

Change Points in Affine Arbitrage-free Term Structure Models*

SIDDHARTHA CHIB[†]

(Washington University in St. Louis)

KYU HO KANG[‡]

(Ajou University)

January 2011

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 macro-economics 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 US 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)

I Introduction

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. From the evidence in these papers it is clear that structural breaks are a feature of

*We thank Taeyoung Doh, Ed Greenberg, Wolfgang Lemke, Hong Liu, James Morley, Srikanth Ramamurthy, Myung Hwan Seo, Yongs Shin, Guofu Zhou, the participants of the 2009 Econometric Society summer meeting, the 2009 Seminar on Bayesian Inference in Econometrics and Statistics, and the 2009 Northern Finance Association Conference, for their thoughtful and useful comments on the paper. 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:* 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 Financial Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suweon, Gyeonggi-do, 443-749, South Korea. E-mail: kyu@kyukang.net.

the US term structure of interest rates. In this paper we provide another approach for modeling structural breaks in affine models that we believe has some advantages over the Markov switching framework. In particular, we model structural breaks through a change-point specification. In such a 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. The change-point formulation can be useful if the conditions that determine a regime are unique and not repeated. Change-point models are also easier to estimate than Markov switching models because they do not suffer from the so-called label switching problem. Finally, if regimes do not actually recur, out-of-sample forecasts based on Markov-switching models can be misleading.

The model we develop is in the family of models that arise from the work of Duffie and Kan (1996) and Dai and Singleton (2000). The model is a contemporary affine term structure model except that all model parameters, including the factor loadings, are subject to regime changes. Thus, in our setting, we do not have to decide which parameters are constant and which break. In addition, our model contains both latent and macro-economic factors, as in, for example, Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007) and Chib and Ergashev (2009).

The data we analyze comprises a collection of 16 yields of US T-bills measured quarterly between 1972:I and 2007:IV. In order to find the appropriate number of change-points, we specify models with different number of change-points and estimate the models by Bayesian techniques. The best fitting model is then selected by the marginal likelihood/Bayes factors criteria.

Our main findings are as follows. The 3 change point model is the one that is most supported by the data. The breaks are inferred to have occurred in 1980:II, 1985:IV and 1995:II. As we discuss below, the model estimation reveals that the parameters across regimes are substantially different, which provides support to our approach of letting all the parameters vary across regimes. We find, for instance, that the mean-reversion parameters in the factor dynamics and the factor loadings are regime-specific.

The rest of the paper is organized as follows. In Section 2 we present our change point term-structure model and derive the resulting bond prices. We outline the prior-

posterior analysis of our model in Section 3, deferring details of the MCMC simulation procedure to the appendix of the paper. Section 4 deals with the empirical analysis of the real data and Section 5 has our conclusions.

II Model Specification

In this section we develop our model of bond pricing under regime changes. It is a standard affine term structure model except that all model parameters are subject to regime changes. Let $\{s_t\}$ denote a discrete-state variable that takes one of the values $\{1, 2, \dots, m + 1\}$ such that $s_t = j$ indicates that the time t observation has been drawn from the j th regime. We refer to the times $\{t_1, t_2, \dots\}$ 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 denote the factors. These consist of one latent variable u_t and two observed macroeconomic variables \mathbf{m}_t . 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

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

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 τ 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.

A 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

$s_t \in \{1, 2, \dots\}$. 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). We prefer the model of Chib (1998) instead of that of Koop and Potter (2007) because bond pricing is substantially more complex in the latter case because the transition probabilities of regime changes depend on the time spent in each regime.

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,

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

This formulation of the change point model in terms of a restricted unidirectional Markov process makes obvious how the change point assumption differs from the Markov-switching regime process in Dai et al. (2007), Bansal and Zhou (2002) and Ang et al. (2008) where the transition probability matrix is unrestricted and previously occupied states can be revisited. As we have argued above, there are strong reasons for looking at the term structure from the change point perspective.

B 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

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

where on letting $\mathcal{N}_3(\cdot, \cdot)$ denote the 3-dimensional normal distribution, $\boldsymbol{\eta}_{t+1} | s_{t+1} \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{\Omega}_{s_{t+1}})$, and for s_t and s_{t+1} ranging from $j = 1$ to $m + 1$, $\boldsymbol{\mu}_j$ is a 3×1 vector and \mathbf{G}_j is a 3×3 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

$$\boldsymbol{\eta}_{t+1} = \mathbf{L}_{s_{t+1}} \boldsymbol{\omega}_{t+1} \quad (3)$$

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 et al. (2007) depend only on s_t whereas those in Bansal and Zhou (2002) and Ang et al. (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 .

C 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

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

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 (3). The SDF is independent of s_{t+1} given s_t as in the model of Dai et al. (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 et al. (2007)).

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

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

where the intercept δ_{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 et al. (2008) and Dai et al. (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 γ_{t,s_t} are governed by

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

where $\tilde{\gamma}_{s_t} : 3 \times 1$ is the regime-dependent expectation of γ_{t,s_t} and $\Phi_{s_t} : 3 \times 3$ is a matrix of regime-specific parameters. We refer to the collection $(\tilde{\gamma}_{s_t}, \Phi_{s_t})$ as the factor-risk parameters. Note that in this specification γ_{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 et al. (2008) and Bansal and Zhou (2002) regime risk cannot also be isolated since it is confounded with the market price of factor risk.

D Bond Prices

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

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

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

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

where R_{t,τ,s_t} is the continuously compounded yield given by

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

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

We find the expressions for the latter functions by the method of undetermined coefficients and the law of the iterated expectation. For $j = 1, 2, \dots$, and $k = j + 1$, the procedure discussed in Appendix A produces the following recursive system for the unknown functions

$$\begin{aligned} a_j(\tau) &= \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \delta_{1,j} - \tilde{\gamma}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1) - \mathbf{b}_j(\tau - 1)' \mathbf{L}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1)/2 + a_j(\tau - 1) \\ \delta_{1,j} - \tilde{\gamma}_j \mathbf{L}'_k \mathbf{b}_k(\tau - 1) - \mathbf{b}_k(\tau - 1)' \mathbf{L}_k \mathbf{L}'_k \mathbf{b}_k(\tau - 1)/2 + a_k(\tau - 1) \end{pmatrix} \\ \mathbf{b}_j(\tau) &= \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{2,j} + (\mathbf{G}_j - \mathbf{L}_j \boldsymbol{\Phi}_j)' \mathbf{b}_j(\tau - 1) \\ \boldsymbol{\delta}_{2,j} + (\mathbf{G}_k - \mathbf{L}_k \boldsymbol{\Phi}_j)' \mathbf{b}_k(\tau - 1) \end{pmatrix} \end{aligned} \quad (10)$$

where τ runs over the positive integers. These recursions are initialized by setting $a_{s_t}(0) = 0$ and $\mathbf{b}_{s_t}(0) = \mathbf{0}_{3 \times 1}$ 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 $m = 0, 1, 2, \dots$. 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 regimes that is possible under that model supposition and that fewer regimes may arise when the states are sampled by the method in Appendix C, step 4. 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. These different models are compared in terms of marginal likelihoods as we discuss below.

