Moderate Time Varying Parameter VARs *

Alessandro Celani †
Örebro University

Luca Pedini ‡
Fondazione ENI Enrico Mattei

April 30, 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. If parameters evolve slowly, as they often do in macroeconomics, this approximation provides forecasting and computational benefits, without compromising structural interpretation. Estimation is framed inside a Bayesian setting, where we further propose a hierarchical step to control for the hyperparameters of state equation prior variances, whose selection is crucial. The resulting approach is fully tractable and computationally efficient. Using U.S. datasets of varying dimensions, we apply the proposed techniques to TVP-VARs, some of the most highly parameterized state space models in macroeconomic practice. 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: .....

*THIS WORK IS PRELIMINARY, PLEASE DO NOT CIRCULATE.

†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 (TVP) models in empirical macroeconometrics. Starting from the seminal works by Cogley and Sargent (2001, 2005) and Primiceri (2005) in Vector Autoregressive models (VAR), time-varying approaches have become a standard tool in forecasting and structural analysis exercises for the main macroeconomic aggregates (Clark, 2011; Koop and Korobilis, 2013; Clark and Ravazzolo, 2015; Carriero et al., 2016, 2019; Kapetanios et al., 2019). Most of their success resides in their flexibility in capturing the dynamics of the economy, often characterized by instabilities and other forms of non-linearities.

However, this flexibility comes at a cost: the TVP scheme postulates new sets of parameters at each point in time, raising potential concerns in terms of overparametrization. Bayesian approaches, via hierarchical priors, allow to mitigate the inherent overfitting they are affected. However, they 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-VARs are commonly used in small-scale setups with the potential shortcoming of omitted variable bias, an issue that cannot be tackled by simply allowing for time variation (Bańbura et al., 2010; Giannone et al., 2014).

To circumvent this issue, some studies in the literature have started to restrict the time variation to only some coefficients of interest. The leading approach focuses solely on time-varying volatilities through stochastic volatility (SV), which is identified as the main source of temporal instability (Clark and Ravazzolo, 2015; Chan, 2023a; Carriero et al., 2024). More flexible alternatives involve intercepts to vary over time, in recognition of the evolving trends in certain key macro variables (Banbura and van Vlodrop, 2018; Götz and Hauzenberger, 2021).

We propose a different perspective: the common denominator of TVP models in the macroeconometric literature is the use of as many coefficients as the length of the relevant time domain T. However, temporal instabilities that economists typically face are quite limited (when present). 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 in this article an efficient and parsimonious alternative parametrization for time-varying dynamics in the realm of State Space (SS) models. Specifically, we decompose the T-dimensional parameter transition equation into a lower dimensional one with deterministic function 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 keeping smooth dynamics.

The strength of the procedure is easily explained: by fixing the number of knots equal to the temporal dimension and with naive splines (0-th degree) it is possible to

recover the canonical TVP pattern commonly used in the literature. However, when fewer knots are used, we effectively impose a constraint on the degree of temporal variability with advantages in terms of estimation efficiency and computational time. Higher-degree splines guarantee smooth dynamics that would otherwise be lost if we merely reduced the parameter space. Indeed, spline functions find their *raison d'être* by ensuring this smoothness, which naturally counters the more rigid and stepwise dynamics induced by a lower-frequency transition.

The proposed method is general and applicable to any linear or linearizable Gaussian SS model, and and can be approached from both a frequentist and a Bayesian perspective. In this paper, we adopt the latter. Due to the linearity of the decomposition, conditional posteriors for the parameters and the latent states can be easily derived using standard procedures from the literature, thus enabling an efficient Gibbs sampling procedure. Nevertheless, our proposed approach does not come without costs: by performing this decomposition, it becomes crucial to develop a strategy for selecting the hyperparameters governing the variances of the new transition equations. To address the point, we propose hierarchical priors. The resulting conditional posteriors come in closed-form, implying minimal adjustments to the Gibbs sampler, and without adding any significant computational burden.

In order to introduce the method as a viable alternative in macroeconomic settings, we apply it to TVP-VARs, the most extensively parametrized SS models, introducing what we will refer to as a Moderate TVP-VAR (MTVP-VAR). The transition equations of our MTVP-VAR are decomposed using third order B-splines and auxiliary transitional dynamics that scales as a fraction of T. By framing the problem as a model selection exercise, we aim to empirically address the question raised before, but we also verify whether an economically-meaningful approximation of evolving macroeconomic dynamics can be achieved with fewer time-varying parameters.

We shed light on these points using a U.S. dataset of varying dimensions and through both full-sample and expanding window forecasting exercises. In particular, the former shows that reducing the number of parameters does not undermine interpretability, even in small-dimensional settings. The estimated structural objects of a benchmark TVP-VAR with T states and a moderate version yield equivalent economic interpretations, despite the latter relying on substantially fewer parameters. Interestingly, this holds not only for lagged coefficients but also for intercepts and stochastic volatilities, components that are expected to exhibit considerable time variation.

In the forecasting exercise, we find that point forecasts from the MTVP-VAR are at-least as competitive as the benchmark, while the density forecasts are clearly superior especially in higher-dimensional settings. Notably, these good forecasting performances are accompanied by significant computational gains, with time savings ranging from 50% to more than 75% compared to the benchmark.

In this way, this article contributes to the expanding literature on developing computationally efficient dimensionality reduction techniques for TVP-VARs. Given the inherently challenging nature of the question, relatively few contributions have moved in this direction from Primiceri (2005) original work. Some notable examples are Koop and Korobilis (2013) with a forgetting factor approach or Chan et al. (2020) with a proper factor model. Recently Chan (2023b) has proposed a Bayesian stochastic search approach to identify which coefficients vary over time in TVP-VARs. Special mention to the strand of literature, which relies on non–likelihood-based methods. Some of these contributions are Giraitis et al. (2018); Kapetanios et al. (2019) with nonparametric local method based on kernel smoothing or Petrova (2019) with quasi-Bayesian treatment.

It is worth mentioning that the MTVP is also related to the literature on using weighting schemes to approximate coefficient evolution, as pionereed by Barnichon and Brownlees (2019), where B-splines combined with second-difference priors achieve smooth dynamics in local projections.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed methodology, beginning with a simple illustrative framework and then generalizing it to a full TVP regression with SV. We also discuss the selection of weights to perform the decomposition. Section 3 outlines the empirical framework and introduces the Moderate TVP-VAR. For this model, we describe the hierarchical prior setup for state variance parameters and detail the estimation procedure. We then present results from both full-sample and forecasting exercises. Finally, Section 4 concludes and outlines directions for future research.

2 Moderate TVP models

To convey the main idea, we start from a simple univariate TVP regression with a single predictor and homoskedastic errors. Let y_t and x_t denote stationary time series observed in $t = 1, \dots, T$. The measurement equation is specified as:

$$y_t = x_t \beta_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

Model 1 commonly poses two specific alternative options to specify the evolution of β_t. First, a constant dynamic over time, such that $\beta_1 = \beta_2 = \cdots = \beta_T$. Although parsimonious, this specification completely disregards potential temporal variation. Alternatively, β_t could vary at every time period, such that $\beta_1 \neq \beta_2 \neq \cdots \neq \beta_T$. Notice that this full-flexible specification introduces as many parameters as observations, making OLS infeasible without additional assumptions. To address this pathology, it is standard to

model β_t using a driftless Random Walk (RW):

$$\beta_t - \beta_{t-1} \sim \mathcal{N}(0, u) \tag{2}$$

with initial condition $\beta_1 \sim \mathcal{N}(0, v_1)$. The system of Equations (1)–(2) defines a SS model, where estimation is typically performed using a combination of MCMC methods and Kalman filter-based techniques. To prevent ambiguity and facilitate comparability, we will refer to the above frameworks as "Const" and "RW_T", respectively.[1]

At a closer look, these two cases appear to be extremes and a third intermediate possibility is viable: in the presence of temporal instability that is not pervasive, dynamic heterogeneity may be effectively captured using a number of parameters lower than T. Motivated by this concept, the MTVP reflects the idea of "moderate time variation", where parameters are allowed to change but not necessarily at each point in time, with both extremes as special cases. In the next section, we will sketch the mathematical and statistical details of the method.

