Efficient Posterior Sampling in Gaussian Affine Term Structure Models

Siddhartha Chib*
(Washington University in St. Louis)
Kyu Ho Kang†
(Korea University)

April 2016

Abstract

This paper proposes an efficient Bayesian estimation method for the fitting of Gaussian affine term structure models. It synthesizes and extends previous work by building broadly on three key themes. The first is in the formulation of a prior distribution (with some Student-t distributed marginals) that is capable of smoothing out irregularities in the posterior distribution. The second is in the adoption of the Chib and Ramamurthy (2010) tailored randomized-blocking Metropolis-Hastings (TaRB-MH) sampling method which has not been investigated in the current context. The third is in the use of the Nelson-Siegel type restrictions to identify the latent factors. These restrictions, which enable interpretation of the latent factors as time-varying level, slope, and curvature of the yield curve, are useful in the prior elicitation step and also promote numerical efficacy. We apply our methods to models containing up to five factors, three latent and two observed. According to our empirical experiments with the U.S. yield curve data, the performance of our approach is more simulation-efficient than the MCMC method that is identical except in the use of the randomized blocking step. An available Matlab toolbox implements this approach with options to choose the number of latent and observed factors, priors for model parameters, and specification of the factor process. (JEL G12, C11, E43)

Keywords: Arbitrage-free Nelson-Siegel restrictions; Irregular likelihood surface, Efficient MCMC sampling, Matlab toolbox; TaRB-MH.

Address for correspondence: Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Bookings Drive, St. Louis, MO 63130. E-mail: chib@wustl.edu.

†*Address for correspondence*: Department of Economics, Korea University, Anamdong, Seongbuk-Gu, Seoul, 136-701, South Korea, E-mail: kyuho@korea.ac.kr.

1 Introduction

Gaussian arbitrage-free affine term structure models (GATSMs) are a class of default-free bond pricing model based on the stochastic discount factor (SDF) approach that produce risk-adjusted bond prices and yields satisfying a no-arbitrage condition. GATSMs are widely used in the fitting of default-free government bond yields, estimating the term premium, analyzing monetary policy effects and forecasting future yield curves. Despite their importance in practice, the estimation of GATSMs is challenging. This is mainly because the likelihood function of these models contains multiple modes and other irregularities that preclude estimation by conventional numerical methods.

Our goal in this paper is to propose an efficient Bayesian estimation method that synthesizes and extends the previous work on these models. The method proposed builds on three key themes. The first is the careful specification of the prior distribution to incorporate available knowledge, which helps to smooth out irregularities in the posterior distribution as demonstrated by Chib and Ergashev (2009). The second is the adoption of the Chib and Ramamurthy (2010) tailored randomized-blocking Metropolis-Hastings (TaRB-MH) sampling method that has been shown to be effective in other situations but has not been investigated in the current context. The third is the use of the Nelson-Siegel type restrictions from Christensen, Diebold, and Rudebusch (2011) to identify the latent factors. The need to restrict the latent factors in such models is comprehensively discussed by Hamilton and Wu (2012). We do not employ the restrictions of Dai and Singleton (2000) where the latent factors are identified by the degree of persistence because the Nelson-Siegel type restrictions enable the explicit interpretation of the latent factors as time-varying level, slope, and curvature of the yield curve, which is useful in the Bayesian prior elicitation step. We have also found that the latter restrictions have better numerical properties. We apply our methods to models containing up to five factors, three latent and two observed. GATSM models of this size are rarely estimated because of the aforementioned numerical difficulties.

According to our empirical experiments with the U.S. yield curve data, the performance of the TaRB-MH based approach is far better than the corresponding MCMC method that is identical except in the use of the randomized blocking step. Evidence of the superior performance of our method is provided below for the computation of

the deep parameters of the model and for the estimation of key quantities such as the term premium and multi-step-ahead predictive yield curves. We have implemented our approach in an available Matlab toolbox with options to choose the number of latent and observed factors, the prior of the model parameters, and the specification of the factor process.

We briefly mention that other MCMC sampling schemes, such as the Hamiltonian Monte Carlo (HMC) and random walk MH, perform poorly. Because of the irregular posterior surface, HMC for example, tends to become trapped in local modes or in regions where the required Hessian is singular. We have largely failed to tune these methods to work in these models.

The reminder of the paper is organized as follows. Section 2 gives a brief description of the Gaussian affine term structure model. In Section 3 we describe the prior, the factor-identifying restrictions, and our posterior sampling method. The results are given in Section 4, and concluding remarks are presented in Section 5.

2 Gaussian Affine Models

2.1 Assumptions and Bond Prices

This section describes the standard Gaussian affine term structure model. The price $P_t(\tau)$ of a default-free zero-coupon bond with τ periods to maturity at time t is the discounted value of its face value

$$P_t(\tau) = \frac{1}{(1 + R_t(\tau))^\tau} \tag{2.1}$$

where the nominal discount rate

$$R_t(\tau) \approx -\frac{1}{\tau} \log P_t(\tau) \tag{2.2}$$

is the yield. The relation between $\{\tau, R_t(\tau)\}$ is the term structure of interest rates. Virtually all modern work on asset pricing in finance is based on the stochastic discount factor (SDF) approach. For an asset such as a bond, the SDF approach requires that

$$P_t(\tau) = \mathbb{E}_t \left[M_{t+1} P_{t+1}(\tau - 1) \right] \tag{2.3}$$

where M_{t+1} is the one-period SDF and $\mathbb{E}_t$ is the conditional expectation given the information at time t. This pricing formula enforces the no-arbitrage condition across maturities. Every valid SDF must satisfy certain conditions. For example, it must be strictly positive, and its time t expectation must equal the non-stochastic discount factor (i.e., $\mathbb{E}_t\left[M_{t+1}\right] = \exp\left(-r_t\right)$, where r_t is the risk-free short rate). An asset pricing model is simply a model of the SDF, followed by a solution for the prices that satisfy the preceding no-arbitrage condition.

For modeling the SDF, suppose that the SDF depends on factors $\boldsymbol{x}_t$ consisting of k latent factors $\boldsymbol{f}_t$ and m observed factors $\boldsymbol{m}_t$

$$\boldsymbol{x}_t = (\boldsymbol{f}_t', \boldsymbol{m}_t')'.$$

Suppose that these factors evolve according to a first-order vector-autoregression

$$\boldsymbol{x}_{t+1} = \boldsymbol{G}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_{t+1} \tag{2.4}$$

where $\boldsymbol{v}_{t+1} \sim \mathcal{N}(0_{k\times 1}, \boldsymbol{I}_k)$ and $\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{L}\boldsymbol{v}_{t+1} \sim \mathcal{N}(0_{k\times 1}, \Omega = \boldsymbol{L}\boldsymbol{L}')$. The model of Duffie (2002), Dai and Singleton (2000), and Ang and Piazzesi (2003) assumes that

$$M_{t+1} = \exp\left(-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{v}_{t+1}\right) : 1 \times 1 \tag{2.5}$$

where $\boldsymbol{\lambda}_t : (k+m) \times 1$ is the market price of risk, which along with r_t, is an affine function of the factors:

$$r_t = \delta + \boldsymbol{\beta}'\boldsymbol{x}_t : 1 \times 1 \tag{2.6}$$

and

$$\boldsymbol{\lambda}_t = \boldsymbol{\lambda} + \Phi\boldsymbol{x}_t : k \times 1. \tag{2.7}$$

Substituting into the no-arbitrage condition we get

$$\begin{aligned} P_t(\tau) \;=\; & \int_{R^k} \exp\left(-r_t - \frac{1}{2}\boldsymbol{\lambda}_t'\boldsymbol{\lambda}_t - \boldsymbol{\lambda}_t'\boldsymbol{v}_{t+1}\right) P_{t+1}(\tau - 1) \\ & \times \mathcal{N}(\boldsymbol{x}_{t+1} | \boldsymbol{G}\boldsymbol{x}_t, \Omega) d\boldsymbol{x}_{t+1} \end{aligned} \tag{2.8}$$

which by simple manipulations can be expressed as

$$P_t(\tau) = e^{-r_t}\mathbb{E}_t^Q[P_{t+1}(\tau - 1)] \tag{2.9}$$

where Q denotes the risk-neutral measure

$$\mathcal{N}(\boldsymbol{x}_{t+1}| - \boldsymbol{L}\boldsymbol{\lambda} + (\boldsymbol{G} - \boldsymbol{L}\Phi)\boldsymbol{x}_t, \Omega)d\boldsymbol{x}_{t+1}. \tag{2.10}$$

We can now obtain a recursive solution for bond-prices by assuming that bond prices take the exponential affine form

$$P_t(\tau) = \mathsf{exp}(-a(\tau) - \boldsymbol{b}'(\tau)\boldsymbol{x}_t) \tag{2.11}$$

where $a(\tau)$ is a scalar and $\boldsymbol{b}(\tau)$ is a $(k+m)\times 1$ vector. Then calculating $e^{-r_t}\mathbb{E}_t^Q[P_{t+1}(\tau-1)]$ and equating coefficients shows that

$$\begin{aligned} a(\tau) &= \delta + a(\tau-1) - \frac{1}{2}\boldsymbol{b}'(\tau-1)\Omega\boldsymbol{b}(\tau-1) - \boldsymbol{b}'(\tau-1)\boldsymbol{L}\boldsymbol{\lambda} \\ \boldsymbol{b}(\tau) &= \boldsymbol{\beta} + \boldsymbol{G}^{Q\prime}\boldsymbol{b}(\tau-1) \end{aligned} \tag{2.12}$$

where $\boldsymbol{G}^Q = \boldsymbol{G} - \boldsymbol{L}\Phi$, $a(0) = 0$, and $\boldsymbol{b}(0) = 0_{(k+m)\times 1}$.

Consequently, the term structure of interest rates as a function of the deep parameters in the SDF and the factor process can be expressed as:

$$R_t(\tau) \approx -\frac{1}{\tau}\log P_t(\tau) = \frac{a(\tau)}{\tau} + \frac{\boldsymbol{b}'(\tau)}{\tau}\boldsymbol{x}_t \tag{2.13}$$

for any non-negative integer τ.

