp \{ -(\chi/z + \psi z)/2 \}$. To sample from this distribution, we use the R package `GIGrvg` (Leydold and Hörmann, 2015) implementing the efficient rejection sampler proposed by Hörmann and Leydold (2013) for each i and j .

4. For each i , the global shrinkage parameter λ_i^2 is sampled from a Gamma distribution given by

$$\lambda_i^2 | \bullet \sim \mathcal{G} \left(b_0 + a_i K_i, b_1 + \frac{a_i}{2} \sum_{j=1}^{K_i} \tau_{ij}^2 \right). \quad (2.17)$$

5. We update the thresholds by applying K_i Griddy Gibbs steps (Ritter and Tanner, 1992) per equation. Due to the structure of the model, the conditional distribution of $\Delta \tilde{\beta}_{ij,1:T}$ is multivariate Gaussian, i.e.

$$p \left(\Delta \tilde{\beta}_{ij,1:T} | d_{ij}, \vartheta_{ij,0}, \vartheta_{ij,1} \right) \propto \prod_{t=1}^T \frac{1}{\sqrt{2\pi\theta_{ij,t}}} \exp \left\{ -\frac{(\tilde{\beta}_{ij,t} - \tilde{\beta}_{ij,t-1})^2}{2\theta_{ij,t}} \right\}. \quad (2.18)$$

This expression can be straightforwardly combined with the prior in Eq. (2.13) to evaluate the conditional posterior of d_{ij} at a given candidate point. The procedure is repeated over a fine grid of values that is determined by the prior and an approximation to the inverse cumulative distribution function of the posterior is constructed.⁵ Finally, this approximation is used to perform inverse transform sampling.

⁵In all applications, we use an evenly spaced grid that contains 150 grid points.

6. The coefficients of each of the the log-volatility equations and the corresponding histories of the log-volatilities are sampled as in [Kastner and Frühwirth-Schnatter \(2014\)](#) through the R package `stochvol` ([Kastner, 2016](#)). Under homoscedasticity, σ_i^{-2} is simulated from $\sigma_i^{-2}|\bullet \sim \mathcal{G}\left(c_0 + T/2, c_1 + \sum_{t=1}^T (y_{it} - \mathbf{z}'_{it}\tilde{\beta}_{it})^2/2\right)$.

After obtaining an appropriate number of draws, we discard the burn-in and base our inference on the remaining draws from the joint posterior.

In comparison with standard TVP-VARs, Step (5) is the only additional MCMC step needed to estimate the proposed TTVP model. Moreover, this update is computationally cheap, increasing the amount of time needed to carry out the analysis conducted in Section 5 by around five percent. For larger models (i.e., with m being around 15) this step becomes slightly more intensive but, relative to the additional computational burden introduced by applying the FFBS algorithm in Step (1), its costs are still comparably small relative to the overall computation time needed.

In the applications that follow, we draw 30 000 samples and discard the first 25 000 draws as burn-in.⁶ Generally speaking, mixing and convergence properties of our proposed algorithm are similar to standard Bayesian TVP-VAR estimators. In other words, the sampling of the thresholds does not seem to substantially increase the autocorrelation of the remaining MCMC draws. Concerning the threshold parameters themselves, we also observe rapid mixing. Appendix A provides some selected convergence criteria for the application to US macroeconomic data.

3 An illustrative example

In this section, we illustrate our approach by means of a rather stylized example that emphasizes how well the mixture innovation component for the state innovations performs when applied to different simulated scenarios.

For demonstration purposes it proves to be convenient to work with the following simple data generating process (DGP) with $K = 1$ and $m = 1$:

$$\begin{aligned} y_t &= x'_{11,t}\beta_{11,t} + u_{1t}, \quad u_{1t} \sim \mathcal{N}(0, 0.1^2), \\ \beta_{11,t} &= \beta_{11,t-1} + e_{11,t}, \quad e_{11,t} \sim \mathcal{N}(0, s_{11,t} \times 0.1^2), \end{aligned}$$

⁶In all applications considered we found that already a small fraction of this burn-in appears to be sufficient to achieve convergence to the stationary distribution.

where $s_{11,t} \in \{0, 1\}$ is chosen to yield paths which are characterized by no ($s_{11,t} \equiv 0$ for all t), few, and many breaks, as well as a standard TVP DGP ($s_{11,t} \equiv 1$ for all t). Finally, independently for all $t = 1, \dots, 500$, we generate $x_{11,t} \sim \mathcal{U}(-1, 1)$.

In order to assess how different models perform in recovering the latent processes, we run a standard TVP model, a mixture innovation model estimated using the algorithm outlined in [Gerlach, Carter, and Kohn \(2000\)](#), and our TTVP model. To ease comparison between the models, we impose a similar prior setup for all models. Specifically, for σ_1^{-2} we set $c_0 = 0.01$ and $c_1 = 0.01$, implying a rather vague prior. For the shrinkage part on $\beta_{11,0}$ we set $\lambda_1^2 \sim \mathcal{G}(0.01, 0.01)$ and $a_1 = 0.1$, effectively applying heavy shrinkage on the initial state of the system. The prior on $\vartheta_{11,1}$ is specified as in [Nakajima and West \(2013a\)](#), i.e., $\vartheta_{11,1}^{-1} \sim \mathcal{G}(3, 0.03)$. To complete the prior setup for the TTVP model we set $\pi_{11,0} = 0.1 \times \sqrt{\vartheta_{11,1}}$ and $\pi_{11,1} = 1.5 \times \sqrt{\vartheta_{11,1}}$. Finally, ξ is set equal to 0.01.

[Fig. 1 about here.]

[Fig. 1](#) displays the evolution of the 98 percent posterior credible intervals of the latent state vectors for standard TVP models (gray), mixture innovation models (blue) and TTVP models (red) along with the actual evolution of the state vector (green). Each panel of [Fig. 1](#) is based on a single realization from the DGP.

At least three interesting findings emerge. First, note that our approach captures parameter movements rather well, signaling large jumps for virtually all time points that feature a structural break in the corresponding parameter. The TVP model also tracks the actual movement of the states well but does so with much more high frequency variation. This is a direct consequence of the inverted Gamma prior on the state innovation variances that, artificially, rules out cases where $\vartheta_{11,1}$ equals zero, irrespective of the information contained in the likelihood (see [Frühwirth-Schnatter and Wagner, 2010](#), for a general discussion of this issue).

Second, investigating posterior uncertainty reveals that our approach succeeds in shrinking the posterior variance. This is due to the fact that in periods where the true value of $\beta_{11,t}$ is constant, our model successfully assumes that the estimate of the coefficient at time t is also constant, whereas the TVP model imposes a certain amount of time variation. This generates additional uncertainty that inflates the posterior variance, possibly leading to imprecise inference. The standard mixture innovation model

is also capable of reducing uncertainty effectively, but at a much larger computational cost as compared to our proposed modeling approach.

Third, contrasting the findings of the TTVP specification with the results obtained from a standard mixture innovation model provides some evidence that our approximation works rather well if the DGP is characterized by not more than a moderate amount of jumps. By contrast, if the DGP features sustained movement, our approach pushes the majority of the high frequency variations to zero. This stems from the fact that we are slightly informative on the specific value of the threshold, effectively ruling out the case where the threshold is zero. Notice, however, that while the posterior distribution of the standard mixture innovation model converges to the posterior distribution of the TVP specification, the corresponding posterior uncertainty also rises sharply. We conjecture that especially in forecasting applications, capturing large swings in the parameters could be sufficient to adequately describe key relations in the data while the reduction in parameter uncertainty could ultimately lead to more precise predictions.

To sum up, the TTVP model detects change points in the parameters in situations where the actual number of breaks is small, moderate and large. In situations where the DGP suggests that the actual threshold equals zero, our approach still captures most of the medium to low frequency noise but shrinks small movements that might, in any case, be less relevant for econometric inference. For cases characterized by a low number of breaks, our proposed approximation closely tracks the standard mixture innovation approach, whereas small but frequent movements tend to be shrunk more rigorously. We conclude by noting that this set of illustrating examples provides evidence in favor of our approach within a small-dimensional setting only. However, comparing TTVP to a standard mixture innovation model in high dimensions is unwieldy due to the tremendous additional computational burden.

4 Forecasting the US term structure of interest rates

The first empirical application deals with predicting the US term structure of interest rates. Forecasting the term structure appears to be an important task for policy makers and practitioners alike. Central banks are interested in how their policy interventions impact the different segments of the term structure and how these movements transmit into the real economy. From a forecasting perspective, precise predictions are necessary for various tasks such as active portfolio management, risk management, as well as

general policy analysis.

