

Change Points in Term-Structure Models: Pricing, Estimation and Forecasting*

SIDDHARTHA CHIB[†]

KYU HO KANG[‡]

(Washington University in St. Louis)

March 2009

Abstract

In this paper we theoretically and empirically examine structural changes in a dynamic term-structure model of zero-coupon bond yields. To do this, we develop a new arbitrage-free one latent and two macro-economics factor affine model to price default-free bonds when the parameters in the dynamics of the factor evolution, in the model of the market price of factor risks, and in the process of the stochastic discount factor, are all subject to change at unknown time points. The bonds in our set-up can be priced straightforwardly once the change-point model is re-formulated in the manner of Chib (1998) as a specific unidirectional Markov process with restricted transition probabilities. We consider four versions of our general model - with 0, 1, 2 and 3 change-points - to a collection of 16 yields measured quarterly over the period 1972:I to 2007:IV. Our empirical approach to inference is fully Bayesian with priors set up to reflect the assumption of a positive term-premium. The use of Bayesian techniques is particularly relevant because the models are high-dimensional (containing 168 parameters in the situation with 3 change-points) and non-linear, and because it is more straightforward to compare our different change-point models from the Bayesian perspective. Our estimation results indicate that the model with 3 change-points is most supported by the data (in comparison with models with 0, 1 and 2 change-points) and that the breaks occurred in 1980:II, 1986:I and 1995:II. These dates correspond (in turn) to the time of a change in monetary policy, the onset of what is termed the great moderation, and the start of technology driven period of economic growth. We also utilize the Bayesian framework to derive the out-of-sample predictive densities of the term-structure. We find that the forecasting performance of our proposed model is substantially better than that of the other models we examine.(JEL G12,C11,E43)

*We thank Ed Greenberg, Wolfgang Lemke, James Morley, Hong Liu, Yongs Shin and Srikanth Ramamurthy for their thoughtful and useful comments on the paper.

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

[‡]*Address for correspondence:* Department of Economics, Washington University in St. Louis, Campus Box 1208, 1 Bookings Drive, St. Louis, MO 63130. E-mail: khkang@artsci.wustl.edu.

1 Introduction

Affine term structure models provide a flexible approach for modeling the dynamics of bond prices and yields. This is especially true of multi-factor affine models (for example, Duffie and Kan (1996) where the factors under the physical measure follow a stationary Gaussian VAR process and Dai and Singleton (2000) where the factors follow a CIR type process) and of multi-factor models that include macro-economic factors (for example, Ang, Dong, and Piazzesi (2007) and Chib and Ergashev (2009)) and/or permit the possibility of regime-changes (for example, Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), Ang, Bekaert, and Wei (2008)). The literature on these topics is quite impressive and the recent strengthening of links between macro-economics and finance in this area is a rather promising development.

One of our main objectives in this paper is to develop a new multi-factor affine model with macro factors within the context of a change-point model of regime-changes, rather than the Markov-switching model of regime-changes that has been used in the existing literature. In our model, all the parameters of the model, including those in the dynamics of the factor evolution, in the model of the market price of factor risks, and in the process of the stochastic discount factor (SDF), are subject to change at unknown time points. The defining feature of the change-point model is that the parameters across change-points are different. We model parameter-changes (equivalently, regime-changes) in this way because we feel that this assumption is particularly appropriate in affine models with macro factors. In such models, when the macro factors are the inflation rate and the growth rate of GDP, the short rate is essentially the Taylor rule of macroeconomics. The Taylor rule reflects the behavior of monetary policy. If one believes that monetary policy is constant over epochs but different across epochs, then it is reasonable to assume that the processes of the factors and the SDF should also be different across epochs. Our second reason for adopting the change-point model is empirical – the filtered regime probabilities in the empirical analysis of Dai et al. (2007) and Ang and Bekaert (2002), strongly suggest that the same regime has prevailed after 1986, a pattern that is more suggestive of a change-point rather than a Markov switching

process.

Our modeling of the factor process, the market price of factor risk and the SDF is of course similar but not identical to that of the existing literature. In fact, these primary building blocks of affine models can be, and have been, specified in different ways. Like Dai et al. (2007), our model of regime-change is in the context of models with Gaussian factors although we go beyond their exclusive reliance on latent factors to include two macro factors. In addition, we follow Bansal and Zhou (2002) to assume that the factor loadings are regime-specific. In this we depart from Dai et al. (2007) and Ang et al. (2008) where the factor loading matrix is assumed to be constant across regimes. Under our time-varying factor loading assumption bond prices can only be derived by an approximate solution to the risk-neutral pricing formula, as in Bansal and Zhou (2002). In our view this is a minor inconvenience since the data seems to support the assumption that the factor loadings vary across regimes. In our work, the dynamics of the factors at time t depend on both the regime s_t at time t and the regime s_{t-1} at time $t-1$. On the other hand, in Dai et al. (2007), the factor dynamics at time t depend on the regime s_{t-1} in period $(t-1)$, rather than on s_t . In Bansal and Zhou (2002) and Ang et al. (2008), the factor dynamics depend on the current regime s_t . Finally, in contrast to Dai et al. (2007), our model does not have regime-shift risk. It is not possible to identify this risk when each regime-shift occurs once. It should be noted that this risk cannot be directly isolated in the models of Ang et al. (2008) and Bansal and Zhou (2002) because their modeling of the SDF is such that this risk 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 et al. (2007).

We apply our model to the largest collection of yields that has been considered in this literature. In particular, we fit up to 4 regime models on 16 quarterly yields containing 168 parameters. Our method of inference is Bayesian with a prior distribution on the parameters that reflects the assumption of a positive term-premium, as in Chib and Ergashev (2009). We adopt the Bayesian perspective because it is virtually impossible to find maximum likelihood estimates given the size of the parameter space, the severe

non-linearities, and potential multi-modalities in the likelihood surface. Our Bayesian approach, on the other hand, is both feasible and reliable. It also offers a formal way to compare different versions of our model and provides the basis for calculating dynamic predictive effects of the macro factors on the yield curve and the out of sample predictive densities, all desirable inferential goals.

We apply our techniques to four versions of the general model - with 0, 1, 2 and 3 change-points - and compare these various versions (that we refer to as the C0L1M2, C1L1M2, C2L1M2 and C3L1M2 models, respectively) in terms of marginal likelihoods and Bayes factors. The main findings from our empirical analysis (from quarterly yields of sixteen US T-bills between 1972:I and 2007:IV) are as follows. The C3L1M2 3 change-point model is the one that is most supported by the data (in comparison with models with 0, 1 and 2 change-points) and that the breaks occurred in 1980:II, 1985:IV and 1995:II. These change-points can be attributed, in turn, to changes in monetary policy, the onset of what is termed the great moderation, and the start of the technology driven period of economic growth. One striking feature emerging from this finding is that the most recent break occurs in 1995, not 1986, as is commonly believed. The differences in the distribution of the term-structure can be seen in Figure 1 where we display the 5%, 50% and 95% quantiles of the yield curve in each of the four regimes. In addition, there are substantial differences in the parameters across regimes. In particular we find support for our assumption that the mean-reversion parameters in the factor dynamics are regime-specific. Finally, we show that the predictive performance of our best model is substantially better than that of the other models we consider.

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.

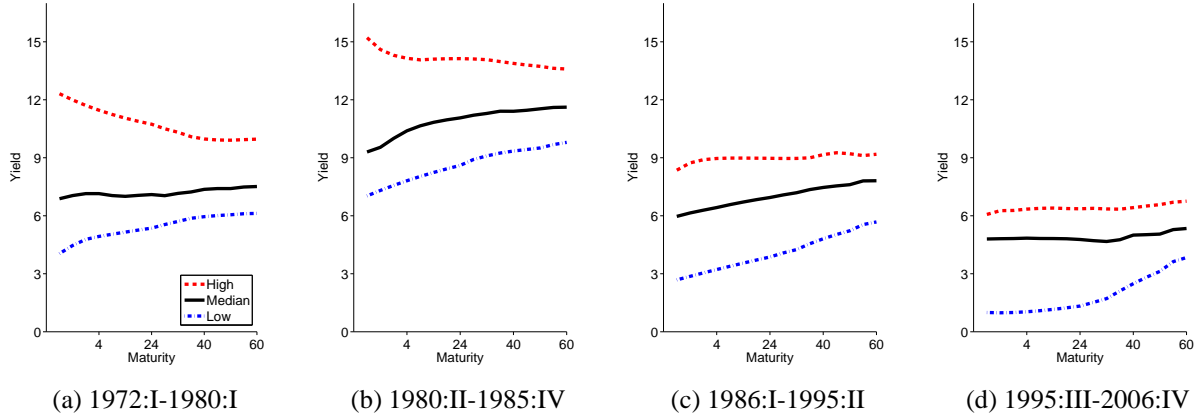


Figure 1: Term structure of interest rates. *Data summary of the term-structure - data obtained from <http://www.federalreserve.gov/econresdata/researchdata.htm>. The graphs display the 5%, 50% and 95% quantiles of the yield curve for bonds of maturity 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 16, 20, 24, 28, 36 and 40 quarters.*

2 Model Specification

We describe our model in two steps. First, we characterize the change-point process and then second, as dictated by the framework of Duffie and Kan (1996), we define the exogenous factors \mathbf{f}_t (containing both latent and observed macro-economic variables), the stochastic evolution equation of the factors, the model of the market price of factor risks $\bar{\gamma}_{t,s_t}$, and the model of the stochastic discount factor $\kappa_{t,t+1}$. Given these ingredients, we then derive the prices of our default-free zero coupon bonds.

2.1 Change-point Process

Suppose that the parameters in the evolution equation of the factors, the market-price of factor risks, and the SDF, are subject to change at the unknown times $\{t_1^*, t_2^*, \dots, t_q^*\}$. These q change-points give rise to $(q + 1)$ distinct regimes. Unlike a Markov switching process, the regimes induced by the change-points are not revisited once vacated.

We now present a reformulation of the change-point model given in Chib (1998) that facilitates risk-neutral pricing and the subsequent Bayesian estimation of the model. Let $\{s_t\}$ be a discrete stochastic process that takes one of the values $\{1, 2, \dots, q + 1\}$ such that $s_t = j$ indicates that the t th observation has been drawn from the j th regime. Now

assume that s_t is Markov and its distribution is governed by the homogenous transition probability matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{11} & 0 & \cdots & 0 \\ 0 & p_{22} & 1 - p_{22} & \cdots & 0 \\ 0 & 0 & p_{33} & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & p_{q+1q+1} \end{bmatrix} \quad (2.1)$$

where $p_{jk} = \Pr[s_{t+1} = k | s_t = j]$. As shown in Chib (1998), this Markov process can be mapped into a change-point process by letting $p_{jk} = 1 - p_{jj}$ ($j = 1, 2, \dots, q$), $k = (j + 1)$ and $p_{q+1q+1} = 1$. Thus, under this specification, $s_t = j$ either stays at the current value j or jumps to the next higher value $(j + 1)$. As required, return visits to a previously occupied state are not possible and the last state is absorbing.

