January 2010

Essays on Macro-Finance Asset Pricing Models and Estimation

Kyu Ho Kang
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Economics

Dissertation Examination Committee:
Siddhartha Chib, Chair
Hong Liu
James C. Morley
Werner Ploberger
Raul Santaeulalia-Llopis
Yongs Shin

Essays on Macro-Finance Asset Pricing Models and Estimation

by

Kyu Ho Kang

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

August 2010

Saint Louis, Missouri

Abstract of The Dissertation

Essays on Macro-Finance Asset Pricing Models and Estimation

by

Kyu Ho Kang

Doctor of Philosophy in Economics

Washington University in St. Louis, 2010

Professor Siddhartha Chib, Chair

In my dissertation, I focus on theoretical asset pricing models and the development of Bayesian econometric methods to estimate them, particularly in the area of bond pricing.

The first essay theoretically and empirically examines 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 all model parameters are subject to change at unknown time points. The bonds in our set-up can be priced straightforwardly once the change-point model is formulated as a specific unidirectional Markov process. We consider five versions of our general model - with 0, 1, 2, 3 and 4 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 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 and that the breaks occurred

in 1980:II, 1985:IV 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 the 3 change-point model is substantially better than that of the other models we examine.

In the second essay, we develop and estimate a model of the term structure of interest rates within the context of a Dynamic Stochastic General Equilibrium model. The model features multiple monetary policy and volatility regimes. We estimate this model by Bayesian methods. Our estimation results reveal that U.S. monetary policy has become "more active" since 1995:Q2, that during this period, the average term premium has fallen, and that the price of regime shift risk is always significantly positive over time. These findings highlight the important role that general equilibrium modeling can play in understanding the complex dynamics of the term structure.

Acknowledgements

I would like thank to thank my advisor Siddhartha Chib, Srikanth Ramamurhty, James Morley, Werner Ploberger, Raul Santaeulalia-Llopis, Yongs Shin and Hong Liu for valuable feedback on the manuscript.

Last, and most, thank my wife Yoo Rim.

Contents

List of Tables

List of Figures

CHAPTER 1

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

Kyu Ho Kang and Siddhartha Chib

1.1 Introduction

In this paper we theoretically and empirically examine structural changes in a dynamic term-structure model of zero-coupon bond yields. We do our analysis in the setting of arbitrage-free multi-factor affine models of the type developed in Duffie and Kan (1996) and Dai and Singleton (2000) though we allow for both latent and macro-economic factors along the lines of Ang and Piazzesi (2003), Ang, Dong, and Piazzesi (2007) and Chib and Ergashev (2009). We depart from the existing modeling of structural changes, however, by relying on a change point process rather than the Markov switching process of Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), and Ang, Bekaert, and Wei (2008).

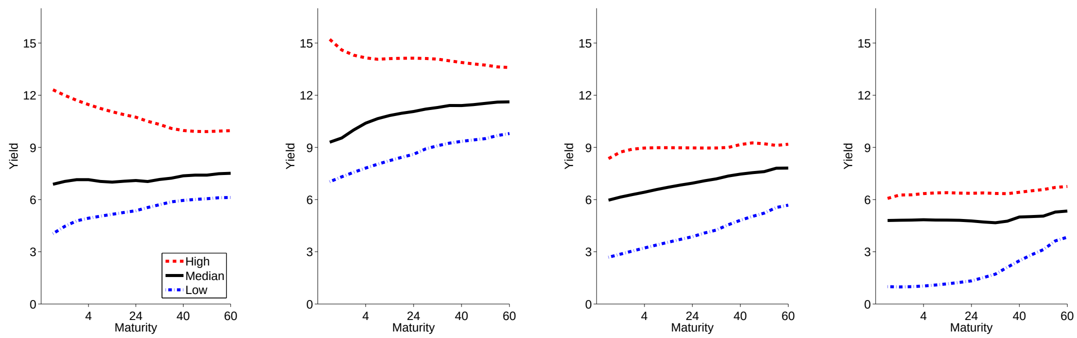
The model we develop and estimate provides a new perspective on the dynamics of zero-coupon bond prices and yields. One reason is because our change-point approach reflects a different view of regime-changes. In a change point specification, a regime once occupied and vacated is never visited again. In contrast, in a Markov switching model, the regimes recur, which implies that a regime occupied in the past (whether distant or near) can occur in the future. The latter assumption may not be germane if one believes that the confluence of conditions that determine a regime are unique and not repeated.

Another reason is because we derive bond prices under the assumption that all parameters in the model can change whereas in previous work some parameters are assumed to be constant across regimes. Thus, in our formulation, we do not have to decide which parameters are constant and which break. As we show, bond prices can be obtained straightforwardly once the change point process is formulated in the manner of Chib (1998) as a specific unidirectional Markov process.

A third reason is because in our empirical analysis we deal with a larger set of maturities than in previous work. This allows us to get finer view of the term-structure than is possible with a smaller set of maturities. In particular, we apply our model to 16 yields of US T-bills measured quarterly between 1972:I and 2007:IV. An added benefit of working with these many yields is that (in comparison with models with fewer yields) the model with 16 yields produces the best forecasts of the term-structure. The reason for this, which apparently has not been documented or exploited before, is that the addition of new yields introduces only the parameters that represent the pricing error variances, but because the parameters are subject to several cross-equation restrictions, the additional outcomes are helpful in estimation and, hence, in predictive inferences.

A notable aspect of our approach is that the prior distribution is motivated by economic considerations. In particular, our prior on the parameters reflects the assumption of a positive term-premium, following Chib and Ergashev (2009). Another aspect is that our estimation approach which is implemented by tuned Markov chain Monte Carlo methods, is both feasible and reliable. We apply this approach successfully to fit a model that has 209 parameters. Models of this size in this context would be difficult to fit by non-Bayesian methods because of the severe non-linearities and the potential multi-modality of the likelihood function. Our Bayesian approach is also relevant in this context because it offers a straightforward way to compare different change point models through marginal likelihoods and Bayes factors.

Our empirical analysis is organized around 5 different versions of the general model. These models, which we label as $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$, contain 0, 1, 2, 3 and 4 change-points, respectively. Our main findings are as follows. The 3 change point model, $\mathcal{M}_3$, is the one that is most supported by the data (in comparison with models with 0, 1, 2 and 4 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. Thus, the most recent break occurs in 1995, not 1985, as is commonly believed. That the underlying distribution of the term-structure is different in the regimes isolated by these change-points can be seen in Figure 1.1 where we display the 5%, 50% and 95% quantiles of the yield curve data categorized by regime. As we discus 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. We conclude our empirical analysis by predicting the yield curve out-of-sample and find that the predictive performance of our best model is substantially better than that of the other models we consider.



(a) 1972:I-1980:I (b) 1980:II-1985:III (c) 1985:IV-1995:I (d) 1995:II-2006:IV

Figure 1.1: Term structure of interest rates

The rest of the paper is organized as follows. In Section 1.2 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 1.3, deferring details of the MCMC simulation procedure to the appendix of the paper. Section 1.4 deals with the empirical analysis of the real data and Section 1.5 has our conclusions.

1.2 Model Specification

In this section we develop our model of bond pricing under regime changes. Essentially, we will explain the dynamics of bond prices in terms of the evolution of a discrete time, discrete-state variable $\{s_t\}$ that takes one of the values $\{1, 2, .., m+1\}$ such that $s_t = j$ indicates that the time t observation has been drawn from the jth regime, and in terms of the evolution of three continuous factors $\mathbf{f}_t$ consisting 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

$$P_t(s_t, \tau) = \mathbf{E}_t \left[\kappa_{t,s_t,t+1} P_{t+1}(s_{t+1}, \tau - 1) \right] \tag{1.2.1}$$

where $\mathbf{E}_t$ 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 thence available as

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

Our goal now is to 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 then derive the prices of our default-free zero coupon bonds that satisfy the preceding risk-neutral pricing condition.

1.2.1 Change Point Process

We assume that the process of regime-changes is governed by $s_t \in \{1, 2, ..., m + 1\}$. When $s_t = j$, the tth observation is assumed to be drawn from regime j. We refer to the times $\{t_1,\ t_2, ...,\ t_m\}$ at which s_t jumps from one value to the next as the change-points. We will suppose that the parameters in the $(m+1)$ regimes induced by these m change-points are different. As mentioned in Section 1, we describe the stochastic evolution of s_t in terms of a change point instead of a Markov switching process. In this we follow Chib (1998). We suppose 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 $(j+1)$. In this sense $\{s_t\}$ can be viewed as a unidirectional process. Thus, in this formulation, return visits to a previously occupied state are not possible. Then, it follows that the jth change point occurs at time (say) t_j when $s_{t_j-1} = j$ and $s_{t_j} = j+1$ $(j = 1, 2, .., m)$. We further assume that s_t follows a Markov process with transition probabilities given by

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1-p_{11} & 0 & \cdots & 0 \\ 0 & p_{22} & 1-p_{22} & \cdots & 0 \\ 0 & 0 & p_{33} & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & p_{m+1,m+1} \end{bmatrix} \tag{1.2.2}$$

where $p_{jk} = \Pr[s_{t+1} = k | s_t = j]$ and, $p_{jk} = 1 - p_{jj}$, $k = j+1$ and $p_{m+1,m+1} = 1$ $(j = 1, 2, .., m)$.

A feature of this specification is an absorbing terminal state. This is intentional because in any setting with a finite observation window one must have an upper limit on the number of change-points (equivalently, the number of possible regimes). An upper limit on the number of change-points does not rule out, however, the possibility of breaks beyond the observation window. Although such breaks can occur it is not possible to make inferences about them from the sample data without making consequential and unverifiable assumptions.

An interesting point is that we can assume that the (infinitely lived) economic agents face a possible infinity of change-points. Regardless of the number of change points, however, as is typical in finance and economic theorizing, we assume that these agents know the parameters in the various regimes. Furthermore, in the asset pricing context, we assume that these agents know the current value 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.

This formulation of the change point model in terms of a restricted unidirectional Markov process facilitates bond 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) 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.

1.2.2 Factor Process

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

$$\mathbf{f}_{t+1} = \boldsymbol{\mu}_{s_{t+1}} + \mathbf{G}_{s_{t+1}}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\eta}_{t+1} \tag{1.2.3}$$

where on letting $\mathcal{N}_3(.,.)$ denote the 3-dimensional normal distribution, $\boldsymbol{\eta}_{t+1}|s_{t+1} \sim \mathcal{N}_3(\mathbf{0}, \Omega_{s_{t+1}})$, 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

$$\boldsymbol{\eta}_{t+1} = \mathbf{L}_{s_{t+1}} \boldsymbol{\omega}_{t+1} \tag{1.2.4}$$

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

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.

1.2.3 Stochastic Discount Factor

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

$$\kappa_{t,s_t,t+1} = \exp\left(-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t} - \boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\omega}_{t+1}\right) \tag{1.2.5}$$

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

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

$$r_{t,s_t} = \delta_{1,s_t} + \boldsymbol{\delta}'_{2,s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \tag{1.2.6}$$

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 $\boldsymbol{\gamma}_{t,s_t}$ are governed by

$$\boldsymbol{\gamma}_{t,s_t} = \tilde{\boldsymbol{\gamma}}_{s_t} + \Phi_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \tag{1.2.7}$$

where $\tilde{\boldsymbol{\gamma}}_{s_t} : 3 \times 1$ is the regime-dependent expectation of $\boldsymbol{\gamma}_{t,s_t}$ and $\Phi_{s_t} : 3 \times 3$ is a matrix of regime-specific parameters. We refer to the collection $(\tilde{\boldsymbol{\gamma}}_{s_t},\ \Phi_{s_t})$ as the factor-risk parameters. Note that in this specification $\boldsymbol{\gamma}_{t,s_t}$ is the same across maturities but different across regimes. A point to note is that negative market prices of risk have the effect of generating a positive term premium. This is important to keep in mind when we construct the prior distribution on the risk parameters.

It is easily checked that $\mathbf{E}\left[\kappa_{t,s_t,t+1}|\mathbf{f}_t, s_t = j\right]$ is equal to the price of a zero coupon bond with $\tau = 1$:

$$\begin{aligned} \mathbf{E}\left[\kappa_{t,s_t,t+1}|\mathbf{f}_t, s_t = j\right] &= \sum_{s_{t+1}=j}^{j+1} p_{js_{t+1}} \mathbf{E}\left[\kappa_{t,s_t,t+1}|\mathbf{f}_t, s_t = j, s_{t+1}\right] \\ &= \exp\left(-r_{t,j}\right), \ j \in \{1, 2, .., m\} \end{aligned} \tag{1.2.8}$$

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

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. Regime risk cannot also be isolated in the models of Ang et al. (2008) and Bansal and Zhou (2002) for the reason that it is confounded with the market price of factor risk.

1.2.4 Bond Prices

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

$$P_t(s_t, \tau) = \mathbf{E}_t\left[\kappa_{t,s_t,t+1} P_{t+1}(s_{t+1}, \tau - 1)\right] \tag{1.2.9}$$

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

$$P_t(s_t,\tau) = \exp(-\tau R_{t,\tau,s_t}) \tag{1.2.10}$$

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

$$R_{t,\tau,s_t} = \frac{1}{\tau}a_{s_t}(\tau) + \frac{1}{\tau}\mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) \tag{1.2.11}$$

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. By the law of the iterated expectation, the risk-neutral pricing formula in (1.2.9) can be expressed as

$$1 = \mathbf{E}_t\left\{\mathbf{E}_{t,s_{t+1}}\left[\kappa_{t,s_t,t+1}\frac{P_{t+1}(s_{t+1},\tau-1)}{P_t(s_t,\tau)}\right]\right\} \tag{1.2.12}$$

where the inside expectation $\mathbf{E}_{t,s_{t+1}}$ is conditioned on s_{t+1}, s_t and $\mathbf{f}_t$. Subsequently, as discussed in Appendix A, one now substitutes $P_t(s_t,\tau)$ and $P_{t+1}(s_{t+1},\tau-1)$ from (1.2.10) and (1.2.11) into this expression, and integrate out s_{t+1} after a log-linearization. We match common coefficients and solve for the unknown functions. When $j \in \{1,..,m\}$ and $k = j+1$, this procedure produces the following recursive system for the unknown functions

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

$$\mathbf{b}_j(\tau) = \begin{pmatrix} p_{jj} & p_{jk} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{2,j} + (\mathbf{G}_j - \mathbf{L}_j \Phi_j)' \, \mathbf{b}_j(\tau - 1) \\ \boldsymbol{\delta}_{2,j} + (\mathbf{G}_k - \mathbf{L}_k \Phi_j)' \, \mathbf{b}_k(\tau - 1) \end{pmatrix} \tag{1.2.13}$$

where τ runs over the positive integers. These recursions are initialized by setting $a_{s_t}(0) = 0$ and $\mathbf{b}_{s_t}(0) = 0_{3\times1}$ 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 $s_t = j$ $(j \leq m)$ 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 given sample of data, and the number of change points m is a finite number, the above recursions are supplemented by the expressions

$$\begin{aligned} a_j(\tau) &= \delta_{1,j} - \tilde{\gamma}_j \mathbf{L}_j' \mathbf{b}_j(\tau - 1) - \mathbf{b}_j(\tau - 1)' \mathbf{L}_j \mathbf{L}_j' \mathbf{b}_j(\tau - 1)/2 + a_j(\tau - 1) \\ \mathbf{b}_j(\tau) &= \boldsymbol{\delta}_{2,j} + (\mathbf{G}_j - \mathbf{L}_j \Phi_j)' \, \mathbf{b}_j(\tau - 1) \end{aligned} \tag{1.2.14}$$

for $j = m + 1$.

Figure 1.2 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 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 (1.2.13) and (1.2.14). Conditioned on the parameters and $\mathbf{f}_{t-1}$, $\mathbf{f}_t$ is generated by the regime-specific autoregressive

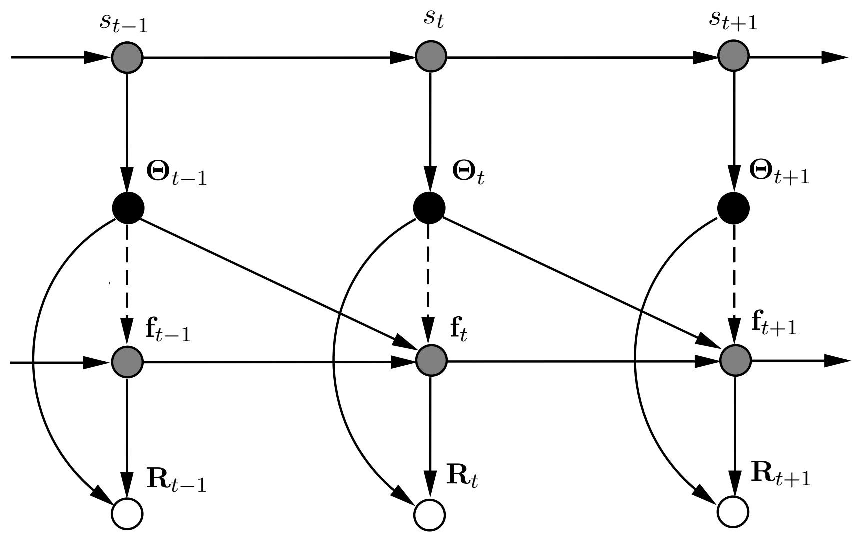


Figure 1.2: Directed graph of model linkages

process in (1.2.3). Finally, from (1.2.11), $a_{s_t}(\tau)$, $\mathbf{b}_{s_t}(\tau)$ and $\mathbf{f}_t$ determine the yields of all maturities. Notice that in Dai et al. (2007) the dashed line in figure 1.2 is absent since $\mathbf{f}_t$ is assumed to be drawn independently of s_t.

1.2.5 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

$$\begin{aligned}\text{Term-premium}_{\tau,t,s_t} &= (\tau-1)\mathsf{Cov}\left(\ln \kappa_{t,s_t,t+1}, R_{t+1,s_{t+1},\tau-1}|\mathbf{f}_t, s_t = j\right) \qquad (1.2.15)\\ &= -p_{jj}\mathbf{b}_j(\tau-1)'\mathbf{L}_j\boldsymbol{\gamma}_{t,j} - p_{jk}\mathbf{b}_k(\tau-1)'\mathbf{L}_k\boldsymbol{\gamma}_{t,j}\end{aligned}$$

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

1.3 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. As far as we know, this is the largest number of yields that have been considered in the setting of affine yield curve models. 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. We fit these various models by tuned Bayesian methods as we discuss below and then compare the competing models through marginal likelihoods, Bayes factors and the predictive performance out of sample.

To begin, let the 16 yields under study be denoted by

$$(R_{t1},\ R_{t2}, .., \ R_{t16})'\ ,\ t = 1, 2, ..., n, \tag{1.3.1}$$

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 ith maturity (in quarters). Let the two macro factors be denoted by

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

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

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

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 $\boldsymbol{\Omega}_{s_t}$ are defined as

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

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

$$\begin{pmatrix} 1/400 & 0 & 0 \\ l_{21,s_t} & \exp(l^*_{22,s_t}) & 0 \\ l_{31,s_t} & l_{32,s_t} & \exp(l^*_{33,s_t}) \end{pmatrix} \tag{1.3.2}$$

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

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

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

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

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

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

1.3.1 Joint distribution of the yields and macro factors

We now derive the joint distribution of the yields and the macro factors conditioned on $\mathbf{S}_n$ and $\boldsymbol{\psi}$. This joint distribution can be obtained without marginalization over $\{u_t\}_{t=1}^n$ if we assume (following, for example, Chen and Scott (2003) and Dai 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 (1.2.13) - (1.2.14). Also let $\bar{a}_{8,s_t}$ ($\bar{\mathbf{a}}_{s_t}$) and $\bar{\mathbf{b}}_{8,s_t}$ ($\bar{\mathbf{b}}_{s_t}$) be the corresponding intercept and factor loadings for R_{t8} ($\mathbf{R}_t$), respectively. Then, since the basis yield is priced without error, if we let

$$\bar{\mathbf{b}}_{8,s_t} = \begin{pmatrix} \bar{b}_{8,u,s_t} \\ \bar{\mathbf{b}}_{8,m,s_t} \end{pmatrix} \tag{1.3.3}$$

we can see from (1.2.11) that R_{t8} is given by

$$R_{t8} = \bar{a}_{8,s_t} + \bar{b}_{8,u,s_t} u_t + \bar{\mathbf{b}}'_{8,m,s_t}(\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t}) \tag{1.3.4}$$

On rewriting this expression, it follows that u_t is

$$u_t = \left(\bar{b}_{8,u,s_t}\right)^{-1} \left(R_{t8} - \bar{a}_{8,s_t} - \bar{\mathbf{b}}'_{8,m,s_t}(\mathbf{m}_t - \boldsymbol{\mu}_{m,s_t})\right) \tag{1.3.5}$$

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

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

$$\begin{aligned} \alpha_{s_t} &= \begin{pmatrix} \left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{\mathbf{b}}'_{8,m,s_t} \boldsymbol{\mu}_{m,s_t} - \left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{a}_{8,s_t} \\ 0_{2\times 1} \end{pmatrix}, \text{ and} \\ \mathbf{A}_{s_t} &= \begin{pmatrix} \left(\bar{b}_{8,u,s_t}\right)^{-1} & -\left(\bar{b}_{8,u,s_t}\right)^{-1} \bar{\mathbf{b}}'_{8,m,s_t} \\ 0_{2\times 1} & \mathbf{I}_2 \end{pmatrix} \end{aligned} \tag{1.3.6}$$

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

$$\mathbf{f}_t = \alpha_{s_t} + \mathbf{A}_{s_t} \mathbf{z}_t \tag{1.3.7}$$

It now follows from equation (1.2.11) 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

$$\mathbf{R}_t = \bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(0, \Sigma_{s_t}) \tag{1.3.8}$$

where

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

In other words,

$$\begin{aligned} p(\mathbf{R}_t|\mathbf{z}_t, s_t, \boldsymbol{\psi}) &= p(\mathbf{R}_t|\mathbf{f}_t, s_t, \boldsymbol{\psi}) \\ &= \mathcal{N}_{15}(\mathbf{R}_t|\bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \Sigma_{s_t}) \end{aligned} \tag{1.3.9}$$

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 (1.2.3) and the linear map between $\mathbf{f}_t$ and $\mathbf{z}_t$ given in equation (1.3.7). In particular,

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{z}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) &= p(\mathbf{f}_t|\mathbf{f}_{t-1}, s_t, s_{t-1}, \boldsymbol{\psi}) \det(\mathbf{A}_{s_t}) \\ &= \mathcal{N}_3(\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t}(\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \Omega_{s_t})| \left(\bar{b}_{8,u,s_t}\right)^{-1} | \end{aligned} \tag{1.3.10}$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi}) = \prod_{t=1}^{n} \mathcal{N}_{15}(\mathbf{R}_t|\bar{\mathbf{a}}_{s_t} + \bar{\mathbf{b}}_{s_t}(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}), \Sigma_{s_t}) \tag{1.3.11}$$

$$\times \mathcal{N}_3(\boldsymbol{\mu}_{s_t} + \mathbf{G}_{s_t}(\mathbf{f}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}), \Omega_{s_t})| \left(\bar{b}_{8,u,s_t}\right)^{-1} | \tag{1.3.12}$$

1.3.2 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 $\boldsymbol{\mu}_{u,s_t} = 0$ 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

$$\mathcal{R} = \{\mathbf{G}_j, \delta_{21,j} | \delta_{21,j} \geq 0,\ 0 \leq p_{jj} \leq 1,\ \ |eig(\mathbf{G}_j)| < 1 \text{ for } j = 1, 2, .., m+1\} \tag{1.3.13}$$

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

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

which we explain as follows.

