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

Siddhartha Chib†
Kyu Ho Kang‡
Srikanth Ramamurthy§

September 2010

Abstract

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, that the price of regime shift risk is significantly positive over time, that 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 regime, and that the volatility of the technology shock accounts for most of the volatility in the term premium. (JEL C11, E43)
Keywords: Bayesian MCMC method, Change point, Term premium

In this paper we provide a detailed analysis of the term structure dynamics in the context of a dynamic stochastic general equilibrium (DSGE) model. We allow for contemporaneous interaction between the bond markets and the real economy in order to analyze the joint dynamics of the term structure and macroeconomic variables. The model we construct is based on a prototypical New Keynesian 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

*We thank Scott Joslin, Taeyoung Doh, Hong Liu, James Morley, and the participants of the 2010 Western Finance Association meeting for their thoughtful and useful comments on the paper.
†*Address for correspondence*: Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Bookings Drive, St. Louis, MO 63130. E-mail: chib@wustl.edu.
‡*Address for correspondence*: Department of Financial Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suweon, Gyeonggi-do, 443-749, South Korea. E-mail: kyu@kyukang.net.
§*Address for correspondence*: Sellinger School of Business, Loyola University Maryland, 4501 N. Charles St., Baltimore, MD 21210. E-mail: sramamurthy@loyola.edu.

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.

A principal goal of our approach is to characterize the interlinkage between the Central bank's monetary policy and the term-structure dynamics. In particular, we are interested in modeling the impact of monetary policy regime changes on the evolution of bond prices. To this end, we specify the central bank's monetary policy function in terms of the generalized Taylor (1993) rule (Davig and Leeper (2007)) that features multiple policy regimes. Following this rule, the central bank adjusts the nominal short rate in response to deviations of inflation and output from their target levels. As in Davig and Doh (2009) and Bikbov and Chernov (2008), an important aspect of this policy function is that the inflation and output coefficients are time varying to capture the possibility of policy changes between active and less active regimes. We show that the bonds in this setup can be priced once the change point model is reformulated in the manner of Chib (1998). 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.

Another goal of the paper is to investigate the role of the aforementioned structural shocks in determining the risk premium. Because the risk premium is potentially determined by both the size of the structural shocks and the sensitivity of the short rate to these shocks, and through changes in the volatilties of these shocks, we model these volatilties, following Ang, Bekaert, and Wei (2008) and Bikbov and Chernov (2008), by a discrete time Markov switching process.

It is important to note that, while a similar setup can be embedded equally well in a partial equilibrium framework, the distinct advantage of the general equilibrium approach is that the nominal pricing kernel and the no arbitrage conditions are determined within the model through the agents' optimization problem (as opposed to being exogenously specified). As a consequence of the general-equilibrium orientation, the macroeconomic aggregates (namely, output and inflation), are determined within the model. The evolution of these quantities depends in part on monetary policy, on expectations of monetary policy changes, and the structural macro-economic shocks. Because the pricing kernel is a function of these quantities, any change in the policy regime impacts the entire term structure. The full term structure is therefore utilized by agents in forming expectations of policy regime shifts.

The empirical implications of our model are isolated by Bayesian techniques, which in recent years have become central for the analysis of DSGE models (Fernandez-Villaverde and Rubio-Ramirez (2009), An and Schorfheide (2007), Smets and Wouters (2007)). Despite the complex nature of the likelihood/posterior surface, our fitting method, which is based on the MCMC simulation methods in Chib and Ergashev (2009) and Chib and Ramamurthy (2010), is efficient in terms of the metrics that are used to evaluate MCMC procedures. Further, our inference is based on priors that reflect the assumption of a positive term-premium (Chib and Ergashev (2009)).

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 Ludvigson and Ng (2009), 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 regime 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 1 we develop the model, discuss the solution procedure and derive the bond prices. Section 2 provides the econometric details and Section 3 contains the empirical results. Concluding remarks are in

Section 4.

1 Model

In this section we discuss the key aspects of our 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 structural 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 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.

1.1 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

$$P_tC_t+\sum_{\tau=1}^{\tau^*}P_t^\tau B_t^\tau+T_t\leq P_tW_tH_t+\sum_{\tau=1}^{\tau^*-1}P_t^\tau B_{t-1}^{\tau+1}+B_{t-1}^1+P_tQ_t. \tag{1.1}$$

The household then maximizes her expected utility function [1]

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} \delta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\gamma} - 1}{1-\gamma} - H_{t+s} \right) \right] \tag{1.2}$$

subject to the intertemporal budget constraint (1.1) 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. Alternatively, preferences could display habit persistence (modeled through a lagged consumption variable), as in Buraschi and Jiltsov (2007), Ludvigson and Ng (2009) 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. We assume that the growth rate of technology $a_t = A_t/A_{t-1}$ follows an autoregressive process

$$\ln a_t = (1 - \phi_a) \ln a^* + \phi_a \ln a_{t-1} + \varepsilon_{a,t} \tag{1.3}$$

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 (Hamilton, 1989, Albert and Chib, 1993, Fruhwirth-Schnatter, 2006). 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

$$\mathbf{Q}^a = \begin{bmatrix} q^a_{11} & 1 - q^a_{11} \\ 1 - q^a_{22} & q^a_{22} \end{bmatrix} \tag{1.4}$$

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

[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).

1.2 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

$$\left(\int_0^1 Y_t(j)^{\frac{\zeta-1}{\zeta}} dj \right)^{\frac{\zeta}{\zeta-1}} \geq Y_t \tag{1.5}$$

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

$$P_t(j) = \left(\frac{Y_t}{Y_t(j)} \right)^{1/\zeta} P_t. \tag{1.6}$$

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

$$P_t = \left(\int_0^1 P_t(j)^{1-\zeta} dj \right)^{\frac{1}{1-\zeta}}. \tag{1.7}$$

1.3 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

$$Y_t(j) = A_t N_t(j). \tag{1.8}$$

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

$$AC_t(j) = \frac{\varphi}{2} \left(\frac{P_t(j)}{\pi^* P_{t-1}(j)} - 1 \right)^2 Y_t \tag{1.9}$$

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

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} \Lambda_{t,t+s} Q_t(j) \right] \tag{1.10}$$

where the real profit at time t is

$$Q_t(j) = \frac{P_t(j)}{P_t} Y_t(j) - W_t N_t(j) - \frac{\varphi}{2} \left(\frac{P_t(j)}{\pi^* P_{t-1}(j)} - 1 \right)^2 Y_t \tag{1.11}$$

and

$$\Lambda_{t,t+s} = \delta^s \left(\frac{C_{t+s}}{A_{t+s}} \right)^{-\gamma} \left(\frac{C_t}{A_t} \right)^{\gamma} \frac{A_t}{A_{t+s}} \tag{1.12}$$

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

1.4 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

$$P_t G_t + \sum_{\tau=1}^{\tau^*-1} P_t^{\tau} B_{t-1}^{\tau+1} + B_{t-1}^{1} = T_t + \sum_{\tau=1}^{\tau^*} P_t^{\tau} B_t^{\tau} \tag{1.13}$$

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

$$\ln g_t = (1 - \phi_g) \ln g^* + \phi_g \ln g_{t-1} + \varepsilon_{g,t} \tag{1.14}$$

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

$$\mathbf{Q}^g = \begin{bmatrix} q_{11}^g & 1 - q_{11}^g \\ 1 - q_{22}^g & q_{22}^g \end{bmatrix} \tag{1.15}$$

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

1.5 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_t^1 has the form

$$P_t^1 = \mathbb{E}_t \left[M_{t,t+1} \right] \tag{1.16}$$

where

$$M_{t,t+1} = \delta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \frac{1}{a_{t+1}} \frac{1}{\pi_{t+1}} \tag{1.17}$$

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 1.9 to price bonds of various maturities.

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

$$1 = \frac{W_t}{A_t} c_t^{-\gamma} \tag{1.18}$$

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

$$1 = \zeta - \zeta c_t^{\gamma} + \varphi \left(\frac{\pi_t}{\pi^*} - 1 \right) \left(\frac{\pi_t}{\pi^*} \right) - \varphi \mathbb{E}_t \left[\Lambda_{t,t+1} \left(\frac{\pi_{t+1}}{\pi^*} - 1 \right) \left(\frac{\pi_{t+1}}{\pi^*} \frac{Y_{t+1}}{Y_t} \right) \right] \tag{1.19}$$

Finally, the aggregate resource constraint must hold in equilibrium:

$$Y_t = C_t + G_t + AC_t \text{ and } H_t = N_t = \int_0^1 N_t(j) dj \tag{1.20}$$

which implies that

$$c_t = \left(\frac{1}{g_t} - \frac{\varphi}{2} \left(\frac{\pi_t}{\pi^*} - 1 \right)^2 \right) x_t \tag{1.21}$$

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

$$R^* = a^* \pi^* / \delta \tag{1.22}$$

Also the equation (1.21) implies that the nonstochastic value of the detrended output is determined by

$$x^* = \frac{c^*}{(1-\rho^*)} \tag{1.23}$$

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

$$\left[\frac{\zeta-1}{\zeta}\right]^{\frac{1}{\gamma}} \tag{1.24}$$

In the absence of shocks, the economy converges to a steady-state growth path along which all the stationary variables are constant over time. It is important to note that in this setup while the steady state values of the aggregated macroeconomic variables (namely inflation, output and the risk-free short rate) are not affected by regime shifts, the steady state values of the long term bond yields are regime specific. As we show in Section (1.10), this is because the term premium is a function of the monetary policy reaction coefficients and the volatilities, both of which are subject to regime shifts.

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 (1.16), (1.19) and (1.20) can be cast in its log-linearized form as follows

$$\hat{\pi}_t = \delta \mathbb{E}_t\left[\hat{\pi}_{t+1}\right] + \kappa \hat{c}_t \text{ with } \kappa = \frac{\zeta\gamma\left(c^*\right)^{-\gamma}}{\varphi} \tag{1.25}$$

$$\hat{c}_t = \mathbb{E}_t\left[\hat{c}_{t+1}\right] - \frac{1}{\gamma}\left(\hat{R}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right] - \mathbb{E}_t\left[\hat{a}_{t+1}\right]\right) \tag{1.26}$$

$$\hat{c}_t = \hat{x}_t - \hat{g}_t. \tag{1.27}$$

1.6 The Central Bank

Given that an important focus of this paper is to analyze the impact of monetary policy regime changes on the dynamics of the bond prices, we model the central bank's policy function following the generalized Taylor (1993) rule (Davig and Leeper (2007)). 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^*

$$\ln R_t = \ln R^* + \alpha_{s_t}\left(\ln \pi_t - \ln \pi^*\right) + \beta_{s_t}\left(\ln x_t - \ln x^*\right) + \ln e_t. \tag{1.28}$$

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

$$\hat{R}_t = \alpha_{s_t}\hat{\pi}_t + \beta_{s_t}\hat{x}_t + \hat{e}_t \tag{1.29}$$

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

$$\hat{e}_t = \phi_e \hat{e}_{t-t} + \varepsilon_{e,t} \tag{1.30}$$

with $\varepsilon_{e,t} \sim \mathcal{N}(0, \sigma^2_{e,v_{t,e}})$ where the volatility $\sigma^2_{e,v_{t,e}}$ 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

$$\mathbf{Q}^e = \begin{bmatrix} q^e_{11} & 1 - q^e_{11} \\ 1 - q^e_{22} & q^e_{22} \end{bmatrix} \tag{1.31}$$

where $q^e_{ij} = \Pr[v_{t+1,e} = j | v_{t,e} = i]$. It is important to clarify that changes in the volatility of the policy shock $\sigma^2_{e,v_{t,e}}$ are distinct from the monetary policy regime change. The volatility regime shift captures the possibility of heteroskedasticity of the short rate process.