Several attempts have been made to predict the term structure of interest rates with a wide range of different models (see, among many others, [Diebold and Li, 2006](#); [Mönch, 2008, 2012](#); [Favero, Niu, and Sala, 2012](#); [Carriero, Kapetanios, and Marcellino, 2012](#); [Carriero, Clark, and Marcellino, 2014](#); [Byrne, Cao, and Korobilis, 2017](#)). The majority of these contributions, however, focuses exclusively on evaluating point predictions while ignoring higher moments of the underlying predictive distribution. In addition, most studies typically assume that the model parameters are constant over time. Some recent exceptions are [Bianchi, Mumtaz, and Surico \(2009\)](#); [Mumtaz and Surico \(2009\)](#); [Koopman, Mallee, and Van der Wel \(2010\)](#); [Carriero, Clark, and Marcellino \(2014\)](#); [Byrne, Cao, and Korobilis \(2017\)](#). In the paper at hand, we apply the TTVP model to predict the term structure of interest rates and benchmark it against various competing model specifications.

4.1 Data overview, model specification, and design of the forecasting exercise

We use monthly Fama-Bliss zero coupon yields obtained from the Chicago Booth Center for Research in Security prices (CRSP) database as well as the dataset described in [Gürkaynak, Sack, and Wright \(2007\)](#). The data spans the period from 1960:M01 to 2014:M12 and the maturities included are 1, 2, 3, 4, 5, 7, and 10 years.⁷ Moreover, we include $P = 3$ lags of the endogenous variables. The prior setup is similar to the one adopted in the previous section. More specifically, for all applicable i and j , we use the following values for the hyperparameters. For the shrinkage part on the initial state of the system, we again set $\lambda_i^2 \sim \mathcal{G}(0.01, 0.01)$ and $a_i = 0.1$. For the parameters of the log-volatility equation we use $\mu_i \sim \mathcal{N}(0, 10^2)$, $\frac{\rho_i+1}{2} \sim \mathcal{B}(25, 5)$, and $\zeta_i \sim \mathcal{G}(1/2, 1/2)$. The prior on the thresholds are set equal to $\pi_{ij,0} = 0.1 \times \sqrt{\vartheta_{ij,1}}$ and $\pi_{ij,1} = 1.5 \times \sqrt{\vartheta_{ij,1}}$. Finally, we assess the impact of different choices for the prior on $\vartheta_{ij,1}$ as well as ξ in [Table 1](#).

Our forecasting design is recursive. For an initial estimation period, in our case 1960:M01 to $T_0 = 1999:M08$, we compute predictions for the next three months via Monte Carlo integration. More concretely, for each of the $l = 1, \dots, 5000$ MCMC draws from the posterior distribution, we start by predicting $\beta_{T_0+n}^{(l)}$ for $n = 1, 2, 3$. This is achieved by first drawing an indicator $s_{ij,T_0+n}^{(l)}$ for all i, j from a Bernoulli distribution

⁷The data for the maturities one up to five years are based on the CRSP data while the data for maturities seven and ten years are taken from the [Gürkaynak, Sack, and Wright \(2007\)](#) database.

with success probability $T_{ij,1}^{(l)}/T_0$. Using this indicator, we compute $\theta_{ij,T_0+n}^{(l)}$ and thus infer whether the predicted change is effectively zero or not. This information enables us to recursively calculate $\beta_{T_0+n}^{(l)}$ utilizing the state evolution in Eq. (2.2). Next, we construct $V_{T_0+n}^{(l)}$ through the corresponding elements in $\beta_{T_0+n}^{(l)}$ and then predict the log-volatilities using Eq. (2.9) to compute $\Sigma_{T_0+n}^{(l)}$. Finally, we obtain predictions for y_{T_0+n} by drawing from $\mathcal{N}(x'_{T_0+n}\beta_{T_0+n}^{(l)}, \Sigma_{T_0+n}^{(l)})$.

After obtaining these, we expand the initial estimation sample by one month and repeat this procedure until the end of the sample is reached. This yields a sequence of 184 multivariate predictive densities for each of the three predictive horizons. Forecasts are then evaluated using log predictive scores (LPSs) for a model of interest and a benchmark model. The LPS is a widely used metric to measure density forecast accuracy (see, e.g., Geweke and Amisano, 2010).

4.2 Competing models

As the benchmark model, we use a TVP-VAR with SV employing the prior setup described in Primiceri (2005). We, moreover, include three additional constant parameter VAR models, namely a Minnesota-type VAR (Doan, Litterman, and Sims, 1984),⁸ a Normal-Gamma (NG) VAR (Huber and Feldkircher, 2018) and a VAR coupled with a stochastic search variable selection (SSVS) prior (George, Sun, and Ni, 2008). For the SSVS VAR, we set the scaling parameters associated with the two Gaussian mixture components of the prior using the semi-automatic approach described in George, Sun, and Ni (2008). This implies that the prior standard deviation for the slab component is ten times the corresponding OLS standard deviation while for the spike component it is one tenth of the OLS standard deviation.

Moreover, and given its success in forecasting the term structure, we also benchmark our unrestricted multivariate model specifications against the model proposed in Diebold and Li (2006) based on the three factor Nelson Siegel (NS) framework (Nelson and Siegel, 1987). The NS approach imposes a factor structure on the yields,

$$i_t(\varrho) = L_t + \frac{1 - e^{-\varrho\alpha}}{\varrho\alpha} S_t + \left(\frac{1 - e^{-\varrho\alpha}}{\varrho\alpha} - e^{-\varrho\alpha} \right) C_t + m_t(\varrho). \quad (4.1)$$

Here, $i_t(\varrho)$ denotes the yield at maturity ϱ , L_t is a factor that controls the level, S_t deter-

⁸This specific implementation follows Koop and Korobilis (2010) but, following Giannone, Lenza, and Primiceri (2012), estimates the hyperparameters using two Metropolis Hastings steps.

mines the slope, and C_t represents the curvature factor of the yield curve. Moreover, we let $m_t(\varrho)$ denote a pricing error. The parameter α controls the shape of the factor loadings in Eq. (4.1) and is set to $\alpha = 0.0609$ to maximize the loading on C_t (for a discussion of this particular choice, see Diebold and Li, 2006). In what follows, we estimate the latent factors L_t , S_t , and C_t by OLS. These factors are then included in $\mathbf{y}_t = (L_t, S_t, C_t)'$ and a VAR with a Minnesota prior is estimated (labeled NS-VAR). Moreover, we also estimate a TTVP-NS-VAR model to assess whether allowing for time-variation in the state equation of the factors pays off.⁹ Notice that since Eq. (4.1) is a standard measurement equation, we forecast the yield curve by using the VAR state equation to compute the predictions in terms of the factors and then map it back to the yields using the factor loadings.

All models, including the random walk, feature stochastic volatility. In order to assess the impact of different prior hyperparameters on $\vartheta_{ij,1}$ and the impact of ξ , we estimate the TTVP model over a grid of meaningful values.

4.3 Forecasting results

Table 1 shows the results for one-month- and three-months-ahead forecasts. We start by considering the joint forecasting performance, provided in the rightmost column of Table 1, for the one-month-ahead forecast horizon. Here we observe that all models improve upon the standard TVP-VAR with SV by large margins. This clearly suggests that a standard TVP-VAR model equipped with inverted Gamma priors on the state innovation variances seems to overfit the data which, in turn, translates into a weak out-of-sample predictive performance. Comparing the predictive performance of our TTVP model across different choices for ξ reveals that this parameter appears to be highly influential. If the hyperparameter is set too large, too little shrinkage is introduced and the forecasting performance deteriorates. Considering the different choices of ξ shows that smaller values are typically accompanied by larger improvements in LPSs. Notice that the choice of r_0 and r_1 tends to play only a minor role compared to the scaling parameter ξ .

We first focus on the evaluation of the joint density forecasts over all maturities and the one-step ahead forecast horizon. Here we see that the TTVP approach improves forecast performance over all constant parameter VARs as well as the models based

⁹For this specification, we use the prior setup described above and set $r_0 = 3$, $r_1 = 0.03$ and $\xi = 0.001$.

Table 1: Log predictive Bayes factors relative to TVP-VAR over the hold-out period 1999:M09 to 2014:M12. Numbers greater than zero indicate that a given model outperforms the benchmark. The final column refers to the joint density forecasting performance, while the other columns refer to the univariate margins. Bold figures indicate the best performing model for each column and horizon.