First, the prior on p_{jj} $(j = 1, .., m)$ is normal with a standard deviation of 0.33, truncated to the interval $(0, 1)$. The mean of these distributions is model-specific. For example, in the $\mathcal{M}_1$ model, the mean 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 (1.3.8) are a function of p_{jj}, which eliminates any benefit from the use of a beta functional form. *Second*, we construct $\bar{\mathbf{g}}_{s_t}$ from the matrix

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

and let $\bar{\mathbf{V}}_{\mathbf{g}_{s_t}}$ be a 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 $\bar{\mathbf{V}}_{\boldsymbol{\mu}_{m,s_t}} \times 400^2 = \mathsf{diag}(25, 1)$. 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

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

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

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

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

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

which tends to imply reasonable prior variation in the implied yield curve.

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

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

$$u_0 \sim \mathcal{N}(0, V_u) \tag{1.3.14}$$

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

To show what these assumptions imply for the outcomes, we simulate the parameters 50,000 times from the prior, and for each drawing of the parameters, we simulate the factors and yields for each maturity and each of 50 quarters. The median, 2.5% and 97.5% quantile surfaces of the resulting term structure in annualized percents are reproduced in Figure 1.3. Because our prior distribution is symmetric among the regimes, the prior distribution of the yield curve is not regime-specific. It can be seen that the simulated prior term structure is gently upward sloping on average. Also the assumed prior allows for considerable a priori variation in the term structure.

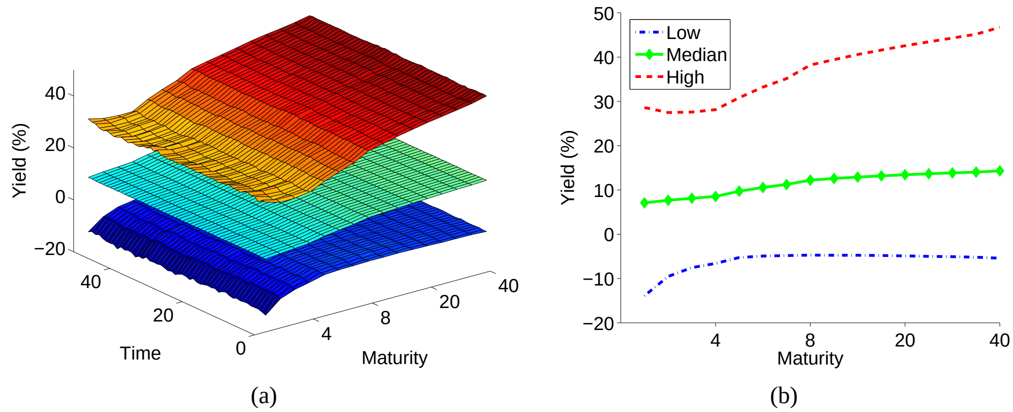


Figure 1.3: The implied prior term structure dynamics

1.3.3 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

$$\pi(\mathbf{S}_n, \boldsymbol{\psi}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})p(\mathbf{S}_n|\boldsymbol{\psi})\pi(\boldsymbol{\psi}) \tag{1.3.15}$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})\pi(u_0|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{1.3.16}$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})p(\mathbf{S}_n|\boldsymbol{\psi})\pi(u_0|\boldsymbol{\theta}) \tag{1.3.17}$$

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

$$p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\psi})\pi(\boldsymbol{\sigma}^{*2}) \tag{1.3.18}$$

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 $u_{\mathbf{0}}$ 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 B.

1.3.4 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 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 marginal likelihood of any given model is obtained as

$$m(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{S}_n, \psi)p(\mathbf{S}_n|\boldsymbol{\psi})\pi(\boldsymbol{\psi})d(\mathbf{S}_n, \boldsymbol{\psi}) \tag{1.3.19}$$

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

$$m(\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\psi}^*)\pi(\boldsymbol{\psi}^*)}{\pi(\boldsymbol{\psi}^*|\mathbf{y})} \tag{1.3.20}$$

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

$$\ln \hat{m}(\mathbf{y}) = \ln p(\mathbf{y}|\boldsymbol{\psi}^*) + \ln \pi(\boldsymbol{\psi}^*) - \ln \hat{\pi}(\boldsymbol{\psi}^*|\mathbf{y}) \tag{1.3.21}$$

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

$$p(\mathbf{y}|\boldsymbol{\psi}) = \sum_{t=0}^{n-1} \ln p\left(\mathbf{y}_{t+1}|I_t, \boldsymbol{\psi}\right) \tag{1.3.22}$$

where

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

is the one-step ahead predictive density of $\mathbf{y}_{t+1}$, and I_t consists of the history of the outcomes R_t and $\mathbf{z}_t$ up to time t. On the right hand side, the first term is the density of $\mathbf{y}_{t+1}$ conditioned on $(I_t, s_t, s_{t+1}, \boldsymbol{\psi})$ which is given in equation (1.3.11), whereas the second term can be calculated from the law of total probability as

$$\mathsf{Pr}[s_t = j, s_{t+1} = k | I_t, \boldsymbol{\psi}] = p_{jk} \, \mathsf{Pr}[s_t = j | I_t, \boldsymbol{\psi}] \tag{1.3.23}$$

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

$$\mathsf{Pr}[s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}] = \sum_{j=1}^{m+1} \mathsf{Pr}[s_t = j, s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}] \tag{1.3.24}$$

where

$$\mathsf{Pr}[s_t = j, s_{t+1} = k|I_{t+1}, \boldsymbol{\psi}] = \frac{p\left[\mathbf{y}_{t+1}|I_t, s_t = j, s_{t+1} = k, \boldsymbol{\psi}\right] \mathsf{Pr}[s_t = j, s_{t+1} = k|I_t, \boldsymbol{\psi}]}{p\left[\mathbf{y}_{t+1}|I_t, \boldsymbol{\psi}\right]} \tag{1.3.25}$$

This completes the calculation of the likelihood function.

1.4 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 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 1.4.4. We measure the predictive accuracy of the forecasts in terms of the posterior predictive criterion (PPC) of Gelfand and Ghosh (1998). For a given model with λ number of the maturities, PPC is defined as

$$\mathsf{PPC} = \mathsf{D} + \mathsf{W} \tag{1.4.1}$$

where

$$\mathsf{D} = \frac{1}{\lambda + 2} \sum_{i=1}^{\lambda+2} \sum_{t=1}^{T} \mathsf{Var}\left(\tilde{y}_{i,t} | \mathbf{y}, \mathcal{M}\right) , \tag{1.4.2}$$

$$\mathsf{W} = \frac{1}{\lambda + 2} \sum_{i=1}^{\lambda+2} \sum_{t=1}^{T} \left[y_{i,t} - E\left(\tilde{y}_{i,t} | \mathbf{y}, \mathcal{M}\right)\right]^2 \tag{1.4.3}$$

$\{\tilde{\mathbf{y}}_t\}_{t=1,2,..,T}$ are the predictions of the yields and macro factors $\{\mathbf{y}_t\}_{t=1,2,..,T}$ under model $\mathcal{M}$, and $\tilde{y}_{i,t}$ and $y_{i,t}$ are the ith 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 1.1, the model with 16 maturities outperforms the models with fewer maturities. The reason

The number	No change point model		
of maturities(λ)	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 1.1: Posterior predictive criterion

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

1.4.1 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

$$1 + 2\sum_{k=1}^{K} \rho(k), \tag{1.4.4}$$

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.

1.4.2 The Number and Timing of Change Points

Table 1.2 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.	$\Pr[\mathcal{M}_m \mid \mathbf{y}]$	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

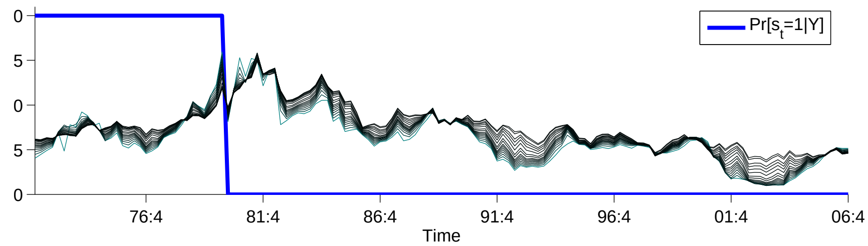
Table 1.2: Log likelihood (lnL) and log marginal likelihood (lnML)

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 1.4. 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. Rudebusch and Wu (2008) also find a change point in the year of 1985. The finding of a break point in 1995 is striking as it has not been isolated from previous regime-change models.

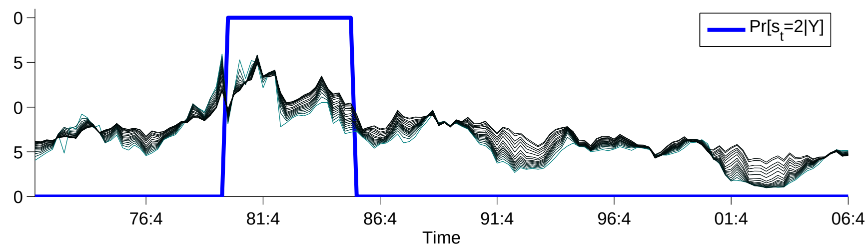
We would like to emphasize that our estimates of the change points from the models without macro factors are exactly the same as those from the change point models with macro factors. We do not report those results in the interest of space. In addition, the results are not sensitive to our choice of 16 maturities, as we have confirmed.

1.4.3 Parameter Estimates

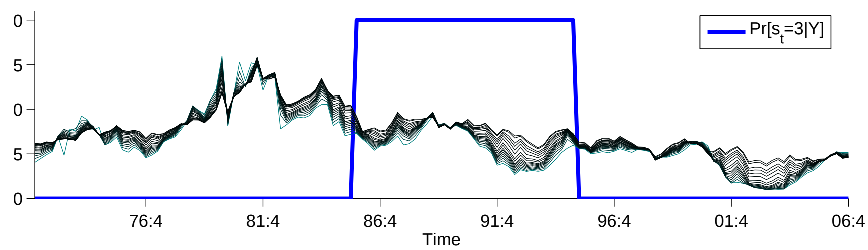
Table 1.3 summarizes the posterior distribution of the parameters. One point to note is that the posterior densities are generally different from the prior given in section 1.3.2, which implies that the data is informative about these parameters. We focus on various



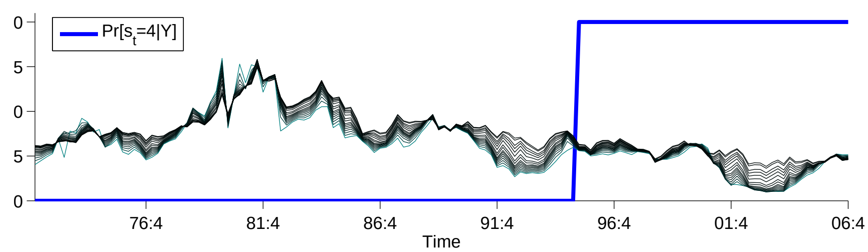
(a) $s_t = 1$



(b) $s_t = 2$



(c) $s_t = 3$



(d) $s_t = 4$

Figure 1.4: Model $\mathcal{M}_3$: $\Pr(s_t = j|\mathbf{y})$

aspects of this posterior distribution in the subsequent subsections.

	Regime 1			Regime 2			Regime 3			Regime 4		
	0.90	0.07	0.15	**0.95**	-0.01	0.03	**0.92**	0.15	0.31	**0.93**	0.04	0.23
	(0.06)	(0.10)	(0.15)	(0.03)	(0.07)	(0.06)	(0.06)	(0.21)	(0.17)	(0.04)	(0.17)	(0.29)
G	-0.24	**0.67**	-0.07	-0.07	**0.73**	-0.10	**0.15**	**0.35**	0.08	0.02	**0.91**	0.01
	(0.26)	(0.23)	(0.12)	(0.05)	(0.05)	(0.03)	(0.06)	(0.14)	(0.08)	(0.02)	(0.13)	(0.06)
	-0.06	-0.16	0.26	0.09	-0.35	**0.52**	-0.04	0.00	**0.34**	-0.03	-0.37	0.19
	(0.25)	(0.23)	(0.17)	(0.17)	(0.24)	(0.17)	(0.09)	(0.21)	(0.13)	(0.08)	(0.26)	(0.15)
μ	0.00	**4.99**	3.54	0.00	**5.88**	**2.63**	0.00	**2.56**	**2.62**	0.00	**1.49**	**3.22**
$\times 400$		(2.17)	(0.90)		(0.41)	(1.00)		(0.41)	(0.49)		(0.80)	(0.53)
	1.00			1.00			1.00			1.00		
L	0.11	**1.72**		0.10	**1.48**		0.11	**0.74**		-0.47	**0.82**	
$\times 400$	(0.40)	(0.19)		(0.44)	(0.13)		(0.34)	(0.13)		(0.59)	(0.12)	
	-0.67	-0.62	**4.28**	0.24	0.27	**4.58**	-0.55	-0.18	**2.00**	-0.13	-0.20	**2.03**
	(0.88)	(0.39)	(0.14)	(0.62)	(0.41)	(0.17)	(0.56)	(0.14)	(0.12)	(0.89)	(0.14)	(0.11)
δ_1		**9.23**			**2.78**			**4.42**			**4.34**	
$\times 400$		(1.69)			(1.60)			(1.18)			(1.00)	
$\boldsymbol{\delta}_2$	**1.16**	0.09	0.17	**1.29**	0.25	0.16	**0.72**	0.31	0.26	**0.57**	0.56	0.10
	(0.13)	(0.23)	(0.22)	(0.16)	(0.23)	(0.15)	(0.09)	(0.26)	(0.21)	(0.07)	(0.37)	(0.25)
$\boldsymbol{\gamma}$	-0.28	-0.40	-0.22	-0.34	-0.65	-0.21	-0.58	-0.56	-0.05	-0.34	-0.25	-0.19
	(0.28)	(0.30)	(0.26)	(0.25)	(0.21)	(0.26)	(0.28)	(0.33)	(0.24)	(0.25)	(0.25)	(0.27)
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 1.3: Model $\mathcal{M}_3$: Parameter estimates

Factor Process

Figure 1.5 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 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 persistence,

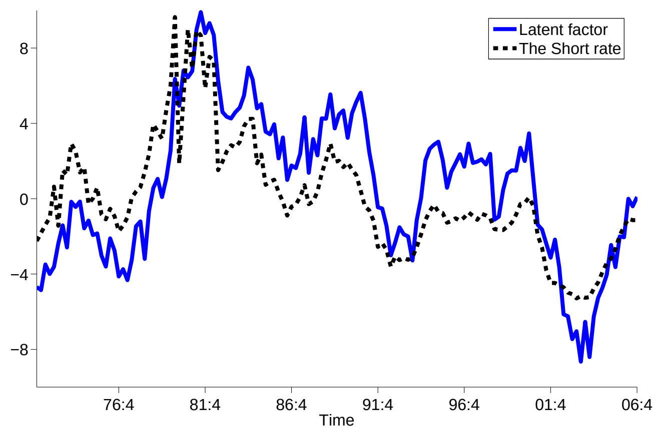


Figure 1.5: Model $\mathcal{M}_3$: Estimates of the latent factor

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

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

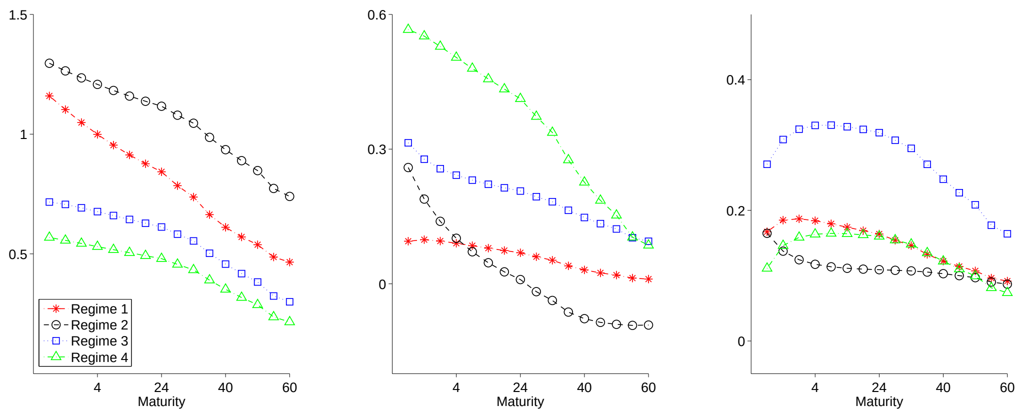
$$eig(\mathbf{G}_3) = \begin{bmatrix} 0.935 \\ 0.312 \\ 0.366 \end{bmatrix}, \; eig(\mathbf{G}_4) = \begin{bmatrix} 0.913 + 0.044i \\ 0.913 - 0.044i \\ 0.204 \end{bmatrix}$$

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 suggest a reduction in factor volatility starting from the middle of the 1980s, which coincides with the period that is called the great moderation (Kim, Nelson, and Piger (2004)).

Factor Loadings

The factor loadings in the short rate equation, $\boldsymbol{\delta}_{2,s_t}$ are all positive, which is consistent with the conventional wisdom that central bankers tend to raise the interest rate in response to a positive shock to the macro factors. It can also be seen that $\boldsymbol{\delta}_{2,s_t}$ along with $\mathbf{G}_{s_t}$ and $\mathbf{L}_{s_t}$ are different across regimes, which makes the factor loadings regime-dependent across the term structure as revealed in figure 1.6. This finding lends support to our assumption of regime-dependent factor loadings.

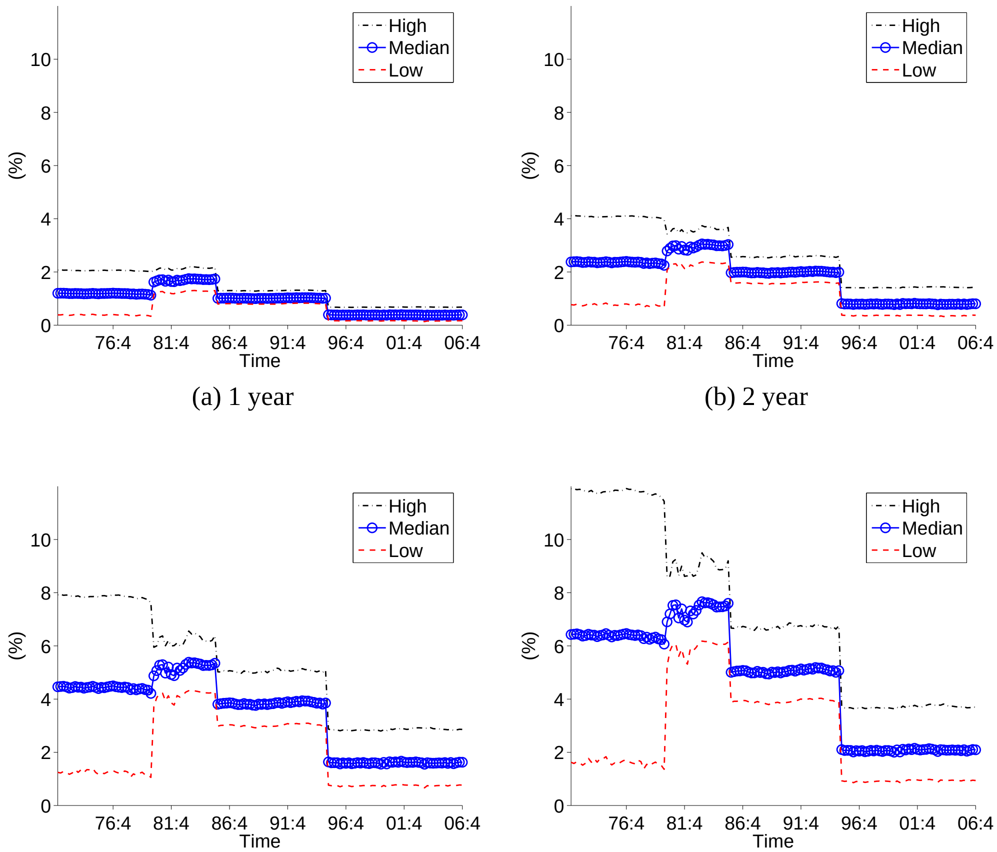


(a) Latent (b) Inflation (c) GDP growth

Figure 1.6: Model $\mathcal{M}_3$: Estimates of the factor loadings, $\bar{\mathbf{b}}_{s_t}$

Term Premium

Figure 1.7 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. A similar finding appears in Rudebusch, Sack, and Swanson (2007).



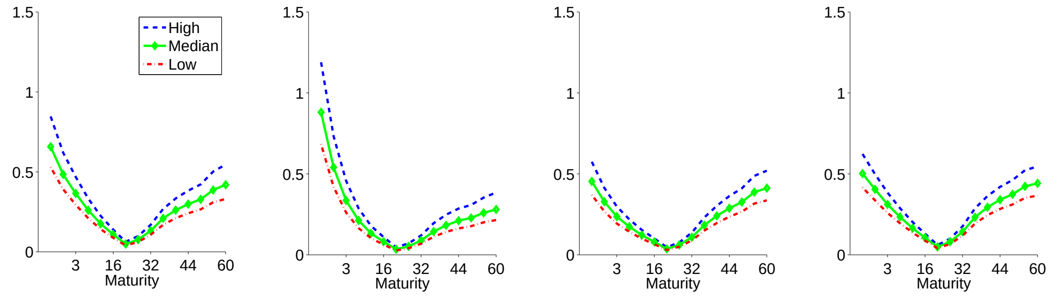
(a) 1 year (b) 2 year

(c) 5 year (c) 10 year

Figure 1.7: Model $\mathcal{M}_3$: Term premium

Pricing Error Volatility

In Figure 1.8 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.



(a) Regime 1 (b) Regime 2 (c) Regime 3 (c) Regime 4

Figure 1.8: Model $\mathcal{M}_3$: Term Structure of the Pricing Error Volatility

1.4.4 Forecasting and Predictive Densities

A principle objective of this paper is to 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 the parameters with respect to the posterior distribution of the parameters. If we let $\mathbf{y}_f$ denote the future observations, the predictive density under model $\mathcal{M}_m$ is given by

$$p(\mathbf{y}_f|\mathcal{M}_m, \mathbf{y}) = \int_{\boldsymbol{\psi}} p(\mathbf{y}_f|\mathcal{M}_m, \mathbf{y}, \boldsymbol{\psi})\pi(\boldsymbol{\psi}|\mathcal{M}_m, \mathbf{y})d\boldsymbol{\psi} \tag{1.4.5}$$

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 (1.2.3). Then given $\mathbf{f}_{n+1}$, the yields $\mathbf{R}_{n+1}$ are drawn using equation (1.3.8).

These two steps are iterated forward to produce the draws $\mathbf{f}_{n+i}$ and $\mathbf{R}_{n+i}$, $i = 1, 2, .., T$. 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.

