*Journal of Financial Econometrics*, 2013, Vol. 11, No. 2, 302–334

# Change-Points in Affine Arbitrage-Free Term Structure Models

SIDDHARTHA CHIB
*Washington University in St. Louis*

KYU HO KANG
*Korea University*



ABSTRACT

In this paper, we investigate the timing of structural changes in yield curve dynamics in the context of an arbitrage-free, one latent and two macroeconomic factors, affine term structure model. We suppose that all parameters in the model are subject to changes at unknown time points. We fit a number of models to the U.S. term structure data and find support for three change-points. We also find that the term structure and the risk premium are materially different across regimes and that the out-of-sample forecasts of the term structure improve from incorporating regime changes. (*JEL*: G12, C11, E43)

KEYWORDS: Bayesian inference, change-points, macro-finance, marginal likelihood, Markov chain Monte Carlo, regime changes, state-space model, stochastic discount factor, term premium, yield curve

In a collection of important papers, Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), Ang and Bekaert (2002), and Ang, Bekaert, and Wei (2008) have developed Markov switching versions of arbitrage-free term structure models of the term structure. The Markov switching approach may be viewed as mainly an attempt to capture the effect of business cycle dynamics on the term structure. In this paper, we provide a new but complementary approach, for extending affine term structure models, through change-point modeling, in order to capture structural

We thank Taeyoung Doh, Ed Greenberg, Wolfgang Lemke, Hong Liu, James Morley, Srikanth Ramamurthy, Myung Hwan Seo, Yongs Shin, Guofu Zhou, and the referees and editor, for their thoughtful and useful comments on the paper. Kyu Ho Kang acknowledges support from the Center for Research in Economics and Strategy at the Olin Business School, Washington University in St. Louis. Address for correspondence: Siddhartha Chib, Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Bookings Drive, St. Louis, MO 63130, or e-mail: chib@wustl.edu.

doi:10.1093/jjfinec/nbs004 Advanced Access publication October 4, 2012

breaks in the spirit of the Lucas critique. In our change-point specification, a regime once occupied and vacated is never visited again, whereas in a Markov switching model, a regime occupied in the past can occur in the future. Change-point modeling can be useful if the conditions that determine a regime are unique and not repeated. In addition, from an econometric perspective, change-point models are somewhat simpler to estimate than Markov switching models because the so-called label switching problem does not arise in change-point models.

The objective, therefore, is to develop the change-point perspective in affine models. We show that by employing tuned Bayesian techniques, and the change-point model of Chib (1998), it is possible to construct and estimate affine term-structure models in which all model parameters, including the factor loadings, vary across regimes. The number of change-points in this general model is determined by model choice methods. Also because all parameters vary, we avoid the question of which parameters are constant and which break. In keeping with the evidence in the recent macro-finance literature (Ang and Piazzesi 2003; Ang, Dong, and Piazzesi 2007; Chib and Ergashev 2009), we specify our model in terms of three factors, one latent and two observed macroeconomic variables.

We apply our change-point model in an (extensive) empirical study of 16 yields of U.S. T-bills measured quarterly between 1972:I and 2007:IV. Because the different models we estimate are high dimensional, and the parameters are subject to complex cross-maturity restrictions, the prior distribution is formulated carefully, in line with the strategy described in Chib and Ergashev (2009). We estimate the models by Markov chain Monte Carlo (MCMC) methods, in particular the tailored randomized block M–H algorithm of Chib and Ramamurthy (2010). The idea behind this MCMC implementation is to update parameters in blocks, where both the number of blocks and the members of the blocks are randomly chosen within each MCMC cycle. In order to determine the number of change-points, we estimate models with different number of change-points and then select the best fitting model by the marginal likelihood/Bayes factors criteria. The marginal likelihoods are calculated by the method of Chib (1995).

The empirical results show that the three change-point model is the one that is best supported by the data. The results indicate that the regime changes occurred at the time points 1980:II, 1985:IV, and 1995:II. These dates roughly correspond to the start of the Volker era of the Federal Reserve, the start of the Greenspan chairmanship in 1987, and the start of the disclosures in 1994 by the Federal Open Market Committee of changes in the target federal funds rate and can be interpreted in the category of structural breaks. The model estimation also reveals that the parameters across regimes are substantially different, which suggests that parameters are indeed regime specific. The evidence shows, for instance, that the mean reversion parameters in the factor dynamics and the factor loadings vary across regimes. As a result, we find that the term structure and the bond premium are materially different across regimes and that the out-of-sample forecasts of the term structure improve from incorporating regime changes. Last, for comparison and scientific completeness, we also estimate a version of a two-state Markov switching model.

Interestingly, the out-of-sample forecast accuracy of this model is worse than the no-change-point model.

The rest of the paper is organized as follows. In Section 1, we present our change-point term structure model and the expression of the resulting bond prices. In Section 2, we outline the Bayesian prior–posterior analysis, and in Section 3, we provide results from our empirical analysis of the real data. Concluding remarks appear in Section 4. The Appendices, split into four parts, contain details regarding the derivation of the bond prices, the formulation of the prior distribution, the MCMC simulation procedure, and the calculation of the marginal likelihood.



## 1 MODEL SPECIFICATION

We start by setting up the affine term structure model in which all model parameters are subject to regime changes. Let $\{s_t\}$ denote a discrete-state variable that takes one of the values $\{1, 2, \ldots,\}$ such that $s_t = j$ indicates that the time $t$ observation has been drawn from the $j$th regime. We refer to the times $\{t_1, t_2, \ldots\}$ at which $s_t$ jumps from one value to the next as the change-points. We will also suppose that the parameters in the regimes induced by these change-points are different. Let

$$\mathbf{f}_t = (u_t, \mathbf{m_t})$$

denote the factors, where $u_t$ is a latent factor and $\mathbf{m}_t$ are two observed macroeconomic variables. Let $P_t(s_t, \tau)$ denote the price of the bond at time $t$ in regime $s_t$ that matures in period $(t + \tau)$. Then, under risk-neutral (or arbitrage-free) pricing, we have that

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

where $\mathbb{E}_t$ is the expectation over $(\mathbf{f}_{t+1}, s_{t+1})$, conditioned on $(\mathbf{f}_t, s_t)$, under the physical measure, and $\kappa_{t,s_t,t+1}$ is the stochastic discount factor (SDF) that converts a time $(t + 1)$ payoff into a payoff at time $t$ in regime $s_t$. The corresponding state-dependent yields for each time $t$ and maturity $\tau$ are given by

$$R_{t,\tau,s_t} = -\frac{\log P_t(s_t, \tau)}{\tau}.$$

We now characterize the stochastic evolution of $s_t$ and the factors $\mathbf{f}_t$ and describe our model of the SDF $\kappa_{t,s_t,t+1}$ in terms of the short rate process and the market price of factor risks. Given these ingredients, we show how one can price default-free zero-coupon bonds that satisfy the preceding risk-neutral pricing condition.

### 1.1 Change-Point Process

We suppose that economic agents are infinitely lived and face a possible infinity of change-points or, equivalently, regime changes. The regime in period $t$ is denoted

by $s_t \in \{1, 2, \ldots\}$. We assume that these agents know the current and past values of the state variable. The central uncertainty from the perspective of these agents is that the state of the next period is random—either the current regime continues or the next possible regime emerges, following the process of change-points in Chib (1998).

Suppose now that from one time period to the next, $s_t$ can either stay at the current value $j$ or jump to the next higher value $(j+1)$. Thus, in this formulation, return visits to a previously occupied state are not possible. Then, the $j$th change point occurs at time (say) $t_j$ when $s_{t_j-1} = j$ and $s_{t_j} = j+1$. Following Chib (1998), $s_t$ is assumed to follow a Markov process with transition probabilities given by $p_{jk} = \Pr[s_{t+1} = k | s_t = j]$ and $p_{jk} = 1 - p_{jj}, k = j+1$. Thus,

$$s_{t+1} = \begin{cases} s_t & \text{with probability} \quad p_{s_t s_t} \\ s_t + 1 & \text{with probability} \quad 1 - p_{s_t s_t}. \end{cases}$$

This formulation of the change-point model in terms of a restricted unidirectional Markov process shows how the change-point assumption differs from the Markov switching regime process in Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), and Ang, Bekaert, and Wei (2008) where the transition probability matrix is unrestricted and previously occupied states can be revisited. Each model offers a different perspective on regime changes. If regime changes are frequent, and states are repeated, then the Markov switching model is more appropriate. If the regimes constitute several distinct epochs, even if there are many such epochs, then the change-point approach should be adopted. As we show below in Table 3 and Figures 3 and 4, the estimation results seem to support the change-point assumption.

## 1.2 Factor Process

Next, suppose that the distribution of $\mathbf{f}_{t+1}$, conditioned on $(\mathbf{f}_t, s_t, s_{t+1})$, is determined by a Gaussian regime-specific mean-reverting first-order autoregression given by

$$\mathbf{f}_{t+1} = \boldsymbol{\mu}_{s_{t+1}} + \mathbf{G}_{s_{t+1}}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\eta}_{t+1}, \tag{2}$$

where on letting $\mathcal{N}_3(\cdot, \cdot)$ denote the three-dimensional normal distribution,

$$\boldsymbol{\eta}_{t+1} | s_{t+1} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Omega}_{s_{t+1}}),$$

and $\boldsymbol{\mu}_{s_{t+1}}$ is a $3 \times 1$ vector and $\mathbf{G}_{s_{t+1}}$ is a $3 \times 3$ matrix. In the sequel, we will express $\boldsymbol{\eta}_{t+1}$ in terms of a vector of i.i.d. standard normal variables $\boldsymbol{\omega}_{t+1}$ as

$$\boldsymbol{\eta}_{t+1} = \mathbf{L}_{s_{t+1}} \boldsymbol{\omega}_{t+1}, \tag{3}$$

where $\mathbf{L}_{s_{t+1}}$ is the lower triangular Cholesky decomposition of $\boldsymbol{\Omega}_{s_{t+1}}$.

Thus, the factor evolution is a function of the current and previous states (in contrast, the dynamics in Dai, Singleton, and Yang, 2007 depend only on $s_t$, whereas those in Bansal and Zhou, 2002 and Ang, Bekaert, and Wei, 2008 depend only on $s_{t+1}$). This means that the expectation of $\mathbf{f}_{t+1}$ conditioned on $(\mathbf{f}_t, s_t = j, s_{t+1} = k)$ is a function of both $\boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_k$. The appearance of $\boldsymbol{\mu}_j$ in this expression is natural because one would like the autoregression at time $(t+1)$ to depend on the deviation of $\mathbf{f}_t$ from the regime in the previous period. Of course, the parameter $\boldsymbol{\mu}_j$ can be interpreted as the expectation of $\mathbf{f}_{t+1}$ when regime $j$ is persistent. The matrices $\{\mathbf{G}_j\}$ can also be interpreted in the same way as the mean-reversion parameters in regime $j$.

## 1.3 Stochastic Discount Factor

We complete our modeling by assuming that the SDF $\kappa_{t,s_t,t+1}$ that converts a time $(t+1)$ payoff into a payoff at time $t$ in regime $s_t$ is given by

$$\kappa_{t,s_t,t+1} = \exp\left(-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t} - \boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\omega}_{t+1}\right), \tag{4}$$

where $r_{t,s_t}$ is the short rate in regime $s_t$, $\boldsymbol{\gamma}_{t,s_t}$ is the vector of time-varying and regime-sensitive market prices of factor risks, and $\boldsymbol{\omega}_{t+1}$ is the i.i.d. vector of regime-independent factor shocks in Equation (3). The SDF is independent of $s_{t+1}$ given $s_t$ as in the model of Dai, Singleton, and Yang (2007). It is easily checked that $\mathbb{E}[\kappa_{t,s_t,t+1}|\mathbf{f}_t, s_t = j]$ is equal to the price of a zero-coupon bond with $\tau = 1$. In other words, the SDF satisfies the intertemporal no-arbitrage condition (Dai, Singleton, and Yang, 2007).

We suppose that the short rate is affine in the factors and of the form

$$r_{t,s_t} = \delta_{1,s_t} + \boldsymbol{\delta}'_{2,s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \tag{5}$$

where the intercept $\delta_{1,s_t}$ varies by regime to allow for shifts in the level of the term structure. The multiplier $\boldsymbol{\delta}_{2,s_t} : 3 \times 1$ is also regime dependent in order to capture shifts in the effects of the macroeconomic factors on the term structure. This is similar to the assumption in Bansal and Zhou (2002) but a departure from both Ang, Bekaert, and Wei (2008) and Dai, Singleton, and Yang (2007) where the coefficient on the factors is constant across regimes. A consequence of our assumption is that the bond prices that satisfy the risk-neutral pricing condition can only be obtained approximately. The same difficulty arises in the work of Bansal and Zhou (2002).

