

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

Manfred M. Fischer*¹, Niko Hauzenberger^{1, 2}
Florian Huber², and Michael Pfarrhofer²

Abstract. This paper proposes a vector autoregressive model with time-varying parameters and stochastic volatility which treats the nature of parameter dynamics as unknown. Coefficients are allowed to evolve according to a random walk, a Markov switching process and/or depend on further predictors. Estimation and inference relies on Bayesian methods. Overfitting issues are tackled through careful model specification and flexible shrinkage priors. As an empirical application, we forecast the US term structure of interest rates and show that our approach performs well relative to a set of competing models, and how the model can be used to explain structural breaks in coefficients related to the US yield curve.

JEL: C11, C30, E37, E43

Keywords: Bayesian shrinkage, interest rate forecasting, latent effect modifiers, MCMC sampling, time-varying parameter regression

May 17, 2021

*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

Time-varying parameter vector autoregressive (TVP-VAR) models are commonly used in finance and macroeconomics to capture dynamic relations across variables, regime shifts and/or structural changes in economic processes (see [Primiceri, 2005](#); [Cogley and Sargent, 2005](#); [Dangl and Halling, 2012](#)). These models typically assume that the parameters evolve over time according to a simple stochastic process such as a random walk. While being rather flexible and parsimonious, this assumption does not allow to examine the extent to which covariates cause changes in the time-varying parameters (TVPs) over time. Moreover, wrongly assuming a random walk state equation could negatively impact predictive accuracy because it essentially implies a smoothness prior on the coefficients. This might be at odds with the rapid shifts we have observed in financial time series such as bond yields and thus could negatively impact predictive accuracy.

The literature has dealt with the issue of selecting the appropriate law of motion for time-varying parameters (TVPs) by estimating different models separately and then using model selection criteria to discriminate between competing specifications (see, e.g., [Sims and Zha, 2006](#); [Koop et al., 2009](#); [Hauzenberger, 2020](#)). But, to the best of our knowledge, no attempt has been made to develop models that rely on a large set of competing laws of motion for the coefficients and decide which one describes the data best.

This paper proposes a flexible approach to TVP-VARs that efficiently integrates out uncertainty with respect to the state evolution equation. The approach assumes that the TVPs depend on a potentially large panel of covariates. These covariates are commonly labeled effect modifiers which can be partially latent and might feature their own state equations. In case they are observed, we obtain a model that is closely related to the varying coefficient model originally proposed in [Hastie and Tibshirani \(1993\)](#). The main advantage of using observed, as opposed to latent effect modifiers, is that we can investigate the driving sources of parameter change. 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.

Careful selection of these effect modifiers is crucial. On the one hand, deciding on the appropriate set of *observed* quantities is difficult, and a large set of candidates could arise. One key objective of this paper is to provide techniques to select promising subsets. On the other hand, appropriately selecting the *latent* effect modifiers allows us to capture situations where parts of the coefficients evolve smoothly whereas others move more abruptly. The latter behavior of the TVPs is often found for US macroeconomic data (see [Sims and Zha, 2006](#)), whereas the former is consistent with financial time series such as bond yields or stock returns (see [Dangl and Halling, 2012](#); [Huber et al., 2019](#)).

Since large VARs often include both macroeconomic as well as financial quantities, a successful model should be able to accommodate both types of structural change or even rely on linear combinations of them. This is precisely what we aim to achieve in this paper. Our approach is capable of answering not only the question why coefficients change, but also to infer an appropriate law of motion using a broad set of latent quantities, each equipped with its own state evolution equation. These unobserved quantities range from latent factors that follow a random walk to Markov switching indicators that allow a subset of parameters to switch between a low number of regimes. Our model approach nests several alternatives proposed in the literature such as the TVP-VAR of [Primiceri \(2005\)](#) and [Cogley and Sargent \(2005\)](#) or the reduced-rank model of [Chan et al. \(2020\)](#).

This large degree of flexibility, however, comes with two concerns. The first is that overfitting problems can easily arise. We overcome these by using Bayesian shrinkage priors. Our prior is a variant of the well-known Horseshoe prior (see [Carvalho et al., 2010](#)) that allows us to shrink coefficients associated with irrelevant effect modifiers towards zero. The second concern relates to computation. Since inclusion of a large number of endogenous variables in a VAR quickly leads to a huge dimensional parameter space, we propose a computationally efficient Markov chain Monte Carlo (MCMC) algorithm. To circumvent mixing issues, we rely on different parameterizations of the model during MCMC sampling. The corresponding

algorithm thus provides a second important contribution of the paper.

In our empirical work, we use the model approach to forecast the US term structure of interest rates. We investigate the empirical properties of our approach using two information sets in the underlying VAR. 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. \(2008\)](#). Adopting a long hold-out period that includes several recessionary episodes, our approach improves upon a wide range of competing models. While improvements for point forecasts are often muted, we find that our proposed model yields favorable density predictions. The predictive exercise is complemented by a comprehensive discussion of patterns in time-variation and their sources. Moreover, how our approach can be used to analyze low frequency relations between the observed quantities in the model over time.

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 effect modifiers which crucially impact the state dynamics. This section, moreover, introduces the Bayesian prior setup and techniques for posterior and predictive inference. [Section 3](#) applies the model approach to the term structure of US interest rates. It also serves to illustrate key model features and to highlight the predictive capabilities of the approach in an out-of-sample forecasting exercise. The last section summarizes and concludes the paper. Additional technical details and further empirical results are provided in the Appendix.

2 Flexible Bayesian Inference in TVP-VARs

2.1 The TVP-VAR

Let \mathbf{y}_t denote an $M \times 1$ vector of macroeconomic and financial quantities at time $t = 1, \dots, T$. We assume that \mathbf{y}_t depends on its P lags which we store in a $K = MP$ -dimensional vector

$\mathbf{x}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-P})'$. Then the basic TVP-VAR can be written as a linear multivariate regression model:

$$\mathbf{y}_t = (\mathbf{I}_M \otimes \mathbf{x}'_t) \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{\beta}_t$ represents a set of $k = MK$ dynamic regression coefficients and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t)$ is a vector Gaussian shock process with time-varying $M \times M$ -dimensional variance-covariance matrix $\boldsymbol{\Sigma}_t$. We assume that $\boldsymbol{\Sigma}_t$ can be decomposed as follows:

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

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

$$h_{jt} = \mu_j + \psi_j (h_{jt-1} - \mu_j) + \nu_{jt}, \quad \nu_{jt} \sim \mathcal{N}(0, \zeta_j^2).$$

μ_j is the unconditional mean, ψ_j the persistence parameter and ζ_j^2 the variance of the log-volatility process for equation j .

In what follows, we rewrite the TVP-VAR using its non-centered parameterization (Frühwirth-Schnatter and Wagner, 2010):

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

$\boldsymbol{\beta}$ denotes a k -dimensional vector of constant coefficients, and $\tilde{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_t - \boldsymbol{\beta}$. This parameterization allows us to disentangle time-invariant (encoded by $\boldsymbol{\beta}$) from time-varying effects (encoded by $\tilde{\boldsymbol{\beta}}_t$) for the regressors. For the decomposed variance-covariance matrix, we have \mathbf{Q} , a lower triangular matrix with ones on the diagonal capturing the constant part of the

covariances. $\tilde{\mathbf{Q}}_t = \mathbf{Q}_t - \mathbf{Q}$ 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 \mathbf{q} and $\tilde{\mathbf{q}}_t$, respectively.

In this paper, the focus is on modeling the $N(= k + v)$ -dimensional vector $\tilde{\boldsymbol{\gamma}}_t = (\tilde{\boldsymbol{\beta}}'_t, \tilde{\mathbf{q}}'_t)'$ and the constant part $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \mathbf{q}')$. The literature typically assumes that the transition distribution $p(\tilde{\boldsymbol{\gamma}}_t | \tilde{\boldsymbol{\gamma}}_{t-1})$ is given by:

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

This law of motion suggests that the expected value of $\tilde{\boldsymbol{\gamma}}_{t-1}$ equals $\tilde{\boldsymbol{\gamma}}_t$ and the amount of time-variation is determined by the $N \times N$ -dimensional process innovation variance-covariance matrix \mathbf{V} . This matrix is often assumed to be diagonal. Notice that if selected elements in \mathbf{V} are equal to zero, the corresponding regression coefficients are constant.

Estimation and inference is typically carried out via Bayesian methods. The recent literature proposes using shrinkage priors to allow for data-based selection of those coefficients which should be time-varying or constant. 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.

2.2 A General Specification for the TVPs

As discussed in the previous sub-section, the typical assumption is that $\tilde{\boldsymbol{\gamma}}_t$ follows a random walk process. In addition, the shocks to the random walk state equation are often assumed to feature a positive error variance. We relax both assumptions to allow for more flexibility.

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

\mathbf{z}_t is given by:

$$\tilde{\boldsymbol{\gamma}}_t = \mathbf{\Lambda} \mathbf{z}_t + \boldsymbol{\eta}_t. \tag{1}$$

$\mathbf{\Lambda}$ denotes an $N \times R$ matrix of regression coefficients, and $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}_N, \mathbf{\Omega})$ is a Gaussian error term with diagonal error variance-covariance matrix $\mathbf{\Omega} = \text{diag}(\omega_1^2, \dots, \omega_N^2)$. If $R \ll N$, the coefficients feature a factor structure and co-move according to the effect modifiers in \mathbf{z}_t . The relationship between $\tilde{\boldsymbol{\gamma}}_t$ and \mathbf{z}_t is determined by the factor loadings in $\mathbf{\Lambda}$. For instance, if the j^{th} column of $\mathbf{\Lambda}$, $\boldsymbol{\lambda}_j$, is equal to zero, the corresponding j^{th} factor in \mathbf{z}_t does not enter the model and thus has no influence on $\tilde{\boldsymbol{\gamma}}_t$.

