Moderate Time Varying Parameter VARs *

Alessandro Celani †
Örebro University

Luca Pedini ‡
Fondazione ENI Enrico Mattei

October 17, 2025

Abstract

This article proposes a new parametric approximation for time-varying parameter models particularly suited for slowly evolving dynamics. Specifically, our "moderate" time-variation approach rewrites the transition equation of a coefficient in terms of B-splines and a low-dimension dynamics, reducing the number of parameters to a scale factor equal to the number of spline knots. Using both simulated and real macroeconomic datasets of varying dimensions, we apply the proposed techniques to Vector Autoregressive models. We show that our methodology not only delivers interpretations consistent with benchmark models, but also improves forecast accuracy and computational times.

Keywords: Time-Varying Parameter models, High-dimension Vector Autoregressions, Stochastic Volatility, B-splines.
JEL classification: .....

*We wish to thank Igor Martins, ..., for their comments, and conference participants at the 3rd edition of the UEA Time Series workshop, the 15th Rimini Bayesian Econometrics workshop, the 31st international conference on Computing in Economics and Finance (CEF), ...

†Alessandro Celani, Örebro University School of Business, Fakultetsgatan 1, 702 81, Örebro, Sweden. Email: alessandro.celani@oru.se.

‡Luca Pedini, Fondazione ENI Enrico Mattei (FEEM), Corso Magenta 63, 20123, Milano, Italy. Email: luca.pedini@feem.it

1 Introduction

It is hard to overstate the role of Time Varying Parameter Vector Autoregressions (TVP VARs) in empirical macroeconometrics. Starting from the seminal works by Cogley and Sargent (2001, 2005) and Primiceri (2005), TVP VARs have become a central tool for analyzing the evolving relationship among macroeconomic variables over time. Their popularity stems from the flexibility to capture evolving, unstable economic dynamics, often featuring various forms of nonlinearities. From an applied perspective, their forecast accuracy is usually superior than their static counterpart, (D'Agostino et al., 2013; Koop and Korobilis, 2013; Clark and Ravazzolo, 2015); likewise, TVP schemes can offer valuable insights for structural analysis as well (Benati and Surico, 2008; Baumeister and Peersman, 2013; Boeck and Mori, 2025).

Similarly to any TVP framework, a TVP VAR postulate new sets of parameters at each point in time, raising concerns in terms of overparametrization. Bayesian methods mitigate but do not solve the dimensionality issue, which is actually magnified by the computational burden of Markov Chain Monte Carlo (MCMC) methods. For this reason, TVP methods are commonly used in small-scale VARs, with the potential shortcoming of omitted variable bias (Bańbura et al., 2010; Giannone et al., 2014). On the other hand, TVP models, by their own nature, usually use of as many coefficients as the length of the relevant time domain T. However, temporal instabilities that economists typically face are quite limited. Most of the unobserved structural dynamics have probably evolved over time, but at a very slow pace. In this context, a key question arises: do we really need T coefficients to capture such evolution?

Moved by this question, we introduce an efficient and parsimonious reparametrization for TVP in the realm of State Space (SS) models. Specifically, we decompose the T-dimensional parameter of the transition equation into a lower-dimensional one via deterministic functions of time, namely *B-spline basis*. In doing so, we reduce the magnitude of the parameter space to a scale factor equal to the number of *spline knots*, while preserving smooth dynamics. The generality is high: the canonical TVP pattern used in the literature can be easily recovered by assuming piecewise-constant (0-th degree) splines with as many knots as the temporal dimension. However, by using fewer knots, we effectively constrain the degree of temporal variability, improving estimation efficiency and computation time. Reducing the knots number is not sufficient, though: 0-th degree splines yield undesirable stepwise paths, which are sensitive to knot positioning. On the other hand, higher-degree splines yield smoother dynamics more resilient to knot positioning. Accordingly, our baseline exploits quadratic (2-nd degree) splines. Because the amount of time variation that can be captured shrinks as the number of knots reduces, we term the approach Moderate TVP (MTVP).

The proposed method is applicable to any linear or linearizable Gaussian SS model.

In this paper, we implement it in the context of TVP-VARs, henceforth Moderate TVP-VARs (MTVP-VARs), which allow us to explore the approach in both forecasting and structural exercises. In this way, our work adds to the growing literature on building efficient dimensionality reduction methods for TVP-VARs. Early Bayesian work includes Koop and Korobilis (2013) or Chan et al. (2020), who reduce the cross-sectional dimension of TVP coefficients via factor methods. More recently, Chan (2023) proposes a Bayesian stochastic-search method to endogenously determine static and TV coefficients. Other noteworthy methods for Bayesian TVP regression include Belmonte et al. (2014); Korobilis (2021); while Giraitis et al. (2018); Kapetanios et al. (2019) stands out for the use of non-likelihood based approaches. We also acknowledge the work of Barnichon and Brownlees (2019) and Tanaka (2020), who pioneered the use of penalized splines for data reduction in macroeconometric settings, although for local projection setups.

Coming to the estimation details of our MTVP-VAR, both frequentist and Bayesian methods are equally applicable: we adopt the latter. Due to the linearity of the decomposition, conditional posterior for the parameters and the latent states can be easily derived, thus enabling an efficient Gibbs sampling procedure. Nevertheless, the choice for valid prior hyperparameters is crucial, especially when comparing models of different dimensions and with different number of knots. To address this, we propose a hierarchical approach to estimate state prior variances jointly with the other parameters.

We run a comprehensive simulation experiment over different moderate setups (both smooth and piecewise-constant) to compare computational time, parameter biases, and sensitivity to knot placement with respect to the literature benchmark T-varying process. The MTVP specifications approximate non-pervasive TVP patterns very well, making the induced bias negligible. In truly non-TVP settings, MTVP tends to dominate, being closer to a limiting static specification. Quadratic splines achieve superior fitting performances, while piecewise-constant splines offer dramatic speed gains. Next, we move to real data, in particular a U.S. dataset of macroeconomic variables and estimate MTVP-VARs of different dimensions. A smooth moderate specification with parameters that vary every two years delivers estimated trends and structural objects statistically indistinguishable from the benchmark. In forecasting, the smoothed versions achieves again superior density forecasts, while point forecasts are comparable to the benchmark. Finally, the predictive marginal likelihood confirms the dominant position of the smooth moderate variant in both small and high-dimensional settings.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed methodology. We first develop the method for TVP regression, and then adapt it to TVP-VARs. We also detail the hierarchical prior framework considered and outline the resulting Gibbs sampler. Section 4 conducts the Monte Carlo exercises and in 5 we discuss the empirical application. Finally, Section 6 concludes and outlines directions for future research.

2 Moderate TVP regression

Consider a linear heteroskedastic TVP regression. This framework is general enough to encompass most of the models commonly employed in empirical macroeconomics (*e.g.*, the TVP-VAR we use in the paper). The measurement equation and the error term are defined as follows:

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

where y_t is the dependent variable, $\mathbf{x}_t$ a $M \times 1$ vector of explanatory variables, $\boldsymbol{\beta}_t$ is the corresponding vector of TV coefficients and $\sigma_t^2 = \exp(h_t)$ is the TV variance governed by the latent state h_t. As the focus of this paper is on the temporal dimension of the model, we find it convenient to adopt the non-centered parameterization pioneered by Frühwirth-Schnatter (2004); Frühwirth-Schnatter and Wagner (2010), which decomposes parameters into a constant and a TV component, the latter with zero initial conditions. The measurement equation rewrites

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

where now $\sigma_t^2 = \exp(h + \tilde{h}_t)$. $\boldsymbol{\beta}$ and h are the static components, whereas $\tilde{\boldsymbol{\beta}}_t$ and $\tilde{h}_t$ are the dynamic ones. The original parameters are recovered as $\boldsymbol{\beta}_t = \boldsymbol{\beta} + \tilde{\boldsymbol{\beta}}_t$, $h_t = h + \tilde{h}_t$. We complete the model by assigning dynamic hierarchical priors to $\tilde{\boldsymbol{\beta}}_t$ and $\tilde{h}_t$ in the form of driftless Random Walk (RW) processes:

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_t - \tilde{\boldsymbol{\beta}}_{t-1} &\sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{Q}}) \\ \tilde{h}_t - \tilde{h}_{t-1} &\sim \mathcal{N}(0, \tilde{s}) \end{aligned} \tag{3}$$

initialized at zero: $\tilde{\boldsymbol{\beta}}_0 = \mathbf{0}$, $\tilde{h}_0 = 0$. To prevent ambiguity and to facilitate comparison with the moderate specifications developed later, we refer to this benchmark framework as "*RW*", highlighting the form of the law of motion for the parameters. This specification represents an extreme case, assigning as many parameter states as time points.[1] Nonetheless, when temporal instability is present but not pervasive, dynamic heterogeneity may be captured with fewer parameters. Building on this idea, the MTVP framework embodies the notion of *moderate time variation*: parameters are allowed to evolve over time, although not necessarily in every period.

2.1 Proposed method

The goal of MTVP is to reduce the number of parameters to be estimated with minimal alteration of their law of motion. This is achieved by rewriting the state equation of

[1] At the opposite extreme is the fully static case obtained with $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ and $h_t = h$, which implies $\tilde{\boldsymbol{\beta}}_1 = \cdots = \tilde{\boldsymbol{\beta}}_T = \mathbf{0}$ and $\tilde{h}_1 = \cdots = \tilde{h}_T = 0$.

a generic parameter as a function of two components: a lower-dimensional state equation and a set of deterministic, time-varying weights. These weights map the lower-dimensional states back to the original time scale, and their functional form determines the shape of the reconstructed TV parameters.

We begin by considering the conditional mean parameter. Suppose to partition the temporal domain into $R = T/K$ non-overlapping intervals of equal length $1 \leq K \leq T$. For example, with quarterly data and $K = 8$ (assuming for the moment that T is a multiple of 8), each interval corresponds to two years. Let $\boldsymbol{\theta}_r$, $r = 1, \dots, R$, be an auxiliary M dimensional vector of parameters associated with the r-th interval. $\boldsymbol{\theta}_r$ captures lower-frequency time variation, as it is allowed to vary "just" R times across the time domain. To maintain coherence with the framework, we model the transition of $\boldsymbol{\theta}_r$ using a driftless RW:

$$\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{4}$$

with $\boldsymbol{\theta}_0 = \mathbf{0}$. The goal is to estimate $\boldsymbol{\theta}_r$ instead of $\tilde{\boldsymbol{\beta}}_t$, while still allowing parameters to evolve over t. The main challenge is to reconcile the mismatch between the two "temporal domains" and to construct a mapping between the two parameter vectors. To bridge this gap, we introduce a set of time-varying weights $w_{t,r}$ satisfying the usual constraints for each t: $w_{t,r} \geq 0$, $\sum_{r=1}^{R} w_{t,r} = 1$. We define the mapping as:

$$\tilde{\boldsymbol{\beta}}_t = \mathbf{W}_{t,r}\boldsymbol{\theta}_1 + \cdots + \mathbf{W}_{t,R}\boldsymbol{\theta}_R \tag{5}$$

where a common weighting scheme is assumed to govern the time variation of each element of $\tilde{\boldsymbol{\beta}}_t$, *i.e.*, $\mathbf{W}_{t,r} = w_{t,r}\mathbf{I}_M$.

Although the discussion of the functional form of the weights is deferred to the next section, one feature is worth emphasizing. It appears counterintuitive that $\tilde{\boldsymbol{\beta}}_t$ depends on all R regions at time t, however, the weights $w_{t,r}$ can be conveniently set to zero to restrict the region of influence.[2] Accordingly, the $T \times R$ matrix $\mathbf{W}$, with generic element $w_{t,r}$, is highly sparse, with nonzero entries concentrated along a narrow band around the main diagonal (see Figure 1 for some examples). Importantly, this general formulation nests the benchmark case and accommodates a range of intermediate specifications. In particular, with $R = T$ and a weighting scheme such that $w_{i,j} = 1$ if $i = j$, 0 otherwise, we recover the *RW*, since (5) reduces to $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\theta}_t$.

The proposed framework naturally extends to the SV, for which we specify the following decomposition of the latent state $\tilde{h}_t$:

$$\begin{aligned} \tilde{h}_t &= w_{t,r}g_1 + \cdots + w_{t,R}g_R \\ g_r - g_{r-1} &\sim \mathcal{N}(0, s) \end{aligned} \tag{6}$$

[2]For example, under the weighting scheme we adopt it effectively depends only on its own region and, at most, a few neighboring ones.

where, again, g_r is an auxiliary parameter initialized at $g_0 = 0$. For clarity, we use the same number of regions R as in the conditional mean specification, although in practice R may differ between the mean and volatility components, a possibility we consider throughout the paper. [3]

2.2 Weight selection and estimation strategy

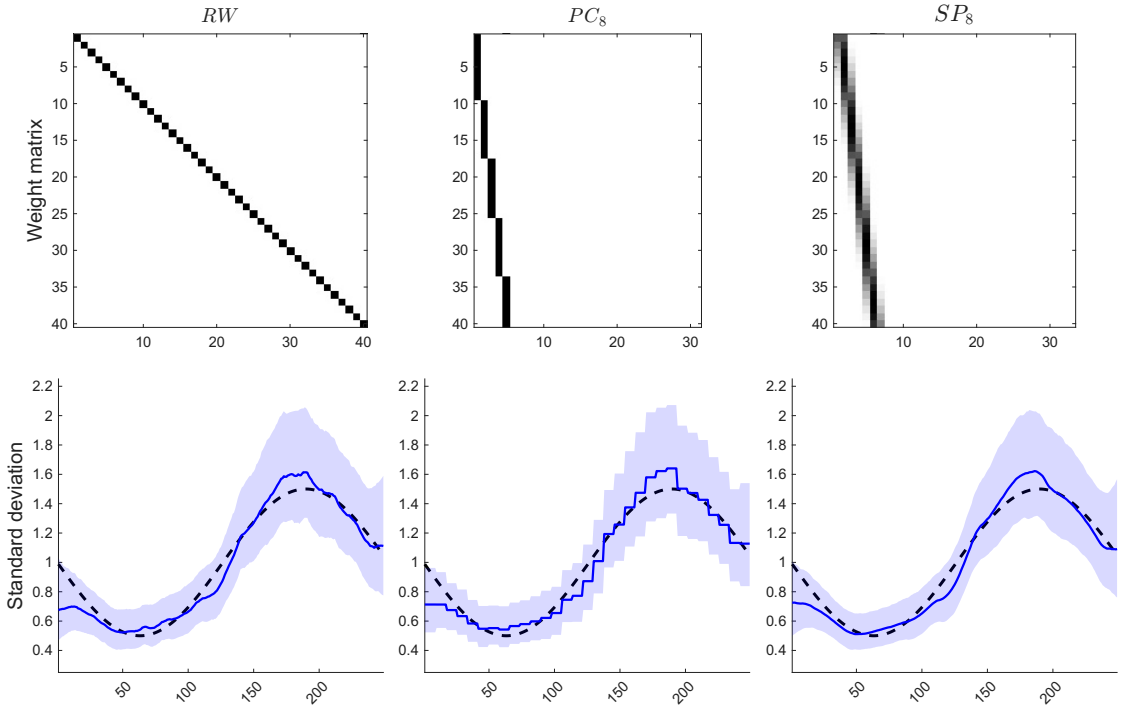
The method crucially depends on the choice of the length of the intervals K and the functional form of the weights $w_{t,r}$, which determine the size and structure of the matrix $\mathbf{W}$, respectively. When $K > 1$, each auxiliary parameter captures the time variation of multiple time points; on the other hand, $w_{t,r}$ defines how these points contribute to the parameter. For example, with the previously mentioned 0-1 weights, the method yields stepwise dynamics. Regardless of the fit such a specification may deliver, it is desirable to have smooth evolution over time to avoid abrupt jumps in law of motion. Intuitively, the columns of $\mathbf{W}$ should gradually vary along the time dimension. Smooth weighting schemes can be generated using standard basis functions such as *B-splines*, *Fourier bases*, or *wavelets*.