2.2 Term Premium

One key quantity of interest in this context is the term premium. The term premium for a τ-period bond at time t is, by definition, the difference between the yield and the yield under the expectation hypothesis (EH). Let

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

denote the one-period expected excess return for holding the τ-period bond. Then, as shown by Cochrane and Piazzesi (2008), the term premium for the τ-period bond is given by

$$R_t(\tau) - \underbrace{\frac{1}{\tau}\sum_{i=0}^{\tau-1}\mathbb{E}_t[r_{t+i}]}_{\text{EH}} = \frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathrm{exr}_t^{(\tau+1-i)} \tag{2.14}$$

This is easily calculated under the affine model and its value studied for each t and τ, because

$$\begin{aligned} \text{exr}_t^{(\tau)} &= [\mathbb{E}_t [\ln P_{t+1}(\tau-1)] - \ln P_t(\tau)] - (-\ln P_t(1)) \\ &= \mathbb{E}_t [-(\tau-1) R_{t+1}(\tau-1) + \tau R_t(\tau)] - r_t \\ &= a(\tau) - a(\tau-1) + [\boldsymbol{b}'(\tau) - \boldsymbol{b}'(\tau-1)\boldsymbol{G}] \boldsymbol{x}_t - r_t \\ &= -\boldsymbol{b}'(\tau-1)\boldsymbol{L}\boldsymbol{\lambda}_t - \frac{1}{2}\boldsymbol{b}'(\tau-1)\Omega\boldsymbol{b}(\tau-1) \end{aligned} \tag{2.15}$$

3 Identifying Restrictions

3.1 Restrictions on $\boldsymbol{G}^Q$ and $\boldsymbol{\beta}$

The set of parameters in the bond prices is $\{\boldsymbol{G}, \boldsymbol{G}^Q, \delta, \boldsymbol{\beta}, \Omega, \boldsymbol{\lambda}\}$. As discussed comprehensively in Hamilton and Wu (2012), the parameters in affine models must be restricted for identification reasons. Here we focus on the arbitrage-free Nelson-Siegel restrictions developed by Christensen, Diebold, and Rudebusch (2009), Christensen et al. (2011), and Niu and Zeng (2012), which we have found are particularly useful in the Bayesian context. Under these restrictions, the no-arbitrage affine model is effectively similar (at least in terms of the evolution of factors) to theDiebold and Li (2006) dynamic Nelson-Siegel model with three latent factors. These factors can be interpreted as the time-varying level, slope, and curvature factors because they are close to their empirical counterpart proxies.

The restrictions are easily described. First, note that the identification of the factors requires restrictions on the factor loadings, and the factor loadings are determined by the factor transition matrix $\boldsymbol{G}^Q = \boldsymbol{G} - \boldsymbol{L}\boldsymbol{\Phi}$ under the risk-neutral measure and by the factor coefficients $\boldsymbol{\beta}$ in the short rate equation. The restrictions, therefore, involve $\boldsymbol{G}^Q$ and $\boldsymbol{\beta}$. In particular, in a model with three latent and two observed factors, the following restrictions are imposed.

Restriction 1 The matrix $\boldsymbol{G}^Q$ is of the type

$$\boldsymbol{G}^Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \exp(-\kappa) & \kappa\exp(-\kappa) & 0 & 0 \\ 0 & 0 & \exp(-\kappa) & 0 & 0 \\ G_{41} & G_{42} & G_{43} & G_{44} & G_{45} \\ G_{51} & G_{52} & G_{53} & G_{54} & G_{55} \end{bmatrix} \tag{3.1}$$

Restriction 2 The vector $\boldsymbol{\beta}$ is of the type

$$\boldsymbol{\beta} = (1, 1, 0, 0_{1\times m})'.$$

Then, as proved by Niu and Zeng (2012), $\boldsymbol{b}'(\tau)$ in equation (2.12) reduces to

$$\begin{aligned} \boldsymbol{b}'(\tau) &= \boldsymbol{\beta}' \times \sum_{j=0}^{\tau-1} \left(\boldsymbol{G}^Q\right)^j \\ &= \left[\ \tau,\ \ \tau - \tfrac{1-\exp(-\kappa\tau)}{\kappa},\ \ \tfrac{1-\exp(-\kappa\tau)}{\kappa} - \tau\exp(-\kappa\tau),\ \ 0_{1\times m}\ \right] \end{aligned} \tag{3.2}$$

which implies that $\boldsymbol{b}'(\tau)/\tau$ in equation (2.13) is given by

$$\left[\ 1,\ \ 1 - \tfrac{1-\exp(-\kappa\tau)}{\kappa\tau},\ \ \tfrac{1-\exp(-\kappa\tau)}{\kappa\tau} - \exp(-\kappa\tau),\ \ 0_{1\times m}\ \right]$$

which is exactly the form of the dynamic Nelson-Siegel factor loadings. Note that the last m columns of $\boldsymbol{b}'(\tau)/\tau$ are zero, which means that the observed factors are allowed to affect the yield to maturity only indirectly through the latent factors. The observed factors cannot influence the shape of the yield curve directly. Hence, it becomes possible to explicitly identify the three latent factors as the level, slope, and curvature effects of the yield curve.

Given that the restrictions are imposed on $\boldsymbol{G}^Q$ explicitly, which has one new parameter κ, it is natural to estimate $(\boldsymbol{G}, \kappa, \boldsymbol{L})$ rather than $(\boldsymbol{G},\ \boldsymbol{L},\ \Phi)$. Because $\boldsymbol{G}^Q$ is restricted, but $\boldsymbol{G}$ and $\boldsymbol{L}$ are unrestricted, Φ is restricted. In addition, as we estimate $(\boldsymbol{G}, \kappa, \boldsymbol{L})$, Φ can be obtained as $\boldsymbol{L}^{-1}(\boldsymbol{G} - \boldsymbol{G}^Q)$.

3.2 Restrictions on $(\boldsymbol{G},\ \boldsymbol{\lambda},\ \Omega,\ \delta)$

Because the shape of the yield curve is determined by the latent factors, it is natural to assume that the average market prices of observed factor risk are zero. Therefore, one can suppose that the last m elements of $\boldsymbol{\lambda}$ are zero.

Restriction 3 The vector $\boldsymbol{\lambda}$ is of the type

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, 0_{1\times m})'.$$

It is convenient to express the variance-covariance matrix of the factor shocks as

$$\Omega = \boldsymbol{V}\Gamma\boldsymbol{V}$$

where $\boldsymbol{V} = \mathsf{diag}(v_1, v_2, .., v_{k+m})$ is the diagonal factor shock volatility matrix and Γ is the factor shock correlation matrix. Suppose that $\Gamma_{i,,j}$ is the (i, j)th element of Γ. The factor volatilities are positive and the conditional correlations of the factors are constrained to be in the interval (-1, 1). In addition, to ensure that the factor process is stationarity, the eigenvalues of $\boldsymbol{G}$ are constrained to lie inside the unit circle. These restrictions can be summarized as follows:

Restriction 4 Support and stationarity restrictions

$$v_i > 0,\ \Gamma_{i,j} \in (-1, 1),\ |\boldsymbol{G}| < 1 \text{ for } i \neq j,\ i, j = 1, 2, .., (k + m)$$

Finally, we fix δ, the intercept term in the short rate dynamics, at the sample mean of the short rate as in the work of Dai, Singleton, and Yang (2007). This is because δ captures the mean of the short rate and it tends to be estimated inefficiently because of the high persistence of the short rate. As a result, the set of parameters of the bond pricing model are

$$\boldsymbol{\theta} = \{\boldsymbol{G},\ \kappa,\ \boldsymbol{V},\ \Gamma,\ \boldsymbol{\lambda}\}$$

of dimension $(k + m)^2$, 1, $(k + m)$, $(k + m) \times (k + m - 1)/2$, and k, respectively.

4 Prior-Posterior Analysis

4.1 State Space Representation

We now express the affine model in the form that is amenable to estimation. Suppose an individual has bonds with N different maturities at time t, whose yields are assembled as

$$\boldsymbol{R}_t = \begin{pmatrix} R_t(\tau_1) & R_t(\tau_2) & \cdots & R_t(\tau_N) \end{pmatrix}'. \tag{4.1}$$

Let the observed yields and factors be stacked as $\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{R}_t' & \boldsymbol{m}_t' \end{pmatrix}' : (N+m) \times 1$. Then, under the assumption that the N yields are measured with independent Gaussian errors and small variances $(\sigma_1^2, \sigma_2^2, .., \sigma_N^2)$, the affine model can be written in state space form.

The measurement equation of the state space model is given by the equations for the bond yields and observed factors:

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

where $\boldsymbol{b}_{1:k}(\tau_i)$ is the first k rows of $\boldsymbol{b}(\tau_i)$. Note that the last m rows of $\boldsymbol{b}(\tau_i)$ are zero. These equations imply that

$$\text{measurement equation: } \boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{a}, \boldsymbol{B}, \Sigma \sim \mathcal{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_t, \Sigma) \tag{4.2}$$

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,

$$\begin{aligned} \boldsymbol{a} &= \begin{pmatrix} \bar{\boldsymbol{a}}' & 0_{1 \times m} \end{pmatrix}' : (N+m) \times 1, \\ \bar{\boldsymbol{a}} &= \begin{pmatrix} \frac{a(\tau_1)}{\tau_1} & \frac{a(\tau_2)}{\tau_2} & \dots & \frac{a(\tau_N)}{\tau_N} \end{pmatrix}' : N \times 1, \\ \bar{\boldsymbol{B}} &= \begin{pmatrix} \frac{\boldsymbol{b}(\tau_1)}{\tau_1} & \frac{\boldsymbol{b}(\tau_2)}{\tau_2} & \dots & \frac{\boldsymbol{b}(\tau_N)}{\tau_N} \end{pmatrix}' : N \times (k+m), \\ \bar{\boldsymbol{B}}_{1:k} &= \text{the first } k \text{ colums of } \bar{\boldsymbol{B}}, \\ \boldsymbol{B} &= \begin{pmatrix} \bar{\boldsymbol{B}} & \\ 0_{m \times k} & \boldsymbol{I}_m \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{B}}_{1:k} & 0_{N \times m} \\ 0_{m \times k} & \boldsymbol{I}_m \end{pmatrix} : (N+m) \times (k+m), \end{aligned}$$

and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.