We summarize the sampled predictive densities in Figure 1.9. The top panel gives

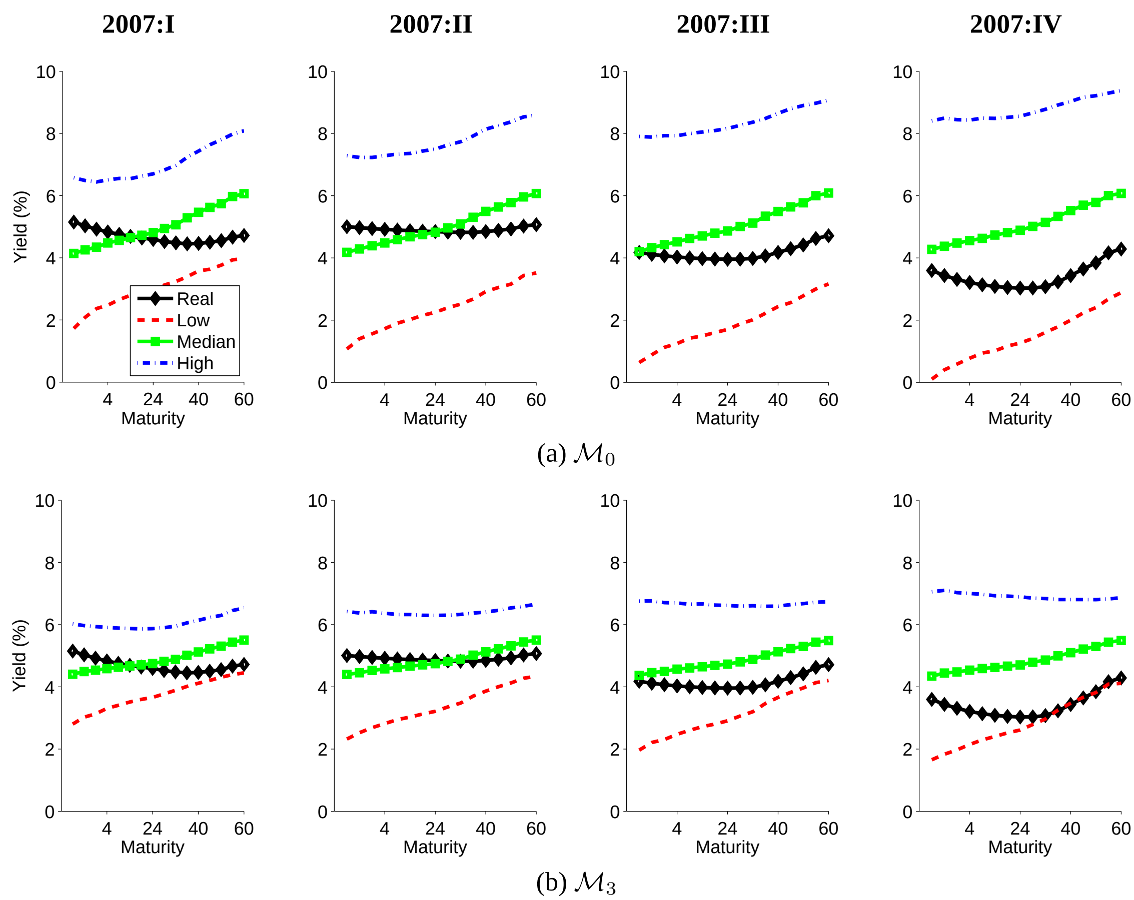


Figure 1.9: Predicted yield curve

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

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 1.4: Posterior predictive criterion

each case in Table 1.4. 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, 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 $\mathcal{M}_3$ 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.

1.5 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. The models we fit involve more bonds than has ever been attempted in the literature. This in turn leads to a better fit to the data. 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 suggests that the term structure has gone through three change points, and that the term structure and the risk premium are 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.

Chapter 2

Term Structure of Interest Rates in a DSGE Model with Regime Changes

Kyu Ho Kang, Siddhartha Chib and Srikanth Ramamurthy

2.1 Introduction

In this paper, we develop and estimate a general equilibrium model of the term structure of interest rates that features regime changes in monetary policy and volatilities of structural shocks. The model we construct is a New Keynesian dynamic stochastic general equilibrium (DSGE) model that comprises a representative household, a continuum of intermediate goods producers, a representative final goods producer, the government sector (which issues bonds of various maturities) and the central bank. The various agents in the model are intertemporal optimizers that face uncertainty arising from exogenous shocks to productivity, monetary policy and government expenditure. Since we are particularly interested in the role that monetary policy plays in shaping the dynamics of the term-structure, an important focus of our model is on the specification of the monetary policy rule. We specify the central bank's monetary policy function in terms of the generalized Taylor (1993) rule (Davig and Leeper (2007)). Following this rule, the central bank adjusts the nominal short rate in response to deviations of inflation and output from their target levels. An important aspect of this policy function is that the inflation and output coefficients are time varying to model the possibility of policy changes between active and less active regimes. Because of the way we formulate the model we are able to isolate the effect of such changes in monetary policy on the term-structure, factor risks and on the bond risk premium.

The second important aspect of our general equilibrium formulation is that the nominal pricing kernel and no arbitrage conditions are derived endogenously in the model. Because this pricing kernel is a function of the underlying model parameters, it (along with the equilibrium dynamics of the other aggregates) is affected by changes in the policy rule. We use this regime-dependent pricing kernel to find the arbitrage-free prices of bonds of various maturities by standard recursive methods.

The empirical implications of our model are isolated by econometric methods for DSGE models based on Bayesian ideas and MCMC simulation techniques that have emerged in the last few years (Chib and Ergashev, 2009, Chib and Ramamurthy, 2010). Despite the complex nature of the likelihood/posterior surface, our fitting method is efficient in terms of the metrics that are used to evaluate MCMC procedures.

The work in this paper can be viewed as a continuation of a recent line of enquiry into general equilibrium modeling of the term structure, as exemplified in Wachter (2006), Rudebusch and Swanson (2008b), and Wu (2006). Unlike these papers, however, we allow for structural changes (a feature that has been shown to be important in the partial equilibrium models of Rudebusch and Wu (2007) and Chib and Kang (2010)), and employ econometric methods to estimate the model, as opposed to calibrating it by simulation methods.

Our estimation results for U.S. quarterly data from 1986:Q4 to 2008:Q4, with bonds of maturities up to 20 quarters, reveal that (a) U.S. monetary policy has become "more active" since 1995:Q2, and that during this period, the average term premium and its volatility have fallen (b) the price of regime shift risk, while small compared to factor risk, is always significantly positive over time (c) although the term premium explains a significant portion of the term spread in the ("less active") first regime, its relative importance has fallen in the second regime2 and (d) the volatility of technology shock accounts for most of the volatility in the term premium.

The rest of the paper is organized as follows. In Section 2.2 we develop the model, discuss the solution procedure and derive the bond prices. Section2.3 provides the econometric details and Section 2.4 contains the empirical results. Concluding remarks are in Section 2.5.

2.2 Model

In this section we discuss the key aspects of our New Keynesian DSGE model with multiple monetary policy and volatility regimes. We present the model, derive the implied pricing kernel and compute the arbitrage-free τ maturity bond prices through the τ-forward iterations of the log-linearized Euler equation.

The model economy comprises a representative household, a continuum of intermediate goods producers indexed by $j \in [0, 1]$, a representative final good producer, the government sector and the central bank. The household maximizes its utility by supplying labor to the intermediate goods sector, consuming the finished good and making a portfolio decision over bonds of various maturities issued by the government. All firms maximize profits. A standard way of introducing market frictions in these models is to assume that the firms in the intermediate good sector face short run nominal rigidities in the form of quadratic price adjustment costs. In its goal to stabilize the economy, the central bank, following the Taylor (1993) rule, adjusts the short interest rate in response to output and inflation. As mentioned earlier, this policy function is time varying, depending on the (stochastic) state of the economy. The aggregate macroeconomic fluctuations in this model are driven by three structural shocks, namely a technology shock, a fiscal shock and a monetary policy shock. To capture the heteroskedastic nature of these shocks, we assume that their volatilities follow a two-state discrete time Markov switching process. As we show later in this section, these shocks play the analogous role of factors in the partial equilibrium framework.

In this economy, therefore, the agents' behavior is shaped by three sources of uncertainty - the policy regime s_t, the volatility regime v_t and the shocks themselves. The fundamental assumption regarding the agents' expectation of the future realizations of the aggregate variables (which are functions of the underlying uncertainties) is that they are based on rational expectations. That is, their expectations at time t, denoted $\mathbb{E}_t$, is based on the complete information set at time t that includes current and past realizations of all decision variables in the model, the regime sequences, $\{s_t,\ s_{t-1},\ s_{t-2},...\}$ and $\{v_t,\ v_{t-1},\ v_{t-2},...\}$, and the shocks. We denote this period-t information set as $\mathbb{I}_t$ and use $\mathbb{E}_t\left[X_{t+j}\right]$ and $\mathbb{E}\left[X_{t+j}|\mathbb{I}_t\right]$ interchangeably throughout the text to denote the j-period ahead expectation of X conditioned on $\mathbb{I}_t$. The agents also know the structural parameters of the model. The only unknowns in their information set are the future realizations of the shocks and the regimes. Given a specific stochastic process for the evolution of these regimes, the agents can form one step ahead expectations of the regimes and thus solve for the growth path of the macroeconomic aggregates as a function of the shocks.

2.2.1 The Representative Household

The representative household faces a consumption-leisure choice, deriving utility from consuming C_t units of the finished good purchased from the final good producer at the nominal price P_t and supplying H_t units of labor to the intermediate goods sector in return for a real wage rate of W_t. In addition to the wage income, the household earns real profits Q_t from the intermediate goods firms. Finally, the household carries a portfolio $\{B_t^\tau\}_{\tau=1}^{\tau^*}$ of nominal τ-quarter maturity zero-coupon bonds B_t^τ with current prices P_t^τ at any time t. We assume that the agent cares only about the time to maturity of the various bonds and not the date at which the bonds are issued. In other words, at time t, she is indifferent between holding a $(\tau+1)$ period maturity bond bought at time $t-1$ and a (τ) period maturity bond bought at time t, so that $B_{t-1}^{\tau+1} = B_t^\tau$. The government issues the multiple maturity bonds at a face value of unity. Current income and financial wealth brought over from the previous period $t-1$ are allocated between consumption, purchases of new bonds and a lumpsum real tax T_t levied by the government. The budget constraint of the household therefore satisfies

$$P_t C_t + \sum_{\tau=1}^{\tau^*} P_t^\tau B_t^\tau + T_t \leq P_t W_t H_t + \sum_{\tau=1}^{\tau^*-1} P_t^\tau B_{t-1}^{\tau+1} + B_{t-1}^1 + P_t Q_t. \tag{2.2.1}$$

The household then maximizes her expected utility function[1]

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty} \delta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\gamma} - 1}{1-\gamma} - H_{t+s}\right)\right] \tag{2.2.2}$$

[1] The simpler log utility function (where γ is fixed at 1) is not meaningful in this context because it generates a bond risk premium that is too small and stable relative to the data (Rudebusch and Swanson, 2008b). Alternatively, preferences could display habit persistence (modeled through a lagged consumption variable), as in Buraschi and Jiltsov (2007), Wachter (2006) and Rudebusch and Swanson (2008b), which can improve the model's ability to to fit the term premium and the nonlinearity of the spot rate process. We leave the examination of this possibility for future work because at the moment DSGE models with both habit persistence and multiple regimes cannot be solved.

subject to the intertemporal budget constraint (2.2.1) and available information up to time t. Here the variable A_t captures the general productivity level or aggregate technology, so that C_t/A_t measures the effective consumption per unit of technology. We assume that the growth rate of technology $a_t = A_t/A_{t-1}$ follows an autoregressive process

$$\ln a_t = (1 - \phi_a) \ln a^* + \phi_a \ln a_{t-1} + \varepsilon_{a,t} \tag{2.2.3}$$

where $|\phi_a| < 1$ and the innovation $\varepsilon_{a,t}$ is normally distributed with mean 0 and a regime-switching volatility process $\sigma^2_{a,v_{t,a}}$. Specifically, we assume that the volatility regime $v_{t,a}$ follows a two-state discrete time Markov process. The economic interpretation of these two regimes is that the economy transits between high volatility and low volatility states. Accordingly, we impose the identification restriction $\sigma_{a,2} > \sigma_{a,1}$, so that $v_{t,a} = 2$ denotes the higher volatility regime. The associated transition probability matrix for the volatility process is given by

$$\mathbf{Q}^a = \begin{bmatrix} q^a_{11} & 1 - q^a_{11} \\ 1 - q^a_{22} & q^a_{22} \end{bmatrix} \tag{2.2.4}$$

where $q^a_{ij} = \Pr[v_{t+1,a} = j | v_{t,a} = i]$.

2.2.2 The Final Good Sector

A representative firm in the finished goods sector combines a continuum of intermediate goods $Y_t(j)$ indexed by $j \in [0, 1]$ using the constant returns to scale production technology

$$\left(\int_0^1 Y_t(j)^{\frac{\zeta-1}{\zeta}} dj \right)^{\frac{\zeta}{\zeta-1}} \geq Y_t \tag{2.2.5}$$

where $\zeta > 1$ measures the elasticity of demand for each intermediate good. In each period $t = 0, 1, 2, \ldots$, it chooses the output level given the price P_t of the finished good and input prices $P_t(j)$. Profit maximization implies that the demand for intermediate goods is given by

$$P_t(j) = \left(\frac{Y_t}{Y_t(j)}\right)^{1/\zeta} P_t. \tag{2.2.6}$$

The aggregate price level is determined by the zero profit condition under competitive equilibrium as

$$P_t = \left(\int_0^1 P_t(j)^{1-\zeta} dj\right)^{\frac{1}{1-\zeta}}. \tag{2.2.7}$$

2.2.3 The Intermediate Good Sector

The intermediate good sector is characterized by a continuum of monopolistically competitive firms. Each firm indexed by j produces a unique, imperfectly substitutable, perishable good $Y_t(j)$ using a linear production technology with respect to the labor input $N_t(j)$ given the exogenous aggregate technology A_t in the economy

$$Y_t(j) = A_t N_t(j). \tag{2.2.8}$$

As mentioned earlier, the firms in the intermediate goods sector face nominal rigidities in the form of an explicit price adjustment cost. As is conventional in the literature, this price adjustment cost takes the quadratic form

$$AC_t(j) = \frac{\varphi}{2}\left(\frac{P_t(j)}{\pi^* P_{t-1}(j)} - 1\right)^2 Y_t \tag{2.2.9}$$

where $\varphi > 0$ measures the degree of price stickiness, $\pi_t = P_t/P_{t-1}$ is the inflation and π^* is the inflation target of the central bank in terms of the price of the final good. When selling its output to the final goods sector, each intermediate-good firm j chooses a sequence of input prices $P_t(j)$ to maximize the expected profits

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\Lambda_{t,t+s}Q_t(j)\right] \tag{2.2.10}$$

where the real profit at time t is

$$Q_t(j)=\frac{P_t(j)}{P_t}Y_t(j)-W_tN_t(j)-\frac{\varphi}{2}\left(\frac{P_t(j)}{\pi^*P_{t-1}(j)}-1\right)^2Y_t \tag{2.2.11}$$

and

$$\Lambda_{t,t+s}=\delta^s\left(\frac{C_{t+s}}{A_{t+s}}\right)^{-\gamma}\left(\frac{C_t}{A_t}\right)^{\gamma}\frac{A_t}{A_{t+s}} \tag{2.2.12}$$

is the representative household's "real" stochastic discount factor.

2.2.4 The Fiscal Authority

In addition to issuing bonds, the fiscal authority consumes a stochastic fraction ρ_t of the aggregate output Y_t. The government also levies a lump-sum tax or issues a subsidy to finance any shortfalls in government revenues. The government's (balanced) budget constraint is therefore given by

$$P_tG_t+\sum_{\tau=1}^{\tau^*-1}P_t^{\tau}B_{t-1}^{\tau+1}+B_{t-1}^{1}=T_t+\sum_{\tau=1}^{\tau^*}P_t^{\tau}B_t^{\tau} \tag{2.2.13}$$

where $G_t=\rho_tY_t$ is the real government expenditure. Here, the aggregate government spending shock is modeled as

$$\ln g_t=(1-\phi_g)\ln g^*+\phi_g\ln g_{t-1}+\varepsilon_{g,t} \tag{2.2.14}$$

where $g_t = 1/(1-\rho_t)$, $|\phi_g| < 1$, and, as in the case of the technology shock $\varepsilon_{a,t}$, the fiscal innovation $\varepsilon_{g,t}$ is assumed to be normally distributed with mean 0 and a regime-switching volatility process $\sigma^2_{g,v_{t,g}}$. We denote the transition probability matrix for the volatility process of the fiscal shock as

$$\mathbf{Q}^g = \begin{bmatrix} q^g_{11} & 1-q^g_{11} \\ 1-q^g_{22} & q^g_{22} \end{bmatrix} \qquad (2.2.15)$$

where $q^g_{ij} = \Pr[v_{t+1,g} = j | v_{t,g} = i]$.

2.2.5 Symmetric Equilibrium, Nonstochastic Values and the Linearized Model

From the utility maximization problem, the first-order condition with respect to the short term bond B^1_t has the form

$$P^1_t = \mathbb{E}_t\left[M_{t,t+1}\right] \qquad (2.2.16)$$

where

$$M_{t,t+1} = \delta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\frac{1}{a_{t+1}}\frac{1}{\pi_{t+1}} \qquad (2.2.17)$$

is the nominal stochastic discount factor (SDF) and $c_t = C_t/A_t$ is the stochastically detrended consumption at time t. Given the form of the SDF derived from our model, we use this condition in section 2.2.9 to price bonds of various maturities.

The aggregate labor supply from the household's problem is derived as

$$1 = \frac{W_t}{A_t}c_t^{-\gamma} \qquad (2.2.18)$$

In this economy, each intermediate goods producer faces the same marginal cost. Hence, in a symmetric equilibrium, $Y_t(j) = Y_t$, $H_t(j) = H_t$, $P_t(j) = P_t$ and $Q_t(j) = Q_t$. Thus, the representative intermediate-goods firm's first order condition for profit maximization implies

$$1 = \zeta - \zeta c_t^{\gamma} + \varphi \left(\frac{\pi_t}{\pi^*} - 1 \right) \left(\frac{\pi_t}{\pi^*} \right) - \varphi \mathbb{E}_t \left[\Lambda_{t,t+1} \left(\frac{\pi_{t+1}}{\pi^*} - 1 \right) \left(\frac{\pi_{t+1}}{\pi^*} \frac{Y_{t+1}}{Y_t} \right) \right] \tag{2.2.19}$$

Finally, the aggregate resource constraint must hold in equilibrium:

$$Y_t = C_t + G_t + AC_t \text{ and } H_t = N_t = \int_0^1 N_t(j) dj \tag{2.2.20}$$

which implies that

$$c_t = \left(1 - \frac{\varphi}{2} \left(\frac{\pi_t}{\pi^*} - 1 \right)^2 \right) x_t \tag{2.2.21}$$

where $x_t = Y_t / A_t$ denotes detrended output.

Further, from the Euler equation, the implied nonstochastic value of the gross nominal interest rate $R_t = 1/P_t^1$ denoted by R^* is

$$R^* = a^* \pi^* / \delta \tag{2.2.22}$$

Also the equation (2.2.21) implies that the nonstochastic value of the detrended output is determined by

$$x^* = \frac{c^*}{(1 - \rho^*)} \tag{2.2.23}$$

where the nostochastic value of the detrended consumption, c^* is

$$\left[\frac{\zeta - 1}{\zeta} \right]^{\frac{1}{\gamma}} \tag{2.2.24}$$

In the absence of shocks, the economy converges to a steady-state growth path along which all the stationary variables are constant over time.

Letting hats denote the percentage deviation of the variables from their respective steady state levels, for instance, $\hat{c}_t = \ln(c_t/c^*)$, the model whose equilibrium dynamics is summarized by the equations (2.2.16), (2.2.20) and (2.2.19) can be cast in its log-linearized form as follows

$$\hat{\pi}_t = \delta \mathbb{E}_t \left[\hat{\pi}_{t+1}\right] + \kappa \hat{c}_t \text{ with } \kappa = \frac{\zeta\gamma \left(c^*\right)^{-\gamma}}{\varphi} \tag{2.2.25}$$

$$\hat{c}_t = \mathbb{E}_t \left[\hat{c}_{t+1}\right] - \frac{1}{\gamma} \left(\hat{R}_t - \mathbb{E}_t \left[\hat{\pi}_{t+1}\right] - \mathbb{E}_t \left[\hat{a}_{t+1}\right]\right) \tag{2.2.26}$$

$$\hat{c}_t = \hat{x}_t - \hat{g}_t. \tag{2.2.27}$$

2.2.6 The Central Bank

We assume that the central bank follows the modified Taylor (1993) rule for conducting monetary policy. According to this rule, the bank adjusts the short term nominal interest rate R_t in response to deviations of inflation π_t from the target π^*, and stochastically detrended output $x_t = Y_t/A_t$ from its non stochastic value x^*

$$\ln R_t = \ln R^* + \alpha_{s_t} \left(\ln \pi_t - \ln \pi^*\right) + \beta_{s_t} \left(\ln x_t - \ln x^*\right) + \ln e_t. \tag{2.2.28}$$

Defining $\hat{e}_t = \ln(e_t)$ and $\hat{R}_t$, $\hat{\pi}_t$ and $\hat{x}_t$ as in linearized model above, this interest rate rule can be written as

$$\hat{R}_t = \alpha_{s_t} \hat{\pi}_t + \beta_{s_t} \hat{x}_t + \hat{e}_t. \tag{2.2.29}$$

where $\hat{e}_t$ is assumed to follow a stationary AR(1) process

$$\hat{e}_t = \phi_e \hat{e}_{t-t} + \varepsilon_{e,t}. \tag{2.2.30}$$

with $\varepsilon_{e,t} \sim \mathcal{N}(0, \sigma^2_{e,v_{t,e}})$. That is, the volatility of the monetary policy shock $\varepsilon_{e,t}$ also follows a two-state Markov switching process. Following the notation for the two other shock volatilities, we denote the transition probability matrix for the volatility process of the monetary shock as

$$\mathbf{Q}^e = \begin{bmatrix} q^e_{11} & 1 - q^e_{11} \\ 1 - q^e_{22} & q^e_{22} \end{bmatrix} \tag{2.2.31}$$

where $q^e_{ij} = \Pr[v_{t+1,e} = j | v_{t,e} = i]$.

Notice that in the above short rate equation the target inflation is assumed to be constant over time[2]. As we show below, the virtue of this simplifying assumption is that it allows us to isolate all monetary policy regime changes solely through changes in the reaction coefficients of the inflation and output gaps.

Except for the fact that the monetary policy coefficients α and β are regime dependent, as indicated by the subscript s_t, this is a standard representation of the Taylor rule. The interpretation of regime dependency is that the response coefficients are allowed to change between active and passive (or less active) regimes. We model these regime changes as a change point process which we characterize in the manner of Chib (1998). Specifically, an m change point process is described in terms of a (m+1)-state unidirectional discrete

[2]In contrast, Moreon, Bekaert, and Cho (2010) and Davig and Doh (2009) assume that the target inflation is a stochastic time-varying process.

time Markov process with transition probability matrix $\mathbf{P}$ of the form

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1-p_{11} & 0 & \cdots & 0 \\ 0 & p_{22} & 1-p_{22} & \cdots & 0 \\ 0 & 0 & p_{33} & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & p_{m+1,m+1} \end{bmatrix} \tag{2.2.32}$$

where $p_{jk} = \Pr[s_{t+1} = k | s_t = j]$, $p_{jk} = 1 - p_{jj}$, $k = j+1$ and $p_{m+1,m+1} = 1$ $(j = 1, 2, .., m)$. Under this process, once a policy regime has been vacated, it is never occupied again. In economic terms, this accommodates, for instance, the realistic belief that the pre-Volker regime will never return, an assumption that is also made by Farmer, Waggoner, and Zha (2008).