In this paper we use B-Splines (BSP). A BSP consists of $q + 1$ polynomial segments of degree q, which are connected at q inner knots. The BSP is positive only within the domain defined by the q knots and is zero outside. This compact support guarantees that each $\boldsymbol{\beta}_t$ depends only on $q+1$ regions, yielding a desirable connection to a few neighboring points. In this way, we enforce a "*forgetfulness*" of distant past observations. Here we use quadratic ($q = 2$) splines as our baseline. They are the lowest-order splines that yield smooth parameter paths, having continuous first derivatives. Thus, they naturally align with the first difference priors commonly used for TVP models as per Equation (3).[4] For comparison purposes, we also analyze the ($q = 0$) degree splines as a reference for stepwise movements. A general moderate smooth configuration is denoted as SP_K, and a piecewise one as PC_K. We remain agnostic about which one is better: neither should be universally preferable *a priori* and the choice depends on the researcher's objectives and application. PC_K is quite sensitive to knot placement, but delivers substantial computational gains even for small K and in low dimensions, making it attractive when speed is prioritized. By contrast, SP_K specifications produce smoother dynamics, which

[3] Empirical evidence indicates that volatilities of macroeconomic variables often display pronounced temporal instability, while the conditional mean parameters tend to be more stable. This suggests that the optimal R is typically larger for volatilities than for conditional means.

[4] In the penalized spline literature, it is common to use cubic ($q = 3$) splines combined with a second-derivative penalization, so that the resulting function also has a continuous second derivative. In our time series context, this choice would correspond to imposing second-difference priors, effectively suggesting $I(2)$ dynamics, which appears unlikely for the macroeconomic dynamics we analyze. An exception arises when modeling trends in nonstationary variables, usually treated as $I(2)$ processes (Hodrick and Prescott, 1997; Grant and Chan, 2017). Moreover, raising the spline degree increases the number of non-zeros in the $\mathbf{W}$ matrix, which in turn worsens computational times.

Figure 1: **The effect of different weighting schemes**



Note: The figure illustrates the effect of different moderate specifications on the estimation of the SV of the residuals for a generic variable in a single realization of a simulated TVP-VAR, as described in Section 4. The sample size is $T = 250$. *First row*: first 40 rows of the three weight matrices considered. The number of columns varies, being equal to $\min(40, R)$, which yields 40 for the *RW* case, 31 for the PC_8, and 33 for the SP_8. The shading ranges from 0 (white) to 1 (black). *Second row*: black dashed lines denote the true process, solid blue lines represent the posterior median, and shaded blue areas indicate the 90% credible intervals.

are less sensitive to knot placement, and generally achieve superior parameter fit relative to PC_K. Their computational advantages, however, become pronounced mainly in high-dimensional settings. A comprehensive assessment of their properties and performance, including sensitivity to knot placement and CPU times, is provided in Section 4. Throughout the paper, we consider intervals of two, four, and six years for quarterly data, namely:

$$K = [8, 16, 24]$$

with equidistant knots starting at $1 - qK$ with spacing K. For each specification, the total number of partitions is $R = \bar{R} + q$, with $\bar{R} = \lfloor T/K \rceil$, where $\lfloor \cdot \rceil$ denotes rounding to the nearest integer for any chosen K. Based on the partitioning and spline degrees described above, we compare the following models alongside the *RW*:

$$\mathcal{M} = \{SP_8, SP_{16}, SP_{24}, PC_8, PC_{16}, PC_{24}\}$$

To provide the reader an idea on how the machinery works, we briefly present a result from the simulation exercise of Section 4, where the MTVP-VAR processes with parame-

ters following deterministic sine and cosine dynamics. While the exact motivation, DGP form, and details can be found in the related section, we simply illustrate here how different weighting schemes affect parameter estimation. In particular, we show the fit of the RW, PC_8, and SP_8, on the TV standard deviation of the residual of a generic equation in the TVP-VAR. In the second row of Figure 1, the black dotted lines denote the true process, while the blue lines and shaded areas represent the posterior median and 90% credible intervals, obtained under the same prior framework but with different weighting schemes. The corresponding weight matrices are shown in the top row. PC_8 and SP_8 use 8 times fewer parameters, but differ in their aggregate dynamics due to the weight functional form: piecewise constant and smooth, respectively. All specifications perform well in capturing the underlying process, even parsimonious ones. Nonetheless, the spline degree is influential: the third column is visually preferable, as it avoids the irregular pattern of the second. [5]

Bayesian estimation. The proposed decomposition preserves linearity in the parameters to be estimated, enabling straightforward Gaussian updates within a Gibbs sampler. As evident from Equations (5) and (6), the multiplicative interaction between the weight matrices and auxiliary parameters poses no difficulty, since the former are fixed prior to estimation. To make this explicit, and to motivate the estimation strategy adopted in the paper, we rewrite the model in compact form. Let $\mathbf{y} = [y_1, \ldots, y_T]'$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]'$, $\tilde{\mathbf{X}} = \text{diag}(\mathbf{x}_1', \ldots, \mathbf{x}_T')$, and $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_T]'$. The stacked measurement equation is a linear regression model with a general error covariance matrix:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{7}$$

where $\tilde{\boldsymbol{\beta}} = [\tilde{\boldsymbol{\beta}}_1', \ldots, \tilde{\boldsymbol{\beta}}_T']'$, and $\boldsymbol{\Sigma} = \text{diag}(\exp(\mathbf{h}))$, $\mathbf{h} = h + [\tilde{h}_1, ..., \tilde{h}_T]'$. Equation (5) can be expressed compactly over T as a Kronecker product:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{W} \otimes \mathbf{I}_M)\boldsymbol{\theta} \tag{8}$$

with $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_R']'$. Substituting Equation (8) into Equation (7) shows the linearity in $\boldsymbol{\theta}$; the combination with the Gaussian prior for $\boldsymbol{\theta}$ leads to the usual result of Gaussian conditional posteriors. Draws from the auxiliary latent states can be efficiently performed using the precision sampler by Chan and Jeliazkov (2009), which allows for a joint sampling procedure in contrast to backward-forward filtering techniques. Similar conclusions apply to the log volatility parameters. Specifically, the vector error term in Equation (7) can be reformulated as a linear function of $\mathbf{g} = [g_1, \ldots, g_R]'$ as follows:

$$\log \boldsymbol{\epsilon}^2 = h + \mathbf{W}\mathbf{g} + \mathbf{u}$$

[5] Figure 8 in the Appendix C shows the same exercise using $K = 16$.

where each component of $\mathbf{u}$ follows a log Chi-Square distribution. Draws from the posterior are obtained by adopting the mixture representation of Kim et al. (1998) in conjunction with the previously mentioned precision sampler.

3 Moderate TVP VAR

We now introduce a class of models we call MTVP-VARs, *i.e.*, TVP-VARs in which the proposed decomposition is applied to both mean coefficients and log volatilities for each equation in the system. For computational convenience, we adopt a slight modification of the original Primiceri (2005) framework by using the equation-by-equation representation of Chan (2023), valid under the recursive identification scheme.

Let $\mathbf{y}_t$ be an N-dimensional vector of endogenous variables whose dynamics are governed by a TVP-VAR with P lags in recursive structural form:

$$\mathbf{\Phi}_{0,t}\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \cdots + \mathbf{\Phi}_{P,t}\mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t), \tag{9}$$

where $\mathbf{\Phi}_{0,t}$ is a lower triangular contemporaneous coefficient matrix with ones on the main diagonal, $\boldsymbol{\mu}_t$ is the vector of intercepts, $\mathbf{\Phi}_{1,t}, \ldots, \mathbf{\Phi}_{P,t}$ are the matrices of lagged coefficients, and $\mathbf{\Sigma}_t = \text{diag}\left(\exp(h_{1,t}), \ldots, \exp(h_{N,t})\right)$. This model can be rewritten as a sequence of N independent non-centered TVP regressions with an expanding set of explanatory variables and SV of the form:

$$y_{i,t} = \mathbf{z}_{i,t}'\boldsymbol{\beta}_i + \mathbf{z}_{i,t}'\tilde{\boldsymbol{\beta}}_{i,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \mathcal{N}(0, \exp(h_i + \tilde{h}_{i,t})) \tag{10}$$

where $\mathbf{x}_t = [1, \mathbf{y}_{t-1'}, ..., \mathbf{y}_{t-P}']$', $\mathbf{z}_{1,t} = \mathbf{x}_t$, and $\mathbf{z}_{i,t} = [\mathbf{x}_t', -y_{1,t}, \ldots, -y_{i-1,t}]'$ for $i > 1$. Our MTVP-VAR is obtained by applying the moderate decomposition to each of these TVP regressions:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_{i,t} &= \mathbf{W}_{1,t}\boldsymbol{\theta}_{i,1} + \cdots + \mathbf{W}_{R,t}\boldsymbol{\theta}_{i,R} \\
\tilde{h}_{i,t} &= w_{1,t}g_{i,1} + \cdots + w_{R,t}g_{i,R} \\
\boldsymbol{\theta}_{i,r} - \boldsymbol{\theta}_{i,r-1} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i) \\
g_{i,r} - g_{i,r-1} &\sim \mathcal{N}(0, s_i)
\end{aligned} \tag{11}$$

where $\mathbf{W}_{t,r} = w_{t,r}\mathbf{I}_{M_i}$, with $M_i = NP + i$ indicating the total number of coefficients in the i-th equation. We assume that the error covariance matrices for the state equations are diagonal, $\mathbf{Q}_i = \text{diag}(q_{i,1}, \ldots, q_{i,M_i})$, with elements following Inverse Gamma distributions. The same holds for the state variance of the log volatilities:

$$\begin{aligned}
q_{i,j} &\sim \mathcal{IG}(\tau_1, (\tau_1 + 1)\xi_{1,j}), \quad j = 1, \ldots, K_i \\
s_i &\sim \mathcal{IG}(\tau_2, (\tau_2 + 1)\xi_2)
\end{aligned} \tag{12}$$

where $\mathcal{IG} \sim (a, b)$ identifies an Inverse-Gamma distribution with shape and scale as a and b, respectively; τ_1 and τ_2 are prior shapes, and $\xi_{1,j}$ and ξ_2 denote prior modes. To distinguish between intercepts and other coefficients, we set $\xi_{1,j} = \bar{\xi}_1$ for VAR coefficients ($j > 1$), and $\xi_{1,j} = 4^2\bar{\xi}_1$ for intercepts ($j = 1$), with $\bar{\xi}_1$ representing the common prior mode.

Choosing state prior hyperparameters is problematic even in the benchmark model. In our framework the problem magnifies, as there is no guarantee that values commonly used in the literature are appropriate for moderate specifications with arbitrary K. To address this, we adopt a simple hierarchical framework by placing a Gamma prior on $\bar{\xi}_1$, which yields an analytical conditional posterior for the parameter:

$$\bar{\xi}_1 \sim \mathcal{G}(2, \gamma/2) \tag{13}$$

with $\mathcal{G}(a, b)$ denoting a Gamma distribution with shape a and scale b, so that $\mathrm{E}(\bar{\xi}_1) = \gamma$. The remaining hyperparameters are adapted across specifications by ensuring the same level of prior shrinkage (through shapes) and expected variance (through modes):

$$\begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} = K^{-1} \begin{bmatrix} 40 \\ 10 \end{bmatrix}, \quad \begin{bmatrix} \gamma \\ \xi_2 \end{bmatrix} = K \begin{bmatrix} .01^2 \\ .1^2 \end{bmatrix}$$

Remaining priors and Gibbs sampler. We employ Gaussian priors for both the mean and log variance static parameters: $\boldsymbol{\beta}_i \sim \mathcal{N}(\mathbf{0}, \bar{\mathbf{Q}}_i)$, $h_i \sim \mathcal{N}(0, 5^2)$. Each $\bar{\mathbf{Q}}_i$ is diagonal and follows a classical Minnesota style (Doan et al., 1984). To this end, let $\bar{\mathbf{Q}} = \mathrm{diag}(\bar{\mathbf{Q}}_1, \dots, \bar{\mathbf{Q}}_N)$ be the full covariance matrix of the static parameters. In particular, its generic j-th diagonal element $\bar{q}_{jj}$ is equal to

$$\bar{q}_{jj} = \begin{cases} \frac{\lambda_1}{l^2}, & \text{own } l\text{-th lag} \\ \frac{\lambda_2 \sigma_i^2}{l^2 \sigma_j^2}, & \text{cross-equation } l\text{-th lag} \\ \frac{\lambda_3 \sigma_i^2}{\sigma_j^2}, & \text{contemporaneous variables} \\ \lambda_4 \sigma_i^2, & \text{intercept} \end{cases}$$

where σ_i^2 denotes the sample variance of the residuals from individual AR(4) of each $y_{i,t}$, $\lambda_3 = .2^2$, $\lambda_4 = 10^2$ are fixed shrinkages, whereas λ_1, λ_2 are hierarchical shrinkages to be estimated. In particular, we follow Chan (2021) and append to both a Gamma prior: $\lambda_1 \sim \mathcal{G}(2, .2^2/2)$, $\lambda_2 \sim \mathcal{G}(2, .1^2/2)$.

