January 2010

Essays on Macro-Finance Relationships

Azamat Abdymomunov
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Economics

Dissertation Examination Committee:
James Morley, Chair
Steven Fazzari
Werner Ploberger
Raul Santaeulalia-Llopis
Andrew Sobel
Guofu Zhou

Essays on Macro-Finance Relationships

by

Azamat Abdymomunov

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

December 2010

Saint Louis, Missouri

Abstract of The Dissertation

Essays on Macro-Finance Relationships

by
Azamat Abdymomunov

Doctor of Philosophy in Economics

Washington University in St. Louis, 2010

Professor James Morley, Chair

In my dissertation, I study relationships between macroeconomics and financial markets. In particular, I empirically investigate the links between key macroeconomic indicators, such as output, inflation, and the business cycle, and the pricing of financial assets. The dissertation comprises three essays.

The first essay investigates how the entire term structure of interest rates is influenced by regime-shifts in monetary policy.[1] 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.

[1]This essay is a joint work with Kyu Ho Kang

The second essay examines the predictive power of the entire yield curve for aggregate output. Many studies find that yields for government bonds predict real economic activity. Most of these studies use the yield spread, defined as the difference between two yields of specific maturities, to predict output. In this paper, I propose a different approach that makes use of information contained in the entire term structure of U.S. Treasury yields to predict U.S. real GDP growth. My proposed dynamic yield curve model produces better out-of-sample forecasts of real GDP than those produced by the traditional yield spread model. The main source of this improvement is in the dynamic approach to constructing forecasts versus the direct forecasting approach used in the traditional yield spread model. Although the predictive power of yield curve for output is concentrated in the yield spread, there is also a gain from using information in the curvature factor for the real GDP growth prediction.

The third essay investigates time variation in CAPM betas for book-to-market and momentum portfolios across stock market volatility regimes[2]. For our analysis, we jointly model market and portfolio returns using a two-state Markov-switching process, with beta and the market risk premium allowed to vary between "low" and "high" volatility regimes. Our empirical findings suggest strong time variation in betas across volatility regimes in most of the cases for which the unconditional CAPM can be rejected. Although the regime-switching conditional CAPM can still be rejected in many cases, the time-varying betas help explain portfolio returns much better than the unconditional CAPM, especially when market volatility is high.

[2]This essay is a joint work with James Morley

Acknowledgements

I would like to express my deepest appreciation to my Advisor, Professor James Morley, for his professional guidance, persistent help, and encouragement. It has been my great privilege and pleasure to learn from him.

I would like to thank Professors Steven Fazzari and Werner Ploberger for their useful suggestions on parts of this dissertation. I would also like to thank Professors Raul Santaeulalia-Llopis, Andrew Sobel, and Guofu Zhou for serving in my defense committee.

I would like to express my appreciation to my wife, Nurzat, for her love, patience, and continuous support. My special thanks go to my daughter, Alima, for making my life joyful over the course of my studying. Finally, I would like to thank my parents, Ajar and Ryskeldi, for their love and support throughout my life. This dissertation is dedicated to them.

Contents

List of Tables

List of Figures

Chapter 1

The Effects of Monetary Policy Regime Shifts on the Term Structure of Interest Rates[1]

1.1 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)).

[1] This essay is a joint work with Kyu Ho Kang

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.

1.2 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.

1.2.1 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:

$$r_t^{m_t} = \overline{r}^{m_t} + \alpha^{m_t}\left(\pi_t - \overline{\pi}^{m_t}\right) + \beta^{m_t} g_t + u_t \ , \tag{1.2.1}$$

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

$$\Pi_m \equiv \begin{bmatrix} 1 - p_m^{12} & p_m^{12} \\ p_m^{21} & 1 - p_m^{21} \end{bmatrix}, \tag{1.2.2}$$

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 (1.2.1) 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.

1.2.2 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

$$\mathbf{f}_{t+1} - d^{m_{t+1}} = G\left(\mathbf{f}_t - d^{m_t}\right) + L^{v_{t+1}}\varepsilon_{t+1}\ , \qquad \varepsilon_{t+1} \sim \mathcal{N}_{3\times 1}(0, I)\ , \tag{1.2.3}$$

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

$$\Pi_v \equiv \begin{bmatrix} 1 - p_v^{12} & p_v^{12} \\ p_v^{21} & 1 - p_v^{21} \end{bmatrix}. \tag{1.2.4}$$

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.[2]

1.2.3 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:

$$\Lambda_t^{l_{t+1}} = \lambda_0^{l_{t+1}} + \lambda_f \mathbf{f}_t \ , \tag{1.2.5}$$

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

$$\Pi_l \equiv \begin{bmatrix} 1 - p_l^{12} & p_l^{12} \\ p_l^{21} & 1 - p_l^{21} \end{bmatrix} . \tag{1.2.6}$$

As we show in Section 1.4, 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]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 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.

1.2.4 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

(1.2.7)

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

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:

$$P^{s_t}_{\tau,t} = \mathbb{E}\left[\kappa_{t,t+1} P^{s_{t+1}}_{\tau-1,t+1} | \mathbf{f}_t, s_t\right], \tag{1.2.8}$$

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[3]:

$$\kappa_{t,t+1} = \exp\left(-r_t^{s_t} - \frac{1}{2}\Lambda_t^{s_{t+1}\prime}\Lambda_t^{s_{t+1}} - \Lambda_t^{s_{t+1}\prime}\varepsilon_{t+1}\right), \tag{1.2.9}$$

where $\Lambda_t^{s_{t+1}}$ is given by equation (1.2.5).

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

$$\log P_{\tau,t}^{s_t} = -A_\tau^{s_t} - B_\tau^{s_t\prime}\mathbf{f}_t \ , \tag{1.2.10}$$

where $A_\tau^{s_t}$ and $B_\tau^{s_t}$ 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 (1.2.1) is transformed to the form:

$$r_t^{s_t} = \delta_0^{s_t} + \delta_f^{s_t\prime}\mathbf{f}_t \ , \tag{1.2.11}$$

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

[3]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.

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 (1.2.8) 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 1.A. The solution has a form of recursive system:

$$\begin{aligned} A_\tau^j &= \delta_0^j + \sum_{k=1}^{S} p^{jk} \Big(A_{\tau-1}^k + \left(d^k - Gd^j - L^k\lambda_0^k\right)' B_{\tau-1}^k \\ &\quad - \frac{1}{2} B_{\tau-1}^{k\prime} L^k L^{k\prime} B_{\tau-1}^k \Big) \qquad (1.2.12) \\ B_\tau^j &= \delta_f^j + \sum_{k=1}^{S} p^{jk} \left(G - L^k\lambda_f\right)' B_{\tau-1}^k \qquad (1.2.13) \end{aligned}$$

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

$$R_{\tau,t}^{s_t} = -\frac{1}{\tau} \log\left(P_{\tau,t}^{s_t}\right) = a_\tau^{s_t} + b_\tau^{s_t\prime}\mathbf{f}_t \ , \qquad (1.2.14)$$

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 (1.2.3);

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 (1.2.12) and (1.2.13);

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 (1.2.10).

1.2.5 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:

$$R_{\tau,t}^{s_t} - r_t^{s_t} = \underbrace{\left[\frac{1}{\tau}\sum_{i=0}^{\tau-1}\mathbb{E}_t\left[r_{t+i}\right] - r_t^{s_t}\right]}_{\text{Expectation Hypothesis Component}} + \underbrace{\frac{1}{\tau}\sum_{i=1}^{\tau-1}\mathrm{ER}_{\tau+1-i,t}^{s_t}}_{\text{Term Premium}}, \tag{1.2.15}$$

where $\mathbb{E}_t$ denotes an expectation operator conditional on s_t and $\mathbf{f}_t$; $\mathrm{ER}_{\tau+1-i,t}^{s_t}$ 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

$$\mathrm{ER}^{j}_{\tau,t} = \mathbb{E}[\overline{p}_{\tau-1,t+1}|s_t = j, \mathbf{f}_t] + \overline{p}^{j}_{1,t} - \overline{p}^{j}_{\tau,t}\ , \tag{1.2.16}$$

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

$$\mathrm{ER}^{j}_{\tau,t} = -\sum_{k=1}^{S} p^{jk} \left(B^{k'}_{\tau-1} L^{k} \Lambda^{k}_{t} + \frac{1}{2} B^{k'}_{\tau-1} L^{k} L^{k\prime} B^{k}_{\tau-1} \right). \tag{1.2.17}$$

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

1.3 Estimation

1.3.1 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

$$g_t = \frac{1}{4} \frac{RGDP_t - RGDP_t^p}{RGDP_t^p} \;, \tag{1.3.1}$$

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.[4] The gap is factored by 1/4 to make estimated coefficients interpretable as coefficients for annualized interest rates.

1.3.2 Identification restrictions

The factor dynamics and Taylor rule equation (1.2.1) 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,

[4]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.

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 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.

1.3.3 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.

1.3.4 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 (1.2.3). 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:

$$R_{12,t} = a_{12}^{s_t} + b_{12}^{s_t\prime}\mathbf{f}_t = a_{12}^{s_t} + b_{u,12}^{s_t} u_t + b_{\overline{m},12}^{s_t\prime}\overline{m}_t \ , \tag{1.3.2}$$

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:

$$u_t = \left(b_{u,12}^{s_t}\right)^{-1} \left(R_{12t} - a_{12}^{s_t} - b_{\overline{m},12}^{s_t\prime}\overline{m}_t\right) . \tag{1.3.3}$$

Thus,

$$\mathbf{f}_t = \begin{pmatrix} u_t \\ \overline{m}_t \end{pmatrix} = \begin{pmatrix} \left(b_{u,12}^{s_t}\right)^{-1} \left(R_{12t} - a_{12}^{s_t} - b_{\overline{m},12}^{s_t\prime}\overline{m}_t\right) \\ \overline{m}_t \end{pmatrix} . \tag{1.3.4}$$

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

$$y_t = \underbrace{\begin{pmatrix} \overline{a}^{s_t} \\ \mathbf{0} \end{pmatrix}}_{\overline{A}^{s_t}} + \underbrace{\begin{pmatrix} \overline{b}^{s_t} \\ I_3 \end{pmatrix}}_{\overline{B}^{s_t}} \mathbf{f}_t + \begin{pmatrix} I_7 \\ 0_{3\times 7} \end{pmatrix} \widetilde{\varepsilon}_t \ , \qquad \widetilde{\varepsilon}_t \sim iidN\left(0, \Sigma\right) \ , \tag{1.3.5}$$

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 $b_{u,12}^{s_t}$.

1.3.5 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 1.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 1.1 displays median, 2.5%, and

97.5% quantile surfaces of simulated yield curves and their time series averages. As Figure 1.1 illustrates, this simulation exercise produces, on average, a slightly upward-sloping yield curves with substantial variation.

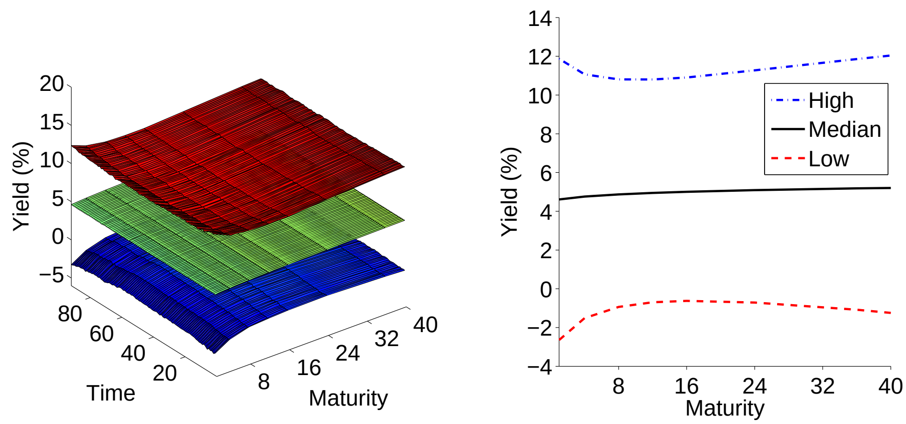


Figure 1.1: **The prior-implied yield curves**

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.

1.3.6 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

$$\pi\left(\theta, \mathbf{S}_T|\mathbf{y}\right) \propto f\left(\mathbf{y}|\theta, \mathbf{S}_T\right) f\left(\mathbf{S}_T|\theta\right) \pi\left(\theta\right) \quad , \tag{1.3.6}$$

where $f\left(\mathbf{y}|\theta, \mathbf{S}_T\right)$ 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 $\mathbf{S}_T = \{s_t\}_{t=0,1,..,T}$; $f\left(\mathbf{S}_T|\theta\right)$ is the density function for regime-indicators given the parameters; $\pi\left(\theta\right)$ is the prior density of the parameters.

The MCMC procedure is discussed in detail in Appendix 1.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 (1.3.3), 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.

1.4 Empirical Results

1.4.1 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).[5] Table 1.1 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.

Table 1.1: **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.

1.4.2 Parameter Estimates and Regimes

Table 1.2 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 MH step of the sampler and the inefficiency factor as dis-

[5] 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.

cussed in Chib (2001).[6] 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

$$r_t^{m_t} = (1-\rho)\left[\widetilde{r}^{m_t} + \widetilde{\alpha}^{m_t}\left(\pi_t - \overline{\pi}^{m_t}\right) + \widetilde{\beta}^{m_t} g_t\right] + \rho r_{t-1}^{m_{t-1}} + \xi_t \ , \tag{1.4.1}$$

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 easy to show that $\rho = G_{1,1}$.[7] After this transformation the coefficients $\widetilde{\alpha}^1$ =3.30 and

[6] 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.

[7] 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.

Table 1.2: **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.

$\widetilde{\alpha}^2$ =16.33 both have values grater than unity, and therefore they do not potentially create a risk of indeterminacy of the equilibrium.[8] Given this result, the regime with

[8]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.

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.[9]

Figure 1.2 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 1.2. 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 out-

[9] 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.

put 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 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 1.3, 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.[10]

The volatility estimates of exogenous shocks to all factors, reported in Table 1.2 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

[10]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".

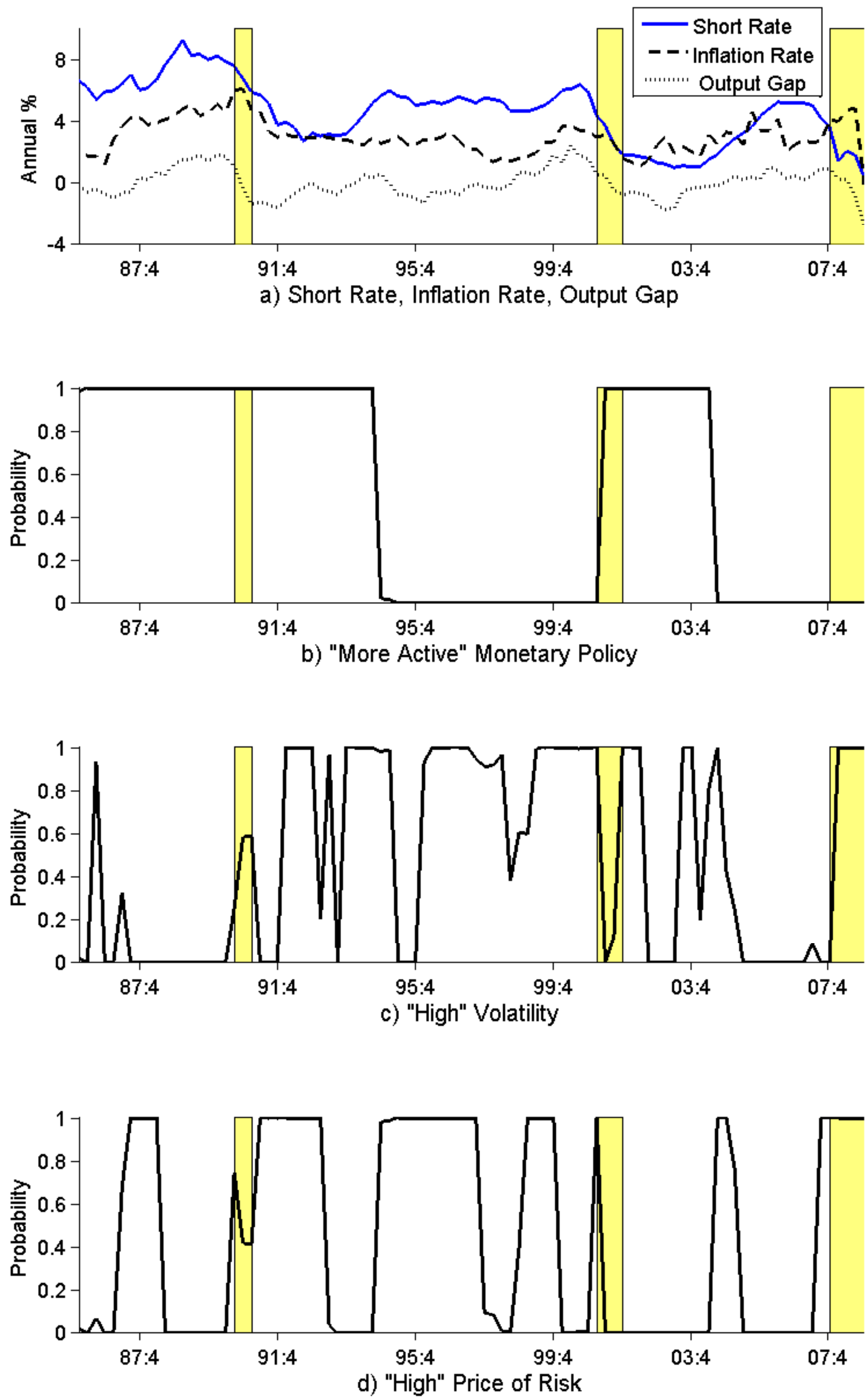


Figure 1.2: **The Probabilities of monetary policy, volatility, and risk regimes**

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.

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 1.2 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 1.3 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.

