Spring 5-15-2019

Essays on Macro-Finance Affine Term Structure Models

Biancen Xie
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Economics

Dissertation Examination Committee:
Siddhartha Chib, Chair
Werner Ploberger, Co-Chair
Gaetano Antinolfi
Yongseok Shin
Guofu Zhou

Essays on Macro-Finance Affine Term Structure Models

by
Biancen Xie

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor Siddhartha Chib, Guofu Zhou, Werner Ploberger, Yongs Shin and Gaetano Antinolfi for valuable feedback on the manuscript. I would also like to thank Kyu Ho Kang and Xiaming Zeng for their help on the finishing dissertation.

Biancen Xie

Washington University in St. Louis

May 2019

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Essays on Macro-Finance Affine Term Structure Models

by

Biancen Xie

Doctor of Philosophy in Economics

Washington University in St. Louis, 2019

Professor Siddhartha Chib, Chair

Professor Werner Ploberger, Co-Chair

In my dissertation, I focus on theoretical affine term structure models and the development of Bayesian econometric methods to estimate them.

In the first Chapter, we address the question of which unspanned macroeconomic factors are the best in the class of macro-finance Gaussian affine term structure models. To answer this question, we extend Joslin, Priebsch, and Singleton (2014) in two dimensions. First, following Ang and Piazzesi (2003) and Chib and Ergashev (2009), three latent factors, instead of the first three principal components of the yield curve, are used to represent the level, slope and curvature of the yield curve. Second we postulate a grand affine model that includes all the macro-variables in contention. Specific models are then derived from this grand model by letting each of the macro-variables play the role of a relevant macro factor (i.e. by affecting the time-varying market price of factor risks), or the role of an irrelevant macro factor (having no effect on the market price of factor risks). The Bayesian marginal likelihoods of the resulting models are computed by an efficient Markov chain Monte Carlo algorithm and the method of Chib (1995) and Chib and Jeliazkov (2001). Given eight common macro factors, our comparison of 2^8=256 affine models shows that the most relevant macro factors for the U.S. yield curve are the federal funds rate, industrial production, total capacity

utilization, and housing sales. We also show that the best supported model substantially improves out-of-sample yield curve forecasting and the understanding of term-premium.

The second Chapter considers the question of which unspanned macro factors can improve prediction in arbitrage-free affine term structure models and convert return forecasts into economic gains. To achieve this, we develop a Bayesian framework for incorporating different combinations of macro variables within an affine term structure framework. Then each specific model within the framework is evaluated statistically and economically. For the statistical evaluation, we examine its out-of-sample yield density forecasting. The economic value of each model is compared in terms of the bond portfolio choice of a Bayesian risk-averse investor. We consider two main kinds of macro factors: representative macro factors in Chib et al. (2019) and principal component macro factors in Ludvigson and Ng (2009b). Our empirical results show that regardless of macro dataset we use(either Chib et al. (2019) or Ludvigson and Ng (2009b)), macro factor in real economic activity, financial sector and price index will help generate notable gains in out-of-sample forecast. Such gains in predictive accuracy translate into higher portfolio returns after accounting for estimation error and model uncertainty. In contrast, incorporating redundant macro variables into the affine term structure models can even decrease utility and prediction accuracy for investors. In addition, given the data sample we consider in the Chapter, we also find that principle component factors can perform relatively better than representative macro factors in terms of certainty equivalence return (CER).

The third Chapter compares the posterior sampling performance of No-U-Turn sampler(NUTS) algorithm and tailored randomized-blocking Metropolis-Hastings (TaRB-MH) for macro-finance affine Term structure models. We conduct empirical experiments on 3 affine term structure models with the U.S. yield curve data. For each experiment, we examine the sampling efficiency of model parameters, factors, term premium, predictive yields,etc. Our emprical results indicate that the TaRB-MH substantially outperforms the NUTS method

in terms of the convergence and efficiency in posterior sampling. Furthermore, we show that NUTS' inefficiency in simulating the affine term structure models will be robust given different initial values for the algorithm.

Chapter 1

Framework for Finding Relevant Macro Factors in Affine Term Structure Models

1.1 Introduction

In recent important work, Duffee (2011) and Joslin et al. (2014) have developed a macro-finance arbitrage-free affine term structure model in which a set of exogenous factors $\mathbf{f}_t$ at time t, consisting of three yield curve factors $\boldsymbol{l}_t : 3 \times 1$ (given by the first three principal components of the yield curve), and pre-selected macro economic factors $\mathbf{m}_t : m \times 1$, are assumed to follow a first-order Gaussian vector-autoregressive (VAR) process given by

$$\mathbf{f}_{t+1} = K + G\mathbf{f}_t + \varepsilon_{t+1} \ , \ \varepsilon_{t+1} \sim \mathcal{N}\left(0_{(3+m)\times 1}, \Omega\right)$$

These factors determine the one-period ahead stochastic discount factor (SDF) M_{t+1} as

$$M_{t+1} = \mathsf{exp}\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' v_{t+1}\right) \tag{1.1.1}$$

where r_t is the short rate that is assumed to be an affine function of $\boldsymbol{l}_t$, but not of $\mathbf{m}_t$, v_{t+1} are the iid factor shocks such that $\varepsilon_{t+1} = \Omega^{1/2} v_{t+1}$, where $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$, and the time-varying market price of factor risks $\lambda_t : (3+m) \times 1$ is assumed to follow the process

$$\lambda_t = \left(\Omega^{1/2}\right)^{-1} \left(\begin{pmatrix} \lambda_l \\ \mathbf{0}_{m\times 1} \end{pmatrix} + \begin{pmatrix} \Lambda_{ll} & \Lambda_{lm} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{l_t} \\ \mathbf{m}_t \end{pmatrix} \right)$$

Joslin et al. (2014) model the role of $\mathbf{m}_t$ in this way to ensure that the bond price for a τ period bond at time t that satisfies the no-arbitrage condition

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

will depend just on $\boldsymbol{l_t}$, but not on $\mathbf{m}_t$. Technically, their modeling ensures that the macro-factors are unspanned by the yield curve, which Duffee (2011) and Joslin et al. (2014) argue is an important requirement. Yet, even though the macro-factors do not directly affect the yield curve at t, Joslin et al. (2014) show that the macro factors, through their influence on λ_t, influence the dynamics of the term-premium, and aid in forecasting future values of the yield curve.

We have two goals in this paper within the context of this model. One is to relax the assumption that $\boldsymbol{l_t}$ are observed factors. Instead, following another branch of the term-structure literature, for example, Ang and Piazzesi (2003), Cochrane and Piazzesi (2005), Chib and Ergashev (2009), Chib and Kang (2013), Ludvigson and Ng (2009b), Cooper and Priestley (2009), Gurkaynak and Wright (2012), Greenwood and D.Vayanos (2014) and Cieslak (2015), we let $\boldsymbol{l_t}$ denote three latent factors. We constrain these in our modeling to represent the level, slope and curvature of the yield curve. The first three principal components of the yield curve are also ostensibly meant to capture the level, slope and curvature of the yield curve, but our approach with latent variables is more flexible. The downside of

modeling the level, slope and curvature by a latent vector is that the econometrics become more complex (essentially because the likelihood function becomes irregular, with separated regions and multiple modes). We address the computational challenge of dealing with such a likelihood function by introducing a prior distribution that smooths out the likelihood irregularities, following Chib and Ergashev (2009) and Abbritti, Gil-Alana, Lovcha, and Moreno (2016). The tailored randomized block Metropolis-Hastings (TaRB-MH) algorithm of Chib and Ramamurthy (2010) is then used to sample the resulting posterior distribution efficiently.

The second goal is to develop a framework for selecting the appropriate unspanned macro factors $\mathbf{m}_t$ on the basis of the data evidence. In the model of Joslin et al. (2014), all included macro-factors affect λ_t, an assumption that the factors are relevant, which may not be justified. It is possible that some macro-factors may not affect λ_t, and thereby be considered irrelevant, or some factors excluded from the model should be included as relevant factors. In the present context, these questions cannot simply be answered by omitting or deleting a macro factor from the model because this changes the set of modeled outcomes, which in the model above, are the yields and the macro factors. Our idea, therefore, is to postulate a grand affine model that includes all the macro-variables in contention. Specific models are then derived from this grand model by letting each of the macro-variables (in a combinatorial fashion) play the role of a relevant macro-factor, or the role of an irrelevant macro-factor. In this way, the number of modeled outcomes remains fixed. In our empirical study with eight candidate macro-factors, this process of considering all possible combinations of relevant and irrelevant macro-factors leads to a trove of $2^8 = 256$ competing affine models, each with its own model parameters, and each with its specific implications for the term-structure of interest rates. We compare each model, in this conglomeration of models, by the Bayesian marginal likelihood criterion. The marginal likelihood has a built-in penalty for complexity, ensuring that more complex models will not rank higher merely because the more flexible

Table 1.1: **Pool of Candidate Macroeconomic Variables**

This table lists the short name of each macro series, the transformation applied to the series, and a brief data description. All series are from the online database housed at the Federal Reserve Bank of St. Louis. In the column labeled trans., 1 indicates that the variable is transformed into annual growth rate, and 0 indicates that the variable is retained in level form.

variable	trans.	description
AHE	1	Average Hourly Earnings:Total Private
CPI	1	Consumer Price Index:All Items
FFR	0	Effective Federal Funds Rate
HSal	1	House Sales-New One Family Houses
IP	1	Industrial Production Index
PCE	1	Personal Consumption Expenditures
TCU	0	Capacity Utilization:Total Industry
Unem	0	Civilian Unemployment Rate

model is capable of fitting the noise in the data.

We find the best supported macro-factors from the composition of relevant factors in the model with the highest marginal likelihood. We show that the best collection of unspanned macro-factors improves yield curve forecasting (which is important, for example, in bond-portfolio construction), and has consequences for term-premium estimation (which is important for extracting the market expectation component from long-term interest rates).

The pool of macro-factors that we consider are listed in Table 1.1. This collection consists of two inflation measures, five real economic activity factors, and one financial factor. The inflation measures are consumer price index (CPI) and average hourly earning (AHE). The real activity is measured by industrial production (IP), personal consumption expenditures (PCE), capacity utilization (CU), unemployment rate (Unem), and housing sales (HSal). The federal fund rate (FFR) is a proxy of the financial factor.

Our main findings from 256 affine models applied to the U.S. yield curve data (over the sample period from January 1990 to December 2017) can be summarized as follows. First, the model with four relevant factors (FFR, HSal, IP, TCU) is found to be best supported by the data. Table 1.2 presents the log Bayes factors of the top 10 models, where the 10th best model is used as the benchmark.

Table 1.2: **Log Bayes Factors of the 10 Best Models**

This table reports the log Bayes factors of the 10 best models (relative to the 10th best model) from the analysis conducted below. The first column reports the rank of each model, the second column reports the set of relevant macro factors in that model and the third column reports the log Bayes Factor of each model. The sample period of the analysis is from January 1990 to December 2017 and the details about the models and the estimation methodology is discussed in the following sections.

rank	relevant macro factors	log Bayes Factor
1	FFR, Hsal, IP, TCU	45.21
2	Hsal, IP, TCU, Unem	23.95
3	FFR, Hsal, TCU, Unem	15.13
4	AHE, CPI, IP, PCE, Unem	13.97
5	CPI, Hsal, IP, Unem	13.06
6	Hsal, IP, PCE, Unem	9.36
7	CPI, Hsal, IP, Unem	5.18
8	IP, PCE, TCU, Unem	1.78
9	AHE, FFR, Hsal, IP, Unem	0.88
10	FFR, Hsal, IP, PCE	0

The evidence in favor of these macro-factors is decisive, given that the log marginal likelihood of the best model exceeds that of the next model by 21.62. Although some of these macro factors have been considered in the work of, for example, Joslin et al. (2014), Diebold, Rudebusch, and Aruoba (2006), and Bansal and Yaron (2004), our comparative analysis shows that these factors in combination are the best supported by the data. Second, we show that the term premium inferred from the best model is strongly counter-cyclical, indicating that investors have to bear more macroeconomic risks during recessions. Third, the best model substantially improves on the out-of-sample density forecasts. It is worth noting that Bauer and Hamilton (2018) have also studied the role of the unspanned macro factors on the yield curve but not in a model such as ours, and, hence, those results are not directly comparable.

The remainder of the paper is organized as follows. In Section 3.2 we present our framework with relevant and irrelevant unspanned macro factors and in Section 2.2 we describe the data that is used in our empirical analysis. Section 2.4 details our Bayesian estimation strategy and marginal likelihood computation method. Section 1.5 provides the empirical

results from the application of our approach. Concluding remarks are presented in Section 2.7 and the Appendix contains supplementary details omitted from the main text.

1.2 Framework

We now describe the macro-finance arbitrage-free affine term structure model, built on the work of Duffee (2011), and Joslin et al. (2014), designed to satisfy the two goals outlined above. Our model allows for three distinct sets of factors: latent factors $\boldsymbol{l}_t$ of size $l\ (= 3) \times 1$, which are present in every model, and two disjoint groups of unspanned macroeconomic factors, relevant factors $\mathbf{m}_t$ of size $m \times 1$, and irrelevant factors $\mathbf{z}_t$ of size $z \times 1$. Note that under this formulation, the total number of factors is $f = l + m + z$ and the size of λ_t is now $f \times 1$ (the size of the other vectors and matrices in the previous section also now changes). The macro factors in $\mathbf{m}_t$ are relevant in the sense that these variables affect λ_t, while the variables in $\mathbf{z}_t$ are constrained to have no effect on λ_t.

A particular model is distinguished by the composition of these groups of macro factors. As a result, there are $2^{(m+z)}$ models of interest. The number of factors is the same across models, though in some models $\mathbf{m}_t$ or $\mathbf{z}_t$ could be empty, but not both at the same time.

1.2.1 Factor Process

We assume that the factors, $\mathbf{f}_t = (\boldsymbol{l}_t, \mathbf{m}_t, \mathbf{z}_t)$, follow a first-order vector-autoregressive (VAR) process given by

$$\underbrace{\begin{pmatrix} \boldsymbol{l}_{t+1} \\ \boldsymbol{m}_{t+1} \\ \boldsymbol{z}_{t+1} \end{pmatrix}}_{\mathbf{f}_{t+1}} = \underbrace{\begin{pmatrix} K_l \\ K_m \\ K_z \end{pmatrix}}_{K} + \underbrace{\begin{pmatrix} G_{ll} & G_{lm} & G_{lz} \\ G_{ml} & G_{mm} & G_{mz} \\ G_{zl} & G_{zm} & G_{zz} \end{pmatrix}}_{G} \underbrace{\begin{pmatrix} \boldsymbol{l}_t \\ \mathbf{m}_t \\ \mathbf{z}_t \end{pmatrix}}_{\mathbf{f}_t} + \underbrace{\begin{pmatrix} \varepsilon_{l,t+1} \\ \varepsilon_{m,t+1} \\ \varepsilon_{z,t+1} \end{pmatrix}}_{\varepsilon_{t+1}}, \tag{1.2.1}$$

where the factor shocks ε_{t+1} are jointly normally distributed as

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

For estimation purposes, it is convenient to express the variance-covariance matrix of the factor shocks as

$$\Omega = V\Gamma V,$$

where $V = \mathsf{diag}(v_1, v_2, .., v_{k+m})$ is the diagonal factor shock volatility matrix and Γ is the factor shock correlation matrix.

1.2.2 SDF

Denote the factor shocks as

$$\varepsilon_{t+1} = \Omega^{1/2} v_{t+1},$$

where

$$v_{t+1} \sim \mathcal{N}\left(0_{f\times 1}, \mathbf{I}_f\right)$$

is the vector of normalized factor shocks and $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$. Then, our assumption is that the one-period ahead SDF is given by

$$M_{t+1} = \mathsf{exp}\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' v_{t+1}\right) \tag{1.2.2}$$

where r_t follows the process

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

and λ_t follows

$$\lambda_t = \left(\Omega^{1/2}\right)^{-1} \left(\lambda + \Lambda \begin{pmatrix} \boldsymbol{l_t} \\ \mathbf{m}_t \\ \mathbf{z}_t \end{pmatrix} \right) \tag{1.2.3}$$

where

$$\lambda = \begin{pmatrix} \lambda_l \\ 0_{m\times 1} \\ 0_{z\times 1} \end{pmatrix}, \text{ and } \Lambda = \begin{pmatrix} \Lambda_{ll} & \Lambda_{lm} & 0_{l\times z} \\ 0_{m\times l} & 0_{m\times m} & 0_{m\times z} \\ 0_{z\times l} & 0_{z\times m} & 0_{z\times z} \end{pmatrix} \tag{1.2.4}$$

similar to the model of Joslin et al. (2014) but allowing for the presence of irrelevant macro-factors $\mathbf{z}_t$.

By the no-arbitrage condition given in the introduction, the price $P_t(\tau)$ of a default-free zero-coupon bond with τ-periods to maturity at time t that pays one unit at time $t+\tau$ is

$$P_t(\tau) = \mathbb{E}_t[M_{t+1} P_{t+1}(\tau - 1)], \tag{1.2.5}$$

where $\mathbb{E}_t(\cdot)$ is the conditional expectation given the information available at time t including the entire history of the factors, $\{\mathbf{f}_i\}_{i=1}^t$. That is,

$$\begin{aligned} P_t(\tau) = \int_{R^k} \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' v_{t+1}\right) P_{t+1}(\tau - 1) \\ \times \mathcal{N}(\boldsymbol{f}_{t+1} | G\boldsymbol{f}_t, \Omega) d\boldsymbol{f}_{t+1}. \end{aligned}$$

By straightforward manipulations, the latter condition can be re-expressed as

$$P_t(\tau) = e^{-r_t}\mathbb{E}_t^{\mathbb{Q}}[P_{t+1}(\tau - 1)], \tag{1.2.6}$$

where the $\mathbb{Q}$-measure is

$$\mathcal{N}(\boldsymbol{f}_{t+1}|K^{\mathbb{Q}} + G^{\mathbb{Q}}\boldsymbol{f}_t, \Omega)d\boldsymbol{f}_{t+1}, \tag{1.2.7}$$

$$K^{\mathbb{Q}} = K - \lambda = \begin{pmatrix} K_l - \lambda_l \\ K_m \\ K_z \end{pmatrix}, \tag{1.2.8}$$

and

$$G^{\mathbb{Q}} = G - \Lambda = \begin{pmatrix} G_{ll} - \Lambda_{ll} & G_{lm} - \Lambda_{lm} & G_{lz} \\ G_{ml} & G_{mm} & G_{mz} \\ G_{zl} & G_{zm} & G_{zz} \end{pmatrix}.$$

Following Joslin et al. (2014), we restrict $G^{\mathbb{Q}}$ as

$$G^{\mathbb{Q}} = \begin{pmatrix} G_{ll}^{\mathbb{Q}} & \mathsf{0}_{l\times m} & \mathsf{0}_{l\times z} \\ G_{ml} & G_{mm} & G_{mz} \\ G_{ml} & G_{mm} & G_{mz} \end{pmatrix},$$

which implies that

$$\begin{aligned} \Lambda_{ll} &= G_{ll} - G_{ll}^{\mathbb{Q}}, \\ \Lambda_{lm} &= G_{lm}, \\ G_{lz} &= \mathsf{0}_{l\times z}. \end{aligned} \tag{1.2.9}$$

We set $G_{lz} = \mathsf{0}_{l\times z}$ to shut down any back-door influence of $\boldsymbol{z_t}$ on λ_t. In contrast, non-zero

Λ_{lm} $(= G_{lm})$ allows $\mathbf{m}_t$ to affect λ_t. We obtain a latent factor only model when all macro variables are regarded as irrelevant and $\mathbf{m}_t$ is empty.

We complete our model by incorporating the Nelson-Siegel type latent-factor identifying restriction of Christensen, Diebold, and Rudebusch (2009) and Christensen, Diebold, and Rudebusch (2011), so that the model-implied risk premiums on exposures to level, slope, and curvature risks can be derived and estimated. Specifically, we restrict β and $G_{ll}^{\mathbb{Q}}$ as follows:

$$\boldsymbol{\beta} = (1, 1, 0)' \tag{1.2.10}$$

and

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

depending on the single parameter κ.

1.2.3 Solution

Assuming that $P_t(\tau)$ is exponentially affine in the latent factors

$$P_t(\tau) = \exp\left(-a(\tau) - b(\tau)'\boldsymbol{l_t}\right)$$

for any non-negative integer value of τ, where $a(\tau)$ is a scalar and $b(\tau)$ is a $l \times 1$ vector, by the method of undetermined coefficients illustrated in Appendix .1, we get that

$$\begin{aligned} a(\tau) &= \delta + a(\tau - 1) - \frac{1}{2} b(\tau - 1)' \Omega_{ll} b(\tau - 1) - b(\tau - 1)' \lambda_l, \\ b(\tau) &= \beta + G_{ll}^{\mathbb{Q}\prime} b(\tau - 1). \end{aligned} \tag{1.2.11}$$

As $P_t(\tau = 0) = 1$ for any $\boldsymbol{l_t}$, this recursive system is initialized by $a(0) = 0$ and $b(0) = 0_{l \times 1}$.

In addition, as proved by Niu and Zeng (2012), $b(\tau)$ reduces to

$$\begin{aligned} b(\tau)' &= \beta' \times \sum_{j=0}^{\tau-1} \left(G_{ll}^{\mathbb{Q}}\right)^j \\ &= \left[\begin{array}{ccc} \tau, & \tau - \frac{1-\exp(-\kappa\tau)}{\kappa}, & \frac{1-\exp(-\kappa\tau)}{\kappa} - \tau\exp(-\kappa\tau) \end{array} \right] \end{aligned} \tag{1.2.12}$$

It follows now that the yield curve can be expressed as a linear function of the latent factors,

$$R_t(\tau) \approx -\frac{1}{\tau}\log P_t(\tau) = \frac{a(\tau)}{\tau} + \frac{b(\tau)'}{\tau}\boldsymbol{l}_t \tag{1.2.13}$$

where $b(\tau)'/\tau$ has exactly the form of the dynamic Nelson-Siegel factor loadings,

$$\left[\begin{array}{ccc} 1, & 1 - \frac{1-\exp(-\kappa\tau)}{\kappa\tau}, & \frac{1-\exp(-\kappa\tau)}{\kappa\tau} - \exp(-\kappa\tau) \end{array} \right].$$

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.

1.2.4 Term Premium and the Role of Relevant Macro Factors

A key quantity of interest in our context is the term premium, which quantifies the compensation required by a risk-averse investor holding long-term government bonds. This section discusses how relevant macro factors affect the term premium. Let

$$\begin{aligned} \text{exr}_t^{(\tau)} &= \left[\mathbb{E}_t\left[\ln P_{t+1}(\tau-1)\right] - \ln P_t(\tau)\right] - \left(-\ln P_t(1)\right) \\ &= -b(\tau-i)'\lambda_t \end{aligned} \tag{1.2.14}$$

denote the one-period expected excess return for holding the τ-period bond. Then, as shown by Cochrane and Piazzesi (2008), the term premium (TP) for a τ-period bond at time t

is defined as the difference between its yield to maturity and corresponding expectation hypothesis (EH) term,

$$\text{TP} = R_t(\tau) - \underbrace{\frac{1}{\tau}\sum_{i=0}^{\tau-1}\mathbb{E}_t\left[r_{t+i}\right]}_{\text{EH}},$$

which reduces to the average expected excess returns,

$$\text{TP} = \frac{1}{\tau}\sum_{i=1}^{\tau-1}\text{exr}_t^{(\tau+1-i)}. \tag{1.2.15}$$

Inserting the expression of $\text{exr}_t^{(\tau)}$ in Equation (1.2.14) into the equation above, we have the term premium of τ-period bond at time t in a closed form,

$$\begin{aligned}\text{TP} &= -\tau^{-1}\sum\nolimits_{i=1}^{\tau-1} b(\tau-i)'\lambda_t \\ &= -\tau^{-1}\sum\nolimits_{i=1}^{\tau-1} b(\tau-i)'\left(\Omega_{ll}^{1/2}\right)^{-1}\left(\lambda_l + \Lambda_{ll}\times \boldsymbol{l_t} + \Lambda_{lm}\times \mathbf{m}_t\right).\end{aligned}$$

This can be decomposed into three components: (i) a time-invariant component,

$$-\tau^{-1}\sum\nolimits_{i=1}^{\tau-1} b(\tau-i)'\left(\Omega_{ll}^{1/2}\right)^{-1}\lambda_l,$$

(ii) a time-varying component by $\boldsymbol{l_t}$,

$$-\tau^{-1}\sum\nolimits_{i=1}^{\tau-1} b(\tau-i)'\left(\Omega_{ll}^{1/2}\right)^{-1}\left(\Lambda_{ll}\times \boldsymbol{l_t}\right),$$

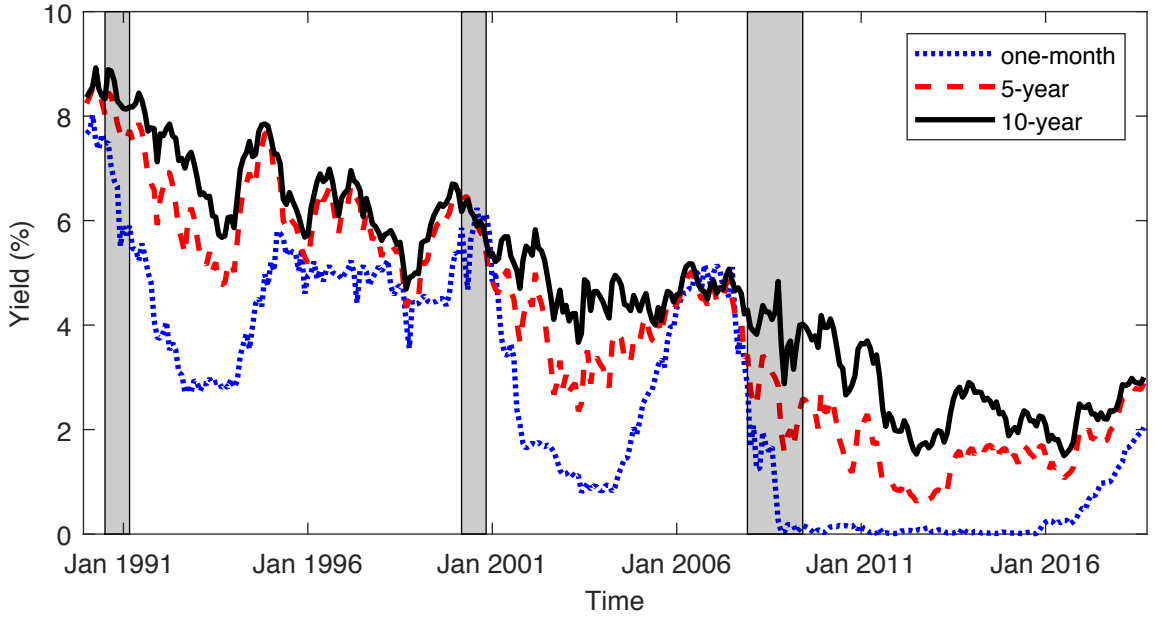
and (iii) a time-varying component by $\mathbf{m}_t$,

$$-\tau^{-1}\sum\nolimits_{i=1}^{\tau-1} b(\tau-i)'\left(\Omega_{ll}^{1/2}\right)^{-1}\left(\Lambda_{lm}\times \mathbf{m}_t\right). \tag{1.2.16}$$

The last term represents the effect of the relevant macro factors on the term premium.

Figure 1.1: **Bond Yields**

This figure plots the monthly time series of government bond yields with 1, 60, and 120 months to maturity from January 2001 to September 2018. Shaded areas indicate recession periods designated by the National Bureau of Economic Research.



1.3 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. Figure 1.1 plots the short-term yield (one month), mid-term yield (60 months), and long-term yield (120 months), indicating that the yield curve is upward sloping on average and highly autocorrelated. Our data on the eight macro factors are plotted in Figure 1.2. It can be seen that IP, PCE, TCU, Hsal, AHE, and CPI are procyclical, while Unem and FFR are counter-cyclical.

The data used in our empirical analysis range from January 1990 to September 2018, corresponding to the period after the Great moderation. Table 1.3 provides summary statistics of the bond yields and macro factors. In addition, as we will discuss later, the data from January 1990 to December 1999 are used as a training sample to arrive at the prior

Figure 1.2: **Macro Factors**

This figure plots the eight macro factors mentioned in Table 1.1 for the sample period from January 1990 to September 2018. Shaded areas indicate recession periods designated by the National Bureau of Economic Research.

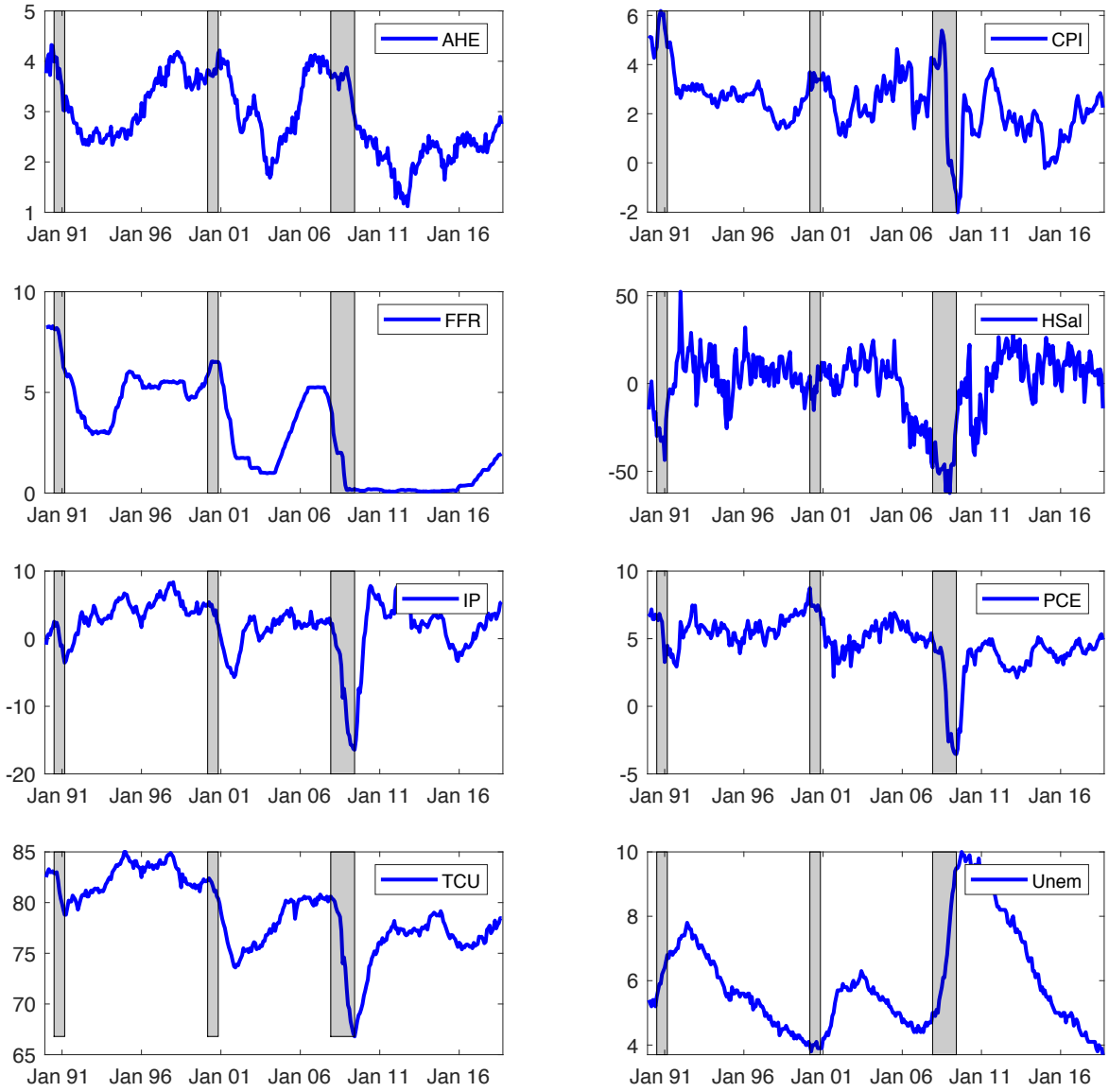


Table 1.3: **Summary Statistics of Yields and Candidate Macro Factors**

The table provides six measures (Mean, St.dev, Skewness, Kurtosis, min and max) of both yields and macro factor time series for the sample periods from January 1990 to September 2018.