E Regime-specific Term Premium

As is well known, under risk-neutral pricing, after adjusting for risk, agents are indifferent between holding a τ -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 τ -period bond can be calculated as

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

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.

III 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 data set of 16 yields of US 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 16 yields under study be denoted by

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

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

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

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 Φ_{s_t} are denoted by

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

where G_{ij,s_t} and Φ_{ij,s_t} denote the (i, j) th element of \mathbf{G}_{s_t} and Φ_{s_t} , respectively. The unknown elements of Ω_{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 $\Omega_{s_t} = \mathbf{L}_{s_t} \mathbf{L}'_{s_t}$ with \mathbf{L}_{s_t} expressed as

$$\begin{pmatrix} 1/400 & 0 & 0 \\ l_{21,s_t} & \exp(l_{22,s_t}^*) & 0 \\ l_{31,s_t} & l_{32,s_t} & \exp(l_{33,s_t}^*) \end{pmatrix} \quad (2)$$

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

$\mathbf{p} = \{p_{jj}, j = 1, 2, \dots, m\}$. Finally, the unknown pricing error variances σ_{i,s_t}^2 are collected in reparameterized form as

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

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

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

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.

A 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 et al. (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 eighth yield in the list above) is the basis yield which is priced exactly by the model. Let \mathbf{R}_t denote the remaining 15 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

$$\bar{\mathbf{b}}_{8,s_t} = \begin{pmatrix} \bar{b}_{8,u,s_t} \\ \bar{\mathbf{b}}_{8,m,s_t} \end{pmatrix} \quad (3)$$

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

$$R_{t8} = \bar{a}_{8,s_t} + \bar{b}_{8,u,s_t} u_t + \bar{\mathbf{b}}_{8,m,s_t}' (\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t}) \quad (4)$$

On rewriting this expression, it follows that u_t is

$$u_t = (\bar{b}_{8,u,s_t})^{-1} (R_{t8} - \bar{a}_{8,s_t} - \bar{\mathbf{b}}'_{8,m,s_t} (\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})) \quad (5)$$

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

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

$$\alpha_{s_t} = \begin{pmatrix} (\bar{b}_{8,u,s_t})^{-1} \bar{\mathbf{b}}'_{8,m,s_t} \boldsymbol{\mu}_{m,s_t} - (\bar{b}_{8,u,s_t})^{-1} \bar{a}_{8,s_t} \\ \mathbf{0}_{2 \times 1} \end{pmatrix}, \text{ and} \quad (6)$$

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

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

$$\mathbf{f}_t = \alpha_{s_t} + \mathbf{A}_{s_t} \mathbf{z}_t \quad (7)$$

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 non-basis yields \mathbf{R}_t in our model are generated according to the process

$$\mathbf{R}_t = \bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{iid} \mathcal{N}(0, \boldsymbol{\Sigma}_{s_t}) \quad (8)$$

where

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

In other words,

$$\begin{aligned} p(\mathbf{R}_t | \mathbf{z}_t, s_t, \boldsymbol{\psi}) &= p(\mathbf{R}_t | \mathbf{f}_t, s_t, \boldsymbol{\psi}) \\ &= \mathcal{N}_{15}(\mathbf{R}_t | \bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \boldsymbol{\Sigma}_{s_t}) \end{aligned} \quad (9)$$

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 (7). In particular,

$$\begin{aligned} p(\mathbf{z}_t | \mathbf{z}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) &= p(\mathbf{f}_t | \mathbf{f}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) \det(\mathbf{A}_{s_t}) \\ &= \mathcal{N}_3(\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t} (\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \boldsymbol{\Omega}_{s_t}) | (\bar{b}_{8,u,s_t})^{-1} | \end{aligned} \quad (10)$$

If we let

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

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi}) = \prod_{t=1}^n \mathcal{N}_{15}(\mathbf{R}_t | \bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \boldsymbol{\Sigma}_{s_t}) \quad (11)$$

$$\times \mathcal{N}_3(\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t}(\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \boldsymbol{\Omega}_{s_t}) | (\bar{b}_{8,u,s_t})^{-1} | \quad (12)$$

B Prior-Posterior Analysis

B.1 Prior Distribution

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. 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.

We begin by recalling the identifying restrictions on the parameters. First, we set $\boldsymbol{\mu}_{u,s_t} = \mathbf{0}$ which implies that the mean of the short rate conditional on s_t is δ_{1,s_t} . Next, the first element of $\boldsymbol{\delta}_{2,s_t}$, namely δ_{21,s_t} , is assumed to be non-negative. 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

$$\mathcal{R} = \{\mathbf{G}_j, \delta_{21,j} | \delta_{21,j} \geq 0, 0 \leq p_{jj} \leq 1, |eig(\mathbf{G}_j)| < 1 \text{ for } j = 1, 2, \dots, m+1\} \quad (13)$$

All the constraints in \mathcal{R} are enforced through the prior distribution. The specific prior distributions can be found in Appendix B.

B.2 Posterior Distribution and MCMC Sampling

Under our assumptions 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.

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

$$\pi(\mathbf{S}_n, \boldsymbol{\psi} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{S}_n, \boldsymbol{\psi}) p(\mathbf{S}_n | \boldsymbol{\psi}) \pi(\boldsymbol{\psi}) \quad (14)$$

where $p(\mathbf{y} | \mathbf{S}_n, \boldsymbol{\psi})$ is the distribution of the data given the regime indicators and the parameters, $p(\mathbf{S}_n | \boldsymbol{\psi})$ is the density of the regime-indicators given the parameters and the initial latent factor, and $\pi(\boldsymbol{\psi})$ 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 $p(\mathbf{y} | \boldsymbol{\psi})$ 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 $\pi(\mathbf{S}_n, \boldsymbol{\psi} | \mathbf{y})$. In particular, after initializing the various unknowns, we go through 4 iterative steps in each MCMC cycle. Briefly, in

Step 2 we sample $\boldsymbol{\theta}$ from the posterior distribution that is proportional to

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})\pi(u_0|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (15)$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})p(\mathbf{S}_n|\boldsymbol{\psi})\pi(u_0|\boldsymbol{\theta}) \quad (16)$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})\pi(\boldsymbol{\sigma}^{*2}) \quad (17)$$

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 given in Appendix C.

IV Results

We apply our modeling approach to analyze US data on quarterly yields of sixteen US 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 8 quarter (or 2 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 16 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 4, 8, and 12 yields to data up to 2006. The last 4 quarters of 2007 are held aside for the validation of the predictions of the yields and the macro factors. These predictions are generated as described in Section D. 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 λ number of the maturities, PPC is defined as

$$\text{PPC} = \text{D} + \text{W} \quad (1)$$

where

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

$$\text{W} = \frac{1}{\lambda + 2} \sum_{i=1}^{\lambda+2} \sum_{t=1}^T [y_{i,t} - E(\tilde{y}_{i,t} | \mathbf{y}, \mathcal{M})]^2 \quad (3)$$

$\{\tilde{\mathbf{y}}_t\}_{t=1,2,\dots,T}$ are the predictions of the yields and macro factors $\{\mathbf{y}_t\}_{t=1,2,\dots,T}$ 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 16 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

The number of maturities(λ)	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

Table 1: Posterior predictive criterion. *PPC is computed by 1 to 3. We use the data from the most recent break time point, 1995:II to 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 et al. (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.*

parameters, the additional outcome helps to improve inferences about the common model parameters, which translates into improved predictive inferences.

