Bayesian Semiparametric estimation of structural VAR models with stochastic volatility*

Matteo Iacopini† Luca Rossini‡

September 30, 2021

Abstract

This paper extends the existing fully parametric Bayesian literature on structural VAR models with stochastic volatility (SVAR-SV) by introducing an innovative Bayesian semiparametric framework to model high-dimensional time series of financial returns. A Bayesian nonparametric (BNP) approach based on a Dirichlet process mixture is used to flexibly model the returns distribution by also accounting for skewness and kurtosis, while the dynamics of each series volatility is modeled with a parametric structure. Our hierarchical prior overcomes over-parametrization and over-fitting issues by clustering the coefficients into groups and shrinking the coefficients of each group toward a common location. An efficient Markov chain Monte Carlo sampling scheme is designed to perform inference in high-dimensional settings and provide a full characterization of parametric and distributional uncertainty. The proposed semiparametric approach is used to investigate returns predictability of the financial series.

Keywords: Bayesian; dynamic shrinkage; Dirichlet Process mixture; stochastic volatility; structural VAR; time-varying parameters

*Matteo Iacopini acknowledges financial support from the EU Horizon 2020 programme under the Marie Skłodowska-Curie scheme (grant agreement no. 887220).

†Vrije Universiteit Amsterdam and Tinbergen Institute, The Netherlands. m.iacopini@vu.nl

‡University of Milan and Ca' Foscari University of Venice, Italy. luca.rossini@unimi.it

1 Introduction

High dimensional models have been introduced and developed in the last period when the availability of large dataset in economics and finance has increased. In particular, large datasets in macroeconomics help to improve forecast, while in finance these datasets have been used to analyse and detect financial crisis, contagion effects, and their impact on the real economy.

The main technique used to deal with multivariate time series is the Vector autoregressive (VAR) model (see Sims, 1980), a multivariate extension of the univariate AR model. Large VAR models have been widely used to analyse and forecast high-dimensional macroeconomic data (e.g., McCracken and Ng, 2016, 2021) and financial panels (e.g., Barigozzi and Brownlees, 2019). Another typical application of VAR models is related to the assessment of the impact and spread of external shocks (i.e. to perform impulse-response analysis), the estimation of Granger-causal networks from observed data, and the study of systemic risk and financial contagion (e.g., Diebold and Yilmaz, 2009; Billio et al., 2012; Barigozzi and Brownlees, 2019; Bianchi et al., 2019).

Despite being a potentially very flexible statistical tools, the high number of parameters and the typical limited length of standard macroeconomic datasets make unrestricted inference daunting as the cross-sectional size increases. The problem of over-parametrization has been extensively studied and the most commonly used solutions rely on penalized regression and Bayesian prior regularization methods. Specifically, within the Bayesian VAR literature, a plethora of different prior distributions have been proposed to reduce the dimensionality of the parameter space or for shrinking to zero the value of irrelevant coefficients. Starting from the well-known Minnesota prior (see Doan et al., 1984; Litterman, 1986), several parametric approaches have been developed exploiting hierarchical structures and finite mixtures (e.g., Gefang, 2014; Huber and Feldkircher, 2019; Cross et al., 2020; Huber et al., 2021). Another interesting stream of the literature deals with Bayesian nonparametric techniques to work with high dimensional models and to cluster coefficients. However, Bayesian semiparametric techniques, despite having become popular in different fields (such as statistics and machine learning), have been scarcely used in econometrics and related fields.

On the other side, the interest of the macroeconomic and financial literature in time-varying parameter (TVP) models has steadily increased over the last decades. Generally speaking, this class includes all the models where the coefficients evolve over time in a discrete (e.g., change point, Markov-switching, and threshold models) or continuous (e.g., autoregressive processes) manner. As an example, Primiceri (2005) uses TVP structural VAR (SVAR) models for studying monetary policy application, Dangl and

Halling (2012) forecast equity returns by mean of TVP models, and Belmonte et al. (2014) study the European inflation via time-varying models. Besides, in a recent contribution, Huber et al. (2021) and Bitto and Frühwirth-Schnatter (2019) show that the flexibility of TVP models has an advantage in capturing gradual changes.

We propose a novel prior specification for time-varying parameters vector autoregressive models (TVP-VAR) with stochastic volatility (SV). The proposed prior has been employed on time variation in both contemporaneous and lagged coefficients through two different components, a Bayesian semiparametric and a dynamic shrinkage component. Based on the semiparametric part we capture both zero and nonzero coefficients, while using the dynamic shrinkage prior allows to capture fully constant, fully time-varying and their combinations in the coefficients. The proposed Bayesian semiparametric prior can be used in economic, financial and climate change literature to capture the impact of covariates, such own lagged observations, that switch from being irrelevant (sparse coefficients) to being relevant (non-sparse coefficients) during time. In the literature the use of dynamic shrinkage prior as been used only in univariate or multivariate cross-section models but the proposed prior is the first one that combine dynamic shrinkage and semiparametric in multivariate models and in particular in econometrics.

In this paper we combine the ideas behind time-varying parameters models and the Bayesian semiparametric techniques to model complex phenomena in a flexible and efficient manner. Following the recent literature on TVP models (e.g., Frühwirth-Schnatter and Wagner, 2010; Huber et al., 2021), we exploit the non-centered parametrization which relies on a suitable transformation of the time-varying coefficients, and we also allow the coefficients to be sparse, meaning that only a fraction of the time-varying parameters have significant effects. To achieve these goals, we define a shrinkage prior on the SVAR coefficients by means of a Bayesian semiparametric prior. This prior can be consider as a spike-and-slab prior characterized by a (parametric) Gaussian spike distribution and a (semiparametric) slab distribution, for which we assume a Bayesian nonparametric Lasso prior as in Billio et al. (2019).

This spike-and-slab prior is different from the usual global-local shrinkage prior (see Cross et al., 2020; Huber et al., 2021) as it groups the time-varying SVAR coefficients into clusters and shrinks the coefficients within a cluster toward common location. Following Billio et al. (2019), our hierarchical prior overcomes over-parametrization and overfitting issues by clustering the SVAR coefficients into groups and by shrinking the coefficients of each group toward a common location. This hierarchical prior allows to contemporaneously estimate the (potentially) sparse time-varying causal structure and to cluster the corresponding coefficients. The research in macroeconomics has also shown that accounting for time-varying volatility or heteroschedasticy in macroeconometric models is an important aspect that leads to strong gains in estimation and forecasting. Consequently, we include stochastic volatility (SV) in the TVP-SVAR model.

One main pitfall of the majority of TVP models is the rigid assumption made on the dynamics of the coefficients, which are most frequently assumed to alternate between different levels, thus implying a piece-wise constant trajectory (as in the case of Markov-switching models), or continuously vary in a random fashion (as in random walk processes). We take a step forward and consider a more general approach and assume a dynamic shrinkage process for the TVP coefficients via the Dynamic Horseshoe prior (DHS) (Kowal et al., 2019). Essentially, this consists in assuming a (autoregressive) time-varying log-variance for the latent time-varying coefficients and allows to learn from the data, at each point in time, whether a coefficient should stay fixed at a constant value or move according to an autoregressive process. This approach has been recently applied in univariate time varying models (e.g., Huber and Pfarrhofer, 2021), but not in a SVAR context.