This formulation of the change-point model in terms of a restricted unidirectional Markov process facilitates pricing (as we show below). It also 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).

2.2 Factor Specification

Following the new affine term-structure literature, we explain the dynamics of bond-prices in terms of the dynamics of both latent and observed macro-economics variables. Let \mathbf{f}_t denote the factors. For concreteness, we assume (as in our empirical work) that these consist of one latent variable u_t and two observed macroeconomic factors \mathbf{m}_t . We next suppose that the evolution of these exogenous factors is governed by the Gaussian regime-specific mean-reverting first-order autoregression

$$\mathbf{f}_{t+1} = \begin{pmatrix} u_{t+1} \\ \mathbf{m}_{t+1} \end{pmatrix} | \mathbf{f}_t, s_{t+1}, s_t \sim \mathcal{N}_3(\boldsymbol{\mu}_{s_{t+1}} + \mathbf{G}_{s_{t+1}} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \Omega_{s_{t+1}}), \quad (2.2)$$

where $\mathcal{N}_3(\cdot, \cdot)$ denotes the 3-dimensional normal distribution, and for $(j = 1, 2, \dots, q + 1)$, $\boldsymbol{\mu}_j$ is a 3×1 vector, \mathbf{G}_j and Ω_j are 3×3 matrices. It is important to note that under this specification

$$E[\mathbf{f}_{t+1} | \mathbf{f}_t, s_{t+1} = k, s_t = j] = \boldsymbol{\mu}_k + \mathbf{G}_k (\mathbf{f}_t - \boldsymbol{\mu}_j)$$

and

$$V[\mathbf{f}_{t+1} | \mathbf{f}_t, s_{t+1} = k, s_t = j] = \Omega_k$$

The conditional expectation, therefore, depends on both $\boldsymbol{\mu}_k$ and $\boldsymbol{\mu}_j$. 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} in regime j . The matrices $\{\mathbf{G}_j\}$ can also be interpreted in the same way: as the mean-reversion parameters in regime j .

2.3 Market Price of Factor Risk

We now move to our model of $\boldsymbol{\gamma}_{t,s_t}$, the vector of market prices of factor risks in regime s_t . Following Dai et al. (2007), we assume that

$$\boldsymbol{\gamma}_{t,s_t} = \bar{\boldsymbol{\gamma}}_{s_t} + \boldsymbol{\Phi}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \quad (2.3)$$

where $\bar{\boldsymbol{\gamma}}_{s_t} : 3 \times 1$ 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 $\{\bar{\boldsymbol{\gamma}}_{s_t}, \boldsymbol{\Phi}_{s_t}\}$ as the factor-risk parameters. Note that in this specification $\boldsymbol{\gamma}_{t,s_t}$ is the same across maturities but different across regimes. Negative market prices of risk have the effect of generating a positive term premium. We exploit this feature to develop a proper prior distribution on the risk parameters.

2.4 Short-rate and the SDF

The instantaneous short rate r_{t,s_t} in regime s_t is also assumed to be an affine function of the factors

$$r_{t,s_t} = \delta_{1,s_t} + \boldsymbol{\delta}'_{2,s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \quad (2.4)$$

It is essential that δ_{1,s_t} be regime-specific if the aim is to allow for shifts in the level of the term structure. We also allow $\boldsymbol{\delta}_{2,s_t} : 3 \times 1$ to be regime-dependent in order to capture shifts in the effects of the macroeconomic factors on the term structure. This is a departure from both Ang et al. (2008) and Dai et al. (2007) where the coefficient on

the factors is constrained to be constant across regimes to get tractable expressions for the bond-prices. We now assume that the SDF $\kappa_{t,t+1}$ in time t is given by

$$\kappa_{t,t+1} = \exp \left(-r_{t,s_t} - \frac{1}{2} \gamma'_{t,s_t} \gamma_{t,s_t} - \gamma'_{t,s_t} \boldsymbol{\omega}_{t+1} \right) \quad (2.5)$$

where $\boldsymbol{\omega}_{t+1}$ is a vector of regime independent normalized $\mathcal{N}_3(0, \mathbf{I}_3)$ factor shocks. Because $\kappa_{t,t+1}$ is independent of s_{t+1} conditional on s_t , it is easily checked that $\mathbf{E}[\kappa_{t,t+1} | \mathbf{f}_t, s_t = j]$ equals the price of a zero coupon bond with ($\tau = 1$):

$$\begin{aligned} \mathbf{E}[\kappa_{t,t+1} | \mathbf{f}_t, s_t = j] &= \sum_{s_{t+1}=j}^{j+1} p_{j s_{t+1}} \mathbf{E}[\kappa_{t,t+1} | \mathbf{f}_t, s_t = j, s_{t+1}] \\ &= \exp(-r_{t,j}) \quad j \in \{1, 2, \dots, q\} \end{aligned} \quad (2.6)$$

Thus, the SDF satisfies the intertemporal no-arbitrage condition (Dai et al. (2007)).

2.5 Bond Prices

We now derive the price of our default-free zero-coupon bonds. Let $P_t(s_t, \tau)$ denote the price of the bond at time t in regime s_t that matures in period $(t + \tau)$. From the risk-neutral pricing formula we have that

$$P_t(s_t, \tau) = E_t[\kappa_{t,t+1} P_{t+1}(s_{t+1}, \tau - 1)] \quad (2.7)$$

where E_t is the expectation under the physical measure conditioned on (\mathbf{f}_t, s_t) . The expectation is over the factor shocks in $(t + 1)$ and the two possible values that s_{t+1} can take given s_t .

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

$$P_t(s_t, \tau) = \exp(-\tau R_{\tau t}) \quad (2.8)$$

where $R_{\tau t}$ is the bond's continuously compounded yield given by

$$R_{\tau t} = \frac{1}{\tau} a_{s_t}(\tau) + \frac{1}{\tau} \mathbf{b}_{s_t}(\tau)' (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \quad (2.9)$$

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 .

By the usual techniques, we find the expressions for the latter functions by the method of undetermined coefficients. By the law of the iterated expectation, the risk-neutral pricing formula in (2.7) can be expressed as

$$1 = E_t \left\{ E \left[\kappa_{t,t+1} \frac{P_{t+1}(s_{t+1}, \tau - 1)}{P_t(s_t, \tau)} \mid \mathbf{f}_t, s_t, s_{t+1} \right] \right\} \quad (2.10)$$

where the inside expectation is conditioned on s_{t+1} as well as s_t . One now substitutes $P_t(s_t, \tau)$ and $P_{t+1}(s_{t+1}, \tau - 1)$ from (2.8) and (2.9) into this expression. We integrate out s_{t+1} after a log-linearization that is discussed in Appendix A. We match common coefficients and solve for the unknown functions. When $j \in \{1, \dots, q\}$ and $k = j + 1$, this procedure produces the following recursive system for the unknown functions

$$\begin{aligned} a_j(\tau) &= \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \delta_{1,j} - \bar{\gamma}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1) - \mathbf{b}_j(\tau - 1)' \mathbf{L}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1)/2 + a_j(\tau - 1) \\ \delta_{1,j} - \bar{\gamma}_j \mathbf{L}'_k \mathbf{b}_k(\tau - 1) - \mathbf{b}_k(\tau - 1)' \mathbf{L}_k \mathbf{L}'_k \mathbf{b}_k(\tau - 1)/2 + a_k(\tau - 1) \end{pmatrix} \\ \mathbf{b}_j(\tau) &= \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{2,j} + (\mathbf{G}_j - \mathbf{L}_j \boldsymbol{\Phi}_j)' \mathbf{b}_j(\tau - 1) \\ \boldsymbol{\delta}_{2,j} + (\mathbf{G}_k - \mathbf{L}_k \boldsymbol{\Phi}_j)' \mathbf{b}_k(\tau - 1) \end{pmatrix} \end{aligned} \quad (2.11)$$

and when $j = q + 1$ we have that

$$\begin{aligned} a_j(\tau) &= \delta_{1,j} - \bar{\gamma}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1) - \mathbf{b}_j(\tau - 1)' \mathbf{L}_j \mathbf{L}'_j \mathbf{b}_j(\tau - 1)/2 + a_j(\tau - 1) \\ \mathbf{b}_j(\tau) &= \boldsymbol{\delta}_{2,j} + (\mathbf{G}_j - \mathbf{L}_j \boldsymbol{\Phi}_j)' \mathbf{b}_j(\tau - 1) \end{aligned} \quad (2.12)$$

where \mathbf{L}_j is the Cholesky decomposition of Ω_j and 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 $s_t = j$ ($j \leq q$) 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.

Figure 2 summarizes the economy that we have just described in terms of a directed acyclic graph. In the beginning of period t , a regime realization occurs. This realization is governed with the regime in the previous period as indicated by the direction of the arrow connecting s_{t-1} to s_t . Then given the regime at time t , the corresponding

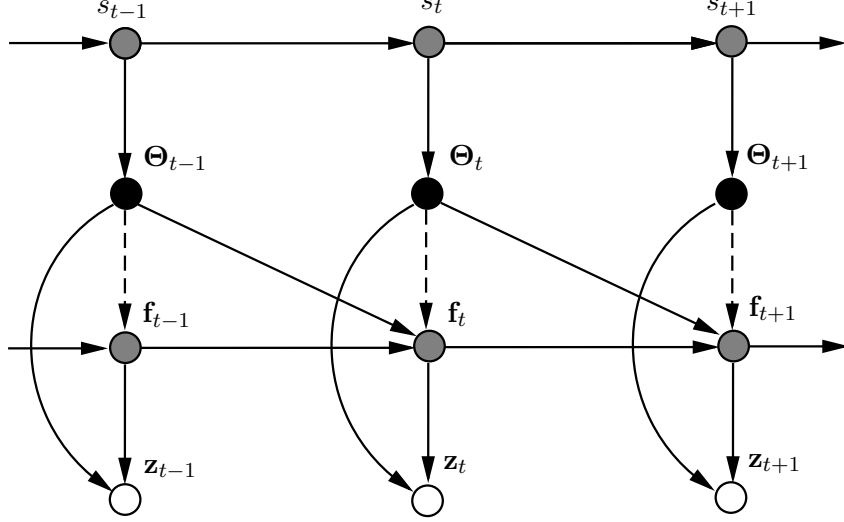


Figure 2: Directed graph of model linkages.

model parameters Θ_t are taken from the full collection of model parameters. These determine the functions $a_{s_t}(\tau)$ and $\mathbf{b}_{s_t}(\tau)$ according to the recursions in (2.11) and (2.12). Conditioned on the parameters and \mathbf{f}_{t-1} , \mathbf{f}_t is generated by the regime-specific autoregressive process in (2.2). Finally, from (2.9), $a_{s_t}(\tau)$, $\mathbf{b}_{s_t}(\tau)$ and \mathbf{f}_t determine the yields of all maturities. Notice that in Dai et al. (2007) the dashed line in figure 2 is absent since \mathbf{f}_t is assumed to be drawn independently of s_t .

