Estimating a Large VAR of the Yield Curve and Macroeconomic Variables with the No-Arbitrage Restriction

Sun Ho Lee*
Kyu Ho Kang†

December 2023
This version is incomplete and preliminary.

Abstract

This paper proposes a comprehensive framework to explore the interaction between the macroeconomy and the government bond market. We integrate a term structure model of government bond yields with 28 macro-financial variables and 17 lags. Challenges in parameter estimation from a high-dimensional model of 16,896 parameters are addressed. Our approach utilizes Bayesian methods to efficiently estimate this extensive model, employing U.S. monthly data from 1987 to 2022. Notably, the inefficiency factors for all parameters are less than 2.38, suggesting that approximately two MCMC iterations are as efficient as one iid sampling. Our scenario analysis method enables the use of arbitrary linear combinations of observed variables at future time points, offering a flexible way to inspect the interactions. Furthermore, most hyperparameters are automatically determined through a data-driven approach, simplifying the process of setting prior distributions. The paper's methodology can be implemented through `GDTSM.jl` package written in Julia. (JEL classification: E43, C11, C58, G12)

Keywords: bond market, term structure model, arbitrage-free, scenario analysis, conditional forecasting, monetary policy, Markov Chain Monte Carlo (MCMC) simulation

*Ph.D. candidate, Department of Economics, Korea University, Seoul, South Korea, Email: preference@korea.ac.kr, Tel: +82-2-3290-2200.

†*Corresponding author*, Professor, Department of Economics, Korea University, Seoul, South Korea, Email: kyuho@korea.ac.kr, Tel: +82-2-3290-5132.

1 Introduction

Research on interest rates and the government bond market resides at the crossroads of macroeconomics and finance. In economics, interest rates determine consumers' intertemporal consumption and firms' investment decisions and influence the burden of government fiscal deficits. Furthermore, short-term interest rates are among the most traditional monetary policy tools. Central banks control short-term interest rates based on macroeconomic shocks, which affect the entire yield curve according to the expectations hypothesis. Additionally, unconventional monetary policies like quantitative easing impact the term premium, responding to macroeconomic shocks. Therefore, government bond investors pay attention to shocks from the macroeconomy while participating in the government bond and related derivative markets. Also, bond investors react directly to economic shocks by adjusting their portfolios. In this context, the government and bond market participants are interested in interactions between the macroeconomy and the bond market.

This paper presents an empirical methodology for analyzing the interactions between the bond market and macroeconomy by incorporating numerous macro-financial variables into a bond market model. We model the government bond market using Joslin, Singleton, and Zhu (2011)'s term structure model, and we enhance estimation efficiency by employing constraints from Christensen, Diebold, and Rudebusch (2011). Additionally, following Joslin, Priebsch, and Singleton (2014), we model the macroeconomy by incorporating 28 macro and financial variables in the form of the unspanned macro risks into the term structure model. Thus, the interaction between the government bond market and the macroeconomy is modeled through the dynamic interplay of three principal components and macro-financial variables in the physical transition equation. To further capture the rich interactions between these two areas, we follow Joslin, Le, and Singleton (2013a) by introducing lags of more than one in the physical transition equation. Consequently, the transition equation is modeled as a monthly VAR with 17 lags, and our state-space model encompasses 16,896 parameters. The number of parameters presents two major challenges. First, the model becomes susceptible to overfitting. Sec-

ond, even if overfitting is addressed, efficiently estimating such many parameters is not straightforward. This paper addresses both of these issues.

We address these problems using Bayesian methodology. The solution to the overfit has seen significant advancements through the Bayesian methods since Giannone, Lenza, and Primiceri (2015). The primary approach involves applying a Ridge-like penalty to the likelihood function using the Minnesota prior(Doan, Litterman, and Sims, 1984), effectively mitigating overfitting. The capability of Bayesian methodology to adeptly resolve overfit issues in a large VAR has been demonstrated in numerous studies(for example, Carriero, Clark, and Marcellino, 2019; Chan, 2021; Cimadomo, Giannone, Lenza, Monti, and Sokol, 2022).

In estimating our high-dimensional model, we employ the multi-block Markov Chain Monte Carlo (MCMC) approach. During this process, we strive to maximize efficiency in every aspect of the algorithm. In doing so, the posterior samples of all parameters, except those related to the transition equations for the bond market, are drawn using Gibbs sampling. Moreover, our proposal distribution for the Metropolis-Hastings (MH) algorithm possesses a closed-form expression and an effective approximation of the target distribution. In our estimation for the U.S. monthly data from 1987 to 2022, the inefficiency factors for all parameters are less than 2.38, indicating that roughly two MCMC iterations are as efficient as one iid sampling.

We present a methodology for scenario analysis in the context of our term structure model, based on Bańbura, Giannone, and Lenza (2015) and Crump, Eusepi, Giannone, Qian, and Sbordone (2021). In our methodology, scenarios for conditional forecasting assume arbitrary linear combinations of observed variables at arbitrary future points in time. To demonstrate practical applications of the scenario analysis method, we use the method to estimate what information about the future is contained in the projections of the Fed rate announced at the FOMC meetings. Additionally, we examine how monetary policy, the government bond market, and the macroeconomy respond if a financial uncertainty of the magnitude of the Global Financial Crisis occurs in 2023.

Our Bayesian methodology can be considered to maximize a penalized likelihood with a Ridge-type penalty. Nevertheless, one of the advantages of conducting esti-

mations within the Bayesian framework is the convenience of tuning hyperparameters compared to the frequentist approach. In a frequentist setting, tuning hyperparameters require a cross-validation process. By contrast, in Bayesian techniques, hyperparameter selection is treated as a part of model selection, and hyperparameters are chosen based on marginal likelihood. For instance, in the context of model averaging, one can assign prior distributions to hyperparameters and combine them with marginal likelihood to obtain a posterior distribution for the hyperparameters. Alternatively, in the context of selecting the best model, one can choose hyperparameters that maximize the marginal likelihood, corresponding to Bayesian model selection. We adopt the latter approach, which sets hyperparameters that maximize the marginal likelihood of the transition equation. Our marginal likelihood has a closed-form expression, allowing numerical maximization for hyperparameter tuning. Additionally, we design rules for setting the other hyperparameters based on data and prior belief conveniently.

Our estimation results show which of the 28 macro and financial variables primarily interact directly with the bond market. Interactions between the two sectors can occur simultaneously or with a lag. To observe simultaneous interactions, we look at the covariances in the physical transition equation, while for lagged interactions, we analyze the slope matrices of the transition equation. In our estimated model, nominal and financial variables mainly interact with the bond market concurrently, whereas interactions with real variables tend to occur with a lag. This indicates that the information immediately reflected in the yield curve primarily comprises nominal and financial information, while unspanned information largely pertains to real activity information.

Before concluding the introduction, we would like to mention two points. Firstly, although we use a large VAR, our methodology is not limited to large datasets. This is because our algorithm automatically adjusts the degree of shrinkage to the model's size, so our algorithm remains effective even for small-size term structure models. Nonetheless, employing many macro variables allows for a more accurate approximation of the macroeconomy, mitigates omitted variable bias, and leads to more accurate predictions based on a wealth of information. Moreover, if we maintain the full information rational expectation (FIRE) condition, including a large number of variables is preferable, as it

allows for calculating risk premiums based on all information available at the time when the expectations are formed.

Second, the reason we employ the Gaussian term structure model is that many term structure models are essentially modifications of the Gaussian models. Therefore, instead of selecting a term structure model with specific characteristics like the zero-lower-bound or trends, we believe that developing efficient algorithms for the universal Gaussian model broadens the applicability of our algorithm.

Notations To prevent an excessive increase in notation, we introduce a notation to indicate an entry in vectors or matrices. For a vector x, $x_{(i)}$ is the i-th entry in x, and $x_{(i:j)}$ is the slice of x from the i-th entry to the j-th entry. For a matrix X, $X_{(i,j)}$ is the entry in the i-th row and j-th column of X. In addition, $X_{(i,:)}$ and $X_{(:,j)}$ are the i-th row and j-th column of X, respectively. $X_{(i:j,m:n)}$ is the submatrix of X starting from the i-th row to the j-th row and from the m-th column to the n-th column.

2 Theoretical Term Structure Model

Our no-arbitrage affine term structure model of bond yields has a standard form of Joslin et al. (2011) (2011, JSZ). However, we have three characteristics different from the original JSZ form. First, we restrict the form of the slope matrix of the transition equation in the risk-neutral measure. Our restricted Gaussian affine term structure model is observationally equivalent to the Arbitrage-Free Nelson Siegel(AFNS) model of Christensen et al. (2011). Second, we introduce unspanned macro risks(Joslin et al., 2014) in the physical dynamics to make interactions between the macroeconomy and the bond market. Lastly, we allow multiple lags(Joslin et al., 2013a) in the physical transition equation to enrich the interaction between the macroeconomy and the bond market. The second and third characteristics introduce many unspanned factors, leading to a high dimensionality problem. We address it in the next section with the Bayesian framework.

2.1 Model in the Latent Factor Space

The short-term interest rate r_t at time t is

$$\underset{(1\times 1)}{r_t} = \underset{(1\times d_{\mathbb{Q}})}{\iota'} \underset{(d_{\mathbb{Q}}\times 1)}{X_t}, \tag{1}$$

where ι is a vector of ones and X_t is a $d_{\mathbb{Q}}$-dimensional vector of state variables. Since our model is a three-factor model, $d_{\mathbb{Q}} = 3$. But a generalization of our method to $d_{\mathbb{Q}} > 3$ is straightforward, so we express the dimension of $\mathbb{Q}$ as $d_{\mathbb{Q}}$. Under our risk-neutral measure $\mathbb{Q}$, the factor dynamics is

$$\begin{aligned}
\underset{(d_{\mathbb{Q}}\times 1)}{X_t} &= \underset{(d_{\mathbb{Q}}\times 1)}{K_X^{\mathbb{Q}}} + \underset{(d_{\mathbb{Q}}\times d_{\mathbb{Q}})}{G_{XX}^{\mathbb{Q}}} X_{t-1} + \underset{(d_{\mathbb{Q}}\times 1)}{\epsilon_{X,t}^{\mathbb{Q}}}, \\
K_X^{\mathbb{Q}} &\equiv \begin{bmatrix} \underset{(1\times 1)}{k_\infty^{\mathbb{Q}}} & O_{1\times(d_{\mathbb{Q}}-1)} \end{bmatrix}', \\
G_{XX}^{\mathbb{Q}} &= diag\left(1, \begin{pmatrix} \exp[-\kappa^{\mathbb{Q}}] & 1 \\ 0 & \exp[-\kappa^{\mathbb{Q}}] \end{pmatrix}\right), \\
\epsilon_{X,t}^{\mathbb{Q}} &\backsim \mathcal{N}(O_{d_{\mathbb{Q}}\times 1}, \Omega_{XX}).
\end{aligned} \tag{2}$$

$O_{m\times n}$ is a $m\times n$ zero matrix and *diag* is the diag operator that generates a block diagonal matrix.

$G_{XX}^{\mathbb{Q}}$ has a three-dimensional Jordan normal form matrix. We justify our specific form of $\mathbb{Q}$-eigenvalues in $G_{XX}^{\mathbb{Q}}$ by the below proposition.

Proposition 1. *When the eigenvalues of $G_{XX}^{\mathbb{Q}} - I_{d_{\mathbb{Q}}}$ are restricted to $\{0, \exp(-\kappa^{\mathbb{Q}}) - 1, \exp(-\kappa^{\mathbb{Q}}) - 1\}$, the restricted JSZ model is observationally equivalent to the AFNS model.*

Proof. See Online Appendix A and Joslin et al. (2011). □

Our AFNS restriction has several advantages. First, $\kappa^{\mathbb{Q}}$ has a direct economic meaning, so imposing prior distribution on this parameter is easy. In our setting, $\kappa^{\mathbb{Q}}$ is the dynamic Nelson-Siegel(DNS) decay parameter. In the DNS literature, the DNS decay parameter is adjusted so that the factor loading of the curvature is maximized on the representative medium maturity. We convert our prior belief about the representative

medium maturity to a prior distribution for $\kappa^{\mathbb{Q}}$. Second, we can utilize the Gibbs sampling algorithm for sampling $G^{\mathbb{Q}}_{XX}$. When we guess the representative medium maturity, the candidates of the maturity would be with a set of integers. Therefore, the prior distributions of the representative medium maturity and the decay parameter $\kappa^{\mathbb{Q}}$ have supports of integer values. We do not need a special treatment for sampling the posterior distribution of $G^{\mathbb{Q}}_{XX}$ because the parameter has a discrete prior distribution. Lastly, we can infer which maturity represents the movement in the curvature of the yield curve from the posterior distribution of $\kappa^{\mathbb{Q}}$.

Despite these advantages, one might like to use the unrestricted JSZ form. In this case, the user can impose their prior belief on $G^{\mathbb{Q}}_{XX}$, and some changes will occur in the empirical methodology part for $G^{\mathbb{Q}}_{XX}$. However, the remaining parts remain the same since our AFNS restriction is not essential for our methodology. Whether or not we adopt the restriction is not an important issue in this paper.

Based on $\mathbb{Q}$-dynamics, the bond yield of maturity τ at time t is

$$R_{\tau,t} = \frac{a_\tau}{\tau} + \frac{b'_\tau}{\tau} X_t + e_{\tau,t}, \tag{3}$$

where

$$\begin{aligned} a_\tau =& a_{\tau-1} + b'_{\tau-1} K^{\mathbb{Q}}_X - 0.5 b'_{\tau-1} \Omega_{XX} b_{\tau-1}, \\ b_\tau =& \iota + G^{\mathbb{Q}\prime}_{XX} b_{\tau-1}, \\ e_{\tau,t} \backsim& \mathcal{N}(0, \sigma^2_{\tau,e}), \end{aligned}$$

(Joslin et al., 2011). Here, $e_{\tau,t}$ is a measurement error.

We complete our model by defining the physical dynamics $\mathbb{P}$. Our $\mathbb{P}$ dynamics is

$$\underbrace{\begin{bmatrix} X_t \\ {\scriptstyle (d_{\mathbb{Q}}\times 1)} \\ M_t \end{bmatrix}}_{F_t : d_{\mathbb{P}}\times 1} = \underbrace{\begin{bmatrix} K^{\mathbb{P}}_X \\ {\scriptstyle (d_{\mathbb{Q}}\times 1)} \\ K^{\mathbb{P}}_M \end{bmatrix}}_{K^{\mathbb{P}}_F : d_{\mathbb{P}}\times 1} + \sum_{l=1}^{p} \underbrace{\begin{pmatrix} G^{\mathbb{P}}_{XX,l} & G^{\mathbb{P}}_{XM,l} \\ {\scriptstyle (d_{\mathbb{Q}}\times d_{\mathbb{Q}})} & \\ G^{\mathbb{P}}_{MX,l} & G^{\mathbb{P}}_{MM,l} \end{pmatrix}}_{G^{\mathbb{P}}_{FF,l} : d_{\mathbb{P}}\times d_{\mathbb{P}}} \begin{bmatrix} X_{t-l} \\ M_{t-l} \end{bmatrix} + \underbrace{\begin{bmatrix} \epsilon^{\mathbb{P}}_{X,t} \\ {\scriptstyle (d_{\mathbb{Q}}\times 1)} \\ \epsilon^{\mathbb{P}}_{M,t} \end{bmatrix}}_{\epsilon^{\mathbb{P}}_{F,t} : d_{\mathbb{P}}\times 1},$$

$$\epsilon^{\mathbb{P}}_{F,t} \backsim \mathcal{N}(O_{d_{\mathbb{P}}\times 1}, \underbrace{\Omega_{FF}}_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \equiv \begin{pmatrix} \underbrace{\Omega_{XX}}_{d_{\mathbb{Q}}\times d_{\mathbb{Q}}} & \Omega'_{MX} \\ \Omega_{MX} & \Omega_{MM} \end{pmatrix}).$$

As we mentioned earlier, there are two generalizations relative to the JSZ in $\mathbb{P}$-dynamics. First, we increase the lag of $\mathbb{P}$-dynamics to p. Second, we increase the dimension of $\mathbb{P}$-dynamics by including the unspanned macro risks. These two generalizations introduce the large-scale Bayesian VAR system in the term structure of interest rates.

2.2 Affine Transformation toward the Principal Component Space

As Joslin et al. (2011) do, we rotate the model to set the principal components to pricing factors. First, the relationship between the principal components and the latent factor X_t from equation (3) is

$$\begin{aligned} \begin{bmatrix} \mathcal{P}_t - c \\ \mathcal{O}_t \end{bmatrix} &:= \begin{pmatrix} W_{\mathcal{P}} \\ W_{\mathcal{O}} \end{pmatrix} R_t - \begin{bmatrix} c \\ O_{(N-d_{\mathbb{Q}})\times 1} \end{bmatrix}, \\ &= \begin{pmatrix} W_{\mathcal{P}}\mathcal{A}_X - c \\ W_{\mathcal{O}}\mathcal{A}_X \end{pmatrix} + \begin{pmatrix} W_{\mathcal{P}} \\ W_{\mathcal{O}} \end{pmatrix} \mathcal{B}_X X_t + \begin{bmatrix} O_{d_{\mathbb{Q}}\times 1} \\ e_{\mathcal{O},t} \end{bmatrix}, \qquad (4) \\ \text{where } R'_t &:= \begin{bmatrix} R_{\tau_1,t} & R_{\tau_2,t} & \cdots & R_{\tau_N,t} \end{bmatrix}, \\ \mathcal{A}'_X &:= \begin{bmatrix} a_{\tau_1}/\tau_1 & a_{\tau_2}/\tau_2 & \cdots & a_{\tau_N}/\tau_N \end{bmatrix}, \qquad (5) \\ \mathcal{B}'_X &:= \begin{bmatrix} b_{\tau_1}/\tau_1 & b_{\tau_2}/\tau_2 & \cdots & b_{\tau_N}/\tau_N \end{bmatrix}, \\ e_{\mathcal{O},t} &\backsim \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, \Sigma_{\mathcal{O}} := diag(\begin{bmatrix} \sigma^2_{O,1} & \sigma^2_{O,2} & \cdots & \sigma^2_{\mathcal{O},N-d_{\mathbb{Q}}} \end{bmatrix})), \end{aligned}$$

where $\mathcal{P}_t$ and $\mathcal{O}_t$ is the first $d_{\mathbb{Q}}$ and remaining principal components and the corresponding eigenvector matrices are $W_{\mathcal{P}}$ and $W_{\mathcal{O}}$, respectively. In addition, we assume that the rotated measurement errors with $W_{\mathcal{P}}$ are zero (Joslin, Le, and Singleton, 2013b). c is a hyperparameter that enhances the stability of estimations. We use $\mathcal{P}_t - c$ as spanned pricing factors and set c as a sample mean of $\mathcal{P}_t$. Later, we use a demeaned macroeconomic dataset, so all pricing factors have zero sample means. The normalization of the dataset is a core process in the machine learning and big data literature.

The above relationship leads to the below affine transformation.

$$X_t = \mathcal{T}_{0,\mathcal{P}} + \mathcal{T}_{1,\mathcal{P}}(\mathcal{P}_t - c), \qquad (6)$$

$$\text{where } \mathcal{T}_{0,\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{XX}) := -\mathcal{T}_{1,\mathcal{P}}(W_{\mathcal{P}}\mathcal{A}_X - c), \tag{7}$$
$$\mathcal{T}_{1,\mathcal{P}}(\kappa^{\mathbb{Q}}) := (W_{\mathcal{P}}\mathcal{B}_X)^{-1}.$$

Expressions $\mathcal{T}_{0,\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{XX})$ and $\mathcal{T}_{1,\mathcal{P}}(\kappa^{\mathbb{Q}})$ mean that they depend on deep parameters($\kappa^{\mathbb{Q}}$, $k_{\infty}^{\mathbb{Q}}$ and Ω_{XX}). When we move from the space of $\{X_t, M_t\}$ to $\{\mathcal{P}_t - c, M_t\}$, $\mathbb{P}$-dynamics is

$$\underbrace{\begin{bmatrix}\mathcal{P}_t - c \\ M_t\end{bmatrix}}_{\mathcal{F}_t} = \underbrace{\begin{bmatrix}\mathcal{T}_{1,\mathcal{P}}^{-1}[K_X^{\mathbb{P}} - (I_{d_{\mathbb{Q}}} - \sum_{l=1}^{p} G_{XX,l}^{\mathbb{P}})\mathcal{T}_{0,\mathcal{P}}] \\ K_M^{\mathbb{P}} + (\sum_{l=1}^{p} G_{MX,l}^{\mathbb{P}})\mathcal{T}_{0,\mathcal{P}}\end{bmatrix}}_{K_{\mathcal{F}}^{\mathbb{P}}}$$
$$+ \sum_{l=1}^{p} \underbrace{\begin{pmatrix}\mathcal{T}_{1,\mathcal{P}}^{-1} G_{XX,l}^{\mathbb{P}} \mathcal{T}_{1,\mathcal{P}} & \mathcal{T}_{1,\mathcal{P}}^{-1} G_{XM,l}^{\mathbb{P}} \\ G_{MX,l}^{\mathbb{P}} \mathcal{T}_{1,\mathcal{P}} & G_{MM,l}^{\mathbb{P}}\end{pmatrix}}_{G_{\mathcal{F}\mathcal{F},l}^{\mathbb{P}}} \mathcal{F}_{t-l} + \underbrace{\begin{bmatrix}\mathcal{T}_{1,\mathcal{P}}^{-1}\epsilon_{X,t}^{\mathbb{P}} \\ \epsilon_{M,t}^{\mathbb{P}}\end{bmatrix}}_{\epsilon_{\mathcal{F},t}^{\mathbb{P}}}.$$

We define new notations to compactly express $\mathbb{P}$-dynamics. Then,

$$\mathcal{F}_t = \underbrace{\begin{bmatrix}K_{\mathcal{P}}^{\mathbb{P}} \\ K_{M|\mathcal{F}}^{\mathbb{P}}\end{bmatrix}}_{K_{\mathcal{F}}^{\mathbb{P}}} + \sum_{l=1}^{p} \underbrace{\begin{pmatrix}G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}} & G_{\mathcal{P}M,l}^{\mathbb{P}} \\ G_{M\mathcal{P},l}^{\mathbb{P}} & G_{MM,l}^{\mathbb{P}}\end{pmatrix}}_{G_{\mathcal{F}\mathcal{F},l}^{\mathbb{P}}} \mathcal{F}_{t-l} + \underbrace{\begin{bmatrix}\epsilon_{\mathcal{P},t}^{\mathbb{P}} \\ \epsilon_{M,t}^{\mathbb{P}}\end{bmatrix}}_{\epsilon_{\mathcal{F},t}^{\mathbb{P}}}, \tag{8}$$
$$\epsilon_{\mathcal{F},t}^{\mathbb{P}} \backsim \mathcal{N}(O_{d_{\mathbb{P}}\times 1}, \Omega_{\mathcal{F}\mathcal{F}} := \begin{pmatrix}\Omega_{\mathcal{P}\mathcal{P}} & \Omega'_{M\mathcal{P}} \\ \Omega_{M\mathcal{P}} & \Omega_{MM}\end{pmatrix} := \begin{pmatrix}\mathcal{T}_{1,\mathcal{P}}^{-1}\Omega_{XX}\mathcal{T}_{1,\mathcal{P}}^{-1\prime} & \mathcal{T}_{1,\mathcal{P}}^{-1}\Omega'_{MX} \\ \Omega_{MX}\mathcal{T}_{1,\mathcal{P}}^{-1\prime} & \Omega_{MM}\end{pmatrix}).$$

This is our transition equation in the state-space form representation. Our measurement equation is derived from equations (4) and (6). That is,

$$\mathcal{O}_t = \underbrace{W_{\mathcal{O}}[\mathcal{A}_X + \mathcal{B}_X\mathcal{T}_{0,\mathcal{P}}]}_{\mathcal{A}_{\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})} + \underbrace{W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}}_{\mathcal{B}_{\mathcal{P}}(\kappa^{\mathbb{Q}})}(\mathcal{P}_t - c) + e_{\mathcal{O},t}. \tag{9}$$

$\mathcal{A}_{\mathcal{P}}$ and $\mathcal{B}_{\mathcal{P}}$ have deep parameters, and $\mathcal{A}_{\mathcal{P}}$ depends $\Omega_{\mathcal{P}\mathcal{P}}$ not Ω_{XX}. Since $\Omega_{XX} \equiv \mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}$ and $\mathcal{T}_{1,\mathcal{P}}$ depends only on $\kappa^{\mathbb{Q}}$, all expressions written with $\{\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{XX}\}$ can be rewritten with $\{\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}\}$. In our paper, we express $\mathbb{Q}$ deep-parameters with $\{\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}\}$, because $\Omega_{\mathcal{P}\mathcal{P}}$ is the one estimated later.