The transition equation is simply the factor dynamics,

$$\text{transition equation: } \boldsymbol{x}_{t+1} | \boldsymbol{x}_t, \boldsymbol{G}, \Omega \sim \mathcal{N}(\boldsymbol{G}\boldsymbol{x}_t, \Omega = \boldsymbol{V}\Gamma\boldsymbol{V}) \tag{4.3}$$

4.2 Likelihood

Let $\boldsymbol{y} = \{\boldsymbol{y}_t\}_{t=1}^T$ be the time series of the observed yields and observable factors. The likelihood $p(\boldsymbol{y}|\boldsymbol{\theta}, \Sigma)$ is the density of the data $\boldsymbol{y}$ given the unknown parameters $(\boldsymbol{\theta}, \Sigma)$. By the law of total probability

$$p(\boldsymbol{y}|\theta, \Sigma) = \prod_{t=1}^{T} p(\boldsymbol{y}_t | \mathcal{F}_{t-1}, \theta, \Sigma) \tag{4.4}$$

$$= \prod_{t=1}^{T} \left[\int p(\boldsymbol{y}_t | \mathcal{F}_{t-1}, \boldsymbol{f}_t, \theta, \Sigma) p(\boldsymbol{f}_t | \mathcal{F}_{t-1}, \theta, \Sigma) d\boldsymbol{f}_t \right] \tag{4.5}$$

where $\mathcal{F}_t$ denotes the history of the observations up to time t. Then, each one-step-ahead predictive density, $p(\boldsymbol{y}_t|\mathcal{F}_{t-1}, \theta, \Sigma)$ is a marginalization of $p(\boldsymbol{y}_t|\mathcal{F}_{t-1}, \boldsymbol{f}_t, \theta, \Sigma)$ over the latent factors $\boldsymbol{f}_t$. This can be done by utilizing the Kalman filter, so that

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \Sigma) = \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{y}_t|\boldsymbol{y}_{t|t-1}, \boldsymbol{V}_{t|t-1})$$

where

$$\begin{aligned}
\mathbb{E}[\boldsymbol{x}_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}, \Sigma] &= \boldsymbol{x}_{t|t-1} = \boldsymbol{G}\boldsymbol{x}_{t-1|t-1}, \\
Var[\boldsymbol{x}_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}, \Sigma] &= Q_{t|t-1} = \boldsymbol{G}Q_{t-1|t-1}\boldsymbol{G}' + \Omega, \\
\mathbb{E}[\boldsymbol{y}_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}, \Sigma] &= \boldsymbol{y}_{t|t-1} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_{t|t-1}, \\
Var[\boldsymbol{y}_t|\mathcal{F}_{t-1}, \boldsymbol{\theta}, \Sigma] &= \boldsymbol{V}_{t|t-1} = \boldsymbol{B}Q_{t|t-1}\boldsymbol{B}' + \Sigma, \\
K_t &= Q_{t|t-1}\boldsymbol{B}'\boldsymbol{V}_{t|t-1}^{-1}, \\
\mathbb{E}[\boldsymbol{x}_t|\mathcal{F}_t, \boldsymbol{\theta}, \Sigma] &= \boldsymbol{x}_{t|t} = \boldsymbol{x}_{t|t-1} + K_t(\boldsymbol{y}_t - \boldsymbol{y}_{t|t-1}), \\
Var[\boldsymbol{x}_t|\mathcal{F}_t, \boldsymbol{\theta}, \Sigma] &= Q_{t|t} = Q_{t|t-1} - K_t\boldsymbol{B}Q_{t|t-1}.
\end{aligned}$$

The initial factors, $\boldsymbol{x}_0$ are integrated out over its ergodic distribution given the parameters:

$$\boldsymbol{x}_0|\mathcal{F}_0, \boldsymbol{\theta}, \Sigma \sim \mathcal{N}\left(\boldsymbol{x}_{0|0} = 0, Q_{0|0}\right)$$

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

$$vec(Q_{0|0}) = \left[I_{k^2} - (\boldsymbol{G} \otimes \boldsymbol{G})\right]^{-1} \times vec(\Omega).$$

4.3 Prior

We now turn to prior distribution of the parameters $\boldsymbol{G}$, κ, $\boldsymbol{V}$, Γ, $\boldsymbol{\lambda}$, and Σ. Following Chib and Ergashev (2009), the prior distribution is set up to reflect the assumption of an upward sloping yield curve. In addition, we suppose that the yields are persistent (Abbritti, Gil-Alana, Lovcha, and Moreno (2015)). We arrive at such a prior by sampling the prior. For given prior means, time series of factors are drawn from the vector-autoregressive process, followed by the yields. If, on average, the yield curve is not upward sloping or too steep, the hyperparameters of the prior are adjusted and the process is repeated until outcomes consistent with these beliefs are realized.

The second step is to specify the strength of our beliefs through an appropriate distribution. If the prior is too tight, then the prior is likely to be in conflict with the likelihood function, whereas if the prior is too weak then it is not likely to smooth the irregularities of the likelihood function, as the likelihood surface plots of Figure 2 demonstrate. Our solution to this problem is to depart from the tradition of a Gaussian prior and employ an independent Student-t distribution for the various parameters. With this choice we are able to incorporate the beliefs mentioned above and, because of the thickness of the prior tails, mitigate any potential conflicts with the data.

Suppose that $G_{i,,j}$ is the (i,j)th element of $\boldsymbol{G}$, and $\boldsymbol{\lambda}_i$ is the ith elements of $\boldsymbol{\lambda}$. We assume that the parameters in $\{\boldsymbol{G},\ \kappa,\ \lambda\}$ have Student-t prior distributions, and the parameters in $\{V,\ \Sigma\}$ have inverse-gamma prior distributions. The specific descriptions are in Tables 1 and 2. Finally, the factor shock correlations have a uniform prior over (-1,1),

$$\Gamma_{i,j} \sim Unif\,(-1,1)$$

for $i \neq j$, $i,j = 1,2,..,(k+m)$.

	Student-t distribution, $St(\mu,\sigma^2,v)$		
parameter	mean (μ)	scale (σ^2)	d. f. (v)
κ	0.935	6×10^{-4}	15
λ_1	-0.170	10^{-4}	15
λ_2	-0.070	10^{-4}	15
λ_3	-0.024	10^{-4}	15
$G_{i,i}$	0.900	10^{-3}	15
$G_{i,j},\ i\neq j$	0.000	10^{-3}	15

Table 1: Prior distributions for κ, λ, and G $St(\mu,\sigma^2,v)$ is the Student-t distribution with the mean μ, scale σ^2 and degree of freedom (d.f.) v.

We now explain how the hyperparameters in our prior distributions are chosen specifically. Figure 1 plots the distribution of the prior-implied yield curve when the parameters are fixed at their prior mean. Figure 1(a) is the prior mean of the yield curve, which is the intercept term, $\bar{\boldsymbol{a}}$ 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(b) presents the 5%, 50%, and 95% quantiles of the ergodic prior-implied yield curve distributions generated from the ergodic distribu-

Inverse-Gamma distribution, $\mathcal{IG}(a,b)$		
parameter	a	b
$10^4 \times \boldsymbol{V}_{1,1}$	66	130
$10^4 \times \boldsymbol{V}_{2,2}$	102	252
$10^4 \times \boldsymbol{V}_{3,3}$	4	1
$10^4 \times \boldsymbol{V}_{4,4}$	146	435
$10^4 \times \boldsymbol{V}_{5,5}$	18	17
$5 \times 10^5 \times \sigma_i^2$	54	260

Table 2: Prior distributions for V and Σ $\mathcal{IG}(a,b)$ is the inverse-gamma distribution whose mean is $b/(a-1)$.

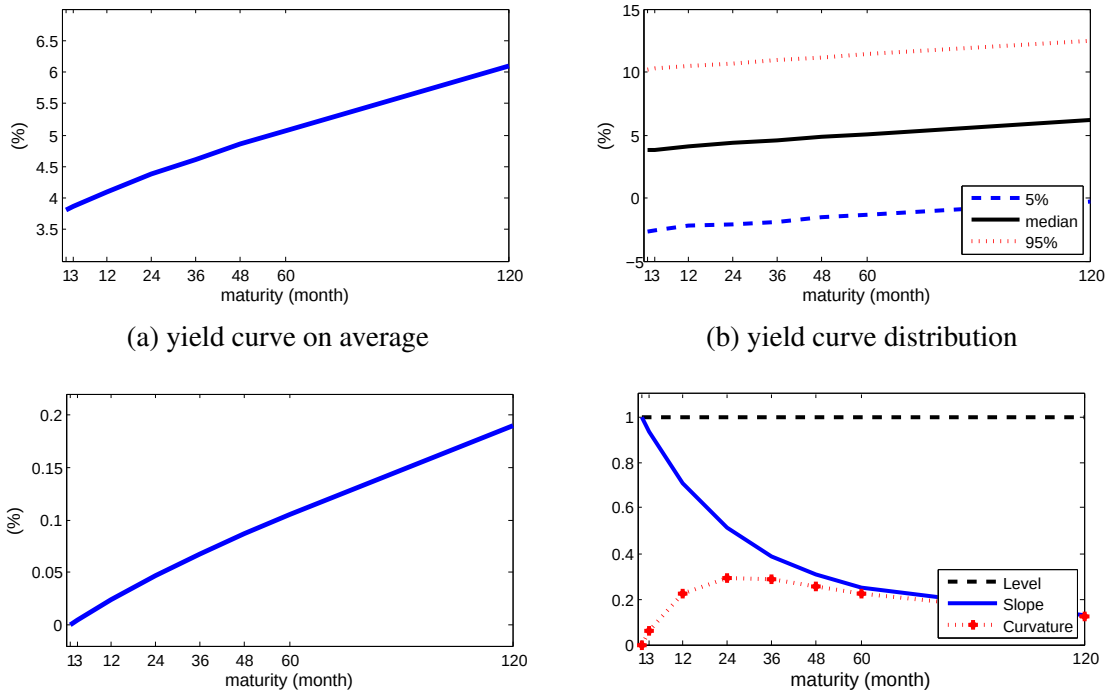
tion of the factors and pricing errors. The prior mean of $G'_{i,i}s$ is 0.9 as the interest rates are very persistent. Our prior means of the factor volatility $\boldsymbol{V}$ allows for yield curve variations between -3% and 11%, which is flexible enough to capture various shapes of yield curves over time.

The prior on the risk parameters λ is more difficult to quantify. 1(c) plots the term structure of the term premiums computed at the prior mean of the parameters based on equation (2.14). Our prior for λ is chosen such that the prior-implied term premium is increasing with the maturity, and it is not too big for long term bonds. Finally, 1(d) is the factor loadings $\bar{\boldsymbol{B}}$, the shape of which is solely determined by the value of κ. The prior factor loadings are the same as in Nelson-Siegel yield curve models, and a priori the factors are supposed to be identified as the level, slope, and curvature effects. The prior mean of κ is specified so that the curvature factor loading is maximized at 24 months. The next subsection demonstrates the role of our prior in improving the efficacy of posterior sampling.

