General Bayesian time-varying parameter VARs for modeling government bond yields

Fischer, Manfred M.; Hauzenberger, Niko; Huber, Florian; Pfarrhofer, Michael

DOI:
10.57938/10106af8-e12a-422e-be86-c465799823aa

Published: 30/05/2022

Document Version
Publisher's PDF, also known as Version of record

General Bayesian time-varying parameter VARs for modeling government bond yields

Manfred M. Fischer*[1], Niko Hauzenberger[1,2]
Florian Huber[2], and Michael Pfarrhofer[2]

May 30, 2022

Abstract. US yield curve dynamics are subject to time-variation, but there is ambiguity about its precise form. This paper develops a vector autoregressive (VAR) model with time-varying parameters and stochastic volatility which treats the nature of parameter dynamics as unknown. Coefficients can evolve according to a random walk, a Markov switching process, observed predictors, or depend on a mixture of these. To decide which form is supported by the data and to carry out model selection, we adopt Bayesian shrinkage priors. Our framework is applied to model the US yield curve. We show that the model forecasts well, and focus on selected in-sample features to analyze determinants of structural breaks in US yield curve dynamics.

JEL: C11, C32, E43, E47
Keywords: Bayesian shrinkage, interest rate forecasting, latent effect modifiers, MCMC sampling

*Corresponding author: Manfred M. Fischer, Vienna University of Economics and Business, Welthandelsplatz 1, A-1020 Vienna, Austria. E-mail: manfred.fischer@wu.ac.at; phone: +43-676-8213-4836.
[1]Vienna University of Economics and Business, [2]University of Salzburg

1 Introduction

The US term structure of interest rates has been subject to structural breaks in the parameters governing its dynamic evolution (see, for example, Rudebusch et al., 2006).[1] For instance, the global financial crisis and the subsequent monetary policy actions triggered a substantial and rapid decline in short-term interest rates, with the policy rate hitting the zero lower bound (ZLB). At the ZLB, macro-financial dynamics suddenly change (Liu et al., 2019). This is reflected in appreciable jumps in the structural parameters which determine yield curve dynamics. Parameter movements of this kind are found in, for example, Rudebusch and Wu (2007), who document changing yield curve dynamics during the late 1980s.

A prominent approach for modeling parameter dynamics in macroeconomics and finance relies on time-varying parameter (TVP) regressions and VARs (Cogley and Sargent, 2005; Primiceri, 2005). Mumtaz and Surico (2009) use TVP-VARs to investigate the relationship between monetary policy shocks and the term structure of interest rates and find evidence in favor of time-variation in the parameters. Standard TVP models typically assume that the parameters evolve according to a simple stochastic process such as a random walk. While being rather flexible and parsimonious, this assumption does not allow for the examination of the extent to which covariates cause changes in the time-varying parameters. Moreover, wrongly assuming a random walk state equation could be deleterious for predictive accuracy because it implies smoothly evolving coefficients. This might be at odds with the rapid shifts that have been observed in the US yield curve (especially on the short end) during the global financial crisis or the Covid-19 pandemic.

A standard approach used to select the state dynamics of the TVPs is based on estimating different models separately and then using model selection criteria to discriminate between competing specifications (see, for example, Sims and Zha, 2006; Koop et al., 2009; Hauzenberger, 2021). This approach has the limitation that only a single distinct model

[1] Reaching the zero lower bound in response to economic slack is one example of a structural break. An overarching, secular change triggered by, for example, movements in the age composition of the population or changing regulations in financial markets, might be another example. Another potential source of time-variation are interaction effects of the structural parameters of the yield curve and observed, exogenous variables. These variables do not necessarily exhibit a direct effect on the yields but impact the yield curve indirectly through their influence on the parameters governing yield curve dynamics.

is selected in the end. But, for financial variables such as bond yields, combinations of different assumptions on parameter dynamics might be more appropriate. Our aim is to develop a model that relies on a large set of competing laws of motion for the coefficients and automatically decides which one describes the data best. We do so by proposing a flexible Bayesian approach to TVP-VARs that efficiently controls for uncertainty with respect to the state evolution equation.

Our model fundamentally builds on the fact that parameters might change either abruptly (such as in the transition period when short-term interest rates hit the ZLB), or smoothly (such as during the Great Moderation, see Bauer and Rudebusch, 2020), or might depend on observed variables and thus interact with exogenous regressors (such as a measure of uncertainty associated with the conduct of monetary policy, see Tillmann, 2020). These three forms of parameter types can be built into the model by assuming that the TVPs depend on a potentially large panel of covariates. These covariates, which we label effect modifiers, can be partially latent and might feature their own state equations. For instance, they could include latent random walk factors or binary Markov switching variables. Inclusion of the former allows for capturing smoothly evolving trends in the parameters of the yield curve (and functions thereof) while including the latter enables us to control for structural breaks (such as those observed in the midst of the 1980s or after hitting the ZLB). In case they are observed, we obtain a model that is closely related to the varying coefficient model originally proposed in Hastie and Tibshirani (1993).

To decide on which law of motion to use, we pursue a Bayesian approach. Our estimation strategy revolves around specifying suitable shrinkage priors that allow us to shrink coefficients associated with irrelevant effect modifiers towards zero. These priors are capable of endogenously selecting the appropriate law of motion for the parameters and, since the TVPs depend linearly on the effect modifiers, also allows for combinations of different law of motions. Moreover, the model enables the researcher to drill into the driving forces of parameter time-variation. This feature is important if the researcher is interested in investigating why relations between variables in a VAR change over time, and to what extent these changes are explained by the observed effect modifiers.

Model estimation is carried out through a computationally efficient Markov chain

Monte Carlo (MCMC) algorithm. The proposed algorithm is scalable to high dimensions and thus applicable to large datasets that commonly arise in macroeconomics and finance.

We then apply this model to analyze the US term structure of interest rates and investigate the empirical properties of our approach using two information sets in the underlying multivariate regression model. First, we include several interest rates at different maturities directly as endogenous variables. Second, we consider a three-factor Nelson-Siegel model for the term structure of interest rates as in Diebold and Li (2006), and Diebold et al. (2008). The dynamics of these three factors are modeled as a TVP-VAR.

These models are then used to predict the term structure of interest rates. Adopting a long hold-out period that includes several recessionary episodes, our approach improves upon a wide range of competing models. Depending on the model class, we find improvements for both point and density forecasts. Model confidence sets provide evidence that our flexible models are very often among the best performing model specifications.

The forecasting exercise provides evidence on out-of-sample model fit. However, our model can also be applied to investigate the key determinants of parameter time-variation. Using all available observations and considering the Nelson-Siegel class of models, we document substantial time-variation in the low frequency relations between key features of the yield curve. These parameter dynamics are strongly influenced by observed quantities such as proxies for uncertainty or recession indicators.

The rest of the paper is structured as follows. Section 2 introduces the econometric framework which includes the general form of the TVP-VAR, a flexible law of motion for the latent states as well as the Bayesian approach used to select appropriate effect modifiers. In Section 3 we illustrate key model features and highlight the predictive capabilities of our approach. The last section concludes the paper. Supplementary materials are provided in the Online Appendix.

2 Flexible Bayesian modeling of the yield curve

As discussed in the introduction, our goal is to develop a model which is capable of simultaneously modeling a panel of yields. Since the yield curve is driven by common

shocks, and yields often co-move, we adopt a general VAR model. In the empirical work, we consider two strategies. The first includes all yields in an otherwise unrestricted VAR with drifting parameters. The second exploits the fact that a low number of latent factors (which evolve according to a TVP-VAR) describe the cross-section of yields well.

2.1 A TVP-VAR for capturing yield curve dynamics

Let $\boldsymbol{y}_t$ denote an $M \times 1$ vector of endogenous variables at time $t = 1, \ldots, T$. If we wish to model the yield curve in an unrestricted manner, $\boldsymbol{y}_t$ consists of M yields of differing maturities. In case we adopt the Nelson-Siegel factor model, $\boldsymbol{y}_t$ contains $M = 3$ factors that describe the level, slope, and curvature of the yield curve.[2]

In both cases, we assume that $\boldsymbol{y}_t$ depends on its P lags which we store in a $K(= MP) \times 1$ vector $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \ldots, \boldsymbol{y}'_{t-P})'$. The TVP-VAR can be written as a multivariate regression model:

$$\boldsymbol{y}_t = (\boldsymbol{I}_M \otimes \boldsymbol{x}'_t)\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{\Sigma}_t),$$

where $\boldsymbol{\beta}_t$ represents a set of $k(= MK)$ dynamic regression coefficients and $\boldsymbol{\epsilon}_t$ is a vector Gaussian shock process with time-varying $M \times M$ variance-covariance matrix $\boldsymbol{\Sigma}_t$. We assume that $\boldsymbol{\Sigma}_t$ can be decomposed as follows:

$$\boldsymbol{\Sigma}_t = \boldsymbol{Q}_t \boldsymbol{H}_t \boldsymbol{Q}'_t.$$

Here, $\boldsymbol{Q}_t$ denotes an $M \times M$ lower triangular matrix with unit diagonal and $v(= M(M-1)/2)$ free elements denoted by $\boldsymbol{q}_t$, and $\boldsymbol{H}_t = \text{diag}\left(\exp(h_{1t}), \ldots, \exp(h_{Mt})\right)$ is a diagonal matrix with h_{jt} $(j = 1, \ldots, M)$ representing time-varying log-volatilities. These are assumed to evolve according to an AR(1) process.

This model assumes that the yields (or the factors describing the yields) depend on their lagged values. This relationship is encoded in $\boldsymbol{\beta}_t$, which changes over time. In principle, without further restrictions on $\boldsymbol{\beta}_t$ we can capture different forms of structural

[2] Consistent with the literature, these factors are first estimated by OLS. More details can be found in Section 3.

breaks. The shocks $\boldsymbol{\epsilon}_t$ are heteroskedastic, and volatilities can evolve smoothly over time. This allows us to capture features such as the decline in the volatility of macroeconomic and financial shocks during the Great Moderation. In addition, time-variation in $\boldsymbol{Q}_t$ allows us to capture changes in the correlation structure of shocks to the yield curve.

It is necessary to introduce regularization in the unrestricted general model to avoid overfitting. Our model specification proceeds along the lines proposed in Frühwirth-Schnatter and Wagner (2010). Specifically, we use a non-centered parameterization and rewrite the TVP-VAR as follows:

$$\boldsymbol{y}_t = (\boldsymbol{I}_M \otimes \boldsymbol{x}_t')(\boldsymbol{\beta} + \tilde{\boldsymbol{\beta}}_t) + (\boldsymbol{Q} + \tilde{\boldsymbol{Q}}_t)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{H}_t).$$

This parameterization allows us to disentangle time-invariant coefficients $\boldsymbol{\beta}$ from the time-varying vector $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t - \boldsymbol{\beta}$. For the decomposed variance-covariance matrix, $\boldsymbol{Q}$ is a lower triangular matrix with ones on the diagonal capturing the constant part of the covariances, and $\tilde{\boldsymbol{Q}}_t = \boldsymbol{Q}_t - \boldsymbol{Q}$ is the corresponding lower triangular matrix with zero-diagonal elements containing the time-varying part. Their free elements are collected in the $v \times 1$ vectors $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}_t$, respectively.

Our focus is on how the $N(= k + v) \times 1$ vector $\tilde{\boldsymbol{\gamma}}_t = (\tilde{\boldsymbol{\beta}}_t', \tilde{\boldsymbol{q}}_t')'$ evolves over time. The literature (see, for example, Mumtaz and Surico, 2009) typically assumes $\tilde{\boldsymbol{\gamma}}_t$ to follow a multivariate random walk process:

$$\tilde{\boldsymbol{\gamma}}_t | \tilde{\boldsymbol{\gamma}}_{t-1} \sim \mathcal{N}(\tilde{\boldsymbol{\gamma}}_{t-1}, \boldsymbol{V}), \text{ with } \tilde{\boldsymbol{\gamma}}_0 = \boldsymbol{0}_N.$$

This law of motion suggests that the expected value of $\tilde{\boldsymbol{\gamma}}_t$ equals $\tilde{\boldsymbol{\gamma}}_{t-1}$, and the amount of time-variation is determined by the $N \times N$ process innovation variance-covariance matrix $\boldsymbol{V}$. Moreover, it implies that parameters evolve smoothly and feature long memory. Another common assumption is that the shocks to the states feature positive error variances. Hence, parameter time-variation is introduced mechanically and this might lead to overfitting issues.

Estimation and inference is typically carried out via Bayesian methods. The recent literature proposes shrinkage priors to allow for data-based selection of the coefficients which

should be time-varying or constant (see, for example, Bitto and Frühwirth-Schnatter, 2019; Huber et al., 2020). This already leads to substantial improvements in predictive accuracy but does not tackle the fundamental question whether the coefficients are better characterized by a random walk, a change-point process, or by mixtures of these. Alternative methods capable of doing so are mixture innovation models (Koop et al., 2009; Maheu and Song, 2018; Huber et al., 2019) or models using flexible mixture priors (Hauzenberger et al., 2021; Lopes et al., 2021). All these methods, however, are not capable of linking coefficient dynamics to observed quantities.

2.2 A general law of motion for the TVPs

The random walk assumption on the state equation of the TVPs is relaxed by assuming that the time-varying part stored in $\tilde{\boldsymbol{\gamma}}_t$ depends on a set of R additional factors $\boldsymbol{z}_t$. These $\boldsymbol{z}_t$ are the effect modifiers mentioned in the introduction, which can be observed or latent. The relationship between the TVPs and $\boldsymbol{z}_t$ is given by:

$$\tilde{\boldsymbol{\gamma}}_t = \boldsymbol{\Lambda}\boldsymbol{z}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Omega}). \tag{1}$$

The matrix $\boldsymbol{\Lambda}$ is an $N \times R$ matrix of regression coefficients, and $\boldsymbol{\eta}_t$ is a Gaussian error term with diagonal error variance-covariance matrix $\boldsymbol{\Omega} = \text{diag}(\omega_1^2, \dots, \omega_N^2)$. If $R \ll N$, the coefficients feature a factor structure and co-move according to the effect modifiers in $\boldsymbol{z}_t$. The presence of uncorrelated measurement errors in $\boldsymbol{\eta}_t$ allows for idiosyncratic deviations from the conditional mean $\boldsymbol{\Lambda}\boldsymbol{z}_t$.

The relationship between $\tilde{\boldsymbol{\gamma}}_t$ and $\boldsymbol{z}_t$ is determined by the factor loadings in $\boldsymbol{\Lambda}$. If the j^{th} column of $\boldsymbol{\Lambda}$, $\boldsymbol{\lambda}_j$, is equal to zero, then the corresponding j^{th} factor in $\boldsymbol{z}_t$ does not enter the model and thus has no influence on $\tilde{\boldsymbol{\gamma}}_t$. Cases in between are possible as well. For example, if the $(i,j)^{\text{th}}$ element in $\boldsymbol{\Lambda}$, λ_{ij}, equals zero, the i^{th} coefficient does not depend on z_{jt} and is exclusively driven by the other factors and potential measurement errors. In the case of the yield curve, we expect the dynamic coefficients associated with the short end of the curve to depend on other quantities than, for example, the coefficients related to longer maturities. This is because short-run interest rates strongly depend on the

monetary policy stance and shifts thereof while longer-term interest rates are driven by expectations about future short-term interest rates, inflation and measures of economic uncertainty.

Selecting $\boldsymbol{z}_t$ is crucial for determining the dynamics of $\tilde{\boldsymbol{\gamma}}_t$. Appropriate choices of $\boldsymbol{z}_t$ yield a variety of important special cases that depend on the specific values of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ as well as on the composition of $\boldsymbol{z}_t$. Two special cases arise independently of $\boldsymbol{z}_t$. If $\boldsymbol{\Lambda} = \mathbf{0}_{N \times R}$, with $\mathbf{0}_{N \times R}$ being an $N \times R$ matrix of zeros, we obtain a random coefficients model that assumes that the regression coefficients follow a white noise process (for some recent papers that follow this approach, see Hauzenberger et al., 2021; Korobilis, 2021); if both $\boldsymbol{\Lambda} = \mathbf{0}_{N \times R}$ and $\boldsymbol{\Omega} = \mathbf{0}_{N \times N}$, we obtain a constant parameter regression model.

Before we discuss the choice of $\boldsymbol{z}_t$, it is worth noting that if $\boldsymbol{z}_t$ is (partially) latent, the model in Eq. (1) is not identified. Since our object of interest is $\tilde{\boldsymbol{\gamma}}_t$, this poses no substantial issues for model estimation and interpretation. If we wish to structurally interpret the loadings $\boldsymbol{\Lambda}$, standard identification strategies from the literature on dynamic factor models can be used (see, for example, Geweke and Zhou, 1996; Aguilar and West, 2000; Stock and Watson, 2011).

2.3 Effect modifiers

Since the literature on yield curve modeling provides evidence favoring different forms of parameter change, we specify $\boldsymbol{z}_t$ appropriately and let the data speak about the form of parameter change.