We also assume that the dynamics of $\boldsymbol{\gamma}_{t,s_t}$ are governed by

$$\boldsymbol{\gamma}_{t,s_t} = \tilde{\boldsymbol{\gamma}}_{s_t} + \boldsymbol{\Phi}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \tag{6}$$

where $\tilde{\boldsymbol{\gamma}}_{s_t} : 3 \times 1$ is the regime-dependent expectation of $\boldsymbol{\gamma}_{t,s_t}$ and $\boldsymbol{\Phi}_{s_t} : 3 \times 3$ is a matrix of regime-specific parameters. We refer to the collection $(\tilde{\boldsymbol{\gamma}}_{s_t}, \boldsymbol{\Phi}_{s_t})$ as the factor risk parameters. Note that in this specification, $\boldsymbol{\gamma}_{t,s_t}$ is the same across maturities but different across regimes. A point to note is that negative market prices

of risk have the effect of generating a positive term premium. This is important to keep in mind when we construct the prior distribution on the risk parameters.

We note that regime-shift risk is equal to zero in our version of the SDF. We make this assumption because it is difficult to identify this risk from our change-point model where each regime-shift occurs once. In the models of Ang, Bekaert, and Wei (2008) and Bansal and Zhou (2002), regime risk cannot also be isolated since it is confounded with the market price of factor risk. We are, however, able to identify the market price of factor risk since we assume that the SDF is independent of $s_{t+1}$ conditioned on $s_t$, as in the model of Dai, Singleton, and Yang (2007). Alternatively, our specification can be more grounded in economic fundamentals by letting the SDF depend on $\gamma_{t+1,s_{t+1}}$ rather than $\gamma_{t,s_t}$ because in a general equilibrium setting (see the Appendix of Bansal and Zhou 2002), the current consumption growth is affected by the current state as well as the past state. Our computational experiments indicate, however, that modifying the SDF in this way does not change the estimation results.

## 1.4 Bond Prices

Under these assumptions, we now solve for bond prices that satisfy the risk-neutral pricing condition

$$P_t(s_t, \tau) = \mathbb{E}_t[\kappa_{t,s_t,t+1} P_{t+1}(s_{t+1}, \tau - 1)]. \tag{7}$$

Following Duffie and Kan (1996), we assume that $P_t(s_t, \tau)$ is a regime-dependent exponential affine function of the factors taking the form

$$P_t(s_t, \tau) = \exp(-\tau R_{t,\tau,s_t}), \tag{8}$$

where $R_{t,\tau,s_t}$ is the continuously compounded yield given by

$$R_{t,\tau,s_t} = \frac{1}{\tau} a_{s_t}(\tau) + \frac{1}{\tau} \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \tag{9}$$

and $a_{s_t}(\tau)$ is a scalar function and $\mathbf{b}_{s_t}(\tau)$ is a $3 \times 1$ vector of functions, both depending on $s_t$ and $\tau$.

We follow Bansal and Zhou (2002) and find the expressions for the latter functions by combining the principles of log-linearization, the method of undetermined coefficients, and the law of the iterated expectation. As we discuss in Appendix A, we are then able to show that for $j = 1, 2, \ldots,$ and $k = j + 1$, the unknown functions satisfy the recursive system

$$a_j(\tau) = \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \delta_{1,j} - \tilde{\boldsymbol{\gamma}}_j' \mathbf{L}_j' \mathbf{b}_j(\tau - 1) - \mathbf{b}_j(\tau - 1)' \mathbf{L}_j \mathbf{L}_j' \mathbf{b}_j(\tau - 1)/2 + a_j(\tau - 1) \\ \delta_{1,j} - \tilde{\boldsymbol{\gamma}}_j' \mathbf{L}_k' \mathbf{b}_k(\tau - 1) - \mathbf{b}_k(\tau - 1)' \mathbf{L}_k \mathbf{L}_k' \mathbf{b}_k(\tau - 1)/2 + a_k(\tau - 1) \end{pmatrix},$$

$$\mathbf{b}_j(\tau) = \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{2,j} + \left(\mathbf{G}_j - \mathbf{L}_j \boldsymbol{\Phi}_j\right)' \mathbf{b}_j(\tau - 1) \\ \boldsymbol{\delta}_{2,j} + \left(\mathbf{G}_k - \mathbf{L}_k \boldsymbol{\Phi}_j\right)' \mathbf{b}_k(\tau - 1) \end{pmatrix}, \tag{10}$$

where $\tau$ runs over the positive integers. These recursions are initialized by setting $a_{s_t}(0) = 0$ and $\mathbf{b}_{s_t}(0) = 0_{3\times 1}$ for all $s_t$. It is readily seen that the resulting intercept and factor loadings are determined by the weighted average of the two potential realizations in the next period where the weights are given by the transition probabilities $p_{jj}$ and $(1 - p_{jj})$, respectively. Thus, the bond prices in regime $s_t = j$ incorporate the expectation that the economy in the next period will continue to stay in regime $j$, or that it will switch to the next possible regime $k = j + 1$, each weighted with the probabilities $p_{jj}$ and $1 - p_{jj}$, respectively.

Note that when we consider inference with a finite sample of data of size $n$, we consider models with finite and different number of change-points. We indicate the number of change-points by $m$, where $m = 0, 1, 2, \ldots$ In that case, when we estimate the $m$ change-point model, state $(m+1)$ is by definition the final state. We then set $p_{m+1,m+1} = 1$ and set $p_{jk} = 0$ in the above recursions once $j = m + 1$. It should also be noted that in the estimation of the $m$ change-point model, the $(m+1)$st regime is the upper limit on the number of possible regimes under that model supposition, and that fewer regimes may arise when the states are sampled by the method in Step 4 of Appendix C. Note also that the final state is only fixed for a given model but is not fixed overall since $m$ varies as we consider models with different number of change-points. As we discuss below, one can find the best fitting model, and hence the number of change-points, from the marginal likelihoods of these different models.

## 1.5 Regime-Specific Term Premium

As is well known, under risk-neutral pricing, after adjusting for risk, agents are indifferent between holding a $\tau$-period bond and a risk-free bond for one period. The risk adjustment is the term premium. In the regime change model, this term premium is regime specific. For each time $t$ and in the current regime $s_t = j$, the term premium for a $\tau$-period bond can be expressed as $(\tau - 1)$ times the conditional covariance at time between the log of the SDF at time $(t + 1)$ and the yield at time $(t + 1)$ on a $(\tau - 1)$ period bond. In particular, this term premium can be calculated as

$$\begin{aligned}\text{Term premium}_{\tau,t,s_t} &= (\tau-1)\text{Cov}(\ln \kappa_{t,s_t,t+1}, R_{t+1,s_{t+1},\tau-1}|\mathbf{f}_t, s_t = j) \qquad (11)\\ &= -p_{jj}\mathbf{b}_j(\tau-1)'\mathbf{L}_j\boldsymbol{\gamma}_{t,j} - p_{jk}\mathbf{b}_k(\tau-1)'\mathbf{L}_k\boldsymbol{\gamma}_{t,j},\end{aligned}$$

where $k = j + 1$. One can see that if $\mathbf{L}_j$, which quantifies the size of the factor shocks in the current regime $s_t = j$, is large, or if $\boldsymbol{\gamma}_{t,j}$, the market prices of factor risk, is highly negative, then the term premium is expected to be large. Even if $\mathbf{L}_j$ in the current regime is small, one can see from the second term in the above expression that the term premium can be big if the probability of jumping to the next possible regime is high and $\mathbf{L}_k$ in that regime is large. In our empirical implementation,

we calculate this regime-specific term premium for each time period in the sample.

## 2 ESTIMATION AND INFERENCE

In this section, we consider the empirical implementation of our yield curve model. In order to get a detailed perspective of the yield curve and its dynamics over time, we operationalize our pricing model on a dataset of sixteen yields of U.S. T-bills measured quarterly between 1972:I and 2007:IV on the maturities given by

$$\{1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 20, 24, 28, 36, 40\}$$

quarters. For these data, we consider five versions of our general model, with 0, 1, 2, 3, and 4 change-points and denoted by $\{\mathcal{M}_m\}_{m=0}^4$. The largest model that we fit, namely $\mathcal{M}_4$, has a total of 209 free parameters. Since the number of change-points are random in our setting, we find the appropriate number of change-points through the computation of marginal likelihoods and Bayes factors, as we discuss below. We also compare the different models in terms of the predictive performance out-of-sample.

To begin, let the sixteen yields under study be denoted by

$$(R_{t1}, R_{t2}, \ldots, R_{t16})', \quad t = 1, 2, \ldots, n, \tag{12}$$

where $R_{t,\tau}$ denotes the yield of $\tau$-period maturity bond at time $t$, $R_{ti} = R_{t,\tau_i}$ and $\tau_i$ is the $i$th maturity (in quarters). Let the two macro factors be denoted by

$$\mathbf{m}_t = (m_{t1}, m_{t2}), \quad t = 1, 2, \ldots, n,$$

where $m_{t1}$ is the inflation rate and $m_{t2}$ is the real GDP growth rate. We also let

$$\mathbf{S}_n = \{s_t\}_{t=1}^n$$

denote the sequence of (unobserved) regime indicators.

We now specify the set of model parameters to be estimated. First, the unknown elements of $\mathbf{G}_{s_t}$ and $\mathbf{\Phi}_{s_t}$ are denoted by

$$\mathbf{g}_{s_t} = \{G_{ij,s_t}\}_{i,j=1,2,3} \quad \text{and} \quad \phi_{s_t} = \{\Phi_{jj,s_t}\}_{j=1,2,3},$$

where $G_{ij,s_t}$ and $\Phi_{ij,s_t}$ denote the $(i,j)$th element of $\mathbf{G}_{s_t}$ and $\mathbf{\Phi}_{s_t}$, respectively. The unknown elements of $\mathbf{\Omega}_{s_t}$ are defined as

$$\boldsymbol{\lambda}_{s_t} = \{l_{21,s_t}, l^*_{22,s_t}, l_{31,s_t}, l_{32,s_t}, l^*_{33,s_t}\},$$

where these are obtained from the decomposition $\mathbf{\Omega}_{s_t} = \mathbf{L}_{s_t}\mathbf{L}'_{s_t}$ with $\mathbf{L}_{s_t}$ expressed as

$$\begin{pmatrix} 1/400 & 0 & 0 \\ l_{21,s_t} & \exp(l^*_{22,s_t}) & 0 \\ l_{31,s_t} & l_{32,s_t} & \exp(l^*_{33,s_t}) \end{pmatrix}. \tag{13}$$

The elements of $\boldsymbol{\lambda}_{s_t}$ are unrestricted. Next, the parameters of the short rate equation are expressed as $\boldsymbol{\delta}_{s_t} = (\delta_{1,s_t} \times 400, \boldsymbol{\delta}'_{2,s_t})'$ and those in the transition matrix $\mathbf{P}$ by $\mathbf{p} = \{p_{jj}, j = 1, 2, \ldots, m\}$. Finally, the unknown pricing error variances $\sigma^2_{i,s_t}$ are collected in reparameterized form as

$$\boldsymbol{\sigma}^{*2} = \{\sigma^{*2}_{i,s_t} = d_i \sigma^2_{i,s_t},\ i = 1, \ldots, 7, 8, \ldots, 16 \quad \text{and} \quad s_t = 1, 2, \ldots, m+1\},$$

where $d_1 = 30, d_2 = d_{16} = 40, d_3 = d_{12} = 200, d_4 = 350, d_5 = d_6 = d_{11} = 500, d_7 = 3000, d_9 = 1500, d_{10} = 1000, d_{13} = d_{14} = d_{15} = 200$. These positive multipliers are introduced to increase the magnitude of the variances.

Under these notations, for any given model with $m$ change-points, the parameters of interest can be denoted as $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, u_0)$ where

$$\boldsymbol{\theta} = \{\mathbf{g}_{s_t}, \boldsymbol{\mu}_{m,s_t}, \boldsymbol{\delta}_{s_t}, \tilde{\gamma}_{s_t}, \boldsymbol{\phi}_{s_t}, \boldsymbol{\lambda}_{s_t}, \mathbf{p}\}_{s_t=1}^{m+1},$$

and $u_0$ is the latent factor at time 0. Note that to economize on notation, we do not index these parameters by a model subscript.

## 2.1 Joint Distribution of the Yields and Macro Factors

We now derive the joint distribution of the yields and the macro factors conditioned on $\mathbf{S}_n$ and $\boldsymbol{\psi}$. This joint distribution can be obtained without marginalization over $\{u_t\}_{t=1}^n$ if we assume, following, for example, Chen and Scott (2003) and Dai, Singleton, and Yang (2007), that one of the yields is priced exactly without error. This is the so-called basis yield. Under this assumption, the latent factor can be expressed in terms of the observed variables and eliminated from the model, as we now describe.