$\mathcal{P}_t - c$ under $\mathbb{Q}$ follows

$$\mathcal{P}_t - c = \underbrace{\mathcal{T}_{1,\mathcal{P}}^{-1}[K_X^{\mathbb{Q}} + (G_{XX}^{\mathbb{Q}} - I_{d_{\mathbb{Q}}})\mathcal{T}_{0,\mathcal{P}}]}_{K_{\mathcal{P}}^{\mathbb{Q}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})} + \underbrace{\mathcal{T}_{1,\mathcal{P}}^{-1}G_{XX}^{\mathbb{Q}}\mathcal{T}_{1,\mathcal{P}}}_{G_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}(\kappa^{\mathbb{Q}})}(\mathcal{P}_{t-1} - c) + \underbrace{\mathcal{T}_{1,\mathcal{P}}^{-1}\epsilon_{X,t}^{\mathbb{Q}}}_{\epsilon_{\mathcal{P},t}^{\mathbb{Q}}}, \tag{10}$$

$$\epsilon^{\mathbb{Q}}_{\mathcal{P},t} \backsim \mathcal{N}(O_{d_{\mathbb{Q}}\times 1}, \Omega_{\mathcal{P}\mathcal{P}}),$$

and the short-term interest rate is

$$r_t = \underbrace{\iota'\mathcal{T}_{0,\mathcal{P}}}_{\delta(\kappa^{\mathbb{Q}},k^{\mathbb{Q}}_{\infty},\Omega_{\mathcal{P}\mathcal{P}})} + \underbrace{\iota'\mathcal{T}_{1,\mathcal{P}}}_{\beta(\kappa^{\mathbb{Q}})'}(\mathcal{P}_t - c).$$

Parameters $K^{\mathbb{Q}}_{\mathcal{P}}$, $G^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$, δ, and β also depend on specific deep parameters in $\mathbb{Q}$. Lastly, we use the difference equations (3) in terms of $\{\kappa^{\mathbb{Q}}, k^{\mathbb{Q}}_{\infty}, \Omega_{\mathcal{P}\mathcal{P}}\}$, so

$$a_\tau = a_{\tau-1} + b'_{\tau-1}K^{\mathbb{Q}}_X - 0.5b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}b_{\tau-1}, \tag{11}$$

$$b_\tau = \iota + G^{\mathbb{Q}\prime}_{XX}b_{\tau-1}. \tag{12}$$

2.3 Risk Premiums

Following Cheridito, Filipovic, and Kimmel (2007), we define the market prices of risks

$$\begin{aligned}
\underset{(d_{\mathbb{P}}\times 1)}{\Omega^{1/2}_{\mathcal{F}\mathcal{F}}\lambda_t} &\equiv \mathbb{E}^{\mathbb{P}}_t[\mathcal{F}_{t+1}] - \mathbb{E}^{\mathbb{Q}}_t[\mathcal{F}_{t+1}], \\
&= \begin{bmatrix} \lambda_{\mathcal{P}} + \sum_{l=1}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}\mathcal{P}_{t+1-l} + \sum_{l=1}^{p}\Lambda_{\mathcal{P}M,l}M_{t+1-l} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times 1} \end{bmatrix}, \\
where\ \underset{(d_{\mathbb{Q}}\times 1)}{\lambda_{\mathcal{P}}} &= K^{\mathbb{P}}_{\mathcal{P}} - K^{\mathbb{Q}}_{\mathcal{P}}, \\
\underset{(d_{\mathbb{Q}}\times d_{\mathbb{Q}})}{\Lambda_{\mathcal{P}\mathcal{P},l}} &= \begin{cases} G^{\mathbb{P}}_{\mathcal{P}\mathcal{P},1} - G^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}} & \text{if } l = 1, \\ G^{\mathbb{P}}_{\mathcal{P}\mathcal{P},l} & \text{otherwise,} \end{cases} \\
\underset{(d_{\mathbb{Q}}\times(d_{\mathbb{P}}-d_{\mathbb{Q}}))}{\Lambda_{\mathcal{P}M,l}} &= G^{\mathbb{P}}_{\mathcal{P}M,l}.
\end{aligned} \tag{13}$$

Also, our specifications for $\mathbb{P}$ and $\mathbb{Q}$ leads to the Radon-Nikodym derivative of

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp[-\frac{1}{2}\lambda'_t\lambda_t - \lambda'_t v^{\mathbb{P}}_{\mathcal{F},t+1}].$$

Therefore, the model-consistent SDF $\mathcal{M}^{\mathbb{P}}_{t+1}$ should be

$$\begin{aligned}
\mathcal{M}^{\mathbb{P}}_{t+1} &\equiv \exp[-r_t]\frac{d\mathbb{Q}}{d\mathbb{P}}, \\
&= \exp[-r_t - \frac{1}{2}\lambda'_t\lambda_t - \lambda'_t v^{\mathbb{P}}_{\mathcal{F},t+1}].
\end{aligned} \tag{14}$$

In appendix B, we derive the term premium formula,

$$\begin{aligned} TP_{\tau,t} = &- \frac{1}{\tau}\sum_{i=1}^{\tau-1}(0.5b_i'\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}_{1,\mathcal{P}}'b_i + b_i'\mathcal{T}_{1,\mathcal{P}}\lambda_{\mathcal{P}}) \\ &- \frac{1}{\tau}\sum_{i=1}^{\tau-1}\sum_{l=1}^{p} b_{\tau-i}'\mathcal{T}_{1,\mathcal{P}}\Lambda_{\mathcal{P}\mathcal{P},l}\mathbb{E}_t^{\mathbb{P}}[\mathcal{P}_{t+i-l} - c] \\ &- \frac{1}{\tau}\sum_{i=1}^{\tau-1}\sum_{l=1}^{p} b_{\tau-i}'\mathcal{T}_{1,\mathcal{P}}\Lambda_{\mathcal{P}M,l}\mathbb{E}_t^{\mathbb{P}}[M_{t+i-l}]). \end{aligned} \tag{15}$$

The above term premium expression has three components. The first line is the constant part. Since $\mathcal{P}_t - c$ and M_t are demeaned, the constant part has an influential role in deciding the average level of $TP_{\tau,t}$. The second and third lines are time-varying components. They are driven by the principal components and macroeconomic variables, respectively. Since $TP_{\tau,t}$ is a scalar, the right-hand side of the equation is a linear combination of constant and all entries in $\mathbb{E}_t^{\mathbb{P}}[\mathcal{P}_{t+k} - c]$ and $\mathbb{E}_t^{\mathbb{P}}[M_{t+k}]$ for all $k = 1 - p, \cdots, \tau - 2$. We group terms in the linear combination according to the type of a variable to calculate the contribution of each variable on $TP_{\tau,t}$. For example, when we denote the first principal component to $\mathcal{P}_{t,(1)}$, we can calculate the contribution of the first principal component by summing up all terms related to $\{\mathbb{E}_t^{\mathbb{P}}[\mathcal{P}_{t+k,(1)} - c]\}_{k=1-p}^{\tau-2}$ in the linear combination. The contribution shows the movement of the term premium generated by the expected path of the first principal component.

3 Empirical Methodology

Equations (8) and (9) constitute the state-space model. Although it is feasible to estimate the model using monthly data in decimals per month, this approach was not adopted due to several reasons. First, a small scale of the data leads to numerical instability easily. Also, since some parameters also have tiny scales due to the data, we have to impose extreme shapes on the prior distributions of those parameters. To detour these issues, we scale up the state-space model and use monthly data in percent per annum. The scaled state-space model is

$$
\begin{aligned}
1200 \cdot \mathcal{O}_t =& W_{\mathcal{O}}(1200 \cdot R_t) = 1200 \cdot \mathcal{A}_{\mathcal{P}} + \mathcal{B}_{\mathcal{P}}(1200 \cdot \mathcal{P}_t - 1200 \cdot c) \\
&+ \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, 1200^2 \cdot \Sigma_{\mathcal{O}}), \qquad (16) \\
\begin{bmatrix} 1200 \cdot \mathcal{P}_t - 1200 \cdot c \\ M_t \end{bmatrix} =& \sum_{l=1}^{p} \begin{pmatrix} G^{\mathbb{P}}_{\mathcal{P}\mathcal{P},l} & 1200 \cdot G^{\mathbb{P}}_{\mathcal{P}M,l} \\ 1200^{-1} \cdot G^{\mathbb{P}}_{M\mathcal{P},l} & G^{\mathbb{P}}_{MM,l} \end{pmatrix} \begin{bmatrix} 1200 \cdot \mathcal{P}_{t-l} - 1200 \cdot c \\ M_{t-l} \end{bmatrix} \\
&+ \begin{bmatrix} 1200 \cdot K^{\mathbb{P}}_{\mathcal{P}} \\ K^{\mathbb{P}}_{M|\mathcal{F}} \end{bmatrix} + \mathcal{N}(O_{d_{\mathbb{P}}\times 1}, \begin{pmatrix} 1200^2 \cdot \Omega_{\mathcal{P}\mathcal{P}} & 1200 \cdot \Omega'_{M\mathcal{P}} \\ 1200 \cdot \Omega_{M\mathcal{P}} & \Omega_{MM} \end{pmatrix}) \qquad (17)
\end{aligned}
$$

with the Jensen's inequality term

$$
\mathcal{J}_\tau(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) := \frac{0.5 b'_{\tau-1} \mathcal{T}_{1,\mathcal{P}}(1200^2 \cdot \Omega_{\mathcal{P}\mathcal{P}}) \mathcal{T}'_{1,\mathcal{P}} b_{\tau-1}}{1200}.
$$

The objects multiplied by 1200 are

$$
\mathcal{P}_t, \mathcal{O}_t, R_t, c, \mathcal{A}_{\mathcal{P}}, K^{\mathbb{P}}_{\mathcal{P}}, G^{\mathbb{P}}_{\mathcal{P}M,l}, \Omega_{M\mathcal{P}}, \mathcal{A}_X, \mathcal{T}_{0,\mathcal{P}}, a_\tau, k^{\mathbb{Q}}_\infty, K^{\mathbb{P}}_{\mathcal{P}}, \lambda_{\mathcal{P}}, \Lambda_{\mathcal{P}M,l}, \text{ and } TP_{\tau,t}.
$$

In addition, $\Sigma_{\mathcal{O}}$ and $\Omega_{\mathcal{P}\mathcal{P}}$ are multiplied by 1200^2 and $G^{\mathbb{P}}_{M\mathcal{P},l}$ is divided by 1200. The full description of the scaled state-space model can be found in Online Appendix C. To simplify notation, we omit the scale(1200, 1200^2 and 1200^{-1}) in all subsequent expressions. For example, we use an expression "$\Omega_{\mathcal{P}\mathcal{P}}$" to indicate $1200^2 \cdot \Omega_{\mathcal{P}\mathcal{P}}$.

3.1 Prior Distribution

This paper estimates the model using the Bayesian methodology. Our estimation starts with imposing the prior distribution on parameters to be estimated.

3.1.1 Error Covariance in the Transition Equation $\Omega_{\mathcal{F}\mathcal{F}}$

We impose

$$
\Omega_{\mathcal{F}\mathcal{F}} \backsim \mathcal{IW}(\nu_0, diag[\Omega_0]).
$$

We set Ω_0 so that the i-th diagonal in $\mathbb{E}[\Omega_{\mathcal{F}\mathcal{F}}]$ is a residual variance from the AR(p) regression for i-th variable in $\mathcal{F}_t$. p is the same one in equation (17). We decide p and ν_0 in a data-driven way in section 3.3.

Instead of using the InverseWishart-InverseWishart update, we convert the InverseWishart

distribution into a series of Normal-InverseGamma distributions using the following proposition.

Proposition 2. *(Chan, 2022) Consider the following normal-inverse-gamma priors on the diagonal elements of Σ and the lower triangular elements of A:*

$$\sigma_i^2 \backsim \mathcal{IG}(\frac{\nu_0 + i - n}{2}, \frac{s_i^2}{2}), i = 1, \cdots, n,$$
$$(A_{i,j}|\sigma_i^2) \backsim \mathcal{N}(0, \frac{\sigma_i^2}{s_j^2}), 1 \leq j < i \leq n, i = 2, \cdots, n.$$

Then $\tilde{\Sigma}^{-1} = A'\Sigma^{-1}A$ has the Wishart distribution $\tilde{\Sigma}^{-1} \backsim \mathcal{W}(v_0, S^{-1})$, where $S = diag(s_1^2, s_2^2 \cdots, s_n^2)$. It follows that $\tilde{\Sigma} \backsim \mathcal{IW}(\nu_0, S)$.

To utilize the proposition, we orthogonalize

$$\mathcal{F}_t = K_{\mathcal{F}}^{\mathbb{P}} + \sum_{l=1}^{p} G_{\mathcal{FF},l}^{\mathbb{P}} \mathcal{F}_{t-l} + \epsilon_{\mathcal{F},t}^{\mathbb{P}},$$

and the orthogonalized system[1] is

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \mathcal{C}_{(2,1)} & 1 & 0 & \cdots & 0 \\ \mathcal{C}_{(3,1)} & \mathcal{C}_{(3,2)} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}_{(d_{\mathbb{P}},1)} & \mathcal{C}_{(d_{\mathbb{P}},2)} & \mathcal{C}_{(d_{\mathbb{P}},3)} & \cdots & 1 \end{pmatrix}}_{\mathcal{C}} \mathcal{F}_t = \mathcal{C}K_{\mathcal{F}}^{\mathbb{P}} + \sum_{l=1}^{p} \mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}} \mathcal{F}_{t-l} + diag(\underbrace{\begin{bmatrix} \sigma_{\mathcal{FF},1} \\ \vdots \\ \sigma_{\mathcal{FF},d_{\mathbb{P}}} \end{bmatrix}}_{\tilde{\Omega}_{\mathcal{FF}}^{1/2}}) v_{\mathcal{F},t}^{\mathbb{P}}. \tag{18}$$

$\mathcal{C}$ is implicitly defined by $\Omega_{\mathcal{FF}}^{-1} = \mathcal{C}' diag[\tilde{\Omega}_{\mathcal{FF}}]^{-1}\mathcal{C}$ (or $\Omega_{\mathcal{FF}} = \mathcal{C}^{-1} diag[\tilde{\Omega}_{\mathcal{FF}}]\mathcal{C}^{-1\prime}$). By the proposition, the InverseWishart prior is the same as

$$\sigma_{\mathcal{FF},i}^2 \backsim \mathcal{IG}(\frac{\nu_0 + i - d_{\mathbb{P}}}{2}, \frac{\Omega_{0,(i)}}{2}), i = 1, \cdots, d_{\mathbb{P}},$$
$$\mathcal{C}_{(i,j)}|\sigma_{\mathcal{FF},i}^2 \backsim \mathcal{N}(0, \frac{\sigma_{\mathcal{FF},i}^2}{\Omega_{0,(j)}}), 1 \leq j < i \leq d_{\mathbb{P}}, i = 2, \cdots, d_{\mathbb{P}}.$$

[1] As Chan (2022) pointed out, the order of variables in $\mathcal{F}_t$ is irrelevant to our estimation results. The orthogonalization is a tool for estimating our model efficiently, not for conducting structural inferences. We translate all posterior samples of $\{\mathcal{C}, \tilde{\Omega}_{\mathcal{FF}}, \mathcal{C}K_{\mathcal{F}}^{\mathbb{P}}, \{\mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}\}$ to that of $\{\Omega_{\mathcal{FF}}, K_{\mathcal{F}}^{\mathbb{P}}, \{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$, and use only the posterior samples of the reduced form parameters for our inference. In addition, we first define the prior distribution for the reduced form parameters and translate it to the prior distribution for the structural parameters. Therefore, the prior distribution for the structural parameters contains the same prior information regardless of the order of the variables.

We use the Normal-InverseGamma distributions when we estimate the model.

3.1.2 VAR Conditional Mean $K_{\mathcal{F}}^{\mathbb{P}}$ and $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$

First, we set prior means of $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$. We set prior mean of $G_{\mathcal{PP},1}^{\mathbb{P}}$ to the prior mean of $G_{\mathcal{PP}}^{\mathbb{Q}}$. For other diagonal elements in $G_{\mathcal{FF},1}^{\mathbb{P}}$, we set to one and zero for level and first-differenced variables, respectively. The other elements in $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$ have zero prior means. The prior means of $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$ set the prior means of $\{\Lambda_{\mathcal{PP},l}, \Lambda_{\mathcal{P}M,l}\}_{l=1}^{p}$ to zeros.

For the efficiency of the sampler, we estimate the orthogonalized equation (18). As Chan (2022) proves, since

$$(\mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}})_{(i,j)} = G_{\mathcal{FF},l,(i,j)}^{\mathbb{P}} + \sum_{k=1}^{i-1} \mathcal{C}_{ik}(G_{\mathcal{FF},l}^{\mathbb{P}})_{(k,j)},$$

equation

$$\mathbb{E}[(\mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}})_{(i,j)}] = \mathbb{E}[G_{\mathcal{FF},l,(i,j)}^{\mathbb{P}}]$$

holds. Therefore, we can use our prior means of $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$ as prior means of $\{\mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$.

For the intercept term, we set a prior mean of $K_{\mathcal{P}}^{\mathbb{P}}$ to be same to that of $K_{\mathcal{P}}^{\mathbb{Q}}$, so $\lambda_{\mathcal{P}}$ has a zero prior mean. We calculate the prior mean of $K_{\mathcal{P}}^{\mathbb{Q}}$ by the Monte-Carlo simulation.[2] Other parameters in $K_{\mathcal{F}}^{\mathbb{P}}$ have zero prior means. By the same logic of Chan (2022),

$$\mathbb{E}[(\mathcal{C}K_{\mathcal{F}}^{\mathbb{P}})_{(i)}] = \mathbb{E}[K_{\mathcal{F},(i)}^{\mathbb{P}}]$$

holds, because

$$(\mathcal{C}K_{\mathcal{F}}^{\mathbb{P}})_{(i)} = K_{\mathcal{F},(i)}^{\mathbb{P}} + \sum_{k=1}^{i-1} \mathcal{C}_{ik}(K_{\mathcal{F}}^{\mathbb{P}})_{(k)}.$$

Therefore, we set a prior mean of $\mathcal{C}K_{\mathcal{F}}^{\mathbb{P}}$ to the prior mean of $K_{\mathcal{F}}^{\mathbb{P}}$.

[2]Since the prior distribution of $K_{\mathcal{P}}^{\mathbb{Q}}$ comes from the prior distribution of $\kappa^{\mathbb{Q}}$, $k_{\infty}^{\mathbb{Q}}$, and $\Omega_{\mathcal{PP}}$ implicitly, it is difficult to derive an analytical form of the prior mean of $K_{\mathcal{P}}^{\mathbb{Q}}$. Therefore, we calculate the prior mean of $K_{\mathcal{P}}^{\mathbb{Q}}$ with the Monte Carlo simulation. The problem is that when we draw a prior sample of $K_{\mathcal{P}}^{\mathbb{Q}}$, we need to set ν_0. However, when we decide ν_0 in section 3.3, we need to set a prior mean of $K_{\mathcal{P}}^{\mathbb{Q}}$. So, the circular reasoning occurs. To address this problem, we fix $\Omega_{\mathcal{PP}}$ to its prior mean when we calculate a prior mean of $K_{\mathcal{P}}^{\mathbb{Q}}$. We also considered the case in which $\Omega_{\mathcal{FF}}$ is a random variable for various values of ν_0 in calculating the prior mean. We found that the prior mean was no different numerically from those obtained when $\Omega_{\mathcal{FF}}$ is fixed.

The remaining thing is to set prior variances. We use the Minnesota prior of Chan (2022), but there are some significant differences from the original one. Our prior variance is

$$Var[(\mathcal{C}G^{\mathbb{P}}_{\mathcal{FF},l})_{(i,j)}] = \frac{q_{slope}}{l^{q_{lag}}} \frac{\sigma^2_{\mathcal{FF},i}}{\Omega_{0,(j)}/(\nu_0 - d_{\mathbb{P}} - 1)},$$

$$\text{where } q_{slope} = \begin{cases} q_{11} & \text{if } i = j \text{ and } i \leq d_{\mathbb{Q}}, \\ q_{12} & \text{if } i = j \text{ and } i > d_{\mathbb{Q}}, \\ q_{21} & \text{if } i \neq j \text{ and } i \leq d_{\mathbb{Q}}, \\ q_{22} & \text{otherwise,} \end{cases}$$

$$q_{lag} = \begin{cases} q_{31} & \text{if } i \leq d_{\mathbb{Q}}, \\ q_{32} & \text{otherwise.} \end{cases}$$

$\{q_{11}, q_{12}\}$ penalizes coefficients on own lags, and $\{q_{21}, q_{22}\}$ penalizes coefficients on lags of other variables. We expect that the own lags have stronger predictive power than the lags of other variables, q_{11} and q_{12} is typically bigger than q_{21} and q_{22}. On the other hand, $\{q_{11}, q_{21}\}$ penalizes coefficients the first $d_{\mathbb{Q}}$ equations of system (18), and $\{q_{12}, q_{22}\}$ penalizes the remaining equations of the system. q_{lag} imposes a soft truncation on the number of lags in system (18). The degrees of truncation are different between the first $d_{\mathbb{Q}}$ and remaining equations. Lastly, $\sigma^2_{\mathcal{FF},i}$ and $\Omega_{0,(j)}/(\nu_0 - d_{\mathbb{P}} - 1)$ in the prior variance reflect scales of dependent variables and regressors. In the case of the intercept,

$$Var[(\mathcal{C}K^{\mathbb{P}}_{\mathcal{F}})_{(i)}] = \begin{cases} q_{41}\sigma^2_{\mathcal{FF},i} & \text{if } i \leq d_{\mathbb{Q}}, \\ q_{42}\sigma^2_{\mathcal{FF},i} & \text{otherwise.} \end{cases}$$

$\{q_{41}, q_{42}\}$ penalizes intercepts in the first $d_{\mathbb{Q}}$ and remaining equations of the system. Also, $\sigma^2_{\mathcal{FF},i}$ reflect the scale of dependent variables.