2.6 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 $s_t = j$, it can be calculated as

$$\begin{aligned}
 \text{Term premium} &= (\tau - 1)\text{Cov}(\ln \kappa_{t,t+1}, R_{\tau-1,t+1} | \mathbf{f}_t, s_t = j) & (2.13) \\
 &= (\tau - 1) \sum_{s_{t+1}=j}^{j+1} p_{j s_{t+1}} \text{Cov}(\ln \kappa_{t,t+1}, R_{\tau-1,t+1} | \mathbf{f}_t, s_{t+1}, s_t = j) \\
 &= -p_{jj} \mathbf{b}_j (\tau - 1)' \mathbf{L}_j \boldsymbol{\gamma}_{t,j} - p_{jk} \mathbf{b}_k (\tau - 1)' \mathbf{L}_k \boldsymbol{\gamma}_{t,j}
 \end{aligned}$$

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 $\gamma_{t,j}$, 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 below we calculate this regime-specific term premium for each time period in the sample.

3 Estimation

One common approach for estimating affine models is by maximum likelihood. This is a reasonable approach when the size of the model (as measured by the number of parameters) is small. In higher-dimensional models, it becomes virtually impossible to find the maximum likelihood estimates due to the complicated nature of the affine model and irregularities (boundary maxima, multi-modalities etc) of the likelihood surface. For this reason, interest in the alternative Bayesian approach implemented by Markov chain Monte Carlo (MCMC) methods (Ang et al. (2007) and Chib and Ergashev (2009)) has grown. This approach is particularly attractive because one of our models contains 168 parameters. Our approach to inference is grounded in the developments that appear in Chib and Ergashev (2009) and Chib and Ramamuthy (2009). The former paper deals with a 3 factor macro factor affine model without regime changes. It introduces the useful idea of building a prior distribution on the parameters that embodies the assumption of a positive term-premium. We follow the same strategy in the regime-change model. 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.

3.1 Empirical State Space Formulation

Let the collection of yields at each time t be denoted by

$$\mathbf{R}_t = (R_{1t}, R_{2t}, \dots, R_{16t})' \quad (3.1)$$

where $R_{it} = R_{\tau_i, t}$ and τ_i is the i th maturity (in quarters). In the application, the set of maturities is given by

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

Following Chen and Scott (2003) and Dai et al. (2007), we assume that one basis yield (the eighth in the above list) is priced exactly by the model.

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 (2.11) - (2.12). Then the measurement equation for the i th yield is given by

$$R_{it} = \bar{a}_{i,s_t} + \bar{\mathbf{b}}_{i,s_t} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\varepsilon}_{it}$$

where $\boldsymbol{\varepsilon}_{it} | \sigma_{i,s_t}^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{i,s_t}^2)$ is the normally distributed pricing error, and the variance for the basis yield σ_{8,s_t}^2 is zero in each regime s_t .

Let $\bar{\mathbf{a}}_{s_t} = (\bar{a}_{1,s_t}, \bar{a}_{2,s_t}, \dots, \bar{a}_{16,s_t})'$, $\bar{\mathbf{B}}_{s_t} = (\bar{\mathbf{b}}_{1,s_t}, \bar{\mathbf{b}}_{2,s_t}, \dots, \bar{\mathbf{b}}_{16,s_t})'$ and

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{R}_t \\ \mathbf{m}_t \end{pmatrix}$$

Then, we express our set of measurement equations in vector-matrix form as

$$\mathbf{y}_t = \underbrace{\begin{pmatrix} \bar{\mathbf{a}}_{s_t} \\ \boldsymbol{\mu}_{m,s_t} \end{pmatrix}}_{\mathbf{a}_{s_t}} + \underbrace{\begin{pmatrix} \bar{\mathbf{B}}_{s_t} \\ \mathbf{J}_{2 \times 3} \end{pmatrix}}_{\mathbf{B}_{s_t}} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \underbrace{\begin{pmatrix} \mathbf{I}_{16} \\ \mathbf{0}_{2 \times 16} \end{pmatrix}}_{\mathbf{T}} \boldsymbol{\varepsilon}_t \quad (3.2)$$

where the yields \mathbf{R}_t are augmented by the identity $\mathbf{m}_t = \mathbf{m}_t$ involving the macro variables \mathbf{m}_t to ensure that the Kalman updates of the factors corresponding to the macro-factors have no error, $\mathbf{J}_{2 \times 3} = (\mathbf{0}_{2 \times 1}, \mathbf{I}_2)$, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}_{16}(0, \boldsymbol{\Sigma}_{s_t})$ and $\boldsymbol{\Sigma}_{s_t} = \text{diag}(\sigma_{1,s_t}^2, \sigma_{2,s_t}^2, \dots, \sigma_{16,s_t}^2)$.

The transition equation is characterized by the evolution of the factors in (2.2). As initial conditions of the factors, we assume that \mathbf{m}_0 , the values of the macro variables

at time 0, is known from the data, and u_0 , the latent factor at time 0, which is assumed to be independent of \mathbf{m}_0 , follows the steady-state distribution in regime 0

$$u_0 \sim N(0, V_u) \quad (3.3)$$

where $V_u = (\mathbf{1} - G_{11,0}^2)^{-1}$. The equations (3.2), and (2.2) complete the state-space form of our model.

3.2 Identification

We impose the standard identifying restrictions on the parameters of our model. First, we set $\boldsymbol{\mu}_{u,s_t} = 0$ which means that δ_{1,s_t} is a free parameter. The mean of the short rate conditional on s_t is thus δ_{1,s_t} . Second, the (1,1) element of \mathbf{L}_{s_t} is also set to be 1/400 for normalization and 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

$$R = \{\mathbf{G}_j, \delta_{21,j} | \delta_{21,j} \geq 0, |eig(\mathbf{G}_j)| < 1 \text{ for } j = 1, 2, \dots, q + 1\} \quad (3.4)$$

All the constraints are enforced through the prior distribution.

3.3 Prior Distribution

Let $\boldsymbol{\theta}$ denote the free parameters in $(\mathbf{G}_{s_t}, \boldsymbol{\mu}_{m,s_t}, \boldsymbol{\delta}_{s_t}, \bar{\gamma}_{s_t}, \boldsymbol{\Phi}_{s_t}, \mathbf{L}_{s_t}, \mathbf{P})$. These along with the diagonal elements of $\{\boldsymbol{\Sigma}_{s_t}\}$ (except for the basis variance in each state which are zero) form the set of unknown parameters. 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 defensible and reasonable prior distribution can be constructed by thinking about 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 within each regime. We arrive at our prior distribution in this way for each of the four models we consider - with 0, 1, 2 and 2 change-points. We feel that this is a satisfactory and sensible solution to an otherwise challenging problem.

Broadly, our prior is constructed along the following lines.

- The parameters in $(\mathbf{G}_{s_t}, \boldsymbol{\mu}_{m,s_t}, \boldsymbol{\delta}_{s_t}, \bar{\boldsymbol{\gamma}}_{s_t}, \boldsymbol{\Phi}_{s_t}, \mathbf{L}_{s_t}, \mathbf{P})$ and those in Σ_{s_t} are assumed to be mutually independent.
- Normal and truncated normal distributions are used to represent the prior uncertainty of $\boldsymbol{\theta}$. For example, the prior on p_{jj} ($j = 1, \dots, q$) is normal with a mean of 0.98 and a standard deviation of 1, truncated to the interval $(0, 1)$, and the distribution on the free parameters in $(\mathbf{G}_{s_t}, \boldsymbol{\delta}_{21,s_t})$ is normal truncated to the region R . Table 1 then summarizes the first and second moment of our normal prior on $\boldsymbol{\theta}$ in the 3-change point model. This prior is modified in obvious ways to arrive at the prior distribution of our model with 0, 1 and 2 change-points.
- Finally, the 15 free parameters of Σ_{s_t} are transformed into $\Sigma_{s_t}^*$ through the transformation $\sigma_{i,s_t}^{*2} = d_{i,s_t} \sigma_{i,s_t}^2$ where the d_{i,s_t} are known multipliers introduced to ensure that the σ_{i,s_t}^{*2} are much bigger than σ_{i,s_t}^2 . We let $\boldsymbol{\sigma}^{*2} = \{\sigma_{i,s_t}^{*2}\}$ denote the entire collection of these transformed variances. In a 4 regime model, the dimension of $\boldsymbol{\sigma}^{*2}$ is 60. We then assume that each σ_{i,s_t}^{*2} has an inverse-gamma prior distribution with a mean of 10 and standard deviation of 14.

To show what these assumptions imply for the outcomes in the 3 change-point model, we simulate the parameters 10,000 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 3. It can be seen that the simulated prior

	Regime 1			Regime 2			Regime 3			Regime 4		
G	diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)		
μ $\times 400$	0.00	6.50 (2.00)	3.00 (1.00)	0.00	8.50 (2.00)	3.00 (1.00)	0.00	5.00 (2.00)	3.00 (1.00)	0.00	2.50 (2.00)	3.00 (1.00)
δ_1 $\times 400$		6.00 (4.00)			8.00 (4.00)			5.00 (4.00)			3.00 (4.00)	
δ_2	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)
γ	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)
Φ	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
λ	(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)		

Table 1: Prior normal distribution for the model parameters in θ for 3 change point model *This table presents the prior mean and standard deviation of the parameters in θ . The prior mean is in bold face and standard deviations are in parenthesis.*

term structure is gently upward sloping on average in each regime. Also the assumed prior allows for considerable a priori variation in the term structure in each regime. The implied prior distribution of the term structure for the other models we consider can be found in appendix 3.3.

3.4 Posterior Distribution and MCMC Sampling

Under our assumptions it is now possible to calculate the posterior distribution of the parameters by MCMC simulation methods. The use of these methods in our high-dimensional problem is made possible by the inclusion of the latent factors and the latent regime indicators in the prior-posterior analysis and by the use of the tailored randomized block (TaRB) MCMC algorithm developed in Chib and Ramamurthy (2009).