Assume that $R_{t8}$ (the 8th yield in the list above) is the basis yield, which is priced exactly by the model. Let $\mathbf{R}_t$ denote the remaining fifteen yields (which are measured with pricing error). Define $\bar{a}_{i,s_t} = a_{s_t}(\tau_i)/\tau_i$ and $\bar{\mathbf{b}}_{i,s_t} = \mathbf{b}_{s_t}(\tau_i)/\tau_i$ where $a_{s_t}(\tau_i)$ and $\mathbf{b}_{s_t}(\tau_i)$ are obtained from the recursive equations in (10). Also let $\bar{a}_{8,s_t}$ ($\bar{\mathbf{a}}_{s_t}$) and $\bar{\mathbf{b}}_{8,s_t}$ ($\bar{\mathbf{b}}_{s_t}$) be the corresponding intercept and factor loadings for $R_{t8}$ ($\mathbf{R}_t$), respectively. Then, since the basis yield is priced without error, if we let

$$\bar{\mathbf{b}}_{8,s_t} = \begin{pmatrix} \bar{b}_{8,u,s_t} \\ \bar{\mathbf{b}}_{8,m,s_t} \end{pmatrix}, \tag{14}$$

we can see from Equation (9) that $R_{t8}$ is given by

$$R_{t8} = \bar{a}_{8,s_t} + \bar{b}_{8,u,s_t} u_t + \bar{\mathbf{b}}'_{8,m,s_t}(\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t}). \tag{15}$$

On rewriting this expression, it follows that $u_t$ is

$$u_t = \left(\bar{b}_{8,u,s_t}\right)^{-1} \left(R_{t8} - \bar{a}_{8,s_t} - \bar{\mathbf{b}}'_{8,m,s_t}(\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})\right). \tag{16}$$

Conditioned on $\mathbf{m}_t$ and $s_t$, this represents a one-to-one map between $R_{t8}$ and $u_t$. If we let

$$\mathbf{z}_t = \begin{pmatrix} R_{t8} \\ \mathbf{m}_t \end{pmatrix},$$

$$\alpha_{s_t} = \begin{pmatrix} \left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{\mathbf{b}}'_{8,m,s_t} \boldsymbol{\mu}_{m,s_t} - \left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{a}_{8,s_t} \\ 0_{2\times 1} \end{pmatrix}, \quad \text{and} \tag{17}$$

$$\mathbf{A}_{s_t} = \begin{pmatrix} \left(\bar{b}_{8,u,s_t}\right)^{-1} - \left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{\mathbf{b}}'_{8,m,s_t} \\ 0_{2\times 1} \qquad\qquad \mathbf{I}_2 \end{pmatrix},$$

then one can check that $\mathbf{f}_t$ can be expressed as

$$\mathbf{f}_t = \alpha_{s_t} + \mathbf{A}_{s_t} \mathbf{z}_t. \tag{18}$$

It now follows from Equation (9) that conditioned on $\mathbf{z}_t$ (equivalently $\mathbf{f}_t$), $s_t$, and the model parameters $\boldsymbol{\psi}$, the nonbasis yields $\mathbf{R}_t$ in our model are generated according to the process

$$\mathbf{R}_t = \bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\varepsilon}_t, \;\; \boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(0, \boldsymbol{\Sigma}_{s_t}), \tag{19}$$

where

$$\boldsymbol{\Sigma}_{s_t} = \text{diag}(\sigma^2_{1,s_t}, \sigma^2_{2,s_t}, \ldots, \sigma^2_{7,s_t}, \sigma^2_{9,s_t},, \sigma^2_{16,s_t}).$$

In other words,

$$\begin{aligned} p(\mathbf{R}_t|\mathbf{z}_t, s_t, \boldsymbol{\psi}) &= p(\mathbf{R}_t|\mathbf{f}_t, s_t, \boldsymbol{\psi}) \\ &= \mathcal{N}_{15}(\mathbf{R}_t|\bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \boldsymbol{\Sigma}_{s_t}). \end{aligned} \tag{20}$$

In addition, the distribution of $\mathbf{z}_t$ conditioned on $\mathbf{z}_{t-1}$, $s_t$ and $s_{t-1}$ is obtained straightforwardly from the process generating $\mathbf{f}_t$ given in Equation (2) and the linear map between $\mathbf{f}_t$ and $\mathbf{z}_t$ given in Equation (18). In particular,

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{z}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) &= p(\mathbf{f}_t|\mathbf{f}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) \det(\mathbf{A}_{s_t}) \\ &= \mathcal{N}_3(\mathbf{f}_t|\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t}(\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \boldsymbol{\Omega}_{s_t})|\left(\bar{b}_{8,u,s_t}\right)^{-1}|. \end{aligned} \tag{21}$$

If we let

$$\mathbf{y}_t = (\mathbf{R}_t, \mathbf{z}_t) \quad \text{and} \quad \mathbf{y} = \{\mathbf{y}_t\}_{t=1}^n,$$

it follows that the required joint density of $\mathbf{y}$ conditioned on $(\mathbf{S}_n, \boldsymbol{\psi})$ is given by

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi}) = \prod_{t=1}^{n} \mathcal{N}_{15}(\mathbf{R}_t|\bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \boldsymbol{\Sigma}_{s_t}) \tag{22}$$

$$\times \mathcal{N}_3(\mathbf{f}_t|\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t}(\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \boldsymbol{\Omega}_{s_t})|\left(\bar{b}_{8,u,s_t}\right)^{-1}|. \tag{23}$$

## 2.2 Prior–Posterior Analysis

Because of the size of the parameter space, and the complex cross-maturity restrictions on the parameters, the formulation of the prior distribution can be a challenge. Chib and Ergashev (2009) have tackled this problem and shown that a reasonable approach for constructing the prior is to think in terms of the term structure that is implied by the prior distribution. The implied yield curve can be determined by simulation: simulating parameters from the prior and simulating yields from the model given the parameters. The prior can be adjusted until the implied term structure is viewed as satisfactory on a priori considerations. Chib and Ergashev (2009) use this strategy to arrive at a prior distribution that incorporates the belief of a positive term premium and stationary but persistent factors. We adapt their approach for our model with change-points, ensuring that the yield curve implied by our prior distribution is upward sloping on average, though the prior-implied yield curve at times can be flat or inverted. We assume, in addition, that the prior distribution of the regime-specific parameters is identical across regimes. We arrive at our prior distribution in this way for each of the five models we consider—with 0, 1, 2, 3, and 4 change-points. Full details of each of model parameters are given in Appendix B.

Under our prior, it is now possible to calculate the posterior distribution of the parameters by MCMC simulation methods. Our MCMC approach is grounded in the recent developments that appear in Chib and Ergashev (2009) and Chib and Ramamurthy (2010). The latter paper introduces an implementation of the MCMC method (called the tailored randomized block M–H algorithm) that we adopt here to fit our model. The idea behind this implementation is to update parameters in blocks, where both the number of blocks and the members of the blocks are randomly chosen within each MCMC cycle. This strategy is especially valuable in high-dimensional problems and in problems where it is difficult to form the blocks on a priori considerations. Appendix C provides the technical details.

# 3 RESULTS

We apply our modeling approach to analyze U.S. data on quarterly yields of sixteen U.S. T-bills between 1972:I and 2007:IV. These data are taken from Gurkaynak, Sack, and Wright (2007). We consider zero-coupon bonds of maturities 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 20, 24, 28, 36, and 40 quarters. We let the basis yield be the eight-quarter (or two-year) bond, which is the bond with the smallest pricing variance. Our macroeconomic factors are the quarterly GDP inflation deflator and the real GDP growth rate. These data are from the Federal reserve bank of St. Louis.

We work with sixteen yields because our tuned Bayesian estimation approach is capable of handling a large set of yields. The involvement of these many yields also tends to improve the out-of-sample predictive accuracy of the yield curve forecasts. To show this, we also fit models with four, eight, and twelve yields to data up to 2006. The last four quarters of 2007 are held aside for the validation of

**Table 1** PPC

| The number of maturities ($\lambda$) | No-change-point model | | |
|---|---|---|---|
| | D | W | PPC |
| 4 | 6.293 | 4.821 | 11.114 |
| 8 | 5.827 | 4.758 | 10.585 |
| 12 | 4.621 | 4.191 | 8.812 |
| 16 | 4.011 | 3.520 | 7.531 |

PPC is computed by Equations (24)–(26). We use the data from the most recent break time point, 1995: II–2006:IV due to the regime shift, and out of sample period is 2007:I–2007:IV. Four yields are of 2, 8, 20, and 40 quarters maturity bonds (used in Dai, Singleton, and Yang 2007). Eight yields are of 1, 2, 3, 4, 8, 12, 16, and 20 quarters maturity bonds (used in Bansal and Zhou 2002). Twelve yields are of 1, 2, 3, 4, 5, 6, 8, 12, 20, 28, 32, and 40 quarters maturity bonds. Sixteen yields are of 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 20, 24, 28, 32, and 40 quarters maturity bonds.

the predictions of the yields and the macro factors. These predictions are generated as described in Section 3.4. We measure the predictive accuracy of the forecasts in terms of the posterior predictive criterion (PPC) of Gelfand and Ghosh (1998). For a given model with $\lambda$ number of the maturities, PPC is defined as

$$\text{PPC} = \text{D} + \text{W}, \tag{24}$$

where

$$\text{D} = \frac{1}{\lambda+2} \sum_{i=1}^{\lambda+2} \sum_{t=1}^{T} \text{Var}(\tilde{y}_{i,t} | \mathbf{y}, \mathcal{M}), \tag{25}$$

$$\text{W} = \frac{1}{\lambda+2} \sum_{i=1}^{\lambda+2} \sum_{t=1}^{T} [y_{i,t} - E(\tilde{y}_{i,t} | \mathbf{y}, \mathcal{M})]^2. \tag{26}$$

$\{\tilde{\mathbf{y}}_t\}_{t=1,2,\ldots,T}$ are the predictions of the yields and macro factors $\{\mathbf{y}_t\}_{t=1,2,\ldots,T}$ under model $\mathcal{M}$, and $\tilde{y}_{i,t}$ and $y_{i,t}$ are the $i$th components of $\tilde{\mathbf{y}}_t$ and $\mathbf{y}_t$, respectively. The term D is expected to be large in models that are restrictive or have redundant parameters. The term W measures the predictive goodness-of-fit. As can be seen from Table 1, the model with sixteen maturities outperforms the models with fewer maturities.

The reason for this behavior is simple. The addition of a new yield introduces only one parameter (namely the pricing error variance) but because of the many cross-equation restrictions on the parameters, the additional outcome helps to improve inferences about the common model parameters, which translates into improved predictive inferences.

## 3.1 Sampler Diagnostics

We base our results on 50,000 iterations of the MCMC algorithm beyond a burn-in of 5000 iterations. We measure the efficiency of the MCMC sampling in terms of

the metrics that are common in the Bayesian literature, in particular, the acceptance rates in the Metropolis–Hastings steps and the inefficiency factors (Chib, 2001), which, for any sampled sequence of draws, are defined as

$$1 + 2\sum_{k=1}^{K} \rho(k), \tag{27}$$

where $\rho(k)$ is the $k$-order autocorrelation computed from the sampled variates and $K$ is a large number, which we choose conservatively to be 500. For our biggest model, the average acceptance rate and the average inefficiency factor in the M–H step are 72.9% and 174.1, respectively. These values indicate that our sampler mixes well. It is also important to mention that our sampler converges quickly to the same region of the parameter space regardless of the starting values.

## 3.2 The Number and Timing of Change-Points

One of our goals was to evaluate the extent to which the regime change model is an improvement over the model without regime changes. We are also interested in determining how many regimes best describe the sample data. Specifically, we are interested in the comparison of six models, which in the introduction were named as $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$. Our most general model is $M_4$ consisting of four possible change-points, one latent factor, and two macro factors. For completeness, we also consider a two-state Markov switching model that we denote by $\mathcal{M}_{\text{MS}}$. In estimating this model, we impose the restriction that the coefficient of the latent factor in regime 2, $\delta_{21,2}$, is bigger than that in regime 1, $\delta_{21,2}$. The prior on the parameters is comparable to that of the change-point model. We compare the collection of models in terms of out-of-sample predictions and the marginal likelihoods. Details regarding the computation of the marginal likelihood are given in Appendix D.

Table 2 contains the marginal likelihood estimates for our six contending models. As can be seen, the $\mathcal{M}_3$ is most supported by the data. We now provide more detailed results for this model.

**Table 2** Loglikelihood (lnL), log marginal likelihood (lnML), numerical standard error (n.s.e.) posterior probability of each model ($\Pr[\mathcal{M}_m|\mathbf{y}]$) under the assumption that the prior probability of each model is 1/6, and change-point estimates

| Model | lnL | lnML | n.s.e. | $\Pr[\mathcal{M}_m\|\mathbf{y}]$ | Change-point |
|---|---|---|---|---|---|
| $\mathcal{M}_0$ | −1488.1 | −1215.5 | 1.39 | 0.00 | |
| $\mathcal{M}_1$ | −1279.4 | −955.5 | 1.77 | 0.00 | 1986:II |
| $\mathcal{M}_2$ | −935.1 | −665.4 | 1.92 | 0.00 | 1985:IV, 1995:II |
| $\mathcal{M}_3$ | −473.4 | −256.1 | 2.27 | 1.00 | 1980:II, 1985:IV, 1995:II |
| $\mathcal{M}_4$ | −313.8 | −281.4 | 2.62 | 0.00 | 1980:II, 1985:IV, 1995:II, 2002:III |
| $\mathcal{M}_{MS}$ | −358.1 | −279.1 | 3.22 | 0.00 | |

Our first set of findings relate to the timing of the change-points. Information about the change-points is gleaned from the sampled sequence of the states. Further details about how this is done can be obtained from Chib (1998). Of particular interest are the posterior probabilities of the timing of the regime changes. These probabilities are given in Figure 1. The figure reveals that the first thirty-two quarters (the first 8 years) belong to the first regime, the next twenty-three quarters (about 6 years) to the second, the next thirty-eight quarters (about 9.5 years) to the third, and the remaining quarters to the 4th regime. Rudebusch and Wu (2008) also find a change-point in the year of 1985. The finding of a break point in 1995 is striking as it has not been isolated from previous regime change models.