The reason why we impose different degrees of shrinkages between the first $d_{\mathbb{Q}}$ and remaining equations is that the shrinkage on the first $d_{\mathbb{Q}}$ equation essentially serves as the shrinkage on the market price of risk equation (13). As noted by Joslin and Le (2021), the $\mathbb{Q}$ parameters are more strongly identified by the data than the $\mathbb{P}$ parameters. Therefore, when we impose the shrinkage on $\mathbb{P}$ parameters, the shrinkage rarely affects the estimates of $\mathbb{Q}$ parameters, and the influences of the shrinkage are directly transferred to the market price of the risk equation. The relative importance of cross-variables may differ

when determining the bond market prices of risks and predicting future macroeconomic variables. Additionally, the degree of truncations on the lags may differ between these two contexts. Furthermore, intercepts in the market prices of risk equation have a special economic role of deciding the time-invariant level of risk premiums. This is why we optimize the hyperparameters differently between the first $d_{\mathbb{Q}}$ and the remaining equations.

Joint Prior Distribution for $\mathcal{C}$, $K_{\mathcal{F}}^{\mathbb{P}}$ and $\{G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$ Later, we estimate $\mathcal{C}K_{\mathcal{F}}^{\mathbb{P}}$ and $\{\mathcal{C}G_{\mathcal{FF},l}^{\mathbb{P}}\}_{l=1}^{p}$ with $\mathcal{C}$ jointly. So, we introduce a new notation

$$\phi_i = \left[\left[(\mathcal{C}K_{\mathcal{F}}^{\mathbb{P}})_{(i)} \quad (\mathcal{C}G_{\mathcal{FF},1}^{\mathbb{P}})_{(i,:)} \quad \cdots \quad (\mathcal{C}G_{\mathcal{FF},p}^{\mathbb{P}})_{(i,:)}\right] \quad \mathcal{C}_{(i:i,1:i-1)}\right]',$$

where $\mathcal{C}_{(1,1:0)}$ is an empty matrix, so we erase it from ϕ_1. In section 3.1.1 and 3.1.2, we describe the prior mean and variance of each entry in ϕ_i. Under the assumption of the prior independence, these prior means and variances are used to express the prior distribution of ϕ_i as

$$\phi_i \backsim \mathcal{N}(m_i, \sigma_{\mathcal{FF},i}^2 \times diag[V_i]).$$

We use the above prior distribution to estimate intercepts and coefficients of system (18). We decide a value of $q \equiv \{q_{11}, q_{12}, q_{21}, q_{22}, q_{31}, q_{32}, q_{41}, q_{42}\}$ in a data-driven way in section 3.3.

3.1.3 Prices of Risks $k_{\infty}^{\mathbb{Q}}$

Joslin and Le (2021) point out that $\mathbb{Q}$ parameters are estimated precisely, and we also found that the posterior distribution of $k_{\infty}^{\mathbb{Q}}$ is insensitive to its prior distribution during our research. Therefore, when we impose

$$k_{\infty}^{\mathbb{Q}} \backsim \mathcal{N}(0, \sigma_{k_{\infty}^{\mathbb{Q}}}^2),$$

an arbitrary number of $\sigma_{k_{\infty}^{\mathbb{Q}}}^2$ is acceptable as long as it is not excessively small. For example, if $\sigma_{k_{\infty}^{\mathbb{Q}}}$ is smaller than the posterior mean of $k_{\infty}^{\mathbb{Q}}$, it would be inappropriate. On the other hand, we found that the prior mean of $K_{\mathcal{P}}^{\mathbb{P}}$ determined by the MonteCarlo

simulation becomes unstable numerically when $\sigma^2_{k^{\mathbb{Q}}_\infty}$ is too large. Therefore, our recommend is to set $\sigma_{k^{\mathbb{Q}}_\infty}$ to the value slightly bigger than the posterior mean of $k^{\mathbb{Q}}_\infty$. To this end, some trial-and-errors are required. For robustness, we set a somewhat large value of $\sigma^2_{k^{\mathbb{Q}}_\infty} = 0.2^2$. Considering the estimated posterior mean of $k^{\mathbb{Q}}_\infty$ is 0.014, it is appropriate.

3.1.4 Decay Parameter $\kappa^{\mathbb{Q}}$

The DNS decay parameter is typically chosen to maximize the curvature factor loading at a specific maturity. The curvature factor loading in the DNS model is

$$\frac{1-\exp(-\kappa^{\mathbb{Q}}\tau)}{\kappa^{\mathbb{Q}}\tau} - \exp(-\kappa^{\mathbb{Q}}\tau).$$

We calculate $\kappa^{\mathbb{Q}}$ that maximize the curvature factor loading at each maturity $\tau \in \{36, 37, \cdots, 42\}$[3]. Suppose the maximizer decay parameters are $\{\kappa^{\mathbb{Q}}_{36}, \kappa^{\mathbb{Q}}_{37} \cdots, \kappa^{\mathbb{Q}}_{42}\}$. We use a uniform distribution on grid $\{\kappa^{\mathbb{Q}}_{36}, \kappa^{\mathbb{Q}}_{37} \cdots, \kappa^{\mathbb{Q}}_{42}\}$ as a prior distribution for $\kappa^{\mathbb{Q}}$.

The maturity at which the factor loading is maximized indicates which bond serves as the representative medium-term bond. Therefore, the grid on which the prior distribution is imposed reflects the researcher's belief about the representative medium-term bond.

3.1.5 Measurement Error $\Sigma_{\mathcal{O}}$

We impose a hierarchical prior structure on the measurement errors so that

$$\sigma^2_{\mathcal{O},i} \backsim \mathcal{IG}(2, \gamma_i),$$
$$\gamma_i \backsim \mathcal{G}(1, \bar{\gamma}).$$

$\mathbb{E}[\sigma^2_{\mathcal{O},i}] = \gamma_i$ and $\mathbb{E}[\gamma_i] = 1/\bar{\gamma}$ hold. We construct the hierarchical structure to set a prior distribution in a data-driven way. We set the hyperparameter $\bar{\gamma}$ to reflect the overall scale of $\{\sigma^2_{\mathcal{O},i}\}_{i=1}^{N-d_{\mathbb{Q}}}$. Specifically, we first regress equation (16) for each maturity separately with unrestricted coefficients. Then, we can obtain $N - d_{\mathbb{Q}}$ OLS residual variances. We set an inverse of $\bar{\gamma}$ to an average of the residual variances.

[3]For robustness, we also tried the grid $\{24, 26, \cdots, 58, 60\}$, but all posterior samples were drawn from grid $\{36, 38, \cdots, 42\}$. So, our prior is consistent with the data.

3.2 Likelihood Function

Our likelihood function is

$$p(\{\mathcal{O}_t, \mathcal{P}_t, M_t\}_{t=1}^T|\theta) = \Pi_{t=1}^T p(\mathcal{O}_t, \mathcal{P}_t, M_t|\mathcal{I}_{t-1}, \theta),$$
$$p(\mathcal{O}_t, \mathcal{P}_t, M_t|\mathcal{I}_{t-1}, \theta) = p(\mathcal{O}_t|\mathcal{P}_t, \theta)p(\mathcal{F}_t|\mathcal{I}_{t-1}, \theta),$$

where θ is a set of parameters and $\mathcal{I}_t$ is an information set at time t, including initial conditions. By

$$p(\mathcal{F}_{t,(1)}, \cdots, \mathcal{F}_{t,(d_{\mathbb{P}})}|\mathcal{I}_{t-1}, \theta) = p(\mathcal{F}_{t,(1)}|\mathcal{I}_{t-1}, \theta)\Pi_{i=2}^{d_{\mathbb{P}}} p(\mathcal{F}_{t,(i)}|\mathcal{F}_{t,(i-1)}, \cdots, \mathcal{F}_{t,(1)}, \mathcal{I}_{t-1}, \theta),$$

the likelihood function is

$$\begin{aligned} p(\{\mathcal{O}_t, \mathcal{P}_t, M_t\}_{t=1}^T|\theta) = \Pi_{t=1}^T p(\mathcal{O}_t|\mathcal{P}_t, \theta) \times \Pi_{t=1}^T p(\mathcal{F}_{t,(1)}|\mathcal{I}_{t-1}, \theta) \\ \times \Pi_{i=2}^{d_{\mathbb{P}}}[\Pi_{t=1}^T p(\mathcal{F}_{t,(i)}|\mathcal{F}_{t,(i-1)}, \cdots, \mathcal{F}_{t,(1)}, \mathcal{I}_{t-1}, \theta)]. \end{aligned}$$

The likelihood of the measurement equation (16) is

$$\Pi_{t=1}^T p(\mathcal{O}_t|\mathcal{P}_t, \theta) = \Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t|\mathcal{A}_{\mathcal{P}}(:\kappa^{\mathbb{Q}}, k_\infty^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) + \mathcal{B}_{\mathcal{P}}(:\kappa^{\mathbb{Q}})(\mathcal{P}_t - c), \Sigma_{\mathcal{O}}).$$

To derive the likelihood of the transition equation (18), we express the i-th equation of system (18) as

$$\mathcal{F}_{t,(i)} = \underbrace{\begin{bmatrix} 1 & \mathcal{F}'_{t-1} & \cdots & \mathcal{F}'_{t-p} & \mathcal{F}'_{t,(1:i-1)} \end{bmatrix}}_{x'_{\phi_i,t}} \phi_i + \mathcal{N}(O_{T\times 1}, \sigma^2_{\mathcal{F}\mathcal{F},i}).$$

In the above expression, $\mathcal{F}'_{t,(1:0)}$ is an empty matrix, so we erase it from the above equation when $i = 1$. The above equation represents $p(\mathcal{F}_{t,(1)}|\mathcal{I}_{t-1}, \theta)$ or $p(\mathcal{F}_{t,(i)}|\mathcal{F}_{t,(i-1)}, \cdots, \mathcal{F}_{t,(1)}, \mathcal{I}_{t-1}, \theta)$ for $i > 1$. It means that when we stack the i-th equations across all periods to construct

$$\underbrace{\begin{bmatrix} \mathcal{F}_{1,(i)} \\ \mathcal{F}_{2,(i)} \\ \vdots \\ \mathcal{F}_{T,(i)} \end{bmatrix}}_{y_{\phi_i}} = \underbrace{\begin{pmatrix} x'_{\phi_i,1} \\ x'_{\phi_i,2} \\ \vdots \\ x'_{\phi_i,T} \end{pmatrix}}_{X_{\phi_i}} \phi_i + \mathcal{N}(O_{T\times 1}, \sigma^2_{\mathcal{F}\mathcal{F},i} I_T),$$

distribution $\mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i}I_T)$ indicates

$$\Pi_{t=1}^{T} p(\mathcal{F}_{t,(i)}|\mathcal{F}_{t,(i-1)}, \cdots, \mathcal{F}_{t,(1)}, \mathcal{I}_{t-1}, \theta),$$

where it becomes $\Pi_{t=1}^{T} p(\mathcal{F}_{t,(1)}|\mathcal{I}_{t-1}, \theta)$ when $i = 1$. Therefore, the likelihood is

$$\begin{aligned} p(\{\mathcal{O}_t, \mathcal{P}_t, M_t\}_{t=1}^{T}|\theta) =& \Pi_{t=1}^{T} \mathcal{N}(\mathcal{O}_t|\mathcal{A}_{\mathcal{P}}(:\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{PP}}) + \mathcal{B}_{\mathcal{P}}(:\kappa^{\mathbb{Q}})(\mathcal{P}_t - c), \Sigma_{\mathcal{O}}) \\ & \times \Pi_{i=1}^{d_{\mathbb{P}}} \mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i}I_T). \end{aligned}$$

3.3 Empirical Bayes and the Marginal Likelihood

There are three types of hyperparameters that we should optimize in a data-driven way: They are p, ν_0, and q. We optimize them from the perspective of the empirical Bayes. Specifically, we maximize the marginal likelihood for the $\mathbb{P}$-transition equation (18) by adjusting p, ν_0, and q. The log marginal likelihood is

$$\begin{aligned} & \sum_{i=1}^{d_{\mathbb{P}}} \log \mathcal{N}(y_{\phi_i}|p, \nu_0, q) \\ =& -\frac{Td_{\mathbb{P}}}{2}\log[2\pi] + \sum_{i=1}^{d_{\mathbb{P}}}\{-\frac{1}{2}(\sum_{j=1}^{i+d_{\mathbb{P}}p} \log[V_{i,(j)}] + \log\det[K_{\phi_i}]) \\ & + \log\Gamma(\nu_i + \frac{T}{2}) + \nu_i \log S_i - \log\Gamma(\nu_i) - (\nu_i + \frac{T}{2})\log[\hat{S}_i]\} \end{aligned} \tag{19}$$

(Chan, 2022). Here, $\nu_i = (\nu_0 + i - d_{\mathbb{P}})/2$, $S_i = \Omega_{0,(i)}/2$, $K_{\phi_i} = diag[V_i]^{-1} + X'_{\phi_i}X_{\phi_i}$, $\hat{\phi}_i = K_{\phi_i}^{-1}(diag[V_i]^{-1}m_i + X'_{\phi_i}y_{\phi_i})$, $\hat{S}_i = S_i + (y'_{\phi_i}y_{\phi_i} + m'_i diag[V_i]^{-1}m_i - \hat{\phi}'_i K_{\phi_i}\hat{\phi}_i)/2$.

We use the differential evolution(DE) optimizer(Feldt, 2018)[4] to maximize our marginal likelihood. Our DE optimizer is a global optimizer similar to Wang, Li, Huang, and Li (2014). But, our algorithm is adaptive, so the setting of the optimizer is adjusted during the optimization. Also, the candidates are only influenced by their neighborhood in the radius-limited topology to increase the diversity of candidates(Spector and Klein, 2006). We use the default setting of Feldt (2018).

[4]Specifically, we use the adaptive DE/rand/1/bin with radius limited sampling. The notation(DE/X/Y/Z) is a summary of the characteristics of a specific differential evolution(DE) algorithm. To understand the notation's meaning and DE itself, see Opara and Arabas (2019).

3.4 Sampling Posterior Samples of Parameters

This section describes how to generate a sample from the posterior distribution of parameters. In our multi-block MCMC algorithm, the starting point is set to the prior mean of the parameters. We then sequentially update the parameter values, moving from block one through block six. During sampling within a single block, the parameters for the other blocks are set to their most recently updated values. Completion of one full loop from block one to block six generates a single joint posterior sample. We obtain 25,000 posterior samples of parameters and discard the first 5,000 as burn-in.

The six blocks are the full-conditional posterior distribution of each parameter. The procedures of deriving the full-conditional distributions are detailed in Appendix D. This section only presents the final expression for the full conditional distribution.

Block 1. $\mathbb{Q}$ **Constant Term** $k_\infty^{\mathbb{Q}}$ We solve the difference equations (11) and (12) forward. And combining them with equation (5) leads to

$$\mathcal{A}_X = \underbrace{\begin{bmatrix} -\sum_{i=2}^{\tau_1} \mathcal{J}_i(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})/\tau_1 \\ -\sum_{i=2}^{\tau_2} \mathcal{J}_i(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})/\tau_2 \\ \vdots \\ -\sum_{i=2}^{\tau_N} \mathcal{J}_i(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})/\tau_N \end{bmatrix}}_{\mathcal{A}_{0,k_\infty^{\mathbb{Q}}}} + \underbrace{\begin{bmatrix} \sum_{i=2}^{\tau_1} \frac{(i-1)}{\tau_1} \\ \sum_{i=2}^{\tau_2} \frac{(i-1)}{\tau_2} \\ \vdots \\ \sum_{i=2}^{\tau_N} \frac{(i-1)}{\tau_N} \end{bmatrix}}_{\mathcal{A}_{1,k_\infty^{\mathbb{Q}}}} k_\infty^{\mathbb{Q}}.$$

where $\sum_{i=2}^{\tau} \mathcal{J}_i(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})/\tau = \sum_{i=2}^{\tau} \frac{(i-1)}{\tau} = 0$ when $\tau = 1$. Therefore, $\mathcal{A}_X$ is an affine function of $k_\infty^{\mathbb{Q}}$, thereby rendering the measurement equation linear in $k_\infty^{\mathbb{Q}}$. It means that we can do Normal-Normal update for $k_\infty^{\mathbb{Q}}$. Specifically, we sample $k_\infty^{\mathbb{Q}}$ from

$$p(k_\infty^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{k_\infty^{\mathbb{Q}}\}) = \mathcal{N}((X'_{k_\infty^{\mathbb{Q}}} X_{k_\infty^{\mathbb{Q}}} + \sigma_{k_\infty^{\mathbb{Q}}}^{-2})^{-1} X'_{k_\infty^{\mathbb{Q}}} Y_{k_\infty^{\mathbb{Q}}}, (X'_{k_\infty^{\mathbb{Q}}} X_{k_\infty^{\mathbb{Q}}} + \sigma_{k_\infty^{\mathbb{Q}}}^{-2})^{-1}),$$

where

$$Y_{k_\infty^{\mathbb{Q}}} := \begin{bmatrix} \Sigma_{\mathcal{O}}^{-1/2}[\mathcal{O}_1 - W_{\mathcal{O}}(I_N - \mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}W_{\mathcal{P}})\mathcal{A}_{0,k_\infty^{\mathbb{Q}}} - W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}c - \mathcal{B}_{\mathcal{P}}(\mathcal{P}_1 - c)] \\ \Sigma_{\mathcal{O}}^{-1/2}[\mathcal{O}_2 - W_{\mathcal{O}}(I_N - \mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}W_{\mathcal{P}})\mathcal{A}_{0,k_\infty^{\mathbb{Q}}} - W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}c - \mathcal{B}_{\mathcal{P}}(\mathcal{P}_2 - c)] \\ \vdots \\ \Sigma_{\mathcal{O}}^{-1/2}[\mathcal{O}_T - W_{\mathcal{O}}(I_N - \mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}W_{\mathcal{P}})\mathcal{A}_{0,k_\infty^{\mathbb{Q}}} - W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}c - \mathcal{B}_{\mathcal{P}}(\mathcal{P}_T - c)] \end{bmatrix},$$

$$X_{k_\infty^{\mathbb{Q}}} := \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}'}_{(T\times 1)} \otimes (\Sigma_{\mathcal{O}}^{-1/2} W_{\mathcal{O}}(I_N - \mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}W_{\mathcal{P}})\mathcal{A}_{1,k_\infty^{\mathbb{Q}}}).$$

Block 2. Decay Parameter $\kappa^{\mathbb{Q}}$ We sample $\kappa^{\mathbb{Q}}$ from the probability mass density of

$$p(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\}) = \frac{p^*(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\})}{\sum_{j \in \{36,37,\cdots,42\}} p^*(\kappa^{\mathbb{Q}} = \kappa_j^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\})}, \text{ for } i \in \{36, 37, \cdots, 42\},$$

where

$$p^*(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\}) = \Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t | \mathcal{A}_{\mathcal{P}}(\kappa_i^{\mathbb{Q}}, k_\infty^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) + \mathcal{B}_{\mathcal{P}}(\kappa_i^{\mathbb{Q}})(\mathcal{P}_t - c), \Sigma_{\mathcal{O}}).$$

Block 3. MH Algorithm for $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_{\mathbb{Q}}}$ Parameters $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_{\mathbb{Q}}}$ affect both the likelihoods of the transition and measurement equations. So, we need to update the parameters using the MH algorithm. For each $i \in \{1, 2, \cdots, d_{\mathbb{Q}}\}$, we make a proposal for $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}$ from our proposal distribution of

$$\phi_i | \sigma^2_{\mathcal{FF},i}, \mathcal{I}_T \backsim \mathcal{N}(b_1, \sigma^2_{\mathcal{FF},i} B_1), \tag{20}$$

$$\sigma^2_{\mathcal{FF},i}, \mathcal{I}_T \backsim \mathcal{IG}(\frac{\nu_0 + i - d_{\mathbb{P}} + T}{2}, \frac{\Omega_{1,(i)}}{2}), \tag{21}$$

where $B_1 = (diag[V_i]^{-1} + X'_{\phi_i} X_{\phi_i})^{-1}$, $b_1 = B_1(diag[V_i]^{-1} m_i + X'_{\phi_i} y_{\phi_i})$, and $\Omega_{1,(i)} = \Omega_{0,(i)} + y'_{\phi_i} y_{\phi_i} + m'_i diag[V_i]^{-1} m_i - b'_1 B_1^{-1} b_1$.

The proposal distribution is the posterior distribution updated by only the transition equation. Our proposal has a good quality because our proposal distribution suggests a sample based on information in the transition equation, and the transition equation has rich information about $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}$. After we draw a proposal, we accept it with a

probability of

$$min[1, \frac{\Pi_{t=1}^{T}\mathcal{N}(\mathcal{O}_t|\mathcal{A}_{\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{PP},proposal}) + \mathcal{B}_{\mathcal{P}}(\kappa^{\mathbb{Q}})\mathcal{P}_t, \Sigma_{\mathcal{O}})}{\Pi_{t=1}^{T}\mathcal{N}(\mathcal{O}_t|\mathcal{A}_{\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{PP},previous}) + \mathcal{B}_{\mathcal{P}}(\kappa^{\mathbb{Q}})\mathcal{P}_t, \Sigma_{\mathcal{O}})}],$$

where $\Omega_{\mathcal{PP},proposal}$ and $\Omega_{\mathcal{PP},previous}$ are computed by

$$\Omega_{\mathcal{PP}} = (\mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})^{-1} diag[\sigma^2_{\mathcal{FF},1}, \sigma^2_{\mathcal{FF},2}, \cdots, \sigma^2_{\mathcal{FF},d_{\mathbb{Q}}}](\mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})^{-1\prime}.$$

In the above formula for $\Omega_{\mathcal{PP}}$, all parameter has the most recent value, except for $\sigma^2_{\mathcal{FF},i}$. When we compute $\Omega_{\mathcal{PP},proposal}$, $\sigma^2_{\mathcal{FF},i}$ is a proposed value. In the case of $\Omega_{\mathcal{PP},previous}$, $\sigma^2_{\mathcal{FF},i}$ is the value of $\sigma^2_{\mathcal{FF},i}$ in the previous loop. The logic is that if a new proposal increases the likelihood of the measurement equation, we accept it. If a new proposal has a lower likelihood value for the measurement equation, we accept the new proposal with a probability corresponding to the likelihood ratio between the new and previous one.