Notice that in the above short rate equation the target inflation is assumed to be constant over time whereas the monetary policy coefficients α and β are regime dependent, as indicated by the subscript s_t.[2] We interpret the regime dependency of the monetary policy coefficients as shifts between relatively more active and less active regimes. A convenient way to model this is to assume a change point process for the regimes. Following Chib (1998), we characterize this as a unidirectional, discrete time Markov process in which (at time $t+1$) the state s_{t+1} can either stay at the current state j with probability p_{jj} or jump to the next state $k = j+1$ with probability p_{jk}. Transitions to states occupied in the past or to states beyond $j+1$ cannot occur at time $t+1$.

[2] In the empirical counterpart of this paper we deal with the time span since the great moderation - a period of relatively low and stable inflation. We therefore attribute the variation in inflation to inflation gap (rather than its trend) by assuming a constant target inflation. As we show below, the virtue of this simplifying assumption is that it allows us to isolate the effect of all monetary policy regime changes solely through changes in the reaction coefficients of inflation and output gaps. In contrast, if one were to analyze a longer time period including the 1960's and 1970's, regime shifts in the target inflation might be essential as in Schorfheide (2005), Moreon, Bekaert, and Cho (2010), Cogley and Sbordone (2008) and Davig and Doh (2009).

Thus, 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) and Clarida, Gali, and Gertler (2000). Agents face this regime uncertainty ad-infinitum and, therefore, the possibility of a infinite number of change-points. The model and the bond prices are solved under this assumption.

When the model is estimated, however, with a finite amount of data, and the objective is to date the change-points, the number of change-points must be finite. In that case, following Chib (1998), we assume a certain number of change-points and assume that the last change-point state is absorbing to ensure that there are at most that many change-points in the sample. We select the number of change-points by estimating the model with different number of change-points and selecting the model that is most supported by the data from a marginal likelihood/Bayes factor perspective. We now turn to a summary of the regime processes.

1.7 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

$$\bar{\mathbf{f}}_t = \begin{bmatrix} \hat{a}_t \\ \hat{g}_t \\ \hat{e}_t \end{bmatrix} = \phi \bar{\mathbf{f}}_{t-1} + \varepsilon_t \tag{1.32}$$

where

$$\phi = \begin{bmatrix} \phi_a & 0 & 0 \\ 0 & \phi_g & 0 \\ 0 & 0 & \phi_e \end{bmatrix}, \text{ and } \varepsilon_t = \begin{bmatrix} \varepsilon_{a,t} \\ \varepsilon_{g,t} \\ \varepsilon_{e,t} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1}, \Omega_{v_t} = \begin{bmatrix} \sigma^2_{a,v_{t,a}} & & \\ & \sigma^2_{g,v_{t,g}} & \\ & & \sigma^2_{e,v_{t,e}} \end{bmatrix} \right).$$

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 1.10, 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.

1.8 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. Our interest lies in the linearized system of equations (1.25)-(1.29) and (1.32). On substituting (1.27) and (1.29) into (1.26), this system collapses to

$$0 = \delta \mathbb{E}_t \left[\hat{\pi}_{t+1}\right] - \hat{\pi}_t + \kappa \left(\hat{x}_t - \hat{g}_t\right) \tag{1.33}$$

$$0 = \mathbb{E}_t \left[\hat{\pi}_{t+1}\right] + \gamma \mathbb{E}_t \left[\hat{x}_{t+1}\right] - \alpha_{s_t} \hat{\pi}_t - (\beta_{s_t} + \gamma)\hat{x}_t + \phi_a \hat{a}_t - \gamma(\phi_g - 1)\hat{g}_t - \hat{e}_t \tag{1.34}$$

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. For this purpose, we adopt the solution method of Davig and Leeper (2007). The solution process rids the system of the unobservable expectational terms by casting them as a linear function of the underlying shock processes. In this paper we restrict our attention to the unique (determinate) solution.[3] A full discussion of the solution algorithm is well beyond the scope of this paper. In terms of the computational details, we begin by casting the endogenous variables as a linear function of the shock processes

$$\underbrace{\begin{bmatrix} \hat{\pi}_{it} \\ \hat{x}_{it} \end{bmatrix}}_{\hat{\mathbf{m}}_{it}} = \underbrace{\begin{bmatrix} h_{\pi}^{a}(s_t = i) & h_{\pi}^{g}(s_t = i) & h_{\pi}^{e}(s_t = i) \\ h_{x}^{a}(s_t = i) & h_{x}^{g}(s_t = i) & h_{x}^{e}(s_t = i) \end{bmatrix}}_{\bar{\mathbf{H}}_{s_t=i}} \bar{\mathbf{f}}_t \tag{1.35}$$

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 (1.33)-(1.34), the conditional expectation of the one-period ahead inflation gap and output gap are

$$\begin{aligned} \mathbb{E}_t \left[\begin{pmatrix} \hat{\pi}_{t+1} & \hat{x}_{t+1} \end{pmatrix}' | s_t = i \right] &= \mathbb{E}_t \left[\bar{\mathbf{H}}_{s_{t+1}} \bar{\mathbf{f}}_{t+1} | s_t = i \right] \\ &= p_{i1} \bar{\mathbf{H}}_{s_{t+1}=1} \phi \bar{\mathbf{f}}_t + p_{i2} \bar{\mathbf{H}}_{s_{t+1}=2} \phi \bar{\mathbf{f}}_t \end{aligned} \tag{1.36}$$

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

$$p_{i1} \left[h_{\pi,1}^{a} \phi_a \hat{a}_t + h_{\pi,1}^{g} \phi_g \hat{g}_t + h_{\pi,1}^{e} \phi_e \hat{e}_t \right] + p_{i2} \left[h_{\pi,2}^{a} \phi_a \hat{a}_t + h_{\pi,2}^{g} \phi_g \hat{g}_t + h_{\pi,2}^{e} \phi_e \hat{e}_t \right] \tag{1.37}$$

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

$$p_{i1} \left[h_{x,1}^{a} \phi_a \hat{a}_t + h_{x,1}^{g} \phi_g \hat{g}_t + h_{x,1}^{e} \phi_e \hat{e}_t \right] + p_{i2} \left[h_{x,2}^{a} \phi_a \hat{a}_t + h_{x,2}^{g} \phi_g \hat{g}_t + h_{x,2}^{e} \phi_e \hat{e}_t \right] \tag{1.38}$$

Next, to compute the regime-dependent solutions $\bar{\mathbf{H}}_{s_t}$, one relies 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 $\bar{\mathbf{H}}_{\mathbf{s_t}}$. Additional computational details of the solution are provided in Appendix A.

Note that because we work with a first-order approximation of the equilibrium conditions of the households and firms, the solution coefficients $\bar{\mathbf{H}}_{s_t}$ depend only on the

[3]Farmer, Zha, and Waggoner (2009) show the existence of general forms of indeterminate equilibria in the quasi-linear system that depend not only on the structural shocks, but also on additional autoregressive shock driven by the structural shocks. The general forms include Davig and Leeper (2007) solutions as special cases.

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 (1.35) can be rewritten as

$$\underbrace{\begin{bmatrix} \ln \pi_t \\ \ln Y_t \end{bmatrix}}_{\mathbf{m}_t} = \underbrace{\begin{bmatrix} \ln \pi^* \\ \ln x^* \end{bmatrix}}_{\mathbf{J}} + \underbrace{\begin{bmatrix} h_\pi^a(d_t = i) & h_\pi^g(d_t = i) & h_\pi^e(d_t = i) & 0 \\ h_x^a(d_t = i) & h_x^g(d_t = i) & h_x^e(d_t = i) & 1 \end{bmatrix}}_{\mathbf{H}_{d_t = i}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} \tag{1.39}$$

As can be seen in section 2.1 below, this representation of the solution turns out to be very convenient in the construction of the empirical model.

It is important to note 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.

1.9 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

$$P_t^\tau = \mathbb{E}_t \left[M_{t,t+\tau} \right] \tag{1.40}$$

where

$$M_{t,t+\tau} = \delta \left(\frac{c_{t+\tau}}{c_t} \right)^{-\gamma} \frac{1}{a_{t+\tau}} \frac{1}{\pi_{t+\tau}} \tag{1.41}$$

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

$$P_t^\tau = \mathbb{E}_t \left[M_{t,t+1} \times M_{t+1,t+\tau} \right] \tag{1.42}$$

$$= \mathbb{E}_t \left[M_{t,t+1} \times \mathbb{E}_{t+1} \left[M_{t+1,t+\tau} \right] \right]$$
$$= \mathbb{E}_t \left[M_{t,t+1} \times P_{t+1}^{\tau-1} \right]$$

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

$$P_{d_t,t}^{(\tau)} = \mathbb{E}\left[M_{t,t+1} P_{d_{t+1},t+1}^{(\tau-1)} | \bar{\mathbf{f}}_t, d_t \right] \tag{1.43}$$

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

$$\ln M_{t,t+1} = m_{t,t+1} \approx c_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}} \bar{\mathbf{f}}_t + \mathbf{L}_{d_{t+1}} \varepsilon_{t+1} \tag{1.44}$$

where

$$c_{d_{t+1}} = -\ln R^* - \frac{1}{2} \mathbf{L}_{d_{t+1}} \Omega_{d_{t+1}} \mathbf{L}_{d_{t+1}}' \tag{1.45}$$

$$\boldsymbol{\lambda}_{d_t,d_{t+1}} = -\begin{pmatrix} 1 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_{t+1}} \phi + \begin{pmatrix} 0 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_t} + \begin{pmatrix} -1 & \gamma & 0 \end{pmatrix} \phi - \begin{pmatrix} 0 & \gamma & 0 \end{pmatrix} \tag{1.46}$$
$$\mathbf{L}_{d_{t+1}} = -\begin{pmatrix} 1 & \gamma \end{pmatrix} \bar{\mathbf{H}}_{d_{t+1}} + \begin{pmatrix} -1 & \gamma & 0 \end{pmatrix} \tag{1.47}$$

