2

The Term Structure of Interest Rates in a Hidden Markov Setting

Robert J. Elliott[1] and Craig A. Wilson[2]

[1] Haskayne School of Business
University of Calgary
Calgary, Alberta, Canada
relliott@ucalgary.ca

[2] College of Commerce
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
cwilson@commerce.usask.ca

Summary. We describe an interest rate model in which randomness in the short-term interest rate is partially due to a Markov chain. We model randomness through the volatility and mean-reverting level as well as through the interest rate directly. The short- term interest rate is modeled in a risk-neutral setting as a continuous process in continuous time. This allows the valuation of interest rate derivatives using the martingale approach. In particular, a solution is found for the value of a zero-coupon bond. This leads to a non-linear regression model for the yield to maturity, which is used to filter the state of the unobservable Markov chain.

Key words: Interest rate modeling, term structure, filtering, Markov chain

2.1 Introduction

Current models of the short-term interest rate often involve treating the short rate as a diffusion or jump diffusion process in which the drift term involves exponential decay toward some value. The basic models of this type are Vasiček [10] and Cox, Ingersoll and Ross [2], where the distinction between these two interest rate models rests with the diffusion term. The drift term, (of both models), tends to cause the short rate process to decay exponentially towards a *constant* level. This feature is responsible for the mean-reverting property exhibited by these processes.

An extension to these models has come in the form of allowing the drift to incorporate exponential decay toward a *manifold*, rather than a constant. This is known as the Hull and White [7] model, and it allows the short rate process the tendency to follow the initial term structure of interest rates. This is an important extension, because with a judicious choice of the manifold, the initial term structure predicted by the model can exactly match the existing term structure, and because of this feature, models of this class are called no arbitrage models. In general, this cannot be done with a constant mean-reverting level, and such models are often called equilibrium models, since they generate stationary interest rate processes. Although there are many other extensions to the basic models—incorporating stochastic volatility, non-linear drift (so decay is no longer exponential), and jumps, for example—the Hull-White extension is the most applicable to the bond pricing component of our study.

The Hull-White model has many advantages: it possesses a closed-form solution for the price of zero-coupon bonds, as well as for call options on such bonds, and it can also be calibrated to fit the initial yield curve exactly. However, one of the disadvantages of the model is that, because there is only one factor of randomness, it only allows parallel shifts in the yield curve through time. Bonds of all maturities are necessarily perfectly correlated with each other. This approach cannot explain the common phenomenon of yield curve twists. This motivates the need to incorporate an additional factor of randomness into the basic model.

The Hull-White model is described under the risk-neutral probability by the stochastic differential equation

$$dr_t = a(t)\{\bar{r}(t) - r_t\}\, dt + \sigma(t) r_t^{\rho} dw_t,$$

where r_t represents the short-term, continuously compounded interest rate, and $\{w_t\}$ is a Brownian motion under the risk-neutral probability. The parameter ρ takes one of the two values 0 or 1/2, depending on whether it extends the Vasiček or Cox-Ingersoll-Ross model. The parameter functions $a(t)$, $\bar{r}(t)$, and $\sigma(t)$ extend the basic models, in which these parameters are just constants. The randomness in this model comes from the Brownian motion, and for the extended Vasiček model when $\rho = 0$, it can be interpreted as adding white noise to the short rate. For the extended Cox-Ingersoll-Ross model the noise is multiplicative, but it is still applied directly to the short rate process.

The main problem with this model is in the way it handles the cyclical nature of interest rates. A time series of interest rates tends to appear cyclical because the supply and demand for money is closely related to income growth, which fluctuates with the business cycle. This has implications for real (adjusted for inflation) interest rates. For example, at a business cycle peak short-term rates should be rising and at a trough rates should be falling. This also has implications for the slope of the term structure—it should be steeper at a peak and flatter at a trough. Roma and Torous [9] find that this property of real interest rates cannot be explained by a simple additive noise type model, such as Vasiček. The Hull-White extension can provide a correction for this problem to a degree, but since the parameter functions are deterministic, it implies that the business cycle effects are known with certainty, which does not allow for the possible variation in length and intensity from what is expected. In addition, when the central bank targets a constant rate of inflation, this fluctuation is transferred to nominal interest rates, so the same characteristics could apply to them.