If we assume that $\boldsymbol{z}_t$ consists exclusively of a sequence of $R = R_\tau$ latent factors $\boldsymbol{\tau}_t$, we obtain a model closely related to the one in Chan et al. (2020). The latent factors evolve according to a multivariate random walk:

$$\boldsymbol{\tau}_t = \boldsymbol{\tau}_{t-1} + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{V}_\tau),$$

where $\boldsymbol{V}_\tau = \text{diag}(v_1^2, \dots, v_{R_\tau}^2)$ denotes a diagonal variance-covariance matrix with v_j^2 being process innovation variances that determine the smoothness of the elements in $\boldsymbol{\tau}_t$. This model implies a latent factor structure in $\tilde{\boldsymbol{\gamma}}_t$ if $R_\tau \ll N$ and that the TVPs evolve

smoothly. For yield curve data, this assumption might be warranted because interest rates strongly co-move and are well captured by few common factors. If we wish to model the term structure without introducing assumptions on the observation equation, this model implies that while coefficients are unrestricted, the TVPs are driven by a few unsobervable factors. Carriero et al. (2021), in a recent paper, follow a similar strategy to model the US term structure of interest rates. As opposed to our specification, they estimate constant coefficients but assume the error variances to be driven by a single latent factor.

It is also worth stressing that if we set $R = R_\tau = N$, $\boldsymbol{\Lambda} = \boldsymbol{I}_N$, we end up with a standard time-varying parameter model. Presuming that the covariances are constant we obtain the model put forth in Cogley and Sargent (2005). These models all rely on parameters that evolve according to a random walk (depending on the innovation variances to the latent states). This is consistent with Bianchi et al. (2009) and Del Negro et al. (2019) who assume the coefficients of the model to evolve smoothly over time.

We obtain a model which allows for sudden shifts in the parameters by setting $\boldsymbol{z}_t = S_t$, with $S_t \in \{0, 1\}$ denoting a binary indicator with transition probabilities given by:

$$p(S_t = i | S_{t-1} = j) = p_{ij}, \quad \text{for } i, j = 0, 1.$$

Here, p_{ij} denotes the $(i, j)^{\text{th}}$ element of a 2×2 transition probability matrix $\boldsymbol{P}$. If $\boldsymbol{\Omega}$ is close to a zero matrix, this model is a Markov switching specification in the spirit of Sims and Zha (2006).[3] The main restriction is that breaks occur for all coefficients in $\tilde{\boldsymbol{\gamma}}_t$ simultaneously. Adding S_t to $\boldsymbol{z}_t$ provides more flexibility and permits sudden changes in yield curve dynamics (see Liu et al., 2019).

Finally, if we include observed variables in $\boldsymbol{z}_t$ we end up with a regression model with interaction effects. For instance, Tillmann (2020) assumes that the yield curve response to monetary policy depends on uncertainty about the conduct of monetary policy. Hence, in this case, $\boldsymbol{z}_t$ could include a measure of monetary policy uncertainty.

These examples show that our model, conditional on choosing a suitable set of effect

[3]In principle, estimating models with more than two regimes would be straightforward and efficient techniques are readily available, see Kaufmann (2015).

modifiers, is capable of mimicking several prominent specifications in the literature. Since the question on the appropriate state evolution equation is essentially a model selection issue, we specify $\boldsymbol{z}_t$ to include most of the modifiers discussed above (with the exception of the $R = N$ setup). More precisely, we set $\boldsymbol{z}_t$ as follows:

$$\boldsymbol{z}_t = (\boldsymbol{r}_t', \boldsymbol{S}_t', \boldsymbol{\tau}_t')'.$$

Here, we let $\boldsymbol{r}_t$ denote a set of R_r observed factors and the dimension of $\boldsymbol{z}_t$ is thus $R = R_r + R_S + R_\tau$. To allow for additional flexibility we assume that $\boldsymbol{\tau}_t = (\boldsymbol{\tau}_{1t}', \ldots, \boldsymbol{\tau}_{Mt}')'$ with $\boldsymbol{\tau}_{jt}$ being equation-specific factors of dimension $R_{\tau j}$ (and thus $R_\tau = \sum_j R_{\tau j}$). Assuming that $R_{\tau i} = R_{\tau j} = \delta$ for all i, j, we have $R_\tau = \delta M$ latent random walk factors. Likewise, we estimate a separate Markov switching indicator S_{jt} per equation. Thus $\boldsymbol{S}_t = (S_{1t}, \ldots, S_{Mt})'$, with the corresponding transition probabilities matrix denoted by $\boldsymbol{P}_j$ such that $R_S = M$.

The loadings matrix $\boldsymbol{\Lambda}$ is structured such that the loadings in equation j associated with the factors $\boldsymbol{\tau}_{it}$ and S_{it} for $i \neq j$ equal zero. This assumption strikes a balance between assuming a large number of latent factors to achieve flexibility (and risk overfitting), and using a rather parsimonious model (with the risk of being too simplistic). Recent contributions using similar assumptions on the state evolutions are Koop et al. (2009) and Maheu and Song (2018). As opposed to these papers, our approach offers more flexibility since the presence of the idiosyncratic shocks to the TVPs allows for deviations in case the factor structure does not represent the data well.

2.4 Stochastic selection of effect modifiers

Our model approach nests a variety of competing models and the specific choice of $\boldsymbol{z}_t$ leads to different assumptions on how the TVPs evolve over time. To single out irrelevant elements in $\boldsymbol{z}_t$, one could simply set the corresponding columns in $\boldsymbol{\Lambda}$ equal to zero. To decide on the appropriate model variant and alleviate over-parameterization concerns, we opt for a Bayesian approach to introduce shrinkage on $\boldsymbol{\Lambda}$.[4]

[4]Since all other priors are relatively standard we provide further details in the Online Appendix.

For illustrating our method we focus on a simple regression case with $M = 1$ and $P = 1$. In that case, we have a time-varying parameter AR(1) model:

$$y_t = \beta_t \, y_{t-1} + \epsilon_t.$$

We assume that $\boldsymbol{z}_t = (r_t, S_t, \tau_t)'$ includes some observed variable, a Markov indicator and a latent random walk factor. The corresponding state equation for β_t is then given by:

$$\beta_t = \beta_0 + \lambda_1 r_t + \lambda_2 S_t + \lambda_3 \tau_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \omega), \tag{2}$$

with ω being a measurement error variance, λ_j $(j = 1, 2, 3)$ are loadings and β_0 is a time-invariant parameter. Considering Eq. (2) reveals that selecting the appropriate law of motion for β_t boils down to testing whether $\lambda_j \neq 0$ for all j. For instance, if $\lambda_2 \neq 0$ and $\lambda_1 = \lambda_3 = \omega = 0$ we have an AR(1) model with Markov switching dynamics for the autoregressive parameter. If we set $\lambda_3 \neq 0$ while zeroing out the other coefficients we obtain an AR(1) model with a time-varying autoregressive parameter that follows a random walk process. Notice that if $\lambda_1 = \lambda_2 = \lambda_3 = 0$, the corresponding model is a random coefficients model which assumes that β_t is centered on a time-invariant component β_0.

From a Bayesian perspective, this problem can be tackled through shrinkage priors. In principle, one could use a spike and slab prior similar to the one proposed in George and McCulloch (1993), see also George and McCulloch (1997). These priors select appropriate covariates by introducing Bernoulli-distributed indicator variables which equal one if a given covariate is included (i.e., $\lambda_j \neq 0$) or zero if the variable is excluded (i.e., $\lambda_j = 0$). However, if the dimensionality of $\boldsymbol{z}_t$ becomes large these priors suffer from mixing issues. As a solution, the statistical literature proposes so-called global-local shrinkage priors (see Polson and Scott, 2010). Recent shrinkage priors such as the Dirichlet-Laplace prior (Bhattacharya et al., 2015), the Normal-Gamma prior (Griffin and Brown, 2010), the horseshoe (Carvalho et al., 2010) or the Triple-Gamma prior (Cadonna et al., 2020) can be used. All of them possess convenient theoretical properties and should be well suited for achieving shrinkage in our model setup. In this paper, we opt for the horseshoe prior

since it does not rely on additional tuning parameters and has been shown to work well in a wide range of different applications (see, for example, Huber et al., 2021; Kowal, 2021).

The horseshoe is a conditionally Gaussian prior on the loadings λ_j:

$$\lambda_j | c_j, d \sim \mathcal{N}(0, c_j^2 \, d^2), \quad c_j \sim \mathcal{C}^+(0,1), \quad d \sim \mathcal{C}^+(0,1). \tag{3}$$

This prior depends on two hyperparameters. The first one, c_j, is a local parameter which controls the amount of shrinkage on λ_j. If it is close to zero, heavy shrinkage is introduced and the corresponding posterior distribution of λ_j will be centered tightly around zero (i.e., the corresponding j^{th} element in $\boldsymbol{z}_t$ does not shape the dynamics of the TVPs). By contrast, if it is large it introduces little prior information and allows the data to speak (i.e., the j^{th} covariate is included in the model). The second hyperparameter, d, is a global shrinkage parameter which forces all coefficients to zero. The local shrinkage parameters allow for non-zero coefficients even if d is close to zero.

In our general model, the prior slightly differs from the simplified version introduced above (exact details can be found in the Online Appendix). Here it suffices to note that the horseshoe is applied on the different columns of $\boldsymbol{\Lambda}$. For each of these R columns, a separate global shrinkage parameter d_i $(i = 1, \ldots, R)$ is introduced. This allows for testing whether z_{it} explains the joint dynamics of $\tilde{\boldsymbol{\gamma}}_t$. Since each of the elements in $\boldsymbol{\Lambda}$ features its own local shrinkage parameter, our model is sufficiently flexible to also allow for deviations from this pattern and to detect appropriate specifications for each equation and parameter separately.

2.5 A brief sketch of the posterior simulator

Our MCMC algorithm is implemented on an equation-by-equation basis. Conditional on $\boldsymbol{Q}_t$ and the block structure of $\boldsymbol{\Lambda}$, one can state the VAR as a system of (conditionally) independent equations. The first equation of this system is given by:

$$y_{1t} = \boldsymbol{x}_t'(\boldsymbol{\beta}_1 + \tilde{\boldsymbol{\beta}}_{1t}) + \varepsilon_{1t},$$

and the j^{th} equation ($j > 1$):

$$y_{jt} = \boldsymbol{x}_t'(\boldsymbol{\beta}_j + \tilde{\boldsymbol{\beta}}_{jt}) + \boldsymbol{u}_{jt}'(\boldsymbol{q}_j + \tilde{\boldsymbol{q}}_{jt}) + \varepsilon_{jt}, \tag{4}$$

where $\boldsymbol{\beta}_j$ and $\tilde{\boldsymbol{\beta}}_{jt}$ denote the j^{th} subvectors of the constant and time-varying parts in $\boldsymbol{\beta}_t$ with $\boldsymbol{\beta}_t = (\boldsymbol{\beta}_{1t}', \ldots, \boldsymbol{\beta}_{Mt}')'$ and $\boldsymbol{u}_{jt} = (y_{1t}, \ldots, y_{j-1,t})'$. The $(j-1)$-dimensional vectors $\boldsymbol{q}_j$ and $\tilde{\boldsymbol{q}}_{jt}$ store the constant and time-invariant part of the free elements in the j^{th} row of $\boldsymbol{Q}_t$.

Using this representation allows us to treat each equation as a separate estimation problem and thus to exploit favorable computational properties. For sampling the parameters and latent states per equation, we rely on several Gibbs updating steps. The single most important feature is that we sample the latent effect modifiers (i.e., the random walk factors and the Markov switching indicators) marginally of the TVPs. This is achieved by integrating out the TVPs analytically, and improves mixing considerably. In the Online Appendix, we provide precise details on all full conditional posterior distributions, show that our algorithm displays favorable convergence properties and that the model is capable of reproducing parameter dynamics accurately.

3 Modeling the term structure of US government bond yields

This section applies the model to the term structure of US interest rates. These time series are characterized by substantial structural breaks (for example, during the period of the ZLB), feature substantial co-movement both in the level of the time series but also in the parameters describing their evolution. Our proposed model framework might thus be well suited to capture such features. We investigate this claim in a thorough forecasting exercise using several established benchmarks. After showing that our approach yields favorable forecasts, we discuss the driving forces behind parameter changes and illustrate how key quantities that shape yield curve dynamics co-move over time at low frequencies.

3.1 Data

Our aim is to model monthly zero-coupon yields of US treasuries at different yearly maturities.[5] Let $i_t(\theta)$ denote the yield at maturity $\theta \in \{1, 3, 5, 7, 10, 15\}$ at time $t = 1, \ldots, T$. As stated in Section 2, we consider two versions of our model. The first one models the yields in an unrestricted manner, and includes all of them as endogenous variables in first differences for stability reasons. That is, for the model labeled VAR, the vector of endogenous variables is $\boldsymbol{y}_t = (\Delta i_t(1), \Delta i_t(3), \ldots, \Delta i_t(15))'$ of size $M = 6$, with Δ referring to the first-difference operator.

The second model class includes specifications based on the three-factor Nelson-Siegel (NS) model as in Diebold and Li (2006) that imposes a factor structure on the yields:

$$i_t(\theta) = L_t + \left(\frac{1 - \exp(-\theta\alpha)}{\theta\alpha}\right) \check{S}_t + \left(\frac{1 - \exp(-\theta\alpha)}{\theta\alpha} - \exp(-\theta\alpha)\right) \check{C}_t. \tag{5}$$

Here, L_t is a factor that controls the level, $\check{S}_t$ determines the slope, and $\check{C}_t$ represents the curvature factor of the yields. The parameter α governs the exponential decay rate, which we set to $\alpha = 0.7308$ (12×0.0609).[6] These factors are extracted from the level of the zero-coupon yields by running OLS on a t-by-t basis.[7] For the model subsequently labeled NS-VAR, we use the latent factors $L_t, \check{S}_t$ and $\check{C}_t$ in differences as endogenous variables in the VAR specifications stated in Section 2 by defining $\boldsymbol{y}_t = (\Delta L_t, \Delta \check{S}_t, \Delta \check{C}_t)'$, resulting in $M = 3$. Figure 1 displays raw yield data, alongside the extracted Nelson-Siegel factors. Charts of the transformed data are provided in the Online Appendix.

For analyzing the role of observed factors governing joint dynamics of the TVPs, we rely on three exogenous variables:

(a) a binary recession indicator (labeled REC) for the US, dated on a monthly basis by the Business Cycle Dating Committee of the National Bureau of Economic Research,

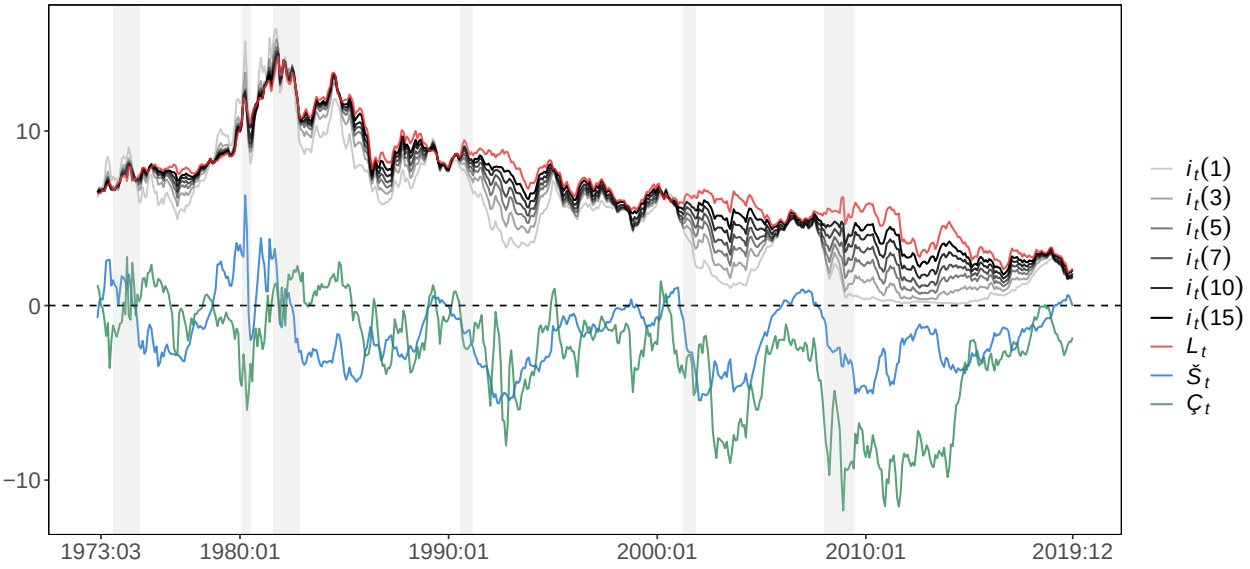
(b) the National Financial Conditions Index (NFCI), maintained by the Federal Reserve

[5] The data we use is described in detail in Gürkaynak et al. (2007), and available online at federalreserve.gov/data/nominal-yield-curve.htm.

[6] See Diebold and Li (2006) for a discussion of this specific choice.

[7] It is noteworthy that the basic Nelson-Siegel model is not arbitrage free. However, using a similar dataset, Coroneo et al. (2011) show that the parameters of an arbitrage free model do not differ from the ones obtained from the standard Nelson-Siegel model in a statistically significant manner.

