Unspanned Macro Factor Selection in Affine Term Structure Models: A Framework*

Siddhartha Chib

Olin Business School, Washington University in St. Louis, E-mail: chib@wustl.edu

Kyu Ho Kang

Department of Economics, Korea University, E-mail: kyuho@korea.ac.kr

Biancen Xie

Wells Fargo, Charlotte, NC, E-mail: biancen.xie@wellsfargo.com

Cheol Woo Lee

Department of Economics, The Ohio State University, E-mail: lee.8968@osu.edu

December 2019

Abstract

Recent empirical studies of macro-finance affine term structure models have used pre-selected macroeconomic factors to estimate term premium dynamics. The choice of macro factors, however, has a bearing on the inferred term premium dynamics. To deal with this problem, we develop a Bayesian framework for selecting the relevant macro factors via marginal likelihood comparison. According to our empirical application with 10 potential relevant macro risk factors, and 1,024 possible models, the Chicago Fed National Activity Index is the factor most supported by the data. This finding seems to be associated with the counter-cyclical nature of the term premium. (JEL classification: G12, C11, E43)

Keywords: Macro-finance affine term structure model, marginal likelihood, term premium, Markov chain Monte Carlo

*We thank the participants of the 2019 North American Summer Meeting of The Econometric Society, 2019 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, Korea University, and University of Washington for valuable comments. We gratefully acknowledge generous financial support from the Center for Research and Strategy at Olin Business School, Washington University in St. Louis. This work is also supported by a Korea University Grant (K1808871).

1. Introduction

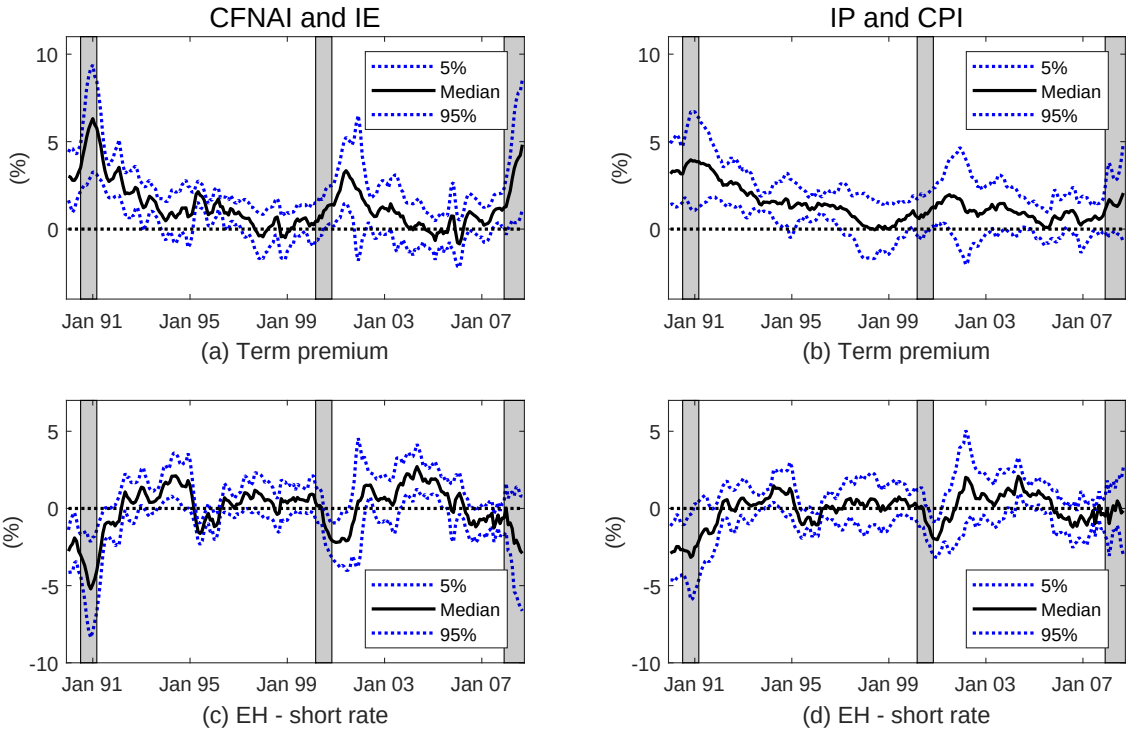
One key objective in yield curve analysis is decomposing the long term interest rate into its term premium and market expectation (EH) components. It has been well established that the dynamics of these components, in particular the dynamics of the term premium, are strongly influenced by macro factors, for example, Ang and Piazzesi (2003), Bernanke et al. (2004), Bikbov and Chernov (2008), and Ludvigson and Ng (2009). In recent important work, Joslin et al. (2014) have argued that these dynamics should be examined within affine models in which the macro-factors are unspanned by the yield curve, which in the context of affine models requires that the macro factors enter the stochastic discount factor (SDF) indirectly through an affect on the market price of unobserved factor risks. Even in this case, the macro factors affect the dynamics of the term premium and (may) improve out-of-sample forecasts of the yield curve. Joslin et al. (2014) use two unspanned macro-factors in their empirical study, and assume that each is relevant (ie., each affects the market price of factor risks). In this paper we develop a framework for validating such an assumption, and for selecting the relevant macro factors given a pool of potential macro factors.

The problem of unspanned macro factor selection is important because the identity of the chosen macro factors can strongly influence the posterior (ex-post) beliefs/inferences about the term premium and EH components. To see this, consider two sets of macro factors: the Chicago Fed National Activity Index (CFNAI) and the inflation expectation survey (IE) factors used in the empirical analysis of Joslin et al. (2014), and the industrial production index (IP) and consumer price index (CPI) inflation rate factors in the analysis of Li et al. (2012). Figure 1 plots and compares the term premium and the EH (minus risk-free short rate) dynamics of a five-year bond over the sample period from January 1990 to September 2008, estimated with (CFNAI, IE), and the other with (IP, CPI), as the relevant unspanned macro factors.[1] The difference is rather significant. For instance, the term premiums during the recessions estimated with (CFNAI, IE) as the

[1]The details behind the model specification and the prior-posterior analysis that leads to these inferences are discussed in Sections 2 and 3.

relevant factors are much higher than those with the (IP, CPI) factors. Around 2008, the model with (CFNAI, IE) suggests that the future short rates are expected to fall on average over the next five years, while the model with (IP, CPI) does not. These varying implications about the term premium and EH components from the two specifications are substantial enough that the resulting economic decisions would likely differ.

Figure 1: Term premium and market expectation: five-year bond The panels (a) and (c) plot the term premium and EH minus the risk-free short rate of the five-year bond, respectively. The dashed lines are 90 percent posterior credibility intervals and solid lines are the posterior means. *CFNAI* and *IE* are the Chicago Fed National Activity Index and inflation expectation survey data by the University of Michigan, respectively. *IP* and *CPI* are the year-to-year industrial production index growth and CPI inflation rate, respectively. Shaded areas indicate recession periods designated by the National Bureau of Economic Research.



The question of which macro factors affect the term premium is also interesting because this question cannot be answered by standard econometric techniques. The set of modeled outcomes in affine models is the yield curve and macro factors, and this set changes depending on which macro factors are withdrawn or introduced. In this paper, we supply a framework for overcoming this problem. The framework, which is inspired by Chib et al. (2018), Chib and Zeng (2018) and Chib et al. (2019), starts

with a model that includes all the macro variables in contention. Specific models (with different combinations of relevant and irrelevant macro factors) are then derived from this grand model, making sure that the total number of factors is the same across these models.

The macro factors considered in our empirical application are listed in Table 1. This collection consists of four inflation measures, five real economic activity indices, and one interest rate factor.

Table 1: Candidate Macroeconomic Factors This table lists the acronyms of each macro series. All series are from the online database housed at the Federal Reserve Bank of St. Louis.

Price	CPI	CPI inflation rate
	CORE	core inflation rate
	PPI	PPI growth rate
	IE	inflation expectation: University of Michigan
Real Activity	IP	industrial production index growth rate
	PCE	personal consumption expenditures growth rate
	TCU	capacity utilization: total industry
	CFNAI	Chicago Fed National Activity Index
	EMP	employment rate
Interest Rate	FFR	effective federal funds rate

Under these 10 candidate macro factors, our method of considering all the possible combinations of relevant and irrelevant macro risk factors leads to $2^{10} = 1,024$ competing affine models, each with its own model parameters and specific implications for the term structure of interest rates. We compare these models in terms of Bayesian marginal likelihoods. The marginal likelihood has a built-in penalty for complexity, ensuring that more complex models will not rank higher merely because the more flexible model is capable of fitting the noise in the data.

Our main findings from the comparison of this trove of 1,024 models, applied to the U.S. yield curve data over the sample period from January 1990 to September 2008, are summarized in Table 2. This table presents the posterior probabilities of the top five models and the corresponding relevant macro risk factors. We find that the model with one relevant factor (CFNAI) and nine irrelevant macro risk factors is the best supported by the data. The evidence in favor of this macro risk factor combination is decisive, given

Table 2: Five best macro factor models This table presents the relevant macro factors in the top five models. *prob.* is the posterior probability of the model, which is computed by applying Bayes theorem on the space of 1024 possible models. *log BF* is the log Bayes factor between each of the best models and the model with no relevant macro risk factor.

ranking	relevant macro factors	prob.	log BF
1	CFNAI	0.7495	13.2345
2	CPI, CFNAI	0.1339	11.5123
3	FFR, TCU, CFNAI	0.0789	10.9832
4	IP, CFNAI, EMP	0.0220	9.7059
5	IP, PCE, CFNAI	0.0048	8.1907

that its posterior probability is much higher than those of the alternatives. CFNAI is a real activity measure, which is one of the two macro risk factors used in the work of Joslin et al. (2014). This finding seems to be strongly attributed to the counter-cyclical nature of the term premium, indicating that investors have to bear more macroeconomic risks during recessions. In addition, our finding suggests that CFNAI is superior to other real activity measures in explaining yield curve dynamics. On the other hand, it is interesting that no price measures are selected, although CPI is included in the second best model. Finally, we note that Bauer and Hamilton (2018) have questioned the spanning hypothesis but not in an affine model and, hence, their findings are not directly relevant or comparable.

The remainder of the paper is organized as follows. In Section 2 we present our framework of affine term structure model with relevant and irrelevant unspanned macro risk factors. Section 3 describes the details of our Bayesian estimation strategy and marginal likelihood computation method. Section 4 provides the empirical results from the application of our approach. Concluding remarks are presented in Section 5.

2. Bond Pricing Model

We now describe the macro-finance arbitrage-free affine term structure model, built on the work of Duffee (2011), and Joslin et al. (2014), that is designed for macro factor selection. Our description focuses on one particular model, say $\mathcal{M}$, but given the 10 potential macro factors in the empirical application, we are implicitly describing each of the 1,024 models in our model space. This is due to the fact that a particular model $\mathcal{M}$

is a function of how the given set of macro factors are split into a disjoint and exhaustive set of relevant and irrelevant macro factors. The 1,024 models emerge as one varies this split in a combinatorial fashion.

A particular model $\mathcal{M}$ in our model space is defined in terms of three distinct sets of factors, $\boldsymbol{f}_t = (\boldsymbol{l}_t, \boldsymbol{m}_t, \boldsymbol{z}_t)'$, where

- $\boldsymbol{l}_t$ are latent factors of size $l\ (=3)\times 1$, which are present in every model, and which capture the level, slope, and curvature of a yield curve
- $\boldsymbol{m}_t$ of size $m\times 1$ are relevant unspanned macroeconomic factors consisting of factors that affect λ_t, the market price of latent factor risks, and
- $\boldsymbol{z}_t$ of size $z \times 1$ are irrelevant factors consisting of factors that are constrained to have no effect on λ_t

The total number of factors is $f = l + m + z$. Thus, model $\mathcal{M}$ is distinguished by the factors that comprise $\boldsymbol{m}_t$ and $\boldsymbol{z}_t$, and as a result, there are $2^{(m+z)}$ models in the model space. By construction, we keep the number of factors the same across models, a necessary condition for the models to be comparable, though in some models $\boldsymbol{m}_t$ or $\boldsymbol{z}_t$ could be empty.