A Sampler Diagnostics

We base our results on 50,000 iterations of the MCMC algorithm beyond a burn-in of 5,000 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

$$1 + 2 \sum_{k=1}^K \rho(k), \quad (4)$$

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.

B The Number and Timing of Change Points

One of our goals is 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 5 models which in the introduction were named as \mathcal{M}_0 , \mathcal{M}_1 , \mathcal{M}_2 , \mathcal{M}_3 and \mathcal{M}_4 . The most general model is \mathcal{M}_4 that has 4 possible change points, 1 latent factor and 2 macro factors. We do the comparison in terms of marginal likelihoods and their ratios which are called Bayes factors. The Appendix D provides the details of the marginal likelihood computation.

Table 2 contains the marginal likelihood estimates for our 5 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.

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

Table 2: Log likelihood (lnL), log marginal likelihood (lnML), posterior probability of each model (Pr[$\mathcal{M}_m|\mathbf{y}$]) under the assumption that the prior probability of each model is 1/5, and change point estimates.

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 32 quarters (the first 8 years) belong to the first regime, the next 23 quarters (about 6 years) to the second, the next 38 quarters (about 9.5 years) to the third, and the remaining quarters to the fourth 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.

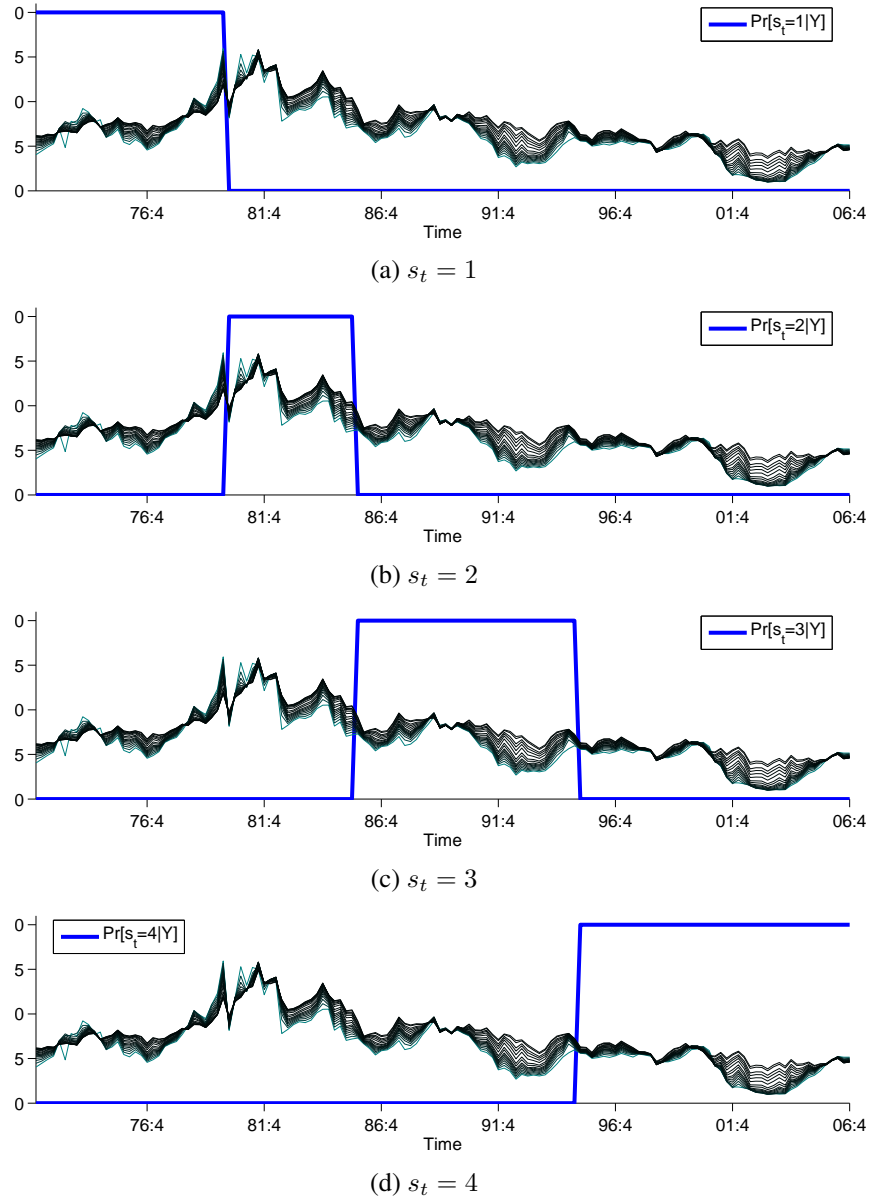


Figure 1: Model \mathcal{M}_3 : $\Pr(s_t = j | \mathbf{y})$. The posterior probabilities for each t are based on 50,000 MCMC draws of s_t - these probabilities are plotted along with the 16 yields in annualized percents (probabilities are multiplied by 20 for legibility).

We would like to emphasize that our estimates of the change points from the models without macro factors are exactly the same as those from the change point models with macro factors. We do not report those results in the interest of space. In addition, the

results are not sensitive to our choice of 16 maturities, as we have confirmed.

C 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 Appendix B, 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.

C.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

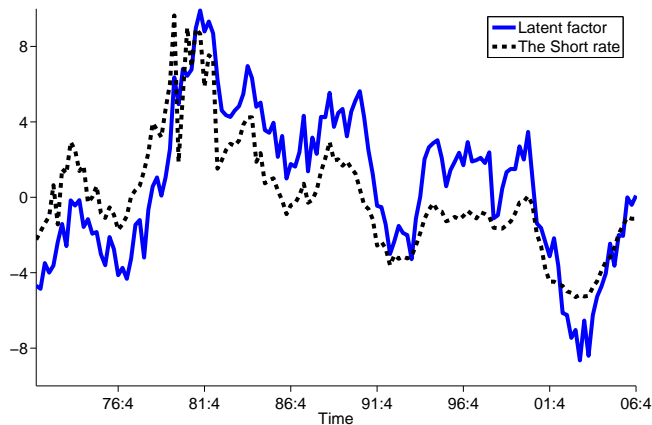


Figure 2: Model \mathcal{M}_3 : Estimates of the latent factor. *The short rate in percent is demeaned and estimates of the latent factor are calculated as the average of factor drawings given the 50,000 MCMC draws of the parameters.*

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 indicates 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

	Regime 1			Regime 2			Regime 3			Regime 4		
G	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)
	-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)
μ $\times 400$	0.00	4.99	3.54	0.00	5.88	2.63	0.00	2.56	2.62	0.00	1.49	3.22
		(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 $\times 400$	0.11	1.72		0.10	1.48		0.11	0.74		-0.47	0.82	
	(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)
δ_1 $\times 400$		9.23			2.78			4.42			4.34	
		(1.69)			(1.60)			(1.18)			(1.00)	
δ_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)
γ	-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)
Φ	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_{22}							0.987					
							(0.003)					