The remainder of this article is organized as follows: Section 2 illustrates the proposed TVP-VAR model with centered and non-centered representation and introduces the dynamic HS prior on the variance coefficients. Section 3 presents the novel semiparametric prior and the details of the Bayesian approach to inference. Section 4 investigates the performance of our method using simulated data. Section 5 shows preliminary results of a macroeconomic application. Section 6 concludes the article and a technical appendix provides further details on the posterior simulation algorithms.

2 TVP-SVAR-SV

Let us define $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})$ as the vector of n variables available at time t and we consider the structural VAR (SVAR) model with 1 lag and with stochastic volatility as

$$A\mathbf{y}_t = B_0 + B_1\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, H_t), \tag{1}$$

$$h_{j,t} = \mu_{h,j} + \phi_{h,j}(h_{j,t-1} - \mu_{h,j}) + \eta_{j,t}, \qquad \eta_{j,t} \sim \mathcal{N}(0, \sigma^2_{h,j}), \tag{2}$$

where B_1 is a $n \times n$ matrix of coefficients, B_0 is a vector of constants, A is an $n \times n$ lower-triangular matrix with ones on the diagonal and $\boldsymbol{\epsilon}_t$ is an $n \times 1$ vector of i.i.d. noise, $\mathcal{N}(\mathbf{0}, H_t)$ denotes a multivariate Gaussian distribution with covariance matrix $H_t = \mathrm{diag}(e^{h_{1,t}}, \ldots, e^{h_{n,t}})$. For the time-varying variance $e^{h_{j,t}}$, $j = 1, \ldots, n$, we assume its logarithm follows an autoregressive AR(1) process, as shown in Equation 2, where $\mu_{h,j}$ is the unconditional mean, $\phi_{h,j}$ is the persistence parameter and $\sigma^2_{h,j}$ is the error variance of the log-volatility process. Following Chan and Eisenstat (2018), we reformulate the

SVAR-SV model as

$$\mathbf{y}_t = X_t\boldsymbol{\beta} + W_t\boldsymbol{\gamma} + \boldsymbol{\epsilon}_t = Z_t\boldsymbol{\theta} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, H_t), \tag{3}$$

where W_t contains the contemporaneous endogenous variables in the appropriate position, $Z_t = (X_t, W_t)$ is an $n \times k_\theta$ matrix, and $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ is a k_θ-dimensional vector, with $k_\theta = k_\beta + k_\gamma$, $k_\beta = n(n+1)$, and $k_\gamma = n(n-1)/2$. Finally, $\boldsymbol{\beta} = \mathrm{vec}([B_0, B_1])$ is a $k_\beta \times 1$ vector of coefficients associated to the lagged observations, with $k_\beta = n(n+1)$, $\boldsymbol{\gamma}$ is a $k_\gamma \times 1$ vector of the coefficients, with $k_\gamma = n(n-1)/2$, that characterize the contemporaneous dependence among the endogenous variables and consists of the free elements of A stacked by rows.

Starting from Equation 3, we include temporal variation of the coefficients by assuming the elements of the coefficient vector $\boldsymbol{\theta}$ follow independent random walk processes. We thus obtaining a time-varying parameter (TVP) SVAR-SV model with *centered* parametrization given sby

$$\begin{aligned}
\mathbf{y}_t &= Z_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, H_t), \\
\boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, V), & (4)\\
\mathbf{h}_t &= \boldsymbol{\mu}_h + \mathrm{diag}(\boldsymbol{\phi}_h)(\mathbf{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_h), & (5)
\end{aligned}$$

Notice that all definitions are the same as in Equation 2 and 3 except that $\boldsymbol{\theta}_t$ are dynamic (time-varying) regression coefficients which follow a random walk with $\boldsymbol{\omega}_t$ being a Gaussian innovation vector with zero mean and diagonal covariance matrix $V = \mathrm{diag}(v_1, \ldots, v_{k_\theta})$. Each v_j, $j = 1, \ldots, k_\theta$, is a process innovation variance associated with the j-th coefficient and controls the amount of time-variation in θ_{jt}. Finally, the noise covariance matrix of the log-volatility process is $\Sigma_h = \mathrm{diag}(\sigma^2_{h,1}, \ldots, \sigma^2_{h,n})$. Defining $\sqrt{V} = \mathrm{diag}(\sqrt{v_1}, \ldots, \sqrt{v_{k_\theta}})$ allows us to derive the corresponding *non-centered* parametrization (Frühwirth-Schnatter and Wagner, 2010) of the TVP model as

$$\begin{aligned}
\mathbf{y}_t &= Z_t\boldsymbol{\theta}_0 + Z_t\sqrt{V}\tilde{\boldsymbol{\theta}}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, H_t), \\
\tilde{\boldsymbol{\theta}}_t &= \tilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{\varpi}_t, \qquad \boldsymbol{\varpi}_t \sim \mathcal{N}(\mathbf{0}, I_{k_\theta}), \\
\mathbf{h}_t &= \boldsymbol{\mu}_h + \mathrm{diag}(\boldsymbol{\phi}_h)(\mathbf{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_h),
\end{aligned} \tag{6}$$

where the j-th element of $\boldsymbol{\theta}_t$ is given by $\theta_{jt} = \frac{\theta_{jt} - \theta_{j0}}{\sqrt{v_j}}$ and $\tilde{\boldsymbol{\theta}}_0 = \mathbf{0}$.

We now extend the non-centered parameterization of the TVP-VAR-SV model in Equation 6 by considering a dynamic shrinkage process for the the time-varying coefficients. Specifically, we consider a dynamic Horseshoe prior (DHS) as in Kowal et al.

(2019), which is obtained by assuming an AR(1) process for the log-variance of the latent states, with precision hyper-parameter following a Pólya-Gamma distribution (Polson et al., 2013), denoted $PG(a,b)$. From Equation 4, the model in *centered* form can be reformulated as

$$
\begin{aligned}
&\mathbf{y}_t = Z_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\exp(h_{1,t}),\dots,\exp(h_{n,t}))),\\
&\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\exp(v_{1,t}),\dots,\exp(v_{n,t}))),\\
&\mathbf{h}_t = \boldsymbol{\mu}_h + \mathrm{diag}(\boldsymbol{\phi}_h)(\mathbf{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\sigma^2_{h,1},\dots,\sigma^2_{h,n})),\\
&\mathbf{v}_t = \boldsymbol{\mu}_v + \mathrm{diag}(\boldsymbol{\phi}_v)(\mathbf{v}_{t-1} - \boldsymbol{\mu}_v) + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\xi^{-1}_{1,t},\dots,\xi^{-1}_{k,t})),\\
&\xi_{i,t} \sim PG(1,0),
\end{aligned}
\tag{7}
$$

where $\mathbf{v}_t$ follows an AR(1) process similar to the one described in Equation 2. The main difference is that the error variance, $\xi^{-1}_{j,t}$, is random but i.i.d. across time and is assumed to follow a Pólya-Gamma distribution. This corresponds to the choice $\alpha = \beta = 1/2$ in Kowal et al. (2019) to which we refer for further details on this construction. Let $\sqrt{V_t} = \mathrm{diag}(\sqrt{v_{1,t}},\dots,\sqrt{v_{k,t}})$ and recall the maps linking the *centered* and the *non-centered* parametrizations as

$$
\boldsymbol{\theta}_t = \boldsymbol{\theta}_0 + \sqrt{V_t}\tilde{\boldsymbol{\theta}}_t, \qquad \tilde{\boldsymbol{\theta}}_t = \frac{\boldsymbol{\theta}_t - \boldsymbol{\theta}_0}{\sqrt{V}_t}, \qquad t = 1,\dots,T.
$$