	Mean	St.dev	Skewness	Kurtosis	min	max
1-month yield	2.24	2.22	0.26	1.76	0.01	8.02
60-month yield	4.01	2.13	0.17	1.91	0.62	8.879
120-month yield	4.75	1.93	0.13	2.05	1.5	8.9
AHE	2.88	0.75	0.07	2.04	1.12	4.32
CPI	2.44	1.25	-0.08	4.26	-2.01	6.18
IP	1.90	4.11	-1.96	8.54	-16.4	8.36
PCE	4.69	1.86	-1.84	8.75	-3.54	8.74
FFR	2.94	2.41	0.27	1.77	0.07	8.29
Unem	5.98	1.56	0.92	3.03	3.7	10
TCU	78.88	3.64	-0.65	3.57	66.8	85
HSal	-0.12	18.62	-1.02	3.87	-62.38	52.22

distributions for each competing model, the data from January 2000 to December 2017 are used for the posterior simulation and marginal likelihood calculation, and the data from January 2018 to September 2018 is reserved for out-of-sample prediction comparisons.

In Table 1.4 we report the sample correlation between macro factors and the yields. Most correlations seem to be high, and the correlations with short-term yields tend to be higher than those with long-term yields. For example, the correlation between AHE and one-month yield is 0.7 and the correlation between TCU and one-month yield is 0.76. Meanwhile, the counterparts for the 120-month yield are 0.53 and 0.62, respectively. Roughly speaking, each macro factor can be seen as potentially informative for the U.S. yield curve.

Table 1.4: **Correlations between Yields and Candidate Macro Factors**

This table is a summary of the sample correlations between the macro factors and selected yields of one-month(short term), 60-month(mid-term), and 120-month(long term) bonds. The sample periods are from January 1990 to September 2018.

	1-month yield	60-month yield	120-month yield
AHE	0.71	0.59	0.53
CPI	0.57	0.56	0.55
IP	0.25	0.18	0.12
PCE	0.58	0.53	0.46
FFR	0.99	0.91	0.84
Unem	-0.48	-0.36	-0.16
TCU	0.76	0.70	0.62
HSal	-0.13	-0.09	-0.11

1.4 Prior-Posterior Analysis

1.4.1 State-Space Form

In order to estimate the affine term structure model, we first represent the yield curve model in the state space form

$$
\begin{aligned}
R_t(\tau_i) &= \frac{a(\tau_i)}{\tau_i} + \frac{b(\tau_i)'}{\tau_i}\boldsymbol{l}_t + e_t(\tau_i),\ e_t(\tau_i) \sim \mathcal{N}(0, \sigma_i^2) \text{ for } i = 1, 2, .., N, \\
\mathbf{m}_t &= \mathbf{m}_t, \\
\mathbf{z}_t &= \mathbf{z}_t,
\end{aligned}
$$

where $e_t(\tau_i)$ is the maturity-specific independent pricing error. We let $\{\tau_i\}_{i=1}^N$ denote N maturities of interest and let

$$
\mathbf{R}_t = \begin{pmatrix} R_t(\tau_1) & R_t(\tau_2) & \cdots & R_t(\tau_N) \end{pmatrix}', \tag{1.4.1}
$$

denote the corresponding bond yields at time t. Then, given the vector of the observations at time t denoted by $\mathbf{y}_t = (\mathbf{R}_t', \mathbf{m}_t', \mathbf{z}_t')'$, the econometric model can be rewritten in vector-matrix

form,

$$\mathbf{y}_t|\mathbf{f}_t, \mathbf{a}, \mathbf{B}, \Sigma \sim \mathcal{N}(\mathbf{a} + \mathbf{B} \times \mathbf{f}_t, \Sigma), \tag{1.4.2}$$

where $\Sigma = \mathsf{diag}(\sigma_1^2, \sigma_2^2, .., \sigma_N^2, 0_{(m+z)\times 1})$ is a $(N + m + z) \times 1$ diagonal matrix, $\mathbf{I}_{m+z}$ is a $(m + z) \times (m + z)$ identity matrix,

$$\begin{aligned}
\mathbf{a} &= \begin{pmatrix} \frac{a(\tau_1)}{\tau_1} & \frac{a(\tau_2)}{\tau_2} & \dots & \frac{a(\tau_N)}{\tau_N} & 0_{1\times(m+z)} \end{pmatrix}' : (N + m + z) \times 1, \\
\bar{\mathbf{B}} &= \begin{pmatrix} \frac{b(\tau_1)}{\tau_1} & \frac{b(\tau_2)}{\tau_2} & \dots & \frac{b(\tau_N)}{\tau_N} \end{pmatrix}' : N \times l, \\
\mathbf{B} &= \begin{pmatrix} \bar{\mathbf{B}} & 0_{N\times(m+z)} \\ 0_{(m+z)\times l} & \mathbf{I}_{m+z} \end{pmatrix} : (N + m + z) \times (l + m + z).
\end{aligned}$$

The transition equation of the state space model is simply given by the factor dynamics,

$$\mathbf{f}_{t+1}|\mathbf{f}_t, G, K, \Omega \sim \mathcal{N}(K + G\mathbf{f}_t, \Omega = V\Gamma V). \tag{1.4.3}$$

where we assume that the eigenvalues of G each lie inside the unit circle for second-order stationarity. In addition, K_l is constrained to zero for identifying the intercept term δ in the short rate process.

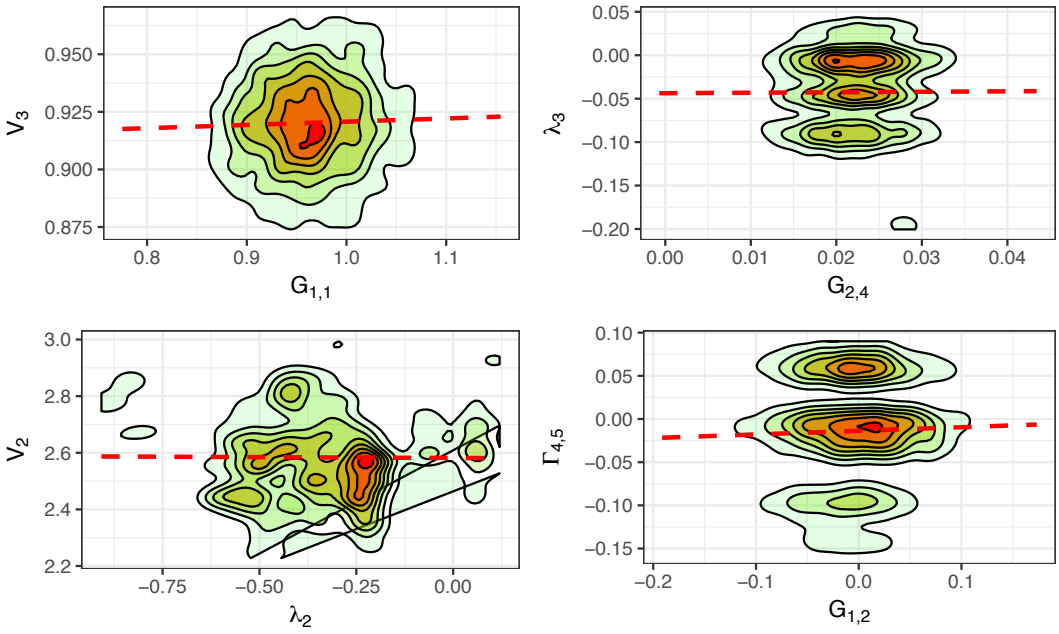
The parameters in the factor dynamics, SDF, and pricing error variances are denoted by

$$\theta_f = (K, G, V, \Gamma),\ \theta_s = (\delta, \lambda_l, \kappa) \text{ and } \theta_e = (\sigma_1^2, \sigma_2^2, .., \sigma_N^2)\ ,$$

respectively. Let $\mathbf{Y} = \{\mathbf{y}_t\}_{t=1}^T$ be the time series of the observed bond yields and macro factors, and $\mathcal{F}_t = \{\mathbf{y}_i\}_{i=1}^t$ denotes the history of observations up to time t. Then, the likelihood function is given by the joint density of the observations given the parameters

Figure 1.3: **Likelihood Surface**

This figure plots bivariate likelihood surfaces for some selected pairs of parameters from the model with the relevant macro factors given by (FFR, IP, Hsal, and TCU). The surface plots are based on $10,000$ draws from the posterior distribution under a flat prior, after a burn-in of $1,000$ draws. The sample periods used in this estimation runs from January 1990 to December 2017.



$\theta = (\theta_f, \theta_s, \theta_e)$

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

which we compute by Kalman filtering, as shown in Appendix .2.

This likelihood function is typically irregular, exhibiting multiple modes. To illustrate this, we consider the model with the best grouping of macro factors in our empirical study, namely $\mathbf{m}_t = \{\text{FFR, Hsal, IP, TCU}\}$ and $\mathbf{z}_t = \{\text{AHE, CPI, PCE, Unem}\}$. For this particular model, we plot in Figure 1.3 the bivariate-likelihood surface for some selected pairs of parameters. These bivariate views are calculated, following Chib and Ergashev (2009), from the output of a MCMC simulation of the posterior distribution under a flat prior.

1.4.2 Prior Specification

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. We now discuss how we specify such a model-specific prior. The set of parameters that appear in any model is given by

$$\theta = (\theta_f,\ \theta_s,\ \theta_e).$$

These parameters, in general, lie in a high dimensional parameter space. For example, in the best model we mentioned above, there are 185 parameters in total. Therefore, the prior, in any given model, should also serve as a suitable regularizer. Note that the competing models with different relevant macro factors are characterized by the zero restrictions on $G_{lz} = 0_{l\times z}$. This means that except for the elements in G_{lz}, the competing models can be given an initial common prior distribution. We determine this initial common prior in Step 1 next. Then in Step 2, this initial common prior distribution is fine-tuned for each model. We now supply the details surrounding these two steps.

Step 1: We develop the initial common prior distribution for all 256 models by incorporating theoretically supported knowledge. In particular, following Chib and Ergashev (2009) and Abbritti et al. (2016), we incorporate in the prior distribution the belief that the yield curve should be gently upward sloping on average and persistent over time. We arrive at this initial 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 prior beliefs. Figure 1.4 plots the distribution of the prior-implied yield curve for model which includes four relevant macro factors. Figure 1.4(a) is the prior mean of the yield curve, which is the intercept term, $\bar{a}$ computed at the prior mean of the parameters. The prior-implied yield curve is gently upward sloping on average, so the first moment of the prior yield curve reflects our prior beliefs. Figure 1.4(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. This process leads to the common initial prior distribution in Table 1.5. This prior distribution is reasonable and can be motivated as follows.

Initial prior for $\theta_f = (K, G, V, \Gamma)$: The parameters of G follow a normal distribution. Let $G_{i,,j}$ denote the (i,j)th element of G. Then, the prior mean of $G_{i,i}$ is 0.9 as the bond yields are known to be persistent. The prior-scale for the remaining elements in G is 0.1. Assuming the independence among the parameters, we have that

$$\pi(G) = \prod_{i,j=1}^{l+m+z} \mathcal{N}\left(G_{i,j} | \bar{G}_{i,j}, V_{G,i,j}\right),$$

where $\mathcal{N}(\cdot|\mu, \sigma^2)$ denotes the normal density of mean μ and variance σ^2. In addition, letting $K_{m,i}$ and $K_{z,i}$ denote the ith elements of K_m and K_z, respectively, the prior densities of K_m and K_z are given by

$$\begin{aligned} \pi(K_m) &= \prod_{i=1}^{m} St\left(K_{m,i} | \bar{K}_{m,i}, V_{K_m,i}, 15\right) \\ \text{and } \pi(K_z) &= \prod_{i=1}^{z} St(K_{z,i} | \bar{K}_{z,i}, V_{K_z,i}, 15), \end{aligned}$$

respectively, where $St(\cdot|\mu, \sigma^2, v)$ denotes the student-t density of mean μ, scale σ^2, and degrees of freedom v. The prior mean for $(K_m,\ K_z)$ is set to be 0, while the prior scale is set to be 0.001. The factor shock correlations, $\Gamma_{i,j}$ have a uniform prior over (-1,1) for $i \neq j$ and $i, j = 1, 2, .., (l + m + z)$. Thus, it follows that the prior density of Γ is $\pi(\Gamma) =$

$(0.5)^{(l+m+z)\times(l+m+z-1)/2}$. In addition, $V = \mathsf{diag}(V_1, V_2, .., V_{l+m+z})$ has an inverse-gamma prior distribution

$$\pi(V) = \prod_{i=1}^{l+m+z} \mathcal{IG}(V_i | a_{V,i}, b_{V,i}),$$

where $\mathcal{IG}(\cdot|a,b)$ is the inverse-gamma density with mean given by $b/(a-1)$. The prior mean and scale for V allows for yield curve variations between -3% and 11%, which is flexible enough to capture various shapes of yield curves over time. Finally, the prior density of θ_f is given by

$$\pi(\theta_f) = \pi(G)\,\pi(K_m)\pi(K_z)\,\pi(\Gamma)\,\pi(V).$$

Initial prior for $\theta_s = (\delta, \lambda_l, \kappa)$: The prior distribution for the parameters in θ_s are Student-t distributions. The prior mean of δ is chosen to be the sample mean of the short term interest rate. Because the value of κ determines the factor loadings $\bar{\mathbf{B}}$, the prior mean of κ makes the curvature factor loading reach a maximum at 24 months. Figure 1.4(d) is the factor loadings $\bar{\mathbf{B}}$. The prior on the factor loadings are the same as in Nelson-Siegel yield curve models, as the latent factors in our model correspond to the level, slope, and curvature of the yield curve. It is critical and challenging to quantify the prior on the risk parameters, λ_l. Given the prior for κ (i.e., $b(\tau)$), our prior for λ_l is chosen such that the prior-implied term premium is increasing 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 is ignored. Under this prior, the prior-implied average term premium of the 10-year bond is about 2.5% in annual terms. Figure 1.4(c) plots the term structure of the term premiums. The scale of parameters of λ_l is set at 0.001 in the Student-t distribution with 15 degrees of freedom. Finally, the prior density of θ_s is given by $\pi(\theta_s) = \pi(\delta)\,\pi(\kappa)\,\pi(\lambda_l)$, where $\lambda_{l,i}$ is the ith

Table 1.5: **Initial Prior Distributions**

$G_{i,j}$ is the (i,j)th element of G. $\lambda_{l,i}$, $K_{m,i}$, and $K_{z,i}$ are the ith elements of λ_l, K_m and K_z, respectively. $V_{i,i}$ is the ith diagonal element of V. $St(\mu, \sigma^2, v)$ is the Student-t distribution with the mean μ, scale σ^2 and degree of freedom (d.f.) v. $\mathcal{IG}(a,b)$ is the inverse-gamma distribution whose mean is $b/(a-1)$.

(a) G

	normal distribution, $\mathcal{N}(\mu, \sigma^2)$	
parameter	mean (μ)	scale (σ^2)
$G_{i,i}$	0.900	0.1
$G_{i,j}$ ($\notin G_{lz}$ and $i \neq j$)	0	0.1

(b) κ, λ and K

	Student-t distribution, $St(\mu, \sigma^2, v)$		
parameter	mean (μ)	scale (σ^2)	d. f. (v)
κ	0.935	6×10^{-3}	15
$\lambda_{l,1}$	-0.170	10^{-3}	15
$\lambda_{l,2}$	-0.070	10^{-3}	15
$\lambda_{l,3}$	-0.024	10^{-3}	15
$K_{m,i}$ $(i = 1, 2, .., m)$	0.000	10^{-3}	15
$K_{z,i}$ $(i = 1, 2, .., z)$	0.000	10^{-3}	15

(c) V and Σ

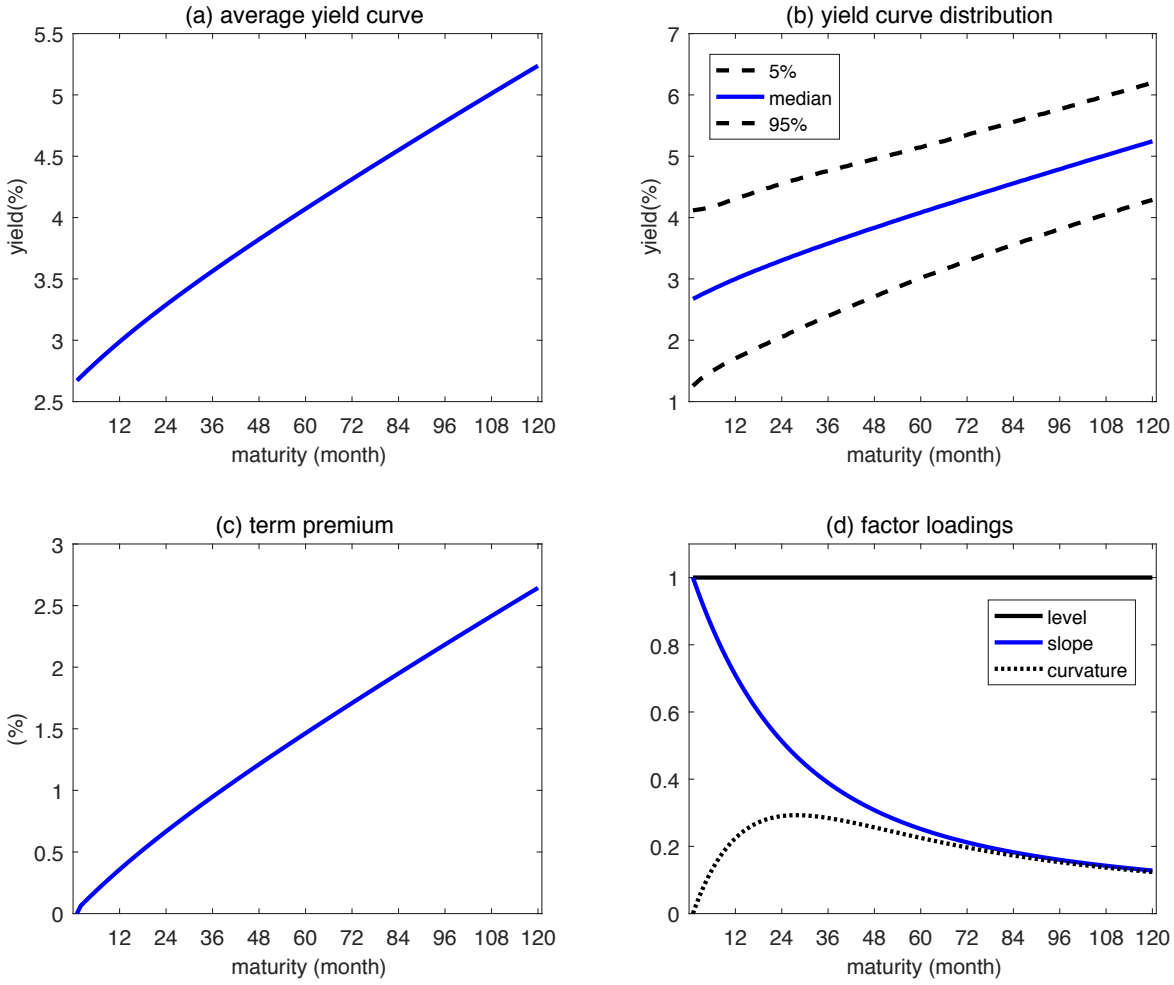
Inverse-Gamma distribution, $\mathcal{IG}(a,b)$		
parameter	a	b
$10^4 \times V_{1,1}$	66	130
$10^4 \times V_{2,2}$	102	252
$10^4 \times V_{3,3}$	4	1
$10^4 \times V_{4,4}$	146	435
$10^4 \times V_{5,5}$	146	435
$10^4 \times V_{6,6}$	146	435
$10^4 \times V_{7,7}$	146	435
$10^4 \times V_{8,8}$	146	435
$10^4 \times V_{9,9}$	146	435
$10^4 \times V_{10,10}$	146	435
$10^4 \times V_{11,11}$	146	435
$5 \times 10^5 \times \sigma_i^2$	54	260

element of λ_l,

$$\begin{aligned} \pi(\delta) &= St(\delta|\bar{\delta}, V_\delta, \nu),\ \pi(\kappa) = St(\kappa|\bar{\kappa}, V_\kappa, \nu), \\ \text{and } \pi(\lambda_l) &= \prod_{i=1}^{3} St(\lambda_{l,i}|\bar{\lambda}_{l,i}, V_{\lambda,i}, \nu). \end{aligned}$$

Figure 1.4: **Prior-implied Yield Curve**

This figure plots average yield curve, yield curve distribution, average term premium, and factor loadings based on 300 periods of simulated prior draws generated from the model which includes four relevant macro factors.



Initial Prior for $\theta_e = (\sigma_1^2, \sigma_2^2, .., \sigma_N^2)$: θ_e governs the size of the pricing errors. In our

initial prior, these error variances are assumed to have inverse-gamma prior distributions with small prior means

$$\pi\left(\theta_e\right) = \prod_{i=1}^{N} \mathcal{IG}\left(\sigma_i^2 | a_i, b_i\right).$$

Given independence across the three blocks of parameters, our initial prior density $\pi(\theta)$ is

$$\pi(\theta) = \pi\left(\theta_f\right) \pi\left(\theta_s\right) \pi(\theta_e).$$

Step 2: This initial prior from Step 1 is updated with training sample data for each of the 256 models by MCMC posterior sampling. The output from these 256 training sample estimations are used to arrive at the model-specific priors. In particular, the means of these draws are used to fix the marginal locations of the prior distributions. Standard deviation of these draws is scaled upwards by a factor of 4 to fix the standard deviation of the model-specific prior. In addition, to further avoid the possibility of likelihood-prior conflict (say on account of incorporating the preceding theoretical beliefs), the distribution of all location-type parameters is taken to be student-t with 2.5 degrees of freedom. Given these model-specific priors, estimation and model comparisons are then based on the remaining portion of the data.

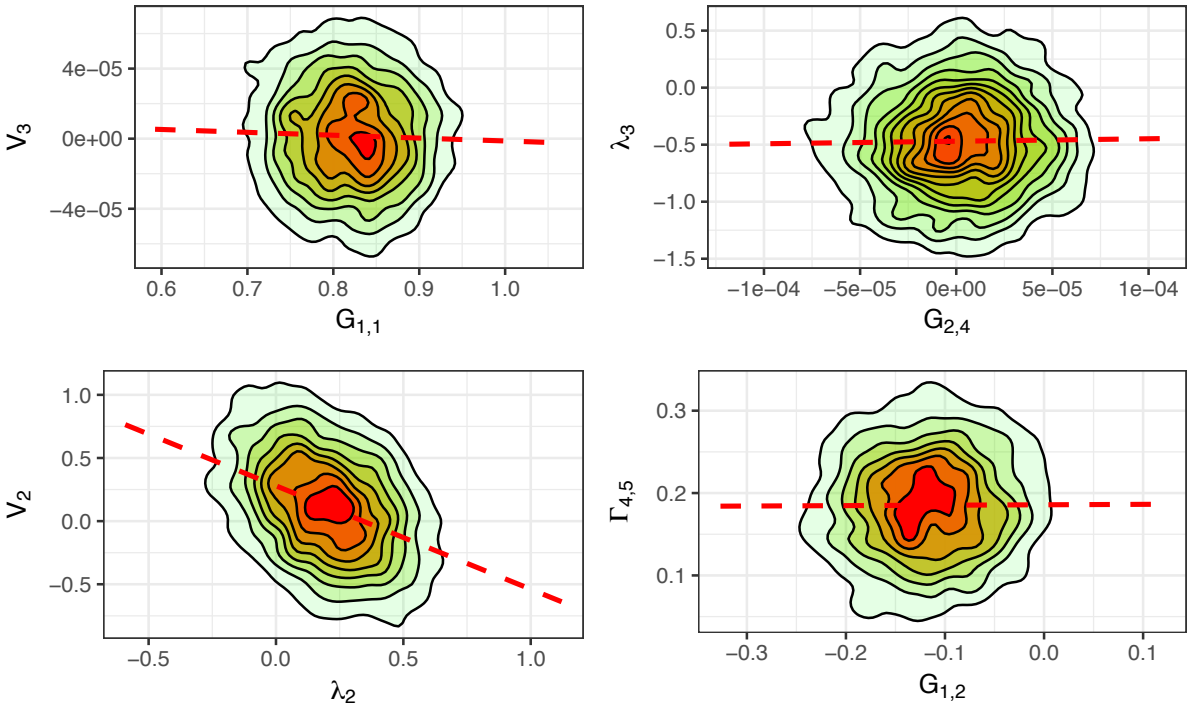
Figure 1.5 shows the posterior surface for the same pair of the parameters in Figure 1.3. This figure clearly reveals that our model-specific prior distributions updated from the training sample mitigate the irregularity remarkably, smoothing out the unlikely regions of the parameter space.

1.4.3 Marginal Likelihood

We compare the models with different relevant macro factors by using their posterior probabilities. Let $\mathsf{Pr}(\mathcal{M}_j)$ and $\mathsf{Pr}(\mathcal{M}_j|\mathbf{Y})$ denote the prior and posterior probability of model $\mathcal{M}_j$,

Figure 1.5: **Posterior Surface**

This figure plots bivariate posterior surfaces for some selected pairs of parameters from the model with the relevant macro factors, (FFR, HSal, IP, and TCU). The surface plots are based on 10,000 posterior draws after a burn-in of 1,000 draws. The sample periods for generating the surface are from January 1990 to December 2017.



respectively. Then the posterior probability of model $\mathcal{M}_j$ is given by

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

We assume the uniform prior on the models, so that the posterior probability of each model is determined by the relative size of the marginal likelihoods. As is well-known, marginal likelihoods automatically penalize the complexity of models. In our framework, a model with more macro factors in $\mathbf{m}_t$ (i.e. fewer zero restrictions on G_{lz}) is not necessarily selected by the marginal likelihood comparison. We calculate the marginal likelihood of each competing model, $\mathcal{M}$ by the method of Chib (1995), which is based on the identity,

$$\log m(\mathbf{Y}|\mathcal{M}) = \log p(\mathbf{Y}|\theta^*, \mathcal{M}) + \log \pi(\theta^*, \mathcal{M}) - \log \pi(\theta^*|\mathbf{Y}, \mathcal{M}). \tag{1.4.4}$$

Technical details of the marginal density computation can be found in Appendix .4.

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 apply the training sample procedure described above to construct the model-specific prior for each candidate model. 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. The simulation results indicate that the marginal likelihood comparison selects the true model with a high probability. Further details are provided in Appendix .5.

1.5 Empirical Results

Our empirical analysis involves the eight macro factors mentioned in Section 1. Instead of a apriori asserting the identity of relevant macro factors, our model comparison approach finds the best supported relevant factors by confronting the $2^8 = 256$ possible model specifications with the data, accessing the relative worth of each model in terms of the marginal likelihood. All marginal likelihoods are computed from 10,000 draws of the MCMC algorithm, collected after an initial burn-in period of 1,000 iterations.

1.5.1 Relevant Macro Factor Selection

Table 2.4 reports the results for the model comparison, where the model with no relevant macro factors, denoted by $\mathcal{M}_0$, is the benchmark, and models $\mathcal{M}_A$, $\mathcal{M}_B$, and $\mathcal{M}_C$ are the three best models selected from the marginal likelihood comparison. These three models are strongly preferred over the benchmark model in terms of the standard Jeffreys' scale, supporting the use of relevant macro-factors in our yield curve modeling. The marginal likelihoods also indicate that the group of the relevant macro factors most supported by the

evidence consist of FFR, Hsal, IP, and TCU.

Table 1.6: **Marginal Likelihood Comparison**

$\mathcal{M}_A$, $\mathcal{M}_B$, and $\mathcal{M}_C$ are the first three best models and $\mathcal{M}_0$ is the benchmark model containing no relevant macro factors. *log Bayes factor* is the log Bayes factor of each model relative to the benchmark model. **m** and **z** indicate relevant and irrelevant macro factors, respectively. The sample periods are from January 2001 to December 2017.

