WORKING PAPER 16/2025 (ECONOMICS/STATISTICS)

Moderate Time-Varying Parameter VARs

Alessandro Celani and Luca Pedini

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Moderate Time-Varying Parameter VARs *

Alessandro Celani †
Örebro University

Luca Pedini ‡
Fondazione ENI Enrico Mattei

December 2, 2025

Abstract

This paper proposes a parsimonious reparametrization for time-varying parameter models that captures smooth dynamics through a low-dimensional state process combined with B-spline weights. We apply this framework to TVP-VARs, yielding Moderate TVP-VARs that retain the interpretability of standard specifications while mitigating overfitting. Monte Carlo evidence shows faster estimation, lower bias, and strong robustness to knot placement. In U.S. macroeconomic data, moderate specifications recover meaningful long-run movements, produce stable impulse responses and deliver superior density forecasts and predictive marginal likelihoods relative to conventional TVP-VARs, particularly in high-dimensional settings.

Keywords: Time-Varying Parameter models, High-dimensional Vector Autoregressions, Stochastic Volatility, B-splines, Macroeconomic Forecasting.
JEL classification: C11, C33, C53

*We wish to thank Sune Karlsson, Igor F. B. Martins, Filippo Ferroni, Chiara Casoli, Joshua Chan, Michele Piffer, Gergely Gánics, and Francesco Valentini for their comments, seminars and conference partecipants 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), and the Università di Bergamo Brown Bag Seminars.

†Örebro University School of Business, Fakultetsgatan 1, 702 81, Örebro, Sweden. Email: alessandro.celani@oru.se. Alessandro Celani receives support from the Jan Wallanders och Tom Hedelius stifelse samt Tore Browaldhs stifelse with grant number B22-0011.

‡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. Since the seminal works by Cogley and Sargent (2005) and Primiceri (2005), TVP-VARs have become a central tool for analyzing the relationship among macroeconomic variables over time. Their popularity stems from the flexibility to capture unstable economic dynamics, often featuring various forms of nonlinearities. From an applied perspective, they typically deliver superior forecast performance than their static counterpart (D'Agostino et al., 2013; Koop and Korobilis, 2013; Clark and Ravazzolo, 2015), and they also offer valuable insights for structural dynamic analysis (Benati and Surico, 2008; Baumeister and Peersman, 2013; Boeck and Mori, 2025).

Similarly to any TVP framework, in TVP-VARs the coefficient dimension grows with the length of the relevant time domain T, raising serious concerns in terms of over-parametrization. Bayesian methods are commonly introduced to mitigate the dimensionality issue. However, the computational burden often remains substantial, as these approaches rely on intensive Markov Chain Monte Carlo (MCMC) methods. An obvious consequence is that time variation is typically applied only to small-scale systems, with the potential shortcoming of omitted variable bias (Bańbura et al., 2010; Giannone et al., 2014). In high-dimensional settings, it is common to allow time variation only in a subset of coefficients, most often volatilities (Chan, 2023a; Carriero et al., 2024) and/or in intercepts (Banbura and van Vlodrop, 2018; Götz and Hauzenberger, 2021).

A different perspective motivates this paper. Temporal instabilities that economists typically face may be more limited than commonly assumed, with underlying dynamics evolving only gradually over time. In this context, a key question arises: are T coefficients truly needed to capture such evolution? Motivated by this question, we introduce a reparametrization of TVP within the state-space realm, with the aim of relaxing the assumption that parameters must evolve at the same frequency as the underlying dataset. Specifically, we linearly decompose the T-dimensional transition equation into two components: *i)* a lower-dimensional process that evolves at a lower frequency than the original data and *ii)* deterministic functions of time, namely B-spline bases, which weight the lower-frequency parameter and provide a mapping to the original temporal dimension. The magnitude of the parameter space becomes proportional to the number of *spline knots*.

The generality is high: the canonical TVP scheme used in the literature can be recovered by employing constant (0-th degree) splines with as many knots as time points. By allowing for fewer knots, we can naturally constrain the degree of temporal variability that the model can capture. However, simply reducing the number of knots is inadequate, as it generates undesirable stepwise dynamics in which parameters remain constant for some periods before abruptly jumping. This is where the spline degree becomes essential:

higher-degree splines produce smoother trajectories which naturally counter stepwise patterns and make knot placement less influential. For this reason, our baseline specification adopts quadratic (2-nd degree) splines, which align with the Random Walk dynamics commonly used in TVP models. Since the amount of time variation that can be represented decreases as the number of knots is reduced, we refer to the resulting specification as the Moderate TVP (MTVP).

The proposed idea can be applied to any state-space model with a linear transition equation, a requirement met by almost all models used in empirical macroeconomics. In this paper, we apply the approach in a TVP-VAR, which we refer to as Moderate TVP-VARs (MTVP-VARs). This setting is appealing because it allows us to assess the method's performance in both forecasting and structural applications, which are equally important.

The resulting estimation is performed via a Bayesian approach. 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 of 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.

If time variation is not pervasive, as we assert, the moderate approach represents an efficient dimensionality reduction technique for TVP-VARs, with benefits in both estimation bias and computational time. To investigate this, we run comprehensive simulation exercises analyzing performances of several moderate specifications. In particular, we compare computational time, parameter bias, and sensitivity to knot placement against the benchmark model (Primiceri, 2005). MTVP-VARs approximate smooth unstable patterns efficiently and for truly constant coefficients they systematically outperform the benchmark, as they are closer to a limiting static specification. Quadratic splines achieve superior fitting performances, while constant splines offer dramatic speed gains. Next, we bring the new models to real data by estimating MTVP-VARs of different dimensions on a quarterly U.S. macro dataset. We begin with a full-sample analysis to assess whether the estimated quantities are economically meaningful. A smooth moderate specification with parameters that vary every two years delivers estimated trends and structural impulse responses statistically indistinguishable from the benchmark. In a formal Bayesian model comparison exercise based on the predictive Bayes factor, we find decisive support for the smoothed MTVP-VAR in both the small-scale and large-scale settings. Finally, the forecasting exercise confirms this fact: all smoothed versions usually outperform the benchmark and the piecewise constant alternatives in terms of density forecasts, with particular reference to interest rate, industrial production and unemployment variables.

Literature review. Our paper contributes to the expanding literature on efficient di-

mensionality reduction techniques for TVP-VARs. Early Bayesian contributions include Koop and Korobilis (2013), who alleviate the computational burden of MCMC estimation through the use of forgetting factors. In a similar vein, Zheng et al. (2023) introduce a recursive, score-driven filtering framework that enables fast sequential estimation of time-varying parameters. Chan et al. (2020) take a different perspective by reducing the cross-sectional dimension of the coefficients via a factor decomposition. In contrast, our approach focuses on the temporal dimension, although the two strategies could be naturally combined. Korobilis (2021) employs factor-graph decompositions to develop efficient Bayesian algorithms for TVP regressions, although their method is not applicable under more classical Random Walk based transitions. Hauzenberger et al. (2022) achieve dimensionality reduction and computational efficiency by grouping the parameters into regimes through mixture models and by compressing the predictor matrix using the singular value decomposition. Instead, a large body of work addresses overfitting without directly tackling the computational issue: some examples include Belmonte et al. (2014); Bitto and Frühwirth-Schnatter (2019); Huber et al. (2021); Chan (2023b), among others. Finally, a separate strand of literature relies on non–likelihood-based methods. Notable contributions are the nonparametric kernel approaches of Giraitis et al. (2018); Kapetanios et al. (2019); Braun et al. (2025), with Petrova (2019) providing a quasi-Bayesian treatment.

It is also worth mentioning that this article contributes to the ongoing research on the use of penalized splines to capture arbitrarily complex patterns, a method originally introduced by Eilers and Marx (1996) and later extended to a Bayesian framework by Lang and Brezger (2004). Notice that the application of P-splines in Economics has recently gained popularity, notably through Barnichon and Brownlees (2019) and Tanaka (2020), who employ them to smooth Local Projections.

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 univariate TVP regression

For simplicity, we introduce our approach in the context of 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 and Wagner (2010), which decomposes parameters into a constant and a TV component, the latter with zero initial conditions. Rewrite the measurement equation as

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

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 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 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 we partition the

[1] 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 $\tilde{h}_1 = \cdots = \tilde{h}_T = 0$.

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 $\boldsymbol{\theta}_r$, $r = 1, \ldots, R$, 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 Random Walk:

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

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:

$$\tilde{\boldsymbol{\beta}}_t = \mathbf{W}_{t,1}\boldsymbol{\theta}_1 + \cdots + \mathbf{W}_{t,R}\boldsymbol{\theta}_R \tag{5}$$

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 a proper discussion on the weights is deferred to the next section, one feature is worth emphasizing. In practice, the weights $w_{t,r}$ can be set to zero to restrict the region of influence. For example, under the weighting scheme we will adopt, parameters effectively depend only on their own region and, at most, a few neighboring ones. Accordingly, the $T \times R$ matrix $\mathbf{W}$, with generic element $w_{t,r}$, is highly sparse (see Figure 1 for some examples).

Importantly, this general formulation nests the benchmark case and accommodates a range of intermediate specifications. In particular, with $R = T$, a weighting scheme such that $w_{i,j} = 1$ if $i = j$ and 0 otherwise, we recover the RW, since Eq. (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 $\tilde{h}_t$:

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

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.[2]

[2]Empirical evidence indicates that volatilities of macroeconomic variables often display pronounced

Compact form. Regardless of the specific form of the weights, the proposed decomposition preserves linearity in the parameters to be estimated. The multiplicative interaction between the weight matrices and the auxiliary parameters does not pose any difficulty, as the former are fixed prior to estimation. Nevertheless, the dependence of each parameter on potentially many auxiliary components complicates the construction of an estimation scheme based on filtering techniques (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). Yet, deriving the conditional posteriors becomes considerably more convenient if the model is expressed in a compact form. This representation enables the use of the efficient sampler of Chan and Jeliazkov (2009), which offers the additional advantage of being computationally more efficient than sequential estimation techniques. To make this explicit, we present the compact representation for both the mean and volatility parameters, and defer the detailed derivation of the conditional posteriors for the TVP-VAR case to Appendix B.