2.2.7 Summary of the Regime Processes

Recall that there are three structural shocks in this model: the technology shock $\varepsilon_{a,t}$, the fiscal shock $\varepsilon_{g,t}$ and the monetary shock $\varepsilon_{e,t}$. We assume that these shocks are independent of one another. Combining this assumption with the notation for the regime-dependent volatilities introduced earlier, we summarize the shock processes as follows

$$\bar{\mathbf{f}}_t = \begin{bmatrix} \hat{a}_t \\ \hat{g}_t \\ \hat{e}_t \end{bmatrix} = \phi \bar{\mathbf{f}}_{t-1} + \varepsilon_t \tag{2.2.33}$$

where

$$\phi = \begin{bmatrix} \phi_a & 0 & 0 \\ 0 & \phi_g & 0 \\ 0 & 0 & \phi_e \end{bmatrix}, \text{ and } \varepsilon_t = \begin{bmatrix} \varepsilon_{a,t} \\ \varepsilon_{g,t} \\ \varepsilon_{e,t} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_{3\times1}, \Omega_{v_t} = \begin{bmatrix} \sigma^2_{a,v_{t,a}} & & \\ & \sigma^2_{g,v_{t,g}} & \\ & & \sigma^2_{e,v_{t,e}} \end{bmatrix} \right).$$

We further assume that the change point process for the policy regimes s_t is independent of the volatility regimes v_t. For notational convenience, we aggregate the regime indicators comprising of both s_t and v_t into d_t as follows (shown here for the number of policy regimes $m = 2$ and the number of volatility regimes $v = 8$).

d_t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
s_t	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
v_t^a	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
v_t^g	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
v_t^e	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2

This aggregation enables us to denote any possible distinct combination of the policy and volatility regimes with a single notation. For instance, $d_t = 1$ captures the first state for the policy regime as well as for each of the three volatility regimes. Thus, the total number of regimes $\mathbf{d}$ equals $(m + 1) \times v$. The corresponding "aggregated" transition probability matrix can therefore be written as $\mathbf{Z} = \mathbf{Q}^e \otimes \mathbf{Q}^g \otimes \mathbf{Q}^a \otimes \mathbf{P}$.

In section 2.2.10, we show that the recurrence of the volatility regimes, combined with the fact that $v_{t,a}$, $v_{t,g}$, $v_{t,e}$ and s_t are independent, implies that both the model-implied term premium and the expected excess returns are time-varying in each monetary policy regime.

2.2.8 Model Solution and Determinacy Restrictions

For concerns of theoretical tractability as well as econometric convenience, we focus on the (local) behavior of the economy around its deterministic, non-stochastic steady state. Hence, our interest lies in the linearized system of equations (2.2.25)-(2.2.29) and (2.2.33). On substituting (2.2.27) and (2.2.29) into (2.2.26), this system collapses to

$$0 = \delta\mathbb{E}_t\left[\hat{\pi}_{t+1}\right] - \hat{\pi}_t + \kappa\left(\hat{x}_t - \hat{g}_t\right) \tag{2.2.34}$$

$$0 = \mathbb{E}_t\left[\hat{\pi}_{t+1}\right] + \gamma\mathbb{E}_t\left[\hat{x}_{t+1}\right] - \alpha_{s_t}\hat{\pi}_t - (\beta_{s_t} + \gamma)\hat{x}_t + \phi_a\hat{a}_t - \gamma(\phi_g - 1)\hat{g}_t - \hat{e}_t \tag{2.2.35}$$

We now have a simultaneous system of two equations in two key aggregated variables of interest (output deviation from its steady state, $\hat{x}_t$, and, deviation of inflation from its target, $\hat{\pi}_t$) and three unobservable shocks (to technology $\hat{a}_t$, government expenditure $\hat{g}_t$ and monetary policy $\hat{e}_t$).

To analyze the evolution of the two variables of interest we first need to solve this model. The solution process rids the system of the unobservable expectational terms by casting them as a linear function of the underlying shock processes. In the context of regime-switching DSGE models, Davig and Leeper (2007) show how to construct a unique (determinate) and bounded solution. Specifically, their approach relies on the minimal state variable (MSV) representation as follows[3]

$$\underbrace{\begin{bmatrix} \hat{\pi}_{it} \\ \hat{x}_{it} \end{bmatrix}}_{\hat{\mathbf{m}}_{it}} = \underbrace{\begin{bmatrix} h_\pi^a(s_t = i) & h_\pi^g(s_t = i) & h_\pi^e(s_t = i) \\ h_x^a(s_t = i) & h_x^g(s_t = i) & h_x^e(s_t = i) \end{bmatrix}}_{\bar{\mathbf{H}}_{s_t=i}} \bar{\mathbf{f}}_t \tag{2.2.36}$$

[3]Farmer, Zha, and Waggoner (2009) derive a class of non-MSV solutions to the quasi-linear system (2.2.25)-(2.2.29). In their approach, determinacy conditions for the quasi-linear model are not feasible unlike in Davig and Leeper (2007)'s method.

where $\hat{\pi}_{it}$ and $\hat{x}_{it}$ denote the state-contingent $(s_t = i)$ values of inflation gap and output gap, respectively.

On inserting this linear solution into the system of equations (2.2.34)-(2.2.35), the conditional expectation of the one-period ahead inflation gap and output gap are

$$\mathbb{E}_t \left[\begin{pmatrix} \hat{\pi}_{t+1} & \hat{x}_{t+1} \end{pmatrix}' | s_t = i \right] = \mathbb{E}_t \left[\bar{\mathbf{H}}_{s_{t+1}} \bar{\mathbf{f}}_{t+1} | s_t = i \right] \tag{2.2.37}$$

$$= p_{i1} \bar{\mathbf{H}}_{s_{t+1}=1} \phi \bar{\mathbf{f}}_t + p_{i2} \bar{\mathbf{H}}_{s_{t+1}=2} \phi \bar{\mathbf{f}}_t$$

Equivalently, on letting $h^j_{\pi,i} \equiv h^j_\pi(s_t = i)$ and $h^j_{x,i} \equiv h^j_x(s_t = i)$, $(j = a,\ g,\ e)$, $\mathbb{E}_t \left[\hat{\pi}_{t+1} | s_t = i\right]$ can be expressed as

$$p_{i1} \left[h^a_{\pi,1} \phi_a \hat{a}_t + h^g_{\pi,1} \phi_g \hat{g}_t + h^e_{\pi,1} \phi_e \hat{e}_t \right] + p_{i2} \left[h^a_{\pi,2} \phi_a \hat{a}_t + h^g_{\pi,2} \phi_g \hat{g}_t + h^e_{\pi,2} \phi_e \hat{e}_t \right] \tag{2.2.38}$$

and $\mathbb{E}_t \left[\hat{x}_{t+1} | s_t = i\right]$ as

$$p_{i1} \left[h^a_{x,1} \phi_a \hat{a}_t + h^g_{x,1} \phi_g \hat{g}_t + h^e_{x,1} \phi_e \hat{e}_t \right] + p_{i2} \left[h^a_{x,2} \phi_a \hat{a}_t + h^g_{x,2} \phi_g \hat{g}_t + h^e_{x,2} \phi_e \hat{e}_t \right] \tag{2.2.39}$$

Next, to compute the regime-dependent solutions $\bar{\mathbf{H}}_{s_t}$, we rely on the method of undetermined coefficients, setting the coefficients of $\hat{a}_t$, $\hat{g}_t$ and $\hat{e}_t$ equal to zero and solving for the resulting solution in terms of the coefficients in $\bar{\mathbf{H}}_{\mathbf{s_t}}$. Further computational details of the solution are in C.

Note that because we use a first-order approximation of the equilibrium conditions of households and firms, the solution coefficients $\bar{\mathbf{H}}_{s_t}$ depend only on the monetary policy regime s_t and not the volatility regimes v_t. In addition, recall that $\ln \pi_t = \hat{\pi}_t + \ln \pi^*$ and $\ln (Y_t/A_t) = \hat{x}_t + \ln x^*$. Hence, the solution for the DSGE model in equation (2.2.36) can be rewritten as

$$\underbrace{\begin{bmatrix} \ln \pi_t \\ \ln Y_t \end{bmatrix}}_{\mathbf{m}_t} = \underbrace{\begin{bmatrix} \ln \pi^* \\ \ln x^* \end{bmatrix}}_{\mathbf{J}} + \underbrace{\begin{bmatrix} h_\pi^a(d_t = i) & h_\pi^g(d_t = i) & h_\pi^e(d_t = i) & 0 \\ h_x^a(d_t = i) & h_x^g(d_t = i) & h_x^e(d_t = i) & 1 \end{bmatrix}}_{\mathbf{H}_{d_t=i}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} \tag{2.2.40}$$

This representation of the solution is needed in the estimation of the model as we show in section 2.3.1.

It is important to note that that the coefficients in $\bar{\mathbf{H}}_{s_t}$ are highly non-linear, complicated mappings of the deep parameters. This mapping can only be calculated numerically given values of the parameters. Because of this complicated nonlinearity, the likelihood function of the model (which we present below) tends to be highly irregular with multiple local maxima, abrupt discontinuities and flat regions. This aspect of the likelihood function is well acknowledged in the DSGE literature and is an important reason why (over the last decade) Bayesian estimation techniques aided by MCMC methods have emerged as the primary tools for estimating DSGE models.

2.2.9 The Bond Prices

The first order conditions for the short and long term bonds B_t^τ $(1 \leq \tau \leq \tau^*)$, which are absent in standard DSGE models without long term bonds, can be shown to have the form

$$P_t^\tau = \mathbb{E}_t \left[M_{t,t+\tau} \right] \tag{2.2.41}$$

where

$$M_{t,t+\tau} = \delta \left(\frac{c_{t+\tau}}{c_t} \right)^{-\gamma} \frac{1}{a_{t+\tau}} \frac{1}{\pi_{t+\tau}} \tag{2.2.42}$$

is the intertemporal marginal rate of substitution between time t and $t+\tau$. These first order conditions provides the demand function for long term bonds. Assuming that the supply of these bonds is perfectly elastic, and using the law of iterated expectation, one has the standard asset-pricing conclusion that

$$\begin{aligned} P_t^{\tau} &= \mathbb{E}_t\left[M_{t,t+1} \times M_{t+1,t+\tau}\right] \\ &= \mathbb{E}_t\left[M_{t,t+1} \times \mathbb{E}_{t+1}\left[M_{t+1,t+\tau}\right]\right] \\ &= \mathbb{E}_t\left[M_{t,t+1} \times P_{t+1}^{\tau-1}\right] \end{aligned} \tag{2.2.43}$$

This equation implies that the equilibrium bond prices at time t, denoted by $P_{d_t,t}^{(\tau)}$, satisfy the following no-arbitrage condition

$$P_{d_t,t}^{(\tau)} = \mathbb{E}\left[M_{t,t+1} P_{d_{t+1},t+1}^{(\tau-1)} | \bar{\mathbf{f}}_t, d_t\right] \tag{2.2.44}$$

and are a function of the model-determined pricing kernel which itself is a function of d_t and the exogenous shocks.

To calculate the form of these prices, we express the nominal pricing kernel in log-linearized form as

$$\ln M_{t,t+1} = m_{t,t+1} \approx c_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}} \bar{\mathbf{f}}_t + \mathbf{L}_{d_{t+1}} \varepsilon_{t+1} \tag{2.2.45}$$

where

$$c_{d_{t+1}} = -\ln R^* - \frac{1}{2} \mathbf{L}_{d_{t+1}} \Omega_{d_{t+1}} \mathbf{L}'_{d_{t+1}} \tag{2.2.46}$$

$$\boldsymbol{\lambda}_{d_t,d_{t+1}} = -\begin{pmatrix} 1 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_{t+1}} \phi + \begin{pmatrix} 0 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_t} + \begin{pmatrix} -1 & \gamma & 0 \end{pmatrix} \phi - \begin{pmatrix} 0 & \gamma & 0 \end{pmatrix} \tag{2.2.47}$$

$$\mathbf{L}_{d_{t+1}} = -\begin{pmatrix} 1 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_{t+1}} + \begin{pmatrix} -1 & \gamma & 0 \end{pmatrix} \tag{2.2.48}$$

Following Ang et al. (2008), we assume that the one period bond is risk-free by augmenting the Jensen's inequality term to equation (2.2.46). This assumption is necessary to generate a positive average term premium in our formulation. Also note that the market price of risk, which is associated with the structural shocks ε_{t+1}, is given by the elements in $\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}^{1/2}$.

Let $p_{d_t,t}^{(\tau)} \equiv \ln P_{d_t,t}^{(\tau)}$ denote the log price of a τ-period maturity bond at time t in regime d_t and suppose that

$$-p_{d_t,t}^{(\tau)} = a_{d_t}(\tau) + \mathbf{b}_{d_t}(\tau)'\bar{\mathbf{f}}_t. \tag{2.2.49}$$

Under this guess and the form of the pricing kernel above we can use the method of undetermined coefficients to derive the following recursive expressions for $i \in \{1, 2, .., \mathbf{d}\}$

$$a_i(\tau) = \ln R^* + \sum_{j=1}^{\mathbf{d}} p_{ij} \left(a_j(\tau-1) + \mathbf{L}_j\Omega_j\mathbf{b}_j(\tau-1)' - \frac{1}{2}\mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) \right) \tag{2.2.50}$$

$$\mathbf{b}_i(\tau)' = \sum_{j=1}^{\mathbf{d}} p_{ij} \left(\mathbf{b}_j(\tau-\mathbf{1})'\phi - \boldsymbol{\lambda}_{i,j} \right). \tag{2.2.51}$$

Further details of this derivation are provided in D. These recursions are initialized by the no-arbitrage condition at $\tau = 0$

$$a_i(0) = \mathbf{b}_i(0) = 0 \text{ for all } i \tag{2.2.52}$$

Then, the continuously compounded yield to maturity $r_{d_t,t}^{(\tau)}$ for the zero-coupon nominal bond is given by

$$r_{d_t,t}^{(\tau)} = \frac{-p_{d_t,t}^{\tau}}{\tau} = \bar{a}_{d_t}(\tau) + \bar{\mathbf{b}}_{d_t}(\tau)'\bar{\mathbf{f}}_t \tag{2.2.53}$$

$$\text{with } \bar{a}_{d_t}(\tau) = \frac{a_{d_t}(\tau)}{\tau} \text{ and } \bar{\mathbf{b}}_{d_t}(\tau) = \frac{\mathbf{b}_{d_t}(\tau)}{\tau}.$$

It is useful to note that the factor loadings $\bar{\mathbf{b}}_{d_t}(\tau)$ are independent of the volatility regimes because $\boldsymbol{\lambda}_{i,j}$ is determined by the parameters in the linearized Euler equation (2.2.45).

Importantly, the equilibrium short rate obtained from these recursions when $\tau = 1$ is exactly the same as the value of the short rate from the Taylor rule at equilibrium (obtained by substituting the equilibrium values of output and inflation into the Taylor rule). This agreement is a consequence of the fact that bond pricing as exemplified here comes from the dynamic general equilibrium solution of the model.

2.2.10 Measures of Long-Term Bond Risk

We focus on three different measures of riskiness of long-term bonds in each regime: the term premium, the expected excess return on the long-term bond and the slope of the yield curve. We now discuss the characteristics of each of these measures.

The term spread is simply the difference between the long-term bond yield and the short rate. As is well-known, it can be rewritten as the sum of two components

$$r_{d_t,t}^{(\tau)} - r_{d_t,t}^{(1)} = \underbrace{\left[\frac{1}{\tau}\sum_{l=0}^{\tau-1}\mathbb{E}_t\left[r_{d_{t+l},t+l}^{(1)}\right] - r_{d_t,t}^{(1)}\right]}_{\text{EH}} + \underbrace{\frac{1}{\tau}\sum_{i=1}^{\tau-1}\text{exr}_{d_t,t}^{(\tau+1-i)}}_{\text{Term Premium}}, \tag{2.2.54}$$

where $\text{exr}_{d_t,t}^{(\tau)}$ denotes the one-period expected excess return to holding the τ-period bond. The first component on the right is the expectation hypothesis. Under risk-neutral pricing, after adjusting for risk, agents are indifferent between holding a long term bond and a one period risk-free bond. The risk adjustment is the term premium, captured by the second term on the right.

Two important points emerge from equation (2.2.54). First, the term spread depends on the expected excess returns as well as the expected average future short rate. Second, the term premium reflects the expected excess return to all bonds of maturities less than τ-periods, not just expected excess return to the τ-period bond.

The one-period expected excess return of the τ-period bond at time t is then defined as

$$\begin{aligned}\text{exr}^{(\tau)}_{d_t,t} &= \left[\mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] - p^{(\tau)}_{d_t,t}\right] - (-p^{(1)}_{d_t,t}) \\ &= \mathbb{E}_t\left[-(\tau-1)\,r^{(\tau-1)}_{d_{t+1},t+1} + \tau r^{(\tau)}_{d_t,t}\right] - r^{(1)}_{d_t,t}\end{aligned} \tag{2.2.55}$$

The first term on the right side of (2.2.55) is the expected one-period return to holding the bond and the second term is the one-period risk-free rate. Importantly, $\text{exr}^{(\tau)}_{d_t,t}$ can be expressed as a sum of the factor risk component $\text{FR}^{(\tau)}_{d_t=i}$ and the regime-shift risk component $\text{RS}^{(\tau)}_{d_t=i,t}$

$$\text{exr}^{(\tau)}_{d_t=i,t} = \text{FR}^{(\tau)}_{d_t=i} + \text{RS}^{(\tau)}_{d_t=i,t} \tag{2.2.56}$$

where

$$\text{FR}^{(\tau)}_{d_t=i} = \sum_{j=1}^{\mathbf{d}} p_{ij}\mathbf{L}_j\Omega_j\mathbf{b}_j(\tau-1) - \frac{1}{2}\sum_{j=1}^{\mathbf{d}} p_{ij}\mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) \tag{2.2.57}$$

$$\begin{aligned}\text{RS}^{(\tau)}_{d_t=i,t} = {} & \left[\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right]\left[\sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}\right] - \sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}K_{j,t} \\ & - \frac{1}{2}\sum_{j=1}^{\mathbf{d}} p_{ij}K^2_{j,t} + \frac{1}{2}\left(\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right)^2\end{aligned} \tag{2.2.58}$$

and

$$W_{d_t,d_{t+1},t} = c_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}} \bar{\mathbf{f}}_t \tag{2.2.59}$$
$$K_{d_{t+1},t} = -a_{d_{t+1}} - \mathbf{b}_{d_{t+1}}(\tau-1)' \phi \bar{\mathbf{f}}_t$$

Similarly, it is straightforward to decompose the term premium, denoted by $\mathrm{TP}^{(\tau)}_{d_t=i,t}$, in equation (2.2.54) as the sum of two averages.

The proof of these results is given in E. Notice that the terms in the factor risk component $\mathrm{FR}^{(\tau)}_{d_t=i}$ are all associated with the structural shocks in the following period. Not surprisingly, the compensation demanded for holding long term bonds depends largely on the size of the factor shocks $\Omega_j^{1/2}\mathbf{b}_j(\tau-1)$ and the price of the risks $\mathbf{L}_j\Omega_j^{1/2}$. This market price of the risks is maturity-independent and determines how much one unit of risk translates into an expected excess return. Meanwhile, the regime-shift risk component $\mathrm{RS}^{(\tau)}_{d_t=i,t}$ will be absent under either a single regime model or a regime switching model with market price of regime shift risk equal to zero as pointed out by Dai et al. (2007). Finally, it is interesting that $\mathrm{FR}^{(\tau)}_{d_t=i}$ is a regime-specific constant, whereas $\mathrm{RS}^{(\tau)}_{d_t=i,t}$ depends on the current values of the time-varying factors. Consequently, the expected excess return is time varying and so is the term premium[4]. Moreover, our regime-dependent factor loadings, generated by the monetary policy regime shifts, allow for the term premium to vary independently of factor volatility. This additional flexibility helps improve the forecast accuracy of future yields, as pointed out in Duffee (2002).

[4]An alternative way of achieving a time-varying term premium is to work with a second-order or third-order approximation of the optimality conditions (Doh (2009) and Bansal and Yaron (2004)). However, a suitable solution method for such non-linear models under a multi-regime specification currently does not exist.

2.3 Estimation methodology

2.3.1 State Space Formulation

We begin by recalling the solution to the DSGE model in equation (2.2.40)

$$\underbrace{\begin{bmatrix} \ln \pi_t \\ \ln Y_t \end{bmatrix}}_{\mathbf{m}_t} = \underbrace{\begin{bmatrix} \ln \pi^* \\ \ln x^* \end{bmatrix}}_{\mathbf{J}} + \underbrace{\begin{bmatrix} h_\pi^a(d_t = i) & h_\pi^g(d_t = i) & h_\pi^e(d_t = i) & 0 \\ h_x^a(d_t = i) & h_x^g(d_t = i) & h_x^e(d_t = i) & 1 \end{bmatrix}}_{\mathbf{H}_{d_t=i}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} \tag{2.3.1}$$

Note that the short rate $r_t^{(1)}$, which is set by the central bank following the Taylor (1993) rule, incorporates the monetary policy shock. Thus, as in the estimation of standard DSGE models, we assume that the final outcomes $(\mathbf{m}_t, \hat{R}_t)$ are generated without additional (measurement) errors. As we show in F, the benefit of this assumption is that, given the regime process $\mathbf{D}_n$ and the initial value of the technology shock $\ln A_0$, the shock process $\bar{\mathbf{f}}_t$ can be solved entirely in terms of the observable quantities $\ln (P_t/P_{t-1})$, $\ln Y_t$ and $\hat{R}_t$, where $\ln A_0$ is treated as an additional parameter to be estimated. This, in turn, substantially simplifies the calculation of the likelihood function conditioned on the regimes.