	$\mathcal{M}_A$	$\mathcal{M}_B$	$\mathcal{M}_C$	$\mathcal{M}_0$
AHE	**z**	**z**	**z**	**z**
CPI	**z**	**z**	**z**	**z**
IP	**m**	**m**	**z**	**z**
PCE	**z**	**z**	**z**	**z**
FFR	**m**	**z**	**m**	**z**
Unem	**z**	**m**	**m**	**z**
TCU	**m**	**m**	**m**	**z**
HSal	**m**	**m**	**m**	**z**
log Bayes factor	245.16	223.90	215.08	0.00

1.5.2 The Source of Marginal Likelihood Improvement

Given that the marginal likelihood is defined as the joint density of the yields and macro factors, marginalized over the parameters, it is useful to consider the source of the marginal likelihood gain from incorporating relevant macro-factors. For this purpose, we decompose the log likelihood gain of a particular model relative to the benchmark, denoted by $\Delta \log lik(\mathbf{R}, \mathbf{M}, \mathbf{Z}|\theta^*)$, as $\Delta \log lik(\mathbf{R}|\mathbf{M}, \mathbf{Z}, \theta^*) + \Delta \log lik(\mathbf{M}, \mathbf{Z}|\theta^*)$, where $\mathbf{R}$, $\mathbf{M}$, and $\mathbf{Z}$ are $\{\mathbf{R}_t\}_{t=1}^T$, $\{\mathbf{m}_t\}_{t=1}^T$, and $\{\mathbf{z}_t\}_{t=1}^T$, respectively. $\Delta \log lik(\mathbf{R}|\mathbf{M}, \mathbf{Z}, \theta^*)$ can be interpreted as the gain from fitting the entire yield curve, and $\Delta \log lik(\mathbf{M}, \mathbf{Z}|\theta^*)$ is the gain from fitting the macro factors. As reported in Table 1.7, the marginal likelihood gain comes mostly from the likelihood rather than the prior information, because the log likelihood gain for $\mathcal{M}_\mathcal{A}$ is 242.47, and this is almost equal to the log marginal likelihood gain of 245.16. More importantly, the log likelihood gains are obtained mostly from the gain of yield curve fit rather than macro factors fit, since $\Delta \log lik(\mathbf{R}|\mathbf{M}, \mathbf{Z}, \theta^*)$ is 244.45, while $\Delta \log lik(\mathbf{M}, \mathbf{Z}|\theta^*)$ is -1.88. Consequently, the large marginal likelihood gains are driven by the better yield curve fit.

Table 1.7: **Likelihood Decomposition**

This table reports the log likelihood differences between each of the three best models and the benchmark model. The total log likelihood difference, $\Delta \log lik(\mathbf{R}, \mathbf{M}, \mathbf{Z}|\theta^*)$ can be decomposed as $\Delta \log lik(\mathbf{R}|\mathbf{M}, \mathbf{Z}, \theta^*) + \Delta \log lik(\mathbf{M}, \mathbf{Z}|\theta^*)$ where $\Delta \log lik(\mathbf{R}|\mathbf{M}, \mathbf{Z}, \theta^*)$ is the log likelihood difference for the yield data conditional on macro data, and $\Delta \log lik(\mathbf{M}, \mathbf{Z}|\theta^*)$ is the log likelihood difference for the macro factors. The sample period is from January 2001 to December 2017.

	$\mathcal{M}_A$	$\mathcal{M}_B$	$\mathcal{M}_C$	$\mathcal{M}_0$
$\Delta lik(\mathbf{R}, \mathbf{M}, \mathbf{Z}\|\theta^*)$	242.47	239.62	243.93	0.00
$\Delta lik(\mathbf{R}\|\mathbf{M}, \mathbf{Z}, \theta^*)$	244.35	243.07	246.91	0.00
$\Delta lik(\mathbf{M}, \mathbf{Z}\|\theta^*)$	-1.88	-3.45	-3.45	0.00

1.5.3 Real Activity and Yield Curve

The most relevant macro factor combination chosen from our model comparison is found to be the IP, FFR, TCU, and HSal. This is a very interesting finding, in sense that three of them are intimately related to the real activity, whereas the inflation factors including the CPI and PCE are not selected. In fact, our finding is consistent with that of Ludvigson and Ng (2009b). Of all the risk factors they estimate, the most important one is highly correlated with measures of real output, not prices.

It is not surprising that the FFR directly controlled by the central bank is chosen, because it moves closely with the short end of the yield curve. The reason that the real activity factors are chosen seems to be related with the stylized fact that the term premium tends to be counter-cyclical, as documented by a number of previous studies such as Ludvigson and Ng (2009b), Joslin et al. (2014), and Wachter (2006). In addition, a key link between the real activity and the yield curve is the monetary policy, so that the current state of real activity affects the market expectation about the future short-term interest rates. However, the business cycle driving variables are latent and impossible to summarize with one or two observable indicators. Among the eight macro factors, the IP, TCU, and HSal turn out to be most relevant, because these macro factors are most successful in detecting changes in the market expectations and the counter-cyclical behaviour of the term premium. This is statistically confirmed by the marginal likelihood comparison.

Regarding the HSal, when the economy is contracting, less people can afford to purchase houses, and home values are expected to fall. This gives home buyers a less incentive to purchase a home. Due to the downward price stickiness, a decline in demand causes the declines in sales rather than falls in house prices. Hence, the decrease in the house sales can be regarded as an early warning sign of a recession. Consequently, the relevant macro factors are useful for forecasting future yields and estimating bond risk premia along with the level, slope, and curvature factors.

We now interpret the finding that the price factors are irrelevant. This finding does not necessarily implies that the market participants do not react changes in inflation. Instead, the information about the inflation or expected inflation are already contained in the latent factors, so that the observed inflation factors can be spanned by the latent yield curve factors. This is because, unlike the real activity, the inflation can be measured by a small number of observations such as CPI or PCE inflation rates. Another reason might be the mixed empirical findings for the inflation risk premium, while the counter-cyclical property of the term premium is robust. For instance, Evans (2003) finds a downward sloping inflation risk premium, but this is negative for most maturities. Hoerdahl, Tristani, and Vestin (2008) argue that inflation risk premium is positive and small around one-year maturity and essentially zero for all other maturities. As a result, the impact of the inflation on the risk premia is relatively unclear compared to the real activity.

Finally, despite the strong evidence of those four relevant macro factors, we would like to remark that one possible reason for the particular macro factor classification is a model misspecification. As Ludvigson and Ng (2009b) point out, all term structure models including our models are imperfect approximation of the true data generating process, so the small number of latent yield curve factors can fail to span the information of the observed macroeconomic factors that bond market participants use in their decision-making. Consequently, depending on how the latent factors are identified and extracted from the yield curve data,

the most relevant macro factors can be different. In practice, it is always unknown how much informative the latent factors are, and we need to examine whether macro variables can help improve the in-sample fitting or out-of-sample prediction.

1.5.4 Mixing

For completeness, we summarize the marginal posterior distributions of the parameters in the best model $\mathcal{M}_A$ in Table 1.8. The TaRB-MH acceptance rate is around 55.89% and the inefficiency factor is 19.87 on average, indicating that our MCMC chains are well mixed.

1.5.5 Term Premium

Using the analytical expression for the term premium in Equation (1.2.15), this section provides an empirical investigation of the term premium dynamics estimated from the best model $\mathcal{M}_A$. Figure 1.6 presents the posterior estimates of the dynamics term structure of term premiums. Note first that the term premiums varies 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, which is consistent with the empirical work of Campbell and Shiller (1991).

Second, more importantly, the term premiums seem to be counter-cyclical. For example, as shown in Figure 1.6, the term premium of the five-year bond yield tends to rise during recessions. This is because bond investors require larger compensation for bearing risks during recessions. This may partly explain the sharp increase in the term premium during the 2007 financial crisis. Another reason for the counter-cyclical term premium is the asymmetric

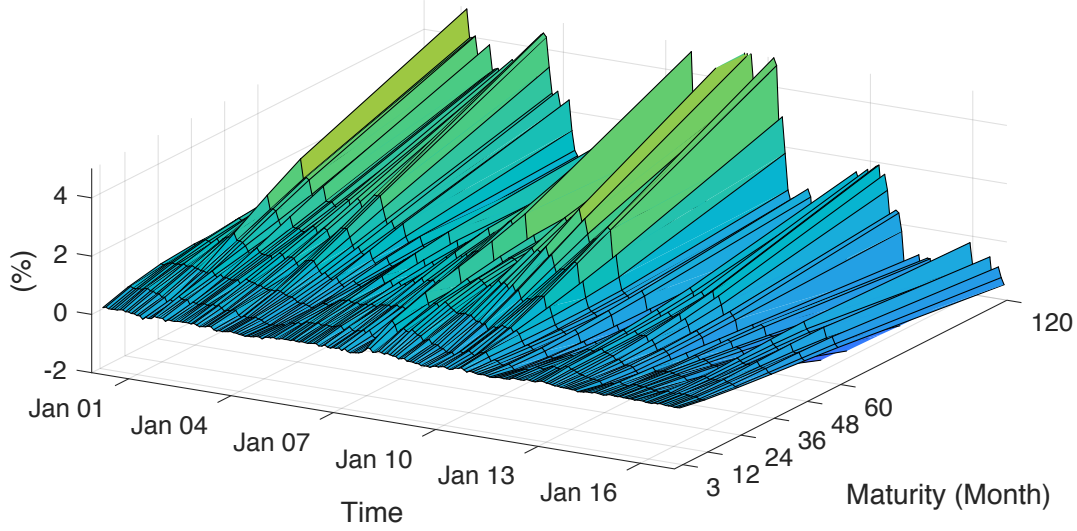
Table 1.8: **Posterior Summary for Selected Parameters in model $\mathcal{M}_A$**

This summary is based on 10,000 MCMC draws beyond a burn-in of 1,000. 2.5% , and 97.5% refer to 0.025 and 0.975 quantiles of the simulated draws, respectively. *mean* is the posterior mean, and *ineff* refers to the inefficiency factor. The sample periods are from January 2001 to December 2017.

Parameters	2.5%	mean	97.5%	accept.rate	ineff
$G_{1,1}$	0.78	0.89	0.93	100%	4.4
$G_{2,2}$	0.81	0.94	1.07	100%	5.80
$G_{3,3}$	0.82	0.89	0.92	100%	4.86
$G_{4,4}$	0.81	0.86	0.91	100%	5.01
$G_{5,5}$	0.79	0.86	0.93	100%	3.86
$G_{6,6}$	0.64	0.70	0.75	100%	6.05
$G_{7,7}$	0.79	0.86	0.92	100%	3.58
$G_{8,8}$	0.77	0.83	0.89	100%	3.44
$G_{9,9}$	0.93	0.97	1.02	100%	5.22
$G_{10,10}$	0.88	0.91	0.94	100%	4.26
$G_{11,11}$	0.94	0.96	0.99	100%	3.49
λ_1	-0.07	-0.07	-0.06	58.68%	15.76
λ_2	-0.22	-0.10	-0.01	57.3%	6.73
λ_3	-0.07	0.01	0.09	50.59%	17.28
$V_{1,1}$	1.88	2.14	2.44	58.55%	20.30
$V_{2,2}$	2.22	2.50	2.84	59.21%	14.80
$V_{3,3}$	5.2	5.8	6.47	59.84%	8.11
$V_{4,4}$	2.68	3.01	4.10	58.99%	5.65
$V_{5,5}$	3.96	4.40	4.93	58.93%	7.80
$V_{6,6}$	0.86	1.01	1.14	53.33%	15.80
$V_{7,7}$	1.27	1.48	1.76	58.96%	5.12
$V_{8,8}$	2.17	2.37	2.62	54.24%	7.86
$V_{9,9}$	5.87	6.48	7.20	56.85%	7.01
$V_{10,10}$	3.12	3.44	3.83	54.23%	6.47
$V_{11,11}$	0.98	1.22	1.47	58.57%	9.68
κ	0.06	0.09	0.07	57.63%	8.76

Figure 1.6: **Term Premium**

This figure plots the term structure of the term premiums estimated from $\mathcal{M}_{\mathcal{A}}$. The term premiums are computed based on Equation (2.4.2). The relevant macro factors in the model are FFR, Hsal, IP, and TCU. The sample periods are from January 2001 to December 2017.

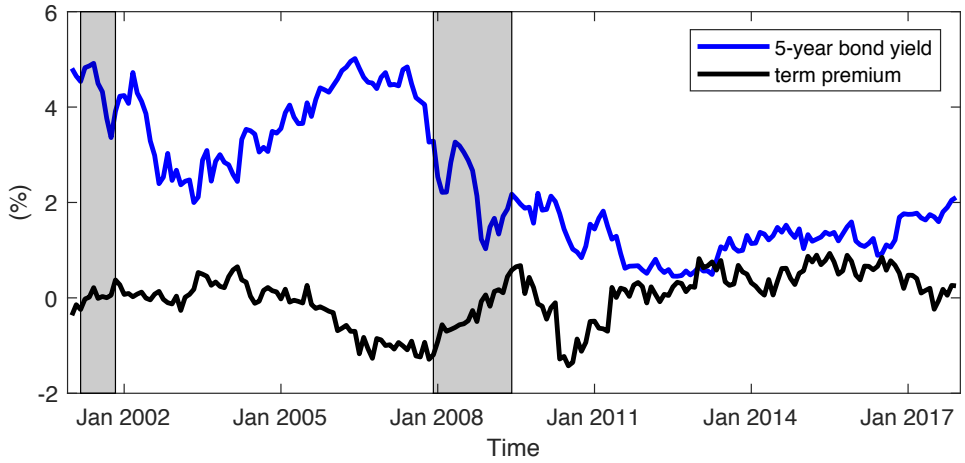


monetary policy. As demonstrated by Kung (2015), more aggressive inflation targeting reduces nominal risks, and it lowers the average nominal term spread. In recession, the inflation targeting tends to be less aggressive, which increases nominal risks and results in a higher term premium.

To quantify the role of relevant macro factors in determining the term premium, we plot the posterior estimates of its time-varying components in Figure 1.8 based on Equation (1.2.16). From the figure, we can see that the time-varying components are also more volatile for longer maturities. In addition, Figure 1.9 presents the time series of the time-varying component for the 60-month yield. Its dynamics appears to be counter-cyclical over time, as its correlation with the IP growth is -0.31. Our findings indicate that the counter-cyclical term premiums do not seem to be sufficiently explained by the level, slope, and curvature factors. In periods of recessions, the relevant macro factors enable us to capture an increase of the term premium, contributing to a steepening of the yield curve. Conversely, the relevant macro factors contribute to a flattening of the yield curve by decreasing the term premium.

Figure 1.7: **60-month Yield and its Term Premium**

This figure plots the 60-month bond yield and its term premium estimated from $\mathcal{M}_{\mathcal{A}}$. The relevant macro factors in the model are FFR, Hsal, IP, and TCU. Shaded areas indicate the NBER recession periods and the sample periods are from January 2001 to December 2017.



Consequently, it is important to incorporate the information of real activity measures that can help properly capture the changes in risk premia.

1.5.6 Out-of-Sample Predictive Performance

We now evaluate the predictive ability of the model with relevant macro factors in terms of out-of-sample yield curve density prediction. The predictive accuracy of density forecasts is typically measured by the log predictive likelihood (PL), which is the sum of log posterior predictive density (PD) evaluated at the realized yield curves. Suppose that $\mathbf{R}^o_{T+1}$ denotes the realized yield curve at time $T+1$, and T_0 and T_1 indicate the starting and end dates of the out-of-sample period, respectively. Then, the log PL of a particular model, $\mathcal{M}$ is calculated as

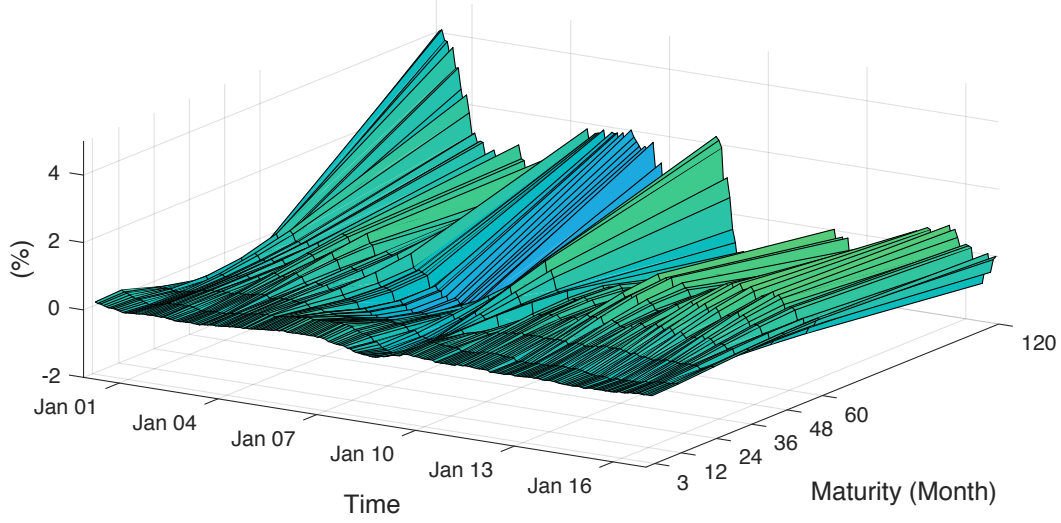
$$\log PL(\mathcal{M}) = \Sigma_{T=T_0}^{T_1} \log p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \mathcal{M}),$$

and the PD, $p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \mathcal{M})$, is defined as

$$p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \mathcal{M}) = \int p(\mathbf{R}^o_{T+1}, \psi, \Sigma, \ \mathbf{L}|\mathbf{Y}, \mathcal{M}) d(\psi, \Sigma, \mathbf{L})$$

Figure 1.8: **Time-varying Component of Term Premium Driven by $\mathbf{m}_t$**

This is a three-dimensional plot of the time-varying component of the term premiums driven by the macro factors. The time-varying component is computed based on Equation (1.2.16). The relevant macro factors are FFR, HSal, IP, and TCU. The sample periods are from January 2001 to December 2017.



$$= \int p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \mathcal{M}, \theta, \Sigma, \mathbf{L}, \mathcal{M})\boldsymbol{\pi}(\psi, \Sigma, \mathbf{L}|\mathbf{Y}, \mathcal{M})d(\psi,\ \Sigma, \mathbf{L}).$$

where $\mathbf{L} = \{\boldsymbol{l}_t\}_{t=1}^T$ is the time series of the latent factors.

The predictive density ordinate is obtained by simulation as

$$p(\mathbf{R}^o_{T+1}|\mathbf{Y},\mathcal{M}) \approx \frac{1}{n_1}\sum_{g=1}^{n_1} p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \psi^{(g)}, \Sigma^{(g)}, \mathbf{L}^{(g)}, \mathcal{M}) \tag{1.5.1}$$

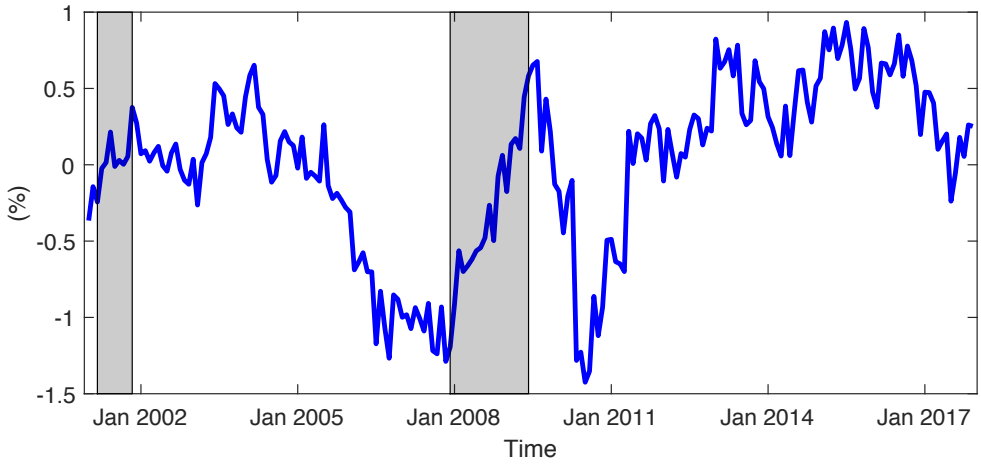
where $(\psi^{(g)}, \Sigma^{(g)},\ \mathbf{f}_T^{(g)})$ are the posterior draws from the gth MCMC cycle, $\bar{\boldsymbol{l}}^{(g)}_{T+1}$ is the first l elements of $\bar{\mathbf{f}}^{(g)}_{T+1} = G^{(g)}\mathbf{f}^{(g)}_T$, $\Omega^{(g)}_{1:k,1:k}$ is the first $k \times k$ elements of $\Omega^{(g)}$, $\Sigma^{(g)}_{1:N,1:N}$ is the first $N \times N$ elements of $\Sigma^{(g)}$, and

$$p(\mathbf{R}^o_{T+1}|\mathbf{Y}, \psi^{(g)},\ \Sigma^{(g)}, \mathbf{L}^{(g)}, \mathcal{M} \equiv \mathcal{N}\left(\mathbf{R}^o_{T+1}|\bar{\boldsymbol{a}}^{(g)} +\ \bar{\mathbf{B}}^{(g)}\bar{\boldsymbol{l}}^{(g)}_{T+1}, \bar{\mathbf{B}}^{(g)}\Omega^{(g)}_{1:k,1:k}\bar{\mathbf{B}}^{(g)\prime} + \Sigma^{(g)}_{1:N,1:N}\right).$$

We consider four models in our out of sample forecasting comparison. The selected models

Figure 1.9: **Time-varying Component of Term Premium of 60-month Yield Driven by $\mathbf{m}_t$**

This figure plots the time series of the time-varying component of the 60-month bond driven by the relevant macro factors. The relevant macro factors are FFR, HSal, IP, and TCU. Shaded areas indicate the NBER recession periods and the sample periods are from January 2001 to December 2017.



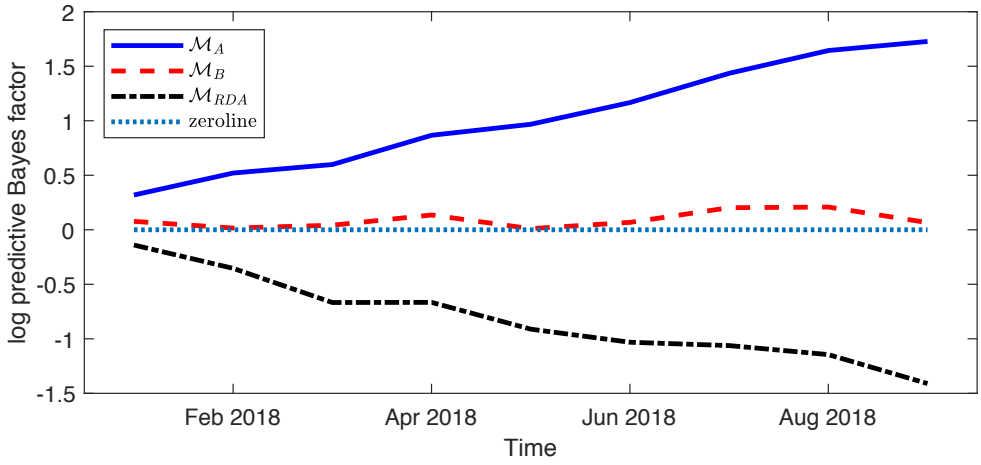
are the best and second best models chosen by the marginal likelihood, the dynamic Nelson-Siegel model from Diebold et al. (2006), which is denoted by $\mathcal{M}_{DRA}$, and the benchmark model $\mathcal{M}_0$. In $\mathcal{M}_{DRA}$, no-arbitrage conditions are not imposed, and the macro factors are CPI, FFR, and TCU. The out of sample period consists of the nine months starting from January 2018 and ending in September 2018.

Using the log PLs of the models, we can compute the log predictive Bayes factor, which is the log difference in the PL between two models. Figure 1.10 reports the cumulative log predictive Bayes factors of the first three models in comparison with $\mathcal{M}_0$ over time. We can see that $\mathcal{M}_A$ consistently outperforms the other models in terms of the out-of-sample predictive density accuracy. The log PL difference can accumulate up to three at the end of out-of-sample periods, showing that the model favored by marginal likelihood also has advantage in out-of-sample prediction. Given that the shape of the yield curve is determined by the three latent yield curve factors, this finding suggests that one-period-ahead latent factors are driven by by the current relevant macro factors as well as by the

current latent factors. Meanwhile, the second best model $\mathcal{M}_{\mathcal{B}}$ can only outperform $\mathcal{M}_0$ for later periods of the sample, while $\mathcal{M}_{DRA}$ offers little prediction gain given the data we analyze here. This empirical finding demonstrates the importance of the relevant macro factor choice in improving the prediction ability of macro-finance term structure models.

Figure 1.10: **Cumulative Log Predictive Bayes Factor**

This figure shows the sequences of the cumulative log predictive bayes factors with respect to the benchmark model, $\mathcal{M}_0$. The blue solid line corresponds to $\mathcal{M}_A$, the red dashed line indicates $\mathcal{M}_B$, and the black dotted line denotes $\mathcal{M}_{RDA}$. The out-of-sample data range from January 2018 to September 2018.



1.6 Conclusion

In this paper, we introduce a Bayesian framework for selecting relevant macro factors in a macro-finance affine term structure model. We show that the models with different set of potentially relevant macro factors can be formulated by different zero restrictions on the vector-autoregressive coefficient matrix, G. By comparing the restrictions on G in terms of the marginal likelihood, we are able to ascertain the identity of the best supported relevant macro factors. To implement this idea in practice, we propose an efficient MCMC algorithms for estimation and marginal likelihood computation. Our empirical analysis based the U.S. yield curve data suggest that the federal funds rate, industrial production, total capacity

utilization and housing sales are the most relevant macro factors among the eight alternative macro factors we consider.

We conclude by pointing out two 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 factors to change over time. Second, it would be worthwhile to allow the factor shock variance-covariance matrix to change over time. Both extensions are challenging, and are left for future research.

Chapter 2

An Economic Evaluation of Unspanned Macro Factors in Affine Term Structure Model

2.1 Introduction

Both central banks and financial investors pay critical attention to term structure of interest rates because the yield curve dynamics can shed light on both monetary policy and portfolio management. While the linkage between predictive yield curve and macro-variables has been widely documented in the literature(See Ludvigson and Ng (2009b),Cooper and Priestley (2009), Greenwood and D.Vayanos (2014),and Cieslak (2015),etc.), there is so far little evidence in support that such predictability could be used to improve investors' economic utility.

This paper contributes to the literature by raising the question of identifying which group of relevant factors can best improve the forecasts of bond returns and generate economic gains for the investors under an affine term structure framework. To address this, we adopt an

idea similar to Chib et al. (2019) to group the macro factors into two categories: The first category contains relevant macro factors that enter into the model while the second contains unnecessary factors hypothesized to be redundant and do not enter into the model. We will consider two main kinds of macro factors in current literature: representative macro factors(Chib et al. (2019)) and macro factors generated using principle component (Ludvigson and Ng (2009b)). More specifically, Chib et al. (2019) considers 8 representative macro factors while Ludvigson and Ng (2009b) considers 8 principle component macro factors generated from macro panel dataset. We will evaluate the macro factors in these two papers in separate experiments. Because 8 candidate macro amounts to a trove of $2^8 = 256$ competing models. In each experiment, we can evaluate macro factors based on both statistical and economic evaluations through an out-of-sample experiment and identify the best model(and thus the best supported composition of macro factors) given 256 different specifications of the model(different macro factors incorporated in the model). In addition, we can also compare across two experiments to see what kind of macro factors(in our context either representative macro factors or principle macro factors) will perform even better.

We demonstrate that the best model includes average hourly earning(AHE) federal fund rate (FFR), consumer price index(CPI) and total capacity utility (TCU) corresponding to macro dataset in Chib et al. (2019) while the best model includes principle component factors F_1(loads heavily on employment and output data),F_2,(loads heavily on interest rate spreads) and F_4(loads on prices) corresponding to macro dataset in Ludvigson and Ng (2009b) . Obviously, the set of relevant factors included in best models for two different datasets resemble to each other, indicating that the most relevant factors are related to real economy, interest rate and price level.

The statistical model comparison is conducted based on the posterior predictive likelihood(PPL), which measures the predictive density accuracy of the yield curve. The data used in the empirical study are U.S monthly yield curve data from January 1990 to Decem-

ber 2017. The yield curves for the most recent 120 months(10 years) are predicted. We repeat the predictive density simulation to compute the PPL by rolling the out-of-sample period by one month and we obtain a sequence of 120 PPLs to examine whether the relative predictive performance is consistent over these 120 out-of-sample periods. The two best models corresponding to two different datasets(Chib, Kang, and Xie (2019) and Ludvigson and Ng (2009b)) mentioned above, consistently exhibits the largest PPLs over the out-of-sample periods among their competing models. In addition, we also find that the best model corresponding to dataset in Chib, Kang, and Xie (2019) performs better (by a margin of $3-5$ in terms of log predictive likelihood) than that in Ludvigson and Ng (2009b).

In additional to statistical valuation, we investigate the economic value of the proposed model. We analyze the portfolio choice problem for a Bayesian risk-averse investor. We will use the certainty equivalent return(CER) to quantify the utility gain of taking different groupings of relevant macro factors into account. As new observations arrive in every time period, an investor with constant relative risk-aversion(CRRA) utility learns about the parameters and factors via the Bayesian update. Then he or she updates the posterior predictive joint distribution of bond returns accounting for uncertainties of parameters and factors. Then the investor optimizes the portfolio by choosing each bond's weight to achieve maximized expected utility. The procedure is repeated in every time period. As the joint density forecasts of bond returns are model-dependent, the optimal portfolio weights and realized portfolio returns are model-dependent,too. Consequently, the CER differs across models. A large CER implies the better portfolio performance and thus the model with the largest CER is preferred in economic evaluation.

We compute and compare the realized CERs for the out-of-sample period, January 2008 to December 2017, among the competing models. Consistent with the prediction comparison result, The two best models corresponding to the two datasets(Chib, Kang, and Xie (2019) and Ludvigson and Ng (2009b)) mentioned above, consistently exhibits the largest PPLs

over the out-of-sample periods among their competing models. In addition, we also find that the best model corresponding to dataset in Chib et al. (2019) performs slightly better than that in Ludvigson and Ng (2009b) (by a margin of $0.5 - 1$ in terms of CER).

The results of our empirical results show that incorporating representative macro factor FFR,AHE,CPI and TCU in Chib, Kang, and Xie (2019) or principle component F_1, F_2, F_4 in Ludvigson and Ng (2009b) into the affine term structure model generates both statistical and economical gains. Furthermore, our empirical result shows that representative macro factors in Chib, Kang, and Xie (2019) can bring slightly better economic gain when compared to the principle component macro factor in Ludvigson and Ng (2009b).

Our paper is related to literature discussing bond return predictability and its economic value. For example, some studies document significant in-sample predictability of treasury bond excess returns for $2 - 5$ year maturities by means of yield spreads(Campbell and Shiller (1991)), a linear combination of forward rates(Cochrane and Piazzesi (2005)) or factors extracted from a cross section of macroeconomic variables(Ludvigson and Ng (2009b)). Both Thornton and Valente (2012) and Sarno, Schneider, and Christian (2016) show that return predictability may not necessarily generate significant economic gains but Gargano, Pettenuzzo, and Timmermann (2017b) reconcile the contradiction by accounting time-varying parameters and stochastic volatility. Building on this literature, our paper will account the no-arbitrage condition and evaluate different macro factors under an arbitrage-free term structure framework. Our paper also contributes to literature on macro-finance term structure models such as Duffee (2002),Duffee (2011), Joslin et al. (2014) and Chib et al. (2019),etc. Our paper extends to discuss the economic value of these models.