The specific selection of \mathbf{z}_t is crucial for determining the dynamics of $\tilde{\boldsymbol{\gamma}}_t$. Appropriate choice of \mathbf{z}_t yields a variety of important special cases that depend on the specific values of $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ as well as on the composition of \mathbf{z}_t . In this sub-section, we briefly focus on special cases that arise independently of the choice of \mathbf{z}_t . The next sub-section deals with cases that arise if \mathbf{z}_t is suitably chosen.

These two cases are the following. If $\mathbf{\Lambda} = \mathbf{0}_{N \times R}$, with $\mathbf{0}_{N \times R}$ being a $N \times R$ 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 [Korobilis, 2019](#); [Hauzenberger et al., 2019](#)). The second special case arises if both $\mathbf{\Lambda} = \mathbf{0}_{N \times R}$ and $\mathbf{\Omega} = \mathbf{0}_{N \times N}$. In this case, we obtain a standard constant parameter regression model.

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

2.3 Possible Choices for the Effect Modifiers

The specific choice of \mathbf{z}_t is crucial in determining how $\tilde{\gamma}_t$ behaves over time. Hence, by suitably choosing the elements in \mathbf{z}_t , our model approach is related to the following specifications:

- [Chan et al. \(2020\)](#): We assume that \mathbf{z}_t consists exclusively of a sequence of $R = R_\tau$ latent factors $\boldsymbol{\tau}_t$, which follow a multivariate random walk:

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

$\mathbf{V}_\tau = \text{diag}(v_1^2, \dots, v_{R_\tau}^2)$ denotes a diagonal variance-covariance matrix with v_j^2 being process innovation variances that determine the smoothness of the elements in $\boldsymbol{\tau}_t$. Note that setting v_j^2 close to zero effectively implies that τ_{jt} , the j^{th} element in $\boldsymbol{\tau}_t$, is constant. This model implies a factor structure in $\tilde{\gamma}_t$ if $R_\tau \ll N$.

- [Primiceri \(2005\)](#): If $R = R_\tau = N$, the elements in \mathbf{z}_t are random walks, and $\boldsymbol{\Lambda} = \mathbf{I}_N$, we obtain a standard time-varying parameter model. Assuming that the covariances are constant we obtain the model put forth in [Cogley and Sargent \(2005\)](#).
- [Sims and Zha \(2006\)](#): A Markov switching model can be obtained by setting $\mathbf{z}_t = S_t$, with $S_t \in \{0, 1\}$ denoting a binary indicator with transition probabilities given by:

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

with p_{ij} denoting the $(i, j)^{\text{th}}$ element of a 2×2 -dimensional transition probability matrix \mathbf{P} . Inclusion of this random quantity allows to capture structural breaks in $\tilde{\gamma}_t$ that are common to all coefficients.

- [Caggiano et al. \(2017\)](#): Assuming that \mathbf{z}_t is exclusively composed of observed quantities we obtain a regression model with interaction terms.

These examples show that our model, conditional on choosing a suitable set of effect 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 simply specify \mathbf{z}_t to include most (with the exception of the $R = N$ setup) of the modifiers discussed above.

More precisely, we set \mathbf{z}_t as follows:

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

Here we let \mathbf{r}_t denote a set of R_r observed factors and the dimension of \mathbf{z}_t is thus $R = R_r + R_S + R_\tau$ with $R_S = M$. To allow for additional flexibility we assume that $\boldsymbol{\tau}_t = (\boldsymbol{\tau}'_{1t}, \dots, \boldsymbol{\tau}'_{Mt})'$ with $\boldsymbol{\tau}_{jt}$ being equation-specific factors of dimension R_{τ_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 (and thus $\mathbf{S}_t = (S_{1t}, \dots, S_{Mt})'$, with the corresponding transition probabilities matrix denoted by \mathbf{P}_j).

The corresponding loadings matrix $\mathbf{\Lambda}$ 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 maximum flexibility (and thus risk overfitting) and using a rather parsimonious model (with the risk of being too simplistic). Recent contributions use similar assumptions on the state evolutions, (see [Koop et al., 2009](#); [Maheu and Song, 2018](#)). As opposed to these papers, our approach offers more flexibility since, if necessary, the presence of the idiosyncratic shocks to the TVPs allows for deviations if the factor structure does not represent the data well.

Our specification implies that, depending on the factor loadings $\mathbf{\Lambda}$, the evolution of $\tilde{\gamma}_t$ might be a combination of a set of random walk factors, a Markov switching process and some observed quantities. To single out irrelevant elements in \mathbf{z}_t , one could simply set the corresponding columns in $\mathbf{\Lambda}$ equal to zero. In this paper, we achieve this through a Bayesian shrinkage prior. The next sub-section discusses our priors in more detail.

2.4 The Prior Setup

The discussion in Sub-section 2.1 shows that our model approach nests a variety of competing models. To select the appropriate model variant and alleviate over-parameterization concerns, we opt for a Bayesian approach to introduce shrinkage. Here, we summarize the priors we impose on key parameters.

In light of the specific choice of \mathbf{z}_t , we introduce some additional notation to clarify details on our prior implementation. Let us assume that $\mathbf{\Lambda}$ is composed of the following matrices:

$$\mathbf{\Lambda} = [\mathbf{\Lambda}_r \quad \mathbf{\Lambda}_S \quad \mathbf{\Lambda}_\tau],$$

where $\mathbf{\Lambda}_r$ is an $N \times R_r$ matrix of loadings related to the observed quantities, $\mathbf{\Lambda}_S$ denotes an $N \times R_S$ matrix of loadings related to \mathbf{S}_t , and $\mathbf{\Lambda}_\tau$ represents an $N \times R_\tau$ factor loadings matrix associated with $\boldsymbol{\tau}_t$.

For imposing shrinkage we rely on variants of the horseshoe prior (Carvalho et al., 2010).¹ While in principle any global-local shrinkage prior may be used, we choose the horseshoe prior due to its excellent shrinkage properties and the lack of tuning parameters. In particular, we specify a column-wise horseshoe prior on the loadings matrix. 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. $\lambda_{ji,\ell}$ refers to the ℓ^{th} element of this vector. The prior is given by:

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

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

To further regularize our potentially huge-dimensional parameter space, we impose an equation-wise horseshoe prior on the constant part of the regression coefficients and covari-

¹In principle, it would be possible to design priors in the Minnesota tradition (see, e.g., Giannone et al., 2015) which introduce more structure in terms of achieving shrinkage. However, since our model nests several competing specifications, more flexible shrinkage priors appear to be better suited.

ances in $\boldsymbol{\gamma}_j$ corresponding to the j^{th} equation. Let γ_{ji} denote the i^{th} element of the vector. 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.

We mentioned earlier that it is often assumed that shocks to the states feature positive error variances. To introduce shrinkage of ω_{ji} towards zero, we impose a horseshoe prior also on the square root of the innovation variances of the measurement errors in Eq. (1). This prior is specified in an equation-specific manner. For equation j , let $\boldsymbol{\omega}_j$ denote a $v_j (= j - 1 + k)$ -dimensional vector which stores the diagonal elements in $\boldsymbol{\Omega}$ 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 $\omega_{ji} | \varpi_{ji}, \vartheta_j \sim \mathcal{G}(1/2, \varpi_{ji}^{-2} \vartheta_j^{-2} / 2)$, 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 $\mathbf{V}_\tau = \mathbf{I}_{R_\tau}$ and thus impose shrinkage through the factor loadings in $\boldsymbol{\Lambda}_\tau$ (see [Chan et al., 2020](#)).

On the parameters of the state equation of the log-volatility processes μ_j, ψ_j and ζ_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, $(\psi_j + 1)/2 \sim \mathcal{B}(5, 1.5)$, a Gamma prior on the state variance, $\zeta_j^2 \sim \mathcal{G}(1/2, 1/2)$, while the prior on the initial state is the unconditional distribution for all

$j = 1, \dots, M$.

For the equation-specific transition probabilities \mathbf{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:

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

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.

2.5 Full Conditional Posterior Simulation

To simulate from the full posterior distribution we develop an efficient MCMC algorithm. Since full-system estimation of the VAR quickly becomes computationally cumbersome, we rely on the equation-by-equation algorithm suggested in [Carriero et al. \(2019\)](#).

Conditional on \mathbf{Q}_t , one can state the VAR as a system of (conditionally) independent equations. The first equation of this system is given by:

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

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

$$y_{jt} = \mathbf{x}'_t(\boldsymbol{\beta}_j + \tilde{\boldsymbol{\beta}}_{jt}) + \mathbf{u}'_{jt}(\mathbf{q}_j + \tilde{\mathbf{q}}_{jt}) + \varepsilon_{jt}. \quad (2)$$

$\boldsymbol{\beta}_j$ and $\tilde{\boldsymbol{\beta}}_{jt}$ denote the j^{th} subvectors of the constant and time-varying parts in $\boldsymbol{\beta}_t$ with $\boldsymbol{\beta}_t = (\boldsymbol{\beta}'_{1t}, \dots, \boldsymbol{\beta}'_{Mt})'$ and $\mathbf{u}_{jt} = (\varepsilon_{1t}, \dots, \varepsilon_{j-1,t})'$. The $(j - 1)$ -dimensional vectors \mathbf{q}_j and $\tilde{\mathbf{q}}_{jt}$ store the constant and time-invariant part of the free elements in the j^{th} row of \mathbf{Q}_t . This approach allows to estimate the different elements of $\boldsymbol{\gamma}_t$ that relate to the M equations independently from each other conditional on the shocks to the preceding $j - 1$ equations. This speeds up computation enormously.