We implement our model on a data set that comprises 5 yields of US T-bills measured on a quarterly basis. We denote these quarterly maturities of interest as

$$\{\tau_1, \tau_2, \tau_3, \tau_4, \tau_5\} = \{1, 2, 4, 8, 20\}$$

and let

$$\mathbf{R}_t = \begin{pmatrix} r_t^{(\tau_1)} & r_t^{(\tau_2)} & r_t^{(\tau_3)} & r_t^{(\tau_4)} & r_t^{(\tau_5)} \end{pmatrix}'$$

where $r_t^{(\tau_i)} = r_{i,t}$. We assume that all bonds with maturity greater than 1 period are priced with errors - that is, the short rate is treated as a basis yield. Let $\bar{\mathbf{a}}_{d_t} = (\bar{a}_{d_t}(\tau_1), \bar{a}_{d_t}(\tau_1), .., \bar{a}_{d_t}(\tau_1))'$ and $\bar{\mathbf{b}}_{d_t} = \left(\bar{\mathbf{b}}_{d_t}(\tau_1), \bar{\mathbf{b}}_{d_t}(\tau_2), .., \bar{\mathbf{b}}_{d_t}(\tau_5)\right)'$. Then the observable quantities $\mathbf{m}_t$ and $\mathbf{R}_t$ are stacked to obtain the measurement equation

$$\underbrace{\begin{bmatrix} \mathbf{m}_t \\ \mathbf{R}_t \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} \mathbf{J} \\ \bar{\mathbf{a}}_{d_t} \end{bmatrix}}_{\mathbf{a}_{d_t}} + \underbrace{\begin{bmatrix} \mathbf{H}_{d_t} & \\ \bar{\mathbf{b}}_{d_t} & \mathbf{0}_{5\times1} \end{bmatrix}}_{\mathbf{b}_{d_t}} \mathbf{f}_t + \underbrace{\begin{bmatrix} \mathbf{0}_{3\times4} \\ \mathbf{I}_4 \end{bmatrix}}_{\mathbf{T}_y} \mathbf{e}_t \tag{2.3.2}$$

where $\mathbf{e}_t \sim \mathcal{N}_4(\mathbf{0}, \mathbf{\Sigma})$; $\mathbf{\Sigma} = \mathsf{diag}(\sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2)$. We complete the state space formulation by combining equation (2.2.33) with the technology shock process $\ln A_t = \ln a^* + \ln A_{t-1} + \hat{a}_t$ and write the transition equation as

$$\underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} = \underbrace{\begin{bmatrix} \mathbf{0}_{3\times1} \\ \ln a^* \end{bmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{bmatrix} \boldsymbol{\phi}_{3\times3} & & & \mathbf{0}_{3\times1} \\ \phi_a & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_{t-1} \\ \ln A_{t-1} \end{bmatrix}}_{\mathbf{f}_{t-1}} + \underbrace{\begin{bmatrix} \mathbf{I}_3 \\ 1 \quad 0 \quad 0 \end{bmatrix}}_{\mathbf{T}_f} \varepsilon_t \tag{2.3.3}$$

with $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{3\times1}, \mathbf{\Omega}_{d_t})$. For notational convenience, we let $\boldsymbol{\theta}$ denote the free parameters in $\mathbf{a}_{d_t}$, $\mathbf{b}_{d_t}$, $\mathbf{\Sigma}$, $\boldsymbol{\mu}$, $\mathbf{G}$ and $\mathbf{\Omega}_{d_t}$.

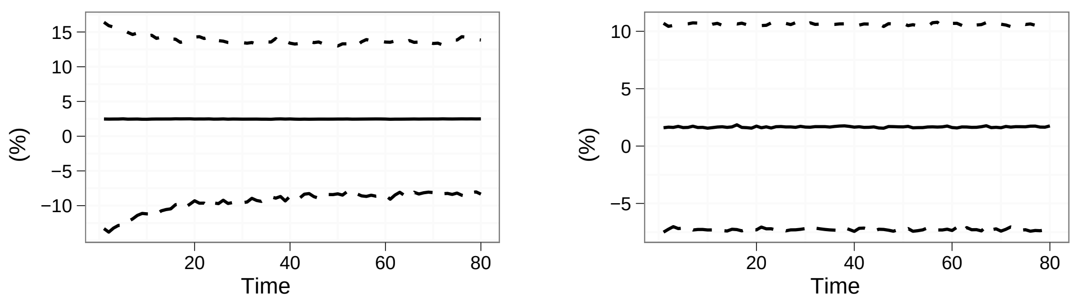
2.3.2 Prior Distribution

We formulate the prior on the parameters to reflect the belief that (under the prior) the average term premium is positive (Chib and Ergashev (2009)). This prior is, of course, restricted to the subset of the parameter space that implies a unique determinate solution to the model. Finally, various blocks of parameters are assumed to be a priori independent. Table 2.1 summarizes our prior.

Parameter	density	mean	S.D.
δ	beta	0.9992	0.0006
ϕ_a	beta	0.3688	0.1189
ϕ_g	beta	0.8472	0.1092
ϕ_e	beta	0.6123	0.1293
p_{11}	beta	0.9745	0.0221
q_{11}^a	beta	0.8995	0.0401
q_{22}^a	beta	0.8997	0.0401
q_{11}^g	beta	0.8997	0.0401
q_{22}^g	beta	0.8997	0.0401
q_{11}^e	beta	0.8997	0.0401
q_{22}^e	beta	0.8997	0.0401
$400 \times \ln R^*$	normal	4.4426	0.3141
κ	gamma	0.4985	0.3036
α_1	normal	1.5154	0.2965
α_2	normal	1.9972	0.3161
β_1	normal	0.9968	0.3139
β_2	normal	1.0067	0.3117
γ	gamma	39.952	10.011
$400 \times \ln a^*$	normal	1.6575	0.3147
$\ln x^*$	gamma	0.9988	0.0978
$\ln A_0$	normal	2.3115	0.1009
$2.0 \times 10^4 \times \sigma_{a,1}^2$	inverse gamma	0.9539	0.1895
$2.0 \times 10^5 \times \sigma_{g,1}^2$	inverse gamma	0.9596	0.1937
$3.0 \times 10^4 \times \sigma_{e,1}^2$	inverse gamma	0.9603	0.1951
$1.0 \times 10^4 \times \sigma_{a,2}^2$	inverse gamma	0.9635	0.1941
$1.0 \times 10^5 \times \sigma_{g,2}^2$	inverse gamma	0.9635	0.1956
$2.5 \times 10^3 \times \sigma_{e,2}^2$	inverse gamma	0.9613	0.1943
$1.4 \times 10^7 \times \sigma_2^2$	inverse gamma	0.9623	0.1927
$3.0 \times 10^6 \times \sigma_3^2$	inverse gamma	0.9620	0.1948
$1.2 \times 10^6 \times \sigma_4^2$	inverse gamma	0.9605	0.1942
$6.0 \times 10^5 \times \sigma_5^2$	inverse gamma	0.9611	0.1979

Table 2.1: Prior distribution for the 16-regime model parameters

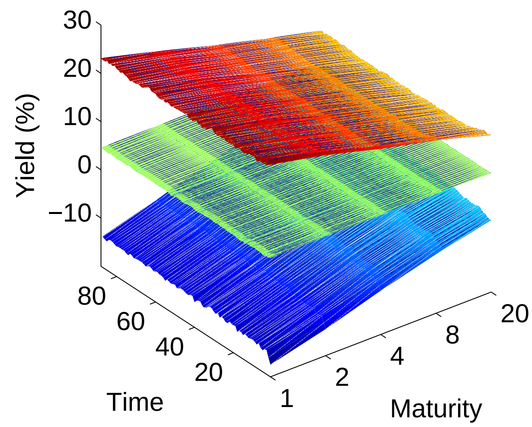
Under this prior, the annual short interest rate is centered at 4.4% with a standard deviation of 0.32%. The steady state technology growth ranges from 1.13% to 2.17%. For the variance of the structural shocks and the risk aversion parameters, the respective marginal prior distributions are set to generate an average positive term premium. The marginal prior distributions of the other parameters are set to be consistent with the existing empirical literature on the term structure and new Keynesian DSGE models. For example, the prior distribution of the slope parameter κ in the Phillips curve is from Lubik and Schorfheide (2004)and the transition probabilities are consistent with Chib and Kang (2010). It is important to note that the values of the hyperparameters in these marginal distributions are chosen to allow the parameters to vary considerably in the domain supported by the determinacy condition. Furthermore, in this change point setup for the policy regime, it is not necessary to impose any restrictions on the relative magnitudes of $\beta_{s_t=1}$, $\beta_{s_t=2}$, $\alpha_{s_t=1}$ and $\alpha_{s_t=2}$. In contrast, we normalize the labels for the volatility regimes by restricting that all diagonal elements in $\Omega_{\mathbf{d}}$ are greater than those in $\Omega_{\mathbf{1}}$. Finally, we note that our prior is quite symmetric across regimes in order to avoid the identification of the regimes through the prior information.



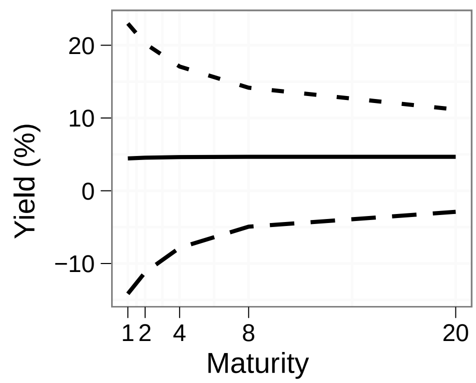
(a) inflation (b) output growth

Figure 2.1: The prior-implied inflation and output growth dynamics

To understand what the prior distribution implies for the outcomes, we sample the parameters 20,000 times from the prior, and then for each drawing of the parameters, we simulate the shocks, macroeconomic variables and yields according to the structural model. The sampled sequences for each macroeconomic variable in annualized percents are shown in Figure 2.1. As one can see from those figures, this prior implies a deviation of roughly 5% for output growth and 7% for inflation. Similarly, the implied term structure



(a)



(b)

Figure 2.2: The prior-implied term structure dynamics

in annualized percents for each time period is reproduced in Figure 2.2. As one can see, the implied average term structure is gently upward sloping in each regime with considerable a priori variation.

2.3.3 Posterior Distribution and MCMC Sampling

We now have the necessary ingredients to calculate the posterior distribution of the parameters. Let $\mathbf{D}_n = \{d_t\}_{t=0,1,..,n}$ denote the sequence of the unobserved regime indicators, $\mathbf{F}_n = \{\mathbf{f}_t\}_{t=0,1,..,n}$ the sequence of the factors and $\mathbf{y} = \{\mathbf{y}_t\}_{t=0,1,..,n}$ the full set of observables (date set). Then, the posterior distribution that we would like to analyze is given by

$$\pi(\boldsymbol{\theta}, \mathbf{F}_n, \mathbf{D}_n|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{F_n}, \mathbf{D}_n)p(\mathbf{F}_n, \mathbf{D}_n|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{2.3.4}$$

where $f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{F}_n, \mathbf{D}_n)$ is the distribution of the data given the regime indicators and the parameters, $p(\mathbf{F}_n, \mathbf{D}_n|\boldsymbol{\theta})$ is the density of the latent factors and the regime-indicators given the parameters, and $\pi(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$. Note that by conditioning on

$\mathbf{D}_n$ we avoid the calculation of the likelihood function $f(\mathbf{y}|\boldsymbol{\theta})$ whose computation is more involved.

We summarize this complex posterior distribution by MCMC simulation methods. The basic idea behind the MCMC approach is to produce correlated (Markov distributed) drawings from the posterior distribution whose invariant distribution is the target density (Chib and Greenberg (1995)). Practically, the sampled draws after a suitably specified burn-in phase are taken as samples from the posterior density. We construct our 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 $\pi(\boldsymbol{\theta}, \mathbf{F}_n, \mathbf{D}_n|\mathbf{y})$. In particular, after initializing the model parameters $\boldsymbol{\theta}$ and the regimes $\mathbf{D}_n$, we go through an iterative sequence of steps in each MCMC cycle. First, we sample $\boldsymbol{\theta}$ from the posterior distribution that is proportional to

$$f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}_n)\pi(\boldsymbol{\theta}) \tag{2.3.5}$$

where $f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}_n)$ is obtained from the standard Kalman filtering recursions given the regime indicators $\mathbf{D}_n$. The sampling of $\boldsymbol{\theta}$ from the latter density is done by the TaRB-MH method following Chib and Ramamurthy (2010). The use of this MCMC method is essential to improve the mixing of the draws when there is no natural way of grouping the parameters. In the next step we solve for $\mathbf{F_n}$ in terms of the observable macro quantities and the short yield. Finally, we sample $\mathbf{D}_n$ conditioned on $\mathbf{F_n}$ and $\boldsymbol{\theta}$ in one block by the algorithm of Chib (1996). These steps of the MCMC algorithm are summarized below. A more detailed description can be found in F.

Algorithm: **MCMC sampling**

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

Step 2 Sample $\boldsymbol{\theta}$ conditioned on $(\mathbf{y}, \mathbf{D}_n)$

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

Step 4 Sample $\mathbf{D}_n$ conditioned on $(\mathbf{y}, \boldsymbol{\theta}, \mathbf{F}_n)$

Step 5 Repeat Steps 2-4, discard the draws from the first n_0 iterations and save the subsequent n_1 draws.

2.3.4 Model Comparison

From the perspective of the data, we are interested in knowing whether a multi-regime model improves on a single regime model. Furthermore, we are also interested in learning which of these multi-regime specifications best describes the data. To address these questions, we compare the following models: a single regime model ($\mathcal{M}_1$), a model with one regime change in monetary policy but no regime shifts in the shock volatilities (2 policy regimes, $\mathcal{M}_2$), a model with one regime change in monetary policy together with simultaneous regime shifts in all three volatilities (2 policy regimes and 2 volatility regimes, $\mathcal{M}_4$), a model with one regime change in monetary policy together with independent regime shifts in each of the three volatilities (2 policy regimes and 8 volatility regimes, $\mathcal{M}_{16}$), and, finally, a model with two regime changes in monetary policy together with independent regime shifts in each of the three volatilities (3 policy regimes and 8 volatility regimes, $\mathcal{M}_{24}$).

$\mathcal{M}_\mathbf{d}$	# of monetary policy regimes$(m+1)$	# of volatility regimes$(\mathbf{v})$
$\mathcal{M}_1$	1	1
$\mathcal{M}_2$	2	1
$\mathcal{M}_4$	2	2
$\mathcal{M}_{16}$	2	8
$\mathcal{M}_{24}$	3	8

Within the Bayesian context, these models are compared in terms of the marginal likelihoods $m(\mathbf{y}|\mathcal{M}_{\mathbf{d}})$ and their ratios (Bayes factors). Following Chib and Jeliazkov (2001) an estimate of the log marginal likelihood can be calculated from the following fundamental identity

$$\ln \hat{m}(\mathbf{y}|\mathcal{M}_{\mathbf{d}}) = \ln f(\mathbf{y}|\boldsymbol{\theta}^*, \mathcal{M}_{\mathbf{d}}) + \ln \pi\,(\boldsymbol{\theta}^*, \mathcal{M}_{\mathbf{d}}) - \ln \hat{\pi}(\boldsymbol{\theta}^*|\mathbf{y},\mathcal{M}_{\mathbf{d}}) \tag{2.3.6}$$

where $\mathbf{d}$=1, 2, 4, 16, and 24, and $\boldsymbol{\theta}^*$ is a high density point in the support of the parameter space. Notice that the first term on the right hand side of this expression is the likelihood ordinate. The second term is the prior ordinate. Both of these are readily available. The third term, the posterior ordinate $\pi(\boldsymbol{\theta}^*|\mathbf{y},\mathcal{M}_{\mathbf{d}})$, is estimated from a marginal-conditional decomposition (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. For details we refer the interested reader to these papers.

2.3.5 Prediction

We are also interested in examining the forecasting performance of the proposed model in relation to other models. Forecasts are generated by sampling the Bayesian predictive density (the density of the future quantities conditioned on the sample data, marginalized over the parameters and other unknowns). This sampling is done by the method of composition. For each draw of the parameters from the posterior distribution, we draw the regimes and the structural shock processes. Then, given the factors and the parameters, we sample the yields and the macroeconomic variables. The resulting sample can be shown to be from the predictive density.

Algorithm: **Sampling the predictive density of the macroeconomic**

variables and yields

Step 1 For $j = 1, 2, .., n_1$:

(a) $t = 1, 2, .., T$:

(i) Compute $\mathbf{Z}^{(j)}$ and draw d_{n+t} given d_{n+t-1} and $\mathbf{Z}^{(j)}$

(ii) Compute $\boldsymbol{\mu}^{(j)}$, $\mathbf{G}^{(j)}$, $\Omega^{(j)}_{d_{n+t}}$, $\Sigma^{(j)}$, $\mathbf{a}^{(j)}_{d_{n+t}}$ and $\mathbf{b}^{(j)}_{d_{n+t}}$

(iii) Compute $\mathbf{f}^{(j)}_{n+t} = \boldsymbol{\mu}^{(j)} + \mathbf{G}^{(j)}\mathbf{f}^{(j)}_{n+t-1} + \mathbf{T}_f \boldsymbol{\varepsilon}^{(j)}_{n+t}$ where $\varepsilon^{(j)}_{n+t} \sim \mathcal{N}_3\left(\mathbf{0}, \Omega^{(j)}_{d_{n+t}}\right)$

(iv) Compute $\mathbf{y}^{(j)}_{n+t} = \mathbf{a}^{(j)}_{d_{n+t}} + \mathbf{b}^{(j)}_{d_{n+t}}\mathbf{f}^{(j)}_{n+t} + \mathbf{T}_y \mathbf{e}^{(j)}_{n+t}$ where $\mathbf{e}^{(j)}_{n+t} \sim \mathcal{N}_4\left(\mathbf{0}, \Sigma^{(j)}\right)$

(b) Set $\mathbf{y}^{(j)}_f = \left\{\mathbf{y}^{(j)}_{n+1}, \mathbf{y}^{(j)}_{n+2}, .., \mathbf{y}^{(j)}_{n+T}\right\}$

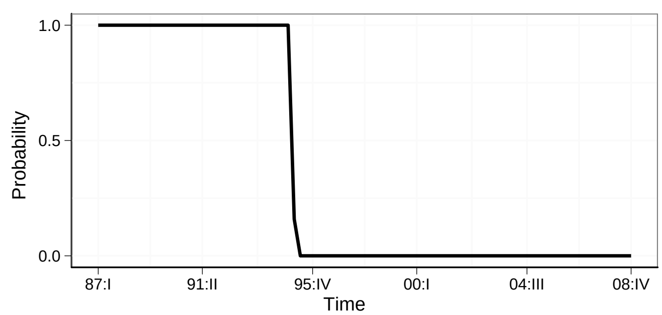
Step 2 Return $\mathbf{y}_f = \left\{\mathbf{y}^{(1)}_f, \mathbf{y}^{(2)}_f, .., \mathbf{y}^{(n_1)}_f\right\}$

2.4 Results

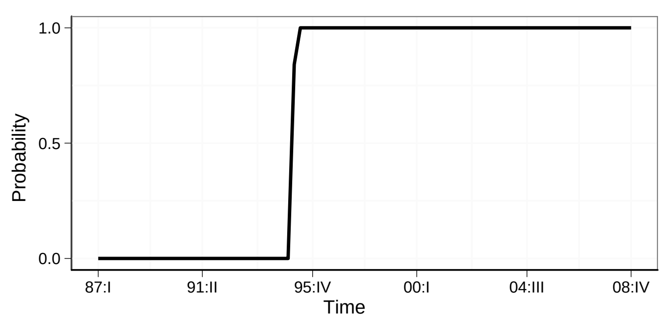
Our empirical results are based on the collection of historical yields of treasury bills with maturities 1, 2, 4, 8 and 20 quarters, GDP per capita and inflation for the sample period 1986:Q4 to 2008:Q4. This data is available online from the Board of Governors of the Federal Reserve System (Gurkaynak et al. (2007)). From the DSGE model perspective, the relevance of this sample period is that it is known for its relative stability compared to the major oil price shocks during the 1970s, the monetary policy experiment and the Volcker disinflation period in the early 1980s.

2.4.1 Change-Point and Structural Shocks

Table 2.2 confirms the existence of one time break in the policy in 1995:Q2 based on the marginal likelihoods. As can be seen in this table, the 4-regime model is most supported by the data. Notice that even the change point model without regime switching variance ($\mathcal{M}_2$) gives the same estimate of the breakpoint. From this we infer that the U.S. monetary policy has been in an absorbing regime since 1995:Q2. Although, it is difficult to pinpoint the specific reason behind the policy change, the timing of this change is consistent with the finding of Chib and Kang (2010) who fit a change point model under the partial equilibrium approach. Figure 2.3 shows the persistence of the policy regimes. In contrast, figures 2.4 and 2.5 reveal that the volatility regimes are far less persistent than the policy regimes. Finally, Figure 2.7 plots the estimated exogenous shock processes $\hat{a}_t$, $\hat{g}_t$ and $\hat{e}_t$.



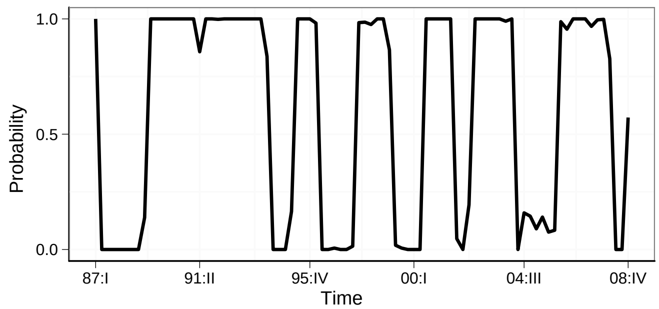
(a) Less active policy regime ($s_t = 1$)



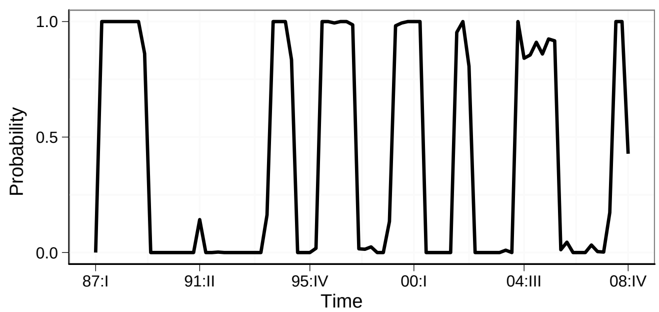
(b) More active policy regime ($s_t = 2$)

Figure 2.3: Model $\mathcal{M}_{16}$: The posterior probability of monetary policy regimes

The coincidence of the technology shock process $\hat{a}_t$ and business cycles is quite striking in this figure.



(a) Low technology volatility regime ($v_t^a = 1$)



(b) High technology volatility regime ($v_t^a = 2$)

Figure 2.4: Model $\mathcal{M}_{16}$: The posterior probability of technology volatility regimes

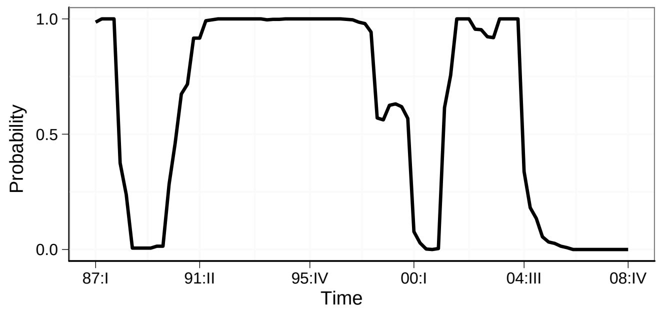
2.4.2 Model Parameters

We next discuss the posterior estimates of the parameters. Table 2.3 summarizes the posterior distribution of the parameters based on 20,000 of the MCMC algorithm beyond a burn-in of 5,000. We measure the efficiency of the MCMC sampling in terms of the acceptance rate in the M-H step and the inefficiency factors[5] (Chib (2001)). These values

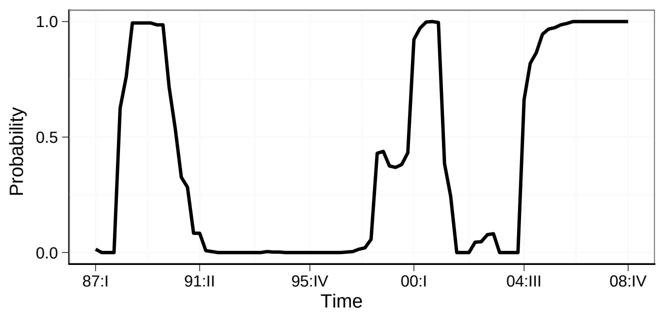
[5]The inefficiency factors approximate the ratio of the numerical variance of the estimate from the MCMC chain relative to that from hypothetical iid draws. For a given sequence of draws the inefficiency factor is computed as

$$1 + 2\sum_{l=1}^{L} \rho_k(l)$$

where $\rho_k(l)$ is the autocorrelation at lag l for the kth sequence, and L is the value at which the autocorrelation function tapers off (the higher order autocorrelations are also downweighted by a



(a) Low government expenditure volatility regime ($v_t^g = 1$)



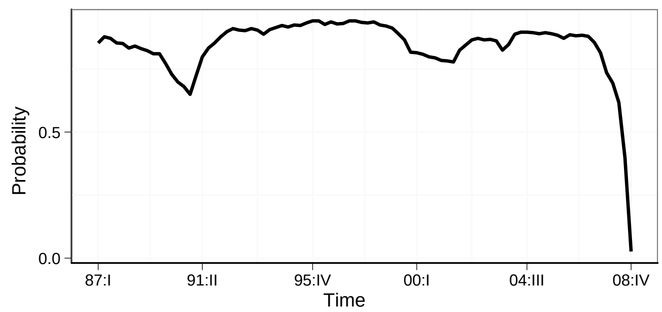
(b) High government expenditure volatility regime ($v_t^g = 2$)

Figure 2.5: Model $\mathcal{M}_{16}$: The posterior probability of government expenditure volatility regimes

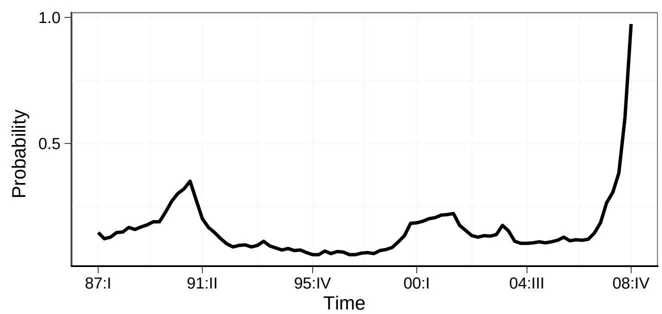
on average are 62.3% and 28.3, respectively, indicating a well mixing, efficient sampler. Also, the sampler converges quickly to the same region of the parameter space regardless of the starting values. Finally, as one can see in Figure 2.8, the posterior densities of the parameters are mostly different from the prior given in Table 2.1. This implies that the data carries information distinct from that contained in the prior distribution.