As is standard in the literature, the factors are assumed to arise from a first-order vector autoregressive (VAR) process. Instead of comprising two distinct sets of factors, the VAR process is defined in terms of our three distinct factors. The process is given by

$$\underbrace{\begin{pmatrix} \boldsymbol{l}_{t+1} \\ \boldsymbol{m}_{t+1} \\ \boldsymbol{z}_{t+1} \end{pmatrix}}_{\boldsymbol{f}_{t+1}} = \underbrace{\begin{pmatrix} G_{ll} & G_{lm} & G_{lz} \\ G_{ml} & G_{mm} & G_{mz} \\ G_{zl} & G_{zm} & G_{zz} \end{pmatrix}}_{G} \underbrace{\begin{pmatrix} \boldsymbol{l}_t \\ \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix}}_{\boldsymbol{f}_t} + \underbrace{\begin{pmatrix} \varepsilon_{l,t+1} \\ \varepsilon_{m,t+1} \\ \varepsilon_{z,t+1} \end{pmatrix}}_{\varepsilon_{t+1}}, \tag{2.1}$$

where the observable macro factors are demeaned and the factor shocks ε_{t+1} are jointly normally distributed as

$$\varepsilon_{t+1} \sim \mathcal{N}\left(0_{f\times 1}, \Omega = \begin{pmatrix} \Omega_{ll} & \Omega_{lm} & \Omega_{lz} \\ \Omega_{lm}' & \Omega_{mm} & \Omega_{mz} \\ \Omega_{lz}' & \Omega_{mz}' & \Omega_{zz} \end{pmatrix}\right). \tag{2.2}$$

2.1. Stochastic Discount Factor and Restrictions on Risk Prices

The key assumptions (also standard in the literature) concern the SDF. In our case, what needs to be noted is the way that we zero out the influence of what ever factors are in $\boldsymbol{z}_t$. Specifically, the one-period ahead SDF M_{t+1} is given by

$$M_{t+1} = \mathsf{exp}\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' v_{t+1}\right), \tag{2.3}$$

where r_t is the yield on a one-period bond, or the risk-free short rate, and λ_t is the time-varying market price of factor risks $v_{t+1} \sim \mathcal{N}\left(\mathsf{0}_{f\times 1}, \boldsymbol{I}_f\right)$. The factor shock is related to the factor risks by $\varepsilon_{t+1} = \Omega^{1/2} v_{t+1}$ with $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$. The short rate is assumed to follow the process

$$r_t = \delta + \begin{pmatrix} \beta' & \mathsf{0}_{1\times m} & \mathsf{0}_{1\times z} \end{pmatrix} \begin{pmatrix} \boldsymbol{l}_t \\ \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix}, \tag{2.4}$$

which is an affine function of $\boldsymbol{l}_t$, but not of $\boldsymbol{m}_t$ or $\boldsymbol{z}_t$. Our process for the market price of risks is

$$\lambda_t = \left(\Omega^{1/2}\right)^{-1}\left(\lambda + \Lambda \begin{pmatrix} \boldsymbol{l}_t \\ \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix}\right), \tag{2.5}$$

where

$$\lambda = \begin{pmatrix} \lambda_l \\ \mathsf{0}_{m\times 1} \\ \mathsf{0}_{z\times 1} \end{pmatrix}, \text{ and } \Lambda = \begin{pmatrix} \Lambda_{ll} & \Lambda_{lm} & \mathsf{0}_{l\times z} \\ \mathsf{0}_{m\times l} & \mathsf{0}_{m\times m} & \mathsf{0}_{m\times z} \\ \mathsf{0}_{z\times l} & \mathsf{0}_{z\times m} & \mathsf{0}_{z\times z} \end{pmatrix}. \tag{2.6}$$

In this way, the macro risk factors directly affect the yield curve factors $\boldsymbol{l}_t$ and also influence the market price of risks λ_t, but the macro factors in $\boldsymbol{z}_t$ do not affect λ_t, in conformity with their status as irrelevant macro factors.

2.2. No-Arbitrage Condition and Bond Prices

Given the SDF, bond prices now emerge from the no-arbitrage condition

$$P_t(\tau) = \mathbb{E}_t[M_{t+1} P_{t+1}(\tau - 1)], \tag{2.7}$$

where $\mathbb{E}_t(\cdot)$ is the conditional expectation given $\{\boldsymbol{f}_i\}_{i=1}^t$.[2] Guessing that the optimal price of the bond with τ-periods to maturity at time t takes the exponential linear form

$$P_t(\tau) = \mathsf{exp}\left(-a(\tau) - b(\tau)'\boldsymbol{l}_t\right),$$

and using the method of undetermined coefficients, the solutions for the no-arbitrage condition are obtained as

$$G_{lm} = \Lambda_{lm}, \tag{2.9}$$

$$G_{lz} = 0_{l\times z}, \tag{2.10}$$

and the recursive system for $a(\tau)$ and $b(\tau)$,

$$\begin{aligned} a(\tau) &= \delta + a(\tau-1) - 0.5b(\tau-1)'\Omega_{ll}b(\tau-1) - b(\tau-1)'\lambda_l, \\ b(\tau) &= \beta + G_{ll}^{\mathbb{Q}\prime}b(\tau-1) \end{aligned}$$

for any non-negative integer value of τ, where $G_{ll}^{\mathbb{Q}} = G_{ll} - \Lambda_{ll}$. As $P_t(0) = 1$ for any $\boldsymbol{l}_t$, this recursive system is initialized by $a(0) = 0$ and $b(0) = 0_{l\times 1}$.

It is important to note that the restrictions on $(G,\, \Lambda)$ (2.9) and (2.10) are required to satisfy the no-arbitrage condition. Conditioned on $\boldsymbol{l}_t$ and $\boldsymbol{m}_t$, $G_{lz} = 0_{l\times z}$ blocks any influence of $\boldsymbol{z}_t$ on λ_t, and $\boldsymbol{z}_t$ has no predictive power for $\boldsymbol{l}_{t+1}$ and $P_{t+1}(\tau)$ for all τ. In contrast, $G_{lm} = \Lambda_{lm}$ implies that the impact of the relevant macro factors on the market price of risks is proportional to G_{lm}, because $P_{t+1}(\tau)$ is determined by $\boldsymbol{l}_{t+1}$ only. When $\boldsymbol{m}_t$ is empty, all macro risk factors are considered irrelevant and the model collapses to a simple term structure model with only yield curve factors. In short, the restrictions in Equations (2.9) and (2.10) clearly distinguish the roles of the relevant and irrelevant macro factors in generating the cross-section and time-series of arbitrage-free bond prices.

[2] Given $M_{t+1} = \mathsf{exp}(-r_t)$ under the risk-neutral $\mathbb{Q}$-measure, the no-arbitrage condition can be re-expressed as

$$P_t(\tau) = e^{-r_t}\mathbb{E}_t^{\mathbb{Q}}[P_{t+1}(\tau-1)], \tag{2.8}$$

where $K^{\mathbb{Q}} = -\lambda$, $G^{\mathbb{Q}} = G - \Lambda$, and the $\mathbb{Q}$-measure is $\mathcal{N}(\boldsymbol{f}_{t+1}|K^{\mathbb{Q}} + G^{\mathbb{Q}}\boldsymbol{f}_t, \Omega)d\boldsymbol{f}_{t+1}$.

2.3. Term Premium and the Role of Relevant Macro Risk Factors

The term premium is regarded as the additional expected return for the risk of holding a given bond for a longer period. Let $TP_t(\tau)$ denote the term premium of a bond with τ period to maturity at time t. As shown by Cochrane and Piazzesi (2008), $TP_t(\tau)$ is equal to the average of the expected excess returns

$$TP_t(\tau) = \tau^{-1} \sum_{i=1}^{\tau-1} \text{exr}_t^{\tau+1-i},$$

where

$$\text{exr}_t^{(\tau)} = [\mathbb{E}_t[\ln P_{t+1}(\tau-1)] - \ln P_t(\tau)] - (-\ln P_t(1))$$

is the one-period expected excess return for holding the τ-period bond at time t.

In our model specification, $\text{exr}_t^{(\tau)}$ can be expressed as

$$\text{exr}_t^{(\tau)} = -0.5 b(\tau-1)' \Omega_{ll} b(\tau-1) - b(\tau-1)' \left(\lambda_l + \Lambda_{ll} \times \boldsymbol{l_t} + \Lambda_{lm} \times \boldsymbol{m}_t\right).$$

and hence we have that

$$TP_t(\tau) = -\tau^{-1} \sum_{i=1}^{\tau-1} [0.5 b(\tau-i)' \Omega_{ll} b(\tau-i) + b(\tau-i)' \left(\lambda_l + \Lambda_{ll} \times \boldsymbol{l_t} + \Lambda_{lm} \times \boldsymbol{m}_t\right)] \tag{2.11}$$

which instructively can be decomposed into three components:

(i) a time-invariant component,

$$-\tau^{-1} \sum\nolimits_{i=1}^{\tau-1} [0.5 b(\tau-i)' \Omega_{ll} b(\tau-i) + b(\tau-i)' \lambda_l],$$

(ii) a time-varying component by the latent yield curve factors $\boldsymbol{l_t}$,

$$-\tau^{-1} \sum\nolimits_{i=1}^{\tau-1} b(\tau-i)' \Lambda_{ll} \boldsymbol{l_t} = -\tau^{-1} \sum\nolimits_{i=1}^{\tau-1} b(\tau-i)' \left(G_{ll} - G_{ll}^{\mathbb{Q}}\right) \boldsymbol{l_t} \tag{2.12}$$

(iii) a time-varying component by the relevant macro factors $\boldsymbol{m}_t$,

$$-\tau^{-1} \sum\nolimits_{i=1}^{\tau-1} b(\tau-i)' \Lambda_{lm} \boldsymbol{m}_t = -\tau^{-1} \sum\nolimits_{i=1}^{\tau-1} b(\tau-i)' G_{lm} \boldsymbol{m}_t \tag{2.13}$$

The last term represents the effect of the relevant macro risk factors on the term premium. We let the yield for the τ-period bond at time t denoted by $R_t(\tau)$. Then, given

the term premium, the EH minus the risk-free short rate is obtained as the term spread (i.e. $R_t(\tau) - r_t$) minus the term premium. That is, we have

$$\tau^{-1} \sum_{i=0}^{\tau-1} \mathbb{E}[r_{t+i}] - r_t = R_t(\tau) - r_t - TP_t(\tau).$$

It immediately follows that a positive (negative) value of the EH minus the current short rate indicates that the average expected short rates over the next τ periods is higher (lower) than the current short rate.

3. Econometric Methodology

The estimation of affine term structure models is in general non-trivial. In particular, since we are taking a Bayesian approach, and each of 1,024 models in the model space has to be estimated, the estimation challenges are considerably greater. Joslin et al. (2014) employed a three-step estimation procedure, which is fast and easy to implement: (i) the principal components are estimated independently of the no-arbitrage condition, and the first three principal components are treated as observed yield curve factors, (ii) the vector-autoregression coefficients are estimated by the least squares given the observable factors, and (iii) the other model parameters are estimated via the maximum likelihood estimation given the vector-autoregression coefficients estimates and all principal components. For our model comparison purpose, it is essential to consider the estimation risk. Thus we estimate the model parameters and yield curve factors simultaneously, and use three basis yields suggested by Bansal and Zhou (2002). As the three basis yields are assumed to be observed without errors, the yield factors can be expressed or inverted in terms of those basis yields. As a result, the likelihood function is calculated without Kalman filtering, which substantially reduces the computational burden.