To finalize the estimation step we define the vectorized version of the TV parameters of interest $\boldsymbol{\theta}_i = [\boldsymbol{\theta}'_{i,1}, \dots, \boldsymbol{\theta}'_{i,R}]'$, and $\mathbf{g}_i = [g_{i,1}, \dots, g_{i,R}]'$. Bayesian estimation is performed

using a Gibbs sampler that iterates over the following conditional posterior distributions: $p(\boldsymbol{\theta}_i|\bullet)$, $i = 1, \ldots, N$, $p(\mathbf{g}_i|\bullet)$, $i = 1, \ldots, N$, $p(q_{i,j}|\bullet)$, $i = 1, \ldots, N$, $j = 1 \ldots, K_i$, $p(s_i|\bullet)$, $i = 1, \ldots, N$, $p(\bar{\xi}_1|\bullet)$, $p(\lambda_1|\bullet)$, $p(\lambda_2|\bullet)$, where $\bullet$ represents the set of conditioning parameters except the one that we sample from. All the details regarding full conditional posteriors and their derivation are given in Appendix B.

4 Monte Carlo Study

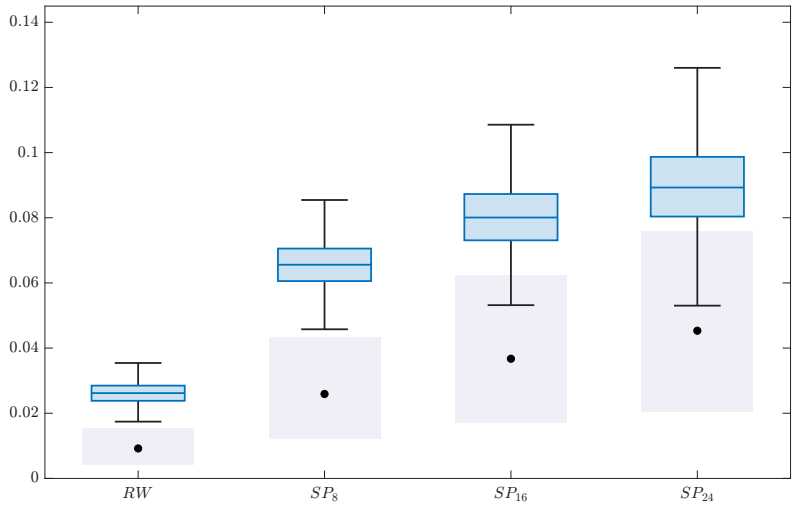
In this section we conduct a simulation study to assess the performance of MTVP-VAR models under parameter instability. The simulation setup closely follows Amir-Ahmadi et al. (2020): the DGP is a TVP-VAR(1) without contemporaneous interactions ($\boldsymbol{\Phi}_{0,t} = \mathbf{I}_N$) and with a lower-triangular structure for $\boldsymbol{\Phi}_{1,t}$, which allows to test the model ability to capture both constant and TV parameters. Both the mean and log variance coefficients evolve according to deterministic sine and cosine functions with varying amplitudes and peak locations, whose details are provided in the Appendix D. All setups consider $T = 250$ observations and $N = 3$, although larger dimensions are also explored when analyzing computational times. We use the prior specification outlined in Section 3.

We divide the Monte Carlo exercise into three sub-experiments: firstly, we evaluate the importance of the hierarchical prior on the latent state variances. Secondly, we analyze the CPU times of the moderate schemes with respect to the *RW*. Finally, we conduct a comprehensive Monte Carlo exercise to assess the MTVP performances, by comparing estimated parameters to the true ones and examining sensitivity to knot placement.

Figure 2 reports the result of the first experiment. The boxplots indicate the posterior distributions of $\sqrt{\bar{\xi}_1}$, while the shaded areas refer to the 90% mass of the prior distribution. As shown in the first column, for the *RW* fixing $\bar{\xi}_1 = .01^2$ (black dot) is clearly suboptimal: the posterior distribution falls outside the prior 90% bands, indicating that a fixed-prior is inadequate. The same pattern repeats across the other specifications. Crucially, the compensation mechanism introduced via the hierarchical prior, which shifts scale and shape, appears necessary: the posterior median and related boxplot bands point to a larger standard deviation. Nonetheless, the magnitude of the adjustment differs by specification, motivating a modular hierarchical approach.

Next, we document the running times of the proposed models. Table 1 reports the runtime for obtaining 1000 posterior draws of the considered specifications. Results are shown for increasing system dimensions: $N = 3$ (small), $N = 14$ (large), and $N = 50$ (very large). Across all cases, moderate specifications deliver substantial time savings relative to the benchmark. As the number of regions decreases, runtime improves at a different speed. The stepwise *PC* specifications show dramatic reductions, whereas the general smooth *SP* exhibits more modest gains. The efficiency gains become more pronounced as the system dimension increases. In the very high-dimensional case ($N =$

Figure 2: **Posterior of $\sqrt{\bar{\xi}_1}$ across specifications.**



Note: The figure shows the posterior distribution of the hierarchical standard deviation $\sqrt{\bar{\xi}_1}$ for the RW and different moderate specifications SP for a single Monte Carlo replication with $N = 3$. Shaded areas indicate the 90% prior probability bands, black dots mark the prior medians, and boxplots summarize the posterior draws.

50), the SP_{24} specification is almost four times faster than the benchmark RW, while the PC_{24} is close to ten times faster. This makes the class of moderate specifications increasingly attractive as N grows, which is particularly relevant for very large-scale applications with hundreds of dependent variables, where estimating TVP models with MCMC methods remains infeasible (Kapetanios et al., 2019). The marked difference between SP and PC may appear counterintuitive at a first glance, but is well motivated by the nature of the weighting matrices: in the PC case, $\mathbf{W}$ is composed by Boolean entries, meaning a parsimonious object in terms of machine memory. On the other hand, SP implies a double-precision matrix, far more computationally intensive.

We now turn to the full Monte Carlo exercise, where we simulate a total of $S = 200$ datasets and evaluate the models using three performance metrics for the estimated parameters: mean squared error (MSE), coverage probability (Coverage), and the average length of credible intervals (Length). Let β_t^{true} denote a generic true parameter and $\hat{\beta}_{m,t}$ the corresponding posterior median estimate from model $m \in \mathcal{M}$ in simulation $s = 1, ..., 200$ [6]. The MSE is defined as the mean of the squared errors:

$$\text{MSE}_m = T^{-1} \sum_{t=1}^{T} \left(\hat{\beta}_{m,t} - \beta_t^{true} \right)^2$$

Coverage is the arithmetic mean of the probability that the true value is within the 90%

[6]For estimated parameters and resulting criteria we avoid the s subscript for ease of exposition.

Table 1: **Computational time**

	RW	SP_8 (PC_8)	SP_{16} (PC_{16})	SP_{24} (PC_{24})
$N = 3$	1.76	23.8% (45.2%)	34.3% (49.4%)	37.3% (50.7%)
$N = 14$	34.81	26.3% (66.7%)	42.1% (71.9%)	46.9% (73.7%)
$N = 50$	1587	44.5% (87.5%)	62.7% (91.6%)	70.5% (93.1%)

Note: The first column reports the runtime required to obtain 100 posterior draws for the benchmark RW. The remaining entries show the percentage reduction in computational time relative to the benchmark model. The RW column reports the nominal runtime in seconds.

credible interval:

$$\text{Coverage}_m = T^{-1} \sum_{t=1}^{T} \mathbb{1}\left(\hat{\beta}_{m,t}^{5\%} < \beta_t^{true}\right) \times \mathbb{1}\left(\beta_t^{true} < \hat{\beta}_{m,t}^{95\%}\right)$$

where $\beta_{m,s,t}^{5\%}$ and $\beta_{m,s,t}^{95\%}$ are the posterior 5-th and 95-th percentile estimates, with $\mathbb{1}(\cdot)$ being the indicator function. Length denotes the arithmetic mean of the length of the 90% credible interval,

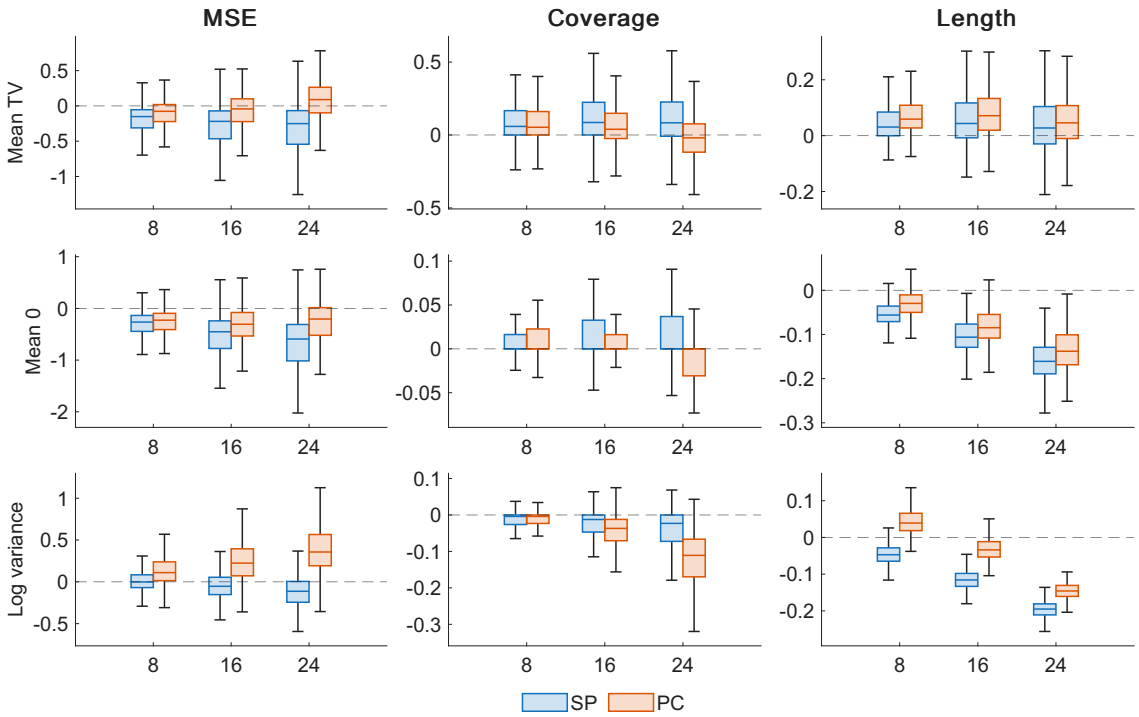
$$\text{Length}_m = T^{-1} \sum_{t=1}^{T} \left(\hat{\beta}_{m,t}^{95\%} - \hat{\beta}_{m,t}^{5\%}\right)$$

For each metric, we report results relative to the benchmark in log differences. [7] Figure 3 reports the results in the form of boxplots across replications. Parameters are grouped by similarity across rows: the $\mathbf{\Phi}_{1,t}$ TV and zero coefficients in the first two rows, and TV log volatilities in the last row.

Starting with MSE, the SP specifications outperform the benchmark across all coefficient classes. As expected, the gains are larger for zero-mean coefficients. For coverage and length, results are mixed. With zero-mean coefficients, SP attains higher coverage at shorter band length, making it universally preferable, as expected. Band length also decreases as K grows, making lower parametrized specifications very precise at little cost, consistent with convergence toward a static specification. For mean-TV coefficients, achieving higher coverage requires longer bands, whereas for log variances SP substantially shortens bands with only a slight drop in coverage. Performance is notably stable across K, indicating robustness to interval-length selection. On the other hand, PCs work adequately well for zero coefficients, while for TV ones their performance deteriorates markedly with longer regions, making PC attractive only with $K = 8$. To sum up, the SP emerges as a clear winner across the three proposed criteria. That said, as Table 1 shows, the PC_8 version remains attractive given its computational advantage.

[7] For example, for the MSE, we have $\log(\text{MSE}_m/\text{MSE}_{RW})$.

Figure 3: **Results of the Monte Carlo simulation**



Note: The figure reports the results of the simulation exercise. Panels are organized by rows according to the type of parameter evolution: TV mean coefficients (first row), zero mean coefficients (second row), and TV log volatilities (third row). Each column corresponds to a different evaluation metric. Blue boxplots denote smooth (SP) specifications, orange boxplots denote piecewise constant (PC) specifications.

Longer-interval PC variants should be considered only when time variation can be nearly ruled out.