Two notable features emerge from the table. First, the estimates indicate that the Fed's response to the macro fundamentals is markedly different across policy regimes.

windowing procedure, but we ignore this aspect for simplicity). A well mixing sampler results in autocorrelations that decay to zero within a few lags (and therefore lead to low inefficiency factors), whereas a poorly mixing sampler exhibits persistent correlations even at large lags. Further details are available in Chib (2001).



(a) Low monetary policy volatility regime ($v_t^e = 1$)



(b) High monetary policy volatility regime ($v_t^e = 2$)

Figure 2.6: Model $\mathcal{M}_{16}$: The posterior probability of monetary policy volatility regimes

The reaction coefficient for the output gap is 0.8 during the policy regime 1 whereas in the second policy regime it is 1.35. At the same time, the short rate adjustment to inflation gap is more aggressive. One possible explanation for this is that because inflation has been reasonably stable during the sample period, the Fed's reaction to output gap became relatively more aggressive, marking the break point.

The second important point to note is that the risk-aversion parameter γ has a large posterior mean of 68. This is closely related to the "bond premium puzzle". Rudebusch and Swanson (2008b) show that many DSGE models with standard macroeconomic parameterizations fail to account for the magnitude of risk premium even with habit formation in the household's utility function. This is often termed the "bond premium puzzle".

model	lnL	lnML	n.s.e.	change point
No change point ($\mathcal{M}_1$)	3025.64	3076.71	0.14	-
2-Regime ($\mathcal{M}_2$)	3114.29	3208.77	0.42	1995:Q2
4-Regime ($\mathcal{M}_4$)	3247.35	3383.52	0.42	1995:Q2
16-Regime ($\mathcal{M}_{16}$)	3344.31	3471.43	0.41	1995:Q2
24-Regime ($\mathcal{M}_{24}$)	3346.31	3468.95	0.42	1995:Q2, 2002:Q2

Table 2.2: Log likelihood (lnL) and log marginal likelihood (lnML)

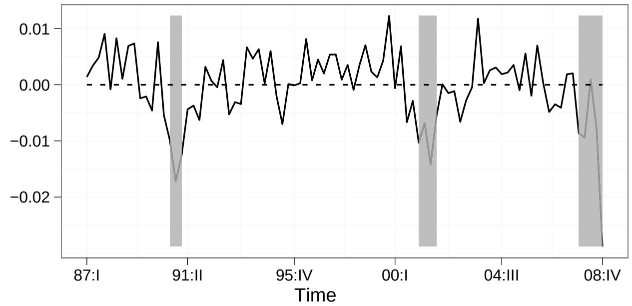
Like in the equity premium puzzle, one possible resolution is a very large value of risk-aversion parameter. Therefore, such large value of γ is essential to account for the level of the term premium.[6]

2.4.3 Changes in the Long Term Bond Risk

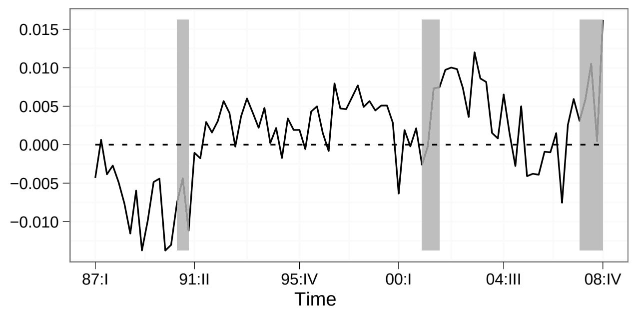
In this paper, the benchmark long-term bond is the five-year Treasury note. Its regime-specific risk is computed by the three different measures as discussed in the section 2.2.10. Figure 2.9 plots the posterior mean of the term premium for the long-term bond over time. Not surprisingly, this risk measure is strictly increasing in maturity (although it is not reported here). It clearly indicates that the average bond risk has diminished since the break, which is consistent with the finding of Chib and Kang (2010).

Moreover, Table 2.4 also reveals that, regardless of the maturity, the average term spread is noticeable lower in the recent regime than in regime 1. Recall that this regime-dependence of the bond risk is solely attributed by the change in the reaction coefficients. This implies that a more active regime on average generates a flatter yield curve. A

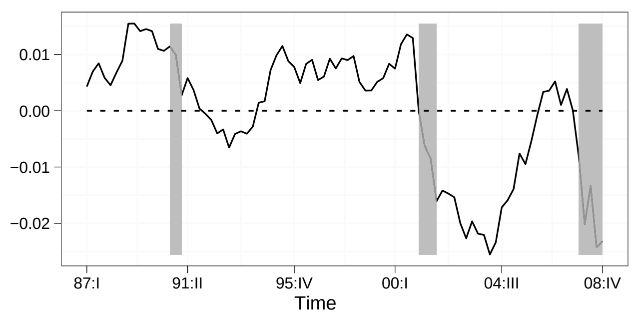
[6] In a standard CRRA preference, high risk aversion (low intertemporal elasticity of substitution) may lead to high real interest rates. However, the average annual real rate implied by our model is 1.884%, which almost matches the observed annual real interest rates of 2.012%. On the other hand, in a calibration exercise, Rudebusch and Swanson (2008a) show that Epstein-Zin preference with a relatively small risk aversion parameter can generate a large risk premium in the context of a single regime DSGE model.



(a) technology ($\hat{a}_t$)



(b) government expenditure ($\hat{g}_t$)



(c) monetary policy ($\hat{e}_t$)

Figure 2.7: Model $\mathcal{M}_{16}$: The exogenous shock process

plausible argument here is that here is that a more aggressive response by the monetary authority can potentially mitigate the effect of the (negative) shocks. This in turn leads the risk-averse agents to expect lower volatility in the macro variables. Hence they price bonds with a smaller market price of risk.

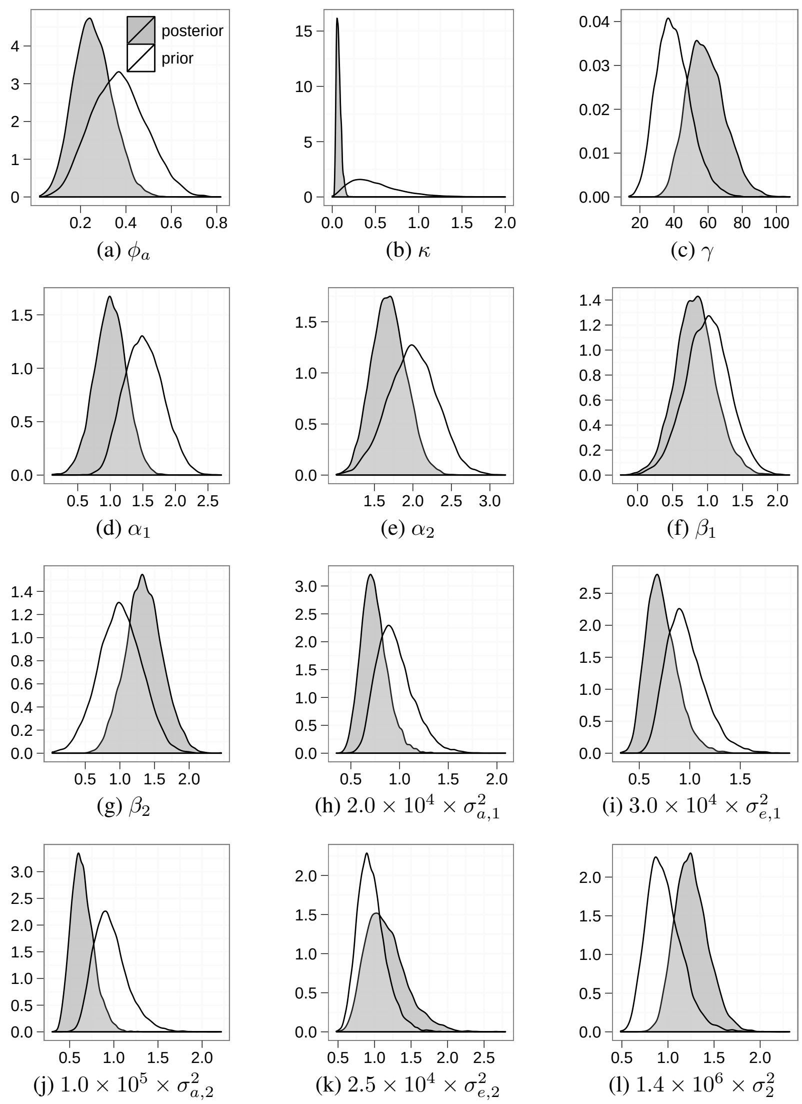


Figure 2.8: Model $\mathcal{M}_{16}$: Marginal prior-posterior plots for some selected parameters

On the other hand, Figure 2.10 presents the result for the decomposition of the term premium of the 5-year bond over time. Interestingly, most of variation of the term premium is explained by the factor risk component. One possible explanation is that sizable factor shocks occur frequently whereas regime shifts happen relatively less frequently. Nevertheless, because the regime shift risk component is consistently positive over time, it should not be neglected.

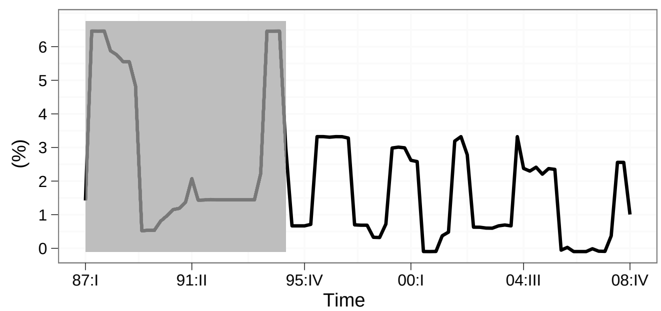
	mean	Numerical S.E.	90% credibility interval	Inefficiency factor	Acceptance rate
δ	0.9989	0.0017	[0.9947, 1.0000]	109.03	42.41
ϕ_a	0.3212	0.0802	[0.1940, 0.4581]	13.93	53.78
ϕ_g	0.9779	0.0080	[0.9637, 0.9902]	413.06	49.66
ϕ_e	0.9546	0.0126	[0.9335, 0.9756]	414.39	50.41
p_{11}	0.9718	0.0174	[0.9405, 0.9996]	184.32	48.29
q_{11}^a	0.8957	0.1072	[0.6684, 0.9941]	13.98	32.16
q_{22}^a	0.8959	0.1072	[0.6706, 0.9938]	9.68	32.22
q_{11}^g	0.9371	0.0348	[0.8733, 0.9855]	128.02	52.60
q_{22}^g	0.9770	0.0220	[0.9310, 0.9987]	33.56	42.60
q_{11}^e	0.8509	0.0274	[0.8025, 0.8911]	134.79	52.00
q_{22}^e	0.7466	0.0875	[0.6009, 0.8865]	184.42	52.40
$400 \times \ln R^*$	4.4463	0.1197	[4.2524, 4.6460]	16.23	53.00
κ	0.0734	0.0583	[0.0129, 0.1897]	400.34	49.91
α_1	0.9590	0.2596	[0.5379, 1.3711]	204.91	51.70
α_2	1.4430	0.2641	[1.0469, 1.9034]	166.47	52.04
β_1	0.7998	0.2804	[0.3634, 1.2874]	113.07	52.50
β_2	1.4834	0.3179	[0.9501, 2.0055]	170.02	51.79
γ	61.678	14.929	[41.883, 90.290]	216.32	51.15
$400 \times \ln a^*$	1.6815	0.1134	[1.4980, 1.8733]	6.26	53.40
$\ln x^*$	0.9817	0.1182	[0.7926, 1.1802]	145.91	49.29
$\ln A_0$	2.2852	0.1182	[2.0867, 2.4750]	146.32	50.10
$2.0 \times 10^4 \times \sigma_{a,1}^2$	1.1002	0.2549	[0.7468, 1.5637]	76.56	52.00
$2.0 \times 10^5 \times \sigma_{g,1}^2$	0.4054	0.0754	[0.2992, 0.5380]	13.73	52.03
$3.0 \times 10^4 \times \sigma_{e,1}^2$	0.4350	0.0870	[0.3119, 0.5879]	25.98	51.73
$1.0 \times 10^4 \times \sigma_{a,2}^2$	0.5221	0.1115	[0.3651, 0.7153]	69.70	51.81
$1.0 \times 10^5 \times \sigma_{g,2}^2$	0.6778	0.1999	[0.3779, 1.0225]	15.35	50.80
$2.5 \times 10^3 \times \sigma_{e,2}^2$	1.8018	1.5077	[0.4647, 4.7115]	276.00	51.62
$1.4 \times 10^7 \times \sigma_2^2$	0.4294	0.1574	[0.2303, 0.7163]	115.12	52.09
$3.0 \times 10^6 \times \sigma_3^2$	0.7350	1.7407	[0.2073, 1.8182]	36.42	48.76
$1.2 \times 10^6 \times \sigma_4^2$	2.4980	1.4150	[0.9006, 5.2584]	199.50	52.40
$6.0 \times 10^5 \times \sigma_5^2$	1.1772	0.5835	[0.5359, 2.4199]	332.15	51.93

Table 2.3: Posterior distribution for the 16-regime model parameters

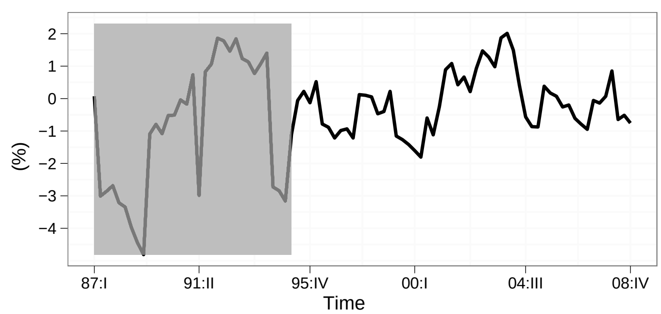
Figure 2.11 indicates the regime-dependence of the factor loadings. The yields in the more active regime are less affected by the shocks to the technological progress and the government expenditure in comparison with those in the less active regime.

	Less active regime ($s_t = 1$) (1987:Q1-1995:Q1)	More active regime ($s_t = 2$) (1995:Q2-2008:Q4)
One-year term spread	0.5000	0.2818
Two-year term spread	1.0051	0.5285
Five-year term spread	1.6133	0.9574

Table 2.4: Regime-specific average term spreads



(a) term premium

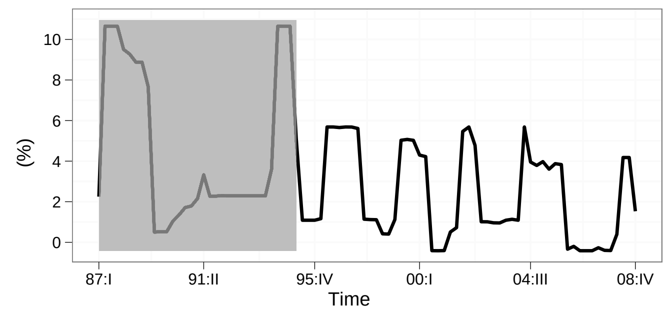


(b) EH component

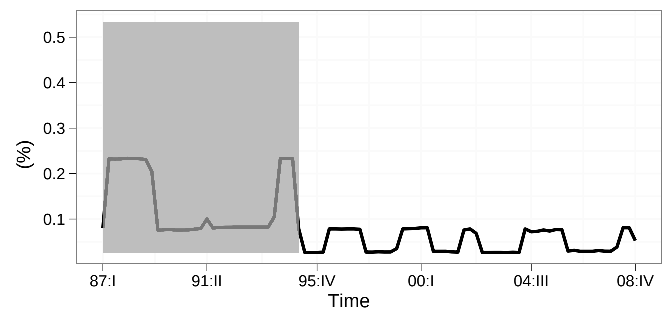
Figure 2.9: The term premium and the EH component of the 5-year bond

2.4.4 Counterfactual Analysis

Since the change point model enables us to estimate the parameters corresponding to each of the regimes, we can perform a time series counterfactual experiment. This exercise is very useful to measure the magnitude of the effect of the monetary policy change on the



(a) factor risk component ($\mathrm{FS}_{d_t,t}^{(\tau)}$)

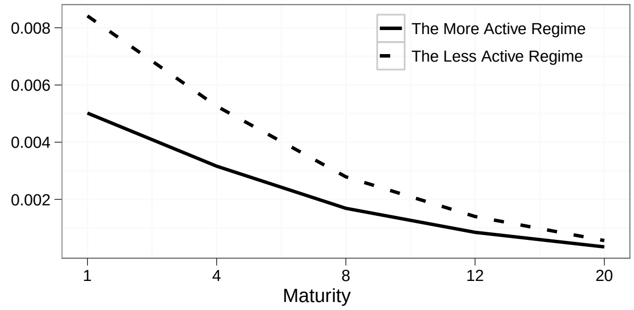


(b) regime shift risk component ($\mathrm{RS}_{d_t,t}^{(\tau)}$)

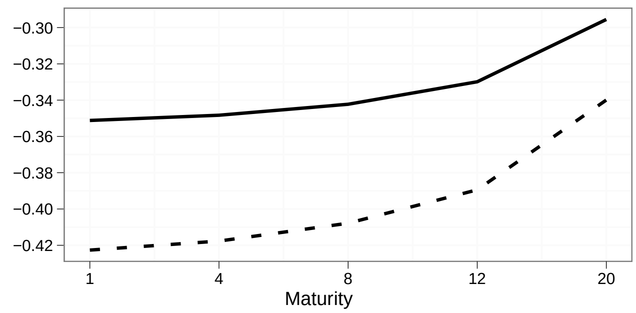
Figure 2.10: Model $\mathcal{M}_{16}$: Decomposition of the term premium of the 5-year bond

macro-economy and the asset prices.

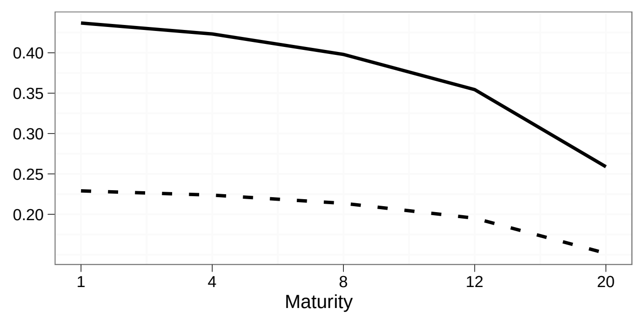
Figure 2.12 plots the results for the short rate and the term spread. As seen in the figure, the short rate would have been more volatile and the slope of the yield curve steeper without the break. On the contrary, if the more active regime prevailed over the entire sample period, then the term spread in regime 1 would have been smaller. As a result, the average yield curve differs across regimes due to the policy change. Figure 2.13 confirms these findings. For instance, the graph on the top clearly shows that the parameters under the more active regime reproduces a much steeper average yield curve than the actual average during the period corresponding to the less active regime. However, Figure 2.14 indicates that inflation and the output growth exhibit little difference, no matter what



(a) technology ($\hat{a}_t$)



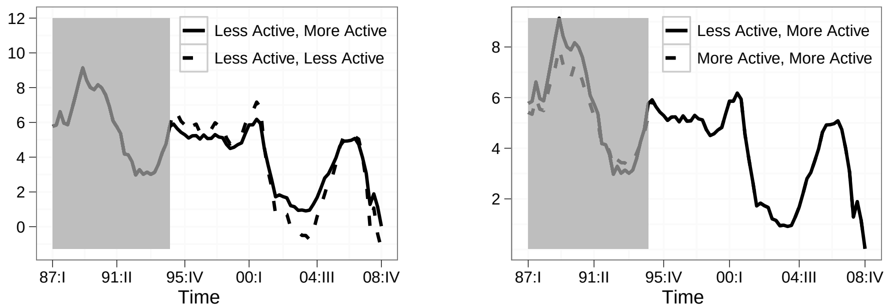
(b) government expenditure ($\hat{g}_t$)



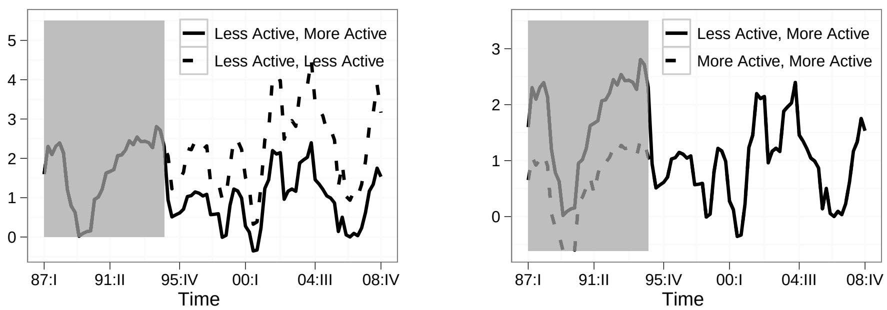
(c) monetary policy ($\hat{e}_t$)

Figure 2.11: Model $\mathcal{M}_{16}$: The factor loadings

policy regime in existence. Therefore, monetary policy regime change mostly impacts the term structure rather than inflation and output growth. This echoes the findings in Gallmeyer, Hollifield, Palomino, and Zin (2008), who also report, within the context of a partial equilibrium model, that the nominal term premium can be highly sensitive to the monetary policy regime.