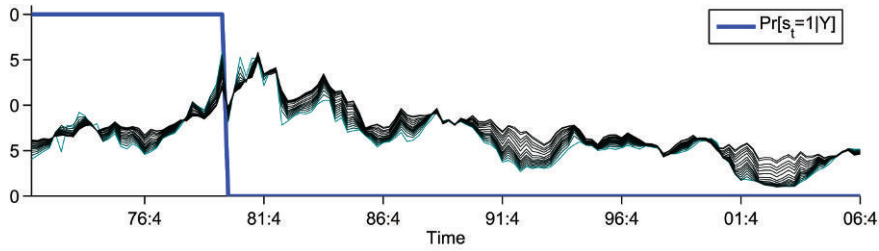
We would like to mention that our estimates of the change-points from the one-latent factor model without macro factors are exactly the same as those from the change point models with macro factors. Therefore, the macro factors do not seem to drive the regime changes. Nonetheless, the general model with macro factors outperforms the one-latent factor model in terms of the out-of-sample forecasts of the term structure. We do not report these results in the interest of space. In addition, none of our results are sensitive to our choice of sixteen maturities, as we have confirmed.

### 3.3 Parameter Estimates

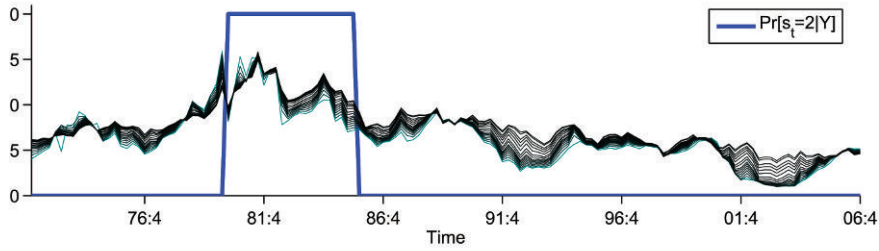
Table 3 summarizes the posterior distribution of the parameters. One point to note is that the posterior densities are generally different from the prior given in Section 4, which implies that the data are informative about these parameters. We focus on various aspects of this posterior distribution in the subsequent subsections. From the estimates of the regime-specific parameters, we can infer the sources of structural changes characterizing the regimes.

**3.3.1 Factor process.** Figure 2 plots the average dynamics of the latent factors along with the short rate. This figure demonstrates that the latent factor movements are very close to those of the short rate. The estimates of the matrix $\mathbf{G}$ for each regime show that the mean reversion coefficient matrix is almost diagonal. The latent factor and inflation rate also display different degrees of persistence across regimes. In particular, the relative magnitudes of the diagonal elements indicate that the latent factor and the inflation factor are less mean reverting in regime 2 and 4, respectively. For a more formal measure of this persistence, we calculate the eigenvalues of the coefficient matrices in each regime. These are given by

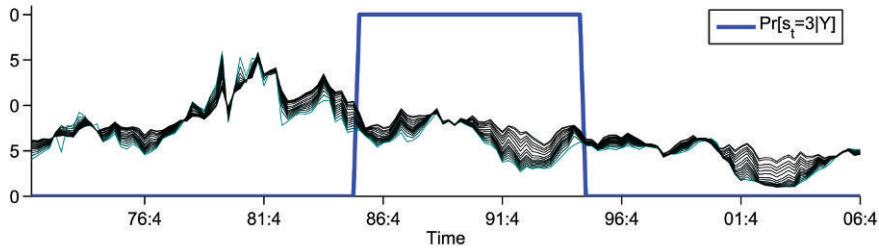
$$\text{eig}(\mathbf{G}_1) = \begin{bmatrix} 0.851 \\ 0.709 \\ 0.267 \end{bmatrix}, \ \text{eig}(\mathbf{G}_2) = \begin{bmatrix} 0.978 \\ 0.814 \\ 0.401 \end{bmatrix}$$



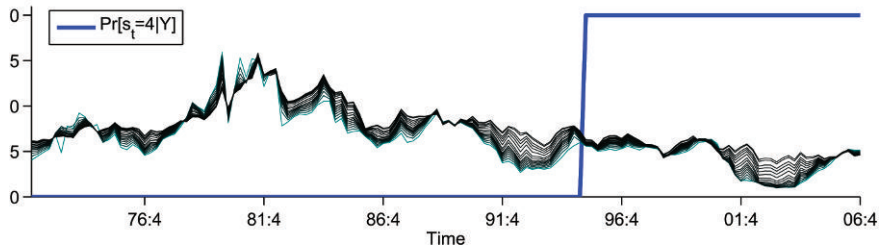

(a) $s_t = 1$




(b) $s_t = 2$




(c) $s_t = 3$




(d) $s_t = 4$

**Figure 1** Model $\mathcal{M}_3$: $\Pr(s_t = j|\mathbf{y})$. The posterior probabilities for each $t$ are based on 50,000 MCMC draws of $s_t$—these probabilities are plotted along with the sixteen yields in annualized percents (probabilities are multiplied by 20 for legibility).

**Table 3** Model $\mathcal{M}_3$: Parameter estimates

| | Regime 1 | | | Regime 2 | | | Regime 3 | | | Regime 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0.90** | 0.07 | 0.15 | **0.95** | −0.01 | 0.03 | **0.92** | 0.15 | 0.31 | **0.93** | 0.04 | 0.23 |
| | (0.06) | (0.10) | (0.15) | (0.03) | (0.07) | (0.06) | (0.06) | (0.21) | (0.17) | (0.04) | (0.17) | (0.29) |
| **G** | −0.24 | **0.67** | −0.07 | −0.07 | **0.73** | −0.10 | **0.15** | **0.35** | 0.08 | 0.02 | **0.91** | 0.01 |
| | (0.26) | (0.23) | (0.12) | (0.05) | (0.05) | (0.03) | (0.06) | (0.14) | (0.08) | (0.02) | (0.13) | (0.06) |
| | −0.06 | −0.16 | 0.26 | 0.09 | −0.35 | **0.52** | −0.04 | 0.00 | **0.34** | −0.03 | −0.37 | 0.19 |
| | (0.25) | (0.23) | (0.17) | (0.17) | (0.24) | (0.17) | (0.09) | (0.21) | (0.13) | (0.08) | (0.26) | (0.15) |
| $\mu$ | 0.00 | **4.99** | 3.54 | 0.00 | **5.88** | **2.63** | 0.00 | **2.56** | **2.62** | 0.00 | **1.49** | **3.22** |
| $\times 400$ | | (2.17) | (0.90) | | (0.41) | (1.00) | | (0.41) | (0.49) | | (0.80) | (0.53) |
| | 1.00 | | | 1.00 | | | 1.00 | | | 1.00 | | |
| **L** | 0.11 | **1.72** | | 0.10 | **1.48** | | 0.11 | **0.74** | | −0.47 | **0.82** | |
| $\times 400$ | (0.40) | (0.19) | | (0.44) | (0.13) | | (0.34) | (0.13) | | (0.59) | (0.12) | |
| | −0.67 | −0.62 | **4.28** | 0.24 | 0.27 | **4.58** | −0.55 | −0.18 | **2.00** | −0.13 | −0.20 | **2.03** |
| | (0.88) | (0.39) | (0.14) | (0.62) | (0.41) | (0.17) | (0.56) | (0.14) | (0.12) | (0.89) | (0.14) | (0.11) |
| $\delta_1$ | | **9.23** | | | **2.78** | | | **4.42** | | | **4.34** | |
| $\times 400$ | | (1.69) | | | (1.60) | | | (1.18) | | | (1.00) | |
| $\boldsymbol{\delta}_2$ | **1.16** | 0.09 | 0.17 | **1.29** | 0.25 | 0.16 | **0.72** | 0.31 | 0.26 | **0.57** | 0.56 | 0.10 |
| | (0.13) | (0.23) | (0.22) | (0.16) | (0.23) | (0.15) | (0.09) | (0.26) | (0.21) | (0.07) | (0.37) | (0.25) |
| $\boldsymbol{\gamma}$ | −0.28 | −0.40 | −0.22 | −0.34 | −0.65 | −0.21 | −0.58 | −0.56 | −0.05 | −0.34 | −0.25 | −0.19 |
| | (0.28) | (0.30) | (0.26) | (0.25) | (0.21) | (0.26) | (0.28) | (0.33) | (0.24) | (0.25) | (0.25) | (0.27) |
| $\boldsymbol{\Phi}$ | 0.99 | 0.98 | 0.93 | 0.53 | 0.89 | 0.65 | 0.91 | 0.94 | 0.98 | 0.98 | 0.93 | 0.98 |
| | (1.08) | (1.09) | (1.08) | (1.07) | (1.08) | (1.12) | (1.08) | (1.09) | (1.09) | (1.09) | (1.10) | (1.09) |
| $p_{00}$ | | | | | | | **0.934** | | | | | |
| | | | | | | | 0.028 | | | | | |
| $p_{11}$ | | | | | | | **0.986** | | | | | |
| | | | | | | | 0.004 | | | | | |
| $p_{12}$ | | | | | | | **0.987** | | | | | |
| | | | | | | | **0.003** | | | | | |

This table presents the posterior mean and standard deviation based on 50,000 MCMC draws beyond a burn-in of 5000. The 95% credibility interval of parameters in boldface does not contain 0. Standard deviations are in parenthesis. The yields are of 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 20, 24, 28, 36, and 40 quarters maturity bonds. Values without standard deviations are fixed by the identification restrictions.

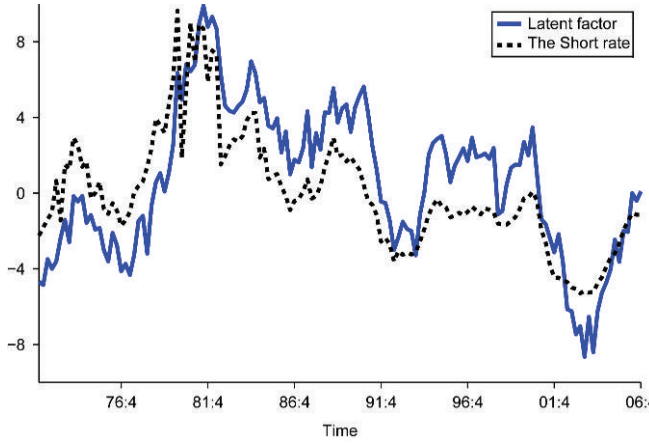



**Figure 2** Model $\mathcal{M}_3$: Estimates of the latent factor. The short rate in percent is demeaned and estimates of the latent factor are calculated as the average of factor drawings given the 50,000 MCMC draws of the parameters.

$$\text{eig}(\mathbf{G}_3) = \begin{bmatrix} 0.935 \\ 0.312 \\ 0.366 \end{bmatrix}, \ \text{eig}(\mathbf{G}_4) = \begin{bmatrix} 0.913 + 0.044i \\ 0.913 - 0.044i \\ 0.204 \end{bmatrix}.$$

It can be seen that the second regime has the largest absolute eigenvalue close to 1. Because the factor loadings for the latent factor $(\delta_{21,s_t})$ are significant whereas those for inflation $(\delta_{22,s_t})$ are not, the latent factor is responsible for most of the persistence of the yields.

Furthermore, the diagonal elements of $\mathbf{L}_3$ and $\mathbf{L}_4$ are even smaller than their counterparts in $\mathbf{L}_1$ and $\mathbf{L}_2$. This suggest a reduction in factor volatility starting from the middle of the 1980s, which coincides with the period that is called the great moderation (Kim, Nelson, and Piger 2004).

**3.3.2 Factor loadings.** The factor loadings in the short rate equation, $\boldsymbol{\delta}_{2,s_t}$, are all positive, which is consistent with the conventional wisdom that central bankers tend to raise the interest rate in response to a positive shock to the macro factors. It can also be seen that $\boldsymbol{\delta}_{2,s_t}$ along with $\mathbf{G}_{s_t}$ and $\mathbf{L}_{s_t}$ are different across regimes, which makes the factor loadings regime dependent across the term structure as revealed in Figure 3. This finding lends support to our assumption of regime-dependent factor loadings.