We approach this problem by modeling the mean-reverting level directly as a random process, and have the short rate chase the mean-reverting level in a linear drift type model. This is similar to the model proposed by Balduzzi, Das, and Foresi [1], except instead of a diffusion process, here the mean-reverting level is assumed to follow a finite-state, continuous-time Markov chain. The switching of the Markov chain to different levels produces a cyclical pattern in the short rate that is consistent with the above effect, and the randomness inherent in the Markov chain prevents the business cycle lengths and intensities from being completely predictable.

The remainder of this paper is organized as follows. Section 2.2 discusses the model, including details about the Markov chain, the short-term interest rate, and the term-structure model. Section 2.3 outlines how the model is implemented and Section 2.4 provides the results of implementing it and discusses some implications. Finally Section 2.5 concludes.

2.2 The Model

In this section we construct the model of the short-term interest rate. This model will be used to derive prices for bonds.

We begin by describing the probability space, denoted by $(\Omega, \mathcal{F}, P)$, that is used to model randomness in this framework. We assume that P is a risk-neutral probability measure, whose existence can be assured by an absence of arbitrage in the underlying economy. Furthermore, we assume that the σ-field over Ω, $\mathcal{F}$, is complete and large enough to support the increasing filtration of sub-σ-fields $\{\mathcal{F}_t\}$ associated with the Markov chain and Brownian motion described below.

2.2.1 The Markov chain

A stochastic process, $\{X_t\}$ satisfies the Markov property (with respect to probability P and filtration $\{\mathcal{F}_t\}$) if

$$P\{X_{s+t} \in B|\mathcal{F}_s\} = P\{X_{s+t} \in B|X_s\}$$

for all $s, t \geq 0$ and all Borel sets, B. If such a stochastic process takes values in a countable set, it is called a Markov chain.

For our purposes, we consider a Markov chain generated by a transition rate matrix $\mathbf{Q}$. Here $\mathbf{Q}$ is an $N \times N$ conservative Q-matrix with non-negative off-diagonal entries and rows that sum to zero. In general, $\mathbf{Q}$ could change with time, but for simplicity and without any direction about how it should change we assume that $\mathbf{Q}$ is constant or homogeneous in time.

A transition function for a Markov chain relates the probability of changing from one state to another within a certain time, and the transition matrix is constructed so that each entry is a transition function

$$\mathbf{P}_{ij}(s,t) = P\{X_{s+t} = j | X_s = i\}.$$

The transition matrix $\mathbf{P}$ for a Markov chain can be generated by the transition rate matrix $\mathbf{Q}$ through the forward Kolmogorov equation

$$\frac{\partial \mathbf{P}(s,t)}{\partial t} = \mathbf{P}(s,t)\mathbf{Q}(s+t).$$

Since $\mathbf{Q}$ is homogeneous, the general solution to the forward Kolmogorov equation is $\mathbf{P}(s,t) = \mathbf{C}(s)e^{\mathbf{Q}t}$ and since $\mathbf{P}(s,0) = \mathbf{I}$, the identity matrix, the constant must also be the identity matrix, $\mathbf{C}(s) = \mathbf{I}$. From this we can conclude that the transition functions are independent of the starting time s, the transition matrix is the matrix exponential of $\mathbf{Q}$

$$\mathbf{P}(t) = e^{\mathbf{Q}t},$$

and the forward Kolmogorov integral equation is

$$\mathbf{P}(t) = \mathbf{I} + \int_0^t \mathbf{P}(u)\mathbf{Q}\, du.$$

Without loss of generality, we assume that the Markov chain is right continuous and it takes values from the set of canonical unit vectors of R^N, $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$, where $\mathbf{e}_i$ is the vector with 1 in the i^{th} entry and 0 elsewhere. To make this clear we will denote the Markov chain by $\{\mathbf{x}_t\}$. In this case we have $E[\mathbf{x}_t] = \mathbf{P}(t)^\mathsf{T}\mathbf{x}_0$, which is the probability distribution for the Markov chain and where $^\top$ denotes the transpose of a vector. More generally we have

$$E[\mathbf{x}_{s+t}|\mathbf{x}_s] = \mathbf{P}(t)^\mathsf{T}\mathbf{x}_s.$$

Putting this together with the forward Kolmogorov equation gives

$$E[\mathbf{x}_{s+t}|\mathbf{x}_s] = \mathbf{x}_s + \int_0^t \mathbf{Q}^\mathsf{T}\mathbf{P}(u)^\mathsf{T}\mathbf{x}_s\, du.$$

It follows directly from this that the stochastic process

$$\mathbf{m}_t = \mathbf{x}_t - \mathbf{x}_0 - \int_0^t \mathbf{Q}^\mathsf{T}\mathbf{x}_u\, du$$

is a square-integrable, right-continuous, zero-mean martingale. Therefore, $\{\mathbf{x}_t\}$ is a semi-martingale

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{Q}^\top \mathbf{x}_u \, du + \mathbf{m}_t.$$

(This derivation is adapted from Elliott [4].)