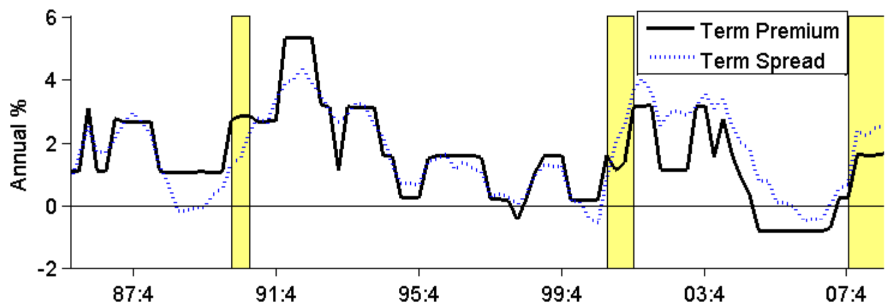


Figure 1.3: **The Term Premium**
The figure displays the model-implied term premium and the term spread for 10-year bonds. Shaded areas corresponds to NBER recession dates.

1.4.3 Monetary Policy Regimes and the Yield Curve

Figure 1.4 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

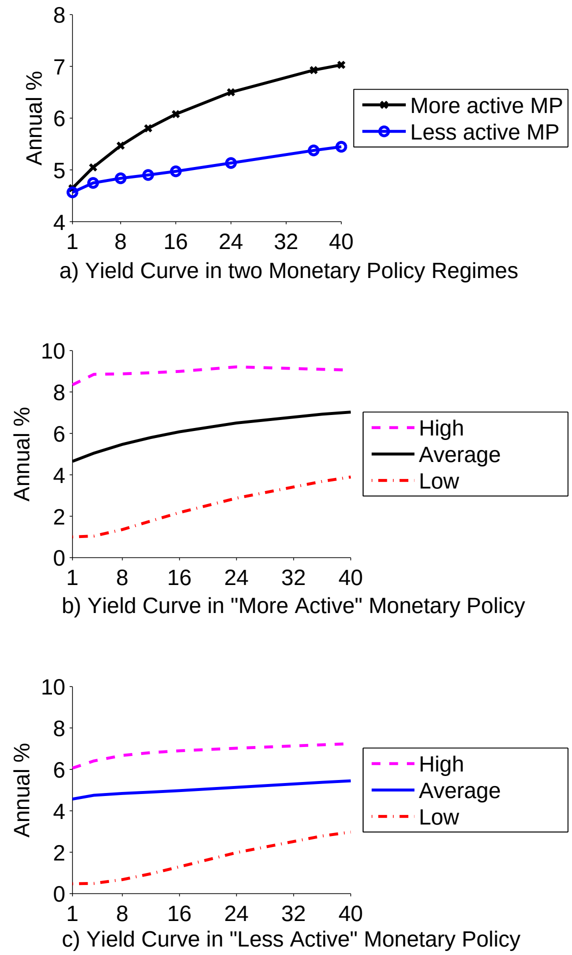


Figure 1.4: **Average realized yield curves**

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.

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 1.4 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 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 rates if a single regime were maintained throughout the sample, we conduct a counterfactual analysis. Figure 1.5 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

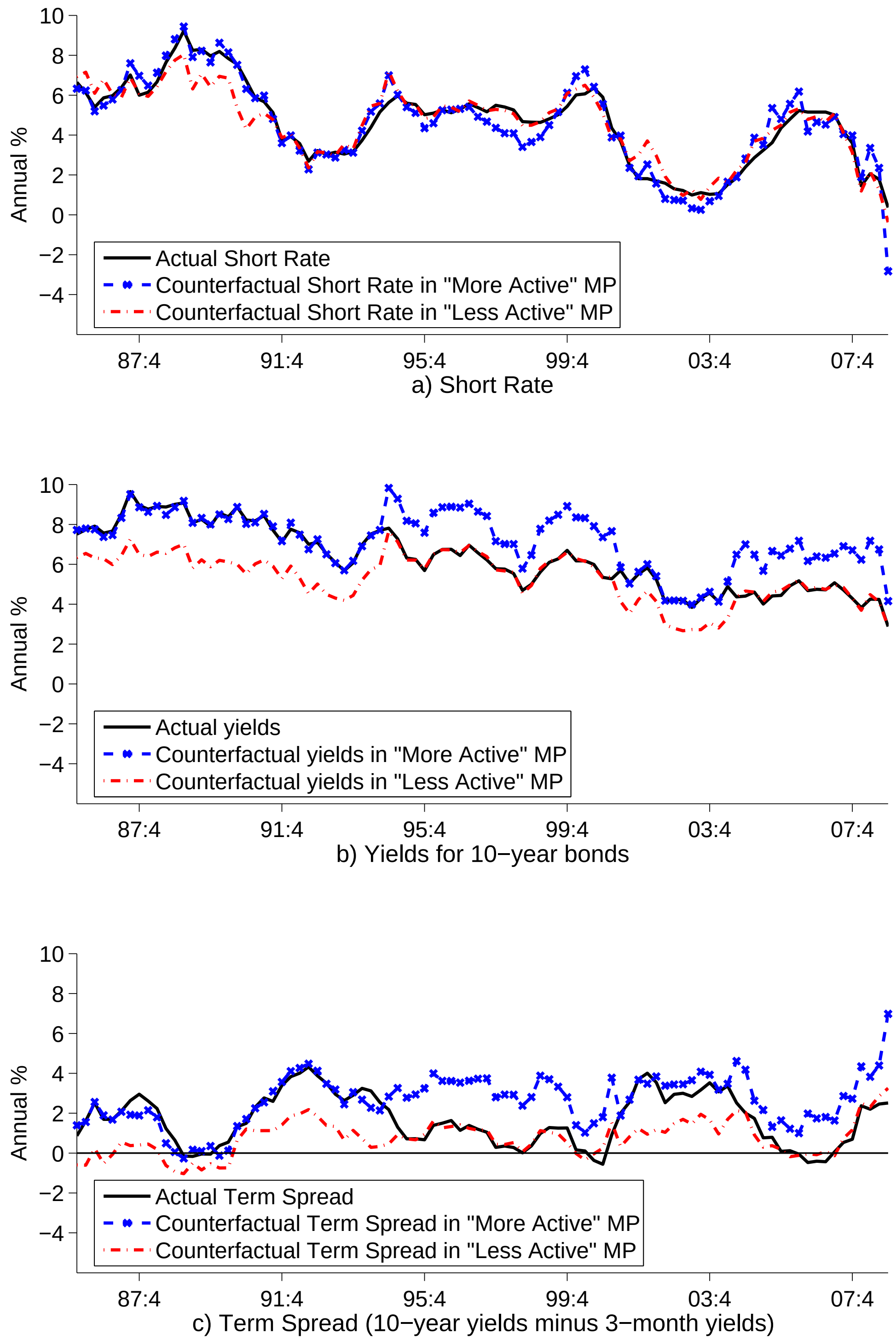


Figure 1.5: **Counterfactual short rates, long rates, and term spreads**
The time series of counterfactual interest rates are simulated by fixing parameters to one of the two monetary policy regimes.

consequently the term spread would have been higher than the actual ones in those periods when the regime was “less active”.

1.5 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

1.A Bond Pricing

We solve for A^j_τ and B^j_τ using the law of iterated expectatios, method of undetermined coefficients, and log-linearization:

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

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

$$\begin{aligned}
&= \sum_{k=1}^{S} p^{jk}\mathbb{E}\left[\exp\begin{pmatrix} -r^j_t - \frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t - \Lambda^{k\prime}_t\varepsilon_{t+1} \\ +A^j_\tau + B^{j\prime}_\tau\mathbf{f}_t - A^k_{\tau-1} - B^{k\prime}_{\tau-1}\mathbf{f}_{t+1}\end{pmatrix}|\mathbf{f}_t, s_t = j, s_{t+1} = k\right] \\
&= \sum_{k=1}^{S} p^{jk}\left\{\begin{array}{c} \exp\left(-r^j_t - \frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t + A^j_\tau - A^k_{\tau-1} + B^{j\prime}_\tau\mathbf{f}_t - B^{k\prime}_{\tau-1}\mu^{j,k}_t\right) \\ \times\mathbb{E}\left[\exp\left(-\left(\Lambda^{k\prime}_t + B^{k\prime}_{\tau-1}L^k\right)\varepsilon_{t+1}\right)|\mathbf{f}_t, s_t = j, s_{t+1} = k\right]\end{array}\right\} \quad (1.A.1) \\
&= \sum_{k=1}^{S} p^{jk}\left\{\exp\begin{pmatrix} -r^j_t - \frac{1}{2}\Lambda^{k\prime}_t\Lambda^k_t + A^j_\tau - A^k_{\tau-1} + B^{j\prime}_\tau\mathbf{f}_t - B^{k\prime}_{\tau-1}\mu^{j,k}_t \\ +\frac{1}{2}\left(\Lambda^{k\prime}_t + B^{k\prime}_{\tau-1}L^k\right)\left(\Lambda^{k\prime}_t + B^{k\prime}_{\tau-1}L^k\right)'\end{pmatrix}\right\} \quad (1.A.2) \\
&= \sum_{k=1}^{S} p^{jk}\exp\begin{pmatrix} -r^j_t + A^j_\tau - A^k_{\tau-1} + B^{j\prime}_\tau\mathbf{f}_t \\ -B^{k\prime}_{\tau-1}\mu^{j,k}_t + B^{k\prime}_{\tau-1}L^k\Lambda^k_t + \frac{1}{2}B^{k\prime}_{\tau-1}L^kL^{k\prime}B^k_{\tau-1}\end{pmatrix} \quad (1.A.3)
\end{aligned}$$

$$\approx \sum_{k=1}^{S} p^{jk} \left\{ \begin{array}{c} -\delta_0^j - \delta_f^{j\prime}\mathbf{f}_t + A_\tau^j - A_{\tau-1}^k + B_\tau^{j\prime}\mathbf{f}_t \\ -B_{\tau-1}^{k\prime} d^k - B_{\tau-1}^{k\prime} G\left(\mathbf{f}_t - d^j\right) \\ +B_{\tau-1}^{k\prime} L^k \left(\lambda_0^k + \lambda_f \mathbf{f}_t\right) + \frac{1}{2} B_{\tau-1}^{k\prime} L^k L^{k\prime} B_{\tau-1}^k + 1 \end{array} \right\}. \tag{1.A.4}$$

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

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

evaluated at $x = -\left(\Lambda_t^{k\prime} + B_{\tau-1}^{k\prime} L^k\right)'$. Following Bansal and Zhou (2002), $(1.A.3)$ is transformed into $(1.A.4)$ using log-approximation $\exp(y) \approx y + 1$ for a sufficiently small y and substituting for r_t^j using equation $(1.2.11)$.

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

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

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:

$$
0 \;=\; \sum_{k=1}^{S} p^{jk} \begin{pmatrix} -\delta_0^j + A_\tau^j - A_{\tau-1}^k - B_{\tau-1}^{k\prime} d^k + B_{\tau-1}^{k\prime} G d^j \\ +B_{\tau-1}^{k\prime} L^k \lambda_0^k + \frac{1}{2} B_{\tau-1}^{k\prime} L^k L^{k\prime} B_{\tau-1}^k \end{pmatrix}
$$

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:

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

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 $B_1^j = \delta_f^j$.

1.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:

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

$$
\begin{aligned}
&= \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right) - r_t^j \\
&\quad - \left\{ -r_t^j + \mathsf{log}\left(\sum_{k=1}^{S} p^{jk}\,\mathsf{exp}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} \end{array}\right)\right)\right\} \\
&= \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right) \\
&\quad - \mathsf{log}\left(\sum_{k=1}^{S} p^{jk}\,\mathsf{exp}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} \end{array}\right)\right) \\
&\approx \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right) \\
&\quad - \mathsf{log}\sum_{k=1}^{S} p^{jk}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} + 1 \end{array}\right) \\
&\approx \sum_{k=1}^{S} p_t^{jk}\left(-A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k}\right) \\
&\quad - \sum_{k=1}^{S} p^{jk}\left(\begin{array}{c} -A_{\tau-1}^{k} - B_{\tau-1}^{k\prime}\mu_t^{j,k} + B_{\tau-1}^{k\prime}L^k\Lambda_t^k \\ +\frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k} \end{array}\right) \\
&= -\sum_{k=1}^{S} p^{jk}\left(B_{\tau-1}^{k\prime}L^k\Lambda_t^k + \frac{1}{2}B_{\tau-1}^{k\prime}L^kL^{k\prime}B_{\tau-1}^{k}\right) .
\end{aligned}
$$

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

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

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

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 1.3 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

$$d_\Omega = \begin{pmatrix} 5 \times 10^5 & 5 \times 10^5 & 7 \times 10^4 \end{pmatrix} \tag{1.C.2}$$

and

$$d_\Sigma = \begin{pmatrix} 7 \times 10^5 & 4 \times 10^6 & 3 \times 10^7 & 6 \times 10^7 & 10^7 & 10^7 & 10^7 \end{pmatrix} . \tag{1.C.3}$$

Table 1.3: **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 (1.C.2) and (1.C.3).

1.D MCMC Sampling

This Section provides details of the MCMC algorithm summarized in Section 1.3.6 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

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

where

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

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

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

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

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

Thus, the conditional density of y_t becomes

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

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

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

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

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

where the first and second terms are given by equations (1.D.1) and (1.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:

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

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

Chapter 2

Predicting Output Using the Entire Yield Curve

2.1 Introduction

There are numerous papers which explore the question: "What information does the yield spread contain about future real economic activity?" These studies are based on the intuition that, when agents price assets, they take into account expectations about future states of the economy, and therefore interest rates potentially contain useful information about future economic growth. Estrella and Hardouvelis (1991) find evidence that the U.S. government bond yield spread contains information about future U.S. real economic activity at horizons of up to four years. Estrella and Mishkin (1997) confirm that the yield spread has the predictive power for real economic activity in the United States and in a number of European countries. Wheelock and Wohar (2009) provide a comprehensive survey of the literature on the predictive power of the term spread for output growth.

In most of the previous literature on the predictive power of yield curve for real economic activity, researchers have considered simple OLS regressions of future output on a yield spread defined as the difference between a specific long-term government bond rate and a short-term T-bill rate. Although this approach has the advantage of its simplicity, it does not have enough flexibility to use the information contained in the entire term structure of interest rates.

In this paper, I propose an approach to predicting output based on information contained in the entire yield curve. In particular, I examine the predictive power of the yield curve for real output by jointly modeling real GDP growth and yield curve using the dynamic yield curve model proposed by Diebold and Li (2006) (hereafter DL(2006)). This model, which I refer to as the "NS dynamic yield curve model" for the purpose of this study, is based on the Nelson and Siegel (1987) three-latent-factor framework. The choice of the NS dynamic model for this study is driven by its relative parsimony compared to other yield curve models and its good out-of-sample forecasting performance for future yields. The model describes the entire term structure of interest rates using only three factors. DL(2006) introduce dynamics to the evolution of these factors and show that the NS dynamic model has more accurate in-sample fit and produces better forecasts of future yields at long horizons relative to other simple models. In terms of predicting output, the NS dynamic model has two advantages over the yield spread framework: (i) the model contains information about the entire term structure of interest rates and (ii) real GDP growth can be modeled jointly with yields in a parsimonious way using the endogenously-defined three factors. Another potential choice of term structure modeling would be the affine arbitrage-free class of models, which is popular in finance literature. However, as reported by Duffee (2002), arbitrage-free models produce poor out-of-sample forecasts of future yields.

Ang, Piazzesi, and Wei (2006) (hereafter APW(2006)) study the predictive power of the short-term yield and yield spread for real GDP growth using an affine arbitrage-free dynamic yield curve model. Their approach is based on modeling real GDP growth jointly with an exogenously-defined short-term yield and yield spread and imposing no-arbitrage constraints on the pricing of bonds. They find, in contrast to the previous findings in the literature on the predictive power of the yield curve for output, that the short-term interest rate has more predictive power for the GDP growth than the yield spread. The authors also report that imposing no-arbitrage restriction only marginally improves forecasts of GDP. Huang, Lee, and Li (2006) also analyze the gains from using information in the entire yield curve for output and inflation in their forecast combination study. They find that combining forecasts, where each individual forecast uses information in the yield curve, can improve forecasts of output growth and inflation. Chauvet and Senyuz (2009) construct a common factor from information in the yield curve to improve forecasts of output and recessions.

The focus of my analysis is to find out whether forecasting output using the entire yield curve is better than using a yield spread forecasting model. For this analysis, I perform pseudo out-of-sample forecast comparisons for real GDP growth based on root mean square errors (RMSEs) for the NS dynamic yield curve model and the yield spread model based on OLS regressions of the GDP growth on a yield spread. I consider various versions of the dynamic yield curve model in which real GDP growth is explained by different yield factors in order to analyze marginal impact of each of the factors on the forecasting performance.

I find that the dynamic yield curve model significantly improves out-of-sample forecasts of real GDP growth at all horizons relative to the yield spread model. The main source of this improvement can be attributed to the dynamic way yield factors and real GDP growth are modeled. Although the predictive power of the yield curve for output is concentrated in the yield spread, there is also a gain from extracting more information from the entire yield curve relative to a specific exogenously-defined yield spread. In particular, there is a gain from using information in the curvature factor for the long horizon prediction.

The rest of this paper is organized as follows. Section 2 describes the data. Section 3 motivates and presents the traditional yield spread model and reports the predictive power of this model for output. Section 4 describes the dynamic yield curve model. Section 5 reports estimation results for the dynamic yield curve model. Section 6 reports out-of-sample forecasting results and compares various versions of the dynamic yield curve and yield spread models. Section 7 concludes.