2.1 Proposed method

Suppose to partition the temporal domain into $1 \leq R \leq T$ non-overlapping intervals of equal length. For example, with quarterly data and $R = T/8$ (assuming for simplicity that T is a multiple of 8), each interval corresponds to two years. Let θ_r, $r = 1, \dots, R$ be an auxiliary parameter associated with the r-th interval, initialized as $\theta_1 \sim \mathcal{N}(0, v_1)$. The parameter θ_r captures lower-frequency time variation, as it is allowed to vary "just" R times across the time domain. To maintain coherence with the standard framework, we model the transition of θ_r using a driftless RW:

$$\theta_r - \theta_{r-1} \sim \mathcal{N}(0, v) \tag{3}$$

The major challenge is to construct a mapping between $\boldsymbol{\theta} = [\theta_1, \dots, \theta_R]' \in \mathbb{R}^R$ and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_T]' \in \mathbb{R}^T$. To achieve this, we introduce a set of fixed weights $w_{t,r}$ satisfying the usual constraints for each t: $w_{t,r} \geq 0, \quad \sum_{r=1}^{R} w_{t,r} = 1$. The mapping is defined as:

$$\beta_t = w_{t,1}\theta_1 + \cdots + w_{t,R}\theta_R$$

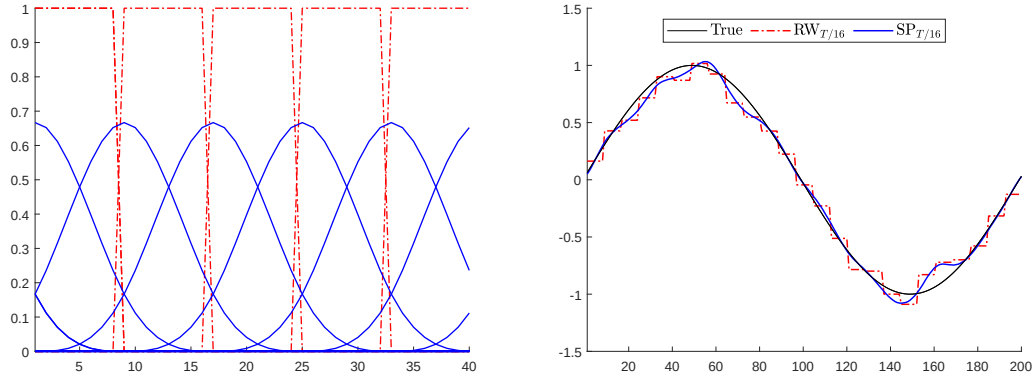
or, compactly:

$$\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\theta} \tag{4}$$

where $\mathbf{W}$ is $T \times R$ weight matrix with generic elements $w_{t,r}$. Equation (4) provides a connection from the high-dimensional $\boldsymbol{\beta}$ to the lower-dimensional $\boldsymbol{\theta}$. While the discussion about specific functional forms of the weights is deferred to the next section, it is worth

[1] The subscript in RW_T emphasizes the number of parameters required.

Figure 1: An example of B-Spline basis and their related fit



Note: Two series of B-Spline basis functions with equal partitions $R = T/8$ and different degrees: $q = 0$ degree (red dashed-dotted lines), $q = 3$ degree (blue lines). Panel (b): estimated parameters under the two different assumptions of a single repetition of the model in (1) - (2). Black line: true β_t, red dashed-dotted line: $RW_{T/8}$ fit, blue line: $SP_{T/8}$ fit.

noting that, at each t, only a subset of the weights is nonzero. Each β_t depends just on parameters from its corresponding region and potentially neighboring regions, making $\mathbf{W}$ highly sparse. Importantly, this hierarchical formulation nests all the previous cases discussed. For example, when $R = 1$, $\beta_1 = \cdots = \beta_T = \theta_1$, and the Const case is reproduced. On the other hand, by using $R = T$, $w_{t,t} = 1$ and $w_{t,r} = 0$ for $r \neq t$, one is able to recover the RW_T case. Intermediate cases, such as the one with $R = T/8$ mentioned before, can be represented as: $w_{t,r} = \mathbb{1}(\lceil t/8 \rceil = r)$ where $\mathbb{1}(\cdot)$ is the indicator function, while $\lceil \cdot \rceil$ is the ceiling operator.[2] To keep consistency with the terminology adopted, we refer to any intermediate case characterized by a plain weighting scheme (either 0 or 1) as RW_R.

2.1.1 Weight selection

The previously discussed 0-1 weights can only produce stepwise functions, which poorly capture the gradual and smooth changes typical of macroeconomic dynamics. However, the equivalence in Equation (4) is more general and applies to many weighting schemes: a variety of basis functions like B-splines, Fourier basis or wavelets can be efficiently represented through the same linear parameterization.

In this study we use B-Splines (BSP).[3] A BSP consists of $q+1$ polynomial segments, each of degree q, which are connected at q inner knots. At these connection points,

[2] In matrix notation the Const, the RW_T, and the $RW_{T/8}$ cases can be obtained by setting, respectively, $\mathbf{W} = 1$, $\mathbf{W} = \mathbf{I}_T$, and $\mathbf{W} = \mathbf{I}_{T/8} \otimes \boldsymbol{\iota}_8$, with $\boldsymbol{\iota}_J$ representing a J dimensional column vector of ones.

[3] See Barnichon and Brownlees (2019) for the application of B-Splines in a Local Projection setting, and DeBoor (2000); Eilers and Marx (1996) for a summary of their properties.

derivatives are continuous up to order q. The BSP is positive only within the domain defined by the q knots and is zero outside this domain. Their advantage can be immediately grasped for two key reasons: firstly, for any $q > 1$ their smooth behavior produces parameter variations that avoid sudden breaks, aligning with the rationale behind the RW-based transitions. Secondly, each BSP basis restricts non-zero influence to a specific area by construction. Consequently, changes affect only part of the curve. Thus, BSPs offer both the features of "smoothness" and "forgetfulness" of far past observations, complying with the observed patterns in macroeconomic dynamics.

In particular, this paper focuses on cubic ($q = 3$) B-Splines with equidistant knots ranging from 1 to T, with increments given by $\bar{R} = T/k$, $\bar{R} = R + q$, and for any chosen $k = 1, \ldots, T$. A general BSP configuration is denoted as SP_R. Notice, however, that BSPs of order 0 allow to incorporate the Const (0 knots, $R = 1$) and the RW_T (Tequally spaced knots, $R = T$) cases. [4]

To give the reader an idea on how the machinery works, we simulate a single replication of the models in (1), setting $x_t = 1$, and specifying a deterministic law of motion for β_t using sine waves: $\beta_t = \sin(2\pi t/T)$, over $T = 200$ periods. [5] On this simulated process, we fit two models: the $RW_{T/8}$ and the $SP_{T/8}$, which differ solely in the polynomial order of the spline: $q = 0$ for the former, and $q = 3$ for the latter. Figure (1) presents the results. Panel (a) illustrates the basis functions employed to fit the models over the initial 40 periods, while panel (b) displays the resulting fit. Both models efficiently interpolate the unobserved sine wave using 8 times fewer parameters than normally required by a standard RW_T. The importance of employing a spline with degree $q > 0$ emerges when comparing the red and blue lines: the former is stepwise, whereas the latter is smooth.

2.2 General framework

The proposed method can be straightforwardly applied to any setting involving TVP parameters, as the decomposition in (2) is variable-specific. We now extend the concept developed earlier to a TVP regression with SV, a framework that includes most of the models employed in empirical macroeconomics. In particular, let $\mathbf{x}_t$ be a $K \times 1$ dimensional vector of explanatory variables, possibly including deterministic terms and lagged values of the dependent one. The general framework can be written as:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \exp(h_t)) \tag{5}$$

[4] In the BSP literature it is actually more common to express the transition equation of the associate coefficient parameters in second difference (Eilers and Marx, 1996; Barnichon and Brownlees, 2019). However, this prior suggests that parameters are I(2), which seems improbable for most of the macroeconomic dynamics we analyze.

[5] This setup is in line with the DGP used by Amir-Ahmadi et al. (2020) in their simulation exercise.

where $\boldsymbol{\beta}_t$ represents a vector of conditional mean parameters, while h_t denotes log volatilities. To keep the exposition as simple as possible, we assume that a common partition R governs the temporal fluctuations of all parameters, although the method readily allows for different values. [6] We postulate the following decomposition for $\boldsymbol{\beta}_t$, and h_t:

$$\begin{aligned} \boldsymbol{\beta}_t =& \mathbf{W}_{t,1}\boldsymbol{\theta}_1 + \ldots + \mathbf{W}_{t,R}\boldsymbol{\theta}_R \\ h_t =& w_{t,1}g_1 + \ldots + w_{t,r}g_R \end{aligned} \tag{6}$$

where, by assuming a common weighting scheme across coefficients, we have $\mathbf{W}_{t,r} = w_{t,r}\mathbf{I}_K$. The model completes by placing the customary RW-based transition dynamics on the auxiliary parameters $\boldsymbol{\theta}_r$, and g_r:

$$\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}), \quad g_r - g_{r-1} \sim \mathcal{N}(\mathbf{0}, s) \tag{7}$$

subject to standard initial conditions $\boldsymbol{\theta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_1)$, $g_1 \sim \mathcal{N}(0, s_1)$.