The remainder of the paper proceeds as follows. In Section 2.2 we describe the data that is used in our empirical analysis. Section 3.2 illustrates our bond pricing model. Section 2.4 discusses detailed Bayesian estimation strategy and predictive likelihood computation method used in this paper. Section 2.5 discusses how to measure portfolio performances.

Section 2.6 provides the empirical results, and concluding remarks are presented in Section 2.7.

2.2 Data

2.2.1 Bond Returns

Previous studies such as Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009b) use overlapping 12-month returns data and this overlap induces strong serial correlation in estimation residuals. to handle this issue, We follow Gargano, Pettenuzzo, and Timmermann (2017a) to reconstruct the yield curve at the daily frequency starting from the parameters estimated by Gurkaynaka, Sack, and Wright (2007). More specifically, the time t zero coupon yield on a bond maturing in τ periods, gets computed as:

$$y_t(\tau+1) = \beta_{0t} + \beta_{1t}\frac{1-exp(-\frac{\tau}{\tau_1})}{\frac{\tau}{\tau_1}} + \beta_{2t}(\frac{1-exp(-\frac{\tau}{\tau_1})}{\frac{\tau}{\tau_1}} - exp(-\frac{\tau}{\tau_1})) \tag{2.2.1}$$

$$+\beta_{3t}(\frac{1-exp(-\frac{\tau}{\tau_2})}{\frac{\tau}{\tau_2}} - exp(-\frac{\tau}{\tau_2})) \tag{2.2.2}$$

The parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$ are provided by Gurkaynaka et al. (2007), who reports daily estimates of the yield curve for the maturity range spanned by outstanding Treasury securities.

In our paper, we consider 14 maturities:$5, 6, 11, 12, 23, 24, 35, 36, 59$
$60, 83, 84, 119$ and 120. The yield data for maturities $6, 12, 24, 36, 60$ and 120 months provided by the U.S treasury department. We construct the rest yields using 2.2.1. Among them, the $5, 11, 23, 35, 59, 83$ and 119 month maturities are used for the posterior estimation of the parameters and we denote τ-month bond yield such that

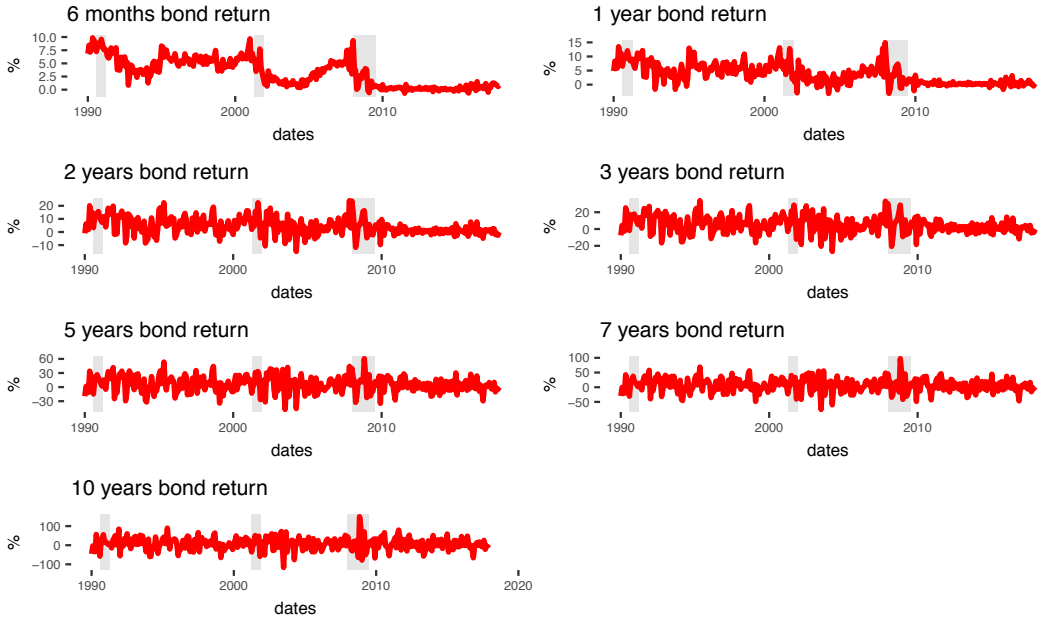
$$\tau \in \tau_{all} = \{5, 11, 23, 59, 83, 119\}$$

When constructing bond portfolio, we invest in the $(\tau + 1)$-period bonds with $\tau \in \tau_{all}$, and use the observed $6, 12, 24, 36, 60, 84$ and 120 month bond yields to obtain the realized bond return.

Figure 2.1 plots monthly bond returns for 6 month, 1 year, 2 year, 3 year, 5 year, 7 year and 10 year maturities. All returns are pretty volatile during early 1990s and year 2007. Table 2.1 presents summary statistics for the bond return. Returns on shorter maturities tend to be right skewed and fat tailed. In addition, the volatility increases with the maturity of bonds.

Figure 2.1: **Bond Return**

This figure shows time series of monthly bond returns (in annualized percentage terms) for maturities ranging from 6 months to 10 years. Bond returns are computed in equation 2.4.4 Shaded areas indicate recession periods designated by the National Bureau of Economic Research. The sample period is from January 1990 to December 2017.



2.2.2 Macro factors

According to the literature, using macro factors as predictors for either yields or bond returns can be contentious. It is because the information in the current yield curve could

Table 2.1: **Summary Statistics of bond returns**

Bond returns are computed in equation 2.4.4 and the maturities range from 6 months to 10 years. The sample period is from January 1990 to December 2017.

Maturity	Mean	St.dev	Skewness	Kurtosis	min	max
6-months	3.20	2.70	0.35	1.95	-0.68	9.85
1-year	3.57	3.54	0.64	2.79	-3.14	14.91
2-year	4.29	6.38	0.41	3.30	-14.89	23.9
3-year	4.95	9.78	0.12	3.30	-26.72	33.79
5-year	6.05	16.48	-0.08	3.43	-47.66	60.54
7-year	6.90	22.69	-0.08	3.85	-75.76	97.27
10 year	7.8	31.42	-0.04	4.63	-116.73	149.65

already subsumes the macro predictors of future yields or returns and macro variables can be irrelevant when added to the model.

To resolve such an issue, we assume macro risk factors are unspanned just as models of Joslin et al. (2014). their modeling of macro factors offers to explain why macro variables could help predict yields and bond returns beyond the information contained in the yield curve. Intuitively speaking, their models suggest that the effect of additional macro variables on future short rates and expected future bond returns cancel out and leave the current yield curve unaffected[1] Such cancellation imply that these variables are "hidden" in bond yields (therefore unspanned by the current yield curve) even though they can have predictive power over the bond returns.

There are two main ways of constructing macro factors. one way is to incorporate representative macro variables into the affine term structure models (such as Chib et al. (2019) and Diebold et al. (2006)). The other way is to use dimension reduction method (such as principle component) to generate macro factors from large panel of macro variables(Ludvigson and Ng (2009b) and Piazzesi (2005)). Our paper will use both types of macro factors for empirical study and the details are discussed in following sections.

[1] current yield curve can be decomposed into two parts: future short rates and expected future bond returns

Table 2.2: **Summary Statistics of Macro Factors(Chib et al. (2019))**

	Mean	St.dev	Skewness	Kurtosis	min	max
AHE	2.89	0.75	0.05	2.01	1.12	4.32
CPI	2.44	1.27	-0.08	4.15	-2.01	6.18
IP	1.90	4.15	-1.92	8.34	-16.4	8.36
PCE	4.69	1.88	-1.81	8.54	-3.54	8.74
FFR	2.96	2.44	0.23	1.73	0.07	8.29
Unem	5.98	1.55	0.92	3.01	3.8	10
TCU	78.91	3.68	-0.67	3.51	66.8	85
HSal	-0.21	18.83	-1.01	3.78	-62.38	52.22

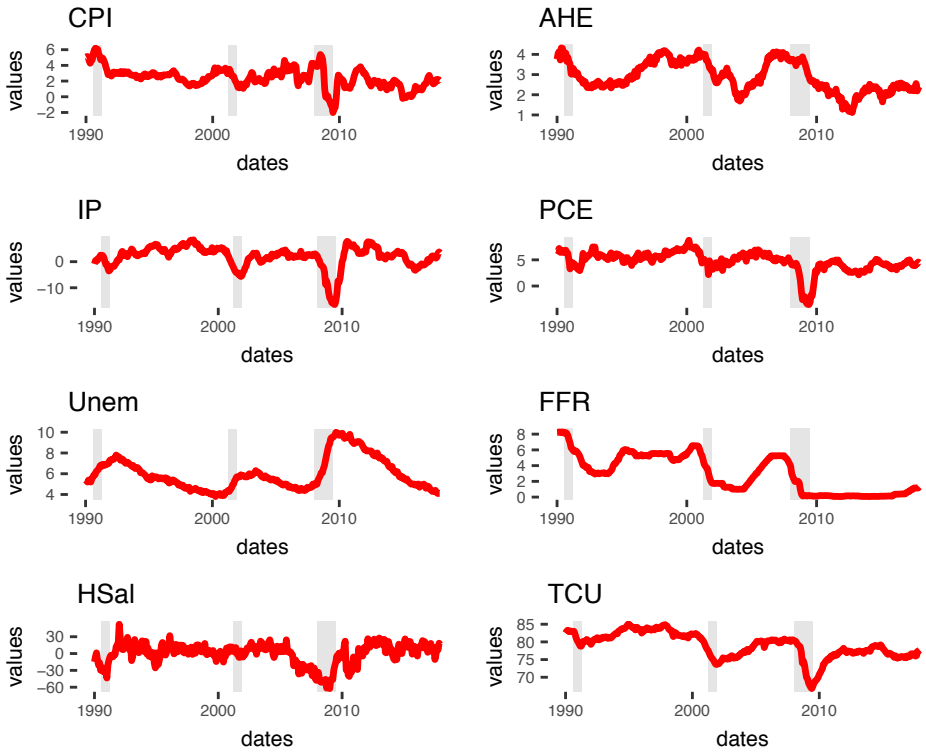
Macro factors in Chib et al. (2019)

As mentioned in the last section, one way of constructing macro factors for term structure model is to use representative macro variables directly. The advantage lies in that each variable is explainable and has economic meaning while its disadvantage is that the macro information can be limited by the number of incorporated macro variables.

Our choice of representative macro factors come from Chib et al. (2019). There are 8 candidate macro factors in this paper: two inflation measures, five real economic activity factors, and one financial factor. The inflation measures are consumer price index (CPI) and average hourly earning (AHE). The real activity is measured by industrial production (IP), personal consumption expenditures (PCE), capacity utilization (CU), unemployment rate (Unem), and housing sales (HSal). The federal fund rate (FFR) is a proxy of the financial factor. Table 2.2 provides a summary statistics of the representative macro factors. Time series of the macro factors are plotted in Figure 2.2. Obviously IP, PCE, TCU, Hsal, AHE, and CPI are procyclical while Unem and FFR are countercyclical.

Figure 2.2: **Macro Factors(Chib et al. (2019))**

This figure plots the 8 representative macro factors used in Chib et al. (2019). The data sample from January 1990 to December 2011. Shaded areas indicate the NBER recession periods.Representative Macro Factors come from macro data set



Macro factors in Ludvigson and Ng (2009b)

Ludvigson and Ng (2009b) proposes another way to construct macro factors by using principle component. More specifically, Ludvigson and Ng (2009b) supposes $T \times M$ panel of macroeconomic variables $\{x_{i,t}\}$ generated by a factor model as follows:

$$x_{i,t} = \kappa F_t + \epsilon_{i,t}$$

Then 8 factors are generated from a panel of 131 macroeconomic variables. Hence $F_t = (F_{t,1}, F_{t,2}, F_{t,3}, F_{t,4}, F_{t,5}, F_{t,6}, F_{t,7}, F_{t,8})$. According to Ludvigson and Ng (2009a), these factors have economic intuition behind them: $F_{t,1}$ is a real activity factor that loads on employment and output data; $F_{t,2}$ loads on interest rate spreads; while $F_{t,3}$ and $F_{t,4}$ load on prices; $F_{t,5}$

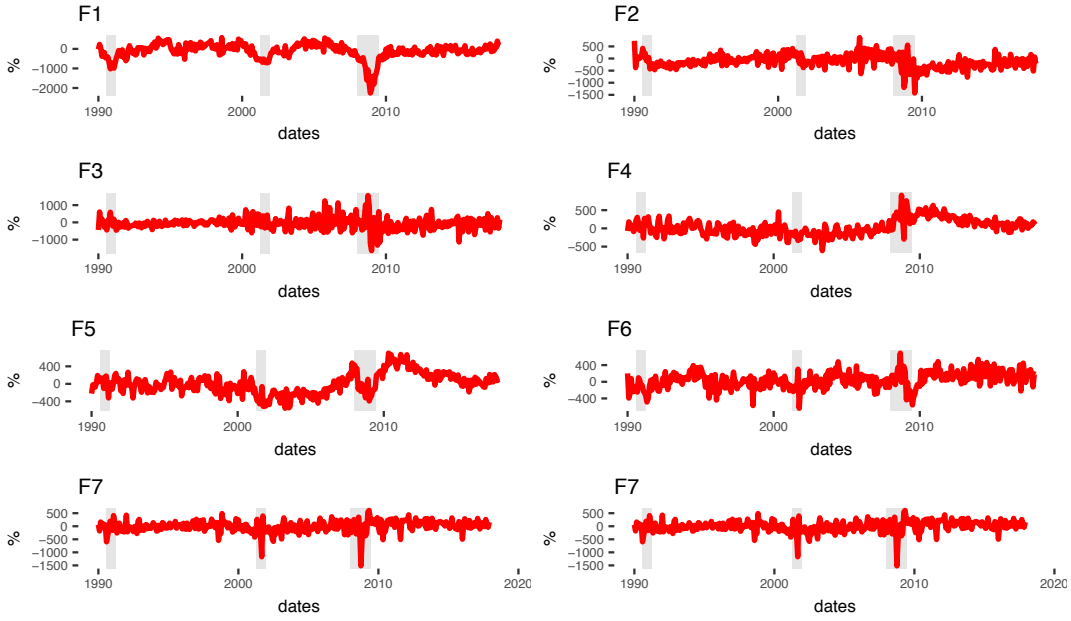
Table 2.3: **Summary Statistics of Macro Factors(Ludvigson and Ng (2009b))**

	Mean	St.dev	Skewness	Kurtosis	min	max
F1	0.08	0.32	1.89	9.48	-0.47	1.88
F2	0.09	0.22	0.28	5.05	-0.72	1.18
F3	0.05	0.29	-0.26	6.58	-1.29	1.36
F4	-0.03	0.18	-0.45	3.42	-0.75	0.5
F5	-0.005	0.19	-0.28	3.07	-0.58	0.46
F6	-0.03	0.16	0.22	3.47	-0.57	0.52
F7	0.00	0.18	1.56	11.62	-0.50	1.26
F8	-0.03	0.14	-0.06	3.31	-0.47	0.37

loads on interest rates and spreads; $F_{t,6}$ loads predominantly on the housing variables; $F_{t,7}$ loads on measures of the money supply; $F_{t,8}$ loads on variables relating to the stock market.

Figure 2.3: **Macro Factor (Ludvigson and Ng (2009b))**

This figure plots the 8 principle component macro factors used in Ludvigson and Ng (2009b). Shaded areas indicate the NBER recession periods. The data sample is from January 1990 to December 2011.



Finally, the data used in our empirical analysis runs from January 1990 to December 2017. The data prior to that period is subject to a structural break, as Chib and Kang (2013) discuss.

2.3 Bond Pricing Model

As mentioned in the previous section, we are focusing on finding the relevant unspanned macro factors that con convert return predictability into economic gains. Hence the arbitrage-free affine term structure model with unspanned macro factors developed by Joslin et al. (2014) and Duffee (2011) will be adopted for our analysis. In their model, factors $\mathbf{f}_t$ at time t, consisting of three yield curve factors $\boldsymbol{l}_t : 3 \times 1$ (given by the first three principal components of the yield curve), and macro economic factors $\mathbf{m}_t : m \times 1$, are assumed to follow a first-order Gaussian vector-autoregressive (VAR) process given by

$$\mathbf{f}_{t+1} = K + G\mathbf{f}_t + \varepsilon_{t+1} \ , \ \varepsilon_{t+1} \sim \mathcal{N}\left(0_{(3+m)\times 1}, \Omega\right)$$

In our paper, we will form $\mathbf{m}_t$ by a combination of macro factors .(macro factors are either chosen from Ludvigson and Ng (2009b) or from Chib et al. (2019)) These factors determine the one-period ahead stochastic discount factor (SDF) M_{t+1} as

$$M_{t+1} = \mathsf{exp}\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' v_{t+1}\right) \tag{2.3.1}$$

where r_t is the short rate that is assumed to be an affine function of $\boldsymbol{l}_t$, but not of $\mathbf{m}_t$, v_{t+1} are the iid factor shocks such that $\varepsilon_{t+1} = \Omega^{1/2} v_{t+1}$, where $\Omega = \Omega^{1/2}\Omega^{1/2\prime}$, and the time-varying market price of factor risks $\lambda_t : (3+m) \times 1$ is assumed to follow the process

$$\lambda_t = \left(\Omega^{1/2}\right)^{-1}\left(\begin{pmatrix} \lambda_l \\ \mathbf{0}_{m\times 1} \end{pmatrix} + \begin{pmatrix} \Lambda_{ll} & \Lambda_{lm} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{l_t} \\ \mathbf{m}_t \end{pmatrix}\right)$$

Joslin et al. (2014) model the role of $\mathbf{m}_t$ in this way to ensure that the bond price for a τ period bond at time t that satisfies the no-arbitrage condition

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

will depend just on $\boldsymbol{l_t}$, but not on $\mathbf{m}_t$. Technically, their modeling ensures that the macro-factors are unspanned by the yield curve. Even though the macro-factors do not directly affect the yield curve at t, Joslin et al. (2014) show that the macro factors, through their influence on λ_t, influence the dynamics of the term-premium, and aid in forecasting future values of the yield curve. Building on Joslin et al. (2014), we also require that the latent factors can be identified as the level, slope, and curvature of the yield curve, and impose the restrictions of Christensen et al. (2009) and Christensen et al. (2011).

Finally, by standard calculations given in Appendix .1, the bond prices that solve the no-arbitrage condition are exponentially affine in the latent factors,

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

for any positive integer value of τ, where $a(\tau)$ is a scalar and $b(\tau)$ is a $l \times 1$ vector given by the recursive difference equations,

$$\begin{aligned} a(\tau) &= \delta + a(\tau - 1) - \frac{1}{2}b(\tau - 1)'\Omega_{ll}b(\tau - 1) - b(\tau - 1)'\lambda_l, \\ b(\tau) &= \beta + G_{ll}^{\mathbb{Q}\prime}b(\tau - 1) \end{aligned} \tag{2.3.2}$$

2.4 Prior-Posterior Analysis

Similar to Chapter 1, in order to estimate the affine term structure model, we first represent the affine model in the following state space form:

$$
\begin{aligned}
R_t(\tau_i) &= \frac{a(\tau_i)}{\tau_i} + \frac{b(\tau_i)'}{\tau_i}\boldsymbol{l}_t + e_t(\tau_i),\ e_t(\tau_i) \sim \mathcal{N}(0, \sigma_i^2) \text{ for } i = 1, 2, .., N, \\
\mathbf{m}_t &= \mathbf{m}_t,
\end{aligned}
$$

or, in matrix form, as

$$
\mathbf{y}_t | \mathbf{f}_t, \mathbf{a}, \mathbf{B}, \Sigma \sim \mathcal{N}(\mathbf{a} + \mathbf{B} \times \mathbf{f}_t, \Sigma), \tag{2.4.1}
$$

where $\Sigma = \mathsf{diag}(\sigma_1^2, \sigma_2^2, .., \sigma_N^2, 0_{m\times 1})$ is a $(N+m)\times 1$ diagonal matrix, $\mathbf{I}_m$ is the $m \times m$ identity matrix,

$$
\begin{aligned}
\mathbf{a} &= \begin{pmatrix} \frac{a(\tau_1)}{\tau_1} & \frac{a(\tau_2)}{\tau_2} & \dots & \frac{a(\tau_N)}{\tau_N} & 0_{1\times m} \end{pmatrix}' : (N+m) \times 1, \\
\bar{\mathbf{B}} &= \begin{pmatrix} \frac{b(\tau_1)}{\tau_1} & \frac{b(\tau_2)}{\tau_2} & \dots & \frac{b(\tau_N)}{\tau_N} \end{pmatrix}' : N \times l, \\
\mathbf{B} &= \begin{pmatrix} \bar{\mathbf{B}} & 0_{N\times m} \\ 0_{m\times l} & \mathbf{I}_m \end{pmatrix} : (N+m) \times (l+m).
\end{aligned}
$$

The transition equation of the state space model is simply given by the assumed factor dynamics,

$$
\mathbf{f}_{t+1} | \mathbf{f}_t, G, K, \Omega \sim \mathcal{N}(K + G\mathbf{f}_t, \Omega = V\Gamma V), \tag{2.4.2}
$$

The prior distribution is developed by incorporating theoretically supported knowledge. In particular, following Chib and Ergashev (2009), Abbritti et al. (2016) and Chib et al. (2019), we incorporate in the prior distribution the belief that yield curve should be gently upward sloping on average and persistent over time. We arrive at the prior by sampling prior similar to analysis in Chapter 1. Since the main goal of this Chapter is to predict bond returns we will focus on how to sample posterior of predictive yields and construct bond returns in this section.

2.4.1 Posterior Predictive Likelihood

We evaluate the predictive ability of the proposed model based on out-of-sample density prediction. Predictive accuracy of density forecasts is typically measured by the log posterior predictive likelihood (PPL), which is the sum of log posterior predictive density (PPD) evaluated at the realized yield curves. m and n indicate the starting and end dates of the out-of-sample, respectively. Suppose that $\boldsymbol{y}_{T+1}^{o}$ denotes the realized yield curve at time $T+1$. Then, the log PPL of a particular model $\mathcal{M}$ is calculated as:

$$\boldsymbol{log}PPL = \Sigma_{T=m-1}^{n-1}\boldsymbol{log}p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\mathcal{M}),$$

and the PPD, $p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\mathcal{M})$ is defined as:

$$\begin{aligned} p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\mathcal{M}) &= \int p(\boldsymbol{y}_{T+1}^{o},\psi,\Sigma,\mathbf{f}|\boldsymbol{X},\mathcal{M})d(\psi,\Sigma,\mathbf{f}) \\ &= \int p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\mathcal{M},\theta,\Sigma,\mathbf{f})\boldsymbol{\pi}(\psi,\Sigma,\mathbf{f}|\boldsymbol{X},\mathcal{M})d(\psi,\Sigma,\mathbf{f}), \end{aligned}$$

Since it is not possible to obtain the predictive density analytically, we rely on the numerical integration. The PPD is obtained as the Monte Carlo averaging of $p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\psi,\Sigma,\mathbf{f})$ over the draws of $(\psi,\Sigma,\mathbf{f})$ from $(\psi,\Sigma,\mathbf{f})|\boldsymbol{X}$

$$p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X}) \approx \frac{1}{n_1}\sum_{g=1}^{n_1} p(\boldsymbol{y}_{T+1}^{o}|\boldsymbol{X},\psi^{(g)},\Sigma^{(g)},\mathbf{f}^{(g)}). \tag{2.4.3}$$

$(\psi^{(g)},\Sigma^{(g)},\mathbf{f}^{(g)})$ are the posterior draws from the gth MCMC cycles, $\bar{\boldsymbol{l}}_{T+1}^{(g)}$ is the first l elements of $\bar{\mathbf{f}}_{T+1}^{(g)} = \boldsymbol{G}^{(g)}\boldsymbol{x}_{T}^{(g)}$, $\Omega_{1:k,1:k}^{(g)}$ is the first $k\times k$ elements of $\Omega^{(g)}$, $\Sigma_{1:N,1:N}^{(g)}$ is the first $N\times N$ elements of $\Sigma^{(g)}$, and

$$p(\boldsymbol{y}_{T+1}^{o}|\mathbf{X},\psi^{(g)},\Sigma^{(g)},\boldsymbol{x}^{(g)}) \equiv \mathcal{N}\left(\boldsymbol{y}_{T+1}^{o}|\bar{\boldsymbol{a}}^{(g)}+\bar{\boldsymbol{B}}^{(g)}\bar{\boldsymbol{l}}_{T+1}^{(g)},\bar{\boldsymbol{B}}^{(g)}\Omega_{1:k,1:k}^{(g)}\bar{\boldsymbol{B}}^{(g)\prime}+\Sigma_{1:N,1:N}^{(g)}\right).$$

2.4.2 Predictive Bond Returns

Let $P_T(\tau) = exp(-\tau \times y_T(\tau))$ denotes the $\tau-$ period bond price at time T. Then, the bond return of the $(\tau+1)-$ period bond at time $T+1$, $y_{T+1}(\tau+1)$ is $logP_{T+1}(\tau) - logP_T(\tau+1)$, which is equal to

$$R_{T+1}(\tau+1) = (\tau+1) \times y_T(\tau+1) - \tau \times y_{T+1}(\tau) \tag{2.4.4}$$

Intuitively, by simulating predictive distribution of $y_{T+1}(\tau)$, we draw the posterior samples of the bond returns. Let $\boldsymbol{y_{T+1}}(\tau_{all})$ is the N-dimensional vector of bond yields, and its predictive distribution to be simulated is defined as $p(\boldsymbol{y_{T+1}}(\tau_{all}))$. The PPD can be obtained similar to equation 2.4.3 which is as follows:

$$p(\boldsymbol{y}_{T+1}(\tau_{all})|\boldsymbol{X}) \approx \frac{1}{n_1}\sum_{g=1}^{n_1} p(\boldsymbol{y}_{T+1}(\tau_{all})|\boldsymbol{X}, \psi^{(g)}, \Sigma^{(g)}, \mathbf{f}^{(g)}).$$

$(\psi^{(g)}, \Sigma^{(g)}, \mathbf{f}^{(g)})$ are the posterior draws from the gth MCMC cycles. It means that we can jointly simulate one-period-ahead bond yields using the posterior draws of the MCMC cycle for the factors and parameters. After sampling the predictive bond yields, we obtain the predictive bond return according to Equation (2.4.4).

2.5 Out-of-Sample Portfolio Choice Evaluation

Following Kandel and Stambaugh (1996), we consider the portfolio choice problem with N different maturities as well as the risk-free rate for a representative Bayesian investor. The investor has a CRRA utility function:

$$U(\omega, R_{T+1}) = \frac{[1 + \Sigma_{i=1}^{N} R_{T+1}(\tau_i+1) - c\Sigma_{i=1}^{N}(\omega_i - \omega_0^{-1})^2]^{1-\rho}}{1-\rho}$$

where ρ is the CRRA coefficient, $-1 \leq \omega_i \leq 2$ is the weight on the ith bond. $\Sigma_{i=1}^{N}\omega_i = 1$ and let ω denotes the vector of weights: $\omega = \{(\omega_i)_{i=1}^{N}\}$. We also denote $\omega_0 = \{\frac{1}{N}, \frac{1}{N}...\frac{1}{N}, \}$ to be the vector of equal weights. Intuitively, we take a simple portfolio with equal weights as the benchmark, and incorporate a cost function of the deviations from benchmark $c\Sigma_{i=1}^{N}(\omega_i - \omega_0^{-1})^2$, in the utility function. This cost function is necessary to take into account the utility loss from the tracking errors(See Choi and Kang (2017)). This also enables us to prevent an extreme allocation like $\omega = (1, 1, 1, 1, -1, -1, -1, 0,)$ which can be occurred by a long-short strategy. The cost coefficient, c is the marginal cost of adjusting the portfolios. Thus, the larger c indicates less incentive to deviate from the previous period's portfolio.

Let $R_{T+1} = (R_{T+1}(\tau_i + 1))_{i=1}^{N}$ be the vector of N bond returns realized at time $T + 1$. $\pi(R_{T+1}|\boldsymbol{Y}, \mathcal{M})$ is the model-dependent predictive density of the bond returns, and

$$\mathbb{E}[U(\omega, R_{T+1})|\boldsymbol{Y}, \mathcal{M}] \;\; = \;\; \int U(\omega, R_{T+1})\pi(R_{T+1}|\boldsymbol{Y}, \mathcal{M})dR_{T+1},$$

is the expected utility. We analyze the asset allocation in a single-period setting. The investor maximizes the one-period-ahead expected utility by solving the optimal allocation problem:

$$\omega_{T,M}^{*} = \underset{\omega}{\mathsf{argmax}}\, \mathbb{E}[U(\omega, R_{T+1})|\boldsymbol{Y}, \mathcal{M}] \tag{2.5.1}$$

with respect to the portfolio weights, ω. However, the integral in the expected utility is not analytically calculated. Instead, using the simulated posterior predictive bond return draws, $\{R_{T+1}^{(g)}(\tau)\}_{g=1}^{n1}$ and method of composition, we can numerically approximate the expected utility as the following:

$$\mathbb{E}[U(\omega, R_{T+1})|\boldsymbol{Y}, \mathcal{M}] = \frac{1}{n_1}\Sigma_{g=1}^{n1}\{\frac{[1 + \omega_f R_{f,T} + \Sigma_{i=1}^{N} R_{T+1}^{g}(\tau_i + 1) - c\Sigma_{i=1}^{N}(\omega_i - N^{-1})^2]^{1-\rho}}{1 - \rho}\} \tag{2.5.2}$$

Our bond portfolio optimization consists of two stages. The first stage is the posterior bond return density simulaion. As new yield curve data arrive every month, the Bayesian investor learns about the parameters and factors, and simulates the predictive bond from an econometrician's point of view accounting for uncertainties in parameter and factors. In the second stage, the investor chooses portfolio weights to maximize the one-month-ahead expected utility of Equation 2.5.1. These stages are repeated for each month.

We then economically evaluate and compare the prediction models in terms of the out-of-sample portfolio performance. Because $R_{T+1}|\boldsymbol{Y}, \mathcal{M}$ is model-specific, the optimal bond portfolio weights, resulting realized portfolio return and the utility generated are also model-specific,too. The utility gain from using a prediction model in the bond portfolio choice is measured by CER as in the works of Johannes (2014). CER is the hypothetical risk freee rate whose utility is equal to the expected utility of the uncertain portfolio return. In other words, it is a certain return the risk-averse bond investor would trade for the risky bond portfolio return here. For a particular model $\mathcal{M}$, the empirical CER can be computed as:

$$CER(\mathcal{M}) = [(1-\rho)\frac{\sum_{t=m-1}^{n-1} U^*_{t=1,\mathcal{M}}}{n-m+1}] - 1 \tag{2.5.3}$$

where R^0_{t+1} is the realized individual bond returns, $\omega^*_t = \{\omega^* t, i\}_{i=1}^N$ is the optimal portfolio weights chosen at time t and

$$U^*_{t=1,\mathcal{M}} = U(\omega^*_t, R^0_{t+1})$$

is the realized utility at time $t+1$.

