The Effects of Monetary Policy Regime Shifts on the Term Structure of Interest Rates

AZAMAT ABDYMOMUNOV*
KYU HO KANG†

Washington University in St. Louis

July 2010

Abstract

We investigate how the entire term structure of interest rates is influenced by changes in monetary policy regimes. To do so, we develop and estimate an arbitrage-free dynamic term-structure model which accounts for regime shifts in monetary policy, volatility, and the price of risk. Our results for U.S. data from 1985-2008 indicate that (*i*) the Fed's reaction to inflation has changed over time, switching between "more active" and "less active" monetary policy regimes, (*ii*) the yield curve in the "more active" regime was considerably more volatile than in the "less active" regime, and (*iii*) on average, the slope of the yield curve in the "more active" regime was steeper than in the "less active" regime. The steeper yield curve in the "more active" regime reflects higher term premia that result from the risk associated with a more volatile future short-term rate given a more sensitive response to inflation.
(JEL G12, C11, E43)

Keywords: Term structure of interest rates, Affine no-arbitrage model, Monetary policy, Markov switching process, Bayesian estimation

---

* *Corresponding author*: Department of Economics, Washington University in St. Louis, Campus Box 1208, One Bookings Drive, St. Louis, MO 63130. E-mail: aabdymom@wustl.edu.

† Department of Economics, Washington University in St. Louis, Campus Box 1208, One Bookings Drive, St. Louis, MO 63130. E-mail: khkang@wustl.edu.

We are especially grateful to James Morley for his valuable discussions and suggestions. We would also like to thank Steve Fazzari, William Gavin, Werner Ploberger, Guofu Zhou, and participants at the 2010 Missouri Economic Conference for their helpful comments. All remaining errors are our own.

1 Introduction

Many empirical studies (e.g. Clarida, Gali, and Gertler (2000); Cogley and Sargent (2005)) focus mainly on the response of output and inflation to monetary policy changes. However, only a few studies (e.g. Bikbov and Chernov (2008) and Ang, Boivin, Dong, and Loo-Kung (2010) hereafter ABDL(2010)) look at the implications of monetary policy changes for the term structure of interest rates.

As discussed in ABDL(2010), the entire term structure of interest rates may respond to the changes in monetary policy in two main ways. First, according to the no-arbitrage condition, the long-term interest rate should be affected by changes in the short-term interest rate caused by monetary policy. Second, the inflation and output fluctuations caused by monetary policy may influence term premia. This effect is supported by many recent studies which provide evidence of the impact of macroeconomic factors on the term structure of interest rates (e.g. Ang and Piazzesi (2003); Ang, Bekaert, and Wei (2008); and Bikbov and Chernov (2010)). At the same time, as discussed in Bikbov and Chernov (2008), if the entire term structure of interest rates responds to the changes in monetary policy, then the term structure may contain more useful information for identifying the monetary policy regimes as compared to only considering the short rate.

The way monetary policy is conducted can have two potential implications for long-term interest rates. First, the monetary authority may influence inflation expectations through aggressively changing the short rate in response to macroeconomic fluctuations. This effect reduces inflation risk premia for long-term interest rates. Second, a more sensitive short rate in response to macroeconomic fluctuations may cause expectations of a more volatile future short rate, which could result in higher risk premia for long-term interest rates. Thus, the monetary authority may face a trade-off between these two opposite effects on long-term interest rates in their choice of how aggressively to respond to macroeconomic fluctuations.

The main objective of this paper is to analyze effects of monetary policy regime changes on the entire term structure of interest rates. Specifically, we aim to identify which of the two above-described effects on long-term rates dominates when the monetary authority responds aggressively to macroeconomic fluctuations. For this analysis, we propose an affine no-arbitrage term structure model with regime shifts in monetary policy, volatility of yield factors, and the market price of risk governed by three separate Markov-switching processes. This framework enables us to identify the effects of monetary policy regime shifts on long rates. In our model, the short-term interest rate, which is considered as the monetary policy instrument, is set by a Taylor (1993) rule with coefficients switching between two monetary policy regimes. These regimes are labeled as "more active" and "less active" regimes, depending on how aggressively the monetary authority changes the short rate in response to inflation and output gap fluctuations.

Our results can be summarized as follows. First, our results indicate that even during "the Great Moderation" period of the past quarter century, the Fed's reaction to inflation has varied over time, switching between "more active" and "less active" regimes. This result concurs with Sims and Zha (2006) and ABDL(2010), who conclude that regime shifts of monetary policy should be considered probabilistically rather than by only a single break in the early 1980s.

Second, monetary policy regime shifts have quantitatively important effects on the term spread and the volatility of the yield curve. For the sample of U.S. data from 1985:Q4 to 2008:Q4, the short rate was considerably more volatile in the "more active" regime than in the "less active" regime, while the average short rates in the two monetary policy regimes were close to each other. The long-term rate was, on average, 129 basis points higher in the "more active" regime than in the "less active" regime, resulting in a steeper slope of the yield curve, on average, in the "more active" regime. In general, the yield curve was more volatile in the "more active" regime than in the "less active" regime. These results can be explained by a more sensitive response of the short rate to inflation fluctuations in the "more active" regime creating higher risk for the future short rate fluctuations. This risk drives up long-term yields. Thus, the Fed appears to face a policy trade-off between a "more active" reaction to the macroeconomic fluctuations and a more volatile yield curve caused by this reaction. This argument is consistent with Woodford (1999), who claims that it may be more optimal for the monetary authority to conduct policies that do not require the short rate to be too volatile.

Our study is distinguished in several dimensions from Bikbov and Chernov (2008) and ABDL(2010), who also investigate the interaction between the term structure of interest rates and monetary policy. In particular, our model employs discrete-time regime-switching processes in contrast to ABDL(2010), who describe monetary policy shifts as continuously changing Taylor rule coefficients. Also, our model is differentiated from ABDL(2010) by incorporating volatility regime shifts, which, as indicated by Sims and Zha (2006), is important for evaluating the impact of monetary policy changes on macroeconomic behavior. Unlike Bikbov and Chernov (2008), who also apply discrete regimes, our model accounts for the regime shifts in the price of risk that are independent of volatility changes. Duffee (2002) reports that it is essential to allow for variation in the price of risk independent of factor volatility for fitting the yield curve and modeling plausible term premium. Also, our study focuses on the interaction between monetary policy and term structure dynamics in the post-1985 period in contrast to the longer periods covered by Bikbov and Chernov (2008) and ABDL(2010). The estimation of the model over the post-1985 period avoids identifying the monetary policy regimes with the major oil shocks in the 1970s, the monetary policy "experiment" in 1979, and the structural break in the monetary policy found by many studies (e.g. Fuhrer (1996) and Clarida et al. (2000)), which is associated with the beginning of the "Volcker" disinflation policy.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 discusses the estimation method. Section 4 presents the empirical results. Section 5 concludes. The Appendices provide details for the model derivation and the estimation method.

2 Model

In this section, we present our model used to quantify effects of monetary policy regime shifts on the term structure of interest rates. In particular, we develop a three-factor affine no-arbitrage term structure model with regime shifts in monetary policy response to macroeconomic fluctuations. The model also accounts for changes in volatility of yield factors and the market price of risk, governed by two other regime-switching processes. This modeling choice allows us to separate the identification of monetary policy changes from changes in volatility of yield factors and the market price of risk. To derive bond prices that account for the effects of monetary policy regime shifts and satisfy no-arbitrage condition, we make assumptions about a monetary policy response function, evolutions of regime processes, dynamics of factor process, and a stochastic discount factor, described in the following subsections.

2.1 Short rate

We assume that the monetary authority use the short rate as their policy instrument and set it according to the Taylor rule (1993) with coefficients subject to regime shifts:

$$r_t^{m_t} = \overline{r}^{m_t} + \alpha^{m_t}\left(\pi_t - \overline{\pi}^{m_t}\right) + \beta^{m_t} g_t + u_t \ , \tag{2.1}$$

where $r_t^{m_t}$ is the short rate, π_t is inflation, $\overline{\pi}^{m_t}$ is the inflation target, g_t is the output gap, $\overline{r}^{m_t}$ is the optimal level of the short rate for the case when inflation and output gaps are zero, α^{m_t} and β^{m_t} are policy response coefficients to inflation and output gaps, respectively, and u_t is a monetary policy shock. Superscript m_t denotes the monetary policy regime.

In this specification of the policy rule, similarly to ABDL(2010), the monetary authority is assumed to respond to contemporaneous inflation and output gap, in contrast to expected inflation and output gap used in some studies on the Taylor rule (e.g. Clarida et al. (2000)). Sims and Zha (2006) argue that using expected inflation in the policy rule may result in distorted conclusions because expected inflation will be measured as a set of all influences on monetary policy and also it has less variation than current nominal variables, potentially causing spuriously scaled up response coefficients.

In our specification of the policy rule, the response coefficients to inflation and output gaps switch between two monetary policy regimes. These monetary policy regimes m_t are governed by a two-state Markov chain with transition matrix

$$\Pi_m \equiv \begin{bmatrix} 1 - p_m^{12} & p_m^{12} \\ p_m^{21} & 1 - p_m^{21} \end{bmatrix}, \tag{2.2}$$

where $p_m^{jk} = \Pr[m_t = k | m_{t-1} = j] \in [0, 1]$.

As pointed out by ABDL(2010), if monetary shocks are correlated with inflation and output, then estimation of the standard Taylor rule equation (i.e. equation (2.1) with single regime) does not produce consistent estimates of the response coefficients. This correlation may be caused by contemporaneous effect of the monetary shocks on macroeconomic variables. However, Ang, Dong, and Piazzesi (2007b), Bikbov and Chernov (2008), and ABDL(2010) show that u_t can be identified by utilizing the information in the entire term structure of interest rates through a no-arbitrage restriction.