	1Y	2Y	3Y	4Y	5Y	7Y	10Y	joint
One-month-ahead								
TTVP-VAR: $\xi = 0.1, r_0 = 3, r_1 = 0.03$	-107.1	-92.1	-67.0	-41.3	11.5	22.0	10.8	2595.2
TTVP-VAR: $\xi = 0.1, r_0 = 1.5, r_1 = 1$	-159.0	-122.5	-79.7	-40.8	15.0	40.5	33.0	2713.0
TTVP-VAR: $\xi = 0.1, r_0 = 0.001, r_1 = 0.001$	-115.4	-100.9	-73.8	-44.7	8.7	23.8	14.0	2696.2
TTVP-VAR: $\xi = 0.01, r_0 = 3, r_1 = 0.03$	-22.3	-10.4	20.2	50.3	97.9	61.1	46.7	2927.3
TTVP-VAR: $\xi = 0.01, r_0 = 1.5, r_1 = 1$	25.3	21.0	40.2	64.9	106.1	70.7	55.3	2943.2
TTVP-VAR: $\xi = 0.01, r_0 = 0.001, r_1 = 0.001$	-6.3	6.2	32.5	59.0	102.4	74.0	59.5	2899.7
TTVP-VAR: $\xi = 0.001, r_0 = 3, r_1 = 0.03$	28.5	20.7	40.3	65.3	108.3	74.0	60.0	2976.1
TTVP-VAR: $\xi = 0.001, r_0 = 1.5, r_1 = 1$	28.2	21.5	40.4	65.4	107.9	71.7	57.8	2953.3
TTVP-VAR: $\xi = 0.001, r_0 = 0.001, r_1 = 0.001$	28.5	22.0	40.9	65.7	108.3	73.0	59.0	2952.9
NS-TTVP-VAR	-26.6	-17.5	20.5	57.7	110.1	101.0	82.8	2655.8
NS-VAR	-61.6	-39.9	8.8	53.1	109.8	125.0	119.3	2681.0
Minnesota-type VAR	22.7	18.8	26.7	3.6	4.5	138.0	128.7	2660.2
NG VAR	35.4	22.5	19.2	-45.0	-50.6	141.9	136.1	2505.9
SSVS VAR	-30.8	-51.9	-38.5	-13.6	37.8	42.3	41.0	1308.8
Random walk	35.3	22.8	31.6	14.8	20.1	141.0	135.0	2656.0
Three-months-ahead								
TTVP-VAR: $\xi = 0.1, r_0 = 3, r_1 = 0.03$	-279.4	-225.4	-181.0	-164.7	-163.8	-175.1	-148.8	625.2
TTVP-VAR: $\xi = 0.1, r_0 = 1.5, r_1 = 1$	-330.0	-257.7	-194.8	-162.2	-160.4	-164.2	-123.5	809.4
TTVP-VAR: $\xi = 0.1, r_0 = 0.001, r_1 = 0.001$	-283.3	-229.3	-182.2	-161.2	-160.6	-168.0	-137.9	786.8
TTVP-VAR: $\xi = 0.01, r_0 = 3, r_1 = 0.03$	-82.7	-34.3	19.2	40.2	35.6	-14.6	16.3	1214.9
TTVP-VAR: $\xi = 0.01, r_0 = 1.5, r_1 = 1$	8.3	36.4	73.8	85.8	82.7	40.6	71.8	1293.2
TTVP-VAR: $\xi = 0.01, r_0 = 0.001, r_1 = 0.001$	-52.7	0.7	50.1	62.6	56.8	14.8	45.2	1194.1
TTVP-VAR: $\xi = 0.001, r_0 = 3, r_1 = 0.03$	14.8	36.4	72.5	87.2	81.7	41.1	71.9	1333.4
TTVP-VAR: $\xi = 0.001, r_0 = 1.5, r_1 = 1$	15.5	39.0	75.1	88.6	84.3	42.8	73.9	1312.2
TTVP-VAR: $\xi = 0.001, r_0 = 0.001, r_1 = 0.001$	16.0	39.1	75.0	88.5	84.2	43.3	74.3	1313.3
NS-TTVP-VAR	-30.6	5.0	56.8	84.3	82.0	39.1	67.6	1038.5
NS-VAR	-74.4	-20.2	43.7	80.1	83.3	46.5	80.6	1016.3
Minnesota-type VAR	-1.2	36.8	69.1	61.4	32.0	29.2	66.5	1005.7
NG VAR	17.3	40.1	58.4	10.4	-29.7	16.0	59.8	863.0
SSVS VAR	-86.5	-74.2	-44.9	-27.7	-28.2	-55.8	-20.3	-614.7
Random walk	17.0	39.4	68.4	54.5	17.9	28.7	69.1	975.7

on the NS approach. The NS-VAR ranks second while the VAR with the hierarchical Minnesota prior ranks third. Using the TTVP framework in combination with the NS factors also produces predictions that are competitive to the remaining constant parameter specifications. Contrasting the differences between the random walk and the Minnesota prior shows that both yield similar predictions. This is because the hierarchical Minnesota prior exerts strong shrinkage towards a random walk process.

Considering the three-months-ahead joint predictive densities yields similar insights. The TTVP models continue to fare well, outperforming both the TVP-VAR with SV, the constant parameter VARs as well as the NS models. It is noteworthy that for multi-step

forecasting, the SSVS VAR shows the weakest performance across all models considered, leading to a forecast performance that is even inferior compared to the benchmark.

Zooming into the marginal results for the different maturities shows that for the short as well as for the long end of the yield curve, most constant parameter models seem to outperform their time-varying parameter competitors and the models based on the NS factors for the one-month-ahead forecast horizon. For 1Y and 2Y maturities, the random walk as well as the NG VAR generate the most precise predictions, improving slightly upon the single best performing TTVP specification. The particularly strong predictive performance of the random walk for the short end of the yield curve has been found in several contributions on predicting interest rates (Diebold and Li, 2006; Carriero, Kapetanios, and Marcellino, 2012). This finding, however, does not carry over to maturities between three and five years. There, we find that our proposed framework as well as the NS models excel, clearly outperforming the competitors. Interestingly, for five year maturities we find that estimating a small-scale factor model with time-varying parameters and mixture innovations yields the strongest performance. Again, when considering multi-step-ahead forecasts, we find a rather similar picture, with models that perform well in terms of one-step-ahead forecasting also doing well when three-step-ahead forecasts are considered.

[Fig. 2 about here.]

Finally, we investigate whether the predictive performance varies with time. Fig. 2 displays the evolution of the one-step-ahead LPSs vis-à-vis the TVP-VAR with SV specification. At least two interesting patterns over time emerge. We find that during the financial crisis in 2008/2009, all competing models' performance increase sharply against the benchmark specification. In addition, a pronounced jump in relative forecasting performance is also visible during the second half of 2011, a period characterized by the US debt ceiling crisis of 2011. Our conjecture is that this is driven by a) the ability to rapidly adjust regression coefficients and thus allow for changing transmission mechanisms in a flexible way and b) the fact that the TVP-VAR with SV overfits the data severely and appears to be incapable of handling large shocks to the term structure.

To sum up, this section highlights that using our TTVP approach generally pays off when used to predict the US term structure of interest rates. When considering the joint density forecasting performance, we find that this model framework improves

sharply against the competing models used. If the forecaster’s goal is to predict only certain segments of the yield curve, we find that for the short and the long end of the term structure, linear VARs with SV as well as the random walk with SV outperform our modeling approach. For three up to five year maturities, TTVP models with and without a factor structure on the yields outperform all remaining models.

5 Structural breaks in US macroeconomic data

We complement the forecasting exercise by investigating the effects of a monetary policy shock. For that purpose we use a standard US macroeconomic data set, employed among others in [Smets and Wouters \(2007\)](#), [Geweke and Amisano \(2012\)](#) and [Amisano and Geweke \(2017\)](#). Data are on a quarterly basis, span the period from 1947Q2 to 2014Q4, and comprise the log differences of consumption, investment, real GDP, hours worked, consumer prices and real wages. Last, and as a policy variable, we include the Federal Funds Rate (FFR) in levels. The prior setup mirrors the one used in the preceding section but sets $\xi = 0.01/6$.¹⁰ Following [Primiceri \(2005\)](#), we include $P = 2$ lags of the endogenous variables.

In the next section, we start by proposing a global measure of time-variation in the VAR coefficients and then move on to analyze macroeconomic relations by means of impulse response analysis in Section 5.2.

5.1 Detecting time-variation in reduced form coefficients

We start by examining the posterior mean of the determinant of the time-varying variance-covariance matrix of the innovations in the state equation ([Cogley and Sargent, 2005](#)). Using the determinant of the process innovation variance-covariance matrix yields a comprehensive measure of time-variation in the autoregressive parameters. Showing the evolution of the states in β_t itself is, given the large dimensionality of the state space, not feasible. The corresponding measure is computed as follows. For each draw of $\Omega_{it} = \text{diag}(\theta_{i1,t}, \dots, \theta_{iK_i,t})$ we compute the exponential of the demeaned logarithm of the determinant. Large values of this measure point towards a pronounced degree of time-variation in the autoregressive coefficients of the corresponding equations. The results are provided in [Fig. 3](#) for each equation and the full system.

¹⁰This value is based on running a forecasting exercise using this dataset and a hold-out period of 45 years. Specific results are available from the authors upon request.

[Fig. 3 about here.]