Table 3: Model M_3 : Parameter estimates. *This table presents the posterior mean and standard deviation based on 50,000 MCMC draws beyond a burn-in of 5,000. The 95% credibility interval of parameters in bold face 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.*

persistence, we calculate the eigenvalues of the coefficient matrices in each regime. These are given by

$$\begin{aligned}
 eig(\mathbf{G}_1) &= \begin{bmatrix} 0.851 \\ 0.709 \\ 0.267 \end{bmatrix}, & eig(\mathbf{G}_2) &= \begin{bmatrix} 0.978 \\ 0.814 \\ 0.401 \end{bmatrix} \\
 eig(\mathbf{G}_3) &= \begin{bmatrix} 0.935 \\ 0.312 \\ 0.366 \end{bmatrix}, & eig(\mathbf{G}_4) &= \begin{bmatrix} 0.913 + 0.044i \\ 0.913 - 0.044i \\ 0.204 \end{bmatrix}
 \end{aligned}$$

It can be seen that the second regime has the largest absolute eigenvalue close to 1. Because the factor loadings for the latent factor (δ_{21,s_t}) are significant whereas those for inflation (δ_{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 suggests 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)).

C.2 Factor Loadings

The factor loadings in the short rate equation, δ_{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 δ_{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.

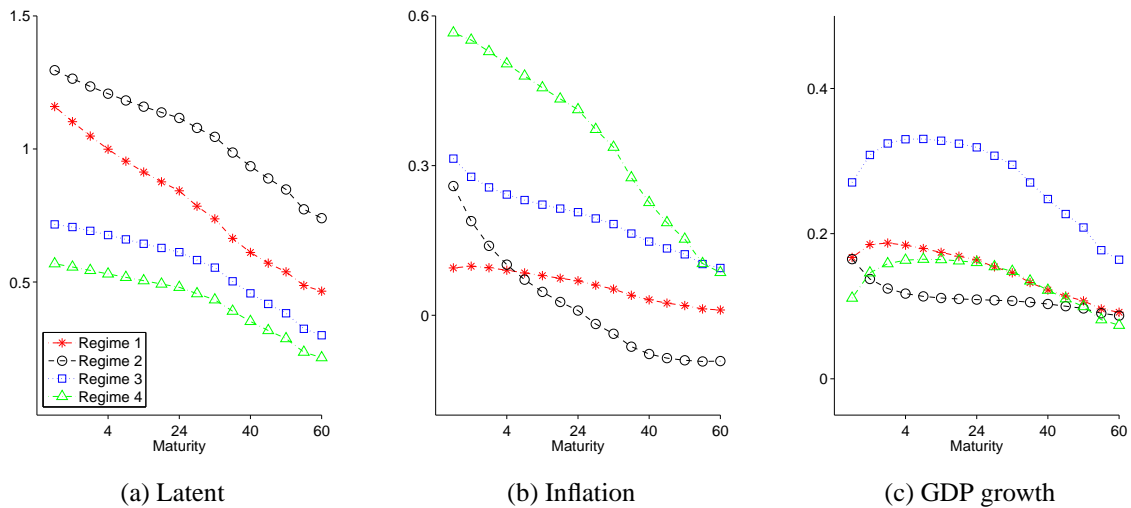


Figure 3: Model \mathcal{M}_3 : Estimates of the factor loadings, $\bar{\mathbf{b}}_{s_t}$. The factor loadings represent the average simulated factor loadings from the retained 50,000 MCMC iterations.

C.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 term premium varies across regimes. In particular, the term premium is the lowest in the most recent regime (although the .025 quantile of the term premium distribution in the first regime is lower than the .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 macro-economic factors although the parameters in Φ_{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).

C.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.

D 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 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,

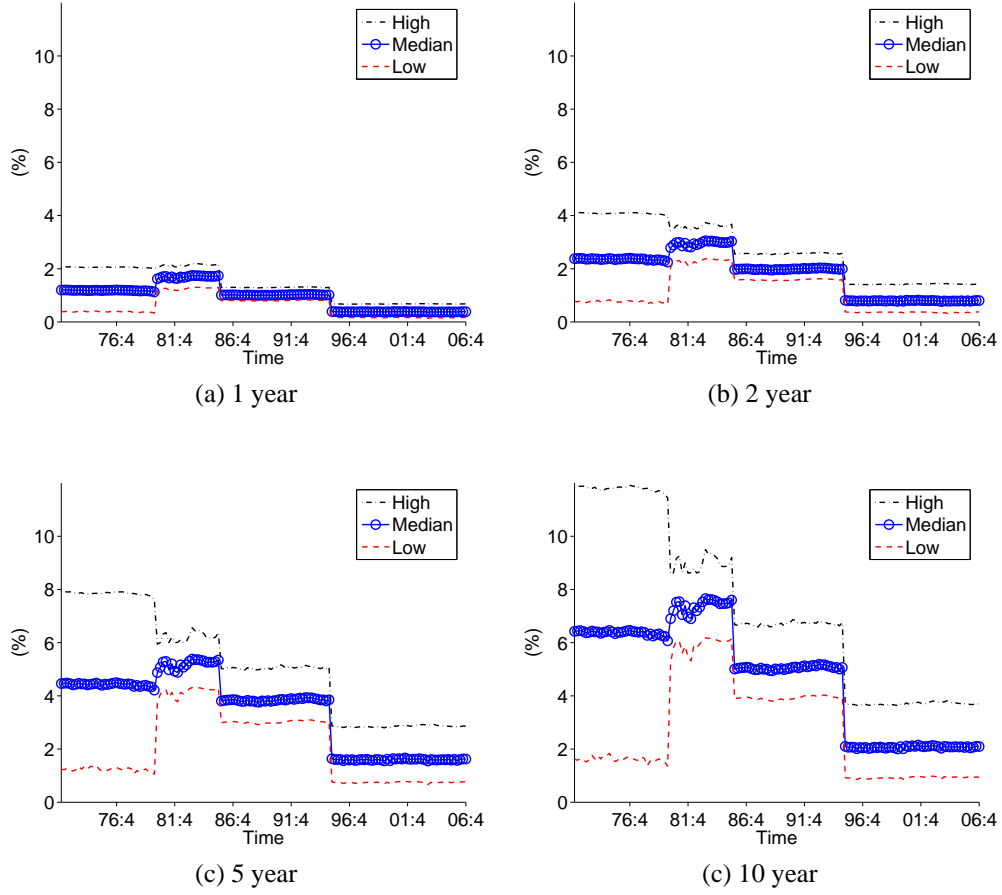


Figure 4: Model \mathcal{M}_3 : Term premium. The figure plots the 2.5%, 50% and 97.5% quantile of the posterior term premium based on 50,000 MCMC draws beyond a burn-in of 5,000 iterations.

the predictive density under model \mathcal{M}_m is given by

$$p(\mathbf{y}_f | \mathcal{M}_m, \mathbf{y}) = \int_{\boldsymbol{\psi}} p(\mathbf{y}_f | \mathcal{M}_m, \mathbf{y}, \boldsymbol{\psi}) \pi(\boldsymbol{\psi} | \mathcal{M}_m, \mathbf{y}) d\boldsymbol{\psi} \quad (5)$$

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 (8). These two steps are iterated forward to produce the draws \mathbf{f}_{n+i} and \mathbf{R}_{n+i} , $i = 1, 2, \dots, T$. Repeated over the course of the MCMC iterations, these steps produce a collection of simulated macro factors and yields that is a sample from the predictive density.