A great advantage of the proposed decomposition is that it retains linearity in the parameters to be estimated. To illustrate this more clearly, let us rewrite the model in compact form. To that end, let $\mathbf{y} = [y_1, \ldots, y_T]'$, $\mathbf{X} = \text{diag}(\mathbf{x}_1', \ldots, \mathbf{x}_T')$, and $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_T]'$. Equation (5) can be seen as a linear regression model with general error covariance matrix:

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

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_T']'$, and $\boldsymbol{\Sigma} = \text{diag}(\exp(h_1), \ldots, \exp(h_T))$. The second ingredient is the vector-valued decomposition of $\boldsymbol{\beta}$, as a function of the weights and $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_R']'$. The first equation of (6) can be now stacked over T and expressed as

$$\boldsymbol{\beta} = (\mathbf{W} \otimes \mathbf{I}_K)\boldsymbol{\theta}$$

Then, the model can expressed as a linear function of $\boldsymbol{\theta}$:

$$\mathbf{y} = \bar{\mathbf{X}}\boldsymbol{\theta} + \mathbf{e} \tag{8}$$

where $\bar{\mathbf{X}} = \mathbf{X}(\mathbf{W} \otimes \mathbf{I}_K)$. In terms of Bayesian estimation, the conditional posterior of $\boldsymbol{\theta}$ can be easily derived combining (8) with the vectorized version of the prior for $\boldsymbol{\theta}$ and using standard linear regression results. Draws from this posterior can be generated using the efficient sampler of Chan and Jeliazkov (2009). The same argument applies to the parameters of stochastic volatility. Specifically, the error term in (5) can be reformulated

[6] In the context of the MTVP-VAR model presented in the next section, we discuss viable alternatives to this choice.

as a function of $\mathbf{g} = [g_1, \ldots, g_R]'$ as follows:

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

where $\mathbf{u} = [u_1, \ldots, u_T]'$. Then, $\mathbf{g}$ can be sampled efficiently using the efficient sampler previously described in combination with the mixture representation of Kim et al. (1998).

All details about posterior derivations and the sampling of parameters are given for the TVP-VAR we use in the empirical application, and presented in Appendix B.

3 Empirical application

We now evaluate our proposed method on real data using TVP-VAR models, probably the most popular TVP modelization in macroeconomics. Our aim is twofold: on the one hand, we provide applied economists with a tool that maintains the same economic interpretation as the literature benchmark. On the other hand, we show the remarkable performance of the method, with the additional advantage of making a previously computationally demanding procedure both feasible and costless.

Specifically, in Section 3.1 we introduce our Moderate TVP-VAR (MTVP-VAR), together with estimation and prior details. In Section 3.2, we present full-sample results in a small dimensional setting, focusing on estimated structural objects and further examining the role of knot positioning. Finally, in Section 3.3, we perform a formal model comparison based on out-of-sample expanding window exercises, showing the advantages of the MTVP with respect to the benchmark RW_T.

Data. The dataset used consists of 14 quarterly variables for the US ranging from Q1 1959 to Q4 2019, sourced from the FRED-QD database at the Federal Reserve Bank of St. Louis. We consider a variety of standard macroeconomic variables, such as real GDP, inflation, unemployment, labor market variables, money supply and interest rates. Non-stationary variables are transformed by taking annualized growth rates, in order to ensure stable dynamics. Results will be presented using different VAR models featuring progressively larger sets of variables, along the same line of Giannone et al. (2015): a *small-scale* (N=3) monetary policy TVP-VAR including real GDP growth, GDP deflator inflation, and the federal funds rate; a *medium-scale* (N=7) one that adds consumption, investment, hours worked, and wages; and a *large-scale* VAR that includes all the 14 variables. The complete list of variables and their transformation are provided in Table 7, Appendix A.

3.1 A Moderate TVP-VAR

Motivated by computational considerations, we consider a slight variation to the original Primiceri (2005) framework by adopting the equation-by-equation representation proposed by Chan et al. (2024). Let $\mathbf{y}_t$ be a N dimensional vector of endogenous variables and assume a TVP-VAR in structural form, namely

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

where $\mathbf{\Phi}_t = [\mathbf{c}_t, \mathbf{\Phi}_{1,t}, \ldots, \mathbf{\Phi}_{P,t}]$ is the $N \times NP+1$ matrix of intercepts and lagged coefficient matrices with P as the lag order, $\mathbf{\Phi}_{0,t}$ is a $N \times N$ lower triangular matrix with ones on the main diagonal, and $\mathbf{\Sigma}_t = \text{diag}(\exp(h_{1,t}), \ldots, \exp(h_{N,t}))$ is a diagonal matrix of stochastic volatilities. The vector of exogenous variables is formed coherently as: $\mathbf{x}_t = [1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-P}]'$. Given the sequential structure, the model in equation (9) can be estimated by running N separate TVP regressions. [7] In particular, let $\mathbf{\Phi}_{i:,t}$ be the i-th row of $\mathbf{\Phi}_t$, and $\mathbf{\Phi}_{i:,0,t} = [\phi_{1,t}, \ldots, \phi_{i-1,t}]'$ collects the free elements of the i-th row of $\mathbf{\Phi}_{0,t}$, with $\mathbf{\Phi}_{1:,0,t} = \emptyset$. Furthermore, let $\boldsymbol{\beta}_{i,t}$ be the stacked version of all coefficients of variable i: $\boldsymbol{\beta}_i = [\mathbf{\Phi}'_{i:,t}, \mathbf{\Phi}'_{i:,0,t}]'$ for $i = 1, \ldots, N$. Clearly, the dimension of the vector of exogenous variables is expanding in i, and equals to $K_i = NP + i$. The total number of parameters to estimate is thus $K = \sum_i K_i$. The TVP-VAR in (9) can be rewritten as a sequence of N independent TVP regressions with SV of the form:

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

where $\mathbf{z}_{1,t} = \mathbf{x}_t$, and $\mathbf{z}_{i,t} = [\mathbf{x}'_t, -y_{1,t}, \ldots, -y_{i-1,t}]'$. We apply our proposed decomposition independently to all these TVP regressions:

$$\begin{aligned} \boldsymbol{\beta}_{i,t} &= \mathbf{W}_{1,t}\boldsymbol{\theta}_{i,1} + \cdots + \mathbf{W}_{R,t}\boldsymbol{\theta}_{i,R} \\ h_{i,t} &= w_{1,t}g_{i,1} + \cdots + w_{R,t}g_{i,R} \end{aligned} \tag{11}$$

Coherently, the auxiliary parameters evolve as driftless RWs:

$$\boldsymbol{\theta}_{i,r} - \boldsymbol{\theta}_{i,r-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_i), \quad g_{i,r} - g_{i,r-1} \sim \mathcal{N}(0, s_i) \tag{12}$$

where the error covariance matrices for the state equations of the mean are diagonal, that is $\mathbf{V}_i = \text{diag}(v_{i,1}, \ldots, v_{i,K_i})$. Anytime $R < T$, this decomposition in conjunction with a third order B-Spline weighting scheme gives rise to a MTVP-VAR. As before, we denote the moderate specifications as SP_R, in line with our convention.

[7] The recursive assumption is used as a computational device and not as an identification scheme. The estimation setup of Primiceri (2005) involves sampling all the parameters of the conditional mean in one step, which is unfeasible even for medium-dimensional models. See the discussion in Chan (2023b).

In this paper, we keep the number of regions R equal between the mean parameters and log volatilities. This is a practical choice, as it allows us to implement a hierarchical prior structure in which one shrinkage governs both intercepts and volatilities prior state variances, while another governs the rest. Using a common R across all parameters ensures that the shrinkage applied to intercepts and volatilities is informed by the same number of observations. A more flexible alternative would involve separating the estimation blocks for the mean parameters, assigning one R for intercepts and volatilities, and a different one for the remaining coefficients. Indeed, even within a hierarchical framework, the optimal choice of R for these two blocks is likely to differ. We leave this extension for future research.

A hierarchical approach for state variances. Eliciting state prior variances in such a model is a daunting task. Standard values commonly used in the literature may be adequate for the benchmark case, *i.e.,* the RW_T, but are unlikely to perform well when the number of regions R is arbitrary. For this reason, we introduce a hierarchical approach to estimate them. The state prior variances for both the mean and log volatilities follow an Inverse Gamma distribution. However, since intercepts and log volatilities typically exhibit a different degree of time variation compared to the other lagged and contemporaneous parameters, we find it convenient to assign them a different scale:

$$s_i, v_{i,1} \sim \mathcal{IG}(\tau_1, (\tau_1 + 1)\xi_1), \quad v_{i,j} \sim \mathcal{IG}(\tau_1, (\tau_1 + 1)\xi_2), \quad j = 2, \dots, K_i \tag{13}$$

where τ_1 represents the shape of the Inverse Gamma. Both are set such such that the prior mode for log volatilities and intercepts is ξ_1 and ξ_2 for all the other coefficients. The hierarchical structure is introduced by estimating both parameters, each of which is assigned an Inverse Gamma prior:

$$\xi_1 \sim \mathcal{IG}(\nu, (\nu + 1)\bar{\xi}_1), \quad \xi_2 \sim \mathcal{IG}(\nu, (\nu + 1)\bar{\xi}_2) \tag{14}$$

As before, $\bar{\xi}_1$ and $\bar{\xi}_2$ represent the prior modes, for which we set $\bar{\xi}_1 = .1^2$ and $\bar{\xi}_2 = .01^2$. However, in our setting, they simply represent the most likely values a priori, which will be updated in light of the data. The hyperparameter ν is set to 1 across all specifications and dimension, providing an overall level of informativeness for these hierarchical priors. Naturally, the degree of informativeness depends on the dimensionality of the system: it is relatively strong when $N = 3$, and becomes progressively weaker as N increases. Notice that the hierarchical step is also applied to the benchmark RW_T in order to guarantee a fair comparison. This deviates from the RW_T commonly encountered in the literature, where prior hyperparameter are fixed by the researcher. This may introduce undesired degrees of subjectivity that we avoid in favor of a more rigorous and more robust specification.