Equation (2) can be simplified to yield:

$$y_{jt} = \mathbf{m}'_{jt}(\boldsymbol{\gamma}_j + \tilde{\boldsymbol{\gamma}}_{jt}) + \varepsilon_{jt}, \quad (3)$$

where $\mathbf{m}_{jt} = (\mathbf{x}'_t, \mathbf{u}'_{jt})'$, $\tilde{\boldsymbol{\gamma}}_{jt} = \boldsymbol{\gamma}_{jt} - \boldsymbol{\gamma}_j$, 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. (3). For notational simplicity, we assume that all elements in \mathbf{z}_t are latent. In light of the discussion in Sub-section 2.3, this implies that $\mathbf{z}_{jt} = (\boldsymbol{\tau}'_{jt}, S_{jt})'$ and the extension to include observed factors is trivial.

Sampling \mathbf{z}_{jt} . Conditional on the remaining quantities of the model, we simulate the latent (random walk and Markov switching) components in \mathbf{z}_{jt} after integrating out $\tilde{\boldsymbol{\gamma}}_{jt}$. This is achieved by rewriting Eq. (3) as:

$$\tilde{y}_{jt} = \mathbf{m}'_{jt}\boldsymbol{\Lambda}_j\mathbf{z}_{jt} + \mathbf{m}'_{jt}\boldsymbol{\eta}_{jt} + \varepsilon_{jt}, \quad (4)$$

with $\tilde{y}_{jt} = y_{jt} - \mathbf{m}'_{jt}\boldsymbol{\gamma}_j$. Defining $\tilde{\mathbf{m}}'_{jt} = \mathbf{m}'_{jt}\boldsymbol{\Lambda}_j$ and $\hat{\varepsilon}_{jt} = \mathbf{m}'_{jt}\boldsymbol{\eta}_{jt} + \varepsilon_{jt}$ allows us to cast Eq. (4) as a simple linear regression model:

$$\tilde{y}_{jt} = \tilde{\mathbf{m}}'_{jt}\mathbf{z}_{jt} + \hat{\varepsilon}_{jt}, \quad \hat{\varepsilon}_{jt} \sim \mathcal{N}(0, \mathbf{m}'_{jt}\text{diag}(\boldsymbol{\omega}_j)\mathbf{m}_{jt} + e^{h_{jt}}). \quad (5)$$

This parameterization has the advantage that it does not depend on $\tilde{\boldsymbol{\gamma}}_{jt}$, and \mathbf{z}_{jt} can thus be sampled marginally of $\tilde{\boldsymbol{\gamma}}_{jt}$. This improves mixing substantially since $\tilde{\boldsymbol{\gamma}}_{jt}$ and \mathbf{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 \mathbf{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).²

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 the factor loadings using a generalized inverse Gaussian distribution (see, among others, Bitto and Frühwirth-Schnatter, 2019). For further details and the moments of this distribution, see Appendix A.

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

$$y_{jt} = \hat{\mathbf{m}}'_{jt} \hat{\gamma}_j + \hat{\varepsilon}_{jt}, \quad (6)$$

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

Sampling the stochastic volatilities. The latent log-volatility processes can again be sampled on an equation-by-equation basis. This step is implemented using the R-package `stochvol`.

²It is worth mentioning that the $O(N^3)$ statement is true for the precision sampler and differs for forward-filtering backward-sampling algorithms.

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 \(2015\)](#). Exact posteriors are provided in [Appendix A](#).

3 Forecasting US Government Bond Yields

This section applies the model to predict the term structure of US interest rates. These time series are characterized by substantial non-linearities (e.g. during the period of the zero lower bound), 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 as well as discuss how key quantities that shape yield curve dynamics co-move over time at low frequencies.

3.1 Data and Design of the Forecasting Exercise

Our aim is to predict monthly zero-coupon yields of US treasuries at different yearly maturities. The data is described in detail in [Gürkaynak et al. \(2007\)](#).³ The target variables are 1, 3, 5, 7, 10 and 15 years maturities. For stability reasons, all variables enter our model in first differences.

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 for 2000:01. After obtaining the predictive distributions we expand the sample and repeat this procedure until we reach 2019:12.⁴ Point forecast performance is measured using Root Mean Squared Errors (RMSEs),

³Available online at federalreserve.gov/data/nominal-yield-curve.htm.

⁴Note 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.

while density forecasts are assessed in terms of Log Predictive Bayes Factors (LPBFs), averaged over the out-of-sample observations (Geweke and Amisano, 2010).

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 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.⁵ Thus, $\mathbf{r}_t = (r_{\text{NGCI},t-1}, r_{\text{REC},t-1}, r_{\text{RF},t-1})'$ and $R_r = 3$. Exogenous variables enter the model as first order lags. Higher-order forecasts involving exogenous variables are based on random walk predictions of these quantities.

We use these three effect modifiers for simple reasons. First, there is strong evidence that yield curve dynamics differ across business cycle phases (see, e.g., 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, e.g., Adrian et al., 2019; Carriero et al., 2020; Adams et al., 2021).

3.2 Competing Model Specifications

The forecast exercise distinguishes between two model classes, with 15 distinct model specifications in each class. The first model class involves TVP-VARs that incorporate the six target variables as endogenous variables, that is $M = 6$. This model class is labeled as VAR.

⁵Available online at mba.tuck.dartmouth.edu/pages/faculty/ken.french.

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:

$$i_t(\theta) = L_t + \left(\frac{1 - \exp(-\theta\alpha)}{\theta\alpha} \right) \check{S}_t + \left(\frac{1 - \exp(-\theta\alpha)}{\theta\alpha} - \exp(-\theta\alpha) \right) \zeta_t,$$

where $i_t(\theta)$ denotes the yield at maturity θ at time t , L_t is a factor that controls the level, \check{S}_t determines the slope, and ζ_t represents the curvature factor of the yields. The parameter α governs the exponential decay rate. To maximize the loading on ζ_t we set $\alpha = 0.7308$ (12×0.0609).⁶ In what follows, we use the latent factors L_t , \check{S}_t and ζ_t as endogenous variables in the VAR specifications by defining $\mathbf{y}_t = (L_t, \check{S}_t, \zeta_t)'$, resulting in $M = 3$. These factors are obtained by running OLS on a t -by- t basis and then used as actual data within our model. This model class is subsequently labeled NS-VAR.

Model specifications are differentiated over a grid of effect modifier combinations. In particular, specifications within a model class differ in terms of three aspects (see Table 1): First, in terms of whether the three exogenous variables (collected in \mathbf{r}_t) are included in \mathbf{z}_t or not ("x" marks inclusion, "-" indicates no observed factors); second, in terms of the number of latent random walk factors R_{τ_j} included in \mathbf{z}_t (which we assume to be equal across equations); and third, in terms of the presence of Markov switching indicators in \mathbf{S}_t , again with "x" marking their inclusion and "-" their absence. This setup implies that we have 15 time-varying parameter NS-VAR and 15 time-varying parameter VAR model specifications. For comparative purposes we also consider the two class-specific constant parameter model variants, labeled "Constant," 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 \mathbf{r}_t and \mathbf{S}_t).⁷ Notice that we also have a specification which includes only \mathbf{S}_t and another one which uses only observed factors. The latter one is closely related to a Markov switching model whereas the second one closely resembles a VAR with interaction terms.

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

⁷It 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.

3.3 Results

Table 1 shows the one-month and one-quarter-ahead out-of-sample forecasting results for US treasury yields at different maturities, using the 30 TVP model specifications and the two constant parameter model variants as described in the previous sub-section. Recall that $M = 3$ (and $K = 9$) in case of the NS-VAR and $M = 6$ (and $K = 18$) in case of the VAR model variants. $\mathbf{z}_t = (\mathbf{r}'_t, \mathbf{S}'_t, \boldsymbol{\tau}'_t)$ where \mathbf{r}_t denotes $R_r (= 3)$ -dimensional vector, \mathbf{S}_t a R_S -dimensional vector, and $\boldsymbol{\tau}_t$ a R_τ -dimensional vector. All specifications feature $P = 3$ lags of the endogenous variables. Diebold and Li (2006) specify VAR(1) transition dynamics for the yield curve factors. Here, we use three lags and rely on the shrinkage prior to select an appropriate lag order by pushing coefficients associated with irrelevant lags towards zero.

The performance of point forecasts is measured in terms of RMSEs and that of density forecasts in terms of LPBFs, relative to the constant parameter VAR model (shaded in yellow) that serves as benchmark. RMSEs are presented in ratios, and LPBFs in differences are given below the RMSEs in parentheses. RMSEs below one indicate superior performance, relative to the benchmark, 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.