2.6 Empirical Results

2.6.1 Competing Models

As mentioned in the previous we consider two datasets: one dataset contains 8 representative macro factors from Chib et al. (2019) while another dataset contains 8 macro factors generated from a large macro panel data set in Ludvigson and Ng (2009b). We will consider the combination of macro factors within each data set[2]. Since each candidate factor can be either irrelevant or relevant, the total number of model specifications of affine term structure models will be $2^8 = 256$ in each dataset. Because it is often believed to be challenging to outperform the non-predictable (NoPred) model which assumes that treasury yields are not predictable , we include NoPred model in our competing models(See de Pooter (2007)). In the following section, we will evaluate the alternative models in terms of out-of-sample performance starting from January 2008 to December 2017, which amounts to 120 out-of-sample periods. We will begin with statistical evaluation and then do economic evaluation of different models.

2.6.2 Predictive Performance Comparison

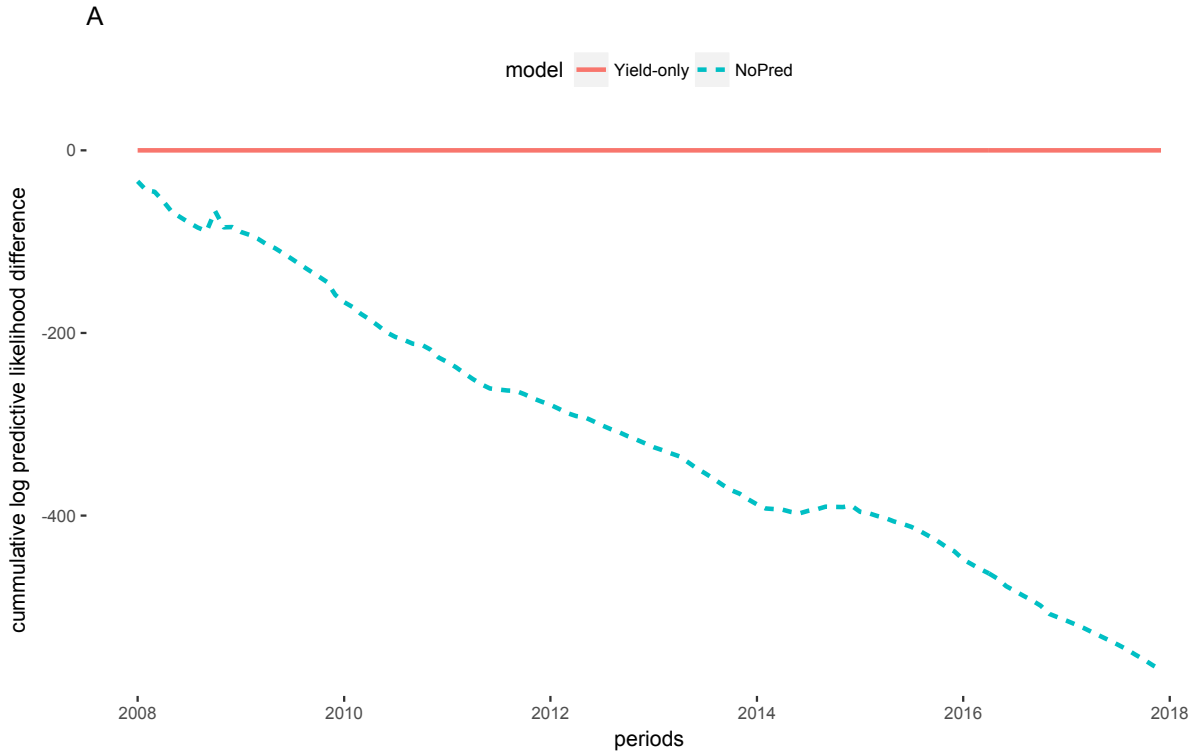
Non-Predictable Model VS Yield-Only Model

In non-predictable model, we assume that each yield is predicted based on historical average independently. Figure 2.4 compares the predictive difference of non-predictable model with respect to yield-only model $\mathcal{M}_0$. Obviously, yield-only affine term structure models can behave significantly better than NoPred model, indicating Gaussian affine term structure models can improve prediction accuracy compared to non-predictable model. Intuitively, the log difference of predictive likelihood comes not only from a more accurate mean pre-

[2]there is no need to mix two datasets since there would be multicollinearity by doing this

Figure 2.4: **Log Posterior Predictive Likelihood Diffrences(NoPred VS Yield-Only)**

This figure shows the sequences of the 120 log PPLs obtained by the rolling windows. The benchmark model is yield-only model $\mathcal{M}_0$ and we plot the log predictive likelihood difference of non-predictable with respect to the benchmark model. Out of sample period is from January 2018 to December 2017.



dictive value but also from an overall more accurate predictive variance-covariance matrix. Non-predictable model assumes independence across yields and thus would have a lower predictive likelihood. However, yield-only model, when making prediction for one specific yield, incorporates a richer information– it not only extract information in that specific yield but also the correlation with other yields.

Yield-Only Model VS Models with unspanned macro factors in Chib et al. (2019)

As mentioned before, we have two macro datasets and we will conduct two separate experiments. In this section we evaluate macro factors in Chib et al. (2019). Figure 2.5 plots the 3 top models (the specifications of the model is listed in the Table 2.4) which outperform

Table 2.4: **Competing Models Summary(Chib et al. (2019))**

m indicates the macro factor is included in the model while **z** indicates the macro factor is excluded from the model. $\mathcal{M}_0$ denotes the model assuming no relevant macro factors. $\mathcal{M}_A$, $\mathcal{M}_B$ and $\mathcal{M}_C$ are the 3 best models outperforming $\mathcal{M}_0$ in terms of out-of-sample cumulative predictive likelihood. Out of sample period is from January 2018 to December 2017.

	$\mathcal{M}_A$	$\mathcal{M}_B$	$\mathcal{M}_C$	$\mathcal{M}_0$
AHE	**m**	**m**	**z**	**z**
CPI	**m**	**z**	**m**	**z**
IP	**z**	**m**	**z**	**z**
PCE	**z**	**z**	**z**	**z**
FFR	**m**	**m**	**m**	**z**
Unem	**z**	**z**	**z**	**z**
TCU	**m**	**z**	**z**	**z**
HSal	**z**	**z**	**z**	**z**

yield-only model $\mathcal{M}_0$ (until the last rolling window). Therefore macro factors from Chib et al. (2019), when chosen carefully into the affine term structure model, can substantially help improve forecasting accuracy relative to yield-only model. In addition, the followings are worth noting:

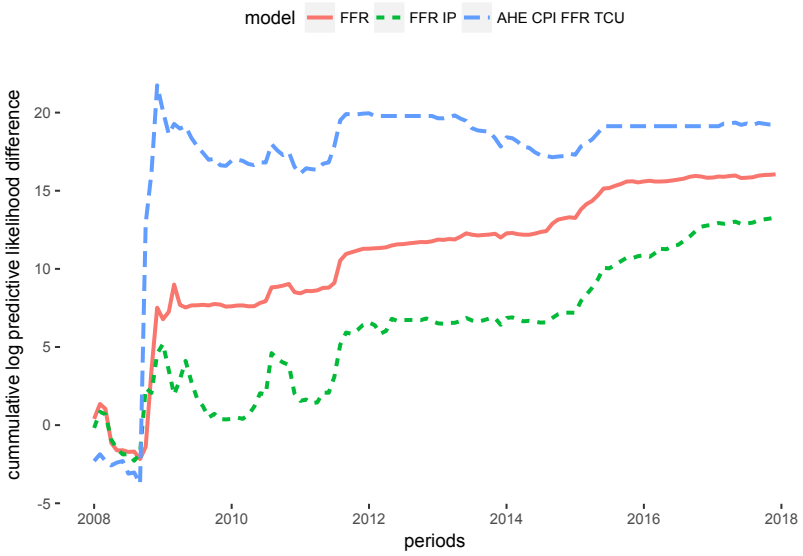
First, the predictive gain over $\mathcal{M}_0$ concentrates during the recession period starting from January,2008 to December 2009. Such predictive gain becomes less significant when economy walks out of recession. This is consistent with the literature showing that either the predictability of equity or bond returns is concentrated in economic recessions. Such finding suggests that bond return predictability can be linked to cyclical variation.

Secondly, it should be noted that additional macro factors might not consistently outperform the yield-only models. Intuitively, unspanned macro factors might provide additional information but they also increase model complexity, lowering the predictive power of the model.

Thirdly, As shown in 2.4, $\mathcal{A}$ which include AHE, CPI,FFR and TCU as relevant macro factors, performs the best. Intuitively, AHE and TCU can be related to real economic activities; CPI can be related to price factors and FFR can be related to macro risks in financial sector.

Figure 2.5: **Log Posterior Predictive Likelihood Diffrences(Chib et al. (2019))**

This figure shows the sequences of the 120 log PPLs obtained by the rolling windows. Benchmark model is $\mathcal{M}_0$ and we plot the cumulative log predictive likelihood difference of $\mathcal{M}_A$, $\mathcal{M}_B$ and $\mathcal{M}_C$ with respect to the benchmark model. Out-of-sample period is from January 2008 to December 2018.



Yield-Only Model VS Models with unspanned macro factors in Ludvigson and Ng (2009b)

Similar to the analysis in the previous section, we now evaluate macro factors from Ludvigson and Ng (2009b). Figure 2.6 plots the 3 top models (the specifications of the model is listed in the Table 2.5) which outperform model $\mathcal{M}_0$ (until the last rolling window). Similar to macro factors in Chib et al. (2019), macro factors selected from Ludvigson and Ng (2009b) can improve predictive accuracy of the yields. Furthermore, similar to the finding in previous section, the predictive gain concentrates largely around 2008 economic recessions.

Finally, the best model include factors F_1, F_2, F_4. Recall that F_1 resembles a real activity factor that loads on employment and output data. F_2 loads on interest rate and spreads, which could be related to financial sector. However, F_4, which mainly loads on prices is also included in the best model. Recall the best model corresponding to dataset in Chib et al. (2019) include AHE, CPI,FFR and TCU. Obviously, the selected macro factors for best

Table 2.5: **Competing Models Summary(Ludvigson and Ng (2009b))**

m indicates the macro factor is included in the model while **z** indicates the macro factor is excluded from the model. $\mathcal{M}_0$ denotes the model assuming no relevant macro factors. $\mathcal{M}_A$, $\mathcal{M}_B$ and $\mathcal{M}_C$ are the 3 best models outperforming $\mathcal{M}_0$ in terms of out-of-sample cumulative predictive likelihood. Out of sample period is from January 2018 to December 2017.

	$\mathcal{M}_A$	$\mathcal{M}_B$	$\mathcal{M}_C$	$\mathcal{M}_0$
F1	**m**	**m**	**m**	**z**
F2	**m**	**m**	**m**	**z**
F3	**z**	**z**	**z**	**z**
F4	**m**	**z**	**z**	**z**
F5	**z**	**m**	**m**	**z**
F6	**z**	**z**	**z**	**z**
F7	**z**	**m**	**m**	**z**
F8	**z**	**z**	**z**	**z**

model corresponding to either Ludvigson and Ng (2009b) or Chib et al. (2019) is similar to each other.

2.6.3 Portfolio Performance

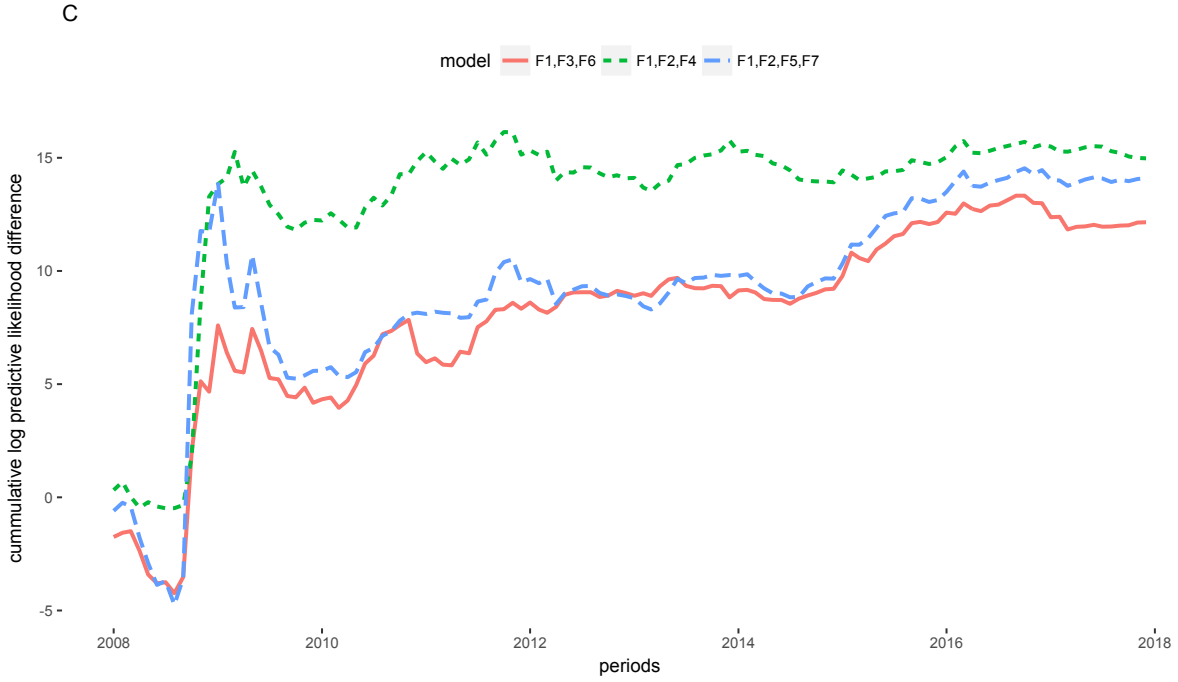
In this section we will consider whether additional macro variables can improve portfolio performances. We will consider macro factors in Chib et al. (2019) and Ludvigson and Ng (2009b) separately just as we did in the previous section. We will use CER as the criteria for comparing the portfolio performance.

More specifically, we will calculate the CERs of the competing models for the two risk aversion coefficient ($\rho = 2$ or 5) and three cost efficients($c = 0.1$ or 0.01) according to Equation 2.5.3 where m is January 2008 and n is December 2017. We are interested in the optimal allocation among the risky assets assuming that the weight on government bond with one month to maturity is exogenously given.

We consider the following portfolio strategies: non-predictability strategy(NoPred model), strategy based on yield-only model($\mathcal{M}_0$), equally weighted portfolio strategy and strategies based on macro-finance models that include macro factors from datasets either from Chib et al. (2019) or Ludvigson and Ng (2009b). In total, we are considering $256 + 2 = 258$ dif-

Figure 2.6: **Log Posterior Predictive Likelihood Diffrences(Ludvigson and Ng (2009b))**

This figure shows the sequences of the 120 log PPLs obtained by the rolling windows. Benchmark model is $\mathcal{M}_0$ and we plot the cumulative log predictive likelihood difference of $\mathcal{M}_A$, $\mathcal{M}_B$ and $\mathcal{M}_C$ with respect to the benchmark model. Out-of-sample period is from January 2008 to December 2018.



ferent kinds of portfolio strategies given the dataset. We include equally-weighted portfolio strategy because it is shown in the literature that outperforming both strategies can be challenging [3] (See DeMiguel, Garlappi, and Uppal (2009)) and we want to evaluate the extent to which utility would increase if the investor forms an optimal portfolio based on macro information. We consider different investors' preference(therefore different ρ) because preferred model can differ depending on the investors' preferences. A more risk averse investor would prefer a model with accurate variance-covariance forecasts whereas a less risk-averse investor who is mainly interested in expected returns would use a model offering accurate point forecast.

Macro factors in Chib et al. (2019)

Table 2.6 reports selected annualized empirical CERs when the relative risk-aversion coefficient is 2 and 5 separately. The number in bold corresponds to the highest economic value achieved among the 258 strategies. Following findings are noteworthy: First, the portfolio strategy based on model $\mathcal{M}_{\mathcal{A}}$, still delivers the best economic performance among all the competing models given any combination of c and ρ considered here. Therefore both measures of method performance, the predictive likelihood and portfolio gains, identify $\mathcal{M}_{\mathcal{A}}$ as the best performing specification.

Second, Portfolio strategies based on $\mathcal{M}_{\mathcal{B}}$ and $\mathcal{M}_{\mathcal{C}}$ do not necessarily outperform yield-only model $\mathcal{M}_0$ indicating that macro variables do not necessarily help improve economic performance in an affine term structure models even they are able to help improve predictive accuracy. Intuitively, recall Panel B in Figure 2.5 $\mathcal{M}_{\mathcal{B}}$ and $\mathcal{M}_{\mathcal{C}}$'s cumulative predictive likelihood relative to $\mathcal{M}_0$ fluctuates over the periods. Such unstable forecasting performance can bring lower utility over certain time and thus bring lower CER (which is the utility averaged over time).

[3] Equally weighted portfolio strategy also indicates 0 transaction costs

Third, table 2.6 shows that $\mathcal{M}_{\mathcal{A}}$ beats other strategies by a greater margin given a lower c (lower transaction cost) and lower ρ (lower risk averse coefficient). For example, if the transaction cost c is lowered to $c = 0.01$, $\mathcal{M}_{\mathcal{A}}$ beats $\mathcal{M}_{\prime}$ by a margin of 0.2 whereas it beats $\mathcal{M}_{\prime}$ only by a margin of 0.1 when the transaction cost $c = 0.1$. Intuitively, a higher cost and risk averse will drive the investors to track closer with the equal weight strategy. Therefore, if the investors are more willing to take risk and care less about transaction cost, they will benefit more from additional macro information in the $\mathcal{M}_{\mathcal{A}}$ strategy.

Fourthly, Figure 2.7 offers a different perspective on the portfolio formed from different

Table 2.6: **Annualized Certainty Equivalence Returns(Chib et al. (2019))**

This table reports the empirical CERs in percentage during out of sample periods. The first column reports the CRRA coefficent ρ used in the calculation, the second column reports the model and $3 - 5$ columns report CER under different transaction cost. Out-of-sample period is from January 2008 to December 2018.

CRRA coefficient	model	$c = 0.1$	$c = 0.01$
$\rho = 2$	Equal Weight	2.74	2.74
	NoPred	2.78	2.99
	$\mathcal{M}_0$	3.01	3.3
	$\mathcal{M}_{\mathcal{A}}$	**3.2**	**3.5**
	$\mathcal{M}_{\mathcal{B}}$	2.55	3.12
	$\mathcal{M}_{\mathcal{C}}$	2.88	2.9
$\rho = 5$	Equal Weight	2.57	2.57
	NoPred	2.59	2.47
	$\mathcal{M}_0$	2.81	3.11
	$\mathcal{M}_{\mathcal{A}}$	**2.9**	**3.21**
	$\mathcal{M}_{\mathcal{B}}$	2.24	2.79
	$\mathcal{M}_{\mathcal{C}}$	2.22	2.75

strategies. It shows the cummulative wealth for 4 portfolios based on equal-weighted strategy(red solid line), $\mathcal{M}_0$(green dashed line),$\mathcal{M}_A$(blue dashed line),NoPred(Non-predictable) strategy(purple dashed line). Both equal-weighted and non-predictable strategy performs much worser than the other two strategies. $\mathcal{M}_A$ generates the highest cumulative wealth for nearly all the time periods. In addition, for most of the time periods, the cumulative wealth curve generated by $\mathcal{M}_A$ has a steeper upward trend than other strategies.

Figure 2.7: **Cumulative wealth of investors forming portfolios under different models(Chib et al. (2019))**

This figure plots cumulative wealth for 4 different strategies during out of sample periods: equal-weighted strategy(purple line), model $\mathcal{M}_0$(green dashed line), model $\mathcal{M}_A$(blue dashed line) and NoPred model(red solid line).Investors have relative risk aversion coefficient $\rho = 2$ and cost $c = 0.01$.Out-of-sample period is from January 2008 to December 2018.

Figures 2.8 and 2.9 present the time series of the optimal portfolio weights over the out-of-sample period of January 2008 to December 2017. It shows the optimal weights for 4 strategies: equally-weighted(red line),$\mathcal{M}_0$(green line),$\mathcal{M}_\mathcal{A}$(blue line) and NoPred(purple line) respectively. Because of the difference in predictive density forecasts, the different optimal allocations across the alternative models arise. First, as shown in these figures. For example, strategy based on NoPred model deviate little from the equal-weighted portfolio strategy whereas both $\mathcal{M}_0$ and $\mathcal{M}_\mathcal{A}$'s optimal allocations are different from the former two strategies. Second, even though both investors' portfolio strategy based on $\mathcal{M}_\mathcal{A}$ and $\mathcal{M}_0$ vary a lot across time, strategy based on $\mathcal{M}_\mathcal{A}$ fluctuate even more (especially during 2008 recession). Intuitively, it is because investors using $\mathcal{M}_\mathcal{A}$ takes into account relevant macro information, bringing additional adjustment for portfolio allocation in each period. Thirdly, a lower risk averse coefficient and transaction cost would encourage investors to take into account additional macro information, therefore bringing a larger variation in portfolio weights across time based on $\mathcal{M}_\mathcal{A}$ than on $\mathcal{M}_0$, (Seen from 2.9 and 2.8)

Macro factors in Ludvigson and Ng (2009b)

Table 2.7 reports selected annualized empirical CERs when the relative risk-aversion coefficient is 2 and 5 separately. The findings is similar to those in Table 2.6. For example, the utility of investors who use $\mathcal{M}_\mathcal{A}$ will increase by a greater margin than other models given a lower transaction cost and lower risk-aversion coefficient The best model based on macro factors from Ludvigson and Ng (2009b) performs better(by a margin of around 0.6) than the best model based on macro factors from Chib et al. (2019) in terms of CER. Figure 2.10 shows the accumulative wealth of different strategies. It can also be seen that $\mathcal{M}_\mathcal{A}$ under Ludvigson and Ng (2009b) performs better than that under Chib et al. (2019). Intuitively, on one hand, macro factors in Ludvigson and Ng (2009b) are generated from a large macro panel dataset and might include a wider range of macro information. On the other hand,

Figure 2.8: **Optimal Bond Portfolio Weights:$\rho = 5$ and $c = 0.1$(Chib et al. (2019))**

This figure plots optimal portfolio weights for 4 cases during out of sample periods for $\rho = 5$ and $c = 0.1$:equal weights(red line),$\mathcal{M}_0$(green line),model $\mathcal{M}_{\mathcal{A}}$(blue line),NoPred(purple line). The out-of-sample periods is from January 2008 to December 2017.

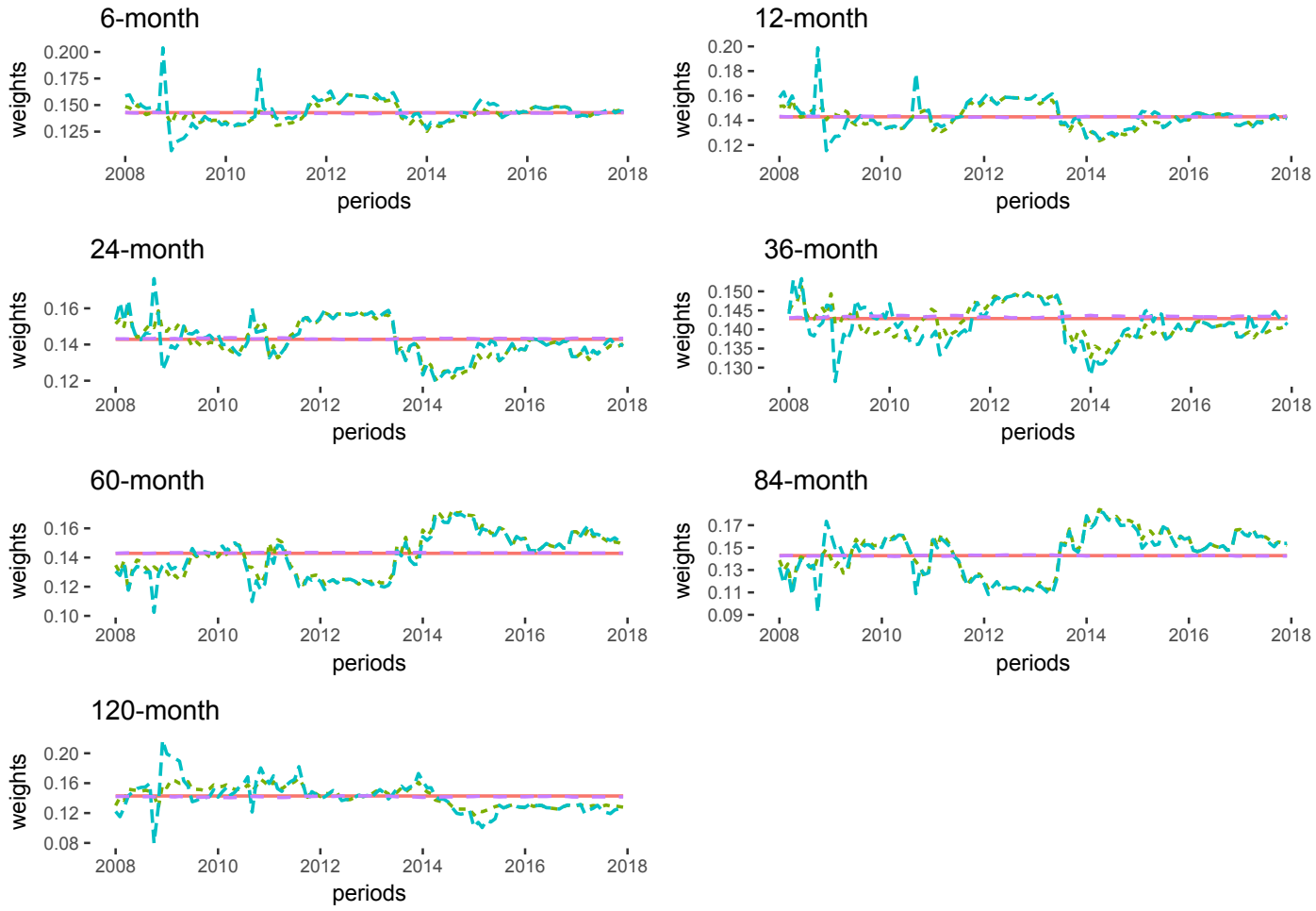
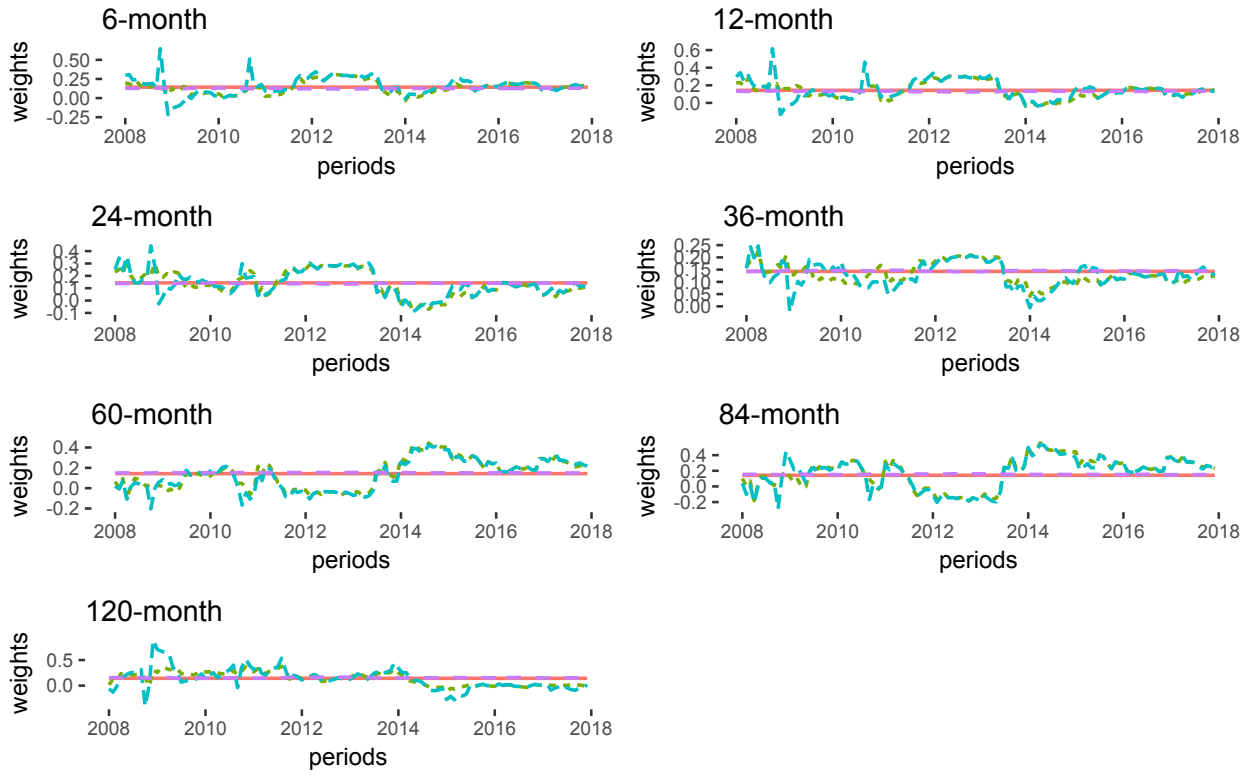


Figure 2.9: **Optimal Bond Portfolio Weights:$\rho = 2$ and $c = 0.01$(Chib et al. (2019))**

This figure plots optimal portfolio weights for 4 cases during out of sample periods for $\rho = 2$ and $c = 0.01$:equal weights(red line),$\mathcal{M}_0$(green line),model $\mathcal{M}_{\mathcal{A}}$(blue line),NoPred(purple line).The out-of-sample periods is from January 2008 to December 2017



it could be possible that some important additional information might be filtered out in the process. In general given the data sample we consider here, principle component macro factors can perform slightly better in terms of CER but performs relatively worse in terms of cumulative returns than representative macro factors. In addition, Figures 2.11 and

Table 2.7: **Annualized Certainty Equivalence Returns(Ludvigson and Ng (2009b))**

This table reports the annualized empirical CERs in percentage over the out of sample period. The first column reports the CRRA coefficent ρ used in the calculation, the second column reports the model $3-5$ columns report the CER under different transaction cost. Out-of-sample period is from January 2008 to December 2018.