2.2 Data

The raw interest rate data are monthly-average yields on U.S. government bonds for maturities 3, 6, 12, 24, 36, 60, 84, 120 months obtained from the FRED database.[1] The yields are constant maturity rates, except for the 3 and 6 month maturities that are secondary market rates.[2] Yield data for the maturities 3, 12, 36, 60, 120 months

[1]Gurkaynak et al. (2007) is another source of publicly available data on the term structure of interest rates, which has yields for long-term bonds. These data are constructed using the Svensson (1994) model, which is an extension of the Nelson and Siegel (1987) model. Since the model used for my study is also based on the Nelson and Siegel (1987) model, I opt not to use these data in order to avoid fitting the data with the approach used to generate data in the first place.

[2]I use secondary market rate data for the 3 and 6 month maturities because the constant maturity rate data for these maturities are available for a substantially shorter sample period than the sample period that I consider for this study. I compared the secondary market 3 and 6 month maturity yield series with the constant maturity rate series for the common sample period and found that the dynamics of the series are close to each other. Therefore, this heterogeneity in data should not significantly affect my results.

cover the period of 1953:04 to 2007:12, for 6 months from 1959:01 to 2007:12, for 24 months from 1976:07 to 2007:12, for 84 months from 1969:07 to 2007:12.[3] Monthly data on yields are transformed to quarterly frequency by using observations from the last month of each quarter. Quarterly data on real GDP from 1952:Q1 to 2007:Q4 are also from the FRED database. Real GDP data are seasonally adjusted and chained in 2000 prices. Annualized real GDP growth is calculated as the difference of natural log output multiplied by 400. Table 2.1 reports descriptive statistics for the yields and real GDP growth.

Table 2.1: **Descriptive statistics for yields and RGDP growth**

Maturities (months)	Period	Mean	Std. Dev.	Min.	Max.	ADF
3	1953-M04 : 2007-M12	5.11	2.79	0.64	16.30	-2.69
6	1959-M01 : 2007-M12	5.63	2.68	0.92	15.52	-2.12
12	1953-M04 : 2007-M12	5.67	2.96	0.82	16.72	-2.04
24[1]	1976-M07 : 2007-M12	6.89	3.15	1.23	16.46	-1.21
36	1953-M04 : 2007-M12	6.07	2.82	1.47	16.22	-1.92
60	1953-M04 : 2007-M12	6.26	2.75	1.85	15.93	-1.77
84[1]	1959-M07 : 2007-M12	7.41	2.56	2.84	15.65	-1.25
120	1953-M04 : 2007-M12	6.46	2.68	2.29	15.32	-1.60
RGDP growth	1953-Q2 : 2007-Q4	3.14	3.66	-11.02	15.46	-10.51*

RGDP growth is calculated as the difference of natural log output multiplied by 400. The Augmented Dickey-Fuller (ADF) unit root test is based on SIC lag selection. The critical values for rejection of hypothesis of a unit root are: -3.44 at 1 percent level and -2.87 at 5 percent level. The hypothesis that yields have unit roots cannot be rejected at 5 percent level. The hypothesis that real GDP growth has a unit root is rejected at the 1 percent level, denoted by an asterisk.

/1 Average yields of 24 and 84 month bonds are higher than those of 36 and 120 month, respectively, because of the difference in sample periods.

[3] The data on yields have different staring dates; however, I do not extrapolate yields with shorter sample periods to the same beginning date, as the focus of this study is predictive power of yields on output using available information.

2.3 Motivation

The standard explanations for why a yield spread might predict economic growth are focused on monetary policy and the expectation hypothesis. Under the expectation hypothesis, the term structure of interest rates is determined by agents' expectations of future short-term interest rates. Therefore, current long-term interest rates are averages of expected future short-term rates. If a monetary contraction sends the current short-term rate higher than the expected future short-term interest rate, then today's investment and consumption will decline causing a decline in future economic growth. Conversely, if a monetary expansion produces low current short-term interest rates leading to higher economic growth in future, then future short-term interest rates are expected to increase.

Harvey (1988) proposes another explanation for why the slope of yield curve and future economic activities can be related, which is based on the theory of smoothing intertemporal consumption and the real term structure of interest rates. In this setting, if agents expect that future economic activity will decline, then they have incentive to save in the current period by selling short-term assets and buying bonds which will pay off in the low-income period. This will lower the yields for the bonds that will mature in the future and increase the short rate. Thus, in theory the yield curve contains information about future economic growth.

The term premium for holding long-term bonds is also a component that contributes to the determination of the term structure of interest rates in addition to the expectation factor. APW(2006) suggest that the expectation hypothesis component of the term structure of interest rates is the main driving force for output predictability. Hamilton and Kim (2002) suggest that the term premium, in addition to the expectation component, is also important for output prediction.

Most previous studies of predictive power of the yield curve for real economic activity have employed OLS regressions of future real GDP growth rate on the yield spread, defined as the difference between interest rates on the long-term (10 year) treasury bond and the short-term (3 month) treasury bill:

$$g_{t,t+k} = \alpha_{0,k} + \alpha_{1,k}\left(y_t\left(120\right) - y_t\left(3\right)\right) + \varepsilon_t \qquad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \tag{2.3.1}$$

where $g_{t,t+k}$ is the annualized real GDP (RGDP) growth rate defined as

$$g_{t,t+k} = 400/k\left(\ln RGDP_{t+k} - \ln RGDP_t\right) \tag{2.3.2}$$

where $y_t\left(120\right)$ and $y_t\left(3\right)$ are interest rates on the 10-year treasury bond and the 3-month treasury bill, respectively.

Figure 2.1 plots the yield spread defined as above, along with the annualized real GDP growth rate over subsequent four quarters. It is evident that real GDP growth and the yield spread are positively correlated. The correlation coefficient is 0.41.

Table 2.2 reports the estimation results for the OLS regressions of future real GDP growth on the yield spread according to equation (2.3.1), the spread and one lag of real GDP growth, the short rate only (defined as the 3-month T-bill interest rate), and the spread and the short rate for the period from 1953:Q2 to 2007:Q4. The estimates for the yield spread coefficient from the yield spread regression are

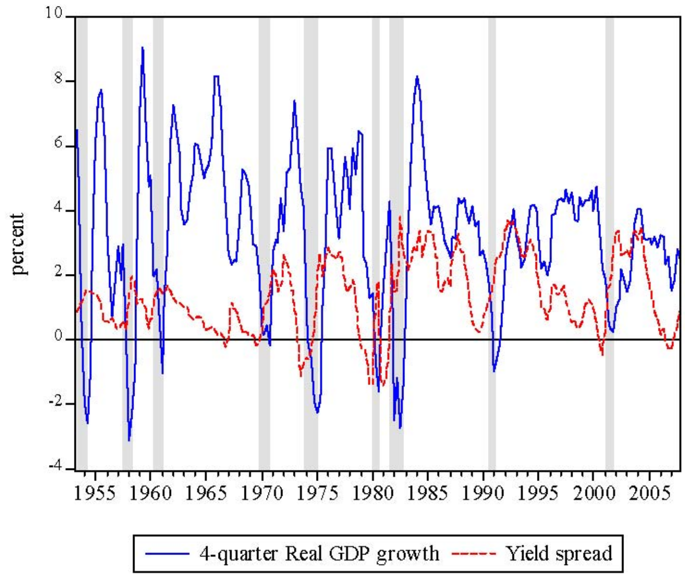


Figure 2.1: **Real GDP growth and Yield Spread**
This figure displays the subsequent four-quarter real GDP growth rate and the yield spread, defined as the difference between interest rates on the 10-year Treasury bond and the 3-month Treasury bill. Shaded areas correspond to NBER recession dates.

statistically significant for all horizons up to 12 quarters ahead and the adjusted-R^2s are considerably higher for 4 and 8 quarter horizons than for 1 and 12 quarter horizons. The estimates for the yield spread coefficient remain robust to controlling for one lag of real GDP growth, with an increase in the adjusted-R^2 only at the one quarter horizon. This increase can be explained by short-term persistence of real GDP growth. I also consider the explanatory power of the short-term interest rate for future real GDP growth. Although the short-term interest rate is statistically significant in the regression with the short-term rate only, the adjusted-R^2 of this regression is lower than for the regression model with the yield spread only. The yield spread remains strongly statistically significant after controlling for the short-term rate up to 8 quarters ahead, while short-term rate remains significant only

Table 2.2: **Parameter estimates for OLS regressions of k-quarter-ahead annualized RGDP growth on the yield spread**

k	Spread $\alpha_{1,k}$	$\overline{R}^2$	Spread $\alpha_{1,k}$	g_{t-1} $\alpha_{2,k}$	$\overline{R}^2$	$y_t(3)$ $\alpha_{3,k}$	$\overline{R}^2$	Spread $\alpha_{1,k}$	$y_t(3)$ $\alpha_{3,k}$	$\overline{R}^2$
1	0.600		0.606	0.171		-0.263		0.640	-0.184	
	(0.239)	0.03	(0.227)	(0.067)	0.06	(0.123)	0.04	(0.276)	(0.133)	0.07
4	0.757		0.756	-0.007		-0.276		0.704	-0.189	
	(0.188)	0.15	(0.189)	(0.051)	0.14	(0.097)	0.10	(0.230)	(0.094)	0.21
8	0.556		0.554	-0.029		-0.173		0.474	-0.113	
	(0.136)	0.16	(0.137)	(0.039)	0.16	(0.081)	0.09	(0.170)	(0.077)	0.18
12	0.313		0.310	-0.032		-0.086		0.255	-0.054	
	(0.119)	0.08	(0.121)	(0.030)	0.09	(0.071)	0.03	(0.138)	(0.070)	0.07

Sample period: 1953:Q2-2007:Q4. Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors are in parentheses. The spread is defined as the difference between yields on the 10-year bond and the 3-month Treasury bill. The short rate is defined as the yield on the 3-month Treasury bill, denoted as $y_t(3)$; $\alpha_{1,k}$, $\alpha_{2,k}$, and $\alpha_{3,k}$ denote the coefficients from respective OLS regressions; g_{t-1} is one lag of the annualized continuously-compounded real GDP growth rate; $\overline{R}^2$ denotes adjusted-R^2.

at 4 quarter ahead. These results, which are in line with previous findings on the predictive power of yield spread for output, confirm that the yield spread may be used to predict real output.

2.4 Model

2.4.1 The Dynamic Yield Curve Model

I consider the three-latent-factor dynamic yield curve model developed by DL(2006) based on the Nelson and Siegel (1987) framework. In this NS dynamic yield curve model, yields are represented by the following functional form:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \underbrace{\left(\frac{1-\exp(-\lambda_t \tau)}{\lambda_t \tau}\right)}_{L_2(\tau,\lambda)} + \beta_{3,t} \underbrace{\left(\frac{1-\exp(-\lambda_t \tau)}{\lambda_t \tau} - \exp(-\lambda_t \tau)\right)}_{L_3(\tau,\lambda)} \quad (2.4.1)$$

where $y_t(\tau)$ is an interest rate of zero-coupon bond with maturity τ at period t; $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ are three latent dynamic factors interpreted as level, slope, and curvature of the yield curve; and λ_t is a parameter responsible for fitting yield curve at different maturities. Small values of λ_t fit the yield curve better at long maturities, while large values produce a better fit at short maturities. In this paper, I follow DL(2006) and, for simplicity, estimate λ_t as a time-invariant parameter. Therefore, its time subscript is dropped in further discussions. $L_2(\tau,\lambda)$ and $L_3(\tau,\lambda)$ denote the loadings for factors $\beta_{2,t}$ and $\beta_{3,t}$, respectively. The loading for factor $\beta_{1,t}$ is 1.

The choice of the NS dynamic model is motivated by its parsimony and good out-of-sample forecasting performance for the future yields. The alternative yield curve model to consider for this study would be the affine arbitrage-free class of yield curve models. However, as reported by Duffee (2002), arbitrage-free yield curve models perform poorly out-of-sample. Also, APW(2006), who study predictive power of the yield curve for output, find that imposing no-arbitrage restriction improves GDP forecasting only marginally over a VAR model. As will be shown in the empirical section, the NS dynamic model does better relative to a VAR model.

In the NS framework, the entire panel of yields is modeled by three latent factors with imposed structure of loadings as follows:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_n) \end{pmatrix} = \begin{pmatrix} 1 & L_2(\tau_1,\lambda) & L_3(\tau_1,\lambda) \\ 1 & L_2(\tau_2,\lambda) & L_3(\tau_2,\lambda) \\ \vdots & \vdots & \vdots \\ 1 & L_2(\tau_n,\lambda) & L_3(\tau_n,\lambda) \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_n) \end{pmatrix} \tag{2.4.2}$$

$$\varepsilon_t \backsim N_n(0, \Sigma)$$

Similarly to DL(2006) and Diebold, Rudebusch, and Aruoba (2006) (hereafter DRA(2006)), the measurement errors of yields of different maturities are assumed to be independent from each other. Therefore, the variance-covariance matrix of measurement errors in this equation, denoted as Σ, is a diagonal.

The latent factors are modeled as Gaussian first-order autoregressive processes:

$$\beta_{i,t} = \mu_i + \phi_i \beta_{1,t-1} + u_t \qquad u_t \backsim N\left(0, \sigma_i^2\right) \quad \text{for } i \in \{1,2,3\} \tag{2.4.3}$$

where σ_i^2 denotes the variance of error-term for the factor process $\beta_{i,t}$.

In their study of the relationship between macro variables and the yield curve, DRA(2006) assume that the factors are governed by a VAR(1) process, allowing for interaction between all three factors and macro variables, and between their shocks. However, DL(2006) report that a model with a VAR(1) factor process forecasts yields poorly compared to a simple AR(1). My result suggests that a model based on independent factor processes also forecasts output better than a model with a VAR(1) factor process.

DL(2006) show that this general model can generate all possible yield curve shapes, has good in-sample fit, and forecasts future yields better out of sample than other models at 6 months or longer horizons. They also show that the $\beta_{1,t}$ factor is highly correlated with yields of different maturities. Therefore, it is interpreted as level factor; $-\beta_{2,t}$ is highly correlated with the yield spread; and $\beta_{3,t}$ is correlated with the curvature. In this model, all three latent factors are assumed to be stationary. As will be shown next, this model is also flexible in terms of incorporating macro variables.

2.4.2 The Dynamic Yield Curve Model with Real GDP growth

This subsection describes how to incorporate real GDP growth into the NS dynamic yield curve model. Since output growth is correlated with yields and yields are described by three factors, output growth should be correlated with the yield factors of the model.[4] Therefore, I modify the yield curve model to jointly model yields with real GDP growth rate using the three yield factors. Previous analysis suggested that adding lagged real GDP growth improves forecasts of output, and therefore the modified model also controls for one lag of the real GDP growth rate. After

[4] DRA(2006) find evidence of interactions between the yield curve and macro variables based on analysis of impulse response functions and variance decompositions. They do not study forecasting performance of the macro-yield-curve model. They model macro variables as additional factors in the state dynamics of the yield curve model.

this modification, equation (2.4.2) has the following form:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_n) \\ g_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mu_g \end{pmatrix} + \begin{pmatrix} 1 & L_2(\tau_1,\lambda) & L_3(\tau_1,\lambda) \\ 1 & L_2(\tau_2,\lambda) & L_3(\tau_2,\lambda) \\ \vdots & \vdots & \vdots \\ 1 & L_2(\tau_n,\lambda) & L_3(\tau_n,\lambda) \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix}$$

$$+ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \gamma_4 g_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_n) \\ \varepsilon_t(g) \end{pmatrix} \tag{2.4.4}$$

$$\varepsilon_t \backsim N_{n+1}\left(0, \widetilde{\Sigma}\right)$$

where g_t denotes real GDP growth defined as

$$g_t = 400\left(\ln RGDP_t - \ln RGDP_{t-1}\right)$$

In this specification, output growth only enters into equation (2.4.4) while the factor dynamics equations remain the same as before. Thus, in this setting, real GDP growth is modeled only by the latent factors, which are mainly identified by the term structure of interest rates due to the rich panel of yields. This approach focuses on the one-way interaction from yields to macro variables. An alternative way of incorporating output growth into the yield curve model would be to follow DRA(2006) and add output growth to the factor process as an additional factor.

This specification would allow for two-way interaction between output growth and other yield factors. However, preliminary results suggested that the forecasts produced by such a model were inferior to those produced by the model in equation (2.4.4).

2.5 In-sample Results

Estimation of the dynamic yield curve model is based on quarterly yield data for the sample period from 1953:Q2 to 2007:Q4. I estimate the model using a one-step Kalman filter maximum-likelihood procedure, which produces more efficient inferences than those from the two-step estimation procedure applied by DL(2006) and APW(2006).

The estimates of the factor process parameters, reported in Table 2.3, suggest that $\beta_{1,t}$ is a very persistent series with an autoregressive coefficient of 0.98 and a standard deviation for its shocks of 0.51. $\beta_{2,t}$ and $\beta_{3,t}$ are less persistent and more

Table 2.3: **Parameter estimates for the factor processes**

	ϕ_i	μ_i	σ_i
for $\beta_{1,t}$	0.981	0.112	0.514
	(0.012)	(0.082)	(0.026)
for $\beta_{2,t}$	0.822	-0.293	0.875
	(0.038)	(0.086)	(0.044)
for $\beta_{3,t}$	0.784	0.027	1.222
	(0.043)	(0.077)	(0.064)

Sample period: 1953:Q2-2007:Q4. The parameters are denoted according to equation (2.4.3). Standard errors of estimates are reported in parentheses.

volatile than the level factor. The Augmented Dickey-Fuller (ADF) tests for unit roots in $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ suggest that $\beta_{1,t}$ may have unit root with p-value 0.575 while $\beta_{2,t}$, $\beta_{3,t}$ appear to be stationary with p-values 0.002 and <0.001, respectively. The ADF tests for unit roots in all yields, reported in the last column of Table 2.1, indicate that all yields may have unit roots.