Let $\mathbf{F}_n = \{\mathbf{f}_t\}_{t=1,\dots,n}$ and $\mathbf{S}_n = \{s_t\}_{t=1,\dots,n}$. Then, the posterior distribution that we would like to explore is given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, u_0, \mathbf{F}_n, \mathbf{S}_n | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, \mathbf{S}_n, \mathbf{F}_n) p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\theta}, u_0) \pi(u_0 | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\sigma}^{*2}) \quad (3.5)$$

where $f(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, \mathbf{S}_n, \mathbf{F}_n)$ is the distribution of the data given the latent factors, the

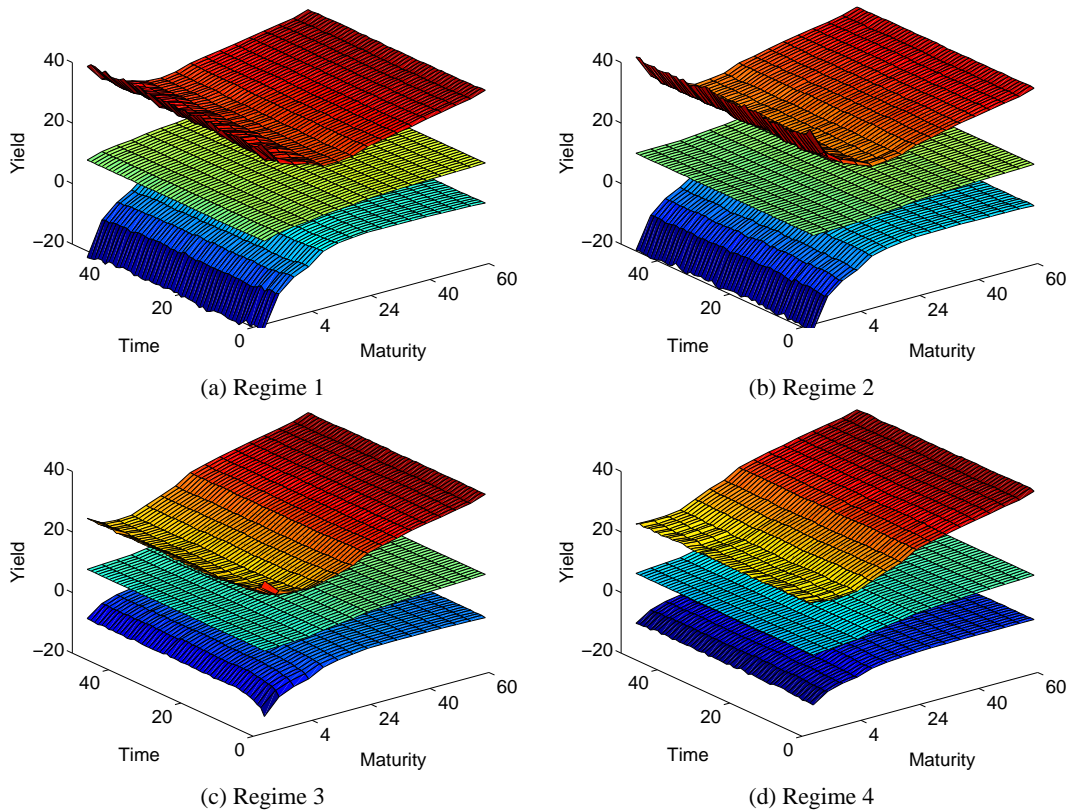


Figure 3: The implied prior term structure dynamics for 3 change point model *These graphs are based on 10,000 simulated draws of the parameters from the prior distribution. In the graphs on the left, the surfaces correspond to the 2.5%, 50%, and 97.5% quantile surfaces of the term structure dynamics in annualized percents implied by the prior distribution for each regime.*

regime indicators and the parameters, $p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\theta}, u_0)$ is the joint density of the regime-indicators and the factors given the parameters and the initial latent factor, $\pi(u_0 | \boldsymbol{\theta})$ is the density of the latent initial factor given the parameters, and $\pi(\boldsymbol{\theta})\pi(\boldsymbol{\sigma}^{*2})$ is the prior density of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2})$.

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 $\pi(\boldsymbol{\theta}, \sigma^{*2}, u_0, \mathbf{F}_n, \mathbf{S}_n | \mathbf{y})$. In particular, after initializing the various unknowns, we go through 4 iterative steps in each MCMC cycle. Briefly, in Step 2 we sample $\boldsymbol{\theta}$ and \mathbf{F}_n in one block. We achieve this by first sampling $\boldsymbol{\theta}$ marginalized over \mathbf{F}_n from the posterior distribution that is proportional to

$$f(\mathbf{y} | \boldsymbol{\theta}, \sigma^{*2}, \mathbf{S}_n) \pi(\boldsymbol{\theta})$$

where $f(\mathbf{y} | \boldsymbol{\theta}, \sigma^{*2}, \mathbf{S}_n)$ is obtained from the standard Kalman filtering recursions given the regime indicators \mathbf{S}_n . Note that by conditioning on \mathbf{S}_n we avoid the calculation of the likelihood function $f(\mathbf{y} | \boldsymbol{\theta}, \sigma^{*2}, u_0)$ 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 sampling of $\boldsymbol{\theta}$ from the latter density is done by the TaRB-MCMC method of Chib and Ramamuthy (2009). Then in Step 2b, given the sampled value of $\boldsymbol{\theta}$ we sample \mathbf{F}_n conditioned on $(u_0, \mathbf{S}_n, \sigma^{*2}, \boldsymbol{\theta})$ in one block by the forward-backward iterations of Carter and Kohn (1994). In Step 3 we sample u_0 from the posterior distribution that is proportional to

$$p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\theta}, u_0) \pi(u_0 | \boldsymbol{\theta})$$

In Step 4, we sample \mathbf{S}_n conditioned on $(\boldsymbol{\theta}, \mathbf{F}_n, u_0, \sigma^{*2})$ in one block by the algorithm of Chib (1996). We finish one cycle of the algorithm by sampling σ^{*2} conditioned on $(\boldsymbol{\theta}, \mathbf{F}_n, \mathbf{S}_n)$ from the posterior distribution that is proportional to

$$f(\mathbf{y} | \boldsymbol{\theta}, \sigma^{*2}, \mathbf{S}_n, \mathbf{F}_n) \pi(\sigma^{*2})$$

Our algorithm can be summarized as follows.

Algorithm: MCMC sampling

Step 1 Initialize $(\boldsymbol{\theta}, u_0, \mathbf{S}_n, \sigma^{*2})$ and fix n_0 (the burn-in) and n_1 (the MCMC sample size)

Step 2 Sample $\boldsymbol{\theta}$ and \mathbf{F}_n in one block by sampling

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

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

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

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

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

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.

3.5 Marginal Likelihood Computation

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 4 models which in the introduction were named as C0L1M2, C1L1M2, C2L1M2 and C3L1M2. The most general model is C3L1M2 that has 3 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 marginal likelihood of any given model is obtained as

$$m(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, \mathbf{S}_n, \mathbf{F}_n) p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\theta}, u_0) \pi(u_0 | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\sigma}^{*2}) d(\boldsymbol{\theta}, \boldsymbol{\sigma}^{*2}, \mathbf{S}_n, \mathbf{F}_n, u_0)$$

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

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta}^*, \boldsymbol{\sigma}^{**2}, u_0^*) \pi(u_0^* | \boldsymbol{\theta}^*) \pi(\boldsymbol{\theta}^*) \pi(\boldsymbol{\sigma}^{**2})}{\pi(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^{**2}, u_0^* | \mathbf{y})}$$

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

$$\ln \hat{m}(\mathbf{y}) = \ln f(\mathbf{y}|\boldsymbol{\theta}^*, \boldsymbol{\sigma}^{**2}, u_0^*) + \ln \{ \pi(u_0^* | \boldsymbol{\theta}^*) \pi(\boldsymbol{\theta}^*) \pi(\boldsymbol{\sigma}^{**2}) \} - \ln \hat{\pi}(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^{**2}, u_0^* | \mathbf{y})$$

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 Ramamuthy (2009) for the case of randomized blocks. We suppress the details.

We do show how the first term in the above expression, the likelihood ordinate would be calculated. To do this, we first begin by re-expressing the measurement equations in (3.2) in an alternative form. Let the basis yield that is priced without error be denoted by R_t^w and the remaining yields(which are measured with error) be denoted by \mathbf{R}_t^e . Also let $\bar{a}_{s_t}^w$ ($\bar{\mathbf{a}}_{s_t}^e$) and $\bar{b}_{s_t}^w$ ($\bar{\mathbf{b}}_{s_t}^e$) be the corresponding intercept and factor loadings for R_t^w (\mathbf{R}_t^e), respectively. Then, from the measurement equations we have that R_t^w is given by

$$R_t^w = \bar{a}_{s_t}^w + \bar{b}_{u,s_t}^w u_t + \bar{\mathbf{b}}_{m,s_t}^w (\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})$$

which implies that the latent factor can be expressed in terms of the observed variables as

$$u_t = (\bar{b}_{u,s_t}^w)^{-1} (R_t^w - \bar{a}_{s_t}^w - \bar{\mathbf{b}}_{m,s_t}^w (\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})) \quad (3.6)$$

Conditioned on \mathbf{m}_t , this represents a one-to-one map between R_t^w and u_t . If we now let

$$\mathbf{f}_t = \begin{pmatrix} u_t \\ \mathbf{m}_t \end{pmatrix} = \begin{pmatrix} (\bar{b}_{u,s_t}^w)^{-1} (R_t^w - \bar{a}_{s_t}^w - \bar{\mathbf{b}}_{m,s_t}^w (\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})) \\ \mathbf{m}_t \end{pmatrix}$$

and define

$$\mathbf{y}_t^e = \begin{pmatrix} \mathbf{R}_t^e \\ \mathbf{f}_t \end{pmatrix}$$

then the distribution of \mathbf{y}_t^e is the same as that of \mathbf{y}_t . The idea now is to calculate the likelihood as

$$\ln L(\mathbf{y}|\boldsymbol{\psi}) = \sum_{t=1}^n \ln f(\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}) \quad (3.7)$$

where $\boldsymbol{\psi}$ denotes the collection of the model parameters and

$$f(\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}) = \sum_{i,j} f(\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j) \Pr[s_{t-1} = i, s_t = j|I_{t-1}, \boldsymbol{\psi}] \quad (3.8)$$

is the one-step ahead predictive density of \mathbf{y}_t^e , and I_{t-1} consists of the history of the outcomes R_{t-1} and \mathbf{m}_{t-1} up to time $(t-1)$. We now show how $f(\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j)$ and $\Pr[s_{t-1} = i, s_t = j|I_{t-1}, \boldsymbol{\psi}]$ can each be computed recursively.