CRRA coefficient	model	$c = 0.1$	$c = 0.01$
$\rho = 2$	Equal Weight	2.74	2.74
	NoPred	2.78	2.99
	$\mathcal{M}_0$	3.01	3.4
	$\mathcal{M}_\mathcal{A}$	**3.5**	**4.10**
	$\mathcal{M}_\mathcal{B}$	2.55	2.99
	$\mathcal{M}_\mathcal{C}$	2.08	2.79
$\rho = 5$	Equal Weight	2.57	2.57
	NoPred	2.59	2.47
	$\mathcal{M}_0$	2.81	3.11
	$\mathcal{M}_\mathcal{A}$	**3.35**	**3.81**
	$\mathcal{M}_\mathcal{B}$	2.07	2.42
	$\mathcal{M}_\mathcal{C}$	2.10	2.33

2.12 present the time series of the optimal portfolio weights over the out-of-sample period of January 2008 to December 2017. The pattern of bond portfolio allocation observed 2.9 and 2.8). Strategy based on $\mathcal{M}_\mathcal{A}$ incorporates additional macro information, induce investors to make necessary adjustment in portfolio allocation compared to

2.6.4 How Unspanned macro factors help predict yield curve

As mentioned in previous section, unspanned macro factors supply additional information for predicting yield curve indirectly through affecting latent factors: level, slope and curvature factors. More specifically, the macro factors' effects on latent factors can be captured by certain coefficients on G matrix. To see the magnitude of the effect, we extract the relevant

Figure 2.10: **Cumulative wealth of investors forming portfolios under different models(Ludvigson and Ng (2009b))**

This figure plots optimal portfolio weights for 4 cases during out of sample periods for $\rho = 5$ and $c = 0.1$:equal weights(red line),$\mathcal{M}_0$(green line),model $\mathcal{M}_{\mathcal{A}}$(blue line),NoPred(purple line). The out-of-sample periods is from January 2008 to December 2017.

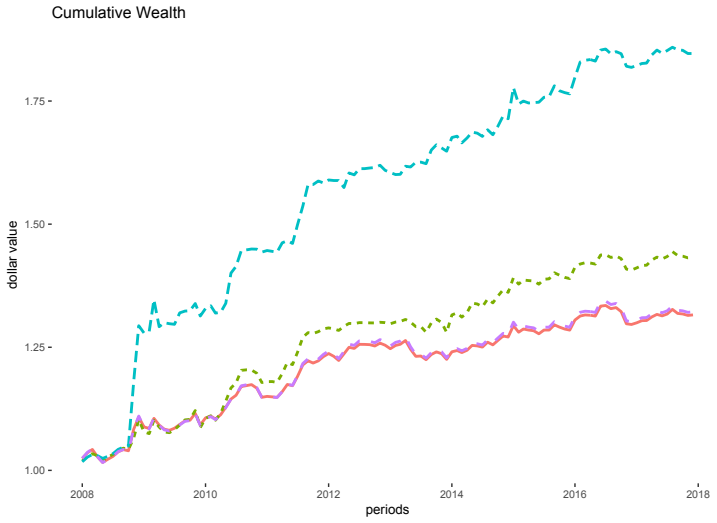


Figure 2.11: **Optimal Bond Portfolio Weights:$\rho = 5$ and $c = 0.1$(Ludvigson and Ng (2009b))**

This figure plots optimal portfolio weights for 4 cases during out of sample periods for $\rho = 5$ and $c = 0.1$:equal weights(red line),$\mathcal{M}_0$(green line),model $\mathcal{M}_{\mathcal{A}}$(blue line),NoPred(purple line). The out-of-sample periods is from January 2008 to December 2017

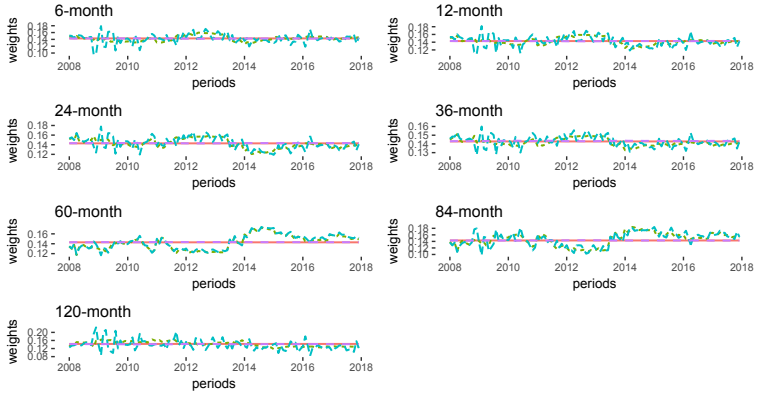


Figure 2.12: **Optimal Bond Portfolio Weights:$\rho = 2$ and $c = 0.01$(Ludvigson and Ng (2009b))**

This figure plots optimal portfolio weights for 4 cases during out of sample periods for $\rho = 2$ and $c = 0.01$:equal weights(red line),model 1(green line),model 12(blue line),NoPred(purple line). The out-of-sample periods is from January 2008 to December 2017

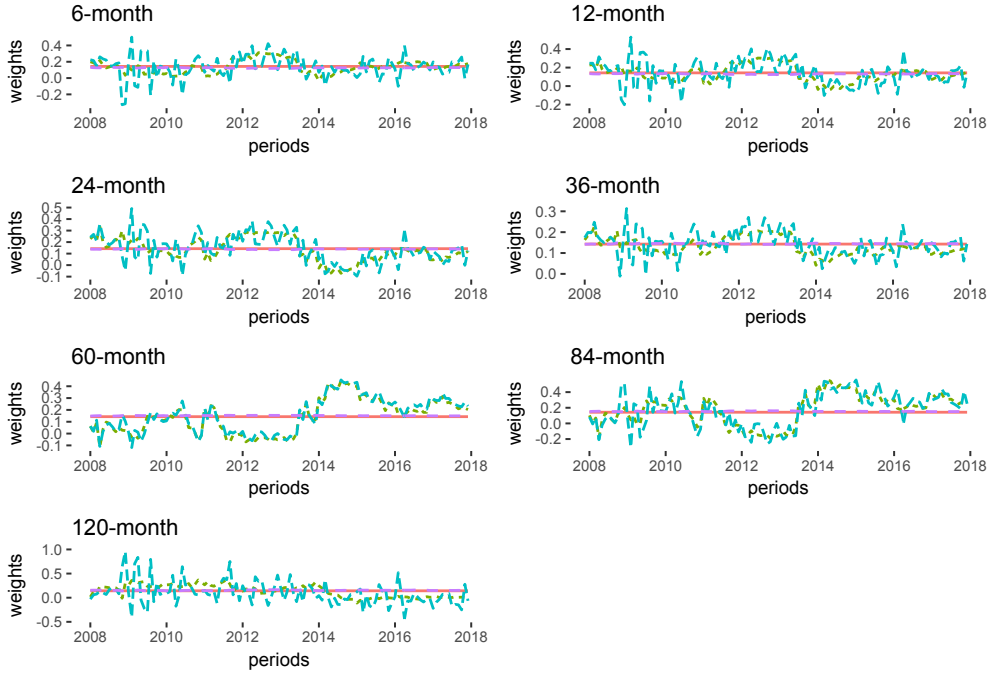
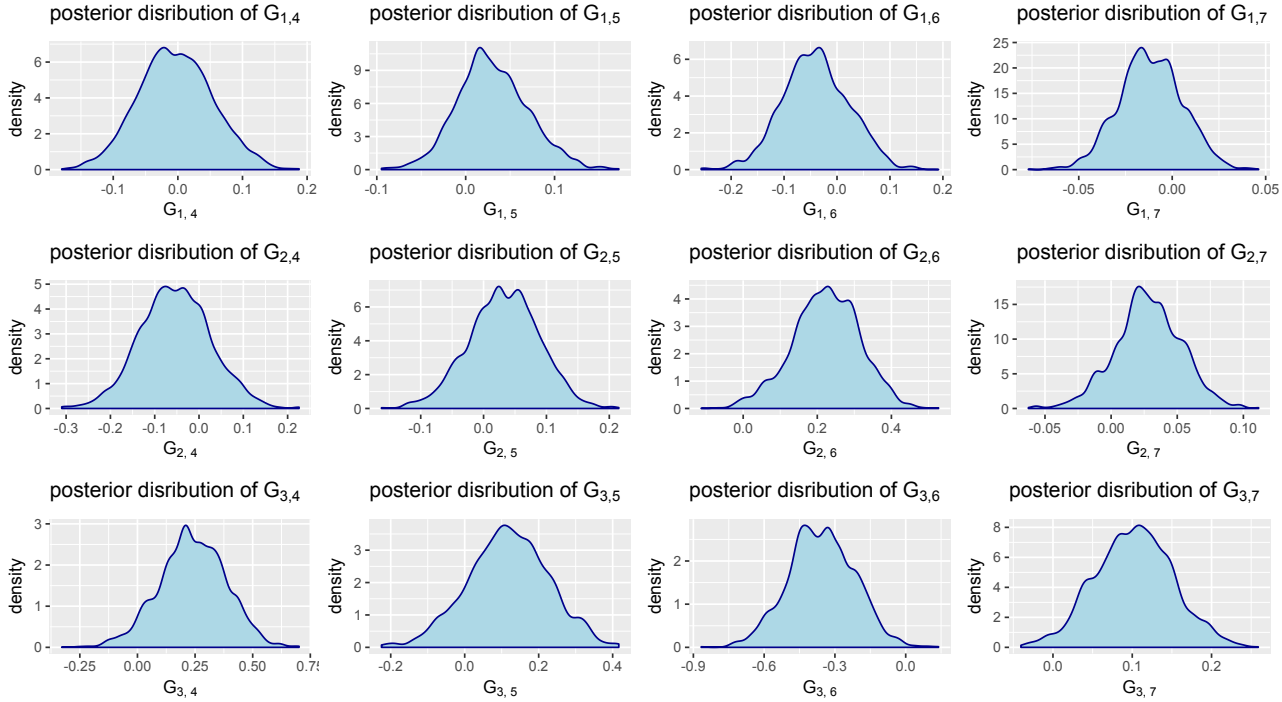


Figure 2.13: **Posterior Density plots for selected parameters for $\mathcal{M}_{\mathcal{A}}$ (Chib et al. (2019))**

This figure plots the corresponding coefficients (in G matrix) capturing unspanned macro factors' effects on latent factors.



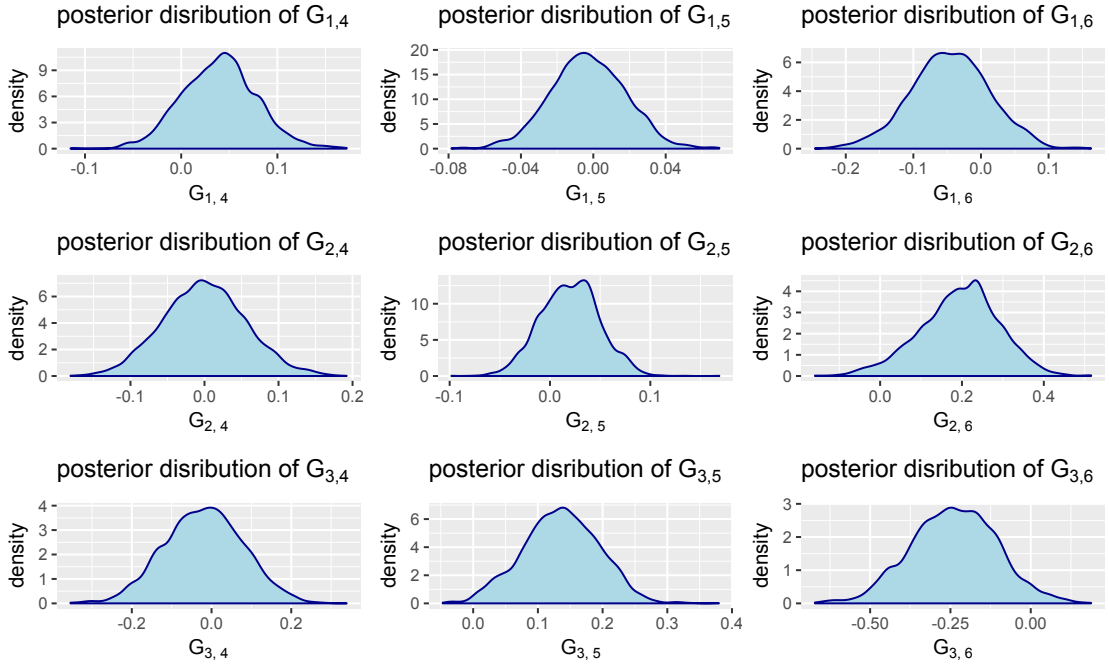
coefficients from the G matrix and plot the posterior density for these parameters. For our purpose, we will use two best models corresponding to macro datasets in Ludvigson and Ng (2009b) and Chib et al. (2019).

Figures 2.13 and 2.14 give the posterior densities of the corresponding coefficients that capture unspanned macro factors' effects on latent factors. More specifically, the 1 column of each Figure (namely $G_{1,4}, G_{1,5}$,$G_{1,6}$ and $G_{1,7}$) captures the macro factor's effect on level factor;the 2 column of each Figure (namely $G_{2,4}, G_{2,5}$,$G_{2,6}$ and $G_{2,7}$) captures the macro factor's effect on slope factor; the 3 column of each Figure (namely $G_{3,4}, G_{3,5}$,$G_{3,6}$ and $G_{3,7}$) captures the macro factor's effect on curvature factor. For example, $G_{3,7}$ indicates that 1 size of change in the macro factor will bring 0.1 size of change in the curvature factor.

It can be observed that the coefficient corresponding to the macro factor's effect on slope

Figure 2.14: **Posterior Density plots for selected parameters for $\mathcal{M}_{\mathcal{A}}$ (Ludvigson and Ng (2009b))**

This figure plots the corresponding coefficients (in G matrix) capturing unspanned macro factors' effects on latent factors.



and curvature factor is generally greater than that on level factor[4]. Thus it can be inferred that unspanned macro factors influence the predictive yields largely through affecting slope and curvature factor rather than level factor. Such a finding is also consistent Ang and Piazzesi (2001) and Wu (2001) that macro variables such as monetary policy might induce a bigger change on slope rather than level of the yield curve. Intuitively in terms of predicting yields, macro variables might be more useful in supplying additional information to predict changes in slope and curvature of yield curve than in level of the yield curve.

[4]For example, the absolute value of $G_{3,7}$(corresponding to curvature factor) is around 0.12 while the absolute value of $G_{1,7}$(corresponding to level factor) is only around 0.01

2.7 Conclusion

In this paper, we introduce a detailed Bayesian framework for assessing the economic value of macro-variables in a macro-finance affine term structure framework. Through an out-of-sample experiment with macro dataset in Chib et al. (2019) and Ludvigson and Ng (2009b), we investigate the bond return forecasting performance and bond portfolio strategies among competing models which include different macro-variables. Regardless of the dataset we use(either Chib et al. (2019) and Ludvigson and Ng (2009b)), the results show that the predictive ability and bond portfolio of the macro-finance affine term structure models can be substantially improved by incorporating relevant macro factors in real economic activity, financial sector and price sector.[5] In contrast , either ignoring relevant factor or including irrelevant ones can lead to inaccurate forecasts and utility loss. In addition, given the data sample we consider in Chapter 2, principle component factors perform slightly better than representative factors in terms of certainty equivalence return (CER). Furthermore, we find that macro factors supply additional information to predictive yields mainly through affecting slope and curvature factors.

In the future, it is possible to incorporate other model specifications such as time-varying factor loadings or stochastic volatility.

[5]More specifically, the best model corresponding to dataset in Chib et al. (2019) includes FFR,AHE,CPI and TCU while the best model corresponding to dataset in Ludvigson and Ng (2009b) includes F_1, F_2, F_4.

Chapter 3

Comparsion of No-U-Turn Sampler and TARB-MH for posterior sampling in Guassian Affine Term Structure Models

3.1 Introduction

No-U-Turn sampler is a widely used automatically tuned Hanmiltonian Monte Carlo (HMC) algorithm. In recent decades, No-U-Turn sampler(NUTS), or generally HMC, proves to better perform the random walks Metro Hasting Sampling algorithm when the parameters are highly correlated.(SeeNeal (2011) and Hoffman and Gelman (2014))it transforms the problem of sampling from a target distribution into a problem of simulating Hamiltonian dynamics. Tailored randomized-blocking Metropolis-Hastings (TaRB-MH) is another sampling algorithm, adopted by Chib and Ramamurthy (2010) and Chib et al. (2019) in order to conduct an efficient posterior sampling of highly non-linear models such as dynamic stochastic general equilibrium (DSGE) and Gaussian arbitrage-free affine term structure models(GATSMs). However, whether No-U-Turn Sampler will be still effective dealing with

complicated models such as GATSMs when compared to alternative algorithm TaRB-MH remains unexplored.

The goal of this paper is to run a horse race between two sampling algorithms: TaRB-MH and NUTS in a set of GATSMs. While there are several different versions of HMC algorithm, we will focus on NUTS used in Hoffman and Gelman (2014) because NUTS gives a solid scheme for automatically tuning the step size parameter ϵ and number of steps parameters l, avoiding a dramatic drop in HMC's efficiency because of a bad choice of these two parameters. Furthermore, NUTS have been fully incorporated in the open-source Bayesian inference package, Stan (Stan Development Team, 2013) and has been widely accepted and used since then.

The specification of GATSMs we consider here are developed in Duffee (2011), Joslin et al. (2014) and Chib et al. (2019). In order to give a through comparison between two algorithms, we will consider both yield-only GATSMs and macro-finance GATSMs(incorporating macro factors in the affine term stucture models). In order to make two algorithms comparable, we impose informative prior adopted in Chib et al. (2019) to address the irregular likelihood surface of high-dimensional and severely nonlinear with respect to the model parameters in GATSMs.

According to our empirical experiments with the U.S. yield curve data from January 1990 to December 2017, the performance of TaRB-MH is far better than the NUTS in two respects. First of all, the Markov chain based on the TaRB-MH method converges so quickly to the target density compared to the NUTS. We find that the NUTS method is highly likely to fail to converge in an GATSM because of the irregular posterior surface. Second, the NUTS tends to generate much less efficient posterior draws for the parameters and key quantities such as the term premiums and predictive yield curves. Thirdly, NUTS' inefficiency in posterior sampling of GATSM becomes even more severe if adding more macro factors in the model. Finally, we find NUTS' inefficiency in posterior sampling of GATSMs

is robust given different initial values.

The remainder of the paper is organized as follows. In Section 3.2 we present the framework for affine term structure models and in Section 2.4 we describe details of prior-posterior analysis. Section 3.4 provides the empirical comparison results from between HMC and TaRB. Concluding remarks are presented in Section 2.7 and the Appendix contains supplementary details omitted from the main text.

3.2 Framework

To compare the performance of HMC and TaRB in terms of affine term structure models, we will focus on GATSMs developed in Duffee (2011) and Joslin et al. (2014). For Details see Section 2 in Chapter 2.

In order to estimate the affine term structure model, we first represent the yield curve model in the state space form similar to Chapter 2:

$$R_t(\tau_i) = \frac{a(\tau_i)}{\tau_i} + \frac{b(\tau_i)'}{\tau_i}\boldsymbol{l}_t + e_t(\tau_i),\ e_t(\tau_i) \sim \mathcal{N}(0, \sigma_i^2) \text{ for } i = 1, 2, .., N,$$
$$\mathbf{m}_t = \mathbf{m}_t$$

where $e_t(\tau_i)$ is the maturity-specific independent pricing error. We let $\{\tau_i\}_{i=1}^N$ denote N maturities of interest and let

$$\mathbf{R}_t = \begin{pmatrix} R_t(\tau_1) & R_t(\tau_2) & \cdots & R_t(\tau_N) \end{pmatrix}', \tag{3.2.1}$$

denote the corresponding bond yields at time t. Then, given the vector of the observations at time t denoted by $\mathbf{y}_t = (\mathbf{R}_t', \mathbf{m}_t')'$, the econometric model can be rewritten in vector-matrix

form,

$$\mathbf{y}_t|\mathbf{f}_t, \mathbf{a}, \mathbf{B}, \Sigma \sim \mathcal{N}(\mathbf{a} + \mathbf{B} \times \mathbf{f}_t, \Sigma), \tag{3.2.2}$$

The parameters need to be estimated would be in the factor dynamics, SDF, and pricing error variances $\{K, G, V, \Gamma, \delta, \lambda_l, \lambda_{ll}, \lambda_{lm}$and $\sigma_1^2, \sigma_2^2, .., \sigma_N^2\}$

Similar to Chapter 1 and Chapter 2, we will sample $\{K, G, \sigma_1^2, \sigma_2^2, .., \sigma_N^2\}$using Gibbs sampling. We will sample $\psi = \{V, \Gamma, \delta, \lambda_l, \lambda_{ll}, \lambda_{lm}\}$ using either TaRB-MH or NUTS and compare the performance of these two algorithms in the following section.

3.3 Posterior Sampling using NUTS

The goal of this paper is to compare the sampling performance of NUTS with that of TaRB-MH. Since the posterior sampling of affine term structure using TaRB-MH (see for a detailed description) is already extensively discussed previous chapters, in this section we will focus on how to use NUTS for posterior sampling and compare it to TaRB-MH algorithm.

3.3.1 A brief description of HMC and NUTS

As discussed in Hoffman and Gelman (2014), NUTS is just a modified version of HMC. The usual HMC introduces an auxiliary vector momentum variable r for model vector variable ψ. r is drawn from standard normal distribution yielding the joint density:

$$p(\psi, r) \propto exp(L(\psi) - \frac{1}{2} r \cdot r)$$

where L denotes the log likelihood (or log posterior) of the joint density of variables of interest ψ and $x \cdot y$ denotes the inner product of vectors x and y. In physical terms as a fictitious Hamiltonian system, ψ can be interpreted as particle's position and r as the momentum of the particle. L can be interpreted as a position-dependent negative potential energy function

and $r \cdot r$ as the kinetic energy of the particle. One can simulate the evolution over time of the Hamiltonian dynamics of this system via leapfrog, which proceeds according to the updates:

$$r^{t+\epsilon/2} = r^t + \epsilon/2 \bigtriangledown_\psi L(\psi^t)$$
$$\psi^{t+\epsilon} = \psi^t + \epsilon r^{t+\epsilon/2}$$
$$r^{t+\epsilon} = r^{t+\epsilon/2} + \epsilon/2 \bigtriangledown_\psi L(\psi^{t+\epsilon})$$

where r^t and ψ^t denote the values of momentum and position variable r and ψ at time t and $\bigtriangledown_\psi$ denotes the gradient with respect to ψ. The standard procedure for Hamiltonian Monte Carlo is described in algorithm 1.

Algorithm 1 HMC algorithm to sample ψ

Precondition: ψ^0, ϵ, l, M,L

for $m \leftarrow 1$ to M **do**
Sample r^0 from $N(0, I)$
    $\psi^m \leftarrow \psi^{m-1}$
    $r^\dagger \leftarrow r^0$
    $\psi^\dagger \leftarrow \psi^{m-1}$
    for $i \leftarrow 1$ to l **do**
        $r^\dagger$,$\psi^\dagger \leftarrow$ Leapfrog$(\psi^\dagger, r^\dagger, \epsilon)$
    end for
With probability $\alpha = min\{1, \frac{exp(L(\psi^\dagger) - \frac{1}{2} r^\dagger \cdot r^\dagger)}{exp(L(\psi^{m-1}) - \frac{1}{2} r^0 \cdot r^0)}\}$:
    $\psi^\dagger \leftarrow \psi^m$
    $-r^\dagger \leftarrow r^m$
end for
function LEAPFROG$(\psi^\dagger, r^\dagger, \epsilon)$
    $r^\dagger \leftarrow r + \epsilon/2 \bigtriangledown_\psi L(\psi)$
    $\psi^\dagger \leftarrow \psi + \epsilon \psi^\dagger$
    $\psi^\dagger \leftarrow \psi + \epsilon/2 \bigtriangledown_\psi L(\psi^\dagger)$
return $\psi^\dagger, r^\dagger$
end function

However, as shown in Algorithm 1, a standard HMC requires that the user specify at least two parameters: step size ϵ and a number of steps l to run the simulated Hamiltonian

system. No-U-Turn Sampler(NUTS) we use in this paper (Details can be found in Hoffman and Gelman (2014)) will automatically tune these parameters as shown in Algorithm 2. The R codes for NUTS are available on https://github.com/kasparmartens/NUTS.

3.3.2 Difference between HMC and TaRB-MH

To compare HMC with TaRB-MH, we also present TaRB-MH in algorithm 2. While two algorithms can be related to the original MH-algorithm, the key difference between two algorithms lies in the generating proposal $\psi^{\dagger}$ for each iteration of the Markov chain. More specifically, HMC incorporates first-order gradient information in the proposal step and jointly draw ψ and r mimicking a Hamiltonian dynamics. Once the proposal is drawn, it is accepted or rejected in the usual MH step. The advantage of HMC, as mentioned in Robert (2001) lies in the fact that Hanmiltonian can change very little during the Metropolis step while possibly generating a very different value of ψ.

In contrast, TaRB belongs to a class of tailored multiple block MH algorithm discussed in Chib and Greenberg (1995).TaRB enhance the MH approach by randomizing formation of blocks in every iteration and generating tailored proposal density by optimization method such as simulated annealing. As discussed in Chib and Ergashev (2009) and Chib and Ramamurthy (2010), such enhancement would be particularly helpful in dealing with irregular distribution, and also for localizing the proposal density far away from the current region to permit large moves.

3.4 Comparison Results

The objective of this paper is to compare the between sampling efficiency between TaRB-MH and HMC. In many affine term structure model literatures, the most of interests are on the model parameters, factors, and model-implied key quantities such as term premium and

Algorithm 2 No-U-Turn-Sampler ψ

Precondition: ψ^0, ϵ,M,L

for $m \leftarrow 1$ to M **do**
Resample r^0 from $N(0, I)$
Resample u^0 from $Uniform([0, exp(L(\psi^\dagger) - \frac{1}{2} r^\dagger \cdot r^\dagger)])$
Initialize $\psi^- = \psi^{m-1}, r^{-1} = r^0, r^+ = r^0, j = 0, \psi^m = \psi^{m-1}, n = 1, s = 1$
    while s=1 **do**
Choose a direction ν_j from Uniform({-1,1}).
        if $nu_j = -1$ **then**
            ψ^-,$r^-, -, -, \psi^{'}, n^{'}, s^{'} \leftarrow$ BuildTree$(\psi^-, r^-, u, \nu_j, j, r^-, \epsilon)$
        else
            ψ^+,$r^+, -, -, \psi^{'}, n^{'}, s^{'} \leftarrow$ BuildTree$(\psi^+, r^+, u, \nu_j, j, r^+, \epsilon)$
        end if
        if $s^{'} = 1$ **then**
With probability $min\{1, \frac{n^{'}}{n}\}$
            $\psi^{'} \leftarrow \psi^{m-1}$
        end if
        $n \leftarrow n + n^{'}$
        $s \leftarrow s^{'} 1[\theta^+ - \theta^- >= 0] \times 1[\theta^+ - \theta^+ >= 0]$
        $j \leftarrow j + 1$
    end while
end for
function BUILDTREE(ψ,r, u, v, j, ϵ)
    if j=0 **then**
Base case - take one leapfrog step in the direction ν
        $r^{'}$,$\psi^{'} \leftarrow$ Leapfrog$(\psi, r^\dagger, \nu\epsilon)$
        $n^{'} \leftarrow 1[u \leq exp(\psi^\dagger - \frac{1}{2} r^\dagger \cdot r^\dagger)]$
        $s^{'} \leftarrow 1[L(\psi^{'}) - \frac{1}{2} r^\dagger \cdot r^\dagger \geq logu - \Delta_{max}]$
    elseRecursion-implicitly build the left and right subtrees.

        ψ^-,$r^-, -, -, \psi^+, r^+, \psi^{'}, n^{'}, s^{'} \leftarrow$ BuildTree$(\psi, r, u, v, j - 1, \epsilon)$
        if $s^{'} = 1$ **then**
            if $\nu = -1$ **then**
                ψ^-,$r^-, -, -, \psi^+, \psi^{''}, n^{''}, s^{''} \leftarrow$ BuildTree$(\psi^-, r^-, u, v, j - 1, \epsilon)$
            else
                ψ^+,$r^+, -, -, \psi^{''}, n^{''}, s^{''} \leftarrow$ BuildTree$(\psi^+, r^+, u, v, j - 1, \epsilon)$
            end if
With probability $\frac{n^{''}}{n^{'} + n^{''}}$
            $\psi^{'} \leftarrow \psi^{''}$
        end ifreturn $\psi^\dagger, r^\dagger$
    end if
end function

Algorithm 3 TaRB algorithm to sample ψ

Precondition: ψ^{0}, M

for $m \leftarrow 1$ to M **do**

Randomly determine the number of blocks and their components in ψ, as $\psi_1,\ \psi_2,\ ..,\ \psi_{B_g}$

for $l \leftarrow 1$ to B_g **do**

Maximize $L(\psi_l, \psi_{-l})$ w.r.t ψ_l to obtain the mode, $\bar{\psi}_l$ and compute the inverted negative Hessians computed at the mode, $V_{\bar{\psi}_l}$.

Draw a proposal for ψ_l, denoted by $\psi_l^{\dagger}$, from the selected multivariate Student-t distribution $\psi_l^{\dagger} \sim St\left(\bar{\psi}_l, V_{\bar{\phi}_l}, 15\right)$

With probability $\alpha = \min\{1, \frac{L\left(\psi_1^{\dagger}, \psi_{-1}|\mathbf{Y}\right)}{L\left(\psi_1^{(g-1)}, \psi_{-1}|\mathbf{Y}\right)}\}$:

$\phi^{\dagger} \leftarrow \phi^{m}$

end for

end for

Table 3.1: **Candidate Models**

The first column reports the model; The second column reports the macro factors incorporated in the model. The third column reports the total size of the parameters and the last column reports the size of NUTS or TaRB.

Model	macro factors	size(Total)	size(NUTS or TaRB)
$\mathcal{M}_0$	None	28	10
$\mathcal{M}_1$	FFR, CPI	53	20
$\mathcal{M}_2$	FFR, CPI, IP, PCE	90	33

predictive yield curves. Hence, in this section we estimate the both yield-only affine term structure models as well as affine term structure models incorporating macro factors.