For all variables, our model produces at least one prominent spike during the sample period indicating a structural break. Most spikes in the determinant occur around 1980, when then Fed chairman Paul Volcker sharply increased short-term interest rates to fight inflation. Other breaks relate to the dot-com bubble in the early 2000s (consumption), the oil price crisis and stock market crash in the early 1970s (hours worked) and another oil price related crisis in the early 1990s. Also, the transition from positive interest rates to the zero lower bound in the midst of the global financial crisis is indicated by a spike in the determinant. That we can relate spikes to historical episodes of financial and economic distress lends further confidence in the modeling approach. Albeit among these periods, the early 1980s seem to have constituted by far the most severe rupture for the US economy, the analysis reveals several further, variable-dependent structural breaks. A model that assumes common dynamics of the coefficients would not be able to pick these up, which emphasizes the flexibility of the proposed approach.

5.2 Impulse responses to a monetary policy shock

In this section, we examine the dynamic responses of a set of macroeconomic variables to a contractionary monetary policy shock. The monetary policy shock is calibrated as a 100 basis point (bp) increase in the FFR and identified using a Cholesky ordering with the variables appearing in exactly the same order as mentioned above. This ordering is in the spirit of [Christiano, Eichenbaum, and Evans \(2005\)](#) and has been subsequently used in the literature (see [Coibion, 2012](#), for an excellent survey).

We proceed in two stages. First, we show slices of impulse responses for the four-step, eight-step and 12-step-ahead forecast horizon. This allows us to get an overall impression of the time variation in the impulse response functions. In the second stage we zoom in and provide the full set of impulse responses for two subsets of the sample, namely the pre-Volcker period from 1947Q4 to 1979Q1 and the rest of the sample.

All impulse response functions are calculated assuming that the shocks to the states are set to their expected value, i.e., zero. Hence, we follow [Primiceri \(2005\)](#); [Cogley and Sargent \(2005\)](#); [Koop, Leon-Gonzalez, and Strachan \(2009\)](#), among many others, and neglect the fact that parameters might be changing over the impulse response horizon. The assumption that parameters remain constant over the impulse response

horizon is, moreover, consistent with the literature on bounded rationality and learning (see [Kreps, 1998](#)). Compared to dynamic forecasts, this simpler strategy is computationally less involved and in light of the fact that our model detects rather few (but large) structural breaks, appears to be reasonable.

[Fig. 4 about here.]

[Fig. 5 about here.]

In [Fig. 4](#) and [Fig. 5](#), we report the overall effects of a monetary tightening. As investment growth decelerates, the growth rate of hours worked and output decreases. These results are reasonable from an economic perspective. Also, estimated effects on output growth and inflation are comparable to those of [Baumeister and Benati \(2013\)](#) who use a TVP-VAR framework and US data. Results on consumption growth and real wages are accompanied by large credible intervals.

Looking at time variation, the results indicate stronger (i.e., more negative) effects of monetary policy for the most recent part of our sample period. More precisely, and starting in the late 1990s, effects on consumption, investment and output growth gradually decrease until the end of our sample period. Most interestingly, though, are the responses of inflation. They sharply increase in the late 1960s, resulting in a pronounced price puzzle, remain constant in the 1970s and start declining strongly with the onset of the Volcker-period. This pattern is also mirrored in the responses of real wage growth, which is only negative during the period of pronounced inflation effects, while positive during the rest of the sample period. The results in [Fig. 4](#) and [Fig. 5](#) thus reveal time variation in the effects of a monetary policy shock and – more importantly – that these are variable specific and can be both gradual and abrupt.

We next zoom in and focus on two sub-sets of the sample, namely the pre-Volcker period from 1947Q4 to 1979Q1 and the rest of the sample.¹¹ The time-varying impulse responses – as functions of horizons – are displayed in [Fig. 6](#). We investigate whether the size and the shape of responses varies between and within the two sub-samples. For that purpose, we show median responses over the first sample split in the top row and for the second part of the sample in the bottom row. Impulse responses that belong to the beginning of a sample split are depicted in light yellow, those that belong to the end

¹¹The split into two sub-sets is conducted for interpretation purposes only. For estimation, the entire sample has been used.

of the sample period in dark red. To fix ideas, if the size of a response increases continuously over time, we should see a smooth darkening of the corresponding impulse from light yellow to dark red.

[Fig. 6 about here.]

Considering the first sub-period from 1947Q4 to 1979Q1, one of the variables that shows a great deal of variation in magnitudes is the response of inflation. Here, effects become increasingly positive the further one moves from 1947Q4 to 1979Q1 and the shades of the responses turn continuously darker. While overall credible sets for the sub-sample are wide, positive responses for inflation and thus the price puzzle are estimated over the period from the mid-1960s to the beginning of the 1980s (see also [Fig. 4](#)). A similar picture arises when looking at consumption growth. During the first sample split, effects become increasingly more negative, but responses are only precisely estimated for the period from the mid-1960s to the beginning of the 1980s. This might be explained by the fact that the monetary policy driven increase in inflation spurs consumption since saving becomes less attractive.

In the bottom panel of [Fig. 6](#), we focus on the results over the more recent second sample split from 1979Q2 to 2014Q4. Paul Volcker's fight against inflation had some bearings on overall macroeconomic dynamics in the USA. With the onset of the 1980s, the aforementioned price puzzle starts to disappear (in the sense that effects are surrounded by wide credible sets and median responses turn increasingly negative). There is also a great deal of time variation evident in other responses which are mostly becoming more negative. Put differently, the effectiveness of monetary policy seems to be higher in the more recent sample period than before. This can be seen by effects on growth in hours worked, investment and output. That the effects of a hypothetical monetary policy shock on output growth are particular strong after the crisis corroborates findings of [Baumeister and Benati \(2013\)](#) and [Feldkircher and Huber \(2016\)](#). The latter argue that this is related to the zero lower bound period: after a prolonged period of unaltered interest rates, a deviation from the (long-run) interest rate mean can exert considerable effects on the macroeconomy.

6 Closing remarks

This paper puts forth a novel approach to estimating large-scale time-varying parameter models with mixture innovations in a Bayesian framework. We propose approx-

imating the indicators that control which mixture component to use by a threshold process where the threshold variable is the absolute period-on-period change of the corresponding states. This implies that if the (proposed) change is sufficiently large, the corresponding variance is set to a value greater than zero. Otherwise, it is set close to zero which implies that the states remains virtually constant between two points in time. Our framework is capable of discriminating between a plethora of competing specifications, most notably models that feature moderately many, few, or even no structural breaks in the regression parameters.

The merits of our approach are illustrated by two applications. The first application serves as a means to assess the forecasting capabilities of the proposed model while the second model illustrates how the framework can be used to perform structural analysis. In the first application, we show that our model performs well when used to predict the US term structure of interest rates. Our results indicate that the model yields precise forecasts, especially so during more volatile times such as witnessed in 2008 and during the debt ceiling crisis in 2011. For that period, the forecast gain over simpler models is particularly high.

For the second application, we turn to US macroeconomic data. We investigate whether reduced-form parameters vary over time by considering the time-varying determinant of the posterior variance-covariance matrix of the state innovations. This analysis suggests several variable specific structural breaks in the reduced form relationships, with the Volcker period marking the most severe rupture for the US economy. Examining the effects of a contractionary monetary policy shock, we see considerable time variation in structural impulse responses. Our results indicate abrupt changes of effects on inflation. More specifically, we find significant evidence for a severe price puzzle during episodes of the pre-Volcker period, whereas the puzzle disappears in the second half of our sample. Effects on other variables such as output and investment growth as well as growth in hours worked change more gradually, reaching a trough during the period after the global financial crisis. For that period, a hypothetical deviation from the zero lower bound would create pronounced effects on the wider economy. These findings highlight the importance to account for different dynamics of the underlying variables in order to adequately capture the complex interaction of the macroeconomy – a salient feature of our modeling framework.

As a potential avenue of further research, the proposed framework can be extended by introducing another prior on the scaling factors that is used to determine the spike

variance. The scaling factor could be defined on a discrete grid and this, combined with a uniform prior, would lead to a tractable conditional posterior distribution. This would further reduce the additional input from the researcher and improve the practicability of our approach for a wide range of different applications in macroeconomics and finance.