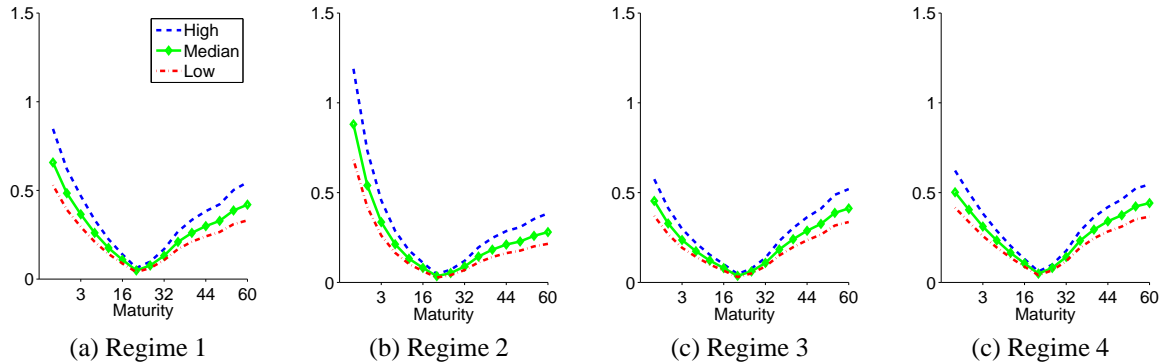


Figure 5: Model \mathcal{M}_3 : Term Structure of the Pricing Error Volatility. *The figures display the 2.5%, 50% and 97.5% quantile of the posterior standard deviation of the pricing errors.*

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. These findings not only reaffirm the finding of structural changes, but also suggest that there are gains to incorporating regime changes when forecasting the term structure of interest rates.

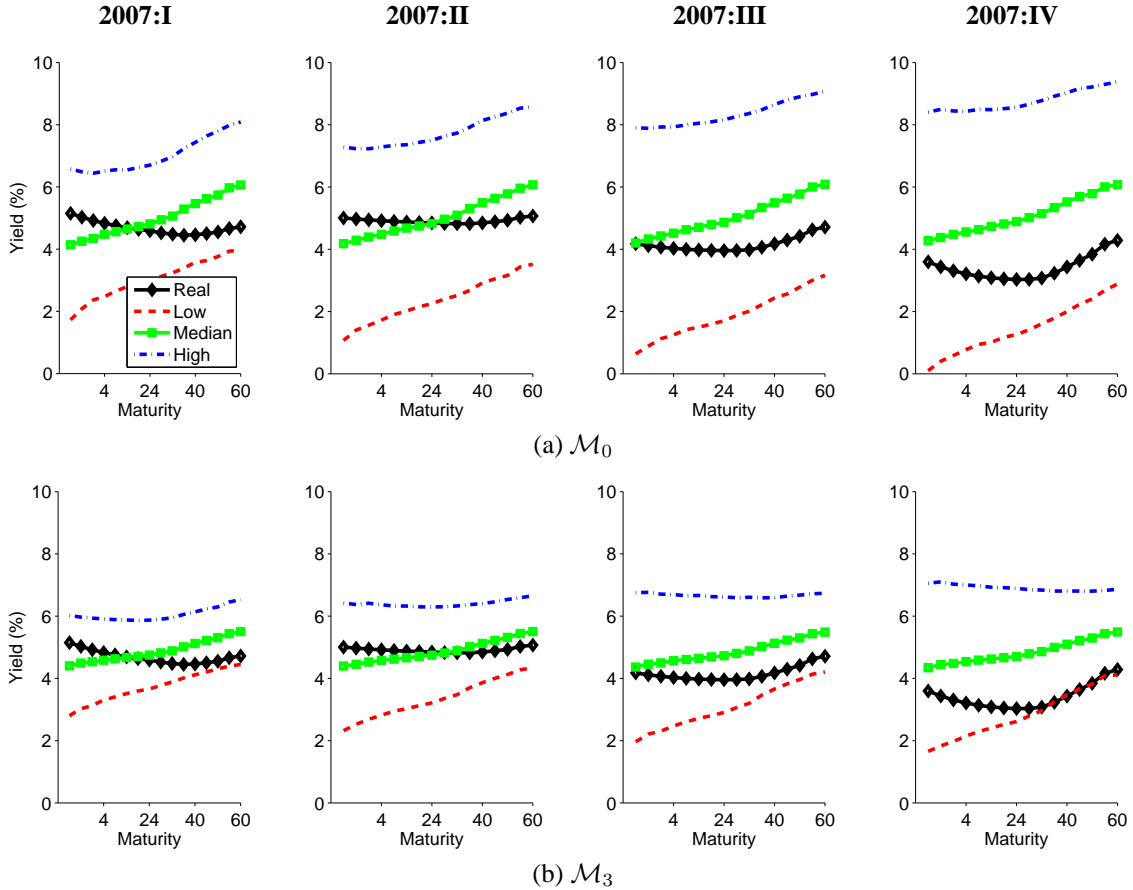


Figure 6: Predicted yield curve. *The figures present four quarters ahead forecasts of the yields on the T-bills. The top panel is based on the no change point model and the bottom panel on the three change point model. In each case, the 2.5%, 50% and 97.5% quantile curves are based on 50,000 forecasted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

V 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

model	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_0
sample period	(1972:I-2006:IV)					(1995:II-2006:IV)
D	12.548	5.401	4.156	4.720	4.599	4.011
W	5.678	4.896	4.201	3.415	2.902	3.520
PPC	18.226	10.297	8.357	8.126	7.501	7.531

(a) *forecast period: 2007:I-2007:IV*

model	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_0
sample period	(1972:I-2005:IV)					(1995:II-2005:IV)
D	12.606	5.799	4.157	4.097	7.011	4.271
W	2.137	5.658	4.432	1.817	3.036	2.390
PPC	14.743	11.457	8.589	5.914	10.047	6.661

(b) *forecast period: 2006:I-2006:IV*

model	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4	\mathcal{M}_0
sample period	(1972:I-2004:IV)					(1995:II-2004:IV)
D	13.474	5.187	3.572	4.609	7.190	3.919
W	2.367	5.787	4.442	1.977	2.657	2.359
PPC	15.841	10.974	8.014	6.587	9.847	6.278

(c) *forecast period: 2005:I-2005:IV*

Table 4: Posterior predictive criterion. PPC is computed by (1) to (3).