Assuming independence across parameters, our prior density $\pi(\boldsymbol{\theta}, \Sigma)$ is given by the product of individual prior densities with a zero density outside the support $\mathcal{R}$ defined by the restrictions on the parameters given above.

4.4 Posterior Sampling

This subsection discusses our posterior MCMC sampling procedure. The target distribution to be simulated is the joint posterior distribution of the parameters, factors, and



(a) yield curve on average

(b) yield curve distribution

(c) term structure of risk premiums

(d) factorloadings

Figure 1: The prior-implied yield curve when the parameters are fixed at the prior mean: $\mathcal{AFNS}$(3,0)

predictive yield curves

$$\boldsymbol{\theta}, \Sigma, \{\boldsymbol{f}_t\}_{t=1}^{T}, \{\boldsymbol{R}_{T+h}\}_{h=1}^{H} | \boldsymbol{y}.$$

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 $\boldsymbol{\theta}$, $\{\boldsymbol{f}_t\}_{t=1}^{T}$, Σ, and $\{\boldsymbol{R}_{T+h}\}_{h=1}^{H}$, as outlined in the following algorithm.

Algorithm: MCMC sampling

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

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

Step 1.(a): Randomly determine the number of blocks and their components in $\boldsymbol{\theta}$,

$$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, .., \boldsymbol{\theta}_{B_g}$$

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

Step 1.(b): Maximize

$$\ln\{p(\boldsymbol{y}|\boldsymbol{\theta}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)}) \times \pi(\boldsymbol{\theta}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\}$$

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

Step 1.(c): Draw a proposal for $\boldsymbol{\theta}_l$, denoted by $\boldsymbol{\theta}_l^{\dagger}$, from the selected multivariate Student-t distribution,

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

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

$$\begin{aligned} \boldsymbol{\theta}_l &= \boldsymbol{\theta}_l^{\dagger} && \text{if } u < \alpha\left(\boldsymbol{\theta}_l^{(g-1)}, \boldsymbol{\theta}_l^{\dagger}|\boldsymbol{y}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)}\right) \\ \boldsymbol{\theta}_l &= \boldsymbol{\theta}_l^{(g-1)} && \text{if } u \geq \alpha\left(\boldsymbol{\theta}_l^{(g-1)}, \boldsymbol{\theta}_l^{\dagger}|\boldsymbol{y}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)}\right) \end{aligned}$$

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

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

Step 3: Draw $\Sigma^{(g)}$ from $\Sigma|\boldsymbol{y}, \{\boldsymbol{f}_t^{(g)}\}_{t=1}^T, \boldsymbol{\theta}^{(g)}$.

Step 4: Generate $\boldsymbol{x}_{T+h}^{(g)}$ from $\boldsymbol{x}_{T+h}|\psi^{(g)}, \boldsymbol{x}_{T+h-1}^{(g)}$ for $h = 1, 2, .., H$ based on the equation (4.3).

Step 5: Generate $\boldsymbol{R}_{T+h}^{(g)}$ from $\boldsymbol{R}_{T+h}|\psi^{(g)}, \boldsymbol{x}_{T+h}^{(g)}$ for $h = 1, 2, .., H$ based on the equation (4.2).

Step 6: $g = g + 1$, 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.

Full details of each stage are as follows.

4.4.1 $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ 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

$$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_{B_g}$$

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 (2.12), it is difficult to form appropriate fixed blocks a priori. A randomized 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 $\boldsymbol{\theta}$ form random blocks and Σ form a fixed block since the full conditional distribution of Σ is tractable. Once the number of blocks and their components are randomly determined, we update $\boldsymbol{\theta}$ 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 in a sequential manner. Consider the lth block of parameters $\boldsymbol{\theta}_l$. Let $\bar{\boldsymbol{\theta}}_l$ and $V_{\bar{\boldsymbol{\theta}}_l}$ denote the mode and the negative inverse Hessian, respectively, defined as

$$\bar{\boldsymbol{\theta}}_l = \arg\max_{\boldsymbol{\theta}_l} \ln\{p(\boldsymbol{y}|\boldsymbol{\theta}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\pi(\boldsymbol{\theta}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\} \text{ subject to } \{\boldsymbol{\theta}_l, \boldsymbol{\theta}_{-l}\} \in \mathcal{R}, \quad (4.6)$$

$$V_{\bar{\boldsymbol{\theta}}_l} = \left(-\frac{\partial^2 \ln\{p(\boldsymbol{y}|\bar{\boldsymbol{\theta}}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\pi(\bar{\boldsymbol{\theta}}_l, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\}}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_l'} \right)^{-1} \quad (4.7)$$

where $\boldsymbol{\theta}_{-l}$ is the parameters in $\boldsymbol{\theta}$ except $\boldsymbol{\theta}_l$. 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 a Student-t distribution, rather than the normal distribution, helps in generating proposal values that are more diverse and more distant because of the fat-tails of the distribution.

In order to find the global mode, within the support $\mathcal{R}$, we utilize a simulated annealing algorithm in combination with the Newton-Raphson method. We refer to this optimizer as the SA-Newton method. Simulated annealing is a stochastic global optimizer that is particularly useful in our context. In finding this mode, we start the SA iterations with a high initial temperature and then use a relatively high cooling factor to reduce the temperature quickly. Specifically, the initial temperature parameter and the cooling factor are set to at 5 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 improve the local search of the modal value.

Block-wise Metropolis-Hastings algorithm Now we move to the third step. Using the mode and Hessian at the mode, we construct a proposal distribution

$$q\left(\boldsymbol{\theta}_l|\boldsymbol{\theta}_{-l}\right) = St\left(\boldsymbol{\theta}_l|\bar{\boldsymbol{\theta}}_l, V_{\bar{\boldsymbol{\theta}}_l}, 15\right).$$

where the St is a multivariate Student-t density with $\nu = 15$ degrees of freedom. Let $\boldsymbol{\theta}_l^{\dagger}$ drawn from

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

denote the proposal value. Then, this proposal value is accepted with the MH probability of the move given by

$$\alpha\left(\boldsymbol{\theta}_l^{(g-1)}, \boldsymbol{\theta}_l^{\dagger}|\boldsymbol{y}, \boldsymbol{\theta}_{-l}\right) = \min\left\{1, \frac{p(\boldsymbol{y}|\boldsymbol{\theta}_l^{\dagger}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\pi(\boldsymbol{\theta}_l^{\dagger}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})}{p(\boldsymbol{y}|\boldsymbol{\theta}_l^{(g-1)}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})\pi(\boldsymbol{\theta}_l^{(g-1)}, \boldsymbol{\theta}_{-l}, \Sigma^{(g-1)})} \frac{q\left(\boldsymbol{\theta}_l^{(g-1)}|\boldsymbol{\theta}_{-l}\right)}{q\left(\boldsymbol{\theta}_l^{\dagger}|\boldsymbol{\theta}_{-l}\right)}\right\} \tag{4.8}$$

where $\boldsymbol{\theta}_l^{(g-1)}$ is the current value. If the proposal is accepted, then $\boldsymbol{\theta}_l^{(g)} = \boldsymbol{\theta}_l^{\dagger}$. Otherwise, $\boldsymbol{\theta}_l^{(g)} = \boldsymbol{\theta}_l^{(g-1)}$. Note that if the proposal value does not satisfy the restriction $\mathcal{R}$, then it is immediately rejected. The remaining blocks of $\boldsymbol{\theta}$ are updated in the same manner.

4.4.2 Factor ($\{\boldsymbol{f}_t\}_{t=1}^T$) sampling

Next, we sample the factors given the data and parameters based on the multi-move method, in which the factors are updated in one block.

Algorithm: Factor sampling

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

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

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

where

$$\begin{aligned}
\boldsymbol{x}_{t|\boldsymbol{x}_{t+1}} &= \boldsymbol{x}_{t|t} + Q_{t|t}\boldsymbol{G}'\left(Q_{t+1|t}\right)^{-1}\left(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1|t}\right), \\
Q_{t|\boldsymbol{x}_{t+1}} &= Q_{t|t} - Q_{t|t}\boldsymbol{G}'\left(P_{t+1|t}\right)^{-1}\boldsymbol{G}Q_{t|t}, \\
\boldsymbol{x}_{t+1|t} &= \mu + \boldsymbol{G}\boldsymbol{x}_{t|t}, \quad Q_{t+1|t} = \boldsymbol{G}Q_{t|t}\boldsymbol{G}' + \Omega,
\end{aligned} \tag{4.9}$$

The first k element of $\boldsymbol{x}_t$ is the vector of latent factors ($\boldsymbol{f}_t$), and the last m elements are the observed factors at time t.

4.4.3 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,

$$\sigma^2_{\tau_i}|\boldsymbol{y}, \{\boldsymbol{x}_t\}_{t=1}^T, \boldsymbol{\theta} \sim \mathcal{IG}(v_1/2, \delta_{1,i}/2)$$

where

$$\begin{aligned}
v_1 &= v_0 + T, \\
e_t(\tau_i) &= R_t(\tau_i) - a(\tau_i) - b(\tau_i) \times \boldsymbol{x}_t, \\
\text{and } \delta_{1,i} &= \sum_{t=1}^{T} e_t(\tau_i)^2 + \delta_0
\end{aligned}$$

for $i = 1, 2, .., N$.

4.4.4 Predictive density simulation

The final stage is to sample predictive yield curves from

$$\boldsymbol{R}_{T+h}|\boldsymbol{\theta}, \Sigma, \boldsymbol{y}.$$

This can be done by the method of composition:

$$\begin{aligned} p\left(\boldsymbol{R}_{T+h}|\boldsymbol{\theta}, \Sigma, \boldsymbol{y}\right) &= \int p\left(\boldsymbol{R}_{T+h}, \boldsymbol{x}_{T+h}|\boldsymbol{\theta}, \Sigma, \boldsymbol{y}\right) d\boldsymbol{x}_{T+h} \\ &= \int p\left(\boldsymbol{R}_{T+h}|\boldsymbol{\theta}, \Sigma, \boldsymbol{y}, \boldsymbol{x}_{T+h}\right) p\left(\boldsymbol{x}_{T+h}|\boldsymbol{\theta}, \Sigma, \boldsymbol{y}\right) d\boldsymbol{x}_{T+h}. \end{aligned}$$

For sampling predictive yield curves, the predictive factors ($\boldsymbol{x}_{T+h}$) are first simulated from

$$\mathcal{N}(\boldsymbol{G}\boldsymbol{x}_{T+h-1}, \Omega)$$

for $h = 1, 2, .., H$. Next, given $\{\boldsymbol{x}_{T+h}\}_{h=1}^{H}$, $\boldsymbol{y}_{T+h}$ is drawn from

$$\mathcal{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}_{T+h}, \Sigma),$$

and $\boldsymbol{R}_{T+h}$ is the first N elements of $\boldsymbol{y}_{T+h}$.