Remaining priors and Bayesian estimation. The priors for the initial conditions are Gaussian: $\boldsymbol{\theta}_{i,1} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{i,1})$, where $\mathbf{V}_{i,1}$ is diagonal and follows a classical Minnesota style (Doan et al., 1984). To this end, let $\mathbf{V}_1 = \text{diag}(\mathbf{V}_{1,1}, \dots, \mathbf{V}_{N,1})$ be the full covariance matrix of the initial conditions parameters. I In particular, its generic j-th element $v_{j,11}$ is equal to

$$v_{1,jj} = \begin{cases} \frac{\lambda_1}{l^{1.5}}, & \text{own } l\text{-th lag} \\ \frac{\lambda_2 \sigma_i^2}{l^{1.5} \sigma_j^2}, & \text{cross-equation } l\text{-th lag} \\ \frac{\lambda_3 \sigma_i^2}{\sigma_j^2}, & \text{contemporaneous variables} \\ \lambda_4 \sigma_i^2, & \text{intercept} \end{cases}$$

where σ_i^2 denotes the sample variance of the residuals from individual AR(4) of each $y_{i,t}$, $\lambda_3 = 1$, $\lambda_4 = 100$ 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: $\lambda_1 \sim \mathcal{G}(1, .4^2)$, $\lambda_2 \sim \mathcal{G}(1, .2^2)$. The scale parameters are such that the prior standard deviation of the initial conditions are, respectively, .4 and 0.2. and $g_{i,1} \sim \mathcal{N}(0, s_{i,1})$. Log volatilities initial conditions $g_{i,1}$ are assumed to be zero mean slab Gaussian distributions, *i.e.*, $g_{i,1} \sim \mathcal{N}(0, 10)$.

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: $p(\boldsymbol{\theta}_i|\bullet)$, $i = 1, \dots, N$, $p(\mathbf{g}_i|\bullet)$, $i = 1, \dots, N$, $p(v_{i,j}|\bullet)$, $i = 1, \dots, N$, $j = 1 \dots, K_i$, $p(s_i|\bullet)$, $i = 1, \dots, N$, $p(\lambda_1|\bullet)$, $p(\lambda_2|\bullet)$, $p(\xi_1|\bullet)$, $p(\xi_2|\bullet)$, 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.

3.2 Full-sample results

In this section, we present results based on full-sample estimations, with the goal of comparing the benchmark specification (RW_T) to the moderate versions of the TVP-VAR. The lower-frequency parameter $\boldsymbol{\theta}_R$ is defined over the following finite grid of partitions:

$$R = \{T/4 \;\; T/8 \;\; T/16 \;\; T/32\}$$

which correspond to a progressive reduction in the number of parameters, ranging from one coefficient per year to the extreme case of one every 8 years.

We begin by analyzing the role played by the hierarchical shrinkage parameters ξ_1 and ξ_2. To that end, Table 1 presents the posterior mean and standard deviation of ξ_1, which governs the overall shrinkage of intercepts and log volatilities. As expected, for each

Table 1: **Posterior means and standard deviations of ξ_1.**

	RW_T	$\text{SP}_{T/4}$	$\text{SP}_{T/8}$	$\text{SP}_{T/16}$	$\text{SP}_{T/32}$
N=3	.271 (.035)	.665 (.107)	.868 (.167)	.801 (.206)	.882 (.238)
N=7	.201 (.019)	.392 (.051)	.490 (.077)	.552 (.104)	.626 (.139)
N=14	.156 (.011)	.353 (.035)	.449 (.053)	.537 (.074)	.637 (.111)

Note: The table shows the posterior means and standard deviations of ξ_1, across different dimensions and TVP-VAR specifications.

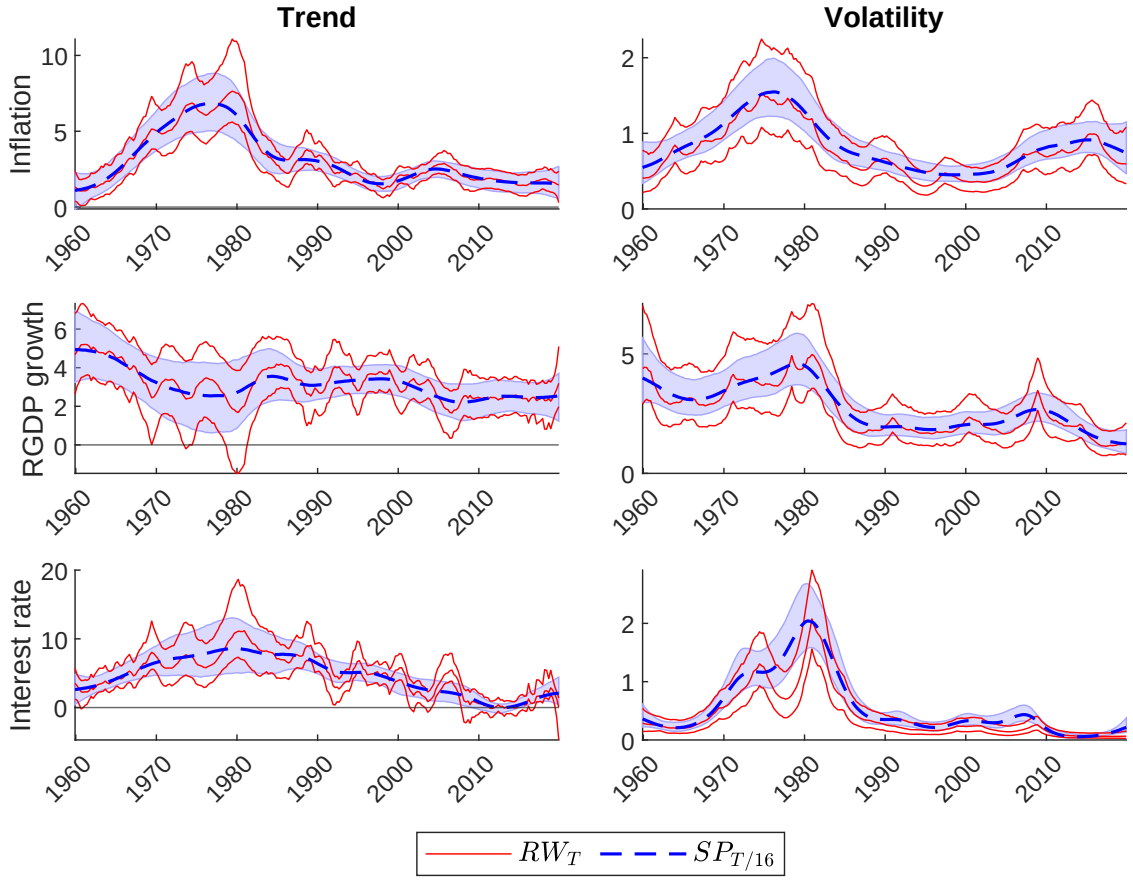
dimension, the overall shrinkage increases as R becomes smaller, indicating that a higher prior variance is needed when the number of parameters decreases. The posterior mean rises monotonically across specifications, and the gap between the most highly parameterized model (RW_T) and the most parsimonious one ($SP_{T/32}$) ranges from approximately 3 for the small model to more than 4 for the large one. The dimensionality plays an important role as well. For any given specification, as the number of variables increases, the posterior mean of ξ_1 decreases, as one would intuitively expect. In small models, time variation tends to capture the effects of omitted variables. This effect diminishes in higher-dimensional settings, as time variation becomes less pervasive. The hierarchical structure is here motivated: ξ_1 exhibits substantial variation across specifications and dimensions, as confirmed by the relatively small posterior standard deviations. Omitting this hierarchical step would likely lead to a deterioration in model fit. Results for ξ_2 are presented in Table 8 in Appendix C. The pattern is very similar, with the exception that the magnitudes are smaller due to the limited temporal variability of the lagged parameters.