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 (7) can be expressed as

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

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 (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

$$P_t(s_t, \tau) = \exp(-a_{s_t}(\tau) - \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})) \quad (\text{A.2})$$

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

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

$$\frac{P_{t+1}(s_{t+1}, \tau - 1)}{P_t(s_t, \tau)} = \exp[-a_{s_{t+1}}(\tau - 1) - \mathbf{b}_{s_{t+1}}(\tau - 1)'(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}}) + a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})] \quad (\text{A.3})$$

It immediately follows from the bond pricing formula that

$$\begin{aligned} 1 &= \mathbb{E}_t \left[\kappa_{t, s_t, t+1} \frac{P_{t+1}(s_{t+1}, \tau - 1)}{P_t(s_t, \tau)} \right] \\ &= \mathbb{E}_t [\kappa_{t, s_t, t+1} h_{\tau, t+1}]. \end{aligned} \quad (\text{A.4})$$

Then by substitution

$$\begin{aligned} &\kappa_{t, s_t, t+1} h_{\tau, t+1} \quad (\text{A.5}) \\ &= \exp[-r_{t, s_t} - \frac{1}{2} \boldsymbol{\gamma}'_{t, s_t} \boldsymbol{\gamma}_{t, s_t} - \boldsymbol{\gamma}'_{t, s_t} \mathbf{L}_{s_{t+1}}^{-1} \boldsymbol{\eta}_{t+1} \\ &\quad - a_{s_{t+1}}(\tau - 1) - \mathbf{b}_{s_{t+1}}(\tau - 1)'(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}}) + a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})] \\ &= \exp[-r_{t, s_t} - \frac{1}{2} \boldsymbol{\gamma}'_{t, s_t} \boldsymbol{\gamma}_{t, s_t} - (\boldsymbol{\gamma}'_{t, s_t} \mathbf{L}_{s_{t+1}}^{-1} + \mathbf{b}_{s_{t+1}}(\tau - 1)') \boldsymbol{\eta}_{t+1} + \zeta_{\tau, s_t, s_{t+1}}] \\ &= \exp[-r_{t, s_t} - \frac{1}{2} \boldsymbol{\gamma}'_{t, s_t} \boldsymbol{\gamma}_{t, s_t} - (\boldsymbol{\gamma}_{t, s_t} + \mathbf{b}_{s_{t+1}}(\tau - 1)' \mathbf{L}_{s_{t+1}}) \boldsymbol{\omega}_{t+1} + \zeta_{\tau, s_t, s_{t+1}}] \\ &= \exp[-r_{t, s_t} - \frac{1}{2} \boldsymbol{\gamma}'_{t, s_t} \boldsymbol{\gamma}_{t, s_t} + \frac{1}{2} \boldsymbol{\Gamma}_{t, \tau} \boldsymbol{\Gamma}'_{t, \tau} + \zeta_{\tau, s_t, s_{t+1}}] \exp[-\frac{1}{2} \boldsymbol{\Gamma}_{t, \tau} \boldsymbol{\Gamma}'_{t, \tau} - \boldsymbol{\Gamma}_{t, \tau} \boldsymbol{\omega}_{t+1}] \end{aligned}$$

where

$$\begin{aligned} \zeta_{\tau, s_t, s_{t+1}} &= a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) - a_{s_{t+1}}(\tau - 1) - \mathbf{b}_{s_{t+1}}(\tau - 1)' \mathbf{G}_{s_{t+1}}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \\ \boldsymbol{\Gamma}_{t, \tau} &= \boldsymbol{\gamma}'_{t, s_t} + \mathbf{b}_{s_{t+1}}(\tau - 1)' \mathbf{L}_{s_{t+1}} \end{aligned}$$

and $\boldsymbol{\omega}_{t+1} = \mathbf{L}_{s_{t+1}}^{-1} \boldsymbol{\eta}_{t+1} \sim \mathcal{N}(0, \mathbf{I}_{k+m})$. 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

$$\mathbb{E}_t \left(\exp[-\frac{1}{2} \boldsymbol{\Gamma}_{t, \tau} \boldsymbol{\Gamma}'_{t, \tau} - \boldsymbol{\Gamma}_{t, \tau} \boldsymbol{\omega}_{t+1}] \right) = 1 \quad (\text{A.6})$$

we have that

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

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

$$\begin{aligned}
& \mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t] & (A.7) \\
& = \exp[-r_{t,s_t} - \frac{1}{2} \gamma'_{t,s_t} \gamma_{t,s_t} + \frac{1}{2} (\gamma'_{t,s_t} + \mathbf{b}_{s_{t+1}}(\tau-1)' \mathbf{L}_{s_{t+1}}) (\gamma'_{t,s_t} + \mathbf{b}_{s_{t+1}}(\tau-1)' \mathbf{L}_{s_{t+1}})' + \zeta_{\tau,s_t,s_{t+1}}] \\
& \approx -r_{t,s_t} + \gamma'_{t,s_t} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}}(\tau-1) + \frac{1}{2} (\mathbf{b}_{s_{t+1}}(\tau-1)' \mathbf{L}_{s_{t+1}} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}}(\tau-1)) + \zeta_{\tau,s_t,s_{t+1}} + 1 \\
& = -(\delta_{1,s_t} + \delta'_{2,s_t} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t})) + (\tilde{\gamma}_{s_t} + \Phi_{s_t} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}))' \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}}(\tau-1) \\
& + \frac{1}{2} (\mathbf{b}_{s_{t+1}}(\tau-1)' \mathbf{L}_{s_{t+1}} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}}(\tau-1)) + \zeta_{\tau,s_t,s_{t+1}} + 1
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_t = j] & = \sum_{s_{t+1}=j,k} p_{js_{t+1}} \mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t = j] & (A.8) \\
& = 1 \text{ where } k = j + 1.
\end{aligned}$$

Thus we have

$$\begin{aligned}
0 & = \sum_{s_{t+1}=j,k} p_{js_{t+1}} \{ \mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t = j] - 1 \} \text{ since } \sum_{s_{t+1}=j,k} p_{js_{t+1}} = 1 & (A.9) \\
& = p_{jj} (\mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1} = j, s_t = j] - 1) + p_{jk} (\mathbb{E}[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1} = k, s_t = j] - 1) \\
& \approx -p_{jj} (\delta_{1,j} + \delta'_{2,j} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t})) + p_{jj} (\tilde{\gamma}_j + \Phi_j (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}))' \mathbf{L}'_j \mathbf{b}_j(\tau-1) \\
& + \frac{1}{2} p_{jj} (\mathbf{b}_j(\tau-1)' \mathbf{L}_j \mathbf{L}'_j \mathbf{b}_j(\tau-1)) + p_{jj} \zeta_{\tau,j,j} \\
& - p_{jk} (\delta_{1,j} + \delta'_{2,j} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t})) + p_{jk} (\tilde{\gamma}_j + \Phi_j (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}))' \mathbf{L}'_k \mathbf{b}_k(\tau-1) \\
& + \frac{1}{2} p_{jk} (\mathbf{b}_k(\tau-1)' \mathbf{L}_k \mathbf{L}'_k \mathbf{b}_k(\tau-1)) + p_{jk} \zeta_{\tau,j,k}
\end{aligned}$$

Matching the coefficients on \mathbf{f}_t and setting the constant terms equal to zero we obtain the recursive equation for $a_{s_t}(\tau)$ 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

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

$$\prod_{s_t=1}^m \mathcal{N}(p_{s_t s_t} | \bar{p}_{s_t s_t}, \bar{\mathbf{V}}_{p_{s_t s_t}})$$

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

which we explain as follows.