Block 4. Other Transition Equation Parameters $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}$ Since $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}$ is relevant for only the transition equation, sampling $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}$ is straightforward. For each $i \in \{d_{\mathbb{Q}}+1, d_{\mathbb{Q}}+2, \cdots, d_{\mathbb{P}}\}$, we sample $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}$ from equations (20) and (21).

Block 5. Measurement Error Variance $\Sigma_{\mathcal{O}}$ For each $i \in \{1, 2, \cdots, N - d_{\mathbb{Q}}\}$, we update $\sigma^2_{\mathcal{O},i}$ by sampling

$$p(\sigma^2_{\mathcal{O},i}|\mathcal{I}_T, \theta\backslash\{\Sigma_{\mathcal{O}}\}) = \mathcal{IG}(\sigma^2_{\mathcal{O},i}|2 + \frac{T}{2}, \gamma_i + \frac{(y_{\sigma^2_{\mathcal{O},i}} - m_{\sigma^2_{\mathcal{O},i}})'(y_{\sigma^2_{\mathcal{O},i}} - m_{\sigma^2_{\mathcal{O},i}})}{2}),$$

where

$$y_{\sigma^2_{\mathcal{O},i}} \equiv \begin{bmatrix}\mathcal{O}_{1,(i)} & \mathcal{O}_{2,(i)} & \cdots & \mathcal{O}_{T,(i)}\end{bmatrix}',$$
$$m_{\sigma^2_{\mathcal{O},i}} \equiv \begin{bmatrix}\mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_1 - c) \\ \mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_2 - c) \\ \vdots \\ \mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_T - c)\end{bmatrix}.$$

Block 6. Mean of Measurement Error Variance $\{\gamma_i\}_{i=1}^{N-d_{\mathbb{Q}}}$ For each $i \in \{1, 2, \cdots, N-d_{\mathbb{Q}}\}$, we update γ_i by

$$p(\gamma_i|\mathcal{I}_T, \theta \backslash \{\gamma_i\}_{i=1}^{N-d_{\mathbb{Q}}}) = \mathcal{G}(3, \frac{1}{\sigma^2_{\mathcal{O},i}} + \bar{\gamma}).$$

3.5 Conditional Forecasting and Scenario Analysis

In this section, we explain how we predict the future bond market and macroeconomy based on specific scenarios. Specifically, we make our conditional scenarios in a form of $\{\mathcal{S}_t, s_t\}_{t=T+1}^{T+h}$, defined implicitly by

$$\underset{(d_{s_t} \times (N+d_{\mathbb{P}}-d_{\mathbb{Q}}))}{\mathcal{S}_t} \begin{bmatrix} R_t \\ M_t \end{bmatrix} = \underset{(d_{s_t} \times 1)}{s_t} \tag{22}$$

for each $t \in [T+1, T+h]$. It says that linear combinations of observables($\mathcal{S}_t \begin{bmatrix} R_t' & M_t' \end{bmatrix}'$) are equal to a specific value in s_t at time t. d_{s_t} can be larger than one, so we can assume multiple linear combinations in one time period. Also, S_t and s_t can be zeros for some t not to make any conditional scenario at time t.

To explain our algorithm for conditional forecasting, we first derive the relationship between observables and $\mathcal{F}_t$. From equations (3) and (6), we derive

$$R_t = \mathcal{A}_X + \mathcal{B}_X \mathcal{T}_{0,\mathcal{P}} + \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} \mathcal{P}_t + e_{R,t}. \tag{23}$$

Therefore, the relationship is

$$\begin{bmatrix} R_t \\ M_t \end{bmatrix} = \begin{bmatrix} \mathcal{A}_X + \mathcal{B}_X \mathcal{T}_{0,\mathcal{P}} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}}) \times 1} \end{bmatrix} + \begin{pmatrix} \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} & O_{N \times (d_{\mathbb{P}}-d_{\mathbb{Q}})} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}}) \times d_{\mathbb{Q}}} & I_{(d_{\mathbb{P}}-d_{\mathbb{Q}})} \end{pmatrix} \mathcal{F}_t + \begin{bmatrix} e_{R,t} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}}) \times 1} \end{bmatrix}. \tag{24}$$

In addition, the relationship between $e_{\mathcal{O},t}$ and $e_{R,t}$ is

$$\begin{bmatrix} O_{d_{\mathbb{Q}} \times 1} \\ e_{\mathcal{O},t} \end{bmatrix} := \begin{pmatrix} W_{\mathcal{P}} \\ W_{\mathcal{O}} \end{pmatrix} e_{R,t} \backsim \mathcal{N}(O_{N \times 1}, \begin{pmatrix} O_{d_{\mathbb{Q}} \times d_{\mathbb{Q}}} & O_{d_{\mathbb{Q}} \times (N-d_{\mathbb{Q}})} \\ O_{(N-d_{\mathbb{Q}}) \times d_{\mathbb{Q}}} & \Sigma_{\mathcal{O}} \end{pmatrix}),$$

so

$$e_{R,t} = \begin{pmatrix} W_{\mathcal{O}} \\ W_{\mathcal{P}} \end{pmatrix}^{-1} \begin{bmatrix} e_{\mathcal{O},t} \\ O_{d_{\mathbb{Q}} \times 1} \end{bmatrix}. \tag{25}$$

When we multiply both sides of equation (24) by $\mathcal{S}_t$,

$$\begin{aligned} s_t =& \mathcal{S}_t \begin{bmatrix} \mathcal{A}_X + \mathcal{B}_X \mathcal{T}_{0,\mathcal{P}} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times 1} \end{bmatrix} + \mathcal{S}_t \begin{pmatrix} \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} & O_{N\times(d_{\mathbb{P}}-d_{\mathbb{Q}})} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times d_{\mathbb{Q}}} & I_{(d_{\mathbb{P}}-d_{\mathbb{Q}})} \end{pmatrix} \mathcal{F}_t \\ &+ \mathcal{S}_{t,(1:d_{s_t},1:N)} \left(\begin{pmatrix} W_{\mathcal{O}} \\ W_{\mathcal{P}} \end{pmatrix}^{-1} \right)_{(1:N,1:N-d_{\mathbb{Q}})} e_{\mathcal{O},t} \end{aligned} \tag{26}$$

holds by equations (22) and (25). Equations (26) and (17) constitute a state-space form model, and we express it in the companion form as

$$\begin{aligned} s_t =& \mathcal{S}_t \begin{bmatrix} \mathcal{A}_X + \mathcal{B}_X \mathcal{T}_{0,\mathcal{P}} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times 1} \end{bmatrix} \\ &+ \mathcal{S}_t \begin{pmatrix} \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} & O_{N\times(d_{\mathbb{P}}-d_{\mathbb{Q}})} & \left(\begin{pmatrix} W_{\mathcal{O}} \\ W_{\mathcal{P}} \end{pmatrix}^{-1} \right)_{(1:N,1:N-d_{\mathbb{Q}})} & O_{N\times(d_{\mathbb{P}}p-d_{\mathbb{P}})} \\ O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times d_{\mathbb{Q}}} & I_{(d_{\mathbb{P}}-d_{\mathbb{Q}})} & O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times(N-d_{\mathbb{Q}})} & O_{(d_{\mathbb{P}}-d_{\mathbb{Q}})\times(d_{\mathbb{P}}p-d_{\mathbb{P}})} \end{pmatrix} \begin{bmatrix} \mathcal{F}_t \\ e_{\mathcal{O},t} \\ \mathcal{F}_{t-1} \\ \mathcal{F}_{t-2} \\ \vdots \\ \mathcal{F}_{t-p+1} \end{bmatrix}, \end{aligned} \tag{27}$$

$$\begin{aligned} \begin{bmatrix} \mathcal{F}_t \\ e_{\mathcal{O},t} \\ \mathcal{F}_{t-1} \\ \mathcal{F}_{t-2} \\ \vdots \\ \mathcal{F}_{t-p+1} \end{bmatrix} =& \begin{pmatrix} G^{\mathbb{P}}_{\mathcal{F}\mathcal{F},1} & O_{d_{\mathbb{P}}\times(N-d_{\mathbb{Q}})} & G^{\mathbb{P}}_{\mathcal{F}\mathcal{F},2} & \cdots & G^{\mathbb{P}}_{\mathcal{F}\mathcal{F},p-1} & G^{\mathbb{P}}_{\mathcal{F}\mathcal{F},p} \\ O_{(N-d_{\mathbb{Q}})\times d_{\mathbb{P}}} & O_{(N-d_{\mathbb{Q}})\times(N-d_{\mathbb{Q}})} & O_{(N-d_{\mathbb{Q}})\times d_{\mathbb{P}}} & \cdots & O_{(N-d_{\mathbb{Q}})\times d_{\mathbb{P}}} & O_{(N-d_{\mathbb{Q}})\times d_{\mathbb{P}}} \\ I_{d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times(N-d_{\mathbb{Q}})} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & \cdots & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \\ O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times(N-d_{\mathbb{Q}})} & I_{d_{\mathbb{P}}} & \cdots & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times(N-d_{\mathbb{Q}})} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & \cdots & I_{d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \end{pmatrix} \begin{bmatrix} \mathcal{F}_{t-1} \\ e_{\mathcal{O},t-1} \\ \mathcal{F}_{t-2} \\ \mathcal{F}_{t-3} \\ \vdots \\ \mathcal{F}_{t-p} \end{bmatrix} \\ &+ \begin{bmatrix} K^{\mathbb{P}}_{\mathcal{F}} \\ O_{(N-d_{\mathbb{Q}})\times 1} \\ O_{d_{\mathbb{P}}\times 1} \\ O_{d_{\mathbb{P}}\times 1} \\ \vdots \\ O_{d_{\mathbb{P}}\times 1} \end{bmatrix} + \begin{bmatrix} \mathcal{N}(O_{d_{\mathbb{P}}\times 1}, \Omega_{\mathcal{F}\mathcal{F}}) \\ \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, \Sigma_{\mathcal{O}}) \\ O_{d_{\mathbb{P}}\times 1} \\ O_{d_{\mathbb{P}}\times 1} \\ \vdots \\ O_{d_{\mathbb{P}}\times 1} \end{bmatrix}. \end{aligned} \tag{28}$$

In this state-space model, s_t is a observable at time t and $\{\{\mathcal{F}_{t-l}\}_{l=0}^{p-1}, e_{\mathcal{O},t}\}$ are state variables at time t. Based on our results in this section, we follow the below steps to make conditional forecasts(Bańbura et al., 2015; Crump et al., 2021).

1. Follow section (3.4) to make a set of posterior samples of parameters.

2. For each posterior sample of parameters,

   (a) Set an initial starting point of the Kalman filter to $\{\{\mathcal{F}_{T-l}\}_{l=0}^{p-1}, e_{\mathcal{O},T}\}$, observed by data. The variance of the initial distribution is zero.

   (b) Do the forward propagation for the state-space model(equations (27) and (28)) with the Kalman filter and obtain the filtered distribution of $\{\{\mathcal{F}_{t-l}\}_{l=0}^{p-1}, e_{\mathcal{O},t}\}_{t=T+1}^{T+h}$. When S_t and s_t are zeros, the Kalman filter only conducts the prediction step, and move on to the next period.

   (c) Do the backward propagation with the Kalman smoother and obtain the smoothed distribution of $\{\{\mathcal{F}_{t-l}\}_{l=0}^{p-1}, e_{\mathcal{O},t}\}_{t=T+1}^{T+h}$.

We use Durbin and Koopman (2002) as the Kalman smoother.

Since there is one smoothed distribution for each posterior sample, we have a set of smoothed distributions. We can derive two versions of conditional forecasts from the set of distributions. The first one is the full Bayesian treatment. It considers both the parameter uncertainty and sampling uncertainty. To this end, we sample one predicted state variable from the smoothed distribution for each posterior sample of parameters. Then, a set of predicted state samples represents a posterior distribution of the conditional forecast. The second version only reflects the parameter uncertainty(Crump et al., 2021). When we do the Kalman smoother, save only the smoothed mean of the state variables. In that case, the number of stored smoothed mean equals the number of posterior samples. We use a set of the stored mean as a distribution of the conditional forecast.

The first method is strict mathematically, but it is difficult to obtain meaningful implications from the prediction because of wide prediction intervals. The second method underestimates the prediction uncertainty, but it is appropriate when the policymaker makes decisions based on the expected path of the future economy. Both methods have pros and cons, and the appropriate method can be chosen depending on the situation. In this study, we opted for the second method.

Lastly, when we make conditional forecasts on future yields or term premiums, we use equations (15) and (23). Specifically, the conditional forecasts for $e_{\mathcal{O},t}$ calculated

using the Kalman smoother are transformed into the conditional forecast for $e_{R,t}$ using Equation (25). Subsequently, by substituting the conditional forecasts for $\mathcal{F}_t$ and $e_{R,t}$ into Equation (23), the conditional forecasts for the yield are computed.

When calculating $\mathbb{E}_T[TP_{\tau,t}|\{\mathcal{S}_j, s_j\}_{j=T+1}^{T+h}]$ using Equation (15), it is necessary to know $\{\mathbb{E}_T[\mathbb{E}_t^{\mathbb{P}}[x_{t+i}]|\{\mathcal{S}_j, s_j\}_{j=T+1}^{T+h}]\}_{i=1-p}^{\tau-2}$, for $x_{t+i} \in \mathcal{F}_{t+i}$. When $i \leq T-t$, observed values are used, while for $T-t < i \leq 0$, the conditional forecast for x_{t+i} is employed. For $i > 0$, the expected values based on the information set comprising both the data and conditional forecasts for $\{\mathcal{F}_k\}_{k=T+1}^{t}$ are used.

4 Empirical Application

Our data has a monthly frequency. The estimation period spans from January 1987 to December 2022. We obtain the U.S. treasury bond yield data from Liu and Wu (2021). We select 18 maturities, so $\tau \in \{1, 3, 6, 9, 12, 18, \cdots, 60, 72, 84, \cdots 120 \text{ months}\}$. Regarding macroeconomic variables for the U.S. economy, we use FRED-MD dataset(McCracken and Ng, 2016) to get 28 variables. We select variables to capture all the information in the FRED-MD dataset while avoiding unnecessary redundancy. The list of macro variables is in Appendix A. For macro variables, our rate variables are in level, while the others are in log level. We transform some rate variables and other variables into stationary time series by applying differencing or log differencing. We use demeaned macro variables in the estimation for stability.

Our affine term structure model(Section 2) assumes stationary, so we have to use parameter values that imply stationary. In our estimation, the probability that the transition equation (17) is stationary is 87.845%. Therefore, the restriction for stationarity is consistent with our data and model. We use only posterior samples that ensure stationarity.

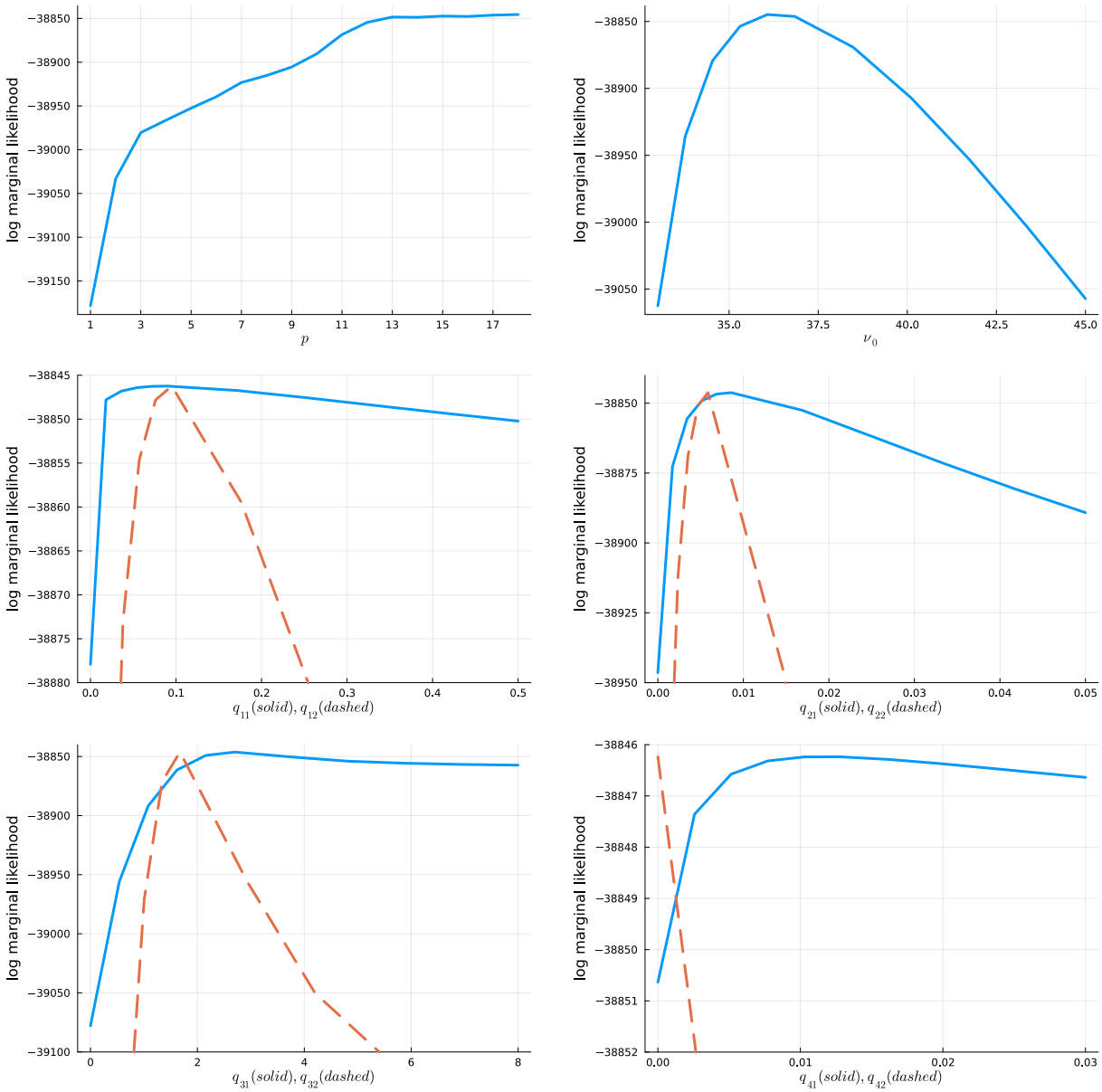
4.1 Optimization on Hyperparameters

The optimization for the hyperparameter is related to choosing a statistically good model that lets the data speak while avoiding overfitting. Our optimization algorithm selects

Table 1: **Optimized Hyperparameters** We maximize the marginal likelihood (19) to optimize the hyperparameters. The optimized values in the table are used to estimate our model.

p	ν_0	q_{11}	q_{12}	q_{21}	q_{22}	q_{31}	q_{32}	q_{41}	q_{42}
17	36.8451	0.0904	0.0951	0.0086	0.0059	2.698	1.675	0.0128	≈ 0

Figure 1: **Marginal Likelihoods around the optimum** The figures depict the behavior of the marginal likelihood as a single hyperparameter parameter value is varied, given the optimized values in Table 1. Our optimized hyperparameter values remain around the optimum.



the model tuned with values in Table 1. The behavior of the criterion(marginal likelihood) around our optimum is displayed in Figure 1. The marginal likelihood values in the figure remain similar even as the lag increases beyond 13. The reason is that coefficients corresponding to lags beyond 13 are heavily penalized, estimated near zero. Consequently, even if those coefficients are precisely constrained to zero through a reduction in the lag, the model effectively remains the same. Therefore, if one desires a parsimonious model, the lags can be constrained to 13. However, since the lag reduction results in a reduced marginal likelihood regardless, this paper maintains 17 lags.

In the Large Bayesian VAR literature, ν_0 is typically set to $d_{\mathbb{P}} + 2$. In our case, this default value is 33. The difference between the default value and our optimal values is not substantial. However, from the perspective of marginal likelihood, our optimal values indicate a superior model. Furthermore, as illustrated in the figures, a reduction in ν_0 from our optimal point leads to a swifter decline in marginal likelihood compared to an increase. Thus, there is a need to set the ν_0 relatively high, implying that some degree of shrinkage is necessary for the non-diagonal elements of the transition equation's covariance matrix.

We penalize parameters differently between the bond market system and the macroeconomy because of the reasons in Section 3.1.2. The differences between $\{q_{11}, q_{21}\}$ and $\{q_{12}, q_{22}\}$ is somewhat subtle, but the behaviors of the marginal likelihood around the optimum are substantially different between them. Our macroeconomic system exhibits a clearly identifiable optimum hyperparameter, whereas the bond system does not. One point is that excessive shrinkage in the bond market system significantly reduces the marginal likelihood. As mentioned in Section 3.1.2, in the bond market system, shrinkage occurs in the direction where the price of risk is zero. Hence, such behavior of the marginal likelihood implies that the price of risk in the bond market has a statistically significant presence. This implication also applied to $\{q_{31}, q_{41}\}$ and $\{q_{32}, q_{42}\}$.

The divergence in the optimal intensity of shrinkage between the macro system and the bond system is evident in the lags and intercepts. That which lags are more heavily penalized in the bond market than the macroeconomy suggests that term structure models have shorter lags. Moreover, as depicted in the figures, an increase in the penalty

Table 2: **Diagnostics for MCMC** We first report the inefficiency factors for six groups of parameters. When the group comprises multiple parameters, the table reports the results based on the parameter with the highest inefficiency factor. To assess convergence, we apply rolling average and standard deviation calculations across the entire set of 20,000 posterior samples, using a window of 100 samples and shifting the window by one sample after each calculation. We then report the smallest and largest values of these calculated means and standard deviations in the third and fourth rows.