Once understood the hierarchical prior impact, we move to the main experiment, where we investigate whether using fewer parameters affect the interpretation of the objects economists are normally interested in. To this end, we use the same small scale model, which includes real GDP growth, inflation, and the interest rate. Figure 2 compares the estimated trends (or infinite-horizon forecasts) and stochastic volatilities for these variables. We compare the benchmark TVP-VAR (RW_T) with a moderate alternative with one coefficient every four years ($SP_{T/16}$). Since the estimation of trends requires global stationarity, we incorporate an accept-reject step to discard explosive draws as in Cogley and Sargent (2005). [8]

What clearly emerges is that the moderate version captures only the very long-run patterns in the coefficients, providing a smoothed version of the more flexible RW_T.

[8] Koop and Potter (2011) criticized this approach, noticing that such a procedure to impose inequality constraints can perform poorly, particularly in high-dimensions. Our approach indirectly addresses this issue: by approximating the dynamics with fewer parameters, locally nonstationary behaviors are limited by construction.

Figure 2: **Trends and volatiltities**

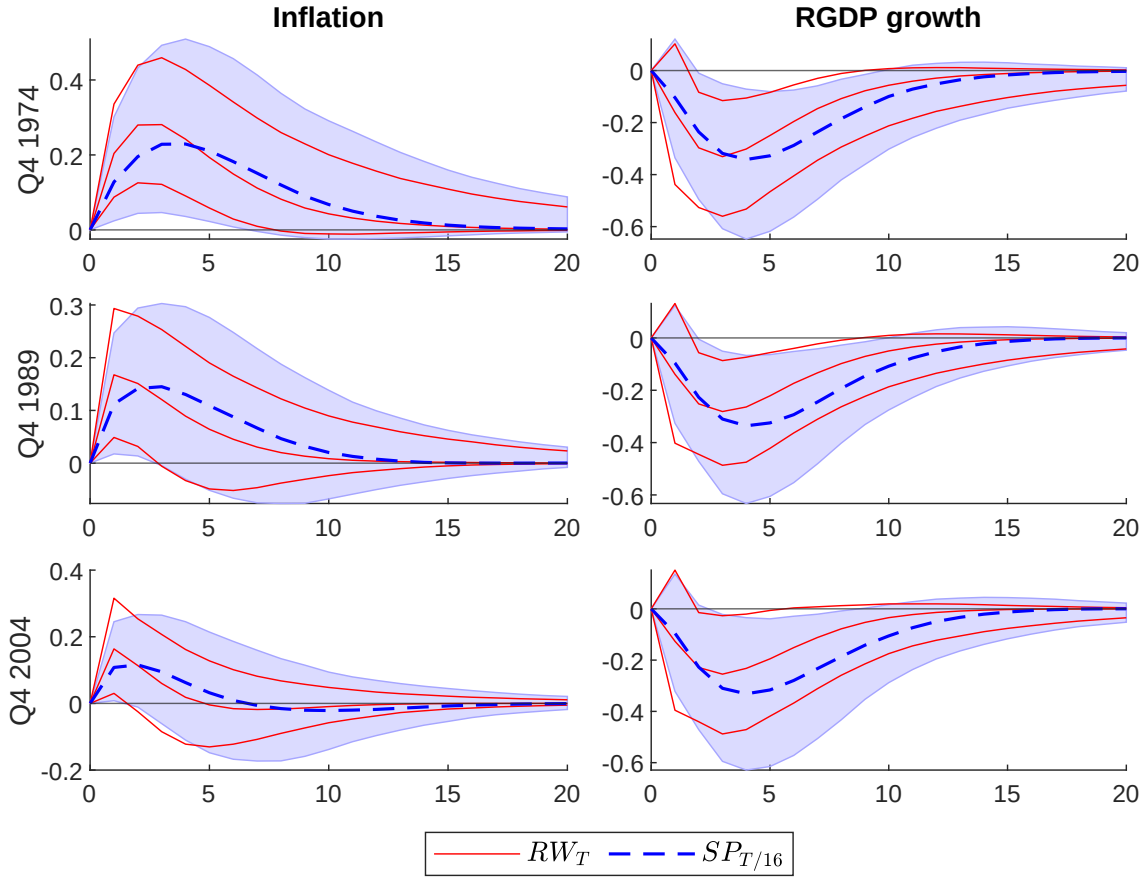


Note: The figure shows the posterior median trends and volatilities $\exp(h_t/2)$ for the two specifications: benchmark RW_T (red lines), and moderate $SP_{T/16}$ (blue dashed lines). External red lines and blue shaded-areas delimit the 90% credible bands.

The latter exhibits many jagged fluctuations, some of which appear to be statistically significant. However, from an interpretative standpoint, an economist would likely reach similar conclusions with both models: both capture the well-known secular decline across all three variables. As for the trends, the only notable difference is for the interest rate, where the posterior distribution of the RW_T oscillates more markedly since the great moderation. For stochastic volatilities we obtain a very similar picture: both models successfully identify the decline in volatility after the Great Inflation period. Again, the two models differ for the interest rate: $SP_{T/16}$ fails to capture the abrupt shift in volatility occurred in the late 70s, but then converge thereafter.

We now present the structural impulse responses of inflation and real GDP growth to an unexpected Monetary Policy (MP) shock that results in an increase in the interest rate. The shock is identified using a standard Cholesky-type recursive ordering, as in Primiceri (2005). IRFs are presented for three selected quarters: Q4 1974, Q4 1989, and Q4 2004. For comparability across time, the shock is renormalized so that its contemporaneous

Figure 3: **IRFs after a monetary policy shock**



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 $SP_{T/16}$ (blue dashed dotted-lines). External red lines and blue shaded-areas delimit the 90% credible bands.

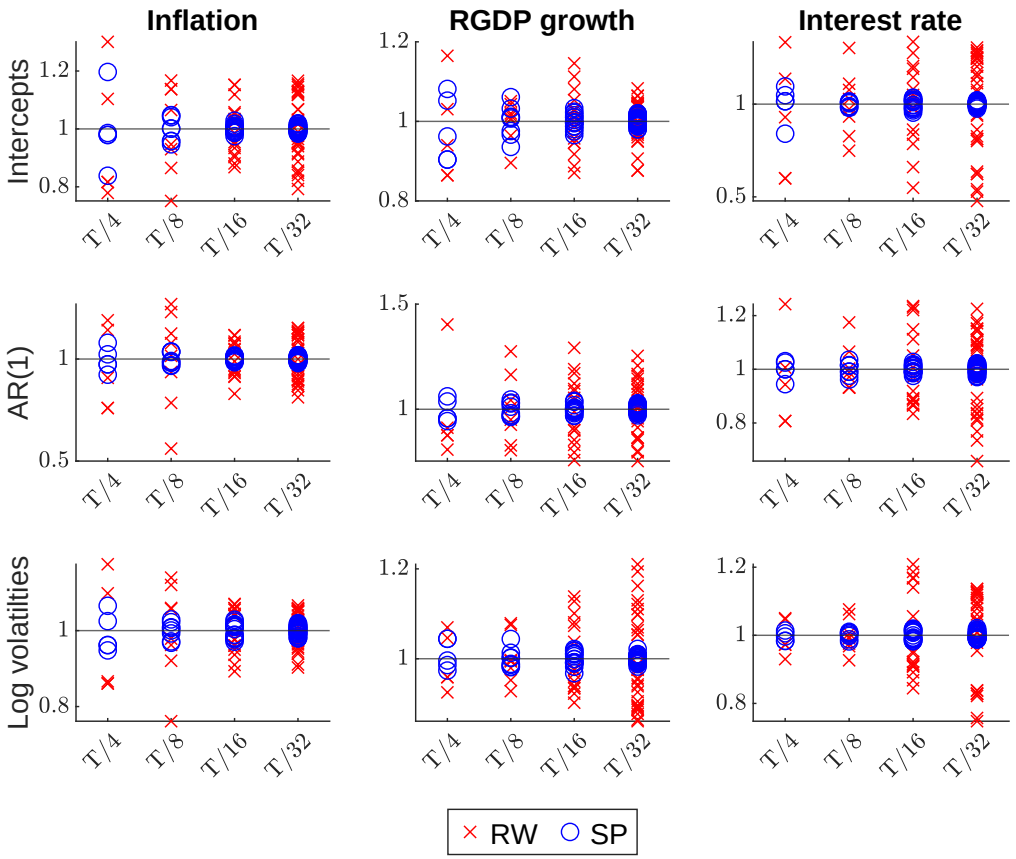
effect on the interest rate is equal to 1 for each t. Figure 3 presents the results of this exercise. In both cases, a quarter after the MP shock, inflation rises and real GDP growth declines, producing a stagflationary effect. For the moderate specification, the peak response occurs one quarter later than in the benchmark model. The IRFs exhibit a hump-shaped pattern, with a more gradual return to the time-varying equilibrium. This is expected, as the estimated trends under $SP_{T/16}$ are smoother, which in turn allows for more persistent cyclical fluctuations. Once again, considering the width of the credible bands, the two models provide a very similar interpretation, suggesting that the moderate specification is a valid modeling alternative even for structural analysis. The results would likely overlap even more if one adopted a time-varying trend-cycle decomposition, allowing for a different number of regions R for trends and volatilities compared to the other coefficients. For the sake of completeness, in Figures 6 and 7 in Appendix C we propose the same using the more parametrized $SP_{T/8}$ specification.