References

- AMISANO, G., AND J. GEWEKE (2017): “Prediction using several macroeconomic models,” *Review of Economics and Statistics*, 99(5), 912–925.
- BAUMEISTER, C., AND L. BENATI (2013): “Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound,” *International Journal of Central Banking*, 9(2), 165–212.
- BELMONTE, M. A., G. KOOP, AND D. KOROBILIS (2014): “Hierarchical Shrinkage in Time-Varying Parameter Models,” *Journal of Forecasting*, 33(1), 80–94.
- BIANCHI, F., H. MUMTAZ, AND P. SURICO (2009): “The great moderation of the term structure of UK interest rates,” *Journal of Monetary Economics*, 56(6), 856–871.
- BITTO, A., AND S. FRÜHWIRTH-SCHNATTER (forthcoming): “Achieving shrinkage in a time-varying parameter model framework,” *Journal of Econometrics*.
- BYRNE, J. P., S. CAO, AND D. KOROBILIS (2017): “Forecasting the term structure of government bond yields in unstable environments,” *Journal of Empirical Finance*, 44, 209–225.
- CARRIERO, A., T. CLARK, AND M. MARCELLINO (2014): “No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates,” *CEPR Discussion Paper No. DP9848*.
- CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2012): “Forecasting government bond yields with large Bayesian vector autoregressions,” *Journal of Banking & Finance*, 36(7), 2026–2047.
- CARTER, C. K., AND R. KOHN (1994): “On Gibbs sampling for state space models,” *Biometrika*, 81(3), 541–553.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 113(1), 1–45.
- COGLEY, T., AND T. J. SARGENT (2002): “Evolving post-world war II US inflation dynamics,” in *NBER Macroeconomics Annual 2001, Volume 16*, pp. 331–388. MIT Press.
- (2005): “Drifts and volatilities: monetary policies and outcomes in the post WWII US,” *Review of Economic Dynamics*, 8(2), 262–302.
- COIBION, O. (2012): “Are the Effects of Monetary Policy Shocks Big or Small?,” *American Economic Journal: Macroeconomics*, 4(2), 1–32.
- D’AGOSTINO, A., L. GAMBETTI, AND D. GIANNONE (2013): “Macroeconomic forecasting and structural change,” *Journal of Applied Econometrics*, 28(1), 82–101.
- DIEBOLD, F. X., AND C. LI (2006): “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130(2), 337–364.
- DOAN, T. R., B. R. LITTERMAN, AND C. A. SIMS (1984): “Forecasting and Conditional Projection Using Realistic Prior Distributions,” *Econometric Reviews*, 3, 1–100.

- EISENSTAT, E., J. C. CHAN, AND R. W. STRACHAN (2016): “Stochastic model specification search for time-varying parameter VARs,” *Econometric Reviews*, 35(8-10), 1638–1665.
- FAVERO, C. A., L. NIU, AND L. SALA (2012): “Term structure forecasting: no-arbitrage restrictions versus large information set,” *Journal of Forecasting*, 31(2), 124–156.
- FELDKIRCHER, M., AND F. HUBER (2016): “Unconventional US Monetary Policy: New Tools, Same Channels?,” Working Papers 208, Oesterreichische Nationalbank (Austrian Central Bank).
- FRÜHWIRTH-SCHNATTER, S. (1994): “Data augmentation and dynamic linear models,” *Journal of time series analysis*, 15(2), 183–202.
- FRÜHWIRTH-SCHNATTER, S., AND H. WAGNER (2010): “Stochastic model specification search for Gaussian and partial non-Gaussian state space models,” *Journal of Econometrics*, 154(1), 85–100.
- GEORGE, E. I., D. SUN, AND S. NI (2008): “Bayesian stochastic search for VAR model restrictions,” *Journal of Econometrics*, 142(1), 553–580.
- GERLACH, R., C. CARTER, AND R. KOHN (2000): “Efficient Bayesian inference for dynamic mixture models,” *Journal of the American Statistical Association*, 95(451), 819–828.
- GEWEKE, J., AND G. AMISANO (2010): “Comparing and evaluating Bayesian predictive distributions of asset returns,” *International Journal of Forecasting*, 26(2), 216–230.
- (2012): “Prediction with misspecified models,” *The American Economic Review*, 102(3), 482–486.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2012): “Prior selection for vector autoregressions,” Discussion paper, National Bureau of Economic Research.
- GIORDANI, P., AND R. KOHN (2008): “Efficient Bayesian inference for multiple change-point and mixture innovation models,” *Journal of Business & Economic Statistics*, 26, 66–77.
- GRIFFIN, J. E., AND P. J. BROWN (2010): “Inference with normal-gamma prior distributions in regression problems,” *Bayesian Analysis*, 5(1), 171–188.
- GÜRKAYNAK, R. S., B. SACK, AND J. H. WRIGHT (2007): “The US Treasury yield curve: 1961 to the present,” *Journal of Monetary Economics*, 54(8), 2291–2304.
- HÖRMANN, W., AND J. LEYDOLD (2013): “Generating generalized inverse Gaussian random variates,” *Statistics and Computing*, 24(4), 1–11.
- HUBER, F., AND M. FELDKIRCHER (2018): “Adaptive shrinkage in Bayesian vector autoregressive models,” *Journal of Business and Economic Statistics*, forthcoming.
- KALLI, M., AND J. E. GRIFFIN (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., AND S. FRÜHWIRTH-SCHNATTER (2014): “Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models,” *Computational Statistics & Data Analysis*, 76, 408–423.
- KIMURA, T., AND J. NAKAJIMA (2016): “Identifying conventional and unconventional monetary policy shocks: a latent threshold approach,” *The BE Journal of Macroeconomics*, 16(1), 277–300.
- KOOP, G., AND D. KOROBILIS (2010): “Bayesian multivariate time series methods for empirical macroeconomics,” *Foundations and Trends in Econometrics*, 3(4), 267–358.

- (2012): “Forecasting inflation using dynamic model averaging,” *International Economic Review*, 53(3), 867–886.
- (2013): “Large time-varying parameter VARs,” *Journal of Econometrics*, 177(2), 185–198.
- KOOP, G., R. LEON-GONZALEZ, AND R. W. STRACHAN (2009): “On the evolution of the monetary policy transmission mechanism,” *Journal of Economic Dynamics and Control*, 33(4), 997–1017.
- KOOP, G., AND S. M. POTTER (2007): “Estimation and forecasting in models with multiple breaks,” *The Review of Economic Studies*, 74(3), 763–789.
- KOOPMAN, S. J., M. I. MALLEE, AND M. VAN DER WEL (2010): “Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters,” *Journal of Business & Economic Statistics*, 28(3), 329–343.
- KREPS, D. M. (1998): “Anticipated utility and dynamic choice,” *Econometric Society Monographs*, 29, 242–274.
- LEYDOLD, J., AND W. HÖRMANN (2015): *GIGrv: Random variate generator for the GIG distribution* R package version 0.4.
- MAHEU, J. M., AND Y. SONG (2018): “An efficient Bayesian approach to multiple structural change in multivariate time series,” *Journal of Applied Econometrics*, 33, 251–270.
- MCCULLOCH, R. E., AND R. S. TSAY (1993): “Bayesian inference and prediction for mean and variance shifts in autoregressive time series,” *Journal of the American Statistical Association*, 88(423), 968–978.
- MÖNCH, E. (2008): “Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach,” *Journal of Econometrics*, 146(1), 26–43.
- (2012): “Term structure surprises: the predictive content of curvature, level, and slope,” *Journal of Applied Econometrics*, 27(4), 574–602.
- MUMTAZ, H., AND P. SURICO (2009): “Time-varying yield curve dynamics and monetary policy,” *Journal of Applied Econometrics*, 24(6), 895–913.
- NAKAJIMA, J., AND M. WEST (2013a): “Bayesian analysis of latent threshold dynamic models,” *Journal of Business & Economic Statistics*, 31(2), 151–164.
- (2013b): “Dynamic factor volatility modeling: A Bayesian latent threshold approach,” *Journal of Financial Econometrics*, 11(1), 116–153.
- NEELON, B., AND D. B. DUNSON (2004): “Bayesian Isotonic Regression and Trend Analysis,” *Biometrics*, 60(2), 398–406.
- NELSON, C. R., AND A. F. SIEGEL (1987): “Parsimonious modeling of yield curves,” *Journal of Business*, 60(4), 473–489.
- PRIMICERI, G. E. (2005): “Time varying structural vector autoregressions and monetary policy,” *The Review of Economic Studies*, 72(3), 821–852.
- RAFTERY, A. E., AND S. LEWIS (1992): “How many iterations in the Gibbs sampler?,” in *Bayesian Statistics 4*, ed. by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, pp. 763–773. Oxford University Press, Oxford.
- RITTER, C., AND M. A. TANNER (1992): “Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler,” *Journal of the American Statistical Association*, 87(419), 861–868.
- ROČKOVÁ, V., AND K. MCALINN (2018): “Dynamic Variable Selection with Spike-and-Slab Process Priors,” *Working Paper*.
- SIMS, C. A., AND T. ZHA (2006): “Were there regime switches in US monetary policy?,”

- The American Economic Review*, 96(1), 54–81.
- SMETS, F., AND R. WOUTERS (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *The American Economic Review*, 97(3), 586–606.
- STOCK, J. H., AND M. W. WATSON (1996): “Evidence on structural instability in macroeconomic time series relations,” *Journal of Business & Economic Statistics*, 14(1), 11–30.
- URIBE, P. V., AND H. F. LOPES (2017): “Dynamic sparsity on dynamic regression models,” Technical report.
- ZHOU, X., J. NAKAJIMA, AND M. WEST (2014): “Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models,” *International Journal of Forecasting*, 30(4), 963–980.