We will use the yield and macro dataset mentioned in Chib, Kang, and Xie (2019).the data runs from January 1990 to December 2017. To get a thorough comparison between two different algorithms, we will consider 3 models in the our experiments. The details for these models can be found in Table 3.1. More specifically, $\mathcal{M}_0$ includes no macro factors and therefore is a yield-only model; $\mathcal{M}_1$ includes 2 macro factors: Federal Fund Rate (FFR) and Consumer Price Index (CPI). $\mathcal{M}_2$ includes 4 macro factors: Federal Fund Rate (FFR),Consumer Price Index (CPI),Personal Consumption Expenditure(PCE) and industrial production (IP). As shown in Table 3.1, an increase in number of macro factors in the

affine term structure models will increase the total number of parameters in the model (size of the model) dramatically. Consequently, the size of parameters corresponding to NUTS or TaRB increases as more macro factors are incorporated. Our biggest model $\mathcal{M}_2$ includes up to 90 parameters and 33 out of which will be sampled using NUTS or TaRB.

We compare the efficiency in terms of the inefficiency factor, defined as $1 + 2\sum_{j=1}^{\infty} \rho(j)$ where $\rho(j)$ is the autocorrelation function at lag j of each of simulated parameters over the Markov chain. One can estimate the inefficiency factor as

$$1 + \frac{400}{199} \sum_{j=1}^{200} K(j/200)\hat{\rho}(j). \tag{3.4.1}$$

Then, $\hat{\rho}(j)$ is the jth-order sample autocorrelation of the MCMC draws with $K(\cdot)$ representing the Parzen kernel (Kim, Shephard, and Chib (1998)). By definition, the smaller inefficiency factor indicates the higher efficiency.

3.4.1 log posterior convergence

To compare the performances of TaRB-MH to NUTS, we compute the log posterior densities over the MCMC draws from TaRB-MH and NUTS after the 1000 burn-in draws. Figure 3.1 plot the results for some selected MCMC iterations for $\mathcal{M}_0$. Obviously, both the Markov chain based on TaRB and the NUTS converges. However, the target posterior NUTS converges to has a much lower posterior likelihood. Figure 3.2 Figure 3.3 plot the results for some selected MCMC iterations for $\mathcal{M}_1$ and $\mathcal{M}_2$. The Markov chain based on TaRB converges while the Markov chain based NUTS does not seem to converge even after 5000 draws. Note that $\mathcal{M}_1$ and $\mathcal{M}_2$ incorporate macro factors into the model, adding to the complexity and size of the model and making the posterior sampling based on NUTS even more difficult.

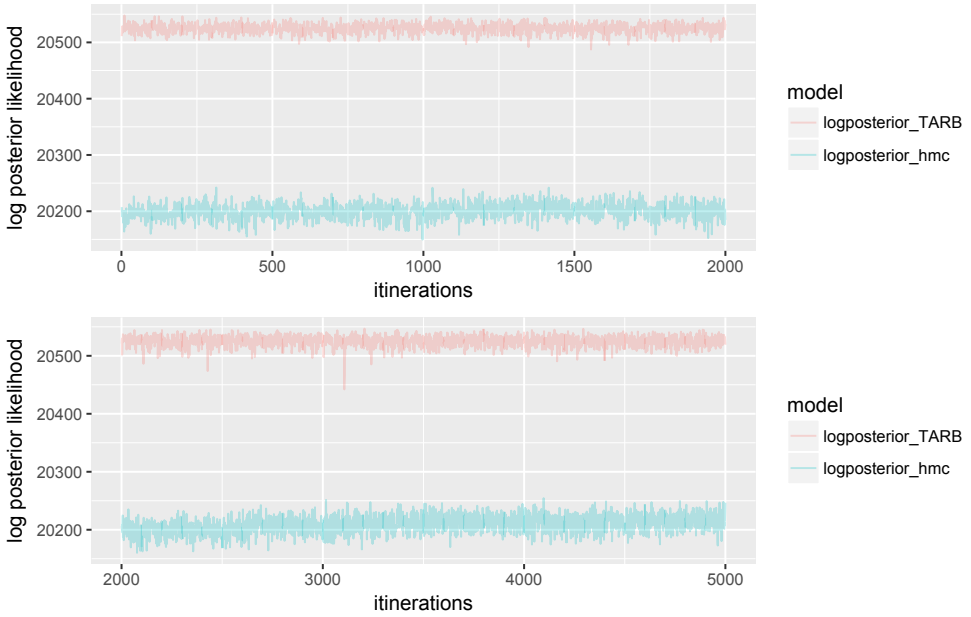


Figure 3.1: **Log posterior densities over MCMC iterations**

This figures plot the log posterior densities (=log likelihood density + log prior density) over selected MCMC iterations for $\mathcal{M}_0$ after the 1000 burn-in draws.

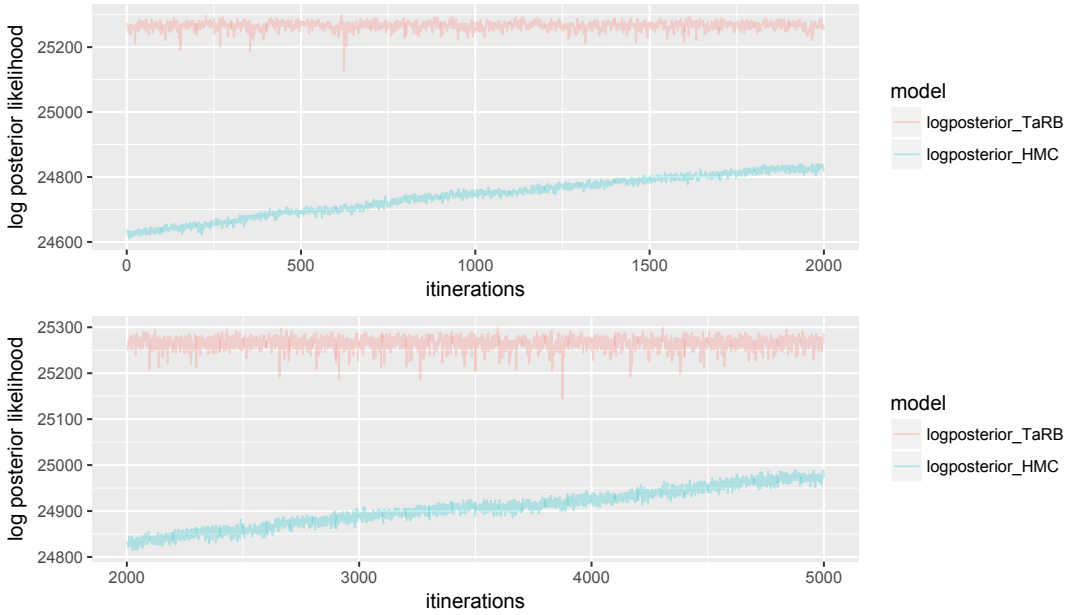


Figure 3.2: **Log posterior densities over MCMC iterations**

This figures plot the log posterior densities (=log likelihood density + log prior density) over selected MCMC iterations for $\mathcal{M}_1$ after the 1000 burn-in draws..

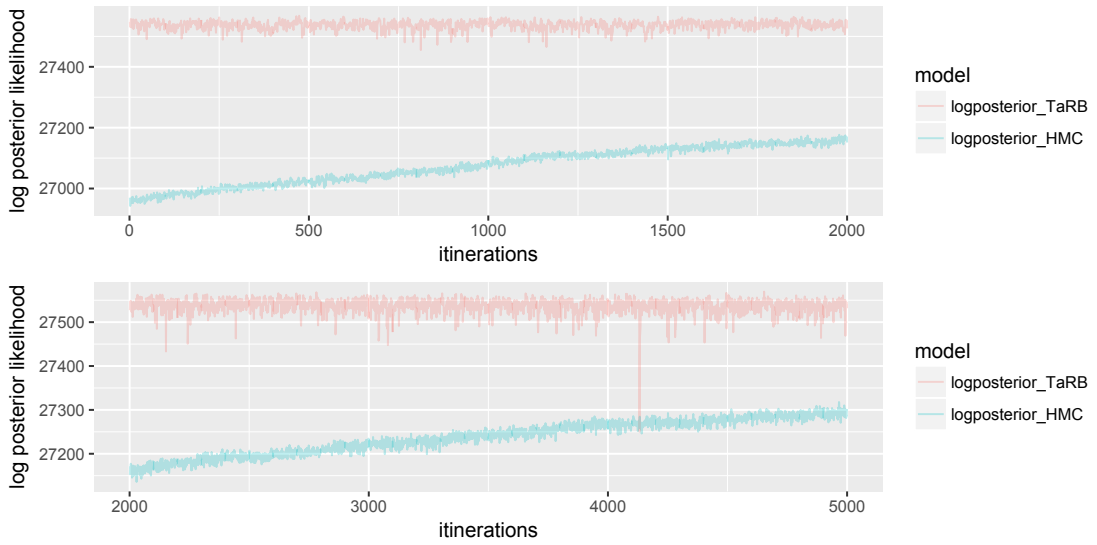


Figure 3.3: **Log posterior densities over MCMC iterations**

This figures plot the log posterior densities (=log likelihood density + log prior density) over selected MCMC iterations for $\mathcal{M}_2$ after the 1000 burn-in draws.

3.4.2 Parameters

Tables 3.2, 3.3 and 3.4 present the summary of posterior distribution of the model parameters and their inefficiency factors for $\mathcal{M}_0, \mathcal{M}_1$ and $\mathcal{M}_2$. The burn-in size is 1,000, and the MCMC size beyond the burn-in is set to be 5,000. The last two columns of the tables are the inefficiency factors from the TaRB-MH and NUTS.

It can be seen from Table 3.2 that the TaRB-MH method simulates the parameters of $\mathcal{M}_0$ more efficiently than the NUTS. More specifically,parameters λ_l,V and κ are sampled using TaRB-MH or NUTS–The inefficiency from TaRB is $5-10$ on average while the inefficiency from NUTS is140 on average. Therefore TaRB-MH is much more efficient when sampling the posterior of affine term structure models.

Table 3.3 and 3.4 describes the inefficiency comparison of NUTS and TaRB when macro factors are included. TaRB-MH maintains lower inefficiency while NUTS's sampling inefficiency becomes even higher. In addition, the sampling inefficiency of some parameters in G and Σ, which are sampled through Gibbs algorithm, is also greatly inflated by the inefficiency

in NUTS part.

Figure 3.4 illustrates the posterior distribution for selected parameters in $\mathcal{M}_2$. The parameters are sampled either through NUTS or TaRB. It should be noted that risk parameters such as λ_l can be difficult to estimate. Obviously, the posterior distributions of parameters sampled through NUTS looks irregular and much tighter than those sampled through TaRB. It could be because the draws from NUTS might be stuck in some local regions and fails to visit other regions of the distribution, bringing a much higher sampling inefficiency compared to TaRB.

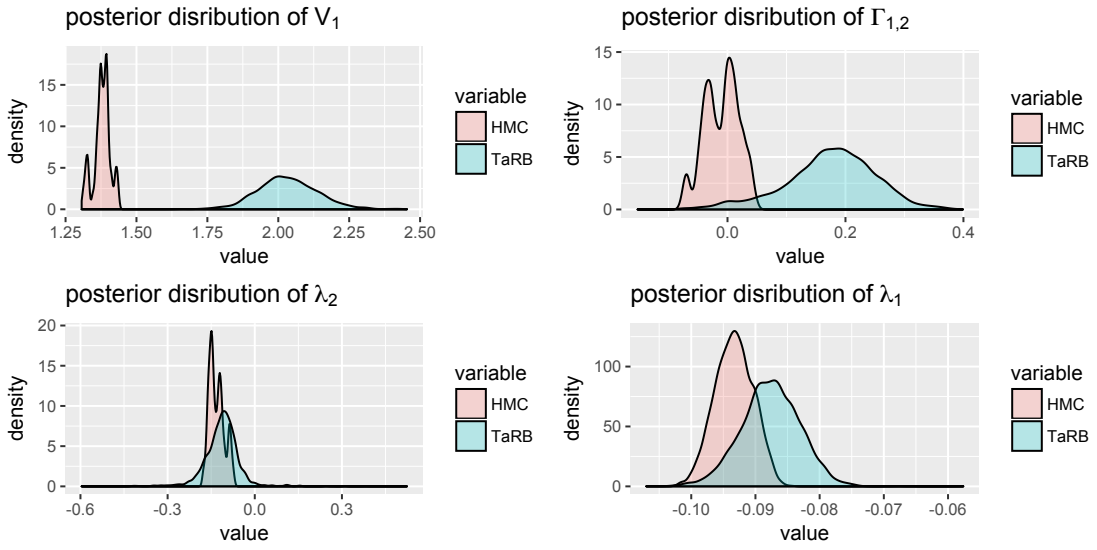


Figure 3.4: **Posterior density plots for selected parameters for $\mathcal{M}_2$**

This figure plots the posterior density of selected parameters in $\mathcal{M}_2$ sampled using either HMC(red) or TaRB(green). The sample period is from January 1990 to December 2017.

3.4.3 Other Key Quantities

In the previous section, we discuss the sampling inefficiency of the parameters. It is also interesting to examine whether TaRB will bring higher sampling efficiency to other key quantities of bond pricing than NUTS. we will investigate the sampling efficiency of latent factors, term premium and predictive yields in this section.

Parameters		TaRB			HMC
	2.5%	mean	97.5%	ineff.	ineff.
$G_{1,1}$	0.90	0.99	1.01	1.2	6.4
$G_{2,2}$	0.81	0.92	1.01	2.3	5.4
$G_{3,3}$	0.83	0.90	0.94	1.5	8.5
λ_1	-0.08	-0.07	-0.065	5.4	144.29
λ_2	-0.22	-0.10	-0.01	9.4	142.08
λ_3	-0.02	0.01	0.09	5.6	146.87
$V_{1,1}$	2.88	3.14	3.44	5.6	149.34
$V_{2,2}$	2.12	2.70	2.84	6.8	40.01
$V_{3,3}$	5.3	5.8	6.47	4.4	141.27
$\sigma_1 \times 1000$	2.43	1.01	4.10	0.9	5.65
$\sigma_2 \times 1000$	0.96	1.40	4.93	0.8	7.80
$\sigma_3 \times 1000$	0.86	0.90	4.14	0.7	5.80
$\sigma_4 \times 1000$	1.27	1.79	2.76	1.3	3.12
$\sigma_5 \times 1000$	2.17	1.23	2.62	1.25	7.86
$\sigma_6 \times 1000$	4.87	5.25	7.20	0.58	8.01
$\sigma_7 \times 1000$	0.12	0.58	0.83	0.79	9.47
$\sigma_8 \times 1000$	0.98	1.12	1.47	0.91	13.68
κ	0.07	0.09	0.10	3.14	136.2

Table 3.2: **Posterior Summary for selected Parameters: $\mathcal{M}_0$**

Inefficiency comparison between No-U-Turn-Sampler (NUTS) and Random blocking scheme (TaRB-MH)

Parameters	TaRB				HMC
	2.5%	mean	97.5%	ineff.	ineff.
$G_{1,1}$	0.86	0.95	0.99	1.9	4.1
$G_{2,2}$	0.82	0.94	1.01	1.3	3.4
$G_{3,3}$	0.81	0.88	0.92	1.0	8.2
$G_{4,4}$	0.79	0.84	0.94	0.98	3.4
$G_{5,5}$	0.83	0.93	0.94	1.4	6.8
λ_1	-0.08	-0.07	-0.065	2.4	144.29
λ_2	-0.3	-0.2	0.01	3.4	142.08
λ_3	-0.05	-0.01	0.2	5.6	146.87
$V_{1,1}$	3.88	4.01	4.44	5.6	149.61
$V_{2,2}$	2.10	2.90	3.02	2.8	148.91
$V_{3,3}$	4.3	4.8	5.5	4.4	145.27
$V_{4,4}$	4.2	6.8	6.47	4.4	149.27
$V_{5,5}$	5.1	5.9	6.40	4.4	149.83
$\sigma_1 \times 1000$	2.2	0.90	3.9	0.9	5.65
$\sigma_2 \times 1000$	1.1	1.8	4.9	0.8	7.80
$\sigma_3 \times 1000$	0.86	0.90	1.14	0.7	45.80
$\sigma_4 \times 1000$	1.27	1.79	1.90	1.3	23.12
$\sigma_5 \times 1000$	2.17	1.23	2.62	1.25	7.86
$\sigma_6 \times 1000$	5.87	1.25	7.20	0.58	27.01
$\sigma_7 \times 1000$	0.12	0.58	0.83	3.70	6.47
$\sigma_8 \times 1000$	0.98	1.32	1.97	1.92	113.68
κ	0.04	0.079	0.12	5.14	132.2

Table 3.3: **Posterior Summary for selected Parameters: $\mathcal{M}_1$**

Inefficiency comparison between Hamiltonian Monte Carlo (HMC) and Random blocking scheme (TaRB-MH). The sample period is from January 1990 to December 2017.

First, Figures 3.5 report the sampling inefficiency of latent factors. Obviously, the TaRB-MH method is superior to NUTS by a margin of more than 100 in terms of efficiency for all model specifications we consider in our experiment.

Second, Figure 3.8 plots the sampling inefficiency of term premium of HMC and TaRB for different model specifications. Again, it can be confirmed that the TaRB-MH method simulates the term premium with the high efficiency while the NUTS' sampling performance is volatile and relatively inefficient. Third, Figure 3.8 plots the sampling inefficiency of one-month-ahead predictive yields for different models. The inefficiency is comparable for

Parameters		TaRB			HMC
	2.5%	mean	97.5%	ineff.	ineff.
$G_{1,1}$	0.86	0.95	1.04	1.9	4.1
$G_{2,2}$	0.82	0.88	0.92	1.8	3.4
$G_{3,3}$	0.84	0.88	0.93	1.7	8.2
$G_{4,4}$	0.	0.84	0.94	2.23	3.4
$G_{5,5}$	0.83	0.93	0.94	1.56	6.8
$G_{6,6}$	0.81	0.92	1.02	2.4	6.8
λ_1	-0.09	-0.08	-0.07	14.4	121.29
λ_2	-0.22	-0.11	-0.007	7.1	146.08
λ_3	-0.045	0.001	0.05	8.6	138.1
$V_{1,1}$	2.83	3.01	3.44	15.6	146.61
$V_{2,2}$	2.10	2.67	2.94	11.8	147.91
$V_{3,3}$	5.03	5.46	5.92	10.4	145.6
$V_{4,4}$	2.7	3.01	3.28	10.59	148.27
$V_{5,5}$	4.1	4.4	4.85	9.48	142.83
$V_{6,6}$	0.9	1.04	1.19	13.45	149.83
$\sigma_1 \times 1000$	2.2	0.90	3.9	0.9	9.65
$\sigma_2 \times 1000$	2.7	3.2	3.7	0.7	7.53
$\sigma_3 \times 1000$	2.86	2.90	3.14	01.3	44.38
$\sigma_4 \times 1000$	3.27	4.29	5.33	1.3	23.12
$\sigma_5 \times 1000$	0.96	1.1	1.29	0.96	7.86
$\sigma_6 \times 1000$	0.87	1.25	2.20	0.58	9.01
$\sigma_7 \times 1000$	0.88	1.02	1.18	2.50	58.47
$\sigma_8 \times 1000$	0.72	0.84	0.98	1.92	44.68
κ	0.039	0.082	0.11	6.8	132.2

Table 3.4: **Posterior Summary for selected Parameters: $\mathcal{M}_2$.**
Inefficiency comparison between Hamiltonian Monte Carlo (HMC) and Random blocking scheme (TaRB-MH). The sample period is from January 1990 to December 2017.

NUTS and TaRB at the very short end of the yield However, as the maturity increases, the predictive yield simulated using draws from NUTS increase dramatically while the predictive yield simulated using draws from TaRB stays low regardless of the yield maturity.

3.4.4 Robust Check on ϵ

Even though NUTS automatically tune step size ϵ and trajectory length(l), we still have to input an initial ϵ(See from Algorithm 2) and the final empirical result might be still sensitive

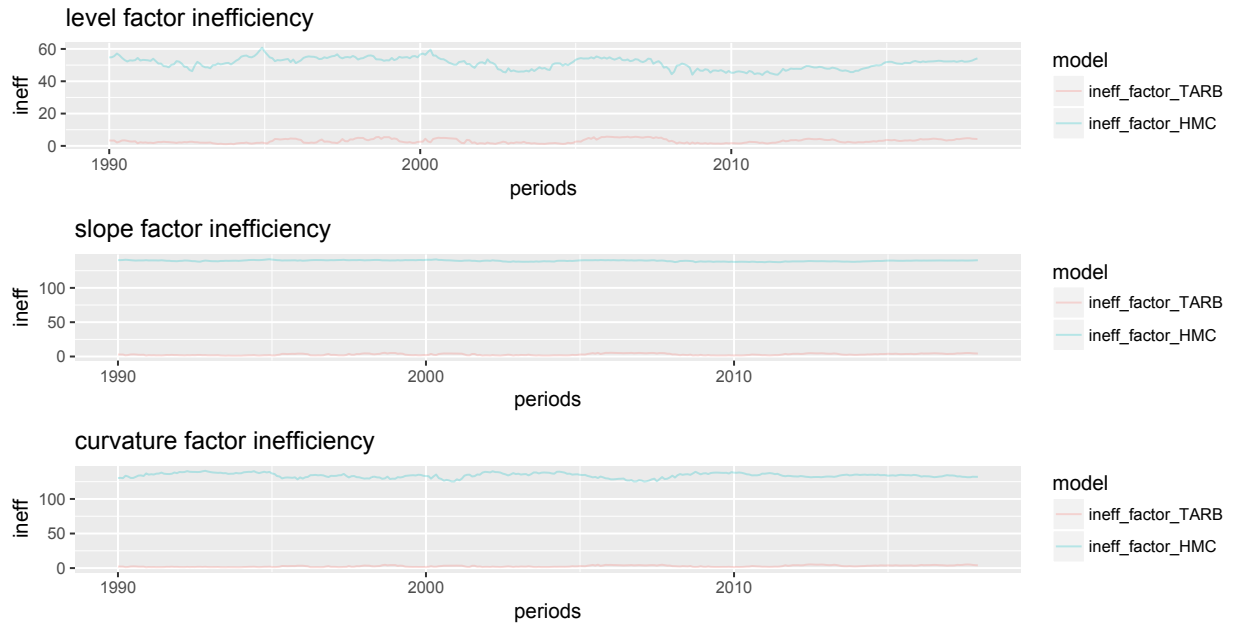


Figure 3.5: **latent factors Inefficiency: $\mathcal{M}_0$**

Inefficiency comparison between the No-U-Turn-Sampler (NUTS) and Random blocking scheme (TaRB-MH)

to this initial value. The default ϵ in the original NUTS codes[1]is set to be 1. Therefore in this section we will use $\mathcal{M}_0$ as an example for robust check on initial value of ϵ and vary the value of ϵ to see how the estimation inefficiency for NUTS parameters will change. As we can see from Table 3.5, the inefficiency centers around 124 while the lowest inefficiency would still be as high as 111.42 given $\epsilon = 0.8$. Therefore we can conclude that our result with respect to NUTS' performance in estimating affine term structure models should be robust with initial values we input.

3.5 Conclusion

This paper evaluates the posterior sampling performance of No-U-Turn sampler(NUTS) algorithm and tailored randomized-blocking Metropolis-Hastings (TaRB-MH) in terms of Gaussian Affine Term structure models. Even NUTS is often argued to be efficient in pos-

[1]check https://github.com/kasparmartens/NUTS

Figure 3.6: **latent factors inefficiency:** $\mathcal{M}_\infty$

Inefficiency comparison of between the Hamiltonian Monte Carlo (HMC) and Random blocking scheme (TaRB-MH).

terior sampling of many models, our empirical experiments with the U.S. yield curve data indicate that the TaRB-MH substantially outperforms the NUTS method in terms of the convergence and efficiency in posterior sampling. Furthermore, Such a result is also robust to different initial step size we input into the NUTS algorithm.

Figure 3.7: **latent factors inefficiency: $\mathcal{M}_2$**

Inefficiency comparison of between the Hamiltonian Monte Carlo (HMC) and Random blocking scheme (TaRB-MH).

Table 3.5: **Robust check on initial value of ϵ**

We use NUTS to estimate $\mathcal{M}_0$ and vary the initial value of ϵ to obtain different average inefficiency of parameters sampled through NUTS. The burn-in size is 1000 and the size of posterior draws is 5000.

initial value of ϵ	average inefficiency (NUTS part)
1.8	124.67
1.6	125.10
1.4	124.62
1.2	124.32
1	124.62
0.8	111.42
0.6	125.34
0.4	124.29
0.2	134.56

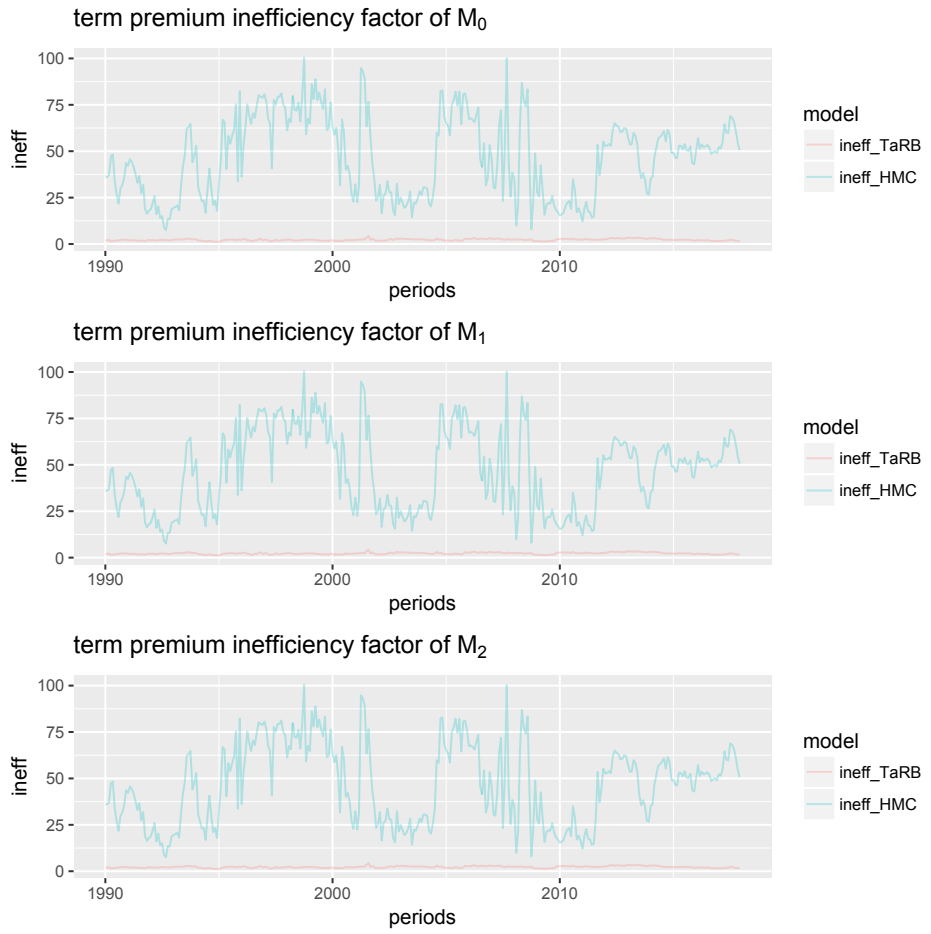


Figure 3.8: **Term structure of term premium Inefficiency**

Inefficiency comparison between the No-U-Turn-Sampler (NUTS) and Random blocking scheme (TaRB-MH)

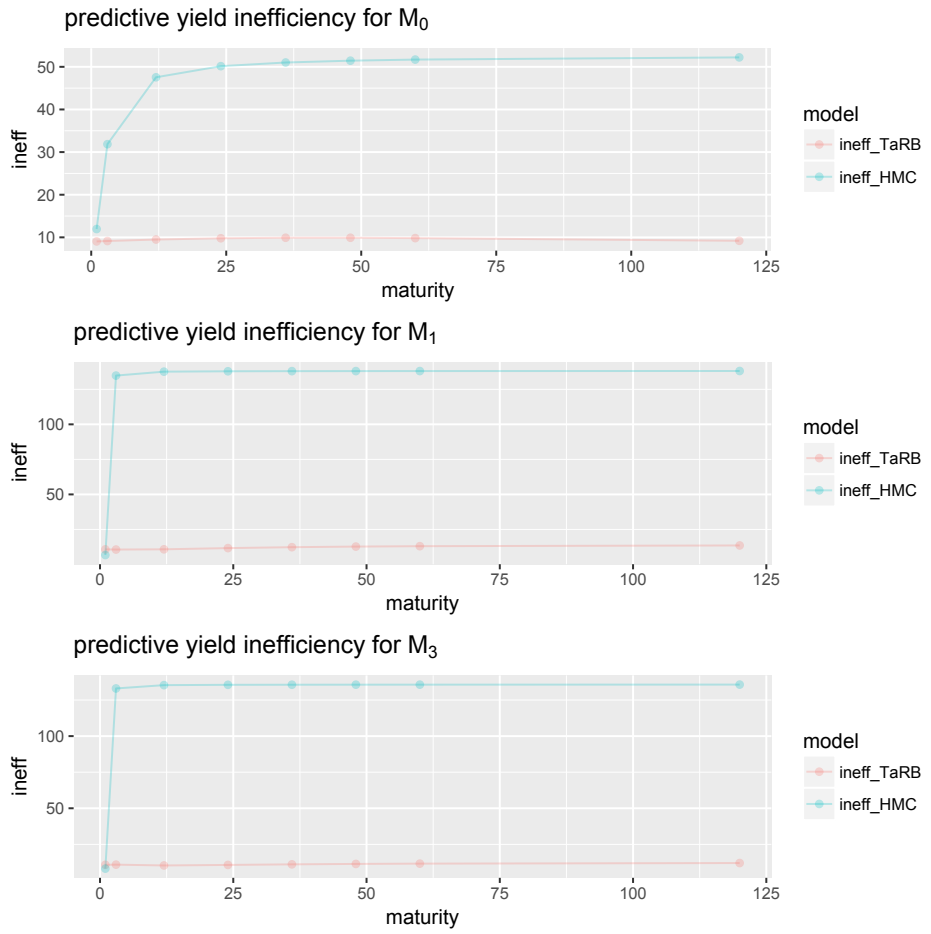


Figure 3.9: **Posterior predictive densities**

Inefficiency comparison between the No-U-Turn-Sampler (NUTS) and Random blocking scheme (TaRB-MH)

Bibliography

Abbritti, M., Gil-Alana, L. A., Lovcha, Y., and Moreno, A. (2016), "Term Structure Persistence," *Journal of Financial Econometrics*, 14(2), 331âĂŞ352.

Ang, A. and Piazzesi, M. (2001), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Working Paper*.

— (2003), "Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variable," *Journal of Monetary Economics*, 50, 745–87.

Bansal, R. and Yaron, A. (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Bauer, M. and Hamilton, J. (2018), "Robust Bond Risk Premia," *Review of Financial Studies*, 31(2), 399âĂŞ448.