2.2 Factor dynamics

Similarly to many studies on the term structure of interest rates in the macro-finance literature (e.g., Ang and Piazzesi (2003); Ang et al. (2007b); and Bikbov and Chernov (2008)), we describe the dynamics of bond prices by three factors $\mathbf{f}_t = (u_t \ , \pi_t, g_t)'$, two of which are observable macro variables and one is a latent variable. The latent variable, denoted by u_t , is interpreted as a monetary policy shock in the Taylor rule equation. The factor dynamics are assumed to follow a regime-dependent Gaussian vector autoregressive process and can be described by

$$\mathbf{f}_{t+1} - d^{m_{t+1}} = G\left(\mathbf{f}_t - d^{m_t}\right) + L^{v_{t+1}} \varepsilon_{t+1} \ , \qquad \varepsilon_{t+1} \sim \mathcal{N}_{3\times 1}(0, I) \ , \tag{2.3}$$

where G is 3×3 matrix; $L^{v_{t+1}}$ is the lower-triangular Cholesky decomposition of $\Omega^{v_{t+1}}$ matrix that denotes the variance-covariance matrix of the factor shocks, d^{m_t} is the mean of factors within each monetary policy regime. We assume that the factors volatilities can change their values between "low" and "high" volatility regimes denoted by v_t and governed by a two-state Markov-switching process with transition probability matrix

$$\Pi_v \equiv \begin{bmatrix} 1 - p_v^{12} & p_v^{12} \\ p_v^{21} & 1 - p_v^{21} \end{bmatrix} . \tag{2.4}$$

By setting the persistence parameter matrix G to be regime-independent we avoid having potential changes in persistence influence the identification of the monetary policy regimes.[1]

[1]The persistence of latent factor and inflation could be assumed to be policy dependent. Watson (1999) finds that persistence of the short rate increased over the two sample periods: 1965-1978 and

2.3 Market price of risk

To model risk premia for long rates, we specify the market price of risk to have a time-varying form. Similarly to Ang et al. (2008), the market price of risk is assumed to have the regime-switching and essentially affine in the factors form:

$$\Lambda_t^{l_{t+1}} = \lambda_0^{l_{t+1}} + \lambda_f \mathbf{f}_t \ , \tag{2.5}$$

where λ_f is a 3×3 matrix and $\lambda_0^{l_{t+1}}$ is 3×1 vector, which switches between "high" and "low" price of risk regimes denoted by l_t and governed by a two-state Markov-switching process with transition matrix

$$\Pi_l \equiv \begin{bmatrix} 1 - p_l^{12} & p_l^{12} \\ p_l^{21} & 1 - p_l^{21} \end{bmatrix} . \tag{2.6}$$

As we show in Section 4, accounting for the regime-switching in $\lambda_0^{l_{t+1}}$ considerably improves the data fitting. It provides grater flexibility for the model to generate plausible time-variation in risk premium in contrast to the time-variation in the price of risk that is originated only from the factors. For tractability we assume that the matrix λ_f is regime independent.

2.4 Bond Prices

The monetary policy (m_t), volatility (v_t), and price of risk (l_t) regime processes are assumed to be independent from each other for the sake of tractability. Because each regime process has two regimes, the aggregate regime process denoted as s_t has eight regimes:

s_t	1	2	3	4	5	6	7	8
l_t	1	1	1	1	2	2	2	2
v_t	1	1	2	2	1	1	2	2
m_t	1	2	1	2	1	2	1	2

(2.7)

where the transition probability matrix of the joint process is given by $\Pi = \Pi_l \otimes \Pi_v \otimes \Pi_m$.

1985-1998. For the sample period considered in our study, preliminary estimates of the model with regime-switches in persistence parameters indicates that the estimates of these parameters are close to each other in the two identified monetary policy regimes.

Bond pricing with a no-arbitrage restriction is derived by assuming the existence of a stochastic discount factor $\kappa_{t,t+1} = \kappa(\mathbf{f}_t, s_t; \mathbf{f}_{t+1}, s_{t+1})$ that establishes a recursion for pricing bonds of different maturities:

$$P^{s_t}_{\tau,t} = \mathbb{E}\left[\kappa_{t,t+1} P^{s_{t+1}}_{\tau-1,t+1} | \mathbf{f}_t, s_t\right], \tag{2.8}$$

where $P^{s_t}_{t,\tau}$ denotes the price of bond at time t in regime s_t that matures at period $(t+\tau)$ and $\mathbb{E}$ is an expectation operator. Note that this expectation is conditional on the current factors and regimes since they are assumed to be known to agents. Meanwhile, the future values of the factors and regimes are unknown and follow the stochastic processes described in the previous subsections, and thus the expectation is over the future uncertainties. However, the whole time path of the factors and regimes (even the past values of the latent factor and regimes) are not observable to econometricians and to be estimated.

In order to impose the no-arbitrage condition, we follow Ang et al. (2008) and assume that the stochastic discount factor has the form[2]:

$$\kappa_{t,t+1} = \exp\left(-r^{s_t}_t - \frac{1}{2}\Lambda^{s_{t+1}\prime}_t \Lambda^{s_{t+1}}_t - \Lambda^{s_{t+1}\prime}_t \varepsilon_{t+1}\right), \tag{2.9}$$

where $\Lambda^{s_{t+1}}_t$ is given by equation (2.5).

The logarithms of bond prices are assumed to be affine in the factors and they depend on three regime processes:

$$\log P^{s_t}_{\tau,t} = -A^{s_t}_\tau - B^{s_t\prime}_\tau \mathbf{f}_t \ , \tag{2.10}$$

where $A^{s_t}_\tau$ and $B^{s_t}_\tau$ are regime specific coefficients a the bond of maturity τ .

In order to represent the continuously-compounded short rate as an affine function of the factors, the Taylor rule equation (2.1) is transformed to the form:

$$r^{s_t}_t = \delta^{s_t}_0 + \delta^{s_t\prime}_f \mathbf{f}_t \ , \tag{2.11}$$

[2]In contrast to Dai, Singleton, and Yang (2007) and Ang et al. (2010), our model specification does not allow us to price the risk of regime shifts explicitly. Explicit pricing the regime-shift risk in our setting would require assuming a factor process in which the next-period-regime uncertainty does not affect the conditional distribution of factors $\mathbf{f}_{t+1}$. As discussed in Bansal and Zhou (2002), the implication of this assumption is not consistent with the evidence reported by Hamilton (1988) and Gray (1996). These two studies empirically show that the short-rate dynamics are successfully described as a mixture of conditional Normal distributions.

where it can easily be seen that $\delta_0^{s_t} = \overline{r}^{s_t} - \alpha^{s_t}\overline{\pi}^{s_t}$ and $\delta_f^{s_t} = \begin{pmatrix} 1 & \alpha^{s_t} & \beta^{s_t} \end{pmatrix}'$.

To solve for A_τ^j and B_τ^j, we substitute for $P_{t,\tau}^{s_t}$ and $P_{t,\tau-1}^{s_{t+1}}$ in equation (2.8) and, following Bansal and Zhou (2002), we use the law of iterated expectations, the method of undetermined coefficients, and log-linearization as discussed in Appendix A. The solution has a form of recursive system:

$$\begin{aligned} A_\tau^j &= \delta_0^j + \sum_{k=1}^{S} p^{jk} \Big(A_{\tau-1}^k + \left(d^k - Gd^j - L^k\lambda_0^k\right)' B_{\tau-1}^k \\ &\quad -\frac{1}{2} B_{\tau-1}^{k\prime} L^k L^{k\prime} B_{\tau-1}^k \Big) \qquad (2.12) \\ B_\tau^j &= \delta_f^j + \sum_{k=1}^{S} p^{jk} \left(G - L^k\lambda_f\right)' B_{\tau-1}^k \qquad (2.13) \end{aligned}$$

with the initial conditions given by $A_1^j = \delta_0^j$ and $B_1^j = \delta_f^j$. Given this recursion, the continuously-compounded yield for a τ-maturity zero-coupon bond is determined by

$$R_{\tau,t}^{s_t} = -\frac{1}{\tau}\log\left(P_{\tau,t}^{s_t}\right) = a_\tau^{s_t} + b_\tau^{s_t\prime}\mathbf{f}_t \ , \qquad (2.14)$$

where $a_\tau^{s_t} = \frac{A_\tau^{s_t}}{\tau}$, $b_\tau^{s_t} = \frac{B_\tau^{s_t}}{\tau}$, and $R_{1,t}^{s_t} = r_t^{s_t}$. This equation and the solution for $a_\tau^{s_t}$ and $b_\tau^{s_t}$ provide a basis for estimating the model and analyzing the effects of monetary policy regime shifts on the term structure of interest rates.

In each time period, the sequence of bond pricing by agents can be described as follows:

Stage 1 At the beginning of time t, agents learn regime s_t, where the realization of s_t depends on s_{t-1} and the transition probabilities;

Stage 2 The regime s_t determines the corresponding model parameters θ^{s_t};

Stage 3 Given θ^{s_t}, the factors $\mathbf{f}_t$ are generated by regime-specific autoregressive process $\mathbf{f}_t = f_{\mathbf{f}}\left(\theta^{s_t}, \mathbf{f}_{t-1}\right)$ in equation (2.3);

Stage 4 Next, given parameters θ^{s_t}, one can calculate the values of $A_\tau^{s_t}$ and $B_\tau^{s_t}$ recursively for all maturities τ based on the recursions in equations (2.12) and (2.13);

Stage 5 Finally using $\mathbf{f}_t$, $A_\tau^{s_t}$, and $B_\tau^{s_t}$ the agents price bonds $P_{t,\tau}^{s_t} = f_P\left(\mathbf{f}_t, A_\tau^{s_t}, B_\tau^{s_t}\right)$ as in equation (2.10).

2.5 Expected Excess Return and Term Premium

This subsection presents the solution for expected excess return and term premium implied by our model. As is well-known, the term spread, which is a difference between long-term and short-term yields, can be decomposed into expectation hypothesis and term premium components:

$$R^{s_t}_{\tau,t} - r^{s_t}_t = \underbrace{\left[\frac{1}{\tau}\sum_{i=0}^{\tau-1}\mathbb{E}_t\left[r_{t+i}\right] - r^{s_t}_t\right]}_{\text{Expectation Hypothesis Component}} + \underbrace{\frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathrm{ER}^{s_t}_{\tau+1-i,t}}_{\text{Term Premium}} \,, \tag{2.15}$$

where $\mathbb{E}_t$ denotes an expectation operator conditional on s_t and $\mathbf{f}_t$; $\mathrm{ER}^{s_t}_{\tau+1-i,t}$ denotes one-period expected excess return for the $(\tau+1-i)$-period bond in regime s_t.

The expected excess returns is derived following the approach of Dai et al. (2007). A risk-neutral agent should be indifferent between two strategies: *i)* holding a bond at time t, which matures at time period $(t+1+\tau-1)$ and *ii)* holding one-period bond at time t and purchasing a bond at time $(t+1)$ that matures at time period $(t+1+\tau-1)$. After accounting for the risk, the difference between these two strategies represents the expected excess return; and therefore the one-period expected excess return on the τ-period bond in regime $s_t = j$ is given by

$$\mathrm{ER}^j_{\tau,t} = \mathbb{E}[\overline{p}_{\tau-1,t+1}|s_t = j, \mathbf{f}_t] + \overline{p}^j_{1,t} - \overline{p}^j_{\tau,t} \,, \tag{2.16}$$

where $\overline{p}^j_{1,t} \equiv \log P^j_{\tau,t}$. Appendix B provides details of the solution for the expected excess return which has the form:

$$\mathrm{ER}^j_{\tau,t} = -\sum_{k=1}^{S} p^{jk}\left(B^{k'}_{\tau-1}L^k\Lambda^k_t + \frac{1}{2}B^{k'}_{\tau-1}L^kL^{k\prime}B^k_{\tau-1}\right). \tag{2.17}$$

The term premium for τ-period holding is simply the average of the expected excess returns over all maturities from 2 to $\tau-$periods.