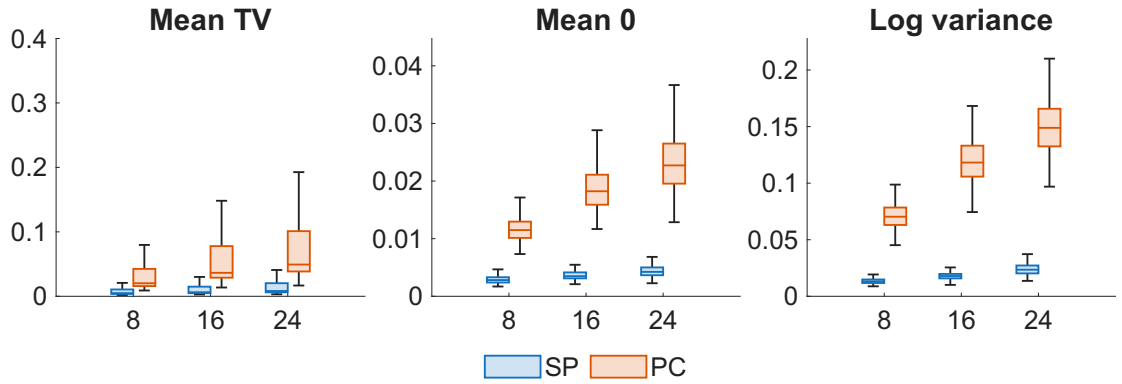
Knot positioning. The main weakness of the MTVP approach is its sensitivity to spline knot placement, a concern that magnifies as K increases. Sensitivity also differs across specifications: PC versions are more vulnerable, whereas SP variants mitigate it by construction through a smooth weighting scheme. In our setting, each moderate specification allows up to K distinct knot locations over time; depending on how much time variation the data truly exhibit, this can be manageable or problematic. Our claim is that, for most macro applications, time variation is not pervasive, so the knot-placement issue is likely second-order relative to the gains. Whether this holds in practice is an empirical question, which we assess via a simulation exercise.

To assess knot placement explicitly, we extend the earlier MSE definition. Previously, the MSE (as the other criteria) were computed for a single, implicit knot position. Since there are up to K admissible placements, there is an equivalent number of MSEs per specification

$$\text{MSE}_{k,m} = T^{-1} \sum_{t} (\hat{\beta}_{k,m,t} - \beta_t^{true})^2$$

where $\hat{\beta}_{k,m,t}$ is the time t estimated parameters under knot positioning $k = 1, ..., K$. The

Figure 4: **Results of the knot positioning exercise**



Note: The figure presents the results of the Monte Carlo based knot positioning exercise. Panels are organized by the type of parameter evolution. Blue boxplots indicate smooth (SP) specifications; orange boxplots indicate piecewise-constant (PC) specifications.

"model level MSE" then marginalizes over all knot placements and is defined as

$$\bar{\text{MSE}}_m = (KT)^{-1} \sum_k \sum_t (\hat{\beta}_{k,m,t} - \beta_t^{true})^2 = K^{-1} \text{MSE}_{m,k} \tag{14}$$

We now introduce $\bar{\hat{\beta}}_{m,t} = K^{-1} \sum_k \hat{\beta}_{k,m,t}$, the average posterior mean across all knot placement specifications. With this in hand, we define the following Bias-Variance style decomposition of eq. (14):

$$\bar{\text{MSE}}_m = \underbrace{T^{-1} \sum_t \left(\beta_t^{true} - \bar{\hat{\beta}}_{m,t}\right)^2}_{\text{Bias}_m^2} + \underbrace{(KT)^{-1} \sum_k \sum_t \left(\hat{\beta}_{k,m,t} - \bar{\hat{\beta}}_{m,t}\right)^2}_{\text{Variance}_m} \tag{15}$$

where the first term is the MSE between the true coefficient and the posterior mean averaged over all knot placements for each model.[8] As shown in Figure 9 in Appendix C, this component is effectively constant across specifications. Intuitively, averaging over placements neutralizes positioning effects, making the result equivalent to estimating *i)* a single specification with $K = 1$ or *ii)* he average of estimates over K knot locations when $K > 1$. Our main concern is the second term, which captures the variability across knot placements around their mean estimate. This quantity is zero for RW and, for the moderate specifications, measures the error incurred by ignoring knot placement. Figure 4 reports the results. Each boxplot shows the empirical distribution of the square root of the variance term in Eq. (15) across Monte Carlo replications. What stands out is the relatively small magnitude of all moderate errors (both SP and PC): the median

[8] Eq. 15 simpliefies because the cross term $2(KT)^{-1} \sum_k \sum_t \left(\beta_t^{true} - \bar{\hat{\beta}}_{m,t}\right)\left(\hat{\beta}_{k,m,t} - \bar{\hat{\beta}}_{m,t}\right) = 0$ due to $\sum_k \hat{\beta}_{k,m,t} - K\bar{\hat{\beta}}_{m,t} = 0 \; \forall t$.

is always below 0.01 for mean-TV coefficients and below 0.02 for log variances. This is striking given that the mean-TV intercepts range of variation (max-min) go from .6 to 2. The direct consequence is the stability across knot placements.

As expected, performances improve for zero-mean coefficients, while log variances are qualitatively similar to the mean-TV errors. Finally, the SP variants seems to be more robust than the PC case: ss shown by the boxplots, the smoother version displays very narrow quantile intervals, as opposed to the large tails of the PC. Figure 12.

5 Empirical application

We now assess the performance of the proposed approach on real data to demonstrate the usefulness of MTVP-VARs. We begin describing the macroeconomic dataset, then we present some full sample results for a small dimensional specification in Section 5.1. Finally, Section 5.2 presents a formal forecasting exercise using an out-of-sample expanding window scheme both in a small and in a high-dimensional setting, highlighting the advantages of the moderate specifications over the benchmark RW.

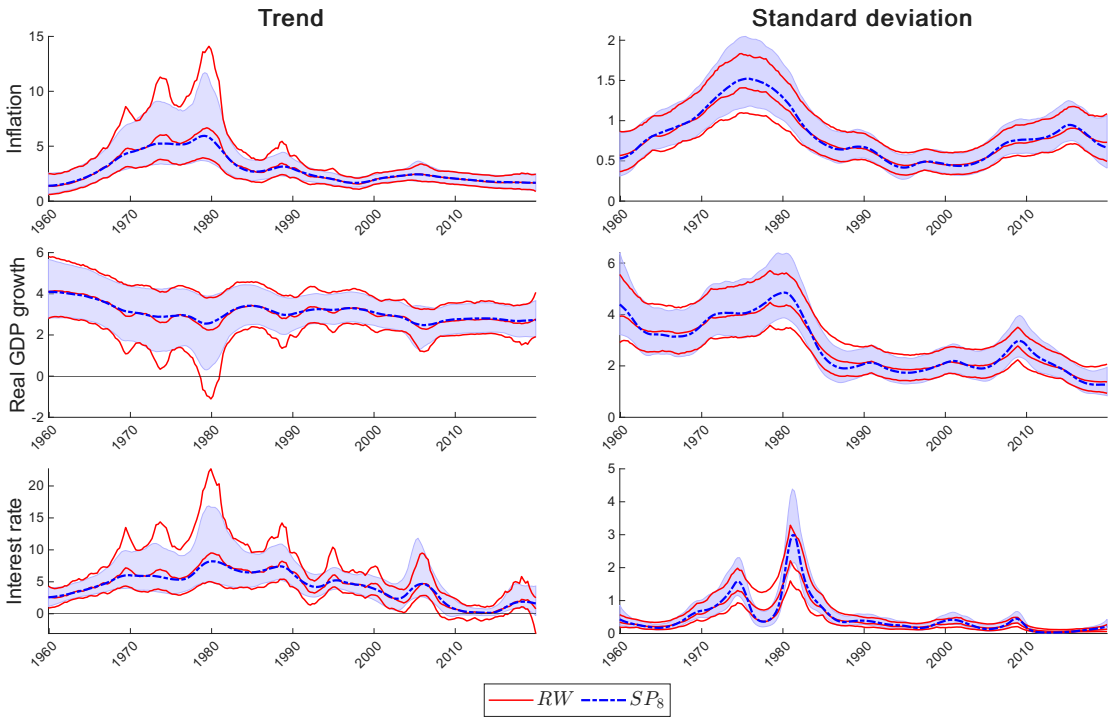
Data. The dataset comprises 14 U.S. quarterly series from 1959Q1 to 2019Q4, obtained from the FRED-QD database (Federal Reserve Bank of St. Louis). We include a range of standard macroeconomic indicators; integrated series are transformed into stationary ones via annualized growth-rate to ensure stable dynamics. Specifically, the small-scale model (N=3), used for both full sample and forecasting exercises, corresponds to the canonical tri-variate monetary-policy VAR, including real GDP growth, GDP deflator inflation, and the federal funds rate. The large-scale specification (N=14), used solely in the forecasting application, augments this set with consumption, investment, unemployment, labor-market indicators, and so on. The complete list of variables and transformations appears in Table 4, Appendix A.

5.1 Full-sample results

In this section, we summarize full-sample results of the moderate approach versus the RW model. For the sake of conciseness, we focus on the SP_8 case, the most parsimonious among the smooth moderate class. We defer the other moderate variants (including the PC cases) to the appendix.

An important aspect of this analysis is to assess whether MTVP-VARs produce economically meaningful quantities from an interpretational perspective, at least in the same way as the RW does. Therefore, we present the model estimated time varying trends (or infinite horizon forecasts) and standard deviations in Figure 5. Because trend estimation requires global stationarity, we incorporate an accept–reject step to discard explosive

Figure 5: **Trends and standard deviations**



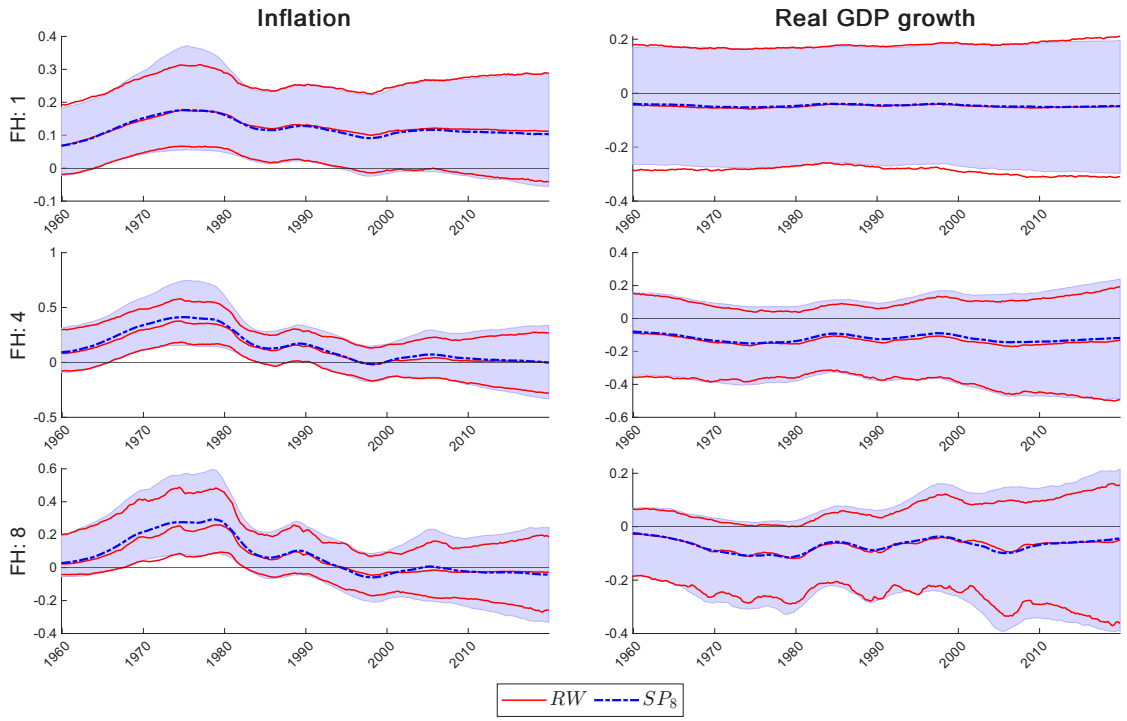
Note: The figure shows the posterior fitted trends and standard deviation for two specifications: benchmark RW (red lines), and moderate SP_8 (blue dashed lines). External red lines and blue shaded-areas delimit the 90% credible bands.

draws following Cogley and Sargent (2005).[9] As expected, SP_8 delivers smoother dynamics with respect to the jagged ones of the RW. Although RW behavior can be hypothetically improved by tightening the state-variance priors on TV coefficients, we exclude this possibility by default thanks to our hierarchical setup. In this way the optimal shrinkage is automatically determined. SP_8 captures the long-run dynamics of trends and standard deviations, while being less sensitive to local fluctuations. This leads to narrower credible bands and more precise estimates. Nevertheless, the moderate process is also capable of detecting relevant shifts in the parameter dynamics, as testified in the abrupt swings of interest-rate volatility in late 1970. This fact confirms that more parsimonious and smooth dynamics can handle strongly time-varying environments without paying excessive bias.

Finally, we investigate the structural Impulse Responses (IRFs) of Inflation and Real GDP growth to a monetary policy shock. Identification follows a standard Cholesky recursive ordering, as in Primiceri (2005). Time-varying IRFs are reported at horizons 1, 4, and 8 in Figure 6. Both IRFs reflects a stagflationary effect with inflation rising and

[9] Koop and Potter (2011) criticize this approach, noting that imposing inequality constraints in this way can perform poorly, especially in high dimensions. Our strategy mitigates this concern: by approximating dynamics with fewer parameters, locally nonstationary behavior is limited by construction. 1/3, 7.54 mins, 67.8% acc. rate 2/3, 6.5 mins, 83.8% acc. rate 3/3, 5.1 mins, 87.7% acc. rate.

Figure 6: **IRFs after a monetary policy shock**



Note: The figure shows the posterior IRFs of Inflation and Real GDP growth after a monetary policy shock for the two TVP VAR specifications: benchmark RW (red lines), and moderate SP_8 (blue dashed dotted-lines). Rows reflect different forecast horizons: 1, 4, 8. External red lines and blue shaded-areas delimit the 90% credible bands.

GDP growth declining. In deatil, the two models yield very similar structural outcomes: median IRFs are qualitatively identical, especially at shorter horizons. The moderate approach confirms the smoother behavior, which efficiently synthesizes the same dynamic of the competitor, but with far less parameters. Interestingly, SP_8 features wider credible bands, especially for the inflation response. The cause can be attributed to the smoother trend estimates, which, in turn, allow for more persistent cyclical fluctuations.

In summary, the full sample experiment reveals how the moderate specification effectively conveys the same dynamics of the RW in terms of TV trends, standard deviations and IRFs as well. However, the advantage, concerning the number of parameters involved, is huge with a proportion of 8 times parameters less with respect the literature benchmark.