The role of knots position. The ability of the proposed approach to reduce the parameter space does not come without costs. As the number of regions R deviates from T, the position of the knots becomes increasingly important. Consequently, it becomes essential to investigate how different knot placements might affect the model's results for varying R. We already know that this pathology is partially mitigated by the weighting scheme, which smooths the otherwise piecewise-constant dynamics. Nevertheless, it is unclear how effective is the spline effect for longer regions, such as the previously discussed case ($R = T/16$). Generally, we expect that as R decreases, estimation variability will increase, potentially undermining estimation stability. Since we do not have a formal test for this issue, we proceed empirically. For each of the four partitions of interest R_i, $i = 1, \dots 4$, we consider a number of MTVP-VARs, each one featuring a different knots placement. For instance, when $R = T/4$, knots are there are four possible way to position the knots: either at time points $1, 5, 9, \dots, T-1$, $2, 6, 10, \dots$, and so on. For the same logic, there can be up eight combination for $R = T/8$ and so forth. For each combination, we estimate a moderate SP_{R_i}, and, for each parameter, take the fitted median value. Then, we compute the Mean Squared Error (MSE) with respect to the fitted median values of the same parameter under RW_T, which by definition has no knots position problem. We then repeat the same procedure with non weighted specifications RW_{R_i}, using them as a benchmark to assess whether the weighting scheme effectively reduces this potential instability. Figure 4 displays the results of three coefficients: intercepts, AR(1) coefficients and log volatilities. Blue circles (red crosses) are, respectively, the MSE related to the SP_{RP_i} (RW_{RP_i}). Since our main goal is to compare the variability between moderate weighted and non-weighted specifications, we rescale the MSE by their mean, ensuring that the average across combinations is 1.

Overall, the SP specifications show relatively small variability, substantially lower than the non-weighted RW cases. As expected, the variability of the latter increases as R_i decreases. This is a logical consequence, as the piecewise-constant specification generates jumps whose location depends on knot placement. The B-spline approach smooths out this behavior, thereby yielding significantly lower estimation variability across different knot configurations. Surprisingly, SP performances do not deteriorate with very low $R_i$values, such as for the $T/32$ case. A declining performance was somehow expected for such extreme values, especially for SVs, where significant time variation can heavily influence knot placement. However, even under these conditions, the MTVP version performs quite well, demonstrating its validity as a parsimonious alternative also for model interpretation.

A numerical comparison. Before presenting the expanding window results, we first report the estimation times for the proposed MTVP-VARs across window sizes R_i and dimensions of the endogenous vector. The MTVP approach is particularly motivated by computational considerations: the reduction in the number of coefficients significantly

Figure 4: **Sensibility on the knots position**



Note: The figure shows the results of the exercise on the knots position variability. For each coefficients, blue circles (red crosses) represent, respectively, the standardized MSE for all the possible combination of knots position between fitted median of the SP_{RP_i} (RW_{RP_i}) and the RW_T.

enhances the estimation feasibility, especially in high-dimensional models where over-parameterization becomes a dominant issue.

We consider the set of partitions introduced above and examine all three dimensions of the endogenous vector. All the TVP-VARs have 2 lags and computations are performed using Matlab on a desktop computer with an Intel Core i7-1185G7 @3.00 GHz processor and 32 GB RAM. Table 2 reports the computational times (in seconds) to obtain 1000 posterior draws.

The SP specifications even though more parsimonious introduce significant CPU time saving only starting from the $SP_{T/8}$ case. In the small-scale model (N = 3) this effect is mitigated by the limited dimension of the VAR, but it becomes evident in larger dimensions. At first glance, this time degradation may appear counterintuitive, but the main motivation lies in the memory allocation of sparse matrices. The weight matrix W of the RW_T pattern can be conveniently represented via Boolean arrays, which allocate small chunks of memory; cubic splines require, on the contrary, double precision matrices,

Table 2: Computational time

	RW_T	$\text{SP}_{T/4}$	$\text{SP}_{T/8}$	$\text{SP}_{T/16}$	$\text{SP}_{T/32}$
N = 3	8	5	3	2	2
N = 7	45	42	22	14	12
N = 14	314	318	196	124	75

Note: The table shows the computational times (in seconds) to obtain 1000 draws for each TVP-VAR specification and dimension of the dataset.

much more memory-consuming.

When the number of knots is not sufficiently low, the excessive memory burden totally offsets the computational advantages of the reduced space. On the contrary, when the number of knots is sufficiently low, the gain is substantial: the $SP_{T/8}$ is twice as fast as the RW_T for the $N = 3$ and $N = 7$, while the $SP_{T/16}$ and $SP_{T/32}$ are four times faster for the small- and medium-scale model. In larger dimension, the advantages become even more limited for the $SP_{T/4}$ and $SP_{T/8}$, with the former being slightly slower than the RW_T. However, the $SP_{T/32}$ is still the fastest by a factor of 4.

3.3 Expanding window exercise

Full-sample exercises are surely necessary to validate the property of our scheme, but determining the advantages of MTVP in terms of forecast is equally vital. As a consequence, in this section we evaluate the forecast performance of the MTVP approach in a high-dimensional (N=14) through a recursive forecasting exercise. We analyze both the point and density forecast accuracy on $\mathbf{y}_{t+h}$, comparing the benchmark with 4 moderate version based on the R grid introduced in 3.2.

The recursive out-of-sample forecast experiment is built upon an expanding window scheme which starts at Q1 1959 - Q1 1980 and progressively add an observation (quarter) up to 2019-Q4. Two main forecast horizons are considered: $h = 1$ and $h = 4$. To evaluate the point forecast on each dependent variable j at horizon h, we employ the Root Mean Square Forecasting Error (RMSFE), defined as,

$$RMSFE_{jh} = \sqrt{\frac{1}{(T-h)-t_0}\sum_{t=t_0}^{T-h}\left(\hat{y}_{j,t+h|t} - y_{j,t+h}\right)^2},$$

where t_0 defines the final observation of the first forecast window, while $\hat{y}_{j,t+h|t}$ and $y_{j,t+h}$ correspond, respectively, to the sample average of the posterior predictive distribution of $y_{j,t+h}$ in the *t-th* window and the true value. Density forecasts, on the other hand, are

compared via the average log-predictive likelihood of variable j,

$$alPL_{jh} = \frac{1}{(T-h)-t_0}\sum_{t=t_0}^{T-h}\log p(y_{j,t+h}|y_{1:t}),$$

where $p(y_{j,t+h}|y_{1:t})$ is the predictive likelihood of variable j at forecast horizon h.[9] To facilitate the comparison with the RW_T model, we consider the relative RMSFE and the relative average log-predictive likelihood, defined as:

$$relRMSFE_{jh} = 100\,(1 - RMSFE^{SP}_{jh}/RMSFE^{RW}_{jh})$$
$$relalPL_{jh} = 100\,(alPL^{SP}_{jh} - alPL^{RW}_{jh})$$

Tables 3 and 4 report the relative RMSFE: at forecast horizon 1 the MTVP approximation uniformly (all specifications) lead to a superior performance for many variables under analysis. Interestingly, the less parametrized models ($T/16$ and $T/32$) generally produce the best results.

In detail, the MTVP outperforms seems to have an advantage in forecasting the GDP deflator, the Federal fund rate, the unemployment rate and the BAA yield.

At forecast horizon 4, the MTVP performance seem to deteriorate: while for the case of inflation, Federal fund rate, unemployment and 10-year Treasury bill yields the advantage has increased with respect the forecast horizon 1, the predictive performance for all other variables is numerically the same or lesser than the benchmark. Moreover, among the spline approximations there is no a clear winner, with some variables better predicted by the less parametrized models and other by the more parametrized ones.

A feature that should be considered, at least in future developments, is surely the forecast instability, which could easily undermine the point forecast examination.

Coming to the density forecast evaluations in Tables 5-6, the results are more consistent: the MTVP for forecast horizon $h = 1$ outperforms the RW model for most variables, with outstanding advantages for the "Federal fund rate", "10 Year Treasury Bond yield" and "BAA yield" (in accordance with the RMSFE). Less evident, but still significant, results are detected for inflation, hours worked, unemployment rate, capacity of utilization, industrial production and nonfarm payroll.

Even though all MTVP specifications have qualitatively similar gains over the benchmark, the less parametrized ones ($T/16$ and $T/32$) achieve a better overall performance. The overall conclusion may suggest that MTVP operates significantly in the variability and stability of the forecast, which is conveniently conveyed by the distribution rather than the point forecast.