The vast range of competing specifications, loss functions used to evaluate forecasts and maturities makes it hard to identify a single best performing model. 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, Table 1 suggests a pronounced degree of heterogeneity in forecast accuracy across models and for both the NS-VAR and VAR specifications. While differences at some maturities are substantial, they are muted or non-existent for others. It is also worth mentioning that the TVP variants of the NS-VAR and VAR models outperform the constant parameter specifications in most cases (apart from one-quarter ahead point forecasts of treasuries with a maturity of five years). This general observation holds true irrespective of

Table 1: Out-of-sample forecasting results for US treasury yields at different maturities using TVP-NS-VAR and TVP-VAR model specifications (Numbers refer to relative RMSEs while those in the parentheses relate to average Log Predictive Bayes Factors between a given model and the unrestricted constant parameter VAR.)

Model	Specification			One-month-ahead							One-quarter-ahead							
	r_t	δ	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-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 (-0.91)	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 (-1.43)	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							
	r_t	δ	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-VAR	x	6	x	0.79 (1.13)	0.92 (0.06)	0.96 (0.03)	0.94 (0.07)	0.89 (0.14)	0.79 (0.33)	0.60 (0.68)	0.94 (0.63)	0.96 (-0.01)	0.99 (-0.01)	0.98 (0.00)	0.97 (0.00)	0.94 (0.01)	0.84 (0.08)	
	x	3	x	0.78 (1.20)	0.92 (0.09)	0.96 (0.03)	0.94 (0.06)	0.89 (0.14)	0.78 (0.33)	0.60 (0.70)	0.94 (0.55)	0.98 (0.01)	1.00 (-0.02)	0.99 (-0.01)	0.98 (0.00)	0.94 (0.02)	0.84 (0.10)	
	x	1	x	0.79 (1.52)	0.93 (0.10)	0.97 (0.03)	0.95 (0.06)	0.89 (0.14)	0.78 (0.33)	0.60 (0.70)	0.93 (0.83)	0.96 (0.02)	0.99 (0.00)	0.98 (0.01)	0.97 (0.01)	0.94 (0.02)	0.84 (0.12)	
	x	6	-	0.79 (1.26)	0.93 (0.06)	0.98 (0.04)	0.95 (0.07)	0.90 (0.15)	0.79 (0.34)	0.60 (0.69)	0.94 (0.69)	0.99 (0.01)	1.00 (0.01)	1.00 (0.02)	0.98 (0.02)	0.95 (0.03)	0.84 (0.10)	
	x	3	-	0.78 (1.20)	0.91 (0.11)	0.96 (0.06)	0.94 (0.09)	0.89 (0.16)	0.79 (0.34)	0.60 (0.71)	0.94 (0.66)	0.98 (0.03)	1.00 (0.02)	1.00 (0.02)	0.98 (0.03)	0.95 (0.04)	0.85 (0.12)	
	x	1	-	0.78 (1.56)	0.91 (0.13)	0.96 (0.06)	0.94 (0.08)	0.89 (0.15)	0.78 (0.34)	0.60 (0.72)	0.92 (0.61)	0.96 (0.07)	0.97 (0.05)	0.97 (0.04)	0.96 (0.04)	0.93 (0.05)	0.83 (0.14)	
	-	6	x	0.81 (0.95)	0.95 (0.02)	0.97 (0.03)	0.95 (0.07)	0.91 (0.13)	0.81 (0.32)	0.63 (0.67)	0.97 (0.45)	1.00 (-0.04)	1.00 (-0.03)	1.00 (-0.01)	1.00 (-0.01)	0.98 (0.00)	0.90 (0.06)	
	-	3	x	0.79 (1.08)	0.95 (0.06)	0.98 (0.02)	0.95 (0.06)	0.90 (0.12)	0.79 (0.31)	0.60 (0.68)	0.93 (0.28)	0.99 (-0.01)	0.99 (-0.02)	0.97 (-0.01)	0.96 (-0.01)	0.93 (0.01)	0.83 (0.08)	
	-	1	x	0.78 (1.29)	0.95 (0.08)	0.96 (0.04)	0.94 (0.07)	0.89 (0.14)	0.79 (0.32)	0.60 (0.69)	0.95 (0.44)	0.97 (0.04)	0.98 (0.00)	0.98 (-0.01)	0.98 (-0.01)	0.96 (0.00)	0.87 (0.09)	
	-	6	-	0.79 (1.04)	0.95 (0.04)	0.96 (0.05)	0.95 (0.08)	0.90 (0.15)	0.79 (0.34)	0.60 (0.69)	0.92 (0.32)	0.99 (-0.03)	0.98 (0.01)	0.97 (0.02)	0.96 (0.03)	0.93 (0.04)	0.83 (0.11)	
	-	3	-	0.79 (1.27)	0.94 (0.08)	0.97 (0.05)	0.94 (0.08)	0.89 (0.15)	0.79 (0.34)	0.61 (0.70)	0.93 (0.61)	0.98 (0.00)	0.98 (0.03)	0.97 (0.04)	0.96 (0.05)	0.93 (0.06)	0.83 (0.13)	
	-	1	-	0.79 (1.58)	0.95 (0.09)	0.96 (0.06)	0.95 (0.08)	0.90 (0.15)	0.79 (0.35)	0.61 (0.72)	0.92 (0.78)	0.99 (0.04)	0.99 (0.04)	0.97 (0.05)	0.95 (0.05)	0.92 (0.06)	0.82 (0.15)	
	x	-	-	0.79 (1.26)	0.93 (0.12)	0.97 (0.04)	0.95 (0.07)	0.90 (0.14)	0.79 (0.35)	0.60 (0.75)	0.93 (0.75)	0.98 (0.07)	0.99 (0.03)	0.98 (0.02)	0.97 (0.01)	0.93 (0.03)	0.83 (0.15)	
	-	-	x	0.79 (0.83)	0.94 (0.07)	0.96 (0.02)	0.93 (0.07)	0.89 (0.14)	0.79 (0.33)	0.61 (0.70)	0.93 (0.49)	0.97 (0.02)	0.97 (0.00)	0.97 (0.01)	0.96 (0.01)	0.94 (0.02)	0.86 (0.11)	
	-	K	-	0.79 (0.28)	0.94 (0.17)	0.96 (0.03)	0.95 (0.04)	0.90 (0.10)	0.80 (0.27)	0.61 (0.67)	0.93 (-0.15)	0.98 (0.12)	0.98 (0.01)	0.97 (-0.02)	0.96 (-0.03)	0.93 (-0.04)	0.83 (0.07)	
	Constant				0.85 (2.04)	0.39 (-0.21)	0.59 (-0.81)	0.72 (-1.08)	0.82 (-1.25)	0.99 (-1.49)	1.28 (-1.87)	0.76 (2.26)	0.43 (-0.30)	0.62 (-0.84)	0.73 (-1.08)	0.81 (-1.19)	0.87 (-1.26)	0.96 (-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 for maturities 1, 3, 5, 7, 10 and 15 years and the corresponding joint measure across all maturities. Specifications of the TVP model variants are differentiated over a grid of effect modifier combinations, in terms of whether the exogenous variables in 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 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. **Root Mean Squared Errors (RMSEs) and Log Predictive Bayes Factors (LPBFs), averaged over the out-of-sample observations are given relative to the constant parameter VAR model (shaded in yellow). 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.**

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 better point forecasting performance of the NS-VAR 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 better density performance of the VAR models is most likely driven by two sources. The first is that, as opposed to the conditional mean, the strong implicit assumptions of the NS-VAR 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 thus yields better density forecasts. The second fact is that the small-scale NS-VARs might feature a too tight predictive variance since relevant information is ignored. And this could harm density forecasting accuracy during turbulent times.

Next, we zoom into 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 is successful, and thus not susceptible to overfitting concerns.

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. The same is true for the best-performing model specification at the one-quarter-ahead horizon, the one-factor NS-VAR without exogenous variables and Markov switching processes, albeit at much smaller margins versus the benchmark. Here, improvements are about nine percent in terms of relative RMSEs. Assessing predictive performance for individual maturities allows to identify which segment of the yield curve drives the overall results. While gains for shorter maturities in the case of one-month-ahead forecasts are muted, we observe large gains at the long end of the yield curve. Relative to the best performing specification, improvements are about 40 percent in RMSE terms for the best performing specification. The same is true for one-quarter-ahead forecasts, however, at smaller margins of about 20 percent.

We proceed with our findings for density forecasts. As mentioned above, the VAR specifications overall exhibit more favorable relative LPBFs compared to the NS-VAR. This is easily observable by noting that most bold values (in parentheses) are located in the lower panel of Table 1. 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. This again serves as an example that even though it is extremely flexible, our approach to shrinking the parameter space avoids overfitting and does not harm predictive performance. In fact, for one-quarter-ahead forecasts, the TVP-VAR with exogeneous variables, one latent factor and Markov switching shows the largest gains in terms of density forecasts. Again, it must be acknowledged that margins within this model class are rather small. As it is the case for point forecasts, these gains are mostly obtained in terms of the long end of the yield curve, while gains in forecast accuracy at shorter horizons are negligible.

Summing up, while improvements relative to established models are often small, our proposed framework is competitive for most maturities at both the one-month and one-quarter ahead horizons. Main differences arise from the considered model class, with the NS-

VAR exhibiting promising results in terms of point forecasts, and superior density forecasts for the VAR. Moreover, we detect the largest improvements at the long end of the yield curve. The additional flexibility of our proposed approach does not lead to severe overfitting since we use shrinkage priors to regularize several parts of the parameter space. And it almost never harms forecast accuracy while improving performance in some cases.