Following Ang et al. (2008), we assume that the one period bond is risk-free by augmenting the Jensen's inequality term to equation (1.45). 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

$$-p_{d_t,t}^{(\tau)} = a_{d_t}(\tau) + \mathbf{b}_{d_t}(\tau)' \bar{\mathbf{f}}_t. \tag{1.48}$$

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 $i \in \{1, 2, .., \mathbf{d}\}$

$$a_i(\tau) = \ln R^* + \sum_{j=1}^{\mathbf{d}} p_{ij} \left(a_j(\tau - 1) + \mathbf{L}_j \Omega_j \mathbf{b}_j(\tau - 1)' - \frac{1}{2} \mathbf{b}_j(\tau - 1)' \Omega_j \mathbf{b}_j(\tau - 1) \right) \tag{1.49}$$

$$\mathbf{b}_i(\tau)' = \sum_{j=1}^{\mathbf{d}} p_{ij} \left(\mathbf{b}_j(\tau - \mathbf{1})'\phi - \boldsymbol{\lambda}_{i,j} \right). \tag{1.50}$$

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

$$a_i(0) = \mathbf{b}_i(0) = 0 \text{ for all } i \tag{1.51}$$

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

$$r_{d_t,t}^{(\tau)} = \frac{-p_{d_t,t}^{\tau}}{\tau} = \bar{a}_{d_t}(\tau) + \bar{\mathbf{b}}_{d_t}(\tau)'\bar{\mathbf{f}}_t \tag{1.52}$$
$$\text{with } \bar{a}_{d_t}(\tau) = \frac{a_{d_t}(\tau)}{\tau} \text{ and } \bar{\mathbf{b}}_{d_t}(\tau) = \frac{\mathbf{b}_{d_t}(\tau)}{\tau}.$$

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 (1.44).

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.

1.10 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

$$r_{d_t,t}^{(\tau)} - r_{d_t,t}^{(1)} = \underbrace{\left[\frac{1}{\tau} \sum_{l=0}^{\tau-1} \mathbb{E}_t \left[r_{d_{t+l},t+l}^{(1)} \right] - r_{d_t,t}^{(1)} \right]}_{\text{EH}} + \underbrace{\frac{1}{\tau} \sum_{i=1}^{\tau-1} \text{exr}_{d_t,t}^{(\tau+1-i)}}_{\text{Term Premium}}, \tag{1.53}$$

where $\text{exr}^{(\tau)}_{d_t,t}$ 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 (1.53). 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

$$\begin{aligned}\text{exr}^{(\tau)}_{d_t,t} &= \left[\mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] - p^{(\tau)}_{d_t,t}\right] - (-p^{(1)}_{d_t,t}) \\ &= \mathbb{E}_t\left[-(\tau-1)\, r^{(\tau-1)}_{d_{t+1},t+1} + \tau r^{(\tau)}_{d_t,t}\right] - r^{(1)}_{d_t,t}\end{aligned} \tag{1.54}$$

The first term on the right side of (1.54) 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 $\text{RS}^{(\tau)}_{d_t=i,t}$

$$\text{exr}^{(\tau)}_{d_t=i,t} = \text{FR}^{(\tau)}_{d_t=i} + \text{RS}^{(\tau)}_{d_t=i,t} \tag{1.55}$$

where

$$\text{FR}^{(\tau)}_{d_t=i} = \sum_{j=1}^{\mathbf{d}} p_{ij}\mathbf{L}_j\Omega_j\mathbf{b}_j(\tau-1) - \frac{1}{2}\sum_{j=1}^{\mathbf{d}} p_{ij}\mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) \tag{1.56}$$

$$\begin{aligned}\text{RS}^{(\tau)}_{d_t=i,t} = &\left[\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right]\left[\sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}\right] - \sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}K_{j,t} \\ &- \frac{1}{2}\sum_{j=1}^{\mathbf{d}} p_{ij}K^2_{j,t} + \frac{1}{2}\left(\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right)^2\end{aligned} \tag{1.57}$$

and

$$W_{d_t,d_{t+1},t} = c_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}}\overline{\mathbf{f}}_t \tag{1.58}$$

$$K_{d_{t+1},t} = -a_{d_{t+1}} - \mathbf{b}_{d_{t+1}}(\tau - 1)'\phi\bar{\mathbf{f}}_t$$

Similarly, it is straightforward to decompose the term premium, denoted by $\text{TP}^{(\tau)}_{d_t=i,t}$, in equation (1.53) as the sum of two averages.

The proof of these results is given in Appendix C. Notice that the terms in the factor risk component $\text{FR}^{(\tau)}_{d_t=i}$ 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 $\text{RS}^{(\tau)}_{d_t=i,t}$ 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, Singleton, and Yang (2007). Finally, it is interesting that $\text{FR}^{(\tau)}_{d_t=i}$ is a regime-specific constant, whereas $\text{RS}^{(\tau)}_{d_t=i,t}$ 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).

2 Estimation methodology

2.1 State Space Formulation

We begin by recalling the solution to the DSGE model in equation (1.39)

$$\underbrace{\begin{bmatrix} \ln \pi_t \\ \ln Y_t \end{bmatrix}}_{\mathbf{m}_t} = \underbrace{\begin{bmatrix} \ln \pi^* \\ \ln x^* \end{bmatrix}}_{\mathbf{J}} + \underbrace{\begin{bmatrix} h^a_\pi(d_t=i) & h^g_\pi(d_t=i) & h^e_\pi(d_t=i) & 0 \\ h^a_x(d_t=i) & h^g_x(d_t=i) & h^e_x(d_t=i) & 1 \end{bmatrix}}_{\mathbf{H}_{d_t=i}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} \quad (2.1)$$

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 stan-

[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.

dard DSGE models, we assume that the final outcomes $(\mathbf{m}_t, \hat{R}_t)$ are generated without additional (measurement) errors. As we show in Appendix D, 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

$$\underbrace{\begin{bmatrix} \mathbf{m}_t \\ \mathbf{R}_t \end{bmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{bmatrix} \mathbf{J} \\ \bar{\mathbf{a}}_{d_t} \end{bmatrix}}_{\mathbf{a}_{d_t}} + \underbrace{\begin{bmatrix} \mathbf{H}_{d_t} & \\ \bar{\mathbf{b}}_{d_t} & \mathbf{0}_{5\times 1} \end{bmatrix}}_{\mathbf{b}_{d_t}} \mathbf{f}_t + \underbrace{\begin{bmatrix} \mathbf{0}_{3\times 4} \\ \mathbf{I}_4 \end{bmatrix}}_{\mathbf{T}_y} \mathbf{e}_t \tag{2.2}$$

where $\mathbf{e}_t \sim \mathcal{N}_4(\mathbf{0}, \mathbf{\Sigma})$; $\mathbf{\Sigma} = \mathrm{diag}(\sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2)$. We complete the state space formulation by combining equation (1.32) with the technology shock process $\ln A_t = \ln a^* + \ln A_{t-1} + \hat{a}_t$ and write the transition equation as

$$\underbrace{\begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix}}_{\mathbf{f}_t} = \underbrace{\begin{bmatrix} \mathbf{0}_{3\times 1} \\ \ln a^* \end{bmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{bmatrix} & \boldsymbol{\phi}_{3\times 3} & & \mathbf{0}_{3\times 1} \\ \phi_a & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \bar{\mathbf{f}}_{t-1} \\ \ln A_{t-1} \end{bmatrix}}_{\mathbf{f}_{t-1}} + \underbrace{\begin{bmatrix} & \mathbf{I}_3 & \\ 1 & 0 & 0 \end{bmatrix}}_{\mathbf{T}_f} \boldsymbol{\varepsilon}_t \tag{2.3}$$

with $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \mathbf{\Omega}_{d_t})$. 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 $\mathbf{\Omega}_{d_t}$.

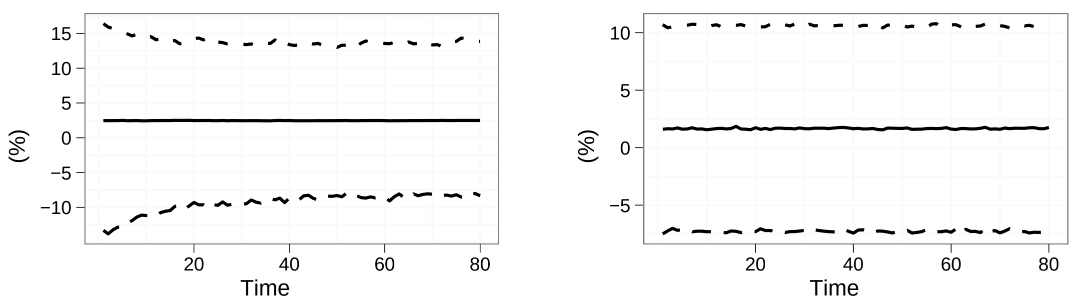
2.2 Prior Distribution

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 I 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^a_{11}	beta	0.8995	0.0401
q^a_{22}	beta	0.8997	0.0401
q^g_{11}	beta	0.8997	0.0401
q^g_{22}	beta	0.8997	0.0401
q^e_{11}	beta	0.8997	0.0401
q^e_{22}	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
$2.0 \times 10^4 \times \sigma^2_{a,1}$	inverse gamma	0.9539	0.1895
$2.0 \times 10^5 \times \sigma^2_{g,1}$	inverse gamma	0.9596	0.1937
$3.0 \times 10^4 \times \sigma^2_{e,1}$	inverse gamma	0.9603	0.1951
$1.0 \times 10^4 \times \sigma^2_{a,2}$	inverse gamma	0.9635	0.1941
$1.0 \times 10^5 \times \sigma^2_{g,2}$	inverse gamma	0.9635	0.1956
$2.5 \times 10^3 \times \sigma^2_{e,2}$	inverse gamma	0.9613	0.1943
$1.4 \times 10^7 \times \sigma^2_2$	inverse gamma	0.9623	0.1927
$3.0 \times 10^6 \times \sigma^2_3$	inverse gamma	0.9620	0.1948
$1.2 \times 10^6 \times \sigma^2_4$	inverse gamma	0.9605	0.1942
$6.0 \times 10^5 \times \sigma^2_5$	inverse gamma	0.9611	0.1979