3 Estimation

3.1 Data

We use quarterly data on yields of zero-coupon bonds and macroeconomic variables for the sample period of 1985:Q4 to 2008:Q4. The term structure data on eight yields of 1, 4, 8, 12, 16, 24, 36, and 40 quarter maturities are obtained from Gurkaynak, Sack, and Wright (2007). The yield for one-quarter Treasury bills is our measure of the short rate. The measure of inflation is the year on year log difference in the CPI. We follow Rudebusch and Swanson (2002) and ABDL(2010) and express the output gap as a percentage of the potential output as

$$g_t = \frac{1}{4}\frac{RGDP_t - RGDP_t^p}{RGDP_t^p}\ , \tag{3.1}$$

where $RGDP_t$ is real GDP in 2005 constant prices obtained from the St. Louis FED database and $RGDP_t^p$ is potential GDP computed similarly to Ang et al. (2007b) by applying the Hodrick and Prescott (1997) filter.[3] The gap is factored by 1/4 to make estimated coefficients interpretable as coefficients for annualized interest rates.

3.2 Identification restrictions

The factor dynamics and Taylor rule equation (2.1) are linked through identification restrictions $\overline{\pi}^{m_t} = d_2^{m_t}$ and $d_3^{m_t} = 0$. The latter of the two restrictions is imposed because the last factor is the output gap and one can reasonably assume that it has to be targeted at zero independently of the monetary policy regimes. For identification of the latent factor, $d_1^{m_t}$ is restricted to zero in both regimes. The inflation target $\overline{\pi}^{m_t}$ and optimal short rate $\overline{r}^{m_t}$ are assumed to be regime-independent, which is a more reasonable assumption for the sample period under consideration than if we had included the 1970s. Setting these parameters to be regime-independent also avoids identifying monetary policy regimes by potential switching in the mean of inflation and/or the short rate rather than switching in the policy reaction coefficients. We also set $\overline{\pi}$ and $\overline{r}$ to their sample average values, as in Dai et al. (2007), Bikbov and Chernov (2008), and Ang et al. (2008). Clarida et al. (2000) also restrict the real rate to its sample average to identify the inflation target.

To reduce the dimension of the parameter space, the variance-covariance matrix Ω^{v_t} is constrained to be a diagonal. In this setting, interactions between factors are determined

[3]We are not claiming that the HP filter actually captures potential output or the output gap. However, we assume that it proxies for the Fed's and the market's perceptions of the output gap. This approach is taken in other papers on Taylor rules, such as Cecchetti, Hooper, Kasman, Schoenholtz, and Watson (2008), which applies the HP filter for real-time data.

by the G matrix. This constraint is not too restrictive given estimation results of many studies that report statistically insignificant and, in most cases, relatively small off-diagonal elements of the variance-covariance matrix (e.g. Ang et al. (2007b), Chib and Kang (2009)).

It is well known that it is hard to estimate the risk parameters in small samples, and therefore, similarly to Ang, Bekaert, and Wei (2007a), for tractability we also constrain λ_f to be a diagonal matrix. This restriction is also in line with the empirical approach of Dai et al. (2007), who constrained most of the off-diagonal elements of the λ_f matrix to zero based on their preliminary estimation results.

In order to label monetary policy regime m_t=1 to be "more active" with respect to response to inflation than regime m_t=2, we restrict $\alpha^1 > \alpha^2$. To label volatility regime v_t=1 to have higher volatility than in regime v_t=2, we restrict $\Omega^1_{i,i} > \Omega^2_{i,i}$ for each diagonal element i. We also label market price of risk regime l_t=1 to have higher price of risk of inflation than in regime l_t=2 by restricting $\lambda^1_{0,2} < \lambda^2_{0,2}$ because more negative value of $\lambda^{s_t}_0$ is associated with higher price of risk.

The factor dynamics are assumed to be a stationary process by constraining all eigenvalues of the G matrix to be less than unity in absolute value. The recursion for $B^{s_t}_\tau$ is also restricted to be stationary to ensure that the implied yields for long-term bonds are non-explosive.

3.3 Estimation Method

No-arbitrage term-structure models are known to have a likelihood surface with many local maxima. The problem becomes more severe in our high dimensional parameter space. Our statistical inference is Bayesian, and to fit such models we use the tailored randomized block Metropolis-Hasting (TaRB-MH) algorithm recently developed by Chib and Ramamurthy (2010). The idea behind this implementation is to update parameters in blocks where both the number of blocks and the members of the blocks are randomly drawn within each MCMC cycle. The use of this MCMC method is essential to improve the mixing of the draws in the context of term structure models in which there is no natural way of grouping the parameters. For more details about the TaRB-MH algorithm, see Chib and Ramamurthy (2010).

One important feature of our estimation method is that proposal densities are constructed from the output of simulated annealing, described in detail in Goffe (1996). For our problem this stochastic optimization method is more reliable than the standard Newton-Raphson class of deterministic optimizers due to high irregularity of the likelihood surface.

3.4 State Space Form

This subsection provides details for the state space form, which comprises the transition and measurement equations and is the basis for model estimation. The transition equation of the state space form is given by equation (2.3). To derive the measurement equation, we follow Dai et al. (2007) and assume that one yield, in particular the 12 quarter maturity yield ($R_{12,t}$), is priced without error. This yield is entitled basis yield. We choose the 12 quarter maturity yield to be priced without error based on the finding in Chib and Kang (2009) that the yields in the middle of the yield curve have the lowest variance of the measurement errors. As a result, the pricing equation for this yield has the form:

$$R_{12,t} = a_{12}^{s_t} + b_{12}^{s_t\prime}\mathbf{f}_t = a_{12}^{s_t} + b_{u,12}^{s_t} u_t + b_{\overline{m},12}^{s_t\prime}\overline{m}_t \ , \tag{3.2}$$

where

$$b_{12}^{s_t} = \begin{pmatrix} b_{u,12}^{s_t} \\ b_{\overline{m},12}^{s_t} \end{pmatrix}$$

and $\overline{m}_t$ denotes the vector of macro factors $(\pi_t, g_t)'$. This assumption allows the latent factor to be expressed in terms of observable yields and macro variables:

$$u_t = \left(b_{u,12}^{s_t}\right)^{-1} \left(R_{12t} - a_{12}^{s_t} - b_{\overline{m},12}^{s_t\prime}\overline{m}_t\right). \tag{3.3}$$

Thus,

$$\mathbf{f}_t = \begin{pmatrix} u_t \\ \overline{m}_t \end{pmatrix} = \begin{pmatrix} \left(b_{u,12}^{s_t}\right)^{-1} \left(R_{12t} - a_{12}^{s_t} - b_{\overline{m},12}^{s_t\prime}\overline{m}_t\right) \\ \overline{m}_t \end{pmatrix}. \tag{3.4}$$

By denoting the vector of all yields other than R_{12t} by R_t and $y_t \equiv (R_t, \mathbf{f}_t)'$, the measurement equation can be expressed as

$$y_t = \underbrace{\begin{pmatrix} \overline{a}^{s_t} \\ \mathbf{0} \end{pmatrix}}_{\overline{A}^{s_t}} + \underbrace{\begin{pmatrix} \overline{b}^{s_t} \\ I_3 \end{pmatrix}}_{\overline{B}^{s_t}} \mathbf{f}_t + \begin{pmatrix} I_7 \\ 0_{3\times 7} \end{pmatrix} \widetilde{\varepsilon}_t \ , \qquad \widetilde{\varepsilon}_t \sim iidN\left(0, \Sigma\right) \ , \tag{3.5}$$

where Σ is the variance-covariance matrix for the measurement errors, which is assumed to be a diagonal and regime independent, and $\overline{a}^{s_t}$ and $\overline{b}^{s_t}$denote the vector and matrix of all stacked $a_\tau^{s_t}$ and $b_\tau^{s_t\prime}$ excluding $a_{12}^{s_t}$ and $b_{u,12}^{s_t}$.

3.5 Prior Distribution

We set the prior distributions of the model parameters based on the general observation that, on average, the yield curve is upward sloping. Following Chib and Ergashev (2009) we simulate parameters and model-implied yield curves from the prior distributions to ensure that our prior produces, on average, a reasonably shaped yield curve. At the same time we set the variances of key parameter distributions to be relatively large so that the distributions cover economically reasonable values of parameters. The prior for the diagonal elements of G is based on the fact that interest rates, inflation, and the output gap are all persistent time series. Since $\lambda_0^{s_t}$ and Ω^{s_t} are key parameters determining the term premium, their means are set based on the simulation outcomes of the model-implied yield curve. Full details of the prior distributions are provided in Appendix C. To show the prior implied outcomes, we sample the parameters 25,000 times from the prior distributions and simulate factor dynamics and yield curves. Figure 1 displays median, 2.5%, and 97.5% quantile surfaces of simulated yield curves and their time series averages. As Figure 1 illustrates, this simulation exercise produces, on average, a slightly upward-sloping yield curves with substantial variation.

3.6 Posterior Distribution

The posterior distributions of parameters are simulated by Markov Chain Monte Carlo (MCMC) methods. The joint posterior distribution to be simulated is described by

$$\pi\left(\theta, \mathbf{S}_T|\mathbf{y}\right) \propto f\left(\mathbf{y}|\theta, \mathbf{S}_T\right) f\left(\mathbf{S}_T|\theta\right) \pi\left(\theta\right) \ , \tag{3.6}$$

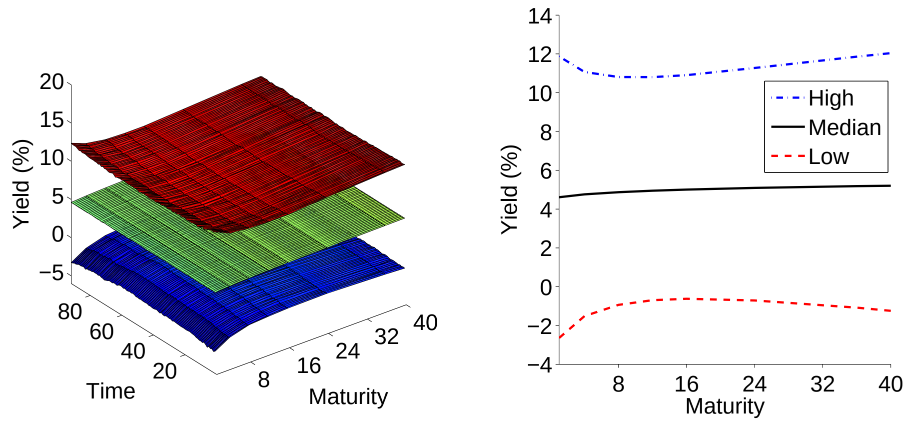

Figure 1: The prior-implied yield curves

The graphs are based on 25,000 simulations of the parameters from the prior distributions. On the left hand side are the 2.5%, 50%, and 97.5% quantile surfaces of the yield curves. The graph on the right hand side is the averaged yield curve quantiles from the graph on the left hand side.

where $f(\mathbf{y}|\theta, \mathbf{S}_T)$ is the likelihood function for data, denoted by $\mathbf{y}$ comprising time series of all yields and macro factors, given all parameters of interest θ and time series of regimes $\mathbf{S}_T = \{s_t\}_{t=0,1,..,T}$; $f(\mathbf{S}_T|\theta)$ is the density function for regime-indicators given the parameters; $\pi(\theta)$ is the prior density of the parameters.