**3.3.3 Term premium.** Figure 4 plots the posterior distribution of the term premium of the two-year maturity bond over time. It is interesting to observe how the

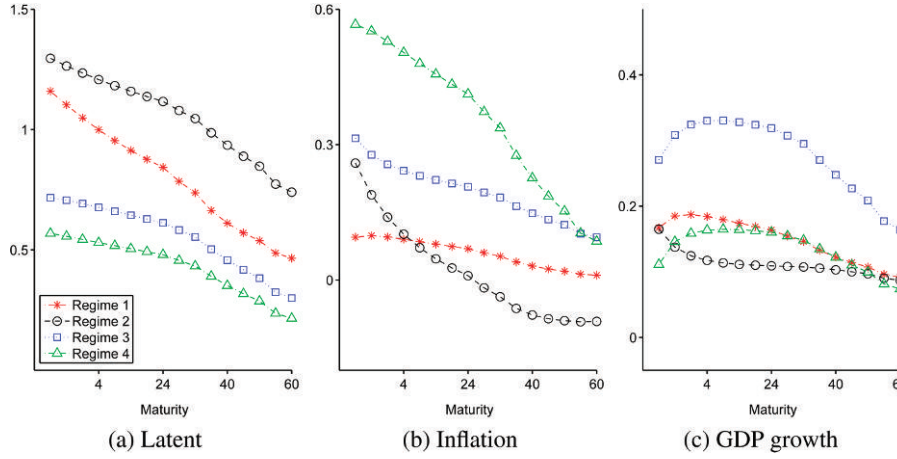



**Figure 3** Model $\mathcal{M}_3$: Estimates of the factor loadings, $\bar{\mathbf{b}}_{s_t}$. The factor loadings represent the average simulated factor loadings from the retained 50,000 MCMC iterations.

term premium varies across regimes. In particular, the term premium is the lowest in the most recent regime (although the 0.025 quantile of the term premium distribution in the first regime is lower than the 0.025 quantile of term premium distribution in the most current regime). This can be attributed to the lower value of factor volatilities in this regime. Moreover, we find that these changes in the term premium are not closely related to changes in the latent and macroeconomic factors, although the parameters in $\mathbf{\Phi}_{s_t}$ tend to be less informed by the data due to the high persistence of the factors. A similar finding appears in Rudebusch, Sack, and Swanson (2007).

**3.3.4 Pricing error volatility.** In Figure 5, we plot the term structure of the pricing error standard deviations. As in the no-change-point model of Chib and Ergashev (2009), these are hump shaped in each regime. One can also see that these standard deviations have changed over time, primarily for the short bonds. These changes in the volatility also help to determine the timing of the change-points.

## 3.4 Forecasting and Predictive Densities

From the posterior distribution of the parameters and regimes, we can confirm that the U.S. yield curve underwent three regime changes and that the various aspects including the factor loadings and the term premium served as sources characterizing the regimes. Now, we find their implications in improving predictive accuracy, which is the principle objective of this paper. To show this, we compare the forecasting abilities of the affine term structure models with and without regime

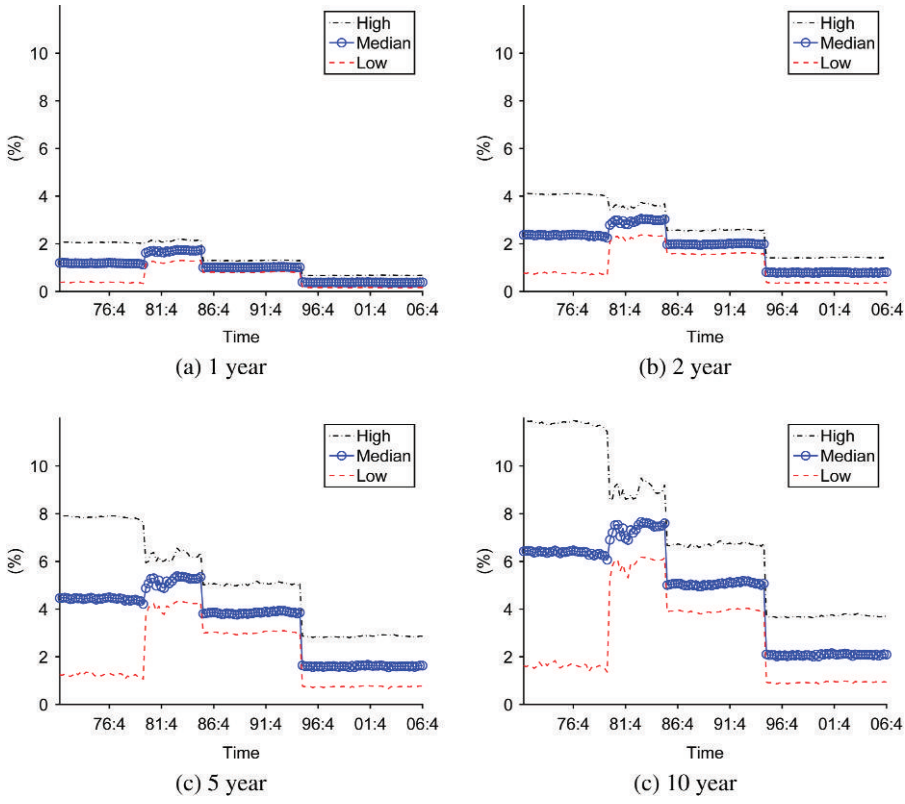



**Figure 4** Model $\mathcal{M}_3$: Term premium. The figure plots the 2.5, 50, and 97.5% quantile of the posterior term premium based on 50,000 MCMC draws beyond a burn-in of 5000 iterations.

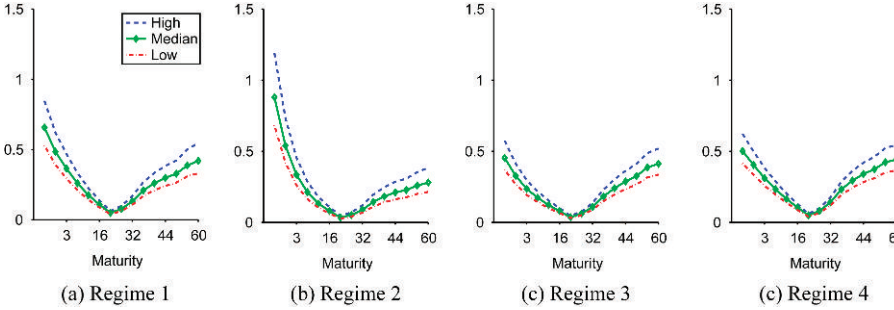



**Figure 5** Model $\mathcal{M}_3$: Term Structure of the Pricing Error Volatility. The figures display the 2.5, 50, and 97.5% quantile of the posterior standard deviation of the pricing errors.

changes. In the Bayesian paradigm, it is relatively straightforward to simulate the predictive density from the MCMC output. By definition, the predictive density of the future observations, conditional on the data, is the integral of the density of the future outcomes given the parameters with respect to the posterior distribution of the parameters. If we let $\mathbf{y}_f$ denote the future observations, the predictive density under model $\mathcal{M}_m$ is given by

$$p(\mathbf{y}_f|\mathcal{M}_m, \mathbf{y}) = \int_{\psi} p(\mathbf{y}_f|\mathcal{M}_m, \mathbf{y}, \psi)\pi(\psi|\mathcal{M}_m, \mathbf{y})d\psi. \tag{28}$$

This density can be sampled by the method of composition as follows. For each MCMC iteration (beyond the burn-in period), conditioned on $\mathbf{f}_n$ and the parameters in the current terminal regime (which is not necessarily regime $m+1$), we draw the factors $\mathbf{f}_{n+1}$ based on the Equation (2). Then, given $\mathbf{f}_{n+1}$, the yields $\mathbf{R}_{n+1}$ are drawn using Equation (19). These two steps are iterated forward to produce

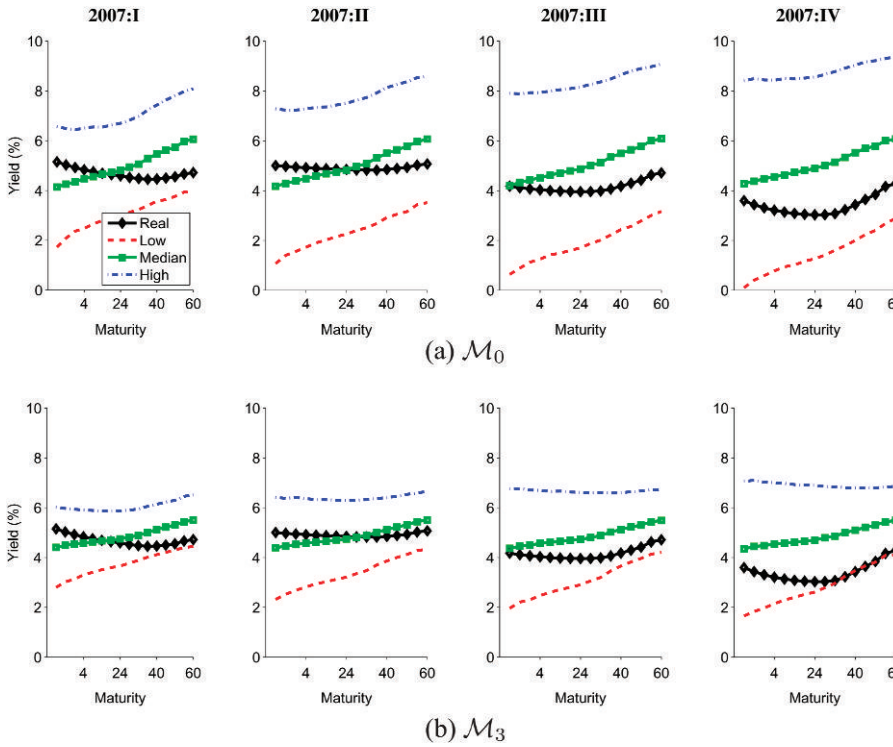



**Figure 6** Predicted yield curve. The figures present four quarters ahead forecasts of the yields on the T-bills. The top panel is based on the no-change-point model and the bottom panel on the three change-point model. In each case, the 2.5, 50, and 97.5% quantile curves are based on 50,000 forecasted values for the period 2007:I–2007:IV. The observed curves are labeled "Real."

**Table 4** PPC

| Model | $M_0$ | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_{\text{MS}}$ | $M_0$ |
|---|---|---|---|---|---|---|---|
| Sample period | (1972:I–2006:IV) | | | | | | (1995:II–2006:IV) |
| $D$ | 12.548 | 5.401 | 4.156 | 4.720 | 4.599 | 15.202 | 4.011 |
| $W$ | 5.678 | 4.896 | 4.201 | 3.415 | 2.902 | 7.397 | 3.520 |
| PPC | 18.226 | 10.297 | 8.357 | 8.126 | 7.501 | 22.598 | 7.531 |
| (a) Forecast period: 2007:I–2007:IV | | | | | | | |
| Model | $M_0$ | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_{\text{MS}}$ | $M_0$ |
| Sample period | (1972:I–2005:IV) | | | | | | (1995:II–2005:IV) |
| $D$ | 12.606 | 5.799 | 4.157 | 4.097 | 7.011 | 16.634 | 4.271 |
| $W$ | 2.137 | 5.658 | 4.432 | 1.817 | 3.036 | 2.969 | 2.390 |
| PPC | 14.743 | 11.457 | 8.589 | 5.914 | 10.047 | 19.602 | 6.661 |
| (b) Forecast period: 2006:I–2006:IV | | | | | | | |
| Model | $M_0$ | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_{MS}$ | $M_0$ |
| Sample period | (1972:I–2004:IV) | | | | | | (1995:II–2004:IV) |
| $D$ | 13.474 | 5.187 | 3.572 | 4.609 | 7.190 | 18.748 | 3.919 |
| $W$ | 2.367 | 5.787 | 4.442 | 1.977 | 2.657 | 2.303 | 2.359 |
| $PPC$ | 15.841 | 10.974 | 8.014 | 6.587 | 9.847 | 21.051 | 6.278 |
| (c) Forecast period: 2005:I–2005:IV | | | | | | | |

PPC is computed by Equations (24)–(26).

the draws $\mathbf{f}_{n+i}$ and $\mathbf{R}_{n+i}$, $i = 1, 2, \ldots, T$. Repeated over the course of the MCMC iterations, these steps produce a collection of simulated macro factors and yields, which is a sample from the predictive density. Note that for model $M_{\text{MS}}$, the future regime $s_{n+i}$ is sequentially simulated conditioned on $s_{n+i-1}$ before drawing $\mathbf{f}_{n+i}$ (Albert and Chib 1993).

We summarize the sampled predictive densities in Figure 6. The top panel gives the forecast intervals from the $\mathcal{M}_0$ model and the bottom panel has the forecast intervals from the $\mathcal{M}_3$ model. Note that in both cases, the actual yield curve in each of the four quarters of 2007 is bracketed by the corresponding 95% credibility interval, though the intervals from the $\mathcal{M}_3$ model are tighter.

For a more formal forecasting performance comparison, we tabulate the PPC for each case in Table 4. We also include in the last column of this table an interesting set of results that make use of the regimes isolated by our $\mathcal{M}_3$ model. In particular, we fit the no-change-point model to the data in the last regime but ending just before our different forecast periods (2005:I-2005:IV, 2006:I-2006:IV,

and 2007:I-2007:IV). As one would expect, the forecasts from the no-change-point model estimated on the sample period of the last regime are similar to those from the $\mathcal{M}_3$ model. Thus, given the regimes we have isolated, an informal approach to forecasting the term structure would be to fit the no-change arbitrage-free yield model to the last regime. Of course, the predictions from the $\mathcal{M}_3$ model produce a smaller value of the PPC than those from the no-change-point model that is fit to the whole sample. This, combined with the in-sample fit of the models as measured by the marginal likelihoods, suggests that the change-point model outperforms the no-change-point version. Finally, it is noticeable that the performance of the $\mathcal{M}_{\text{MS}}$ model is worse than the no-change-point model. The assumption that the Markov switching observed in the past would persist into the future is apparently a worse assumption than assuming that the current regime would persist into the future. These findings not only reaffirm the finding of structural changes but also suggest that there are gains to incorporating change-points when forecasting the term structure of interest rates.