Figure 1: Zero-coupon yields and Nelson-Siegel factors.



Notes: $i_t(\theta)$ denotes the zero-coupon yields at maturity $\theta = \{1, 3, 5, 7, 10, 15\}$ years, L_t is a factor that controls the level, $\check{S}_t$ determines the slope, and $\check{C}_t$ represents the curvature factor of the yields. The black dashed line denotes the zero line, while the gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period: 1973:01 to 2019:12. Vertical axis: zero-coupon yields and Nelson-Siegel factors. Horizontal axis: months.

Bank of Chicago, downloaded from the FRED database of the Federal Reserve Bank of St. Louis (available online at fred.stlouisfed.org), and

(c) the Risk-Free (RF) interest rate from the Fama-French Portfolios and Factors database, provided on the web page of Kenneth R. French as another important early warning indicator.[8]

Thus, $\boldsymbol{r}_t = (r_{\text{REC},t-1}, r_{\text{NFCI},t-1}, r_{\text{RF},t-1})'$ and $R_r = 3$. The lagged effect modifiers enter the model in levels. Plots of these series are provided in Section 3.4 in the context of our discussion of the resulting effect modifier estimates.

We use these three effect modifiers for simple reasons. First, there is strong evidence that yield curve dynamics differ across business cycle phases (see, for example, Hevia et al., 2015). Second, the RF interest rate serves as an early warning indicator which possesses predictive power for changes in the shape of the yield curve. Finally, the inclusion of the NFCI is motivated by the recent literature on forecasting tail risks in macroeconomic and financial time series (see, for example, Adrian et al., 2019; Carriero et al., 2020; Adams et al., 2021).

[8] Available online at mba.tuck.dartmouth.edu/pages/faculty/ken.french.

3.2 Design of the forecasting exercise

Estimation and forecasting is carried out recursively. Using data from 1973:01 to 1999:12, we produce one-month-ahead and three-months-ahead forecasts.[9] After obtaining the predictive distributions for 2000:01 we expand the sample and repeat this procedure until we reach 2019:12.[10] This yields 240 out-of-sample observations. Point forecast performance is measured using Root Mean Squared Errors (RMSEs), while density forecasts are assessed in terms of Log Predictive Bayes Factors (LPBFs), averaged over the out-of-sample observations (see also Geweke and Amisano, 2010). We moreover consider model confidence sets (MCS) proposed in Hansen et al. (2011) as a measure of absolute model performance. The MCS procedure constructs a superior set of models which includes the best performing model with a certain significance level (set to 0.25 in our case).

The forecast exercise distinguishes between two model classes. In our forecasting horse race we estimate two versions of all models: one based on using the unrestricted yields, and one relying on the NS factors.[11] We have 15 distinct model specifications in each class. These range from nested alternatives (such as constant parameter VARs, TVP-VARs in the spirit of Primiceri, 2005, or models akin to a Markov switching specification) to the most general model which includes all effect modifiers discussed above. In particular, specifications within a model class differ in terms of three aspects (see Table 1): First, whether the three exogenous variables (collected in $\boldsymbol{r}_t$) are included in $\boldsymbol{z}_t$ or not ("x" marks inclusion, "–" indicates no observed factors); second, in terms of the number δ of latent random walk factors $R_{\tau j}$ included in $\boldsymbol{z}_t$ (which we assume to be equal across equations); and third, in terms of the presence of Markov switching indicators in $\boldsymbol{S}_t$, again with "x" marking their inclusion and "–" their absence.

This setup implies that we have 15 TVP-NS-VAR and 15 TVP-VAR specifications. For comparative purposes we also consider the two class-specific constant parameter model

[9] Higher-order forecasts involving the effect modifiers are based on random walk predictions of these quantities. For the Markov indicators and the random walk factors, this is achieved by exploiting the dynamic structure of the model, producing forecasts within a unified econometric framework.

[10] Note that for our set of financial indicators, data revisions and ragged edges arising from delays in the publication of the series do not matter. This is due to financial market data being available almost instantaneously, and the published quotes are not subject to revisions at later dates.

[11] In the latter case we forecast the factors and then, using the implied matrix of factor loadings, map the factor forecasts into the space of observed yields.

variants, labeled "Constant" (which serves as the benchmark), and a conventional independent random walk specification of the TVPs (implying that we set the number of random walk factors equal to K and exclude $\boldsymbol{r}_t$ and $\boldsymbol{S}_t$).[12] Notice that we also have a specification which includes only $\boldsymbol{S}_t$ and another one which uses only observed factors.

Finally, we also consider variants of two simpler models. The first is a standard AR(P) model with random walk TVPs or constant parameters, labeled TVP-AR in Table 1. The constant parameter AR(P) is an often used benchmark when assessing the accuracy of more intricate models for predicting government bond yields, which typically performs similar to no-change random walk models at shorter horizons and yields improvements for long-run forecasts (see, for example, Diebold and Li, 2006; Carriero et al., 2012). The second one, labeled TVP-NS-AR, is used to model the state dynamics of the NS factors (similar to what our VAR does), but disregards contemporaneous correlations between the factors.

All model specifications considered include stochastic volatility (see Section 2.1) and $P = 3$ lags of the endogenous variables.

3.3 Forecasting results

Table 1 shows the one-month-ahead and one-quarter-ahead forecasting results for US treasury yields at different maturities. RMSEs are shown as ratios, and LPBFs in differences are given below the RMSEs in parentheses. RMSEs below one indicate superior performance, relative to the benchmark "Constant," as LPBF figures greater than zero do. The best performing model specification by column is given in bold, highlighting the specification with the smallest RMSE ratio and the largest positive LPBF difference, respectively. Gray shaded cells indicate models which are excluded from the MCS based on a significance level of 0.25, while non-shaded cells refer to models which are among the superior model set.

We first provide a general overview on model performance and then zoom into differences in predictive accuracy for point and density forecasts. At a very general level,

[12] It is worth stressing that the constant parameter VAR is equipped with a horseshoe prior and thus allows for flexible shrinkage towards simpler model specifications such as a random walk with stochastic volatility.

Table 1: Out-of-sample forecasting results for US treasury yields at different maturities using TVP-NS-VAR and TVP-VAR model specifications alongside simple univariate TVP-NS-AR and TVP-AR models (numbers refer to relative RMSEs while those in the parentheses relate to average LPBFs between a given model and the unrestricted constant parameter VAR.)

Model	Specification			One-month-ahead							One-quarter-ahead						
	$\boldsymbol{r}_t$	δ	$\boldsymbol{S}_t$	Joint	1 year	3 year	5 year	7 year	10 year	15 year	Joint	1 year	3 year	5 year	7 year	10 year	15 year
TVP-NS-AR	–	3	–	0.79 (-1.78)	1.02 (-1.07)	0.98 (-0.45)	0.95 (-0.15)	0.90 (0.03)	0.78 (0.27)	0.60 (0.67)	0.92 (-2.18)	1.01 (-1.02)	0.98 (-0.47)	0.97 (-0.21)	0.96 (-0.10)	0.92 (-0.04)	0.83 (0.07)
	Constant			0.79 (-1.76)	1.04 (-1.06)	0.98 (-0.45)	0.95 (-0.15)	0.89 (0.03)	0.78 (0.28)	0.60 (0.68)	0.91 (-2.14)	0.99 (-1.01)	0.97 (-0.45)	0.96 (-0.19)	0.95 (-0.09)	0.92 (-0.03)	0.82 (0.08)
TVP-NS-VAR	x	6	x	0.78 (-0.79)	0.91 (-0.28)	0.95 (-0.11)	0.94 (0.03)	0.89 (0.14)	0.79 (0.34)	0.61 (0.72)	0.95 (-1.42)	0.99 (-0.35)	1.04 (-0.20)	1.01 (-0.06)	0.98 (-0.01)	0.94 (0.01)	0.85 (0.10)
	x	3	x	0.79 (-0.66)	0.96 (-0.21)	1.00 (-0.09)	0.96 (0.04)	0.90 (0.14)	0.78 (0.35)	0.60 (0.73)	0.95 (-1.35)	1.03 (-0.27)	1.06 (-0.14)	1.02 (-0.03)	0.98 (0.00)	0.94 (0.02)	0.84 (0.11)
	x	1	x	**0.77** (-0.67)	0.94 (-0.14)	0.96 (-0.04)	**0.93** (0.06)	**0.88** (0.15)	**0.77** (0.34)	**0.59** (0.73)	0.95 (-1.19)	1.04 (-0.21)	1.05 (-0.10)	1.01 (-0.01)	0.98 (0.01)	0.94 (0.02)	0.84 (0.11)
	x	6	–	0.80 (-0.89)	0.93 (-0.32)	0.98 (-0.16)	0.96 (0.00)	0.91 (0.13)	0.79 (0.34)	0.61 (0.72)	0.94 (-1.50)	0.96 (-0.39)	1.00 (-0.23)	0.99 (-0.07)	0.97 (-0.01)	0.94 (0.02)	0.85 (0.11)
	x	3	–	0.79 (-0.66)	0.92 (-0.22)	0.98 (-0.10)	0.96 (0.03)	0.91 (0.14)	0.79 (0.35)	0.60 (0.72)	0.92 (-1.26)	0.95 (-0.27)	0.99 (-0.16)	0.97 (-0.03)	0.95 (0.02)	0.92 (0.04)	0.83 (0.13)
	x	1	–	0.78 (-0.63)	0.92 (-0.18)	0.98 (-0.08)	0.95 (0.05)	0.89 (0.15)	0.77 (**0.36**)	0.59 (0.74)	0.93 (-1.26)	0.96 (-0.24)	0.99 (-0.13)	0.97 (-0.02)	0.96 (0.02)	0.93 (0.04)	0.83 (0.13)
	–	6	x	0.79 (-0.86)	0.96 (-0.30)	0.97 (-0.11)	0.95 (0.04)	0.90 (0.15)	0.79 (0.35)	0.60 (0.72)	0.94 (-1.49)	1.04 (-0.37)	1.03 (-0.19)	1.00 (-0.05)	0.97 (0.00)	0.93 (0.01)	0.84 (0.09)
	–	3	x	0.78 (-0.70)	0.95 (-0.21)	0.95 (-0.06)	0.93 (0.06)	0.89 (0.15)	0.78 (0.34)	0.60 (0.72)	0.94 (-1.25)	1.02 (-0.28)	1.03 (-0.13)	0.99 (-0.02)	0.96 (0.02)	0.93 (0.03)	0.83 (0.11)
	–	1	x	0.78 (-0.70)	0.98 (-0.18)	0.97 (-0.04)	0.94 (0.07)	0.89 (**0.17**)	0.78 (0.36)	0.60 (0.73)	0.92 (-1.16)	0.99 (-0.22)	0.99 (-0.09)	0.97 (0.00)	0.95 (0.02)	0.92 (0.03)	0.83 (0.12)
	–	6	–	0.79 (-0.92)	0.94 (-0.33)	0.97 (-0.15)	0.95 (0.00)	0.90 (0.12)	0.79 (0.33)	0.61 (0.70)	0.93 (-1.44)	0.98 (-0.40)	0.99 (-0.23)	0.97 (-0.07)	0.96 (-0.01)	0.93 (0.02)	0.84 (0.10)
	–	3	–	0.79 (-0.86)	0.94 (-0.23)	0.96 (-0.10)	0.95 (0.04)	0.90 (0.15)	0.79 (0.34)	0.60 (0.72)	0.92 (-1.37)	0.98 (-0.29)	0.98 (-0.15)	0.96 (-0.02)	0.95 (0.02)	0.92 (0.03)	0.83 (0.12)
	–	1	–	0.78 (-0.72)	0.95 (-0.20)	0.97 (-0.06)	0.94 (0.05)	0.89 (0.15)	0.78 (0.34)	0.60 (0.72)	**0.91** (-1.30)	0.97 (-0.22)	0.97 (-0.11)	0.96 (0.00)	0.95 (0.03)	**0.91** (0.04)	**0.82** (0.11)
	x	–	–	0.78 (-0.36)	0.93 (0.05)	0.99 (0.00)	0.95 (0.07)	0.89 (0.16)	0.77 (0.35)	0.59 (0.74)	0.91 (-0.95)	**0.95** (0.02)	0.97 (-0.02)	0.96 (0.03)	0.95 (**0.05**)	0.91 (0.06)	0.82 (**0.15**)
	–	–	x	0.78 (-0.76)	0.95 (-0.24)	0.98 (-0.07)	0.95 (0.04)	0.89 (0.14)	0.78 (0.33)	0.60 (0.71)	0.94 (-1.26)	1.05 (-0.29)	1.03 (-0.12)	0.99 (-0.03)	0.96 (0.00)	0.93 (0.01)	0.83 (0.11)
	–	K	–	0.79 (-0.42)	0.96 (0.11)	0.98 (0.01)	0.95 (0.07)	0.90 (0.15)	0.78 (0.34)	0.60 (0.71)	0.92 (-1.02)	0.99 (0.07)	1.00 (-0.01)	0.97 (0.02)	0.95 (0.03)	0.92 (0.03)	0.83 (0.11)
	Constant			0.79 (-0.82)	1.01 (-0.37)	0.99 (-0.11)	0.96 (0.03)	0.90 (0.13)	0.78 (0.33)	0.60 (0.71)	0.91 (-1.34)	0.99 (-0.36)	0.97 (-0.13)	0.96 (-0.01)	0.95 (0.02)	0.92 (0.03)	0.82 (0.12)

Table 1 continued

Model	Specification			One-month-ahead							One-quarter-ahead						
	$\boldsymbol{r}_t$	δ	$\boldsymbol{S}_t$	Joint	1 year	3 year	5 year	7 year	10 year	15 year	Joint	1 year	3 year	5 year	7 year	10 year	15 year
TVP-AR	–	3	–	0.79	0.92	**0.95**	0.95	0.91	0.79	0.61	0.91	0.98	0.97	0.96	0.94	0.92	0.82
				(-7.29)	(**0.18**)	(0.05)	(0.07)	(0.14)	(0.34)	(0.70)	(-7.91)	(0.10)	(0.03)	(0.03)	(0.04)	(0.05)	(0.13)
	Constant			0.79	0.92	0.95	0.95	0.90	0.79	0.60	0.91	0.98	**0.96**	**0.95**	**0.94**	0.92	0.83
				(-7.32)	(0.14)	(0.04)	(0.07)	(0.14)	(0.34)	(0.71)	(-7.94)	(0.08)	(0.02)	(0.03)	(0.04)	(0.04)	(0.13)
TVP-VAR	x	6	x	0.78	0.92	0.96	0.94	0.89	0.79	0.60	0.94	0.96	0.99	0.98	0.97	0.94	0.84
				(1.11)	(0.05)	(0.03)	(0.07)	(0.14)	(0.33)	(0.68)	(0.62)	(-0.01)	(-0.01)	(0.00)	(0.00)	(0.01)	(0.08)
	x	3	x	0.78	0.92	0.96	0.94	0.89	0.78	0.60	0.94	0.98	1.00	0.99	0.98	0.94	0.84
				(1.19)	(0.09)	(0.03)	(0.06)	(0.14)	(0.33)	(0.70)	(0.54)	(0.01)	(-0.02)	(-0.01)	(0.00)	(0.02)	(0.10)
	x	1	x	0.79	0.93	0.97	0.95	0.89	0.78	0.60	0.93	0.96	0.99	0.98	0.97	0.94	0.84
				(1.49)	(0.10)	(0.03)	(0.06)	(0.13)	(0.33)	(0.70)	(**0.80**)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)	(0.12)
	x	6	–	0.79	0.93	0.98	0.95	0.90	0.79	0.60	0.94	0.99	1.00	1.00	0.98	0.95	0.84
				(1.26)	(0.06)	(0.04)	(0.07)	(0.15)	(0.34)	(0.69)	(0.69)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.10)
	x	3	–	0.78	0.91	0.96	0.94	0.89	0.79	0.60	0.94	0.98	1.00	1.00	0.98	0.95	0.85
				(1.20)	(0.11)	(**0.06**)	(**0.09**)	(0.16)	(0.34)	(0.71)	(0.66)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.12)
	x	1	–	0.78	**0.91**	0.96	0.94	0.89	0.78	0.60	0.92	0.96	0.97	0.97	0.96	0.93	0.83
				(1.56)	(0.13)	(0.06)	(0.08)	(0.15)	(0.34)	(0.72)	(0.61)	(0.07)	(**0.05**)	(0.04)	(0.04)	(0.05)	(0.14)
	–	6	x	0.81	0.95	0.97	0.95	0.91	0.81	0.63	0.97	1.00	1.00	1.00	1.00	0.98	0.90
				(0.94)	(0.02)	(0.03)	(0.07)	(0.13)	(0.32)	(0.67)	(0.45)	(-0.04)	(-0.03)	(-0.01)	(-0.01)	(0.00)	(0.06)
	–	3	x	0.79	0.95	0.98	0.95	0.90	0.79	0.60	0.93	0.99	0.99	0.97	0.96	0.93	0.83
				(1.07)	(0.05)	(0.02)	(0.06)	(0.12)	(0.31)	(0.68)	(0.27)	(-0.01)	(-0.02)	(-0.01)	(-0.01)	(0.01)	(0.08)
	–	1	x	0.78	0.95	0.96	0.94	0.89	0.79	0.60	0.95	0.97	0.98	0.98	0.98	0.96	0.87
				(1.26)	(0.08)	(0.04)	(0.07)	(0.14)	(0.32)	(0.69)	(0.41)	(0.03)	(0.00)	(-0.01)	(-0.01)	(0.00)	(0.09)
	–	6	–	0.79	0.95	0.96	0.95	0.90	0.79	0.60	0.92	0.99	0.98	0.97	0.96	0.93	0.83
				(1.04)	(0.04)	(0.05)	(0.08)	(0.15)	(0.34)	(0.69)	(0.32)	(-0.03)	(0.01)	(0.02)	(0.03)	(0.04)	(0.11)
	–	3	–	0.79	0.94	0.97	0.94	0.89	0.79	0.61	0.93	0.98	0.98	0.97	0.96	0.93	0.83
				(1.27)	(0.08)	(0.05)	(0.08)	(0.15)	(0.34)	(0.70)	(0.61)	(0.00)	(0.03)	(0.04)	(0.05)	(0.06)	(0.13)
	–	1	–	0.79	0.95	0.96	0.95	0.90	0.79	0.61	0.92	0.99	0.99	0.97	0.95	0.92	0.82
				(**1.58**)	(0.09)	(0.06)	(0.08)	(0.15)	(0.35)	(0.72)	(0.78)	(0.04)	(0.04)	(**0.05**)	(0.05)	(**0.06**)	(0.15)
	x	–	–	0.79	0.93	0.97	0.95	0.90	0.79	0.60	0.93	0.98	0.99	0.98	0.97	0.93	0.83
				(1.26)	(0.12)	(0.04)	(0.07)	(0.14)	(0.35)	(**0.75**)	(0.75)	(0.07)	(0.03)	(0.02)	(0.01)	(0.03)	(0.15)
	–	–	x	0.79	0.94	0.96	0.93	0.89	0.79	0.61	0.93	0.97	0.97	0.97	0.96	0.94	0.86
				(0.77)	(0.07)	(0.02)	(0.07)	(0.14)	(0.33)	(0.70)	(0.42)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)	(0.11)
	–	K	–	0.79	0.94	0.96	0.95	0.90	0.80	0.61	0.93	0.98	0.98	0.97	0.96	0.93	0.83
				(0.28)	(0.17)	(0.03)	(0.04)	(0.10)	(0.27)	(0.67)	(-0.15)	(**0.12**)	(0.01)	(-0.02)	(-0.03)	(-0.04)	(0.07)
	Constant*			0.85	0.39	0.59	0.72	0.82	0.99	1.28	0.76	0.43	0.62	0.73	0.81	0.87	0.96
				(2.04)	(-0.21)	(-0.81)	(-1.08)	(-1.25)	(-1.49)	(-1.87)	(2.26)	(-0.30)	(-0.84)	(-1.08)	(-1.19)	(-1.26)	(-1.35)