Cointegration tests using the Johansen (1998) method suggest that the yields are cointegrated with each other.[5] Based on these results, I also considered a version of the model where yields are assumed to be cointegrated unit root processes.[6] Forecast results for real GDP growth in the stationary and unit root specifications are close to each other and there is no dominant model; therefore, I focus only on the model with the stationary specification in the remaining analysis.[7]

Table 2.4 reports estimates of the factor loadings for real GDP growth in the dynamic yield curve model. In contrast to the estimation results for the yield spread model, all estimates of the factor loadings for real GDP growth are statistically insignificant, although they are economically significant given their point estimates are considerably different from zero. The negative sign of the slope coefficient γ_2 for real GDP growth is consistent with the interpretation of $\beta_{2,t}$ as minus the slope of the yield curve.

The estimate of the coefficient for the lagged real GDP growth rate, denoted as γ_4, is statistically significant and its value is comparable with the estimate in AR(1) model suggesting that the autocorrelation component remains important

[5] The cointegration test suggests that elements of the vector of 3, 12, 36, 120 month yields are cointegrated with each other at a 5 percent level.

[6] In the unit root specification, it is assumed that $\beta_{1,t}$ is unit root process by restricting ϕ_1 to unity.

[7] The unit root dynamic yield curve model produces lower RMSEs of yields than the stationary model at long horizons. Forecasting performances of the models for real GDP growth relative to each other are mixed.

Table 2.4: **Parameter estimates for RGDP growth**

γ_1	γ_2	γ_3	γ_4	μ_g
-0.145	-0.344	0.253	0.303	0.303
(0.109)	(0.190)	(0.162)	(0.064)	(0.064)

Sample period: 1953:Q2-2007:Q4. The parameters correspond to equation (2.4.4). Standard errors of estimates are reported in parentheses.

Table 2.5: **Statistics for measurement errors of yields and RGDP growth**

maturity	Dynamic Model		OLS		AR(1)	
and RGDP	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
3	-0.01	1.02				
12	0.16	1.02				
36	0.01	0.84				
60	-0.01	0.75				
120	0.02	0.63				
RGDP growth						
1 quarter ahead	0.01	3.41	0.00	3.55	0.00	3.48
4 quarter ahead	0.07	2.20	0.00	2.16	0.06	2.30
8 quarter ahead	0.10	1.57	0.00	1.48	0.09	1.64
12 quarter ahead	0.10	1.28	0.00	1.48	0.09	1.64

Sample period: 1953:Q2-2007:Q4. The dynamic yield curve and the yield spread models include one lag of real GDP growth. OLS denotes a regression of RGDP growth on the yield spread and one lag of RGDP growth.

after controlling for the yield factors.

Table 2.5 reports statistics for the measurement errors of yields and real GDP growth based on the in-sample fit of the dynamic yield curve model, OLS yield spread model, and an AR(1) model. All these models control for one lag of real GDP growth.

The dynamic yield curve model has a better fit for real GDP growth at the one-quarter horizon, while the OLS yield spread model has a better fit at most of the other horizons. The fit of real GDP growth by the OLS yield spread model at long horizons is explained by the forecasting specification of the model and the nature of OLS regression, which is to minimize squared residuals. Specifically, the OLS yield spread model has an advantage in terms of in-sample fit over the dynamic yield curve model, because the former is a forecasting model at targeted horizons, while the dynamic yield curve model fits the current data.[8] Meanwhile, both the dynamic yield curve model and the OLS yield spread model have a better fit than the univariate autoregressive model because they nest the AR(1) model.

2.6 Out-of-sample Forecasting Results

2.6.1 Forecasting Procedure and Notation

Pseudo out-of-sample forecasts of real GDP growth are performed for the period from 1990:Q1 through 2007:Q4. The forecast performance of models is compared using root mean square errors (RMSEs) relative to a benchmark model. Following Stock and Watson (2003), I use the RMSEs for the AR(1) model at different horizons as benchmarks.

The RMSE statistic for the dynamic yield curve model is generated using the following procedure. First, the parameters of the state-space model are estimated using Kalman filter method and then yields and real GDP growth are forecasted

[8] To check this point, I estimated a dynamic yield curve model with the specification changed to be similar to a direct forecasting model. Even with a forecasting specification at one period ahead and iterating for longer horizon forecasts, the in-sample fit for the forecasting dynamic yield curve model improved over the results of the OLS yield spread models for most horizons. Despite the obvious advantage of the forecasting specification of the dynamic yield curve model, I use the contemporaneous version of the model for this study as it uses all available current information for out-of-sample forecasting. Also, out-of-sample forecasting results suggest that the contemporaneous model outperforms the model with a forecasting specification.

for 1 to 12 quarters ahead. Next, one more observation is added to the in-sample data and the estimation and forecasting are repeated. This procedure produces 73-k observations of k-quarter-ahead out-of-sample forecasts for k from 1 to 12 quarter horizons.

For the forecast performance comparisons at a given horizon, I use a cumulative real GDP growth averaged for the whole horizon rather than marginal one-period forecasts at that horizon. This choice of the forecast comparison is explained by the iterative approach to constructing forecasts using the NS dynamic yield curve model. For this approach, the quality of forecasts at a given horizon depends directly on the quality of the forecasts at all previous periods.

I compare the out-of-sample forecast performance of the two classes of models: the NS dynamic yield curve models and the OLS yield spread models. For each class of models, I consider several specifications of the models with different explanatory variables for real GDP growth. I denote the class of dynamic yield curve models as NS and the class of yield spread models as PR, which stands for "predictive regression". To denote the specification of a model in each class of models, the explanatory variables used to model real GDP growth are listed in parentheses. For example, the notation $NS(g(\beta_2, \beta_3, g_{t-1}))$ means that this is the NS dynamic yield curve model with real GDP growth modeled by $\beta_{2,t}$, $\beta_{3,t}$ factors and one lag of real GDP growth g_{t-1}.

2.6.2 Forecasts of Real GDP Growth

In this subsection, I analyze effects of different explanatory factors for the real GDP growth forecasts. Table 2.6 reports RMSE results for different versions of the NS dynamic yield curve and the OLS yield spread models.

The dynamic yield curve model with lagged real GDP growth has lower RMSEs than models without lagged real GDP growth. Most of the improvement is observed at short horizons. Similarly, adding lagged real GDP growth in the OLS yield spread model improves forecasts at short horizons. The positive effect of the autoregressive component in the short-term horizon forecasts reflects the short-term persistence of real GDP growth.

The RMSEs for models with a curvature factor $\beta_{3,t}$ are smaller than for models without this factor at long horizons. At short horizons, RMSEs for both models are close to each other. Although the gain at the long horizon from adding the curvature factor to the slope factor for real GDP growth forecasting is relatively small, it still extracts additional information contained in yield curve for real GDP modeling, while the OLS yield spread model does not contain this information. This result concurs with Huang et al. (2006) who find some usefulness of the curvature factor for forecasting output in their study of forecast combination using information in the yield curve.

Adding the level factor $\beta_{1,t}$ as an explanatory variable for real GDP growth in the dynamic yield curve model produces lower RMSEs at short horizons. However, as will be shown below, this result is not robust to the choice of the out-of-sample

Table 2.6: **Out-of-sample forecasts of RGDP growth rates: Root Mean Square Error Ratios. Out-of-sample period 1990:Q1-2007:Q4**

	Forecast horizon k-quarters ahead			
	1	4	8	12
Dynamic Yield Curve Models				
NS($g(\beta_2)$)	1.046	1.048	1.022	1.007
NS($g(\beta_2, g_{t-1})$)	1.009	1.010	1.002	0.996
NS($g(\beta_2, \beta_3, g_{t-1})$)	1.006	1.014	**0.996**	**0.981**
NS($g(\beta_1, \beta_2, \beta_3, g_{t-1})$)	**0.999**	**1.008**	1.013	1.021
VAR($Spread, g_{t-1}$)	1.007	1.016	1.013	1.012
Yield Spread Models				
PR($Spread$)	1.130	1.365	**1.366**	**1.197**
PR($Spread, g_t$)	**1.076**	**1.323**	1.382	1.208
PR($Shrt.Rate, Spread, g_t$)	1.099	1.420	1.525	1.307

NS and PR denote the dynamic yield curve and OLS yield spread models, respectively. The denominators are the RMSEs for an AR(1) model. The lowest RMSE ratios within each class of models are in bold.

period. Also, at long horizons, adding the level factor has a negative effect on the forecasting performance of the model. Similarly, adding the short rate to the yield spread model increases the RMSEs.

2.6.3 Does the Dynamic Yield Curve Model Forecast Output better than the Yield Spread Model?

To answer the question of whether the dynamic yield curve model improves forecasts of real GDP growth over the OLS yield spread model, I compare RMSEs for the following pairs of models with comparable explanatory variables for real GDP growth: $NS(g(\beta_2))$ and $PR(Spread)$; $NS(g(\beta_2, g_{t-1}))$ and $PR(Spread, g_t)$; $NS(g(\beta_2, \beta_3, g_{t-1}))$ and $PR(Spread, g_t)$. Table 2.6 reports noticeably lower RMSEs for the dynamic yield curve models than for OLS yield spread models at all horizons. The Diebold and Mariano (1995) (hereafter DM(1995)) test of forecast accuracy comparison, reported in Table 2.7, suggests that these differences in RMSEs are statistically significant.[9] This result remains robust to controlling for lagged real GDP growth in both models. Thus, the dynamic yield curve model outperforms the OLS yield spread model in forecasting real GDP growth.

Table 2.7: **Diebold-Mariano tests for comparative forecast accuracy**

Models	Forecast horizon k-quarters ahead			
	1	4	8	12
NS($g(\beta_2)$)	-0.756	-1.439	-0.863	-0.308
against PR($Spread$)	(0.025)	(0.032)	(0.011)	(0.047)
NS(g(β_2,g_{t-1}))	-0.584	-1.373	-0.951	-0.343
against PR($Spread, g_t$)	(0.013)	(0.021)	(0.004)	(0.021)
NS(g(β_2,β_3,g_{t-1}))	-0.606	-1.357	-0.964	-0.366
against PR($Spread, g_t$)	(0.023)	(0.026)	(0.004)	(0.021)
NS(g(β_2,β_3,g_{t-1}))	0.052	0.053	-0.009	-0.028
against AR(1)	(0.660)	(0.673)	(0.889)	(0.441)
PR($Spread, g_t$)	0.658	1.410	0.955	0.338
against AR(1)	(0.040)	(0.042)	(0.010)	(0.061)

NS and PR denote the dynamic yield curve and OLS yield spread models, respectively. The null hypothesis of the test is that the mean of square loss-differential of two models is zero, against alternative that it is not zero. Negative (positive) value of the estimate indicates that the first model produces more (less) accurate forecasts than compared model. The p-values for the test are reported in parentheses. The test is based on the Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors.

Given that the dynamic NS model with real GDP growth modeled by the slope factor outperforms the yield spread model with a direct forecasting approach at all horizons, there are two potential sources for the forecast improvement. In particular, the improvement could originate from i) an iterative forecasting scheme versus the

[9]While there are several tests of forecast accuracy (e.g. West (1996) and Giacomini and White (2006)), the choice of the Diebold and Mariano (1995) test for this study is explained by the focus on out-of-sample performance and simplicity of the test application.

direct forecasting approach used in the yield spread model, and/or ii) from using an endogenously estimated slope factor versus a particular observable yield spread. To check these two possibilities, I evaluate out-of-sample forecasts using a simple VAR(1) model with an observable term spread and real GDP growth as variables:

$$\begin{pmatrix} Spread_t \\ g_{t-1} \end{pmatrix} = \begin{pmatrix} \mu_s \\ \mu_g \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} Spread_{t-1} \\ g_{t-2} \end{pmatrix} \tag{2.6.1}$$

The RMSEs for the VAR model, denoted as $VAR\,(Spread, g_{t-1})$ and reported in Table 2.6, are considerably smaller than those from the yield spread model $PR(Spread, g_t)$. Thus, the dynamic approach for forecasting real GDP appears to be the main source of the forecast improvement over the direct forecasting approach. As noted earlier, the OLS regression for a targeted forecasting horizon may cause overfitting of in-sample data due to the "least squares" nature of the inferences. This point is supported by the fact that the yield spread model performs considerably worse than the dynamic yield curve model in out-of-sample forecasts, while it has the best in-sample fit. Thus, poor out-of-sample performance of the yield spread model indicates that the yield curve is less useful for GDP forecasting than suggested by the in-sample OLS regression.

In addition, while the RMSEs for the $VAR\,(Spread, g_{t-1})$ and $NS(g(\beta_2, g_{t-1}))$ models are close to each other at the short horizons, the NS dynamic yield curve model outperforms the VAR at long horizons. Thus, there is also a gain from using the endogenously-estimated slope factor versus the observable yield spread. Modeling real GDP growth by endogenously determined factors avoids the problem of dependence of results on the choice of the maturities for the yield spread.

2.6.4 Do the Dynamic Yield Curve and Yield Spread Models Forecast Output Better than an AR(1)?

It is important to note that the yield spread model cannot beat the AR(1) model in the out-of-sample period and the dynamic NS yield curve model improves forecasts over the AR(1) model only marginally at long horizons. The DM(1995) test, reported in Table 2.7, suggests that the AR(1) model produces significantly smaller forecast errors than the OLS yield spread model. The differences in RMSEs for the dynamic yield curve model and the AR(1) model are small in magnitude and statistically insignificant. These results can be explained in two ways.

Table 2.8: **Out-of-sample forecasts of RGDP growth rate: Root Mean Square Error Ratios. Different in-sample and out-of-sample periods**

	Forecast horizon k-quarters			
	1	4	8	12
In-Sample period: 1953:Q2-1997:Q4; Out-of-sample period 1998-2007				
NS($g(\beta_2, g_{t-1})$)	1.002	0.997	0.990	0.988
NS($g(\beta_2, \beta_3, g_{t-1})$)	0.994	0.998	0.978	0.959
NS($g(\beta_1, \beta_2, \beta_3, g_{t-1})$)	0.997	1.045	1.068	1.107
PR($Spread, g_t$)	1.045	1.251	1.235	1.152
VAR($Spread, g_{t-1}$)	1.000	1.007	1.010	1.021
In-Sample period: 1985:Q1-1997:Q4; Out-of-sample period 1998-2007				
PR($Spread, g_t$)	1.008	1.036	1.030	0.935
In-Sample period: 1953:Q2-1970:Q4; Out-of-sample period 1971-1984				
PR($Spread, g_t$)	0.981	0.805	0.741*	0.909

NS and PR denote the dynamic yield curve and OLS yield spread models, respectively. Denominators are RMSEs for the AR(1) model from respective samples. One asterisk indicates statistical significance of forecast improvements compared to the AR(1) model at 5 percent level based on the Diebold-Mariano (1995) test.

First, there is some evidence for structural instability in the yield curve and output relationship reported in the literature. Haubrich and Dombrosky (1996) and Dotsey (1998) find a decline in predictive ability of the yield curve for output in the period after 1985. Estrella, Rodrigues, and Schich (2003), using the test for unknown break date, also find some evidence of structural instability in the yield spread and industrial production relationship in 1983. To analyze the effect of this structural instability on the forecast performance of the OLS yield spread model, I perform out-of-sample forecasts of real GDP using the in-sample period from 1985:Q1 to 1997:Q4. The beginning of this period is chosen based on the previous literature and the end of period is extended from 1989:Q4, used in my previous analysis, to 1997:Q4 to allow for a sufficient number of observations for in-sample estimation. It leaves the period 1998-2007 for out-of-sample forecasts. The first two panels in Table 2.8 report RMSE ratios for the period 1998-2007 from the OLS yield spread model based on two in-sample periods: 1953:Q2-1997:Q4 and 1985:Q1-1997:Q4. The RMSEs for the OLS yield spread model based on the post 1985 in-sample period is noticeably smaller, suggesting a possible structural break in the relationship between the yield curve and output. However, the yield spread model still cannot improve real GDP forecasts relative to the AR(1) model at most of horizons.

I also estimate the dynamic yield curve model based on the sample period of 1953-1997 and perform out-of-sample forecasts for the period of 1998-2007. Although this sample period change does not fully address structural change in parameters, it should still reduce any bias of parameter estimates given that the sample period contains more post regime-shift observations.[10] Even after this partial ad-

[10]Forecasting using the dynamic yield curve model based on the in-sample period 1985-1997 is not performed due to the high number of model parameters relative to the small number of in-sample observations. The large standard errors of parameter estimates based on this short in-sample period overweigh the potential benefit from just using the post-structural-break data.

justment in parameters for a possible regime shift, the dynamic yield curve model forecasts output better than the AR(1) model at all horizons.[11] The dynamic yield curve model also outperforms the OLS yield spread model at most of horizons.

Second, the predictive power of the yield spread for output is mainly concentrated in periods of large changes in economic conditions. Previous research findings show that the yield spread is a relatively good predictor of recessions (e.g.Estrella and Mishkin (1996) and Estrella and Mishkin (1998)). Meanwhile, the AR(1) model has a good predictive performance in periods of low volatility. The period 1990-2007 had considerably more observations with relatively low volatility in real GDP growth than in previous years. Even the two recessions within this period were not as deep as those in preceding years. Thus, the AR(1) model has an advantage over the yield models in this out-of-sample period. The results are opposite if 1971-1984 is considered as the out-of-sample period for the OLS yield spread model. This period is characterized by high volatility in the business cycle and substantial changes in real GDP growth. The ratios of RMSEs for the OLS yield spread model to those for the AR(1) model, reported in the third panel of Table 2.8, suggest that the OLS model produces better forecasts than the AR(1) model in periods of large fluctuations in real GDP. Since the dynamic yield curve also uses yield information for predicting output, presumably it would have outperformed the AR(1) model in that out-of-sample period.[12]

[11] The DM(1995) test of forecast accuracy suggests statistical insignificance of all improvements over AR(1) model performance, which might be related to weak power due to the short out-of-sample period.

[12] A similar comparison for the period 1970-1990 with the dynamic yield curve model is not preformed because of the short in-sample size relative to the number of parameters in the dynamic yield curve model. The model estimation based on this short in-sample period would produce highly inefficient parameter estimates, negatively affecting the quality of forecasts.