Table I: 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 $\mathbf{\Omega_d}$ are greater than those in $\mathbf{\Omega_1}$. 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 1: The prior-implied inflation and output growth dynamics *These graphs are based on 50,000 simulated draws of the parameters from the prior distribution. In the graphs on the left, the surfaces correspond to the 2.5%, 50%, and 97.5% quantile surfaces of the term structure dynamics in annualized percents implied by the prior distribution for each regime.*

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

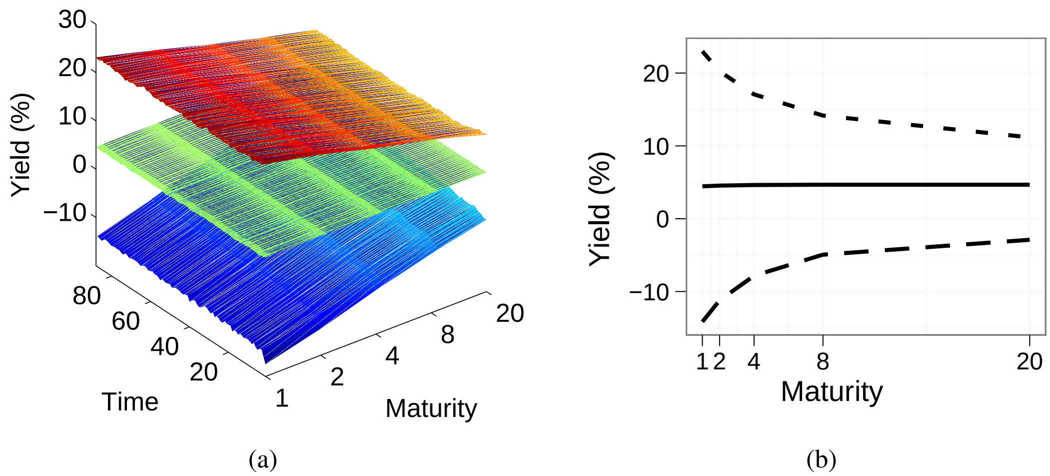


Figure 2: The prior-implied term structure dynamics *These graphs are based on 50,000 simulated draws of the parameters from the prior distribution. In the graphs on the left, the surfaces correspond to the 2.5%, 50%, and 97.5% quantile surfaces of the term structure dynamics in annualized percents implied by the prior distribution for each regime.*

model. The sampled sequences for each macroeconomic variable in annualized percents are shown in Figure 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 in annualized percents for each time period is reproduced in Figure 2. As one can see, the implied average term structure is gently upward sloping in each regime with considerable a priori variation.

2.3 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

$$\pi(\boldsymbol{\theta}, \mathbf{F}_n, \mathbf{D}_n|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{F_n}, \mathbf{D}_n)p(\mathbf{F}_n, \mathbf{D}_n|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (2.4)$$

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

$$f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}_n)\pi(\boldsymbol{\theta}) \tag{2.5}$$

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 Appendix D.

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.

2.4 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

$$\ln \hat{m}(\mathbf{y}|\mathcal{M}_{\mathbf{d}}) = \ln f(\mathbf{y}|\boldsymbol{\theta}^*, \mathcal{M}_{\mathbf{d}}) + \ln \pi\left(\boldsymbol{\theta}^*, \mathcal{M}_{\mathbf{d}}\right) - \ln \hat{\pi}(\boldsymbol{\theta}^*|\mathbf{y},\mathcal{M}_{\mathbf{d}}) \tag{2.6}$$

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.

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

3 Results

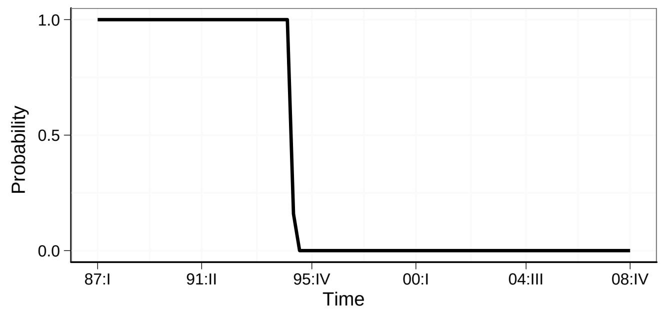
Our empirical results are based on the collection of historical yields of treasury bills with maturities 1, 2, 4, 8 and 20 quarters, real GDP per capita and inflation for the sample period 1986:Q4 to 2008:Q4. The inflation is calculated as a quarterly decimal change in the GDP deflator. This data is available online from the Board of Governors of the Federal Reserve System (Gurkaynak, Sack, and Wright (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.

3.1 Change-Point and Structural Shocks

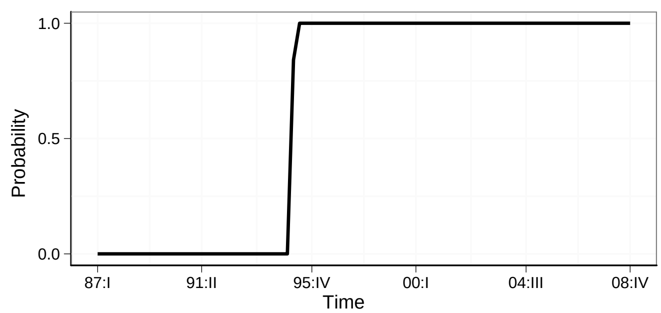
Table II unambiguously confirms the presence of a regime shift in monetary policy dated around the second quarter of 1995. In particular, it is interesting to note that both models $\mathcal{M}_2$, that focuses purely on just the regime change in monetary policy without any regime shifts in the volatilities, and $\mathcal{M}_4$ which incorporates simultaneous shifts in both policy and volatility regimes, provide the same estimate of the breakpoint as model $\mathcal{M}_{16}$, which is the model with independent policy and volatility regimes. Based on marginal likelihoods, however, it is clear that the best fitting model is $\mathcal{M}_{16}$ that edges out even the model with 3 policy regimes.

As mentioned earlier, in this general equilibrium setup, both the structural shocks and the policy reaction coefficients drive output, inflation and the term premium dynamics, which is distinct from that in a partial equilibrium approach. This points to the fundamental notion that the macroeconomic fundamentals and the entire term structure, not just the short-term rate, contain valuable information about monetary policy regime shifts. This distinction also helps explain why our finding of the breakpoint is different from that in Ang, Boivin, Dong, and Loo-Kung (2010) and Bikbov and Chernov (2008).

Figure 3 shows the persistence of the policy regimes. In contrast, figures 4 and 5 reveal that the volatility regimes are far less persistent than the policy regimes. Finally, Figure 7 plots the estimated exogenous shock processes $\hat{a}_t$, $\hat{g}_t$ and $\hat{e}_t$. The coincidence



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



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

Figure 3: Model $\mathcal{M}_{16}$: The posterior probability of monetary policy regimes *These graphs are based on 20,000 simulated draws of the posterior simulation.*

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

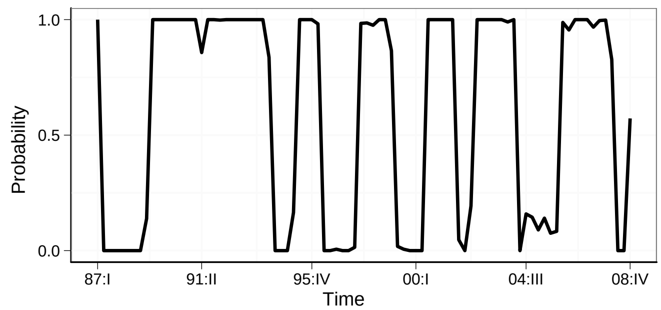
3.2 Model Parameters

We next discuss the posterior estimates of the parameters. Table III 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

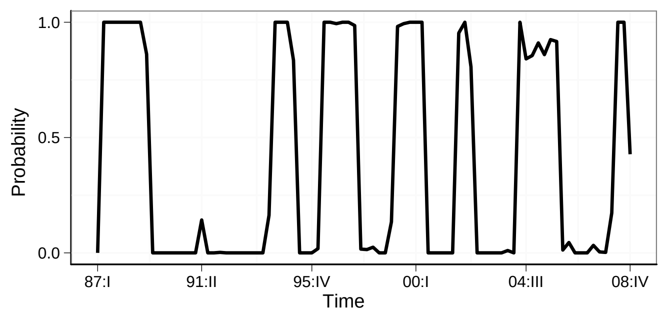
[5]The inefficiency factors approximate the ratio of the numerical variance of the estimate from the MCMC chain relative to that from hypothetical iid draws. For a given sequence of draws the inefficiency factor is computed as

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

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 windowing procedure, but we ignore this aspect for simplicity). A well mixing sampler results in autocorrelations



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



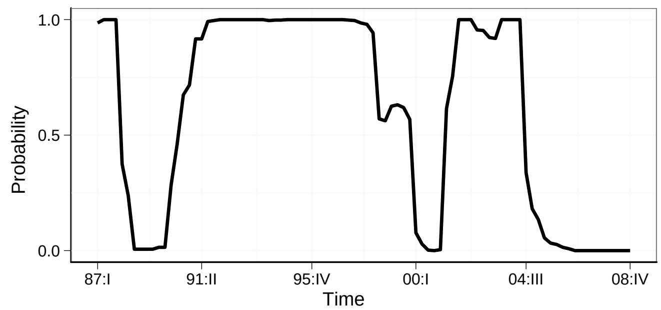
(b) High technology volatility regime ($v_t^a = 2$)

Figure 4: Model $\mathcal{M}_{16}$: The posterior probability of technology volatility regimes
These graphs are based on 20,000 simulated draws of the posterior simulation.