		$\kappa^{\mathbb{Q}}$	$k_{\infty}^{\mathbb{Q}}$	$\{\gamma_i\}_{i=1}^{N-d_{\mathbb{Q}}}$	$\Sigma_{\mathcal{O}}$	$\{\sigma^2_{\mathcal{FF},i}\}_{i=1}^{N-d_{\mathbb{Q}}}$	$\{\phi\}_{i=1}^{N-d_{\mathbb{Q}}}$
inefficiency factor		1.196	2.3806	≤ 1.0574	≤ 1.4668	≤ 1.9363	≤ 1.7332
mean	minimum	0.0462	0.0137	5.7417×10^{-5}	0.0067	0.8158	-0.1489
	maximum	0.047	0.0144	9.0399×10^{-5}	0.0071	0.8708	-0.1346
standard deviation	minimum	0.0008	0.0006	2.9313×10^{-5}	0.0004	0.0349	0.0109
	maximum	0.0014	0.001	6.3528×10^{-5}	0.0007	0.079	0.0203

for lags does not significantly reduce the marginal likelihood. Hence, there is a clear necessity to differentiate the lags between the macro and bond systems. Lastly, $q_{41} \neq 0$ indicates that we cannot constrain the intercept term in the prices of risk equation (13) to zero.

4.2 Diagnostics for the Posterior Sampler

We need to design the sampling algorithm delicately so that our finite posterior samples represent the posterior distribution well. This section checks the efficiency of our posterior sampling algorithm to ascertain whether our posterior samples have such representativeness. To this end, we use a rolling window of 100 observations to calculate sample means and standard deviations sequentially, starting with the first 100 observations and moving one observation forward at each step. Consequently, this generates a continuous series of sample means and standard deviations. The sample means and standard deviations should display stability throughout the sequence if our algorithm is efficient.

Table 2 presents the maximum and minimum values of the means and standard deviations throughout the series, along with the inefficiency factors.[5] The results dis-

[5]The calculations for the table include posterior samples that imply nonstationarity. When only using posterior samples that satisfy the stationarity condition, the results are almost similar, or our algorithm appears more efficient.

Table 3: **Bond pricing errors** We derived the pricing error(or measurement error) by calculating the difference between the observed yield and the fitted yield. The sample mean and standard deviation of that pricing error are presented in the table.

maturity τ	1	3	6	9	12	18	24	30	36
mean	-0.038	-0.013	0.017	0.031	0.034	0.019	-0.008	-0.021	-0.021
SD	0.101	0.04	0.042	0.073	0.082	0.061	0.036	0.029	0.036
maturity τ	42	48	54	60	72	84	96	108	120
mean	-0.012	-0.008	-0.007	-0.007	0.006	0.008	0.008	0.004	0.008
SD	0.045	0.053	0.053	0.052	0.05	0.032	0.03	0.046	0.072

played in the table suggest that our algorithm can generate a representative sample of the posterior distribution, even with a small number of samples. Considering that the maximum inefficiency factor is 2.38, it indicates that our algorithm approximates one *iid* sample for every two or three MCMC samples.

We end this section by mentioning some results related to the algorithm. We use the MH algorithm(section 3.4) to sampling $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_\mathbb{Q}}$, related to $\Omega_{\mathcal{PP}}$. To make a proposal sample, we update these parameters only using the transition equation, and then we accept the proposal stochastically. The acceptance rates in this MH algorithm is 83.056%, 83.064%, and 79.608% for $i = 1, 2,$ and $3(= d_\mathbb{Q})$, respectively. The high acceptance rates with our approximations on the target distributions indicate that the transition equation has most information in infer $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_\mathbb{Q}}$. This result justifies the methodology of Adrian, Crump, and Moench (2013) partially. However, considering that a proposal is rejected with a probability of about 20%, it suggests that the measurement equation also contains distinctive information about those parameters.

4.3 Parameter estimates

We start our evaluation of the estimation results by looking at the measurement error (or pricing error) in the measurement equation to show that the yields are well-fitted. According to Table 3, these errors average near zero, and taking the standard deviation into account, they consistently hover around zero. Since our term structure model is statistically similar to the DNS model, it is natural.

Table 4: **Posterior Distribution of the Decay Parameter** $\kappa^{\mathbb{Q}}$ The table displays the posterior distribution of $\kappa^{\mathbb{Q}}$ and indicates the maturity at which the factor loading of the curvature is maximized when calculated based on the value of $\kappa^{\mathbb{Q}}$.

$\kappa^{\mathbb{Q}}$	0.0427	0.0437	0.0448	0.046	0.0472	0.0485	0.0498
maturity τ	42	41	40	39	38	37	36
probability(%)	0.03	1.1	11.25	37.87	38.48	10.68	0.57

Table 4 shows the posterior distribution of the representative medium-term bond's maturity. It is shown that the bond has a maturity falling between 3 and 3.5 years, with 38 months identified as the posterior mode. In addition, $k_{\infty}^{\mathbb{Q}}$ has a posterior mean and standard deviation of 0.014 and 0.0007, respectively.

The remaining parameters we are interested in are the things in the transition equation. In Figure 2, the correlation matrix, corresponding to $\Omega_{\mathcal{FF}}$, reveals clustering among financial, real, and nominal variables while also indicating interactions between these clusters. We are interested in the contemporaneous interactions between the principal component clusters and other variables. Interestingly, nearly all variables exhibiting statistically significant interactions with the principal component clusters are nominal or financial variables. Consequently, the new information to which the yield curve reacts immediately primarily originates from nominal or financial variables. Conversely, this implies that the new information stemming from real variables is predominantly unspanned information.

Lastly, while it is possible to evaluate the statistical significance of each entry in $G_{\mathcal{FF},i}^{\mathbb{P}}$ directly, this approach is not suitable as it only assesses the predictive power of individual lag terms for a specific variable. An individual lag term may not have a statistically significant coefficient on its own, but all the lag terms of a particular variable combined may have jointly statistically significant predictive power for a variable. Therefore, we assess the statistical significance of the transition equation's coefficient matrix($\{G_{\mathcal{FF},i}^{\mathbb{P}}\}_{i=1}^{p}$) based on its contribution to the term premium in the next section(Section 4.4).

Figure 2: **Error Correlation Matrix in the Transition Equation** (17) The heat map displays the posterior mean of the correlation matrix derived from $\Omega_{\mathcal{FF}}$. Entries not statistically significant based on the 95% posterior interval are marked in white(= 0). "PC1", "PC2", and "PC3" represent the first three principal components. For the descriptions of the remaining macroeconomic variable names, refer to the Appendix A.

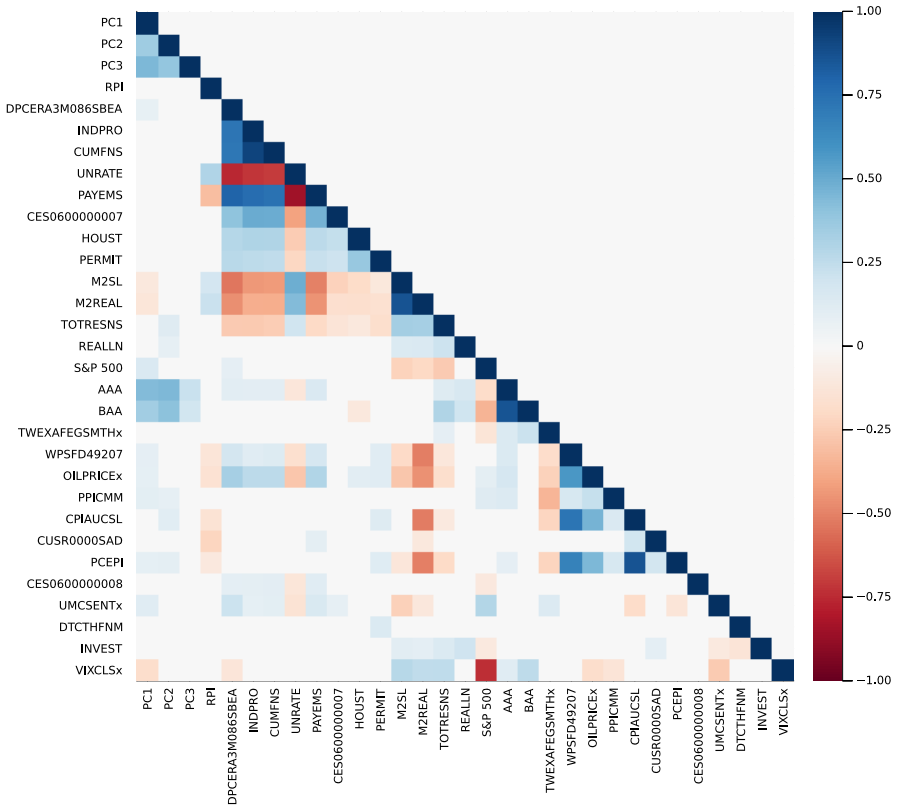
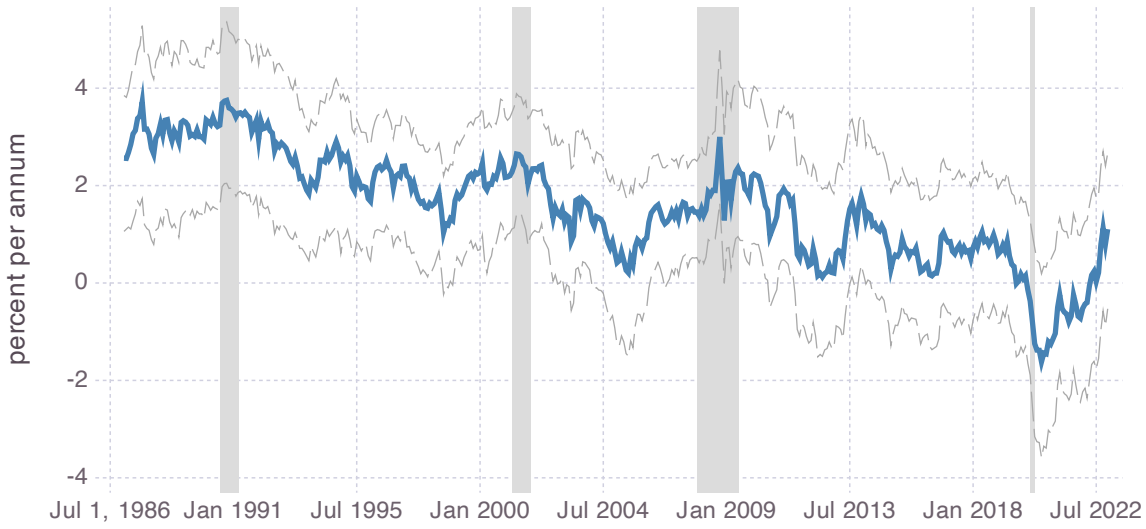
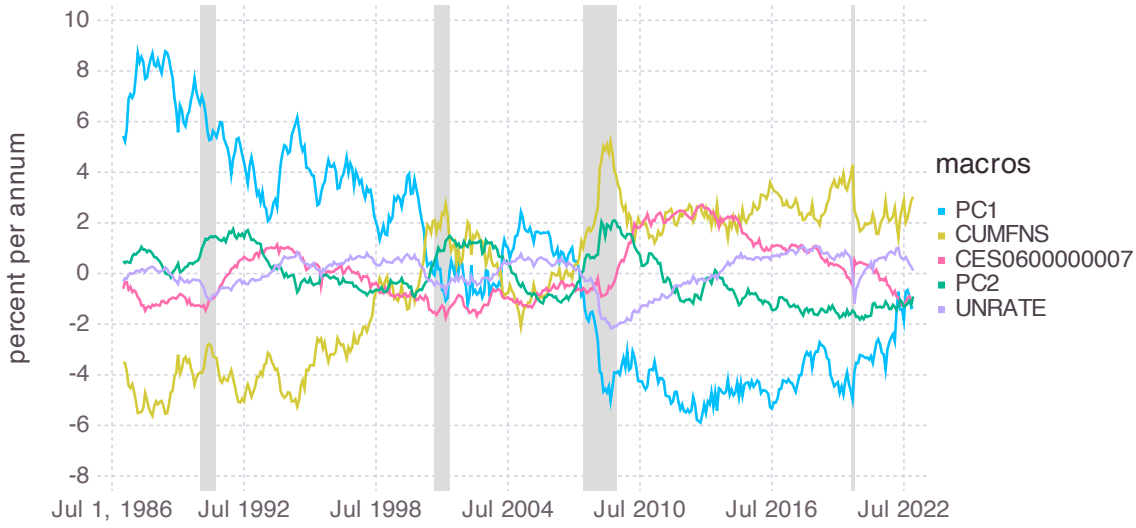


Figure 3: **Ten-Year Term Premium Estimates** The estimated term premium for the 10-year yield is presented in Subfigure (a). The posterior mean and the 95% posterior interval are represented by solid and dashed lines, respectively. Subfigure (b) displays the movements in the term premium according to the expected paths of each factor(Section 2.3), focusing on the top five variations with the largest variance. The corresponding legend is ordered by the size of the variance. In the legend, "PC1" and "PC2" are the first two principal components of bond yields, "CUMFNS" is the capacity utilization: manufacturing, "CES0600000007" is average weekly hours : goods-producing, and "UNRATE" is the civilian unemployment rate.

(a) 10-year Term Premium



(b) Decomposition of the Term Premium



4.4 Term Premium

As explained in Section 2.3, the expected paths of the 31(= $d_{\mathbb{P}}$) variables contribute to the fluctuations in the term premium, resulting in the countercyclical term premium as shown in Figure 3-(a). When we decompose the contributions of each variable's expected path to the term premium, Figure 3-(b) emerges. If we sum up all the variable contributions and then add the time-invariant component, the term premium displayed in Figure 3-(a) would be obtained. We plot only those contributions with a standard deviation greater than 0.5 in Figure 3-(b).

Obviously, the most important factor for the term premium is the first principal component(PC1). The yield curve has exhibited a downward trend for an extended period, and PC1 considers part of this downtrend as the trend in the term premium. However, as Bauer and Rudebusch (2020) points out, traditional term structure models tend to overstate the variations of the term premium. In our model, the capacity utilization: manufacturing(CUMFNS) transforms the portion that PC1 attributes to fluctuations and trends in the term premium into those of the expectation hypothesis component. The second principal component(PC2) contributes to the variation in the term premium as the fourth-largest factor and increases the term premium in the event of a recession.

The average weekly hours:goods-producing(CES0600000007) contributes to an increase in the term premium after the end of a recession, as a decrease in risk aversion occurs with the onset of an economic expansion, leading to a reduced preference for safe assets. The information indicating an impending economic expansion phase is reflected in the term premium through the projected path of CES0600000007. Additionally, the expected trajectory of the civilian unemployment rate(UNRATE) appears to reduce the term premium around recession periods. This occurs as flight-to-safety emerges during these times, increasing the demand for U.S. bonds.

Although only five variables create quantitatively significant fluctuations in the term premium, the remaining variables cannot be deemed irrelevant in the term structure model as long as we adhere to the FIRE condition. Should these additional variables be excluded from the transition equation, the expected paths of the five key variables

would lack a sufficient information base. Hence, to ensure the term premium is derived from accurately projected paths, the inclusion of the remaining variables is essential.

We conclude this section with an analysis comparing Figures 2 and 3. In Figure 2, we observe that the new information spanned by the contemporaneous yield curve primarily relates to nominal and financial variables. Conversely, in Figure 3, the variables contributing significant fluctuations to the term premium are predominantly real variables. Hence, it can be inferred that unspanned macro risk, which is unspanned by the contemporaneous yield curve yet possesses predictive power over the future yield curve, predominantly relates to information concerning real activity. The information contained in financial and nominal variables related to the bond market is reflected swiftly in the contemporaneous yield curve as spanned information.

5 Scenario Analysis

This section demonstrates our methodology for scenario analysis(Section 3.5). Specifically, we estimate the posterior distribution of

$$\mathbb{E}_t[obs_{t+h}|\text{scenario}, \theta] - \mathbb{E}_t[obs_{t+h}|\theta],$$

where obs_{t+h} is a variable of interest at time $t+h$, and h is a forecasting horizon. It is similar to the generalized impulse response function of Koop, Pesaran, and Potter (1996) and Pesaran and Shin (1998). However, the scenarios are more flexible in our method as they are conditional upon arbitrary linear combinations of variables assuming certain values at arbitrary future points. Furthermore, we allow the linear combinations of variables to vary over time.

We report the results of our scenario analysis using the difference between conditional and unconditional forecasts. Conditional forecasts are formed using both past and future information in the scenario, while unconditional forecasts are based solely on past information. Therefore, by examining the difference between these two types of forecasts, we can see how the new economic shocks(including policy shocks) in the scenario alters the predicted economic trajectory. However, in practical applications where the goal is

to accurately predict the level of economic variables through conditional forecasting, it may be preferable to use only conditional forecasts.

One important note is that conditional scenarios do not represent effects from the exogenous shocks. Rather, conditional forecasts illustrate the path of variables likely to occur alongside the scenario, assuming the scenario is realized in the future. In this sense, conditional forecasts can be seen as equilibrium paths or results from a bidirectional interaction between the scenario and the variables.[6]

In this section, we derive conditional forecasts based on the below two scenarios.

- Scenario 1: the one-month yield is 5.1% in December 2023, 4.1% in December 2024, and 3.1% in December 2025.
- Scenario 2: From January 2023 to October 2023, the VIX trajectory mirrors that observed from September 2008 to June 2009.

The point of making conditional forecasts is December 2022, and January 2023 corresponds to a forecasting horizon of one. The first scenario is based on the FOMC Projections released in December 2022. The path we assumed is the median of the projected appropriate Fed rate path. Through this scenario analysis, we can deduce the expectations for future shocks that the FOMC participants incorporated into their reported projections. The second scenario examines how the macroeconomy, financial markets, and monetary policy would react in the event of financial uncertainty similar to the Global Financial Crisis.

5.1 Scenario 1: Information in the FOMC's Fed Rate Projections

The new economic shocks deduced from the FOMC projection modify the path of future economic development as Figures 4 and 5. In macroeconomy(Figure 4), inflation pressure is expected to occur over the next one to two years. In particular, there are pronounced cost shocks, such as in metals and oil prices. The contractionary monetary policy(Figure

[6]It is also possible to replicate exogenous structural shocks by fixing the paths of variables in a specific way (see Crump et al., 2021).

Figure 4: **Information in the FOMC's Fed Rate Projections: Prediction for the Macroeconomy** The figures display predictions for macroeconomic variables conditional on the Federal Open Market Committee's projection for the federal funds rate, as released in December 2022. The scenario is as follows: the federal funds rate is projected to be 5.1% in December 2023, 4.1% in December 2024, and 3.1% in December 2025. The X-axis represents the forecast horizon, with each point indicating the prediction for a specific month after December 2022. The Y-axis displays "conditional forecast - unconditional forecast". The blue shaded areas represent the 68% and 95% posterior intervals. "INDPRO" is the industrial production index, "PERMIT" is new private housing permits, "M2SL" is the M2 money stock,"OILPRICEx" is to the WTI oil price, "PPICMM" is the producer price index: metals and metal products, "PCEPI" is the personal consumption expenditures price index, "DTCTHFNM" is the total consumer loans and leases outstanding, and "VIXCLSx" is the CBOE volatility index. UNRATE is in level, and the other variables are presented as YoY growth rates.

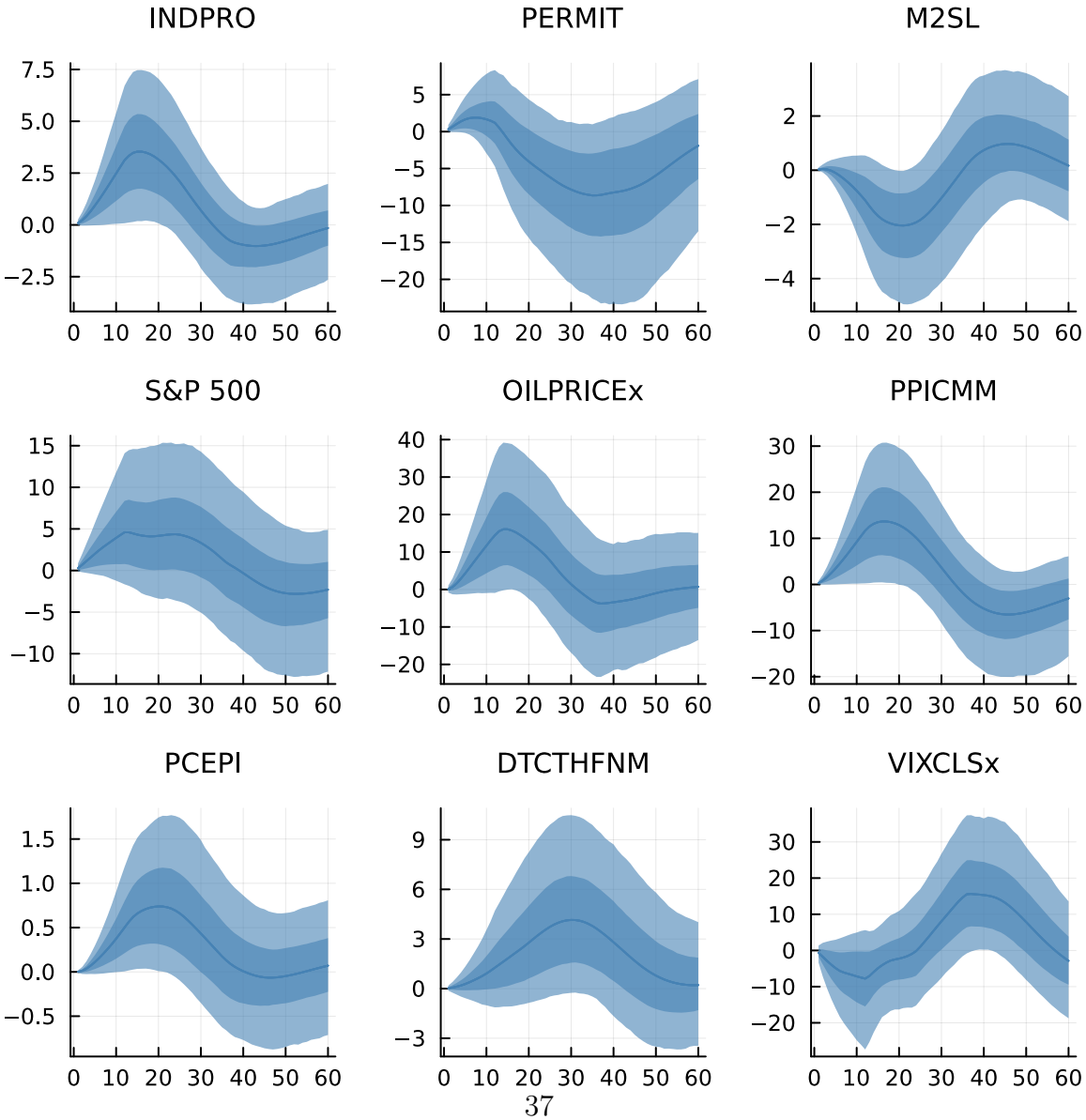
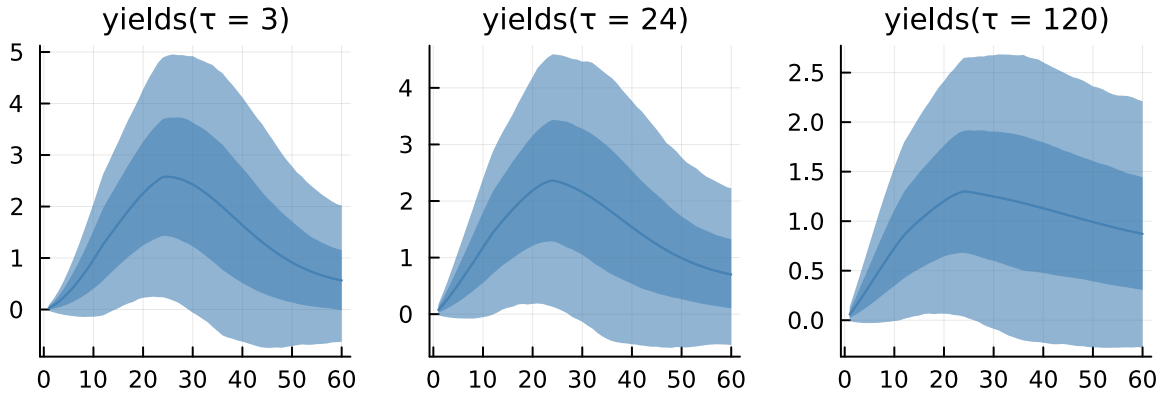
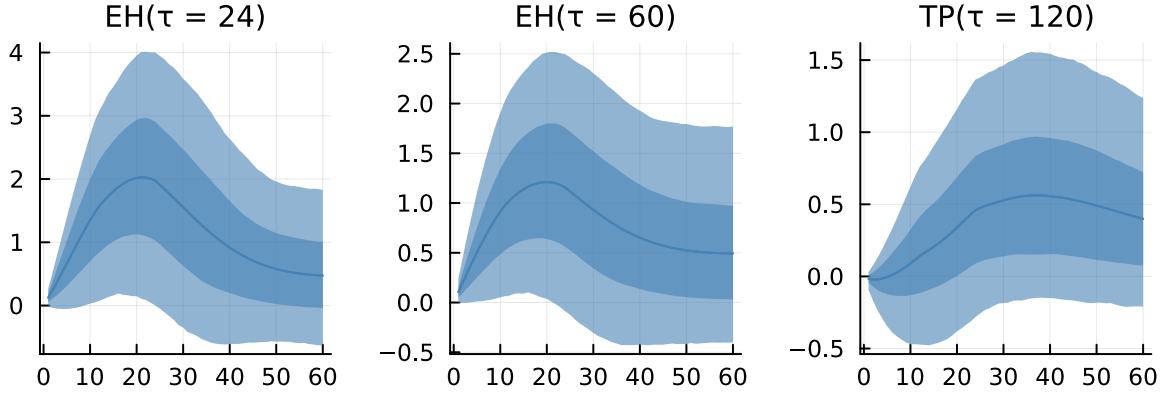