3.1. Likelihood

The yield curve can be expressed as an affine function of the latent factors,

$$R_t(\tau) \approx -\frac{1}{\tau} \ln P_t(\tau) = \frac{a(\tau)}{\tau} + \frac{b(\tau)'}{\tau} \boldsymbol{l}_t. \tag{3.1}$$

For reasons just mentioned, suppose as in Bansal and Zhou (2002) that three yields are measured without any pricing errors. Suppose there are N maturities of interest. Let

$(\bar{\tau}_1, \bar{\tau}_2, \bar{\tau}_3)$ denote the maturities of the three basis bonds, corresponding to short-, mid-, and long-term basis yields, respectively. Since these yields are assumed to be measured without error, we have that

$$\boldsymbol{R}_t = \boldsymbol{a} + \boldsymbol{b}\boldsymbol{l}_t, \tag{3.2}$$

where

$$\begin{aligned} \boldsymbol{a} &= \begin{pmatrix} a(\tau_1)/\tau_1 & a(\tau_2)/\tau_2 & \dots & a(\tau_N)/\tau_N \end{pmatrix}' : N \times 1, \\ \boldsymbol{b} &= \begin{pmatrix} b(\tau_1)/\tau_1 & b(\tau_2)/\tau_2 & \dots & b(\tau_N)/\tau_N \end{pmatrix}' : N \times 3. \end{aligned} \tag{3.3}$$

Along with the relevant and irrelevant macro risk factors, the one-to-one mapping between $\boldsymbol{l}_t$ and $\bar{\boldsymbol{R}}_t$ can be rewritten as

$$\begin{pmatrix} \bar{\boldsymbol{R}}_t \\ \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{a} \\ 0_{m\times 1} \\ 0_{z\times 1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{b} & & \\ & I_m & \\ & & I_z \end{pmatrix} \begin{pmatrix} \boldsymbol{l}_t \\ \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix}. \tag{3.4}$$

On the other hand, $(\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_{N-3})$ are the maturities of the remaining $(N-3)$ non-basis bonds, which we express as

$$\hat{\boldsymbol{R}}_t = \begin{pmatrix} R_t(\hat{\tau}_1) \\ R_t(\hat{\tau}_2) \\ \vdots \\ R_t(\hat{\tau}_{N-3}) \end{pmatrix} = \hat{\boldsymbol{a}} + \hat{\boldsymbol{b}}\boldsymbol{l}_t + e_t,\ e_t \sim \mathcal{N}(0, \Sigma), \tag{3.5}$$

where $\Sigma = diag(\sigma_1^2, \sigma_2^2, .., \sigma_{N-3}^2)$ is a diagonal pricing error variance-covariance and

$$\begin{aligned} \hat{\boldsymbol{a}} &= \begin{pmatrix} a(\hat{\tau}_1)/\hat{\tau}_1 & a(\hat{\tau}_2)/\hat{\tau}_2 & \cdots & a(\hat{\tau}_{N-3})/\hat{\tau}_{N-3} \end{pmatrix}' : (N-3) \times 1, \\ \hat{\boldsymbol{b}} &= \begin{pmatrix} b(\hat{\tau}_1)/\hat{\tau}_1 & b(\hat{\tau}_2)/\hat{\tau}_2 & \cdots & b(\hat{\tau}_{N-3})/\hat{\tau}_{N-3} \end{pmatrix}' : (N-3) \times 3. \end{aligned} \tag{3.6}$$

Under this error structure, the likelihood of $\boldsymbol{y}_t = \left(\hat{\boldsymbol{R}}_t, \bar{\boldsymbol{R}}_t, \boldsymbol{m}_t, \boldsymbol{z}_t\right)$ is

$$p\left(\boldsymbol{Y}|\theta\right) = \prod_{t=1}^{T} p(\boldsymbol{y}_t|\mathcal{F}_{t-1}, \theta), \tag{3.7}$$

where $\boldsymbol{y}_0$ is observed from the data and $\mathcal{F}_{t-1}$ is the observations up to $(t-1)$. The conditional joint density of the observations $p(\boldsymbol{y}_t|\mathcal{F}_{t-1},\theta)$ is obtained from

$$\begin{aligned} p(\boldsymbol{y}_t|\mathcal{F}_{t-1},\theta) &= p\left(\bar{\boldsymbol{R}}_t,\boldsymbol{m}_t,\boldsymbol{z}_t|\mathcal{F}_{t-1},\theta\right)\times p\left(\hat{\boldsymbol{R}}_t|\bar{\boldsymbol{R}}_t,\boldsymbol{m}_t,\boldsymbol{z}_t,\mathcal{F}_{t-1},\theta\right) \\ &= \mathcal{N}(\boldsymbol{f}_t|G\boldsymbol{f}_{t-1},\Omega)\times\left|\bar{\boldsymbol{b}}^{-1}\right|\times\mathcal{N}\left(\hat{\boldsymbol{R}}_t|\hat{\boldsymbol{a}}+\hat{\boldsymbol{b}}\boldsymbol{l}_t,\Sigma\right), \end{aligned} \tag{3.8}$$

given $\boldsymbol{l}_t=\bar{\boldsymbol{b}}^{-1}(\bar{\boldsymbol{R}}_t-\bar{\boldsymbol{a}})$.

3.2. Identification Restrictions

We complete our model by imposing the Nelson-Siegel type of identification restrictions on the yield curve factors $\boldsymbol{l}_t$ so that the model-implied risk premiums on exposure to level, slope, and curvature risks of the yield curves can be derived and estimated. Specifically, such restrictions are as follow:

$$\beta=(1,1,0)' \tag{3.9}$$

and

$$G_{ll}^{\mathbb{Q}}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(-\kappa) & \kappa\exp(-\kappa) \\ 0 & 0 & \exp(-\kappa) \end{bmatrix}. \tag{3.10}$$

Note that $G_{ll}^{\mathbb{Q}}$ only depends on κ. See Christensen et al. (2009) and Christensen et al. (2011) for details on theoretical backgrounds of the arbitrage-free Nelson-Siegel model.

According to Niu and Zeng (2012), $b(\tau)$ can be reduced to

$$\begin{aligned} b(\tau)' &= \beta'\times\sum_{j=0}^{\tau-1}\left(G_{ll}^{\mathbb{Q}}\right)^j \\ &= \left[\ \tau,\ \ \tau-(1-\exp(-\kappa\tau))/\kappa,\ \ (1-\exp(-\kappa\tau))/\kappa-\tau\exp(-\kappa\tau)\ \right]. \end{aligned}$$

Then, the factor loading $b(\tau)'/\tau$ exhibits the form of the dynamic Nelson-Siegel factor loadings

$$\left[\ 1,\ \ 1-(1-\exp(-\kappa\tau))/(\kappa\tau),\ \ (1-\exp(-\kappa\tau))/(\kappa\tau)-\exp(-\kappa\tau)\ \right].$$

Due to this functional form of the factor loadings, the latent factors in $\boldsymbol{l}_t$ can be interpreted as dynamic level, slope, and curvature factors of yield curves.

For computational convenience, we impose two additional restrictions on the model parameters for identification and efficient estimation. First, we assume $\Omega_{lm} = \Omega_{lz} = 0$. This restriction allows $\Omega_{MZ} = \begin{pmatrix} \Omega_{mm} & \Omega_{mz} \\ \Omega_{mz}' & \Omega_{zz} \end{pmatrix}$ to be fast and efficiently estimated via Gibbs sampling without affecting the bond pricing. If this restriction is not imposed, the whole variance-covariance matrix Ω has to be sampled by a Metropolis-Hastings (M-H) algorithm, which is substantially inefficient compared to the Gibbs sampler. This issue is especially crucial when the dimension of the parameter space gets higher as more candidates of macro risk factors are considered. Second, δ in the short rat e dynamics is fixed at the sample mean of the short rates because the short rate is so persistent that δ tends to be inefficiently estimated.

In sum, the parameters to be estimated are $\theta = \{\kappa, \lambda_l, \Omega_{ll}, G, \Sigma, \Omega_{MZ}\}$ and the size of θ is 1 (from κ) + 3 (from λ_l) + 6 (from Ω_{ll}) + $(3 + m + z)^2$ (from G) + $(N - 3)$ (from Σ) + $0.5(m+z)(m+z+1)$ (from Ω_{MZ}). For example, if there are ten risk factors ($m + z = 10$) and twelve maturities of bonds ($N = 12$), the number of parameters for the most general model specification is 266 $(= 1 + 3 + 6 + 169 + 9 + 78)$.

3.3. Prior

For effective model comparison, the prior distributions should be comparable across the different models so that the differences in marginal likelihoods are not merely caused by the differences in priors. At the same time, the prior in any given model should also serve as a suitable regularizer since the parameters lie in a high dimensional parameter space. We develop such a prior distribution for all 1,024 models by incorporating theoretically supported knowledge. Following Chib and Ergashev (2009) and Abbritti et al. (2016) in particular, we specify the prior means based on the belief that (i) the yield curve factors are identified as level, slope, and curvature effects, (ii) the yield curve and term premium are gently upward sloping and concave with maturity, (iii) the term premium are time-invariant, (iv) the macro risk factors are persistent.

We arrive at this prior by sampling a default prior. Given draws from this default prior, we draw the time series of factors from the assumed VAR process, followed by the draws of the yield curve from the measurement equation given the factors. We repeat these steps many times, generating an ensemble of yield curves by time. If these

simulated yield curves are not upward sloping on average for each time point, we adjust the hyperparameters of our default prior, and sample again. This procedure is repeated until the outcomes are consistent with our theoretical prior beliefs. The specific values for the prior means are as follows;

(i) Level, slope, and curvature yield curve factors The factor loadings, $(\bar{\boldsymbol{b}}, \hat{\boldsymbol{b}})$ are solely determined by the value of κ, and the prior mean of κ, $\bar{\kappa} = 0.0672$, is chosen to make the curvature factor loading maximized at 36 months.

(ii) Upward sloping and concave yield curve on average Given $\bar{\kappa}$ (i.e., $b(\tau)$), our prior mean for λ_l, denoted by $\bar{\lambda}_l$, is chosen such that the prior-implied term premium is increasing and concave with maturity, and is not too big for long term bonds. The term premium averaged over time,

$$-\tau^{-1}\sum_{i=1}^{\tau-1}[b(\tau-i)'\lambda_l]$$

is the time-invariant component of the term premium once the Jensen's inequality part is ignored. Our prior mean of the 10-year term premium is about 1.8% in annual terms, which leads to $\bar{\lambda}_l = (-0.0086, -0.1026, 0)'/1200$. Given $(\bar{\kappa}, \bar{\lambda}_l)$. The prior mean of Ω_{ll}, $\bar{\Omega}_{ll} = diag(0.0410, 0.0526, 0.2501)/1200$ is chosen for the unconditionnal mean of the yield curve $\boldsymbol{a} = (a(\tau_1)/\tau_1, a(\tau_2)/\tau_2, \cdots, a(\tau_N)/\tau_N)$ to be increasing and gently concave with maturity.

(iii) Time-invariant term premium The term premium changes over time according to the yield curve factors and relevant macro factors. As we have no prior belief whether the factors have a positive or negative effect on the market price of risks, we assume their impacts are zero on average *a priori*. Suppose that $g_{i,j}$ and $\bar{g}^Q_{i,j}$ are the (i,j)th element of G and $\bar{G}^{\mathbb{Q}}_{ll}$, respectively, and $\bar{g}_{i,j}$ is the prior mean of $g_{i,j}$, where

$$\bar{G}^{\mathbb{Q}}_{ll} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \exp(-\bar{\kappa}) & \bar{\kappa}\exp(-\bar{\kappa}) \\ 0 & 0 & \exp(-\bar{\kappa}) \end{bmatrix}.$$

Then, we have $\bar{g}_{i,j} = \bar{g}^Q_{i,j}$ for all $i, j \leq 3$, and $\bar{g}_{i,j} = 0$ for all $(i,j) \in \{(i,j) | g_{i,j} = (i,j)\text{th element of } G_{lm}\}$. That is, $\Lambda_{ll} = 0_{l\times l}$ and $\Lambda_{lm} = 0_{l\times m}$ at the prior mean.

(iv) Persistence and volatility of macro factors The persistence is prevalent in most macroeconomic variables such as real activity measures, inflation rates, and interest rates. To accommodate this property in the prior setting, we specify the prior mean of the diagonal elements of G as $\bar{g}_{i,i} = 0.9$ for all $i \geq 4$, while the prior means of $(G_{ml}, G_{zl}, G_{zm}, G_{mz})$ and non-diagonal elements of G_{mm} and G_{zz} are assumed to be zero. Given that we use the standardized macro risk factors with zero-mean and unit-variance in annual terms, the prior mean of Ω_{MZ}, denoted by $\bar{\Omega}_{MZ}$, is set to be $(1-0.9^2) \times \boldsymbol{I}_{m+z}$. Finally, the pricing error variance is assumed to follow an inverse-gamma distribution with small prior means.