Notes: We present the results of out-of-sample one-month-ahead and one-quarter-ahead forecasting from the 15 TVP-NS-VAR and 15 TVP-VAR model variants alongside from the two NS-AR(3) and two AR(3) specifications for maturities 1, 3, 5, 7, 10 and 15 years and the corresponding joint measure across all maturities. Specifications of the TVP-VAR model variants are differentiated over a grid of effect modifier combinations, in terms of whether the exogenous variables in $\boldsymbol{r}_t$ are included or not ("x" marks inclusion, "–" indicates absence), the number $\delta = R_T/M$ of latent factors per equation (with $\delta = K$ yielding conventional independent random walk specifications for the TVPs) and the presence of Markov switching factors in $\boldsymbol{S}_t$, again with "x" marking inclusion and "–" absence. "Constant" labels a conventional constant parameter VAR. $M = 3$ and $K = 9$ in the case of NS-VAR, and $M = 6$ and $K = 18$ in the case of VAR. We estimate and forecast recursively, using data from 1973:01 to the time that the forecast is made, beginning in 2000:01 through 2019:12 leading to 240 out-of-sample observations. RMSEs and LPBFs, averaged over the out-of-sample observations are given relative to the constant parameter VAR model (marked by "*"). The best performing model specification by column is given in bold, highlighting the specification with the smallest RMSE ratio and the largest positive LPBF difference, respectively. The gray shaded cells indicate models which are excluded from the model confidence set (MCS, Hansen et al., 2011) based on a 25% significance level.

Table 1 suggests a pronounced degree of heterogeneity in forecast accuracy across models and for both the NS-AR/-VAR and AR/VAR specifications. While differences at some maturities are substantial (especially at the long end of the yield curve), they are muted or non-existent for others. The univariate AR(3) models, with and without TVPs, are competitive benchmarks for point forecasts, particularly for the one-quarter-ahead horizon. For joint densities, the univariate AR(3) models perform poorly, indicating that modeling the covariance structure across the yield curve is beneficial to forecast accuracy. It is also worth mentioning that the TVP variants of the multivariate NS-VAR and VAR models outperform the respective constant parameter specifications in most cases. This general observation holds true irrespective of whether only point forecasts or the full predictive distribution are considered. These accuracy premia point towards the necessity of addressing structural breaks in the dynamic evolution of the yield curve.

Comparing the NS-VAR and VAR specifications indicates that the latter usually perform better for density forecasts, while the former are often superior in terms of point forecasts. The MCS procedure confirms the significance of these results, with NS-VAR models being included in the set of superior models for point forecasts in almost all cases, often showing the smallest numerical values. The better point forecasting performance of the NS-VARs suggests that the three factors contain relevant information for the first moment of the predictive distribution. When we also consider higher order moments, this story changes. The significantly better density performance of the VAR models, signaled by a large number of inclusions in the MCS, is most likely driven by two sources. First, as opposed to the conditional mean, the strong implicit assumptions of the NS-VARs on the prediction error variance-covariance matrix seems overly restrictive. Allowing for richer dynamics in the covariances by explicitly modeling the shocks to a panel of yields improves density forecasts. Second, the small-scale NS-VARs might feature a too tight predictive variance since relevant information is ignored. And this harms density forecasting accuracy during turbulent times.

Next, we focus on specific model classes. Within these, differences in performance along sub-divisions (provided by the inclusion/exclusion of the exogenous variables or the Markov switching processes and the number of latent random walk factors) are often

negligible. This finding indicates that our flexible approach to probabilistically selecting the most adequate state evolution via Bayesian shrinkage successfully avoids overfitting. The MCS procedure suggests that improvements of using our proposed approach with respect to predictive accuracy, while being small in absolute value, are significant in many cases. We observe a lack of gray cells that signal exclusion from the MCS particularly for TVP-VAR in the respective upper part of the table. Two models worth being mentioned explicitly are the TVP-VAR featuring the observed and three latent effect modifiers and Markov switching, and the one without observed but six latent effect modifiers alongside Markov switching indicators. These two models are among the superior model set for both forecast horizons, point and density forecasts, and across all maturities (apart from one exclusion from the set for the first, and two for the latter). This points towards very consistent and versatile performance of the proposed model.

We now turn to considering model-specific forecasting accuracy with the aim to find the best performing models for point and density forecasts and for the two different forecast horizons. The overall winner for point forecasts at the one-month horizon, on average, is the flexible NS-VAR model specification featuring the exogenous variables, one latent factor per equation and the Markov switching processes. While this specification yields RMSEs that are 23 percent lower than those of the benchmark specification, it must be acknowledged that most competing specifications exhibit values that are similar in magnitude. At the one-quarter-ahead horizon, the one-factor NS-VAR without exogenous variables and Markov switching processes also improves upon the benchmark, albeit at a much smaller margin. It is worth stressing that these predictive gains are mostly found for longer maturities while for shorter maturities, accuracy gains are muted.

We proceed with our findings for density forecasts. As mentioned above, the VAR specifications overall exhibit more favorable relative LPBFs compared to the NS-VARs. This is easily observable by noting that most bold values (in parentheses) are located in the lower panel of Table 1 and the corresponding gray shaded cells mostly in the upper panel. In terms of average performance at the one-month horizon, we find that the TVP-VAR specification with one factor but no exogenous variables or Markov switching performs best, closely followed by the most flexible specification with one unobserved

factor including Markov switching. The MCS procedure confirms these findings, with many of our proposed specifications being included in the superior model set. For one-quarter-ahead forecasts, the TVP-VAR with exogenous variables, one latent factor and Markov switching shows the largest gains in terms of density forecasts. Similarly to the point forecasting results, these gains are mostly obtained in terms of the long end of the yield curve, while improvements in forecast accuracy at shorter horizons are negligible. It is also worth mentioning that while the joint forecast performance of the constant parameter AR(3) is poor, it is a strong competitor with respect to density forecasts of individual maturities.

The main take away from this discussion is that our approach is highly competitive to established models. Especially for longer maturities and both forecast horizons, it pays off to control for structural breaks using our general framework for modeling TVPs in multivariate regression models. Given the substantial flexibility our approach offers, the results clearly show that we do not face overfitting issues, suggesting that our horseshoe prior successfully shrinks irrelevant effect modifiers towards zero.

3.4 Determinants of time-variation in the coefficients

Including observed and latent effect modifiers allows for investigating the sources of time-variation in coefficients, and thus the driving factors of improvements in predictive accuracy. We carry out a detailed analysis of these determinants below.

Because of its favorable forecasting properties, we choose the NS-VAR with one latent factor per equation to investigate the driving forces of parameter variation over the full estimation sample. Recall that the choice of $\boldsymbol{z}_t$ for this specification translates into a single equation-specific latent factor (τ_{jt}), an equation-specific Markov switching indicator S_{jt} and the three observed early warning indicators.

To illustrate the observed and latent factors, we transform each column vector in $\boldsymbol{z}_t$ ex-post such that it is bounded between zero and one. This allows for comparisons even though some blocks of the respective matrices are not econometrically identified.[13] We

[13] In particular, this is the case for the product $\boldsymbol{\Lambda}_\tau \boldsymbol{\tau}_t$. Note that we do not face this issue for $\boldsymbol{\Lambda}_r \boldsymbol{r}_t$ and $\boldsymbol{\Lambda}_S \boldsymbol{S}_t$ because both are either observed or already bounded between zero and one, and their scale and sign are identified.

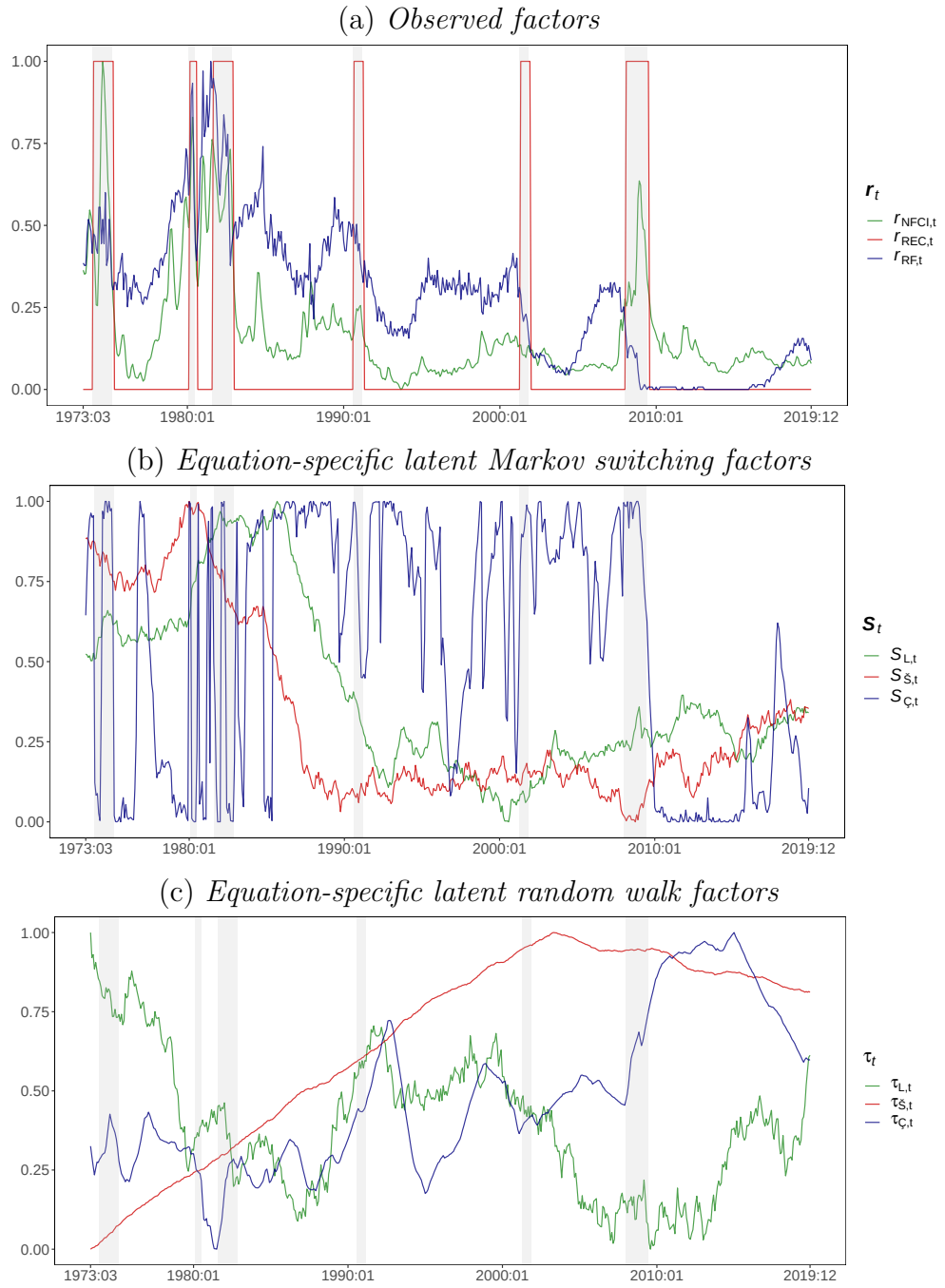
therefore exploit the fact that introducing any invertible $R \times R$ matrix does not alter the likelihood of the model, since $\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{z}_t = \boldsymbol{\Lambda}\boldsymbol{z}_t$. Define $\boldsymbol{U}$ as diagonal matrix such that the maximum of each $\boldsymbol{z}_j$ (for $j = 1, \ldots, R$) corresponds to one, the minimum is zero, $\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\boldsymbol{U}^{-1}$ and $\tilde{\boldsymbol{z}}_t = \boldsymbol{U}\boldsymbol{z}_t$. This simple linear transformation allows for assessing the relative movement of the indicators in $\tilde{\boldsymbol{z}}_t$ without affecting overall dynamics.

Figure 2 displays the evolution of the normalized indicators in $\tilde{\boldsymbol{z}}_t$ over time. First, we focus on the features of the observed effect modifiers depicted in the upper panel (a). The normalized NFCI index peaks during the oil crisis in the mid 1970s, while exhibiting a stable evolution at a quite low level during the Great Moderation (the period from 1990 until to the onset of the global financial crisis in 2007). Spikes in $r_{\text{NFCI},t}$ tend to coincide with recessionary episodes, indicated by the binary recession indicator $r_{\text{REC},t}$. The risk-free interest rate $r_{\text{RF},t}$ can be related to the monetary policy stance. Early in the sample we observe substantial increases, peaking during the Volcker disinflation in the early 1980s. Subsequently, large and abrupt decreases are notable during recessions, while an overall decreasing trend is observable. Before turning to the latent indicators in $\boldsymbol{S}_t$ and $\boldsymbol{\tau}_t$, note that the respective elements S_{jt} and τ_{jt} are equation-specific with $j \in \{L, \check{S}, \acute{C}\}$ referring to the level, slope and curvature of the yield curve. The middle panel (b) indicates the posterior median of the three Markov switching factors collected in $\boldsymbol{S}_t$. The lower panel (c) shows the posterior medians of the three (transformed) gradually changing latent random walk factors in $\boldsymbol{\tau}_t$.

Several features of the latent indicators are worth highlighting. Each of the latent quantities exhibits distinct dynamics and thus carries information in addition to the observed indicators. Examining the posterior means, which are essentially the unconditional posterior probabilities that a given Markov indicator equals one, we observe that both $S_{L,t}$ and $S_{\check{S},t}$ evolve comparatively gradually over time. Before the Volcker disinflation, both tend to increase with posterior medians above 0.5, indicating that regime 1 is more likely. After 1985, we observe a major shift towards regime 0. Interestingly, for $S_{\check{S},t}$ this transition appears immediately after 1985, while we detect a notable delay in $S_{L,t}$.

During the Great Moderation both indicators tend to remain associated with regime 0. The indicator associated with the middle segment of the yield curve, $S_{\acute{C},t}$, by con-

Figure 2: Evolution of the normalized effect modifiers $\tilde{\boldsymbol{z}}_t = \boldsymbol{U}\boldsymbol{z}_t$ over time.