(a) short rate

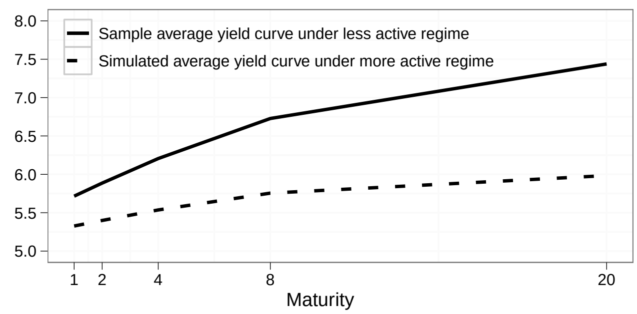


(b) term spread

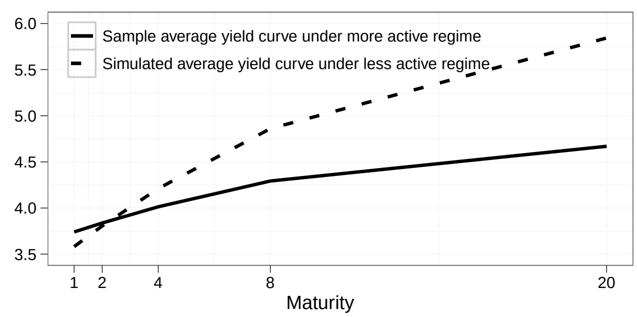
Figure 2.12: Model $\mathcal{M}_{16}$: Counterfactual analysis: interest rates

2.4.5 Out-of-Sample Prediction

As described in section 2.3.5, we forecast the four quarters in 2008 using the data up to 2007:Q4. Following Chib and Kang (2010), the predictive accuracy is measured in terms of the posterior predictive criterion (PPC, Gelfand and Ghosh (1998)). PPC favors models to minimize a sum of goodness-of-fit and penalty term on model complexity. Table 2.5 clearly displays that the proposed model outperforms the alternatives, which is consistent with the marginal likelihood results (i.e. in-sample forecasting).



(a) Less Active Regime (Policy regime 1, 1987:Q1-1995:Q1)



(b) More Active Regime (Policy regime 2, 1995:Q2-2008:Q4)

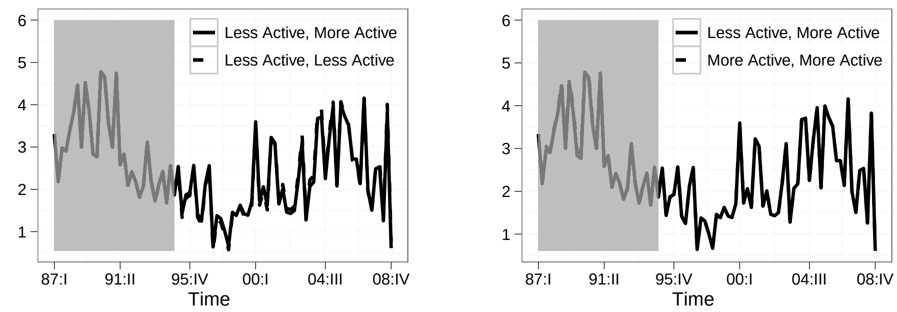
Figure 2.13: Model $\mathcal{M}_{16}$: Counterfactual analysis: average yield curve

model	PPC
No change point ($\mathcal{M}_1$)	84.299
2-Regime ($\mathcal{M}_2$)	56.202
4-Regime ($\mathcal{M}_4$)	50.451
16-Regime ($\mathcal{M}_{16}$)	48.845
24-Regime ($\mathcal{M}_{24}$)	50.467

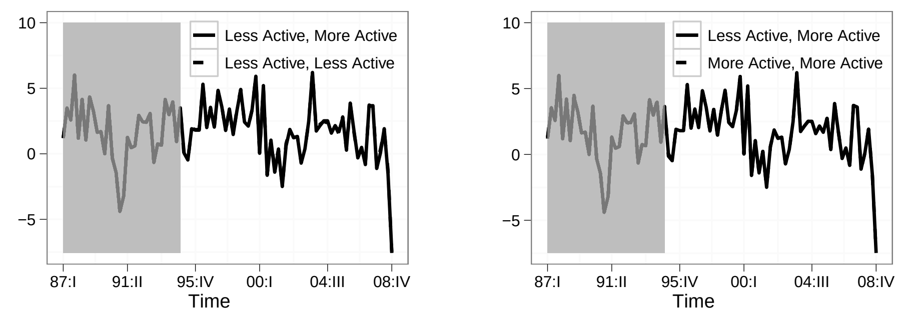
Table 2.5: Posterior predictive criterion

2.5 Conclusion

In this paper we propose and estimate a general equilibrium model of the term structure of interest rates with regime changes. The main goal of our work is to examine the term structure of interest rates from a combined macro-finance perspective. Interest in such



(a) inflation



(b) output growth

Figure 2.14: Model $\mathcal{M}_{16}$: Counterfactual analysis: inflation and output growth

combined modeling is growing and the general equilibrium model we have described, the solution method we have used, and the econometrics we have employed, can all be adapted for other similar purposes. Such work should appear quite rapidly.

Our empirical results reveal that, in its goal of stabilizing the economy, monetary policy has been more responsive to the macro fundamentals since 1995:Q2 with important effects on the dynamics of the term structure. Because in a more active regime agents anticipate less volatility in the macro variables, bonds are priced with a lower market price of risk. As a result, the average term premium is smaller in this regime and the slope of the yield curve is flatter on average.

Furthermore, during the more active policy period, both the average term premium and its volatility have fallen, and consequently, whereas the term premium explains a significant portion of the term spread in the less active regime, the relative share of the market expectations has increased in the second regime.

Finally, our empirical findings reveal the important role that incorporating regime changes can play in improving forecasts of the term structure.

Appendix A

Essay 1: Bond Prices under Regime Changes

By the assumption of the affine model, we have

$$P_t(s_t,\tau) = \exp\left(-a_{s_t}(\tau) - \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})\right) \tag{A.0.1}$$

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

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

$$\begin{aligned}&\frac{P_{t+1}(s_{t+1},\tau-1)}{P_t(s_t,\tau)} \\ &= \exp\left[-a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}}) + a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'(\mathbf{f}_t - \boldsymbol{\mu}_{s_t})\right]\end{aligned} \tag{A.0.2}$$

It immediately follows from the bond pricing formula that

$$\begin{aligned}1 &= \mathbf{E}_t\left[\kappa_{t,s_t,t+1}\frac{P_{t+1}(s_{t+1},\tau-1)}{P_t(s_t,\tau)}\right] \\ &= \mathbf{E}_t\left[\kappa_{t,s_t,t+1}h_{\tau,t+1}\right].\end{aligned} \tag{A.0.3}$$

Then by substitution

$$\begin{aligned}&\kappa_{t,s_t,t+1}h_{\tau,t+1} \\ &= \exp[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t} - \boldsymbol{\gamma}'_{t,s_t}\mathbf{L}^{-1}_{s_{t+1}}\boldsymbol{\eta}_{t+1} \\ &- a_{s_{t+1}}(\tau-1) - \mathbf{b}_{s_{t+1}}(\tau-1)'\left(\mathbf{f}_{t+1} - \boldsymbol{\mu}_{s_{t+1}}\right) + a_{s_t}(\tau) + \mathbf{b}_{s_t}(\tau)'\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)]\end{aligned} \tag{A.0.4}$$

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

where

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

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

$$\mathbf{E}_t\left(\mathsf{exp}[-\frac{1}{2}\Gamma_{t,\tau}\Gamma'_{t,\tau} - \Gamma_{t,\tau}\boldsymbol{\omega}_{t+1}]\right) = 1 \tag{A.0.5}$$

we have that

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

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

$$\begin{aligned}
&\mathbf{E}\left[\kappa_{t,s_t,t+1}h_{\tau,t+1}|\mathbf{f}_t, s_{t+1}, s_t\right] \qquad \text{(A.0.6)} \\
&= \mathsf{exp}[-r_{t,s_t} - \frac{1}{2}\boldsymbol{\gamma}'_{t,s_t}\boldsymbol{\gamma}_{t,s_t} \\
&+ \frac{1}{2}\left(\boldsymbol{\gamma}'_{t,s_t} + \mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\right)\left(\boldsymbol{\gamma}'_{t,s_t} + \mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\right)' + \zeta_{\tau,s_t,s_{t+1}}] \\
&\approx -r_{t,s_t} + \boldsymbol{\gamma}'_{t,s_t}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1) + \frac{1}{2}\left(\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\right) + \zeta_{\tau,s_t,s_{t+1}} + 1 \\
&= -\left(\delta_{1,s_t} + \boldsymbol{\delta}'_{2,s_t}\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right) + \left(\tilde{\boldsymbol{\gamma}}_{s_t} + \boldsymbol{\Phi}_{s_t}\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right)'\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1) \\
&+ \frac{1}{2}\left(\mathbf{b}_{s_{t+1}}(\tau-1)'\mathbf{L}_{s_{t+1}}\mathbf{L}'_{s_{t+1}}\mathbf{b}_{s_{t+1}}(\tau-1)\right) + \zeta_{\tau,s_t,s_{t+1}} + 1
\end{aligned}$$

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

$$\mathbf{E}\left[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_t = j\right] = \sum_{s_{t+1}=j,k} p_{j s_{t+1}} \mathbf{E}\left[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t = j\right] \tag{A.0.7}$$
$$= 1 \text{ where } k = j + 1.$$

Thus we have

$$0 = \sum_{s_{t+1}=j,k} p_{j s_{t+1}} \left\{\mathbf{E}\left[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1}, s_t = j\right] - 1\right\} \text{ since } \sum_{s_{t+1}=j,k} p_{j s_{t+1}} = 1 \tag{A.0.8}$$
$$= p_{jj}\left(\mathbf{E}\left[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1} = j, s_t = j\right] - 1\right)$$
$$+ p_{jk}\left(\mathbf{E}\left[\kappa_{t,s_t,t+1} h_{\tau,t+1} | \mathbf{f}_t, s_{t+1} = k, s_t = j\right] - 1\right)$$
$$\approx -p_{jj}\left(\delta_{1,j} + \boldsymbol{\delta}'_{2,j}\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right) + p_{jj}\left(\tilde{\gamma}_j + \Phi_j\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right)' \mathbf{L}'_j \mathbf{b}_j(\tau - 1)$$
$$+ \frac{1}{2} p_{jj}\left(\mathbf{b}_j(\tau-1)' \mathbf{L}_j \mathbf{L}'_j \mathbf{b}_j(\tau-1)\right) + p_{jj} \zeta_{\tau,j,j}$$
$$- p_{jk}\left(\delta_{1,j} + \boldsymbol{\delta}'_{2,j}\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right) + p_{jk}\left(\tilde{\gamma}_j + \Phi_j\left(\mathbf{f}_t - \boldsymbol{\mu}_{s_t}\right)\right)' \mathbf{L}'_k \mathbf{b}_k(\tau - 1)$$
$$+ \frac{1}{2} p_{jk}\left(\mathbf{b}_k(\tau-1)' \mathbf{L}_k \mathbf{L}'_k \mathbf{b}_k(\tau-1)\right) + p_{jk} \zeta_{\tau,j,k}$$

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

Essay 1: 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 62-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 (2009)).

Step 2 Sampling $\boldsymbol{\theta}$

We sample $\boldsymbol{\theta}$ conditioned on $(\mathbf{S}_n, u_0, \boldsymbol{\sigma}^{*2})$ 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 $\boldsymbol{\theta}$ 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 gth iteration, we have h_g sub-blocks of $\boldsymbol{\theta}$

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

If $\boldsymbol{\psi}_{-i}$ denotes the collection of the parameters in $\boldsymbol{\psi}$ except $\boldsymbol{\theta}_i$, then the proposal density $q\left(\boldsymbol{\theta}_i|\mathbf{y},\boldsymbol{\psi}_{-i}\right)$ for the ith block conditioned on $\boldsymbol{\psi}_{-i}$ is constructed by a quadratic approximation at the mode of the current target density $\pi\left(\boldsymbol{\theta}_i|\mathbf{y},\boldsymbol{\psi}_{-i}\right)$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q\left(\boldsymbol{\theta}_i|\mathbf{y},\boldsymbol{\psi}_{-i}\right) = St\left(\boldsymbol{\theta}_i|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15\right) \tag{B.0.1}$$

where

$$\hat{\boldsymbol{\theta}}_i = \arg\max_{\boldsymbol{\theta}_\mathbf{i}} \ln\{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i)\} \tag{B.0.2}$$

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

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

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

$$\begin{aligned}&\alpha(\boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\theta}_i^\dagger|\mathbf{y}, \boldsymbol{\psi}_{-i}) \\ &= \min\left\{\frac{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^\dagger, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^\dagger)}{p(\mathbf{y}|\mathbf{S}_n, \boldsymbol{\theta}_i^{(g-1)}, \boldsymbol{\psi}_{-i})\pi(\boldsymbol{\theta}_i^{(g-1)})}\frac{St(\boldsymbol{\theta}_i^{(g-1)}|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, 15)}{St(\boldsymbol{\theta}_j^\dagger|\hat{\boldsymbol{\theta}}_\mathbf{i}, \mathbf{V}_{\hat{\boldsymbol{\theta}}_\mathbf{i}}, 15)}, 1\right\}.\end{aligned} \tag{B.0.3}$$

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

$$\pi\left(\boldsymbol{\theta}_1|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-1}\right), \ \pi\left(\boldsymbol{\theta}_2|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-2}\right), \ldots, \ \pi(\boldsymbol{\theta}_{h_g}|\mathbf{y}, \mathbf{S}_n, \boldsymbol{\psi}_{-h_g}) \tag{B.0.4}$$

are sequentially updated as above.

Step 3 Sampling the initial factor

Given the prior in equation (1.3.14), 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 (1.3.5). In the following, it is assumed that all the underlying coefficients are those in regime 0. Then

$$u_0|\mathbf{f}_1, \boldsymbol{\theta} \sim \mathcal{N}_1 \left(\bar{u}_0, \mathbf{U}_0\right) \tag{B.0.5}$$

where

$$\bar{u}_0 = \mathbf{U}_0 \left(\Sigma_u^{-1} + \mathbf{H}^{*\prime}\Omega_{11,0}^{*} u_1^{*}\right), \quad \mathbf{U}_0 = \left(\Sigma_u^{-1} + \mathbf{H}^{*\prime}\Omega_{11,1}^{*}\mathbf{H}^{*}\right)$$

and on letting

$$\begin{aligned}
\mathbf{G}_0 &= \begin{pmatrix} \mathbf{G}_{11,1} & \mathbf{G}_{12,1} \\ \mathbf{G}_{21,1} & \mathbf{G}_{22,1} \end{pmatrix}, \ \Omega_1 = \begin{pmatrix} \Omega_{11,1} & \Omega_{12,1} \\ \Omega_{21,1} & \Omega_{22,1} \end{pmatrix} \\
\mathbf{H}^{*} &= \mathbf{G}_{11,1} - \Omega_{12,1}\Omega_{22,1}^{-1}\mathbf{G}_{21,1}, \ \ \Omega_{11,1}^{*} = \Omega_{11,1} - \Omega_{12,1}\Omega_{22,1}^{-1}\Omega_{21,1} \\
u_1^{*} &= u_1 - \Omega_{12,1}\Omega_{22,1}^{-1}(\mathbf{m}_1 - \boldsymbol{\mu}_{m,1}) + \left(\Omega_{12,1}\Omega_{22,1}^{-1}\mathbf{G}_{22,1} - \mathbf{G}_{12,1}\right)(\mathbf{m}_0 - \boldsymbol{\mu}_{m,1})
\end{aligned}$$

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) 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 $\Pr[s_1 = 1|I_1, \boldsymbol{\psi}] = 1$. Then one first obtains $\Pr[s_t = j|I_t, \boldsymbol{\psi}]$ for all $j = 1, 2, .., m+1$ and $t = 1, 2, .., n$ by calculating

$$\Pr[s_t = j|I_t, \boldsymbol{\psi}] = \sum_{i=j-1}^{j} \Pr[s_{t-1} = i, s_t = j|I_t, \boldsymbol{\psi}] \tag{B.0.6}$$

where

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

This can be done by the equations (1.3.22)-(1.3.25).

In the backward pass, one simulates $\mathbf{S}_n$ by the method of composition. One samples s_n from $\Pr[s_n = 1 | I_n, \boldsymbol{\psi}]$. We remark that in this sampling step, s_n can take any value in $\{1, 2, .., m+1\}$. For instance, if s_n turns out to be m and not $(m+1)$, then m is taken to be the absorbing regime and 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 $\Pr[s_n = 1 | I_n, \boldsymbol{\psi}]$ has almost a unit mass on $(m+1)$. Then for $t = 1, 2, .., n-1$ we sequentially calculate

$$\begin{aligned} \Pr[s_t = j | I_t, s_{t+1} = k, S^{t+2}, \boldsymbol{\psi}] &= \Pr[s_t = j | I_t, s_{t+1} = k, \boldsymbol{\psi}] \\ &= \frac{\Pr[s_{t+1} = k | s_t = j] \Pr[s_t = j | I_t, \boldsymbol{\psi}]}{\sum_{j=k-1}^{k} \Pr[s_{t+1} = k | s_t = j] \Pr[s_t = j | I_t, \boldsymbol{\psi}]} \end{aligned} \tag{B.0.7}$$

where $S^{t+1} = \{s_{t+1}, .., s_n\}$ 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 $i \in \{1, 2, .., 7, 9, .., 16\}$ and $j = 1, 2, .., m+1$, $\sigma^{*2}_{i,j}$ is sampled from

$$\mathbf{IG}\left\{ \frac{\bar{v} + \sum_{t=1}^{n} I(s_t = j)}{2}, \frac{\bar{d} + \sum_{t=1}^{n} d_{i,j} I(s_t = j)(R_{ti} - \bar{a}_{i,j} - \bar{\mathbf{b}}'_{i,j}(\mathbf{f}_t - \boldsymbol{\mu}_j))^2}{2} \right\} \tag{B.0.8}$$

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

Appendix C

Essay 2: Solution

When solving the model we enforce the condition that the stable solution is unique and bounded. Our model solution method relies on the approach of Davig and Leeper (2007). For this, we construct the auxiliary representation of the linearized equilibrium dynamics or the stacked system which is available for any purely forward-looking rational expectations model with regime changes. We begin by defining the state-contingent forecast error as

$$\eta_{jt+1}^{\pi} = \hat{\pi}_{jt+1} - \mathbb{E}_t\left(\hat{\pi}_{jt+1}\right) \text{ and } \eta_{jt+1}^{x} = \hat{x}_{jt+1} - \mathbb{E}_t\left(\hat{x}_{jt+1}\right), \quad j = 1, 2 \tag{C.0.1}$$

where $\hat{y}_{jt+1}$ denotes the value of $\hat{y}_{t+1}$ conditioned on $s_{t+1} = j$. Then substituting the conditional expectations in equations (2.2.38) and (2.2.39) into the system of equations (2.2.34)-(2.2.35) yields the following stacked system

$$\mathbf{A}\begin{bmatrix} \hat{\pi}_{1t+1} \\ \hat{\pi}_{2t+1} \\ \hat{x}_{1t+1} \\ \hat{x}_{2t+1} \end{bmatrix} = \mathbf{B}\begin{bmatrix} \hat{\pi}_{1t} \\ \hat{\pi}_{2t} \\ \hat{x}_{1t} \\ \hat{x}_{2t} \end{bmatrix} + \mathbf{A}\begin{bmatrix} \eta_{1t+1}^{\pi} \\ \eta_{2t+1}^{\pi} \\ \eta_{1t+1}^{x} \\ \eta_{2t+1}^{x} \end{bmatrix} + \mathbf{C}\bar{\mathbf{f}}_t \tag{C.0.2}$$

where

$$\mathbf{A} = \begin{bmatrix} \delta \otimes \mathbf{P} & 0_{2\times 2} \\ \mathbf{P} & \gamma \otimes \mathbf{P} \end{bmatrix}, \tag{C.0.3}$$

$$\mathbf{B}_{11} = \mathbf{I}_{m+1},\ \mathbf{B}_{12} = -\kappa \times \mathbf{I}_{m+1},\ \mathbf{B}_{21} = \mathsf{diag}(\alpha_1, \alpha_2, .., \alpha_{m+1}), \tag{C.0.4}$$

$$\mathbf{B}_{22} = \mathsf{diag}(\beta_1 + \gamma, \beta_2 + \gamma, .., \beta_{m+1} + \gamma),$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \tag{C.0.5}$$

$$\text{and } \mathbf{C} = \begin{bmatrix} 0 & \kappa & 0 \\ 0 & \kappa & 0 \\ -\phi_a & \gamma(\phi_g - 1) & 1 \\ -\phi_a & \gamma(\phi_g - 1) & 1 \end{bmatrix} \tag{C.0.6}$$

Uniqueness and boundedness of the MSV solution are equivalent to the determinacy restriction of the solution space of this stacked system (Davig and Leeper (2007)). In terms of the computational details, this restriction requires that all the generalized eigenvalues of **A** and **B** lie outside the unit circle. For further discussion about necessary and sufficient conditions for determinacy, we refer the interested reader to Davig and Leeper (2007).