Begin by writing

$$\mathbf{y}_t^e = \begin{pmatrix} \bar{\mathbf{a}}_{s_t}^e \\ \boldsymbol{\mu}_{s_t} \end{pmatrix} + \begin{pmatrix} \bar{\mathbf{b}}_{s_t}^e \\ \mathbf{I}_3 \end{pmatrix} (\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \begin{pmatrix} \boldsymbol{\varepsilon}_t^e \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_t^e \sim iidN(\mathbf{0}, \Sigma_{s_t}^e)$$

From this it is easy to see that \mathbf{y}_t^e is Gaussian, conditioned on $(I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j)$. If we let

$$\begin{aligned} \mathbf{E}[\mathbf{f}_t|I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j] &= \boldsymbol{\mu}_j + \mathbf{G}_j (\mathbf{f}_{t-1} - \boldsymbol{\mu}_i) \\ &\equiv \mathbf{f}_{t|t-1}^{ij} \end{aligned}$$

then it follows that the needed moments of the latter Gaussian density are given by

$$E(\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j) = \begin{pmatrix} \bar{\mathbf{a}}_j^e \\ \boldsymbol{\mu}_j \end{pmatrix} + \begin{pmatrix} \bar{\mathbf{b}}_j^e \\ \mathbf{I}_3 \end{pmatrix} (\mathbf{f}_{t|t-1}^{ij} - \boldsymbol{\mu}_j)$$

and

$$Var[\mathbf{y}_t^e|I_{t-1}, \boldsymbol{\psi}, s_{t-1} = i, s_t = j] = \begin{pmatrix} \bar{\mathbf{b}}_j^e \\ \mathbf{I}_3 \end{pmatrix} \Omega_j \begin{pmatrix} \bar{\mathbf{b}}_j^e \\ \mathbf{I}_3 \end{pmatrix}' + \begin{pmatrix} \Sigma_j^e & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Next, from the law of total probability we have that

$$\Pr[s_{t-1} = i, s_t = j|I_{t-1}, \boldsymbol{\psi}] = p_{ij} \Pr[s_{t-1} = i|I_{t-1}, \boldsymbol{\psi}]$$

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

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

where

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

This completes the calculation of the likelihood function.

4 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 since it is the bond with the smallest pricing variance. Our macroeconomic factors are the quarterly GDP inflation deflator and the real GDP growth rate in annualized percents. These data are from the Federal reserve bank of St. Louis.

Before proceeding we would like to explain our reasons for modeling 16 yields since these many yields have not been used in previous work. One reason is that our Bayesian estimation approach is capable of handling a large set of yields, more so than is possible by maximum-likelihood methods or less-tuned MCMC implementations than ours. The other reason is that in comparison with models with fewer yields, the model with 16 yields has the best out-of-sample predictive accuracy. To show this we fit the model up to 2006 and predict the yields and macro factors for each of the 4 quarters of 2007. We measure the predictive accuracy in terms of the posterior predictive criterion of Gelfand and Ghosh (1998) (PPC, henceforth). The technical details about simulating the Bayesian predictive density are given in section 4.6. For any given model M_j , the PPC criterion is defined as

$$PPC_j = D_j + W_j \quad (4.1)$$

$$\text{where } D_j = \frac{1}{\lambda + 2} \sum_{i=1}^{\lambda+m} \sum_{t=1}^T \text{Var}(\tilde{y}_{i,t} | \mathbf{y}, M_j) \quad (4.2)$$

$$\text{and } W_j = \frac{1}{\lambda + 2} \sum_{i=1}^{\lambda+m} \sum_{t=1}^T [y_{i,t} - E(\tilde{y}_{i,t} | \mathbf{y}, M_j)]^2 \quad (4.3)$$

where λ is the number of the maturities, $\{\tilde{\mathbf{y}}_t\}_{t=1,2,\dots,T}$ are the predictions of the actual yields and macro factors $\{\mathbf{y}_t\}_{t=1,2,\dots,T}$ under model M_j , and $y_{i,t}$ and $\tilde{y}_{i,t}$ are the i th component of \mathbf{y}_t and $\tilde{\mathbf{y}}_t$, respectively. The first term has large values for models that are very simple or complex. The second term is a sum of squared residuals and measures goodness-of-fit in terms of how well the forecasts under model M_j fit the actual ob-

servations. Table 2 clearly shows that the model with 16 maturities out performs the models with fewer maturities. The reason for this superior performance is simple. The

The number of maturities(λ)	No change point model		
	D_j	W_j	PPC_j
4	6.188	4.441	10.629
8	5.459	4.352	9.811
12	4.140	3.984	8.124
16	3.831	3.950	7.780

Table 2: Posterior predictive criterion PPC is computed by 4.1 to 4.3. We use the data from the most recent break time point, 1996:I 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 *DSY (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.

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

4.1 Sampler Diagnostics

We base our results on 10,000 iterations of the MCMC algorithm beyond a burn-in of 2,000 iterations. We measure the efficiency of the MCMC sampling in terms of the metrics that are common in the Bayesian literature, in particular, in terms of the acceptance rates in the Metropolis-Hastings steps and the inefficiency factors (Chib (2001)) which, for any sampled sequence of draws, is defined as

$$1 + 2 \sum_{k=1}^M \rho(k), \quad (4.4)$$

where $\rho(k)$ is the k -order autocorrelation computed from the sampled variates and M 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 54.4% and 160.0, respectively. These values indicate that our sampler mixes well. It

is also important to note that our sampler converges quickly to the same region of the parameter space regardless of the starting values.

4.2 The Number and Timing of Change Points

Table 3 contains the marginal likelihood estimates for our 4 contending models. As can be seen, the 3 change point model, C3L1M2, is the model that gets the most support from the data. We now provide more detailed results for this model.

sample period	Model	lnL	lnML	change point
1972:I-2006:IV	C0L1M2	-1487.0	-1657.3	
	C1L1M2	-1154.3	-1507.5	1986:II
	C2L1M2	-774.5	-1297.9	1985:IV, 1995:II
	C3L1M2	-445.5	-1107.1	1980:I, 1986:II, 1995:II

Table 3: Log likelihood (lnL), log marginal likelihood (lnML) 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 each of the states by time. These probabilities are given in Figure 4. 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. It is striking finding that this analysis picks up a breakpoint in 1995 since this has not been detected in previous regime-change models.

We would like to emphasize that our estimates of the change points from the models without macro factors (i.e. C1L1M0, C2L1M0 and C3L1M0 models) 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.

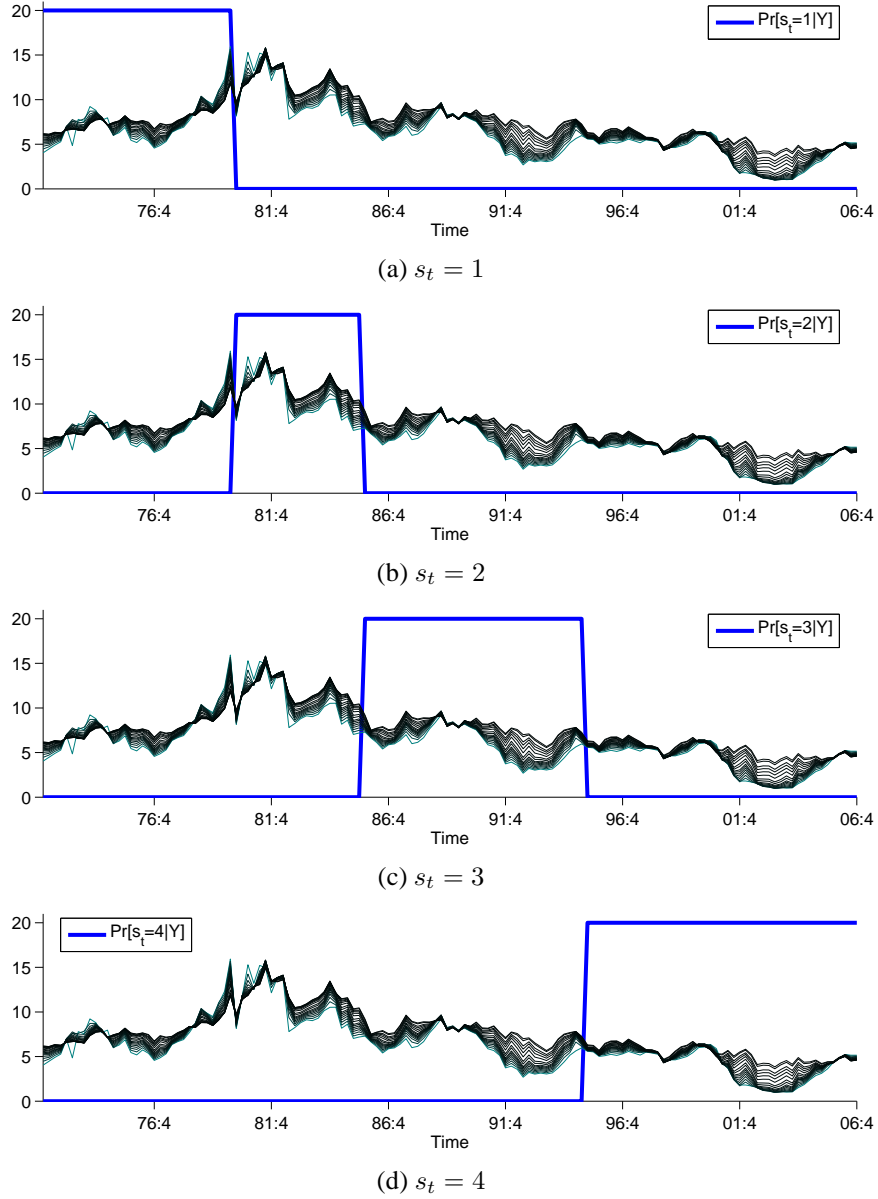


Figure 4: Posterior probability of $s_t = 1$, $s_t = 2$, $s_t = 3$ and $s_t = 4$. The figure plots the average of 10,000 sets of sampled $\{s_t\}$ against the sixteen yields in annualized percents. Each posterior probability is multiplied by 20

4.3 Parameter Estimates

Table 4 summarizes the posterior distribution of the parameters. One point to note is that the posterior densities are generally different from the prior given in table 1, which implies that the data is informative about these parameters. We focus on various aspects

of this posterior distribution in the subsequent subsections.