Figure 5: **Information in the FOMC's Fed Rate Projections: Prediction for the Yield Curve** The figures display predictions conditional on the Federal Open Market Committee's projection for the federal funds rate, as released in December 2022. The scenario is as follows: the one-month yield is projected to be 5.1% in December 2023, 4.1% in December 2024, and 3.1% in December 2025. The X-axis represents the forecast horizon, with each point indicating the prediction for a specific month after December 2022. The Y-axis displays "conditional forecast - unconditional forecast". The blue shaded areas represent the 68% and 95% posterior intervals. In Subfigure (a), the bond yields are presented as monthly yields in percent per annum. Subfigure (b) depicts the expectation hypothesis components for two-year and five-year bonds, as well as the term premium for the ten-year bond yield.

(a) Bond Yields



(b) Expectation Hypothesis Components(EH) or Term Premium(TP)



5-(a)) reduces the M2 monetary base to stabilize inflation in this period, and stable growth is also observed in both financial markets and the real economy. While prolonged high-interest rates are expected to stabilize inflation, they may also lead to an increase in household debt and are anticipated to result in financial uncertainty and stock market losses in the distant future. Additionally, the contractionary policy negatively impacts the construction industry.

Due to the tightening monetary policy, the overall yield curve shifts upwards in Figure 5-(a), but the rise is more pronounced in short-term and medium-term yields. This is because the factors driving the yield curve's movement are primarily attributed to the contractionary medium-term monetary policy. This is reflected in the rise of the expectation hypothesis component of medium-term bonds in Figure 5-(b). However, as time progresses and financial uncertainties increase, it has been observed that the term premium subsequently rises.

5.2 Scenario 2: Financial Crisis Starting with January 2023

Figure 6 illustrates the characteristics of the crisis as we have assumed in our second scenario. Despite the financial uncertainty(VIX) following the same trajectory as during the Global Financial Crisis, the deterioration in the real economy is not as severe. Furthermore, post-2023, a rapid recovery in the real activity is evident, attributed to the expectation of a swift yet prolonged accommodative monetary policy. As seen in Figure 7-(a), the yield curve across all maturities shifts horizontally downwards. Additionally, the decrease in short-term interest rates persists for an extended period.

Market participants also anticipate that medium-term monetary policy will remain accommodative for a considerable time(Figure 7-(b)). Also, the accommodative supply of liquidity leads to an increase in M2 and a decrease in the term premium. As a secondary effect, the inflation level based on PCEPI is expected to stabilize, dropping below 1.5% after 2023 and eventually converging to a range of 2% to 3% in the long run.

Figure 6: **Financial Crisis Starting with January 2023: Prediction for the Macroeconomy** The figures display predictions for macroeconomic variables conditional on the scenario in which the trajectory of the VIX from September 2008 to June 2009 is replicated in the period from January 2023 to October 2023. The X-axis represents the forecast horizon, with each point indicating the prediction for a specific month after December 2022. The Y-axis displays "conditional forecast - unconditional forecast". The blue shaded areas represent the 68% and 95% posterior intervals. "INDPRO" is the industrial production index, "CUMFNS" is the capacity utilization: manufacturing, "UNRATE" is the civilian unemployment rate, "PERMIT" is new private housing permits, "M2SL" is the M2 money stock, "PCEPI" is the personal consumption expenditures price index, "UMCSENTx" is the consumer sentiment index, and "VIXCLSx" is the CBOE volatility index. UNRATE and CUMFNS are in level, and the other variables are presented as YoY growth rates.

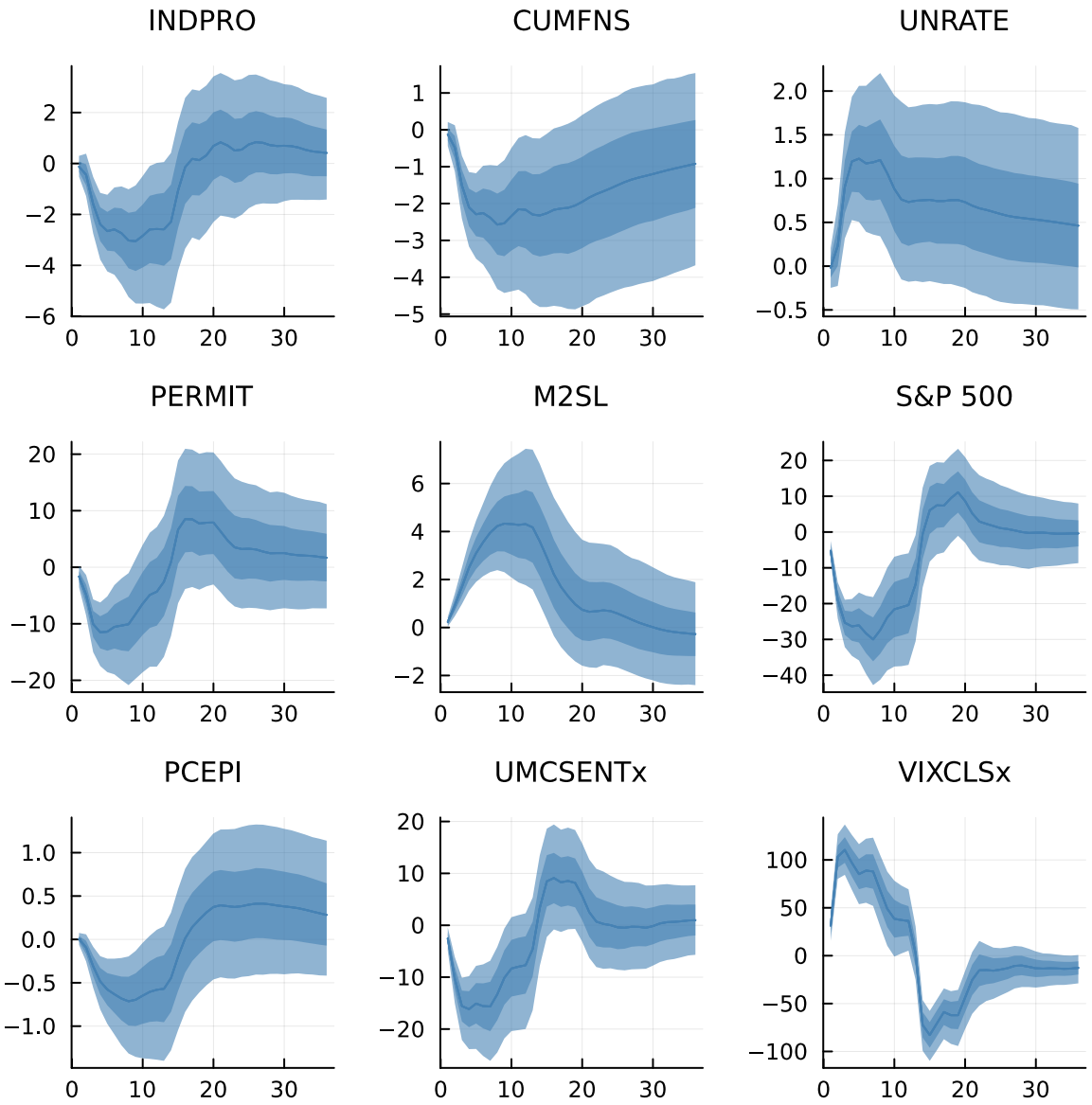
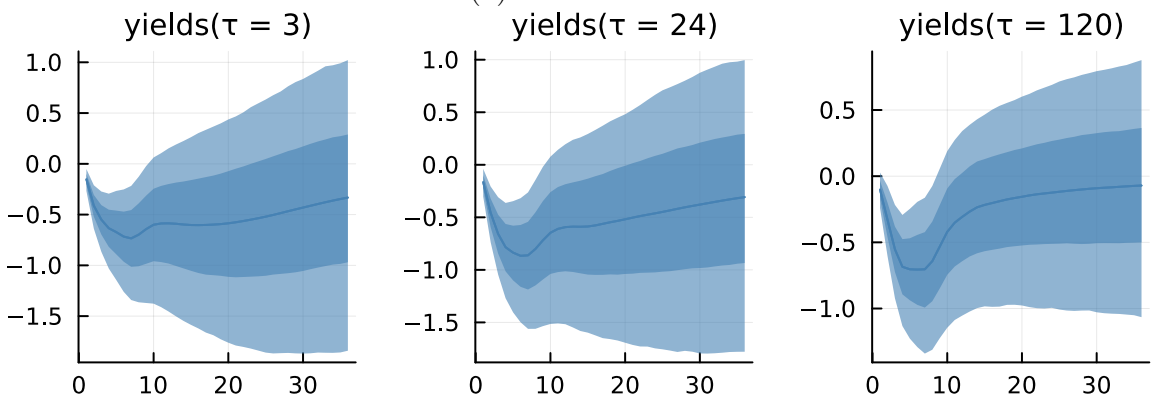
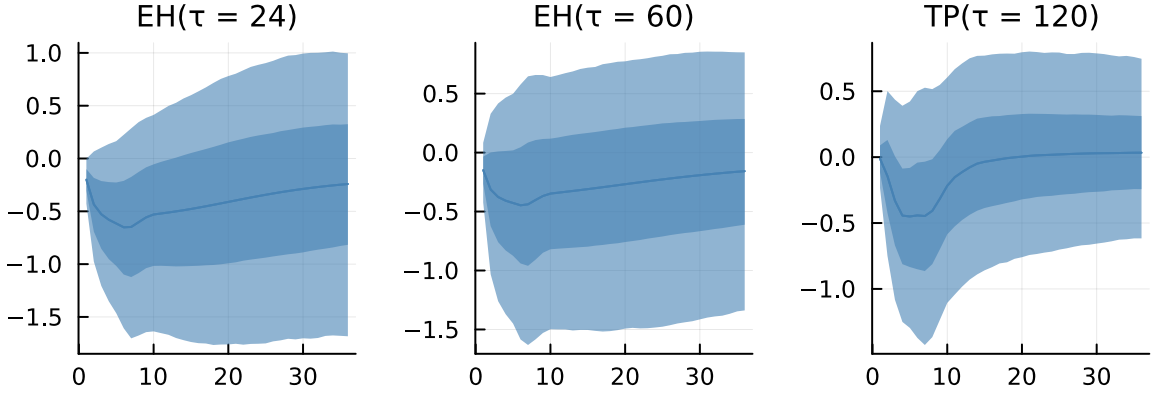


Figure 7: **Financial Crisis Starting with January 2023: Prediction for the Yield Curve** The figures display predictions conditional on the scenario in which the trajectory of the VIX from September 2008 to June 2009 is replicated in the period from January 2023 to October 2023. The X-axis represents the forecast horizon, with each point indicating the prediction for a specific month after December 2022. The Y-axis displays "conditional forecast - unconditional forecast". The blue shaded areas represent the 68% and 95% posterior intervals. In Subfigure (a), the bond yields are presented as monthly yields in percent per annum. Subfigure (b) depicts the expectation hypothesis components for two-year and five-year bonds, as well as the term premium for the ten-year bond yield.

(a) Bond Yields



(b) Expectation Hypothesis Components or Term Premium



6 Conclusion

Our paper has developed a methodology for efficiently estimating the term structure of interest rates. In this process, most hyperparameters are automatically determined by the data, reducing the need for fine-tuning the model. Our algorithm's ability to efficiently estimate the model can be attributed to four main reasons.

Firstly, by introducing the asymmetric conjugate priors of Chan (2022) into the transition equation, we estimate the transition equation on an equation-by-equation basis. As a result, equations requiring the MH algorithm are distinctly separated from those that do not. We also introduce an efficient proposal distribution for the MH algorithm, combined with the prior distribution and the likelihood of the transition equation. Additionally, the prior distribution of Chan (2022) allow for the data-based hyperparameter tuning with less burden due to the absence of cross-validation. It simply involves maximizing the marginal likelihood of the transition equation, which possesses a closed-form expression, by fine-tuning the hyperparameters.

Secondly, the Minnesota-type prior distribution enables stable estimation even with numerous macro variables and lag terms by imposing a Ridge-type penalty on coefficients of the transition equation. Third, by incorporating AFNS constraints into the model, we express our belief about the representative medium-term maturity through the prior distribution of the $\mathbb{Q}$ slope matrix, $G^{\mathbb{Q}}_{\mathcal{PP}}$. This justified the use of discrete prior distributions for $G^{\mathbb{Q}}_{\mathcal{PP}}$, allowing us to estimate these parameters without complex sampling algorithms. Finally, we solve the difference equations for bond pricing to render the measurement equation linear to the $\mathbb{Q}$ intercept term, $k^{\mathbb{Q}}_{\infty}$. This enables the estimation of the intercept term through efficient Normal-Normal updates. These four lessons can be applied when estimating other forms of term structure models.

References

Adrian, T., Crump, R. K., and Moench, E. (2013), "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 110, 110–138, publisher: Elsevier.

Bańbura, M., Giannone, D., and Lenza, M. (2015), "Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections," *International Journal of Forecasting*, 31, 739–756.

Bauer, M. D. and Rudebusch, G. D. (2020), "Interest Rates Under Falling Stars," *American Economic Review*, 110, 01–47, publisher: American Economic Association.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017), "Julia: A fresh approach to numerical computing," *SIAM Review*, 59, 65–98, publisher: SIAM.

Carriero, A., Clark, T. E., and Marcellino, M. (2019), "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154, publisher: Elsevier.

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

— (2022), "Asymmetric Conjugate Priors for Large Bayesian VARs," *Quantitative Economics*.

Cheridito, P., Filipovic, D., and Kimmel, R. L. (2007), "Market price of risk specifications for affine models: Theory and evidence," *Journal of Financial Economics*, 83, 123–170.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2011), "The affine arbitrage-free class of Nelson – Siegel term structure models," *Journal of Econometrics*, 164, 4–20, publisher: Elsevier B.V.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., and Sokol, A. (2022), "Nowcasting with large Bayesian vector autoregressions," *Journal of Econometrics*, 231, 500–519.

Cochrane, J. H. and Piazzesi, M. (2008), "Decomposing the Yield Curve," Tech. rep., publication Title: SSRN.

Crump, R. K., Eusepi, S., Giannone, D., Qian, E., and Sbordone, A. M. (2021), "A Large Bayesian VAR of the United States Economy," .

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

Durbin, J. and Koopman, S. J. (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89, 603–615, publisher: [Oxford University Press, Biometrika Trust].

Feldt, R. (2018), "BlackBoxOptim.jl," Commit: b8bb17b.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015), "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

Joslin, S. and Le, A. (2021), "Interest rate volatility and no-arbitrage affine term structure models," *Management Science*, 67, 7391–7416, publisher: INFORMS.

Joslin, S., Le, A., and Singleton, K. J. (2013a), "Gaussian macro-finance term structure models with lags," *Journal of Financial Econometrics*, 11, 581–609.

— (2013b), "Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs," *Journal of Financial Economics*, 109, 604–622.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014), "Risk premiums in dynamic term structure models with unspanned macro risks," *Journal of Finance*, 69, 1197–1233.

Joslin, S., Singleton, K. J., and Zhu, H. (2011), "A new perspective on Gaussian dynamic term structure models," *The Review of Financial Studies*, 24, 926–970, publisher: Oxford University Press.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996), "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74, 119–147.

Liu, Y. and Wu, J. C. (2021), "Reconstructing the yield curve," *Journal of Financial Economics*, publisher: North-Holland.

McCracken, M. W. and Ng, S. (2016), "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business and Economic Statistics*, 34, 574–589.

Niu, L. and Zeng, G. (2012), "The Discrete-Time Framework of the Arbitrage-Free Nelson-Siegel Class of Term Structure Models," Tech. rep., publication Title: Ssrn.

Opara, K. R. and Arabas, J. (2019), "Differential Evolution: A survey of theoretical analyses," *Swarm and Evolutionary Computation*, 44, 546–558.

Pesaran, H. and Shin, Y. (1998), "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, 58, 17–29.

Spector, L. and Klein, J. (2006), "Trivial Geography in Genetic Programming," in *Genetic Programming Theory and Practice III*, eds. Yu, T., Riolo, R., and Worzel, B., Boston: Kluwer Academic Publishers, vol. 9, pp. 109–123, series Title: Genetic Programming.

Wang, Y., Li, H.-X., Huang, T., and Li, L. (2014), "Differential evolution based on covariance matrix learning and bimodal distribution parameter setting," *Applied Soft Computing*, 18, 232–247.

A Appendices: Macroeconomic Variables

name	mnemonic	transformation
Real Personal Income	RPI	1200Δlog
Real personal consumption expenditures	DPCERA3M086SBEA	1200Δlog
IP Index	INDPRO	1200Δlog
Capacity Utilization: Manufacturing	CUMFNS	level
Civilian Unemployment Rate	UNRATE	level
All Employees: Total nonfarm	PAYEMS	1200Δlog
Avg Weekly Hours : Goods-Producing	CES0600000007	100 log
Housing Starts: Total New Privately Owned	HOUST	1200Δlog
New Private Housing Permits (SAAR)	PERMIT	1200Δlog
M2 Money Stock	M2SL	1200Δlog
Real M2 Money Stock	M2REAL	1200Δlog
Total Reserves of Depository Institutions	TOTRESNS	1200Δlog
Real Estate Loans at All Commercial Banks	REALLN	1200Δlog
S&P's Common Stock Price Index: Composite	S&P 500	1200Δlog
Moody's Seasoned Aaa Corporate Bond Yield	AAA	Δlevel
Moody's Seasoned Baa Corporate Bond Yield	BAA	Δlevel
Trade Weighted U.S. Dollar Index	TWEXAFEGSMTHx	1200Δlog
PPI: Finished Goods	WPSFD49207	1200Δlog
Crude Oil, spliced WTI and Cushing	OILPRICEx	1200Δlog
PPI: Metals and metal products:	PPICMM	1200Δlog
CPI : All Items	CPIAUCSL	1200Δlog
CPI : Durables	CUSR0000SAD	1200Δlog
Personal Cons. Expend.: Chain Index	PCEPI	1200Δlog
Avg Hourly Earnings : Goods-Producing	CES0600000008	1200Δlog
Consumer Sentiment Index	UMCSENTx	1200Δlog
Total Consumer Loans and Leases Outstanding	DTCTHFNM	1200Δlog
Securities in Bank Credit at All Commercial Banks	INVEST	1200Δlog
CBOE Volatility Index: VIX	VIXCLSx	100 log

Online Appendices: Estimating a Large Vector Autoregression of Yield Curve and Macroeconomic Variables with No-Arbitrage Restriction

A Proof of Proposition 1

Our restricted JSZ form is

$$
\begin{aligned}
r_t &= \iota' X_t, \\
X_t &= \underbrace{\begin{bmatrix} k_\infty^{\mathbb{Q}} \\ 0 \\ 0 \end{bmatrix}}_{K_X^{\mathbb{Q}}} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \exp[-\kappa^{\mathbb{Q}}] & 1 \\ 0 & 0 & \exp[-\kappa^{\mathbb{Q}}] \end{pmatrix}}_{G_{XX}^{\mathbb{Q}}} X_{t-1} + \epsilon_{X,t}^{\mathbb{Q}}, \\
X_t &= K_X^{\mathbb{P}} + \sum_{l=1}^{p} G_{XX,l}^{\mathbb{P}} X_{t-l} + \sum_{l=1}^{p} G_{XM,l}^{\mathbb{P}} M_{t-l} + \epsilon_{X,t}^{\mathbb{P}}, \\
M_t &= K_M^{\mathbb{P}} + \sum_{l=1}^{p} G_{MX,l}^{\mathbb{P}} X_{t-l} + \sum_{l=1}^{p} G_{MM,l}^{\mathbb{P}} M_{t-l} + \epsilon_{M,t}^{\mathbb{P}}.
\end{aligned}
$$