The prior variances are assumed to be large enough so that we can avoid the possibility of likelihood-prior conflict and allow for the parameters to be sufficiently updated by the information in the data. Given the independence across the blocks of parameters, the prior density $\pi(\theta)$ can be summarized as

$$\pi(\theta) = \pi(\kappa)\,\pi(G)\,\pi(\lambda_l)\,\pi(\Omega_{ll})\,\pi(\Omega_{MZ})\,\pi(\Sigma),$$

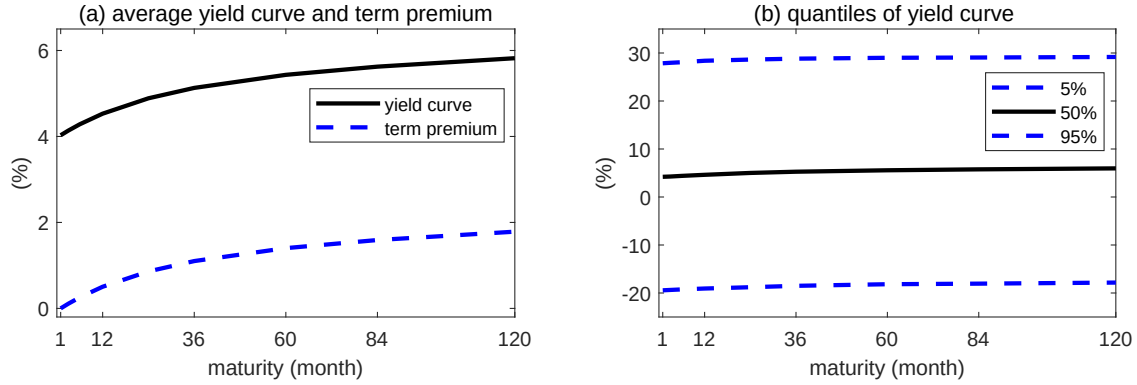
where $\mathcal{IW}(\cdot|a,b)$ is the inverse Wishart density of $p \times p$ covariance matrix with mean $b/(a-p-1)$, $\mathcal{IG}(\cdot|a,b)$ is the inverse gamma density with mean $b/(a-1)$,

$$\begin{aligned}
\pi(\kappa) &= \mathcal{N}(\kappa|\bar{\kappa},\ V_\kappa = 0.01^2), \\
\pi(G) &= \prod\nolimits_{i,j} \mathcal{N}\left(g_{i,j}|\bar{g}_{i,j},\ V_{G,i,j} = 0.3^2\right), \\
\pi(\lambda_l) &= 1200 \times \mathcal{N}(1200 \times \lambda_l | 1200 \times \bar{\lambda}_l,\ V_\lambda = 0.05^2), \\
\pi(\Omega_{ll}) &= 1200^2 \times \mathcal{IW}(1200^2 \times \Omega_{ll} | v_0 = 50,\ \Gamma_{ll,0}^{-1} = (v_0 - 4)\bar{\Omega}_{ll}), \\
\pi(\Omega_{MZ}) &= 1200^2 \times \mathcal{IW}(1200^2 \times \Omega_{MZ} | v_0 = 50,\ \Gamma_{MZ,0}^{-1} = (v_0 - 11)\bar{\Omega}_{MZ}), \\
\text{and } \pi(\Sigma) &= \prod\nolimits_{i=1}^{N-3} \left[1200^2 \times \mathcal{IG}\left(1200^2 \times \sigma_i^2 | \alpha_0/2 = 25,\ \gamma_0/2 = 0.0625\right)\right].
\end{aligned}$$

Figure 2 plots the distribution of the prior-implied yield curve for model which includes 10 macro risk factors. Figure 2 (a) is the yield curve averaged over time computed at the prior mean of the parameters. The prior-implied yield curve is gently upward sloping on average, thus the first moment of the prior yield curve reflects our prior beliefs. Figure 2(b) presents the 5%, 50%, and 95% quantiles of the ergodic prior-implied yield

curve distributions generated from the ergodic distribution of the factors and pricing errors.

Figure 2: Prior-implied Yield Curve This figure plots average yield curve, average term premium, and yield curve distribution based on 300 periods of simulated prior draws generated from the model which includes two relevant macro factors.



3.4. Posterior Sampling

By Bayes' rule, the posterior distribution of θ is

$$\pi(\theta|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\theta)\pi(\theta),$$

where $p(\boldsymbol{Y}|\theta)$ is as given in Equation (3.8). The Markov chain Monte Carlo (MCMC) algorithm implemented to simulate the posterior distribution is comprised of four blocks,

$$\psi = (\kappa, \lambda_l, \Omega_{ll}),\ G,\ \Omega_{MZ},\ \Sigma$$

in which ψ is sampled by M-H algorithm and the rest by Gibbs sampler. The following summarizes the posterior sampling algorithm. For the details of each MCMC step, see Appendix A.

MCMC Sampling Algorithm

Step 0: Initialize the parameters $(\psi^{(0)}, G^{(0)}, \Omega_{MZ}^{(0)}, \Sigma^{(0)})$ and set $g = 1$.

Step 1: At the gth MCMC iteration, draw $\psi^{(g)}$ from $\psi|\boldsymbol{Y}, \psi^{(g-1)}$ as follows:

Step 1.(a): Maximize

$$\ln\{p(\boldsymbol{Y}|\psi, G^{(g-1)}, \Omega_{MZ}^{(g-1)}, \Sigma^{(g-1)}) \times \pi(\psi)\}$$

with respect to ψ to obtain the mode, $\bar{\psi}$, and compute the inverted negative Hessians computed at the mode, $V_{\bar{\psi}}$.

Step 1.(b): Draw a proposal for ψ, denoted by $\psi^{\dagger}$, from the selected multivariate Student-t distribution,

$$\psi^{\dagger} \sim St\left(\bar{\psi}, V_{\bar{\psi}}, 15\right).$$

Step 1.(c): Draw a sample u from uniform distribution over $[0, 1]$. Then, ψ is updated as

$$\begin{aligned} \psi &= \psi^{\dagger} \qquad \text{if } u < \alpha\left(\psi^{(g-1)}, \psi^{\dagger}|\boldsymbol{Y}, G^{(g-1)}, \Omega_{MZ}^{(g-1)}, \Sigma^{(g-1)}\right) \\ \psi &= \psi^{(g-1)} \quad \text{if } u \geq \alpha\left(\psi^{(g-1)}, \psi^{\dagger}|\boldsymbol{Y}, G^{(g-1)}, \Omega_{MZ}^{(g-1)}, \Sigma^{(g-1)}\right) \end{aligned}$$

Step 3: Draw $G^{(g)}$ from $G|\boldsymbol{Y}, \psi^{(g)}, \Sigma^{(g-1)}, \Omega_{MZ}^{(g-1)}$.

Step 4: Draw $\Omega_{MZ}^{(g)}$ from $\Omega|\boldsymbol{Y}, \psi^{(g)}, G^{(g)}, \Sigma^{(g-1)}$.

Step 4: Draw $\Sigma^{(g)}$ from $\Sigma|\boldsymbol{Y}, \psi^{(g)}, G^{(g)}, \Omega_{MZ}^{(g)}$.

Step 5: $g = g + 1$, and go to Step 1 while $g \leq n_0 + n_1$. Then, discard the draws from the first n_0 iterations and save the subsequent n_1 draws.

3.5. Marginal Likelihood

Competing term structure models with different relevant macro risk factors are compared based on their marginal likelihoods. Denoting the prior and posterior probability of model $\mathcal{M}_j$ by $\Pr(\mathcal{M}_j)$ and $\Pr(\mathcal{M}_j|\boldsymbol{Y})$, respectively, the posterior probability of model $\mathcal{M}_j$ is given by

$$\Pr(\mathcal{M}_j|\boldsymbol{Y}) = \frac{\Pr(\mathcal{M}_j) m(\boldsymbol{Y}|\mathcal{M}_j)}{\sum_i \Pr(\mathcal{M}_i) m(\boldsymbol{Y}|\mathcal{M}_i)}.$$

Assuming the uniform prior on the competing models, the posterior probability of the each model is determined by the relative size of the marginal likelihoods. Marginal likelihood is a statistically robust criterion which automatically penalizes the complexity

of models, which implies that a model with more macro risk factors in $\boldsymbol{m}_t$ (i.e. fewer zero restrictions on G_{lz}) is not necessarily selected by the marginal likelihood comparison. We calculate the marginal likelihoods of each model by the method proposed by Chib (1995) and Chib and Jeliazkov (2001). The Chib method starts from the identity

$$\log m(\boldsymbol{Y}) = \log p(\boldsymbol{Y}|\theta^*) + \log \pi(\theta^*) - \log \pi(\theta^*|\boldsymbol{Y}),$$

where θ^* is a point in the parameter space that is a high density point, say the value that maximizes the sum of the log likelihood and log prior functions. The computation of the last term, the posterior ordinate, is by a marginal-conditional decomposition, following Chib (1995).

Specifically, we split the log posterior ordinate into four terms:

$$\begin{aligned}\log \pi(\theta^*|\boldsymbol{Y}) = \log \pi(\psi^*|\boldsymbol{Y}) + \log \pi(\Sigma^*|\boldsymbol{Y}, \psi^*) \\ + \log \pi(\Omega^*_{MZ}|\boldsymbol{Y}, \Sigma^*, \psi^*) + \log \pi(G^*|\boldsymbol{Y}, \Sigma^*, \Omega^*_{MZ}, \psi^*).\end{aligned}$$

Each posterior ordinate can be numerically computed very efficiently by using parallel processing, and the reduced MCMC run algorithms are provided in Appendix B

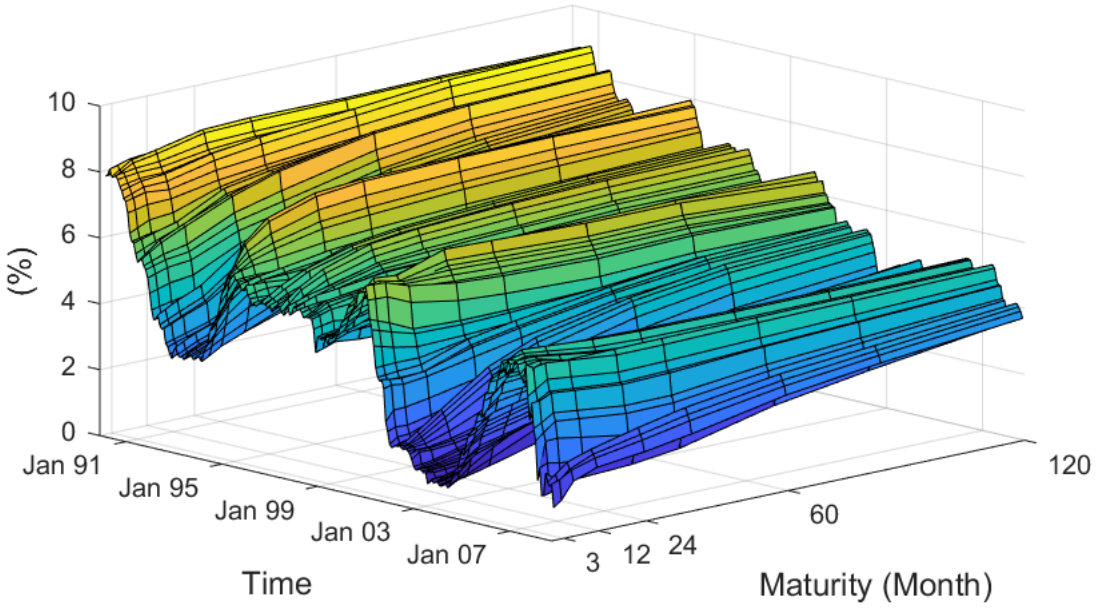
4. Empirical Study

4.1. Data

Our monthly yield curve data for 1, 3, 12, 24, 36, 48, 60, and 120 month maturities are obtained from the constant maturity treasury securities that are provided by the U.S treasury department. The maturities of the basis yields are (3, 36, 120) months. Figure 3 plots the term structure of interest rates, indicating that the yield curve is upward sloping on average and highly autocorrelated. Our data on the 10 macro factors are plotted in Figure 4. It can be seen that all real activity measures are pro-cyclical.

The data used in our empirical analysis range from January 1990 to September 2008, corresponding to the period from the great moderation to the great recession. Following Bauer (2018) and Joslin et al. (2014), we use the data up to the beginning of the financial crisis, because this sample period seems to be the longest recent sample that can be considered as a single regime, not subject to a structural break. As Bauer

Figure 3: Bond Yields This figure plots the monthly time series of government bond yields with 1, 3, 12, 24, 36, 48, 60, and 120 months to maturity from January 1990 to September 2008. Shaded areas indicate recession periods designated by the National Bureau of Economic Research.

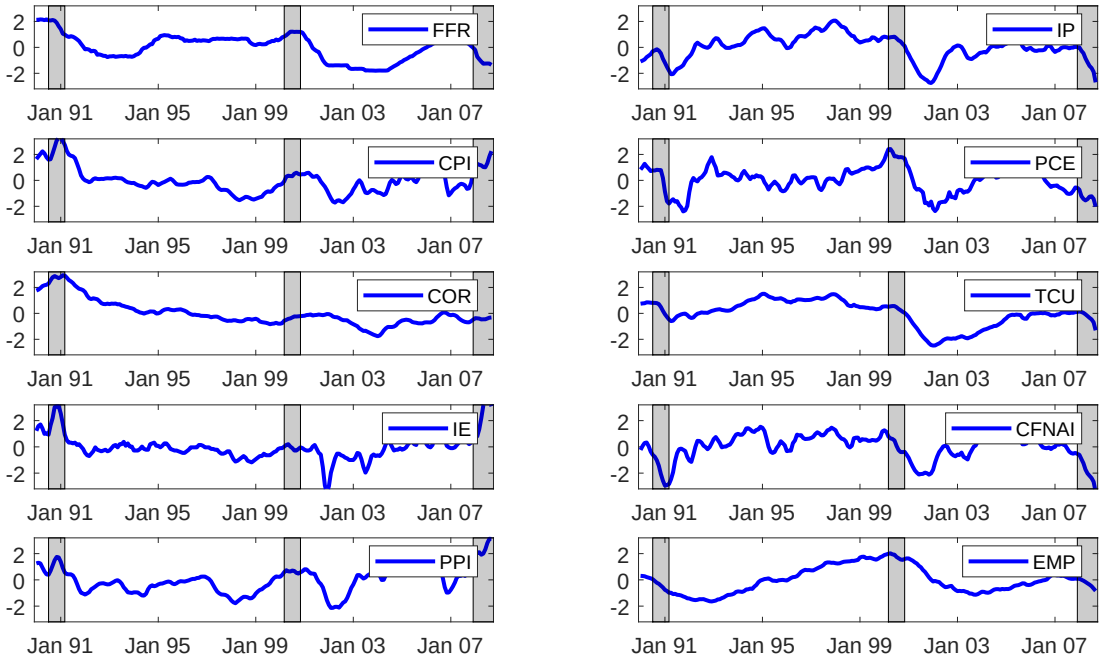


(2018) and Joslin et al. (2014) have already pointed out, it is obvious that the short term bond yields show different dynamics since 2009, and that the more recent data is highly likely to cause biases. In Table 3, we report the sample correlations among the macroeconomic variables and the first three principal components of the bond yields. Most of the correlations seem to be high, and each macroeconomic variables might be seen as potentially informative for the U.S. yield curve.