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 $\boldsymbol{\Sigma} = \text{diag}(\exp(\mathbf{h}))$, $\mathbf{h} = h + [\tilde{h}_1, ..., \tilde{h}_T]'$. Eq. (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 Eq. (9) into Eq. (7) shows that the model is linear in $\boldsymbol{\theta}$:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

where $\mathbf{L} = \tilde{\mathbf{X}}(\mathbf{W} \otimes \mathbf{I}_M)$. Combining this expression with the stacked prior for $\boldsymbol{\theta}$ makes it straightforward to show that the conditional posterior is Gaussian. Similar conclusions apply to the log volatility parameters. Specifically, the vector error term in Eq. (7) can be reformulated as a linear function of $\mathbf{g} = [g_1, \ldots, g_R]'$ as follows:

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

where each component of $\mathbf{u}$ follows a log Chi-Square distribution. By adopting the mixture representation of Kim et al. (1998) and combining it with the stacked Gaussian prior for $\mathbf{g}$, we obtain a Gaussian conditional posterior, from which we can sample using the precision sampler discussed above.

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.

2.2 Weight selection

The proposed method crucially depends on the choice of K and on the functional form of the weights $w_{t,r}$, which determine the size and structure of the matrix $\mathbf{W}$. For example, with the previously mentioned $0-1$ weights, the method yields stepwise dynamics characterized by regular abrupt jumps. This motivates the need for a smoother temporal evolution. Smooth weighting schemes can be generated using standard basis functions such as *B-splines*, *Fourier bases*, or *wavelets*.

In this paper we adopt 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.[3] For comparison purposes, we also consider constant ($q=0$) splines as a reference for stepwise movements. A general moderate smooth configuration is denoted as SP_K, and a piecewise one as PC_K.[4] 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. As we will see, 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 are less sensitive to knot placement, and generally achieve superior parameter fitting 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:

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

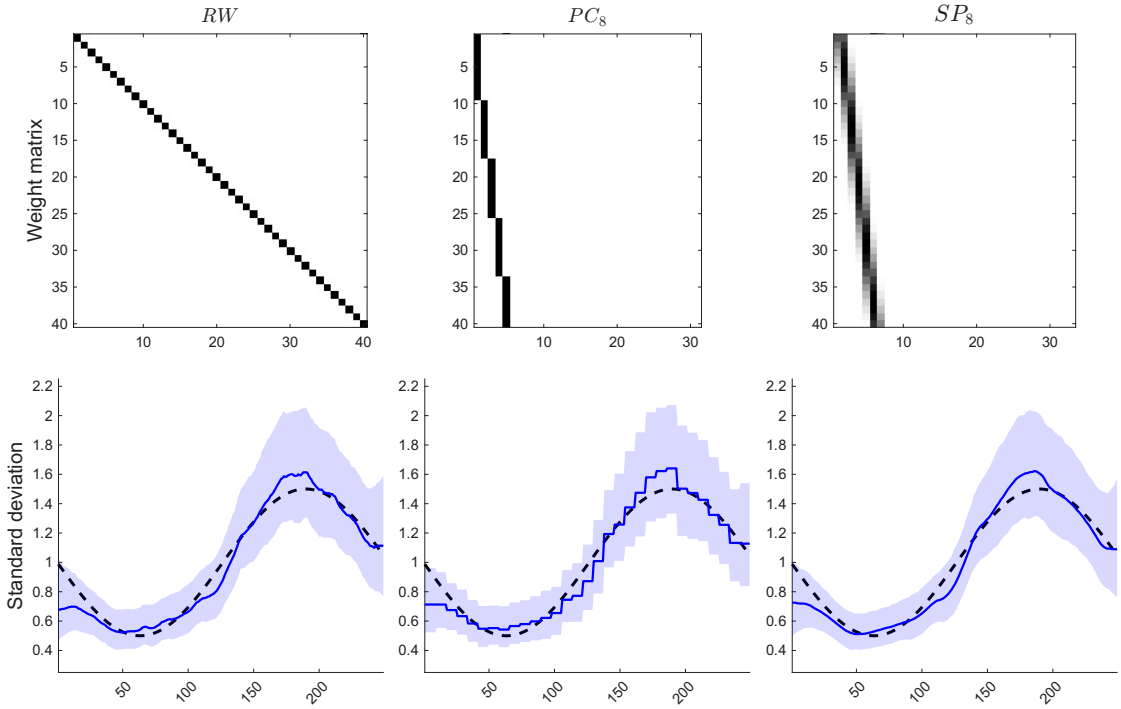
with *equidistant* knots starting at $1-qK$ [5] with spacing K. For each specification, the

[3] 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 of nonstationary variables, usually treated as $I(2)$ processes (Hodrick and Prescott, 1997). Moreover, raising the spline degree increases the number of non-zeros in the $\mathbf{W}$ matrix, which in turn worsens computational times.

[4] Note that, as previously discussed, the PC specification is a natural generalization of the RW, which is recovered by setting $K=1$, so that $RW = PC_1$. Indeed, the PC_K dynamics can be interpreted as a "multi-period Random Walk" that evolves at a lower frequency than the data.

[5] Placing the first knot at this point in the negative domain ensures that the corresponding basis

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.

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 consider the following models:

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

To provide the reader an idea on how the machinery works, we briefly present a result from the simulation exercise of Section 4. In particular, we show the fit of the *RW*, PC_8 and SP_8 on the TV standard deviation of the residual from 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, respectively, 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

function peaks at the first observation.

well in capturing the underlying process. Nonetheless, the spline degree is influential: the third column is visually preferable, as it avoids the irregular pattern of the second.[6]

3 Moderate TVP-VAR

We now introduce a class of models we call MTVP-VARs, *i.e.*, TVP-VARs in which the moderate 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 (2023b), 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 $\mathbf{\Sigma}_t = \operatorname{diag}(\exp(h_{1,t}), \ldots, \exp(h_{N,t}))$. 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, $\mathbf{Q}_i = \operatorname{diag}(q_{i,1}, \ldots, q_{i,M_i})$, with elements following Inverse Gamma distributions.

[6] Figure 7 in the Appendix C shows an analogous result for the four-year interval, corresponding to $K = 16$.

The same holds for the state variance of the log volatilities:

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

where τ_1 and τ_2 are prior shapes, while $\xi_{1,j}$ and ξ_2 denote prior modes. To distinguish between intercepts and lagged 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 mean 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 is motivated by having conditional analytical posterior for this coefficient:

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

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 a common expected total variance for the auxiliary processes (through modes):

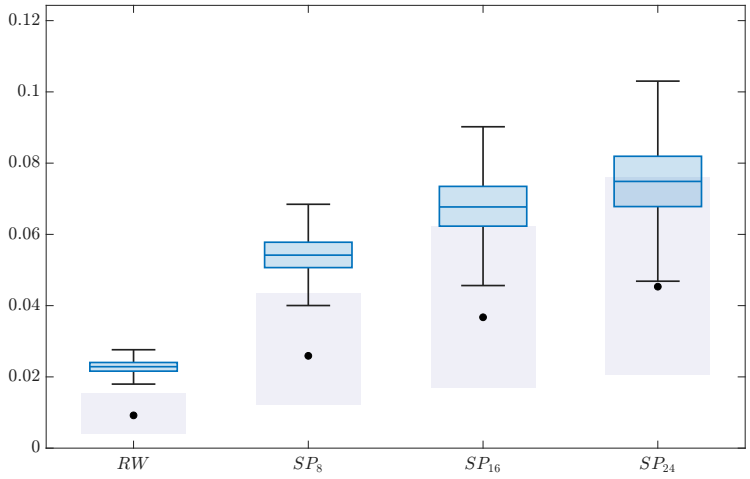
$$\begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} = K^{-1} \begin{bmatrix} 40 \\ 10 \end{bmatrix}, \quad \begin{bmatrix} \gamma \\ \xi_2 \end{bmatrix} = K \begin{bmatrix} 0.01^2 \\ 0.1^2 \end{bmatrix}$$

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 $y_{i,t}$, $\lambda_3 = 0.2^2$, $\lambda_4 = 10^2$ 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:

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.

$\lambda_1 \sim \mathcal{G}(2, 0.2^2/2)$, $\lambda_2 \sim \mathcal{G}(2, 0.1^2/2)$.

To finalize the estimation step we define the vectorized version of the TV parameters of interest $\boldsymbol{\theta}_i = [\boldsymbol{\theta}'_{i,1}, \ldots, \boldsymbol{\theta}'_{i,R}]'$, and $\mathbf{g}_i = [g_{i,1}, \ldots, g_{i,R}]'$. Bayesian estimation is performed using a Gibbs sampler that iterates over the following conditional posterior distributions: $p(\boldsymbol{\beta}_i|\bullet)$, $p(\boldsymbol{\theta}_i|\bullet)$, $p(h_i|\bullet)$, $p(\mathbf{g}_i|\bullet)$, $p(s_i|\bullet)$, $i = 1, \ldots, N$, $p(q_{i,j}|\bullet)$, $i = 1, \ldots, N$, $j = 1 \ldots, K_i$, $p(\bar{\xi}_1|\bullet)$, $p(\lambda_1|\bullet)$, $p(\lambda_2|\bullet)$, 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 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 and sampler outlined in Section 3.