5.2 Forecasting exercise

While full-sample analyses are essential to validate our framework's properties, assessing forecasting performance is equally crucial. Indeed, TVP-VARs are widely used for prediction in the literature. Accordingly, this section evaluates the MTVP approach in both the small (N=3) and high-dimensional (N=14) TVP-VAR via a recursive forecast-

ing exercise. We examine point and density forecast accuracy, comparing the benchmark model to all moderate variants previously introduced (*i.e.*, SP and PC).

The out-of-sample forecasting experiment follows an expanding-window scheme that begins with 1959Q1 – 1980Q1 and then adds one quarter at a time up to 2019Q4. At each recursion, we estimate the model using data only up to time t, denoted $\mathbf{y}_{1:t}$. We consider two forecast horizons, $h = 1$ and $h = 4$. Point-forecast accuracy for variable i is evaluated using the Mean Squared Forecasting Error (MSFE), defined as:

$$\text{MSFE}_{i,h}^{m} = \frac{1}{T-h-t_0} \sum_{t=t_0}^{T-h} \left(\hat{y}_{i,t+h}|\mathbf{y}_{1:t} - y_{i,t+h}\right)^2$$

where t_0 is the final observation of the first forecast window, $y_{j,t+h}$ the observed value, and $\hat{y}_{i,t+h}|\mathbf{y}_{1:t}$ is the mean of the posterior predictive mean under model m. Density forecasts, on the other hand, are compared via the Average Log Predictive Likelihood (ALPL):

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

where $p(\hat{y}_{i,t+h} = y_{i,t+h}|\mathbf{y}_{1:t})$ is the predictive likelihood of variable i at forecast horizon h for model m.[10] To formally compare model m against the RW, we follow Carriero et al. (2015) and report percentage gains both in terms of MSFE and ALPL:

$$100\left(1 - \text{MSFE}_{i,h}^{m}/\text{MSFE}_{i,h}^{RW}\right)$$
$$100\left(\text{ALPL}_{i,h}^{m} - \text{ALPL}_{i,h}^{RW}\right)$$

For both measures, values greater than zero indicate that model m outperforms the RW.

Tables 2 display the results for the small-scale setup. At forecast horizon $FH = 1$, the MSFE indicates that the SP specifications perform better than the PC alternatives almost uniformly. Moreover, all three SP specifications outperform the benchmark on average, with SP_8 being the preferred option for forecasting the Federal Funds rate, and SP_{24} for GDP growth (even though not statistical significant). The PC alternatives, on the other hand, show an overall deterioration relative to the RW, except for the point forecast improvement of PC_{24} for the Federal Funds rate. At $FH = 4$, the SP models confirm their positive performance, with SP_8 emerging as the best model on average. Interestingly, with the exception of the aforementioned SP_8, all models experience a marked deterioration in forecasting the interest rate, while performance improves for the GDP deflator, particularly for SP_{24}. Statistical significance is once again not achieved in all cases under examination. Turning to the ALPL results, the SP specifications again

[10] For further details we refer to Geweke and Amisano (2010); Chan et al. (2012); Chan and Eisenstat (2018).

outperform both the benchmark and the PC alternatives. In particular, SP_8 shows a clear advantage in density forecasting for inflation (a 2% gain at $FH = 4$) and the interest rate variable (a 9% and 7% gain at $FH = 1$ and $FH = 4$, respectively, with the former statistically significant at size 0.10). However, at $FH = 1$, all SP models deliver qualitatively similar results. Among the PC specifications, the more parameterized one (PC_8) stands out when considering the average performance.

Table 2: **Point and density forecast (N=3)**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
Rel. MSFE, FH: 1						
GDP deflator	**0.45**	0.21	−0.34	−1.25	−8.43	−11.21
Real GDP	−0.04	0.33	**1.31**	0.06	−0.77	−1.07
Federal funds rate	3.32	1.70	2.52	0.68	0.47	**3.51**
Average	**1.25**	0.75	1.16	−0.17	−2.91	−2.92
Rel. MSFE, FH: 4						
GDP deflator	2.13	6.38	**9.07**	−4.99	−8.78	−19.87
Real GDP	−0.04	0.94	−4.96	0.71	0.69	**1.05**
Federal funds rate	**4.68**	−1.77	−8.49	−3.95	−4.99	−5.61
Average	**2.26**	1.85	−1.46	−2.74	−4.36	−8.15
Rel. ALPL, FH: 1						
GDP deflator	**0.77**	0.45	−0.68	−0.09	-2.50^{l}	$-3.34^{\star}$
Real GDP	0.39	0.13	**0.44**	0.30	0.04	−0.48
Federal funds rate	**8.79***	7.33	7.86	3.24	−0.32	$-1.33^{\star}$
Average	**3.32**	2.63	2.54	1.15	−0.93	−1.71
Rel. ALPL, FH: 4						
GDP deflator	**2.26**	1.38	0.45	−0.58	$-4.28^{\star\star}$	$-8.51^{\star\star\star}$
Real GDP	0.06	0.34	−0.08	2.14	**2.28**	1.82
Federal funds rate	**6.75**	1.13	−0.95	0.28	−4.78	−5.37
Average	**3.02**	0.95	−0.19	0.61	−2.26	−4.02

Note: Point and density forecast N:3. Bold identifies the best performance across the considered specification - *, **, *** identify significance of the Diebold-Mariano test (Diebold and Mariano, 1995) for point forecast at 10%, 5% and 1%, respectively; The symbols *, **, and *** indicate the three significance levels for the Amisano–Giacomini test (Amisano and Giacomini, 2007) with normal weights for the center of the distribution.

Moving on to the large-scale setup, Table 3 reports both the relative MSFE and ALPL for some main macroeconomic variable of interest, namely GDP deflator, Real GDP growth, interest rate, industrial production and unemployment rate.[11] Similarly to

[11]The full tables are reported in Appendix E.

the small-scape case, the SP specification not only outperforms the competitor PC, but also the benchmark for both forecast horizons. Specifically, the best model on average appears to be SP_{24} with remarkable performance in forecasting the GDP deflator (though not statistically significant) and the unemployment rate (statistically significant at level 0.10 for $FH = 4$). In terms of density forecast, most of the conclusion previously reached for the MSFE remain valid for $FH = 1$, though with some noticeable modifications: all SP models yield similar results, with SP_8 showing superior performance for the federal funds rate and the unemployment (the latter statistically significant at 0.01), while SP_{24} for industrial production and unemployment as well (the latter statistically significant at 0.05 level). At $FH = 4$, SP_8 maintains an overall positive performance; the same goes for SP_{16} and SP_{24}, which reach an outstanding advantage in terms of unemployment rate (in magnitude). Interestingly, the PC is also able to recover positive performances, especially for GDP growth (statistically significant in case of PC_{16} and PC_{24}, industrial production (significant for PC_{24} and unemployment rate. This result may indicate that with larger dimensions and higher forecast horizons the leading pattern in parameters may be even captured via stepwise dynamics with few turning points.

To summarize the results, the MTVP seem to achieve indisputable advantages, especially in terms of density forecast, with respect the RW benchmark. This seems particularly evident for the SP cases which display superior performances in both the small-scale and in the large-scale setups for most of the macroeconomic variables, with particular reference to the federal funds rate, the industrial production and the unemployment rate. Most of the results achieved at $FH = 1$ tends to consolidate and magnify at $FH = 4$ (even though not always with statistical significance). Moreover, the performance of the PC also seems to improve at higher forecast horizon: this may suggest that the time variability of such quantities is much more modest.

Model comparison. Before concluding, we leverage the expanding window scheme to perform a model comparison exercise, based on the log predictive Marginal Likelihood (ML). Following Geweke and Amisano (2010), we derive the ML from the sequence of one-step-ahead joint log predictive densities up to time t. Specifically, the log predictive ML can be factored as

$$\log \text{ML}^m \approx \sum_{t=t_0}^{T} \log p(\hat{\mathbf{y}}_{t+1}^m = \mathbf{y}_{t+1} | \mathbf{y}_{1:t})$$

where $p(\hat{\mathbf{y}}_{t+h}^m = \mathbf{y}_{t+h} | \mathbf{y}_{1:t})$ is the one step ahead joint predictive density of model m.[12] The resulting log predictive Bayes Factor between model m and the RW is simply

$$\log \text{BF}^m = \log \text{ML}^m - \log \text{ML}^{RW}$$

[12]The approximation arises because the sum starts at $t_0 > 1$ rather than the first observation, thereby reducing the effect of the prior.

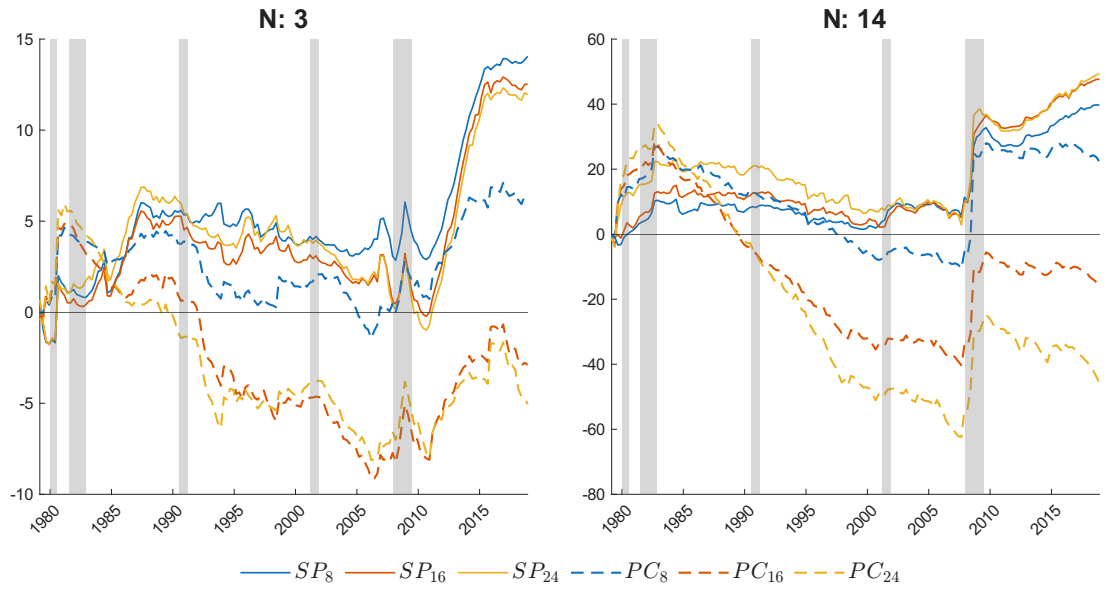
Table 3: **Point and density forecast (N=14), selected series**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
Rel. MSFE, FH: 1						
GDP deflator	2.16	4.65	**6.85**	−6.43	−6.61*	−6.09*
Real GDP	0.02	−0.26	**0.55**	−0.39	−0.52	−0.66
Federal funds rate	1.34	1.50	1.45	0.24	−0.76	**2.09**
Industrial production	−1.18	−1.21	**0.95**	−6.33	−9.72	−1.50
Unemployment rate	1.41	1.85	**5.21**	−2.18	−1.27	−1.19
Tot. Average	−0.38	0.17	**1.28**	−2.65	−3.08	−1.24
Rel. MSFE, FH: 4						
GDP deflator	7.20	8.28	**12.52**	−17.65	−18.64	−22.98
Real GDP	**−2.66**	−2.71	**−2.66**	−3.83	−4.45	−3.07
Federal funds rate	0.64	−3.35	1.47	4.27	−2.61	**4.91**
Industrial production	−0.72	−1.10	−2.01	−0.80	−0.30	**1.71**
Unemployment rate	4.22	10.60*	**14.37***	−1.48	−3.23	−7.96
Tot. Average	0.84	1.31	**2.18**	−1.23	−2.21	−1.42
Rel. ALPL, FH: 1						
GDP deflator	0.81	**1.42**	0.71	−1.24	−2.14	−3.07⋆
Real GDP	−0.07	−0.12	**−0.05**	−0.46	−0.49	−0.76
Federal funds rate	**5.67**	4.63	3.49	−0.56	−1.93	−7.63***
Industrial production	1.45	1.70	**2.69**	−0.5	−3.15**	−1.07**
Unemployment rate	3.35***	3.92**	**4.79****	−3.85	−8.41***	−15.53***
Tot. Average	1.96	2.12	**2.36**	−0.15	−1.64	−3.42
Rel. ALPL, FH: 4						
GDP deflator	**2.4** ⋆	2.27	1.07	−2.14	−5.56***	−11.14***
Real GDP	**1.09**	0.46	0.72	0.75*	0.12**	0.03***
Federal funds rate	**2.16**	−1.12	−2	1.25	−0.34	−1.15
Industrial production	6.02	4.2	**6.8**	5.65	5.52	6.73**
Unemployment rate	2.69	**17.94**	17.51	−1.86	17.22	10.98
Tot. Average	4.99	5.79	**6.17**	4.22	5.28	3.71

Note: Point and density forecast N:14 for selected series. Bold identifies the best performance across the considered specifications. The entry "Tot. Average" refers to the average computed over all 14 variables. - *, **, *** identify significance of the Diebold-Mariano test (Diebold and Mariano, 1995) for point forecast at 10%, 5% and 1%, respectively; The symbols ⋆, ⋆⋆, and ⋆⋆⋆ indicate the three significance levels for the Amisano–Giacomini test (Amisano and Giacomini, 2007) with normal weights for the center of the distribution.

Figure 7 reports the log predictive Bayes factors for each model in both the small and high-dimensional settings. To provide context to the plot, a $\log \mathrm{BF}^m$ greater than $\approx$

Figure 7: **Cumulative log predictive Bayes Factor**



Note: The figure shows the evolution of the log predictive BF for all moderate specifications relative to the *RW* for the small dimensional case (first panel) and the high dimensional case (second panel).