Notes: Results are based on a TVP-NS-VAR model specification $\boldsymbol{z}_t = (\boldsymbol{r}_t', \boldsymbol{S}_t', \boldsymbol{\tau}_t')'$, $\delta = R_\tau/M = 1$, $P = 3$, and using 15,000 MCMC draws. Panel (a) shows the normalized observed factors $r_{\text{NFCI},t}, r_{\text{REC},t}$ and $r_{\text{RF},t}$ (collected in $\boldsymbol{r}_t$), panel (b) the posterior mean of the latent Markov switching factors, $S_{L,t}, S_{\check{S},t}$ and $S_{\c{C},t}$ (collected in $\boldsymbol{S}_t$) and panel (c) the posterior median of the latent random walk factors, $\tau_L, \tau_{\check{S}}$ and $\tau_{\c{C}}$ (collected in $\boldsymbol{\tau}_t$). Note that S_{jt} and τ_{jt} are equation-specific latent quantities with $M(=3)$ endogenous variables, $j \in \{L, \check{S}, \c{C}\}$. $S_{L,t}$ and $\tau_{L,t}$ correspond to the first equation, $S_{\check{S},t}$ and $\tau_{\check{S},t}$ to the second, while $S_{\c{C},t}$ and $\tau_{\c{C},t}$ to the third. The gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period: 1973:01 to 2019:12. Vertical axis: normalized values. Horizontal axis: months.

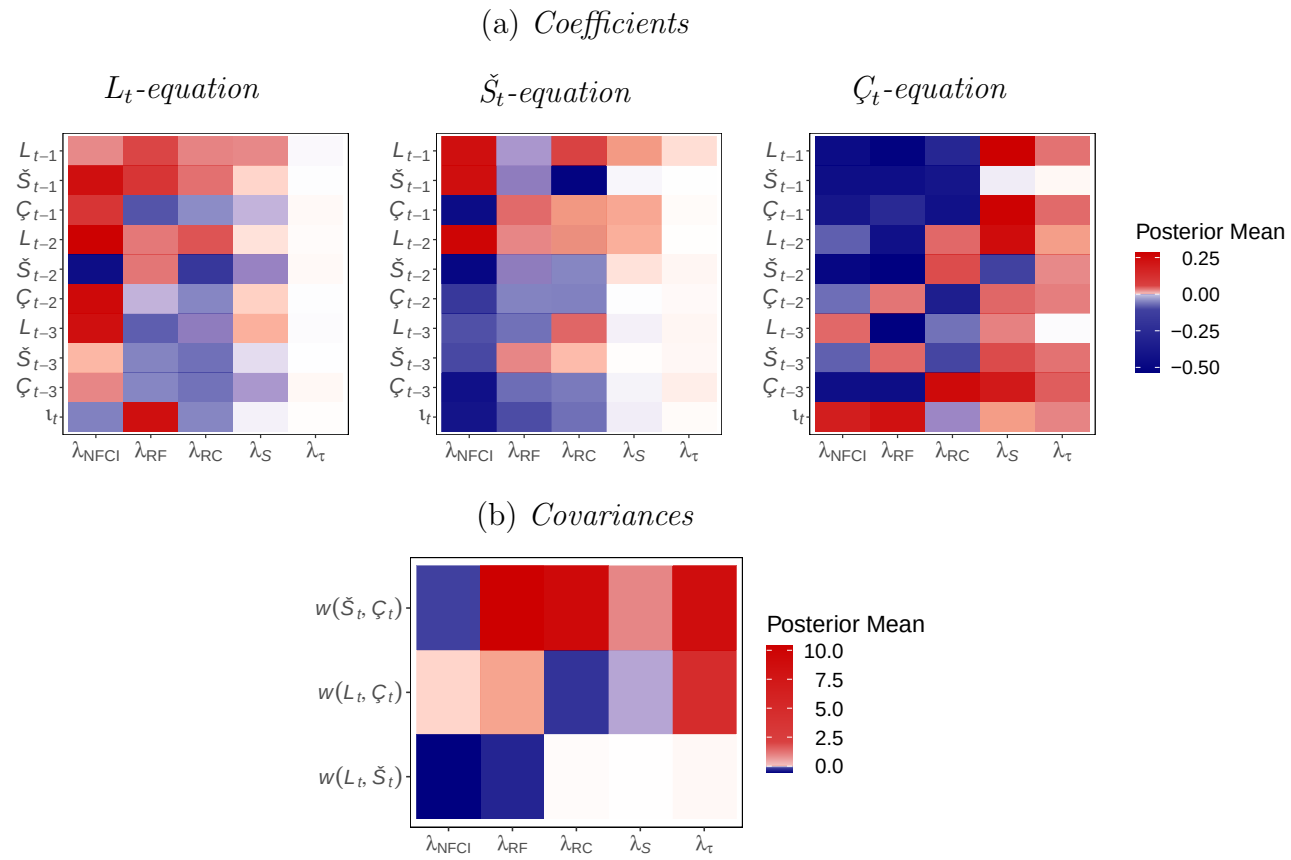
trast, transitions between regimes at a higher frequency. Particularly during the Volcker disinflation we observe mixed patterns and no clear or steady tendency towards a single regime. This changes between 1990 and 1997, where the posterior mean of $S_{\zeta,t}$ is consistently above 0.5, albeit with several high-frequency movements. In the aftermath of the global financial crisis, with short-term rates approaching the zero lower bond, $S_{\zeta,t}$ switches abruptly into regime 0. Conditional on the respective loadings in $\mathbf{\Lambda}_S$ being non-zero, this feature would directly relate to the observed narrowing spread of the yield curve and accompanying structural breaks in coefficients of the equation related to the curvature of the yield curve.

We observe several interesting features of the unobserved factors in $\boldsymbol{\tau}_t$. While $\tau_{L,t}$ is noisy and indicates substantial high-frequency movements, $\tau_{\check{S},t}$ and $\tau_{\zeta,t}$ are much smoother. The factor governing coefficients in the level-equation of the yield curve peaks early in the sample, followed by a decline between 1980 and 1990. After a brief increase and stabilization between 1990 and 2000, we see gradual declines until the global financial crisis starting in 2007. Since then, the factor shows upward trending movement, with several high-frequency troughs. By contrast, the unobserved factor related to $\check{S}_t$ exhibits approximately linear trending behavior from the beginning of the sample until the early 2000s, where it plateaued. After 2010, a gradual but moderate decrease is visible. The factor $\tau_{\zeta,t}$ is comparable to $\tau_{L,t}$, albeit with several differences. While several peaks coincide, we also find adverse movements, for instance in the brief early 1980s recession and after 2000. Interestingly, high-frequency movements are muted when compared to $\tau_{L,t}$.

The preceding discussion of $\tilde{\boldsymbol{z}}_t$ must be considered in light of the rescaled loadings in $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\boldsymbol{U}^{-1}$. The matrix $\tilde{\mathbf{\Lambda}}$ translates the law of motion captured in $\tilde{\boldsymbol{z}}_t$ to the coefficients in $\boldsymbol{\beta}_t$ by acting either as amplifier or attenuator. Figure 3 shows the posterior mean of the rescaled loadings $\mathbf{\Lambda}\boldsymbol{U}^{-1}$ and allows to assess which elements in $\tilde{\boldsymbol{z}}_t$ determine time-variation in the TVPs. We differentiate the loadings along two dimensions. Panel (a) shows the effect modifiers related to the VAR coefficients, while panel (b) depicts the block of $\tilde{\mathbf{\Lambda}}$ related to the covariances (stored in $\boldsymbol{q}_t$).

Assessing the loadings in $\mathbf{\Lambda}_r$ for the L_t-equation reveals that a major part of the coefficients loads strongly positive on $r_{\text{NFCI},t}$. In the case of $r_{\text{RF},t}$ and $r_{\text{REC},t}$ the patterns

Figure 3: Heat maps for rescaled loadings in $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\boldsymbol{U}^{-1}$.



Notes: $\tilde{\mathbf{\Lambda}}$ translates the law of motion captured in $\tilde{\boldsymbol{z}}_t$ to (a) the VAR coefficients in $\boldsymbol{\beta}_t$, and (b) the covariances stored in $\boldsymbol{q}_t$, based on a TVP-NS-VAR model specification with $\boldsymbol{z}_t = (\boldsymbol{r}_t', \boldsymbol{S}_t', \boldsymbol{\tau}_t')'$, $\delta = R_\tau/M = 1$ and $P = 3$. ι_t denotes the time-varying intercept, and $L_{t-p}, \check{S}_{t-p}, \check{C}_{t-p}$ for $p = 1, 2, 3$ the lagged variables $L_t, \check{S}_t$ and $\check{C}_t$, respectively. $\boldsymbol{\lambda}_r$ relates to elements of $\tilde{\mathbf{\Lambda}}$ associated with $r_{\text{NFCI},t}, r_{\text{REC},t}$ and $r_{\text{RF},t}$ and the respective equations, $\boldsymbol{\lambda}_S$ denotes loadings related to a single Markov switching factor collected in $\boldsymbol{S}_t$, and $\boldsymbol{\lambda}_\tau$ loadings corresponding to latent random walk factors in $\boldsymbol{\tau}_t$. Note that S_{jt} and τ_{jt} for $j \in \{L, \check{S}, \check{C}\}$ are equation-specific quantities, while $\boldsymbol{r}_t$ stays fixed across equations. $w(\check{S}_t, \check{C}_t)$ denotes the contemporaneous relation between the $\check{S}_t$- and $\check{C}_t$-equations, $w(L_t, \check{C}_t)$ and $w(L_t, \check{S}_t)$ are defined analogously. Sample period: 1973:01 to 2019:12.

are more mixed. For $r_{\text{NFCI},t}$ we find a maximum loading of around 0.25 for $\check{S}_{t-1}$, L_{t-2}, $\check{C}_{t-2}$ and L_{t-3}, while for $r_{\text{RF},t}$ and $r_{\text{REC},t}$ some coefficients load moderately positive (for example, the first own lag L_{t-1}) and others moderately negative (for example, loadings related to lags of the curvature factor). The loadings $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_\tau$ related to estimated latent factors exhibit only modest relevance for defining the law of motion in coefficients related to L_t. In particular, the amplifiers for $\boldsymbol{\tau}_t$ are shrunk heavily towards zero, implying that low frequency movements in the respective coefficients are either already captured by other measures, or irrelevant for coefficients in the L_t-equation. The same is true for the indicators in $\boldsymbol{S}_t$ and $\boldsymbol{\tau}_t$ for the case of the $\check{S}_t$-equation.

Turning to observed factors, the corresponding factor loadings are positive for lower-order lags while strongly negative for higher-order lags. By contrast, we find mostly negative loadings for observed factors for first-lags in the $\acute{C}_t$-equation. Loadings related to higher-order lags are mixed, and no clear patterns are observable. Interestingly, the curvature factor is the only equation where we detect substantial loadings on the latent factors, with most loadings showing a positive sign.

While unobserved factors appear to be less important for coefficients in the conditional mean of the model, they play an important role for the covariances, as indicated in panel (b). The covariance between $\check{S}_t$ and $\acute{C}_t$, $w(\check{S}_t, \acute{C}_t)$, shows pronounced loadings on all other effect modifiers than NFCI, where we observe a modest negative loading. A mixed pattern emerges for the L_t and $\acute{C}_t$ covariance, $w(L_t, \acute{C}_t)$, with some positive and negative measures. The contemporaneous relationship $w(L_t, \check{S}_t)$ between L_t and $\check{S}_t$, marks a particularly interesting case, with modestly negative loadings on NFCI and RF, while REC and the unobserved factor loadings are close to zero.

Summarizing, we find that observed factors often load strongly on the coefficients of all equations. While the Markov switching indicator appears to be important particularly for the L_t- and $\acute{C}_t$-equations, its loadings are muted for the coefficients of $\check{S}_t$. Another interesting aspect is that conditional on observed effect modifiers, the gradually evolving coefficients captured by $\boldsymbol{\tau}_t$ are mostly irrelevant for the L_t- and $\check{S}_t$-equations, different to strong positive loadings in the case of the $\acute{C}_t$-equation and the covariances. Additional results showing the actual regression coefficients are provided in the Online Appendix.

3.5 Low frequency relations between the Nelson-Siegel factors

To assess what our framework implies on the relations between the factors that determine the yield curve, we compute the low-frequency relationship between the L_t, $\check{S}_t$ and $\acute{C}_t$. We choose this long-run correlation measure for two reasons. First, it allows to illustrate movements in time-varying coefficients (i.e., in the transmission channels) and changes in the error variances in a single indicator over time. Second, this measure compresses information of all coefficients in a structurally meaningful way since it isolates long-run trends and correlations from short-run fluctuations (Sargent and Surico, 2011; Kliem et al.,

2016). The Online Appendix collects additional results for the reduced form coefficients.

To construct the measure, in terms of the generic notation in Eq. (1), we rewrite the VAR(P) in its state-space VAR(1) form. In what follows, the observation equation is given by $\boldsymbol{y}_t = \boldsymbol{J}\boldsymbol{Y}_t$, while the state equation is defined as $\boldsymbol{Y}_t = \boldsymbol{B}_t\boldsymbol{Y}_{t-1} + \boldsymbol{v}_t$ with $\boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Xi}_t)$. Here, $\boldsymbol{J}$ maps a $K \times 1$ vector $\boldsymbol{Y}_t = (\boldsymbol{y}_t', \ldots, \boldsymbol{y}_{t-P+1}')'$ to $\boldsymbol{y}_t$, $\boldsymbol{B}_t$ collects the elements in $\boldsymbol{\beta}_t$ in the upper $M \times K$ block and defines identities otherwise. Similarly, the $K \times K$ variance-covariance matrix $\boldsymbol{\Xi}_t$ collects the elements of $\boldsymbol{\Sigma}_t$ in the upper-left $M \times M$ block and is zero otherwise. We follow Sargent and Surico (2011) and first calculate the spectral density $\boldsymbol{\Phi}_t(0)$ of $\boldsymbol{y}_t$ at a zero frequency which coincides with the unconditional variance-covariance matrix of $\boldsymbol{y}_t$:

$$\boldsymbol{\Phi}_t(0) = \boldsymbol{J}\left(\boldsymbol{I}_K - \boldsymbol{B}_t\right)\boldsymbol{\Xi}_t(\boldsymbol{I}_K - \boldsymbol{B}_t')^{-1}\boldsymbol{J}', \quad \text{for } t = 1, \ldots, T.$$

Next, we transform the covariances of $\boldsymbol{\Phi}_t$ into a correlation measure for each period t and each variable combination i, j ($i \neq j$ and $i, j = 1, \ldots, M$):

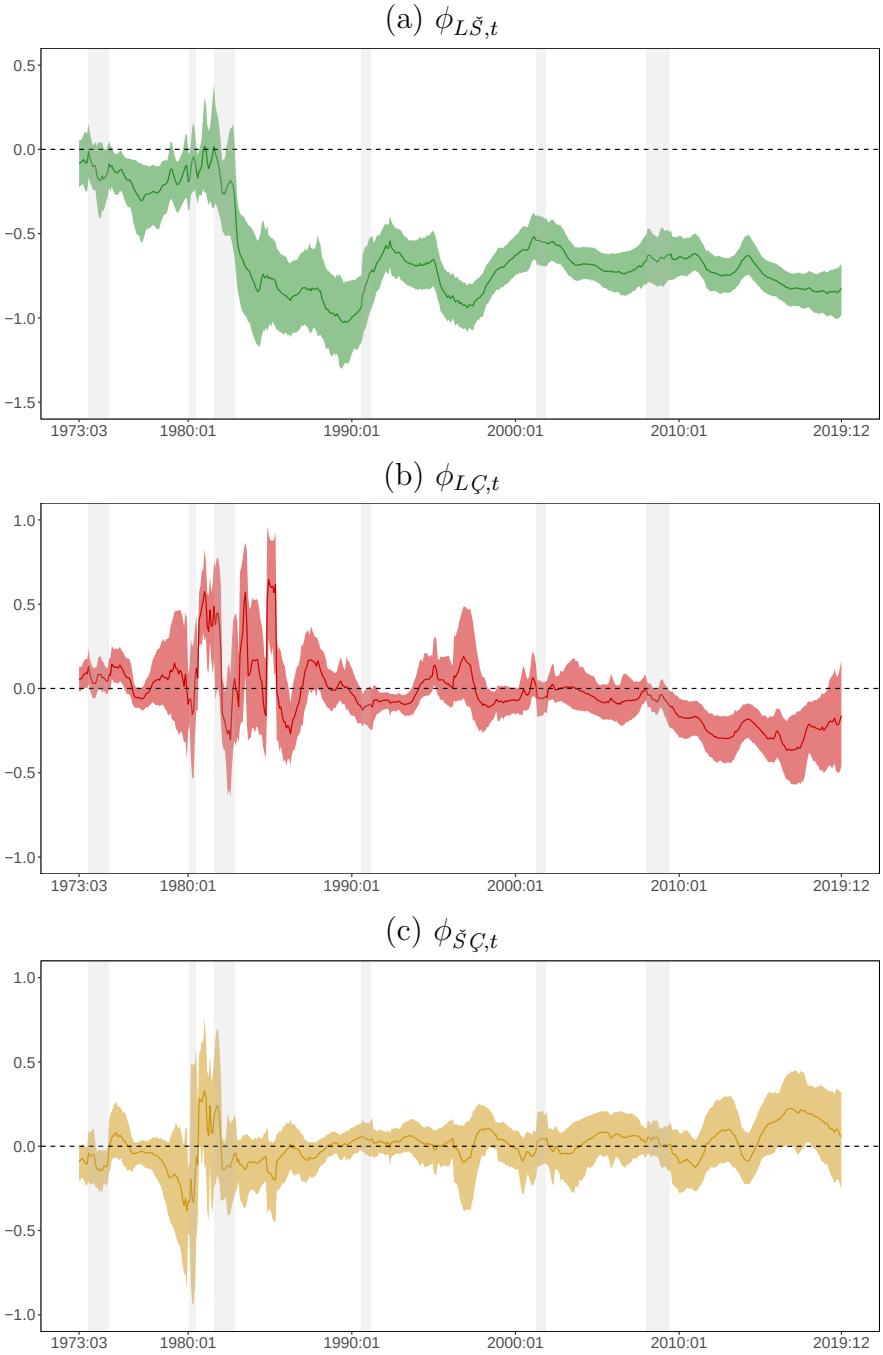
$$\phi_{ij,t} = \frac{\Phi_{ij,t}(0)}{\Phi_{jj,t}(0)}.$$

The measure $\phi_{ij,t}$ describes the long-run relations between variables i and j at each point in time, and is displayed in Figure 4 for the model specification TVP-NS-VAR with $\boldsymbol{z}_t = (\boldsymbol{r}_t', \boldsymbol{S}_t', \boldsymbol{\tau}_t')'$, $R_{\tau j} = 1$ and $P = 3$. Note that the variables enter our model in differences. Hence, Figure 4 depicts the low frequency relations of *changes* in the level, slope and curvature of the yield curve.