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

Carter, C. and Kohn, R. (1994), "On Gibbs sampling for state space models," *Biometrika*, 81, 541–53.

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., Kang, H. K., and Xie, B. (2019), "Macro Factor Selection in Gaussian Affine Term Structure Models via Marginal Likelihood," *Working paper*.

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.

Choi, A. and Kang, H. K. (2017), "Stochastic Volatility Dynamic Nelson-Siegel Model with Time-Varying Factor Loadings and Correlated Factor Shocks," *Working Papers*.

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.

Cieslak, A. P. (2015), "Expected Returns in Treasury Bonds," *Review of Financial Studies*, 28, 2858–901.

Cochrane, J. H. and Piazzesi, M. (2005), "Bond Risk Premia," *American Economic Review*, 95, 138–160.

— (2008), "Decomposing the Yield Curve," *Unpublished manuscript*.

Cooper, I. and Priestley, R. (2009), "Time-Varying Risk Premiums and the Output Gap," *The Review of Financial Studies*, 22, 2802–2833.

de Pooter, M. (2007), "Examining the Nelson-Siegel Class of Term Structure Models," *Tinbergen Institute Discussion Paper*.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009), "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

Diebold, F. X., Rudebusch, G., and Aruoba, S. B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics*, 131, 309–338.

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

Duffee, G. R. (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57(1), 405–443.

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

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2017a), "Bond return predictability: Economic Value and links to the macro economy," *Management Science*, Articles in Advance, 1–33.

— (2017b), "Bond Return Predictability:Economic Value and Links to Macroeconomy," *Management Science*, Published online in Articles in Advance 22 Sep 2017.

Greenwood, R. and D.Vayanos (2014), "Bond Supply and Excess Bond Returns," *Review of Financial Studies*, 27, 663–713.

Gurkaynak, R. S. and Wright, J. H. (2012), "Macroeconomics and the Term Structure," *Journal of Economic Literature*, 50, 331–67.

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

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

Hoffman, M. and Gelman, A. (2014), "The No-U-Turn Sampler: Adaptively Setting Path lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, 15, 1351–1381.

Johannes, M., K. A. a. P. N. (2014), "Sequential Learning,Predictability and Optimal Portfolio Returns," *Journal of Finance*, 69, 611–644.

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.

Kandel, S. and Stambaugh, R. F. (1996), "On the predictability of stock returns: an asset-allocation perspective," *The journal of Finance*, 51, 385–424.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361–393.

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

Ludvigson, S. and Ng, S. (2009a), "A factor analysis of bond risk premia," *NBER working paper*.

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

Neal, R. (2011), "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo, Chapter 5*.

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

Piazzesi, M. (2005), "Bond yields and the federal reserve," *Journal of Political Economy*, 113, 311–344.

Robert, C. (2001), "The Metropois-Hastings Algorithm," *Monte Carlo Statistical Methods*.

Sarno, L., Schneider, P., and Christian, W. (2016), "The economic value of predicting bond risk premia," *Journal of Empirical Finance*, 37, 247–267.

Thornton, D. and Valente, G. (2012), "Out of sample prediction of bond excess returns and foward rates: An asset allocation persfective," *Review of Financial Studies*, 10, 3141–3168.

Wachter, J. A. (2006), "A consumption-based model of the term structure of interest rates," *Journal of Financial Economics*, 79(2), 365–399.

Wu, T. (2001), "Monetary Policy and the Slope Factor in Empirical Term Structure Estimation," *Working paper*.

Appendix

.1 Arbitrage-free Bond Prices

In the class of affine term structure models, the τ-period bond price at time t is assumed to be exponentially linear to the latent factors,

$$P_t(\tau) = \exp\left[-a(\tau) - b(\tau)'\boldsymbol{l_t}\right].$$

By the no-arbitrage restriction we have

$$\exp\left[-a(\tau) - b(\tau)'\boldsymbol{l_t}\right] = \exp\left[-\delta - \beta'\boldsymbol{l_t}\right]\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left[-a(\tau-1) - b(\tau-1)'\boldsymbol{l_{t+1}}\right]\right]$$

or

$$\begin{aligned}&\exp\left[-a(\tau) - b(\tau)'\boldsymbol{l_t}\right]\\ &= \exp\left[-\delta - \beta'\boldsymbol{l_t} - a(\tau-1) - b(\tau-1)'[K_l^{\mathbb{Q}} + G_{ll}^{\mathbb{Q}}\boldsymbol{l_t}] + \frac{1}{2}b(\tau-1)'\Omega_{ll}b(\tau-1)\right].\end{aligned}$$

Therefore,

$$-a(\tau) - b(\tau)'\,\boldsymbol{l_t} = -\delta - a(\tau-1) - b(\tau-1)'K_l^{\mathbb{Q}} + \frac{1}{2}b(\tau-1)'\Omega_{ll}b(\tau-1) - (\beta' + b(\tau-1)'G_{ll}^{\mathbb{Q}})\boldsymbol{l}_t$$

Matching coefficients on the left and right hands yields the recursive system for the coefficients $a(\tau)$ and $b(\tau)$,

$$a(\tau) = \delta + a(\tau - 1) + b(\tau - 1)' K_l^{\mathbb{Q}} - \frac{1}{2} b(\tau - 1)' \Omega_{ll} b(\tau - 1),$$
$$b(\tau) = \beta + G_{ll}^{\mathbb{Q}\prime} b(\tau - 1),$$

where this system is initialized at $a(0) = 0$ and $b(0) = 0_{l \times 1}$, given $P_t(0) = 1$ for all $t = 1, 2, .., T$.

.2 Likelihood

Denote the history of the observations up to time t by $\mathcal{F}_t = \{\mathbf{y}_i\}_{i=1}^t$. Then, the likelihood is given by the joint density of the observations,

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

Using the Kalman filter, the likelihood density at each data point can be obtained as

$$p(\mathbf{y}_t|\mathcal{F}_{t-1}, \theta) = \mathcal{N}(\mathbf{y}_t|\mathbf{y}_{t|t-1}, V_{t|t-1}),$$

where

$$\mathbb{E}[\mathbf{f}_t|\mathcal{F}_{t-1}, \theta] = \mathbf{f}_{t|t-1} = G\mathbf{f}_{t-1|t-1} + K$$
$$Var[\mathbf{f}_t|\mathcal{F}_{t-1}, \theta] = Q_{t|t-1} = GQ_{t-1|t-1}G' + \Omega,$$

$$\mathbb{E}[\mathbf{y}_t|\mathcal{F}_{t-1}, \theta] = \mathbf{y}_{t|t-1} = \mathbf{a} + \mathbf{B}\mathbf{f}_{t|t-1},$$

$$Var[\mathbf{y}_t|\mathcal{F}_{t-1},\theta] = V_{t|t-1} = \mathbf{B}Q_{t|t-1}\mathbf{B}' + \Sigma,$$

$$\begin{aligned}
\mathbf{K}_t &= Q_{t|t-1}\mathbf{B}'V_{t|t-1}^{-1}, \\
\mathbb{E}[\mathbf{f}_t|\mathcal{F}_t,\theta] &= \mathbf{f}_{t|t} = \mathbf{f}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{y}_{t|t-1}), \\
Var[\mathbf{f}_t|\mathcal{F}_t,\theta] &= Q_{t|t} = Q_{t|t-1} - \mathbf{K}_t\mathbf{B}Q_{t|t-1}.
\end{aligned}$$

Notice that only the upper $l \times l$ block of $Q_{t|t}$ is non-zero, while the other blocks are all zeros, because the macro factors are observed without errors. The initial factors, $\mathbf{f}_0$ are unobserved and assumed to follow its ergodic distribution given the parameters,

$$\mathbf{f}_0|\mathcal{F}_0,\psi \sim \mathcal{N}\left(\mathbf{f}_{0|0} = 0, Q_{0|0}\right),$$

where $Q_{0|0}$ is the unconditional covariance matrix of $\mathbf{f}_t$ such that

$$vec(Q_{0|0}) = \left[\mathbf{I}_{(l+m+z)} - (G \otimes G)\right]^{-1} \times vec(\Omega).$$

.3 Posterior Sampling Algorithm

The model parameters to be estimated reduce to

$$\theta = \{\kappa,\ G,\ K_m,\ K_z,\ \delta,\ V,\ \Gamma,\ \Sigma,\ \lambda\}.$$

To facilitate our discussion, we denote $\psi = \{G_{ll}^{\mathbb{Q}},\ K_m,\ K_z,\ \delta,\ \beta,\ \Omega,\ \lambda\}$, and we have $\theta = \{\psi,\ \Sigma,\ G\}$. Unlike Σ and G, the parameters in ψ can not be sampled by Gibbs-sampling, because their full conditional distributions are not tractable. The target distribution to be simulated

is the joint posterior distribution of the parameters and latent factors,

$$\theta, \{\boldsymbol{l}_t\}_{t=1}^T | \mathbf{Y}.$$

Using the posterior draws for the parameters and factors, we can infer the term premiums and factor loadings.

In each MCMC iteration, they are updated in the order of ψ, $\{\boldsymbol{l}_t\}_{t=1}^T$, G, and Σ. We use the TaRB-MH of Chib and Ramamurthy (2010) to implement the sampling of ψ and Gibbs sampling for G and Σ. For sampling $\{\boldsymbol{l}_t\}_{t=1}^T$ we employ the multi-move method of Carter and Kohn (1994). We outline our MCMC algorithm as follows:

Algorithm 3: MCMC sampling

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

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

Step 1.(a): Randomly determine the number of blocks and their components in ψ,

$$\psi_1,\ \psi_2,\ ..,\ \psi_{B_g}$$

where B_g is the number of blocks at the gth MCMC iteration. Set $l = 1$.

Step 1.(b): Maximize

$$\ln\{p(\mathbf{Y}|\psi_l, \psi_{-l}, G^{(g-1)}, \Sigma^{(g-1)}) \times \pi(\psi_l, \psi_{-l}, G^{(g-1)}, \Sigma^{(g-1)})\}$$

w.r.t ψ_l to obtain the mode, $\bar{\psi}_l$ and compute the inverted negative Hessians computed at the mode, $V_{\bar{\psi}_l}$.

Step 1.(c): Draw a proposal for ψ_l, denoted by $\psi_l^\dagger$, from the selected multivariate

Student-t distribution,

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

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

$$\begin{aligned}
\psi_l &= \psi_l^{\dagger} \qquad \text{if } u < \alpha\left(\psi_l^{(g-1)}, \psi_l^{\dagger} | \mathbf{y}, \psi_{-l}, G^{(g-1)}, \Sigma^{(g-1)}\right) \\
\psi_l &= \psi_l^{(g-1)} \quad \text{if } u \geq \alpha\left(\psi_l^{(g-1)}, \psi_l^{\dagger} | \mathbf{y}, \psi_{-l}, G^{(g-1)}, \Sigma^{(g-1)}\right)
\end{aligned}$$

Step 1.(e): $l = l + 1$, and go to Step 1.(b) while $l \leq B_g$.

Step 2: Draw $\{\boldsymbol{l}_t\}_{t=1}^T$ from $\{\boldsymbol{l}_t\}_{t=1}^T | \mathbf{Y}, \psi^{(g)}, G^{(g)}, \Sigma^{(g)}$ based on the Carter and Kohn method.

Step 3: Draw $G^{(g)}$ from $G | \mathbf{Y}, \{\boldsymbol{l}_t\}_{t=1}^T, \psi^{(g)}, \Sigma^{(g)}$.

Step 4: Draw $\Sigma^{(g)}$ from $\Sigma | \mathbf{Y}, \{\boldsymbol{l}_t\}_{t=1}^T, \psi^{(g)}, G^{(g)}$.

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.

Now we discuss the posterior sampling algorithm in the order of ψ, $\{\boldsymbol{l}_t\}_{t=1}^T$, G, and Σ.

.3.1 ψ Sampling via TaRB-MH

Given the target distribution, the efficacy of the MH sampling depends on the way of grouping parameters into multiple blocks and constructing proposal distributions. In each MCMC cycle, sampling ϕ consists of three steps. The first step is to randomly choose the number of blocks and their components. These blocks are sequentially updated given the other parameters. The second step is to construct Student-t proposal distribution of each block using the output of stochastic optimization given the other parameters. The third step is to

draw from the proposal distribution and update the block based on the MH algorithm. We discuss each of the steps in details as follows.

Randomizing Blocks

In the gth MCMC iteration, the sampler begins by randomly grouping the parameters into B_g blocks

$$\psi_1,\ \psi_2,\ ...,\ \psi_{B_g}$$

with randomly selected components in each block. To maximize the efficacy in a high-dimensional problem, we should simulate highly correlated parameters in one block and the remaining parameters in separate blocks. However, when the likelihood is severely nonlinear as shown in Equation (2.3.2), it is difficult to form appropriate fixed blocks a priori. Randomizing blocking scheme is particularly valuable in this case. This blocking scheme enables us to avoid the pitfalls from a poor choice of fixed blocks. We let the parameters in ϕ form random blocks and Σ form a fixed block since the full conditional distribution of Σ is tractable.

Algorithm 2: Randomizing blocks

Step 0: Let the number of the parameters denoted by $n_\psi = \mathsf{dim}(\psi)$. $Idx = (1, 2, .., n_\psi)'$ is a $n_\psi \times 1$ vector. Then, an one-to-one mapping between Idx and the parameters in ψ exists such that $\psi(Idx(j))$ is the $Idx(j)$th parameter in ψ where $Idx(j)$ is the jth element of Idx.

Step 1: At the gth MCMC iteration, randomize the permutation of Idx and denote it by Idx^*.

Step 1(a): set $i = 2$, $b = 1$, and $cp_0 = 1$.

Step 1(b): draw u from $Uniform\,(0, 1)$.

Step 1(c): if $u > tp$, then $cp_b = i$ and $b = b + 1$.

Step 1(d): $i = i + 1$ and go to Step 1(a) while $i \leq n_\theta$.

Step 2: $B_g = b$ is the number of blocks. For $1 \leq j \leq B_g - 1$,

$$\theta_j = \{\theta(Idx^*(i)) | cp_{j-1} \leq i \leq cp_j - 1\}$$

is the jth block, and $\phi_{B_g} = \{\psi(Idx^*(i)) | cp_{B_g - 1} \leq i\}$ is the last block.

Note that the average block size (i.e., the number of components in a block) is $1/(1-tp)$. If the average block size is too small or too big, the efficacy of the MCMC simulation reduces. In our Matlab toolbox, the default argument for tp is 0.8, so that the average block size is set to be five. However, researchers are allowed to control tp depending on their preference. This is appropriate according to our extensive empirical experiments. Once the number of blocks and their components are randomly determined, then we update ψ by sequentially iterating through those blocks with the help of the multiple-block MH algorithm (Chib and Greenberg (1995)).

Block-wise Stochastic Optimization

The second stage is to construct a proposal distribution of the first block ψ_1 by finding the global mode ($\bar{\psi}_1$) of its full conditional density and the inverted negative Hessians computed at the mode:

$$\bar{\psi}_1 = \arg\max_{\psi_1} \log\{p(\mathbf{Y}|\phi_1, \psi_{-1})\pi(\psi_1, \psi_{-1})\} \text{ subject to } \{\psi_1, \psi_{-1}\} \in \mathcal{R}, \tag{.3.1}$$

$$V_{\bar{\psi}_1} = \left(-\frac{\partial^2 \log\{p(\mathbf{Y}|\bar{\psi}_1, \psi_{-1})\pi(\bar{\theta}_1, \psi_{-1})\}}{\partial\psi_1 \partial\psi_1'} \right)^{-1}, \tag{.3.2}$$

where ψ_{-l} is the parameters in ψ except θ_l. 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 particular, using Student-t distribution rather than normal distribution helps move from the global region to a local region and vice versa because of its fat tail property.

In order to find the mode satisfying the restrictions on the parameters, we utilize a simulated annealing algorithm combining the Newton-Raphson method. We refer to this optimizer as the SA-Newton method. Simulated annealing algorithms are a stochastic global optimizer that is particularly useful in the presence of a multi-modality problem. In finding the mode, we try with a high initial temperature and a relatively low 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 five and 0.01, respectively. Then, the probability of accepting a point with a lower function value that is proportional to the current temperature is high, which helps move to another search regions with higher function values. Once the simulated annealing stages are complete, the optimizer switches to the Newton-Raphson method to converge to the mode quickly.

Once the stochastic optimizer is implemented, the probability of finding the global mode, not a local mode, increases and there is a chance to find the global mode even when the starting value is located near a local mode. This never happens when only deterministic optimizations such as the Newton-Raphson method are used. Moreover, we set the total number of simulated annealing stages to be two, so the search for the global mode through the SA-Newton method is not quite extensive. Even if the maximum simulated annealing stage is set to be three or larger, the efficacy gain is found to be marginal whereas the computing

time substantially increases according to our comprehensive empirical experiments. In our Matlab toolbox, users can try different simulated annealing parameters including the number of stages, initial temperature, and temperature reduction factor.

Block-wise Metropolis-Hastings Algorithm

Now we move to the third stage. Using the mode and Hessian at the mode, we construct a multivariate Student-t proposal distribution,

$$q\left(\psi_1|\psi_{-1}\right) = St\left(\psi_1|\bar{\psi}_1, V_{\bar{\psi}_1}, 15\right).$$

where the St is a multivariate Student-t density with $\nu = 15$ degrees of freedom. Let the proposal value, denoted by $\theta_1^{\dagger}$, drawn from

$$\psi_1^{\dagger} \sim St\left(\bar{\psi}_1, V_{\bar{\psi}_1}, 15\right).$$

Then, given the current value, $\psi_1^{(g-1)}$, this proposal value is accepted with the usual MH probability of the move,

$$\begin{aligned}\alpha\left(\psi_1^{(g-1)}, \psi_1^{\dagger}|\mathbf{Y}, \psi_{-1}\right) &= \min\left\{1, \frac{\pi\left(\psi_1^{\dagger}, \psi_{-1}|\mathbf{Y}\right)}{\pi\left(\psi_1^{(g-1)}, \psi_{-1}|\mathbf{Y}\right)} \frac{q\left(\psi_1^{(g-1)}|\psi_{-1}\right)}{q\left(\psi_1^{\dagger}|\psi_{-1}\right)}\right\} \\ &= \min\left\{1, \frac{p(\mathbf{Y}|\psi_1^{\dagger}, \psi_{-1})\pi(\psi_1^{\dagger}, \psi_{-1})}{p(\mathbf{Y}|\psi_1^{(g-1)}, \psi_{-1})\pi(\psi_1^{(g-1)}, \psi_{-1})} \frac{q\left(\psi_1^{(g-1)}|\psi_{-1}\right)}{q\left(\psi_1^{\dagger}|\psi_{-1}\right)}\right\}.\end{aligned} \tag{.3.3}$$

If the proposal is accepted, then $\psi_1^{(g)} = \psi_1^{\dagger}$. Otherwise, $\psi_1^{(g)} = \psi_1^{(g-1)}$. Note that if the proposal value does not satisfy the restrictions on the parameters, then it is immediately rejected because its prior density $\pi(\psi_1^{\dagger}, \psi_{-1})$ is zero.

The remaining blocks, (ψ_2, ψ_3, ..., ψ_{B_g}) are sequentially updated within the MCMC itera-

tion in the same manner as the first block. Chib and Ramamurthy (2010) have already shown that the TaRB-MH method is scalable and reliable in cases of irregular likelihood/posterior despite a relatively high computational burden in the application of dynamic stochastic general equilibrium models.

.3.2 Latent Factor Sampling

Next, we sample the factors, $\{\boldsymbol{l}_t\}_{t=1}^{T}$ given the data and parameters based on the multi-move method, in which the factors are updated in one block.

Algorithm 3: Factor sampling

Step 1: Sample $\mathbf{f}_T$ from $\mathcal{N}\left(\mathbf{f}_{T|T}, Q_{T|T}\right)$

Step 2: For $t = T-1, T-2, \ldots, 1$, sample $\mathbf{x}_t$ from

$$\mathbf{f}_t|\mathcal{F}_t, \mathbf{f}_{t+1}, \theta \sim N\left(\mathbf{f}_{t|\mathbf{f}_{t+1}}, Q_{t|\mathbf{f}_{t+1}}\right)$$

where

$$\mathbf{f}_{t|\mathbf{f}_{t+1}} = \mathbf{f}_{t|t} + Q_{t|t}G\prime\left(GQ_{t|t}G' + \Omega\right)^{-1}\left(\mathbf{f}_{t+1} - K - G\mathbf{f}_{t|t}\right), \tag{.3.4}$$

$$Q_{t|\mathbf{f}_{t+1}} = Q_{t|t} - Q_{t|t}G'\left(P_{t+1|t}\right)^{-1}G^{Q}_{t|t}. \tag{.3.5}$$

The first l element of $\mathbf{f}_t$ is the vector of latent factors, $\boldsymbol{l}_t$, and the last $m+z$ elements are the observed factors at time t.

.3.3 G Sampling

Now we turn to the sampling of G. Suppose that $\boldsymbol{F}_{\boldsymbol{a}:\boldsymbol{b}} = (\mathbf{f}_{\boldsymbol{a}}, \mathbf{f}_{\boldsymbol{a}+1}, .., \mathbf{f}_{\boldsymbol{b}})'$ is a $(b-a+1) \times (l+m+z)$ matrix. G_0 indicates the prior mean matrix of G, and G_{v0} is the corresponding

prior variance matrix of G. Then, G is updated from its full conditional distribution,

$$vec(G)|\mathbf{Y}, \{\mathbf{f}_t\}_{t=1}^T, \theta \sim \mathcal{N}(vec(\hat{G}), G_v),$$

where $G_v = G_{v0} \times \Omega^{-1} \otimes (\boldsymbol{F}'_{1:\boldsymbol{T}-1} \times \boldsymbol{F}_{1:\boldsymbol{T}-1})$ and

$$vec(\hat{G}) = G_v \times (G_{v0} \times vec(G_0) + (\Omega^{-1} \otimes \boldsymbol{F}_{1:\boldsymbol{T}-1})' \times vec(\boldsymbol{F}_{2:\boldsymbol{T}}).$$

.3.4 Σ Sampling

The full conditional distribution of the error variances is tractable. Given the factors, parameters, and data, the error variances are updated via inverse gamma distributions,

$$\sigma^2_{\tau_i}|\mathbf{Y}, \{\mathbf{f}_t\}_{t=1}^T, \theta \sim \mathcal{IG}((v_0 + T)/2,\ \delta_{1,i}/2),$$

where $\delta_{1,i} = \sum_{t=1}^T (R_t(\tau_i) - a(\tau_i) - b(\tau_i) \times \mathbf{f}_t)^2 + \delta_0$ for $i = 1, 2, .., N$.

.4 Marginal Density Computation

By the identity in Equation (1.4.4), the log marginal likelihood can be expressed as

$$\log m(\mathbf{Y}) = \log p(\mathbf{Y}|\theta^*, \Sigma^*, G^*) + \log \pi(\psi^*, \Sigma^*, G^*) - \log \pi(\theta^*, \Sigma^*, G^*|\mathbf{Y}).$$

The first term is the likelihood evaluated at the mode and the second term can be directly calculated according to our prior specification. The last term can be decomposed as

$$\log \pi(\psi^*, G^*, \Sigma^*|\mathbf{Y}) = \log \pi(\Sigma^*|\mathbf{Y}) + \log \pi(G^*|\Sigma^*, \mathbf{Y}) + \log \pi(\psi^*|\mathbf{Y}, G^*, \Sigma^*).$$

In the above equation, the first term is the log marginal posterior density of Σ^*, which is numerically computed as

$$\begin{aligned}\log \pi(\Sigma^*|\mathbf{Y}) &= \sum_{j=1}^{N} \log \pi(\sigma_j^{*2}|\mathbf{Y}) \\ &= \sum_{j=1}^{N} \left(\frac{1}{n_1} \sum_{g=1}^{G} \log \pi(\sigma_j^{*2}|\mathbf{Y}, \mathbf{L}^{(g)}, \psi^{(g)}) \right),\end{aligned}$$

where $\mathbf{L} = \{\boldsymbol{l}_t\}_{t=1}^{T}$,

$$\pi(\sigma_j^{*2}|\mathbf{Y}, \mathbf{L}^{(g)}, \theta^{(g)}) = \mathcal{IG}(\sigma_j^{*2}|v_1^{(g)}/2, \delta_{1,j}^{(g)}/2),$$

and $(\psi^{(g)}, v_1^{(g)}, \delta_{1,j}^{(g)})$ are obtained from the $g$the MCMC cycle. For the second term, we simulate the reduced MCMC run, in which $\mathbf{L}, G, \psi|\mathbf{Y}, \Sigma^*$ are sampled. Using the output of the reduced run, we can compute the full conditional mean and variance of $vec(G)$ (i.e., $vec(\hat{G}^{(g)}), G_v^{(g)}$) over the reduced run cycles. Then, the second term is obtained as

$$\log \pi(G^*|\Sigma^*, \mathbf{Y}) = \frac{1}{n_1} \sum_{g=1}^{n_1} \log \pi(G^*|\mathbf{Y}, \mathbf{f}^{(g)}, \psi^{(g)}, \Sigma^*),$$

where

$$\log \pi\left(G^*|\mathbf{Y}, \mathbf{f}^{(g)}, \psi^{(g)}, \Sigma^*\right) = \mathcal{N}\left(vec(G^*)|vec(\hat{G}^{(g)}), G_v^{(g)}\right).$$

Finally, we compute the last term, $\log \pi(\psi^*|\mathbf{Y}, \Sigma^*, G^*)$. Since ψ is sampled through the TaRB-MH, we follow Chib and Ramamurthy (2010) and fix the number of blocks in the posterior estimation step at the average number of blocks (say B). Then, we construct the blocks, ψ_1, ψ_2, ..., ψ_B where each block consists randomly chosen components from ψ. Then, $\log \pi(\psi^*|\mathbf{Y}, \Sigma^*, G^*)$ can be decomposed as

$$\begin{aligned}\log \pi(\theta^*|\mathbf{Y}, \Sigma^*, G^*) &= \log \pi(\psi_1^*|\mathbf{Y}, \Sigma^*, G^*) + \log \pi(\psi_2^*|\mathbf{Y}, \psi_1^*, \Sigma^*, G^*) \\ &\quad ... + \log \pi(\psi_B^*|\mathbf{Y}, \psi_1^*, \psi_2^*, ...\psi_{B-1}^* \Sigma^*, G^*).\end{aligned}$$

This conditional density computation is completed by each ordinate, $\pi(\psi_b^*|\mathbf{Y}, \psi_b^*, ...\psi_{b-1}^*)$ for $b = 1, ..., B$, where ψ_0^* is empty.

As shown by Chib and Jeliazkov (2001), the simulation-consistent estimate of the bth ordinate is given by

$$\begin{aligned}&\hat{\pi}(\psi_b^*|\mathbf{Y}, \Sigma^*, \psi_1^*, ...\psi_{b-1}^*) \\ &= \frac{n_1^{-1}\sum_{g=1}^{n_1}\alpha(\psi_b^{(g)}, \psi_b^*|\mathbf{Y}, \Sigma^*, \Psi_{b-1}^*, \Psi^{b+1,(g)})q(\psi_b^*|\mathbf{Y}, \Sigma^*, \Psi_{b-1}^*, \Psi^{b+1,(g)})}{n_1^{-1}\sum_{j=1}^{n_1}\alpha(\psi_b^{(g)}, \psi_b^*|\mathbf{Y}, \Sigma^*, \Psi_{b-1}^*, \Psi^{b+1,(g)})},\end{aligned}$$

with $\Psi_{b-1} = (\psi_1, ..., \psi_{b-1})$ and $\Psi^{b+1} = (\psi_{b+1}, ..., \psi_B)$. The denominator is the average acceptance rate with respect to the gth draws, $\psi_b^{(g)}$ and $\Psi^{b+1,(g)}$, generated from the conditional distribution, $\psi_b, \Psi^{b+1}|\mathbf{Y}, \Psi_{b-1}^*$. That is, these draws are obtained from the bth reduced MCMC run, in which the parameters at the preceding blocks are sampled conditioned on Ψ_{b-1}^*. On the other hand, the numerator is the conditional expectation of the product measure,

$$\alpha(\psi_b, \psi_b^*|\mathbf{Y}, \Sigma^*, G^*, \Psi_b^*) \times q(\psi_b^*|\mathbf{Y}, \Sigma^*, G^*, \Psi_{b-1}^*, \Psi^{b+1}).$$

Conveniently, draws from $\psi^{b+1}|\mathbf{Y}, \Sigma^*, G^*, \Psi_b^*$ are available from the calculation of the $(b+1)$st stage numerator which are supplemented with draws from $\psi_b|\mathbf{Y}, \Sigma^*, G^*, \Psi_{b-1}^*, \Psi^{b+1}$. In the final stage, we generate n_1 posterior draws from $\psi_B|\mathbf{Y}, \Sigma^*, \Psi_{B-1}^*$, and use them to compute the denominator.

.5 Simulation Study

In our framework, a different collection of relevant macro factors correspond to different zero restrictions on $G^{\mathbb{P}}$, and hence, a different model. Therefore, comparing different models in terms of the Bayesian model choice provides the means to identify the relevant macro

Table 6: **Model selection frequency based on 20 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.

True relevant factors	selection frequency by simulation
IP, PCE	65%
Φ	90%

factors. In this section, we use simulations to assess how well the framework performs in identifying the relevant macro factors. The simulation studies are designed to mimic the real world features of the factors in our setting. We will consider five candidate factors in our simulation study: CPI, IP, PCE, FFR, TCU. We consider two true models in our simulation study: one model includes IP and PCE as the relevant macro factor, while the other includes no relevant macro factors. We compare the true model with other 31 models and see whether the marginal likelihoods correctly picks the true one. To generate data from the true models, we need to fix the parameters at the posterior mode to ensure that the generated data on the yield curve and driving factors resemble real data. In our simulation study, we simulate the model for 300 periods. We use the first 200 periods to train the prior of 32 different models and then we use the recent 100 observations for the posterior sampling and marginal likelihood computation.

Specifically, our simulation study consists of two stages. The first stage is to simulate imaginary dataset, $\{\hat{\mathbf{y}}_t\}_{t=1}^T$, from the true models using the corresponding posterior parameter mode, the measurement equation (3.2.2), and transition equation (2.4.2). In the second stage, we use $\{\ \hat{\mathbf{y}}_t\}_{t=1}^T$ to estimate and compare the marginal likelihoods of all 32 competing models. These two stages are repeated 20 times.

The model selection frequencies for each true model are given in Table 6. The selection of the true model by the marginal likelihood criterion is quite high. Even in cases with two relevant factors, the frequency of selecting the correct model is about 65%. This is a substantial improvement over picking a model at random, in which case the chance of finding

the correct model would be just around 3%.