A Convergence and mixing properties

Here, we assess convergence of our proposed algorithm for the US macroeconomic dataset. As mentioned in the main part of the paper, convergence characteristics closely resemble those typically reported when standard TVP-VARs with SV are used. To assess mixing and convergence properties of the thresholds, [Table 2](#) shows the empirical distribution of inefficiency factors and the [Raftery and Lewis \(1992\)](#) diagnostic of the total number of runs required to achieve a certain level of precision. The parameters of the diagnostic are specified as in [Primiceri \(2005\)](#).¹²

The table indicates that inefficiency factors across equations appear to be favorable (i.e., well below 50) for all covariates. Notice that the marginal posteriors of selected thresholds feature estimated inefficiency factors of one, indicating virtually no autocorrelation. Considering the required number of runs to achieve a certain level of precision reveals that this is far below the actual number of iterations in practically all cases.

Table 2: Empirical distribution across covariates within an equation of selected convergence metrics for the thresholds: US macroeconomic data.

	Inefficiency factors					Required number of runs				
	Low ₁₀	Median	High ₉₀	Min	Max	Low ₁₀	Median	High ₉₀	Min	Max
consumption	1	1	5	1	7	928	1375	2913	907	3945
investment	1	7	13	1	31	2222	3112	4725	1162	4746
output	1	1	1	1	2	922	968	2788	907	6360
hours	1	1	2	1	7	928	2129	3538	907	5064
inflation	1	1	14	1	16	1094	2270	4319	922	4780
real wage	1	9	19	1	49	1415	2825	4606	1242	5415
interest rate	1	1	10	1	19	922	1242	2896	922	3390

¹²The quantiles are set equal to 0.025, the desired degree of accuracy is 0.025, and the probability of achieving the required accuracy is 0.95.

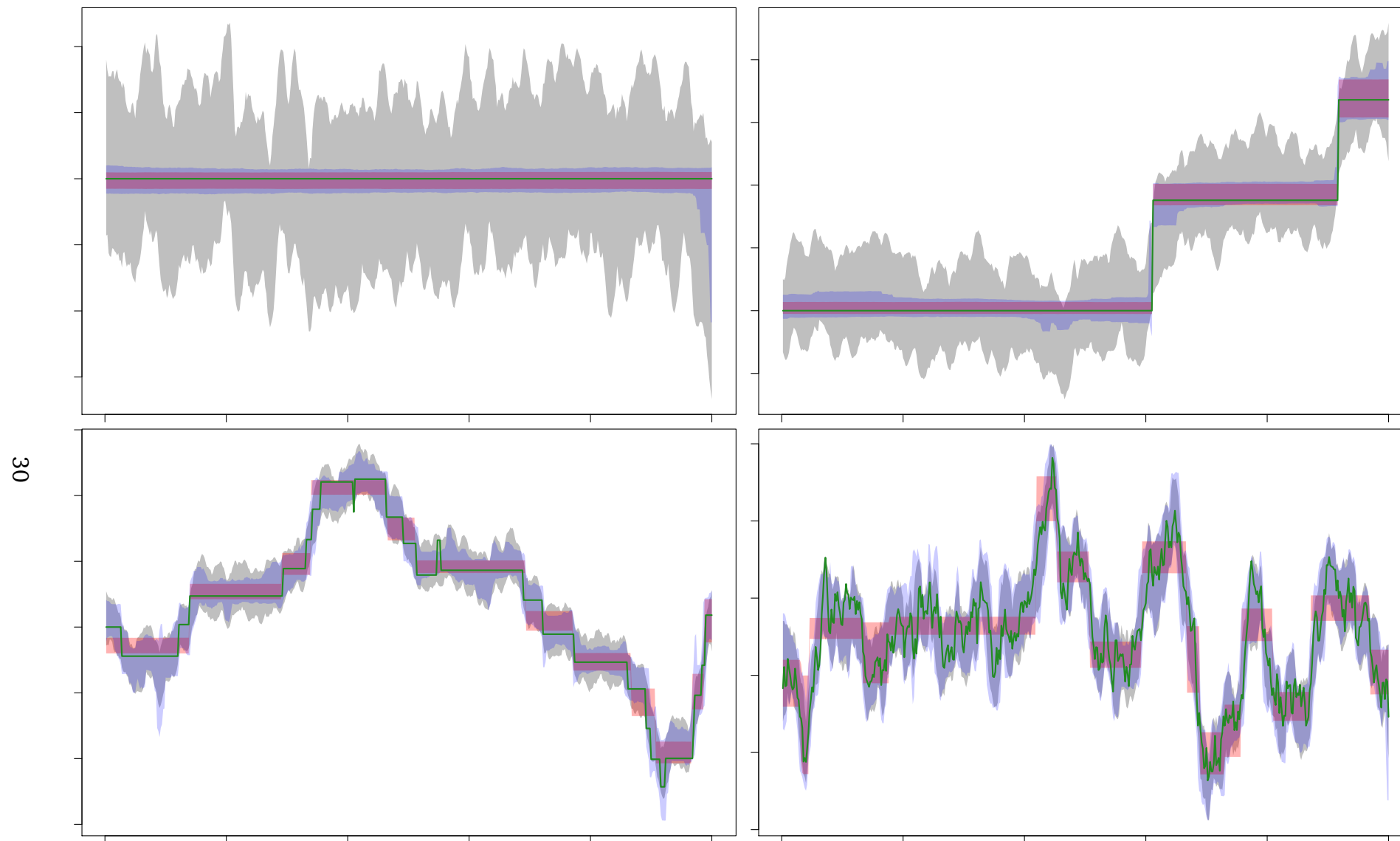


Fig. 1: Evolution of the actual state vector (sold green) along with the 98 percent posterior credible intervals of the TVP model (gray shaded area), the TTVP model (red shaded area) and a standard mixture innovation model (blue shaded area).

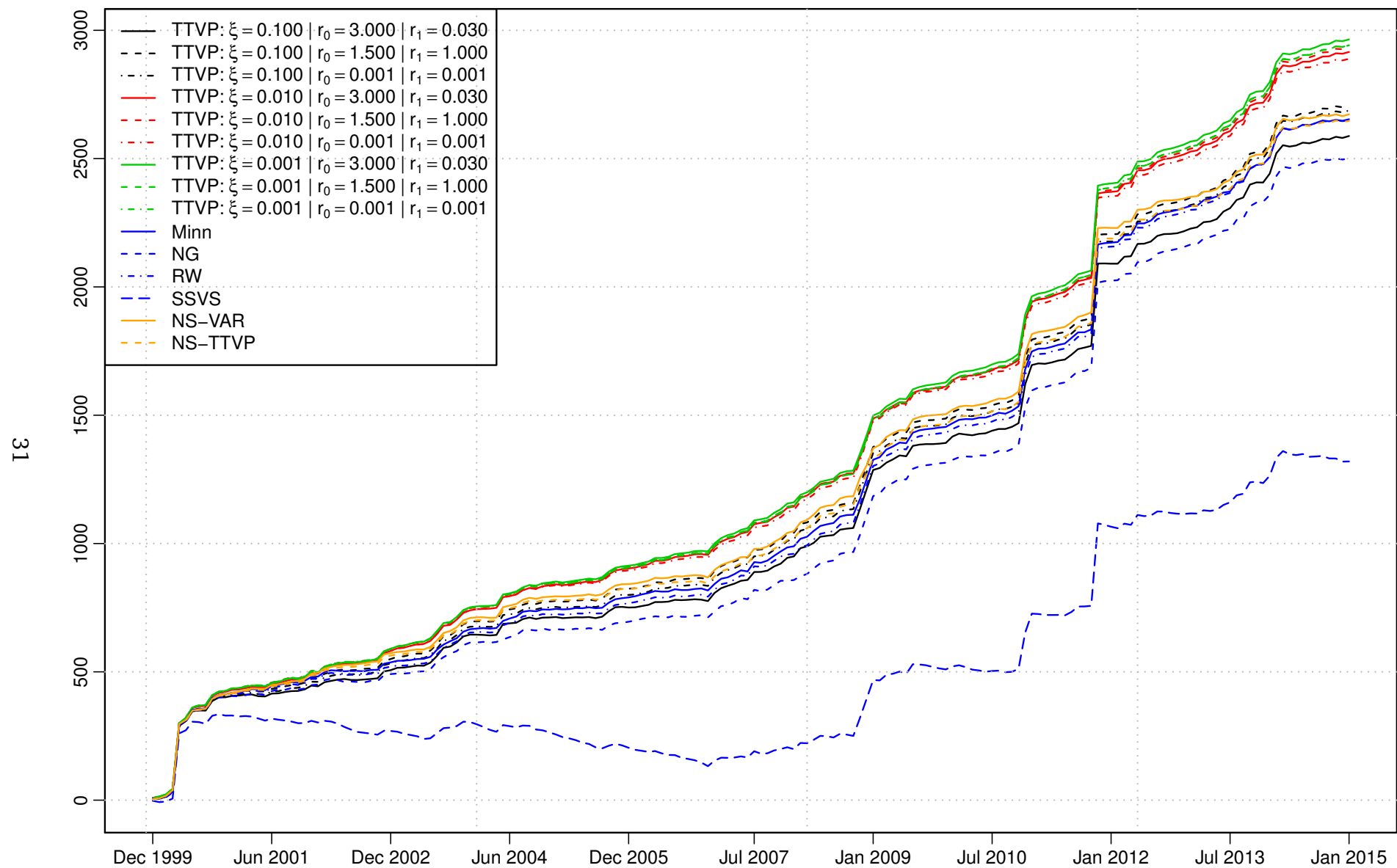


Fig. 2: One-month-ahead log predictive Bayes factor relative to a TVP-VAR model with SV.