To better understand the joint contribution, Figure 5 reports the joint average pre-

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

	T/4	T/8	T/16	T/32
GDPDEF	**3.09**	**3.2**	**4.21**	**3.42**
GDPC1	**0.55**	**0.5**	**1.11**	**1.39**
FEDFUNDS	**2.86**	**2.41**	**3.12**	**2.53**
DPIC96	**0.18**	**0.33**	**0.72**	**0.86**
PCECC96	**0.92**	**0.46**	**1.17**	**1.04**
CES0600000007	-0.6	**0.67**	**2.03**	**4.22**
CES0600000008	**1.25**	**1.7**	**1.98**	-0.47
CUMFNS	**2.45**	**1.86**	**1.97**	**1.36**
PAYEMS	**2.84**	**1.93**	**2.72**	**2.69**
INDPRO	**3.05**	**2.35**	**2.84**	**2.82**
UNRATE	**0.43**	**2.73**	**5.81**	**5.24**
M2REAL	**0.44**	**0.2**	-0.15	-0.02
GS10	-1.16	-1.11	**1.2**	**1**
BAAFFM	**3.71**	**0.98**	**2.58**	**4**

Table 3: Relative RMSFE with respect to the random-walk model at forecast horizon $h = 4$, bold indicates a superior performance.

	T/4	T/8	T/16	T/32
GDPDEF	**13.53**	**14.02**	**14.71**	**15.1**
GDPC1	-2.69	-2.94	-1.76	-2.79
FEDFUNDS	**5.74**	**2.16**	**4.24**	**3.06**
DPIC96	-0.64	-0.67	-0.38	-0.93
PCECC96	-0.89	-0.82	**0.08**	-0.19
CES0600000007	-4.03	-2.67	-2.3	**2.68**
CES0600000008	**5.66**	**6.44**	**6.52**	**1.29**
CUMFNS	**2.08**	**0.88**	**0.25**	-1.46
PAYEMS	-1.49	-2.5	-2.87	-4.14
INDPRO	-1.43	-2.45	-2.93	-3.45
UNRATE	**2.85**	**6.28**	**8.05**	**3.9**
M2REAL	**0.52**	-1.32	-0.14	**0.77**
GS10	-1.17	-0.94	**7.19**	**5.46**
BAAFFM	**4.16**	-1.86	-1.12	**0.49**

Table 4: Relative RMSFE with respect to the random-walk model at forecast horizon $h = 4$, bold indicates a superior performance.

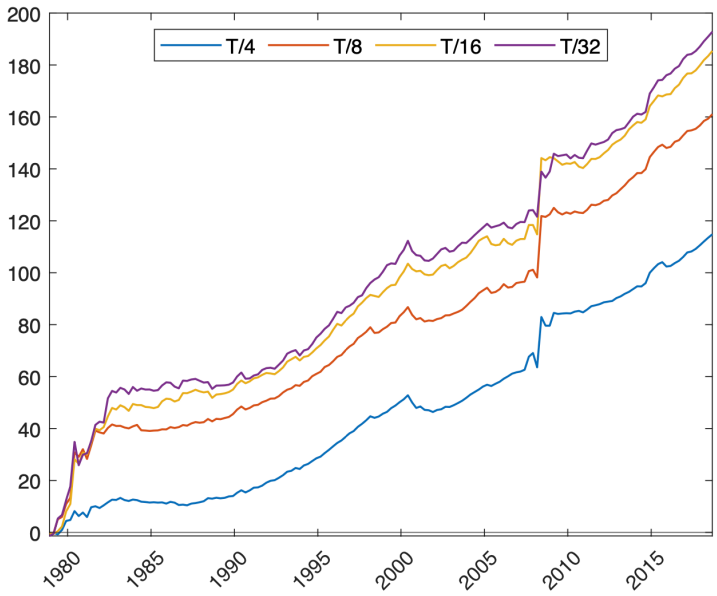


Figure 5: Relative Joint Average Log Predictive Likelihood at forecast horizon $h = 1$ with respect to the random-walk model.

dictive likelihood of the MTVP relative to RW_T: as it can be grasped the advantage is positive for all the spline specifications, but particularly marked for the $T/16$ and $T/32$ ones.

At forecast horizon $h = 4$, the relative alPLs of the MTVP worsen for many variables, similarly to the RMSE case: positive (and marked) gains are still detected for inflation, the interest rate, hours worked and 10-year Treasury bill, but the overall tendency leads to a deterioration of the overall prediction. Interestingly, the $T/16$ and $T/32$ retain significant advantages over the other spline approximations when the RW_T is outperformed, while $T/4$ and $T/8$ produce better results when RW_T wins.

In conclusion, the exercise suggests that time-variation in the parameters under investigation is not so pervasive to require a random-walk pattern,[10] and probably a smoother dynamics which slowly evolves over time is more appropriate. As a consequence, our MTVP model, which proposes a less parametrized specification, is a valid alternative to the random-walk pattern, *i.e.* superior or at least equal performances are detected in most cases for point forecast, while remarkable improvements are detected for the density forecast. Finally, the comparison has been performed considering hierarchical prior for the benchmark model too: this fact clearly reflects a more rigorous and fair analysis with respect to the literature benchmark. In Appendix D we report additional results with non-hiearchical priors.

[10]This finding is confirmed in a recent contribution by Chan (2023b).

	T/4	T/8	T/16	T/32
GDPDEF	**2.03**	**2.75**	**5.38**	**5.74**
GDPC1	**1.07**	**1.42**	**3.14**	**3.7**
FEDFUNDS	**51.24**	**63.8**	**60.68**	**60.34**
DPIC96	-0.03	**0.18**	**0.61**	**0.97**
PCECC96	**1.29**	**1**	**2.21**	**2.82**
CES0600000007	-0.93	**1.55**	**4.01**	**6.46**
CES0600000008	**1.78**	**2.31**	**3.81**	**2.24**
CUMFNS	**4.12**	**4.66**	**7.68**	**9.04**
PAYEMS	**3.07**	**3.07**	**5.28**	**7.02**
INDPRO	**4.42**	**4.57**	**6.61**	**7.62**
UNRATE	**2.26**	**5.64**	**9.17**	**8.76**
M2REAL	**1.69**	**2.75**	**2.36**	**1.24**
GS10	**4.67**	**8.42**	**14.01**	**14.42**
BAAFFM	**19.78**	**27.77**	**25.07**	**26.15**

Table 5: Average Log Predictive Likelihood (ALPL) at forecast horizon $h = 1$, bold indicates a superior performance with respect to the random-walk model.

4 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 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

	T/4	T/8	T/16	T/32
GDPDEF	**9.57**	**10.73**	**13.7**	**11.96**
GDPC1	-1.3	-1.02	-1.23	-3.18
FEDFUNDS	**15.4**	**17.88**	**15.89**	**17.9**
DPIC96	-1.05	-1.49	-1.29	-2.19
PCECC96	-0.97	-1.45	-1.51	-0.89
CES0600000007	-1.53	**2.85**	**7.09**	**12.24**
CES0600000008	**5.77**	**7.66**	**9.64**	**4.68**
CUMFNS	**7.1**	-0.59	-1.26	-22.3
PAYEMS	**1.86**	-0.48	-2.62	-5.44
INDPRO	-0.86	-1.73	-5.66	-7.1
UNRATE	**5.72**	**4.85**	**7.35**	**3.48**
M2REAL	-1.18	-2.59	-2.78	-4.28
GS10	**4.88**	**8.49**	**20.68**	**21.07**
BAAFFM	**3.43**	-0.76	-2.48	-5.29

Table 6: Average Log Predictive Likelihood (ALPL) at forecast horizon $h = 4$, bold indicates a superior performance with respect to the random-walk model.

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.

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

Banbura, M. and A. van Vlodrop (2018): "Forecasting with Bayesian vector autoregressions with time variation in the mean," Tech. rep., Tinbergen Institute Discussion Paper.

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

Carriero, A., T. E. Clark, and M. Marcellino (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

——— (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2024): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, 106, 1403–1417.

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

——— (2023a): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 235, 1419–1446.

——— (2023b): "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. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business & Economic Statistics*, 29, 327–341.

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.

DeBoor, C. (2000): *A Practical Guide to Splines*, Springer New York.

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

Eilers, P. H. and B. D. Marx (1996): "Flexible smoothing with B-splines and penalties," *Statistical science*, 11, 89–121.

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.

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

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.

Götz, T. B. and K. Hauzenberger (2021): "Large mixed-frequency VARs with a parsimonious time-varying parameter structure," *The Econometrics Journal*, 24, 442–461.