4.2. Relevant Macro Factor Selection

Instead of asserting the identity of relevant macro risk factors *a priori*, our model comparison approach finds the best supported relevant factors by confronting the 2^{10} possible model specifications with the data, assessing the relative worth of each model in terms of the marginal likelihood. All the marginal likelihoods are computed from 5,000 draws of the MCMC algorithm, collected after an initial burn-in period of 1,000 iterations. The findings, already presented in Section 1, Table 2, show that the best model, with CFNAI as the sole relevant factor, dominates the other models in terms of the standard Jeffreys' scale. Models with inflation and federal funds rate factors are

Figure 4: Macro Factors This figure plots the 10 macro factors mentioned in Table 1 for the sample period from January 1991 to September 2008. Shaded areas indicate recession periods designated by the National Bureau of Economic Research.



strictly dominated by this best model. This result is perhaps due to the fact that the term premium dynamics exhibit strong counter-cyclical trends, which are well explained by the CFNAI factor.

Our finding that the inflation factors are not relevant has some parallel to findings in the literature. For instance, Evans (2003) finds evidence of a downward sloping inflation risk premium (IRP), but the premiums are negative for most maturities. Hoerdahl et al. (2008) argue that the IRP is positive and small around one-year maturity and essentially zero for all other maturities. Joyce et al. (2010) reports that the IRP is inverted and hump-shaped across maturities and has been negative for long maturities on average. Moreover, IRP dynamics differ across countries, as aversion to inflation risks and inflation predictability appears to be country-specific, as documented by Buraschi and Jiltsov (2005), D'Amico et al. (2014) and Hordahl and Tristani (2012).

Table 3: Sample Correlations between Yield Curve and Macro Factors This table is a summary of the sample correlations between the macro factors and the first three principal components of the bond yields (PC_1, PC_2, and PC_3). The sample periods are from January 1990 to September 2008.

	PC_1	PC_2	PC_3	CPI	FFR	IP	PCE	TCU	CFNAI	CORE	EMP	PPI	IE
PC_1	1.00												
PC_2	-0.32	1.00											
PC_3	-0.18	-0.19	1.00										
CPI	0.25	-0.22	0.25	1.00									
FFR	0.71	-0.86	0.05	0.36	1.00								
IP	0.31	-0.44	-0.28	-0.38	0.37	1.00							
PCE	0.25	-0.40	-0.18	0.07	0.34	0.56	1.00						
TCU	0.58	-0.72	-0.25	-0.03	0.75	0.82	0.52	1.00					
CFNAI	0.17	-0.20	-0.49	-0.45	0.13	0.83	0.53	0.66	1.00				
CORE	0.18	-0.12	0.36	0.76	0.26	-0.59	-0.33	-0.22	-0.63	1.00			
EMP	0.45	-0.62	-0.04	-0.18	0.67	0.37	0.34	0.50	0.12	-0.26	1.00		
PPI	0.20	-0.22	0.14	0.84	0.30	-0.12	0.38	0.12	-0.18	0.39	-0.13	1.00	
IE	0.38	-0.40	0.12	0.81	0.51	0.01	0.26	0.31	-0.08	0.61	-0.12	0.75	1.00

Table 4: Posterior summary for the risk parameters in model $\mathcal{M}_{CKXL}$ This summary is based on 20,000 MCMC draws beyond a burn-in of 1,000. The numbers are the posterior means, and the standard errors are in the parenthesis.

$10^3 \times \lambda_l$	$10^2 \times \Lambda_{ll}$			$10^2 \times \Lambda_{lm}$
-0.007	-1.221	-1.492	0.056	-1.064
(0.001)	(0.716)	(0.788)	(0.949)	(1.574)
-0.086	0.488	-0.839	-0.150	8.540
(0.015)	(0.680)	(0.764)	(0.922)	(1.846)
-0.013	-0.067	0.352	-0.634	4.229
(0.015)	(0.649)	(0.754)	(0.952)	(4.086)

4.3. The Source of Marginal Likelihood Improvement

In order to investigate the role of CFNAI and the source of the marginal likelihood improvement, we summarize the marginal posterior distributions of the risk parameters in the best model, denoted by $\mathcal{M}_{CKXL}$, in Table 4. The effect of CFNAI on the market price of risks is captured by $\Lambda_{lm}(= G_{lm})$. As shown in this table, the impact of CFNAI on the second yield curve factor risk is positive, with a small posterior standard deviation. This implies that this factor has a negative effect on the term premium, and drives the counter-cyclical behavior of the term premium.

Suppose that $\boldsymbol{R}$, $\boldsymbol{M}$, and $\boldsymbol{Z}$ denote $\{\boldsymbol{R}_t\}_{t=1}^T$, $\{\boldsymbol{m}_t\}_{t=1}^T$, and $\{\boldsymbol{z}_t\}_{t=1}^T$, respectively. Given that the marginal likelihood is defined as the joint density of the yields and macro risk factors, marginalized over the parameters, it is useful to consider the source

of the marginal likelihood gain from incorporating relevant macro risk factors. Notice that $\log lik(\boldsymbol{R}, \boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}, \theta)$ can be decomposed into $\log lik(\boldsymbol{R}|\boldsymbol{M}, \boldsymbol{Z}, \mathcal{M}, \theta)$ and $\log lik(\boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}, \theta)$. Then, the log likelihood gain of the best model relative to the benchmark is given by

$$\log lik(\boldsymbol{R}, \boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}_{CKXL}, \theta^*) - \log lik(\boldsymbol{R}, \boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}_0, \theta_0^*),$$

where θ^* is the posterior mode of the model $\mathcal{M}_{CKXL}$ and θ_0^* is equal to θ^* except $G_{lm} = 0$. Because $\log lik(\boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}_{CKXL}, \theta^*) = \log lik(\boldsymbol{M}, \boldsymbol{Z}|\mathcal{M}_0, \theta_0^*))$, the likelihood gain is equal to ($\log lik(\boldsymbol{R}|\boldsymbol{M}, \boldsymbol{Z}, \mathcal{M}_{CKXL}, \theta^*)$ - $\log lik(\boldsymbol{R}|\boldsymbol{M}, \boldsymbol{Z}, \mathcal{M}_0, \theta_0^*)$), whose value is found to be 19.2 and it is close to the log Bayes factor. Therefore, the marginal likelihood gain comes mostly from the likelihood rather than the prior information. More importantly, the log likelihood gain is obtained mostly from the improved fit of the yield curve, rather than the fit of the macro factors.

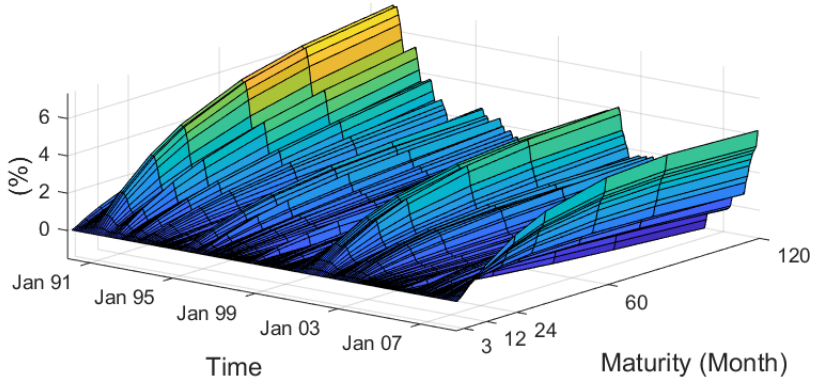
4.4. Term Premium and Market Expectation

Using the analytical expression for the term premium in Equation (2.11), we report on our posterior estimates of the term premium and (EH – risk-free short rate) dynamics estimated from the best model. In panel (a) of Figure 5 we plot the posterior estimates of the term structure of term premium dynamics. Note first that the term premiums vary considerably over time and that these variations are higher for the longer maturities. In particular, the term premium of the 10-year bond tends to be negative during recessions, perhaps because of the flight-to-safety effect. The expectations hypothesis of the term structure of interest rates states that the term premium is constant. However, it is clear that such a hypothesis is not supported by our term premium dynamics, in conformity with Campbell and Shiller (1991).

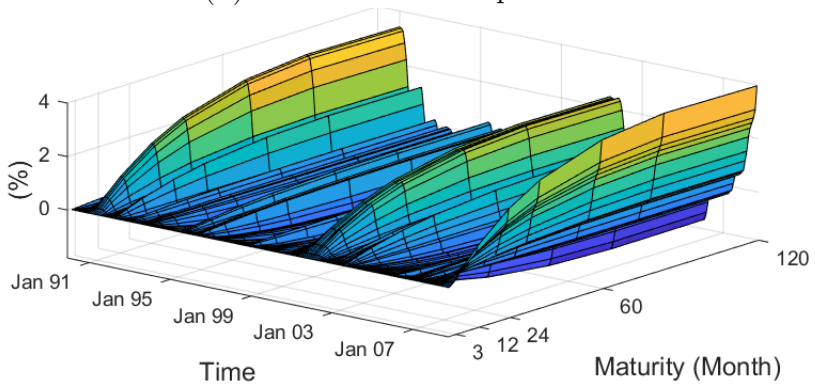
On the other hand, the panels (b) and (c) of Figure 5 displays the time-varying components of the term premiums driven by the yield curve factors and macro risk factors, respectively, based on Equations (2.12) and (2.13). This figure quantifies the roles of the two types of the factors in determining the term premium. It can be seen that the yield curve factor component has been decreasing over time while the macro risk factor component has fluctuated counter-cyclically. Our finding indicates that the

Figure 5: Term structure of term premiums The panel (a) plots the term structure of the term premiums estimated from the best model, $\mathcal{M}_{\mathcal{CKXL}}$. The panels (b) and (c) display the time-varying components of the term premium driven by CFNAI and the yield curve factors, respectively.