3.4 Determinants of Time-Variation in the Coefficients

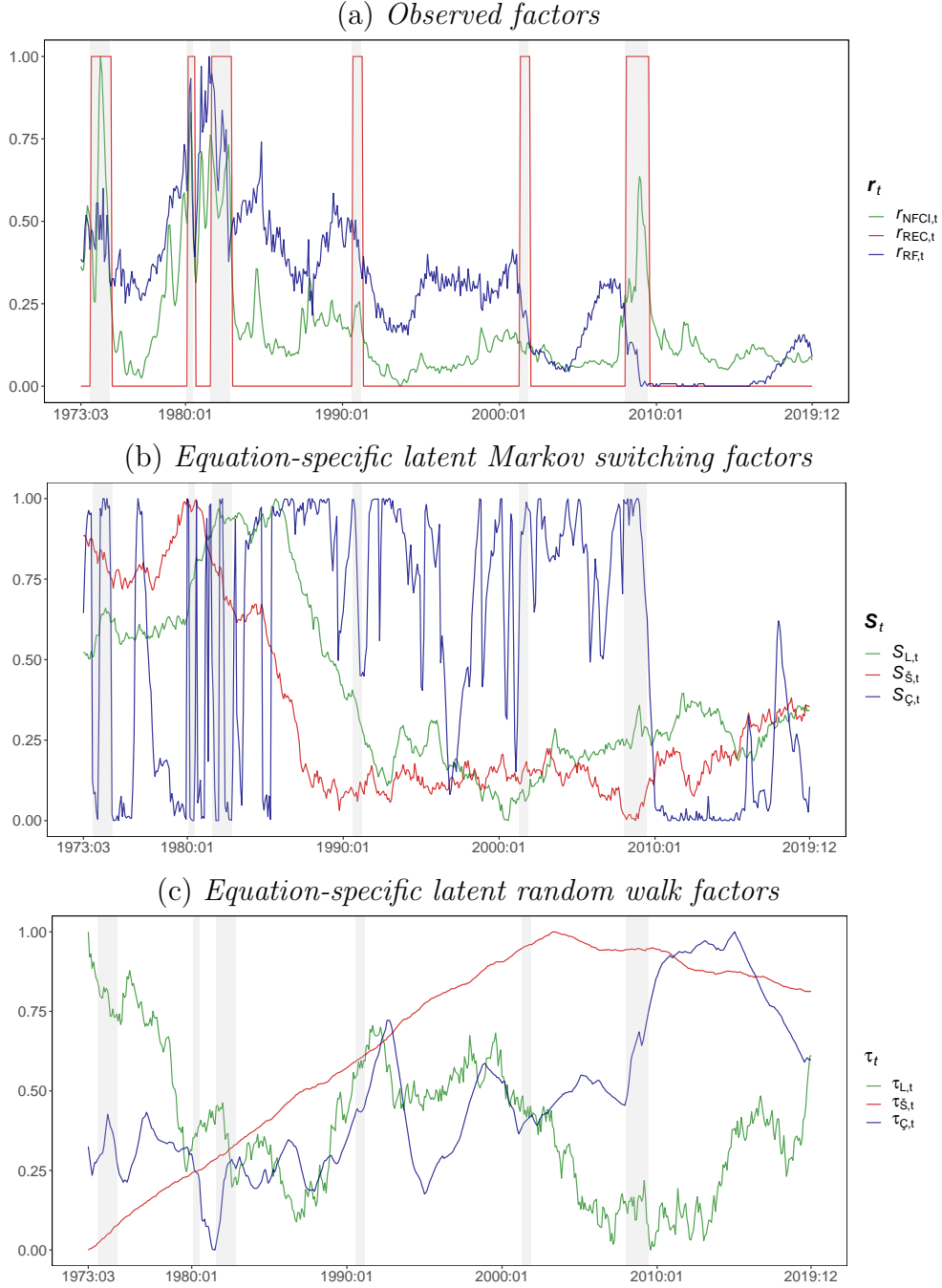
The previous sub-section established that the proposed approach yields more favorable predictions than conventional TVP-VARs. Including observed and latent effect modifiers allows for investigating the sources of time-variation in coefficients, and thus the driving factors of improvements in predictive accuracy. We carry out a detailed analysis of these determinants in this sub-section.

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 \mathbf{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 \mathbf{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.⁸ We therefore exploit the fact that introducing any invertible $R \times R$ -dimensional matrix does not alter the likelihood of the model, since $\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{U}\mathbf{z}_t = \mathbf{\Lambda}\mathbf{z}_t$. Define \mathbf{U} as diagonal matrix such that the maximum of each \mathbf{z}_j (for $j = 1, \dots, R$) corresponds to one, the minimum is zero, $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\mathbf{U}^{-1}$ and $\tilde{\mathbf{z}}_t = \mathbf{U}\mathbf{z}_t$. This simple linear transformation allows for assessing the relative movement of the indicators in $\tilde{\mathbf{z}}_t$ without affecting overall dynamics.

⁸In particular, this is the case for the product $\mathbf{\Lambda}_\tau\boldsymbol{\tau}_t$. Note that we do not face this issue for $\mathbf{\Lambda}_\tau\mathbf{r}_t$ and $\mathbf{\Lambda}_S\mathbf{S}_t$ because both are either observed or already bounded between zero and one, and their scale and sign are identified.

Figure 1: Evolution of the normalized effect modifiers $\tilde{\mathbf{z}}_t = \mathbf{U}\mathbf{z}_t$ over time.



Notes: Results are based on a TVP-NS-VAR model specification $\mathbf{z}_t = (\mathbf{r}'_t, \mathbf{S}'_t, \boldsymbol{\tau}'_t)'$ and $\delta = R_\tau/M = 1$ and $P = 3$. Panel (a) shows the normalized observed factors $r_{\text{NFCI},t}$, $r_{\text{REC},t}$ and $r_{\text{RF},t}$ (collected in \mathbf{r}_t), panel (b) the posterior mean of the latent Markov switching factors, $S_{L,t}$, $S_{\dot{S},t}$ and $S_{C,t}$ (collected in \mathbf{S}_t) and panel (c) the posterior median of the latent random walk factors, $\tau_L, \tau_{\dot{S}}$ and τ_C (collected in $\boldsymbol{\tau}_t$). Note that S_{jt} and τ_{jt} are equation-specific latent quantities with $M(= 3)$ endogenous variables, $j \in \{L, \dot{S}, C\}$. $S_{L,t}$ and $\tau_{L,t}$ correspond to the first equation, $S_{\dot{S},t}$ and $\tau_{\dot{S},t}$ to the second, while $S_{C,t}$ and $\tau_{C,t}$ to the third. Results are based on the TVP-NS-VAR model variant with $\delta = R_\tau/M = 1$ and using 15,000 MCMC draws. 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. Front axis: months.

Figure 1 displays the evolution of the normalized indicators in $\tilde{\mathbf{z}}_t$ 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 \mathbf{S}_t and $\boldsymbol{\tau}_t$, note that the respective elements S_{jt} and τ_{jt} are equation-specific with $j \in \{L, \check{S}, \check{C}\}$ 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 \mathbf{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, $S_{\check{C},t}$, by contrast, 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 $S_{\check{C},t}$ is consistently above 0.5,