We observe several interesting periods characterized by structural breaks. First, the relationship between the level and slope of the yield curve was close to zero until the Volcker disinflation of the 1980s. After this period, we can identify an abrupt decrease to significantly negative values. The long-run correlation stays negative until the end of the sample, with minor low-frequency movements.

Second, for most of the sample the long-run coefficient between level and curvature of the yield curve is insignificant. Substantial structural breaks are detectable in the early 1980s. Again, this coincides with a shift in the US monetary policy regime. Be-

Figure 4: Posterior median and the 68 percent credible set for pairwise long-run correlations.



Notes: Panel (a) $\phi_{L\check{S},t}$, (b) $\phi_{L\acute{C},t}$ and (c) $\phi_{\check{S}\acute{C},t}$, based on a TVP-NS-VAR with $\boldsymbol{z}_t = (\boldsymbol{r}_t', \boldsymbol{S}_t', \boldsymbol{\tau}_t')'$, $\delta = R\tau/M = 1$ and $P = 3$. The solid lines denote the posterior medians, the black dashed line the zero line, and the color shaded bands the 68 percent posterior coverage interval, while the gray vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. $\phi_{ij,t}$ is the long-run correlation for the variables i and j at time t, for $i, j \in \{L, \check{S}, \acute{C}\}$. Results are based on 15,000 MCMC draws. Sample period: 1973:01 to 2019:12. Vertical axis: correlation measurements. Horizontal axis: months.

tween the two recessions in the early 1980s and in the recovery period afterwards, we note a strong positive relationship between the level and curvature of the yield curve. Afterwards, during the Great Moderation, there are mostly insignificant values. This ends during the Great Recession, where the relationship is estimated to be significantly negative. This finding may be linked to short-term interest rates approaching the zero lower bound rapidly. The trend reverses later in the sample, with the Federal Reserve conducting several subsequent rate hikes starting at the end of 2015. Turning to the relationship between the slope and curvature of the yield curve, we again find an insignificant relationship for most of the sample. Large breaks are observable in the period between the two 1980s recessions, but different to the relationship between the level and curvature of the yield curve, this is not visible in the subsequent recovery period. Interestingly, we estimate a significantly positive relationship in the period of the aforementioned rate hikes by the Federal Reserve starting in 2015.

4 Conclusions

In this paper we develop a flexible multivariate econometric framework and use it to model the term structure of US interest rates. The recent literature on yield curve dynamics has highlighted different forms of parameter change to be of relevance. To cope with this inherent uncertainty with respect to how coefficients behave over time we propose methods for automatically selecting adequate state equations in TVP-VAR models in a data-driven manner. The TVPs are assumed to depend on a set of observed and unobserved covariates, also known as effect modifiers. As unobserved covariates, we consider a set of low dimensional latent factors that follow a random walk, alongside Markov switching indicators that allow for abrupt structural breaks. Our model nests several alternatives commonly used in the literature on modeling macroeconomic and financial time series. To choose between state equations, we use a hierarchical Bayesian global-local shrinkage prior on the most flexible specification.

To assess whether our methods pay off relative to simpler alternatives (which are nested in our general framework), we carry out a thorough predictive exercise. We show

that our techniques produce favorable point and density forecasts. The performance is specific to the information set used in the underlying TVP-VAR and appears to be more pronounced for density forecasts. Using the MCS procedure we also show that our set of general models is very often included in the superior model set. This finding is particularly pronounced for the long end of the yield curve.

The forecasting exercise illustrates that our approach produces very competitive forecasts without increasing the risk of overfitting, while providing a framework to trace the sources of time-variation. To illustrate this we consider a full-sample analysis of structural breaks in the relationship between the level, slope and curvature of the US yield curve. We detect several interesting patterns in abrupt and gradual time-variation patterns in long-run cross-variable relations. These changes appear to be specific to the monetary regime and the state of the business cycle.

It is worth stressing that the methods have been applied specifically to US interest rate data. These time series feature appreciable structural breaks (such as during the Volcker disinflation or the recent period at the ZLB) and thus call for models which flexibly capture parameter time-variation. Nevertheless, our approach is very general and widely applicable to other datasets such as a standard macroeconomic dataset in the spirit of McCracken and Ng (2016).

Finally, it is worth stressing that we have focused on a particular prior, the horseshoe, in this paper. Depending on the dataset under consideration the empirical performance of the approach might be improved further by considering other priors. Since prior robustness is an important issue we leave a systematic analysis of alternative priors to be used in our general model for future research.

Acknowledgments: The authors thank the co-editor Herman van Dijk, an anonymous associate editor, three anonymous referees, and participants at the 4th Annual Workshop on Financial Econometrics in Örebro (in particular Barbara Rossi and Gary Koop) for helpful comments and suggestions. This research was supported by the Jubiläumsfonds of the Oesterreichische Nationalbank (OeNB, project 18127) and by the Austrian Science Fund (FWF, project ZK 35).

References

Adams, P. A., T. Adrian, N. Boyarchenko, and D. Giannone (2021): "Forecasting macroeconomic risks," *International Journal of Forecasting*, 37(3), 1173–1191.

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable growth," *American Economic Review*, 109(4), 1263–1289.

Aguilar, O., and M. West (2000): "Bayesian dynamic factor models and portfolio allocation," *Journal of Business & Economic Statistics*, 18(3), 338–357.

Bauer, M. D., and G. D. Rudebusch (2020): "Interest rates under falling stars," *American Economic Review*, 110(5), 1316–1354.

Bhattacharya, A., D. Pati, N. S. Pillai, and D. B. Dunson (2015): "Dirichlet–Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, 110(512), 1479–1490.

Bianchi, F., H. Mumtaz, and P. Surico (2009): "The great moderation of the term structure of UK interest rates," *Journal of Monetary Economics*, 56(6), 856–871.

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

Cadonna, A., S. Frühwirth-Schnatter, and P. Knaus (2020): "Triple the gamma—A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models," *Econometrics*, 8(2), 20.

Carriero, A., T. E. Clark, and M. Marcellino (2020): "Capturing macroeconomic tail risks with Bayesian vector autoregressions," *Federal Reserve Bank of Cleveland*, Working Paper 20-02R.

——— (2021): "No-arbitrage priors, drifting volatilities, and the term structure of interest rates," *Journal of Applied Econometrics*, 36(5), 495–516.

Carriero, A., G. Kapetanios, and M. Marcellino (2012): "Forecasting government bond yields with large Bayesian vector autoregressions," *Journal of Banking & Finance*, 36(7), 2026–2047.

Carter, C. K., and R. Kohn (1994): "On Gibbs sampling for state space models," *Biometrika*, 81(3), 541–553.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010): "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), 465–480.

Chan, J. C., E. Eisenstat, and R. Strachan (2020): "Reducing the state space dimension in a large TVP-VAR," *Journal of Econometrics*, 218(1), 105–118.

Cogley, T., and T. J. Sargent (2005): "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8(2), 262–302.

Coroneo, L., K. Nyholm, and R. Vidova-Koleva (2011): "How arbitrage-free is the Nelson–Siegel model?," *Journal of Empirical Finance*, 18(3), 393–407.

Del Negro, M., D. Giannone, M. P. Giannoni, and A. Tambalotti (2019): "Global trends in interest rates," *Journal of International Economics*, 118, 248–262.

Diebold, F. X., and C. Li (2006): "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 130(2), 337–364.

Diebold, F. X., C. Li, and V. Z. Yue (2008): "Global yield curve dynamics and interactions: A dynamic Nelson–Siegel approach," *Journal of Econometrics*, 146(2), 351–363.

Frühwirth-Schnatter, S. (1994): "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, 15(2), 183–202.

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

George, E., and R. McCulloch (1993): "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

——— (1997): "Approaches for Bayesian variable selection," *Statistica Sinica*, 7(2), 339–373.

Gerlach, R., C. Carter, and R. Kohn (2000): "Efficient Bayesian inference for dynamic mixture models," *Journal of the American Statistical Association*, 95, 819–828.

Geweke, J., and G. Amisano (2010): "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26(2), 216–230.

Geweke, J., and G. Zhou (1996): "Measuring the pricing error of the arbitrage pricing theory," *Review of Financial Studies*, 9(2), 557–587.

Giordani, P., and R. Kohn (2008): "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business & Economic Statistics*, 26(1), 66–77.

Griffin, J., and P. Brown (2010): "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5(1), 171–188.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007): "The US Treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54(8), 2291–2304.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): "The model confidence set," *Econometrica*, 79(2), 453–497.

Hastie, T., and R. Tibshirani (1993): "Varying-coefficient models," *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–779.

Hauzenberger, N. (2021): "Flexible mixture priors for time-varying parameter models," *Econometrics and Statistics*, 20, 87–108.

Hauzenberger, N., F. Huber, G. Koop, and L. Onorante (2021): "Fast and flexible Bayesian inference in time-varying parameter regression models," *Journal of Business & Economic Statistics*, in press, doi: 10.1080/07350015.2021.1990772.

Hevia, C., M. Gonzalez-Rozada, M. Sola, and F. Spagnolo (2015): "Estimating and forecasting the yield curve using a Markov switching dynamic Nelson and Siegel model," *Journal of Applied Econometrics*, 30(6), 987–1009.

Huber, F., G. Kastner, and M. Feldkircher (2019): "Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models," *Journal of Applied Econometrics*, 34(5), 621–640.

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

Huber, F., G. Koop, and M. Pfarrhofer (2020): "Bayesian inference in high-dimensional time-varying parameter models using integrated rotated Gaussian approximations," *arXiv preprint*, 2002.10274.

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

Kaufmann, S. (2015): "K-state switching models with time-varying transition distributions—Does loan growth signal stronger effects of variables on inflation?," *Journal of Econometrics*, 187(1), 82–94.

Kim, C.-J., and C. R. Nelson (1999): *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications*. MIT Press, Cambridge, MA.

Kliem, M., A. Kriwoluzky, and S. Sarferaz (2016): "On the low-frequency relationship between public deficits and inflation," *Journal of Applied Econometrics*, 31(3), 566–583.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009): "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33(4), 997–1017.

Korobilis, D. (2021): "High-dimensional macroeconomic forecasting using message passing algorithms," *Journal of Business & Economic Statistics*, 39(2), 493–504.

Kowal, D. R. (2021): "Dynamic regression models for time-ordered functional data," *Bayesian Analysis*, 16(2), 459–487.

Liu, P., K. Theodoridis, H. Mumtaz, and F. Zanetti (2019): "Changing macroeconomic dynamics at the zero lower bound," *Journal of Business & Economic Statistics*, 37(3), 391–404.

Lopes, H. F., R. E. McCulloch, and R. S. Tsay (2021): "Parsimony inducing priors for large scale state–space models," *Journal of Econometrics*, in press, doi: 10.1016/j.jeconom.2021.11.005.

Maheu, J. M., and Y. Song (2018): "An efficient Bayesian approach to multiple structural change in multivariate time series," *Journal of Applied Econometrics*, 33(2), 251–270.

Makalic, E., and D. F. Schmidt (2015): "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23(1), 179–182.

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

Mumtaz, H., and P. Surico (2009): "Time-varying yield curve dynamics and monetary policy," *Journal of Applied Econometrics*, 24(6), 895–913.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140(2), 425–449.

Polson, N. G., and J. G. Scott (2010): "Shrink globally, act locally: Sparse Bayesian regularization and prediction," in *Bayesian Statistics*, ed. by J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, vol. 9, pp. 501–538, Oxford, UK. Oxford University Press.

Primiceri, G. (2005): "Time varying structural autoregressions and monetary policy," *Review of Economic Studies*, 72(3), 821–852.

Raftery, A. E., and S. Lewis (1992): "How many iterations in the Gibbs sampler?," in *Bayesian Statistics*, ed. by J. Bernardo, J. Berger, A. Dawid, and A. Smith, vol. 4, pp. 763–773, Oxford, UK. Oxford University Press.

Rudebusch, G. D., E. T. Swanson, and T. Wu (2006): "The bond yield "conundrum" from a macro-finance perspective," *Federal Reserve Bank of San Francisco*, Working Paper 2006-16.

Rudebusch, G. D., and T. Wu (2007): "Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models," *Journal of Money, Credit and Banking*, 39(2-3), 395–422.

Sargent, T. J., and P. Surico (2011): "Two illustrations of the quantity theory of money: Breakdowns and revivals," *American Economic Review*, 101(1), 109–128.

Sims, C. A., and T. Zha (2006): "Were there regime switches in US monetary policy?," *American Economic Review*, 96(1), 54–81.

Stock, J. H., and M. Watson (2011): "Dynamic factor models," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, pp. 35–60. Oxford University Press, Oxford, UK.

Tillmann, P. (2020): "Monetary policy uncertainty and the response of the yield curve to policy shocks," *Journal of Money, Credit and Banking*, 52(4), 803–833.

Online Appendix to: "General Bayesian time-varying parameter VARs for modeling government bond yields"

Manfred M. FISCHER[1], Niko HAUZENBERGER[1,2], Florian HUBER[2] and Michael PFARRHOFER[2]

[1] *Vienna University of Economics and Business*
[2] *University of Salzburg*

A Technical appendix

A.1 Details on the prior setup

Stochastic selection of effect modifiers. Let $\mathbf{\Lambda}_j$ denote a sub-matrix of the free elements in $\mathbf{\Lambda}$ corresponding to the j^{th} equation, and $\boldsymbol{\lambda}_{ji}$ mark the i^{th} column of this matrix. We let $\lambda_{ji,\ell}$ denote the ℓ^{th} element of this vector. The prior is given by:

$$\lambda_{ji,\ell}|c_{ji,\ell}, d_{ji} \sim \mathcal{N}(0, c_{ji,\ell}^2\, d_{ji}^2), \quad c_{ji,\ell} \sim \mathcal{C}^+(0,1), \quad d_{ji} \sim \mathcal{C}^+(0,1).$$

Here, $\mathcal{C}^+(0,1)$ denotes the half Cauchy distribution, and d_{ji} is an equation- and column-specific global shrinkage factor, while $c_{ji,\ell}$ is a local scaling parameter.

Priors on the constant parameters. Let γ_{ji} denote the i^{th} element of the vector $\boldsymbol{\gamma}_j$. The setup is similar to the one of the loadings matrix, and given by the hierarchical structure:

$$\gamma_{ji}|\xi_{ji}, \zeta_j \sim \mathcal{N}(0, \xi_{ji}^2\, \zeta_j^2), \quad \xi_{ji} \sim \mathcal{C}^+(0,1), \quad \zeta_j \sim \mathcal{C}^+(0,1).$$

The hyperparameter ζ_j is an equation-specific global shrinkage factor, while the ξ_{ji}'s are local scalings.

Priors on the state innovation variances. For equation j, let $\boldsymbol{\omega}_j$ denote a $v_j(= j-1+k)$-dimensional vector which stores the diagonal elements in $\mathbf{\Omega}$ associated with the j^{th} equation. This includes the process innovation variances on the k regression coefficients and the $j-1$ covariance parameters in $\tilde{\boldsymbol{Q}}_t$. The square root of the i^{th} element of $\boldsymbol{\omega}_j$, $\sqrt{\omega_{ji}}$, features the following prior hierarchy:

$$\sqrt{\omega_{ji}}|\varpi_{ji}, \vartheta_j \sim \mathcal{N}(0, \varpi_{ji}^2\, \vartheta_j^2), \quad \varpi_{ji} \sim \mathcal{C}^+(0,1), \quad \vartheta_j \sim \mathcal{C}^+(0,1).$$

Choosing a Gaussian prior on the square root of the variance in the first level of the hierarchy implies a Gamma prior on ω_{ji}, with $\omega_{ji}|\varpi_{ji}, \vartheta_j \sim \mathcal{G}\left(0.5, \varpi_{ji}^{-2}\, \vartheta_j^{-2}/2\right)$, see also Frühwirth-Schnatter and Wagner (2010). The hyperparameters ϑ_j and ϖ_{ji} are again equation-specific global and local shrinkage parameters. Furthermore, we set $\boldsymbol{V}_\tau = \boldsymbol{I}_{R_\tau}$

and thus impose shrinkage through the factor loadings in $\mathbf{\Lambda}_\tau$ (see Chan et al., 2020).