The MCMC procedure is discussed in detail in Appendix D and summarized as follows:

Step 1: Initialize $(\theta, \mathbf{u}_T, \mathbf{S}_T)$; where $\mathbf{u}_T = \{u_t\}_{t=0...T}$ is the time series of the latent factor and $\mathbf{S}_T = \{s_t\}_{t=0...T}$ is the time series of regimes;

Step 2: Sample θ conditional on $(\mathbf{S}_T, \mathbf{F}_T, \mathbf{R}_T)$, where $\mathbf{F}_T = \{\mathbf{f}_t\}_{t=0...T}$ is the time series of factors and $\mathbf{R}_T = \{R_t\}_{t=0...T}$ is the time series of yields;

Step 3: Sample $\mathbf{S}_T$ conditional on $(\theta, \mathbf{F}_T, \mathbf{R}_T)$;

Step 4: Compute $\mathbf{u}_T$ conditional on $(\theta, \mathbf{S}_T, \overline{\mathbf{m}}_T, \mathbf{R}_{12,T})$ using equation (3.3), where $\overline{\mathbf{m}}_T = \{\overline{m}_t\}_{t=0...T}$ is the time series of macro factors and $\mathbf{R}_{12,T} = \{R_{12,t}\}_{t=0...T}$ is the time series of basis yield;

Step 5: Repeat Steps 2-4 $(n_0 + n)$ times, then disregard the first n_0 iterations, which are burn-in iterations, and save n draws of the parameters.

4 Empirical Results

4.1 Model comparisons

To confirm an importance of accounting for regime shifts in the monetary policy, volatility of yield factors, and market price of risk for fitting the data, we estimate models with different combinations of regime-processes and conduct model comparisons. We compare the model with the three regime-switching processes and models with all combination of two regime-switching processes out of the three processes using the deviance information criterion (DIC) proposed by Spiegelhalter, Best, Carlin, and van der Linde (2002).[4] Table 1 confirms that the model with the three regime-switching processes is the most supported by the data.The following subsections discuss estimation results for this model and analyze the effects of monetary policy regime shifts on the term structure of interest rates.

4.2 Parameter Estimates and Regimes

Table 2 reports the parameter estimates of the model. Specifically, the table reports the posterior means of parameters and their standard deviations in parentheses based on 15,000 iterations of the MCMC algorithm beyond a burin-in of 5,000 iterations. To evaluate the efficiency of the MCMC-produced results, we use the acceptance rates in the

---

[4]The deviance information criterion (DIC) is defined as: $DIC = 2\frac{1}{n}\sum_{i=1}^{n} D(\mathbf{y},\theta^{(i)}) - D(y,\overline{\theta})$, where $D(\mathbf{y},\theta) = -2\log f(\mathbf{y}|\theta)$, $\theta^{(i)}$ is the vector of parameters from the posterior distribution, and $\overline{\theta}$ is the mean of the posterior distribution of parameters. The model with the smallest value of DIC is the most supported by the data. The Bayesian information criterion (BIC) gives the consistent result for a model comparison. Alternative criterion for a model comparison, used widely in the Bayesian literature, is the Bayes factor, which is based on the marginal likelihood. However, given big values of log likelihoods due to the scale of the data and the model specification used for this study, the computation of values of likelihoods is numerically infeasible. Therefore, the majority of methods to compute marginal likelihoods based on values of likelihoods (for example, harmonic mean estimator) cannot be used for this study. The method for estimating the marginal log likelihood proposed by Chib and Jeliazkov (2001) is computationally costly for our study.

Table 1: The deviance information criterion (DIC) and Log likelihood

Model	DIC	LnL
Regimes: m_t, v_t, l_t	-11618.7	5830.4
Regimes: m_t, v_t	-11513.6	5675.8
Regimes: m_t, l_t	-11115.7	5608.7
Regimes: v_t, l_t	-11446.8	5696.6

m_t, v_t, and l_t denote regimes of monetary policy, volatility, and the market price of risk, respectively. The model with the smallest value of the DIC is the most supported by the data. LnL denotes log likelihood evaluated at the mode of the posterior distribution.

MH step of the sampler and the inefficiency factor as discussed in Chib (2001).[5] These parameters have, on average, values of 53.7 percent and 180.0 respectively indicating good mixing.

We start the interpretation of the estimation results with analysis of the parameter estimates in the two monetary policy regimes. The inflation coefficients α^1 and α^2, which have values of 0.18 and 0.88, respectively, are considerably different in the two monetary policy regimes. The output gap coefficients β^1 =0.63 and β^2 =0.75 are also different in the two monetary policy regimes; however, this difference is not as strong as for the inflation coefficients. Thus, the monetary policy regimes are mainly identified by switching in the Fed's reaction to inflation.

These coefficients are not directly comparable to those from a single-equation Taylor rule that accounts for interest rate smoothing. The single-equation Taylor rule with interest rate smoothing is specified as a linear combination of the target rate and past value of the short rate as

$$r_t^{m_t} = (1-\rho)\left[\widetilde{r}^{m_t} + \widetilde{\alpha}^{m_t}\left(\pi_t - \overline{\pi}^{m_t}\right) + \widetilde{\beta}^{m_t} g_t\right] + \rho r_{t-1}^{m_{t-1}} + \xi_t \ , \qquad (4.1)$$

where ξ_t denotes monetary policy shocks for this specification of the policy rule. It is easy to see that $\widetilde{r}^{m_t} = \frac{\overline{r}^{m_t}}{(1-\rho)}$, $\widetilde{\alpha}^{m_t} = \frac{\alpha^{m_t}}{(1-\rho)}$, $\widetilde{\beta}^{m_t} = \frac{\beta^{m_t}}{(1-\rho)}$, $u_t = \rho r_{t-1}^{m_{t-1}} + \xi_t$, and it is

[5] The inefficiency factor is defined as $1+2\sum_{k=1}^{M}\rho(k)$, where $\rho(k)$ is the k-order autocorrelation computed from the sampled distribution and M is a large number, which we set to be 500. Thus, if the sampler did not mix at all then the inefficiency factor would have a value of 500. Given this choice for M, empirically, a value of the inefficiency factor of 250 is usually considered as an upper-bound for a reasonable level of mixing.

Table 2: Parameter estimates

(a) Monetary Policy

α^1	α^2	β^1	β^2
0.178	0.882	0.628	0.750
(0.098)	(0.164)	(0.167)	(0.226)

(b) G matrix

G		
0.946	0.006	0.016
(0.030)	(0.010)	(0.026)
-0.039	0.958	0.041
(0.036)	(0.023)	(0.046)
0.133	0.014	0.838
(0.036)	(0.030)	(0.039)

(c) Factors' Volatilities $\times 400$

L^1			L^2		
0.692	0.688	0.411	0.739	1.154	0.668
(0.065)	(0.060)	(0.053)	(0.068)	(0.164)	(0.097)

(d) Measurement Errors' Volatilities $\times 400$

σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
0.438	0.174	0.052	0.026	0.064	0.115	0.132
(0.038)	(0.015)	(0.004)	(0.002)	(0.006)	(0.009)	(0.011)

(e) Market Price of Risks

λ_0^1			λ_0^2			λ_f		
0.237	-0.342	-0.374	0.193	-0.442	-0.498	0.314	0.733	0.251
(0.060)	(0.086)	(0.155)	(0.076)	(0.103)	(0.176)	(1.997)	(1.974)	(1.837)

(f) Transition Probabilities

p_m^{11}	p_m^{22}	p_v^{11}	p_v^{22}	p_l^{11}	p_l^{22}
0.988	0.986	0.943	0.959	0.978	0.975
(0.004)	(0.004)	(0.016)	(0.016)	(0.007)	(0.009)

The Table reports posterior means and their standard deviations in parentheses based on 15,000 posterior draws beyond 5,000 draws as a burn-in.

easy to show that $\rho = G_{1,1}$.[6] After this transformation the coefficients $\widetilde{\alpha}^1$ =3.30 and $\widetilde{\alpha}^2$ =16.33 both have values grater than unity, and therefore they do not potentially

[6] We do not use the specification of the Taylor with smoothing because, in our structure, the short rate has an affine form in the factors and also the latent factor is identified from the VAR(1) dynamics rather than from the single short-rate equation.

create a risk of indeterminacy of the equilibrium.[7] Given this result, the regime with the smaller inflation coefficient is entitled a “less active” monetary policy regime and the one with the bigger coefficient, a “more active” regime. The transformed coefficients for the output gap $\widetilde{\beta}^1$ and $\widetilde{\beta}^2$ have values of 11.63 and 13.89, respectively. In our model structure, the policy response coefficients are responsible for fitting the short rate as well as the long-term interest rate through a no-arbitrage restriction rather than only the short rate in the single-equation Taylor rule. Therefore, this model structure can lead to different estimates of the coefficients than those from the single-equation model.[8]

Figure 2 displays the probabilities of regimes for all three regime processes. In general, the monetary policy regimes are well-identified and very persistent throughout the sample period with 99 percent probabilities of staying in the same regime from quarter to quarter, as reported in Table 2. The period from 1986 through 1994 is characterized by the “more active” monetary policy regime. In this period, inflation was, on average, relatively high and the Fed was adjusting the short rate relatively close to inflation and output gap dynamics. The period from 1995 through 2000, where the “less active” monetary policy regime prevails, is characterized by the relatively stable short rate and inflation, while the output gap was steadily increasing in magnitude. At the beginning of 2001, when the recession hit the U.S. economy, the Fed responded to the decline in output and inflation by reducing the short rate and switching to the “more active” policy regime, which lasted until 2004. In the period from 2002 through 2004, inflation remained, on average, relatively low and the Fed kept the short rate at a low level to accommodate the still low output gap. The identification of the monetary policy regime in this period as “more active” is also affected by the increased term spread. As we noted above, in the no-arbitrage framework, the Taylor rule coefficients are identified by the short rate as well as the slope of the yield curve.

Identification of monetary policy as “less active” for the period from the middle

[7] As discussed in Clarida et al. (2000), if the inflation coefficients are below unity, then increase in expected inflation causes a decline in the real interest rate. The decline in the real interest rate leads to growth in aggregate consumption, which consequently leads to further increase in inflation.

[8] Although the estimates of the policy response coefficients for inflation and output gap after transformation are higher than those often reported from a single-equation Taylor rule model, they are of the same magnitude as those reported by ABDL(2010) for their specification of a no-arbitrage model.