First, the prior on p_{jj} ($j = 1, \dots, m$) 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 (8) are a function of p_{jj} , which eliminates any benefit from the use of a beta functional form. *Second*, we construct a 9×1 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 9×9 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% produces 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

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

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

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

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

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

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 σ^{*2} . Each σ_{i,s_t}^{*2} 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

$$u_0 \sim \mathcal{N}(0, V_u) \tag{B.1}$$

where $V_u = (1 - G_{11,1}^2)^{-1}$.

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 50 quarters. The median, 2.5% 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.

Appendix C MCMC Sampling

This section provides the details of the MCMC algorithm given in section 3.4. 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 3 change-point model. In contrast, a random-walk M-H algorithm takes about 2 days to complete 1 million iterations but with unknown reliability and much less efficient exploration (Chib and Ramamurthy (2010)).

Step 2 Sampling θ

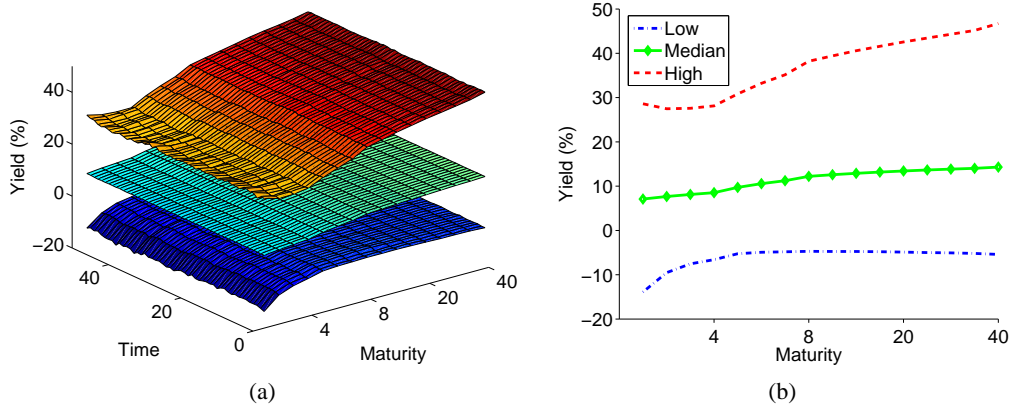


Figure 7: The implied prior term structure dynamics. *These graphs are based on 50,000 simulated draws of the parameters from the prior distribution. In the graphs on the left, the “Low”, “Median”, and “High” surfaces correspond to the 2.5%, 50%, and 97.5% quantile surfaces of the term structure dynamics 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 50 quarters.*

We sample θ conditioned on $(\mathbf{S}_n, u_0, \sigma^{*2})$ by the tailored randomized block M-H (TaRB-MH) algorithm introduced in Chib and Ramamurthy (2010). The schematics of the TaRB-MH algorithm are as follows. The parameters in θ are first randomly partitioned into various sub-blocks at the beginning of an iteration. Each of these sub-blocks 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 sub-blocks of θ

$$\theta_1, \theta_2, \dots, \theta_{h_g}$$

If ψ_{-i} denotes the collection of the parameters in ψ except θ_i , then the proposal density $q(\theta_i | \mathbf{y}, \psi_{-i})$ for the i th block conditioned on ψ_{-i} is constructed by a quadratic approximation at the mode of the current target density $\pi(\theta_i | \mathbf{y}, \psi_{-i})$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q(\theta_i | \mathbf{y}, \psi_{-i}) = St\left(\theta_i | \hat{\theta}_i, \mathbf{V}_{\hat{\theta}_i}, 15\right) \quad (\text{C.1})$$

where

$$\hat{\theta}_i = \arg \max_{\theta_i} \ln\{p(\mathbf{y} | \mathbf{S}_n, \theta_i, \psi_{-i})\pi(\theta_i)\} \quad (\text{C.2})$$

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

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

$$\begin{aligned} & \alpha(\boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\theta}_i^\dagger | \mathbf{y}, \boldsymbol{\psi}_{-i}) \\ &= \min \left\{ \frac{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^\dagger, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^\dagger)}{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^{(g-1)})} \frac{St(\boldsymbol{\theta}_i^{(g-1)} | \hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15)}{St(\boldsymbol{\theta}_i^\dagger | \hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15)}, 1 \right\}. \end{aligned} \quad (\text{C.3})$$

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 sub-blocks

$$\pi(\boldsymbol{\theta}_1 | \mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-1}), \pi(\boldsymbol{\theta}_2 | \mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-2}), \dots, \pi(\boldsymbol{\theta}_{h_g} | \mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-h_g}) \quad (\text{C.4})$$

are sequentially updated as above.

Step 3 Sampling the initial factor

Given the prior in equation (B.1), 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 (5). In the following, it is assumed that all the underlying coefficients are those in regime 0. Then

$$u_0 | \mathbf{f}_1, \boldsymbol{\theta} \sim \mathcal{N}_1(\bar{u}_0, \mathbf{U}_0) \quad (\text{C.5})$$

where

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

and on letting

$$\begin{aligned} \mathbf{G}_0 &= \begin{pmatrix} \mathbf{G}_{11,1} & \mathbf{G}_{12,1} \\ \mathbf{G}_{21,1} & \mathbf{G}_{22,1} \end{pmatrix}, \quad \boldsymbol{\Omega}_1 = \begin{pmatrix} \boldsymbol{\Omega}_{11,1} & \boldsymbol{\Omega}_{12,1} \\ \boldsymbol{\Omega}_{21,1} & \boldsymbol{\Omega}_{22,1} \end{pmatrix} \\ \mathbf{H}^* &= \mathbf{G}_{11,1} - \boldsymbol{\Omega}_{12,1} \boldsymbol{\Omega}_{22,1}^{-1} \mathbf{G}_{21,1}, \quad \boldsymbol{\Omega}_{11,1}^* = \boldsymbol{\Omega}_{11,1} - \boldsymbol{\Omega}_{12,1} \boldsymbol{\Omega}_{22,1}^{-1} \boldsymbol{\Omega}_{21,1} \\ u_1^* &= u_1 - \boldsymbol{\Omega}_{12,1} \boldsymbol{\Omega}_{22,1}^{-1} (\mathbf{m}_1 - \boldsymbol{\mu}_{m,1}) + (\boldsymbol{\Omega}_{12,1} \boldsymbol{\Omega}_{22,1}^{-1} \mathbf{G}_{22,1} - \mathbf{G}_{12,1}) (\mathbf{m}_0 - \boldsymbol{\mu}_{m,1}) \end{aligned}$$

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 $j = 1, 2, \dots, m + 1$ and $t = 1, 2, \dots, n$ by calculating

$$\Pr[s_t = j|I_t, \boldsymbol{\psi}] = \sum_{i=j-1}^j \Pr[s_{t-1} = i, s_t = j|I_t, \boldsymbol{\psi}] \quad (\text{C.6})$$

where

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

This can be done by the equations (D.4)-(D.7).