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

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

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

Appendix

A Description of the dataset

Variables	FRED-QD code	Transformation	Small (N=3)	Medium (N=7)	Large (N=14)
Real GDP	GDPC1	$400 \cdot \Delta \log$	x	x	x
GDP deflator	GDPDEF	$400 \cdot \Delta \log$	x	x	x
Federal funds rate	FEDFUNDS	Raw	x	x	x
Real income	DPIC96	$400 \cdot \Delta \log$		x	x
Real consumption	PCEC96	$400 \cdot \Delta \log$		x	x
Hours worked	CES0600000007	Raw		x	x
Hourly earnings	CES0600000008	$400 \cdot \Delta \log$		x	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

Table 7: Description of the Database

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 $\mathbf{s} = [s_1, \ \ldots, \ s_T]$,

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

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 $\mathbf{V}_{g_i} = \text{diag}(v_{g_{i,1}}, v_{g_i}\mathbf{I}_{R-1})$. 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 (13) and (14), the conditional posterior for $\bar{\xi}_1$ can be written as

$$\begin{aligned} p(\xi_1|\bullet) \propto & \xi_i^{2N\tau_1} \exp\left[-\xi_1(\tau_1+1)\Big(\sum_i s_i^{-1} + \sum_i v_{i,1}^{-1}\Big)\right] \\ & \times \xi_1^{-(\nu+1)} \exp\Big(-(\nu+1)\bar{\xi}_1\xi_1^{-1}\Big) \\ = & \xi_1^{2N\tau_1-\nu-1} \exp\left[-\xi_1(\tau_1+1)\Big(\sum_i s_i^{-1} + \sum_i v_{i,1}^{-1}\Big) - \xi_1^{-1}(\nu+1)\bar{\xi}_1\right] \end{aligned}$$

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.

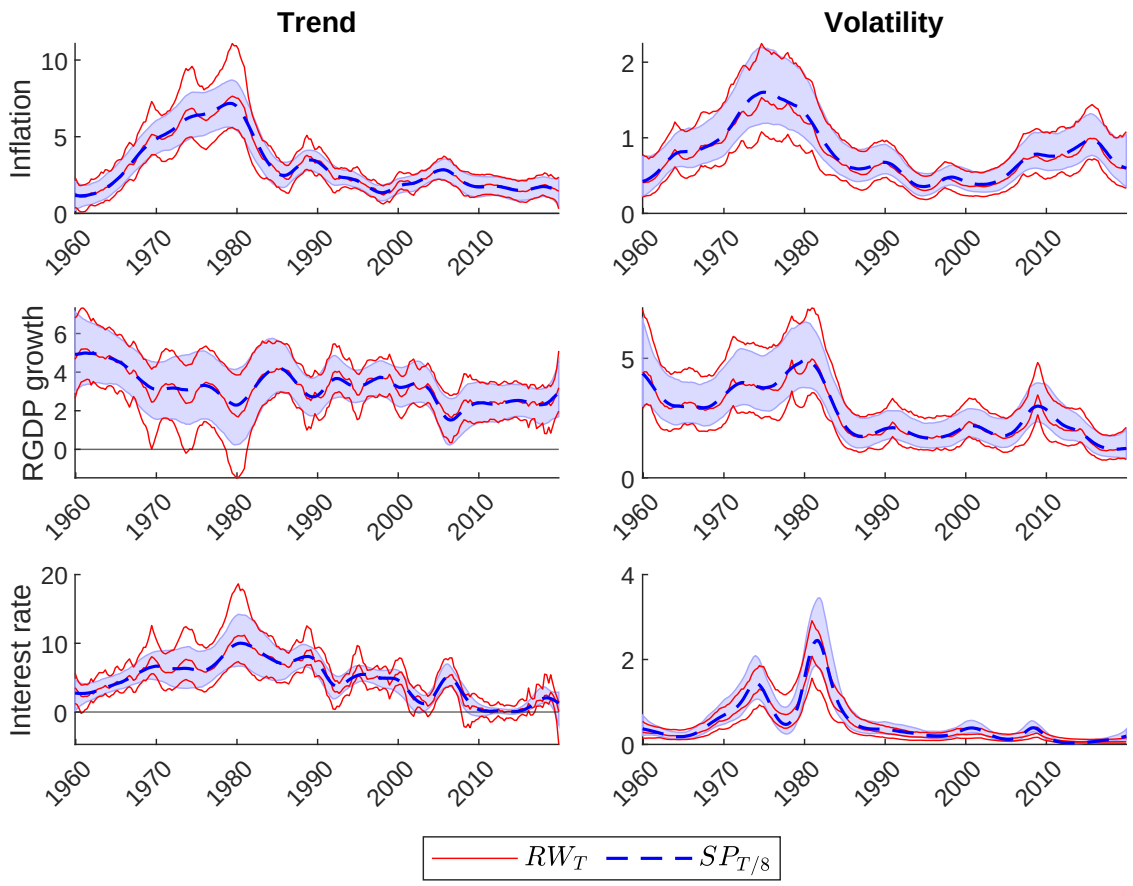
C Additional results

Table 8: **Posterior means and standard deviations of ξ_2.**

	RW_T	$\text{SP}_{T/4}$	$\text{SP}_{T/8}$	$\text{SP}_{T/16}$	$\text{SP}_{T/32}$
N=3	.771 (.144)	.854 (.164)	1.032 (.255)	1.316 (.382)	1.557 (.493)
N=7	.251 (.019)	.367 (.033)	.442 (.055)	.528 (.076)	.628 (.097)
RN=14	.130 (.003)	.230 (.013)	.282 (.018)	.326 (.031)	.382 (.033)

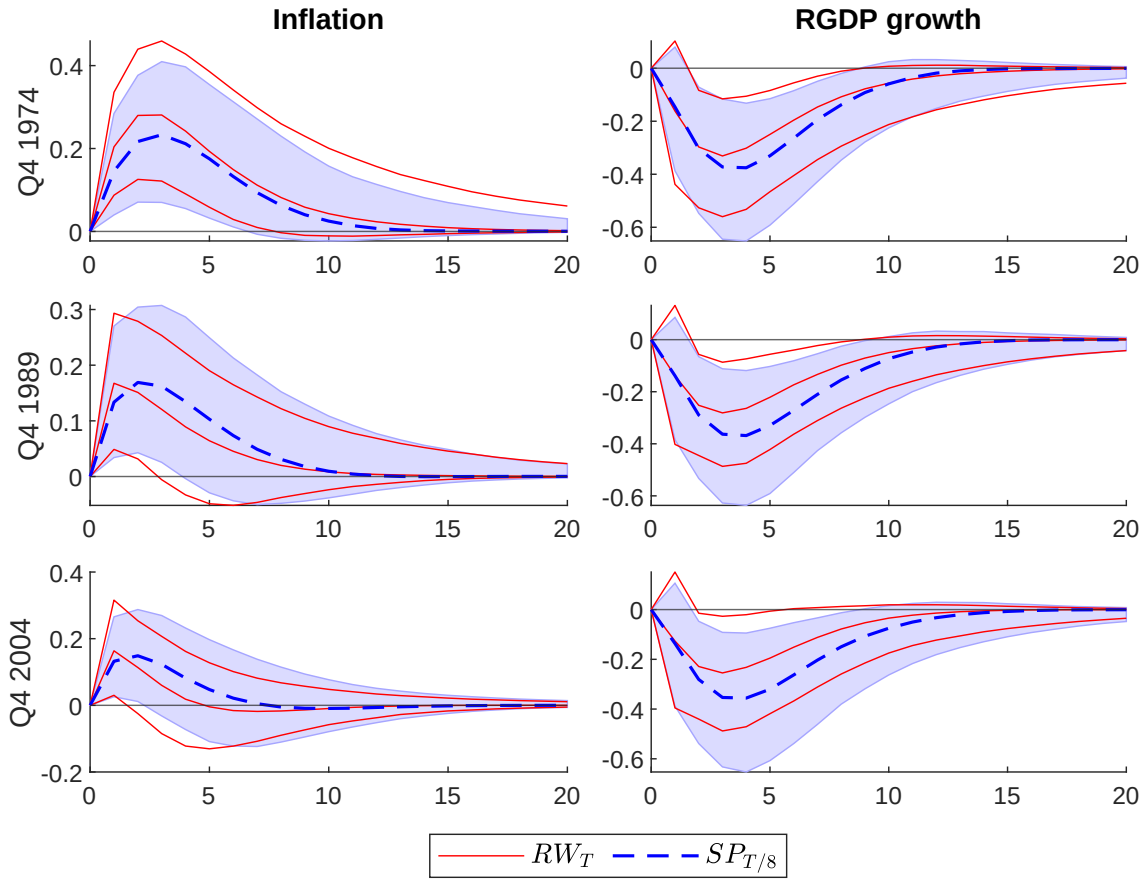
Note: The table shows the posterior means and standard deviations of ξ_1, across different dimensions and TVP-VAR specifications. Entries are multiplied by 100 to help interpretability.

Figure 6: **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 7: **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 Additional expanding-window results

	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 9: 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 10: 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 11: 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 12: 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.**