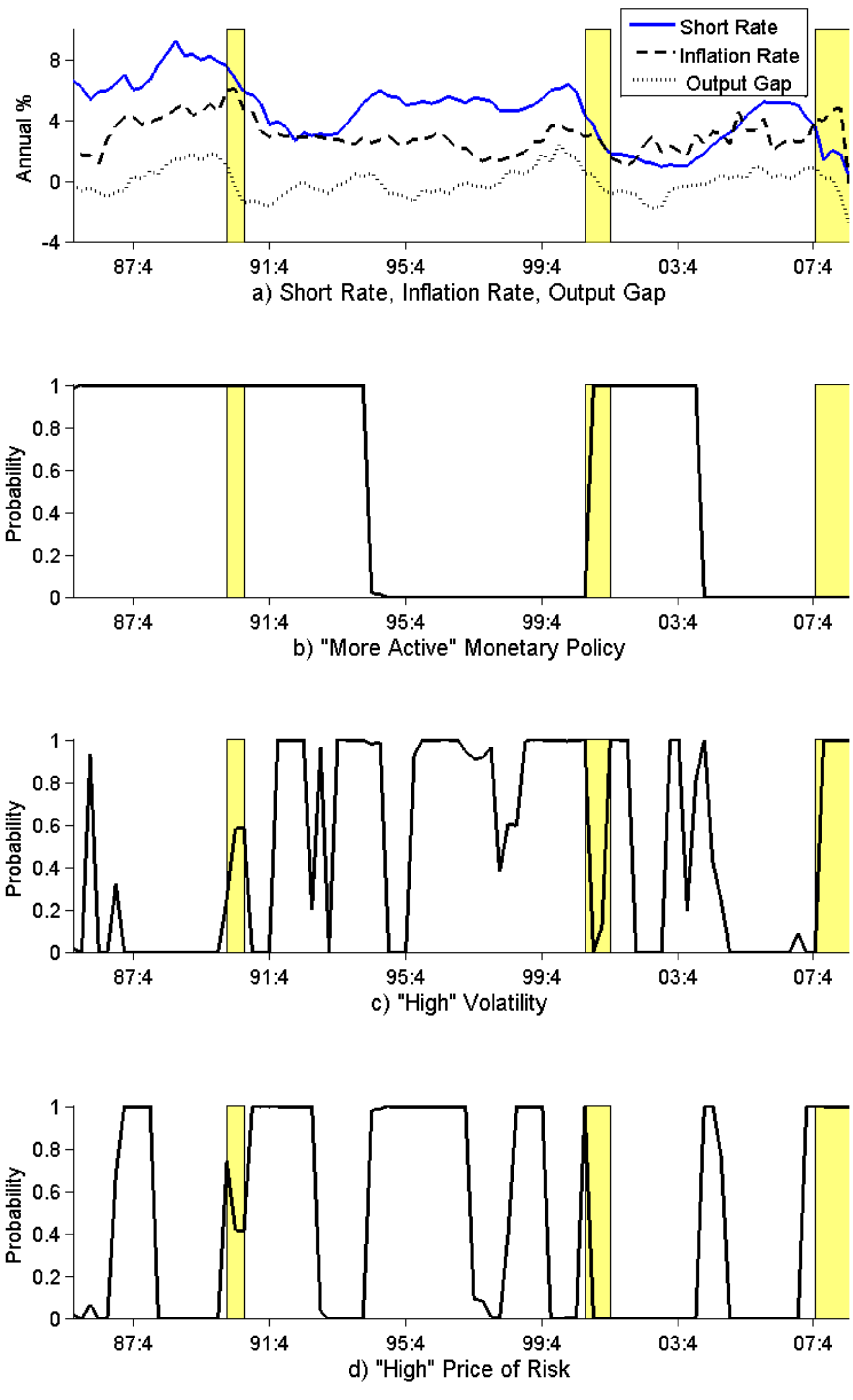

Figure 2: The Probabilities of monetary policy, volatility, and risk regimes
Graph (a) displays the time series of the short rate, inflation and the output gap; graphs (b), (c), and (d) display probabilities of regimes in "more active" monetary policy, "high" volatility, and "high" price of risk, respectively. Shaded areas correspond to NBER recession dates.

of 2004 through 2005 is also affected by the slope of the yield curve. In this period, entitled a "conundrum" by then-Fed Chairman Alan Greenspan, the long-term yields slightly declined while the short rate was steadily increasing from 1 percent to around 4 percent. These dynamics of the yield curve, as discussed by Rudebusch, Swanson, and Wu (2006) in detail, are perceived to be unusual given economic expansion, the falling unemployment rate, and the increasing fiscal gap, which all normally correspond a higher long rate. Similar to Kim and Wright (2005), our results suggest that the term premium, displayed in Figure 3, was low in this period. While this result suggests that part of the "conundrum" can be related to a decline in the term premium, full assessment of its contribution to the pricing anomaly is beyond the scope of this study.[9]

The volatility estimates of exogenous shocks to all factors, reported in Table 2 suggest that identification of the volatility regimes is presumably driven by the volatility of inflation shocks. The volatility estimates for the inflation shocks factored by 400 have values of 0.69 and 1.15 - the values with the largest difference in the two volatility regimes among all factors. The transition probabilities of staying in the same volatility regime are estimated at 94 and 98 percent for the "low" and "high" volatility regimes, respectively.

The bottom graph of Figure 2 displays probabilities of the "high" price of risk regime based on switching of risk parameters $\lambda_0^{s_{t+1}}$. While risk parameter $\Lambda_{t,t+1}$ has the continuously time-varying component as a function of the factors, one can see from this graph and Figure 3 that the regime-switching of the risk parameters is closely related to the term spread dynamics, indicating the importance of its regime-switching for better fitting of the term structure of interest rates. Also, as we pointed earlier, the model comparisons suggest that accounting for the regime-switching of the risk parameters considerably improves the data fitting by the model.

[9]Kim and Wright (2005) finds that the decline in term premium is a key factor explaining the "conundrum". In contrast, Rudebusch et al. (2006) find that no arbitrage macro-finance models are not able to explain it. They consider macroeconomic factors other than those included in the macro-finance models and find that declines in long-term bond volatility may explain a part of the "conundrum".

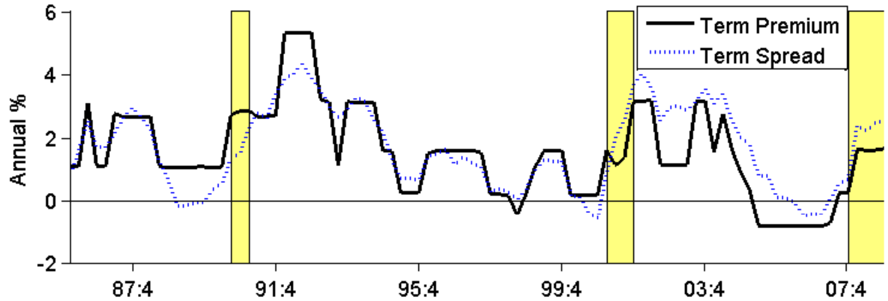

Figure 3: The Term Premium

The figure displays the model-implied term premium and the term spread for 10-year bonds. Shaded areas corresponds to NBER recession dates.

4.3 Monetary Policy Regimes and the Yield Curve

Figure 4 displays the average realized yield curves in the two monetary policy regimes. The left-hand-side graph demonstrates that the average yield curves in the two regimes mainly differ in terms of their long rates and slopes. In particular, while the average short rates in two regimes are close to each other, the long rate in the "more active" regime is, on average, 129 basis points higher than in the "less active" regime, resulting in a considerably steeper sloped yield curve, on average, in the "more active" regime. This result suggests that long-term yields are more sensitive to monetary policy shifts than the short-rate, which is in line with findings of ABDL(2010) and can be explained as follows. Because the policy coefficients switch to higher values in response to greater macroeconomic factor risk in the "more active" regime, they also magnify this risk for the long-term yields through a no-arbitrage restriction. The middle- and right-hand-side graphs of Figure 4 demonstrate that the short rate in the "more active" regime was considerably more volatile than in the "less active" regime. The sample standard deviation of the short rate in the "more active" regime is 2.48 percent compared to 1.39 percent in the "less active" regime. In general, the yield curve in the "more active" regime is more volatile than in the "less active" regime with the standard deviations of the long-term yields of 1.65 and 1.10 percent in the "more active" and "less active" regimes, respectively. In summary, these results can be explained by a more sensitive response of the short rate to inflation in the "more active" regime that creates higher

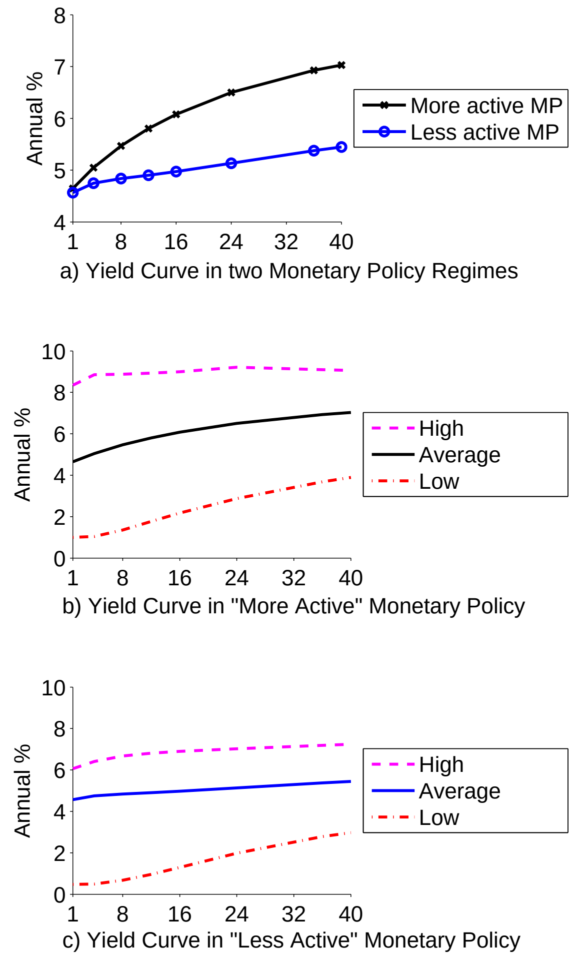

Figure 4: Average realized yield curves

The graphs are constructed using the term structure of interest rates computed at each iteration of the posterior distribution and then separately averaging them over the two monetary policy regimes. Graphs (b) and (c) display the average and 2.5%, and 97.5% quantile yield curves in the two monetary policy regimes.

risk for the future short rate fluctuations. This risk drives the higher long-term rate relative to the short rate. Thus, the Fed faces a policy trade-off between a "more active" reaction to macroeconomic fluctuations and a more volatile yield curve caused by this reaction. This argument is consistent with Woodford (1999), who claims that it may be more optimal for the monetary authority to conduct policies that do not require the short rate to be too volatile.