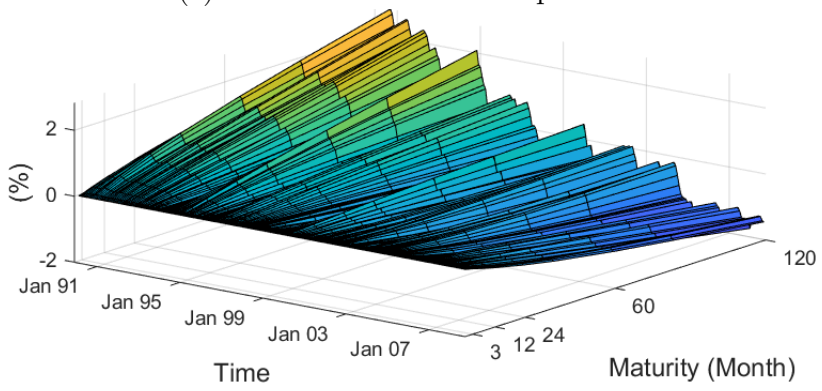
(a) Term premium



(b) Macro factor component



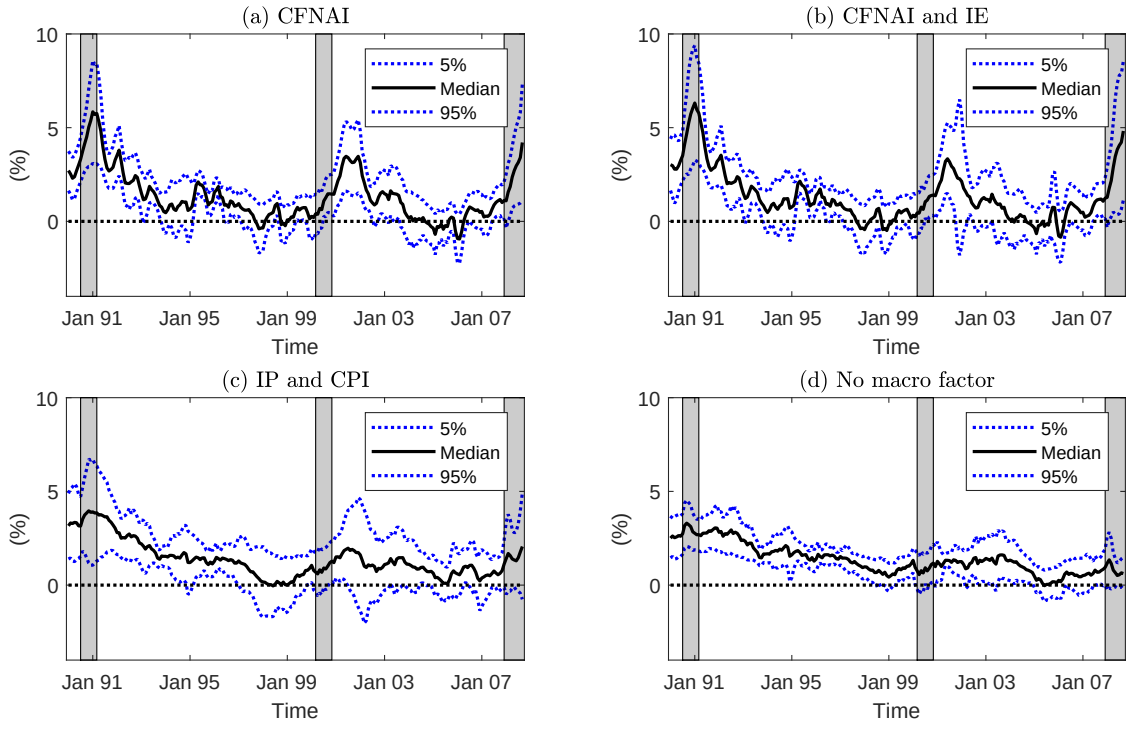
(c) Yield curve factor component



counter-cyclical term premiums do not seem to be sufficiently explained by the level, slope, and curvature factors. The counter-cyclical behavior of the term premium arises because bond investors require larger compensation for bearing risks during recessions.

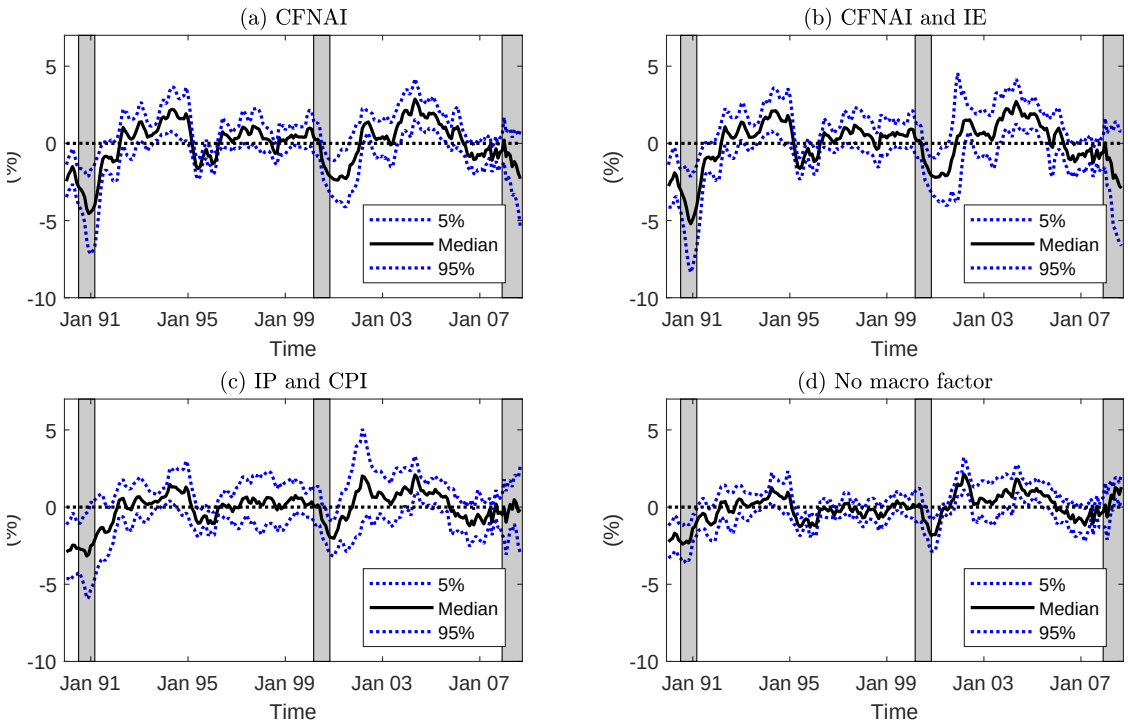
Another reason for the counter-cyclical term premium may stem from asymmetric monetary policy. As demonstrated in Kung (2015), more aggressive inflation targeting reduces nominal risks and lowers the average nominal term spreads. In recessions, inflation targeting tends to be less aggressive and credible, which heightens the nominal risks and results in a higher term premium, and the real activity risk factor enables us to capture such increase of the term premium contributing to the steepening of the yield curve. Conversely, the real activity risk factor contributes to a flattening of the yield curve by decreasing the term premium. Therefore, it is important to incorporate the information of the real activity measure that can help properly detect the changes in risk premia.

Figure 6: Term premium: 5-year This figure plots the posterior median of the 5-year term premiums over time. The dotted lines are 90 percent credibility intervals.



As a comparison with different macro risk factor choices, Figures 6 and 7 plot the time

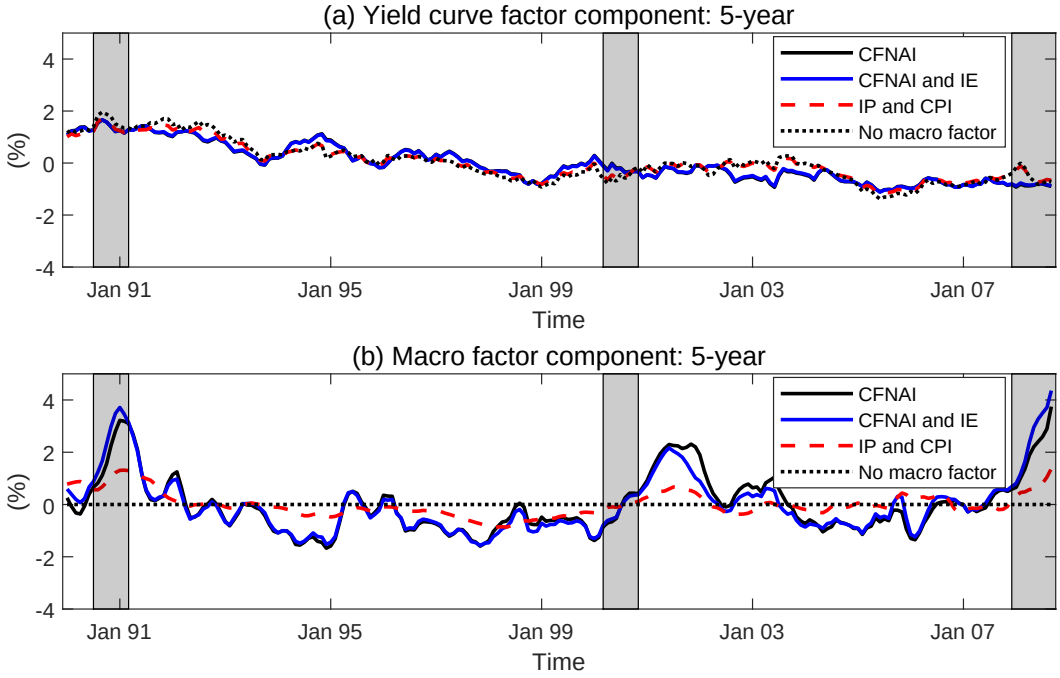
Figure 7: Market expectation: 5-year This figure plots the posterior median of the 5-year market expectations minus short rate over time. The dotted lines are 90 percent credibility intervals.



serious of the term premiums and (EH – risk-free short rate) of 5-year bond obtained from four different relevant macro risk factor combinations. Our selection produces more volatile outputs than no relevant macro risk factor or (IE, CPI) does. More importantly, the posterior estimates are remarkably similar between CFNAI and (CFNAI, IE), which proves that IE is redundant and its effect on the market price of risk is negligible even though it is included in $\boldsymbol{m}_t$. This is also confirmed by Figure 8, which displays the yield curve factor and macro risk factor components of the 5-year term premium. The yield curve factor component is very similar across different macro risk factor combinations while the macro risk factor component shows the sensitivity to the chosen relevant factors.

Finally, to show the importance of the macro risk factor selection in inferencing the EH, we plot the term structure of the EH minus risk-free short rate in January 2001 and January 2004 in Figures 9 and 10, respectively. During the recession in January

Figure 8: Time-varying components of the five-year term premium This figure plots the time series of the time-varying components of the five-year bond driven by the yield curve factors and relevant macro factor. The relevant macro factor are CFNAI. Shaded areas indicate the NBER recession periods and the sample periods are from January 1990 to September 2008.



2001, the EH minus risk-free short rate is negative, which reflects the market belief that the central bank will lower the short-term interest rate. The strength of the belief is measured by the 90% credibility intervals. In contrast, the posterior distributions of the EH minus risk-free short rate during the recovery are strongly positive and different across the macro risk factors. This also shows that a deficient choice of macro risk factors can yield misleading estimates of the EH.

5. Conclusion

In this paper, we introduce a Bayesian framework for selecting relevant unspanned macro factors in a macro-finance affine term structure model. We show that the models with different sets of relevant macro factors arise from different zero restrictions on the vector-autoregressive coefficients. By comparing the various restricted models that emerge in this way, we are able to ascertain the identity of the best supported relevant macro risk factors. To implement this framework in practice, we propose efficient MCMC

Figure 9: Term structure of market expectation - short rate: January 2001 The solid line is the posterior estimates of the term structure of (EH – risk-free short rate) and the dashed lines are the 90 percent credibility intervals. A negative value indicates that a decline in the short rate is expected over the next five years.

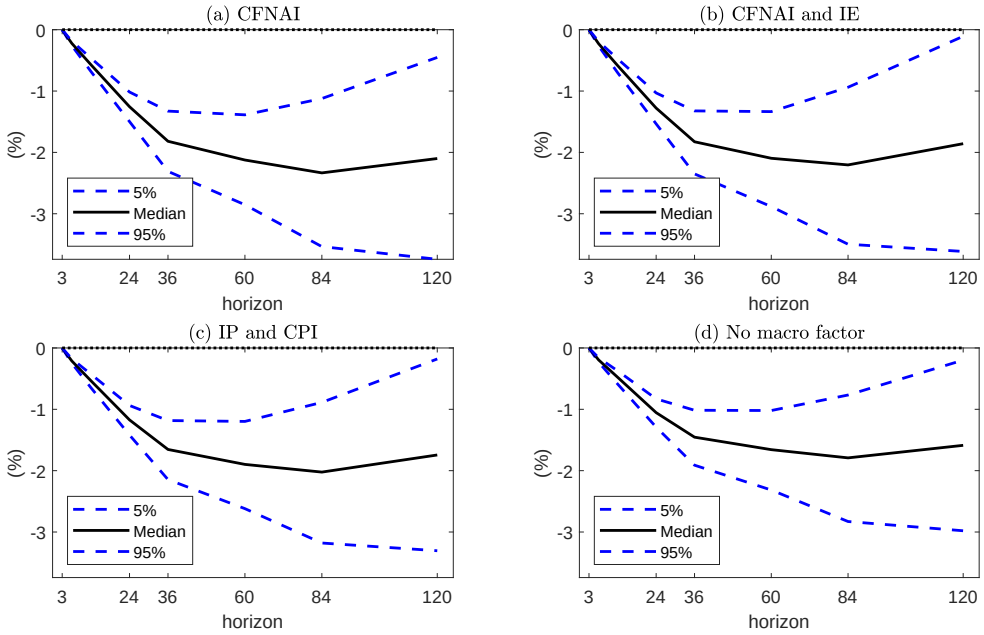
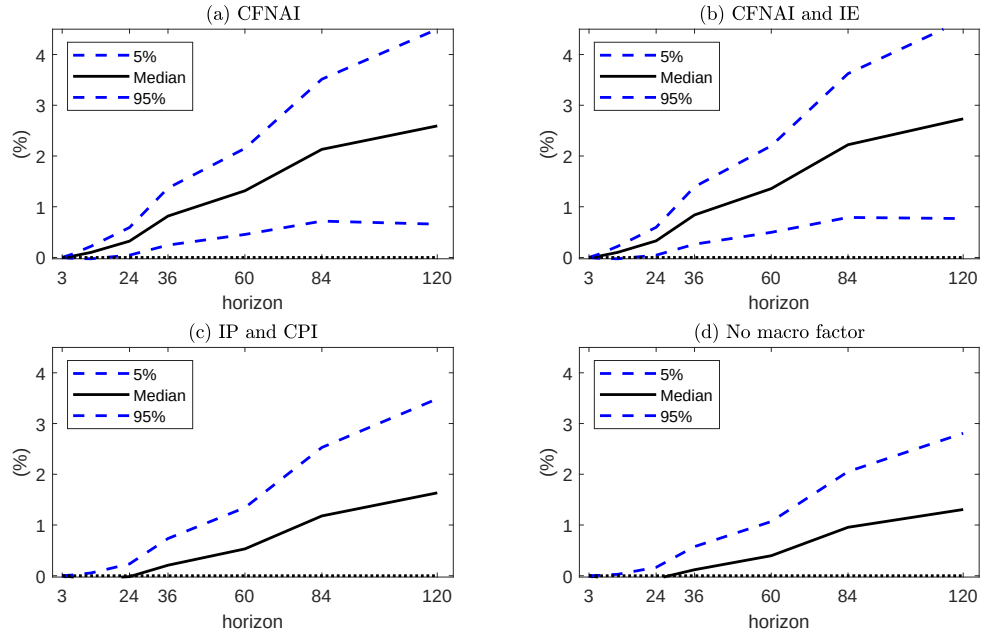