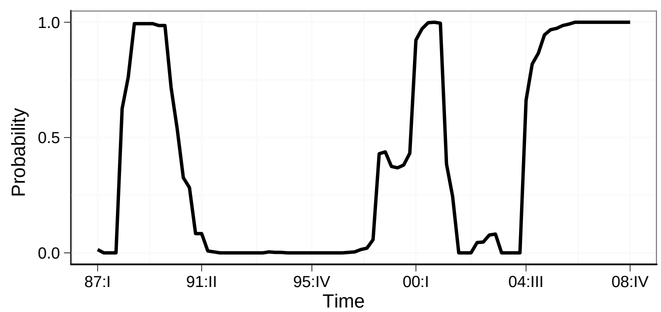
on average are 62.3% and 28.3, respectively, indicating a well mixing, efficient sampler. Also, the sampler converges quickly to the same region of the parameter space regardless of the starting values. Finally, as one can see in Figure 8, the posterior densities of the parameters are mostly different from the prior given in Table I. 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. 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

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



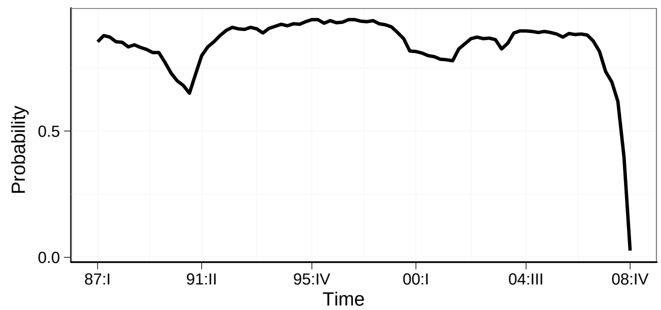
(b) High government expenditure volatility regime ($v_t^g = 2$)

Figure 5: Model $\mathcal{M}_{16}$: The posterior probability of government expenditure volatility regimes *These graphs are based on 20,000 simulated draws of the posterior simulation.*

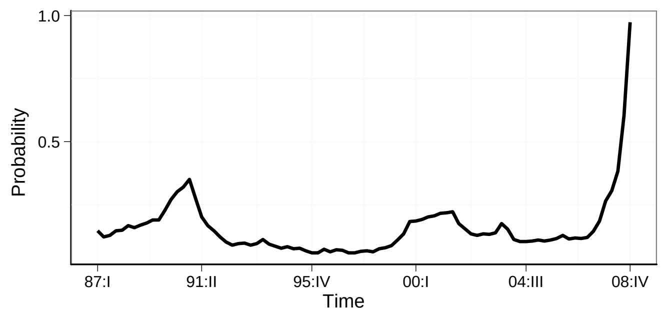
inflation has been reasonably stable during the sample period, the Fed's reaction to output gap became relatively more aggressive, marking the break point.

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". 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]

[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,



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



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

Figure 6: Model $\mathcal{M}_{16}$: The posterior probability of monetary policy volatility regimes *These graphs are based on 20,000 simulated draws of the posterior simulation.*

3.3 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 1.10. Figure 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 IV also reveals that, regardless of the maturity, the average term spread is noticeable lower in the recent policy regime than in policy regime 1. Recall that this regime-dependence of the bond risk is solely attributed by the change in the

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.

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 II: Log likelihood (lnL), log marginal likelihood (lnML), numerical standard error(n.s.e) and change point estimates

reaction coefficients. This implies that a more active regime on average generates a flatter yield curve. A 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.

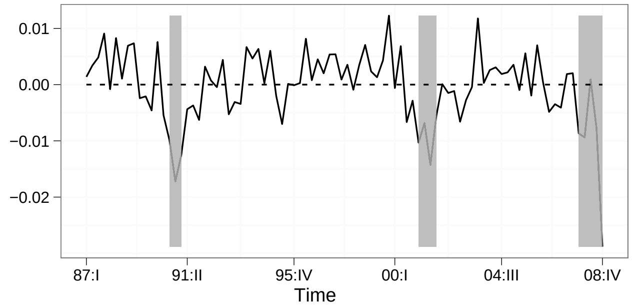
On the other hand, Figure 10 presents the result for the decomposition of the term premium of the 5-year bond over time. Interestingly, most of variation of the term premium is explained by the factor risk component. One possible explanation is that sizable factor shocks occur frequently whereas regime shifts happen relatively less frequently. Nevertheless, because the regime shift risk component is consistently positive over time, it should not be neglected.

Figure 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.

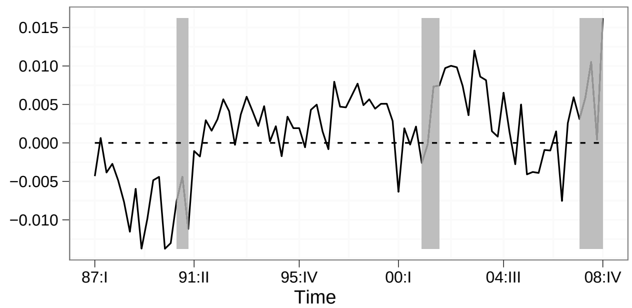
3.4 Counterfactual Analysis

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 macro-economy and the asset prices.

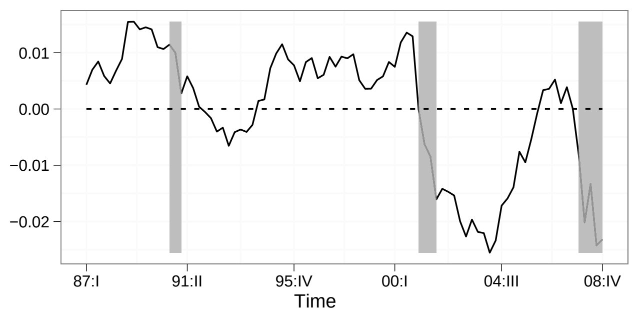
Figure 12 plots the results for the short rate and the term spread. As seen in the figure, the short rate would have been more volatile and the slope of the yield curve steeper without the break. On the contrary, if the more active regime prevailed over



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



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



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

Figure 7: Model $\mathcal{M}_{16}$: The exogenous shock process *These graphs are based on 20,000 simulated draws of the posterior simulation. NBER recessions are shaded*

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 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 14 indicates that inflation and the output growth exhibit little

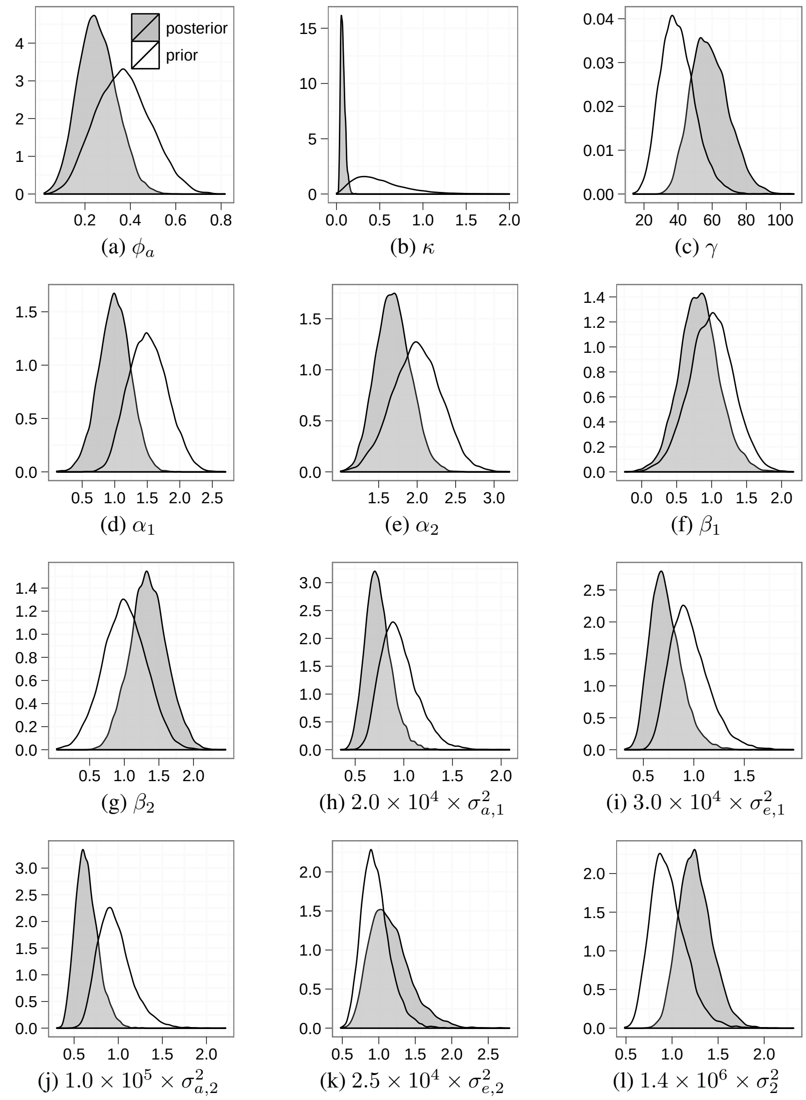


Figure 8: Model $\mathcal{M}_{16}$: Marginal prior-posterior plots for some selected parameters
These graphs are based on 20,000 simulated draws of the posterior simulation.

difference, no matter what 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.

	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

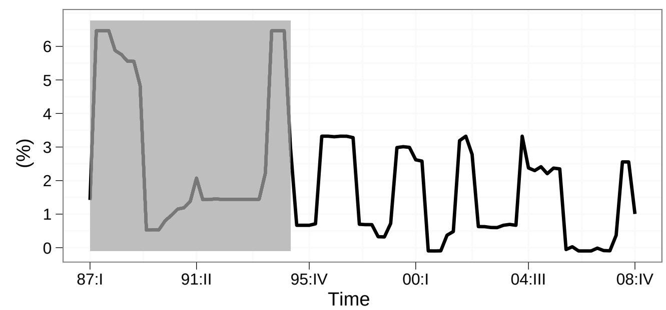
Table III: Posterior distribution for the 16-regime model parameters *This table presents the posterior mean, standard deviation, 90 percent interval and inefficiency factor based on 20,000 posterior draws beyond 5,000 burn-in.*

3.5 Out-of-Sample Prediction

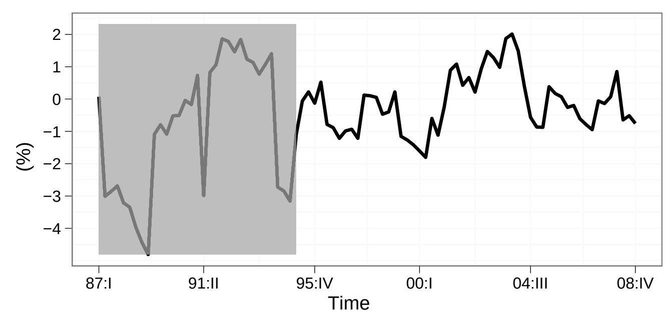
As described in section 2.5, we forecast the four quarters in 2008 using the data up to 2007:Q4. Following Chib and Kang (2010), the predictive accuracy is measured in terms of the posterior predictive criterion (PPC, Gelfand and Ghosh (1998)). PPC favors

	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 IV: Regime-specific average term spreads *These term spreads are in annualized percents.*



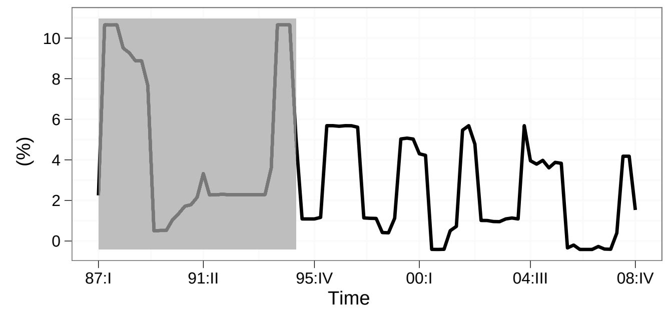
(a) term premium



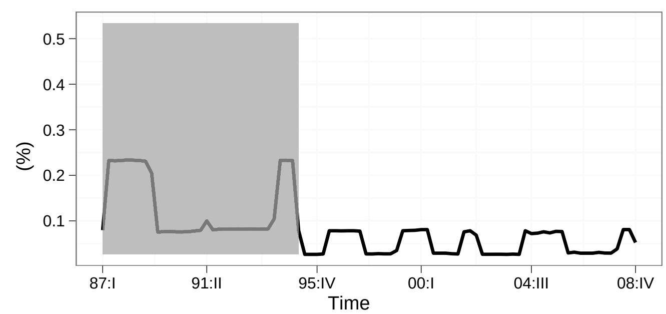
(b) EH component

Figure 9: The term premium and the EH component of the 5-year bond *These graphs are based on 20,000 simulated draws of the posterior simulation. The term premium, the expected excess return and the EH component are computed by 1.53 and 1.55, respectively. These are in annualized percents. The shaded area represents the less active policy regime.*

models to minimize a sum of goodness-of-fit and penalty term on model complexity. Table V clearly displays that the proposed model outperforms the alternatives, which is consistent with the marginal likelihood results (i.e. in-sample forecasting).