Thus, Equation 6 corresponds to the *non-centered* form

$$
\begin{aligned}
&\mathbf{y}_t = Z_t\boldsymbol{\theta}_0 + Z_t\sqrt{V_t}\tilde{\boldsymbol{\theta}}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\exp(h_{1,t}),\dots,\exp(h_{n,t}))),\\
&\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_{t-1} + \boldsymbol{\varpi}_t, \qquad \boldsymbol{\varpi}_t \sim \mathcal{N}(\mathbf{0}, I_k),\\
&\mathbf{h}_t = \boldsymbol{\mu}_h + \mathrm{diag}(\boldsymbol{\phi}_h)(\mathbf{h}_{t-1} - \boldsymbol{\mu}_h) + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\sigma^2_{h,1},\dots,\sigma^2_{h,n})),\\
&\mathbf{v}_t = \boldsymbol{\mu}_v + \mathrm{diag}(\boldsymbol{\phi}_v)(\mathbf{v}_{t-1} - \boldsymbol{\mu}_v) + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\xi^{-1}_{1,t},\dots,\xi^{-1}_{k,t})),\\
&\xi_{i,t} \sim PG(1,0).
\end{aligned}
\tag{8}
$$

3 Bayesian Inference

3.1 A Semiparametric Prior specification

Concerning the initial value of the TVP vector, $\boldsymbol{\theta}_0$, we assume a mixture prior independently for each coefficient $\theta_{j,0}$, $j = 1, \dots, k_\theta$, as follows

$$
\begin{aligned}
P(\boldsymbol{\theta}_0) &= \prod_{j=1}^{k_\theta} P(\theta_{j,0}|\mu_j, \tau_j, \pi, \lambda_0) \\
\theta_{j,0}|\mu_j, \tau_j, \pi &\sim \pi \mathcal{N}(\theta_{j,0}|0, \lambda_0) + (1-\pi)\mathcal{DE}(\theta_{j,0}|\mu_j, 1/\sqrt{\tau_j}),
\end{aligned} \tag{9}
$$

where $\mathcal{DE}(x|a, s)$ denotes a Double Exponential (DE) distribution[1] with a equal to the location and s equal to the scale(MacLehose and Dunson, 2010). This mixture prior can be considered as a spike-and-slab distribution, with Gaussian spike and Double Exponential slab, which allows for shrinkage to zero of the (irrelevant) coefficients. Exploiting the Gaussian scale-mixture representation of the DE distribution, one obtains a (conditional) mixture of Gaussians

$$
\theta_{j,0}|\mu_j, \lambda_j, \pi, \lambda_0 \sim \pi \mathcal{N}(\theta_{j,0}|0, \lambda_0) + (1-\pi)\mathcal{N}(\theta_{j,0}|\mu_j, \lambda_j) \tag{10}
$$

$$
\lambda_j \sim \mathcal{E}xp(\lambda_j|2/\tau_j). \tag{11}
$$

A standard approach in Bayesian nonparametrics relies on the specification of a Dirichlet Process (DP, see Ferguson, 1973) prior for the distribution of the parameters of interest. We adopt a semiparametric approach and assume a Dirichlet Process Mixture (DPM, see Lo, 1984) prior for the location and scales of the Double Exponential slab distribution, μ_j, τ_j, as follows[2]

$$
\begin{aligned}
(\mu_j, \tau_j)|P &\sim P, \\
P &\sim DP(\alpha, P_0) \\
P_0(\mu_j, \tau_j) &\sim \mathcal{N}(\mu_j|c, d)\mathcal{G}a(\tau_j|a_1, b_1)
\end{aligned} \tag{12}
$$

[1]The Double Exponential (or Laplace) distribution is a (exponential) scale mixture of Normals

$$
x|\lambda \sim \mathcal{N}(x|\mu, \lambda), \quad \lambda \sim \mathcal{E}xp(\lambda|2\tau^2) \implies x \sim \mathcal{DE}(x|\mu, \tau),
$$

and has probability density function

$$
x \sim \mathcal{DE}(x|\mu, \tau) \iff p(x|\mu, \tau) = \frac{1}{2\tau} \exp\Big(-\frac{|x-\mu|}{\tau} \Big).
$$

[2]We use the shape-scale parametrisation of the Gamma distribution (thus, if $x \sim \mathcal{G}a(x|a, b)$, then $\mathbb{E}[x] = ab$ and $\mathbb{V}[x] = ab^2$). The exponential distribution is obtained when $a = 1$, that is $\mathcal{E}xp(x|b) = \mathcal{G}a(x|1, b)$.

where α and P_0 are the DP concentration parameter and base measure, respectively. Moreover, c and d are the hyperparameters of the Gaussian distribution representing the mean and the standard deviation and $a_1.b_1$ are the hyperparameters of the Gamma distribution.

The hierarchical prior structure for the coefficients $\theta_{j,0}$, $j = 1, \ldots, k_\theta$, is completed by assuming standard prior distributions for the remaining hyperparameters, and can be summarized as follows

$$\begin{aligned}
\theta_{j,0}|\mu_j, \lambda_j, \pi &\sim \pi\mathcal{N}(\theta_{j,0}|0, \lambda_0) + (1-\pi)\mathcal{N}(\theta_{j,0}|\mu_j, \lambda_j) \\
\lambda_0|\tau_0 &\sim \mathcal{E}xp(\lambda_0|2/\tau_0) \\
\tau_0 &\sim \mathcal{G}a(\tau_0|a_0, b_0) \\
\lambda_j|\tau_j &\sim \mathcal{E}xp(\lambda_j|2/\tau_j) \\
(\mu_j, \tau_j)|P &\sim P, \\
P &\sim DP(\alpha, P_0), \\
P_0(\mu_j, \tau_j) &\sim \mathcal{N}(\mu_j|c, d)\mathcal{G}a(\tau_j|a_1, b_1) \\
\pi &\sim \mathcal{B}e(1, \eta_0),
\end{aligned} \tag{13}$$

where π is the mixing probability parameter, which assumes a Beta distribution. Obviously the parameter of the spike component have some hyperparameters to be considering, which are a_0 and b_0 for the Gamma distribution. Note that the specification of DP prior for a random probability measure P allows for clustering of the variables drawn from that distribution, μ_j and τ_j in our case.