Figure 10: Term structure of market expectation - short rate: January 2004 The solid line is the posterior estimates of the term structure of (EH – risk-free short rate) and the dashed lines are the 90 percent credibility intervals. A positive value indicates that an increase in the short rate is expected over the next five years.



algorithms for estimation and marginal likelihood computation. Our empirical analysis based on the U.S. yield curve data suggests that the Chicago Fed National Activity Index is the sole relevant macro risk factor in our pool of 10 macro factors.

We conclude by pointing out possible extensions of our model and framework. First, it would be interesting to allow for change-points in yield curve dynamics, as in Chib and Kang (2013), with the aim of letting the number and components of the relevant macro risk factors change over time. Second, it would be worthwhile to also allow the factor shock variance-covariance matrix to change over time. Both extensions are challenging, and are left for future research.

References

Abbritti, M., Gil-Alana, L. A., Lovcha, Y., and Moreno, A. (2016), "Term Structure Persistence," *Journal of Financial Econometrics*, 14(2), 331–352.

Ang, A. and Piazzesi, M. (2003), "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50, 745–787.

Bansal, R. and Zhou, H. (2002), "Term structure of interest rates with regime shifts," *Journal of Finance*, 57(5), 1997–2043.

Bauer, M. (2018), "Restrictions on Risk Prices in Dynamic Term Structure Models," *Journal of Business and Economic Statistics,*, 36(2), 196–211.

Bauer, M. and Hamilton, J. (2018), "Robust Bond Risk Premia," *Review of Financial Studies*, 31(2), 399–448.

Bernanke, B. S., Reinhart, V. R., and Sack, B. P. (2004), "Monetary Policy Alternatives at the Zero Bound: An Empirical Assessment," *Brookings Papers on Economic Activity*, 2, 1–78.

Bikbov, R. and Chernov, M. (2008), "Monetary Policy Regimes and the Term Structure of Interest Rates," *CEPR Discussion Papers 7096*.

Buraschi, A. and Jiltsov, A. (2005), "Inflation Risk Premia and the Expectations Hypothesis," *Journal of Financial Economics*, 75(2), 429–490.

Campbell, J. and Shiller, R. (1991), "Yield Spreads and Interest Rates movements," *Review of Financial Studies*, 58, 495–514.

Chib, S. (1995), "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. and Ergashev, B. (2009), "Analysis of multi-factor affine yield curve Models," *Journal of the American Statistical Association*, 104(488), 1324–1337.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm," *American Statistician*, 49, 327–335.

Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

Chib, S. and Kang, K. H. (2013), "Change Points in Affine Arbitrage-free Term Structure Models," *Journal of Financial Econometrics*, 11(2), 302–334.

Chib, S. and Ramamurthy, S. (2010), "Tailored randomized-block MCMC methods for analysis of DSGE models," *Journal of Econometrics*, 155(1), 19–38.

Chib, S., Shin, M., and Simoni, A. (2018), "Bayesian Estimation and Comparison of Moment Condition Models," *Journal of the American Statistical Association*, 113, 1656–1668.

Chib, S. and Zeng, X. (2018), "Which Factors are Risk Factors in Asset Pricing? A Model Scan Framework," *Journal of Business and Economic Statistics*, in press.

Chib, S., Zeng, X., and Zhao, L. (2019), "On Comparing Asset Pricing Models," *The Journal of Finance*, in press.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2009), "An arbitrage-free generalized Nelson-Siegel term structure model," *Econometrics Journal*, 12(3), 33–64.

— (2011), "The affine arbitrage-free class of Nelson-Siegel term structure models," *Journal of Econometrics*, 164, 4–20.

Cochrane, J. H. and Piazzesi, M. (2008), "Decomposing the Yield Curve," *Unpublished manuscript*.

Duffee, G. (2011), "Information in (and not in) the term structure," *Review of Financial Studies*, 57, 2895–2934.

D'Amico, S., Kim, D. H., and Wei, M. (2014), "Tips from TIPS: the Informational Content of Treasury Inflation-Protected Security Prices," *Finance and Economics Discussion Series 2014-24*, 1–88.

Evans, M. D. D. (2003), "Real risk, inflation risk, and the term structure," *Economic Journal*, 113, 345–389.

Hoerdahl, P., Tristani, O., and Vestin, D. (2008), "The yield curve and macroeconomic dynamics," *Economic Journal*, 118, 1937–1970.

Hordahl, P. and Tristani, O. (2012), "The Inflation Risk Premium in the Term Structure of Interest Rates," *Journal of the European Economic Association*, 10, 634–657.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014), "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finace*, 69(3), 1197–1233.

Joyce, M. A., Lildholdt, P., and Sorensen, S. (2010), "Extracting inflation expectations and inflation risk premia from the term structure: A joint model of the UK nominal and real yield curves," *Journal of Banking and Finance*, 34(2), 281–294.

Kung, H. (2015), "Macroeconomic linkages between monetary policy and the term structure of interest rates," *Journal of Financial Economics*, 115, 42–57.

Li, C., Niu, L., and Zeng, G. (2012), "A Generalized Arbitrage-Free Nelson-Siegel Term Structure Model with Macroeconomic Fundamentals," *manuscript*, 1–41.

Ludvigson, S. C. and Ng, S. (2009), "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22(12), 5027–5067.

Niu, L. and Zeng, G. (2012), "The Discrete-Time Framework of Arbitrage-Free Nelson-Siegel Class of Term Structure Models," *manuscript*, 1–68.

Appendix A. MCMC Simulation

Here we discuss the posterior sampling algorithm in the order of ψ, G, Ω_{MZ}, and Σ.

Appendix A.1. Sampling $\psi = (\kappa, \lambda_l, \Omega_{ll})$

Given the target distribution, the efficacy of the MH sampling depends on the way of constructing proposal distributions. In each MCMC cycle, sampling ψ consists of two steps. The first step is to construct a Student-t proposal distribution using the output of stochastic optimization given the other parameters. The second step is to draw from the proposal distribution and update the block based on the M-H algorithm. We discuss each of the steps in details as follows.

Proposal Distribution via Stochastic Optimization. The first stage is to construct a proposal distribution of ψ by finding the global mode $\bar{\psi}$ of its full conditional density and the inverted negative Hessians $V_{\bar{\psi}}$ computed at the mode:

$$\bar{\psi} = \underset{\psi}{\mathsf{argmax}} \left[\log p(\boldsymbol{Y}|\theta_{-\psi}, \psi) + \log \pi(\psi)\right] \text{ subject to } \psi \in \mathcal{R},$$

and

$$V_{\bar{\psi}} = \left(-\frac{\partial^2 \log\{p(\boldsymbol{Y}|\bar{\psi}, \theta_{-\psi})\pi(\bar{\psi})\}}{\partial\psi\partial\psi'} \right)^{-1}, \tag{A.1}$$

where $\theta_{-\psi}$ is the parameters in θ except ψ. The key idea of our proposal distribution is to approximate the full conditional distribution of the block by a multivariate Student-t distribution. The first and second moments of the Student-t distributions are obtained from the global mode and the Hessian computed at the global mode. The Student-t distributions are informative enough to approximate the full conditional distribution of the parameters when the posterior surface is irregular. In addition, with the Student-t distribution, rather than normal distribution, larger moves arise, which helps in the exploration of the posterior density.

In order to find the mode satisfying the restrictions on the parameters, we utilize a simulated annealing algorithm that is combined with the Newton-Raphson method. We refer to this optimizer as the SA-Newton method. Simulated annealing is a stochastic algorithm for finding the global mode. It is particularly useful when the objective function is suspected to have multiple modes. In our search for the global mode, we start with a high initial temperature and a relatively small temperature reduction factor, which means that the temperature reduction is fast. Specifically, the initial temperature parameter and the temperature reduction factor are set to be 5 and 0.01, respectively. Then, the probability of accepting a point with a lower function value (proportional to the current temperature) is high, which helps the search process to traverse quickly to regions with higher function values. Once the simulated annealing stage is complete, our SA-Newton method switches to the Newton-Raphson method to locate the mode.

Our SA-Newton method has been extensively tested and performs well in our affine models. We have found that local optimizers perform poorly and remain stuck near a local mode of the likelihood function. In our implementation of the SA-Newton method, which has to be employed repeatedly across blocks of parameters, and across the MCMC iterations, it is important to make use of the power of the SA steps efficiently. We do this by setting the total number of simulated annealing stages to be two, so the search for the global mode through the SA-Newton method is not onerous. In our Matlab toolbox, it is possible to vary the SA design parameters, including the number of stages, initial temperature, and temperature reduction factor. Further details of the SA formulation in the context of the TaRB-MH algorithm are given in Chib and Ramamurthy (2010).

M-H Move. Given the mode and Hessian at the mode, we now draw a proposal value, denoted by $\psi^\dagger$, from the multivariate Student-t distribution with $\nu = \mathsf{15}$ degrees of freedom,

$$\psi^\dagger \sim St\left(\bar{\psi}, V_{\bar{\psi}}, \mathsf{15}\right).$$

Then, given the current value $\psi^{(g-1)}$, this proposal value is accepted with the usual M-H probability of the move (Chib and Greenberg, 1995),

$$\alpha\left(\psi^{(g-1)}, \psi^{\dagger}|\boldsymbol{Y}, \theta_{-\psi}\right) = \min\left\{1, \frac{p\left(\boldsymbol{Y}|\theta_{-\psi}, \psi^{\dagger}\right)\pi(\psi^{\dagger})}{p\left(\boldsymbol{Y}|\theta_{-\psi}, \psi^{(g-1)}\right)\pi(\psi^{(g-1)})}\frac{St\left(\psi^{(g-1)}|\bar{\psi}, V_{\bar{\psi}}, 15\right)}{St\left(\psi^{\dagger}|\bar{\psi}, V_{\bar{\psi}}, 15\right)}\right\}. \tag{A.2}$$

If the proposal is accepted, then $\psi^{(g)} = \psi^{\dagger}$. Otherwise, $\psi^{(g)} = \psi^{(g-1)}$. Note that if the proposal value does not satisfy the restrictions on the parameters, then it is immediately rejected because its prior density $\pi(\psi)$ is zero.