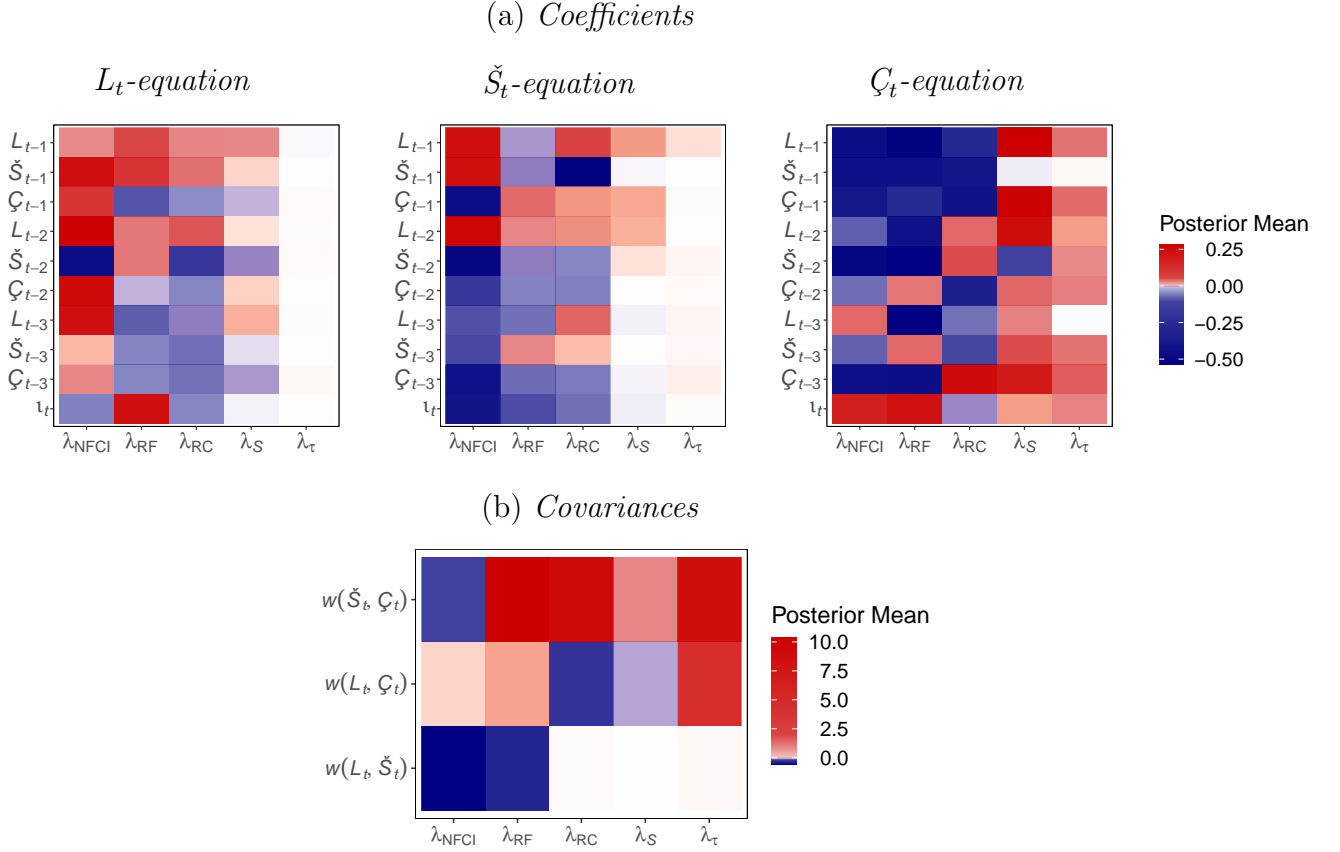
albeit with several high-frequency movements. In the aftermath of the global financial crisis, with short-term rates approaching the zero lower bound, $S_{C,t}$ switches abruptly into regime 0. Conditional on the respective loadings in $\mathbf{\Lambda}_S$ being non-zero, this feature would directly 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, $\tau_{S,t}$ and $\tau_{C,t}$ 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 \mathcal{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. $\tau_{C,t}$ 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 $\tilde{\mathbf{z}}_t$ must be considered in light of the rescaled loadings in $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\mathbf{U}^{-1}$. $\tilde{\mathbf{\Lambda}}$ translates the law of motion captured in $\tilde{\mathbf{z}}_t$ to the coefficients in $\boldsymbol{\beta}_t$ by acting either as amplifier or attenuator. Figure 2 shows the posterior mean of the rescaled loadings $\mathbf{\Lambda}\mathbf{U}^{-1}$ and allows to assess which elements in $\tilde{\mathbf{z}}_t$ determine the 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 \mathbf{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 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} , C_{t-2} and L_{t-3} ,

Figure 2: Heat maps for rescaled loadings in $\tilde{\Lambda} = \Lambda U^{-1}$.



Notes: $\tilde{\Lambda}$ translates the law of motion captured in \tilde{z}_t to (a) the VAR coefficients in β_t , and (b) the covariances stored in q_t , based on a TVP-NS-VAR model specification with $z_t = (r'_t, S'_t, \tau'_t)'$, $\delta = R_\tau/M = 1$ 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. λ_r relates to elements of $\tilde{\Lambda}$ associated with $r_{NFCI,t}, r_{REC,t}$ and $r_{RF,t}$ and the respective equations, λ_S denotes loadings related to a single Markov switching factor collected in S_t , and λ_τ loadings corresponding to latent random walk factors in τ_t . Note that S_{jt} and τ_{jt} for $j \in \{L, \check{S}, \check{C}\}$ are equation-specific quantities, while 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.

while for $r_{RF,t}$ and $r_{REC,t}$ some coefficients load moderately positive (e.g., the first own lag L_{t-1}) and others moderately negative (e.g., loadings related to lags of the curvature factor). The loadings Λ_S and Λ_τ 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 τ_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 S_t and τ_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 ζ_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 ζ_t , $w(\check{S}_t, \zeta_t)$, 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 ζ_t covariance, $w(L_t, \zeta_t)$, 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 ζ_t -equations, 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 ζ_t -equation and the covariances. Additional results showing the actual regression coefficients are provided in Appendix B.

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 ζ_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). Additional results for the reduced form coefficients are provided in Appendix B.

To construct the measure, we transform the TVP-VAR(P) in Eq. (1) to its state-space TVP-VAR(1) form. In what follows, the observation equation is given by $\mathbf{y}_t = \mathbf{J}\mathbf{Y}_t$, while the state equation is defined as $\mathbf{Y}_t = \mathbf{B}_t\mathbf{Y}_{t-1} + \boldsymbol{\omega}_t$ with $\boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t)$. Here, \mathbf{J} maps a K -dimensional vector $\mathbf{Y}_t = (\mathbf{y}'_t, \dots, \mathbf{y}'_{t-P+1})'$ to \mathbf{y}_t , \mathbf{B}_t collects the elements in $\boldsymbol{\beta}_t$ in the upper $M \times K$ block and defines identities otherwise. Similar, the $K \times K$ -dimensional variance-covariance matrix $\boldsymbol{\Omega}_t$ collects elements in $\boldsymbol{\Sigma}_t$ 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 \mathbf{y}_t at a zero frequency which coincides with the unconditional variance-covariance matrix of \mathbf{y}_t :

$$\boldsymbol{\Phi}_t(0) = \mathbf{J}(\mathbf{I} - \mathbf{B}_t)\boldsymbol{\Omega}_t(\mathbf{I} - \mathbf{B}'_t)^{-1}\mathbf{J}', \quad \text{for } t = 1, \dots, T.$$

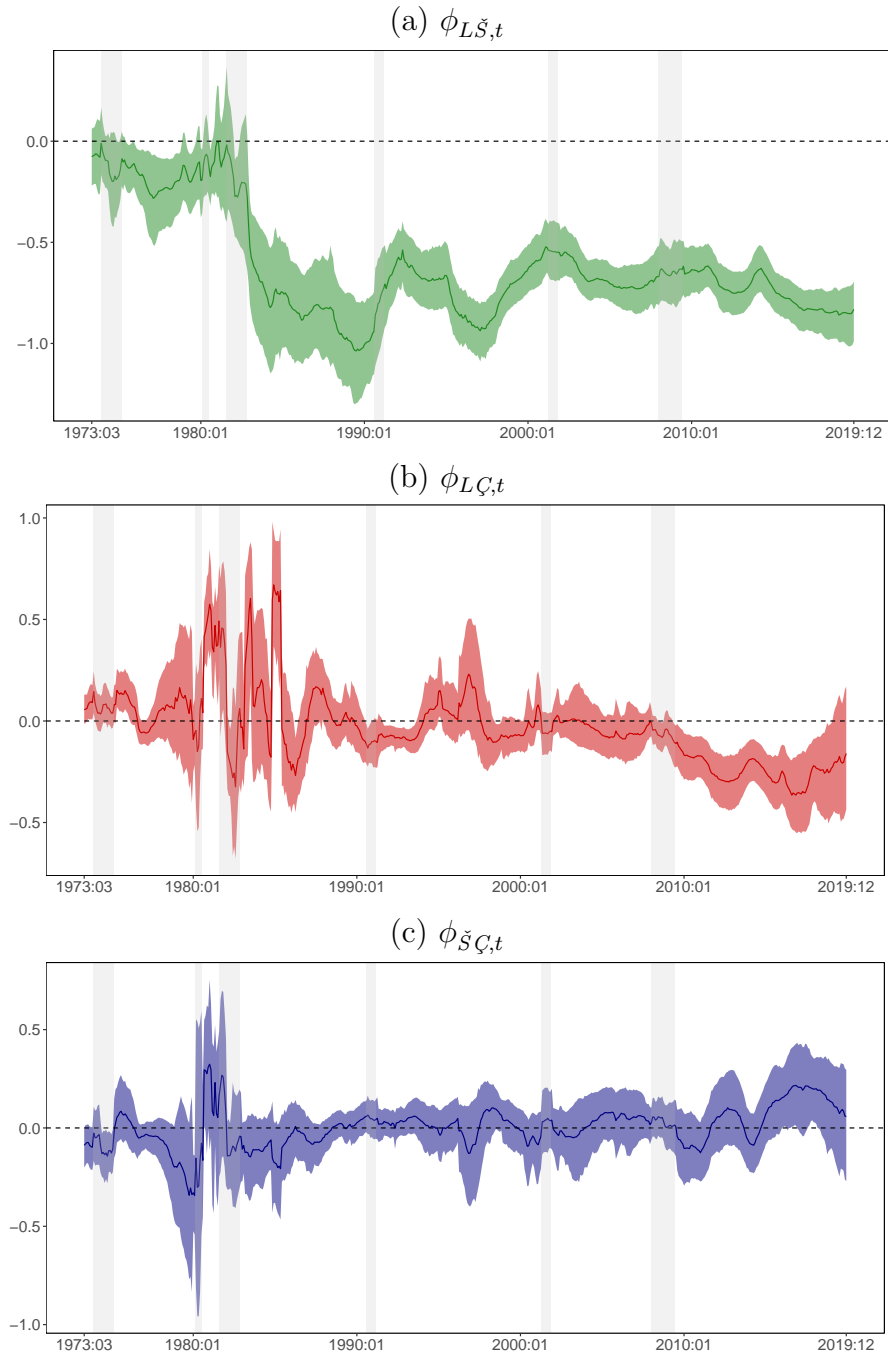
Next, we transform the covariances of $\boldsymbol{\Phi}_t$ into a correlation measure for each period t and each variable combination i, j ($i \neq j$ and $i, j = 1, \dots, M$):

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

The measure $\phi_{ij,t}$ describes the long-run relations between variable i and j at each point in time, and is displayed in Figure 3 for the model specification TVP-NS-VAR with $\mathbf{z}_t = (\mathbf{r}'_t, \mathbf{S}'_t, \boldsymbol{\tau}'_t)'$, $R_{\tau_j} = 1$ and $P = 3$. Note that the variables enter our model in differences. Hence, Figure 3 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

Figure 3: Posterior median and the 68 percent credible set for pairwise long-run correlations.