We divide the Monte Carlo exercise into three sub-experiments: firstly, we evaluate

the importance of the hierarchical prior with adaptive hyperparameters 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 $\bar{\xi}_1 = 0.01^2$ (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 examine the computational performances 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 length of the regions increases, 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 setting ($N = 50$), the *SP* specification is roughly two to more than three times faster than the benchmark *RW* model. The *PC* variants achieve larger gains, ranging from nearly eight to almost fifteen 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. 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 $S = 200$ small ($N = 3$) 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

Table 1: **Computational time of MTVP in different dimensions**

	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: For each dimension N (across rows), the RW column reports the runtime in seconds to obtain 1000 posterior draws for the benchmark. The remaining entries display the percentage time savings achieved by each MTVP specification relative to it.

simulation $s = 1, \dots, 200$.[7] 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% 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,

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

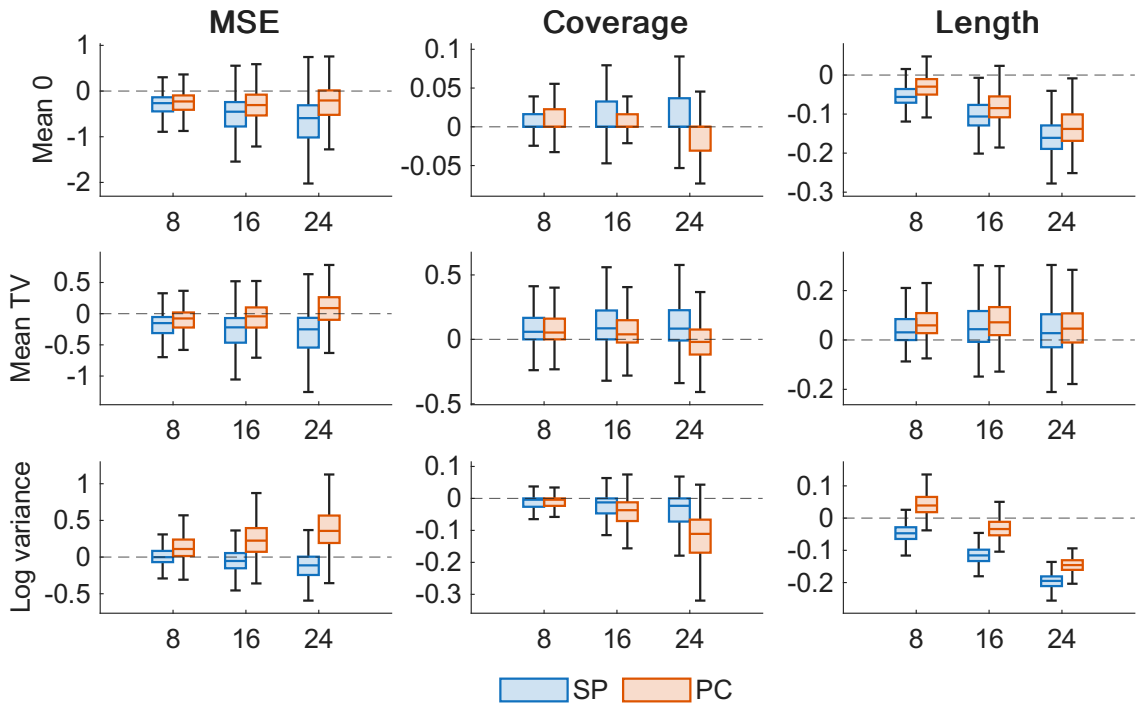
For each metric, we report results relative to the benchmark in log differences.[8]

Figure 3 reports the results in the form of boxplots across replications. Parameters are grouped by similarity across rows and the related metrics averaged by groups: the time-varying and zero entries of $\mathbf{\Phi}_{1,t}$ are displayed in the first two rows; TV log variances in the last row.

Starting with MSE, the SP specifications outperform the benchmark across all coefficient classes. The gains are consistent across the number of regions, not only for the coefficients that are constant over time (as one would expect) but also for the TV ones. For the log variances, performance improves as dimensionality decreases, with SP_{24} delivering the best results in this class. By contrast, SP_8 and SP_{16} yield estimates that are statistically equivalent to the benchmark. For coverage and length, results are mixed.

[7] For estimated parameters and resulting criteria we avoid the s subscript for ease of exposition.
[8] For example, the relative MSE is defined as $\log(\text{MSE}_m / \text{MSE}_{RW})$.

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: zero mean coefficients (first row), TV 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.

For 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 TV mean 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 the performance deteriorates markedly at higher K, 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. 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 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 with equidistant knots, each moderate specification allows up to K distinct knot locations over time; depending

Table 2: **Monte Carlo knot positioning exercise**

	SP_8	PC_8	SP_{16}	PC_{16}	SP_{24}	PC_{24}
Mean 0	0.008	0.029	0.011	0.053	0.014	0.007
	(0.006)	(0.018)	(0.008)	(0.034)	(0.011)	(0.045)
Mean TV	0.003	0.012	0.004	0.019	0.005	0.023
	(0.001)	(0.002)	(0.001)	(0.004)	(0.001)	(0.005)
Log variance	0.014	0.072	0.018	0.120	0.025	0.151
	(0.003)	(0.013)	(0.004)	(0.020)	(0.007)	(0.024)

Note: The table reports the results of the Monte Carlo based knot placement exercise. Columns report the Monte Carlo mean (with standard deviations in parentheses) of the square root of the variance component in Eq. (15) for the various MTVP specifications.

on how much time variation the data truly exhibit, this can be manageable or problematic. Our claim is that whenever time variation is not pervasive, the knot-placement issue is likely second-order relative to the gains.

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

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

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):

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

where the first term is the MSE between the true coefficient and the posterior mean averaged over all knot placements for each model.[9] As shown in Figure 8 in Appendix C, this component is effectively constant across specifications. Intuitively, averaging over

[9] Eq. (15) simplifies 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 $\sum_k \hat{\beta}_{k,m,t} - K\bar{\hat{\beta}}_{m,t} = 0 \; \forall t$.

placements neutralizes positioning effects, making the result equivalent to estimating *i)* a single specification with $K = 1$ or *ii)* the 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.

Table 2 reports the results, organized by rows according to the coefficient class. Columns report the averages of the Monte Carlo empirical distribution of the square root of the variance component in Eq. (15).

As expected, the magnitude of the estimated variability mirrors the underlying variability of the true coefficients: it remains modest for the mean coefficients, while it becomes substantially larger for the log-variances, which exhibit the strongest time variation. The SP specifications display markedly smaller magnitudes than the PC alternatives, with differences of two to six times. Importantly, the smooth specifications also remain stable across different numbers of regions: even when selecting only one coefficient every 24 observations, the SP produces, on average, less than twice the error associated with more parametrized specifications.

Summing up, while the proposed moderate approach inherently entails a knot selection variability, this has very limited practical implications for both non-TV and TV coefficients when smooth weighting schemes are used. In contrast, with stepwise dynamics the issue becomes pronounced, requiring estimation across multiple combinations and thereby reducing efficiency.

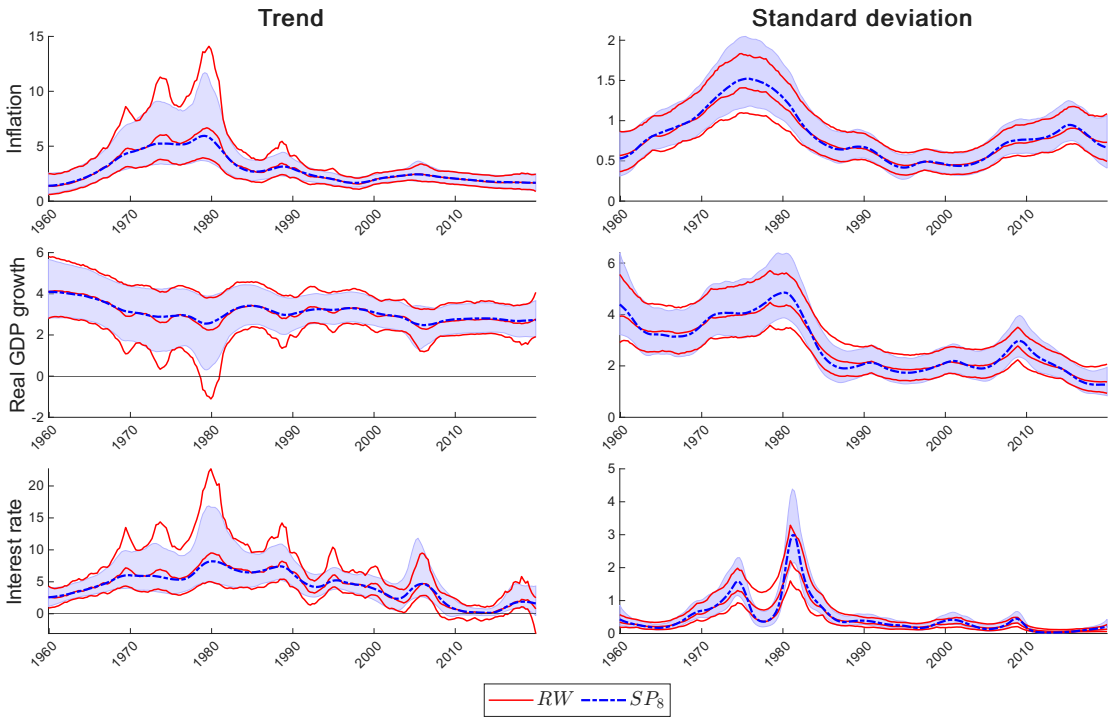
5 Empirical application

We now assess the performance of the proposed approach on real data. 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.

In all empirical applications of the MTVP-VAR, we fix $K = 8$ for the variance component, while the conditional mean parameters are allowed to follow larger values of K. Specifically, all SPs and PCs with $K = 16, 24$ apply exactly these values to the mean component, whereas for the variance we fix it at $K = 8$. Naturally, the RW and $K = 8$ models are unaffected by this choice. We adopt this restriction because SV is known to be the most TV part in the parameter space and from our experiments it emerges that setting $K > 8$ for the variance oversmooths the volatility dynamics.

Data. The dataset comprises 14 U.S. quarterly series from 1959Q1 to 2019Q4, obtained from the FRED-QD database at the Federal Reserve Bank of St. Louis. In

Figure 4: **Trends and standard deviations** (SP_8)