Unlike the 1998-2007 period, the "Great Recession" in 2007-2009 involved large movements in real GDP. The question then is whether the yield curve information improves forecasts of real GDP over the benchmark AR(1) model for this period. In order to answer this question, I consider predictions of real GDP implied by the NS dynamic model, the yield spread model, and the AR(1) model for the period of 2007:Q1-2009:Q4. Figure 2.2 displays the one-year-ahead and two-year-ahead forecasts of log real GDP for the period of 2007-2009 using three models: AR(1), PR($Spread, g_t$), and NS($g(\beta_2, \beta_3, g_{t-1})$). None of the thee models predicts the severe decline in real GDP prior to the occurrence of the recession. However, the NS dynamic model and the yield spread model performed better than the AR(1) model. The RMSE ratios for the NS dynamic yield curve model relative to those for the AR(1) model for the period of 2007-2009 have values of 0.944 and 0.875 for 4-quarter and 8-quarter-ahead, respectively. The RMSE ratios for the yield spread model have values of 0.976 and 0.879 for 4-quarter and 8-quarter-ahead, respectively. These results suggest that the NS dynamic model predicted real GDP in this period better than the yield spread model. Also, these results confirm that the yield curve is more useful for forecasting output when there are large changes in output than when it is relatively stable.

2.7 Conclusion

Most studies that investigate the predictive power of the yield curve for real GDP growth consider a simple direct forecasting structure with yield spread as the predictive variable. In this paper, I have considered a different approach. In particular, I have jointly modeled real GDP growth and yields using the dynamic three-factor

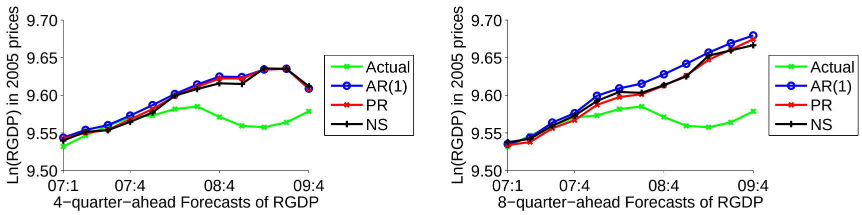


Figure 2.2: **Forecasts of Real GDP for 2007-2009**
This figure displays the 4-quarter-ahead and 8-quarter-ahead forecasts of logarithm of real GDP in 2005 prices for the period of 2007-2009 using three models: AR(1), PR($Spread, g_t$), and NS($g(\beta_2, \beta_3, g_{t-1})$). All forecasts are based on in-sample periods starting from 1985:Q1 and ending 4 quarters and 8 quarters prior to the forecasted period.

yield curve model.

My empirical findings suggest that the dynamic yield curve model produces better out-of-sample forecasts of real GDP growth than the traditional yield spread model. This result is mainly attributed to the dynamic structure of the yield curve model. Although the predictive power of yield curve is concentrated in the yield spread, there is also a gain from extracting more information from the term structure of interest rates versus an exogenously-defined yield spread used in the yield spread model. In particular, there is a gain from using information in the curvature factor for the long horizon prediction. In general, through, the yield curve is less useful for out-of-sample prediction of real GDP than the predictive power suggested by in-sample OLS regression analysis.

Chapter 3

Time Variation of CAPM Betas across Market Volatility Regimes[1]

3.1 Introduction

The Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) remains a benchmark asset pricing models in the academic literature. According to the CAPM, the risk of an asset is measured by its "beta", which is the covariance between the asset's return and the return on the market portfolio per unit of variance for the market return. A number of studies (e.g., Fama and French (1992, 1993, 1996)) have examined the CAPM with constant betas (i.e., the unconditional CAPM) and reported that the model performs poorly and is unable to explain certain asset pricing anomalies. In particular, they find that the unconditional CAPM cannot explain why i) portfolios of small firms outperform those of large firms (the "size" effect), ii) portfolios of firms with high Book-to-market (B/M) ratios outperform those for firms with low B/M ratios (the "B/M" effect), and iii) portfolios of firms with relatively high returns in the past year outperform those for firms with relatively low past returns (the "Momentum" effect).

[1]This essay is a joint work with James Morley

One of the explanations for the failure of the CAPM is its assumption that beta for a given portfolio and the market risk premium are constant over time. Many papers report that betas are time varying (e.g., Jagannathan and Wang (1996); Lettau and Ludvigson (2001); Fama and French (1997, 2006); Lewellen and Nagel (2006); and Ang and Chen (2007)). Jagannathan and Wang (1996) show that "alpha" from test regressions for the unconditional CAPM, where "alpha" corresponds to the expected excess return for the portfolio over what would be predicted by the unconditional CAPM, is theoretically related to the covariance between a time-varying beta and a time-varying market risk premium. They and several other studies (e.g., Lettau and Ludvigson (2001)) argue that capturing this covariance can help to explain the size and B/M anomalies.

In order to estimate time variation in CAPM betas, most previous studies use rolling windows of historical data, as proposed by Fama and MacBeth (1973). However, Lewellen and Nagel (2006) argue that these studies of the conditional CAPM based on cross-sectional regressions do not impose important theoretical restrictions in the estimation of the covariance between beta and the market risk premium. Also, Ang and Kristensen (2010) argue that the Fama and MacBeth (1973) method produces inconsistent estimates of the standard errors for average and conditional alphas. To address these limitations, some papers (e.g., Ang and Chen (2007) and Adrian and Franzoni (2009)) study the conditional CAPM by modeling time variation in betas as stationary latent variables. This is the approach we take in this paper, except that we model large and discrete changes in betas rather than assuming smooth and continuous changes, as has been previously done in the literature. In particular, we focus on investigating time variation in betas for the B/M and momentum portfolios across states of the economy corresponding to discrete changes in the level of stock market volatility and the market risk premium.

Our consideration of discrete changes in CAPM betas is motivated by numerous previous studies that find large discrete changes in the level of stock market volatility. For example, Hamilton and Susmel (1994) find that most of the ARCH effects in weekly stock returns vanish at the monthly horizon and the remaining persistent low frequency changes in volatility can be captured by a discrete Markov-switching process that appears somewhat related to discrete changes in business cycle phases between periods of expansion and recession.[2] Thus, if low frequency changes in volatility are abrupt and priced by market participants, one might expect the market risk premium to change in a discrete way too. Then, according to the idea of the conditional CAPM, if any changes in beta coincide with the discrete changes in market volatility, they could explain the empirical failure of the unconditional CAPM.

For our analysis, we follow Turner, Startz, and Nelson (1989) and Kim et al. (2004) by assuming that i) stock market volatility follows a two-state Markov-switching process, with the market risk premium varying across these "low" and "high" volatility regimes and ii) the processing of information about the prevailing volatility regime generates a volatility feedback effect that needs to be accounted for in order to reveal a positive underlying relationship between market volatility and

[2]Schwert (1989); Chu, Santoni, and Liu (1996); Schaller and van Norden (1997); Assoe (1998); Kim, Nelson, and Startz (1998); Kim, Morley, and Nelson (2001, 2004); Hess (2003); and Mayfield (2004), among many others, have modeled monthly stock return volatility using a Markov-switching specification, with high volatility regimes typically corresponding to periods of recession and low volatility regimes typically corresponding to periods of expansion. Perez-Quiros and Timmermann (2000) and Guidolin and Timmermann (2008) also find evidence of discrete changes in stock return risk across business cycle phases. Huang (2000) considers a Markov-switching beta for a single stock, but he does not relate it to market volatility regimes or business cycle phases.

the market risk premium. According to the idea of volatility feedback, an exogenous and persistent increase in the volatility of market news leads to additional return volatility as stock prices adjust in response to higher future expected returns.[3] We then jointly model the market return and the conditional CAPM, with time variation in beta driven by the market volatility regimes.

The Markov-switching specification for the conditional CAPM has two benefits over the traditional approaches taken in the literature. First, we do not have to find exogenous variables to try to identify time variation in the market risk premium, thus helping us to avoid any data mining concerns with an instrumental variables approach. Second, the timing of changes in beta, which correspond to changes in the market risk premium, is determined directly by the data, rather than imposed exogenously. For example, this has a benefit over a rolling window approach, which will naturally smooth out discrete changes in beta and the results for which will depend highly on the choice of window length.

Consistent with the basic idea of the conditional CAPM, our empirical findings suggest strong time variation in betas across market volatility regimes in most of the cases for which the unconditional CAPM can be rejected. For "value" portfolios of stocks for firms, which have relatively high B/M ratios, and "winner" portfolios of stocks, which have relatively strong returns over the previous year, the regimes alternate between periods of low market volatility/high beta and periods of high market volatility/low beta. For "loser" portfolios of stocks, which have relatively weak returns over the previous year, the regimes alternate between periods of low

[3] French, Schwert, and Stambaugh (1987), Turner et al. (1989), Campbell and Hentschell (1992), and Kim et al. (2004), among many others, account for volatility feedback to study the relationship between stock returns and volatility.

market volatility/low beta and periods of high market volatility/high beta. Although the regime-switching conditional CAPM can still be rejected in many cases, the time-varying betas help explain portfolio returns much better than the unconditional CAPM, especially when market volatility is high.

The rest of this paper is organized as follows. Section 2 presents the model. Section 3 describes the data and reports the empirical results. Section 4 concludes.

3.2 Model

According to the Sharpe-Lintner Capital Asset Pricing Model (CAPM), the expected excess return on a portfolio of assets over a risk-free rate depends on a simple measure of the portfolio's risk relative to the market portfolio:

$$E\left[r_{i,t}\right] = \beta_i E\left[r_{m,t}\right], \tag{3.2.1}$$

where $r_{i,t}$ is the excess return for portfolio i, $r_{m,t}$ is the market excess return, and β_i is the measure of the portfolio's risk defined as

$$\beta_i = \frac{cov\left(r_{i,t}, r_{m,t}\right)}{var\left(r_m\right)}. \tag{3.2.2}$$

Fama and French (1992) examine the performance of the unconditional CAPM and find that estimated betas do not explain variation in average returns across different portfolios. One of the explanations for the failure of the CAPM is its assumption that the market risk premium and beta are both constant over time.

Relaxing this assumption we get the conditional CAPM, which holds period by period:

$$E_{t-1}\left[r_{i,t}\right] = \beta_{i,t-1} E_{t-1}\left[r_{m,t}\right]. \tag{3.2.3}$$

In this equation, subscript $t-1$ indicates that everything in the model is conditional on information available to market participants in the previous period. Following Jagannathan and Wang (1996) and applying iterated expectations on both sides of equation (3.2.3), we get

$$E\left[r_{i,t}\right] = \overline{\beta}_i E\left[r_{m,t}\right] + cov(\beta_{i,t-1}, E_{t-1}\left[r_{m,t}\right]), \tag{3.2.4}$$

where $\overline{\beta}_i$ is the unconditional expectation of beta. Thus, it is straightforward to see that the unconditional CAPM would fail if beta were correlated with the market risk premium.

Lewellen and Nagel (2006) analytically decompose the covariance term in equation (3.2.4) and compute the upper bound of this term based on various assumptions about parameter values. However, equation (3.2.4) and the proposed upper bound of the covariance are valid only if beta is stationary. Instead, if beta were an integrated process, then the unconditional expectation of beta and the covariance term in equation (3.2.4) would not exist, which means that the upper bound could not be computed. In this case, the failure of the unconditional CAPM would increase as the number of observations goes to infinity. In this paper, we assume beta is stationary. However, by allowing it to follow a Markov-switching process, we can capture very persistent changes in beta.

The conditional CAPM requires specifying the information available to market participants when they form conditional expectations of the market risk premium and beta. In this paper, we assume that market participants know that risk changes in a discrete way, distinguishing only between "good" and "bad" states of the economy related to market volatility. Following many studies, including Turner et al. (1989) and Kim et al. (2004), we model states of the economy with a two-state Markov-switching variance for the market excess return:

$$\begin{aligned}
\varepsilon_{m,t} &\sim N\left(0, \sigma^2_{m,S_{m,t}}\right), && (3.2.5)\\
\sigma^2_{m,s_{m,t}} &= \sigma^2_{m,0}\left(1 - S_{m,t}\right) + \sigma^2_{m,1} S_{m,t}\ ,\ \sigma^2_{m,0} < \sigma^2_{m,1}, && (3.2.6)\\
\Pr\left[S_{m,t} = 0 | S_{m,t-1} = 0\right] &= q_m \text{ and } \Pr\left[S_{m,t} = 1 | S_{m,t-1} = 1\right] = p_m, && (3.2.7)
\end{aligned}$$

where $\varepsilon_{m,t}$ denotes the market news at time t, $\sigma^2_{S_{m,t}}$ is the variance of $\varepsilon_{m,t}$, $S_{m,t}$ is a Markov-switching state variable that takes value 0 in the low volatility regime and 1 in the high volatility regime, and q_m and p_m are continuation probabilities for the regimes.

In this two-state specification of market volatility, one possible informational assumption is that market participants perfectly observe the current state of market volatility. Under this assumption, the period-by-period market risk premium can be expressed as

$$E\left[r_{m,t} | S_{m,t}\right] = \mu_{m,0} + \mu_{m,1} S_{m,t}, \tag{3.2.8}$$

where $\mu_{m,0}$ denotes the market risk premium in the low volatility regime and $\mu_{m,1}$ determines the marginal effect of the high volatility regime on the market risk premium. However, consistent with past findings, we find a negative estimate for $\mu_{m,1}$ in our empirical analysis. This result runs contrary to the theoretical positive relationship between risk and return, suggesting that, although market participants react to information inherent in the true volatility regimes, they may take time to process information about the prevailing volatility regime.

Campbell and Hentschell (1992) and Kim et al. (2004), among many others, account for volatility feedback, which helps to reveal a positive relationship between volatility and return. According to the idea of volatility feedback, an exogenous and persistent increase in the volatility of market news generates additional return volatility as stock prices adjust in response to higher future expected returns. We follow Kim et al. (2004) and consider a Markov-switching model of the market excess return with volatility feedback, which is specified as

$$r_{m,t} = E\left[r_{m,t}|S_{m,t-1}\right] + f_{m,t} + \varepsilon_{m,t}, \tag{3.2.9}$$

where

$$\begin{aligned} E\left[r_{m,t}|S_{m,t-1}\right] &= \mu_{m,0} + \mu_{m,1}\,\mathsf{Pr}\left[S_{m,t} = 1|S_{m,t-1}\right], & (3.2.10)\\ f_{m,t} &= \delta\left\{S_{m,t} - \mathsf{Pr}\left[S_{m,t} = 1|S_{m,t-1}\right]\right\}. & (3.2.11)\end{aligned}$$

The $f_{m,t}$ term captures an unpredictable volatility feedback effect on the market return due to period-by-period revisions in future expected returns, where $E\left[f_{m,t}|S_{m,t-1}\right] = 0$. The δ coefficient in the volatility feedback term is related to the other model parameters based on a discounted sum of log-linear future expected returns, as shown in Kim et al. (2004). Specifically, the coefficient is equal to $\delta = \frac{-\mu_{m,1}}{1-\rho\lambda}$, where $\lambda = p_m + q_m - 1$ and ρ denotes the parameter of linearization for the log-linear present value model, which is the average ratio of the stock price to the sum of the stock price and the dividend and, in practice, has the value of 0.997, as reported in Kim et al. (2004). In this specification, it is assumed that market participants observe the previous volatility regime $S_{m,t-1}$ at the beginning of the current period, but learn about the current volatility regime $S_{m,t}$ during the current period.

Similar to the market excess return, we assume that the portfolio excess return is specified as

$$r_{i,t} = E\left[r_{i,t}|S_{m,t-1}\right] + f_{i,t} + \varepsilon_{i,t}, \tag{3.2.12}$$

where $E\left[r_{i,t}|S_{m,t-1}\right]$ is defined by the conditional CAPM, $f_{i,t}$ is the volatility feedback term for the portfolio return, and $\varepsilon_{i,t}$ is news about portfolio i. Because the conditional CAPM time-varying beta may covary with the time-varying market risk premium, which in our setting takes on two discrete values conditional on the market volatility regimes, we allow for different values of beta in these two regimes.[4] Thus, the regime-switching conditional CAPM is given by

$$E\left[r_{i,t}|S_{m,t-1}\right] = \beta_{i,S_{m,t-1}} E\left[r_{m,t}|S_{m,t-1}\right], \tag{3.2.13}$$

where $\beta_{i,S_{m,t-1}}$ takes on two values depending on the market volatility regime at period $t-1$.[5] Also, substituting for $r_{i,t}$ and $r_{m,t}$ in equation (3.2.13) based on

[4]An alternative approach would be to assume that beta has its own Markov-switching process. However, from the theory of the conditional CAPM, the relevant issue for the failure of the unconditional CAPM is whether beta covaries with the market risk premium, which in this case is driven by market volatility. Therefore, for simplicity, we consider a specification with common regimes for beta and market volatility.

[5]The beta used to price a portfolio depends on expectations of $S_{m,t}$. Thus, the beta will depend on the sensitivity of the portfolio to market news in both regimes, with the weights on the two regimes depending on the continuation probabilities for the Markov-switching state variable. Analytically, given fixed continuation probabilities, this assumption is equivalent to specifying beta to be a function of $S_{m,t-1}$, as this will capture the weighted-average sensitivity for the portfolio

equations (3.2.12) and (3.2.9), we can show that

$$E\left[f_{i,t}|S_{m,t-1}\right] = \beta_{i,S_{m,t-1}} E\left[f_{m,t}|S_{m,t-1}\right] = 0,$$

which is consistent with the CAPM notion that the expected excess return for a portfolio depends only on its beta and the market risk premium.