Rotating X_t using

$$
X_t^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\kappa^{\mathbb{Q}}}{1-\exp[-\kappa^{\mathbb{Q}}]} & \frac{1+2\kappa^{\mathbb{Q}}-\exp[\kappa^{\mathbb{Q}}]-\kappa^{\mathbb{Q}}\exp[-\kappa^{\mathbb{Q}}]}{(1-\exp[-\kappa^{\mathbb{Q}}])^2} \\ 0 & 0 & \frac{\exp[\kappa^{\mathbb{Q}}]}{1-\exp[-\kappa^{\mathbb{Q}}]} \end{pmatrix} X_t
$$

yields

$$
\begin{aligned}
r_t &= \underbrace{\begin{bmatrix} 1 & \frac{1-\exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} & \frac{1-\exp(-\kappa^{\mathbb{Q}})}{\kappa^{\mathbb{Q}}} - \exp(-\kappa^{\mathbb{Q}}) \end{bmatrix}}_{\beta^{*\prime}} X_t^*, \\
X_t^* &= K_X^{\mathbb{Q}} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \exp[-\kappa^{\mathbb{Q}}] & \kappa^{\mathbb{Q}}\exp[-\kappa^{\mathbb{Q}}] \\ 0 & 0 & \exp[-\kappa^{\mathbb{Q}}] \end{pmatrix}}_{G_{X^*X^*}^{\mathbb{Q}}} X_{t-1}^* + \epsilon_{X^*,t}^{\mathbb{Q}},
\end{aligned}
$$

In a companion form, the $\mathbb{P}$-dynamics is

$$\begin{bmatrix} \begin{bmatrix} X_t^* \\ M_t \end{bmatrix} \\ \begin{bmatrix} X_{t-1}^* \\ M_{t-1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} X_{t-p+1}^* \\ M_{t-p+1} \end{bmatrix} \end{bmatrix} = \underbrace{\begin{bmatrix} \begin{bmatrix} K_{X^*}^{\mathbb{P}} \\ K_M^{\mathbb{P}} \end{bmatrix} \\ O_{d_{\mathbb{P}}(p-1)\times 1} \end{bmatrix}}_{K^{\mathbb{P}}} + G^{\mathbb{P}} \begin{bmatrix} \begin{bmatrix} X_{t-1}^* \\ M_{t-1} \end{bmatrix} \\ \begin{bmatrix} X_{t-2}^* \\ M_{t-2} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} X_{t-p}^* \\ M_{t-p} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} \epsilon_{X^*,t}^{\mathbb{P}} \\ \epsilon_{M,t}^{\mathbb{P}} \end{bmatrix} \\ O_{d_{\mathbb{P}}(p-1)\times 1} \end{bmatrix},$$

where

$$G^{\mathbb{P}} \equiv \begin{pmatrix} \begin{pmatrix} G_{X^*X^*,1}^{\mathbb{P}} & G_{X^*M,1}^{\mathbb{P}} \\ G_{MX^*,1}^{\mathbb{P}} & G_{MM,1}^{\mathbb{P}} \end{pmatrix} & \begin{pmatrix} G_{X^*X^*,2}^{\mathbb{P}} & G_{X^*M,2}^{\mathbb{P}} \\ G_{MX^*,2}^{\mathbb{P}} & G_{MM,2}^{\mathbb{P}} \end{pmatrix} & \cdots & \begin{pmatrix} G_{X^*X^*,p-1}^{\mathbb{P}} & G_{X^*M,p-1}^{\mathbb{P}} \\ G_{MX^*,p-1}^{\mathbb{P}} & G_{MM,p-1}^{\mathbb{P}} \end{pmatrix} & \begin{pmatrix} G_{X^*X^*,p}^{\mathbb{P}} & G_{X^*M,p}^{\mathbb{P}} \\ G_{MX^*,p}^{\mathbb{P}} & G_{MM,p}^{\mathbb{P}} \end{pmatrix} \\ I_{d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & \cdots & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \\ O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & I_{d_{\mathbb{P}}} & \cdots & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} & \cdots & I_{d_{\mathbb{P}}} & O_{d_{\mathbb{P}}\times d_{\mathbb{P}}} \end{pmatrix}.$$

And then, relocating the factors using

$$\underbrace{\begin{bmatrix} \begin{bmatrix} X_t^{**} \\ M_t^{**} \end{bmatrix} \\ \begin{bmatrix} X_{t-1}^{**} \\ M_{t-1}^{**} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} X_{t-p+1}^{**} \\ M_{t-p+1}^{**} \end{bmatrix} \end{bmatrix}}_{F_t^{**}} = \begin{bmatrix} \begin{bmatrix} X_t^* \\ M_t \end{bmatrix} \\ \begin{bmatrix} X_{t-1}^* \\ M_{t-1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} X_{t-p+1}^* \\ M_{t-p+1} \end{bmatrix} \end{bmatrix} + \underbrace{(G^{\mathbb{P}} - I)^{-1} K^{\mathbb{P}}}_{z^{\mathbb{P}}}$$

leads to

$$F_t^{**} - (G^{\mathbb{P}} - I)^{-1} K^{\mathbb{P}} = K^{\mathbb{P}} + (G^{\mathbb{P}} - I + I)(F_{t-1}^{**} - (G^{\mathbb{P}} - I)^{-1} K^{\mathbb{P}}) + \underbrace{\begin{bmatrix} \begin{bmatrix} \epsilon_{X^*,t}^{\mathbb{P}} \\ \epsilon_{M,t}^{\mathbb{P}} \end{bmatrix} \\ O_{d_{\mathbb{P}}(p-1)\times 1} \end{bmatrix}}_{\epsilon_t^{\mathbb{P}**}},$$

$$F_t^{**} = G^{\mathbb{P}} F_{t-1}^{**} + \epsilon_t^{\mathbb{P}**}. \tag{29}$$

Therefore, we can erase the intercept in $\mathbb{P}$-dynamics.

Except for the case of a singularity, $z^{\mathbb{P}}$ would be a dense vector. Therefore, we can

express the transformation as

$$X_t^{**} = X_t^* + \underbrace{\begin{bmatrix} z_1^{\mathbb{P}} \\ z_2^{\mathbb{P}} \\ z_3^{\mathbb{P}} \end{bmatrix}}_{z_{(1:3)}^{\mathbb{P}}}.$$

The short-term interest rate in terms of X_t^{**} is

$$\begin{aligned} r_t &= \beta^{*\prime} X_t^*, \\ &= \underbrace{-\beta^{*\prime} z_{(1:3)}^{\mathbb{P}}}_{\delta^{**}} + \beta^{*\prime} X_t^{**}, \end{aligned} \tag{30}$$

so it has an affine form.

Lastly, $\mathbb{Q}$-dynamics in terms of X_t^{**} is

$$\begin{aligned} X_t^{**} - z_{(1:3)}^{\mathbb{P}} &= K_X^{\mathbb{Q}} + G_{X^*X^*}^{\mathbb{Q}} (X_{t-1}^{**} - z_{(1:3)}^{\mathbb{P}}) + \epsilon_{X^*,t}^{\mathbb{Q}}, \\ X_t^{**} &= \underbrace{(K_X^{\mathbb{Q}} + (I_{d_{\mathbb{Q}}} - G_{X^*X^*}^{\mathbb{Q}}) z_{(1:3)}^{\mathbb{P}})}_{K_{X^{**}}^{\mathbb{Q}}} + G_{X^*X^*}^{\mathbb{Q}} X_{t-1}^{**} + \epsilon_{X^*,t}^{\mathbb{Q}}. \end{aligned} \tag{31}$$

Vector $K_{X^{**}}^{\mathbb{Q}}$ is a dense vector, because

$$\begin{aligned} K_{X^{**}}^{\mathbb{Q}} &= K_X^{\mathbb{Q}} + (I_{d_{\mathbb{Q}}} - G_{X^*X^*}^{\mathbb{Q}}) z_{(1:3)}^{\mathbb{P}}, \\ &= \begin{bmatrix} k_\infty^{\mathbb{Q}} \\ 0 \\ 0 \end{bmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 - \exp(-\kappa^{\mathbb{Q}}) & -\kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ 0 & 0 & 1 - \exp(-\kappa^{\mathbb{Q}}) \end{pmatrix} \begin{bmatrix} z_1^{\mathbb{P}} \\ z_2^{\mathbb{P}} \\ z_3^{\mathbb{P}} \end{bmatrix} \\ &= \begin{bmatrix} k_\infty^{\mathbb{Q}} \\ z_2^{\mathbb{P}}[1 - \exp(-\kappa^{\mathbb{Q}})] - z_3^{\mathbb{P}} \kappa^{\mathbb{Q}} \exp(-\kappa^{\mathbb{Q}}) \\ z_3^{\mathbb{P}}[1 - \exp(-\kappa^{\mathbb{Q}})] \end{bmatrix}. \end{aligned}$$

Since there are sufficient degrees of freedom in expressions of δ^{**} and $K_{X^{**}}^{\mathbb{Q}}$, these parameters are free parameters.

Given the constrained forms of β^* and $G_{X^*X^*}^{\mathbb{Q}}$, the term structure model composed of equations (29), (30), and (31) is identified as the AFNS model(Niu and Zeng, 2012). Also, check that we cannot make $K_{X^{**}}^{\mathbb{Q}}$ to a zero vector, because no such $z^{\mathbb{P}}$ makes the first element in $K_{X^{**}}^{\mathbb{Q}}$ zero. In other words, $K_{X^{**}}^{\mathbb{Q}} = O_{d_{\mathbb{Q}} \times 1}$ is an over-identification.

B Derivation of the Term Premium Formula

According to Cochrane and Piazzesi (2008), the term premium is

$$TP_{\tau,t} = \frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathbb{E}_t^{\mathbb{P}} exr_{\tau-i+1,t+i-1}, \tag{32}$$
$$where\ exr_{\tau,t} \equiv \mathbb{E}_t^{\mathbb{P}}[\ln P_{\tau-1,t+1} - \ln P_{\tau,t} - r_t].$$

By equations (1), (6), and $P_{\tau,t} = \exp[-a_\tau - b_\tau' X_t]$,

$$exr_{\tau,t} = \mathbb{E}_t^{\mathbb{P}}[-a_{\tau-1} - b_{\tau-1}'\mathcal{T}_{0,\mathcal{P}} - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(\mathcal{P}_{t+1} - c) + a_\tau + (b_\tau' - \iota')\mathcal{T}_{0,\mathcal{P}} + (b_\tau' - \iota')\mathcal{T}_{1,\mathcal{P}}(\mathcal{P}_t - c)].$$

By $\mathbb{P}$-dynamics $\mathcal{P}_{t+1} - c = K_{\mathcal{P}}^{\mathbb{P}} + \sum_{l=1}^{p} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}}(\mathcal{P}_{t+1-l} - c) + \sum_{l=1}^{p} G_{\mathcal{P}M,l}^{\mathbb{P}} M_{t+1-l} + \epsilon_{\mathcal{P},t+1}^{\mathbb{P}}$,

$$\begin{aligned}
exr_{\tau,t} =& \mathbb{E}_t^{\mathbb{P}}[-a_{\tau-1} - b_{\tau-1}'\mathcal{T}_{0,\mathcal{P}} + a_\tau + (b_\tau' - \iota')\mathcal{T}_{0,\mathcal{P}} + (b_\tau' - \iota')\mathcal{T}_{1,\mathcal{P}}(\mathcal{P}_t - c) \\
& - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(K_{\mathcal{P}}^{\mathbb{P}} + \sum_{l=1}^{p} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}}(\mathcal{P}_{t+1-l} - c) + \sum_{l=1}^{p} G_{\mathcal{P}M,l}^{\mathbb{P}} M_{t+1-l} + \epsilon_{\mathcal{P},t+1}^{\mathbb{P}})], \\
=& - a_{\tau-1} - b_{\tau-1}'\mathcal{T}_{0,\mathcal{P}} + a_\tau + (b_\tau' - \iota')\mathcal{T}_{0,\mathcal{P}} + (b_\tau' - \iota')\mathcal{T}_{1,\mathcal{P}}(\mathcal{P}_t - c) \\
& - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}K_{\mathcal{P}}^{\mathbb{P}} - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=1}^{p} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}}(\mathcal{P}_{t+1-l} - c) - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=1}^{p} G_{\mathcal{P}M,l}^{\mathbb{P}} M_{t+1-l}, \\
=& (a_\tau - a_{\tau-1}) + (b_\tau' - \iota' - b_{\tau-1}')\mathcal{T}_{0,\mathcal{P}} + (b_\tau'\mathcal{T}_{1,\mathcal{P}} - \iota'\mathcal{T}_{1,\mathcal{P}} - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}G_{\mathcal{P}\mathcal{P},1}^{\mathbb{P}})(\mathcal{P}_t - c) \\
& - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}K_{\mathcal{P}}^{\mathbb{P}} - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=2}^{p} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}}(\mathcal{P}_{t+1-l} - c) - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=1}^{p} G_{\mathcal{P}M,l}^{\mathbb{P}} M_{t+1-l}.
\end{aligned}$$

By equations (11), (12), and (10),

$$\begin{aligned}
a_\tau - a_{\tau-1} &= b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}K_{\mathcal{P}}^{\mathbb{Q}} + b_{\tau-1}'(I_{d_{\mathbb{Q}}} - G_{XX}^{\mathbb{Q}})\mathcal{T}_{0,\mathcal{P}} - 0.5 b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}_{1,\mathcal{P}}' b_{\tau-1}, \\
b_\tau' - \iota' - b_{\tau-1}' &= b_{\tau-1}'(G_{XX}^{\mathbb{Q}} - I_{d_{\mathbb{Q}}}), \\
b_\tau'\mathcal{T}_{1,\mathcal{P}} - \iota'\mathcal{T}_{1,\mathcal{P}} - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}G_{\mathcal{P}\mathcal{P},1}^{\mathbb{P}} &= b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(G_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} - G_{\mathcal{P}\mathcal{P},1}^{\mathbb{P}}),
\end{aligned}$$

and these results simplify $exr_{\tau,t}$ to

$$\begin{aligned}
exr_{\tau,t} =& - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(K_{\mathcal{P}}^{\mathbb{P}} - K_{\mathcal{P}}^{\mathbb{Q}}) - 0.5 b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}_{1,\mathcal{P}}' b_{\tau-1} + b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(G_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}} - G_{\mathcal{P}\mathcal{P},1}^{\mathbb{P}})(\mathcal{P}_t - c) \\
& - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=2}^{p} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}}(\mathcal{P}_{t+1-l} - c) - b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}\sum_{l=1}^{p} G_{\mathcal{P}M,l}^{\mathbb{P}} M_{t+1-l}.
\end{aligned}$$

Using notations defined in equation (13),

$$\begin{aligned} exr_{\tau,t} =& - b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\lambda_{\mathcal{P}} - 0.5b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}b_{\tau-1} - b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\Lambda_{\mathcal{P}\mathcal{P},1}(\mathcal{P}_t - c) \\ & - b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\sum_{l=2}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}(\mathcal{P}_{t+1-l} - c) - b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\sum_{l=1}^{p}\Lambda_{\mathcal{P}M,l}M_{t+1-l}, \\ =& - 0.5b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}b_{\tau-1} - b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}(\lambda_{\mathcal{P}} + \sum_{l=1}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}(\mathcal{P}_{t+1-l} - c) + \sum_{l=1}^{p}\Lambda_{\mathcal{P}M,l}M_{t+1-l}). \end{aligned} \tag{33}$$

Therefore,

$$\begin{aligned} TP_{\tau,t} =& \frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathbb{E}_t^{\mathbb{P}}exr_{\tau-i+1,t+i-1}, \\ =& - \frac{1}{\tau}\sum_{i=1}^{\tau-1}(0.5b'_i\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}b_i + b'_i\mathcal{T}_{1,\mathcal{P}}\lambda_{\mathcal{P}}) \\ & - \frac{1}{\tau}\sum_{i=1}^{\tau-1}b'_{\tau-i}\mathcal{T}_{1,\mathcal{P}}(\sum_{l=1}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}\mathbb{E}_t^{\mathbb{P}}[\mathcal{P}_{t+i-l} - c] + \sum_{l=1}^{p}\Lambda_{\mathcal{P}M,l}\mathbb{E}_t^{\mathbb{P}}[M_{t+i-l}]). \end{aligned}$$

C Scaled State-Space Form Model

Our measurement equation is

$$\begin{aligned} \mathcal{O}_t &= W_{\mathcal{O}}R_t = \mathcal{A}_{\mathcal{P}} + \mathcal{B}_{\mathcal{P}}(\mathcal{P}_t - c) + \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, \Sigma_{\mathcal{O}}), \\ \mathcal{A}_{\mathcal{P}} &= W_{\mathcal{O}}[\mathcal{A}_X + \mathcal{B}_X\mathcal{T}_{0,\mathcal{P}}], \\ \mathcal{B}_{\mathcal{P}} &= W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}, \\ \mathcal{T}_{1,\mathcal{P}} &= (W_{\mathcal{P}}\mathcal{B}_X)^{-1}, \\ \mathcal{T}_{0,\mathcal{P}} &= -\mathcal{T}_{1,\mathcal{P}}(W_{\mathcal{P}}\mathcal{A}_X - c) = -\mathcal{T}_{1,\mathcal{P}}\mathcal{T}_{0,X}, \\ \mathcal{A}'_X &= \begin{bmatrix} a_{\tau_1}/\tau_1 & a_{\tau_2}/\tau_2 & \cdots & a_{\tau_N}/\tau_N \end{bmatrix}, \\ \mathcal{B}'_X &= \begin{bmatrix} b_{\tau_1}/\tau_1 & b_{\tau_2}/\tau_2 & \cdots & b_{\tau_N}/\tau_N \end{bmatrix}, \\ a_\tau &= a_{\tau-1} + b'_{\tau-1}K_X^{\mathbb{Q}} - 0.5b'_{\tau-1}\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}'_{1,\mathcal{P}}b_{\tau-1}, \\ b_\tau &= \iota + G_{XX}^{\mathbb{Q}\prime}b_{\tau-1}, \end{aligned}$$

and the transition equation is

$$\begin{bmatrix} \mathcal{P}_t - c \\ M_t \end{bmatrix} = \begin{bmatrix} K_{\mathcal{P}}^{\mathbb{P}} \\ K_{M|\mathcal{F}}^{\mathbb{P}} \end{bmatrix} + \sum_{l=1}^{p} \begin{pmatrix} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}} & G_{\mathcal{P}M,l}^{\mathbb{P}} \\ G_{M\mathcal{P},l}^{\mathbb{P}} & G_{MM,l}^{\mathbb{P}} \end{pmatrix} \begin{bmatrix} \mathcal{P}_{t-l} - c \\ M_{t-l} \end{bmatrix} + \mathcal{N}(O_{d_{\mathbb{P}}\times 1}, \begin{pmatrix} \Omega_{\mathcal{P}\mathcal{P}} & \Omega'_{M\mathcal{P}} \\ \Omega_{M\mathcal{P}} & \Omega_{MM} \end{pmatrix}).$$

The term premium is

$$\begin{aligned} TP_{\tau,t} = &-\frac{1}{\tau}\sum_{i=1}^{\tau-1}(0.5b_i'\mathcal{T}_{1,\mathcal{P}}\Omega_{\mathcal{P}\mathcal{P}}\mathcal{T}_{1,\mathcal{P}}'b_i + b_i'\mathcal{T}_{1,\mathcal{P}}\lambda_{\mathcal{P}}) \\ &-\frac{1}{\tau}\sum_{i=1}^{\tau-1} b_{\tau-i}'\mathcal{T}_{1,\mathcal{P}}(\sum_{l=1}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}\mathbb{E}_t^{\mathbb{P}}[\mathcal{P}_{t+i-l} - c] + \sum_{l=1}^{p}\Lambda_{\mathcal{P}M,l}\mathbb{E}_t^{\mathbb{P}}[M_{t+i-l}]). \end{aligned}$$

If we use data in percent per annum so yields are $1200 \times R_t$, the state space should be modified as

$$\begin{aligned} \mathbf{1200}\cdot\mathcal{O}_{\mathbf{t}} &= W_{\mathcal{O}}(\mathbf{1200}\cdot\mathbf{R_t}) = \mathbf{1200}\cdot\mathcal{A}_{\mathcal{P}} + \mathcal{B}_{\mathcal{P}}(\mathbf{1200}\cdot\mathcal{P}_{\mathbf{t}} - \mathbf{1200}\cdot\mathbf{c}) + \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, \mathbf{1200^2}\cdot\mathbf{\Sigma}_{\mathcal{O}}), \\ \mathbf{1200}\cdot\mathcal{A}_{\mathcal{P}} &= W_{\mathcal{O}}[\mathbf{1200}\cdot\mathcal{A}_{\mathbf{X}} + \mathcal{B}_X(\mathbf{1200}\cdot\mathcal{T}_{\mathbf{0},\mathcal{P}})], \\ \mathcal{B}_{\mathcal{P}} &= W_{\mathcal{O}}\mathcal{B}_X\mathcal{T}_{1,\mathcal{P}}, \\ \mathcal{T}_{1,\mathcal{P}} &= (W_{\mathcal{P}}\mathcal{B}_X)^{-1}, \\ \mathbf{1200}\cdot\mathcal{T}_{\mathbf{0},\mathcal{P}} &= -\mathcal{T}_{1,\mathcal{P}}[\mathbf{W}_{\mathcal{P}}(\mathbf{1200}\cdot\mathcal{A}_{\mathbf{X}}) - \mathbf{1200}\cdot\mathbf{c}] = -\mathcal{T}_{1,\mathcal{P}}(\mathbf{1200}\cdot\mathcal{T}_{\mathbf{0},\mathbf{X}}), \\ \mathbf{1200}\cdot\mathcal{A}_{\mathbf{X}}' &= \begin{bmatrix}(\mathbf{1200}\cdot\mathbf{a_{\tau_1}})/\tau_1 & (\mathbf{1200}\cdot\mathbf{a_{\tau_2}})/\tau_2 & \cdots & (\mathbf{1200}\cdot\mathbf{a_{\tau_N}})/\tau_N\end{bmatrix}, \\ \mathcal{B}_X' &= \begin{bmatrix} b_{\tau_1}/\tau_1 & b_{\tau_2}/\tau_2 & \cdots & b_{\tau_N}/\tau_N\end{bmatrix}, \\ \mathbf{1200}\cdot\mathbf{a}_{\tau} &= \mathbf{1200}\cdot\mathbf{a_{\tau-1}} + b_{\tau-1}'(\begin{bmatrix}\mathbf{1200}\cdot\mathbf{k}_{\infty}^{\mathbb{Q}} \\ O_{(d_{\mathbb{Q}}-1)\times 1}\end{bmatrix}) - \frac{0.5b_{\tau-1}'\mathcal{T}_{1,\mathcal{P}}(\mathbf{1200^2}\cdot\mathbf{\Omega}_{\mathcal{P}\mathcal{P}})\mathcal{T}_{1,\mathcal{P}}'b_{\tau-1}}{1200}, \\ b_{\tau} &= \iota + G_{XX}^{\mathbb{Q}\prime}b_{\tau-1}, \end{aligned}$$

and

$$\begin{aligned} \begin{bmatrix}\mathbf{1200}\cdot\mathcal{P}_{\mathbf{t}} - \mathbf{1200}\cdot\mathbf{c} \\ M_t\end{bmatrix} = &\begin{bmatrix}\mathbf{1200}\cdot\mathbf{K}_{\mathcal{P}}^{\mathbb{P}} \\ K_{M|\mathcal{F}}^{\mathbb{P}}\end{bmatrix} + \sum_{l=1}^{p}\begin{pmatrix} G_{\mathcal{P}\mathcal{P},l}^{\mathbb{P}} & \mathbf{1200}\cdot\mathbf{G}_{\mathbf{\mathcal{P}M,l}}^{\mathbb{P}} \\ \mathbf{1200^{-1}}\cdot\mathbf{G}_{\mathbf{M\mathcal{P},l}}^{\mathbb{P}} & G_{MM,l}^{\mathbb{P}}\end{pmatrix}\begin{bmatrix}\mathbf{1200}\cdot\mathcal{P}_{\mathbf{t-l}} - \mathbf{1200}\cdot\mathbf{c} \\ M_{t-l}\end{bmatrix} \\ &+\mathcal{N}\left(O_{d_{\mathbb{P}}\times 1}, \begin{pmatrix}\mathbf{1200^2}\cdot\mathbf{\Omega}_{\mathcal{P}\mathcal{P}} & \mathbf{1200}\cdot\mathbf{\Omega}_{\mathbf{M\mathcal{P}}}' \\ \mathbf{1200}\cdot\mathbf{\Omega}_{\mathbf{M\mathcal{P}}} & \Omega_{MM}\end{pmatrix}\right). \end{aligned}$$