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

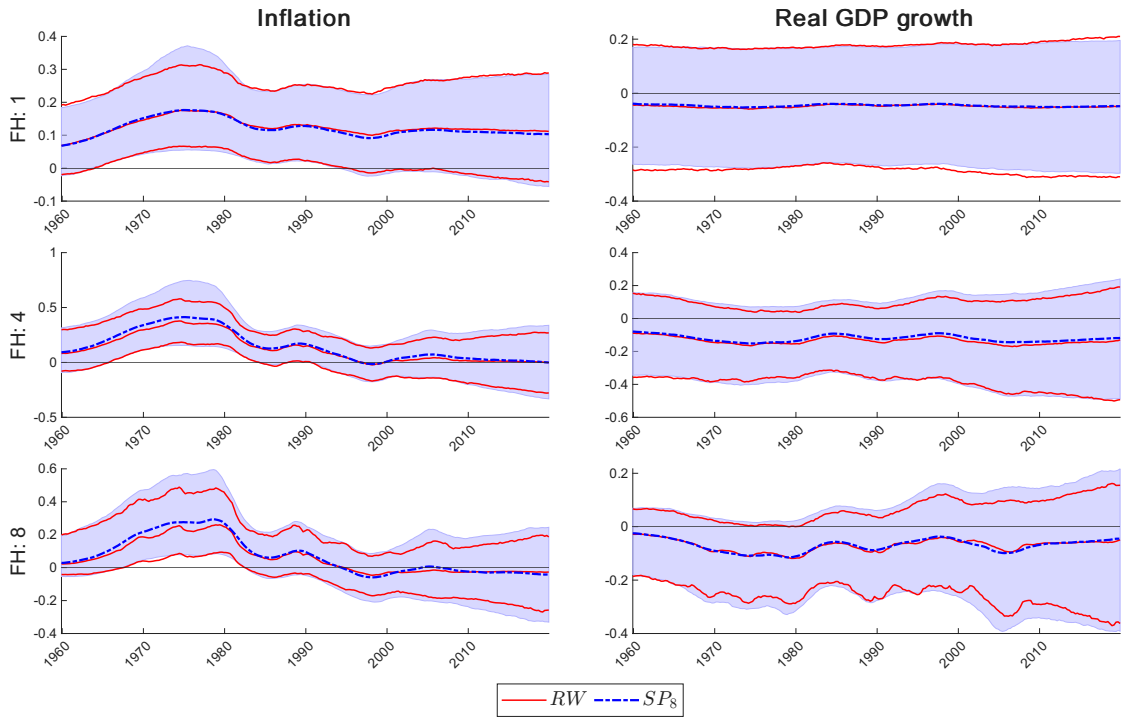
the empirical analysis, we consider two VAR specifications: a small-scale model with $N = 3$ variables and a larger one with $N = 14$. The small-scale VAR is the canonical tri-variate monetary policy system, including real GDP growth, GDP deflator inflation and the federal funds rate. Given the convenient dimension, we use this specification to illustrate both full-sample and forecasting results. The large-scale VAR, employed exclusively in the forecasting application, augments the small model with, among the others, consumption, investment, industrial production, unemployment and other labor-market and real-activity variables. Integrated series are transformed into stationary ones via annualized growth rate to ensure stable dynamics. The full list of variables and data transformations is reported in Table 5 in Appendix A.

5.1 Full sample results

An important aspect in our 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. To address this, in this section we summarize full-sample results of the moderate approach versus the benchmark model. For the sake of conciseness, we focus on the SP_8 case, while we defer the other moderate variants to the Appendix C.

We present the model estimated TV trends (or infinite horizon forecasts) and standard

Figure 5: **IRFs after a monetary policy shock (SP_8)**



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.

deviations in Figure 4.[10] At a general level, SP_8 delivers smoother dynamics compared to the more jagged ones of the RW. While the latter could, in principle, be smoothed by imposing tighter state-variance priors on the time-varying coefficients, our hierarchical specification rules out such ad-hoc tuning, since the amount of shrinkage is itself estimated.[11] The SP_8 effectively 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, as particularly evident in the trend estimates. Nevertheless, when the data demand it, the moderate specification is fully capable of capturing rapid swings in parameter dynamics. This is evident from the sharp movements of interest-rate volatility in the late 1970s, which SP_8 captures accurately. This demonstrates that a more parsimonious and smoother specification can still accommodate highly time-varying environments without incurring in 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

[10]Because trend estimation requires global stationarity, we incorporate an accept–reject step to discard explosive draws following Cogley and Sargent (2005).

[11]Figure 9 in Appendix C shows the estimated hierarchical shrinkages for the empirical application. The results underscore the importance of the adaptive mechanism also with real data, mirroring the simulation results.

recursive ordering, as in Primiceri (2005). TV IRFs are reported for horizons 1, 4, and 8 in Figure 5. This shock produces an inflationary response, while the effect on real GDP is statistically equal to zero. The two models deliver very similar results, particularly at short horizons, where the median IRFs are virtually identical. The moderate specification again exhibits smoother dynamics, efficiently summarizing the same patterns captured by the benchmark. At the same time, the SP_8 model displays wider credible intervals, especially at longer horizons, a consequence of its smoother trend estimates, which in turn permit more persistent cyclical fluctuations.

In summary, the full-sample experiment shows that the moderate SP_8 specification successfully reproduces the qualitative dynamics of the RW model while using far fewer parameters (an eight-fold reduction in this case). Although this result is strongest for SP_8, which is the least parsimonious moderate setup, the other moderate specifications also deliver robust and satisfactory outcomes. To illustrate this, we report the corresponding results for the PC_8 and SP_{16} in Figures 10–13 in Appendix C.

5.2 Forecasting exercise

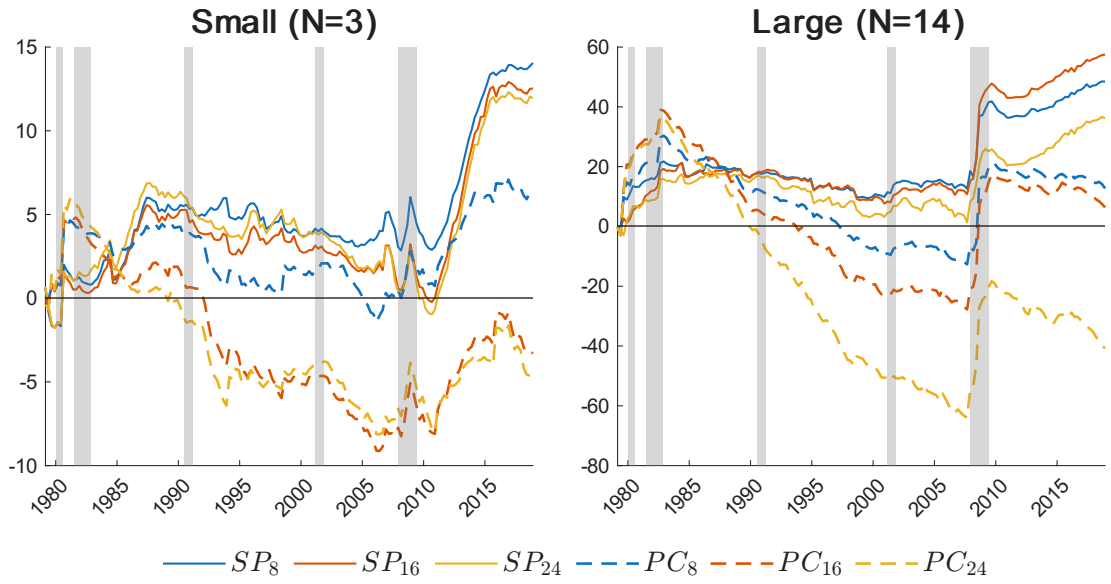
While full-sample analyses help validate the framework from an interpretational standpoint, evaluating forecasting performance is equally important, as it provides an additional form of model validation. Out-of-sample forecasting allows us to assess whether the gains from the induced parsimony of the proposed approach outweigh any advantage the more parametrized RW may have in capturing unstable features of the data. Accordingly, this section evaluates the MTVP approach in both a Small ($N = 3$) and Large ($N = 14$) TVP-VAR via a recursive forecasting exercise. Specifically, 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 perform model comparison via the predictive Marginal Likelihood (ML) and examine point and density forecast accuracy as well. Even in this case, we compare all moderate variants previously introduced to the benchmark RW.

Model comparison. 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

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

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, with

Figure 6: **Cumulative log predictive BF**



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

$t_0 = 1980Q1$. [12] Given the log ML, computing the resulting log predictive Bayes Factor (BF) between model m and the *RW* is straightforward

$$\log \text{BF}_m = \log \text{ML}_m - \log \text{ML}_{RW}$$

Figure 6 reports the evolution of log predictive Bayes factors over time for each model, in both the small and high-dimensional settings. To provide context to the plot, a $\log \text{BF}_m$ greater than ≈ 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 system, all *SP* specifications exhibit very similar dynamics, with significant results emerging from 1985 onward and becoming particularly pronounced toward the end of the sample ($\log \text{BF}_m > 10$), with SP_8 as the best performing model. Among the *PC* specifications, only the more parameterized one, namely PC_8 shows a positive and significant pattern. Interestingly, at the very beginning of the sample, all *PCs* displays positive $\log BF$, even superior to those of the *SP* counterpart. However, as the sample grows, the advantage is progressively lost for PC_{16} and PC_{24}, while PC_8 more than offsets this tendency. The *SP* paths, instead, appear more stable across time with the only exception of the sudden rise after 2010. The same conclusions, though magnified, apply to the large VAR system. Values of $\log \text{BF}_m > 20$ appear from the beginning of the sample and exceed approximately 40 toward the end for all *SP* models. This time data

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

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

favor SP_{16}, which approaches a $\log \mathrm{BF}_m$ of 60 in the last window, with SP_8 immediately after. Among the PC specifications, both PC_8 and PC_{16} attain positive results at the end of the sample, even though markedly inferior to the SP alternatives. Once again PCs exhibit an initial positive spike in BF immediately followed by a persistent decline up to the 2007 financial crisis, after which they partially recover. On the other hand, all SPs are more stable with positive $\log BF$ throughout the sample and a pronounced, steady increase after the financial crisis.

Point and density forecast. Point-forecast accuracy for variable i is evaluated using the Mean Squared Forecasting Error (MSFE), defined as:

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

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):

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

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.We consider two forecast horizons, $h = 1$ and $h = 4$. 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:

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

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

Table 3 displays the results for the small-scale setup. At forecast horizon $h = 1$, 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 out-of-sample forecasting the federal funds rate, and SP_{24} for GDP growth. The PC alternatives, on the other hand, show an overall deterioration relative to the RW, apart from the case of PC_{24} on the federal funds rate. At $h = 4$, 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 worsening in forecasting the interest rate, while perfor-

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

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
			Rel. MSFE, $h = 1$			
GDP deflator	**0.45**	0.21	−0.34	−1.25	−8.29	−11.22
Real GDP	−0.04	0.33	**1.31**	0.06	−0.77	−1.06
Federal funds rate	3.32	1.70	2.52	0.68	0.47	**3.51**
Average	**1.25**	0.75	1.16	−0.17	−2.87	−2.92
			Rel. MSFE, $h = 4$			
GDP deflator	2.13	6.38	**9.07**	−4.98	−9.17	−19.87
Real GDP	−0.04	0.94	−4.96	0.71	0.68	**1.03**
Federal funds rate	**4.68**	−1.77	−8.49	−3.95	−4.96	−5.55
Average	**2.26**	1.85	−1.46	−2.74	−4.48	−8.13
			Rel. ALPL, $h = 1$			
GDP deflator	**0.77**	0.45	−0.68	−0.09	−2.44*	−3.34*
Real GDP	0.39	0.13	**0.44**	0.30	0.00	−0.44
Federal funds rate	**8.79***	7.33	7.86	3.23	−0.46	−1.33*
Average	**3.32**	2.63	2.54	1.15	−0.97	−1.70
			Rel. ALPL, $h = 4$			
GDP deflator	**2.26**	1.38	0.45	−0.59	−4.57**	−8.50***
Real GDP	0.06	0.34	−0.08	2.14	**2.23**	1.85
Federal funds rate	**6.75**	1.13	−0.95	0.25	−4.58	−5.28
Average	**3.02**	0.95	−0.19	0.60	−2.31	−3.98

Note: Point and density forecast for the small setup ($N = 3$). 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; The symbols *, **, and *** indicate the three significance levels for the Amisano–Giacomini test (Amisano and Giacomini, 2007) for log predictive likelihoods with normal weights based on the center of the distribution.

mance 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 outperform both the benchmark and the PC alternatives. In particular, SP_8 shows a clear advantage in density forecasting inflation (a 2% gain at $h = 4$) and the interest rate variable (a 9% and 7% gain at $h = 1$ and $h = 4$, respectively, with the former statistically significant at size 0.10). Notice that, at $h = 1$, all SP models deliver qualitatively similar results. Among the PC specifications, only PC_8 stands out when considering the average performance. This fact seems to suggest that in a small-dimensional setup the temporal variability of parameters is high and a stepwise dynamics becomes clearly problematic when few knots are used. On the other hand, the quadratic splines of SP models tackle the issue entirely.