For the parameters governing the stochastic volatility processes for the observations, we follow the common practice in the literature (e.g., see Kastner and Frühwirth-Schnatter, 2014) and assume

$$\mu_{h,j} \sim \mathcal{N}(0, \sigma^2_{h,j}B_0), \qquad \frac{\phi_{h,j}-1}{2} \sim \mathcal{B}e(a_h, b_h), \qquad \sigma^2_{h,j} \sim \mathcal{IG}(c_h, d_h). \tag{14}$$

Finally, for the dynamic shrinkage prior for the time-varying volatility, we build on Kowal et al. (2019) and assume the following prior structure for the parameters driving the log-variance process of the TVP parameters

$$\mu_{v,j}|\xi_{\mu,j} \sim \mathcal{N}(m_0, \xi^{-1}_{\mu,j}), \quad \frac{\phi_{v,j}-1}{2} \sim \mathcal{B}e(a_v, b_v), \quad \xi_{j,t} \sim PG(1,0), \quad \xi_{\mu,j} \sim PG(1,0). \tag{15}$$

3.2 Posterior approximation via MCMC

Since the joint posterior distribution is not tractable and it is complex to be sample from, Bayesian estimator cannot be obtained analytically. In this paper, we rely on simulation based inference methods, and develop a Gibbs sampler algorithm for approximating the posterior distribution.

To deal with the mixture provided by the spike-and-slab prior and the infinite mixture given by the DPM, we exploit a data augmentation approach. For each $j = 1, \ldots, k_\theta$ we introduce two sets of allocation variables, γ_j, d_j; a set of stick-breaking variables, $\mathbf{w} = \{w_i : i = 1, 2, \ldots\}$ and a set of slice variables, u_j. The allocation variables, γ_j, select between the spike and slab components of the mixture for $\theta_{j,0}$, whereas the allocation variable d_j selects the component of the Dirichlet Process Mixture to which each single coefficient $\theta_{j,0}$ is allocated to. The sequence of stick-breaking variables, $\mathbf{w}$, defines the mixture weights, whereas the slice variable, u_j, allow to deal with the infinite mixture components by identifying a finite number of stick-breaking variables to be sampled and an upper bound for the allocation variables d_j.

Let $(\boldsymbol{\mu}, \boldsymbol{\tau}) = \{(\mu_k, \tau_k) : k = 1, \ldots, k^*\}$ denote the atoms, where k ranges from 1 to the number k^* of allocated DP components, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{k_\theta})$. Define the collection of latent variables as follows: $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_T)$, $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_T)$, $\mathbf{u} = (u_1, \ldots, u_{k_\theta})$, $\mathbf{d} = (d_1, \ldots, d_{k_\theta})$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{k_\theta})$. Then, denote the collection of all parameters and latent variables with $\boldsymbol{\vartheta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T, \pi, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \tau_0, \lambda_0, \mathbf{V}, \boldsymbol{\mu}_v, \boldsymbol{\phi}_v, \boldsymbol{\Xi}, \boldsymbol{\xi}_\mu, \mathbf{H}, \boldsymbol{\mu}_h, \boldsymbol{\phi}_h, \boldsymbol{\sigma}_h^2)$. We obtain the following joint posterior distribution

$$\begin{aligned} P(\boldsymbol{\vartheta}|\mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\theta}_0, \mathbf{H}) \cdot \prod_{k=1}^{k^*} P(\mu_k) P(\tau_k) \cdot \prod_{i=1} P(u_i|w_i) P(w_i) \\ \cdot P(\pi) \prod_{j=1}^{k_\theta} P(\theta_{j,0}|\pi, \lambda_j, \boldsymbol{\mu}) P(\lambda_j|\boldsymbol{\tau}) P(d_j|w_j, u_j) P(\gamma_j|\pi) \\ \cdot P(\boldsymbol{\mu}_h, \boldsymbol{\phi}_h, \boldsymbol{\sigma}_h^2) \prod_{t=1}^{T} P(\mathbf{h}_t|\mathbf{h}_{t-1}, \boldsymbol{\mu}_h, \boldsymbol{\phi}_h, \boldsymbol{\sigma}_h^2) \\ \cdot P(\boldsymbol{\mu}_v, \boldsymbol{\phi}_v, \boldsymbol{\Xi}) \prod_{t=1}^{T} P(\mathbf{v}_t|\mathbf{v}_{t-1}, \boldsymbol{\mu}_v, \boldsymbol{\phi}_v, \boldsymbol{\Xi}) P(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \mathbf{v}_t), \end{aligned} \tag{16}$$

Due to the intractability of the joint posterior distribution, we design an Markov Chain Monte Carlo (MCMC) algorithm based on a Gibbs sampler to approximate the posterior distribution. The Gibbs sampler is based on the algorithm of Hatjispyros et al. (2011) and on the slice sampler approach (Walker, 2007; Kalli et al., 2011) for estimating the weights and locations of each random measure P. The re-sampling of the time-varying coefficients, $(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$, using alternatively the non-centered and the centered parametrization helps improving the mixing of the MCMC. Hereafter, we show the iterative steps by using the conditional independence between variables, for $k = 1, \ldots, k^*$ and $j = 1, \ldots, k_\theta$:

(1) the slice and stick-breaking variables u_j and w_i are updated given $[d_j, \gamma_j]$;
(2) the latent scale variables λ_j are updated given $[\boldsymbol{\mu}, \boldsymbol{\tau}, \theta_{j,0}, d_j, \gamma_j]$;
(3) the parameters of the stick-breaking locations (μ_k, τ_k) are updated given $[\boldsymbol{\lambda}, \boldsymbol{\theta}_0, \mathbf{d}, \boldsymbol{\gamma}]$;
(4) the allocation variables d_j, γ_j are jointly updated given $[\mu_k, \tau_k, \theta_{j,0}, u_j, w_i, \pi]$;
(5) the mixing probability π of having sparse coefficients is updated given $[(\gamma_j)_j]$;
(6) the path of the SVAR coefficients $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_T)$ is jointly updated given $[\mathbf{H}, \mathbf{V}, \boldsymbol{\theta}_0, \mathbf{Y}]$, in the non-centered parametrization following Chan and Jeliazkov (2009);
(7) the initial value of the SVAR coefficients $\boldsymbol{\theta}_0$ is updated given $[\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{H}, \mathbf{d}, \boldsymbol{\gamma}, \boldsymbol{\theta}_1, \mathbf{Y}]$, in the non-centered parametrization;
- let $\sqrt{V_t} = \mathrm{diag}(\sqrt{v_{1,t}}, \ldots, \sqrt{v_{k_\theta,t}})$ and move to the centered parametrization: $\boldsymbol{\theta}_t = \boldsymbol{\theta}_0 + \sqrt{V_t}\tilde{\boldsymbol{\theta}}_t$;