Appendix D

Essay 2: Bond Prices

This section provides the details on the derivation of the bond prices in (2.2.50) and (2.2.51). We begin by letting $\mathbf{E}^{d_{t+1}}$ denote an expectation conditioned on d_{t+1}. Then the equation (2.2.44) can be expressed as

$$P^{(\tau)}_{d_t,t} = \mathbb{E}^{d_{t+1}}\left[P^{(\tau)}_{d_t,d_{t+1},t}\right] \text{ where } P^{(\tau)}_{d_t,d_{t+1},t} \equiv \mathbb{E}\left[M_{t,t+1}P^{(\tau-1)}_{d_{t+1},t+1}|\bar{\mathbf{f}}_t, d_t, d_{t+1}\right] \quad \text{(D.0.1)}$$

or

$$1 = \mathbb{E}^{s_{t+1}}\left[\mathbb{E}\left[M_{t,t+1}h_{\tau,t+1}|\bar{\mathbf{f}}_t, d_t, d_{t+1}\right]\right] \quad \text{(D.0.2)}$$

where

$$\begin{aligned} h_{\tau,t+1} &= P^{(\tau-1)}_{d_{t+1},t+1}/P^{(\tau)}_{d_t,t} \quad \text{(D.0.3)} \\ &= \exp\left[-\mathbf{a}_{d_{t+1}}(\tau-1) - \mathbf{b}_{d_{t+1}}(\tau-1)'\bar{\mathbf{f}}_{t+1} + \mathbf{a}_{d_t}(\tau) + \mathbf{b}_{d_t}(\tau)'\bar{\mathbf{f}}_t\right]. \end{aligned}$$

If we define

$$\Theta_{d_t,d_{t+1}} = -\mathbf{a}_{d_{t+1}}(\tau-1) + \mathbf{a}_{d_t}(\tau) + \left(\mathbf{b}_{d_t}(\tau)' - \mathbf{b}_{d_{t+1}}(\tau-1)'\phi\right)\bar{\mathbf{f}}_t \quad \text{(D.0.4)}$$

and $\Gamma_{\tau,d_{t+1}} = \mathbf{L}_{d_{t+1}} - \mathbf{b}_{d_{t+1}}(\tau-1)'$, then $M_{t,t+1}h_{\tau,t+1}$ can be rewritten as

$$\begin{aligned}
&\exp\left[-\ln R^{*}-\frac{1}{2}\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}'_{d_{t+1}}+\boldsymbol{\lambda}_{d_t,d_{t+1}}\overline{\mathbf{f}}_t+\left(\mathbf{L}_{d_{t+1}}-\mathbf{b}_{d_{t+1}}(\tau-1)'\right)\varepsilon_{t+1}+\Theta_{d_t,d_{t+1}}\right]\\
&=\exp\left[-\ln R^{*}-\frac{1}{2}\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}'_{d_{t+1}}+\boldsymbol{\lambda}_{d_t,d_{t+1}}\overline{\mathbf{f}}_t+\Gamma_{\tau,d_{t+1}}\varepsilon_{t+1}+\Theta_{d_t,d_{t+1}}\right]\\
&=\exp\left[-\ln R^{*}-\frac{1}{2}\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}'_{d_{t+1}}+\boldsymbol{\lambda}_{d_t,d_{t+1}}\overline{\mathbf{f}}_t+\frac{\mathbf{1}}{\mathbf{2}}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}}+\Theta_{d_t,d_{t+1}}\right]\\
&\times\exp\left[-\frac{1}{2}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}}+\boldsymbol{\Gamma}_{\tau,d_{t+1}}\varepsilon_{t+1}\right] \qquad \text{(D.0.5)}
\end{aligned}$$

Since

$$\mathbb{E}\left[\exp\left[-\frac{1}{2}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}}+\Gamma'_{\tau,d_{t+1}}\varepsilon_{t+1}\right]|\overline{\mathbf{f}}_t,d_t,d_{t+1}\right]=1 \qquad \text{(D.0.6)}$$

the log-approximation gives

$$\begin{aligned}
&\mathbb{E}\left[M_{t,t+1}h_{\tau,t+1}|\overline{\mathbf{f}}_t,d_t,d_{t+1}\right] \qquad \text{(D.0.7)}\\
&\approx-\ln R^{*}+\boldsymbol{\lambda}_{d_t,d_{t+1}}\overline{\mathbf{f}}_t-\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1)'+\frac{1}{2}\mathbf{b}_{d_{t+1}}(\tau-1)'\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1)\\
&+\Theta_{d_t,d_{t+1}}+1
\end{aligned}$$

The next step is integrating out d_{t+1} for $d_t=i$ $(i=1,2,3,4)$. Then the equation (D.0.1) implies that

$$0=\sum_{j=1}^{\mathbf{d}}p_{ij}\left(-\ln R^{*}+\boldsymbol{\lambda}_{i,j}\overline{\mathbf{f}}_t-\mathbf{L}_j\Omega_j\mathbf{b}_j(\tau-1)'+\frac{1}{2}\mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1)+\Theta_{i,j}\right) \qquad \text{(D.0.8)}$$

Matching the coefficients for constant and $\overline{\mathbf{f}}_t$ completes the derivation of the bond prices.

Appendix E

Essay 2: Proof of the Term Premium and the Expected Excess Return

This appendix provides the proof of the term premium and the expected excess return in the equation (2.2.54) and (2.2.56).

By definition, the term spread of τ-period bond yield is given by

$$r_{d_t,t}^{(\tau)} - r_{d_t,t}^{(1)} \tag{E.0.1}$$

Let $x_t^{(\tau)} = p_{d_{t+1},t+1}^{\tau-1} - p_{d_t,t}^{\tau} - r_{d_t,t}^{(1)}$ denote the excess return. Then we have

$$\begin{aligned} r_{d_t,t}^{(\tau)} - r_{d_t,t}^{(1)} &= \frac{1}{\tau}\sum_{l=0}^{\tau-1} \mathbb{E}_t\left[r_{d_{t+l},t+l}^{(1)}\right] - r_{d_t,t}^{(1)} + \frac{1}{\tau}\sum_{i=1}^{\tau-1} \mathbb{E}_t\left[x_t^{(\tau+1-i)}\right] \\ &= \frac{1}{\tau}\sum_{l=0}^{\tau-1} \mathbb{E}_t\left[r_{d_{t+l},t+l}^{(1)}\right] - r_{d_t,t}^{(1)} + \frac{1}{\tau}\sum_{i=2}^{\tau} \mathrm{exr}_{d_t,t}^{(i)} \\ &= \frac{1}{\tau}\sum_{l=0}^{\tau-1} \mathbb{E}_t\left[r_{d_{t+l},t+l}^{(1)}\right] - r_{d_t,t}^{(1)} + \mathrm{TP}_{d_t,t}^{(\tau)} \end{aligned} \tag{E.0.2}$$

where

$$\mathrm{TP}_{d_t,t}^{(\tau)} = \frac{1}{\tau}\sum_{i=2}^{\tau} \mathrm{exr}_{d_t,t}^{(i)} = \frac{1}{\tau}\left(\mathrm{exr}_{d_t,t}^{(2)} + \mathrm{exr}_{d_t,t}^{(3)} + ... + \mathrm{exr}_{d_t,t}^{(\tau)}\right) \tag{E.0.3}$$

Now we prove the equation (2.2.56). We begin by noting that the risk-neutral pricing formula in the equation (2.2.44) implies

$$p_{d_t,t}^{(\tau)} = \mathbb{E}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] \tag{E.0.4}$$

This equation holds exactly when the conditional distribution of bond prices and the pricing kernel are jointly log-normal. Then it follows that

$$\begin{aligned} p_{d_t,t}^{(\tau)} &= \mathbb{E}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] \\ &= \mathbb{E}_t\left[m_{t,t+1}\right] + \mathbb{E}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] \\ &= p_{d_t,t}^{(1)} - \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1}\right] + \mathbb{E}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p_{d_{t+1},t+1}^{(\tau-1)}\right] \\ \text{since } p_{d_t,t}^{(1)} &= \mathbb{E}_t\left[m_{t,t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1}\right] \end{aligned} \tag{E.0.5}$$

and thus

$$p_{d_t,t}^{(\tau)} = p_{d_t,t}^{(1)} + \mathbb{E}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] + \frac{1}{2}\mathbb{V}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] + \mathsf{Cov}_t\left[m_{t,t+1}, p_{d_{t+1},t+1}^{(\tau-1)}\right] \tag{E.0.6}$$

This implies that

$$\begin{aligned} \mathrm{exr}_{d_t,t}^{(\tau)} &= \left[\mathbb{E}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] - p_{d_t,t}^{(\tau)}\right] - (-p_{d_t,t}^{(1)}) \\ &= -\mathsf{Cov}_t\left[m_{t,t+1}, p_{d_{t+1},t+1}^{(\tau-1)}\right] - \frac{1}{2}\mathbb{V}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] \end{aligned} \tag{E.0.7}$$

The covariance term is compensation for holding long term bond risk associated with the macro structural shocks, and the variance term is the convexity effect (Jensen's inequality).

The remaining is to compute the two terms in the equation (E.0.7). We begin by expressing the pricing kernel and the log of bond price as

$$m_{t,t+1} \approx W_{d_t,d_{t+1},t} + \mathbf{L}_{d_{t+1}}\varepsilon_{t+1} \tag{E.0.8}$$

$$p^{(\tau-1)}_{d_{t+1},t+1} = -a_{d_{t+1}}(\tau-1) - \mathbf{b}_{d_{t+1}}(\tau-1)' \left(\phi\bar{\mathbf{f}}_t + \varepsilon_{t+1}\right) \tag{E.0.9}$$
$$= K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}$$

where

$$W_{d_t,d_{t+1},t} = c_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}}\bar{\mathbf{f}}_t \text{ and } K_{d_{t+1},t} = -a_{d_{t+1}} - \mathbf{b}_{d_{t+1}}(\tau-1)'\phi\bar{\mathbf{f}}_t$$

We first compute the conditional covariance between $m_{t,t+1}$ and $p^{(\tau-1)}_{d_{t+1},t+1}$ using the law of iterative expectation as follows.

$$\mathbb{E}_t[p^{(\tau-1)}_{d_{t+1},t+1}] = \mathbb{E}_t\left(\mathbb{E}_t[p^{(\tau-1)}_{t+1}|d_{t+1}]\right) = \mathbb{E}_t\left(K_{d_{t+1},t}\right) = \sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t} \tag{E.0.10}$$

$$\mathbb{E}_t[m_{t,t+1}] = \mathbb{E}_t[W_{d_t,d_{t+1},t}] = \sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t} \tag{E.0.11}$$

$$\mathbb{E}_t[m_{t,t+1}p^{(\tau-1)}_{d_{t+1},t+1}] = \mathbb{E}_t[\left(W_{d_t,d_{t+1},t} + \mathbf{L}_{d_{t+1}}\varepsilon_{t+1}\right)\left(K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}\varepsilon_{t+1}\right)]$$
$$= \mathbb{E}_t[W_{d_t,d_{t+1},t}K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}_{d_{t+1}}]$$
$$= \sum_{j=1}^{\mathbf{d}} p_{ij}\left(W_{i,j,t}K_{j,t} - \mathbf{b}_j\Omega_j\mathbf{L}_j\right) \tag{E.0.12}$$

Therefore,

$$-\mathsf{Cov}_t(m_{t,t+1}, p^{(\tau-1)}_{d_{t+1},t+1}) = \mathbb{E}_t[p^{(\tau-1)}_{d_{t+1},t+1}]\mathbb{E}_t[m_{t,t+1}] - \mathbb{E}_t[m_{t,t+1}p^{(\tau-1)}_{d_{t+1},t+1}] \tag{E.0.13}$$
$$= \left(\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right)\left(\sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}\right) - \sum_{j=1}^{\mathbf{d}} p_{ij}\left(W_{i,j,t}K_{j,t} - \mathbf{b}_j\Omega_j\mathbf{L}_j\right)$$

For the conditional variance of $p^{(\tau-1)}_{d_{t+1},t+1}$,

$$\mathbb{E}_t\left[\left(p^{(\tau-1)}_{d_{t+1},t+1}\right)^2\right] = \mathbb{E}_t\left[\left(K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}\right)^2\right] \tag{E.0.14}$$
$$= \mathbb{E}_t\left[K^2_{d_{t+1},t} - 2K_{d_{t+1},t}\mathbf{b}_{d_{t+1}}\varepsilon_{t+1} + \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}\varepsilon'_{t+1}\mathbf{b}_{d_{t+1}}(\tau-1)\right]$$

$$= \mathbb{E}_t \left[K^2_{d_{t+1},t} + \mathbf{b}_{d_{t+1}}(\tau-1)'\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1) \right]$$
$$= \sum_{j=1}^{\mathbf{d}} p_{ij} \left(K^2_{j,t} + \mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) \right)$$

and thus

$$\mathbb{V}_t \left[p^{(\tau-1)}_{d_{t+1},t+1} \right] = \mathbb{E}_t \left[\left(p^{(\tau-1)}_{d_{t+1},t+1} \right)^2 \right] - \left(\mathbb{E}_t \left[p^{(\tau-1)}_{d_{t+1},t+1} \right] \right)^2 \tag{E.0.15}$$
$$= \sum_{j=1}^{\mathbf{d}} p_{ij} \left(K^2_{j,t} + \mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) \right) - \left(\sum_{j=1}^{\mathbf{d}} p_{ij} K_{j,t} \right)^2$$

which completes the proof.

Appendix F

Essay 2: MCMC Sampling

Step 2 Sampling $\boldsymbol{\theta}$

Integrating out $\mathbf{F}_n$, we sample $\boldsymbol{\theta}$ conditioned on $\mathbf{D}_n$ by using the tailored randomized block M-H (TaRB-MH) algorithm. In the gth iteration, we have h_g sub-blocks of $\boldsymbol{\theta}$

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

The variance of pricing errors $\left\{\sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2\right\}$ and the initial technology level $\ln A_0$ form two fixed blocks ($\boldsymbol{\theta}_{h_g-1}$ and $\boldsymbol{\theta}_{h_g}$), and the others are randomly grouped ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, .., $\boldsymbol{\theta}_{h_g-2}$). Then the proposal density $q\left(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y}\right)$ for the ith block, conditioned on the most current value of the remaining blocks $\boldsymbol{\theta}_{-i}$, is constructed by a quadratic approximation at the mode of the current target density $\pi\left(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y}\right)$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q\left(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y}\right) = St\left(\boldsymbol{\theta}_i|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, \mathbf{15}\right) \tag{F.0.1}$$

where

$$\hat{\boldsymbol{\theta}}_i = \arg\max_{\boldsymbol{\theta}_\mathbf{i}} \ln\{f(\mathbf{y}|\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}, \mathbf{D}_n)\pi(\boldsymbol{\theta}_i)\} \tag{F.0.2}$$

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

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

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

$$\alpha\left(\boldsymbol{\theta}_i^{(g-1)},\boldsymbol{\theta}_i^{\dagger}|\boldsymbol{\theta}_{-i},\mathbf{y}\right) \tag{F.0.3}$$
$$= \min\left\{\frac{f\left(\mathbf{y}|\boldsymbol{\theta}_i^{\dagger},\boldsymbol{\theta}_{-i},\mathbf{D}_n\right)\pi\left(\boldsymbol{\theta}_i^{\dagger}\right)}{f\left(\mathbf{y}|\boldsymbol{\theta}_i^{(g-1)},\boldsymbol{\theta}_{-i},\mathbf{D}_n\right)\pi\left(\boldsymbol{\theta}_i^{(g-1)}\right)}\frac{St\left(\boldsymbol{\theta}_i^{(g-1)}|\hat{\boldsymbol{\theta}}_i,\mathbf{V}_{\hat{\boldsymbol{\theta}}_i},\mathbf{15}\right)}{St\left(\boldsymbol{\theta}_j^{\dagger}|\hat{\boldsymbol{\theta}}_{\mathbf{i}},\mathbf{V}_{\hat{\boldsymbol{\theta}}_{\mathbf{i}}},\mathbf{15}\right)},1\right\}.$$

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

$$\pi\left(\boldsymbol{\theta}_1|\boldsymbol{\theta}_{-1},\mathbf{y},\mathbf{D}_n\right),\pi\left(\boldsymbol{\theta}_2|\boldsymbol{\theta}_{-2},\mathbf{y},\mathbf{D}_n\right),\ldots,\ \pi\left(\boldsymbol{\theta}_{h_g}|\boldsymbol{\theta}_{-h_g},\mathbf{y},\mathbf{D}_n\right) \tag{F.0.4}$$

are sequentially updated as above.

Now we explain how to calculate $f\left(\mathbf{y}|\boldsymbol{\theta},\mathbf{D}_n\right)$ integrating out $\mathbf{F}_n$ where I_t is the history of the outcomes up to time t. The first step is to solve for the shock process $\mathbf{f}_t$ in terms of the observable quantities, $\ln\left(P_t/P_{t-1}\right)$, $\ln Y_t$ and R_t given $\boldsymbol{\theta}$ and $\mathbf{D}_n$. Since there is no measurement error for inflation, output and the short rate, we have

$$\underbrace{\begin{bmatrix}\ln\left(P_t/P_{t-1}\right)\\ \ln Y_t\end{bmatrix}}_{\mathbf{m}_t}=\underbrace{\begin{bmatrix}\ln\pi^*\\ \ln x^*+\ln A_t\end{bmatrix}}_{\bar{\mathbf{J}}_t}+\underbrace{\begin{bmatrix}h_\pi^a(d_t) & h_\pi^g(d_t) & h_\pi^e(d_t)\\ h_x^a(d_t) & h_x^g(d_t) & h_x^e(d_t)\end{bmatrix}}_{\bar{\mathbf{H}}_{d_t}}\bar{\mathbf{f}}_t \tag{F.0.5}$$

$$= \underbrace{\begin{bmatrix} \ln \pi^* \\ \ln x^* + \ln a^* + \ln A_{t-1} \end{bmatrix}}_{\mathbf{J}_{t-1}} + \underbrace{\begin{bmatrix} h_\pi^a(d_t) & h_\pi^g(d_t) & h_\pi^e(d_t) \\ 1 + h_x^a(d_t) & h_x^g(d_t) & h_x^e(d_t) \end{bmatrix}}_{\tilde{\mathbf{H}}_{d_t}} \bar{\mathbf{f}}_t \tag{F.0.6}$$

and thus

$$\begin{bmatrix} \mathbf{m}_t \\ r_{1t} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{J}}_t \\ \bar{a}_{d_t}(\tau_1) \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{H}}_{d_t} \\ \bar{\mathbf{b}}_{d_t}(\tau_1)' \end{bmatrix} \bar{\mathbf{f}}_t \tag{F.0.7}$$

$$= \begin{bmatrix} \mathbf{J}_{t-1} \\ \bar{a}_{d_t}(\tau_1) \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{H}}_{d_t} \\ \bar{\mathbf{b}}_{d_t}(\tau_1)' \end{bmatrix} \bar{\mathbf{f}}_t \tag{F.0.8}$$

For $t = 0$, the vector of the initial state variables, $\bar{\mathbf{f}}_0$ is straightforwardly calculated by $\mathbf{m}_0$ and r_{10} conditioned on $\ln A_0$ and s_0 where $\mathbf{m}_0$ and r_{10} are observed in the data.

$$\bar{\mathbf{f}}_0 = \begin{bmatrix} \bar{\mathbf{H}}_{s_0} \\ \bar{\mathbf{b}}_{s_0}(\tau_1)' \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{m}_0 \\ r_{10} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{J}}_0 \\ \bar{a}_{s_0}(\tau_1) \end{bmatrix} \right) \tag{F.0.9}$$

For $t = 1, 2, .., n-1$,

$$\mathbf{f}_t = \begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix} \tag{F.0.10}$$

where

$$\bar{\mathbf{f}}_t = \begin{bmatrix} \tilde{\mathbf{H}}_{d_t} \\ \bar{\mathbf{b}}_{d_t}(\tau_1)' \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{m}_t \\ r_{1t} \end{bmatrix} - \begin{bmatrix} \mathbf{J}_{t-1} \\ \bar{a}_{d_t}(\tau_1) \end{bmatrix} \right) \tag{F.0.11}$$

and

$$\ln A_t = \ln A_{t-1} + \ln a^* + \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \bar{\mathbf{f}}_t \tag{F.0.12}$$

Notice that conditioned on $\mathbf{y}_t$, $\bar{\mathbf{f}}_t$(or $\hat{a}_t$) depends on $\ln A_{t-1}$ and d_t, and $\ln A_{t-1} = (t-1)\ln a^* + \sum_{i=1}^{t-1} \hat{a}_i$. Thus $\ln A_{t-1}$ is affected by the path of regime process up to time $(t-1)$. Therefore, in the time updates of $\mathbf{f}_t$ it is very difficult to integrate out the regime path. This is the main reason for sampling $\boldsymbol{\theta}$ conditioned on $\mathbf{D}_n$.

The second step, which is prediction error decomposition, completes the likelihood function conditioned on $\mathbf{D}_n$

$$\ln f(\mathbf{y}|\boldsymbol{\theta},\mathbf{D}_n) = \sum_{t=1}^{n} \ln f[\mathbf{y}_t|I_{t-1},d_t,\boldsymbol{\theta}] \tag{F.0.13}$$

where

$$f[\mathbf{y}_t|I_{t-1},d_t,\boldsymbol{\theta}] = -(2\pi)^{-7/2}\left|\Lambda^{d_t}\right|^{-1/2} \times \exp\left[-\frac{1}{2}\eta_{t|t-1}^{d_t\prime}\left(\Lambda^{d_t}\right)^{-1}\eta_{t|t-1}^{d_t}\right] \tag{F.0.14}$$

$$\mathbf{f}_{t|t-1} = \boldsymbol{\mu} + \mathbf{G}\mathbf{f}_{t-1}$$

$$\eta_{t|t-1}^{d_t} = \mathbf{y}_t - \mathbf{a}_{d_t} - \mathbf{b}_{d_t}\mathbf{f}_{t|t-1}$$

$$\text{and } \Lambda^{d_t} = \mathbf{b}_{d_t}\mathbf{T}_f\Omega_{d_t}\mathbf{T}_f'\mathbf{b}_{d_t}' + \mathbf{T}_y\Sigma_{d_t}\mathbf{T}_y'$$

Step 3 Sampling factors

Conditioned on $\boldsymbol{\theta}$ and $\mathbf{D}_n$, the equations (F.0.9) - (F.0.12) give $\mathbf{F}_n$.

Step 4 Sampling regimes

In this step one samples the states from $p[\mathbf{D}_n|I_n,\boldsymbol{\theta}]$. This is done according to the method of Chib (1996) and Chib (1998) by sampling $\mathbf{D}_n$ in a single block from the output of one forward and backward pass through the data. We remark that in this sampling step, s_n can take any value in {1,2}. For instance, if s_n turns out to be 1 and not 2, then $s_n = 1$ is taken to be the absorbing regime and the parameters of $s_n = 2$ are drawn from the prior in that iteration. In our data, however, $s_n = 2$ is always drawn because the last change-point occurs in the interior of the sample and, therefore, the distribution $\Pr[s_n = 2|I_n,\boldsymbol{\theta}]$ has almost a unit mass on 2.

Bibliography

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

Bansal, R. and Zhou, H. (2002), "Term structure of interest rates with regime shifts," *Journal of Finance*, 57(5), 1997–2043.

Buraschi, A. and Jiltsov, A. (2007), "Habit Formation and Macroeconomic Models of the Term Structure of Interest Rates," *Journal of Finance*, 62(6), 3009–3063.

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 Kang, K. H. (2010), "Change Points in Affine Term-Structure Models: Pricing, Estimation and Forecasting," *Manuscript*.

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.

Davig, T. and Doh, T. (2009), "Monetary policy regime shifts and inflation persistence," The federal reserve bank of Kansas city RWP08-16.

Davig, T. and Leeper, E. M. (2007), "Generalizing the Taylor Principle," *American Economic Review*, 97(3), 607–635.

Doh, T. (2009), "Yield curve in an estimated nonlinear macro model," Federal Reserve Bank of Kansas City RWP 09-04.

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

Duffie, G. and Kan, R. (1996), "A yield-factor model of interest rates," *Mathematical Finance*, 6, 379–406.

Farmer, R. E., Waggoner, D. F., and Zha, T. (2008), "Minimal state variable solutions to Markov-switching rational expectations models," , Federal Reserve Bank of Atlanta Working paper No. 2008–23.

Farmer, R. E., Zha, T., and Waggoner, D. F. (2009), "Understanding Markov-switching rational expectations models," *Journal of Economic Theory*, 144-5, 1849–1867.

Gallmeyer, M., Hollifield, B., Palomino, F., and Zin, S. (2008), "Term premium dynamics and the Taylor rule," *Manuscript*.

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.

Lubik, T. A. and Schorfheide, F. (2004), "Testing for indeterminacy: an application to U.S. monetary policy," *American Economic Review*, 94(1), 190–217.

Moreon, A., Bekaert, G., and Cho, S. (2010), "New-Keynesian Macroeconomics and the Term Structure," *Journal of Money, Credit and Banking*, 42, 33–62.

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 Swanson, E. T. (2008a), "The Bond Premium in a DSGE model with Long-Run and Nominal Risks," Federal Reserve Bank of San Francisco Working Paper.

— (2008b), "Examining the bond premium puzzle with a DSGE model," *Journal of Monetary Economics*, 55, 111–126.

Rudebusch, G. and Wu, T. (2007), "Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models," *Journal of Money Credit and Banking*, 39, 395–422.

— (2008), "A macro-finance model of the term structure, monetary policy and the economy," *Economic Journal*, 118, 906–926.

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

Wu, T. (2006), "Macro Factors and the Affine Term Structure of Interest Rates," *Journal of Money, Credit and Banking*, 38, 1847–1875.