To see what effect monetary policy would have had on the term structure of interest

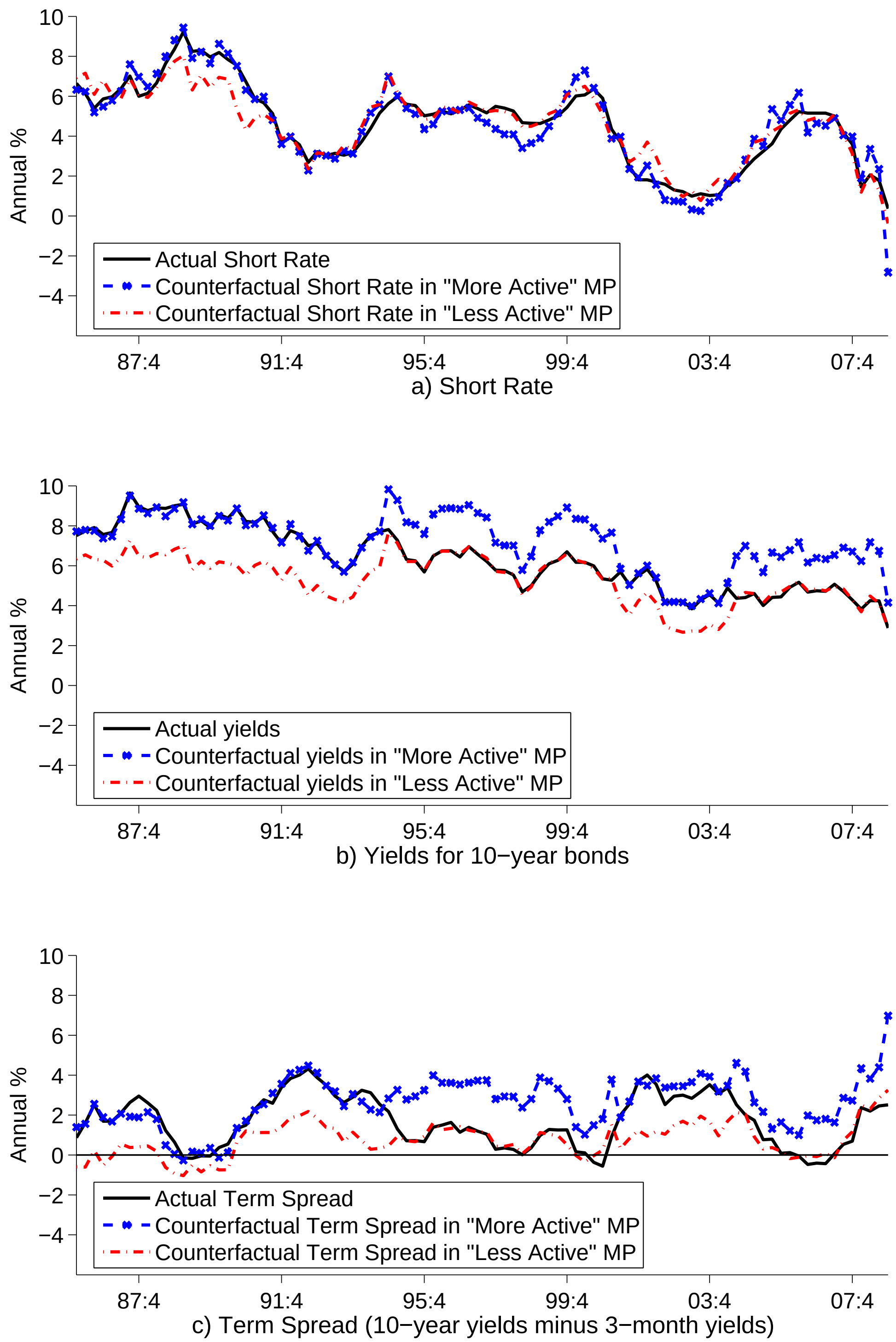

Figure 5: Counterfactual short rates, long rates, and term spreads
The time series of counterfactual interest rates are simulated by fixing parameters to one of the two monetary policy regimes.

rates if a single regime were maintained throughout the sample, we conduct a counterfactual analysis. Figure 5 displays the short and long rates and the term spreads generated by fixing parameters to one of the two monetary policy regimes. Throughout most of the sample, the short rate in the "more active" regime would have been more volatile than in the "less active" regime. The long rate and consequently the term spread would have been higher than the actual ones in those periods when the regime was "less active".

5 Conclusion

In this paper, we proposed a no-arbitrage affine term structure model with regime shifts in monetary policy, factor volatilities, and the price of risk. This model allowed us to quantitatively assess the influence of monetary policy regime shifts on the entire term structure of interest rates.

We found that, in the "more active" monetary policy regime, the slope of the yield curve was steeper than in the "less active" regime. Also, the short rate and the entire yield curve in general were more volatile in the "more active" regime than in the "less active" regime. The explanation for these results is that a higher sensitivity of the short rate in response to inflation fluctuations in the "more active" regime leads to a higher term premium in anticipation of a more volatile future short rate. These results also suggest that the Fed faces a policy trade-off between a "more active" reaction to macroeconomic fluctuations and a more volatile yield curve caused by this reaction.

Appendices

A Bond Pricing

We solve for A_τ^j and B_τ^j using the law of iterated expectatios, method of undetermined coefficients, and log-linearization:

$$P_{t,\tau}^{s_t} = \mathbb{E}\left[\exp\left(-r_t^{s_t} - \frac{1}{2}\Lambda_t^{s_{t+1}\prime}\Lambda_t^{s_{t+1}} - \Lambda_t^{s_{t+1}\prime}\varepsilon_{t+1}\right) P_{\tau-1,t+1}^{s_{t+1}}|\mathbf{f}_t, s_t\right]$$

$$\begin{aligned}
1 &= \mathbb{E}\left[\exp\left(-r_t^j - \frac{1}{2}\Lambda_t^{s_{t+1}\prime}\Lambda_t^{s_{t+1}} - \Lambda_t^{s_{t+1}\prime}\varepsilon_{t+1}\right) \frac{P_{\tau-1,t+1}^{s_{t+1}}}{P_{\tau,t}^j}|\mathbf{f}_t, s_t = j\right] \\
&= \sum_{k=1}^S p^{jk}\mathbb{E}\left[\exp\left(-r_t^j - \frac{1}{2}\Lambda_t^{k\prime}\Lambda_t^k - \Lambda_t^{k\prime}\varepsilon_{t+1}\right) \frac{P_{\tau-1,t+1}^{k}}{P_{\tau,t}^j}|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^S p^{jk}\mathbb{E}\left[\exp\left(\begin{array}{c} -r_t^j - \frac{1}{2}\Lambda_t^{k\prime}\Lambda_t^k - \Lambda_t^{k\prime}\varepsilon_{t+1} \\ +A_\tau^j + B_\tau^{j\prime}\mathbf{f}_t - A_{\tau-1}^k - B_{\tau-1}^{k\prime}\mathbf{f}_{t+1}\end{array}\right)|\mathbf{f}_t, s_t = j, s_{t+1} = k\right] \\
&= \sum_{k=1}^S p^{jk}\left\{\begin{array}{c} \exp\left(-r_t^j - \frac{1}{2}\Lambda_t^{k\prime}\Lambda_t^k + A_\tau^j - A_{\tau-1}^k + B_\tau^{j\prime}\mathbf{f}_t - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right) \\ \times\mathbb{E}\left[\exp\left(-\left(\Lambda_t^{k\prime} + B_{\tau-1}^{k\prime}L^k\right)\varepsilon_{t+1}\right)|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]\end{array}\right\} && \text{(A.1)} \\
&= \sum_{k=1}^S p^{jk}\left\{\exp\left(\begin{array}{c} -r_t^j - \frac{1}{2}\Lambda_t^{k\prime}\Lambda_t^k + A_\tau^j - A_{\tau-1}^k + B_\tau^{j\prime}\mathbf{f}_t - B_{\tau-1}^{k\prime}\mu_t^{j,k} \\ +\frac{1}{2}\left(\Lambda_t^{k\prime} + B_{\tau-1}^{k\prime}L^k\right)\left(\Lambda_t^{k\prime} + B_{\tau-1}^{k\prime}L^k\right)^\prime\end{array}\right)\right\} && \text{(A.2)} \\
&= \sum_{k=1}^S p^{jk}\exp\left(\begin{array}{c} -r_t^j + A_\tau^j - A_{\tau-1}^k + B_\tau^{j\prime}\mathbf{f}_t \\ -B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k + \frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k\end{array}\right) && \text{(A.3)} \\
&\approx \sum_{k=1}^S p^{jk}\left\{\begin{array}{c} -\delta_0^j - \delta_f^{j\prime}\mathbf{f}_t + A_\tau^j - A_{\tau-1}^k + B_\tau^{j\prime}\mathbf{f}_t \\ -B_{\tau-1}^{k\prime}d^k - B_{\tau-1}^{k\prime}G\left(\mathbf{f}_t - d^j\right) \\ +B_{\tau-1}^{k\prime}L^k\left(\lambda_0^k + \lambda_f\mathbf{f}_t\right) + \frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k + 1\end{array}\right\}. && \text{(A.4)}
\end{aligned}$$

$(A.1)$ is transformed into $(A.2)$ using the property of moment generating function for Normally distributed ε_{t+1}:

$$\varphi_t^{jk}(x) \equiv \mathbb{E}\left[\exp\left(x'\varepsilon_{t+1}\right)|\mathbf{f}_t, s_t = j, s_{t+1} = k\right] = \exp(\frac{x'x}{2})\ ,\ x \in \mathbb{R}^3$$

evaluated at $x = -\left(\Lambda_t^{k\prime} + B_{\tau-1}^{k\prime}L^k\right)^\prime$. Following Bansal and Zhou (2002), $(A.3)$ is transformed into $(A.4)$ using log-approximation $\exp(y) \approx y + 1$ for a sufficiently small y and substituting for r_t^j using equation (2.11).