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 $\{1, 2, \dots, m + 1\}$. 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 $t = 1, 2, \dots, n - 1$ we sequentially calculate

$$\begin{aligned} \Pr[s_t = j|I_t, s_{t+1} = k, S^{t+2}, \boldsymbol{\psi}] &= \Pr[s_t = j|I_t, s_{t+1} = k, \boldsymbol{\psi}] \quad (\text{C.7}) \\ &= \frac{\Pr[s_{t+1} = k|s_t = j] \Pr[s_t = j|I_t, \boldsymbol{\psi}]}{\sum_{j=k-1}^k \Pr[s_{t+1} = k|s_t = j] \Pr[s_t = j|I_t, \boldsymbol{\psi}]} \end{aligned}$$

where $S^{t+1} = \{s_{t+1}, \dots, s_n\}$ 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

$i \in \{1, 2, \dots, 7, 9, \dots, 16\}$ and $j = 1, 2, \dots, m + 1$, $\sigma_{i,j}^{*2}$ is sampled from

$$\mathbf{IG} \left\{ \frac{\bar{v} + \sum_{t=1}^n I(s_t = j)}{2}, \frac{\bar{d} + \sum_{t=1}^n d_{i,j} I(s_t = j) (R_{ti} - \bar{a}_{i,j} - \bar{\mathbf{b}}'_{i,j} (\mathbf{f}_t - \boldsymbol{\mu}_j))^2}{2} \right\} \quad (\text{C.8})$$

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

Appendix D Marginal Likelihood Computation

The marginal likelihood of any given model is obtained as

$$m(\mathbf{y}) = \int p(\mathbf{y} | \mathbf{S}_n, \boldsymbol{\psi}) p(\mathbf{S}_n | \boldsymbol{\psi}) \pi(\boldsymbol{\psi}) d(\mathbf{S}_n, \boldsymbol{\psi}) \quad (\text{D.1})$$

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

$$m(\mathbf{y}) = \frac{p(\mathbf{y} | \boldsymbol{\psi}^*) \pi(\boldsymbol{\psi}^*)}{\pi(\boldsymbol{\psi}^* | \mathbf{y})} \quad (\text{D.2})$$

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

$$\ln \hat{m}(\mathbf{y}) = \ln p(\mathbf{y} | \boldsymbol{\psi}^*) + \ln \pi(\boldsymbol{\psi}^*) - \ln \hat{\pi}(\boldsymbol{\psi}^* | \mathbf{y}) \quad (\text{D.3})$$

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 $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ is, by definition,

$$p(\mathbf{y} | \boldsymbol{\psi}) = \sum_{t=0}^{n-1} \ln p(\mathbf{y}_{t+1} | I_t, \boldsymbol{\psi}) \quad (\text{D.4})$$

where

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

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 (11), whereas the second term can be calculated from the law of total probability as

$$\Pr[s_t = j, s_{t+1} = k | I_t, \boldsymbol{\psi}] = p_{jk} \Pr[s_t = j | I_t, \boldsymbol{\psi}] \quad (\text{D.5})$$

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

$$\Pr[s_{t+1} = k | I_{t+1}, \boldsymbol{\psi}] = \sum_{j=1}^{m+1} \Pr[s_t = j, s_{t+1} = k | I_{t+1}, \boldsymbol{\psi}] \quad (\text{D.6})$$

where

$$\Pr[s_t = j, s_{t+1} = k | I_{t+1}, \boldsymbol{\psi}] = \frac{p[\mathbf{y}_{t+1} | I_t, s_t = j, s_{t+1} = k, \boldsymbol{\psi}] \Pr[s_t = j, s_{t+1} = k | I_t, \boldsymbol{\psi}]}{p[\mathbf{y}_{t+1} | I_t, \boldsymbol{\psi}]} \quad (\text{D.7})$$

This completes the calculation of the likelihood function.

References

- Ang, A. and Bekaert, G. (2002), “Regime switches in interest rates,” *Journal of Business and Economic Statistics*, 20, 163–82.
- Ang, A., Bekaert, G., and Wei, M. (2008), “The term structure of real rates and expected inflation,” *Journal of Finance*, 63, 797–849.
- Ang, A., Dong, S., and Piazzesi, M. (2007), “No-arbitrage Taylor rules,” Columbia University working paper.
- Ang, A. and Piazzesi, M. (2003), “A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables,” *Journal of Monetary Economics*, 50, 745–787.
- Bansal, R. and Zhou, H. (2002), “Term structure of interest rates with regime shifts,” *Journal of Finance*, 57(5), 1997–2043.
- Chen, R. and Scott, L. (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.
- (1996), “Calculating posterior distributions and modal estimates in Markov mixture models,” *Journal of Econometrics*, 75, 79–97.
- (1998), “Estimation and comparison of multiple change-point models,” *Journal of Econometrics*, 86, 221–241.
- (2001), “Markov chain Monte Carlo methods: computation and inference,” in *Handbook of Econometrics*, eds. Heckman, J. and Leamer, E., North Holland, Amsterdam, vol. 5, pp. 3569–3649.
- Chib, S. and Ergashev, B. (2009), “Analysis of multi-factor affine yield curve Models,” *Journal of the American Statistical Association*, 104, 1324–1337.
- Chib, S. and Greenberg, E. (1995), “Understanding the Metropolis-Hastings algorithm,” *American Statistician*, 49, 327–335.
- Chib, S. and Jeliazkov, I. (2001), “Marginal likelihood from the Metropolis-Hastings output,” *Journal of the American Statistical Association*, 96, 270–281.
- Chib, S. and Ramamurthy, S. (2010), “Tailored Randomized-block MCMC Methods with Application to DSGE Models,” *Journal of Econometrics*, 155, 19–38.
- 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.
- Duffie, G. and Kan, R. (1996), “A yield-factor model of interest rates,” *Mathematical Finance*, 6, 379–406.
- Gelfand, A. E. and Ghosh, S. K. (1998), “Model choice: A minimum posterior predictive loss approach,” *Biometrika*, 85, 1–11.
- Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007), “The U.S. treasury yield curve: 1961 to the present,” *Journal of Monetary Economics*, 54, 2291–2304.

- Kim, C. J., Nelson, C. R., and Piger, J. (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.
- Koop, G. and Potter, S. (2007), “Forecasting and estimating multiple change-point models with an unknown number of change points,” *Review of Economic Studies*, 74(3), 763–89.
- Rudebusch, G., Sack, B. P., and Swanson, E. T. (2007), “Macroeconomic implications of changes in the term premium,” *Federal Reserve Bank of St Louis Review*, 89, 241–269.
- Rudebusch, G. and Wu, T. (2008), “A macro-finance model of the term structure, monetary policy and the economy,” *Economic Journal*, 118, 906–926.