## 4 CONCLUDING REMARKS

In this paper, we have developed a new model of the term structure of zero-coupon bonds with regime changes. This paper complements the recent developments in this area because it is organized around a different model of regime changes than the Markov switching model that has been used to date. It also complements the recent work on affine models with macro factors, which has been done in settings without regime changes. Furthermore, we incorporate some recent developments in Bayesian econometrics that make it possible to estimate the large-scale models in this paper.

Our empirical analysis demonstrates that three change-points characterize the data well, and that the term structure and the risk premium are materially different across regimes. We also show that out-of-sample forecasts of the term structure improve from incorporating regime changes.

## APPENDIX A: BOND PRICES UNDER REGIME CHANGES

By the law of the iterated expectation, the risk-neutral pricing formula in Equation (7) can be expressed as

$$1 = \mathbb{E}_t \left\{ \mathbb{E}_{t,s_{t+1}} \left[ \kappa_{t,s_t,t+1} \frac{P_{t+1}(s_{t+1}, \tau - 1)}{P_t(s_t, \tau)} \right] \right\}, \tag{A1}$$

where the inside expectation $\mathbb{E}_{t,s_{t+1}}$ is conditioned on $s_{t+1}$, $s_t$ and $\mathbf{f}_t$. Subsequently, as discussed below, one now substitutes $P_t(s_t, \tau)$ and $P_{t+1}(s_{t+1}, \tau - 1)$ from Equations (8) and (9) into this expression and integrate out $s_{t+1}$ after a log-linearization. We match common coefficients and solve for the unknown functions. The detailed procedures are as follows. By the assumption of the affine model, we have

$$P_t(s_t,\tau) = \exp(-a_{s_t}(\tau) - \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})) \tag{A2}$$

$$\text{and } P_{t+1}(s_{t+1},\tau-1) = \exp(-a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}})).$$

Let $h_{\tau,t+1}$ denote

$$\frac{P_{t+1}(s_{t+1},\tau-1)}{P_t(s_t,\tau)} = \exp\big[ - a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}})$$

$$+a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})\big]. \tag{A3}$$

It immediately follows from the bond pricing formula that

$$1 = \mathbb{E}_t\left[\kappa_{t,s_t,t+1}\frac{P_{t+1}(s_{t+1},\tau-1)}{P_t(s_t,\tau)}\right] \tag{A4}$$

$$= \mathbb{E}_t[\kappa_{t,s_t,t+1}h_{\tau,t+1}].$$

Then, by substitution

$$\kappa_{t,s_t,t+1}h_{\tau,t+1} \tag{A5}$$

$$= \exp\Big[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}_{t,s_t}'\boldsymbol{\gamma}_{t,s_t} - \boldsymbol{\gamma}_{t,s_t}'\mathbf{L}_{s_{t+1}}^{-1}\boldsymbol{\eta}_{t+1}$$

$$- a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'\left(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}}\right) + a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\Big]$$

$$= \exp\left[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}_{t,s_t}'\boldsymbol{\gamma}_{t,s_t} - \left(\boldsymbol{\gamma}_{t,s_t}'\mathbf{L}_{s_{t+1}}^{-1} + \mathbf{b}_{s_{t+1}}(\tau-1)'\right)\boldsymbol{\eta}_{t+1} + \zeta_{\tau,s_t,s_{t+1}}\right]$$

$$= \exp\left[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}_{t,s_t}'\boldsymbol{\gamma}_{t,s_t} - \left(\boldsymbol{\gamma}_{t,s_t} + \mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\right)\boldsymbol{!}_{t+1} + \zeta_{\tau,s_t,s_{t+1}}\right]$$

$$= \exp\left[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}_{t,s_t}'\boldsymbol{\gamma}_{t,s_t} + \frac{1}{2}\boldsymbol{\Gamma}_{t,\tau}\boldsymbol{\Gamma}_{t,\tau}' + \zeta_{\tau,s_t,s_{t+1}}\right]\exp\left[-\frac{1}{2}\boldsymbol{\Gamma}_{t,\tau}\boldsymbol{\Gamma}_{t,\tau}' - \boldsymbol{\Gamma}_{t,\tau}\boldsymbol{!}_{t+1}\right],$$

where

$$\zeta_{\tau,s_t,s_{t+1}} = a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) - a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{G}_{s_{t+1}}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}),$$

$$\boldsymbol{\Gamma}_{t,\tau} = \boldsymbol{\gamma}_{t,s_t}' + \mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}},$$

and $\boldsymbol{\omega}_{t+1} = \mathbf{L}_{s_{t+1}}^{-1}\eta_{t+1} \sim \mathcal{N}(0,\mathbf{I}_{k+m})$. Given $\mathbf{f}_t, s_{t+1}$ and $s_t$, the only random variable in $\kappa_{t,t+1}h_{\tau,t+1}$ is $\boldsymbol{\omega}_{t+1}$. Then, since

$$\mathbb{E}_t\left(\exp\left[-\frac{1}{2}\boldsymbol{\Gamma}_{t,\tau}\boldsymbol{\Gamma}_{t,\tau}' - \boldsymbol{\Gamma}_{t,\tau}\boldsymbol{!}_{t+1}\right]\right) = 1, \tag{A6}$$

we have that

$$\mathbb{E}\left[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1},s_t\right]=\exp\left[-r_{t,s_t}-\frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t}+\frac{1}{2}\boldsymbol{\Gamma}_{t,\tau}\boldsymbol{\Gamma}'_{t,\tau}+\zeta_{\tau,s_t,s_{t+1}}\right].$$

Using log-approximation $\exp(y)\approx y+1$ for a sufficiently small $y$ leads to

$$\begin{aligned}
&\mathbb{E}[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1},s_t]\\
&=\exp\Big[-r_{t,s_t}-\frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t}+\frac{1}{2}\big(\boldsymbol{\gamma}'_{t,s_t}+\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\big)\big(\boldsymbol{\gamma}'_{t,s_t}+\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\big)'\\
&\quad+\zeta_{\tau,s_t,s_{t+1}}\Big]\approx-r_{t,s_t}+\boldsymbol{\gamma}'_{t,s_t}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\\
&\quad+\frac{1}{2}\big(\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\big)+\zeta_{\tau,s_t,s_{t+1}}+1\\
&=-\left(\delta_{1,s_t}+\boldsymbol{\delta}'_{2,s_t}(\mathbf{f}_t-\boldsymbol{\mu}_{s_t})\right)+\left(\tilde{\boldsymbol{\gamma}}_{s_t}+\Phi_{s_t}(\mathbf{f}_t-\boldsymbol{\mu}_{s_t})\right)'\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\\
&\quad+\frac{1}{2}\big(\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\big)+\zeta_{\tau,s_t,s_{t+1}}+1. \qquad \text{(A7)}
\end{aligned}$$

Given the information at time $t$ (i.e., $\mathbf{f}_t$ and $s_t=j$), integrating out $s_{t+1}$ yields

$$\begin{aligned}
\mathbb{E}[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_t=j]&=\sum_{s_{t+1}=j,k}p_{js_{t+1}}\mathbb{E}[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1},s_t=j] \qquad \text{(A8)}\\
&=1,\ \text{where } k=j+1.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
0&=\sum_{s_{t+1}=j,k}p_{js_{t+1}}\left\{\mathbb{E}\left[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1},s_t=j\right]-1\right\}\ \text{since}\ \sum_{s_{t+1}=j,k}p_{js_{t+1}}=1\\
&=p_{jj}(\mathbb{E}[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1}=j,s_t=j]-1)\\
&\quad+p_{jk}(\mathbb{E}[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t,s_{t+1}=k,s_t=j]-1)\\
&\approx-p_{jj}(\delta_{1,j}+\boldsymbol{\delta}'_{2,j}(\mathbf{f}_t-\boldsymbol{\mu}_{s_t}))+p_{jj}(\tilde{\boldsymbol{\gamma}}_j+\Phi_j(\mathbf{f}_t-\boldsymbol{\mu}_{s_t}))'\mathbf{L}'_j\mathbf{b}_j(\tau-1)\\
&\quad+\frac{1}{2}p_{jj}(\mathbf{b}_j(\tau-1)'\mathbf{L}_j\mathbf{L}'_j\mathbf{b}_j(\tau-1))+p_{jj}\zeta_{\tau,j,j}\\
&\quad-p_{jk}(\delta_{1,j}+\boldsymbol{\delta}'_{2,j}(\mathbf{f}_t-\boldsymbol{\mu}_{s_t}))+p_{jk}(\tilde{\boldsymbol{\gamma}}_j+\Phi_j(\mathbf{f}_t-\boldsymbol{\mu}_{s_t}))'\mathbf{L}'_k\mathbf{b}_k(\tau-1)\\
&\quad+\frac{1}{2}p_{jk}(\mathbf{b}_k(\tau-1)'\mathbf{L}_k\mathbf{L}'_k\mathbf{b}_k(\tau-1))+p_{jk}\zeta_{\tau,j,k}. \qquad \text{(A9)}
\end{aligned}$$

Matching the coefficients on $\mathbf{f}_t$ and setting the constant terms equal to zero, we obtain the recursive equation for $a_{s_t}(\boldsymbol{\omega})$ and $\mathbf{b}_{s_t}(\tau)$ given the initial conditions

$a_{s_t}(0) = 0$ and $\mathbf{b}_{s_t}(0) = \mathbf{0}_{3\times 1}$ implied by the no-arbitrage condition. Finally, imposing the restrictions on the transition probabilities establishes the proof.

## APPENDIX B: PRIOR DISTRIBUTION

We begin by recalling the identifying restrictions on the parameters. First, we set $\boldsymbol{\mu}_{u,s_t} = 0$, which implies that the mean of the short rate conditional on $s_t$ is $\delta_{1,s_t}$. Next, the first element of $\boldsymbol{\delta}_{2,s_t}$, namely $\delta_{21,s_t}$, is assumed to be nonnegative. Finally, to enforce stationarity of the factor process, we restrict the eigenvalues of $\mathbf{G}_{s_t}$ to lie inside the unit circle. Thus, under the physical measure, the factors are mean reverting in each regime. These constraints are summarized as

$$\mathcal{R} = \left\{\mathbf{G}_j, \delta_{21,j} | \delta_{21,j} \geqslant 0,\ 0 \leqslant p_{jj} \leqslant 1, \quad \left|\text{eig}(\mathbf{G}_j)\right| < 1 \text{ for } j = 1, 2, .., m+1\right\}. \quad \text{(B1)}$$

All the constraints in $\mathcal{R}$ are enforced through the prior distribution.

We now provide the specific prior distributions as follows. The free parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^{*2}$ are assumed to be mutually independent. Our prior distribution on $\boldsymbol{\theta}$ is normal $\mathcal{N}(\bar{\boldsymbol{\theta}}, \bar{\mathbf{V}}_{\boldsymbol{\theta}})$ truncated by the restrictions in $\mathcal{R}$. In particular, the $\mathcal{N}(\bar{\boldsymbol{\theta}}, \bar{\mathbf{V}}_{\boldsymbol{\theta}})$ distribution has the form

$$\begin{aligned}
&\prod_{s_t=1}^{m} \mathcal{N}(p_{s_t s_t} | \bar{p}_{s_t s_t}, \bar{\mathbf{V}}_{p_{s_t s_t}}) \\
&\times \prod_{s_t=1}^{m+1} \left\{\mathcal{N}(\mathbf{g}_{s_t} | \bar{\mathbf{g}}_{s_t}, \bar{\mathbf{V}}_{\mathbf{g}_{s_t}}) \mathcal{N}(\boldsymbol{\mu}_{m,s_t} | \bar{\boldsymbol{\mu}}_{m,s_t}, \bar{\mathbf{V}}_{\boldsymbol{\mu}_{m,s_t}}) \mathcal{N}(\boldsymbol{\delta}_{s_t} | \bar{\boldsymbol{\delta}}_{s_t}, \bar{\mathbf{V}}_{\boldsymbol{\delta}_{s_t}})\right\} \\
&\times \prod_{s_t=1}^{m+1} \left\{\mathcal{N}(\tilde{\gamma}_{s_t} | \bar{\gamma}_{s_t}, \bar{\mathbf{V}}_{\tilde{\gamma}_{s_t}}) \mathcal{N}(\phi_{s_t} | \bar{\phi}_{s_t}, \bar{\mathbf{V}}_{\phi_{s_t}}) \mathcal{N}(\breve{}_{s_t} | \bar{}_{s_t}, \bar{\mathbf{V}}_{\breve{}_{s_t}})\right\},
\end{aligned}$$

which we explain as follows.