Based on equations (3.2.9) and (3.2.13), our joint model of market and portfolio excess returns is given as follows:

$$\begin{aligned} r_{m,t} &= \mu_{m,0} + \mu_{m,1} \Pr\left[S_{m,t} = 1|S_{m,t-1}\right] \\ &\quad + \delta\left\{S_{m,t} - \Pr\left[S_{m,t} = 1|S_{m,t-1}\right]\right\} + \varepsilon_{m,t} \qquad (3.2.14) \\ r_{i,t} &= \alpha_{i,S_{m,t-1}} + \beta_{i,S_{m,t-1}} r_{m,t} + u_t \qquad (3.2.15) \\ \varepsilon_{m,t} &\sim N\left(0, \sigma^2_{m,S_{m,t}}\right) \quad \text{and} \quad u_t \sim N\left(0, \sigma^2_{i,S_{i,t}}\right), \end{aligned}$$

where u_t denotes idiosyncratic news for portfolio i, which according to the CAPM should be uncorrelated with market news. In this model, the regime-switching process for market volatility and the alpha and beta for portfolio i is driven by a common unobservable state variable $S_{m,t}$ that takes on discrete values of 0 in the low market volatility regime and 1 in the high market volatility regime. If the conditional CAPM holds, $\alpha_{i,S_{m,t-1}} = 0$ in both regimes. In addition to regime-switching market volatility, we also control for heteroskedasticity in the residual for the portfolio return by assuming that the variance $\sigma^2_{i,S_{i,t}}$ of idiosyncratic news u_t follows a two-state Markov-switching process that is assumed to be independent of the process for market volatility. It should be noted that, in principle, we could consider more

conditional on $S_{m,t-1}$.

regimes for the Markov-switching parameters in the model. However, as we show in the empirical analysis, the residual diagnostics suggest that two-regime processes are sufficient to address heteroskedasticity in the residuals for the sample period under consideration. Meanwhile, given the Markov-switching structure, we estimate the model by applying maximum likelihood based on the procedure developed by Hamilton (1989).

3.3 Empirical results

3.3.1 Data

We consider monthly data for stock returns on value-weighted decile portfolios of all stocks listed on the NYSE, AMEX, and NASDAQ sorted separately by book-to-market ratios (B/M portfolios) and by the previous year's returns (Momentum portfolios).[6] The B/M portfolios are constructed at the end of June each year based on the ratio of the book equity of stocks for the previous fiscal year to their market capitalization in December of the previous year. The portfolios are formed annually by sorting stocks using decile breakpoints of B/M ratios for the NYSE stocks only. Momentum portfolios are constructed each month using the previous 11-month-return decile breakpoints for NYSE stocks. The portfolio returns are value-weighted monthly average returns on the stocks in deciles. We define the "market" return by considering the return on a value-weighted portfolio of all stocks listed on the NYSE, AMEX, and NASDAQ. All returns are continuously compounded in excess

[6]We are grateful to Kenneth French for making these data available at his data library at dhttp://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/. Detailed description of portfolio formation are provided in Fama and French (2006).

of the continuously compounded one-month Treasury bill rate and expressed in percentage terms. Most previous empirical studies only consider data for July 1963 and afterwards in order to focus on a period for which the unconditional CAPM fails to explain B/M and momentum effects (e.g., Ang and Chen (2007) find that the unconditional CAPM cannot be rejected for B/M portfolios over the longer sample period of 1926-2001). Therefore, we consider the sample period of July 1963 to December 2007 in our analysis. For this sample period, we do not observe a strong size effect for portfolio returns double-sorted by size and B/M ratios, which is common way of sorting portfolios in the literature, so we consider the overall B/M sorting.[7]

Table 3.1 reports summary statistics for the returns on B/M and momentum portfolios and estimates for the unconditional CAPM regression model. The results suggest a pattern of increasing average returns with increasing B/M ratios and momentum. Based on the estimated alphas, the unconditional CAPM can be rejected for the last five deciles of the B/M portfolios and for the first two and last three deciles of the momentum portfolios.

[7]By considering overall B/M-sorted portfolios, we are following Ang and Chen (2007). Although the average returns for the largest size portfolios are always smaller than average returns for other size portfolios, the average returns for size portfolios other than largest size portfolio are not always decreasing with size. Also, preliminary analysis, not reported to conserve space, suggests that the unconditional CAPM cannot be rejected for the size-sorted portfolios for the sample period of July 1963 to December 2007.

Table 3.1: **Summary statistics for book-to-market and momentum portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
	Panel A: B/M portfolios										
$\bar{r}_i$	0.24	0.35	0.41	0.44	0.41	0.53	0.60	0.64	0.69	0.77	0.53
std.dev.	(5.14)	(4.72)	(4.69)	(4.62)	(4.37)	(4.32)	(4.22)	(4.22)	(4.56)	(5.27)	(4.40)
α_i	-0.17	-0.04	0.02	0.08	0.07	**0.19**	**0.29**	**0.32**	**0.35**	**0.40**	**0.58**
std.error	(0.10)	(0.07)	(0.07)	(0.10)	(0.10)	(0.08)	(0.11)	(0.11)	(0.11)	(0.16)	(0.23)
β_i	1.09	1.03	1.02	0.98	0.91	0.90	0.84	0.84	0.90	0.98	-0.11
std.error	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)
	Panel B: Momentum portfolios										
$\bar{r}_i$	-0.59	0.07	0.24	0.31	0.23	0.33	0.37	0.59	0.64	0.99	1.58
std.dev.	(7.29)	(5.81)	(4.95)	(4.57)	(4.29)	(4.43)	(4.35)	(4.40)	(4.82)	(6.20)	(6.16)
α_i	**-1.10**	**-0.35**	-0.12	-0.04	-0.11	-0.02	0.03	**0.24**	**0.27**	**0.53**	**1.64**
std.error	(0.18)	(0.14)	(0.11)	(0.11)	(0.09)	(0.06)	(0.07)	(0.08)	(0.09)	(0.14)	(0.26)
β_i	1.36	1.12	0.97	0.93	0.90	0.93	0.91	0.92	1.00	1.21	-0.15
std.error	(0.07)	(0.06)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.11)

Data are for the value-weighted portfolios sorted into deciles of B/M ratios and the previous 12-month returns for the sample period of July 1963 to December 2007. HML denotes a "High minus Low" portfolio; $\bar{r}_i$ denotes the average excess return for portfolio i. Sample standard deviations for excess returns are reported in parentheses. Estimates of α_i and β_i for the unconditional CAPM regression model are based on OLS. Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors are reported in parentheses for α and β. Statistically significant alphas at the 5% level are in bold.

3.3.2 Regime-switching volatility and the estimated market risk premium

Table 3.2 reports estimates for regime-switching volatility and market risk premium based on the model of the market return given in equation (3.2.14). In this case, the model is estimated separately from consideration of portfolio returns and we consider both a restricted version of the model without volatility feedback (i.e., $\delta = 0$) and a version that allows for volatility feedback. The model without volatility feedback has a negative estimated market risk premium in the high volatility regime.

Notably, whether or not the market risk premium is actually negative, its estimate is significantly lower in the high volatility regime than in the low volatility regime. This result does not accord with a basic theoretical positive relationship between risk and return.[8] However, after accounting for volatility feedback, the estimates are consistent with a positive relationship. Meanwhile, a likelihood ratio (LR) test rejects the restricted model without volatility feedback with a p-value of <0.001 based on an asymptotic $\chi^2(1)$ distribution, suggesting that volatility feedback is an important feature of stock returns. Also, the estimates for market volatility are quite different across the two regimes. Thus, taken together, these results provide evidence of significant time variation in the market risk premium related to changes in market volatility.

Table 3.2: **Parameter estimates for regime-switching volatility and market risk premium**

Model	$\mu_{m,0}$	$\mu_{m,1}$	δ	$\sigma_{m,0}$	$\sigma_{m,1}$	q_m	p_m	$logL$
Model without	0.95	-1.61		3.07	5.91	0.96	0.93	-1512.07
volatility feedback	(0.21)	(0.85)		(0.20)	(0.48)	(0.02)	(0.05)	
Model with	0.24	0.76	-7.61	2.72	5.21	0.96	0.94	-1505.87
volatility feedback	(0.22)	(0.27)	(1.20)	(0.31)	(0.41)	(0.02)	(0.02)	

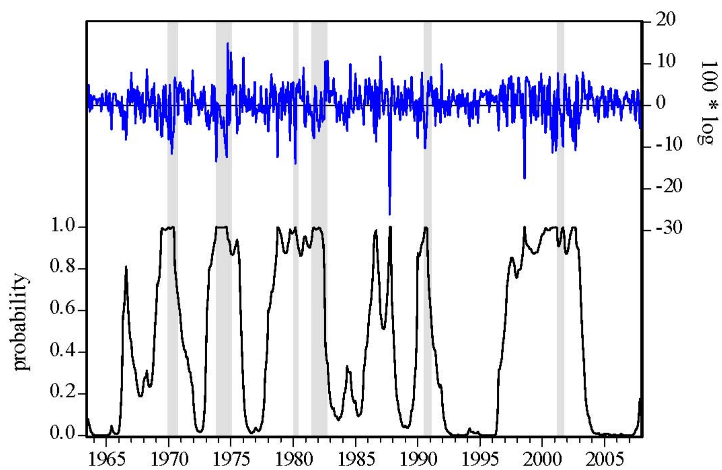
The model for the market return is described by equation (3.2.14), where $\delta = 0$ for the version of the model without volatility feedback. The standard error for the volatility feedback parameter estimate was obtained using the Delta method. $logL$ denotes the log likelihood.

Despite the differences in estimates of the market risk premium for the two models, the estimates related to volatility are quite similar. Also, smoothed probabilities of the volatility regimes for both models are similar, with a correlation of 0.90, suggesting that the regimes are mainly identified by changes in variance rather

[8]Breen, Glosten, and Jagannathan (1989), Campbell (1987), Nelson (1991), and Glosten, Jagannathan, and Runkle (1993), among many others, find a negative relationship between market volatility and the market risk premium. Glosten et al. (1993) argue that market participants may not require a larger risk premium in more risky periods because they may need to save relatively more for a future that may be even riskier.

than by changes in the mean of excess returns. However, given the significance of the feedback parameter, the remaining analysis in this paper is based on models with volatility feedback.

Figure 3.1: **Monthly stock market returns and smoothed probabilities of the high volatility regime**



Returns are continuously compounded monthly value-weighted returns for all stocks listed on the NYSE, AMEX, and NASDAQ in excess of continuously compounded one-month Treasury bill yields for the sample period of July 1963 to December 2007. Shaded areas correspond to NBER recessions.

In terms of the volatility regimes, the estimates of the continuation probabilities suggest that both regimes are very persistent, with 96% and 94% month-to-month probabilities of remaining in the low and high volatility regimes, respectively. From Figure 3.1, which displays the smoothed probabilities of the high volatility regime over the sample period, we observe that the periods of high stock market volatility include all of the NBER recessions and major stock market crashes.

3.3.3 Regime-switching betas for book-to-market portfolios

Do the betas for the B/M portfolios vary across market volatility regimes? To test for a regime-switching beta for a given portfolio, we consider another LR test. In this case, the LR test statistic is constructed based on the likelihood for a restricted version of the joint model of market and portfolio returns described by equations (3.2.14) and (3.2.15) in which alpha is allowed to be regime switching and beta is assumed to be constant across volatility regimes relative to the likelihood for a less restrictive version of the model in which both alpha and beta are allowed to be regime switching. Because both models have Markov-switching market volatility under the null hypothesis, they are nested without nuisance parameters and the LR statistic should have an asymptotic $\chi^2(1)$ distribution. The test results, reported in Table 3.3, support regime-switching betas for most of the B/M portfolios. Amongst the B/M portfolios for which the unconditional CAPM is rejected, the LR tests support regime-switching betas for all but the 7th decile portfolio at the 10% level. Notably, the LR tests support regime-switching betas for the 9th and 10th decile portfolios at the 1% level. The LR tests also support regime-switching betas for three of the B/M portfolios for which the unconditional CAPM cannot be rejected. It should be noted that the fact that the LR tests cannot reject a constant beta for the 2nd, 4th, and 7th decile portfolios only suggests that the betas for these portfolios do not have large changes over the market volatility regimes, but they may still be time varying. However, importantly for the conditional CAPM, they appear not be time varying in a way that corresponds to changes in the market risk premium.

Table 3.3: **Likelihood ratio tests for regime-switching betas and residual diagnostics for book-to-market portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
LR stat.	11.40	3.75	12.55	0.27	12.01	3.42	0.13	3.27	6.61	15.14	18.47
p-value	(0.00)	(0.05)	(0.00)	(0.60)	(0.00)	(0.06)	(0.72)	(0.07)	(0.01)	(0.00)	(0.00)
	Residual diagnostic tests: portfolio return with constant beta and variance										
ARCH-LM	0.67	23.24	21.40	89.49	40.68	7.08	26.48	19.20	15.15	17.05	9.78
p-value	(0.41)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
JB stat.	4.76	26.61	376.77	492.48	245.81	199.24	322.26	343.62	136.31	139.05	35.52
p-value	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Residual diagnostic tests: portfolio return with regime-switching beta and variance										
ARCH-LM	1.21	1.13	0.18	0.66	0.25	2.12	1.35	0.51	0.71	0.06	0.29
p-value	(0.27)	(0.29)	(0.67)	(0.42)	(0.62)	(0.15)	(0.25)	(0.47)	(0.40)	(0.80)	(0.59)
JB stat.	0.86	7.15	9.32	1.39	0.12	0.12	0.95	3.60	1.81	10.10	6.11
p-value	(0.65)	(0.03)	(0.01)	(0.50)	(0.94)	(0.94)	(0.62)	(0.17)	(0.40)	(0.01)	(0.05)

To test for a regime-switching β, we use likelihood ratio (LR) test statistics constructed based on the likelihood for the joint model of market and portfolio returns with regime-switching α and constant β (null) and the likelihood for the model with regime-switching α and β (alternative). The residual diagnostic tests are conducted for the residuals in the portfolio return equation of the joint model. The ARCH-LM statistics are constructed using R^2 from an auxiliary regression of squared standardized residuals on their lag and have a $\chi^2(1)$ asymptotic distribution under the null of no ARCH effects. The Jarque and Bera (1980) (JB) test statistics of Normality of residuals have a $\chi^2(2)$ asymptotic distribution under the null of Normality. HML denotes a "High minus Low" portfolio.

The residual diagnostics, also reported in Table 3.3, suggest that, after accounting for time variation in beta and a regime-switching variance, there are no remaining significant ARCH effects in the portfolio residuals and, for the most of the B/M portfolios, conditional Normality cannot be rejected based on the Jarque and Bera (1980) test. These results are consistent with Hamilton and Susmel (1994). They find that most of the ARCH effects in weekly stock returns die out at the monthly horizon and the remaining volatility changes that persist over longer period of time can be captured by a Markov-switching process. For comparison, the residuals for the unconditional CAPM regression model exhibit strong ARCH effects and Normality is strongly rejected.

Table 3.4: **Estimates for the regime-switching model of market and portfolio returns for book-to-market portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
	Panel A: Regime-switching alphas										
$\alpha_{i,0}$	-0.13	**-0.16**	-0.12	0.00	-0.02	0.15	**0.29**	0.22	**0.25**	0.29	0.41
std.error	(0.10)	(0.07)	(0.06)	(0.05)	(0.04)	(0.10)	(0.10)	(0.15)	(0.12)	(0.16)	(0.25)
$\alpha_{i,1}$	-0.13	0.16	**0.33**	0.05	0.00	0.09	-0.12	0.28	0.25	0.24	0.42
std.error	(0.19)	(0.33)	(0.13)	(0.13)	(0.05)	(0.15)	(0.19)	(0.39)	(0.16)	(0.23)	(0.36)
	Panel B: Regime-switching betas										
$\beta_{i,0}$	1.00	1.10	1.09	1.04	1.04	0.99	0.90	0.93	1.13	1.21	0.24
std.error	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.07)
$\beta_{i,1}$	1.16	1.01	0.97	1.02	0.89	0.91	0.93	0.81	0.84	0.75	-0.38
std.error	(0.03)	(0.07)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.09)	(0.04)	(0.05)	(0.07)
	Panel C: Other parameters										
$\mu_{m,0}$	0.15	0.27	0.29	0.24	0.21	0.30	0.29	0.27	0.32	0.30	0.23
std.error	(0.23)	(0.22)	(0.22)	(0.22)	(0.34)	(0.20)	(0.21)	(0.21)	(0.19)	(0.19)	(0.20)
$\mu_{m,1}$	0.94	0.96	1.00	0.75	0.69	0.73	0.80	0.86	0.42	0.73	0.80
std.error	(0.32)	(1.20)	(0.37)	(0.26)	(0.28)	(0.23)	(0.30)	(0.38)	(0.19)	(0.24)	(0.25)
δ	-7.25	-6.72	-7.42	-7.63	-7.94	-8.02	-7.58	-7.34	-6.75	-6.67	-6.54
std.error	(1.16)	(4.48)	(1.27)	(1.22)	(1.09)	(1.26)	(1.41)	(1.37)	(1.46)	(1.19)	(1.16)
$\sigma_{m,0}$	2.75	2.89	3.06	2.69	2.57	2.80	2.86	2.94	2.96	2.91	2.83
std.error	(0.22)	(0.50)	(0.21)	(0.30)	(0.18)	(0.27)	(0.20)	(0.41)	(0.18)	(0.18)	(0.18)
$\sigma_{m,1}$	5.18	5.54	5.61	5.19	5.02	5.27	5.37	5.45	5.28	5.41	5.33
std.error	(0.37)	(1.37)	(0.42)	(0.40)	(0.26)	(0.37)	(0.35)	(0.52)	(0.33)	(0.31)	(0.33)
$\sigma_{i,0}$	1.12	1.14	1.18	1.24	1.41	1.35	1.40	1.52	1.79	2.32	3.41
std.error	(0.10)	(0.05)	(0.04)	(0.06)	(0.06)	(0.07)	(0.08)	(0.10)	(0.10)	(0.10)	(0.16)
$\sigma_{i,1}$	2.10	2.43	3.22	3.68	3.80	3.16	3.87	3.72	4.37	5.15	6.24
std.error	(0.09)	(0.31)	(0.45)	(0.39)	(0.57)	(0.41)	(0.44)	(0.48)	(0.71)	(0.70)	(0.59)

Data are for value-weighted book-to-market decile portfolios for the sample period of July 1963 to December 2007. HML denotes a "High minus Low" portfolio. Panels A and B report alphas and betas from the regime-switching model of market and portfolio returns described by equations (3.2.14) and (3.2.15). Statistically significant alphas at the 5% level are in bold.