Table 4: **Point and density forecast ($N = 14$), selected series**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
Rel. MSFE, $h = 1$						
GDP deflator	1.48	4.98	**6.75**	−6.31	−5.38	−5.49*
Real GDP	0.07	**0.11**	−0.13	−0.37	−0.65	−0.54
Federal funds rate	0.61	1.05	1.28	0.24	−1.12	**2.36**
Industrial production	−1.03	−0.79	**2.09**	−5.94	−9.64*	−2.40
Unemployment rate	1.30	4.26	**6.90**	−2.73	−1.01	−1.49
Tot. Average	−0.35	0.66	**1.52**	−2.68	−3.07	−1.19
Rel. MSFE, $h = 4$						
GDP deflator	5.20	7.96	**14.65**	−17.59	−16.66	−22.48
Real GDP	**−2.51**	−3.17	−3.70	−3.44	−3.66	−3.84
Federal funds rate	1.08	−1.95	2.13	3.74	−2.86	**2.37**
Industrial production	−0.58	−0.92	−2.88	−1.13	0.35	**1.01**
Unemployment rate	4.98*	10.37*	**14.87***	−1.84	−2.87	−8.25
Tot. Average	0.74	1.30	**2.89**	−1.62	−1.78	−2.06
Rel. ALPL, $h = 1$						
GDP deflator	0.88	**1.38**	0.50	−1.40	−1.57	−2.77
Real GDP	**−0.11**	−0.18	−0.36	−0.36	−0.59	−0.40
Federal funds rate	**5.37**	4.88	3.80	−0.65	−1.81	−7.31**
Industrial production	1.56	1.69	**3.09**	−0.72	−3.32**	−1.33**
Unemployment rate	3.02**	4.69***	**5.32****	−3.45	−8.38**	−15.91***
Tot. Average	1.93	2.36	**2.46**	−0.21	−1.54	−3.37
Rel. ALPL, $h = 4$						
GDP deflator	2.00	**2.47**	1.00	−2.49*	−5.27***	−10.73***
Real GDP	**0.74**	0.55	0.11	0.51	0.66*	−0.07***
Federal funds rate	**2.58**	0.75	−1.97	0.54	−1.55	−1.19
Industrial production	5.37	5.75	4.47	5.88	**8.59**	7.66**
Unemployment rate	6.39	**17.77**	11.11	8.39	1.68	8.00
Tot. Average	5.12	**6.71**	6.20	6.51	4.65	3.64

Note: Point and density forecast large setup ($N = 14$), 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) for log predictive likelihoods with normal weights on the center of the distribution.

Moving on to the large-scale setup, Table 4 reports both the relative MSFE and ALPL for a selection of macroeconomic variables of interest, namely the previous three from the

small-dimensional setup, plus industrial production and unemployment rate.[14] Again, the SP specifications not only outperform the competitor PC, but also the benchmark for both forecast horizons. Specifically, the best model on average in terms of MSFE appears to be SP_{24} with remarkable results for the GDP deflator (percentage gains over the RW of 7% at $h = 1$ and 15% at $h = 4$, though not statistically significant) and the unemployment rate (relative gains of 7% and 15% at the two forecast horizons respectively, with statistical significance at the 0.10 level for $h = 4$). Notably, all SP models display substantial advantage in forecasting the unemployment rate at further horizons (all cases are statistically significant at 0.10).

Analogously to the $N = 3$ scenario, all SP models yield broadly similar conclusions for density forecast at $h = 1$. A net superiority with respect to the benchmark is exhibited for the federal funds rate (4-5% gains), the industrial production (2-3% gains) and the unemployment rate (3-5% gains, with statistically significance at 0.05 for SP_8 and SP_{24}; at 0.01 for SP_{16}). Consistently with the MSFE conclusions, SP_{24} has the best overall performance among the MTVP-VARs. At $h = 4$, the positive performance of the SPs is corroborated again. Coming to the individual predicted densities, SP_{16} is preferable for GDP deflator (2% gain) and the unemployment rate (18% advantage, though not statistically significant). The PCs are also able to recover positive performances at $h = 4$, especially for the GDP growth (statistically significant in case of PC_{16} at 0.10), industrial production (statistically significant at 0.05 for PC_{24}) and unemployment rate. Following these results, we suspect that, at higher forecast horizon in a high-dimensional system, parameter time variation becomes less pronounced, favoring MTVP schemes, even with stepwise dynamics.

6 Conclusion

In this paper we have proposed a new way to model gradual time variation in TVP models by decoupling the evolution of coefficients from the frequency of the data. Rather than allowing parameters to vary at each period, our MTVP framework represents time variation through a low-dimensional process combined with B-spline weights. This reduces the number of states required to capture smooth dynamics and nests the standard benchmark TVP scheme as a special case.

Our simulations show that both piecewise constant and smooth variants deliver substantial computational gains relative to the benchmark, with the reduction becoming larger in higher dimensions. Smooth specifications also achieve lower estimation bias and remain remarkably robust to knot placement, confirming that when time variation is modest, the exact knot positions have little impact. Using US data, we show that mod-

[14] For brevity, we report only a subset of the results; the complete set of tables can be found in the Appendix E.

erate specifications produce economically meaningful estimates, making them a credible tool for structural analysis. Coming to forecast performances, MTVP-VARs are at least as competitive as the standard TVP-VAR in terms of point forecasting, but have a clear advantage in terms of density forecasting, especially in high-dimensional setup.

The proposed approach is not without limitations. When time variation is pervasive, the choice of the number of knots becomes critical, and using too few knots may introduce misspecification. A promising direction for future work is to endogenize both the number and placement of knots, for example through Reversible Jump MCMC methods (Green, 1995), which can accommodate model spaces of different dimensions. Moreover, the TVP framework adopted here assumes uncorrelated state innovations. This restriction can be relaxed by allowing for a full hierarchical covariance matrix, as in Amir-Ahmadi et al. (2020), or by using a factor decomposition, as in Korobilis (2022). Finally, the implicit variable ordering required by the benchmark TVP-VAR can be avoided by extending the moderate approach to alternative TVP covariance structures that are not subject to this limitation, such as those in Kastner (2019); Chan et al. (2024).

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.

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.

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.

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

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

Braun, R., G. Kapetanios, and M. Marcellino (2025): "Nonparametric Time Varying IV-SVARs: Estimation and Inference," *Review of Economics and Statistics*, 1–47.

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

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.

Carter, C. K. and R. Kohn (1994): "On Gibbs sampling for state space models," *Biometrika*, 81, 541–553.

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

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 (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.

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

Frühwirth-Schnatter, S. (1994): "Data augmentation and dynamic linear models," *Journal of time series analysis*, 15, 183–202.

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

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

Green, P. J. (1995): "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Hauzenberger, N., F. Huber, G. Koop, and L. Onorante (2022): "Fast and flexible Bayesian inference in time-varying parameter regression models," *Journal of Business & Economic Statistics*, 40, 1904–1918.

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

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

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.

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.

Lang, S. and A. Brezger (2004): "Bayesian P-splines," *Journal of computational and graphical statistics*, 13, 183–212.

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.

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

Zheng, T., S. Ye, and Y. Hong (2023): "Fast estimation of a large TVP-VAR model with score-driven volatilities," *Journal of Economic Dynamics and Control*, 157, 104762.

Appendix

A Description of the dataset

Table 5: **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 Eq. (10) with the corresponding prior framework.

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

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\beta}_i + \tilde{\mathbf{Z}}_i\tilde{\boldsymbol{\beta}}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i) \tag{16}$$

  where $\mathbf{y}_i = [y_{i,1}, \ldots, y_{i,T}]'$, $\mathbf{Z}_i = [\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,T}]'$, $\tilde{\mathbf{Z}}_i = \text{diag}(\mathbf{z}'_{i,1}, \ldots, \mathbf{z}'_{i,T})$, $\boldsymbol{\beta}_i = [\boldsymbol{\beta}'_{i,1}, \ldots, \boldsymbol{\beta}_{i,T}]'$, $\boldsymbol{\epsilon}_i = [\epsilon_{i,1}, \ldots, \epsilon_{i,T}]$', and $\boldsymbol{\Sigma}_i = \text{diag}[\exp(h_{i,1}), \ldots, \exp(h_{i,T})]$.