*First*, the prior on $p_{jj}$ $(j = 1, \ldots, m)$ is truncated normal, truncated to the interval $(0, 1)$, with a standard deviation of 0.28. The mean of these distributions is model specific. For example, in the $\mathcal{M}_1$ model, the mode is 0.986, so that the a priori expected duration of stay in regime 1 is about 70 quarters in relation to a sample period of 140 quarters. In the $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ models, the prior mean of the transition probabilities is specified to imply 50, 40, and 33 quarters of expected duration in each regime. It is important to note that we work with a truncated normal prior distribution on these transition probabilities instead of the more conventional beta distribution because $\bar{\mathbf{a}}_{s_t}$ and $\bar{\mathbf{b}}_{s_t}$ in the Equation (19) are a function of $p_{jj}$, which eliminates any benefit from the use of a beta functional form. *Second*, we construct a $9 \times 1$ vector $\bar{\mathbf{g}}_{s_t}$ from the matrix

$$\bar{\mathbf{G}}_{s_t} = \text{diag}(0.95, 0.8, 0.4),$$

and let $\bar{\mathbf{V}}_{\mathbf{g}_{s_t}}$ be a $9 \times 9$ diagonal matrix with each diagonal element equal to 0.1. This choice of prior incorporates the prior belief that the latent factor is more persistent

than the macro factors. *Third*, we assume that $\bar{\boldsymbol{\mu}}_{m,s_t} \times 400 = (4,3)'$ and $\bar{\mathbf{V}}_{\boldsymbol{\mu}_{m,s_t}} \times 400^2 = \text{diag}(25,1)$. Thus, the prior mean of inflation is assumed to be 4% and that of real GDP growth rate to be 3%. The standard deviations of 5 and 1% produce a distribution that covers the most likely values of these rates. *Fourth*, based on the Taylor rule intuition that the response of the short rate to an increase of inflation and output growth tend to be positive, we let

$$\bar{\boldsymbol{\delta}}_{s_t} = (6, 0.8, 0.4, 0.4)',$$

and the let the prior standard deviations be $(5, 0.4, 0.4, 0.4)$. *Fifth*, we assume that

$$\bar{\gamma}_{s_t} = (-0.5, -0.5, -0.5)' \quad \text{and} \quad \bar{\mathbf{V}}_{\tilde{\gamma}_{s_t}} = \text{diag}(0.1, 0.1, 0.1),$$

where the prior mean of $\tilde{\gamma}_{s_t}$ is negative in order to suggest an upward sloping average yield curve in each regime. *Sixth*, we assume that

$$\bar{\phi}_{s_t} = (1,1,1)' \quad \text{and} \quad \bar{\mathbf{V}}_{\phi_{s_t}} = \text{diag}(1,1,1),$$

where the positive prior is justified from the intuition that positive shocks to macroeconomic fundamentals should tend to decrease the overall risk in the economy. *Seventh*, we let

$$\bar{\lambda}_{s_t} = (0,0,0,0,1)' \quad \text{and} \quad \bar{\mathbf{V}}_{s_t} = \text{diag}(4,4,4,4,4),$$

so that the prior of $\mathbf{L}_{s_t}$ is relatively weak. This leads to considerable prior variation in the implied yield curve.

Next, we place the prior on the $15 \times (m+1)$ free parameters of $\boldsymbol{\sigma}^{*2}$. Each $\sigma^{*2}_{i,s_t}$ is assumed to have an inverse gamma prior distribution $\mathcal{IG}(\bar{v}, \bar{d})$ with $\bar{v} = 4.08$ and $\bar{d} = 20.80$, which implies a mean of 10 and standard deviation of 14.

Finally, we assume that the latent factor $u_0$ at time 0 follows the steady-state distribution in regime 1

$$u_0 \sim \mathcal{N}(0, V_u), \tag{B2}$$

where $V_u = \left(1 - G^2_{11,1}\right)^{-1}$.

To show what these assumptions imply for the outcomes, we simulate the parameters 50,000 times from the prior, and for each drawing of the parameters, we simulate the factors and yields for each maturity and each of fifty quarters. The median, 2.5% quantile, and 97.5% quantile surfaces of the resulting term structure in annualized percents are reproduced in Figure 7. Because our prior distribution is symmetric among the regimes, the prior distribution of the yield curve is not regime specific. It can be seen that the simulated prior term structure is gently upward sloping on average. Also the assumed prior allows for considerable a priori variation in the term structure.

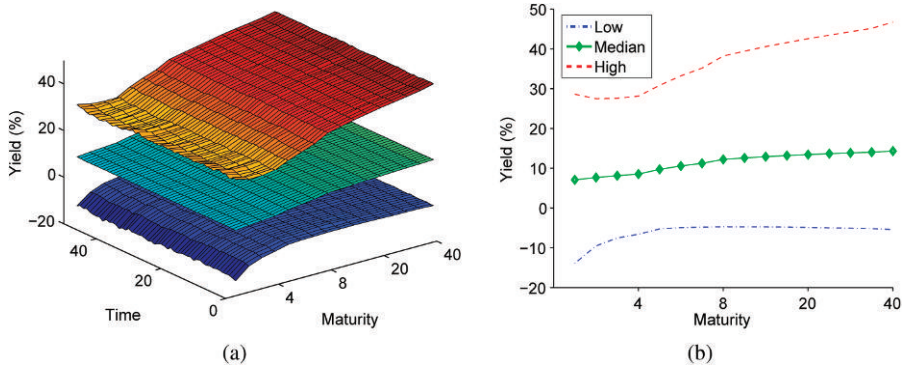



**Figure 7** The implied prior term structure dynamics. These graphs are based on 50,000 simulated draws of the parameters from the prior distribution. In the graphs on the left, the "Low," "Median," and "High" surfaces correspond to the 2.5, 50, and 97.5% quantile surfaces of the term structure dynamics in annualized percents implied by the prior distribution. In the second graph, the surfaces of the first graph are averaged over the entire period of fifty quarters.

## APPENDIX C: POSTERIOR DISTRIBUTION AND MCMC SAMPLING

The posterior distribution that we would like to explore is given by

$$\pi(\mathbf{S}_n, \psi|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{S}_n, \psi)p(\mathbf{S}_n|\psi)\pi(\psi), \tag{C1}$$

where $p(\mathbf{y}|\mathbf{S}_n, \psi)$ is the distribution of the data given the regime indicators and the parameters, $p(\mathbf{S}_n|\psi)$ is the density of the regime indicators given the parameters and the initial latent factor, and $\pi(\psi)$ is the joint prior density of $u_0$ and the parameters. Note that by conditioning on $\mathbf{S}_n$, we avoid the calculation of the likelihood function $p(\mathbf{y}|\psi)$ whose computation is more involved. We discuss the computation of the likelihood function in the next section in connection with the calculation of the marginal likelihood.

The idea behind the MCMC approach is to sample this posterior distribution iteratively, such that the sampled draws form a Markov chain with invariant distribution given by the target density. Practically, the sampled draws after a suitably specified burn-in are taken as samples from the posterior density. We construct our MCMC simulation procedure by sampling various blocks of parameters and latent variables in turn within each MCMC iteration. The distributions of these various blocks of parameters are each proportional to the joint posterior $\pi(\mathbf{S}_n, \psi|\mathbf{y})$. In particular, after initializing the various unknowns, we go through four iterative steps in each MCMC cycle. Briefly, in Step 2, we sample $\boldsymbol{\theta}$ from the posterior distribution that is proportional to

$$p(\mathbf{y}|\mathbf{S}_n, \psi)\pi(u_0|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{C2}$$

The sampling of $\boldsymbol{\theta}$ from the latter density is done by the tailored randomized block M–H (TaRB-MH) method of Chib and Ramamurthy (2010). In Step 3, we sample $u_0$ from the posterior distribution that is proportional to

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})p(\mathbf{S}_n|\boldsymbol{\psi})\pi(u_0|\boldsymbol{\theta}). \tag{C3}$$

In Step 4, we sample $\mathbf{S}_n$ conditioned on $\boldsymbol{\psi}$ in one block by the algorithm of Chib (1996). We finish one cycle of the algorithm by sampling $\boldsymbol{\sigma}^{*2}$ conditioned on $(\mathbf{S}_n, \boldsymbol{\theta})$ from the posterior distribution that is proportional to

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})\pi(\boldsymbol{\sigma}^{*2}). \tag{C4}$$

Our algorithm can be summarized as follows.

**Algorithm**: **MCMC sampling**

Step 1 Initialize $(\mathbf{S}_n, \boldsymbol{\psi})$ and fix $n_0$ (the burn-in) and $n_1$ (the MCMC sample size).

Step 2 Sample $\boldsymbol{\theta}$ conditioned on $(\mathbf{y}, \mathbf{S}_n, u_0, \boldsymbol{\sigma}^{*2})$.

Step 3 Sample $u_0$ conditioned on $(\mathbf{y}, \boldsymbol{\theta}, \mathbf{S}_n)$.

Step 4 Sample $\mathbf{S}_n$ conditioned on $(\mathbf{y}, \boldsymbol{\theta}, u_0, \boldsymbol{\sigma}^{*2})$.

Step 5 Sample $\boldsymbol{\sigma}^{*2}$ conditioned on $(\mathbf{y}, \boldsymbol{\theta}, \mathbf{S}_n)$.

Step 6 Repeat Steps 2–6, discard the draws from the first $n_0$ iterations, and save the subsequent $n_1$ draws.

Full details of each of these steps are the following.

Step 2 **Sampling $\boldsymbol{\theta}$.**

We sample $\boldsymbol{\theta}$ conditioned on $(\mathbf{S}_n, u_0, \boldsymbol{\sigma}^{*2})$ by the TaRB-MH algorithm introduced in Chib and Ramamurthy (2010). The schematics of the TaRB-MH algorithm are as follows. The parameters in $\boldsymbol{\theta}$ are first randomly partitioned into various subblocks at the beginning of an iteration. Each of these subblocks is then sampled in sequence by drawing a value from a tailored proposal density constructed for that particular block; this proposal is then accepted or rejected by the usual M–H probability of move (Chib and Greenberg 1995). For instance, suppose that in the $g$th iteration, we have $h_g$ subblocks of $\boldsymbol{\theta}$

$$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{h_g}.$$

If $\boldsymbol{\psi}_{-i}$ denotes the collection of the parameters in $\boldsymbol{\psi}$ except $\boldsymbol{\theta}_i$, then the proposal density $q(\boldsymbol{\theta}_i|\mathbf{y}, \boldsymbol{\psi}_{-i})$ for the $i$th block conditioned on $\boldsymbol{\psi}_{-i}$ is constructed by a quadratic approximation at the mode of the current target

density $\pi(\boldsymbol{\theta}_i|\mathbf{y},\boldsymbol{\psi}_{-i})$. In our case, we let this proposal density take the form of a student $t$ distribution with fifteen degrees of freedom

$$q(\boldsymbol{\theta}_i|\mathbf{y},\boldsymbol{\psi}_{-i}) = St(\boldsymbol{\theta}_i|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15), \tag{C5}$$

where

$$\hat{\boldsymbol{\theta}}_i = \arg \max_{\boldsymbol{\theta}_i} \ln\{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i)\} \tag{C6}$$

$$\text{and } \mathbf{V}_{\hat{\boldsymbol{\theta}}_i} = \left( -\frac{\partial^2 \ln\{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i'} \right)^{-1}_{|\boldsymbol{\theta}_i = \hat{\boldsymbol{\theta}}_i}.$$

Because the likelihood function tends to be ill-behaved in these problems, we calculate $\hat{\boldsymbol{\theta}}_i$ using a suitably designed version of the simulated annealing algorithm. In our experience, this stochastic optimization method works better than the standard Newton–Raphson class of deterministic optimizers.

We then generate a proposal value $\boldsymbol{\theta}_i^{\dagger}$, which, upon satisfying all the constraints, is accepted as the next value in the chain with probability

$$\alpha\left(\boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\theta}_i^{\dagger}|\mathbf{y},\boldsymbol{\psi}_{-i}\right) \tag{C7}$$

$$= \min\left\{ \frac{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^{\dagger}, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^{\dagger})}{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^{(g-1)})} \frac{St(\boldsymbol{\theta}_i^{(g-1)}|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15)}{St(\boldsymbol{\theta}_j^{\dagger}|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15)}, 1 \right\}.$$

If $\boldsymbol{\theta}_i^{\dagger}$ violates any of the constraints in $\mathcal{R}$, it is immediately rejected. The simulation of $\boldsymbol{\theta}$ is complete when all the subblocks

$$\pi(\boldsymbol{\theta}_1|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-1}), \quad \pi(\boldsymbol{\theta}_2|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-2}), \ldots, \pi(\boldsymbol{\theta}_{h_g}|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-h_g}) \tag{C8}$$

are sequentially updated as above.

Step 3 **Sampling the initial factor**.