	Regime 1			Regime 2			Regime 3			Regime 4		
G	0.89	0.05	0.05	0.97	-0.04	0.01	0.90	0.20	0.29	0.92	0.07	0.10
	(0.05)	(0.08)	(0.06)	(0.02)	(0.05)	(0.03)	(0.06)	(0.17)	(0.10)	(0.06)	(0.19)	(0.16)
	-0.27	0.70	-0.04	-0.07	0.74	-0.10	0.16	0.42	0.10	0.05	0.83	-0.01
	(0.24)	(0.23)	(0.11)	(0.06)	(0.06)	(0.04)	(0.11)	(0.19)	(0.09)	(0.06)	(0.16)	(0.08)
	-0.09	-0.19	0.20	0.04	-0.27	0.50	-0.03	-0.01	0.31	-0.05	-0.41	0.17
	(0.28)	(0.27)	(0.17)	(0.23)	(0.30)	(0.22)	(0.14)	(0.25)	(0.17)	(0.12)	(0.29)	(0.16)
μ $\times 400$	0.00	5.59	3.45	0.00	5.98	2.75	0.00	2.65	2.64	0.00	1.66	3.17
		(1.63)	(0.96)		(0.51)	(1.22)		(0.51)	(0.57)		(0.84)	(0.60)
	1.00			1.00			1.00			1.00		
L $\times 400$	0.23	1.73		0.20	1.49		0.23	0.77		-0.18	0.72	
	(0.58)	(0.19)		(0.43)	(0.18)		(0.29)	(0.18)		(0.43)	(0.21)	
	-0.15	-0.76	3.99	0.51	0.38	4.42	-1.30	-0.39	1.48	0.27	-0.43	1.75
	(1.03)	(0.80)	(0.17)	(1.00)	(0.83)	(0.20)	(0.67)	(0.55)	(0.27)	(0.74)	(0.56)	(0.19)
δ_1 $\times 400$		9.46			3.34			4.25			4.11	
		(1.92)			(1.93)			(1.47)			(1.20)	
δ_2	1.28	0.04	0.01	1.43	0.20	0.12	1.08	0.49	0.41	0.78	0.49	0.02
	(0.17)	(0.24)	(0.10)	(0.21)	(0.24)	(0.09)	(0.25)	(0.33)	(0.18)	(0.21)	(0.48)	(0.16)
γ	-0.19	-0.49	-0.29	-0.25	-0.52	-0.24	-0.55	-0.42	-0.19	-0.20	-0.10	-0.24
	(0.26)	(0.35)	(0.40)	(0.19)	(0.23)	(0.40)	(0.29)	(0.41)	(0.32)	(0.27)	(0.20)	(0.42)
Φ	0.97	0.79	0.92	0.73	0.89	0.66	0.86	0.93	0.91	0.96	0.90	0.88
	(1.20)	(1.21)	(1.32)	(1.30)	(1.27)	(1.30)	(1.24)	(1.26)	(1.26)	(1.23)	(1.24)	(1.26)
p_{00}						0.94						
						(0.03)						
p_{11}						0.98						
						(0.01)						
p_{22}						0.98						
						(0.00)						

Table 4: Estimates of model parameters *This table presents the posterior mean and standard deviation based on 10,000 posterior draws beyond 2,000 burn-in. 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.*

4.3.1 Factor Process

Figure 5 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 **G** for each regime show that the mean-reversion coef-

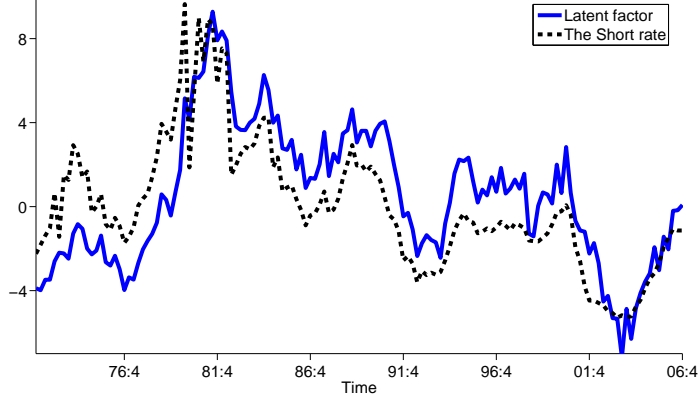


Figure 5: Latent factor *The short rate in percents is demeaned. The latent factor represents the average simulated latent factor from retained 10,000 MCMC iterations.*

ficient matrix is almost diagonal. The latent factor and inflation rate also display very high and considerably different 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 3, respectively. For a more formal measure of this persistence, we calculate the eigenvalues of the coefficient matrices in each regime. These are given by

$$\begin{aligned}
 eig(\mathbf{G}_1) &= \begin{bmatrix} 0.795 + 0.033i \\ 0.795 - 0.033i \\ 0.196 \end{bmatrix}, \quad eig(\mathbf{G}_2) = \begin{bmatrix} 0.986 \\ 0.801 \\ 0.420 \end{bmatrix} \\
 eig(\mathbf{G}_3) &= \begin{bmatrix} 0.948 \\ 0.358 \\ 0.319 \end{bmatrix}, \quad eig(\mathbf{G}_4) = \begin{bmatrix} 0.918 \\ 0.834 \\ 0.164 \end{bmatrix}
 \end{aligned}$$

It can be seen that the second regime has the largest absolute eigenvalue close to 1. Another point to note is that the factors in regime 1 have oscillatory dynamics under the physical measure. Since 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 that a reduction in factor volatility starting from the middle of the 1980s, which coincides with the period that is called the great moderation.

4.3.2 Term Premium

Figure 6 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.

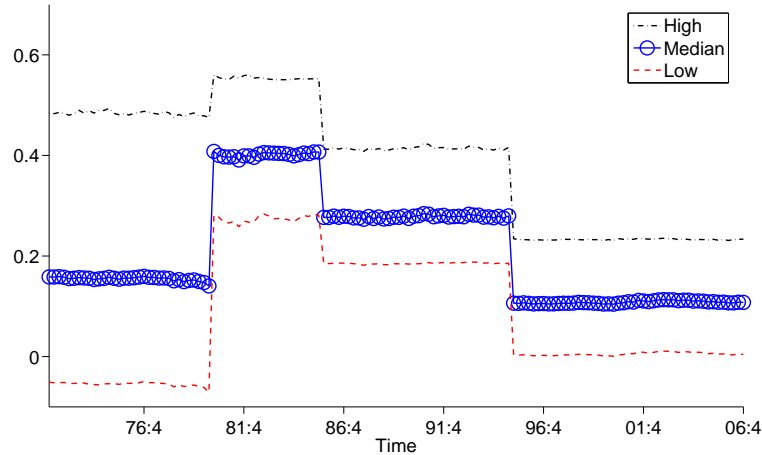


Figure 6: Term premium. The figure plots the 2.5%, 50% and 97.5% quantile of the posterior term premium which correspond to “Low,” “Median” and “high” based on 10,000 draws beyond a burn-in of 2,000 iterations.

4.3.3 Pricing Error Variances

In Figure 7 we plot the term structure of the pricing error variances. As in no-change point model of Chib and Ergashev (2009), these are hump-shaped in each regime. One can also see that these variances have changed over time, primarily for the short-bonds. These changes in the variances also help to determine the timing of the change-points.

4.4 Forecasting and Predictive Densities

A principle objective of this paper is to compare the forecasting abilities of the L1M2 model with and without regime changes. In the Bayesian paradigm, it is relatively

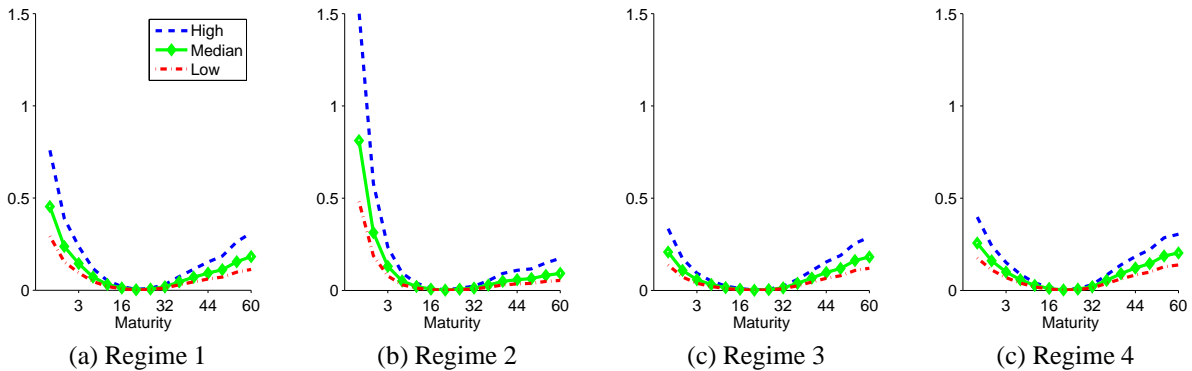


Figure 7: Term Structure of the Pricing Error Variances *The figures display the 2.5%, 50% and 97.5% quantile of the posterior draws which correspond to "low", "median" and "high". Regime 1 ranges from 1972:I to 1980:I, regime 2 from 1980:II to 1985:IV, regime 3 from 1986:I-1995:II, and regime 4 from 1995:III-2006:IV*

straightforward to calculate the predictive density during the course of the MCMC iterations. This is because the predictive density of the future observations, conditional on the data, is obtained by simply integrating out the parameters with respect to the posterior distribution. Denoting \mathbf{y}_f as the future observations, the predictive density can be calculated as

$$f(\mathbf{y}_f|M_i, \mathbf{y}) = \int_{\Psi} f(\mathbf{y}_f|M_i, \mathbf{y}, \Psi)\pi(\Psi|M_i, \mathbf{y})d\Psi \quad (4.5)$$

where the predictive draws are sampled under the terminal regime $q + 1$.

Specifically, note that each MCMC iteration (beyond the burn-in period) provides us with the factors \mathbf{F}_n and the parameters of the model from the posterior distribution. Therefore, conditioned on \mathbf{f}_n and the underlying parameters in regime $q + 1$, we draw the forecasts of factors \mathbf{f}_{n+1} based on the transition equation. Then given \mathbf{f}_{n+1} , the yields in the forecast period are drawn using the relationship described in the measurement equation. The resulting collection of the simulated macro factors and yields is taken as a sample from the Bayesian predictive density.