Notes: Panel (a) $\phi_{L\check{S},t}$, (b) $\phi_{L\check{C},t}$ and (c) $\phi_{\check{S}\check{C},t}$, based on a TVP-NS-VAR with $\mathbf{z}_t = (\mathbf{r}'_t, \mathbf{S}'_t, \boldsymbol{\tau}'_t)'$, $\delta = R\tau/M = 1$ and $P = 3$. The colored solid lines denote the posterior medians, the black dashed line the zero line, and the colored shaded band 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 $i, j \in \{L, \check{S}, \check{C}\}$. Results are based on 15,000 MCMC draws. Sample period: 1973:01 to 2019:12. Vertical axis: correlation measurements. Front axis: months.

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. Between 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 late 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

This paper proposes methods for automatically selecting adequate state equations in TVP-VAR models in a data-driven fashion. 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.

We apply our econometric framework to US yield curve data. Carrying out a thorough predictive exercise, we show that our techniques produce favorable point and density forecasts vis-à-vis a set of established benchmark models (which are nested variants of our proposed modeling approach). The performance is specific to the information set used in the underlying TVP-VAR and appears to be more pronounced for density forecasts. This 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 – a key advantage compared to conventional TVP-VARs. This predictive exercise is complemented by 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.

Acknowledgments: The authors gratefully acknowledge financial support 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*, in press, published online 6 Feb, doi: 10.1016/j.ijforecast.2021.01.003.
- ADRIAN, T., N. BOYARCHENKO, AND D. GIANNONE (2019): “Vulnerable growth,” *American Economic Review*, 109(4), 1263–1289, doi: 10.1257/aer.20161923.
- AGUILAR, O., AND M. WEST (2000): “Bayesian dynamic factor models and portfolio allocation,” *Journal of Business & Economic Statistics*, 18(3), 338–357, doi: 10.2307/1392266.
- BITTO, A., AND S. FRÜHWIRTH-SCHNATTER (2019): “Achieving shrinkage in a time-varying parameter model framework,” *Journal of Econometrics*, 210(1), 75–97, doi: 10.1016/j.jeconom.2018.11.006.
- CAGGIANO, G., E. CASTELNUOVO, AND G. PELLEGRINO (2017): “Estimating the real effects of uncertainty shocks at the zero lower bound,” *European Economic Review*, 100, 257–272, doi: 10.1016/j.eurocorev.2017.08.008.
- CARRIERO, A., T. CLARK, AND M. MARCELLINO (2019): “Large Vector Autoregressions with stochastic volatility and flexible priors,” *Journal of Econometrics*, 212(1), 137–154, doi: 10.1016/j.jeconom.2019.04.024.
- CARRIERO, A., T. E. CLARK, AND M. G. MARCELLINO (2020): “Capturing macroeconomic tail risks with Bayesian vector autoregressions,” Working Paper 202002R, Federal Reserve Bank of Cleveland.
- CARTER, C. K., AND R. KOHN (1994): “On Gibbs sampling for state space models,” *Biometrika*, 81(3), 541–553, doi: 10.2307/2337125.
- CARVALHO, C. M., N. G. POLSON, AND J. G. SCOTT (2010): “The horseshoe estimator for sparse signals,” *Biometrika*, 97(2), 465–480, doi: 10.1093/biomet/asq017.
- 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, doi: 10.1016/j.jeconom.2019.11.006.
- 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, doi: 10.1016/j.red.2004.10.009.
- DANGL, T., AND M. HALLING (2012): “Predictive regressions with time-varying coefficients,” *Journal of Financial Economics*, 106, 157–181, doi: 10.1016/j.jfineco.2012.04.003.
- DIEBOLD, F. X., AND C. LI (2006): “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130(2), 337–364, doi: 10.1016/j.jeconom.2005.03.005.
- 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, doi: 10.1016/j.jeconom.2008.08.017.
- FRÜHWIRTH-SCHNATTER, S. (1994): “Data augmentation and dynamic linear models,” *Journal of Time Series Analysis*, 15(2), 183–202, doi: 10.1111/j.1467-9892.1994.tb00184.x.
- 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, doi: 10.1016/j.jeconom.2009.07.003.
- GERLACH, R., C. CARTER, AND R. KOHN (2000): “Efficient Bayesian inference for dynamic mixture models,” *Journal of the American Statistical Association*, 95, 819–828, doi: 10.2307/2669465.
- GEWEKE, J., AND G. AMISANO (2010): “Comparing and evaluating Bayesian predictive distributions of asset returns,” *International Journal of Forecasting*, 26(2), 216–230, doi: 10.1016/j.ijforecast.2009.10.007.
- GEWEKE, J., AND G. ZHOU (1996): “Measuring the pricing error of the arbitrage pricing theory,” *Review of Financial Studies*, 9(2), 557–587, doi: 10.1093/rfs/9.2.557.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2015): “Prior selection for vector autoregressions,” *Review of Economics and Statistics*, 97(2), 436–451, doi: 10.1162/REST_a.00483.
- 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, doi: 10.1198/073500107000000241.

- 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, doi: 10.1016/j.jmoneco.2007.06.029.
- HASTIE, T., AND R. TIBSHIRANI (1993): “Varying-coefficient models,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–779, doi: 10.1111/j.2517-6161.1993.tb01939.x.
- HAUZENBERGER, N. (2020): “Flexible mixture priors for time-varying parameter models,” *arXiv preprint*, 2006.10088.
- HAUZENBERGER, N., F. HUBER, G. KOOP, AND L. ONORANTE (2019): “Fast and flexible Bayesian inference in time-varying parameter regression models,” *arXiv preprint*, 1910.10779.
- 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, doi: 10.1002/jae.2399.
- 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, doi: 10.1002/jae.2680.
- 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, doi: 10.1016/j.csda.2013.01.002.
- 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, doi: 10.1002/jae.2427.
- 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, doi: 10.1016/j.jedc.2008.11.003.
- KOROBILIS, D. (2019): “High-dimensional macroeconomic forecasting using message passing algorithms,” *Journal of Business & Economic Statistics*, in press, published online 22 Nov, doi: 10.1080/07350015.2019.1677472.
- 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, doi: 10.1002/jae.2606.
- MAKALIC, E., AND D. F. SCHMIDT (2015): “A simple sampler for the horseshoe estimator,” *IEEE Signal Processing Letters*, 23(1), 179–182, doi: 10.1109/LSP.2015.2503725.
- PRIMICERI, G. (2005): “Time varying structural autoregressions and monetary policy,” *Review of Economic Studies*, 72(3), 821–852, doi: 10.1111/j.1467-937X.2005.00353.x.
- 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, doi: 10.1257/aer.101.1.109.
- SIMS, C. A., AND T. ZHA (2006): “Were there regime switches in US monetary policy?,” *American Economic Review*, 96(1), 54–81, doi: 10.1257/000282806776157678.
- 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.

Appendices

A Technical appendix

A.1 Sampling the state innovation variances

For sampling the state innovation variances based on Eq. (1), we let $\eta_{ji,t}$ denote the shock to the i^{th} coefficient in $\tilde{\gamma}_t$ with respect to the j^{th} equation. The posterior of the state innovation variances is a generalized inverse Gaussian (GIG) distribution:⁹

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

A.2 Posterior for the horseshoe prior

Our specification of the horseshoe prior on Λ_j , γ_j and the square root of ω_j in Section 2.4 for a generic parameter b_i for $i = 1, \dots, K$ is:

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

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

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

⁹The 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 $x \sim \mathcal{GIG}(\lambda, \chi, \psi)$.