2.5 indicates a strong evidence in favor of model m, whereas $\log \text{BF}^m > 7$ indicates a decisive evidence in favor of model m.[13] In the small VAR system, all SP models exhibit very similar dynamics, with significant results emerging from 1985 onward and becoming particularly pronounced toward the end of the sample ($\log BF_m > 10$). Among the PC specifications, only the more parameterized one, namely PC_8m shows a positive and significant pattern.

The same conclusions, though magnified, apply to the large VAR system, where $\log BF_m > 20$ values appear from the beginning of the sample and exceed 40 toward the end for all SP models and PC_8 (although the latter performs slightly worse than the former).

6 Conclusion

This article has introduced a moderate time variation model based on spline approximations: in doing so, we have proposed an alternative to the common-use TVP dynamics, which is flexible enough to be applied to any time variational pattern and parsimonious enough to be used in large-scale models.

Using data from the US economy, we have shown that our approach is at least as competitive as the standard TV- VAR with RW_T in terms of point forecasting accuracy, but has a clear advantage in terms of density forecasting and computational time, especially

[13]The threshold values are mutated from Jeffreys (1998), but in natural-log form.

in high dimensions. In the full sample exercise, similar results are achieved in terms of posterior medians trend and volatilities as well as impulse responses.

The proposed approach is not without limitations: the choice of the number of knots is crucial and can be a source of misspecification. At the same time, the consequences of the proposed method on structural analysis are still unclear.

Finally, recall that the MTVP method here proposed, can be equally applied to any time varying dynamics, both in conditional mean and volatility. In this regard, the treatment of the SV component deserves some considerations: in accordance with many applied works we have followed Primiceri (2005) introducing a variable-order dependence. Despite this has proven to be a minor problem, potential attenuations or solutions such as the order invariance methods by Kastner (2019) or Chan et al. (2024) can be easily incorporated in the MTVP. Similarly, in our applications we postulate diagonal covariances matrices for the state transition equations, feature that could be generalized by incorporating a factor structure similar to that used by Korobilis (2022) for VAR errors.

References

Amir-Ahmadi, P., C. Matthes, and M.-C. Wang (2020): "Choosing prior hyperparameters: With applications to time-varying parameter models," *Journal of Business & Economic Statistics*, 38, 124–136.

Amisano, G. and R. Giacomini (2007): "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business & Economic Statistics*, 25, 177–190.

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

Barnichon, R. and C. Brownlees (2019): "Impulse Response Estimation by Smooth Local Projections," *The Review of Economics and Statistics*, 101, 522–530.

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

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

Benati, L. and P. Surico (2008): "Evolving US monetary policy and the decline of inflation predictability," *Journal of the European Economic Association*, 6, 634–646.

Boeck, M. and L. Mori (2025): "Has globalization changed the international transmission of US monetary policy?" *Journal of International Economics*, 157, 104139.

Carriero, A., T. E. Clark, and M. Marcellino (2015): "Bayesian VARs: specification choices and forecast accuracy," *Journal of Applied Econometrics*, 30, 46–73.

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

——— (2023): "Large hybrid time-varying parameter VARs," *Journal of Business & Economic Statistics*, 41, 890–905.

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

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

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

Chan, J. C., G. Koop, and X. Yu (2024): "Large order-invariant Bayesian VARs with stochastic volatility," *Journal of Business & Economic Statistics*, 42, 825–837.

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

Clark, T. E. and F. Ravazzolo (2015): "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30, 551–575.

Cogley, T. and T. J. Sargent (2001): "Evolving post-world war II US inflation dynamics," *NBER macroeconomics annual*, 16, 331–373.

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

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

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13.

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

Frühwirth-Schnatter, S. (2004): *Efficient Bayesian parameter estimation for state space models based on reparameterizations*, United Kingdom: Cambridge University Press, 123 – 151.

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

Geweke, J. and G. Amisano (2010): "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26, 216–230.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014): "Short-term inflation projections: A Bayesian vector autoregressive approach," *International journal of forecasting*, 30, 635–644.

Giraitis, L., G. Kapetanios, and T. Yates (2018): "Inference on multivariate heteroscedastic time varying random coefficient models," *Journal of Time Series Analysis*, 39, 129–149.

Grant, A. L. and J. C. Chan (2017): "Reconciling output gaps: Unobserved components model and Hodrick–Prescott filter," *Journal of Economic Dynamics and Control*, 75, 114–121.

Hodrick, R. J. and E. C. Prescott (1997): "Postwar US business cycles: an empirical investigation," *Journal of Money, credit, and Banking*, 1–16.

Jeffreys, H. (1998): *The theory of probability*, OuP Oxford.

Kapetanios, G., M. Marcellino, and F. Venditti (2019): "Large time-varying parameter VARs: A nonparametric approach," *Journal of Applied Econometrics*, 34, 1027–1049.

Kastner, G. (2019): "Sparse Bayesian time-varying covariance estimation in many dimensions," *Journal of Econometrics*, 210, 98–115.

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

Koop, G. and D. Korobilis (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177, 185–198, dynamic Econometric Modeling and Forecasting.

Koop, G. and S. M. Potter (2011): "Time varying VARs with inequality restrictions," *Journal of Economic Dynamics and Control*, 35, 1126–1138.

Korobilis, D. (2021): "High-Dimensional Macroeconomic Forecasting Using Message Passing Algorithms," *Journal of Business & Economic Statistics*, 39, 493–504.

——— (2022): "A new algorithm for structural restrictions in Bayesian vector autoregressions," *European Economic Review*, 148, 104241.

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

Tanaka, M. (2020): "Bayesian inference of local projections with roughness penalty priors," *Computational Economics*, 55, 629–651.

Appendix

A Description of the dataset

Table 4: **Dataset**

Variables	FRED-QD code	Transformation	Small (N=3)	Large (N=14)
Real GDP	GDPC1	$400 \cdot \Delta \log$	x	x
GDP deflator	GDPDEF	$400 \cdot \Delta \log$	x	x
Federal funds rate	FEDFUNDS	Raw	x	x
Real income	DPIC96	$400 \cdot \Delta \log$		x
Real consumption	PCEC96	$400 \cdot \Delta \log$		x
Hours worked	CES0600000007	Raw		x
Hourly earnings	CES0600000008	$400 \cdot \Delta \log$		x
Capacity utilization	CUMFNS	Raw		x
Nonfarm payrolls	PAYEMS	$400 \cdot \Delta \log$		x
Industrial production	INDPRO	$400 \cdot \Delta \log$		x
Unemployment rate	UNRATE	Raw		x
Real M2 money stock	M2REAL	$400 \cdot \Delta \log$		x
10 year yield	GS10	Raw		x
BAA spread	BAAFM	Raw		x

B Estimation details

In this Appendix, we outline the estimation details for the MTVP VAR in Equation (10) with the corresponding prior framework.

- In order to sample from each $p(\boldsymbol{\theta}_i|\bullet)$, we first stack Equation (10) over T:

$$\mathbf{y}_i = \mathbf{Z}_i(\mathbf{W}_i \otimes \mathbf{I}_{K_i})\boldsymbol{\theta}_i = \bar{\mathbf{Z}}_i\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$

  where $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,T}]'$, $\mathbf{Z}_i = \text{diag}(\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,T})$, $\boldsymbol{\epsilon}_i = [\epsilon_{i,1}, \ldots, \epsilon_{i,T}]'$, and $\boldsymbol{\Sigma}_i = \text{diag}[\exp(h_{i,1}), \ldots, \exp(h_{i,T})]$. Next, let $\mathbf{H}_{\theta_i}$ denote a first difference matrix:

$$\mathbf{H}_{\theta_i} = \begin{bmatrix} \mathbf{I}_{K_i} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_{K_i} & \mathbf{I}_{K_i} & & \\ \vdots & & \ddots & \\ \mathbf{0} & & -\mathbf{I}_{K_i} & \mathbf{I}_{K_i} \end{bmatrix}$$

  Thus, we can rewrite compactly the prior for $\boldsymbol{\theta}_i$ as:

$$\mathbf{H}_{\theta_i}\boldsymbol{\theta}_i = \boldsymbol{\xi}_i, \quad \boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\theta_i})$$

  where $\mathbf{V}_{\theta_i} = \text{diag}(\mathbf{V}_{\theta_{i,1}}, \mathbf{I}_{R-1} \otimes \mathbf{V}_{\theta_i})$. Using standard linear regression results, it can be shown that

$$p(\boldsymbol{\theta}_i|\bullet) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_i, \mathbf{B}_{\theta_i}^{-1})$$

  where

$$\mathbf{B}_{\theta_i} = \mathbf{H}_{\theta_i}'\mathbf{V}_{\theta_i}\mathbf{H}_{\theta_i} + \bar{\mathbf{Z}}_i'\boldsymbol{\Sigma}_i\bar{\mathbf{Z}}_i, \quad \hat{\boldsymbol{\theta}}_i = \mathbf{B}_{\theta_i}^{-1}(\bar{\mathbf{Z}}_i'\boldsymbol{\Sigma}_i\mathbf{y}_i)$$

  We obtain draws from this high dimensional full conditional distribution using the efficient sampler of Chan and Jeliazkov (2009).

- In order to sample from each $p(\mathbf{g}_i|\bullet)$, we start by stacking over T the error term of Equation (10):

$$\boldsymbol{\epsilon}_i = \exp(\mathbf{W}_i\mathbf{g}_i/2)\mathbf{u}_i, \quad \mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_T)$$

  By squaring and subsequently taking logarithms of each element of $\boldsymbol{\epsilon}_i$ we obtain a linear state space model:

$$\tilde{\boldsymbol{\epsilon}}_i = \mathbf{W}_i\mathbf{g}_i + \tilde{\mathbf{u}}_i$$

  where $\tilde{\boldsymbol{\epsilon}}_i = \log \boldsymbol{\epsilon}_i^2$, and each element of $\tilde{\mathbf{u}}_i$ follows a log Chi-Square distribution with one degree of freedom. Following Kim et al. (1998), we approximate $\tilde{\mathbf{u}}_i$ with

a mixture of 7 Gaussian distributions

$$\tilde{\mathbf{u}}^* \sim \pi_1 \mathcal{N}(\mathbf{a}_1, \mathbf{C}_1) + \cdots + \pi_7 \mathcal{N}(\mathbf{a}_7, \mathbf{C}_7)$$

where $\mathbf{a}_j$, $\mathbf{C}_j$ are predetermined mixtures mean and variance components and π_j the mixture weight, , with $j = 1, \ldots, 7$. Given a draw of the component density indicators $\mathbf{s} = [s_1, \ \ldots, \ s_T]$,

$$\begin{aligned} \tilde{\mathbf{u}}^* | s_j = j &\sim \mathcal{N}(\mathbf{a}_j, \mathbf{C}_j) \\ \mathrm{P}(s_j = j) &= \pi_j \end{aligned}$$

the model can be recast in linear and Gaussian form:

$$\tilde{\boldsymbol{\epsilon}}_i | s_j = \mathbf{W}_i \mathbf{g}_i + \mathbf{a}_j + \tilde{\mathbf{u}}^* | s_j, \quad \mathbf{u}^* | s_j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_j)$$

Let $\mathbf{H}_{g_i}$ be another first difference matrix:

$$\mathbf{H}_{g_i} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & & \\ \vdots & & \ddots & \\ 0 & & -1 & 1 \end{bmatrix}$$

Then, the prior for $\mathbf{g}_i$ can be written compactly as

$$\mathbf{H}_{g_i} \mathbf{g}_i = \boldsymbol{\nu}_i, \quad \boldsymbol{\nu}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{g_i})$$

where $\mathbf{V}_{g_i} = \mathrm{diag}(v_{g_{i,1}}, v_{g_i} \mathbf{I}_{R-1})$. Using standard linear regression results, we have:

$$p(\mathbf{g}_i | \bullet) \sim \mathcal{N}(\hat{\mathbf{g}}_i, \mathbf{B}_{g_i}^{-1})$$

where

$$\mathbf{B}_{g_i} = \mathbf{H}_{g_i}' \mathbf{V}_{g_i} \mathbf{H}_{g_i} + \mathbf{W}_i' \mathbf{C}_i \mathbf{W}_i, \quad \hat{\mathbf{g}}_i = \mathbf{B}_{g_i}^{-1} [\mathbf{W}_i' \mathbf{C}_i (\mathbf{y}_i - \mathbf{a}_s)]$$

Again, draws from this full conditional distribution are obtained via the precision sampler of Chan and Jeliazkov (2009).