For each of the B/M portfolios, Table 3.4 reports estimates for the regime-switching model of market and portfolio returns described by equations (3.2.14) and (3.2.15). The estimates of the betas for the three portfolios with the highest B/M ratios (i.e., "value" portfolios) vary considerably across the two market volatility regimes; in particular, the betas for these portfolios in the low volatility regime are higher than in the high volatility regime. This result appears contrary to some theoretical models (e.g. Zhang (2005)) that suggest betas for value portfolios should be higher during bad times when marginal utility is high than in good times. However, our findings are similar to Ang and Kristensen (2010), who find using nonparametric estimates that betas for value portfolios are higher during bad times than in good times. Our findings are also consistent with Lakonishok, Shleifer, and Vishny (1994), who report that betas for value portfolios are higher (lower) than betas for growth portfolios (i.e., portfolios with low B/M ratios) in good times (bad times). They explain the B/M anomaly by "contrarian" investment behavior, whereby certain market participants overinvest in stocks that are "underpriced" and underinvest in stock that are "overpriced". By contrast, Petkova and Zhang (2005) find a positive relationship between betas for value portfolios and the market risk premium. However, as discussed by Ang and Kristensen (2010), this result is presumably driven by the specification of both beta and the market risk premium as linear functions of the same instrumental variables. Meanwhile, our finding that the dispersion of betas for B/M portfolios is considerably higher in the high volatility regime than in the low volatility regime is consistent with the theoretical findings in Gomes, Kogan, and Zhang (2003), who show that the dispersion of conditional betas should be countercyclical to the business cycle.

From Table 3.1, the unconditional CAPM can be rejected for the value portfolios, while we find that the regime-switching alphas for these portfolios are closer to zero in both regimes. The beta for the 1st decile portfolio also demonstrates statistically significant regime switching; however, its beta is lower in the low volatility regime than in the high volatility regime. Although the alphas for the 2nd, 7th, and 9th decile portfolios remain statistically significant at the 5% level in the low volatility regime, only the alpha for the 3rd decile portfolio is statistically significant in the high volatility regime. We note that, for the 2nd and 3rd decile portfolios, the unconditional CAPM regression model has economically and statistically insignificant alphas, while the regime-switching model has statistically significant alphas in one of the regimes. This result illustrates that, while CAPM may appear to hold unconditionally, it could still fail in some states of the economy.

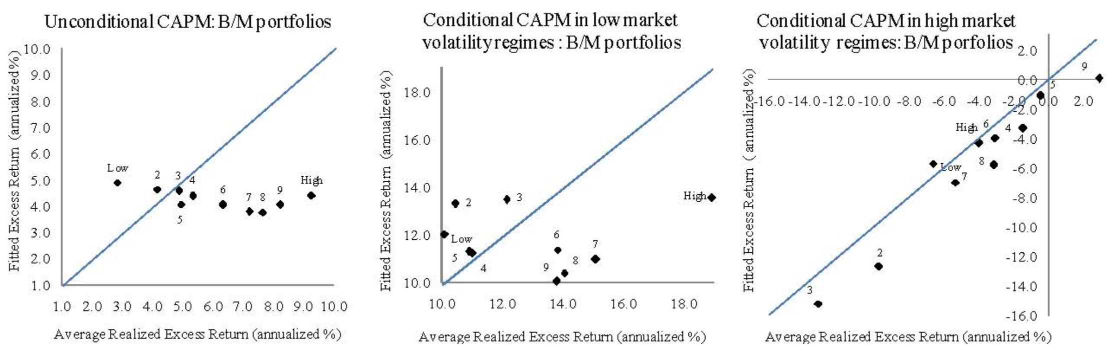
Table 3.5: **Long-run expected alphas for book-to-market portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
	Panel A: Alphas for the unconditional CAPM regression model										
α_i	-0.17	-0.04	0.02	0.08	0.07	**0.19**	**0.29**	**0.32**	**0.35**	**0.40**	**0.58**
std.error	(0.10)	(0.07)	(0.07)	(0.10)	(0.10)	(0.08)	(0.11)	(0.11)	(0.11)	(0.16)	(0.23)
	Panel B: Long-run expected alphas for the regime-switching model										
α_i	-0.13	-0.05	0.00	0.02	-0.01	0.13	0.14	**0.24**	**0.25**	**0.27**	**0.42**
std.error	(0.08)	(0.06)	(0.06)	(0.04)	(0.03)	(0.07)	(0.09)	(0.09)	(0.09)	(0.11)	(0.18)

Panel A repeats estimates of α from the unconditional CAPM regression model, also reported in Table 3.1, for comparison purposes. HML denotes a "High minus Low" portfolio. Panel B reports estimates of long-run expected alphas for the regime-switching model of market and portfolio returns described by equations (3.2.14) and (3.2.15). The long-run expected alpha for each portfolio is constructed as the weighted average of alphas in the two market volatility regimes, with weights equal to the steady-state probabilities of each regime. The standard errors for these expected alphas are computed using the Delta method.

Table 3.5 reports estimates of the long-run expected alphas for the B/M portfolios. These long-run alphas are computed as weighted-averages of alpha in the two market volatility regimes, with weights equal to the steady-state probabilities of the regimes. The estimates suggest that the values of most of the long-run alphas are closer to zero than the alphas for the unconditional CAPM regression model, although some of them are still statistically significant. To be clear, then, we do not claim that the conditional CAPM explains the entire behavior of excess returns for the B/M portfolios; point estimates of alphas for the last three portfolios are still large. Yet, we find evidence that portfolios with high B/M return premia demonstrate strong time variation of betas in the two market volatility regimes. We also find that accounting for time variation in the betas for the B/M portfolios over different states of the economy helps to explain some of the excess returns not captured by the unconditional CAPM. For example, the long-run expected alpha for the "High minus Low" portfolio strategy declined from 0.58 for the unconditional CAPM regression model to 0.42 for the regime-switching model.

Figure 3.2: **CAPM fitted excess returns versus average realized excess returns for book-to-market portfolios**



The returns are expressed as annualized percentages. The left scatter plot displays points with the average realized excess returns on the horizontal axis and the fitted excess returns from the unconditional CAPM on the vertical axis. The scatter plot in the middle (on right) displays points with the average realized excess returns conditional on smoothed probabilities of the high market volatility regime being lower (higher) than 0.5 on the horizontal axis and the fitted excess returns in the low (high) volatility regime from the regime-switching conditional CAPM on the vertical axis. The fitted excess return for each portfolio at low (high) market volatility regime is computed as an average of fitted excess returns calculated as a product of estimated betas in a previous period regime and realized market excess returns for observations with smoothed probabilities of high market volatility lower (higher) than 0.5. The straight lines on the graphs are 45 degree lines from the origins.

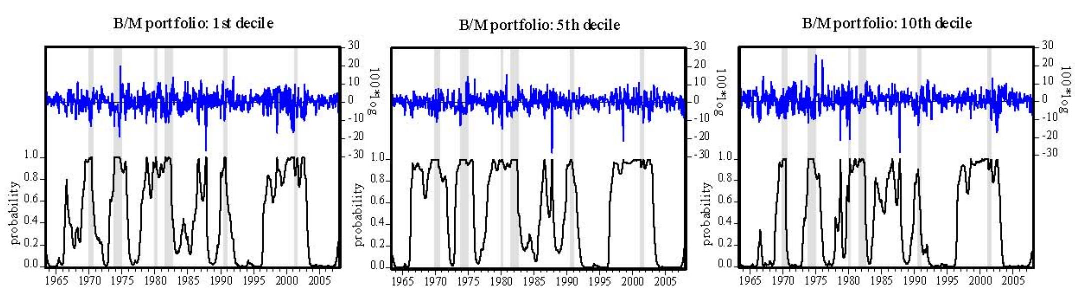
Figure 3.2 illustrates the relative performance of the unconditional CAPM and the regime-switching conditional CAPM for the B/M portfolios. If the CAPM provided a useful qualitative prediction for the behavior of returns, then one should observe points scattered along the 45 degree line, which corresponds to excess returns fitted by the CAPM being equal to average realized excess returns. As reported in many studies (e.g., Fama and French (1992), Jagannathan and Wang (1996)), the unconditional CAPM performs very poorly. The unconditional CAPM predicts flat excess returns for the B/M portfolios, while the average realized excess returns vary significantly across the portfolios.[9] The correlation coefficient between the excess returns predicted by the unconditional CAPM and average realized excess returns for the different portfolios has a value of -0.67, confirming the poor performance of the unconditional CAPM.

The performance of the regime-switching conditional CAPM for the B/M portfolios is different across the two regimes. Although there is not much visual improvement in the regime-switching conditional CAPM performance in the low volatility regime, there is an apparent improvement in the high volatility regime, where we can observe a fairly linear relationship between the CAPM-predicted excess returns and the average realized excess returns. The correlation coefficients between the excess returns fitted by the conditional CAPM and the average realized excess returns for different B/M portfolios have values of 0.05 and 0.97 in the "low" and "high" market volatility regimes, respectively. This result suggests that, in the high volatility regime at least, the regime-switching conditional CAPM provides a much better qualitative prediction for excess returns on the B/M portfolios than provided by the unconditional CAPM.

[9] The fitted excess returns for B/M portfolios in high market volatility regime are negative because they are computed based on realized market excess returns (see details in the note to Figure 3.2), which are negative. We use realized market excess returns to compute fitted excess returns because we compare them with actual realized excess returns of portfolios.

Although there is some variation in the evidence for regime-switching betas across different B/M portfolios, the estimates of parameters related to the market return when considered jointly with the different B/M portfolios are in the same range as those for a model of the market return with regimes identified using only market volatility and not jointly estimated with portfolio betas. The correlation coefficients between smoothed probabilities of the high volatility regime for the market-only model and the joint market/CAPM model for the different deciles of the B/M portfolios range from 0.81 to 1.00. This finding suggests that the regimes are mainly identified by changes in market volatility rather than by changes in the betas. Figure 3.3 displays portfolio excess returns and smoothed probabilities of the high market volatility regime for the 1st, 5th, and 10th B/M decile portfolios. Consistent with the regimes being identified by changes in market volatility, the smoothed probabilities appear quite similar to those in Figure 3.1 and to each other across the different portfolios.

Figure 3.3: **Monthly returns for selected book-to-market portfolios and smoothed probabilities of the high market volatility regime**



Returns are continuously compounded monthly value-weighted returns for B/M portfolios in excess of continuously compounded one-month Treasury bill yields for the sample period of July 1963 to December 2007. Shaded areas correspond to NBER recessions.

3.3.4 Regime-switching betas for momentum portfolios

For the analysis of the momentum portfolios, we proceed as before with the B/M portfolios. The LR tests for the null hypothesis of a constant beta, the results for which are reported in Table 3.6, support regime-switching betas at the 5% level for all but the 6th decile portfolio. Indeed, the tests are significant at the 1% level in the majority of cases. Thus, there is evidence for regime-switching betas for all of the momentum portfolios for which the unconditional CAPM can be rejected, as well as for some of the portfolios for which it cannot be rejected.

Table 3.6: **Likelihood Ratio tests for regime-switching betas and residual diagnostics for momentum portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
LR stat.	7.09	55.03	51.94	21.44	4.13	2.40	5.97	6.41	72.25	39.83	16.16
p-value	(0.01)	(0.00)	(0.00)	(0.00)	(0.04)	(0.12)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
	Residual diagnostic tests: portfolio return with constant beta and variance										
ARCH-LM	36.83	54.50	94.40	64.79	5.71	44.45	32.61	27.92	29.20	14.62	31.73
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
JB stat.	435.49	372.53	293.57	501.56	2433.05	701.79	1200.03	110.67	362.54	173.62	380.92
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Residual diagnostic tests: portfolio return with regime-switching beta and variance										
ARCH-LM	2.82	8.69	0.01	0.05	0.02	0.93	0.56	0.58	0.22	0.11	0.89
p-value	(0.09)	(0.00)	(0.90)	(0.82)	(0.88)	(0.33)	(0.46)	(0.45)	(0.64)	(0.74)	(0.35)
JB stat.	31.12	34.10	3.85	1.21	2.88	1.41	1.57	1.85	122.89	7.87	32.40
p-value	(0.00)	(0.00)	(0.15)	(0.55)	(0.24)	(0.49)	(0.46)	(0.40)	(0.00)	(0.02)	(0.00)

To test for a regime-switching β, we use likelihood ratio (LR) test statistics constructed based on the likelihood for the joint model of market and portfolio returns with regime-switching α and constant β (null) and the likelihood for the model with regime-switching α and β (alternative). The residual diagnostic tests are conducted for the residuals in the portfolio return equation of the joint model. The ARCH-LM statistics are constructed using R^2 from an auxiliary regression of squared standardized residuals on their lag and have a $\chi^2(1)$ asymptotic distribution under the null of no ARCH effects. The Jarque and Bera (1980) (JB) test statistics of Normality of residuals have a $\chi^2(2)$ asymptotic distribution under the null of Normality. HML denotes a "High minus Low" portfolio.

The residual diagnostic tests, also reported in Table 3.6, suggest that, for most of the momentum portfolios, there are no remaining ARCH effects in the portfolio residuals. The Normality of the residuals cannot be rejected for the majority of the portfolios based on the Jarque and Bera (1980) test and the test statistics for other portfolios declined considerably relative to those for the unconditional CAPM regression model. Again, the residuals for the unconditional CAPM regression model exhibit strong ARCH effects and their Normality is strongly rejected.

Table 3.7 reports estimates for the regime-switching model of the market and portfolio returns for each of the momentum portfolios. The estimates of the betas for most of the momentum portfolios vary considerably across the two volatility regimes. Betas for the four portfolios of stocks, which have relatively strong returns in the previous year (the "winner" portfolios), are higher in the low volatility regime than in the high volatility regime. By contrast, betas for the four portfolios of stocks, which have relatively weak returns in the previous year (the "loser" portfolios), are lower in the low volatility regime than in the high volatility regime. Jegadeesh and Titman (1993) find that the profitability of short-term "winners-minus-losers" strategy cannot be explained by systematic risk of the trading strategy or delayed stock price reactions to information about a common factor. They and Fama and French (1996) suggest a possible explanation for momentum anomaly is that investors underreact to short-term past information and overreact to long-term past information.

Table 3.8 reports estimates of the long-run expected alphas for the momentum portfolios. Unlike with the B/M portfolios, we do not observe tangible improvements compared to the unconditional CAPM. Therefore, we do not argue that the

Table 3.7: **Estimates for the regime-switching model of market and portfolio returns for momentum portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
	Panel A: Regime-switching alphas										
$\alpha_{i,0}$	**-0.36**	-0.15	-0.13	-0.12	-0.02	-0.13	-0.08	**0.26**	**0.19**	**0.59**	**1.08**
std.error	(0.18)	(0.10)	(0.08)	(0.08)	(0.05)	(0.07)	(0.08)	(0.09)	(0.09)	(0.14)	(0.34)
$\alpha_{i,1}$	**-1.96**	**-0.69**	0.05	0.14	**-1.04**	0.40	0.12	0.12	**0.32**	0.39	**2.50**
std.error	(0.24)	(0.22)	(0.28)	(0.20)	(0.41)	(0.31)	(0.13)	(0.13)	(0.13)	(0.26)	(0.44)
	Panel B: Regime-switching betas										
$\beta_{i,0}$	1.13	0.91	0.85	0.86	0.90	0.96	1.03	1.04	1.20	1.34	0.36
std.error	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.06)	(0.03)	(0.04)	(0.09)
$\beta_{i,1}$	1.35	1.49	1.23	1.13	1.03	1.03	0.91	0.86	0.82	0.82	-0.41
std.error	(0.05)	(0.05)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.06)	(0.14)
	Panel C: Other parameters										
$\mu_{m,0}$	0.22	0.38	0.28	0.19	0.13	0.22	0.27	0.41	0.32	0.36	0.23
std.error	(0.22)	(0.19)	(0.23)	(0.18)	(0.26)	(0.24)	(0.21)	(0.20)	(0.19)	(0.18)	(0.23)
$\mu_{m,1}$	0.68	0.91	1.16	1.27	3.16	1.53	0.89	0.66	0.66	1.00	0.84
std.error	(0.23)	(0.28)	(0.38)	(0.37)	(1.64)	(0.81)	(0.29)	(0.38)	(0.20)	(0.34)	(0.25)
δ	-7.79	-6.24	-6.13	-5.41	-6.47	-5.90	-6.50	-4.84	-5.26	-6.16	-6.42
std.error	(1.24)	(1.10)	(1.10)	(1.06)	(1.74)	(1.36)	(1.23)	(2.07)	(1.11)	(1.16)	(1.56)
$\sigma_{m,0}$	2.64	3.25	3.28	3.24	3.45	3.13	2.87	3.03	2.99	3.35	2.88
std.error	(0.17)	(0.15)	(0.15)	(0.16)	(0.16)	(0.18)	(0.27)	(0.18)	(0.16)	(0.15)	(0.30)
$\sigma_{m,1}$	5.04	5.82	5.91	5.97	7.01	6.14	5.48	5.86	5.64	6.01	5.30
std.error	(0.25)	(0.40)	(0.47)	(0.49)	(0.84)	(0.67)	(0.46)	(0.41)	(0.33)	(0.46)	(0.49)
$\sigma_{i,0}$	2.23	1.45	1.35	1.30	1.21	1.13	1.30	1.15	1.45	2.01	3.90
std.error	(0.12)	(0.10)	(0.07)	(0.06)	(0.05)	(0.04)	(0.06)	(0.07)	(0.06)	(0.16)	(0.30)
$\sigma_{i,1}$	6.46	3.94	4.61	4.38	3.64	3.66	4.82	2.56	4.40	4.60	9.59
std.error	(0.45)	(0.25)	(0.40)	(0.45)	(0.39)	(0.34)	(0.68)	(0.23)	(0.71)	(0.45)	(0.98)

Data are for value-weighted momentum decile portfolios for the sample period of July 1963 to December 2007. HML denotes a "High minus Low" portfolio. Panels A and B report alphas and betas from the regime-switching model of market and portfolio returns described by equations (3.2.14) and (3.2.15). Statistically significant alphas at the 5% level are in bold.

regime-switching conditional CAPM explains the failure of the unconditional CAPM for the momentum portfolios. However, as shown in Figure 3.4, allowing for changes in beta still helps the CAPM in terms of its qualitative predictions. In particular, similar to Figure 3.2 for the B/M portfolios, Figure 3.4 illustrates the relative performance of the unconditional CAPM and the regime-switching conditional CAPM

Table 3.8: **Long-run expected alphas for momentum portfolios**

	Low	2	3	4	5	6	7	8	9	High	HML
	Panel A: Alphas for the unconditional CAPM regression model										
α_i	**-1.10**	**-0.35**	-0.12	-0.04	-0.11	-0.02	0.03	**0.24**	**0.27**	**0.53**	**1.64**
std.error	(0.18)	(0.14)	(0.11)	(0.11)	(0.09)	(0.06)	(0.07)	(0.08)	(0.09)	(0.14)	(0.26)
	Panel B: Long-run expected alphas for the regime-switching model										
α_i	**-1.17**	**-0.28**	-0.09	-0.06	-0.12	-0.02	-0.01	**0.22**	**0.23**	**0.55**	**1.64**
std.error	(0.22)	(0.10)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.08)	(0.07)	(0.12)	(0.26)

Panel A repeats estimates of α from the unconditional CAPM regression model, also reported in Table 3.1, for comparison purposes. HML denotes a "High minus Low" portfolio. Panel B reports estimates of long-run expected alphas for the regime-switching model of market and portfolio returns described by equations (3.2.14) and (3.2.15). The long-run expected alpha of each portfolio is constructed as the weighted average of alphas in the two market volatility regimes, with weights equal to the steady-state probabilities of each regime. The standard errors for these expected alphas are computed using the Delta method.