4.5 Inefficiency Factor

We compare the efficacy 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 the simulated parameters over the Markov chain. One can estimate the inefficiency factor as

$$1 + \frac{400}{199} \sum_{j=1}^{200} K(j/200)\hat{\rho}(j). \tag{4.10}$$

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, a smaller inefficiency factor indicates higher efficacy.

4.6 Posterior Predictive Density Accuracy Measure

In a Bayesian approach the accuracy of density forecasts is typically measured by the posterior predictive density (PPD). In particular, we concentrate on the PPD of the

yield curves, not the vector of yields and macro factors. This is because the PPD of the yield curves can be used for model comparison regardless of the number and choice of factors. Suppose that $\boldsymbol{R}^o_{T+1}$ denotes the realized yield curve at time $T+1$. Then, the PPD of the yield curves is defined as

$$\begin{aligned} p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}) &= \int p(\boldsymbol{R}^o_{T+1}, \boldsymbol{\theta}, \Sigma, \boldsymbol{x}|\boldsymbol{y}) d(\boldsymbol{\theta}, \Sigma, \boldsymbol{x}) \\ &= \int p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}, \boldsymbol{\theta}, \Sigma, \boldsymbol{x}) \boldsymbol{\pi}(\boldsymbol{\theta}, \Sigma, \boldsymbol{x}|\boldsymbol{y}) d(\boldsymbol{\theta}, \Sigma, \boldsymbol{x}), \end{aligned}$$

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 $p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}, \boldsymbol{\theta}, \Sigma, \boldsymbol{x})$ over the draws of $(\boldsymbol{\theta}, \Sigma, \boldsymbol{x})$ from $(\boldsymbol{\theta}, \Sigma, \boldsymbol{x})|\boldsymbol{y}$

$$p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}) \approx \sum_{g=1}^{n_1} p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}, \boldsymbol{\theta}^{(g)}, \Sigma^{(g)}, \boldsymbol{x}^{(g)}).$$

$(\boldsymbol{\theta}^{(g)}, \Sigma^{(g)}, \boldsymbol{x}^{(g)})$ are the posterior draws from the gth MCMC cycles, $\bar{\boldsymbol{f}}^{(g)}_{T+1}$ is the first k elements of $\bar{\boldsymbol{x}}^{(g)}_{T+1} = \boldsymbol{G}^{(g)}\boldsymbol{x}^{(g)}_T$, $\Omega^{(g)}_{1:k,1:k}$ 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

$$p(\boldsymbol{R}^o_{T+1}|\boldsymbol{y}, \boldsymbol{\theta}^{(g)}, \Sigma^{(g)}, \boldsymbol{x}^{(g)}) \equiv \mathcal{N}\left(\boldsymbol{R}^o_{T+1}|\bar{\boldsymbol{a}}^{(g)} + \bar{\boldsymbol{B}}^{(g)}\bar{\boldsymbol{f}}^{(g)}_{T+1}, \bar{\boldsymbol{B}}^{(g)}\Omega^{(g)}_{1:k,1:k}\bar{\boldsymbol{B}}^{(g)\prime} + \Sigma^{(g)}_{1:N,1:N}\right).$$

5 Examples

The objective of our work is to propose an efficient Bayesian MCMC simulation scheme and Matlab toolbox for Gaussian affine term structure models.[1] In much of the affine term structure model literature, most of the interest is on the model parameters, factors, and model-implied key quantities such as term premium and predictive yield curves. Our Matlab toolbox produces all these quantities.

Let the affine model with k latent factors and m macro factors denoted by $\mathcal{AFNS}(k, m)$. In this section we estimate the models $\mathcal{AFNS}(3, 0)$ and $\mathcal{AFNS}(3, 2)$ as examples and demonstrate the sampling performance of our toolbox with the TaRB-MH method for simulating the key quantities in comparison with the tailored fixed-blocking MH (TaFB-MH) method. In the fixed-blocking scheme the parameters in $\{\boldsymbol{G}, \kappa, \boldsymbol{V}, \Gamma, \boldsymbol{\lambda}, \Sigma\}$ are

[1] Our Matlab toolbox is available upon request from the authors.

grouped in separate blocks. This fixed block version of the tailored multiple-block MH method was suggested by Chib and Ergashev (2009) for Gaussian affine term structure model estimation. We use yields of constant maturity treasury securities computed by the U.S. Treasury. The time span ranges from January 1971 to December 2007. The yield data consists of eight time series comprising the short rate (approximated by the one month yield) and the yields of the following maturities: 3, 12, 24, 36, 48, 60, and 120 months. For the estimation of the model $\mathcal{AFNS}(3,2)$, the observed factors are the U.S. CPI inflation and federal funds rate with the same sample period as the yields. For convenience, the observed factors are demeaned.

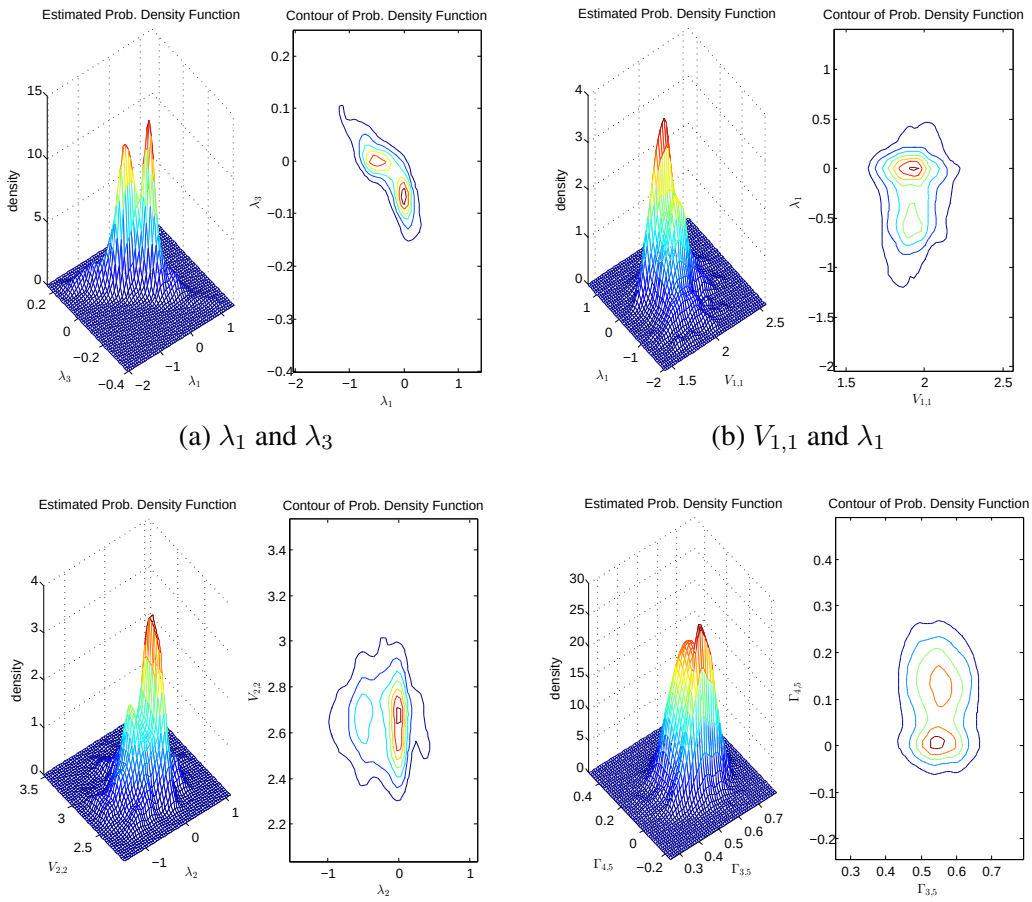
It is interesting to examine this likelihood surface to get a sense of why inference for affine term structure models is difficult. To do this, we simulate the likelihood of $\mathcal{AFNS}(3,2)$ model as an example and plot bivariate likelihood surfaces for some selected pairs of the parameters in Figure 2.

Figure 2 clearly reveals that the likelihood surface has multiple modes, which makes maximum likelihood estimation based on a numerical optimization unreliable. The reason is the severe nonlinearity of the model with respect to the parameters. One of the advantages of the Bayesian approach is that it is possible to mitigate the multi-modality problem by specifying a proper prior and smoothing out the posterior surface.

Figure 3 plots the posterior surface for the same pairs of the parameters in Figure 2. This figure demonstrates the role of our prior in the posterior simulation. Evidently, the multi-modality problem becomes less serious, which helps us estimate the models in a more efficient and reliable way. This is because the prior densities allocate little weight to the infeasible parameter region against our prior beliefs. In addition, the parameters are highly correlated, which makes estimating GATSMs challenging. This is the reason why simple MCMC methods such as the random-walk MH algorithm and Hamiltonian Monte Carlo do not mix very well.