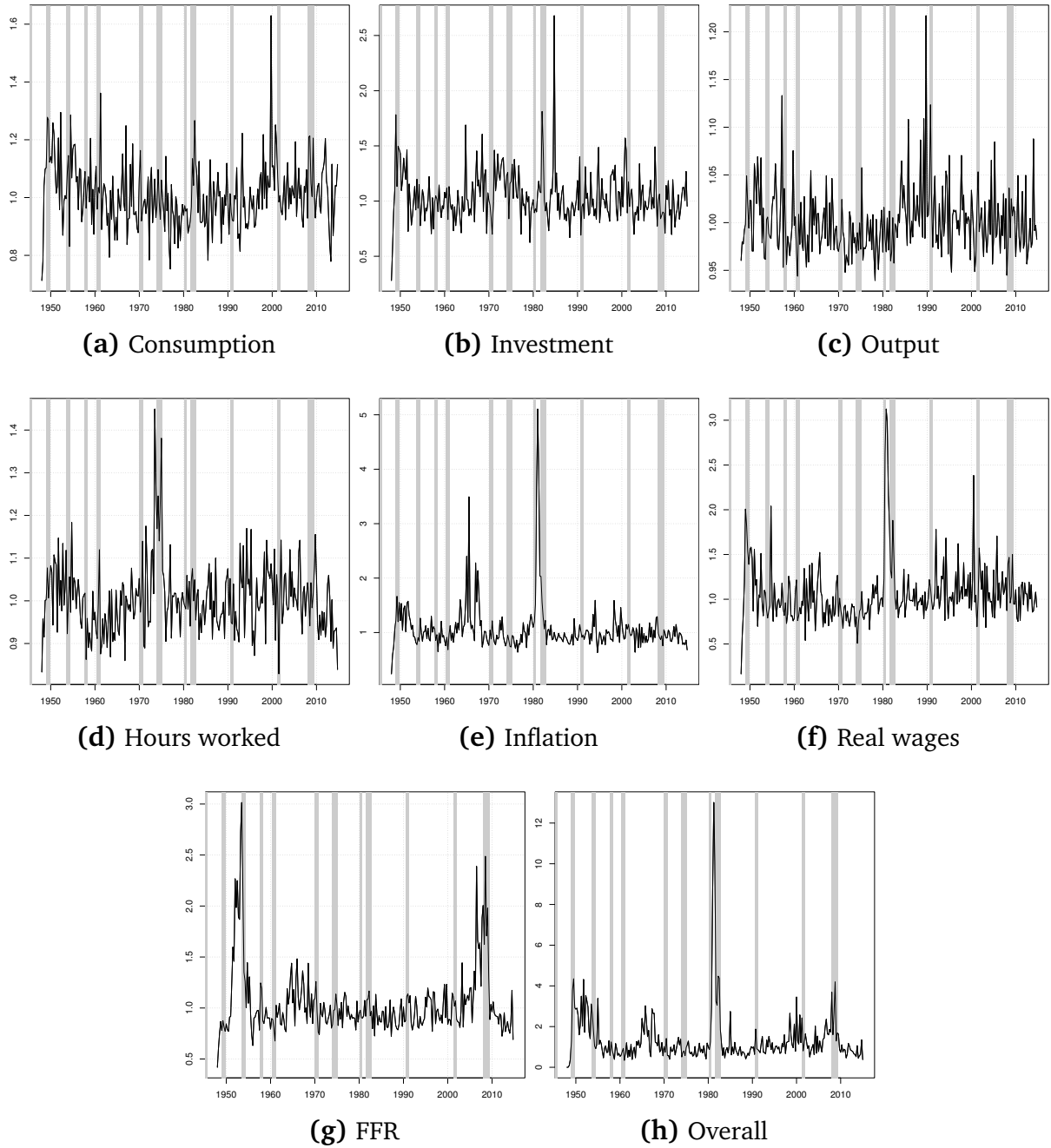


Fig. 3: Determinant of time-varying variance-covariance matrix of the innovations to the state equation. Values are obtained by taking the mean exponential of the demeaned log-determinant across equations. Gray shaded areas refer to US recessions dated by the NBER business cycle dating committee.

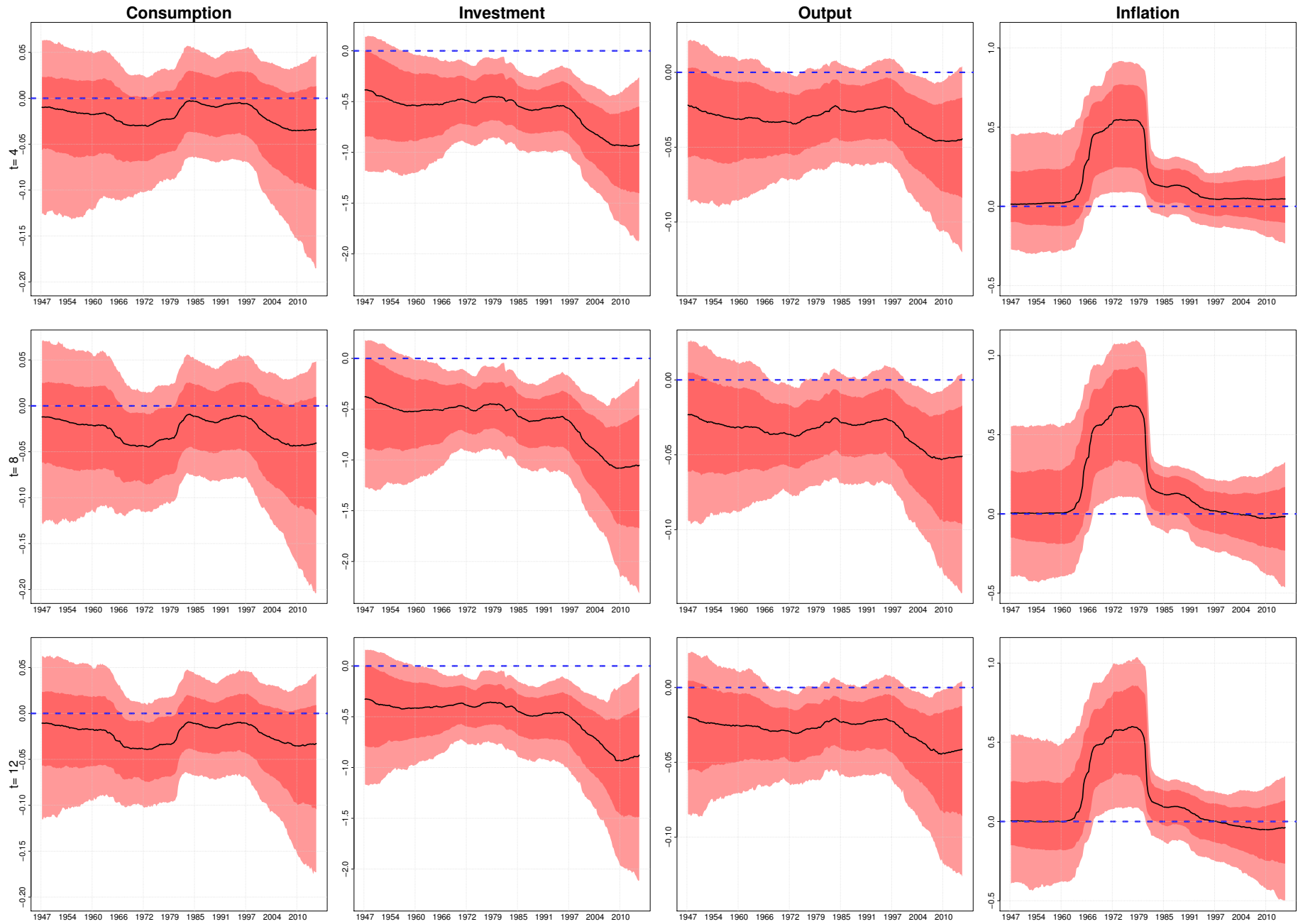


Fig. 4: Posterior median responses to a +100 bp monetary policy shock, after 4 (top panels), 8 (middle panels) and 12 (bottom panels) quarters. Shaded areas correspond to 90% (dark red) and 68% (light red) credible sets.