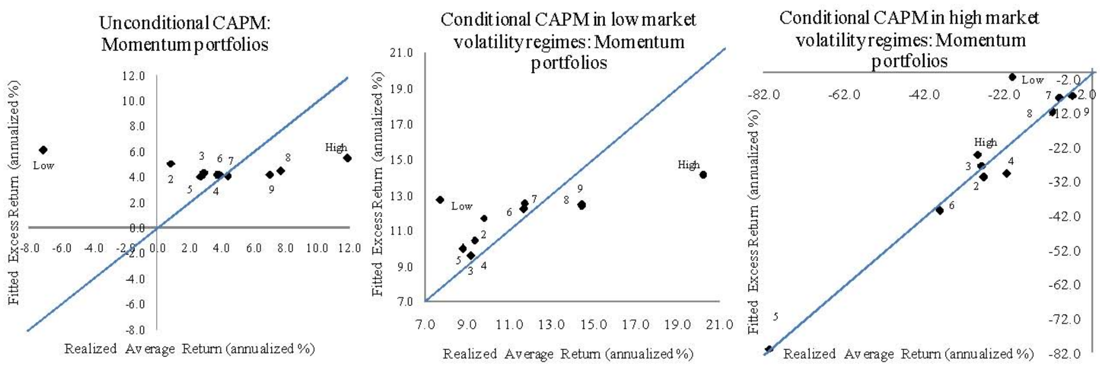
for the momentum portfolios. As with the B/M portfolios, the unconditional CAPM predicts nearly the same excess returns for the various momentum portfolios, while there is significant variation in average realized excess returns across the portfolios. When the two volatility regimes are considered separately, there appears to be a positive linear relation between the excess returns fitted by the conditional CAPM and the averaged realized excess returns.[10]

The correlation coefficient between the excess returns fitted by the unconditional CAPM and average realized excess returns for the different portfolios has a value of -0.39, confirming a similarly poor performance of the unconditional CAPM as was found for the B/M portfolios. Meanwhile, the correlation coefficients between the excess returns fitted by the conditional CAPM and the average realized excess returns for different momentum portfolios have values of 0.71 and 0.95 in the "low"

[10]The 1st decile portfolio appears to be an outlier from the linear relationship in both volatility regimes. However, the average returns for this portfolio are negative, while it has the highest volatility amongst all momentum portfolios. Because this portfolio comprises assets under financial stress and limited borrowing, we should probably not expect the CAPM to explain the returns for this decile.

and “high” market volatility regimes, respectively. Thus, the regime-switching conditional CAPM provides much better qualitative predictions for excess returns on the momentum portfolios than provided by the unconditional CAPM.

Figure 3.4: **CAPM fitted excess returns versus average realized excess returns for momentum portfolios**



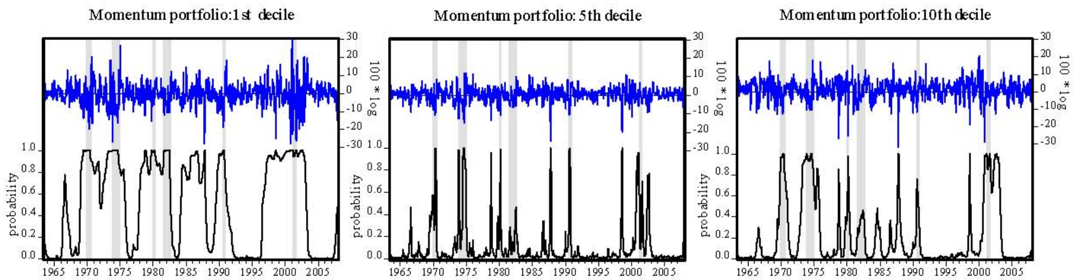
The returns are expressed as annualized percentages. The left scatter plot displays points with the average realized excess returns on the horizontal axis and the fitted excess returns from the unconditional CAPM on the vertical axis. The scatter plot in the middle (on right) displays points with the average realized excess returns conditional on smoothed probabilities of the high market volatility regime being lower (higher) than 0.5 on the horizontal axis and the fitted excess returns in the low (high) volatility regime from the regime-switching conditional CAPM on the vertical axis. The fitted excess return for each portfolio at low (high) market volatility regime is computed as an average of fitted excess returns calculated as a product of estimated betas in a previous period regime and realized market excess returns for observations with smoothed probabilities of high market volatility lower (higher) than 0.5. The straight lines on the graphs are 45 degree lines from the origins.

Given the lack of improvement in the long-run expected alphas, it might seem surprising that there is such an improvement in the qualitative predictions of the conditional CAPM. This result can be explained by the fact that the alphas, while apparently not equal to zero, are responsible for only relatively small portions of the overall portfolio returns. By contrast, variation in the market return explains sizable portions of the portfolio returns, especially in the high volatility regime.

In this sense, the conditional CAPM, while not strictly holding for all portfolios, appears to provide a reasonable approximation of asset pricing behavior.

The correlation coefficients between smoothed probabilities of the high volatility regime from the market-only model and the joint market/CAPM model for different deciles of the momentum portfolios range from 0.53 to 0.96. Evidently, in some cases, changes in betas are not so strongly related to changes in market volatility. In principle, to resolve this issue, we could consider a joint model that imposes the same market volatility regimes for all momentum portfolios. However, in practice, this is not feasible since it is important to allow for heteroskedasticity in idiosyncratic news for each portfolio, which would require incorporating 2^{11} (i.e., 2048) regime processes in the joint model for all momentum portfolios. In some cases, then, the joint market/CAPM model for each momentum portfolio identifies regimes as joint market volatility/beta regimes rather than as market volatility regimes. For the "loser" portfolios (1st, 2nd, 3rd, and 4th deciles), the joint market volatility/beta regimes are identified as low volatility/low beta and high volatility/high beta regimes. For the "winner" portfolios (7th, 8th, 9th, and 10th deciles), the regimes are identified as low volatility/high beta and high volatility/low beta. Figure 3.5 displays the excess portfolio returns and smoothed probabilities of the high market volatility regime for the 1st, 5th, and 10th momentum decile portfolios. Although changes in beta for the 1st momentum decile portfolio do not appear to significantly alter the identification of volatility regimes, as the smoothed probabilities are similar to those in Figure 3.1, changes in beta appear to strongly affect the identification of regimes for the 5th and 10th momentum decile portfolios. This lack of correspondence may also explain why the regime-switching conditional CAPM can still be rejected for a majority of the momentum portfolios.

Figure 3.5: **Monthly returns for selected momentum portfolios and smoothed probabilities of the high market volatility regime**



Returns are continuously compounded monthly value-weighted returns for momentum portfolios in excess of continuously compounded one-month Treasury bill yields for the sample period of July 1963 to December 2007. Shaded areas correspond to NBER recessions.

3.4 Conclusion

In this paper, we allowed for time variation in Capital Asset Pricing Model (CAPM) betas for book-to-market and momentum portfolios according to a two-state Markov-switching process driven by stock market volatility. Our empirical findings suggest strong time variation in betas across volatility regimes in most of cases for which the unconditional CAPM can be rejected. Somewhat supportive of the regime-switching conditional CAPM, we found that accounting for this time variation in betas helps explain some of the portfolio excess returns that are not captured by the unconditional CAPM. Thus, although the regime-switching conditional CAPM can still be rejected in many cases, it provides much better qualitative predictions about the relationship between risk and return compared to the unconditional CAPM, especially when market volatility is high.

Bibliography

Adrian, T. and Franzoni, F. (2009), "Learning about beta: time-varying factor loadings, expected returns, and the conditional CAPM," *Journal of Empirical Finance*, 16, 537–556.

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. and Chen, J. (2007), "CAPM over the long run: 1926-2001," *Journal of Empirical Finance*, 14, 1–40.

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

Ang, A. and Kristensen, D. (2010), "Testing conditional factor models," *Working Paper*.

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

Ang, A., Piazzesi, M., and Wei, M. (2006), "What does the yield curve tell us about GDP growth?" *Journal of Econometrics*, 131, 359–403.

Assoe, K. (1998), "Regime-switching in emerging stock market returns," *Multinational Finance Journal*, 2, 101–132.

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.

Breen, W., Glosten, L. R., and Jagannathan, R. (1989), "Economic significance of predictable variations in stock index returns," *Journal of Finance*, 44(5), 1177–1189.

Campbell, J. (1987), "Stock returns and the term structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. and Hentschell, L. (1992), "No news is good news: an asymmetric model of changing volatility in stock returns," *Journal of Financial Economics*, 31, 281–318.

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

Chauvet, M. and Senyuz, Z. (2009), "A Joint Dynamic Bi-Factor Model of the Yield Curve and the Economy as a Predictor of Business Cycles," *MPRA Paper*, 15076.

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.

Chu, C.-S. J., Santoni, G., and Liu, T. (1996), "Stock Market Volatility and Regime Shifts in Returns," *Information Sciences*, 94, 179–190.

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.

Diebold, F. X. and Li, C. L. (2006), "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 130, 337–364.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics*, 131, 309–338.

Dotsey, M. (1998), “The predictive content of the interest rate term spread for future economic growth,” *Federal Reserve Richmond Economic Quarterly*, 84(3), 31–51.

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

Estrella, A. and Hardouvelis, G. A. (1991), “The term structure a predictor of real economic activity,” *Journal of Finance*, 46(2), 555–576.

Estrella, A. and Mishkin, F. S. (1996), “The yield curve as a predictor of U.S. recessions,” *Current Economic Issues and Economics and Finance, Federal Reserve Bank of New York*, 2(7), 1–6.

— (1997), “The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank,” *European Economic Review*, 41, 1375–1401.

— (1998), “Predicting US recessions: Financial variables as leading indicators,” *Review Of Economics And Statistics*, 80, 45–61.

Estrella, A., Rodrigues, A. P., and Schich, S. (2003), “How stable is the predictive power of the yield curve? Evidence from Germany and the United States,” *Review Of Economics And Statistics*, 85, 629–644.

Fama, E. and French, K. (1992), “The cross-section of expected stock returns,” *Journal of Finance*, 47(2), 427–465.

— (1993), “Common risk factor in the returns on stock and bonds,” *Journal of Financial Economics*, 33, 3–56.

— (1996), “Multifactor explanations of asset pricing anomalies,” *Journal of Finance*, 51, 55–84.

— (1997), "Industry cost of equity," *Journal of Financial Economics*, 43, 153–193.

— (2006), "The value premium and the CAPM," *Journal of Finance*, 61(5), 2163–2185.

Fama, E. and MacBeth, J. (1973), "Risk, return, and equilibrium: empirical test," *Journal of Political Economy*, 81, 607–636.

French, K., Schwert, G., and Stambaugh, R. (1987), "Expected stock returns and volatility," *Journal of Financial Economics*, 19, 3–29.

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

Giacomini, R. and White, H. (2006), "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48(5), 1779–1801.

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

Gomes, J., Kogan, L., and Zhang, L. (2003), "Equilibrium cross section of returns," *Journal of Political Economy*, 111, 693–732.

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

Guidolin, M. and Timmermann, A. (2008), "Size and value anomalies under regime shifts," *Journal of Financial Econometrics*, 6(1), 1–48.

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.

Hamilton, J. and Susmel, R. (1994), "Autoregressive conditional heteroskedasticity and changes in regime," *Journal of Econometrics*, 64, 307–333.

Harvey, C. R. (1988), "The real term structure and consumption growth," *Journal of Financial Economics*, 22, 305–333.

Haubrich, J. G. and Dombrosky, A. M. (1996), "Predicting real growth using the yield curve," *Federal Reserve Bank Cleveland Economic Review*, 32(1), 26–35.

Hess, M. (2003), "What drives Markov regime-switching behavior of stock markets? The Swiss case," *International Review of Financial Analysis*, 12, 527–543.

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.

Huang, H., Lee, T.-H., and Li, C. (2006), "Forecasting output and inflation: how to use information in the yield curve," *Unpublished manuscript*.

Huang, H.-C. (2000), "Tests of regimes-switching CAPM," *Applied Financial Economics*, 10, 573–578.

Jagannathan, R. and Wang, Z. (1996), "The conditional CAPM and the cross-section of expected returns," *Journal of Finance*, 51(1), 3–53.

Jarque, C. M. and Bera, A. K. (1980), "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, 6, 255–259.

Jegadeesh, N. and Titman, S. (1993), "Returns to buying winners and selling losers: implications for stock market efficiency," *Journal of Finance*, 48(1), 65–91.

Johansen, S. (1998), "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

Kim, C., Nelson, C., and Startz, R. (1998), "Testing for mean reversion in heteroskedastic data based on Gibbs-sampling-augmented randomization," *Journal of Empirical Finance*, 5, 131–154.

Kim, C. J., Morley, J. C., and Nelson, C. R. (2001), "Does an intertemporal tradeoff between risk and return explain mean reversion in stock prices?" *Journal of Empirical Finance*, 8, 403–426.

— (2004), "Is there a positive relationship between stock market volatility and the equity premium?" *Journal of Money, Credit, and Banking*, 36(3), 339–360.

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.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994), "Contrarian investment, extrapolation, and risk," *Journal of Finance*, 49(5), 1541–1578.

Lettau, M. and Ludvigson, S. (2001), "Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109(6), 1238–1287.

Lewellen, J. and Nagel, S. (2006), "The conditional CAPM does not explain asset-pricing anomalies," *Journal of Financial Economics*, 82, 289–314.

Lintner, J. (1965), “The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets,” *Review of Economics and Statistics*, 47, 13–37.

Mayfield, E. (2004), “Estimating the market risk premium,” *Journal of Financial Economics*, 73(3), 465–496.

Nelson, C. R. and Siegel, A. F. (1987), “Parsimonious modeling of yield curves,” *Journal of Business*, 60, 473–490.

Nelson, D. (1991), “Conditional heteroskedasticity in asset returns: a new approach,” *Econometrica*, 59, 347–370.

Newey, W. K. and West, K. D. (1987), “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55(3), 703–708.

Perez-Quiros and Timmermann, A. (2000), “Firm size and cyclical variations in stock returns,” *Journal of Finance*, 55, 1229–1262.

Petkova, R. and Zhang, L. (2005), “Is value riskier than growth?” *Journal of Financial Economics*, 78(1), 187–202.

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.

Schaller, H. and van Norden, S. (1997), “Regime switching in stock market returns,” *Applied Financial Economics*, 7, 177–191.

Schwert, G. (1989), “Business cycles, financial crises, and stock volatility,” *Carnegie-Rochester Conference series on public policy*, 31, 83–126.

Sharpe, W. (1964), "Capital asset prices: a theory of market equilibrium under conditions of risk," *Journal of Finance*, 19, 425–442.

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.

Stock, J. H. and Watson, M. W. (2003), "Forecasting output and inflation: the role of asset prices," *Journal of Economic Literature*, 41, 788–829.

Svensson, L. (1994), "Estimating and interpreting forward rates: Sweden 1992-1994," *National Bureau of Economic Research Working Paper Series*, 4871.

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

Turner, C., Startz, R., and Nelson, C. (1989), "A Markov model of heteroskedasticity, risk, and learning in stock market," *Journal of Financial Economics*, 25, 3–22.

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

West, K. D. (1996), "Asymptotic inference about predicitve ability," *Econometrica*, 64(5), 1067–1084.

Wheelock, D. and Wohar, M. (2009), "Can the term spread predict output growth and recessions? A survey of the literature," *Federal Reserve Bank of St. Louis Review*, 91(5, Part 1), 419–440.

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

Zhang, L. (2005), "The value premium," *Journal of Finance*, 60, 67–103.