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

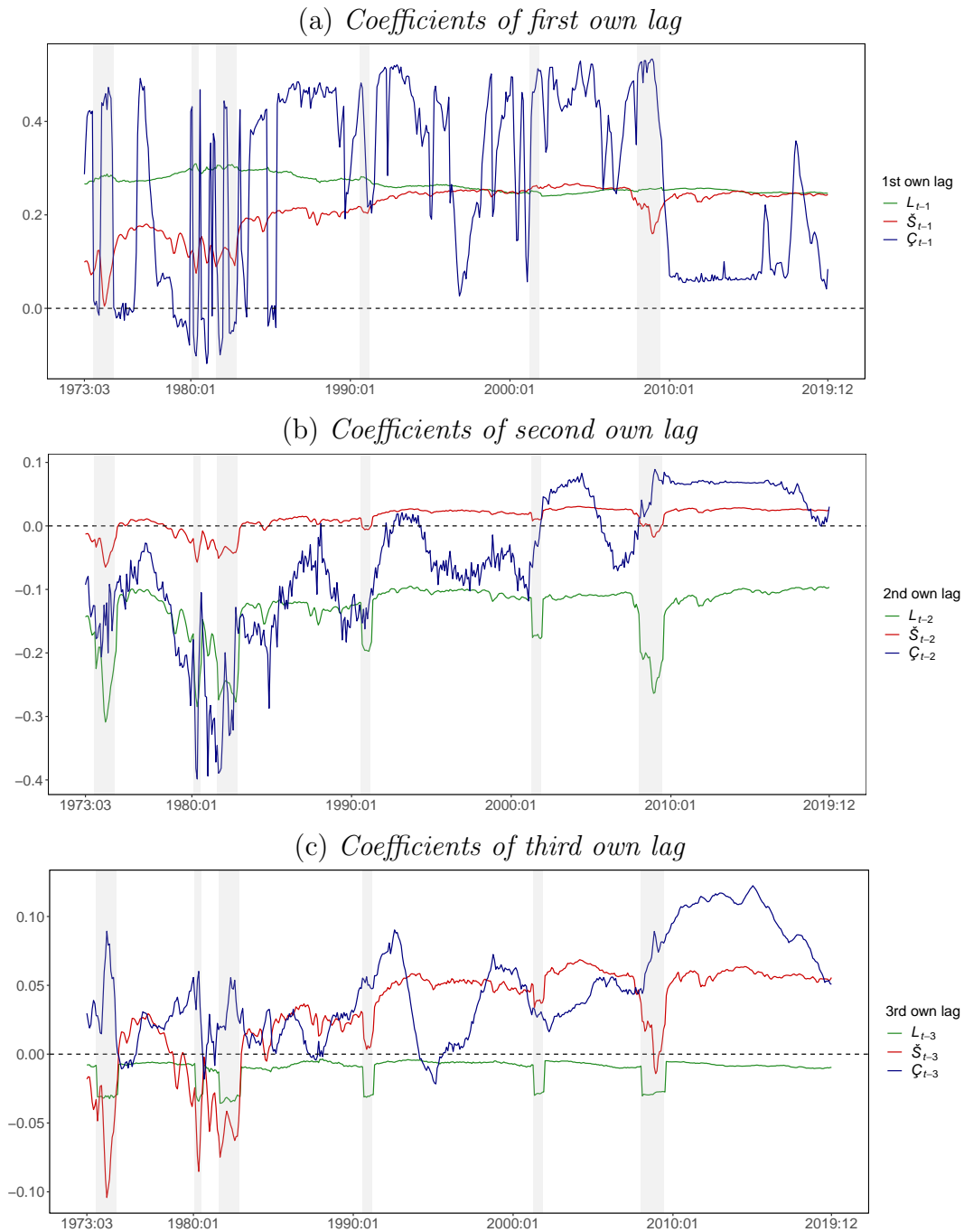
$$c_i^2|b_i, d, e_i \sim \mathcal{G}^{-1}\left(1, \frac{1}{e_i} + \frac{b_i^2}{2d^2}\right), \quad d^2|b_i, c_i, f \sim \mathcal{G}^{-1}\left(\frac{K+1}{2}, \frac{1}{f} + \sum_{i=1}^K \frac{b_i^2}{2c_i^2}\right),$$

$$e_i|c_i \sim \mathcal{G}^{-1}(1, 1 + c_i^{-2}), \quad f|d \sim \mathcal{G}^{-1}(1, 1 + d^{-2}).$$

B Further empirical results

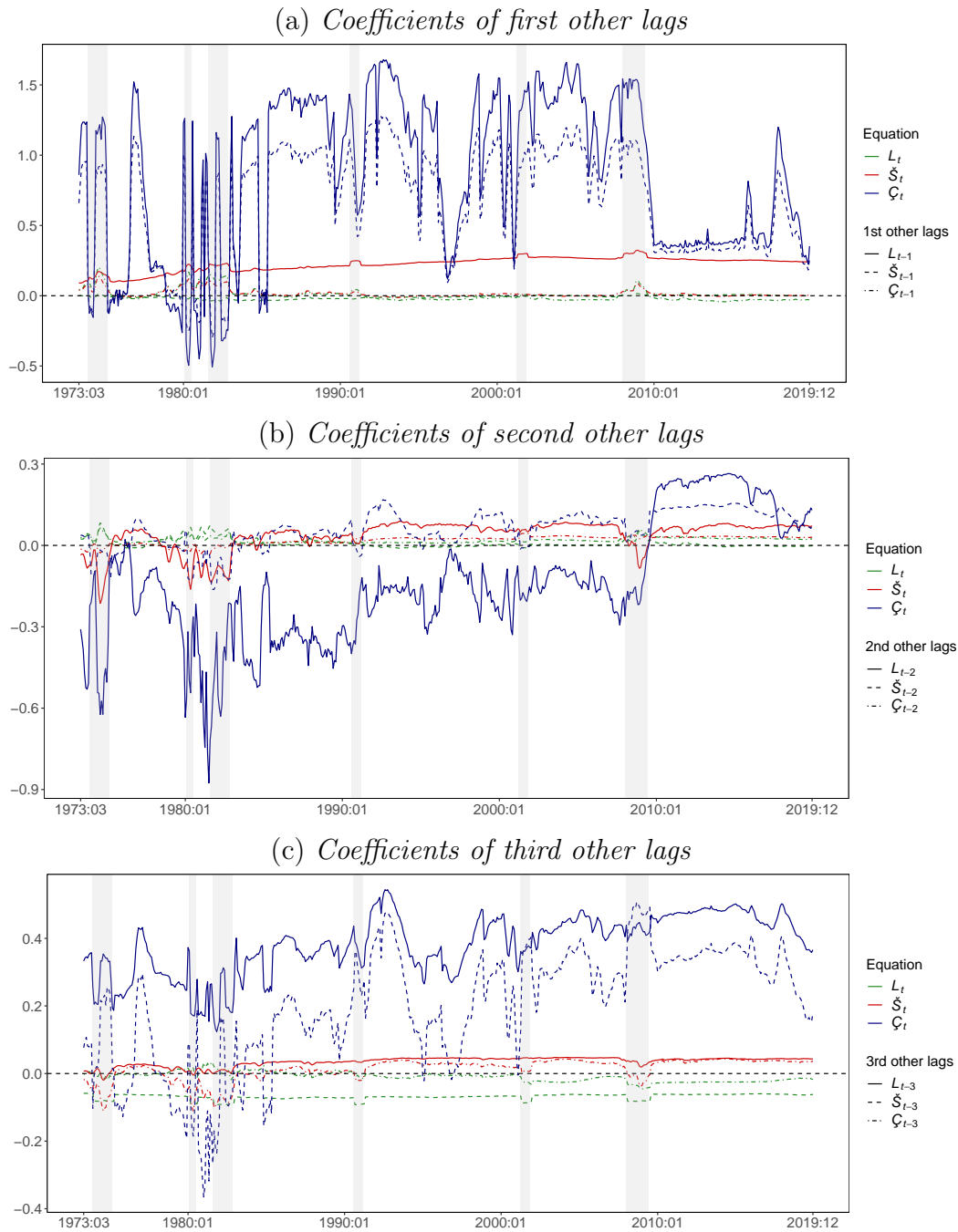
This Appendix contains additional results for the reduced form coefficients. While Subsection 3.4 provides posterior estimates of the lower dimensional effect modifiers, Figs. B.1–B.3 display the coefficients obtained by multiplying the factor loadings with the observed/latent factors based on the relationship established in Eq. (1).

Figure B.1: Posterior median of the coefficients associated with the own lags of L_t , \check{S}_t and ζ_t .



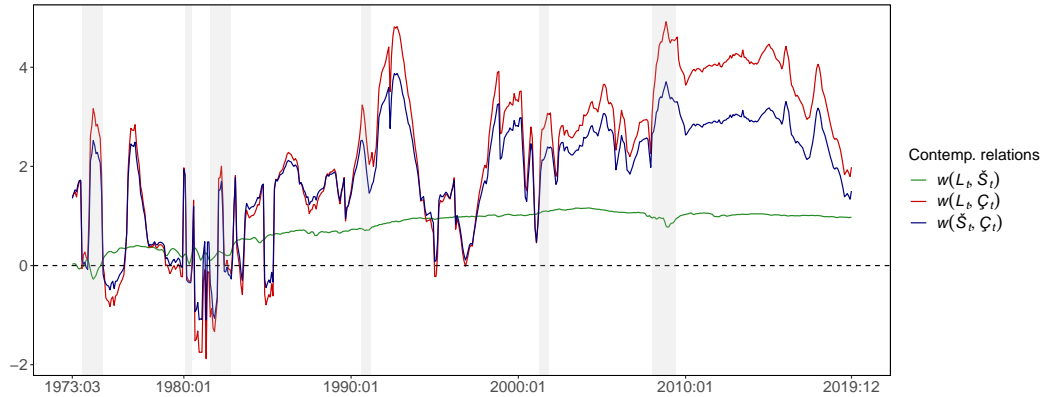
Notes: Panels (a), (b) and (c) show the dynamic evolution of the coefficients related to the variables' own lags $p \in \{1, 2, 3\}$ of the respective equation for L_t , \check{S}_t and ζ_t . Results are based on the TVP-NS-VAR model variant with $\delta = R_\tau/M = 1$ and using 15,000 MCMC draws. The 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. Front axis: months.

Figure B.2: Posterior median of the coefficients associated with cross-variable lags of L_t , \check{S}_t and ζ_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 ζ_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. Front axis: months.

Figure B.3: Posterior median of the contemporaneous relationships between L_t , \check{S}_t and \check{C}_t .



Notes: $w(\check{S}_t, \check{C}_t)$ 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. 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. Front axis: months.