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



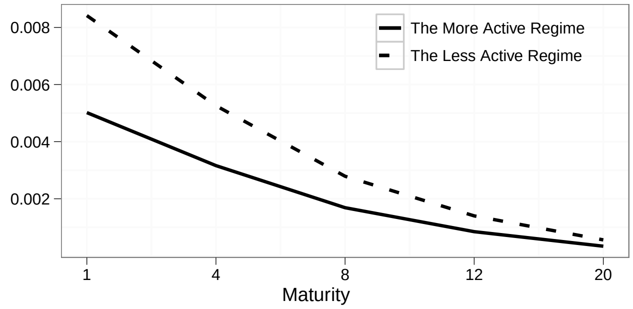
(b) regime shift risk component ($\text{RS}_{d_t,t}^{(\tau)}$)

Figure 10: Model $\mathcal{M}_{16}$: Decomposition of the term premium of the 5-year bond *These graphs are based on 20,000 simulated draws of the posterior simulation. These two components in annualized percents are computed by (1.55). The shaded area represents the less active policy regime.*

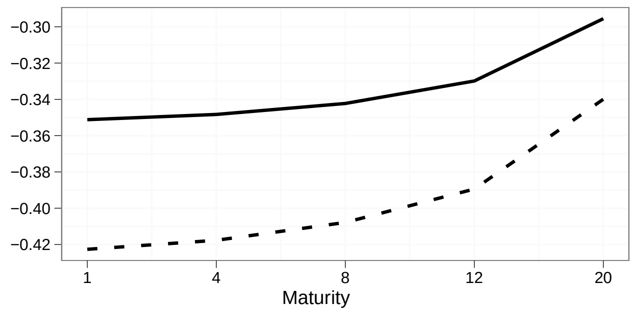
4 Conclusion

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 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.

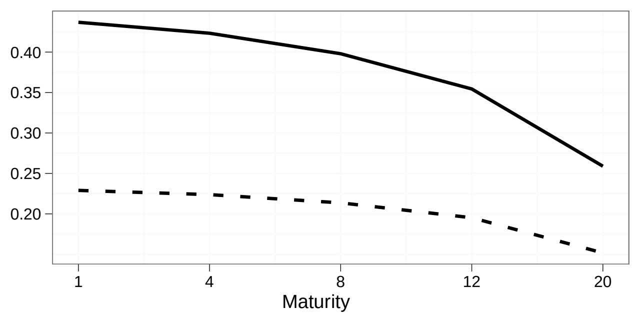
Our empirical results reveal that, in its goal of stabilizing the economy, monetary policy has been more responsive to the macro fundamentals since 1995:Q2 with important effects on the dynamics of the term structure. Because in a more active regime



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



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

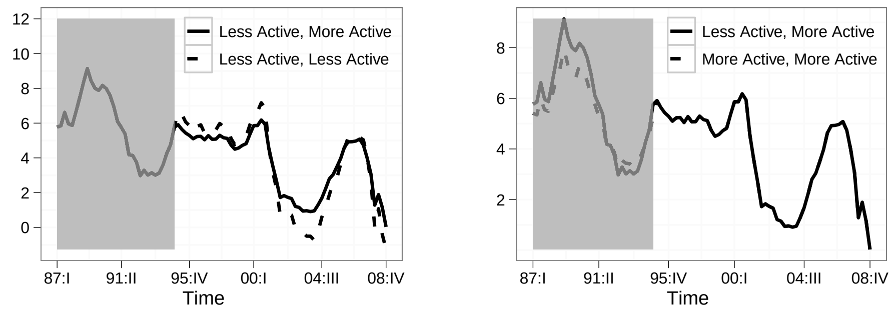


(c) monetary policy ($\hat{e}_t$)

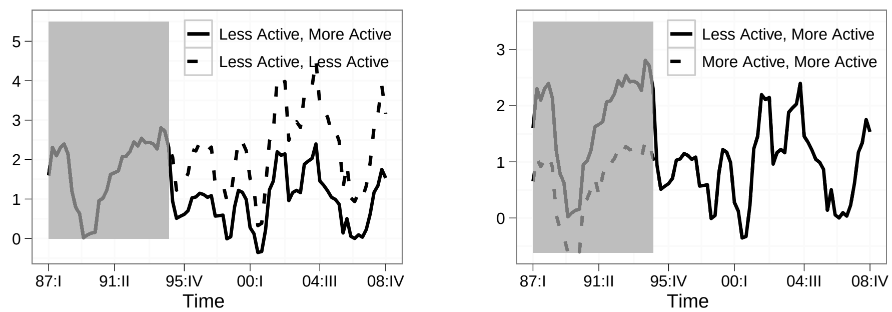
Figure 11: Model $\mathcal{M}_{16}$: The factor loadings *These graphs plot the estimates of the factor loadings on each of the exogenous processes. These graphs are based on on 20,000 simulated draws of the posterior simulation.*

agents anticipate less volatility in the macro variables, bonds are priced with a lower market price of risk. At the same time, the economy becomes less vulnerable to the inflation risk, and then investors require lower compensations for risk to hold long term bonds. 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



(a) short rate



(b) term spread

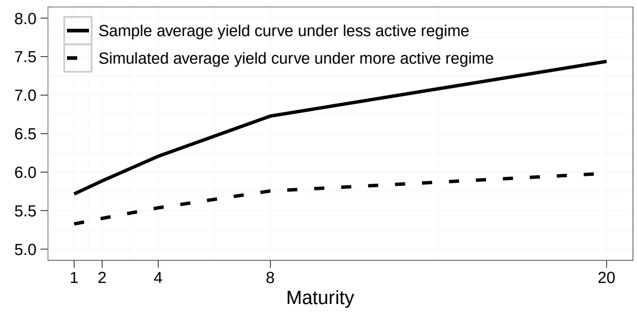
Figure 12: Model $\mathcal{M}_{16}$: Counterfactual analysis: interest rates *The top panel graphs the results for the short rate and the bottom one is for the term spread of 20 quarter bond. The shaded area represents the less active policy regime.*

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.

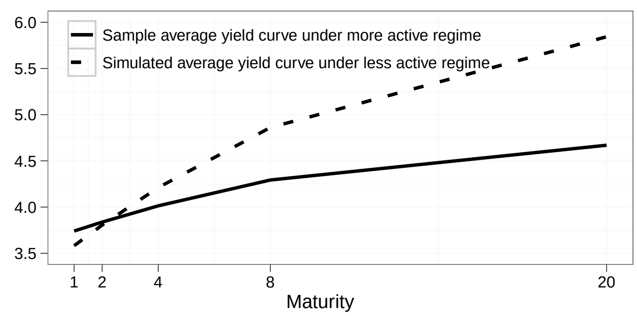
Finally, our empirical findings reveal the important role that incorporating regime changes can play in improving forecasts of the term structure.

A 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



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



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

Figure 13: Model $\mathcal{M}_{16}$: Counterfactual analysis: average yield curve

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

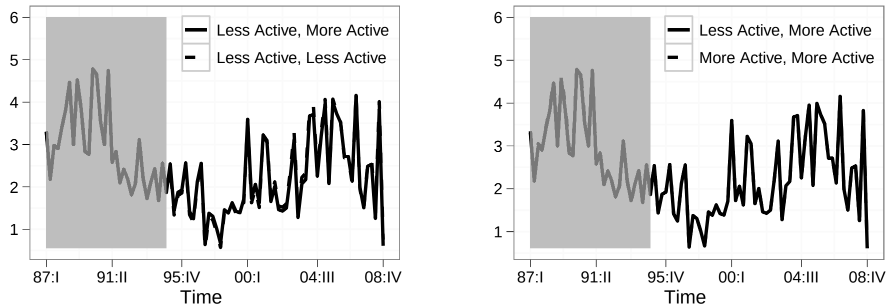
$$\eta^{\pi}_{jt+1} = \hat{\pi}_{jt+1} - \mathbb{E}_t\left(\hat{\pi}_{jt+1}\right) \text{ and } \eta^{x}_{jt+1} = \hat{x}_{jt+1} - \mathbb{E}_t\left(\hat{x}_{jt+1}\right), \quad j = 1, 2 \tag{A.1}$$

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 (1.37) and (1.38) into the system of equations (1.33)-(1.34) yields the following stacked system

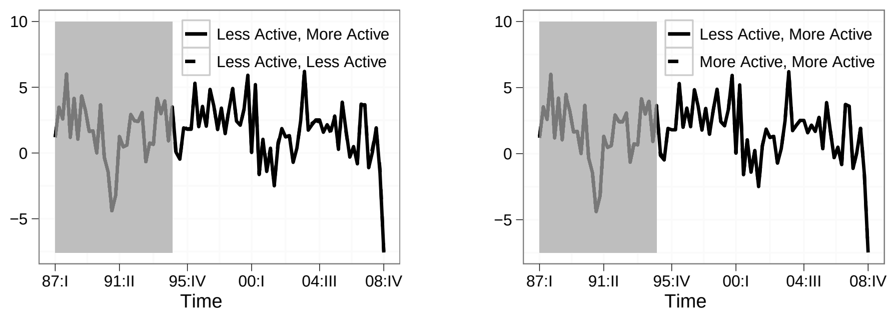
$$\mathbf{A}\begin{bmatrix} \hat{\pi}_{1t+1} \\ \hat{\pi}_{2t+1} \\ \hat{x}_{1t+1} \\ \hat{x}_{2t+1} \end{bmatrix} = \mathbf{B}\begin{bmatrix} \hat{\pi}_{1t} \\ \hat{\pi}_{2t} \\ \hat{x}_{1t} \\ \hat{x}_{2t} \end{bmatrix} + \mathbf{A}\begin{bmatrix} \eta^{\pi}_{1t+1} \\ \eta^{\pi}_{2t+1} \\ \eta^{x}_{1t+1}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

where

$$\mathbf{A} = \begin{bmatrix} \delta \otimes \mathbf{P} & 0_{2\times 2} \\ \mathbf{P} & \gamma \otimes \mathbf{P} \end{bmatrix}, \tag{A.3}$$



(a) inflation



(b) output growth

Figure 14: Model $\mathcal{M}_{16}$: Counterfactual analysis: inflation and output growth *The shaded area represents the less active policy regime.*

$$\mathbf{B}_{11} = \mathbf{I}_{m+1},\ \mathbf{B}_{12} = -\kappa \times \mathbf{I}_{m+1},\ \mathbf{B}_{21} = \mathsf{diag}(\alpha_1, \alpha_2, .., \alpha_{m+1}), \tag{A.4}$$

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

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \tag{A.5}$$

$$\text{and } \mathbf{C} = \begin{bmatrix} 0 & \kappa & 0 \\ 0 & \kappa & 0 \\ -\phi_a & \gamma\left(\phi_g - 1\right) & 1 \\ -\phi_a & \gamma\left(\phi_g - 1\right) & 1 \end{bmatrix} \tag{A.6}$$

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.