Appendix A.2. Sampling G

Recall that $G_{lz} = 0_{l\times z}$ is imposed as it relates the irrelevant macro risk factors to the yield curve factors. Using the same $\boldsymbol{l}_t$ as in sampling Σ, we construct $\boldsymbol{f}_t = (\boldsymbol{l}_t', \boldsymbol{m}_t', \boldsymbol{z}_t')'$ and $\boldsymbol{F}_{t_0:t_1} = (\boldsymbol{f}_{t_0}, \boldsymbol{f}_{t_0+1}, \cdots, \boldsymbol{f}_{t_1})'$. With G_0 and V_{G_0} being the corresponding prior mean and variance matrix of G, respectively, G is updated from its full conditional distribution,

$$vec(G)|\boldsymbol{Y}, \theta_{-G} \sim \mathcal{N}(vec(G_1), V_{G_1}), \tag{A.3}$$

where

$$V_{G_1} = \left(V_{G_0}^{-1} + \Omega^{-1} \otimes (\boldsymbol{F}_{1:T-1}' \times \boldsymbol{F}_{1:T-1})\right)^{-1}$$

and

$$vec(G_1) = V_{G_1} \times \left[V_{G_0}^{-1} \times vec(G_0) + (\Omega^{-1} \otimes \boldsymbol{F}_{1:T-1})' \times vec(\boldsymbol{F}_{2:T})\right].$$

Appendix A.3. Sampling Ω_{MZ}

The full conditional of Ω_{MZ} is also tractable, which follows the inverse Wishart distribution

$$1200^2 \times \Omega_{MZ} \sim \mathcal{IW}(v_{MZ,1}, \Gamma_{MZ,1}),$$

where

$$\begin{aligned}
\boldsymbol{l}_t = \bar{\boldsymbol{b}}^{-1}(\bar{\boldsymbol{R}}_t - \bar{\boldsymbol{a}}),\ \boldsymbol{x}_t = I_{m+z} \otimes \boldsymbol{f}_t',\ v_{MZ,1} = v_{MZ,0} + T,\\
G_{MZ} = \begin{pmatrix} G_{ml} & G_{mm} & G_{mz} \\ G_{zl} & G_{zm} & G_{zz} \end{pmatrix},\ u_t = \begin{pmatrix} \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix} - \boldsymbol{x}_{t-1} \times vec\left((G_{MZ})'\right),\ \text{and}\\
\Gamma_{MZ,1} = \left[\Gamma_{MZ,0}^{-1} + 1200^2 \times \sum\nolimits_{t=1}^{T} u_t u_t'\right]^{-1}.
\end{aligned}$$

Appendix A.4. Sampling Σ

Since the full conditional distribution of the error variance is tractable, Σ is updated by the inverse gamma distributions,

$$1200^2 \times \sigma_i^2 | \boldsymbol{Y}, \theta \sim \mathcal{IG}(\alpha_1/2,\ \gamma_{i,1}/2), \tag{A.4}$$

where $\boldsymbol{l}_t = \bar{\boldsymbol{b}}^{-1}(\bar{\boldsymbol{R}}_t - \bar{\boldsymbol{a}})$, $\alpha_1 = \alpha_0 + T$ and

$$\gamma_{i,1} = \gamma_0 + 1200^2 \times \sum_{t=1}^{T} \left(\hat{R}_t(\hat{\tau}_i) - \hat{a}(\hat{\tau}_i)/\hat{\tau}_i - \left(\hat{b}(\hat{\tau}_i)/\hat{\tau}_i \right) \times \boldsymbol{l}_t \right)^2$$

for $i = 1, 2, \cdots, N-3$.

Appendix B. Marginal Likelihood Calculation

Appendix B.1. Posterior Ordinates

The first ordinate $\pi(\psi^*|\boldsymbol{Y})$ is obtained by

$$\pi(\psi^*|\boldsymbol{Y}) = \frac{\int \alpha\left(\psi, \psi^*\right) \times q(\psi^*|\psi) \times \pi(\psi|\boldsymbol{Y}) d\psi}{\int \alpha\left(\psi^*, \psi\right) \times q\left(\psi|\psi^*\right) d\psi},$$

which is a rearrangement of the reversibility condition for M-H sampling. Given $(\boldsymbol{Y}, \theta^*)$ and the posterior draws $\{\theta^{(g)}\}_{g=1}^{n_1}$, it can be numerically approximated by

$$\pi(\psi^*|\boldsymbol{Y}) \simeq \frac{\sum_{g=1}^{n_1} \alpha\left(\psi^{(g)}, \psi^*\right) \times q(\psi^*|\psi^{(g)})}{\sum_{j=1}^{n_1} \alpha\left(\psi^*, \hat{\psi}^{(j)}\right)},$$

where $q(\psi^*|\psi^{(g)}) = St\left(\psi^*|\bar{\psi}, V_{\bar{\psi}}, 15\right)$ for the numerator and $\hat{\psi}^{(j)}$ is drawn from the proposal distribution, $St\left(\bar{\psi}, V_{\bar{\psi}}, 15\right)$.

The second ordinate $\pi(\Sigma^*|\boldsymbol{Y}, \psi^*)$ can be computed by

$$\begin{aligned}
\pi(1200^2 \times \Sigma^*|\boldsymbol{Y}, \psi^*) &= \int \pi(1200^2 \times \Sigma^*, G, \Sigma, \Omega_{MZ}|\boldsymbol{Y}, \psi^*) d(G, \Sigma, \Omega_{MZ}) \\
&= \int \pi(1200^2 \times \Sigma^*|\boldsymbol{Y}, G, \Sigma, \Omega_{MZ}, \psi^*) \pi(G, \Sigma, \Omega_{MZ}|\boldsymbol{Y}, \psi^*) d(G, \Sigma, \Omega_{MZ}).
\end{aligned}$$

The usual way to approximate the above quantity, $\{G^{(g)}, \Omega_{MZ}^{(g)}, \Sigma^{(g)}\}_{g=1}^{n_1}$ is simulated from $G, \Omega_{MZ}, \Sigma|\boldsymbol{Y}, \psi^*$ by reduced MCMC runs. For each g, it can be numerically ap-

proximated by

$$\pi(1200^2 \times \Sigma^* | \boldsymbol{Y}, \psi^*) \approx \frac{1}{n_1} \sum_{g=1}^{n_1} \pi\left(1200^2 \times \Sigma^* | \boldsymbol{Y}, G^{(g)}, \Sigma^{(g)}, \Omega_{MZ}^{(g)}, \psi^*\right),$$

where $\pi\left(1200^2 \times \Sigma^* | \boldsymbol{Y}, G^{(g)}, \Sigma^{(g)}, \Omega_{MZ}^{(g)}, \psi^*\right) = \prod_{i=1}^{N-k} \mathcal{IG}\left(1200^2 \times \sigma_i^{*2} | \alpha_1/2, \gamma_{i,1}^{(g)}/2\right)$. However, since $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ are a function of ψ, not of (G, Ω_{MZ}), it can be immediately obtained that $\pi\left(1200^2 \times \Sigma^* | \boldsymbol{Y}, G^{(g)}, \Sigma^{(g)}, \Omega_{MZ}^{(g)}, \psi^*\right) = \pi\left(1200^2 \times \Sigma^* | \boldsymbol{Y}, \psi^*\right)$ only given ψ^* without reduced MCMC runs.

Analogously, the third ordinate $\pi(\Omega_{MZ}^* | \boldsymbol{Y}, \Sigma^*, \psi^*)$ is computed by

$$\begin{aligned}
\pi(\Omega_{MZ}^* | \boldsymbol{Y}, G^*, \Sigma^*) &= \int \pi(\Omega_{MZ}^*, G | \boldsymbol{Y}, \Sigma^*, \psi^*) dG \\
&= \int \pi(\Omega_{MZ}^* | \boldsymbol{Y}, G, \Sigma^*, \psi^*) \pi(G | \boldsymbol{Y}, \Sigma^*, \psi^*) dG.
\end{aligned}$$

For each $\{\Omega_{MZ}^{(g)}, G^{(g)}\}_{g=1}^{n_1}$ simulated by reduced MCMC runs, $\bar{\boldsymbol{a}}^{(g)}$ and $\bar{\boldsymbol{b}}^{(g)}$ can be calculated and so are the following:

$$\begin{aligned}
\boldsymbol{l}_t^{(g)} &= \bar{\boldsymbol{b}}^{(g)-1}(\bar{\boldsymbol{R}}_t - \bar{\boldsymbol{a}}^{(g)}), \ \boldsymbol{f}_t^{(g)} = \left(\boldsymbol{l}_t^{(g)\prime}, \boldsymbol{m}_t', \boldsymbol{z}_t'\right)', \\
\boldsymbol{x}_t^{(g)} &= I_{m+z} \otimes \boldsymbol{f}_t^{(g)\prime}, \\
u_t &= \begin{pmatrix} \boldsymbol{m}_t \\ \boldsymbol{z}_t \end{pmatrix} - \boldsymbol{x}_{t-1}^{(g)} \times vec\left(G_{MZ}^{(g)}\right), \\
\Gamma_{MZ,1}^{(g)} &= \left[\Gamma_{MZ,0}^{-1} + 1200^2 \times \sum\nolimits_{t=2}^{T} u_t' u_t\right]^{-1}.
\end{aligned}$$

From the full conditional density of Ω_{MZ}^*,

$$\pi\left(1200^2 \times \Omega_{MZ}^* | \boldsymbol{Y}, G^{(g)}, \Sigma^*, \psi^*\right) = \mathcal{IW}\left(1200^2 \times \Omega_{MZ}^* | v_{MZ,1}, \Gamma_{MZ,1}^{(g)}\right),$$

it can be numerically approximated by

$$\pi(1200^2 \times \Omega_{MZ}^* | \boldsymbol{Y}, \Sigma^*, \psi^*) \approx \frac{1}{n_1} \sum_{g=1}^{n_1} \pi\left(1200^2 \times \Omega_{MZ}^* | \boldsymbol{Y}, G^{(g)}, \Sigma^*, \psi^*\right).$$

The last ordinate $\pi(G^* | \boldsymbol{Y}, \Sigma^*, \Omega_{MZ}^*, \psi^*)$ does not require any reduced MCMC runs.

Table B.5: Model selection frequency based on 50 simulated datasets. Each simulation dataset consists of 300 periods of data. The first column reports the true relevant macro factors in the DGP. The second column reports the frequency with which the true model has the highest marginal likelihood. $\mathcal{M}_i$ denotes the model with i relevant macro factors and $(10-i)$ irrelevant macro factors.

true model	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$	$\mathcal{M}_8$	$\mathcal{M}_9$	$\mathcal{M}_{10}$
$\mathcal{M}_0$	62.8	31.9	5.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\mathcal{M}_3$	0.0	0.0	0.0	75.5	17.5	6.9	0.0	0.0	0.0	0.0	0.0
$\mathcal{M}_5$	0.0	0.0	0.0	0.0	0.0	61.2	27.1	9.3	2.4	0.0	0.0

It can be directly calculated by

$$\pi\left(G^*|\boldsymbol{Y},\Sigma^*,\Omega^*_{MZ},\psi^*\right)=\mathcal{N}\left(vec(G^*)|vec(G^*_1),V^*_{G_1}\right),$$

where

$$\begin{aligned}
V^*_{G_1} &= \left(V^{-1}_{G_0}+(\Omega^*)^{-1}\otimes(\boldsymbol{F}'_{1:T-1}\times\boldsymbol{F}_{1:T-1})\right)^{-1},\\
vec(G^*_1) &= V_{G_1}\times\left[V^{-1}_{G_0}\times vec(G_0)+((\Omega^*)^{-1}\otimes\boldsymbol{F}_{1:T-1})'\times vec(\boldsymbol{F}_{2:T})\right],\\
\boldsymbol{l}_t &= \bar{\boldsymbol{b}}^{-1}(\bar{\boldsymbol{R}}_t-\bar{\boldsymbol{a}}),\ \text{and}\ \boldsymbol{f}_t=(\boldsymbol{l}'_t,\boldsymbol{m}'_t,\boldsymbol{z}'_t)'
\end{aligned}$$

for all $t=1,2,..,T$, given $(\boldsymbol{Y},\Sigma^*,\Omega^*_{MZ},\psi^*)$.

Appendix B.2. Simulation Study

In our framework, a different collection of relevant macro risk factors correspond to different zero restrictions on G_{lz}, and hence, a different model. Therefore, comparing different models in terms of the Bayesian model choice provides the means to identify the relevant macro risk factors. We examine the reliability of our marginal likelihood comparison approach by using a simulation study. After specifying a true model, we generate artificial data for both bond yields and macro factors. We then estimate all competing models and calculate their marginal likelihoods to see if the model with the largest marginal likelihood is the true model specification.

Specifically, our simulation study consists of two stages. The first stage is to generate artificial data from the true models according to the transition equation (2.1) and the measurement equations (3.2) and (3.4). The parameters are fixed at the prior mean except $G_{lm}=0.2\times 1_{l\times m}$ where $1_{l\times m}$ is a matrix with all elements equal to one. We

consider 10 candidates of macro risk factors, as in the real data analysis and three true models; an yield-only model, a model with three relevant macro risk factors, and a model with five relevant macro risk factors. In the second stage, we use the simulated bond yields and macro risk factors to estimate and compare the marginal likelihoods of the 11 competing model, $\{\mathcal{M}_i\}_{i=0}^{10}$, where $\mathcal{M}_i$ denotes the model with i relevant macro risk factors and $(10-i)$ irrelevant macro factors. As we use simulated data, the 11 competing models, rather than 2^{10} models, are compared. These two stages are repeated 50 times. The model selection frequencies for each true model are given in Table B.5. The selection of the true model by the marginal likelihood criterion is quite high, demonstrating the reliability of our marginal likelihood comparison.