(8) the path of the SVAR coefficients $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)$ is re-sampled jointly given $[\mathbf{H}, \mathbf{V}, \boldsymbol{\theta}_0, \mathbf{Y}]$, in the centered parametrization following Chan and Jeliazkov (2009);
(9) the path of $(\mathbf{v}_1, \ldots, \mathbf{v}_T)$ is jointly updated given $[\boldsymbol{\mu}_v, \boldsymbol{\phi}_v, \Xi, \boldsymbol{\theta}_0, \boldsymbol{\theta}]$ following Kowal et al. (2019);
(10) the parameters driving the log-variance of the TVP coefficients, $\boldsymbol{\mu}_v, \boldsymbol{\phi}_v, \Xi$, given $[\mathbf{v}]$ from their full conditional distributions (using a slice sampler for drawing $\boldsymbol{\phi}_v$);
(11) the initial value of the SVAR coefficients $\boldsymbol{\theta}_0$ is re-sampled jointly given $[\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\lambda}, \mathbf{H}, \mathbf{d}, \boldsymbol{\gamma}, \boldsymbol{\theta}_1, \mathbf{Y}]$, in the centered parametrization;
- let $\sqrt{V_t} = \mathrm{diag}(\sqrt{v_{1,t}}, \ldots, \sqrt{v_{k_\theta,t}})$ and move to the non-centered parametrization: $\tilde{\boldsymbol{\theta}}_t = (\sqrt{V_t})^{-1}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_0)$;

(12) the path of observations' stochastic volatility $(\mathbf{h}_1, \ldots, \mathbf{h}_T)$ is sampled jointly given $[\boldsymbol{\mu}_h, \boldsymbol{\phi}_h, \boldsymbol{\sigma}_h^2, \boldsymbol{\theta}_0, \boldsymbol{\theta}, \mathbf{Y}]$;
(13) the parameters driving the observations' stochastic volatility, $\boldsymbol{\mu}_h, \boldsymbol{\phi}_h, \boldsymbol{\sigma}_h^2$, given $[\mathbf{h}]$ from their full conditional distributions (using a slice sampler for drawing $\boldsymbol{\phi}_h$);

The detailed Gibbs sampler is described in the Appendix.

4 Evidence using Artificial Data

In this section, we test the ability of the model proposed in Section 2 to recover the path of time-varying parameters on synthetic data generated from different TVP-VAR-SV models. Across the different data generating processes (DGP), the path of the time-varying coefficients follows different specifications: constant; time-varying with piece-wise constant trajectory, and time-varying with alternating random walk and constant trajectory. Before presenting the results, we describe the three different setting used for generating the data:[3]

(i) $n = 3$, $T = 150$;
(ii) $n = 6$, $T = 200$;
(iii) $n = 12$, $T = 200$.

We start by considering a small-dimensional TVP-VAR-SV model, with $n = 3$ and $T = 150$ observations. Figure 1 shows the estimated path of the stochastic volatility process, $\mathbf{h}$. We find that the true value is always within the 90% credible interval (C.I.), and most of the time even within the 67% C.I.. Figure 2 reports the same type of plots for a subset of the time-varying coefficients, $\boldsymbol{\theta}$. Several comments are in order. Overall, we find evidence of good performance of the proposed model in recovering the true path of the variables. Specifically, this is the case when the true coefficient is constant (e.g., see the plot in row 1, column 2), time-varying with piece-wise constant trajectory (plot in row 1, column 3), or time-varying with alternating random walk and constant trajectory (plot in row 2, column 2).

We replicate the exercise for a TVP-VAR-SV model with $n = 6$ and $T = 200$ observations. Again, Figure 3 highlights the correct estimation of the stochastic volatility process, $\mathbf{h}$. More interestingly, Figure 4 confirms the previous results in terms of the ability of the model to recover the path of the time-varying coefficients, $\boldsymbol{\theta}$, in presence of constant true coefficient (e.g, see plot in row 1, column 1), time-varying with piece-wise constant trajectory (plot in row 2, columns 2), random walk behavior (plot in row 1, columns 3-4), or time-varying with alternating random walk and constant trajectory (plot in row 3, columns 2-3).

Figures 5 and 6 shows the results for the TVP-VAR-SV model with $n = 12$ and $T = 200$. As in previously DGP, the results for the stochastic volatility process and for the time-varying parameters are in line with the previous results and thus confirm the goodness of the proposed prior.

[3]The computing time for a single iteration in the setting $n = 12$, $T = 200$ is 0.89 seconds, using MATLAB on a laptop MacBookPro 3,1 GHz Intel Core i7.

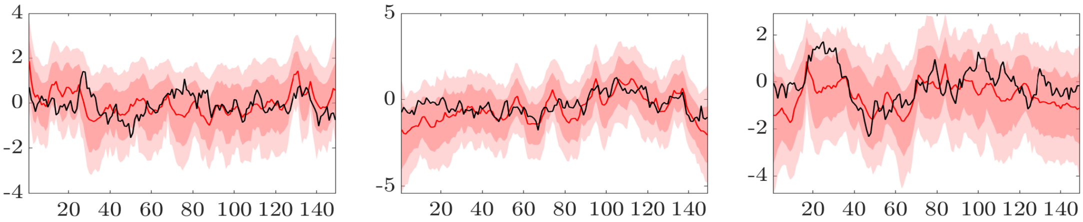

Figure 1: Path of the observations' log-volatility process, $\mathbf{h}$: true value (black line), posterior median (red line), 90% and 67% credible interval (light and dark red shading, respectively). Dataset: $n = 3$, $T = 150$.

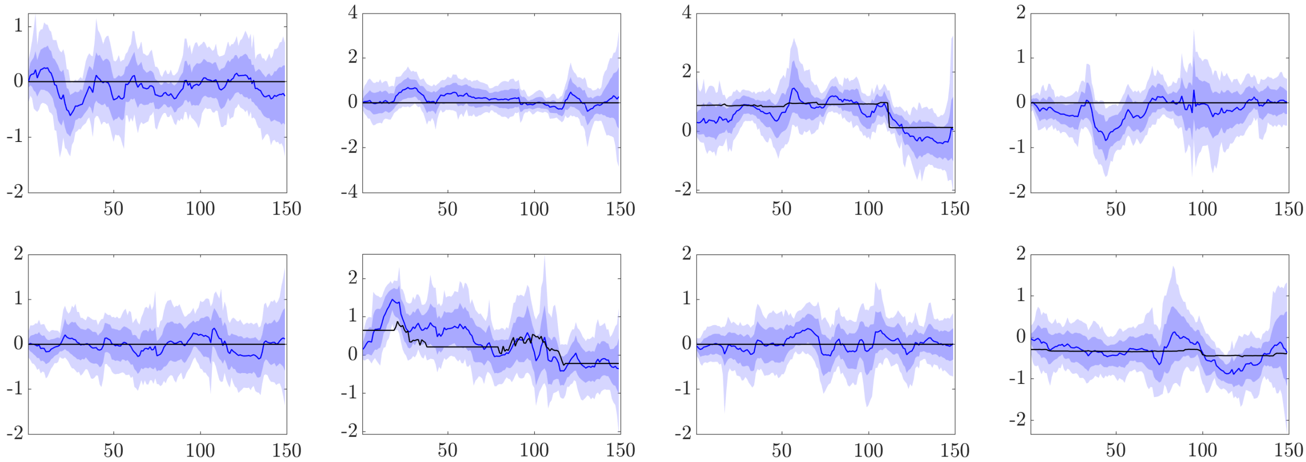

Figure 2: Path of selected time-varying coefficients', $\boldsymbol{\theta}$: true value (black line), posterior median (blue line), 90% and 67% credible interval (light and dark blue shading, respectively). Dataset: $n = 3$, $T = 150$.