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 V: Posterior predictive criterion (PPC) *The forecast period is 2008:Q1-2008:Q4. We use the data from 1986:Q4 to 2007:Q4. The five yields are of 1, 2, 3, 8 and 20 quarters maturity bonds*

B Bond Prices

This section provides the details on the derivation of the bond prices in (1.49) and (1.50). We begin by letting $\mathbf{E}^{d_{t+1}}$ denote an expectation conditioned on d_{t+1}. Then the equation (1.43) can be expressed as

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

or

$$1 = \mathbb{E}^{s_{t+1}}\left[\mathbb{E}\left[M_{t,t+1}h_{\tau,t+1}|\bar{\mathbf{f}}_t, d_t, d_{t+1}\right]\right] \tag{B.2}$$

where

$$\begin{aligned} h_{\tau,t+1} &= P^{(\tau-1)}_{d_{t+1},t+1}/P^{(\tau)}_{d_t,t} \\ &= \exp\left[-\mathbf{a}_{d_{t+1}}(\tau-1) - \mathbf{b}_{d_{t+1}}(\tau-1)'\bar{\mathbf{f}}_{t+1} + \mathbf{a}_{d_t}(\tau) + \mathbf{b}_{d_t}(\tau)'\bar{\mathbf{f}}_t\right]. \end{aligned} \tag{B.3}$$

If we define

$$\begin{aligned} \Theta_{d_t,d_{t+1}} &= -\mathbf{a}_{d_{t+1}}(\tau-1) + \mathbf{a}_{d_t}(\tau) + \left(\mathbf{b}_{d_t}(\tau)' - \mathbf{b}_{d_{t+1}}(\tau-1)'\boldsymbol{\phi}\right)\bar{\mathbf{f}}_t \\ \text{and } \Gamma_{\tau,d_{t+1}} &= \mathbf{L}_{d_{t+1}} - \mathbf{b}_{d_{t+1}}(\tau-1)', \end{aligned} \tag{B.4}$$

then $M_{t,t+1}h_{\tau,t+1}$ can be rewritten as

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

$$= \exp\left[-\ln R^* - \frac{1}{2}\mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}'_{d_{t+1}} + \boldsymbol{\lambda}_{d_t,d_{t+1}}\bar{\mathbf{f}}_t + \frac{\mathbf{1}}{\mathbf{2}}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}} + \Theta_{d_t,d_{t+1}}\right]$$

$$\times \exp\left[-\frac{1}{2}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}} + \boldsymbol{\Gamma}_{\tau,d_{t+1}}\varepsilon_{t+1}\right] \quad \text{(B.5)}$$

Since

$$\mathbb{E}\left[\exp\left[-\frac{1}{2}\Gamma_{\tau,d_{t+1}}\Omega_{d_{t+1}}\Gamma'_{\tau,d_{t+1}} + \Gamma'_{\tau,d_{t+1}}\varepsilon_{t+1}\right] |\bar{\mathbf{f}}_t, d_t, d_{t+1}\right] = 1 \quad \text{(B.6)}$$

the log-approximation gives

$$\mathbb{E}\left[M_{t,t+1}h_{\tau,t+1}|\bar{\mathbf{f}}_t, d_t, d_{t+1}\right] \quad \text{(B.7)}$$

$$\approx -\ln R^* + \boldsymbol{\lambda}_{d_t,d_{t+1}}\bar{\mathbf{f}}_t - \mathbf{L}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1)' + \frac{1}{2}\mathbf{b}_{d_{t+1}}(\tau-1)'\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1) + \Theta_{d_t,d_{t+1}} + 1$$

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

$$0 = \sum_{j=1}^{\mathbf{d}} p_{ij}\left(-\ln R^* + \boldsymbol{\lambda}_{i,j}\bar{\mathbf{f}}_t - \mathbf{L}_j\Omega_j\mathbf{b}_j(\tau-1)' + \frac{1}{2}\mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1) + \Theta_{i,j}\right) \quad \text{(B.8)}$$

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

C 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 (1.53) and (1.55).

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

$$r^{(\tau)}_{d_t,t} - r^{(1)}_{d_t,t} \quad \text{(C.1)}$$

Let $x^{(\tau)}_t = p^{\tau-1}_{d_{t+1},t+1} - p^{\tau}_{d_t,t} - r^{(1)}_{d_t,t}$ denote the excess return. Then we have

$$r^{(\tau)}_{d_t,t} - r^{(1)}_{d_t,t} = \frac{1}{\tau}\sum_{l=0}^{\tau-1}\mathbb{E}_t\left[r^{(1)}_{d_{t+l},t+l}\right] - r^{(1)}_{d_t,t} + \frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathbb{E}_t\left[x^{(\tau+1-i)}_t\right] \quad \text{(C.2)}$$

$$= \frac{1}{\tau}\sum_{l=0}^{\tau-1}\mathbb{E}_t\left[r^{(1)}_{d_{t+l},t+l}\right] - r^{(1)}_{d_t,t} + \frac{1}{\tau}\sum_{i=2}^{\tau}\mathrm{exr}^{(i)}_{d_t,t}$$

$$= \frac{1}{\tau}\sum_{l=0}^{\tau-1}\mathbb{E}_t\left[r^{(1)}_{d_{t+l},t+l}\right] - r^{(1)}_{d_t,t} + \mathrm{TP}^{(\tau)}_{d_t,t}$$

where

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

Now we prove the equation (1.55). We begin by noting that the risk-neutral pricing formula in the equation (1.43) implies

$$p^{(\tau)}_{d_t,t} = \mathbb{E}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] \tag{C.4}$$

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

$$\begin{aligned} p^{(\tau)}_{d_t,t} &= \mathbb{E}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] \\ &= \mathbb{E}_t\left[m_{t,t+1}\right] + \mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] \\ &= p^{(1)}_{d_t,t} - \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1}\right] + \mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1} + p^{(\tau-1)}_{d_{t+1},t+1}\right] \end{aligned} \tag{C.5}$$

since $p^{(1)}_{d_t,t} = \mathbb{E}_t\left[m_{t,t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[m_{t,t+1}\right]$

and thus

$$p^{(\tau)}_{d_t,t} = p^{(1)}_{d_t,t} + \mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] + \frac{1}{2}\mathbb{V}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] + \mathsf{Cov}_t\left[m_{t,t+1}, p^{(\tau-1)}_{d_{t+1},t+1}\right] \tag{C.6}$$

This implies that

$$\begin{aligned} \mathrm{exr}^{(\tau)}_{d_t,t} &= \left[\mathbb{E}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] - p^{(\tau)}_{d_t,t}\right] - (-p^{(1)}_{d_t,t}) \\ &= -\mathsf{Cov}_t\left[m_{t,t+1}, p^{(\tau-1)}_{d_{t+1},t+1}\right] - \frac{1}{2}\mathbb{V}_t\left[p^{(\tau-1)}_{d_{t+1},t+1}\right] \end{aligned} \tag{C.7}$$

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 (C.7). We begin by expressing the pricing kernel and the log of bond price as

$$m_{t,t+1} \approx W_{d_t,d_{t+1},t} + \mathbf{L}_{d_{t+1}}\varepsilon_{t+1} \tag{C.8}$$

$$p_{d_{t+1},t+1}^{(\tau-1)} = -a_{d_{t+1}}(\tau-1) - \mathbf{b}_{d_{t+1}}(\tau-1)'\left(\phi\bar{\mathbf{f}}_t + \varepsilon_{t+1}\right) \tag{C.9}$$
$$= K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}$$

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 $p_{d_{t+1},t+1}^{(\tau-1)}$ using the law of iterative expectation as follows.

$$\mathbb{E}_t[p_{d_{t+1},t+1}^{(\tau-1)}] = \mathbb{E}_t\left(\mathbb{E}_t[p_{t+1}^{(\tau-1)}|d_{t+1}]\right) = \mathbb{E}_t\left(K_{d_{t+1},t}\right) = \sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t} \tag{C.10}$$

$$\mathbb{E}_t[m_{t,t+1}] = \mathbb{E}_t[W_{d_t,d_{t+1},t}] = \sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t} \tag{C.11}$$

$$\begin{aligned}
\mathbb{E}_t[m_{t,t+1}p_{d_{t+1},t+1}^{(\tau-1)}] &= \mathbb{E}_t[(W_{d_t,d_{t+1},t} + \mathbf{L}_{d_{t+1}}\varepsilon_{t+1})\left(K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}\varepsilon_{t+1}\right)] \\
&= \mathbb{E}_t[W_{d_t,d_{t+1},t}K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}\Omega_{d_{t+1}}\mathbf{L}_{d_{t+1}}] \\
&= \sum_{j=1}^{\mathbf{d}} p_{ij}\left(W_{i,j,t}K_{j,t} - \mathbf{b}_j\Omega_j\mathbf{L}_j\right)
\end{aligned} \tag{C.12}$$