Using above result for the bond pricing equation and collecting terms for $\mathbf{f}_t$:

$$\begin{aligned}
0 &= \sum_{k=1}^{S}\left\{p^{jk}\mathbb{E}\left[\exp\left(-r_t^j-\frac{1}{2}\Lambda_t^{k\prime}\Lambda_t^k-\Lambda_t^{k\prime}\varepsilon_{t+1}\right)\frac{P_{t+1,k}^{\tau-1}}{P_{t,j}^{\tau}}|\mathbf{f}_t,s_t=j,s_{t+1}=k\right]\right\}-1 \\
&\approx \sum_{k=1}^{S}p^{jk}\left(\begin{array}{c}-\delta_0^j-\delta_f^{j\prime}\mathbf{f}_t+A_\tau^j-A_{\tau-1}^k+B_\tau^{j\prime}\mathbf{f}_t-B_{\tau-1}^{k\prime}d^k-B_{\tau-1}^{k\prime}G\left(\mathbf{f}_t-d^j\right)\\ +B_{\tau-1}^{k\prime}L^k\left(\lambda_0^k+\lambda_f\mathbf{f}_t\right)+\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k\end{array}\right) \\
&= \sum_{k=1}^{S}p^{jk}\left(\begin{array}{c}-\delta_0^j+A_\tau^j-A_{\tau-1}^k-B_{\tau-1}^{k\prime}d^k+B_{\tau-1}^{k\prime}Gd^j\\ +B_{\tau-1}^{k\prime}L^k\lambda_0^k+\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k\end{array}\right) \\
&\quad+\sum_{k=1}^{S}p^{jk}\left(-\delta_f^{j\prime}+B_\tau^{j\prime}-B_{\tau-1}^{k\prime}G+B_{\tau-1}^{k\prime}L^k\lambda_f\right)\mathbf{f}_t\ .
\end{aligned}$$

The above identity has to be true for every value of $\mathbf{f}_t$, which will be the case only if the first and second terms are 0:

$$0 = \sum_{k=1}^{S}p^{jk}\left(\begin{array}{c}-\delta_0^j+A_\tau^j-A_{\tau-1}^k-B_{\tau-1}^{k\prime}d^k+B_{\tau-1}^{k\prime}Gd^j\\ +B_{\tau-1}^{k\prime}L^k\lambda_0^k+\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k\end{array}\right)$$

and

$$0 = \sum_{k=1}^{S}p^{jk}\left(-\delta_f^{j\prime}+B_\tau^{j\prime}-B_{\tau-1}^{k\prime}\left(G-L^k\lambda_f\right)\right).$$

This leads to the solution for A_τ^j and B_τ^j in the form of recursive system:

$$\begin{aligned}
A_\tau^j &= \delta_0^j+\sum_{k=1}^{S}p^{jk}\left(A_{\tau-1}^k+\left(d^k-Gd^j-L^k\lambda_0^k\right)'B_{\tau-1}^k-\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^k\right) \\
B_\tau^j &= \delta_f^j+\sum_{k=1}^{S}p^{jk}\left(G-L^k\lambda_f\right)'B_{\tau-1}^k\ .
\end{aligned}$$

To derive the initial conditions for A_0^j and B_0^j, we let $\tau=0$. Given $P_{\tau,t}^j=\exp(-\tau r_t^j)$, we have $P_{0,t}^j=\exp(-0\times r_t^j)=1$. From $P_{j,t}^\tau=\exp(-A_\tau^j-B_\tau^{j\prime}\mathbf{f}_t)$ for $\tau=0:1=P_{0,t}^j=\exp(-A_0^j-B_0^{j\prime}\mathbf{f}_t)$ has to be true for every $\mathbf{f}_t$, therefore $A_0^j=0$ and $B_0^j=\mathbf{0}$, consequently $A_1^j=\delta_0^j$ and $B_1^j=\delta_f^j$.

B Expected Excess Return

The one-period expected excess return on the n-period bond:

$$\mathrm{ER}_{\tau,t}^j=\mathbb{E}[\overline{p}_{\tau-1,t+1}|\mathbf{f}_t,s_t=j]+\overline{p}_{1,t}^j-\overline{p}_{\tau,t}^j\ ,$$

where $\overline{p}^j_{\tau,t}$ and $\overline{p}^j_{1,t}$ are log prices of bonds derived in the following ways:

$$\begin{aligned}
\overline{p}^j_{\tau,t} &= \mathsf{log}\, P^j_{\tau,t} = \mathsf{log}\,\mathbb{E}\left[\mathsf{exp}\left(-r^j_t - \frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t - \Lambda^{k\prime}_t\varepsilon_{t+1}\right)P_{\tau-1,t+1}|\mathbf{f}_t, s_t = j\right] \\
&= -r^j_t + \mathsf{log}\left(\sum_{k=1}^S p^{jk}\mathbb{E}\left[\mathsf{exp}\left(-\frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t - \Lambda^{k\prime}_t\varepsilon_{t+1}\right)P^k_{t+1,\tau-1}|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]\right) \\
&= -r^j_t + \mathsf{log}\left(\sum_{k=1}^S p^{jk}\mathbb{E}\left[\begin{array}{c}\mathsf{exp}(-\frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t - \Lambda^{k\prime}_t\varepsilon_{t+1} \\ -A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mathbf{f}_{t+1})|\mathbf{f}_t, s_t = j, s_{t+1} = k\end{array}\right]\right) \\
&= -r^j_t + \mathsf{log}\left(\begin{array}{c}\sum_{k=1}^S p^{jk}\,\mathsf{exp}\left(-\frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t - A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t\right) \\ \times\mathbb{E}\left[\mathsf{exp}(-\left(\Lambda^{k\prime}_t + B^{k\prime}_{\tau-1}L^k\right)\varepsilon_{t+1})|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]\end{array}\right) \\
&= -r^j_t + \mathsf{log}\left(\sum_{k=1}^S p^{jk}\,\mathsf{exp}\left(\begin{array}{c}-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t \\ +B^{k\prime}_{\tau-1}L^k\Lambda^k_t + \frac{1}{2}B^{k\prime}_{\tau-1}L^kL^{k\prime}B^k_{\tau-1}\end{array}\right)\right)
\end{aligned}$$

and

$$\overline{p}^j_{t,1} = \mathsf{log}\left(\mathsf{exp}\left(-r^j_t\right)\right) = -r^j_t\ .$$

Then the expected value of the log price is given by

$$\begin{aligned}
\mathbb{E}[\overline{p}_{\tau-1,t+1}|\mathbf{f}_t, s_t &= j] = \sum_{k=1}^S p^{jk}\mathbb{E}[\overline{p}^k_{\tau-1,t+1}|\mathbf{f}_t, s_t = j, s_{t+1} = k] \\
&= \sum_{k=1}^S p^{jk}\left(-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mathbb{E}\left[\mathbf{f}_{t+1}|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]\right) \\
&= \sum_{k=1}^S p^{jk}\left(-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t\right)\ .
\end{aligned}$$

Next, the expected excess return is derived in the following way:

$$\begin{aligned}
\mathbb{E}[\overline{p}_{\tau-1,t+1}|\mathbf{f}_t, s_t &= j] + \overline{p}^j_{1,t} - \overline{p}^j_{\tau,t} \\
&= \sum_{k=1}^S p^{jk}_t\left(-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t\right) - r^j_t \\
&\quad - \left\{-r^j_t + \mathsf{log}\left(\sum_{k=1}^S p^{jk}\,\mathsf{exp}\left(\begin{array}{c}-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t + B^{k\prime}_{\tau-1}L^k\Lambda^k_t \\ +\frac{1}{2}B^{k\prime}_{\tau-1}L^kL^{k\prime}B^k_{\tau-1}\end{array}\right)\right)\right\} \\
&= \sum_{k=1}^S p^{jk}_t\left(-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t\right) \\
&\quad - \mathsf{log}\left(\sum_{k=1}^S p^{jk}\,\mathsf{exp}\left(\begin{array}{c}-A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mu^{j,k}_t + B^{k\prime}_{\tau-1}L^k\Lambda^k_t \\ +\frac{1}{2}B^{k\prime}_{\tau-1}L^kL^{k\prime}B^k_{\tau-1}\end{array}\right)\right)
\end{aligned}$$

$$\approx \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right)$$

$$- \log \sum_{k=1}^{S} p^{jk}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} + 1 \end{array}\right)$$

$$\approx \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right)$$

$$- \sum_{k=1}^{S} p^{jk}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} \end{array}\right)$$

$$= -\sum_{k=1}^{S} p^{jk}\left(B_{\tau-1}^{k\prime}L^k\Lambda_t^k + \frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k}\right) .$$

To derive the above result, we applied log-linearization for $\exp(y)$ and $\log(x)$. The argument of the exponent is a return, which is a sufficiently small number, therefore it can be approximated as $\exp(y) \approx y + 1$. $\sum_{k=1}^{S} p^{jk}(y+1) \equiv x$ is a number sufficiently close to 1, therefore it can be approximated as $\log(x) \approx x - 1$.

C Details for the Prior Distributions

First, we describe the approach for estimating the transition probabilities.We estimate the transition probabilities separately for each regime process as functions of Normally distributed parameters

$$p_{rg}^{jk} = \frac{1}{1 + \exp\left(\eta_{rg}^{jk}\right)}, j \neq k \ , \tag{C.1}$$

which truncates the transition probability values to be within 0 and 1 bounds.

We assume that all parameters, denoted as θ, are distributed independently from each other. Table 3 provides detail for the prior distributions of the parameters. We set the prior for all variances to be defuse to ensure that the prior implied yield curve and the factor processes have considerable variations. Parameters Ω^1, Ω^2, Σ are reparameterized using coefficients

$$d_\Omega = \left(\begin{array}{ccc} 5\times 10^5 & 5\times 10^5 & 7\times 10^4 \end{array}\right) \tag{C.2}$$

and

$$d_\Sigma = \left(\begin{array}{ccccccc} 7\times 10^5 & 4\times 10^6 & 3\times 10^7 & 6\times 10^7 & 10^7 & 10^7 & 10^7 \end{array}\right) . \tag{C.3}$$

Table 3: Prior distributions

Parameter	density	mean			Std.		
α^1, α^2	normal	0.40	0.40		1.00	1.00	
β^1, β^2	normal	0.30	0.30		1.00	1.00	
G	normal	0.80	0.00	0.00	0.20	0.10	0.10
		0.00	0.80	0.00	0.10	0.20	0.10
		0.00	0.00	0.80	0.10	0.10	0.20
λ_0^1	normal	-0.10	-0.10	-0.10	0.30	0.30	0.30
λ_0^2	normal	-0.10	-0.10	-0.10	0.30	0.30	0.30
λ_f	normal	1.00	1.00	1.00	2.00	2.00	2.00
η_m^{12}, η_m^{21}	normal	3.48	3.48		0.50	0.50	0.50
η_v^{12}, η_v^{21}	normal	3.48	3.48		0.50	0.50	0.50
$\eta_\lambda^{12}, \eta_\lambda^{21}$	normal	3.48	3.48		0.50	0.50	0.50
$d_\Omega \times \Omega^1, d_\Omega \times \Omega^2$	defuse prior	1.10	1.10		0.23	0.23	
$d_\Sigma \times \Sigma$	defuse prior	1.00			0.17		

All elements of the reparameterized $d_\Omega \times \Omega^1, d_\Omega \times \Omega^2$, and $d_\Sigma \times \Sigma$ matrices have the same prior means and standard deviations within each matrix stated in the Table, where d_Ω and d_Σ are defined by (C.2) and (C.3).

D MCMC Sampling

This Section provides details of the MCMC algorithm summarized in Section 3.6 and the construction of the likelihood function.