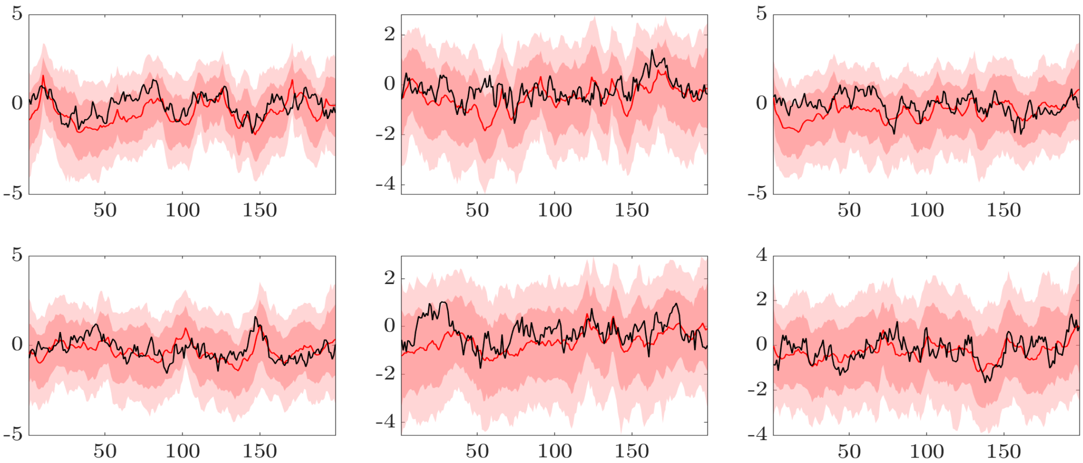

Figure 3: Path of the observations' log-volatility process, $\mathbf{h}$: true value (black line), posterior median (red line), 90% and 67% credible interval (light and dark red shading, respectively). Dataset: $n = 6$, $T = 200$.

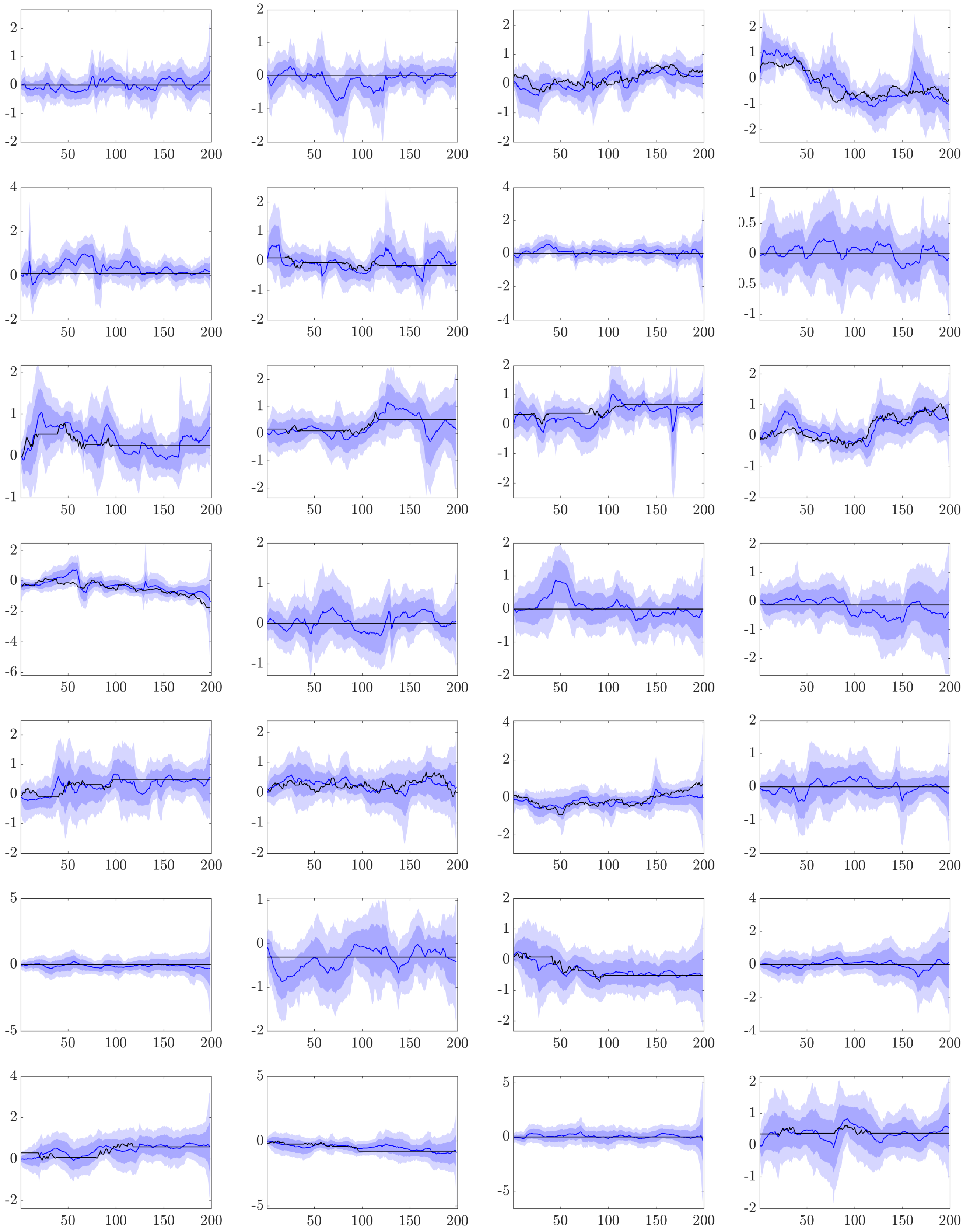

Figure 4: Path of selected time-varying coefficients', $\boldsymbol{\theta}$: true value (black line), posterior median (blue line), 90% and 67% credible interval (light and dark blue shading, respectively). Dataset: $n = 6$, $T = 200$.