  1. Using standard linear regression results, it can be shown that the conditional posterior of $\boldsymbol{\beta}_i$ is Gaussian

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

     with mean and covariance:

$$\mathbf{B}_{\boldsymbol{\beta}_i} = \bar{\mathbf{Q}}_i^{-1} + \mathbf{Z}_i'\boldsymbol{\Sigma}_i\mathbf{Z}_i, \quad \hat{\boldsymbol{\theta}}_i = \mathbf{B}_{\beta_i}^{-1}\left[\mathbf{Z}_i'\boldsymbol{\Sigma}_i(\mathbf{y}_i - \tilde{\mathbf{Z}}_i\tilde{\boldsymbol{\beta}}_i)\right]$$

  2. For $p(\boldsymbol{\theta}_i|\bullet)$, we rewrite Eq. (16) as a function of $\boldsymbol{\theta}_i$:

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

     where $\mathbf{L}_i = \tilde{\mathbf{Z}}_i(\mathbf{W}_i \otimes \mathbf{I}_{K_i})$. 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:

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

     where $\mathbf{V}_{\theta_i} = \text{diag}(\mathbf{I}_R \otimes \mathbf{Q}_i)$. It can be shown that

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

where

$$\mathbf{B}_{\theta_i} = \mathbf{H}'_{\theta_i}\mathbf{V}_{\theta_i}\mathbf{H}_{\theta_i} + \mathbf{L}'_i\boldsymbol{\Sigma}_i\mathbf{L}_i, \quad \hat{\boldsymbol{\theta}}_i = \mathbf{B}_{\theta_i}^{-1}\left[\mathbf{L}'_i\boldsymbol{\Sigma}_i(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\beta}_i)\right]$$

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 $p(h_i|\bullet)$ and $p(\mathbf{g}_i|\bullet)$, we express the error term of Eq. (16) as:

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

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

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

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 10 Gaussian distributions

$$\tilde{\mathbf{u}}^* \sim \pi_1\mathcal{N}(\mathbf{a}_1, \mathbf{C}_1) + \cdots + \pi_{10}\mathcal{N}(\mathbf{a}_{10}, \mathbf{C}_{10})$$

where $\mathbf{a}_j$, $\mathbf{C}_j$ are predetermined mixtures mean and variance components and π_j the mixture weight, , with $j = 1, \ldots, 10$. Given a draw of the auxiliary component density indicators $\zeta = [\zeta_1, \ \ldots, \ \zeta_T]$,

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

the model can be recast in linear and Gaussian form:

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

with $\mathbf{a}_i = [\mathbf{a}_{j,1}, \ldots, \mathbf{a}_{j,T}]'$ and $\mathbf{C}_i = \mathrm{diag}[\mathbf{C}_{j,1}, \ldots, \mathbf{C}_{j,T}]$.

1. The conditional posterior of h_i has a Gaussian form

$$p(h_i|\bullet) \sim \mathcal{N}(\hat{h}_i, b_{h_i}^{-1})$$

with mean and variance

$$b_{h_i} = 5^{-2} + \boldsymbol{\iota}'\mathbf{C}_i\boldsymbol{\iota}, \quad \hat{\mathbf{g}}_i = [\boldsymbol{\iota}'(\tilde{\boldsymbol{\epsilon}}_i|\zeta - \mathbf{a}_i - \mathbf{W}_i\mathbf{g}_i)]/b_{h_i}$$

where $\boldsymbol{\iota}$ is a conformable vector of ones.

2. To sample from $p(\mathbf{g}_i|\bullet)$, we introduce 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 $\mathbf{V}_{g_i} = \text{diag}(s_i\mathbf{I}_R)$. 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

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

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

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

$$q_{i,j}|\bullet \sim \mathcal{IG}\left(\tau_1 + \frac{R}{2}, (\tau_1+1)\xi_{1,j} + \frac{\sum_{r=1}^{R}\Delta\theta_{i,j,r}^2}{2}\right)$$

$$s_i|\bullet \sim \mathcal{IG}\left(\tau_2 + \frac{R}{2}, (\tau_2+1)\xi_2 + \frac{\sum_{r=1}^{R}\Delta g_{i,r}^2}{2}\right)$$

- In order to sample from $p(\bar{\xi}_1|\bullet)$, we first note that these parameters only appear in their respective priors, so

$$p(\bar{\xi}_1|\bullet) \propto \prod_{i=1}^{N} p(\mathbf{Q}_i|\bar{\boldsymbol{\xi}}_1)p(\bar{\boldsymbol{\xi}}_1)$$

After manipulating the expression and noticing that the mode $\bar{\boldsymbol{\xi}}_1$ is simply a rescaling of the rate parameter, we obtain,

$$\bar{\xi}_1|\bullet \sim \mathcal{G}\left(2 + \tau_1\sum_{i=1}^{N} M_i, \left[(1+\tau_1)\left(\frac{1}{\sum_i 4^{-2}q_{i1}\sum_{j>1}q_{ij}}\right) + \frac{2}{\gamma}\right]^{-1}\right)$$

where the Gamma parameters have to be intended as shape and scale.

- The conditional posteriors of λ_1 and λ_2 are Generalized Inverse Gaussians (Chan, 2021):

$$\lambda_1|\bullet \sim \mathcal{GIG}\left(2 - \frac{NP}{2}, \frac{2}{0.2^2}, \sum_{ij \in S_1} \frac{\beta_{ij}^2}{\tilde{q}_{ij}}\right)$$

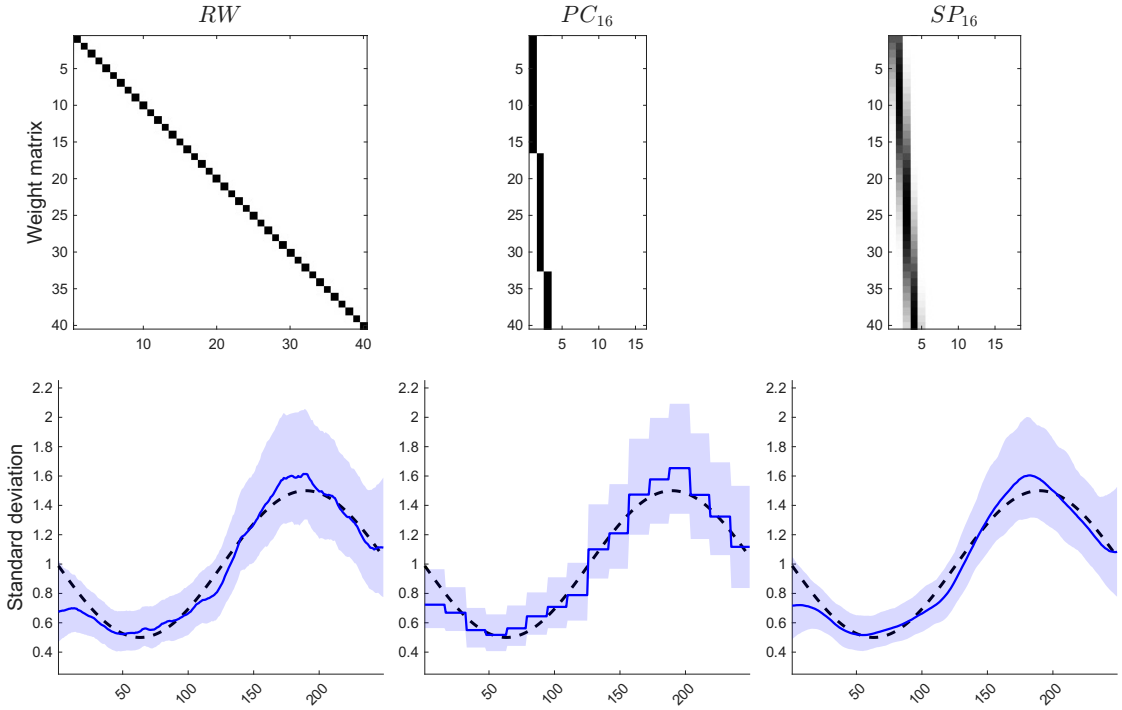
and similarly for λ_2:

$$\lambda_2|\bullet \sim \mathcal{GIG}\left(2 - \frac{(N-1)NP}{2}, \frac{2}{0.1^2}, \sum_{ij \in S_2} \frac{\beta_{ij}^2}{\tilde{q}_{ij}}\right)$$

where $\tilde{q}_{ij}$ denotes the Minnesota prior element $\bar{q}_{ij}$ without λ_1, λ_2, while S_1 and S_2 denote, respectively, the set of indices (i, j) associated with own lags and other coefficient lags

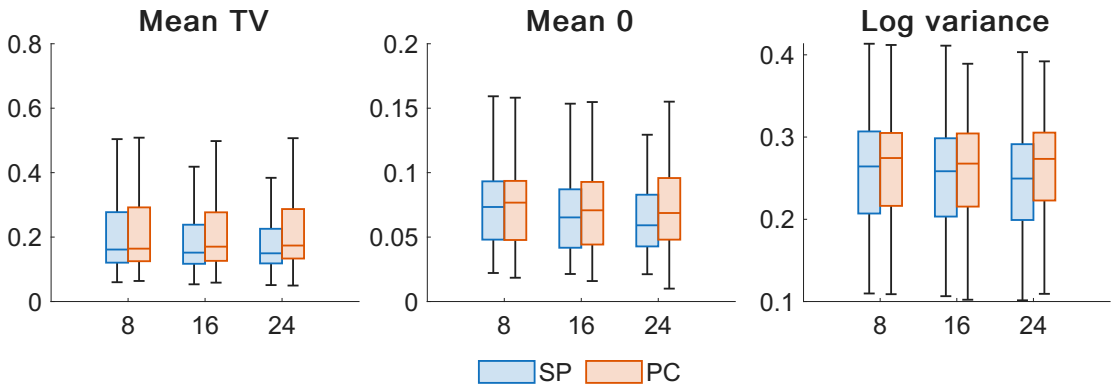
C Additional results

Figure 7: **The effect of different weighting schemes** ($K = 16$)



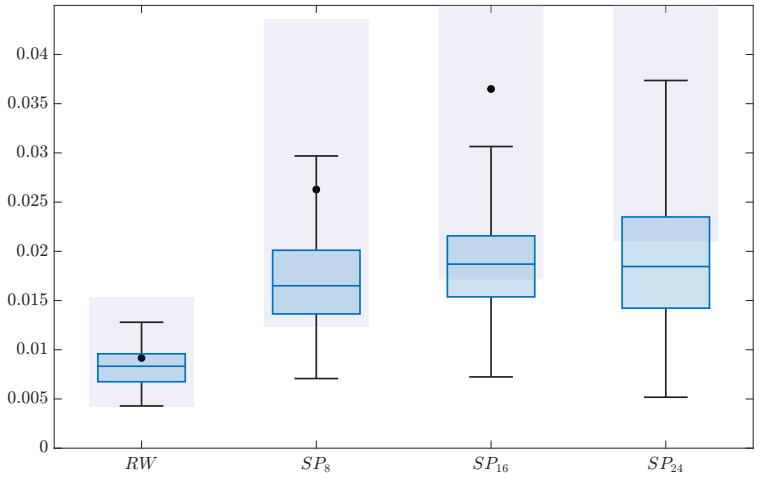
Note: The figure illustrates the effect of different moderate specifications ($K = 16$) on the estimation of the SV of the residuals for a generic variable in a single realization of a simulated TVP-VAR, as described in Section 4. The sample size is $T = 250$. *First row*: first 40 rows of the three 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 8: **Monte Carlo knot positioning exercise** ($Bias^2$)