5.1 Computational burden

The two MCMC posterior sampling schemes for $\mathcal{AFNS}(3,0)$ and $\mathcal{AFNS}(3,2)$ were coded in Matlab R2014a and executed on a Windows 7 64-bit machine with a 3.30 GHz Intel Core(TM) CPU. Each code was run for 11,000 iterations, and the first 1,000



(a) λ_1 and λ_3 (b) $V_{1,1}$ and λ_1

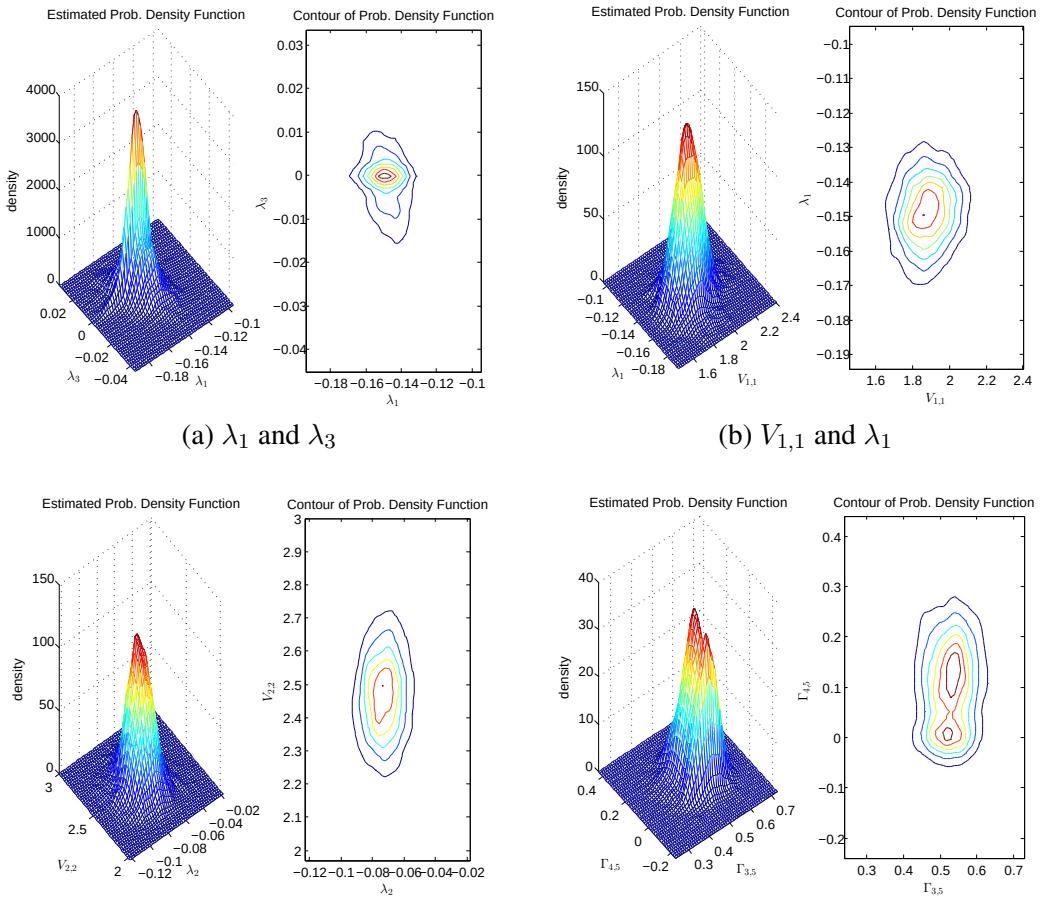
(c) λ_2 and $V_{2,2}$ (d) $\Gamma_{3,5}$ and $\Gamma_{4,5}$

Figure 2: Likelihood surface: $\mathcal{AFNS}$(3,2)

posterior draws were discarded as burn-in. The TaRB-MH algorithm for $\mathcal{AFNS}(3,0)$ and $\mathcal{AFNS}(3,2)$ took approximately 3.5 and 7.5 hours, respectively, to generate 11,000 draws. The TaRB-MH algorithm required approximately 30 minutes longer than the TaFB-MH algorithm. The computing time cost from using the randomizing-blocking scheme rather than the fixed-blocking scheme is negligible considering the substantial efficacy gains which are shown in the next subsections.

5.2 Parameters

To find out whether the Markov chains based on TaRB-MH and TaFB-MH converge, we compute the log posterior densities over the MCMC draws. Figure 4 plots the results for the first 200 MCMC iterations for $\mathcal{AFNS}(3,0)$ and $\mathcal{AFNS}(3,2)$. Obviously, both



(a) λ_1 and λ_3 (b) $V_{1,1}$ and λ_1

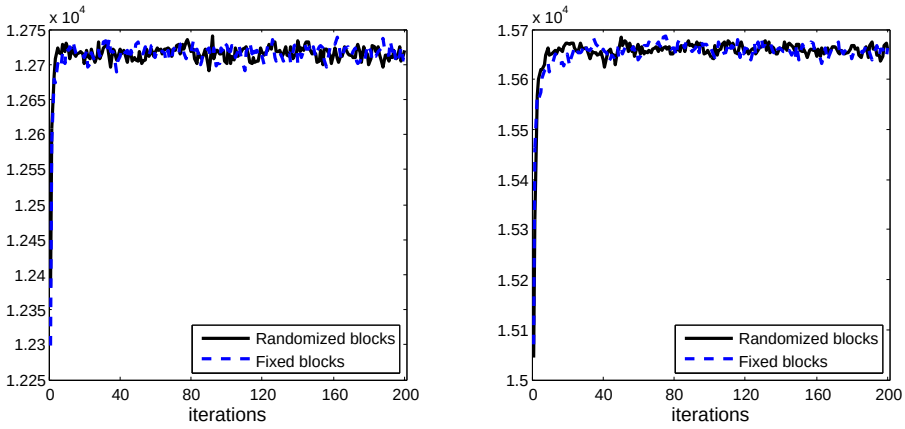
(c) λ_2 and $V_{2,2}$ (d) $\Gamma_{3,5}$ and $\Gamma_{4,5}$

Figure 3: Posterior surface: $\mathcal{AFNS}$(3,2)

of the Markov chains converge to the target distribution before the 100th iteration.

Tables 3 and 4 present the summary of the posterior distribution of the model parameters and their inefficiency factors. The burn-in size is 1,000, and the MCMC size beyond the burn-in is set to be 10,000. The output of the TaRB-MH reveals the smaller serial dependence compared to the TaFB-MH. This shows that randomizing blocks can improve the efficacy of the MCMC posterior sampling.

Note that the risk parameters are difficult to estimate precisely. As shown in equation (2.12), $\boldsymbol{\lambda}$ appears in the intercept term, not in the factor loadings. Therefore, the information that is necessary for estimating $\boldsymbol{\lambda}$ is from the cross-sectional information of the yield curve such as the slope or curvature. The high persistence of the slope and curvature effects makes estimating the risk parameters difficult. This is why in the



(a) $\mathcal{AFNS}$(3,0) (b) $\mathcal{AFNS}$(3,2)

Figure 4: Log posterior densities over MCMC iterations This figures plot the log posterior densities (=log likelihood density + log prior density) over the MCMC iterations for $\mathcal{AFNS}(3,0)$ and $\mathcal{AFNS}(3,2)$. The solid line and dotted lines are the densities at the TaRB-MH draws and TaFB-MH draws, respectively.

previous works of Moench (2008) and Duffee (2012) the risk parameters are estimated with large standard errors and in our empirical works these parameters are simulated relatively less efficiently compared to the other parameters regardless of the sampling method. Bauer (2014) resolves this problem by restricting most market price of risk parameters to zero.

We would like to emphasize that when the target density is high-dimensional and irregular, there are four key characteristics that can make our posterior MCMC sampler more efficient: informative prior, randomizing blocks, Student-t proposal distribution, and the use of block-wise stochastic optimization. Our work shows that without such carefully tuned implementation of the MH algorithm, it is not easy to achieve efficient sampling. Meanwhile, the Markov chain based on a random-walk MH or Hamiltonian Monte Carlo method tends to fail to escape from a local region or cannot explore local regions, which causes either high inefficiency or failure of convergence.

5.3 Factors

Figures 5 and 6 plot the time series of the level, slope, and curvature factors along with 95% credibility intervals over time. The factors are precisely estimated and identified as the dynamics of the level, slope, and curvature factors are similar to those of the

Parameter	TaRB-MH					TaFB-MH	
	mean	2.5%	97.5%	ineff.	accept. rate	ineff.	accept. rate
$G_{1,1}$	0.98	0.96	0.98	4.21	36.63	8.71	28.95
$G_{2,2}$	0.97	0.96	0.98	5.01	46.65	10.68	28.95
$G_{3,3}$	0.92	0.88	0.96	7.09	58.55	22.72	28.95
κ	0.07	0.06	0.07	6.78	58.66	3.54	93.22
λ_1	-0.15	-0.17	-0.13	14.15	57.88	7.99	80.96
λ_2	-0.07	-0.09	-0.05	14.28	57.57	5.46	80.96
λ_3	0.00	-0.02	0.02	15.01	52.65	6.30	80.96
$V_{1,1}$	1.84	1.61	2.07	15.56	58.01	21.54	92.15
$V_{2,2}$	2.58	2.30	2.88	12.76	58.64	6.21	92.15
$V_{3,3}$	5.12	4.66	5.57	6.21	58.55	3.63	92.15
$\Gamma_{1,2}$	-0.67	-0.76	-0.57	13.73	57.32	15.23	83.15
$\Gamma_{1,3}$	0.36	0.20	0.51	8.79	57.79	3.71	83.15
$\Gamma_{2,3}$	-0.03	-0.21	0.09	10.76	54.63	8.07	83.15

Table 3: Parameters: $\mathcal{AFNS}(3,0)$ Inefficiency comparison with the fixed blocking scheme (TaFB-MH)

observed proxies (i.e., the long rate, (short rate - long rate), (short rate + long rate - 2×midterm rate), respectively.

5.4 Term Structure of Term Premiums

Figure 7 plots the posterior mean of the term premium dynamics across maturities over time, which are simulated from the three- and five-factor models. We find that the term premiums are more volatile for longer maturities. The term premium of the 10-year bond tends to be negative during recessions because of the flight-to-safety effect.

5.5 Posterior Predictive Distributions

In constructing a bond portfolio, the posterior predictive density simulation of yield curves is essential because the expected returns and covariance structure of the future yields are critical for decision-making. Figures 8 and 9 provide plots of the one-month-ahead and six-month-ahead predictive yield curves, respectively. Not surprisingly, the six-month-ahead predictive yields have greater variance than the one-month-ahead yields because the predictive yields at a longer horizon involve greater uncertainty.