There are a number of other benefits that arise from associating states of the Markov chain with unit vectors. First note that the inner product of the Markov chain at any time is always equal to 1

$$\mathbf{x}_t^\top \mathbf{x}_t = 1,$$

the inner product between $\mathbf{1}$, the vector with 1 in each entry, and the Markov chain is also always 1

$$\mathbf{1}^\top \mathbf{x}_t = 1,$$

and the outer product of the Markov chain is the diagonal matrix of the Markov chain

$$\mathbf{x}_t \mathbf{x}_t^\top = \mathrm{diag}[\mathbf{x}_t].$$

Furthermore, any real-valued function of the Markov chain has a linear representation

$$f(\mathbf{x_t}) = \mathbf{f}^\top \mathbf{x}_t$$

where $\mathbf{f}_i = f(\mathbf{e}_i)$ and any vector-valued function of the Markov chain also has a linear representation

$$\mathbf{f}(\mathbf{x_t}) = \mathbf{F}^\top \mathbf{x}_t$$

where $\mathbf{F}_{ij} = \mathbf{f}_j(\mathbf{e}_i)$. Finally, notice that iterated multiples of the Markov chain have the following idempotency property

$$(\mathbf{f}^\top \mathbf{x}_t)\mathbf{x}_t = \mathrm{diag}[\mathbf{f}]\mathbf{x}_t.$$

The linear representations are also useful for describing the dynamics of certain stochastic processes. Consider the stochastic process $\{\mathbf{f}_t\}$ where $\mathbf{f}_t = \mathbf{F}_t^\top \mathbf{x}_t$, and $\mathbf{F}_t$ is continuous and adapted to $\{\mathcal{F}_t\}$. Then applying Itô's integration by parts for general semi-martingales allows the semi-martingale decomposition

$$\mathbf{f}_t = \mathbf{f}_0 + \int_0^t \mathbf{F}_u^\top \mathbf{Q}^\top \mathbf{x}_u \, du + \int_0^t \{d\mathbf{F}_u^\top \mathbf{x}_u\} + \int_0^t \mathbf{F}_u^\top \, d\mathbf{m}_u.$$

For the special case where $\mathbf{F}_u^\top$ commutes with $\mathbf{Q}^\top$ and $d\mathbf{F}_u^\top = \mathbf{G}_u^\top \mathbf{F}_u^\top \, du$, the semi-martingale representation can be written as

$$\mathbf{f}_t = \mathbf{f}_0 + \int_0^t \{\mathbf{Q} + \mathbf{G}_u\}^\top \mathbf{f}_u \, du + \int_0^t \mathbf{F}_u^\top \, d\mathbf{m}_u.$$

The following particular example arises in the context of bond pricing. Consider the processes $f_t = \exp(\int_0^t \mathbf{g}_u^\top \mathbf{x}_u \, du)$, where $\{\mathbf{g}_t\}$ is adapted, and $\mathbf{F}_t = f_t \mathbf{I}$. Clearly $\{\mathbf{F}_t\}$ is continuous, adapted, and it commutes with any $N \times N$ matrix. Furthermore, $d\mathbf{F}_t = \mathbf{F}_t \mathbf{G}_t \, dt$, where $\mathbf{G}_t = \mathbf{g}_t^\top \mathbf{x}_t \mathbf{I}$, so the dynamics of $\mathbf{f}_t = \mathbf{F}_t^\top \mathbf{x}_t$ have the above semi-martingale form. Moreover, $\mathbf{G}_t^\top \mathbf{f}_t = \mathrm{diag}[\mathbf{g}_t]\mathbf{f}_t$, so we have the semi-martingale representation

$$\mathbf{f}_t = \mathbf{f}_0 + \int_0^t \{\mathbf{Q} + \mathrm{diag}[\mathbf{g}_u]\}^\top \mathbf{f}_u \, du + \int_0^t \mathbf{F}_u^\top \, d\mathbf{m}_u.$$

If $\mathbf{g}_t$ and $\mathbf{f}_t$ are independent, (for example if $\mathbf{g}_t$ is deterministic), then we can find $E[\mathbf{f}_t]$ by solving a homogeneous linear system of ordinary differential equations. Since we can equivalently write $\mathbf{f}_t = f_t \mathbf{x}_t$, we have $f_t = \mathbf{1}^\top \mathbf{f}_t$, and therefore $E[f_t] = \mathbf{1}^\top E[\mathbf{f}_t]$.