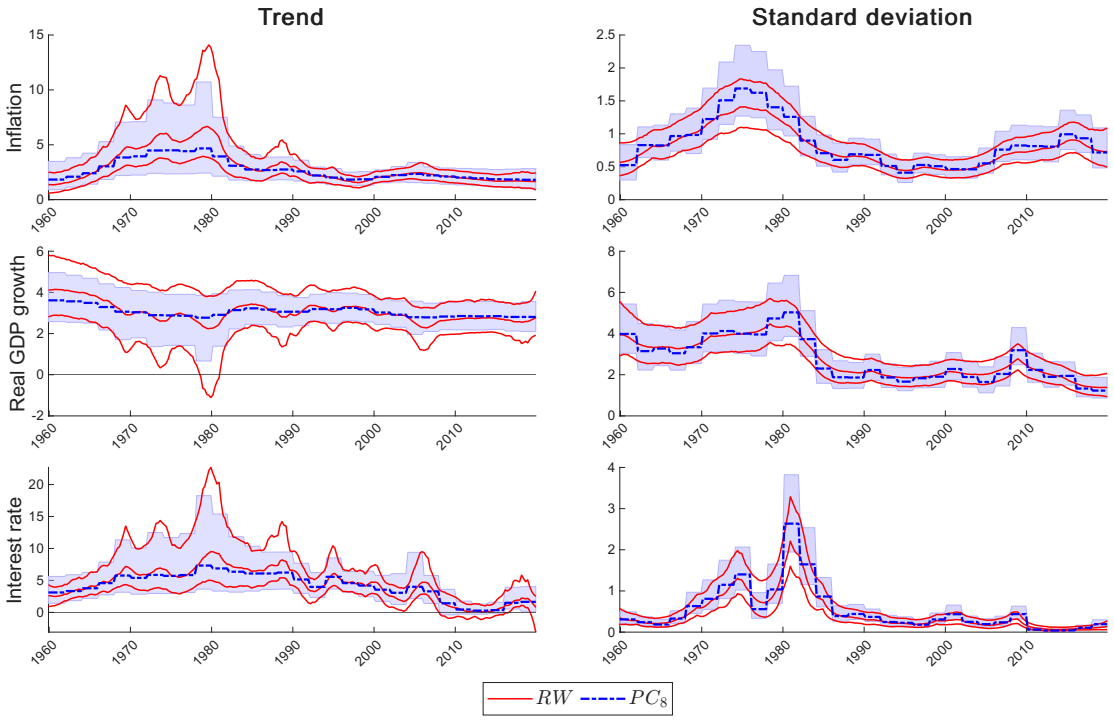
Note: The figure shows more results from the Monte Carlo knot positioning study. Each panel displays the $Bias^2$ component in Eq. (15) for the various MTVP specifications.

Figure 9: **Posterior of $\sqrt{\bar{\xi}_1}$ across specifications (real data).**



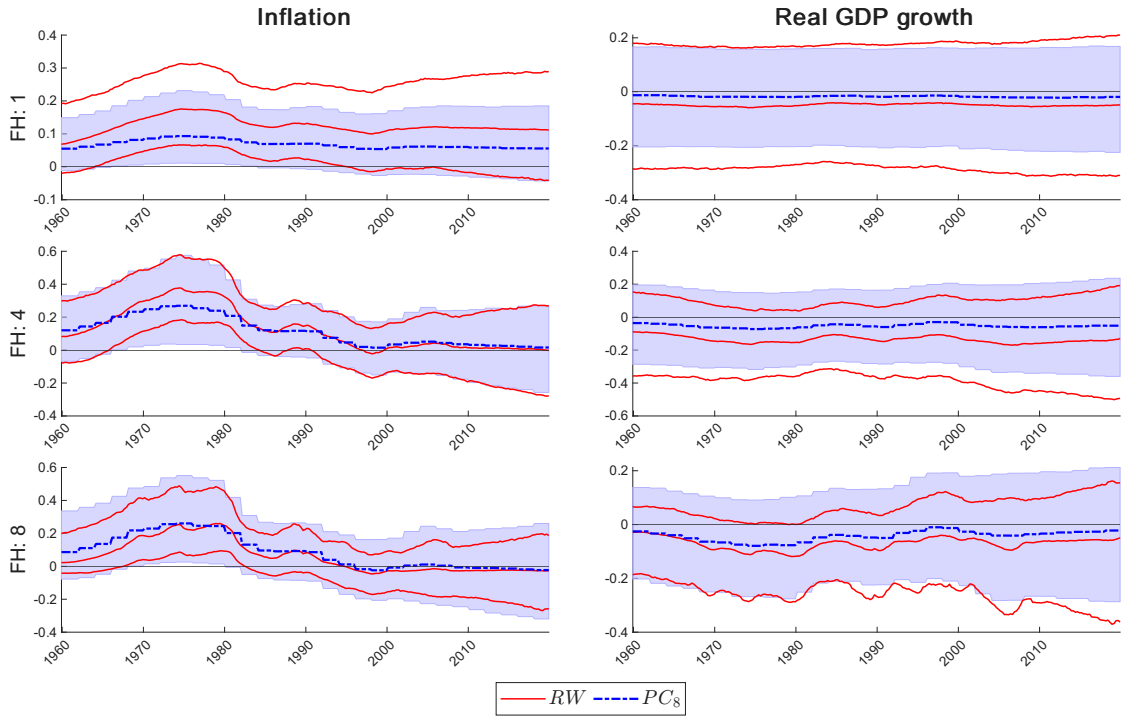
Note: The figure shows the posterior distribution of the hierarchical standard deviation $\sqrt{\bar{\xi}_1}$ for the *RW* and different moderate specifications *SP* in the small-dimensional full-sample application with real data. Shaded areas indicate the 90% prior probability bands, black dots mark the prior medians, and boxplots summarize the posterior draws.

Figure 10: **Trends and standard deviations** (PC_8)



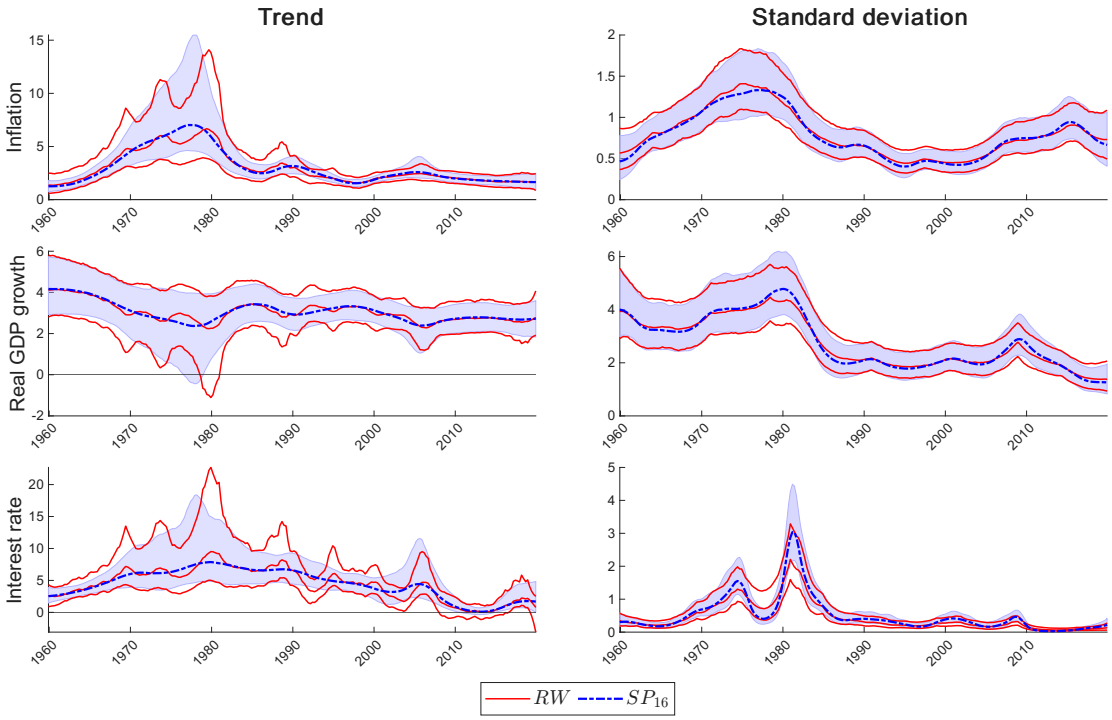
Note: The figure shows the posterior trends and standard deviation for two specifications: benchmark RW (red lines), and moderate PC_8 (blue dashed-dotted lines). External red lines and blue shaded-areas delimit the 90% credible bands for the two specifications, respectively.