The term premium is also modified to

$$\mathbf{1200}\cdot\mathbf{TP}_{\tau,\mathbf{t}} = -\frac{1}{\tau}\sum_{i=1}^{\tau-1}[\frac{0.5b_i'\mathcal{T}_{1,\mathcal{P}}(\mathbf{1200^2}\cdot\mathbf{\Omega}_{\mathcal{P}\mathcal{P}})\mathcal{T}_{1,\mathcal{P}}'b_i}{1200} + b_i'\mathcal{T}_{1,\mathcal{P}}(\mathbf{1200}\cdot\lambda_{\mathcal{P}})]$$

$$- \frac{1}{\tau}\sum_{i=1}^{\tau-1} b'_{\tau-i}\mathcal{T}_{1,\mathcal{P}}(\sum_{l=1}^{p}\Lambda_{\mathcal{P}\mathcal{P},l}\mathbb{E}_t^{\mathbb{P}}[\mathbf{1200}\cdot\mathcal{P}_{\mathbf{t+i-l}} - \mathbf{1200}\cdot\mathbf{c}] + \sum_{l=1}^{p}\mathbf{1200}\cdot\mathbf{\Lambda_{\mathcal{P}M,l}}\mathbb{E}_t^{\mathbb{P}}[M_{t+i-l}]),$$

and

$$\mathbf{1200}\cdot\lambda_{\mathcal{P}} = \mathbf{1200}\cdot\mathbf{K}_{\mathcal{P}}^{\mathbb{P}} - \mathcal{T}_{1,X}[\begin{bmatrix}\mathbf{1200}\cdot\mathbf{k}_{\infty}^{\mathbb{Q}}\\ O_{(d_{\mathbb{Q}}-1)\times 1}\end{bmatrix} + (G_{XX}^{\mathbb{Q}} - I_{d_{\mathbb{Q}}})(\mathbf{1200}\cdot\mathcal{T}_{\mathbf{0},\mathcal{P}})].$$

D Posterior Sampler

We use the multi-block MCMC algorithm to sample our joint posterior distribution. Here, θ is a set containing all parameters in the model.

D.1 $\mathbb{Q}$ Constant Term $k_{\infty}^{\mathbb{Q}}$

The full-conditional posterior distribution is

$$p(k_{\infty}^{\mathbb{Q}}|\mathcal{I}_T, \theta\backslash\{k_{\infty}^{\mathbb{Q}}\}) \propto p(k_{\infty}^{\mathbb{Q}})\Pi_{t=1}^{T}\mathcal{N}(\mathcal{O}_t|\mathcal{A}_{\mathcal{P}}(\kappa^{\mathbb{Q}}, k_{\infty}^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) + \mathcal{B}_{\mathcal{P}}(\kappa^{\mathbb{Q}})(\mathcal{P}_t - c), \Sigma_{\mathcal{O}}).$$

We first simplify the measurement equation. A close inspection on the difference equations (11) and (12) leads to

$$a_\tau = a_{\tau-1} - \mathcal{J}_\tau(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) + (\tau-1)k_{\infty}^{\mathbb{Q}}.$$

So,

$$\begin{aligned} a_\tau &= 0, \text{for } \tau < 2 \\ a_\tau &= \underbrace{-\sum_{i=2}^{\tau}\mathcal{J}_i(\kappa^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}})}_{a_{0\tau}} + \underbrace{\sum_{i=2}^{\tau}(i-1)}_{a_{1\tau}}k_{\infty}^{\mathbb{Q}}, \text{for } \tau \geq 2. \end{aligned}$$

Based on the above solution for a_τ, we can express

$$\mathcal{A}_X = \underbrace{\begin{bmatrix} a_{0\tau_1}/\tau_1 \\ a_{0\tau_2}/\tau_2 \\ \vdots \\ a_{0\tau_N}/\tau_N \end{bmatrix}}_{\mathcal{A}_{0,k_{\infty}^{\mathbb{Q}}}} + \underbrace{\begin{bmatrix} \sum_{i=2}^{\tau_1}\frac{(i-1)}{\tau_1} \\ \sum_{i=2}^{\tau_2}\frac{(i-1)}{\tau_2} \\ \vdots \\ \sum_{i=2}^{\tau_N}\frac{(i-1)}{\tau_N} \end{bmatrix}}_{\mathcal{A}_{1,k_{\infty}^{\mathbb{Q}}}} k_{\infty}^{\mathbb{Q}}.$$

So, $\mathcal{A}_X$ is an affine function of $k_\infty^{\mathbb{Q}}$.

Based on equations (9), (7), and the above result for $\mathcal{A}_X$, the measurement equation can be rephrased as

$$\begin{aligned}
\mathcal{O}_t &= \mathcal{A}_{\mathcal{P}} + \mathcal{B}_{\mathcal{P}}(\mathcal{P}_t - c) + \mathcal{N}(O_{N-d_{\mathbb{Q}}}, \Sigma_{\mathcal{O}}), \\
&= W_{\mathcal{O}}(I_N - \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} W_{\mathcal{P}})\mathcal{A}_X + W_{\mathcal{O}}\mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} c + \mathcal{B}_{\mathcal{P}}(\mathcal{P}_t - c) + \mathcal{N}(O_{N-d_{\mathbb{Q}}}, \Sigma_{\mathcal{O}}), \\
&= W_{\mathcal{O}}(I_N - \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} W_{\mathcal{P}})(\mathcal{A}_{0,k_\infty^{\mathbb{Q}}} + \mathcal{A}_{1,k_\infty^{\mathbb{Q}}} k_\infty^{\mathbb{Q}}) + W_{\mathcal{O}}\mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} c + \mathcal{B}_{\mathcal{P}}(\mathcal{P}_t - c) + \mathcal{N}(O_{N-d_{\mathbb{Q}}}, \Sigma_{\mathcal{O}}).
\end{aligned}$$

By reallocating terms in the above equation, we arrive at

$$\begin{aligned}
&\underbrace{\Sigma_{\mathcal{O}}^{-1/2}[\mathcal{O}_t - W_{\mathcal{O}}(I_N - \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} W_{\mathcal{P}})\mathcal{A}_{0,k_\infty^{\mathbb{Q}}} - W_{\mathcal{O}}\mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} c - \mathcal{B}_{\mathcal{P}}(\mathcal{P}_t - c)]}_{y_{k_\infty^{\mathbb{Q}},t}} \\
=&\underbrace{\Sigma_{\mathcal{O}}^{-1/2} W_{\mathcal{O}}(I_N - \mathcal{B}_X \mathcal{T}_{1,\mathcal{P}} W_{\mathcal{P}})\mathcal{A}_{1,k_\infty^{\mathbb{Q}}}}_{x_{k_\infty^{\mathbb{Q}}}} k_\infty^{\mathbb{Q}} + \mathcal{N}(O_{(N-d_{\mathbb{Q}})\times 1}, I_{N-d_{\mathbb{Q}}}).
\end{aligned}$$

Stacking up $y_{k_\infty^{\mathbb{Q}},t}$ leads to

$$\underbrace{\begin{bmatrix} y_{k_\infty^{\mathbb{Q}},1} \\ y_{k_\infty^{\mathbb{Q}},2} \\ \vdots \\ y_{k_\infty^{\mathbb{Q}},T} \end{bmatrix}}_{Y_{k_\infty^{\mathbb{Q}}}} = \underbrace{\begin{bmatrix} x_{k_\infty^{\mathbb{Q}}} \\ x_{k_\infty^{\mathbb{Q}}} \\ \vdots \\ x_{k_\infty^{\mathbb{Q}}} \end{bmatrix}}_{X_{k_\infty^{\mathbb{Q}}}} k_\infty^{\mathbb{Q}} + \mathcal{N}(O_{T(N-d_{\mathbb{Q}})\times 1}, I_{T(N-d_{\mathbb{Q}})}).$$

This derivation says that we can do the Normal-Normal update for $k_\infty^{\mathbb{Q}}$, so

$$p(k_\infty^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{k_\infty^{\mathbb{Q}}\}) = \mathcal{N}((X'_{k_\infty^{\mathbb{Q}}} X_{k_\infty^{\mathbb{Q}}} + \sigma_{k_\infty^{\mathbb{Q}}}^{-2})^{-1} X'_{k_\infty^{\mathbb{Q}}} Y_{k_\infty^{\mathbb{Q}}}, (X'_{k_\infty^{\mathbb{Q}}} X_{k_\infty^{\mathbb{Q}}} + \sigma_{k_\infty^{\mathbb{Q}}}^{-2})^{-1}).$$

We do not care about the Jacobian term because it does not depend on $k_\infty^{\mathbb{Q}}$.

D.2 Decay Parameter $\kappa^{\mathbb{Q}}$

The unnormalized kernel of the full-conditional posterior distribution is

$$p^*(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\}) = p(\kappa_i^{\mathbb{Q}}) \Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t | \mathcal{A}_{\mathcal{P}}(\kappa_i^{\mathbb{Q}}, k_\infty^{\mathbb{Q}}, \Omega_{\mathcal{P}\mathcal{P}}) + \mathcal{B}_{\mathcal{P}}(\kappa_i^{\mathbb{Q}})(\mathcal{P}_t - c), \Sigma_{\mathcal{O}}).$$

Therefore, the posterior is

$$p(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\}) = \frac{p^*(\kappa^{\mathbb{Q}} = \kappa_i^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\})}{\sum_{j \in \{36,37,\cdots,42\}} p^*(\kappa^{\mathbb{Q}} = \kappa_j^{\mathbb{Q}} | \mathcal{I}_T, \theta \backslash \{\kappa^{\mathbb{Q}}\})},$$

for $i \in \{36, 37, \cdots, 42\}$.

D.3 Parameters in the Transition Equation

From the recursive structure that $\Omega_{\mathcal{FF}}$ has (see section 3.1.1), we can derive that

$$\Omega_{\mathcal{PP}} = (\mathcal{C}^{-1})_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})} diag[\tilde{\Omega}_{\mathcal{FF},(1:d_{\mathbb{Q}})}]((\mathcal{C}^{-1})_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})'.$$

Since

$$\mathcal{C} = \begin{pmatrix} \mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})} & O_{d_{\mathbb{Q}} \times (d_{\mathbb{P}} - d_{\mathbb{Q}})} \\ \mathcal{C}_{(d_{\mathbb{Q}}+1:d_{\mathbb{P}},1:d_{\mathbb{Q}})} & \mathcal{C}_{(d_{\mathbb{Q}}+1:d_{\mathbb{P}},d_{\mathbb{Q}}+1:d_{\mathbb{P}})} \end{pmatrix},$$

$(\mathcal{C}^{-1})_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})} = (\mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})^{-1}$ holds, so

$$\Omega_{\mathcal{PP}} = (\mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})^{-1} diag[\tilde{\Omega}_{\mathcal{FF},(1:d_{\mathbb{Q}})}](\mathcal{C}_{(1:d_{\mathbb{Q}},1:d_{\mathbb{Q}})})^{-1\prime}.$$

By the above expression, $\sigma^2_{\mathcal{FF},1}$, and $\{C_{(i,1:i-1)}, \sigma^2_{\mathcal{FF},i}\}_{i=2}^{d_{\mathbb{Q}}}$ decide $\Omega_{\mathcal{PP}}$. Since all entries in $\Omega_{\mathcal{PP}}$ are in both the transition and measurement equations, $\sigma^2_{\mathcal{FF},1}$, and $\{C_{(i,1:i-1)}, \sigma^2_{\mathcal{FF},i}\}_{i=2}^{d_{\mathbb{Q}}}$ affect both the likelihood of the transition and measurement equations. In contrast, the other elements in $\mathcal{C}$ and $\tilde{\Omega}_{\mathcal{FF}}$ are only relevant to the likelihood of the transition equation. Therefore, we conduct the Metropolis-Hastings(MH) algorithm for $\sigma^2_{\mathcal{FF},1}$, and $\{C_{(i,1:i-1)}, \sigma^2_{\mathcal{FF},i}\}_{i=2}^{d_{\mathbb{Q}}}$, and impose conjugate prior distributions on the other parameters.

Before deriving the full-conditional posterior distributions of the parameters, we describe how the Normal-InverseGamma-Normal-InverseGamma update is conducted. In the general case of

$$\begin{aligned} y|\beta,\sigma^2 &= X\beta + \mathcal{N}(O, \sigma^2 I_T), \\ \beta|\sigma^2 &\backsim \mathcal{N}(\beta_0, \sigma^2 B_0), \\ \sigma^2 &\backsim \mathcal{IG}(\alpha_0, \delta_0), \end{aligned}$$

the Normal-InverseGamma-Normal-InverseGamma update leads to the posterior distribution of

$$\begin{aligned}\beta|\sigma^2, y &\backsim \mathcal{N}(\beta_1, \sigma^2 B_1),\\ \sigma^2|y &\backsim \mathcal{IG}(\alpha_0 + \frac{T}{2}, \delta_1),\end{aligned}$$

where $B_1 = (B_0^{-1} + X'X)^{-1}$, $\beta_1 = B_1(B_0^{-1}\beta_0 + X'y)$, and $\delta_1 = \delta_0 + (y'y + \beta_0' B_0^{-1}\beta_0 - \beta_1' B_1^{-1}\beta_1)/2$. We use this result to conduct the update.

D.3.1 MH Algorithm for $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_\mathbb{Q}}$

The full-conditional posterior is

$$\begin{aligned}&\Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t|\mathcal{A}_\mathcal{P}(\kappa^\mathbb{Q}, k_\infty^\mathbb{Q}, \Omega_{\mathcal{PP}}) + \mathcal{B}_\mathcal{P}(\kappa^\mathbb{Q})(\mathcal{P}_t - c), \Sigma_\mathcal{O}) \times \Pi_{i=1}^{d_\mathbb{P}} \mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i} I_T)\\ &\times p(\{\sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_\mathbb{P}}|\nu_0, \Omega_0) p(\{\phi_i\}_{i=1}^{d_\mathbb{P}}|p, \{\sigma^2_{\mathcal{FF},i}\}_{i=1}^{d_\mathbb{P}}, q, \nu_0, \Omega_0).\end{aligned}$$

In this subsection, we sampling $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}$ for each $i \in \{1, \cdots, d_\mathbb{Q}\}$ separately. For specific i, the targeted $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}$ has the full-conditional posterior distribution of

$$\begin{aligned}&\Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t|\mathcal{A}_\mathcal{P}(\kappa^\mathbb{Q}, k_\infty^\mathbb{Q}, \Omega_{\mathcal{PP}}) + \mathcal{B}_\mathcal{P}(\kappa^\mathbb{Q})(\mathcal{P}_t - c), \Sigma_\mathcal{O})\\ &\times \mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i} I_T) \times \mathcal{N}(\phi_i|m_i, \sigma^2_{\mathcal{FF},i} \cdot diag[V_i]) \times \mathcal{IG}(\sigma^2_{\mathcal{FF},i}|\frac{\nu_0 + i - d_\mathbb{P}}{2}, \frac{\Omega_{0,(i)}}{2}).\end{aligned}$$

The second line

$$\mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i} I_T) \times \mathcal{N}(\phi_i|m_i, \sigma^2_{\mathcal{FF},i} \cdot diag[V_i]) \times \mathcal{IG}(\sigma^2_{\mathcal{FF},i}|\frac{\nu_0 + i - d_\mathbb{P}}{2}, \frac{\Omega_{0,(i)}}{2})$$

corresponds to the Normal-InverseGamma-Normal-InverseGamma update. This update only uses information in the transition equation. We use the above expression as a kernel of the proposal distribution.

After we draw a proposal from the updated Normal-InverseGamma distribution, we accept the proposal with a probability of

$$min[1, \frac{\Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t|\mathcal{A}_\mathcal{P}(\kappa^\mathbb{Q}, k_\infty^\mathbb{Q}, \Omega_{\mathcal{PP},proposal}) + \mathcal{B}_\mathcal{P}(\kappa^\mathbb{Q})\mathcal{P}_t, \Sigma_\mathcal{O})}{\Pi_{t=1}^T \mathcal{N}(\mathcal{O}_t|\mathcal{A}_\mathcal{P}(\kappa^\mathbb{Q}, k_\infty^\mathbb{Q}, \Omega_{\mathcal{PP},current}) + \mathcal{B}_\mathcal{P}(\kappa^\mathbb{Q})\mathcal{P}_t, \Sigma_\mathcal{O})}],$$

where $\Omega_{\mathcal{PP},proposal}$ and $\Omega_{\mathcal{PP},current}$ are a new proposal and the current value, respectively.

If a new proposal increases the likelihood of the measurement equation, we accept it. If a new proposal has a lower likelihood value for the measurement equation, we accept the new proposal with a probability corresponding to the likelihood ratio between the new and existing proposals.

D.3.2 Other Transition Equation Parameters $\{\phi_i, \sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}$

The full-conditional distribution is

$$\begin{aligned}&\Pi_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}\mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i}I_T) \times p(\{\sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}|\nu_0, \Omega_0)\\ \times &p(\{\phi_i\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}|p, \{\sigma^2_{\mathcal{FF},i}\}_{i=d_{\mathbb{Q}}+1}^{d_{\mathbb{P}}}, q, \nu_0, \Omega_0).\end{aligned}$$

For $i \in \{d_{\mathbb{Q}}+1, \cdots, d_{\mathbb{P}}\}$, the kernel is

$$\mathcal{N}(y_{\phi_i}|X_{\phi_i}\phi_i, \sigma^2_{\mathcal{FF},i}I_T) \times \mathcal{N}(\phi_i|m_i, \sigma^2_{\mathcal{FF},i}diag[V_i]) \times \mathcal{IG}(\sigma^2_{\mathcal{FF},i}|\frac{\nu_0 + i - d_{\mathbb{P}}}{2}, \frac{\Omega_{0,(i)}}{2}),$$

so it is the Normal-InverseGamma-Normal-InverseGamma update.

D.4 Measurement Error Variance $\Sigma_{\mathcal{O}}$

The full-conditional distribution is

$$\begin{aligned}&\Pi_{i=1}^{N-d_{\mathbb{Q}}}[\{\Pi_{t=1}^{T}\mathcal{N}(\mathcal{O}_{t,(i)}|\mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_t - c), \sigma^2_{\mathcal{O},i})\}\mathcal{IG}(\sigma^2_{\mathcal{O},i}|2, \gamma_i)]\\ =&\Pi_{i=1}^{N-d_{\mathbb{Q}}}[\mathcal{N}(\underbrace{\begin{bmatrix}\mathcal{O}_{1,(i)}\\ \mathcal{O}_{2,(i)}\\ \vdots\\ \mathcal{O}_{T,(i)}\end{bmatrix}}_{y_{\sigma^2_{\mathcal{O},i}}} | \underbrace{\begin{bmatrix}\mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_1 - c)\\ \mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_2 - c)\\ \vdots\\ \mathcal{A}_{\mathcal{P},(i)} + \mathcal{B}_{\mathcal{P},(i,:)}(\mathcal{P}_T - c)\end{bmatrix}}_{m_{\sigma^2_{\mathcal{O},i}}}, \sigma^2_{\mathcal{O},i}I_T)\mathcal{IG}(\sigma^2_{\mathcal{O},i}|2, \gamma_i)].\end{aligned}$$

Therefore, there is a posterior independence for each i. We can update $\sigma^2_{\mathcal{O},i}$ for each i by

$$p(\sigma^2_{\mathcal{O},i}|\mathcal{I}_T, \theta\backslash\{\Sigma_{\mathcal{O}}\}) = \mathcal{IG}(\sigma^2_{\mathcal{O},i}|2 + \frac{T}{2}, \gamma_i + \frac{(y_{\sigma^2_{\mathcal{O},i}} - m_{\sigma^2_{\mathcal{O},i}})'(y_{\sigma^2_{\mathcal{O},i}} - m_{\sigma^2_{\mathcal{O},i}})}{2}).$$

D.5 Mean of Measurement Error Variance $\{\gamma_i\}_{i=1}^{N-d_{\mathbb{Q}}}$

The corresponding full-conditional kernel is

$$\begin{aligned}&\Pi_{i=1}^{N-d_{\mathbb{Q}}} p(\sigma^2_{\mathcal{O},i}|\gamma_i)p(\gamma_i)\\ =&\Pi_{i=1}^{N-d_{\mathbb{Q}}} \mathcal{IG}(\sigma^2_{\mathcal{O},i}|2,\gamma_i)\mathcal{G}(\gamma_i|1,\bar{\gamma}).\end{aligned}$$

Therefore, there is a posterior independence for each i. For each i,

$$\begin{aligned}&\mathcal{IG}(\sigma^2_{\mathcal{O},i}|2,\gamma_i)\mathcal{G}(\gamma_i|1,\bar{\gamma})\\ =&\frac{\gamma_i^2}{\Gamma(2)}(\sigma^2_{\mathcal{O},i})^{-(2+1)}\exp[-\sigma^{-2}_{\mathcal{O},i}\gamma_i]\frac{\bar{\gamma}}{\Gamma(1)}\exp[-\bar{\gamma}\gamma_i],\\ \propto&\gamma_i^{3-1}\exp[-(\sigma^{-2}_{\mathcal{O},i}+\bar{\gamma})\gamma_i]\end{aligned}$$

holds. Therefore, we update γ_i by

$$p(\gamma_i|\mathcal{I}_T,\theta\backslash\{\gamma_i\}_{i=1}^{N-d_{\mathbb{Q}}}) = \mathcal{G}(3,\frac{1}{\sigma^2_{\mathcal{O},i}}+\bar{\gamma}).$$