Given the prior in Equation (B2), $u_0$ is updated conditioned on $\boldsymbol{\theta}$, $\mathbf{m}_0$, and $\mathbf{f}_1 = (u_1 \ \mathbf{m}_1')'$, where $\mathbf{m}_0$ is given by data and $u_1$ is obtained from the Equation (16). In the following, it is assumed that all the underlying coefficients are those in regime 0. Then,

$$u_0|\mathbf{f}_1, \boldsymbol{\theta} \sim \mathcal{N}_1(\bar{u}_0, \mathbf{U}_0), \tag{C9}$$

where

$$\bar{u}_0 = \mathbf{U}_0\left(\boldsymbol{\Sigma}_u^{-1} + \mathbf{H}^{*\prime}\boldsymbol{\Omega}_{11,0}^{*} u_1^{*}\right), \quad \mathbf{U}_0 = \left(\boldsymbol{\Sigma}_u^{-1} + \mathbf{H}^{*\prime}\boldsymbol{\Omega}_{11,1}^{*}\mathbf{H}^{*}\right)$$

and on letting

$$\mathbf{G}_0 = \begin{pmatrix} \mathbf{G}_{11,1}\, \mathbf{G}_{12,1} \\ \mathbf{G}_{21,1}\, \mathbf{G}_{22,1} \end{pmatrix}, \quad \mathbf{\Omega}_1 = \begin{pmatrix} \mathbf{\Omega}_{11,1}\, \mathbf{\Omega}_{12,1} \\ \mathbf{\Omega}_{21,1}\, \mathbf{\Omega}_{22,1} \end{pmatrix},$$

$$\mathbf{H}^* = \mathbf{G}_{11,1} - \mathbf{\Omega}_{12,1}\mathbf{\Omega}_{22,1}^{-1}\mathbf{G}_{21,1},\ \mathbf{\Omega}_{11,1}^* = \mathbf{\Omega}_{11,1} - \mathbf{\Omega}_{12,1}\mathbf{\Omega}_{22,1}^{-1}\mathbf{\Omega}_{21,1},$$

$$u_1^* = u_1 - \mathbf{\Omega}_{12,1}\mathbf{\Omega}_{22,1}^{-1}(\mathbf{m}_1 - \boldsymbol{\mu}_{m,1}) + \left(\mathbf{\Omega}_{12,1}\mathbf{\Omega}_{22,1}^{-1}\mathbf{G}_{22,1} - \mathbf{G}_{12,1}\right)(\mathbf{m}_0 - \boldsymbol{\mu}_{m,1}).$$

Step 4 **Sampling regimes**.

In this step, one samples the states from $p[\mathbf{S}_n|I_n, \boldsymbol{\psi}]$ where $I_n$ is the history of the outcomes up to time $n$. This is done according to the method of Chib (1996) by sampling $\mathbf{S}_n$ in a single block from the output of one forward and backward pass through the data.

The forward recursion is initialized at $t = 1$ by setting $\Pr[s_1 = 1|I_1, \boldsymbol{\psi}] = 1$. Then, one first obtains $\Pr[s_t = j|I_t, \boldsymbol{\psi}]$ for all $j = 1, 2, \ldots, m+1$ and $t = 1, 2, \ldots, n$ by calculating

$$\Pr[s_t = j|I_t, \boldsymbol{\psi}] = \sum_{i=j-1}^{j} \Pr[s_{t-1} = i, s_t = j|I_t, \boldsymbol{\psi}], \tag{C10}$$

where

$$\Pr[s_{t-1} = i, s_t = j|I_t, \boldsymbol{\psi}] = \frac{p\left[\mathbf{y}_t|I_{t-1}, s_{t-1} = i, s_t = j, \boldsymbol{\psi}\right] \Pr[s_{t-1} = i, s_t = j|I_{t-1}, \boldsymbol{\psi}]}{p\left[\mathbf{y}_t|I_{t-1}, \boldsymbol{\psi}\right]}.$$

This can be done by the Equations (D4–D7).

In the backward pass, one simulates $\mathbf{S}_n$ by the method of composition. One samples $s_n$ from $\Pr[s_n|I_n, \boldsymbol{\psi}]$. We remark that in this sampling step, $s_n$ can take any value in $\{1, 2, \ldots, m+1\}$. For instance, if $s_n$ turns out to be $m$ and not $(m+1)$, then the parameters of regime $(m+1)$ are drawn from the prior in that iteration. In our data, however, $(m+1)$ is always drawn because the last change-point occurs in the interior of the sample and, therefore, the distribution $\Pr[s_n|I_n, \boldsymbol{\psi}]$ has almost a unit mass on $(m+1)$. Then, for $t = 1, 2, \ldots, n-1$, we sequentially calculate

$$\Pr[s_t = j|I_t, s_{t+1} = k, S^{t+2}, \boldsymbol{\psi}] = \Pr[s_t = j|I_t, s_{t+1} = k, \boldsymbol{\psi}] \tag{C11}$$

$$= \frac{\Pr[s_{t+1} = k|s_t = j] \Pr[s_t = j|I_t, \boldsymbol{\psi}]}{\sum_{j=k-1}^{k} \Pr[s_{t+1} = k|s_t = j] \Pr[s_t = j|I_t, \boldsymbol{\psi}]},$$

where $S^{t+1} = \{s_{t+1}, \ldots, s_n\}$ denotes the set of simulated states from the earlier steps. A value $s_t$ is drawn from this distribution and it is either the value $k$ or $(k-1)$ conditioned on $s_{t+1} = k$.

Step 5 **Sampling the variances of the pricing errors**.

A convenient feature of our modeling approach is that, conditional on the history of the regimes and factors, the joint distribution of the parameters in $\boldsymbol{\sigma}^{*2}$ is analytically tractable and takes the form of an inverse gamma density. Thus, for $i \in \{1,2,\ldots,7,9,\ldots,16\}$ and $j = 1,2,\ldots,m+1$, $\sigma^{*2}_{i,j}$ is sampled from

$$\mathbf{IG}\left\{\frac{\bar{v}+\sum_{t=1}^{n} I(s_t=j)}{2}, \frac{\bar{d}+\sum_{t=1}^{n} d_{i,j} I(s_t=j)(R_{ti}-\bar{a}_{i,j}-\bar{\mathbf{b}}'_{i,j}(\mathbf{f}_t-\boldsymbol{\mu}_j))^2}{2}\right\}, \tag{C12}$$

where $I(\cdot)$ is the indicator function.

The algorithm is coded in Gauss 9.0 and executed on a Windows Vista 64-bit machine with a 2.66 GHz Intel Quad Core2 CPU. About 12 days are needed to generate 50,000 MCMC draws in the three change-point model. In contrast, a random-walk M–H algorithm takes about two days to complete 1 million iterations but with unknown reliability and much less efficient exploration (Chib and Ramamurthy, 2010).

## APPENDIX D: MARGINAL LIKELIHOOD COMPUTATION

The marginal likelihood of any given model is obtained as

$$m(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{S}_n,\boldsymbol{\psi})p(\mathbf{S}_n|\boldsymbol{\psi})\pi(\boldsymbol{\psi})d(\mathbf{S}_n,\boldsymbol{\psi}). \tag{D1}$$

This integration is obviously infeasible by direct means. It is possible, however, by the method of Chib (1995), which starts with the recognition that the marginal likelihood can be expressed in equivalent form as

$$m(\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\psi}^*)\pi(\boldsymbol{\psi}^*)}{\pi(\boldsymbol{\psi}^*|\mathbf{y})}, \tag{D2}$$

where $\boldsymbol{\psi}^* = (\boldsymbol{\theta}^*, \boldsymbol{\sigma}^{**2}, u_0^*)$ is some specified (say high-density) point of $\boldsymbol{\psi} = (\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, u_0)$. Provided we have an estimate of posterior ordinate $\pi(\boldsymbol{\psi}^*|\mathbf{y})$, the marginal likelihood can be computed on the log scale as

$$\ln \hat{m}(\mathbf{y}) = \ln p(\mathbf{y}|\boldsymbol{\psi}^*) + \ln \pi(\boldsymbol{\psi}^*) - \ln \hat{\pi}(\boldsymbol{\psi}^*|\mathbf{y}). \tag{D3}$$

Notice that the first term in this expression is the likelihood. It has to be evaluated only at a single point, which is highly convenient. The calculation of the second term is straightforward. Finally, the third term is obtained from a marginal–conditional decomposition following Chib (1995). The specific implementation in this context requires the technique of Chib and Jeliazkov (2001) as modified by Chib and Ramamurthy (2010) for the case of randomized blocks.

As for the calculation of the likelihood, the joint density of the data $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is, by definition,

$$p(\mathbf{y}|\boldsymbol{\psi}) = \sum_{t=0}^{n-1} \ln p\left(\mathbf{y}_{t+1}|I_t, \boldsymbol{\psi}\right), \tag{D4}$$

where

$$p\left(\mathbf{y}_{t+1}|I_t, \boldsymbol{\psi}\right) = \sum_{s_{t+1}=1}^{m+1} \sum_{s_t=1}^{m+1} p\left(\mathbf{y}_{t+1}|I_t, s_t, s_{t+1}, \boldsymbol{\psi}\right) \Pr[s_t, s_{t+1}|I_t, \boldsymbol{\psi}]$$

is the one-step ahead predictive density of $\mathbf{y}_{t+1}$, and $I_t$ consists of the history of the outcomes $R_t$ and $\mathbf{z}_t$ up to time $t$. On the right-hand side, the first term is the density of $\mathbf{y}_{t+1}$ conditioned on $(I_t, s_t, s_{t+1}, \boldsymbol{\psi})$, which is given in Equation (22), whereas the second term can be calculated from the law of total probability as

$$\Pr[s_t = j, s_{t+1} = k|I_t, \boldsymbol{\psi}] = p_{jk} \Pr[s_t = j|I_t, \boldsymbol{\psi}], \tag{D5}$$

where $\Pr[s_t = j|I_t, \boldsymbol{\psi}]$ is obtained recursively starting with $\Pr[s_1 = 1|I_0, \boldsymbol{\psi}] = 1$ by the following steps. Once $\mathbf{y}_{t+1}$ is observed at the end of time $t+1$, the probability of the regime $\Pr[s_{t+1} = k|I_t, \boldsymbol{\psi}]$ from the previous step is updated to $\Pr[s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}]$ as

$$\Pr[s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}] = \sum_{j=1}^{m+1} \Pr[s_t = j, s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}], \tag{D6}$$

where

$$\Pr[s_t = j, s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}] = \frac{p\left[\mathbf{y}_{t+1}|I_t, s_t = j, s_{t+1} = k, \boldsymbol{\psi}\right] \Pr[s_t = j, s_{t+1} = k|I_t, \boldsymbol{\psi}]}{p\left[\mathbf{y}_{t+1}|I_t, \boldsymbol{\psi}\right]}. \tag{D7}$$

This completes the calculation of the likelihood function.

*Received March 30, 2011; revised 15 February, 2012; accepted 24 February, 2012.*

## REFERENCES

Albert, J., and S. Chib. 1993. Bayesian Analysis via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts. *Journal of Business and Economic Statistics* 11(1): 1–15.

Ang, A., and G. Bekaert. 2002. Regime Switches in Interest Rates. *Journal of Business and Economic Statistics* 20: 163–182.

Ang, A., G. Bekaert, and M. Wei. 2008. The Term Structure of Real Rates and Expected Inflation. *Journal of Finance* 63: 797–849.

Ang, A., S. Dong, and M. Piazzesi. (2007). "No-Arbitrage Taylor Rules." Columbia University working paper.

Ang, A., and M. Piazzesi. 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50: 745–787.

Bansal, R., and H. Zhou. 2002. Term Structure of Interest Rates with Regime Shifts. *Journal of Finance* 57(5): 1997–2043.

Chen, R., and L. Scott. 2003. ML Estimation for a Multifactor Equilibrium Model of the Term Structure. *Journal of Fixed Income* 27: 14–31.

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

Chib, S. 1996. Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics* 75: 79–97.

Chib, S. 1998. Estimation and Comparison of Multiple Change-Point Models. *Journal of Econometrics* 86: 221–241.

Chib, S. 2001. "Markov Chain Monte Carlo Methods: Computation and Inference." In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, 3569–3649. Amsterdam, The Netherlands: North Holland.

Chib, S., and B. Ergashev. 2009. Analysis of Multi-Factor Affine Yield Curve Models. *Journal of the American Statistical Association* 104: 1324–1337.

Chib, S., and E. Greenberg. 1995. Understanding the Metropolis-Hastings Algorithm. *American Statistician* 49: 327–335.

Chib, S., and I. Jeliazkov. 2001. Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association* 96: 270–281.

Chib, S., and S. Ramamurthy. 2010. Tailored Randomized-Block MCMC Methods with Application to DSGE Models. *Journal of Econometrics* 155: 19–38.

Dai, Q., K. J. Singleton, and W. Yang. 2007. Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields. *Review of Financial Studies* 20: 1669–1706.

Duffie, G., and R. Kan. 1996. A Yield-Factor Model of Interest Rates. *Mathematical Finance* 6: 379–406.

Gelfand, A. E., and S. K. Ghosh. 1998. Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika* 85: 1–11.

Gurkaynak, R. S., B. Sack, and J. H. Wright. 2007. The U.S. Treasury Yield Curve: 1961 to the Present. *Journal of Monetary Economics* 54: 2291–2304.

Kim, C. J., C. R. Nelson, and J. Piger. 2004. The Less Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations. *Journal of Business and Economic Statistics* 22: 80–93.

Rudebusch, G., B. P. Sack, and E. T. Swanson. 2007. Macroeconomic Implications of Changes in the Term Premium. *Federal Reserve Bank of St Louis Review* 89: 241–269.

Rudebusch, G., and T. Wu. 2008. A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy. *Economic Journal* 118: 906–926.