We plot the out of sample forecasts in figure 8. The top panel gives the forecast intervals from the C0L1M2 model. The bottom panel has the forecast intervals from the model averaged predictive distribution that is obtained by averaging the 4 predictive distributions (one from each candidate model) with weights given by the posterior prob-

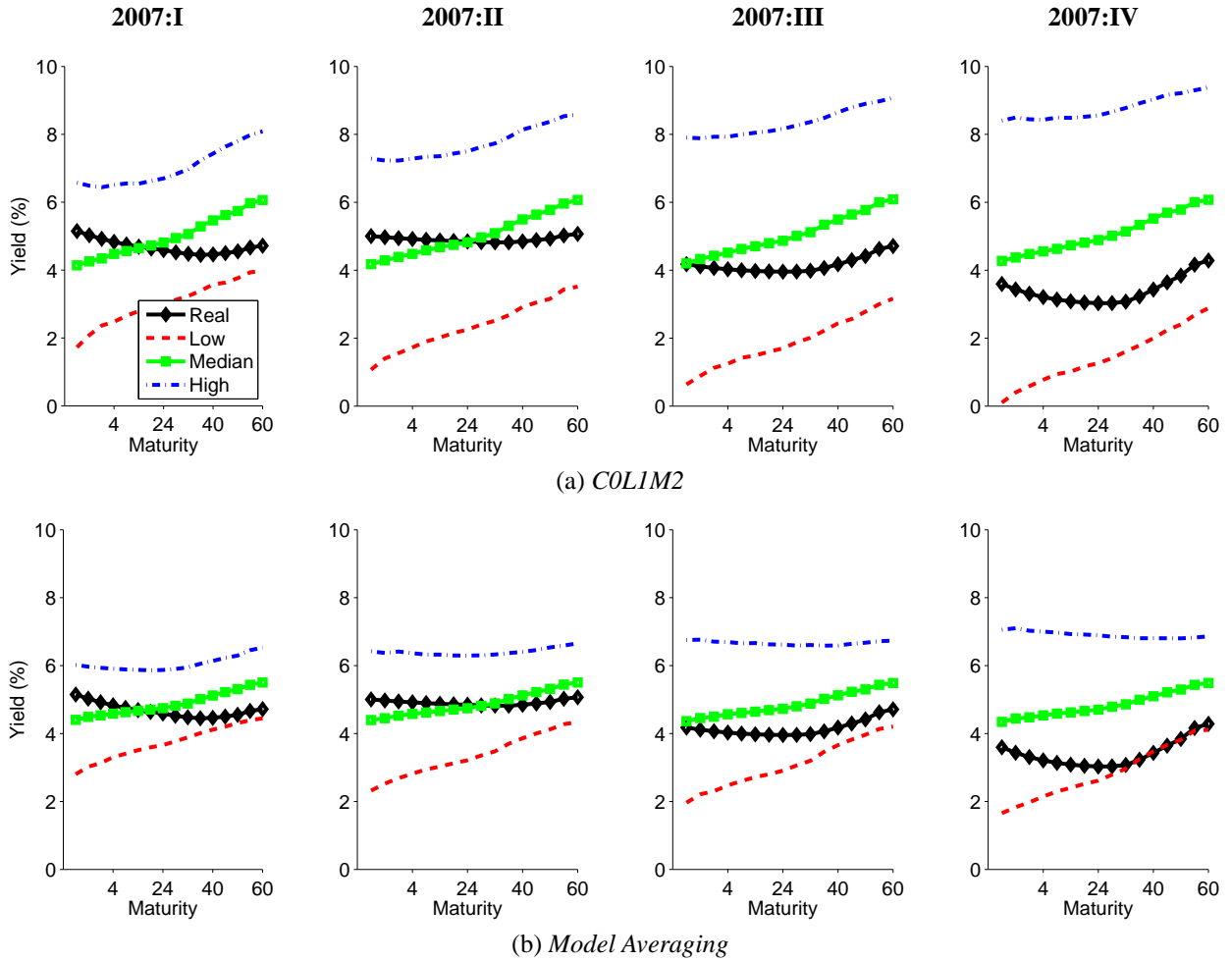


Figure 8: The MCMC forecasts of the yield curve *The figures present four quarters ahead forecasts of the yields on the T-bills. The left column panel is based on no change point model and the right column panel shows model averaged forecasts from COL1M2, C1L1M2, C2L1M2 and C3L1M2. In each case, the 2.5%, 50% and 97.5% quantile curves, labeled “Low”, “Median” and “High” respectively, are based on 10,000 forecasted values for the period of 2007:I-2007:IV. The observed curves are labeled “Real.”*

ability of each model (these being derived from the marginal likelihood of each 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 model averaged distribution are tighter.

For a more formal forecasting performance comparison, we tabulate the PPC for each case in Table 5. We also include in the last column of this table an interesting set of results that make use of the regimes isolated by our 3 change-point model. In

Model	C0L1M2	Model averaging	C0L1M2
Sample period	(1972:I-2006:IV)		(1996:I-2006:IV)
D_j	11.091	4.011	3.831
W_j	5.637	3.703	3.950
PPC_j	16.728	7.714	7.781

Table 5: Posterior predictive criterion *PPC is computed by 4.1 to 4.3.*

particular, we fit the no-change point model to the data in the last regime but ending just before the forecast period. This is the period 1996:I-2006:IV. We would expect the no-change point model fit to this sample to produce forecasts that are similar to those from our model averaged distribution. The results in the table bear this out. Thus, given the regimes we have isolated, a poor-man’s 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 model averaged distribution produce a smaller value of the PPC than 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 it is essential to incorporate regime changes when forecasting the term structure of interest rates.

5 Concluding Remarks

In this paper we have developed a new model of the term structure of zero-coupon bonds with regime changes. Our work complements the recent work in this area since it is organized around a different model of regime changes than the Markov switching model that has been used to date. Our work also complements the recent work on affine models with macro factors which has been done in settings without regime changes. The models we fit involve more bonds than in previous work which allows us to capture more of the term structure. This enlargement of the model is made possible by our tuned econometric methods which rely on some recent developments in Bayesian econometrics.

Our empirical analysis suggest that the term structure has gone through three change

points, and that the term structure and the risk premium is materially different across regimes. Our analysis also shows that there are gains in predictive accuracy by incorporating regime changes when forecasting the term structure of interest rates.

A Bond Prices under Regime Changes

By the assumption of the affine model, we have

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

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

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

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

It immediately follows from the bond pricing formula that

$$\begin{aligned} 1 &= E \left[\kappa_{t, t+1} \frac{P_{t+1}(s_{t+1}, \tau - 1)}{P_t(s_t, \tau)} \middle| \mathbf{f}_t, s_t \right] \\ &= E [\kappa_{t, t+1} h_{\tau, t+1} | \mathbf{f}_t, s_t]. \end{aligned}$$

Then by substitution

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

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 $\boldsymbol{\omega}_{t+1} = \mathbf{L}_{s_{t+1}}^{-1} \eta_{t+1} \sim iidN(0, I_{k+m})$. 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

$$E \left(\exp \left[-\frac{1}{2} \Gamma_{t,\tau} \Gamma'_{t,\tau} - \Gamma_{t,\tau} \boldsymbol{\omega}_{t+1} \right] \right) = 1$$

we have that

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

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

$$\begin{aligned} & E [\kappa_{t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t] \\ &= \exp \left[-R_{t,s_t} - \frac{1}{2} \boldsymbol{\gamma}'_{t,s_t} \boldsymbol{\gamma}_{t,s_t} + \frac{1}{2} (\boldsymbol{\gamma}'_{t,s_t} + \mathbf{b}_{s_{t+1}} (\tau - 1)' \mathbf{L}_{s_{t+1}}) (\boldsymbol{\gamma}'_{t,s_t} + \mathbf{b}_{s_{t+1}} (\tau - 1)' \mathbf{L}_{s_{t+1}})' + \zeta_{\tau,s_t,s_{t+1}} \right] \\ &\approx -R_{t,s_t} + \boldsymbol{\gamma}'_{t,s_t} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}} (\tau - 1) + \frac{1}{2} (\mathbf{b}_{s_{t+1}} (\tau - 1)' \mathbf{L}_{s_{t+1}} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}} (\tau - 1)) + \zeta_{\tau,s_t,s_{t+1}} + 1 \\ &= -(\delta_{1,s_t} + \boldsymbol{\delta}'_{2,s_t} \mathbf{f}_t) + (\bar{\boldsymbol{\gamma}}_{s_t} + \boldsymbol{\Phi}_{s_t} \mathbf{f}_t)' \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}} (\tau - 1) \\ &+ \frac{1}{2} (\mathbf{b}_{s_{t+1}} (\tau - 1)' \mathbf{L}_{s_{t+1}} \mathbf{L}'_{s_{t+1}} \mathbf{b}_{s_{t+1}} (\tau - 1)) + \zeta_{\tau,s_t,s_{t+1}} + 1 \end{aligned}$$

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

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

Thus we have

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

Matching the coefficients on \mathbf{f}_t and constant terms equal to zero we obtain the recursive equation for $a_{s_t}(\boldsymbol{\tau})$ and $\mathbf{b}_{s_t}(\boldsymbol{\tau})$ given the initial condition of $a_{s_t}(0) = 0$ and $\mathbf{b}_{s_t}(0) = \mathbf{0}_{3 \times 1}$ implied by no arbitrage condition. Finally imposing the restriction on the transition probability establishes the proof.

B Prior distribution

	No change point			One change point					
				Regime 0			Regime 1		
\mathbf{G}	diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)		
$\boldsymbol{\mu}$ $\times 400$	0.00 (1.00)	5.00 (2.00)	3.00 (1.00)	0.00 (1.00)	6.50 (2.00)	3.00 (1.00)	0.00 (1.00)	4.00 (2.00)	3.00 (1.00)
δ_1 $\times 400$		5.00 (4.00)			6.00 (4.00)			5.00 (4.00)	
δ_2	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)
γ	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)
Φ	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
λ	(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)		

Table 6: Prior distribution of the parameters in the two-regime change point model This table presents the prior mean and standard deviation of the parameters in θ . The prior mean are indicated in bold face and the standard deviations are in parenthesis.

C MCMC Sampling

This section provides the details of the MCMC algorithm (steps 2-5) outlined in section 3.4.