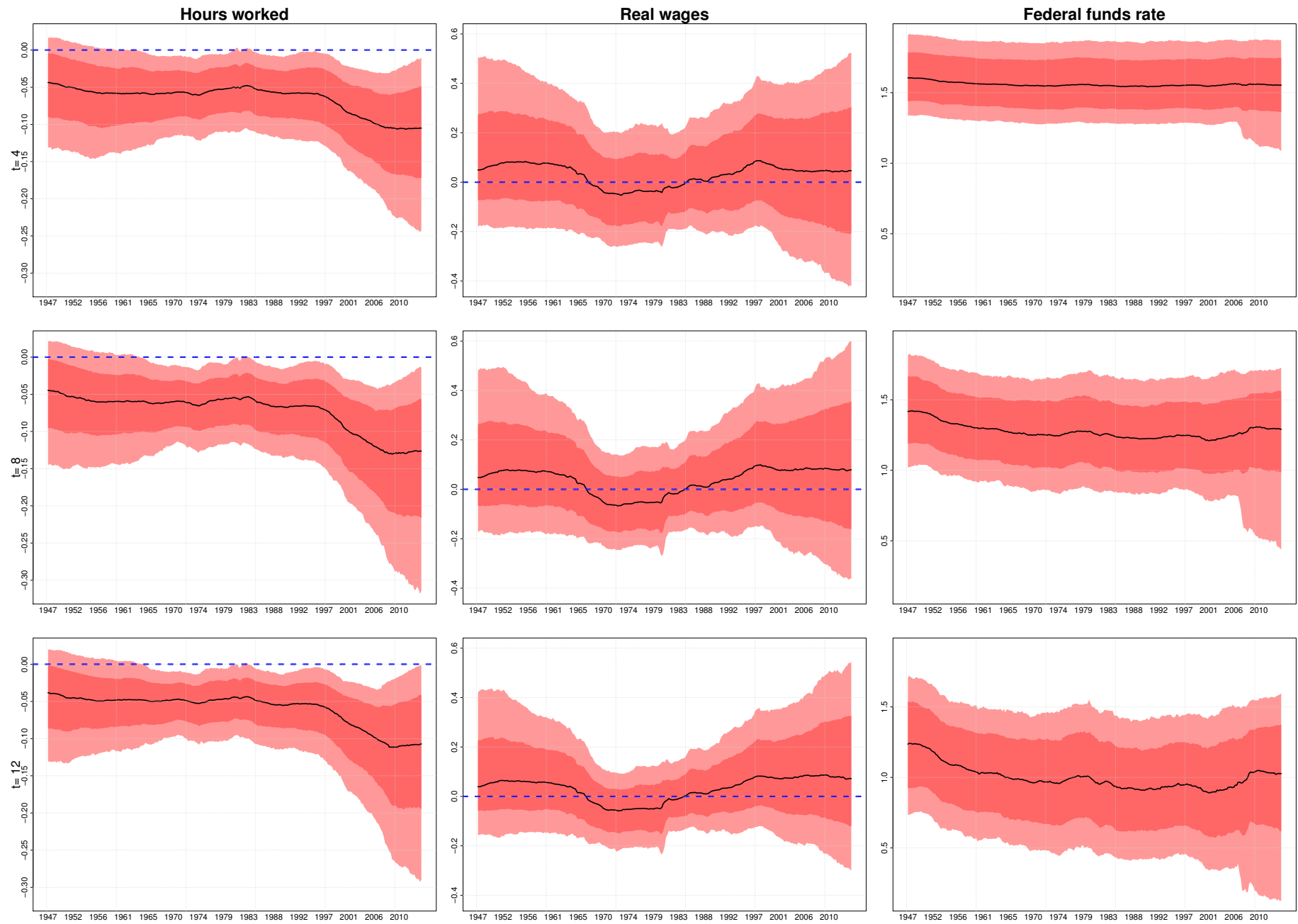


Fig. 5: Posterior median responses to a +100 bp monetary policy shock, after 4 (top panels), 8 (middle panels) and 12 (bottom panels) quarters. Shaded areas correspond to 90% (dark red) and 68% (light red) credible sets.

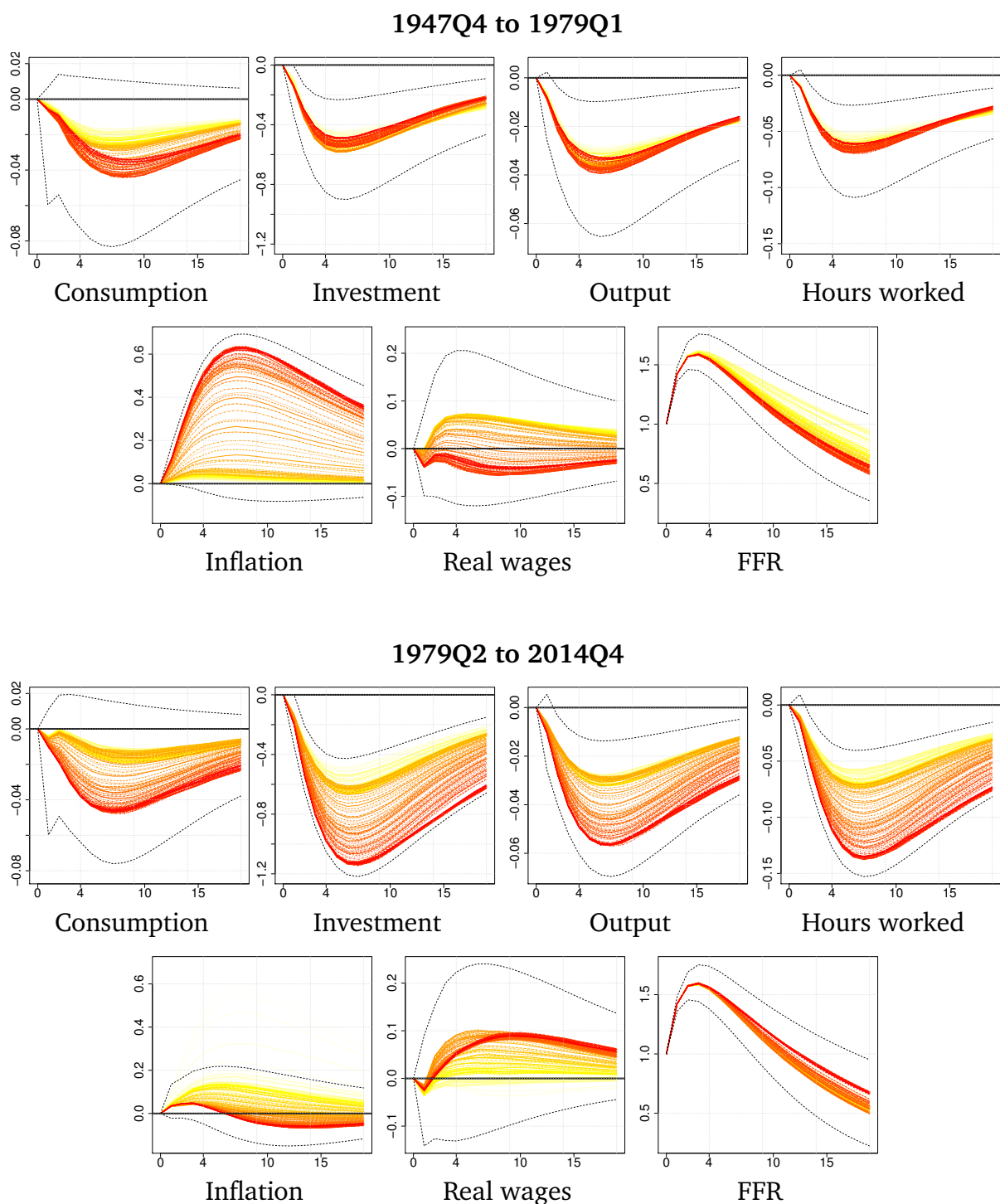


Fig. 6: Posterior median impulse response functions over two sample splits, namely the pre-Volcker period (1947Q4 to 1979Q1) and the rest of the sample period (1979Q2 to 2014Q4). The coloring of the impulse responses refer to their timing: light yellow stands for the beginning of the sample split, dark red stands for the end of sample split. For reference, 68% credible intervals over the average of the sample period provided (dotted black lines).