Therefore,

$$\begin{aligned}
-\mathsf{Cov}_t(m_{t,t+1}, p_{d_{t+1},t+1}^{(\tau-1)}) &= \mathbb{E}_t[p_{d_{t+1},t+1}^{(\tau-1)}]\mathbb{E}_t[m_{t,t+1}] - \mathbb{E}_t[m_{t,t+1}p_{d_{t+1},t+1}^{(\tau-1)}] \\
&= \left(\sum_{j=1}^{\mathbf{d}} p_{ij}K_{j,t}\right)\left(\sum_{j=1}^{\mathbf{d}} p_{ij}W_{i,j,t}\right) - \sum_{j=1}^{\mathbf{d}} p_{ij}\left(W_{i,j,t}K_{j,t} - \mathbf{b}_j\Omega_j\mathbf{L}_j\right)
\end{aligned} \tag{C.13}$$

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

$$\begin{aligned}
\mathbb{E}_t\left[\left(p_{d_{t+1},t+1}^{(\tau-1)}\right)^2\right] &= \mathbb{E}_t\left[\left(K_{d_{t+1},t} - \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}\right)^2\right] \\
&= \mathbb{E}_t\left[K_{d_{t+1},t}^2 - 2K_{d_{t+1},t}\mathbf{b}_{d_{t+1}}\varepsilon_{t+1} + \mathbf{b}_{d_{t+1}}(\tau-1)'\varepsilon_{t+1}\varepsilon_{t+1}'\mathbf{b}_{d_{t+1}}(\tau-1)\right] \\
&= \mathbb{E}_t\left[K_{d_{t+1},t}^2 + \mathbf{b}_{d_{t+1}}(\tau-1)'\Omega_{d_{t+1}}\mathbf{b}_{d_{t+1}}(\tau-1)\right] \\
&= \sum_{j=1}^{\mathbf{d}} p_{ij}\left(K_{j,t}^2 + \mathbf{b}_j(\tau-1)'\Omega_j\mathbf{b}_j(\tau-1)\right)
\end{aligned} \tag{C.14}$$

and thus

$$\mathbb{V}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right] = \mathbb{E}_t\left[\left(p_{d_{t+1},t+1}^{(\tau-1)}\right)^2\right] - \left(\mathbb{E}_t\left[p_{d_{t+1},t+1}^{(\tau-1)}\right]\right)^2 \tag{C.15}$$

$$= \sum_{j=1}^{\mathbf{d}} p_{ij} \left(K_{j,t}^2 + \mathbf{b}_j(\tau-1)'\Omega_j \mathbf{b}_j(\tau-1)\right) - \left(\sum_{j=1}^{\mathbf{d}} p_{ij} K_{j,t}\right)^2$$

which completes the proof.

D 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 $\{\sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_5^2\}$ 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 $q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y})$ 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 $\pi(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y})$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \mathbf{y}) = St\left(\boldsymbol{\theta}_i|\hat{\boldsymbol{\theta}}_i, \mathbf{V}_{\hat{\boldsymbol{\theta}}_i}, \mathbf{15}\right) \tag{D.1}$$

where

$$\hat{\boldsymbol{\theta}}_i = \arg\max_{\boldsymbol{\theta}_i} \ln\{f(\mathbf{y}|\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}, \mathbf{D}_n)\pi(\boldsymbol{\theta}_i)\} \tag{D.2}$$

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

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

$$\alpha\left(\boldsymbol{\theta}_i^{(g-1)},\boldsymbol{\theta}_i^{\dagger}|\boldsymbol{\theta}_{-i},\mathbf{y}\right) \tag{D.3}$$
$$= \min\left\{\frac{f\left(\mathbf{y}|\boldsymbol{\theta}_i^{\dagger},\boldsymbol{\theta}_{-i},\mathbf{D}_n\right)\pi\left(\boldsymbol{\theta}_i^{\dagger}\right)}{f\left(\mathbf{y}|\boldsymbol{\theta}_i^{(g-1)},\boldsymbol{\theta}_{-i},\mathbf{D}_n\right)\pi\left(\boldsymbol{\theta}_i^{(g-1)}\right)}\frac{St\left(\boldsymbol{\theta}_i^{(g-1)}|\hat{\boldsymbol{\theta}}_i,\mathbf{V}_{\hat{\boldsymbol{\theta}}_i},\mathbf{15}\right)}{St\left(\boldsymbol{\theta}_j^{\dagger}|\hat{\boldsymbol{\theta}}_{\mathbf{i}},\mathbf{V}_{\hat{\boldsymbol{\theta}}_{\mathbf{i}}},\mathbf{15}\right)},1\right\}.$$

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

$$\pi\left(\boldsymbol{\theta}_1|\boldsymbol{\theta}_{-1},\mathbf{y},\mathbf{D}_n\right),\pi\left(\boldsymbol{\theta}_2|\boldsymbol{\theta}_{-2},\mathbf{y},\mathbf{D}_n\right),\ldots,\ \pi\left(\boldsymbol{\theta}_{h_g}|\boldsymbol{\theta}_{-h_g},\mathbf{y},\mathbf{D}_n\right) \tag{D.4}$$

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

$$\underbrace{\begin{bmatrix}\ln\left(P_t/P_{t-1}\right)\\ \ln Y_t\end{bmatrix}}_{\mathbf{m}_t} = \underbrace{\begin{bmatrix}\ln\pi^*\\ \ln x^*+\ln A_t\end{bmatrix}}_{\bar{\mathbf{J}}_t}+\underbrace{\begin{bmatrix}h_\pi^a(d_t) & h_\pi^g(d_t) & h_\pi^e(d_t)\\ h_x^a(d_t) & h_x^g(d_t) & h_x^e(d_t)\end{bmatrix}}_{\bar{\mathbf{H}}_{d_t}}\bar{\mathbf{f}}_t \tag{D.5}$$

$$= \underbrace{\begin{bmatrix}\ln\pi^*\\ \ln x^*+\ln a^*+\ln A_{t-1}\end{bmatrix}}_{\mathbf{J}_{t-1}}+\underbrace{\begin{bmatrix}h_\pi^a(d_t) & h_\pi^g(d_t) & h_\pi^e(d_t)\\ 1+h_x^a(d_t) & h_x^g(d_t) & h_x^e(d_t)\end{bmatrix}}_{\tilde{\mathbf{H}}_{d_t}}\bar{\mathbf{f}}_t \tag{D.6}$$

and thus

$$\begin{bmatrix}\mathbf{m}_t\\ r_{1t}\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{J}}_t\\ \bar{a}_{d_t}(\tau_1)\end{bmatrix}+\begin{bmatrix}\bar{\mathbf{H}}_{d_t}\\ \bar{\mathbf{b}}_{d_t}(\tau_1)'\end{bmatrix}\bar{\mathbf{f}}_t \tag{D.7}$$
$$= \begin{bmatrix}\mathbf{J}_{t-1}\\ \bar{a}_{d_t}(\tau_1)\end{bmatrix}+\begin{bmatrix}\tilde{\mathbf{H}}_{d_t}\\ \bar{\mathbf{b}}_{d_t}(\tau_1)'\end{bmatrix}\bar{\mathbf{f}}_t \tag{D.8}$$

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.

$$\bar{\mathbf{f}}_0 = \begin{bmatrix}\bar{\mathbf{H}}_{s_0}\\ \bar{\mathbf{b}}_{s_0}(\tau_1)'\end{bmatrix}^{-1}\left(\begin{bmatrix}\mathbf{m}_0\\ r_{10}\end{bmatrix}-\begin{bmatrix}\bar{\mathbf{J}}_0\\ \bar{a}_{s_0}(\tau_1)\end{bmatrix}\right) \tag{D.9}$$

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

$$\mathbf{f}_t = \begin{bmatrix} \bar{\mathbf{f}}_t \\ \ln A_t \end{bmatrix} \qquad \text{(D.10)}$$

where

$$\bar{\mathbf{f}}_t = \begin{bmatrix} \tilde{\mathbf{H}}_{d_t} \\ \bar{\mathbf{b}}_{d_t}(\tau_1)' \end{bmatrix}^{-1} \left(\begin{bmatrix} \mathbf{m}_t \\ r_{1t} \end{bmatrix} - \begin{bmatrix} \mathbf{J}_{t-1} \\ \bar{a}_{d_t}(\tau_1) \end{bmatrix} \right) \qquad \text{(D.11)}$$

and

$$\ln A_t = \ln A_{t-1} + \ln a^* + \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \bar{\mathbf{f}}_t \qquad \text{(D.12)}$$

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 $\mathbf{D}_n$

$$\ln f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{D}_n) = \sum_{t=1}^{n} \ln f[\mathbf{y}_t | I_{t-1}, d_t, \boldsymbol{\theta}] \qquad \text{(D.13)}$$

where

$$f[\mathbf{y}_t | I_{t-1}, d_t, \boldsymbol{\theta}] = -(2\pi)^{-7/2} \left| \Lambda^{d_t} \right|^{-1/2} \times \exp\left[-\frac{1}{2} \eta_{t|t-1}^{d_t \prime} \left(\Lambda^{d_t} \right)^{-1} \eta_{t|t-1}^{d_t} \right] \qquad \text{(D.14)}$$

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

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

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

Step 3 Sampling factors

Conditioned on $\boldsymbol{\theta}$ and $\mathbf{D}_n$, the equations (D.9) - (D.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.

References

Albert, J. H. and Chib, S. (1993), "Bayes inference via gibbs sampling of autoregressive time-series subject to markov mean and variance shifts," *Journal of Business & Economic Statistics*, 11, 1–15.

An, S. and Schorfheide, F. (2007), "Bayesian analysis of DSGE models," *Econometric Reviews*, 26, 113–172.

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., Boivin, J., Dong, S., and Loo-Kung, R. (2010), "Monetary Policy Shifts and the Term Structure," *Review of Economic Studies*, in press.

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.

Bikbov, R. and Chernov, M. (2008), "Monetary Policy Regimes and the Term Structure of Interest Rates," *CEPR Discussion Papers 7096*.

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.

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*, forthcoming.

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 for analysis of DSGE models," *Journal of Econometrics*, forthcoming.

Clarida, R., Gali, J., and Gertler, M. (2000), "Monetary policy rules and macroeconomic stability: evidence and some theory," *Quarterly Journal of Economics*, 65, 147–180.

Cogley, T. and Sbordone, A. (2008), "Dynamic Equilibrium and the Structure of Premiums In A Reinsurance Market," *American Economic Review*, 98, 2101–26.

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.

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.

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. (2009), "Two Books on the New Macroeconometrics," *Econometric Reviews*, 28, 376–387.

Fruhwirth-Schnatter, S. (2006), "Finite Mixture and Markov Switching Models," *Springer*.

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.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–84.

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

Ludvigson, S. C. and Ng, S. (2009), "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22(12), 5027–5067.

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. 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.

Schorfheide, F. (2005), "Learning and monetary policy shifts," *Review of Economic Dynamics*, 8, 392–419.

Smets, F. and Wouters, R. (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, 97, 586–606.

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