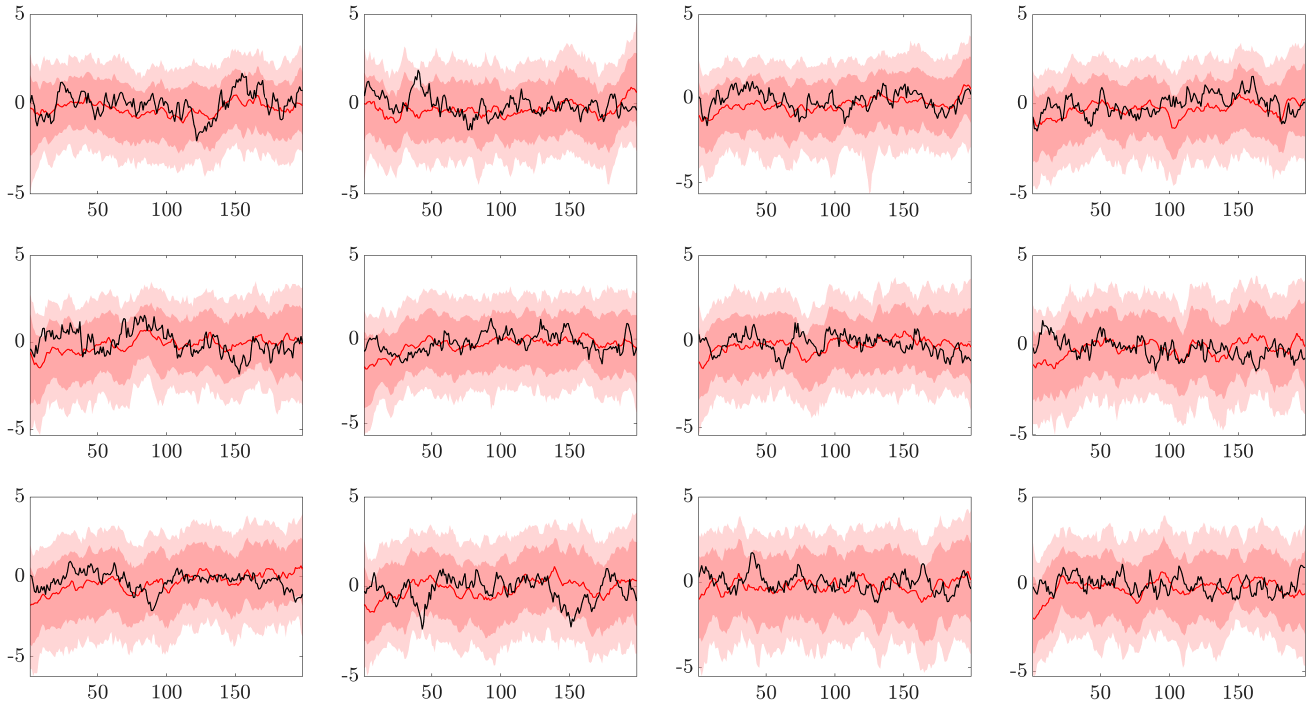

Figure 5: Path of the observations' log-volatility process, $\mathbf{h}$: true value (black line), posterior median (red line), 90% and 67% credible interval (light and dark red shading, respectively). Dataset: $n = 12$, $T = 200$.

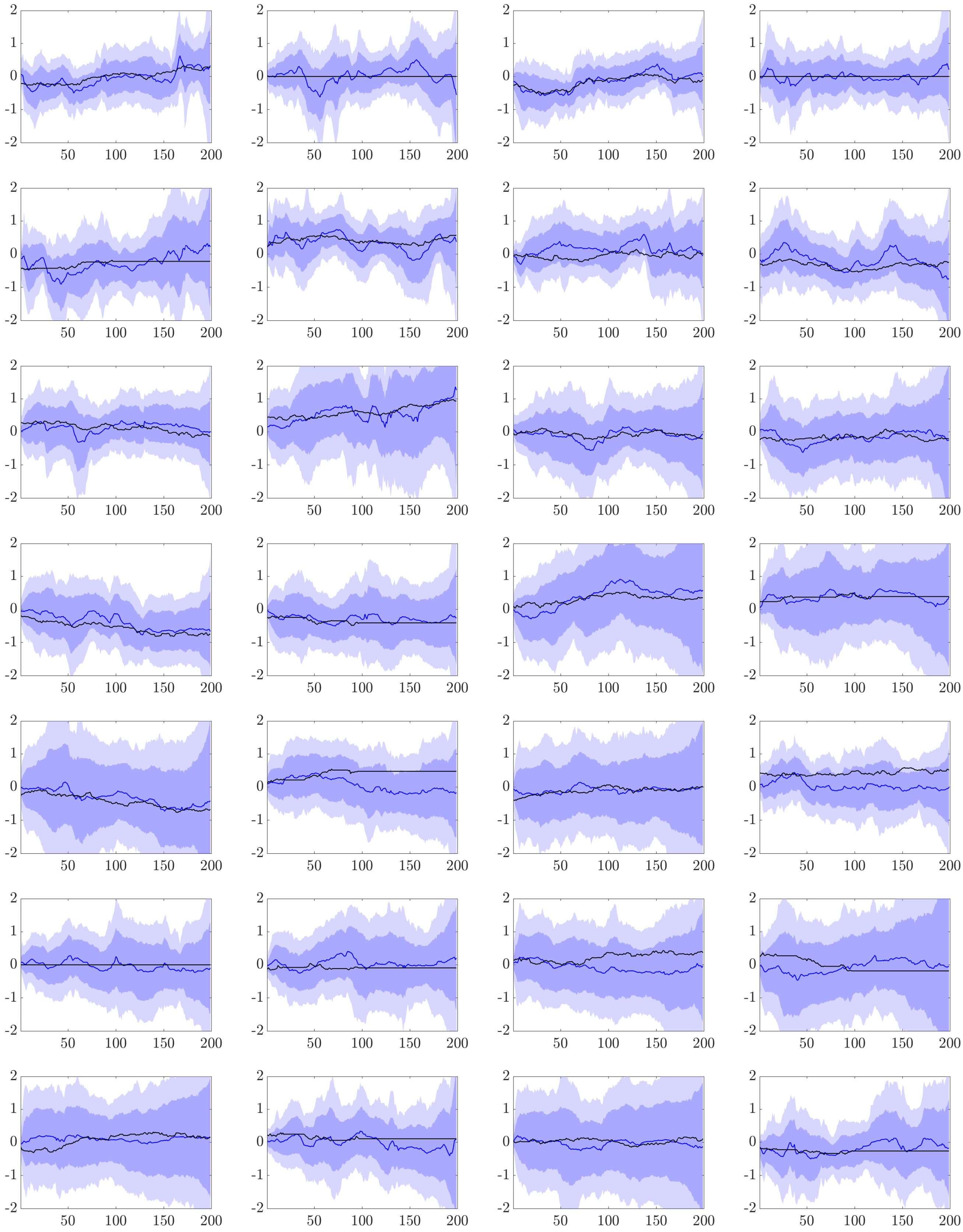

Figure 6: Path of selected time-varying coefficients', $\boldsymbol{\theta}$: true value (black line), posterior median (blue line), 90% and 67% credible interval (light and dark blue shading, respectively). Dataset: $n = 12$, $T = 200$.