Priors for the stochastic volatilities. On the parameters of the state equation of the log-volatility processes μ_j, ψ_j and ς_j^2, we use the setup proposed in Kastner and Frühwirth-Schnatter (2014). That is, we assume a Gaussian prior on the unconditional mean, $\mu_j \sim \mathcal{N}(0, 10^2)$, a Beta prior on the transformed autoregressive parameter, $(\psi_j + 1)/2 \sim \mathcal{B}(5, 1.5)$, and a Gamma prior on the state variance, $\varsigma_j^2 \sim \mathcal{G}(0.5, 0.5)$, while the prior on the initial state is the unconditional distribution for all $j = 1, \ldots, M$.

Priors for the Markov switching indicators. For the equation-specific transition probabilities $\boldsymbol{P}_j$ of Markov switching indicators, we assume that the $(i,i)^{\text{th}}$ element $p_{j,ii}$ arises from a Beta distribution given by:

$$p_{j,ii} \sim \mathcal{B}(e_{i0}, e_{i1}), \quad \text{for } i = 0, 1;\ j = 1, \ldots, M,$$

and hence $p_{j,i\ell} = 1 - p_{j,i\ell}$ for $i \neq \ell$. In the empirical application, we define $e_{00} = e_{11} = 10$ and $e_{01} = e_{10} = 1$, in order to weakly push each S_{jt} towards a single-state a priori.

A.2 Full conditional posterior simulation

To simulate from the full posterior distribution we develop an efficient MCMC algorithm. Since full-system estimation quickly becomes computationally cumbersome, we rely on the structural form of the VAR to allow for an equation-by-equation algorithm used in many recent papers (see also Chan et al., 2020; Huber et al., 2021).

Equation (4) can be simplified to yield:

$$y_{jt} = \boldsymbol{m}_{jt}'(\boldsymbol{\gamma}_j + \tilde{\boldsymbol{\gamma}}_{jt}) + \varepsilon_{jt}, \tag{A.1}$$

where $\boldsymbol{m}_{jt} = (\boldsymbol{x}_t', \boldsymbol{u}_{jt}')'$, $\tilde{\boldsymbol{\gamma}}_{jt} = \boldsymbol{\gamma}_{jt} - \boldsymbol{\gamma}_j$, and $\boldsymbol{\gamma}_{jt}$ refers to the TVPs associated with the j^{th} equation in $\boldsymbol{\gamma}_t$, and $\boldsymbol{\gamma}_j$ denotes the corresponding constant part. All the following steps will be carried out on an equation-by-equation basis and making use of the regression form in Eq. (A.1). For notational simplicity, we assume that all elements in $\boldsymbol{z}_t$ are latent. In light of the discussion in Section 2.3, this implies that $\boldsymbol{z}_{jt} = (\boldsymbol{\tau}_{jt}', S_{jt})'$ and the extension

to include observed factors is trivial.

Sampling $\boldsymbol{z}_{jt}$. Conditional on the remaining quantities of the model, we simulate the latent (random walk and Markov switching) components in $\boldsymbol{z}_{jt}$ after integrating out $\tilde{\boldsymbol{\gamma}}_{jt}$. This is achieved by rewriting Eq. (A.1) as:

$$\tilde{y}_{jt} = \boldsymbol{m}_{jt}' \, \boldsymbol{\Lambda}_j \, \boldsymbol{z}_{jt} + \boldsymbol{m}_{jt}' \, \boldsymbol{\eta}_{jt} + \varepsilon_{jt}, \tag{A.2}$$

with $\tilde{y}_{jt} = y_{jt} - \boldsymbol{m}_{jt}' \, \boldsymbol{\gamma}_j$. Defining $\tilde{\boldsymbol{m}}_{jt}' = \boldsymbol{m}_{jt}' \, \boldsymbol{\Lambda}_j$ and $\hat{\varepsilon}_{jt} = \boldsymbol{m}_{jt}' \, \boldsymbol{\eta}_{jt} + \varepsilon_{jt}$ allows us to cast Eq. (A.2) as a simple linear regression model:

$$\tilde{y}_{jt} = \tilde{\boldsymbol{m}}_{jt}' \, \boldsymbol{z}_{jt} + \hat{\varepsilon}_{jt}, \quad \hat{\varepsilon}_{jt} \sim \mathcal{N}(0, \boldsymbol{m}_{jt}' \, \text{diag}(\boldsymbol{\omega}_j) \, \boldsymbol{m}_{jt} + \exp(h_{jt})). \tag{A.3}$$

This parameterization has the advantage that it does not depend on $\tilde{\boldsymbol{\gamma}}_{jt}$, and $\boldsymbol{z}_{jt}$ can thus be sampled marginally of $\tilde{\boldsymbol{\gamma}}_{jt}$. This improves mixing substantially since $\tilde{\boldsymbol{\gamma}}_{jt}$ and $\boldsymbol{z}_{jt}$ will often be highly correlated (for a detailed discussion of this issue, see Gerlach et al., 2000; Giordani and Kohn, 2008).

Depending on the precise laws of motion for the elements in $\boldsymbol{z}_{jt}$, standard algorithms can now be used. In this paper, we use two different law of motions. For the latent random walk factors in $\boldsymbol{\tau}_{jt}$, we use the forward filtering backward sampling algorithm outlined in Carter and Kohn (1994), and Frühwirth-Schnatter (1994). In case of the latent Markov switching factors in S_{jt}, we use the algorithm outlined in Kim and Nelson (1999). Both algorithms are well known and relevant details may be found in the original papers. Here, it suffices to note that in both cases, sampling the latent states is computationally easy since the state-space is low dimensional with $R \ll N$. In this setting, sampling the factors equation-wise can be carried out in $O(R^3)$ steps, a substantial computational improvement relative to the $O(N^3)$ steps necessary to estimating an unrestricted TVP regression (see also the discussion in Chan et al., 2020).[14]

Sampling the state innovation variances. To obtain draws for the state innovation variances, reconsider Eq. (1) and draw them conditional on the observed/latent states and

[14]It is worth mentioning that the $O(N^3)$ statement is true for the precision sampler and differs for forward-filtering backward-sampling algorithms.

the factor loadings. We let $\eta_{ji,t}$ denote the shock to the i^{th} coefficient in $\tilde{\boldsymbol{\gamma}}_t$ with respect to the j^{th} equation. The posterior of the state innovation variances is a generalized inverse Gaussian (GIG) distribution:[15]

$$\omega_{ji}|\varpi_{ji}, \vartheta_j \sim \mathcal{GIG}\left(\frac{1-T}{2}, \sum_{t=1}^{T} \eta_{ji,t}^2, \varpi_{ji}\,\vartheta_j\right).$$

Sampling $\boldsymbol{\Lambda}_j$ and $\boldsymbol{\gamma}_j$ jointly. Similarly to $\boldsymbol{z}_{jt}$, we sample the non-zero loadings in $\boldsymbol{\Lambda}$ and the time-invariant coefficients marginally of $\boldsymbol{\gamma}_{jt}$ by using equation-by-equation estimation. The observation equation for equation j (conditional on $\boldsymbol{z}_{jt}$) can be written as a standard regression model:

$$y_{jt} = \hat{\boldsymbol{m}}_{jt}'\,\hat{\boldsymbol{\gamma}}_j + \hat{\varepsilon}_{jt}, \tag{A.4}$$

where $\hat{\boldsymbol{m}}_{jt} = (\boldsymbol{m}_{jt}', (\boldsymbol{z}_{jt} \otimes \boldsymbol{m}_{jt})')'$ is an Rv_j^2-dimensional vector of covariates, and $\hat{\boldsymbol{\gamma}}_j = (\boldsymbol{\gamma}_j', \text{vec}(\boldsymbol{\Lambda}_j)')'$ denoting an Rv_j^2-dimensional coefficient vector. The posterior of $\hat{\boldsymbol{\gamma}}_j$ is Gaussian with well known moments.

Sampling the stochastic volatilities. The latent log-volatility processes can again be sampled equation-by-equation via standard state-space methods based on an auxiliary mixture approximation of the respective log χ_1^2 error term (see Omori et al., 2007, for details). This step is implemented using the `R`-package `stochvol`.

Sampling the horseshoe prior hyperparameters. Our assumptions imply analogous horseshoe priors for the factor loadings matrix, the constant part of the coefficients and the square root of the state innovation variances. Sampling is carried out using the algorithm described in Makalic and Schmidt (2015).

Our specification of the horseshoe prior on $\boldsymbol{\Lambda}_j$, $\boldsymbol{\gamma}_j$ and the square root of $\boldsymbol{\omega}_j$ in Section 2.4 for a generic parameter b_j for $j = 1, \ldots, K$ is:

$$b_j|c_j, d \sim \mathcal{N}(0, c_j^2\, d^2), \quad c_j \sim \mathcal{C}^+(0,1), \quad d \sim \mathcal{C}^+(0,1).$$

[15] The generalized inverse Gaussian distribution is specified such that its density function is proportional to $f(x) = x^{\lambda-1}\exp(-(\chi/x + \psi x)/2)$ for a random variable $x \sim \mathcal{GIG}(\lambda, \chi, \psi)$.

We rely on this prior in its auxiliary representation as in Makalic and Schmidt (2015) for efficient sampling of the local (c_j) and global (d) shrinkage parameters:

$$c_j^2|e_j \sim \mathcal{G}^{-1}(0.5, 1/e_j), \quad d^2|f \sim \mathcal{G}^{-1}(0.5, f^{-1}), \quad e_j \sim \mathcal{G}^{-1}(0.5, 1), \quad f \sim \mathcal{G}^{-1}(0.5, 1).$$

Here, $\mathcal{G}^{-1}$ denotes the inverse Gamma distribution. This setup yields the following conditional posterior distributions:

$$c_j^2|b_j, d, e_j \sim \mathcal{G}^{-1}\left(1, \frac{1}{e_j} + \frac{b_j^2}{2d^2}\right), \quad d^2|b_j, c_j, f \sim \mathcal{G}^{-1}\left(\frac{K+1}{2}, \frac{1}{f} + \sum_{j=1}^{K} \frac{b_j^2}{2c_j^2}\right),$$
$$e_j|c_j \sim \mathcal{G}^{-1}\left(1, 1 + c_j^{-2}\right), \quad f|d \sim \mathcal{G}^{-1}\left(1, 1 + d^{-2}\right).$$

B Additional results

This Appendix contains additional results based on synthetic data and for the reduced form coefficients. Section B.1 presents simulation based evidence for our proposed econometric framework. Section B.2 display the coefficients obtained by multiplying the factor loadings with the observed/latent factors based on the relationship established in Eq. (1).

B.1 Simulation evidence

To evaluate the merits of our proposed approach, we conduct a simulation exercise. We simulate data from the following data generating process (DGP):

$$y_t = \sum_{i \in \{r,S,\tau,c\}} \overbrace{(\beta_i + \tilde{\beta}_{it})}^{=\beta_{it}} x_{1t} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, 0.1), \qquad \text{for } t = 1, \ldots, 200,$$

$$\tilde{\beta}_{it} = \lambda_i z_{it} + \eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, \varsigma_i), \qquad \text{for } i \in \{r, S, \tau, c\}.$$

The constant part of the parameters (initial state) is set to $\boldsymbol{\beta} = (0.5, 0, 2, 1)'$. The latent effect modifiers z_{it} with indicator $i \in \{r, S, \tau, c\}$ joint with the loadings λ_i determine the type of variation in the coefficient governed by observed variables (r), a binary indicator (S), an unknown random walk motion (τ) or a constant parameter (c). They are simulated as follows:

$$z_{rt} = z_{rt-1} + \varepsilon_{rt}, \quad \varepsilon \sim \mathcal{N}(0, 0.1),$$

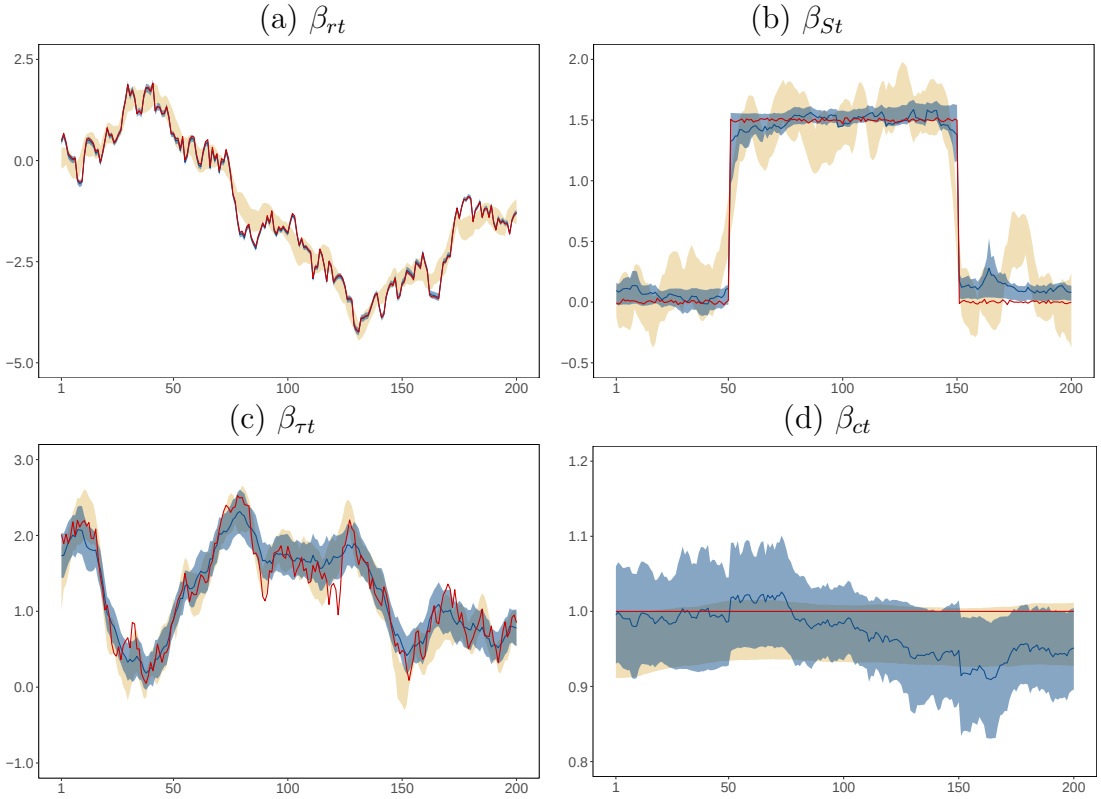
$$z_{\tau t} = z_{\tau t-1} + \varepsilon_{\tau t}, \quad \varepsilon \sim \mathcal{N}(0, 0.1),$$

$$z_{St} = \begin{cases} 1 & \text{if } 50 < t \leq 150 \\ 0 & \text{otherwise} \end{cases}$$

$$z_{ct} = 0.$$

The loadings are set to the following values: $\lambda_r = -1$, $\lambda_S = 1.5$, $\lambda_\tau = 0.5$ and $\lambda_c = 0$. The state innovation variances are set to $\varsigma_i = 0.01^2$ for $i \in \{r, S, \tau\}$ and $\varsigma_c = 0$. The predictor variables x_{it} are simulated from independent standard normal distributions.

Figure B.1: Posterior median and the 68 percent credible set of coefficients.



Notes: Panel (a) β_{rt}, (b) β_{St}, (c) $\beta_{\tau t}$ and (d) β_{ct} with the type of variation in the coefficient governed by observed variables (r), a binary indicator (S), an unknown random walk motion (τ) or a constant parameter (c). The red solid lines denote the true paths of the coefficients, the blue lines and blue areas represent the posterior median and the 68 percent posterior credible set of the proposed model using one observed, one latent random walk effect modifier and a latent Markov switching indicator, while the yellow-orange shaded areas refer to the 68 percent posterior credible set of a standard TVP regression equipped with a random walk state equation. Results are based on 15,000 MCMC draws. Vertical axis: estimates and true paths. Horizontal axis: observations.

Table B.1: Summary of MCMC diagnostics of posterior estimates.