Tables 5 and 6 present the predictive correlations among the yields, and indicate

Parameter	TaRB-MH					TaFB-MH	
	mean	2.5%	97.5%	ineff.	accept. rate	ineff.	accept. rate
$G_{1,1}$	0.97	0.96	0.98	7.84	46.43	12.89	35.25
$G_{2,2}$	0.97	0.96	0.98	7.29	56.08	14.95	35.25
$G_{3,3}$	0.92	0.88	0.95	10.03	56.68	24.33	35.25
$G_{4,4}$	0.92	0.87	0.96	5.08	56.99	14.94	35.25
$G_{5,5}$	0.98	0.97	0.98	6.87	31.03	6.33	35.25
κ	0.07	0.06	0.07	6.90	57.62	5.00	93.35
λ_1	-0.15	-0.17	-0.13	20.84	56.72	5.91	82.41
λ_2	-0.07	-0.09	-0.05	20.78	56.44	6.59	82.41
λ_3	0.00	-0.02	0.01	17.37	47.23	7.86	82.41
$V_{1,1}$	1.86	1.66	2.08	14.20	57.40	29.23	93.37
$V_{2,2}$	2.46	2.22	2.71	14.52	57.11	14.81	93.37
$V_{3,3}$	5.14	4.70	5.58	6.20	57.68	3.90	93.37
$V_{4,4}$	2.86	2.60	3.14	4.85	57.66	3.59	93.37
$V_{5,5}$	1.55	1.39	1.71	7.23	57.60	12.05	93.37
$\Gamma_{1,2}$	-0.71	-0.78	-0.63	16.98	55.71	50.70	23.39
$\Gamma_{1,3}$	0.35	0.19	0.50	16.36	55.59	23.76	23.39
$\Gamma_{1,4}$	-0.05	-0.20	0.06	21.39	50.65	24.74	23.39
$\Gamma_{1,5}$	0.11	-0.03	0.28	27.13	52.51	31.04	23.39
$\Gamma_{2,3}$	-0.02	-0.15	0.10	16.85	49.15	23.21	23.39
$\Gamma_{2,4}$	-0.02	-0.17	0.10	7.76	48.36	15.17	23.39
$\Gamma_{2,5}$	-0.01	-0.12	0.12	24.05	49.04	34.64	23.39
$\Gamma_{3,4}$	0.53	0.43	0.62	17.70	55.86	32.37	23.39
$\Gamma_{3,5}$	0.21	0.01	0.36	17.53	55.89	42.58	23.39
$\Gamma_{4,5}$	0.09	-0.04	0.26	24.18	51.50	55.68	23.39

Table 4: Parameters: $\mathcal{AFNS}(3,2)$ Inefficiency comparison with the fixed blocking scheme (TaFB-MH)

that the yields are more highly correlated in the long-run than in the short-run. This is because the level factor appearing across all maturities with equal factor loading is more persistent than the slope and curvature factors whose weights are decreasing in maturities beyond the midterm. Interestingly, the estimated correlation structure differs across model specifications. The five-factor model reveals a higher correlation than the three-factor model. We evaluate the alternative predictive models in terms of out-of-sample density forecasts. Table 7 reports the log PPDs for the twelve months in 2007 along with the log posterior predictive likelihoods (PPL), which is the sum of the log PPDs. Over the out-of-sample periods, the three- and five-factor models produce similar

Maturity	1	3	12	24	36	48	60	120
1	1.00	0.93	0.91	0.90	0.85	0.77	0.74	0.65
3	0.93	1.00	0.86	0.86	0.82	0.76	0.73	0.65
12	0.91	0.86	1.00	0.91	0.89	0.84	0.81	0.72
24	0.90	0.86	0.91	1.00	0.97	0.92	0.90	0.84
36	0.85	0.82	0.89	0.97	1.00	0.97	0.96	0.91
48	0.77	0.76	0.84	0.92	0.97	1.00	0.94	0.91
60	0.74	0.73	0.81	0.90	0.96	0.94	1.00	0.94
120	0.65	0.65	0.72	0.84	0.91	0.91	0.94	1.00

(a) One-month-ahead

Maturity	1	3	12	24	36	48	60	120
1	1.00	0.99	0.97	0.95	0.92	0.89	0.86	0.78
3	0.99	1.00	0.97	0.95	0.92	0.89	0.87	0.78
12	0.97	0.97	1.00	0.98	0.96	0.94	0.92	0.84
24	0.95	0.95	0.98	1.00	0.99	0.98	0.97	0.90
36	0.92	0.92	0.96	0.99	1.00	0.99	0.99	0.93
48	0.89	0.89	0.94	0.98	0.99	1.00	0.99	0.95
60	0.86	0.87	0.92	0.97	0.99	0.99	1.00	0.97
120	0.78	0.78	0.84	0.90	0.93	0.95	0.97	1.00

(b) Six-month-ahead

Table 5: Posterior predictive correlation structure of the yields: $\mathcal{AFNS}(3,0)$

values regardless of the form of the $\boldsymbol{G}$ matrix. The macro factors that we choose do not seem to help improve the predictive accuracy because they have little additional information on the past yield curve in forecasting.

6 Conclusion

In this paper we have provided a detailed Bayesian analysis of affine arbitrage-free term-structure models with careful elaboration of the required factor identifying restrictions and the MCMC simulation techniques for overcoming the computational challenges. With the help of the available Matlab toolbox the entire approach can be applied to practical problems.

Further work is possible, for instance, connected to the development and fitting of affine models that incorporate zero lower bound restrictions on yields and models

Maturity	1	3	12	24	36	48	60	120
1	1.00	0.69	0.52	0.42	0.35	0.30	0.27	0.17
3	0.69	1.00	0.74	0.64	0.57	0.51	0.48	0.34
12	0.52	0.74	1.00	0.86	0.83	0.79	0.76	0.61
24	0.42	0.64	0.86	1.00	0.94	0.92	0.89	0.76
36	0.35	0.57	0.83	0.94	1.00	0.95	0.93	0.82
48	0.30	0.51	0.79	0.92	0.95	1.00	0.94	0.86
60	0.27	0.48	0.76	0.89	0.93	0.94	1.00	0.88
120	0.17	0.34	0.61	0.76	0.82	0.86	0.88	1.00

(a) One-month-ahead

Maturity	1	3	12	24	36	48	60	120
1	1.00	0.90	0.72	0.57	0.48	0.42	0.38	0.25
3	0.90	1.00	0.85	0.72	0.64	0.58	0.54	0.40
12	0.72	0.85	1.00	0.95	0.91	0.87	0.84	0.71
24	0.57	0.72	0.95	1.00	0.98	0.96	0.94	0.84
36	0.48	0.64	0.91	0.98	1.00	0.98	0.97	0.90
48	0.42	0.58	0.87	0.96	0.98	1.00	0.99	0.93
60	0.38	0.54	0.84	0.94	0.97	0.99	1.00	0.95
120	0.25	0.40	0.71	0.84	0.90	0.93	0.95	1.00

(a) Six-month-ahead

Table 6: Posterior predictive correlation structure of the yields: $\mathcal{AFNS}(3,2)$

that allow for the short-term interest rate to be close to zero for extended time spans. Although we have explicated the importance of incorporating economic information into the prior formulation it is possible that the prior in this paper can be further refined to incorporate other beliefs, such as those concerning the term structure of yield variances and the variance of the term premium. These various economic priors can be compared using posterior predictive likelihoods or marginal likelihoods.

References

Abbritti, M., Gil-Alana, L. A., Lovcha, Y., and Moreno, A. (2015), "Term Structure Persistence," *Journal of Financial Econometrics*, Forthcoming, 1–43.

Ang, A. and Piazzesi, M. (2003), "A no-arbitrage vector autoregression of term structure

	G is Full		G is Diagonal	
log PPD	$\mathcal{AFNS}(3,0)$	$\mathcal{AFNS}(3,2)$	$\mathcal{AFNS}(3,0)$	$\mathcal{AFNS}(3,2)$
Jan-07	64.99	64.85	64.90	65.00
Feb-07	64.15	63.80	64.13	64.02
Mar-07	64.53	64.14	64.38	64.37
Apr-07	63.49	63.48	63.21	63.31
May-07	62.72	62.97	62.22	61.93
Jun-07	63.68	63.62	63.56	63.59
Jun-07	61.38	61.42	61.16	61.29
Aug-07	57.74	58.55	58.41	58.03
Sep-07	59.21	59.28	59.58	58.72
Oct-07	63.93	64.01	63.92	64.02
Nov-07	55.50	54.82	56.34	56.12
Dec-07	56.21	56.27	56.23	56.37
log PPL	737.52	737.20	738.05	736.75

Table 7: Posterior predictive densities (PPD) and likelihoods (PPL)

dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50, 745–787.

Bauer, M. D. (2014), "Bayesian Estimation of Dynamic Term Structure Models under Restrictions on Risk Pricing," *Federal Reserve Bank of San Francisco Working Paper Series*, 2011-03, 1–43.

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

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

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

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

Dai, Q. and Singleton, K. J. (2000), "Specification analysis of affine term structure models," *Journal of Finance*, 55, 1943–1978.

Dai, Q., Singleton, K. J., and Yang, W. (2007), "Regime shifts in a dynamic term structure model of U.S. treasury bond yields," *Review of Financial Studies*, 20, 1669–1706.

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

Duffee, G. R. (2012), "Estimation of Dynamic Term Structure Models," *Quarterly Journal of Finance*, 02, 1–51.

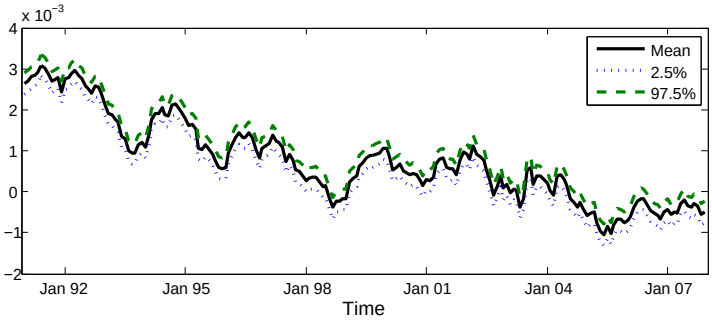
Duffie, G. (2002), "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57, 405–43.

Hamilton, J. D. and Wu, J. C. (2012), "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168, 315–331.

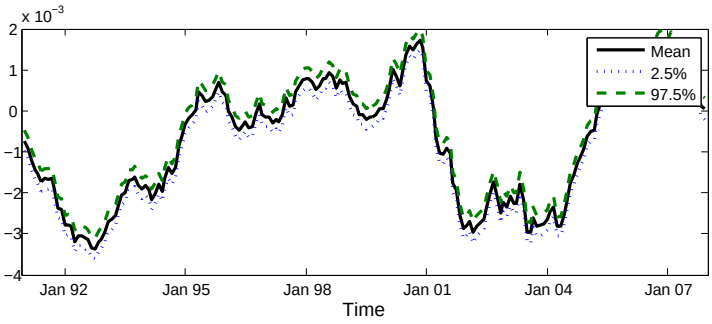
Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361–393.

Moench, E. (2008), "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach," *Journal of Econometrics*, 146, 26–43.

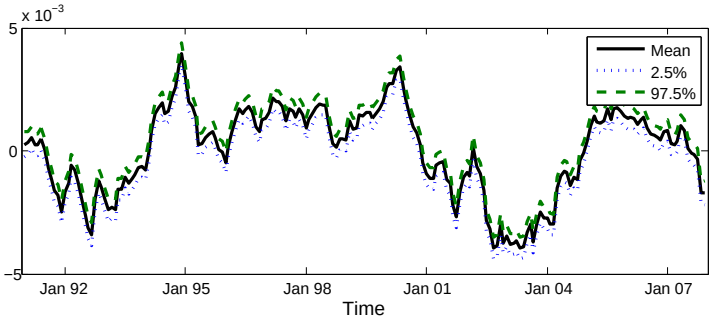
Niu, L. and Zeng, G. (2012), "The Discrete-Time Framework of Arbitrage-Free Nelson-Siegel Class of Term Structure Models," *manuscript*, 1–68.