Step 2a Sampling θ

We sample θ conditioned on $(u_0, \mathbf{S}_n, \sigma^{*2})$ by the tailored randomized block M-H (TaRB-MH) algorithm introduced in Chib and Ramamurthy (2009). 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

	Regime 0			Regime 1			Regime 2		
G	diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)			diag(0.9, 0.8, 0.4) (0.33)		
μ $\times 400$	0.00	6.50 (2.00)	3.00 (1.00)	0.00	8.50 (2.00)	3.00 (1.00)	0.00	5.00 (2.00)	3.00 (1.00)
δ_1 $\times 400$		6.00 (4.00)			8.00 (4.00)			5.00 (4.00)	
δ_2	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)	0.60 (1.00)	0.40 (1.00)	0.40 (1.00)
γ	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)	-0.50 (0.33)
Φ	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
λ	(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)			(0, 0, 0, 0, 1)' (1.00)		

Table 7: Prior distribution of the parameters in the three-regime change point model. *This table presents the prior mean and standard deviation of the parameters in θ . The prior means are indicated in bold face and standard deviations are in parenthesis.*

θ

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

Then the proposal density $q(\theta_i | \theta_{-i}, \mathbf{y})$ for the i th block, conditioned on the most current value of the remaining blocks θ_{-i} , is constructed by a quadratic approximation at the mode of the current target density $\pi(\theta_i | \theta_{-i}, \mathbf{y})$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q(\theta_j | \theta_{-i}, \mathbf{y}) = St(\theta_i | \hat{\theta}_i, \mathbf{V}_{\hat{\theta}_i}, 15)$$

where

$$\hat{\theta}_i = \arg \max_{\theta_i} \ln \{ f(\mathbf{y} | \theta_i, \theta_{-i}, \mathbf{S}_n) \pi(\theta_i) \}$$

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

Because the likelihood function tends to be ill-behaved in these problems, we calculate $\hat{\theta}_i$ 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

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

If $\boldsymbol{\theta}_i^\dagger$ violates any of the constraints, it is immediately rejected. The simulation of $\boldsymbol{\theta}$ is complete when all the sub-blocks

$$\pi \left(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_{-1}, \mathbf{y}, \mathbf{S}_n \right), \pi \left(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_{-2}, \mathbf{y}, \mathbf{S}_n \right), \dots, \pi \left(\boldsymbol{\theta}_{h_g} | \boldsymbol{\theta}_{-h_g}, \mathbf{y}, \mathbf{S}_n \right)$$

are sequentially updated as above.

Step 2b Sampling the factors

Our sampling of the factors is based on the method of Carter and Kohn (1994). Conditioned on \mathbf{S}_n , we obtain $\mathbf{f}_{t|t} = E(\mathbf{f}_t | I_t, \boldsymbol{\psi})$, $\mathbf{R}_{t+1|t} = Var(\mathbf{f}_{t+1} | I_t, \boldsymbol{\psi})$ and $\mathbf{R}_{t|t} = Var(\mathbf{f}_t | I_t, \boldsymbol{\psi})$ through the Kalman filter. In the backward recursions, we first draw \mathbf{f}_n from $\mathcal{N}_3(\mathbf{f}_{n|n}, \mathbf{R}_{n|n})$. Then for $t = 1, 2, \dots, n-1$, we sample \mathbf{f}_t given the \mathbf{f}_{t+1} from $\mathcal{N}_3(\hat{\mathbf{f}}_t, \hat{\mathbf{R}}_t)$ where

$$\begin{aligned} \hat{\mathbf{f}}_t &= \mathbf{f}_{t|t} + \mathbf{M}_t \left(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}} - \mathbf{G}_{s_{t+1}} \left(\mathbf{f}_{t|t} - \boldsymbol{\mu}_{s_t} \right) \right) \\ \hat{\mathbf{R}}_t &= \mathbf{R}_{t|t} - \mathbf{M}_t \mathbf{R}_{t+1|t}^{-1} \mathbf{M}_t' \end{aligned}$$

and

$$\mathbf{M}_t = \mathbf{R}_{t|t} \mathbf{G}_{s_{t+1}}' \mathbf{R}_{t+1|t}^{-1}$$

It is important to note that because of the way the measurement equations are set up in our model, the macro factors are updated without error.

Step 3 Sampling the initial factor

Given the prior in section 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 \mathbf{f}_1 is obtained from step 2b above. In the following, it is assumed that all the underlying coefficients are those in regime 0. Then

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

where

$$\begin{aligned} u_0 &= \mathbf{U}_0 (\boldsymbol{\Sigma}_u^{-1} + \mathbf{H}^{*'} \boldsymbol{\Omega}_{11,0}^* u_1^*) \\ \mathbf{U}_0 &= (\boldsymbol{\Sigma}_u^{-1} + \mathbf{H}^{*'} \boldsymbol{\Omega}_{11,0}^* \mathbf{H}^*) \end{aligned}$$

and on letting

$$\mathbf{G}_0 = \begin{pmatrix} \mathbf{G}_{11,0} & \mathbf{G}_{12,0} \\ \mathbf{G}_{21,0} & \mathbf{G}_{22,0} \end{pmatrix}, \boldsymbol{\Omega}_0 = \begin{pmatrix} \boldsymbol{\Omega}_{11,0} & \boldsymbol{\Omega}_{12,0} \\ \boldsymbol{\Omega}_{21,0} & \boldsymbol{\Omega}_{22,0} \end{pmatrix}$$

$$\begin{aligned} \mathbf{H}^* &= \mathbf{G}_{11,0} - \boldsymbol{\Omega}_{12,0} \boldsymbol{\Omega}_{22,0}^{-1} \mathbf{G}_{21,0} \\ \boldsymbol{\Omega}_{11,0}^* &= \boldsymbol{\Omega}_{11,0} - \boldsymbol{\Omega}_{12,0} \boldsymbol{\Omega}_{22,0}^{-1} \boldsymbol{\Omega}_{21,0} \\ u_1^* &= u_1 - \mathbf{G}_{12,0} (\mathbf{m}_0 - \boldsymbol{\mu}_{m,0}) - \boldsymbol{\Omega}_{12,0} \boldsymbol{\Omega}_{22,0}^{-1} (\mathbf{m}_1 - \boldsymbol{\mu}_{m,0}) + \boldsymbol{\Omega}_{12,0} \boldsymbol{\Omega}_{22,0}^{-1} \mathbf{G}_{22,0} (\mathbf{m}_0 - \boldsymbol{\mu}_{m,0}) \end{aligned}$$

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) and Chib (1998) 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 $\Pr[s_1 = 1 | I_1, \boldsymbol{\psi}] = 1$. Let $\mathbf{f}_{t|t-1}^{jk}$ denote $E[\mathbf{f}_t | I_{t-1}, s_{t-1} = j, s_t = k]$. Then one first obtains $\Pr[s_t = k | I_t, \boldsymbol{\psi}]$ for all $k = 1, 2, \dots, q + 1$ and $t = 1, 2, \dots, n - 1$ by calculating

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

where

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

and the three terms on the right hand side of this expression are obtained as

$$f[\mathbf{y}_t | I_{t-1}, s_{t-1} = j, s_t = k, \boldsymbol{\psi}] = -\frac{1}{\sqrt{2\pi}} |\boldsymbol{\Lambda}^{jk}|^{-1/2} \times \exp \left[-\frac{1}{2} \boldsymbol{\eta}_{t|t-1}^{jk'} (\boldsymbol{\Lambda}^{jk})^{-1} \boldsymbol{\eta}_{t|t-1}^{jk} \right]$$

for

$$\mathbf{f}_{t|t-1}^{jk} = \boldsymbol{\mu}_j + \mathbf{G}_j (\mathbf{f}_{t-1} - \boldsymbol{\mu}_i)$$

$$\eta_{t|t-1}^{jk} = \mathbf{y}_t - \mathbf{a}_j - \mathbf{b}_j \left(\mathbf{f}_{t|t-1}^{jk} - \boldsymbol{\mu}_j \right)$$

$$\text{and } \Lambda^{jk} = \mathbf{b}_j \Omega_j \mathbf{b}_j' + \mathbf{T} \Sigma_j \mathbf{T}'$$

$$\Pr[s_{t-1} = j, s_t = k | I_{t-1}, \boldsymbol{\psi}] = p_{jk} \Pr[s_{t-1} = j | I_{t-1}, \boldsymbol{\psi}]$$

and

$$f[\mathbf{y}_t | I_{t-1}, \boldsymbol{\psi}] = \sum_{k=j}^{j+1} \sum_{j=1}^{q+1} f[\mathbf{y}_t | I_{t-1}, s_{t-1} = j, s_t = k, \boldsymbol{\psi}] \Pr[s_{t-1} = j, s_t = k | I_{t-1}, \boldsymbol{\psi}]$$

In the backward pass, one simulates \mathbf{S}_n by the method of composition. One samples s_n from $\Pr[s_n = 1 | I_n, \boldsymbol{\psi}]$. Then for $t = 1, 2, \dots, n-1$ we sequentially calculate

$$\begin{aligned} \Pr[s_t = j | I_t, s_{t+1} = k, S^{t+2}, \boldsymbol{\psi}] &= \Pr[s_t = j | I_t, s_{t+1} = k, \boldsymbol{\psi}] \\ &= \frac{\Pr[s_{t+1} = k | s_t = j] \Pr[s_t = j | I_t, \boldsymbol{\psi}]}{\sum_{k=j}^{j+1} \Pr[s_{t+1} = k | s_t = j] \Pr[s_t = j | I_t, \boldsymbol{\psi}]} \end{aligned}$$

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. It is either the value k or $(k-1)$. We remark that in these sampling steps, for instance, if s_n turns out to be q and not $(q+1)$, then q is taken to be the absorbing regime and the parameters of regime $(q+1)$ are drawn from the prior in that iteration. In our data, however, $(q+1)$ is always drawn because the last change-point occurs in the interior of the sample and, therefore, the distribution $\Pr[s_n = 1 | I_n, \boldsymbol{\psi}]$ has almost a unit mass on $(q+1)$.

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 Σ^* is analytically tractable and takes the form of an Inverse Gamma density. Thus, for $i \in \{1, \dots, 7, 9, \dots, 16\}$ and $j = 1, 2, \dots, q+1$, $\sigma_{i,j}^{*2}$ is sampled from

$$\mathbf{IG} \left\{ \frac{\bar{v} + \sum_{t=1}^n I(s_t = j)}{2}, \frac{\bar{d} + \sum_{t=1}^n d_{i,j} I(s_t = j) (R_{it} - \bar{a}_{s_t}^w(i) - \bar{\mathbf{B}}_{s_t}^w(i)(\mathbf{f}_t - \boldsymbol{\mu}_j))^2}{2} \right\}$$

where $I(\cdot)$ is the indicator function, and $\bar{a}_{s_t}^w(i)$ and $\bar{\mathbf{B}}_{s_t}^w(i)$ denote the i th rows of $\bar{\mathbf{a}}_{s_t}^w$ and $\bar{\mathbf{B}}_{s_t}^w$, respectively.

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*.
- Bansal, R. and Zhou, H. (2002), “Term structure of interest rates with regime shifts,” *Journal of Finance*, LVII, 463–473.
- Carter, C. K. and Kohn, R. (1994), “On Gibbs sampling for state space models,” *Biometrika*, 81, 541–553.
- 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*, in press.

- 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. (2009), “Tailored randomized-block MCMC methods for analysis of DSGE models,” *working paper*.
- 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.