Figure 11: **IRFs after a monetary policy shock (PC_8)**



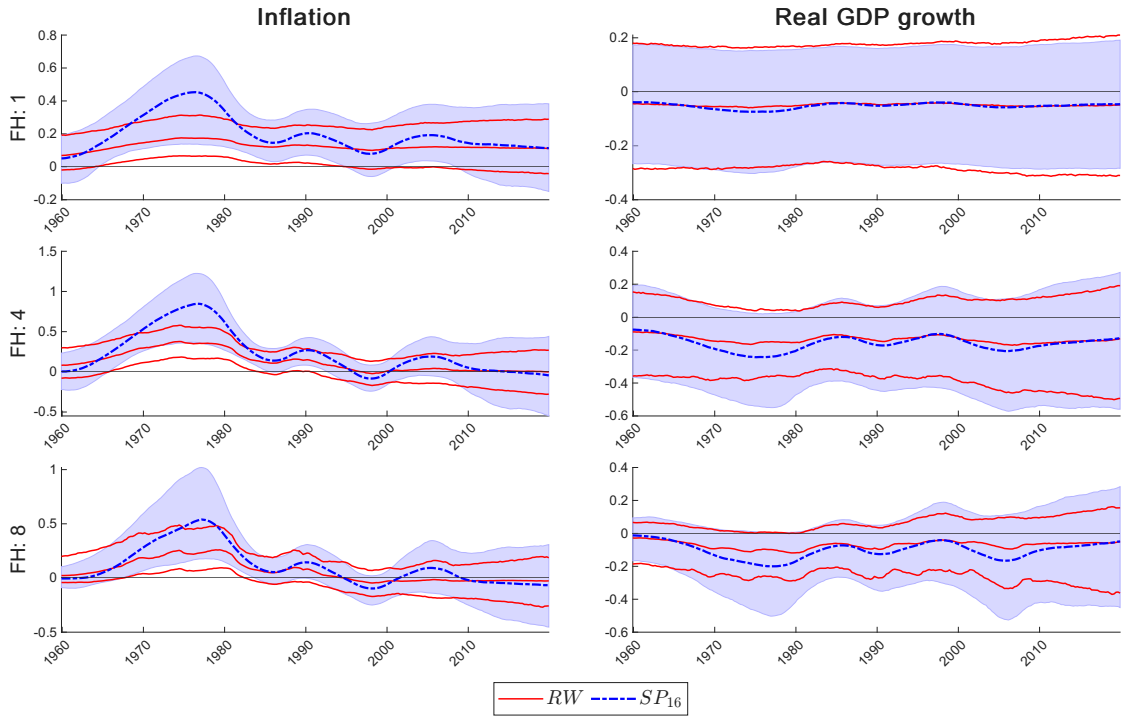
Note: The figure shows the posterior IRFs of Inflation and real GDP growth after a monetary policy shock for the two TVP-VAR specifications: benchmark RW (red lines), and moderate PC_8 (blue dashed dotted-lines). Rows reflect different forecast horizons: 1, 4, 8. External red lines and blue shaded-areas delimit the 90% credible bands.

Figure 12: **Trends and standard deviations** (SP_{16})



Note: The figure shows the posterior trends and standard deviation for two specifications: benchmark RW (red lines), and moderate SP_{16} (blue dashed-dotted lines). External red lines and blue shaded-areas delimit the 90% credible bands for the two specifications, respectively.

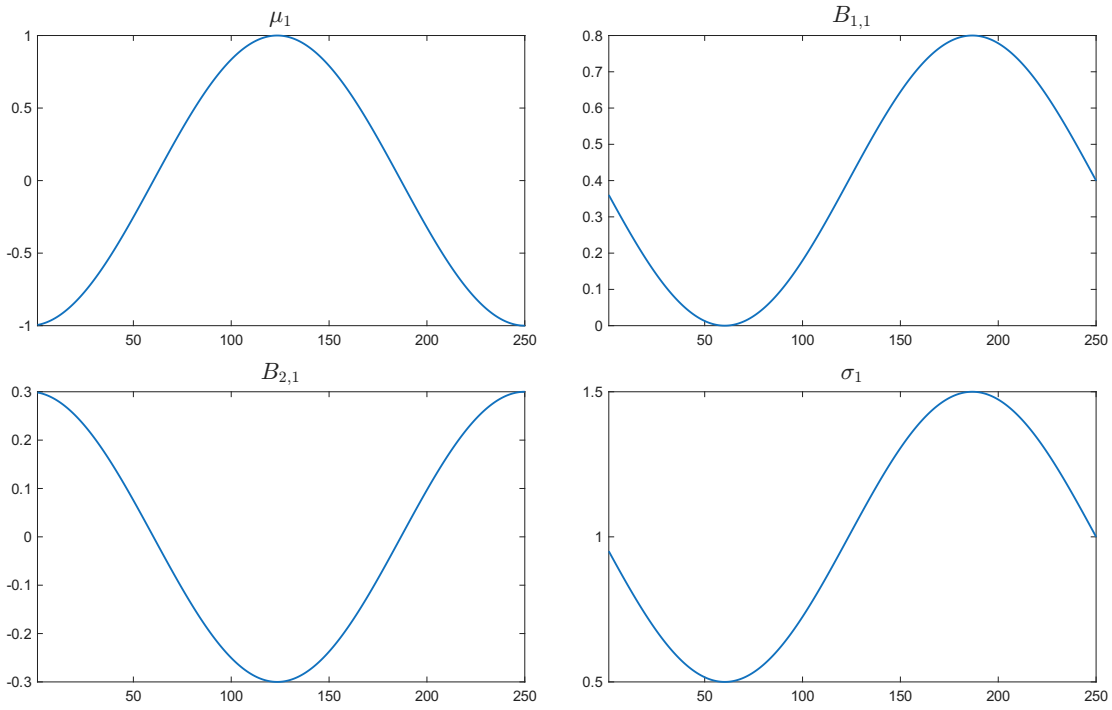
Figure 13: **IRFs after a monetary policy shock (SP_{16})**



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_{16} (blue dashed dotted-lines). Rows reflect different forecast horizons: 1, 4, 8. External red lines and blue shaded-areas delimit the 90% credible bands.

D Simulation details: Data Generating Process

Figure 14: **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

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

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

E Additional expanding-window results

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

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
			FH: 1			
GDP deflator	1.48	4.98	**6.75**	−6.31	−5.38	−5.49*
Real GDP	0.07	**0.11**	−0.13	−0.37	−0.65	−0.54
Federal funds rate	0.61	1.05	1.28	0.24	−1.12	**2.36**
Real income	−1.39*	−1.26*	**−1.07**	−3.34**	−3.48**	−3.26*
Real consumption	−0.69	−1.27**	**−0.53**	−2.88*	−3.52*	−3.73*
Hours worked	1.64	3.22	5.21	3.11	2.90	**8.04***
Hourly earnings	−0.26	1.69	**1.86**	−5.43**	−6.44**	−6.87***
Capacity utilization	−1.36	−2.41	**−0.94**	−2.08	−5.21	−2.37
Nonfarm payrolls	−1.02	−0.08	**2.07**	−5.10	−7.08*	−2.21
Industrial production	−1.03	−0.79	**2.09**	−5.94	−9.64*	−2.40
Unemployment rate	1.30	4.26	**6.90**	−2.73	−1.01	−1.49
Real M2 money stock	−0.45	**0.67**	−3.60	−2.33*	−1.69	0.55
10 year yield	1.54	2.09	2.04	0.33	**4.56**	2.41
BAA spread	−5.32	−3.06	**−0.59**	−4.67	−5.30	−1.68
Average	−0.35	0.66	**1.52**	−2.68	−3.07	−1.19
			FH: 4			
GDP deflator	5.20	7.96	**14.65**	−17.59	−16.66	−22.48
Real GDP	**−2.51**	−3.17	−3.70	−3.44	−3.66	−3.84
Federal funds rate	1.08	−1.95	2.13	3.74	−2.86	**2.37**
Real income	**0.12**	−0.21	−0.20	−0.48	−0.19	−0.60
Real consumption	**−0.91**	−1.22	−1.35	−4.00	−3.17	−3.58
Hours worked	3.45	4.64	7.22*	20.57***	**23.08*****	21.03**
Hourly earnings	−0.27	4.33	**9.67**	−19.26	−19.06**	−17.89*
Capacity utilization	−1.40	−2.01	−1.79	**−1.09**	−2.51	−2.09
Nonfarm payrolls	−1.50	−1.33	**−0.16**	−2.78	−3.26	−5.02
Industrial production	−0.58	−0.92	−2.88	−1.13	0.35	**1.01**
Unemployment rate	4.98*	10.37*	**14.87***	−1.84	−2.87	−8.25
Real M2 money stock	0.30	−0.31	**3.07**	−5.85	−1.06	−2.16
10 year yield	7.03**	6.52**	9.69**	2.91	**11.83**	3.22
BAA spread	−4.61	−4.54	−10.67	7.51	−4.94	**9.50**
Average	0.74	1.30	**2.89**	−1.62	−1.78	−2.06

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 7: **Density forecast - ALPL**

	SP_8	SP_{16}	SP_{24}	PC_8	PC_{16}	PC_{24}
			FH: 1			
GDP deflator	0.88	**1.38**	0.50	−1.40	−1.57	−2.77
Real GDP	**−0.11**	−0.18	−0.36	−0.36	−0.59	−0.40
Federal funds rate	**5.37**	4.88	3.80	−0.65	−1.81	−7.31**
Real income	0.49	0.65	0.55	1.78	1.70	**1.95**
Real consumption	**−0.07**	−0.41*	−0.16	−1.12**	−1.10**	−1.33***
Hours worked	2.97	3.78	**4.80***	1.07	−3.16	−6.65**
Hourly earnings	0.06	0.90	**0.91**	−2.88**	−3.65***	−5.69***
Capacity utilization	**1.55****	0.98	1.28	0.56	−2.02	−2.93**
Nonfarm payrolls	2.00**	2.33**	**3.20****	0.25	−1.93	−2.29**
Industrial production	1.56	1.69	**3.09**	−0.72	−3.32**	−1.33**
Unemployment rate	3.02**	4.69***	**5.32****	−3.45	−8.38**	−15.91***
Real M2 money stock	1.54	2.33	1.54	1.53	2.06	**2.15**
10 year yield	3.62	**4.19**	3.37	2.12	4.00	1.39**
BAA spread	4.15	5.86	**6.57**	0.28*	−1.83**	−6.12**
Average	1.93	2.36	**2.46**	−0.21	−1.54	−3.37
			FH: 4			
GDP deflator	2.00	**2.47**	1.00	−2.49*	−5.27***	−10.73***
Real GDP	**0.74**	0.55	0.11	0.51	0.66*	−0.07***
Federal funds rate	**2.58**	0.75	−1.97	0.54	−1.55	−1.19
Real income	0.34	0.14	0.52	**2.19**	1.79	1.55
Real consumption	**−1.29**	−1.60**	−1.36**	−2.42**	−1.99***	−2.86***
Hours worked	4.04*	8.65**	**10.18****	8.15***	2.49	−4.28
Hourly earnings	0.76	2.22	**2.39**	−5.60***	−7.86***	−10.67***
Capacity utilization	16.66	21.2	21.40	**35.76**	28.87*	28.04***
Nonfarm payrolls	6.18	8.50	**12.57**	10.55	9.30**	8.32***
Industrial production	5.37	5.75	4.47	5.88	**8.59**	7.66**
Unemployment rate	6.39	**17.77**	11.11	8.39	1.68	8.00
Real M2 money stock	**−1.01**	−1.14*	−2.14	−2.13	−1.33	−2.20*
10 year yield	7.61	7.60	9.27	8.28	**10.46**	6.08
BAA spread	21.28	20.92	19.22	**23.51**	19.29	23.35
Average	5.12	6.71	**6.20**	6.51	4.65	3.64

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.