	Summary Statistics					
	Mean	Median	Min	Max	10^{th} Perc.	90^{th} Perc.
Inefficiency factors (IF)						
$\{\beta_{rt}\}_{t=1}^{T}$	5.09	5.34	2.42	8.67	2.75	7.09
$\{\beta_{St}\}_{t=1}^{T}$	6.98	4.90	1.13	20.16	1.98	16.19
$\{\beta_{\tau t}\}_{t=1}^{T}$	3.36	2.03	0.98	61.09	1.20	5.82
$\{\beta_{ct}\}_{t=1}^{T}$	2.73	2.38	1.24	5.47	1.60	4.76
$\{\sigma_t^2\}_{t=1}^{T}$	10.38	9.21	4.69	29.12	6.75	15.88
Raftery and Lewis (1992) diagnostics						
$\{\beta_{rt}\}_{t=1}^{T}$	463	413	186	976	346	756
$\{\beta_{St}\}_{t=1}^{T}$	401	406	151	1072	166	596
$\{\beta_{\tau t}\}_{t=1}^{T}$	234	158	148	2688	152	369
$\{\beta_{ct}\}_{t=1}^{T}$	230	178	156	418	164	376
$\{\sigma_t^2\}_{t=1}^{T}$	919	872	211	2212	558	1321

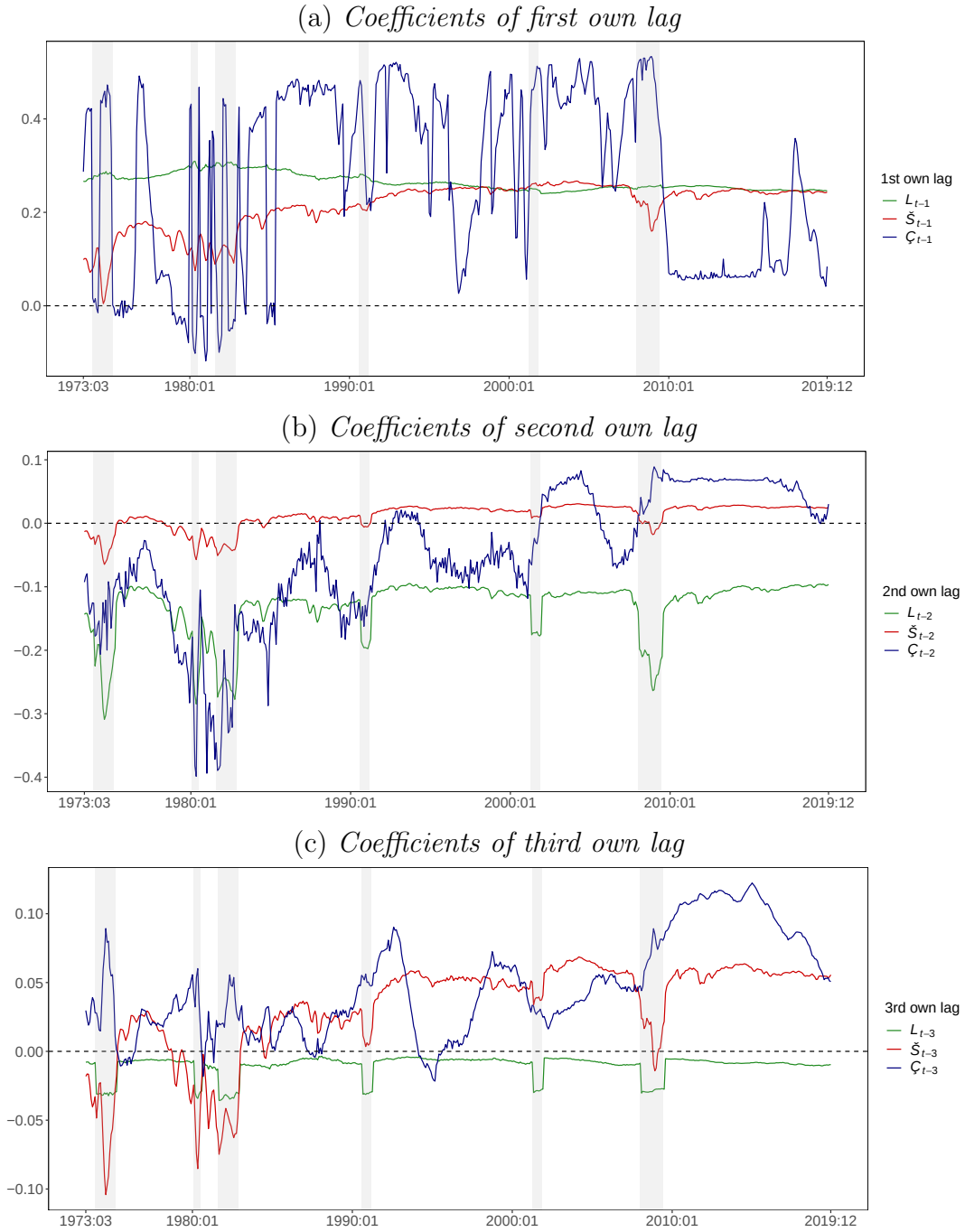
Notes: The table shows the inefficiency factors, specified as the inverse of the relative effective sample size, and the Raftery and Lewis (1992) diagnostics of the number of runs to obtain the 2.5^{th} percentile with 95% probability and 2.5% accuracy.

Figure B.1 shows the resulting "true" paths of the coefficients in red, alongside estimates from using a TVP regression equipped with a random walk state equation (with the yellow-orange shaded area referring to the 68 percent posterior credible set) and our proposed model (the posterior median is the dark blue line with the blue shaded area referring to the 68 percent posterior credible set). We estimate the proposed model using one observed and one latent random walk effect modifier alongside a latent Markov switching indicator. Both competing models allow for stochastic volatility.

To assess convergence properties of our algorithm, we show inefficiency factors (IF) and the Raftery and Lewis (1992) diagnostics in Table B.1. Inefficiency factors below 20 are considered suitable in TVP regressions (see, for example, Primiceri, 2005). The Raftery and Lewis metric refers to the number of iterations from the algorithm necessary to achieve a certain level of precision. Both metrics signal satisfactory convergence properties of our algorithm.

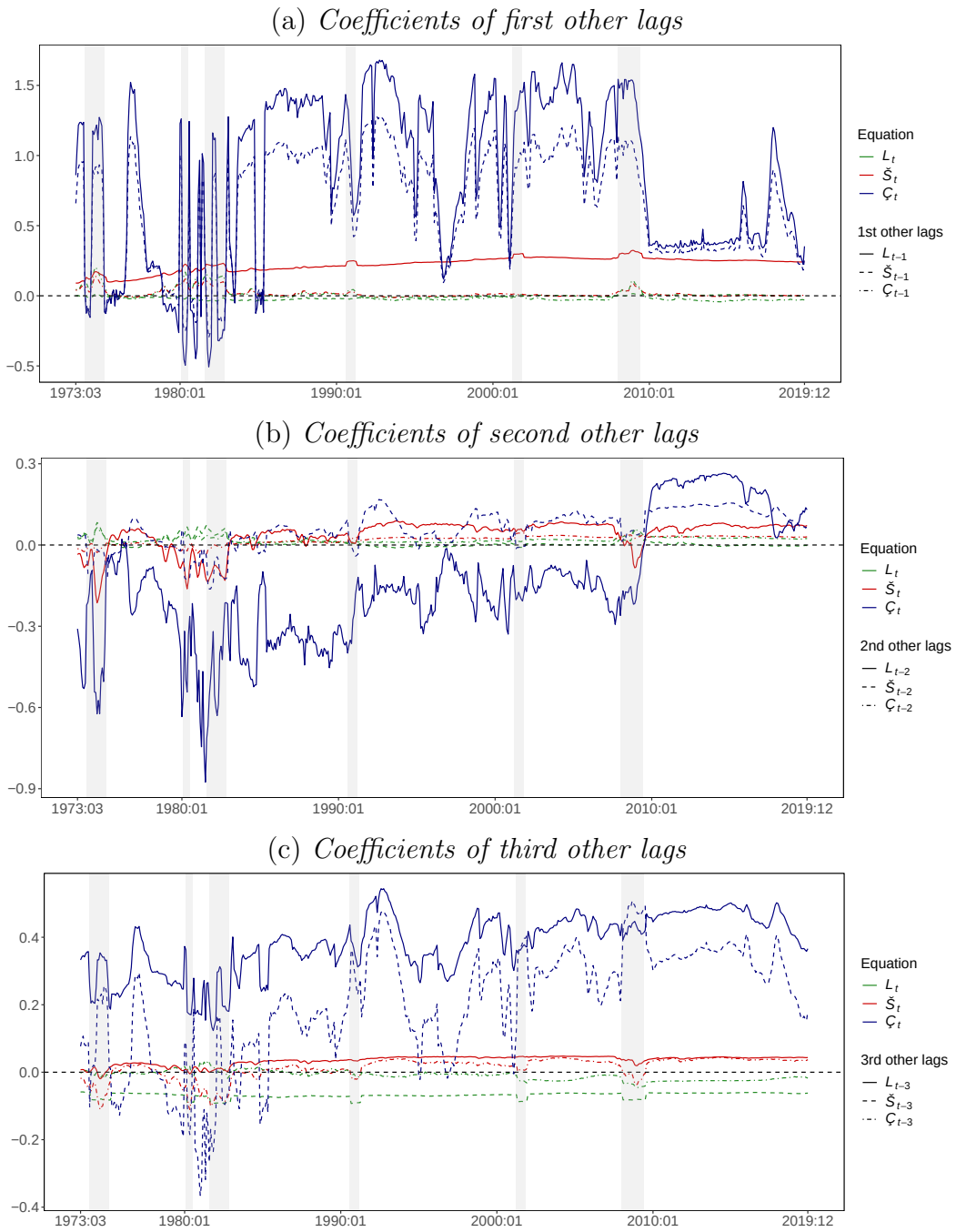
B.2 Further empirical results

Figure B.2: Posterior median of the coefficients associated with the own lags of L_t, $\check{S}_t$ and $\check{C}_t$.



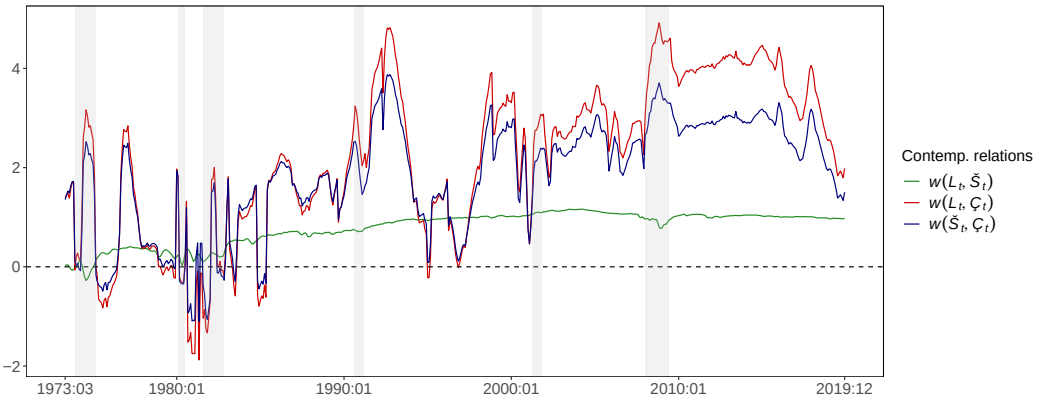
Notes: Panels (a), (b) and (c) show the dynamic evolution of the coefficients related to the variables' own lags $p \in \{1, 2, 3\}$ of the respective equation for L_t, $\check{S}_t$ and $\check{C}_t$. Results are based on the TVP-NS-VAR model variant with $\delta = R_\tau/M = 1$ and using 15,000 MCMC draws. The black dashed line denotes the zero line, while the gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period 1973:01 to 2019:12. Vertical axis: posterior median estimate. Horizontal axis: months.

Figure B.3: Posterior median of the coefficients associated with cross-variable lags of L_t, $\check{S}_t$ and $\check{C}_t$.



Notes: Panels (a), (b) and (c) show the dynamic evolution of the coefficients related to cross-variable lags $p \in \{1, 2, 3\}$ of the respective equation for L_t, $\check{S}_t$ and $\check{C}_t$. Results are based on the TVP-NS-VAR model variant with $\delta = R_\tau/M = 1$ and using 15,000 MCMC draws. The black dashed line denotes the zero line, while the gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period 1973:01 to 2019:12. Vertical axis: posterior median estimate. Horizontal axis: months.

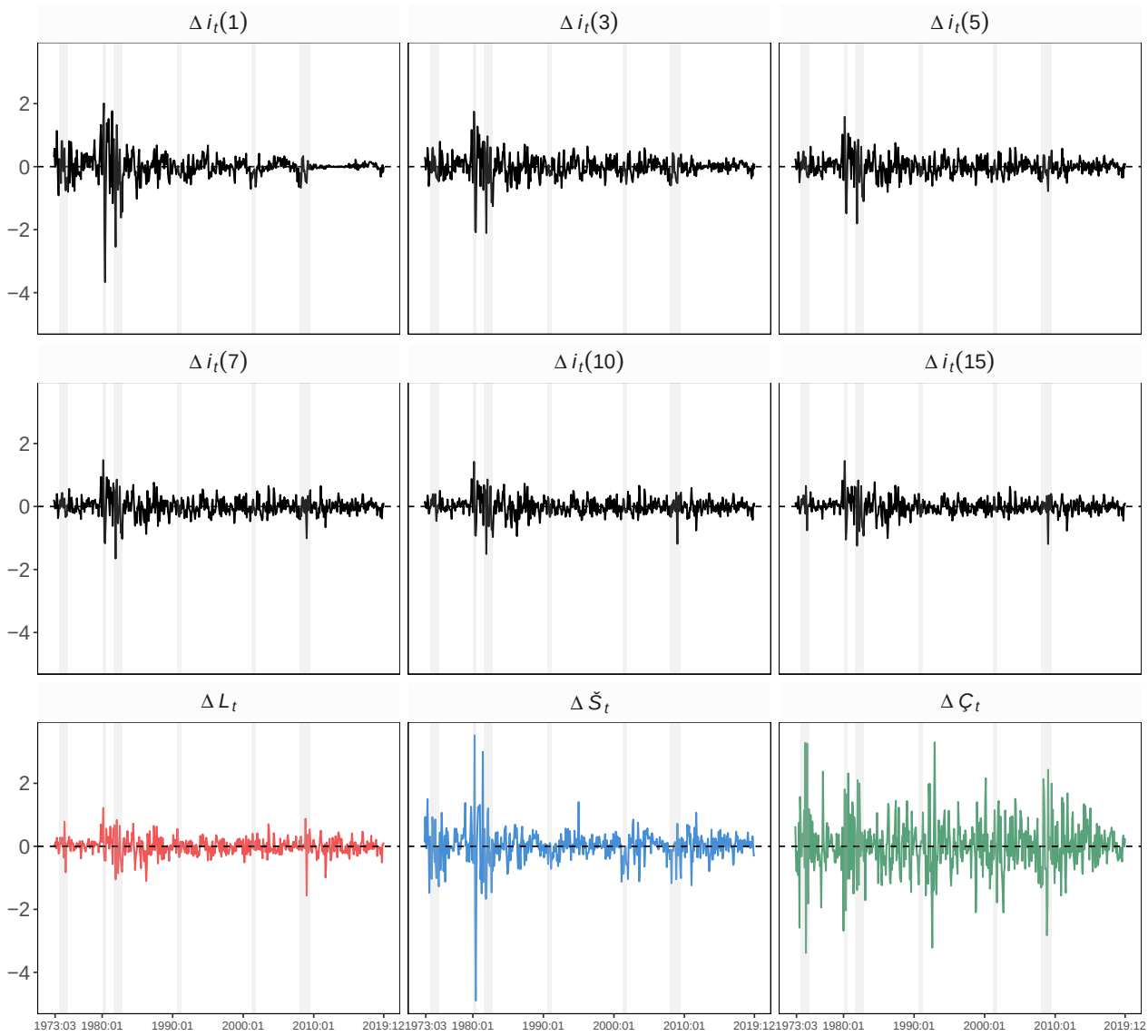
Figure B.4: Posterior median of the contemporaneous relationships between L_t, $\check{S}_t$ and $\acute{C}_t$.



Notes: $w(\check{S}_t, \acute{C}_t)$ denotes the contemporaneous relation between the $\check{S}_t$- and $\acute{C}_t$-equations, $w(L_t, \acute{C}_t)$ and $w(L_t, \check{S}_t)$ are defined analogously. Results are based on the TVP-NS-VAR model variant with $\delta = R_\tau/M = 1$ and using 15,000 MCMC draws. The black dashed line denotes the zero line, while the gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period 1973:01 to 2019:12. Vertical axis: posterior median estimate. Horizontal axis: months.

C Data

Figure C.1: Zero-coupon yields and Nelson-Siegel factors in first differences.



Notes: $\Delta i_t(\theta)$ denotes the zero-coupon yields in first differences at maturity $\theta = \{1, 3, 5, 7, 10, 15\}$ years (with Δ referring to the first-difference operator), ΔL_t is a factor that controls the level, $\Delta \check{S}_t$ determines the slope, and $\Delta \check{C}_t$ represents the curvature factor of the yields in first differences. The black dashed line denotes the zero line, while the gray shaded vertical bars represent recessions dated by the NBER Business Cycle Dating Committee. Sample period 1973:01 to 2019:12. Vertical axis: zero-coupon yields and Nelson-Siegel factors in first differences. Horizontal axis: months.