- All the conditional posteriors for $p(v_{i,j}|\bullet)$, and $p(s_i|\bullet)$ are Inverse Gamma:

$$v_{i,j} | \bullet \sim \mathcal{IG}\left(\tau_1 + \frac{R}{2}, \tau_2 + \frac{\sum_t \theta_{i,j,t}^2}{2}\right)$$

$$s_i|\bullet \sim \mathcal{IG}\left(\tau_1 + \frac{R}{2}, \tau_2 + \frac{\sum_t g_{i,t}^2}{2}\right)$$

- In order to sample from $p(\xi_1|\bullet)$ and $p(\xi_2|\bullet)$, we first note that these parameters only appear in their respective priors. By combining equations (**??**) and (**??**), the conditional posterior for $\bar{\xi_1}$ can be written as

$$\begin{aligned} p(\xi_1|\bullet) \propto & \xi_i^{2N\tau_1} \exp\left[-\xi_1(\tau_1+1)\Big(\sum_i s_i^{-1} + \sum_i v_{i,1}^{-1}\Big)\right] \\ & \times \xi_1^{-(\nu+1)} \exp\Big(-(\nu+1)\bar{\xi_1}\xi_1^{-1}\Big) \\ = & \xi_1^{2N\tau_1-\nu-1} \exp\Big[-\xi_1(\tau_1+1)\Big(\sum_i s_i^{-1} + \sum_i v_{i,1}^{-1}\Big) - \xi_1^{-1}(\nu+1)\bar{\xi_1}\Big] \end{aligned}$$

  which is the kernel of a Generalized Inverse Gaussian distribution:

$$\xi_1|\bullet \sim \mathcal{GIG}\left(2N\tau_1 - \nu, 2(\tau_1+1)\Big(\sum_i s_i^{-1} + \sum_i v_{i,1}^{-1}\Big), 2(\nu+1)\bar{\xi_1}\right)$$

  The conditional posterior of ξ_2 is derived along the same line:

$$\xi_2|\bullet \sim \mathcal{GIG}\left((N^2P - N)\tau_1 - \nu, 2(\tau_1+1)\sum_i\sum_{j=2}^{K_i} v_{i,j}^{-1}, 2(\nu+1)\bar{\xi_2}\right)$$

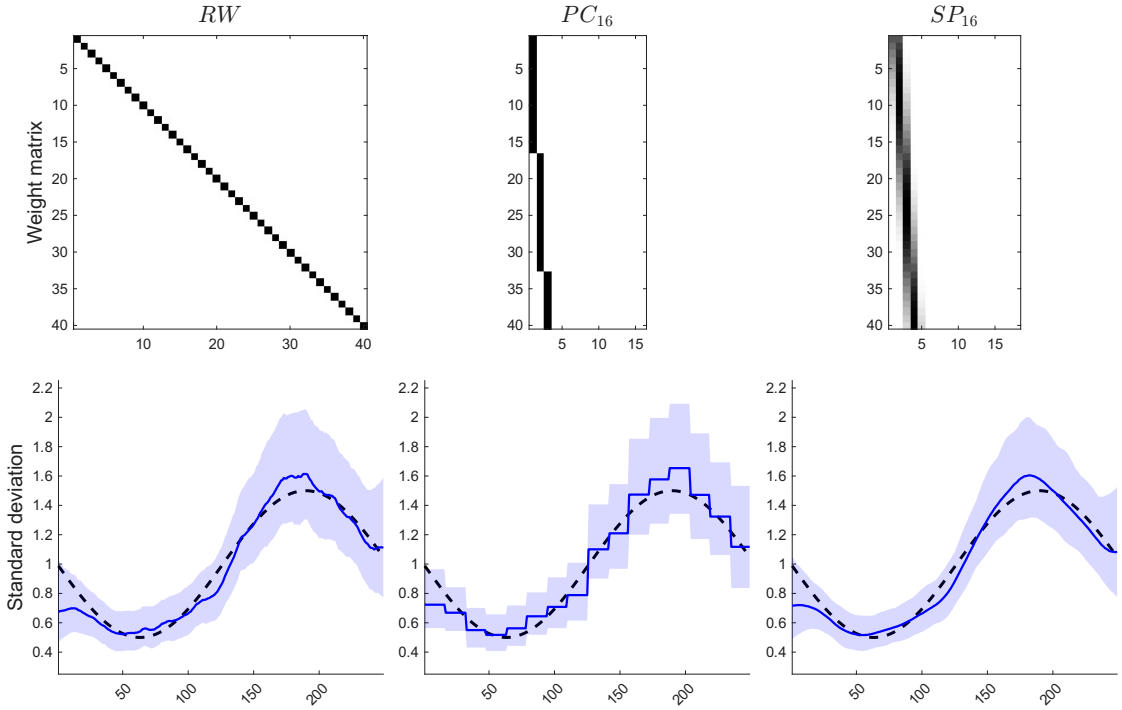
- The conditional posteriors of λ_1 and λ_2 are also Generalized Inverse Gaussians:

$$\lambda_1|\bullet \sim \mathcal{GIG}()$$

Estimation proceeds via a Gibbs sampler which loops through the full conditional distributions. Notice that both $\mathbf{H}_{\theta_i}$ and $\mathbf{H}_{g_i}$ are banded, so sparse matrix libraries can be conveniently employed to speed up the algorithm.

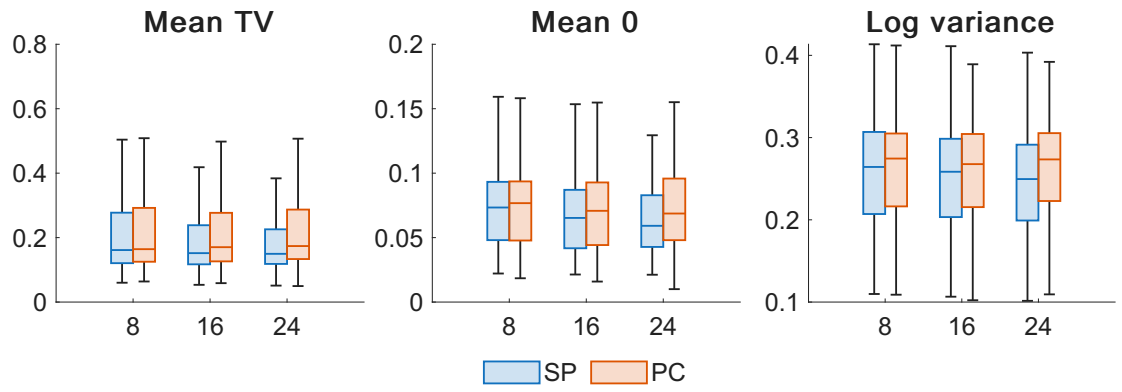
C Additional results

Figure 8: **The effect of different weighting schemes** ($K = 16$)



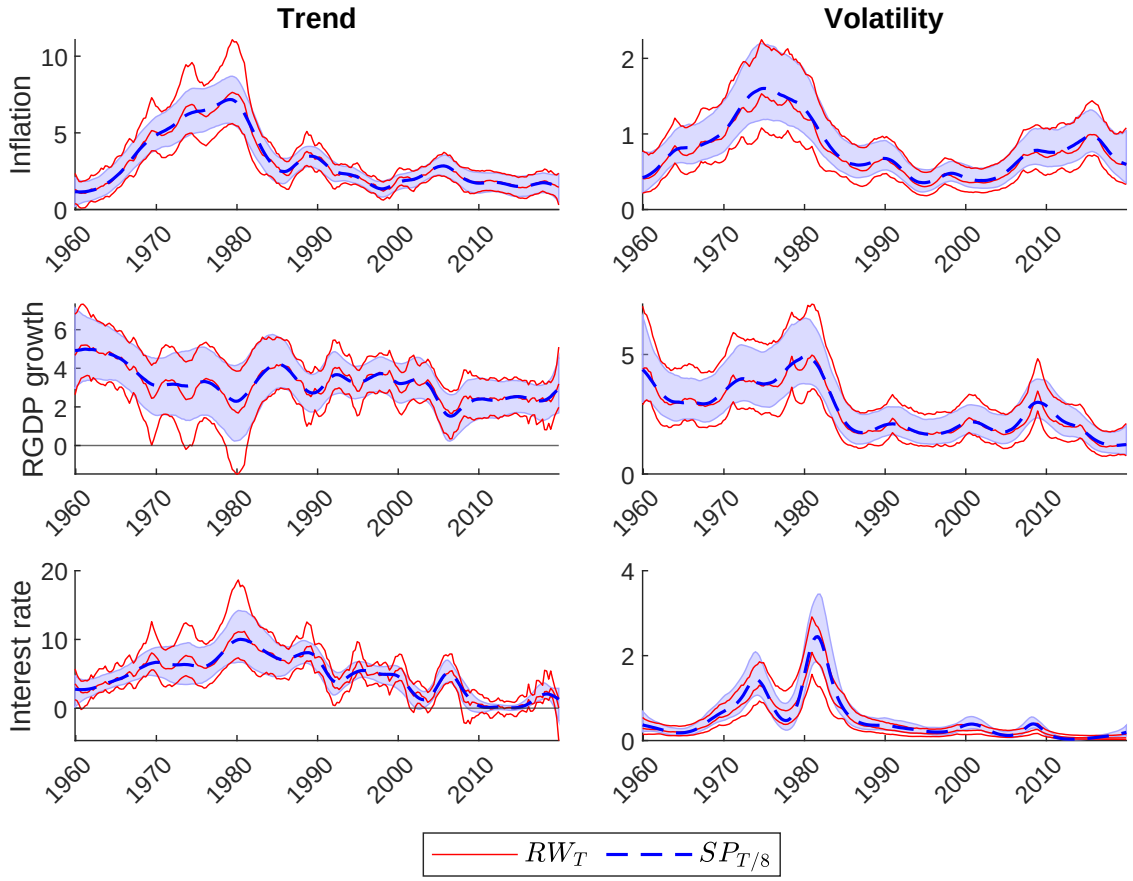
Note: The figure illustrates the effect of different moderate specifications ($K = 16$) on the estimation of the SV of the residuals for a generic variable in a single realization of a simulated TVP-VAR, as described in Section 4. The sample size is $T = 250$. *First row*: first 40 rows of the three weight matrices considered. The number of columns varies, being equal to $\min(40, R)$, which yields 40 for the RW case, 16 for the PC_8, and 18 for the SP_{16}. The shading ranges from 0 (white) to 1 (black). *Second row*: black dashed lines denote the true process, solid blue lines represent the posterior median, and shaded blue areas indicate the 90% credible intervals.

Figure 9: **Results of the knot positioning exercise**



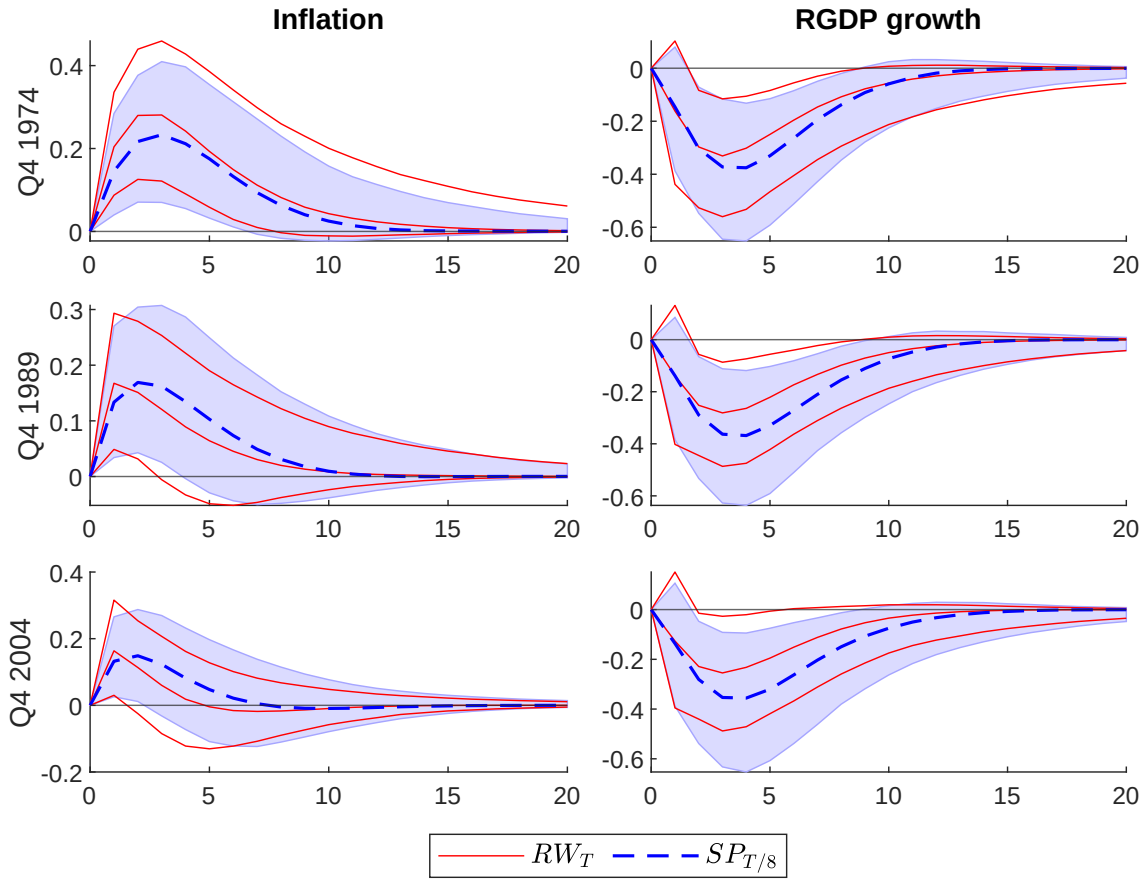
Note: .....

Figure 10: **Trends and volatiltities**



Note: Note: The figure shows the posterior median trends and volatilities $\exp(h_t/2)$ for the two specifications: benchmark RW_T (red lines), and moderate $SP_{T/8}$ (blue dashed lines). External red lines and blue shaded-areas delimit the 90% credible bands.

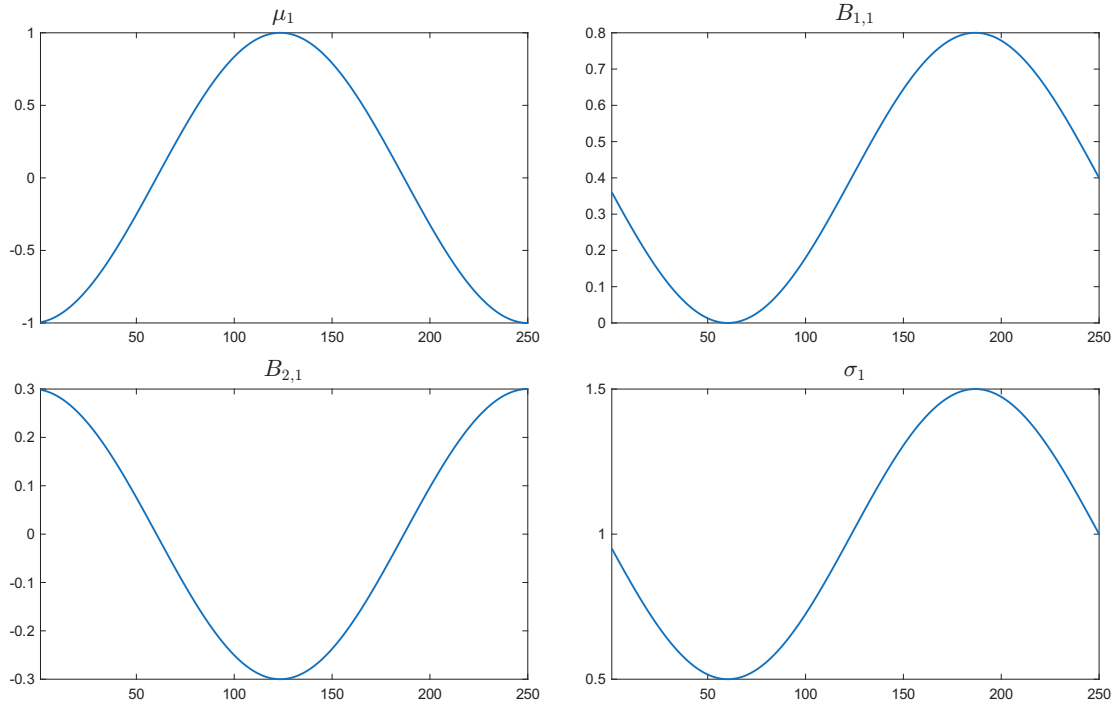
Figure 11: **IRFs after a monetary policy shock**



Note: Note: The figure shows the posterior median impulse responses of inflation and real GDP growth after a monetary policy shock identified with a Cholesky-type recursive ordering for the two TVP VAR specifications: benchmark RW_T (red line), and moderate $SP_{T/8}$ (blue dashed dotted-lines). External red lines and blue shaded-areas delimit the 90% credible bands.