Step 2: Sampling θ

Parameters θ conditional on $(\mathbf{S}_T, \mathbf{F}_T, \mathbf{R}_T)$ are sampled using the Metropolis-Hastings (MH) algorithm. Because it is difficult to find an optimal parameter blocking scheme due to the high dimension of parameter space of the model, we use the tailored randomized block M-H (TaRB-MH) method developed by Chib and Ramamurthy (2010). The general idea of this method is in setting a number and composition of blocks randomly in each sampling iteration. We let the proposal density $q(\theta_i|\theta_{-i}, \mathbf{y})$ for parameters θ_i in the ith block, conditional on the value of parameters in the remaining blocks θ_{-i} to take the form of a multivariate student t distribution with 15 degrees of freedom

$$q(\theta_i|\theta_{-i}, \mathbf{y}) = St\left(\theta_i|\widehat{\theta}_i, V_{\widehat{\theta}_i}, 15\right) \ ,$$

where

$$\widehat{\theta}_i = \arg\max_{\theta_i} \ln\{f(\mathbf{y}|\theta_i, \theta_{-i}, \mathbf{S_T})\pi(\theta_i)\}$$
$$\text{and } V_{\widehat{\theta}_i} = \left(-\frac{\partial^2 \ln\{f(\mathbf{y}|\theta_i, \theta_{-i}, \mathbf{S_T})\pi(\theta_i)\}}{\partial\theta_i\partial\theta_i'}\right)^{-1}_{|\theta_i=\widehat{\theta}_i}.$$

Following Chib and Kang (2009) and Chib and Ergashev (2009), we solve numerical optimization problem using the simulated annealing algorithm, which has better performance in this problem than deterministic optimization routines due to high irregularity of the likelihood surface.

Next, we draw a proposal value $\theta_i^{\dagger}$ from the multivariate student t distribution with 15 degrees of freedom, mean $\widehat{\theta}_i$ and variance $V_{\widehat{\theta}_i}$. If the proposed value does not satisfy the model imposed constrains, then it is immediately rejected. The proposed value, satisfying the constraints, is accepted as the next value in the Markov chain with probability

$$\alpha\left(\theta_i^{(g-1)},\theta_i^{\dagger}|\theta_{-i},\mathbf{y}\right)$$
$$= \min\left\{\frac{f\left(\mathbf{y}|\theta_i^{\dagger},\theta_{-i},\mathbf{S}_T\right)\pi\left(\theta_i^{\dagger}\right)}{f\left(\mathbf{y}|\theta_i^{(g-1)},\theta_{-i},\mathbf{S}_T\right)\pi\left(\theta_i^{(g-1)}\right)}\frac{St\left(\theta_i^{(g-1)}|\widehat{\theta}_i,V_{\widehat{\theta}_i},15\right)}{St\left(\theta_i^{\dagger}|\widehat{\theta}_i,V_{\widehat{\theta}_i},15\right)},1\right\},$$

where g is an index for the current iteration. The completed simulation of θ in the gth iteration with h_g blocks produces sequentially updated parameters in all blocks:

$$\pi\left(\theta_1|\theta_{-1},y,\mathbf{S}_T\right),\pi\left(\theta_2|\theta_{-2},y,\mathbf{S}_T\right),...,\pi\left(\theta_{h_g}|\theta_{-h_g},y,\mathbf{S}_T\right).$$

Now we derive the log-likelihood function conditional on θ and $\mathbf{S}_T$, which has the form:

$$\log f\left(\mathbf{y}|\theta,\mathbf{S}_T\right) = \sum_{t=1}^{T}\log f(y_t|I_{t-1},\theta,\mathbf{S}_T),$$

where $I_{t-1} = \{y_n\}_{n=0}^{t-1}$ denotes the information set available for the econometricians at time t-1. Given the model specification, y_t conditional on $s_{t-1} = j, s_t = k, I_{t-1}$, and θ is distributed Normally with the mean and variance defined as

$$y_{t|t-1}^{jk} \equiv E\left[y_t|s_{t-1}=j,s_t=k,I_{t-1},\theta\right] = \overline{A}^k + \overline{B}^k\mu_{t-1}^{j,k}$$

$$V_{t|t-1}^{jk} \equiv Var\left[y_t|s_{t-1}=j, s_t=k, I_{t-1}, \theta\right] = \overline{B}^k L^k L^{k\prime} \overline{B}^{k\prime} + \underbrace{\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}}_{\overline{W}} .$$

Thus, the conditional density of y_t becomes

$$f\left(y_t|s_{t-1}=j, s_t=k, I_{t-1}, \theta\right) = \frac{1}{(2\pi)^{10/2}\,|V_{t|t-1}^{jk}|^{1/2}} \left(-\frac{1}{2}\left(y_t - y_{t|t-1}^{jk}\right)' \left[V_{t|t-1}^{jk}\right]^{-1}\left(y_t - y_{t|t-1}^{jk}\right)\right) . \quad \text{(D.1)}$$

Step 3: Sampling regimes $\mathbf{S}_T$

Regimes $\mathbf{S}_T$ are sampled from $f\left(\mathbf{S}_T|I_T, \theta\right)$ in a single block in backward order. First, the regime probabilities conditional on I_t and θ are obtained by applying the filtering procedure developed by Hamilton (1989) as follows:

Step 1: Probabilities of regime s_0 conditional on available information at time $t = 0$ and parameters are initialized at unconditional probabilities of regimes denoted by $p_{steady-state}$:

$$\Pr\left(s_0|I_0, \theta\right) = p_{steady-state} .$$

Step 2: The joint density of s_{t-1} and s_t conditional on information at time $t-1$ and parameters is given by

$$\Pr\left(s_{t-1}=j, s_t=k|I_{t-1}, \theta\right) = p^{jk} \Pr\left(s_{t-1}=j|I_{t-1}, \theta\right) . \quad \text{(D.2)}$$

Step 3: Then, the density of y_t conditional on information at time $t-1$ and parameters is given by

$$f\left(y_t|I_{t-1}, \theta\right) = \sum_{j,k} f\left(y_t|s_{t-1}=j, s_t=k, I_{t-1}, \theta\right) \Pr\left(s_{t-1}=j, s_t=k|I_{t-1}, \theta\right), \text{(D.3)}$$

where the first and second terms are given by equations (D.1) and (D.2), respectively.

Step 4: The joint density of s_{t-1} and s_t conditional on information at time t and parameters is obtained by using the Bayes rule:

$$\mathsf{Pr}\left(s_{t-1}=j,s_t=k|I_t,\theta\right)=\frac{f\left(y_t,s_{t-1}=j,s_t=k|I_{t-1},\theta\right)}{f\left(y_t|I_{t-1},\theta\right)}$$
$$=\frac{f\left(y_t|s_{t-1}=j,s_t=k,I_{t-1},\theta\right)\mathsf{Pr}\left(s_{t-1}=j,s_t=k|I_{t-1},\theta\right)}{f\left(y_t|I_{t-1},\theta\right)}\ ,$$

where the first and second terms of the nominator are given by equations (D.1) and (D.2) and the denominator is given by equation (D.3).

Step 5: By integrating out regime s_{t-1} we obtain the probabilities of regime s_t conditional of information at time t and parameters:

$$\mathsf{Pr}\left(s_t=k|I_t,\theta\right)=\sum_j \mathsf{Pr}\left(s_{t-1}=j,s_t=k|I_t,\theta\right)\ .$$

Next, the regimes are drawn backward based on regime probabilities. In particular, regime s_T is sampled from $\mathsf{Pr}\left(s_T|I_T,\theta\right)$ and then for t from T-1 to 1 regimes are sampled from probabilities computed sequentially backward as

$$\mathsf{Pr}\left(s_t=j|I_t,s_{t+1}=k,\theta\right)=\frac{\mathsf{Pr}\left(s_{t+1}=k|s_t=j\right)\mathsf{Pr}\left(s_t=j|I_t,\theta\right)}{\sum_{j=1}^{n}\mathsf{Pr}\left(s_{t+1}=k|s_t=j\right)\mathsf{Pr}\left(s_t=j|I_t,\theta\right)}\ ,$$

where n is the total number of regimes.

References

Ang, A., Bekaert, G., and Wei, M. (2007a), "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 54, 1163–1212.

— (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*, forthcoming.

Ang, A., Dong, S., and Piazzesi, M. (2007b), "No-arbitrage Taylor rules," *NBER Working paper*, 13448.

Ang, A. and Piazzesi, M. (2003), "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50, 745–787.

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

Bikbov, R. and Chernov, M. (2008), "Monetary policy regimes and the term structure of interest rates," *CEPR Discussion Papers 7096*.

— (2010), "No-arbitrage macroeconomic determinants of the yield curve," *Journal of Econometrics*, forthcoming.

Cecchetti, S., Hooper, P., Kasman, B., Schoenholtz, K., and Watson, M. (2008), "Understanding the evolving inflation process," *U.S. Monetary Policy Forum*.

Chib, S. (2001), "Markov chain Monte Carlo methods: computation and inference," in *Handbook of Econometrics*, eds. Heckman, J. and Leamer, E., North Holland, Amsterdam, vol. 5, pp. 3569–3649.

Chib, S. and Ergashev, B. (2009), "Analysis of multi-factor affine yield curve models," *Journal of the American Statistical Association*, 104(488), 1324–1337.

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. (2009), "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*, 155(1), 19–38.

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 Sargent, T. (2005), "Drift and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302.

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.

Duffee, G. R. (2002), "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57(1), 405–443.

Fuhrer, J. C. (1996), "Monetary policy shifts and long-term interest rates," *Quarterly Journal of Economics*, 111, 1183–1209.

Goffe, W. (1996), "SIMANN: a global optimization algorithm using simulated annealing," *Studies in Nonlinear Dynamics and Econometrics*, 1(3), 169–176.

Gray, S. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27–62.

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. (1988), "Rational expectation econometric analysis of changes in regimes: an investigation of the term structure of interest rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

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

Hodrick, R. J. and Prescott, E. C. (1997), "Postwar U.S. business cycles: an empirical investigation," *Journal of Money, Credit and Banking*, 29, 1–16.

Kim, D. H. and Wright, J. (2005), "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates," *Finance and Economics Discussion Paper. Federal Reserve Board*, 33, 83–128.

Rudebusch, G. D. and Swanson, E. T. (2002), "Eurosystem monetary targeting: lessons from U.S. data," *European Economic Review*, 46, 417–442.

Rudebusch, G. D., Swanson, E. T., and Wu, T. (2006), "The bond yield conundrum from a macro-finance perspective," *Monetary and Economic Studies*, 55, 83–128.

Sims, C. and Zha, T. (2006), "Where there regime switches in U.S. monetary policy," *American Economic Review*, 96, 54–81.

Spiegelhalter, D. J., Best, N. J., Carlin, B. P., and van der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society. Series B.*, 64(4), 583–639.

Taylor, J. (1993), "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

Watson, M. W. (1999), "Explaining the increased variability in long-term interest rates," *Federal Reserve Bank of Richmond Economic Quarterly*, 85(4), 71–96.

Woodford, M. (1999), "Optimal monetary policy inertia," *NBER Working paper series.*