5 Real Data Application

In this section, we present some preliminary results from an exercise using U.S. quarterly macroeconomic data taken from the FRED-QD dataset (see McCracken and Ng, 2016). The data spans from 1959Q1 to 2019Q4 and we focus on GDP, inflation (based on the GDP deflator), and the Fed Funds rate (FFR). All the variables considered are standardized and we consider a TVP-VAR-SV model with 1 lag.

Figure 8 shows the estimated path of the time-varying coefficients, $\boldsymbol{\theta}$ jointly with the 95% and 67% credible intervals. The main findings concern (i) the time variation of almost all the coefficients and (ii) the significance of only some of the coefficients across time (as given by the C.I. not including the 0). The results show different path across the coefficients, with some of them varying across the time, while other remaining constant or close to zero, thus providing signal for sparsity.

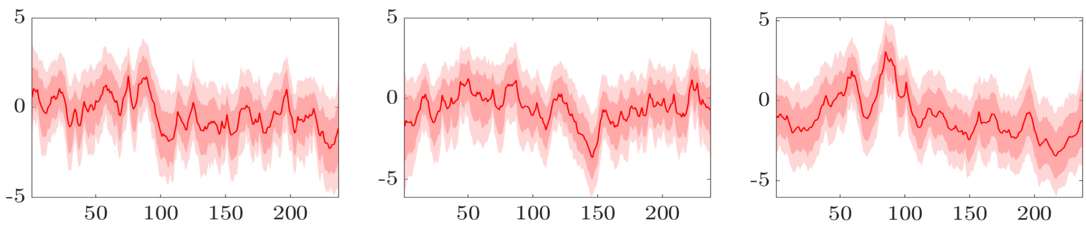

Figure 7: Path of the observations' log-volatility process, $\mathbf{h}$: posterior median (red line), 90% and 67% credible interval (light and dark red shading, respectively). FRED dataset, with $n = 3$, $T = 238$.

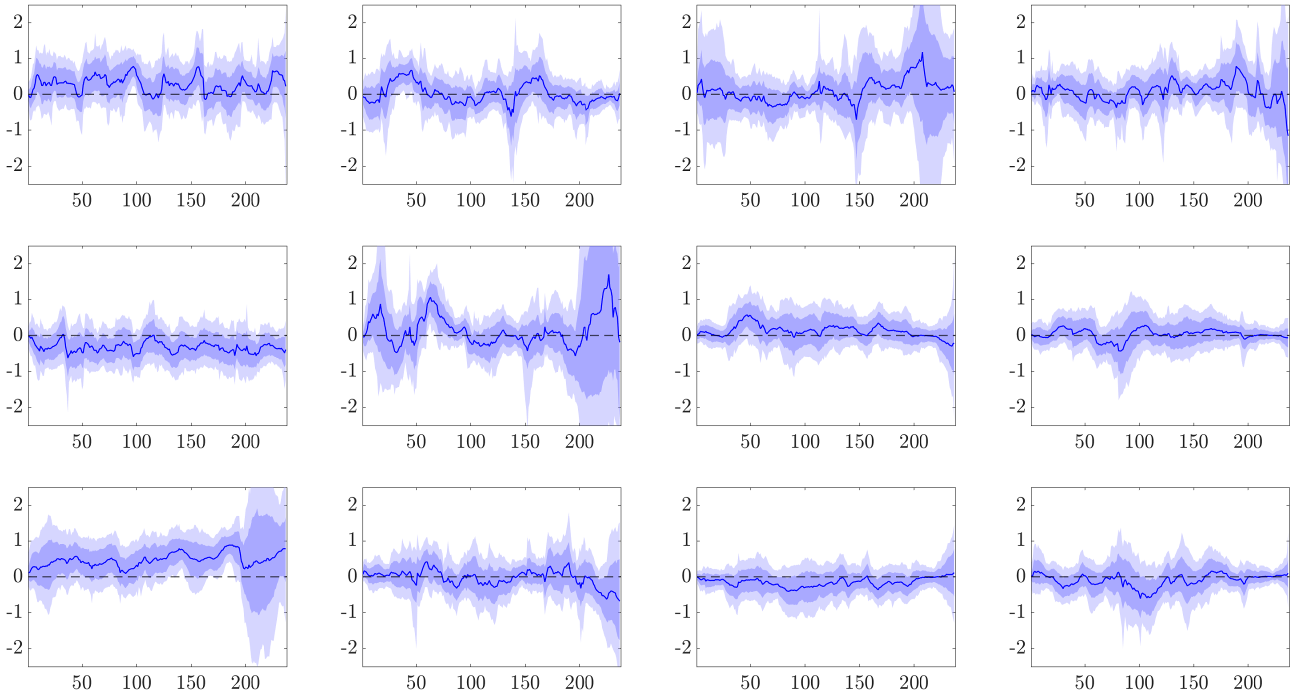

Figure 8: Path of selected time-varying coefficients', $\boldsymbol{\theta}$: zero value (black-dashed line), posterior median (blue line), 90% and 67% credible interval (light and dark blue shading, respectively). FRED dataset, with $n = 3$, $T = 238$.

6 Conclusions

Time-varying parameters models have increased the popularity in economics and finance although they deal with over-parametrization issues. In the current literature, different prior specifications, such as global-local shrinkage prior have proposed to induce sparsity. In this paper we have defined a new semiparametric TVP-SVAR model with stochastic volatility that is able to recover several types of trajectories for the time-varying (contemporaneous and lagged) coefficients, including constant, piecewise constant, random walk, and their combinations. Differently from the standard choice of prior distributions, our proposed prior allows to shrink the coefficients to zero and to cluster them. Moreover, the use of a dynamic horseshoe prior for the time-varying coefficients allows to capture a wide range of possible trajectories, including (piece-wise) constant and random walk behaviours. The performance of the model has been tested in several experiments on synthetic data.

References

Barigozzi, M. and Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364.

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

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics*, 2010(1):58–74.

Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *Journal of Econometrics*, 212(1):97–115.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

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

Chan, J. C. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter VARs with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532.

Chan, J. C. and 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.

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

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Diebold, F. X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

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

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230.

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

Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, 30(1):1–11.

Hatjispyros, S. J., Nicoleris, T., and Walker, S. G. (2011). Dependent mixtures of Dirichlet processes. *Computational Statistics & Data Analysis*, 55(6).

Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, 37(1):27–39.

Huber, F., Koop, G., and Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, 39(3):669–683.

Huber, F. and Pfarrhofer, M. (2021). Dynamic shrinkage in time-varying parameter stochastic volatility in mean models. *Journal of Applied Econometrics*, 36(2):262–270.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and computing*, 21(1):93–105.

Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, 76:408–423.

Kowal, D. R., Matteson, D. S., Ruppert, D., et al. (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society Series B*, 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.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, pages 351–357.

MacLehose, R. F. and Dunson, D. B. (2010). Bayesian semiparametric multiple shrinkage. *Biometrics*, 66(2):455–462.

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

McCracken, M. W. and Ng, S. (2021). FRED-QD: A quarterly database for macroeconomic research. *Federal Reserve Bank of St. Louis*, 103(1):1–44.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.

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

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, 36(1):45–54.