D Simulation details: Data Generating Process

Figure 12: **True simulation parameters**



Note: The figure displays four parameters of the DGP shown below. Panel 1: intercept, panel 2: own-lag coefficient, panel 3: cross-lag coefficient, and panel 4: standard deviation.

The simulation design follows the tri-variate VAR by Amir-Ahmadi et al. (2020) with some minor changes introduced to align TVP dynamics for block of coefficients. The corresponding DGP is reported below:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \end{bmatrix} + \begin{bmatrix} \beta_{11,t} & 0 & 0 \\ \beta_{21,t} & \beta_{22,t} & 0 \\ \beta_{31,t} & \beta_{32,t} & \beta_{33,t} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix} \qquad t = 1, \dots, T,$$

where

$$\begin{aligned} \mu_{i,t} &= \cos(x_t) \\ \beta_{ii,t} &= s_i \cdot 0.4 \sin(x_t) \\ \beta_{ij,t} &= s_i \cdot 0.3 \cos(x_t), \quad i > j \\ s_i &\sim \text{Rademacher} \\ \epsilon_{i,t} &\sim \mathcal{N}(0, \sigma_{i,t}^2) \\ \sigma_{i,t} &= 1 + \frac{\sin(x_t)}{2} \end{aligned}$$

with x_t as a vector of T evenly spaced points in $[-\pi, \pi]$.

E Additional expanding-window results

Table 5: **Point forecast - Large-scale VAR**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
			FH: 1			
GDP deflator	2.16	4.65	**6.85**	−6.43	−6.61*	−6.09*
Real GDP	0.02	−0.26	**0.55**	−0.39	−0.52	−0.66
Federal funds rate	1.34	1.50	1.45	0.24	−0.76	**2.09**
Real income	−1.34**	−1.41*	**−1.24***	−2.90**	−3.17**	−2.97**
Real consumption	−1.18**	−1.33***	**−0.12**	−2.99*	−3.06	−3.58*
Hours worked	1.35	2.60	4.34	2.71	3.05	**7.70**
Hourly earnings	0.46	1.05	**2.73**	−5.25**	−6.69*	−6.11**
Capacity utilization	**−1.67**	−3.25	−2.41	−2.47	−5.02	−2.01
Nonfarm payrolls	−1.38	−0.88	**1.04**	−5.13	−7.19	−1.87
Industrial production	−1.18	−1.21	**0.95**	−6.33	−9.72	−1.50
Unemployment rate	1.41	1.85	**5.21**	−2.18	−1.27	−1.19
Real M2 money stock	**−0.36**	−0.67	−1.80	−2.4 *5	−1.75	−1.87
10 year yield	0.27	1.97	1.57	0.45	**5.06**	2.59
BAA spread	−5.15	−2.28	**−1.22**	−3.92	−5.50	−1.93
Average	−0.38	0.17	**1.28**	−2.65	−3.08	−1.24
			FH: 4			
GDP deflator	7.20	8.28	**12.52**	−17.65	−18.64	−22.98
Real GDP	**−2.66**	−2.71	**−2.66**	−3.83	−4.45	−3.07
Federal funds rate	0.64	−3.35	1.47	**4.27**	−2.61	4.91
Real income	−0.10	**0.14**	−0.28	−0.43	−0.03	0.12
Real consumption	−0.94	**−0.37**	−1.54	−4.53*	−4.55	−3.74
Hours worked	3.32	4.90	5.58	20.09***	**22.35*****	21.32**
Hourly earnings	2.42	4.96	**10.06**	−18.23*	−18.33**	−16.91*
Capacity utilization	−1.71	−2.39	−2.13	**−0.37**	−3.35	−1.44
Nonfarm payrolls	−1.46	−1.03	**−0.53**	−2.06	−3.99	−4.72
Industrial production	−0.72	−1.10	−2.01	−0.80	−0.30	**1.71**
Unemployment rate	4.22	10.60*	**14.37***	−1.48	−3.23	−7.96
Real M2 money stock	**0.42**	−0.16	−1.27	−5.04	−2.34	−2.37
10 year yield	6.32**	7.10**	7.99*	4.35	**13.22**	4.97
BAA spread	−5.21	−6.52	−11.02	8.42	−4.61	**10.26**
Average	0.84	1.31	**2.18**	−1.23	−2.21	−1.42

Note: Relative MSFE for point forecasts, N:14 - full results. Bold identifies the best performance across the considered specifications. - *, **, *** identify significance of the Diebold-Mariano test (Diebold and Mariano, 1995) for point forecast at 10%, 5% and 1%, respectively.

Table 6: **Density forecast - Large-scale VAR**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
			FH: 1			
GDP deflator	0.81	**1.42**	0.71	−1.24	−2.14	−3.07*
Real GDP	−0.07	−0.12	**−0.05**	−0.46	−0.49	−0.76
Federal funds rate	**5.67**	4.63	3.49	−0.56	−1.93	−7.63***
Real income	0.71	0.46	0.63	**2**	1.93	1.91
Real consumption	−0.49	−0.30	**0.04**	−1.19**	−1 **	−1.56***
Hours worked	2.62	3.75	**4.57***	0.95	−3.21	−6.91**
Hourly earnings	0.45	0.75	**1.43**	−2.73**	−3.66***	−5.44***
Capacity utilization	**1.52****	0.98	0.79	0.9	−1.88	−2.69**
Nonfarm payrolls	1.85***	2.15**	**2.98****	0.26	−1.81	−2.15**
Industrial production	1.45	1.70	**2.69**	−0.5	−3.15**	−1.07**
Unemployment rate	3.35***	3.92**	**4.79****	−3.85	−8.41***	−15.53***
Real M2 money stock	2.03	1.02	**2.41**	1.95	2.18	1.83
10 year yield	2.92	**3.62**	3.32	2.18	2.81	1.61*
BAA spread	4.58	5.65	5.22	0.16*	−2.13**	−6.4 **
Average	1.96	2.12	2.36	−0.15	−1.64	−3.42
			FH: 4			
GDP deflator	**2.4** *	2.27	1.07	−2.14	−5.56***	−11.14***
Real GDP	**1.09**	0.46	0.72	0.75*	0.12**	0.03***
Federal funds rate	**2.16**	−1.12	−2	1.25	−0.34	−1.15
Real income	0.58	0.46	0.53	**1.96**	1.88	1.7
Real consumption	**−1.16**	−1.35*	−2.09**	−2.43***	−2.38***	−3.23***
Hours worked	3.53*	7.28*	**10.17****	6.96**	2.27	−4.98*
Hourly earnings	1.5	2.27	**2.74**	−5.46***	−7.61***	−10.37***
Capacity utilization	13.97	12.04	16.89	19.5	23.91*	**27.79*****
Nonfarm payrolls	**11.47**	7.86	10.55	7.49	10.58*	9.56***
Industrial production	6.02	4.2	**6.8**	5.65	5.52	6.73***
Unemployment rate	2.69	**17.9**	17.51	−1.86	17.22	10.98
Real M2 money stock	−1.36*	−1.83**	−1.49	−2.15*	**−1.18**	−2.05**
10 year yield	7.31	7.56	9.22	8.8	**10.72**	5.26
BAA spread	19.59	**22.99**	15.72	20.74	18.73	22.78
Average	4.99	5.79	**6.17**	4.22	5.28	3.71

Note: Relative ALPL for point forecasts, N:14 - full results. Bold identifies the best performance across the considered specifications. - *, **, *** identify significance of the Amisano-Giacomini test (Amisano and Giacomini, 2007) for density forecast at level 10%, 5% and 1%, respectively, using normal distribution weights for the distribution center.

	T/4	T/8	T/16	T/32
GDPDEF	**4.72**	**4.01**	**0.66**	-1.72
GDPC1	**1.98**	**1.85**	-0.22	-0.67
FEDFUNDS	**9.15**	**8.37**	**9.27**	**10.32**
DPIC96	**3.88*****	**5.31*****	**6.35*****	**5.5****
PCECC96	**3.11**	**0.37**	-6.13	-6.63
CES0600000007	-3.55	**14.21*****	**16.17*****	**15.65****
CES0600000008	**2.79**	-4.68	-7.91	-10.9*
CUMFNS	**1.9**	**4.89**	**0.64**	-0.79
PAYEMS	**1.47**	**2.76**	-4.07	-4.81
INDPRO	**2.76**	**3.99**	-2.03	-3.03
UNRATE	-4.09	**15.57*****	**14.12***	**13.74***
M2REAL	-0.12	-1.19	-3.43	-3.88
GS10	-1.84	**8.2***	**8.23***	**8.8***
BAAFFM	**4.6**	**5.49**	**8.35**	**9.38**

Table 7: Relative RMSFE with respect to the random-walk model at forecast horizon $h = 1$, bold indicates a superior performance. *, **, *** indicate significance for the Diebold-Mariano test at size 0.10, 0.05 and 0.01, respectively. **Note: here the hierarchical prior on state variances parameters is replaced with fixed-value hyperparameters, uninformative ones.**

	T/4	T/8	T/16	T/32
GDPDEF	**2.44**	-7.71	-14.3*	-20.98*
GDPC1	**0.29**	**3.08**	**2.04**	**1.6**
FEDFUNDS	**0.6**	-4.42	**2.06**	**3.42**
DPIC96	**4.76*****	**6.3****	**6.14****	**5.41***
PCECC96	-1.74	-2.46	-3.14	-2.99
CES0600000007	**9.62**	**33.31**	**43.03**	**43.29**
CES0600000008	-0.33*	-17.31***	-21.97***	-25.03***
CUMFNS	-1.07	**3.59**	**4.69**	**3.22***
PAYEMS	**0.88**	**1.52**	**1.34**	**1.11**
INDPRO	**3.94**	**7.12**	**10.28**	**9.7**
UNRATE	-5.68	**11.8****	-0.1	-1.63
M2REAL	**5.65***	**4.54**	**7.5**	**5.55**
GS10	-5.65	**6.57**	**7.41**	**5.4**
BAAFFM	-0.04	-5.69	**9.29**	**13.79**

Table 8: Relative RMSFE with respect to the random-walk model at forecast horizon $h = 4$, bold indicates a superior performance. *, **, *** indicate significance for the Diebold-Mariano test at size 0.10, 0.05 and 0.01, respectively. **Note: here the hierarchical prior on state variances parameters is replaced with fixed-value hyperparameters, uninformative ones.**

	T/4	T/8	T/16	T/32
GDPDEF	**18.04**	**17.76**	**14.18**	**10.55**
GDPC1	**1.9**	**1**	-3.12	-4.32
FEDFUNDS	**73.13**	**76.75**	**74.29**	**72.86**
DPIC96	**6.38**	**8.17**	**2.64**	**2.6**
PCECC96	**3.01**	-0.7	-9.2	-9.77
CES0600000007	**92.83**	**118.33**	**123.91**	**123.33**
CES0600000008	**14.85**	**10.3**	**4.36**	**0.19**
CUMFNS	**46.17**	**49.85**	**43.15**	**40.7**
PAYEMS	**39.17**	**41.91**	**33.46**	**31.39**
INDPRO	**35.95**	**36.83**	**27.23**	**24.91**
UNRATE	**128.23**	**151.09**	**149.99**	**148.88**
M2REAL	**10.95**	**9.02**	-0.85	-2.83
GS10	**75.19**	**83.93**	**82.34**	**82.99**
BAAFFM	**89.59**	**97.11**	**92.91**	**92.96**

Table 9: Average Log Predictive Likelihood (ALPL) at forecast horizon $h = 1$, bold indicates a superior performance with respect to the random-walk model. **Note: here the hierarchical prior on state variances parameters is replaced with fixed-value hyperparameters, uninformative ones.**

	T/4	T/8	T/16	T/32
GDPDEF	**56.36**	**59.27**	**54.53**	**49.32**
GDPC1	**45.42**	**47.3**	**43.62**	**42.75**
FEDFUNDS	**59.99**	**59.52**	**55.54**	**61.49**
DPIC96	**35.44**	**37.29**	**34.04**	**34.37**
PCECC96	**46.15**	**45.08**	**42.86**	**43.4**
CES0600000007	**125.29**	**179.66**	**187.88**	**187.53**
CES0600000008	**56.36**	**57.21**	**50.9**	**43.05**
CUMFNS	**68.81**	**85.67**	**63.8**	**61.59**
PAYEMS	**67.72**	**80.55**	**70.08**	**68.99**
INDPRO	**76.39**	**86.71**	**82.37**	**79.28**
UNRATE	**112.17**	**139.85**	**84.67**	**89.18**
M2REAL	**53.72**	**54.38**	**50.82**	**48.71**
GS10	**84.41**	**99.26**	**105.86**	**103.95**
BAAFFM	**89.4**	**95.76**	**80.5**	**84.95**

Table 10: Average Log Predictive Likelihood (ALPL) at forecast horizon $h = 4$, bold indicates a superior performance with respect to the random-walk model. **Note: here the hierarchical prior on state variances parameters is replaced with fixed-value hyperparameters, uninformative ones.**