(a) Level factor

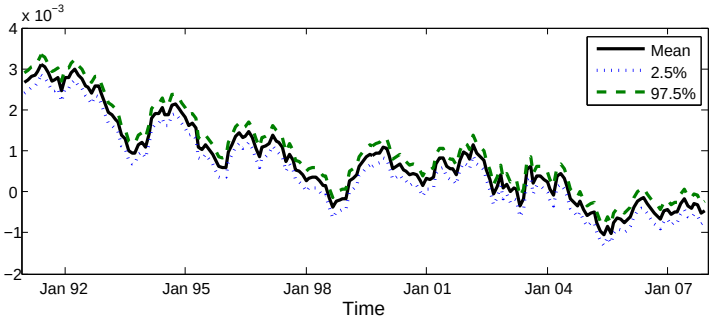


(b) Slope factor

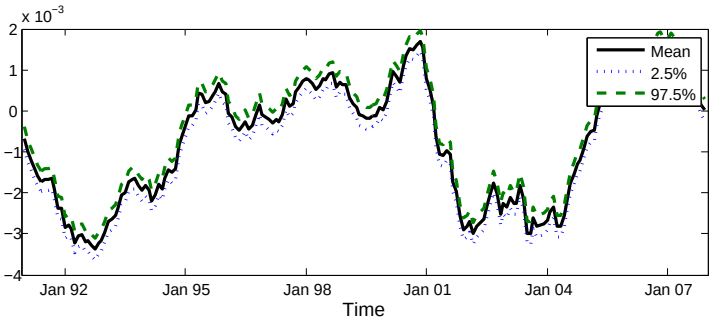


(c) Curvature factor

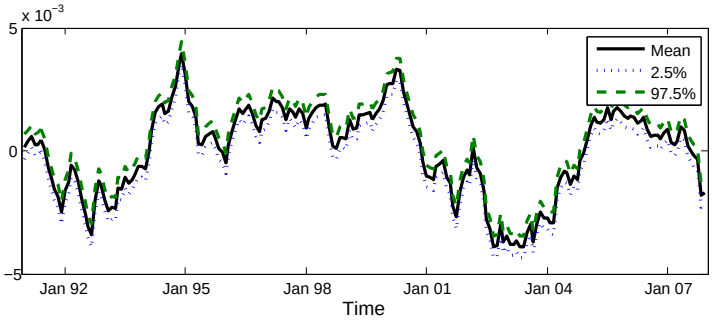
Figure 5: Factors: $\mathcal{AFNS}(3,0)$



(a) Level factor

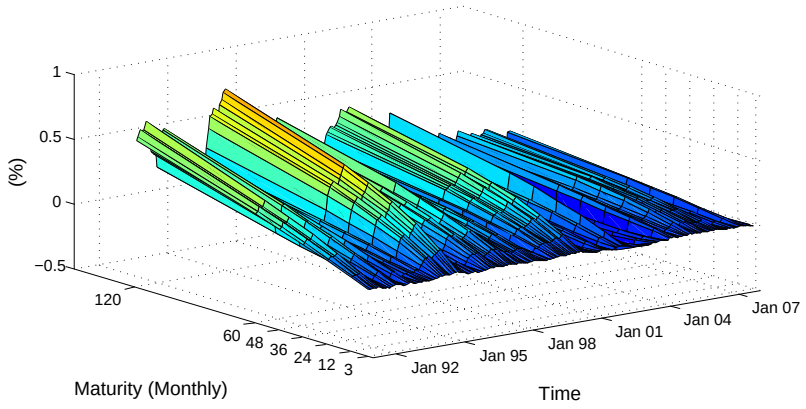


(b) Slope factor

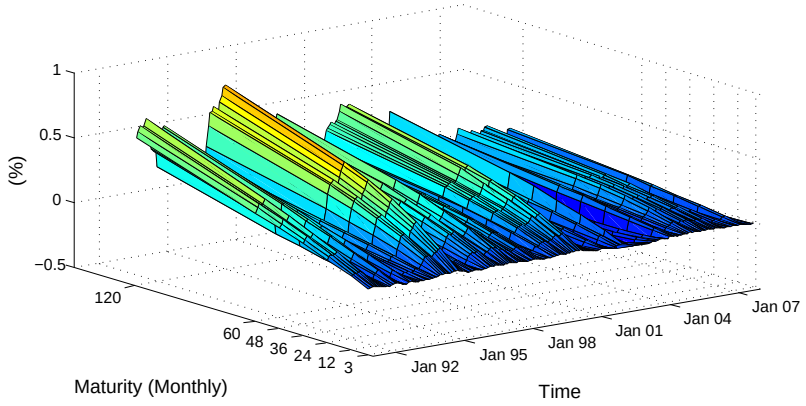


(c) Curvature factor

Figure 6: Factors: $\mathcal{AFNS}(3,2)$



(a) $\mathcal{AFNS}$(3,0)



(b) $\mathcal{AFNS}$(3,2)

Figure 7: Term structure of term premium

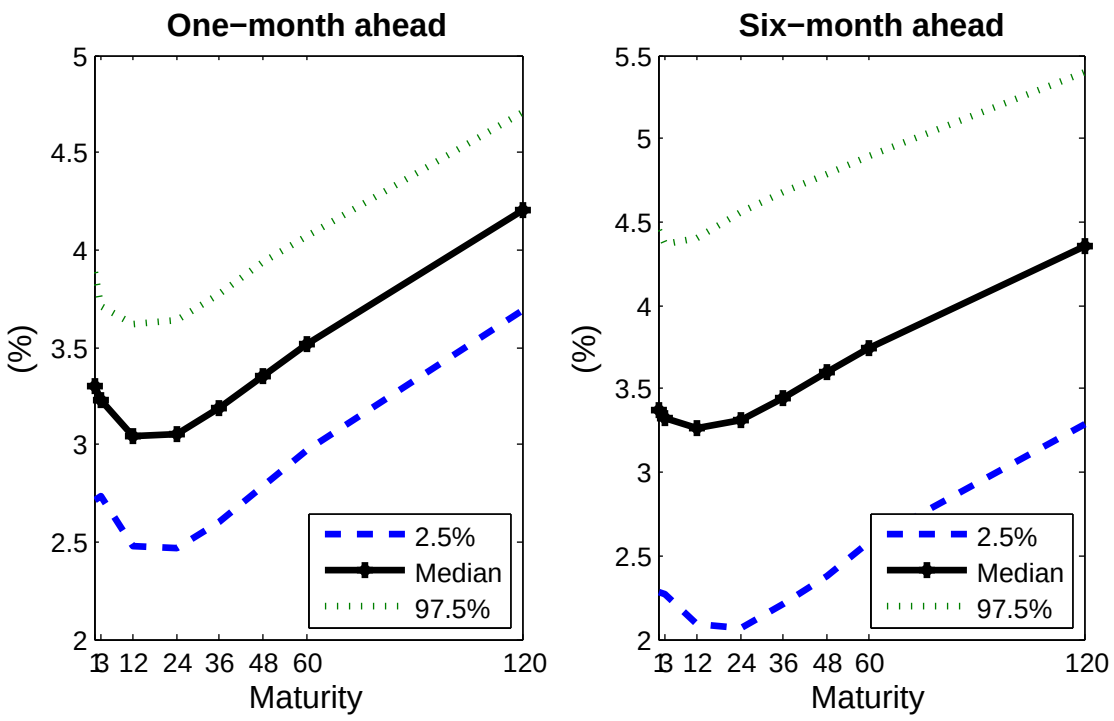


Figure 8: Posterior predictive densities: $\mathcal{AFNS}(3,0)$

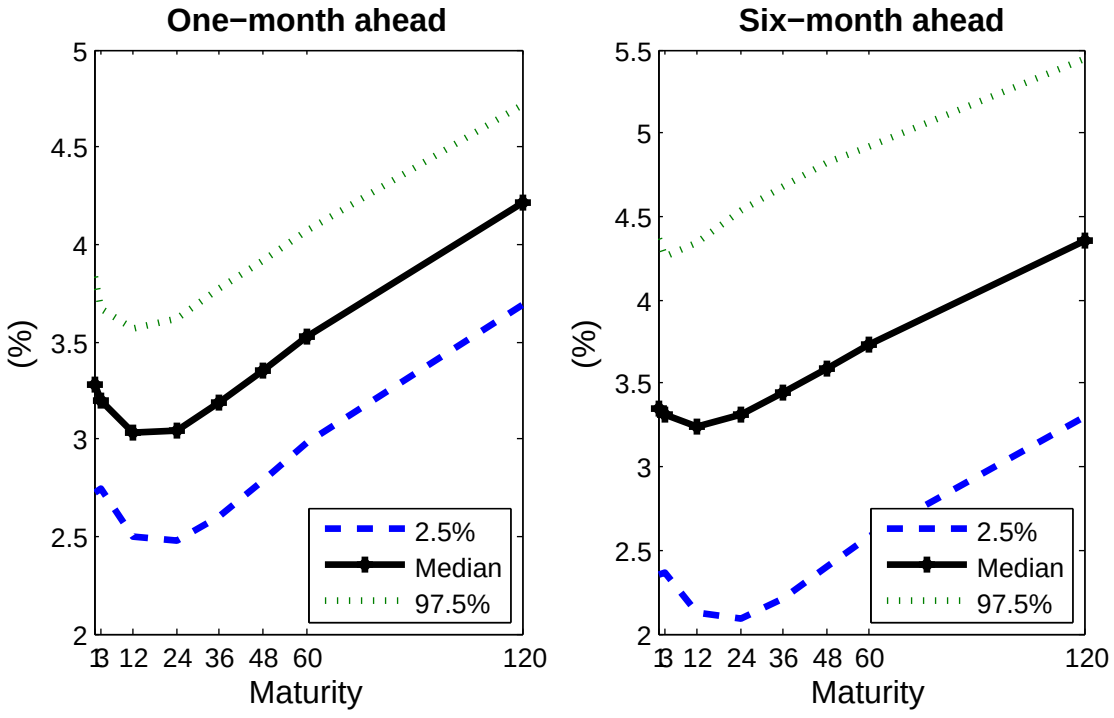


Figure 9: Posterior predictive densities: $\mathcal{AFNS}(3,2)$