2.2.2 The short-term interest rate

We now consider the model for the short-term interest rate. The short rate dynamics are defined through a stochastic differential equation, so a priori we require that a Brownian motion denoted $\{w_t\}$ exists for our probability space and filtration. In fact, since Brownian motion is a martingale, it is straight forward to show by taking $\mathbf{F}_t = w_t \mathbf{I}$ above that if it exists, it must be uncorrelated with the Markov chain. However, we require that the Markov chain and Brownian motion be independent, so we will assume that this stronger condition is satisfied. In this case, by defining $\{\mathcal{F}_t^x\}$ to be the filtration generated by the Markov chain, $\{w_t\}$ is still a Brownian motion with respect to the larger filtration $\{\mathcal{F}_t \vee \mathcal{F}_T^x\}$ for fixed T.

Following Naik and Lee [8], we model the short rate dynamics denoted $\{r_t\}$ using the equation

$$dr_t = a(\bar{r}_t - r_t) \, dt + \sigma_t \, dw_t. \tag{2.1}$$

This model suggests that the short rate is expected to decay exponentially toward the level $\bar{r}_t$ at the rate a, but it is subjected to additive noise modulated by the volatility σ_t. The level and volatility parameters are permitted to switch from time to time according to the state of the Markov chain, so we have

$$\bar{r}_t = \bar{\mathbf{r}}^\top \mathbf{x}_t \quad \text{and} \quad \sigma_t = \sigma^\top \mathbf{x}_t.$$

For simplicity and estimation purposes, we take the parameters to be constant, but the analysis follows identically if these are functions of time. This specification has two benefits over the basic models of Vasiček [10] and Hull and White [7]. It allows a better fit to the term structure and it has the potential to resolve the difficulty with accurately estimating the mean r eversion rate a.

The solution to the SDE in (2.1) is

$$r_t = \frac{1}{A_t}\left\{r_0 + \int_0^t A_u a\bar{r}_u\,du + \int_0^t A_u\sigma_u\,dw_u\right\} \tag{2.2}$$

where

$$A_t = e^{\int_0^t a\,du}.$$

The more general version of (2.2) is

$$r_t = \frac{1}{A_t}\left\{A_s r_s + \int_s^t A_u a\bar{r}_u\,du + \int_s^t A_u\sigma_u\,dw_u\right\} \quad \text{for } s \le t$$

From this, we can see that conditional on the information $\mathcal{F}_t^x$, r_t is normally distributed. Furthermore, by changing the order of integration we have

$$\int_0^t r_u\,du = r_0\int_0^t \frac{A_0}{A_s}\,ds + \int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)a\bar{r}_u\,du + \int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)\sigma_u\,dw_u.$$

Again, conditional on $\mathcal{F}_t^x$, $\int_0^t r_u\,du$ has a normal distribution with mean and variance

$$E\left[\int_0^t r_u\,du \,\middle|\, \mathcal{F}_t^x\right] = r_0\int_0^t \frac{A_0}{A_s}\,ds + \int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)a\bar{r}_u\,du$$
$$\operatorname{var}\left[\int_0^t r_u\,du \,\middle|\, \mathcal{F}_t^x\right] = \int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)^2\sigma_u^2\,du.$$

2.2.3 The zero-coupon bond value

Since we are working under the risk-neutral probability, the value of a zero-coupon bond maturing in t years is

$$B(t) = E\left[\exp\left(-\int_0^t r_u\,du\right)\right].$$

We determine this expectation in two stages, by first conditioning on the σ-field $\mathcal{F}_t^x$. Because the integral is conditionally normal, it is straightforward to get the conditional expectation

$$\begin{aligned}
E\left[\exp\left(-\int_0^t r_u\,du\right) \,\middle|\, \mathcal{F}_t^x\right] &= \exp\left\{\frac{1}{2}\int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)^2\sigma_u^2\,du\right.\\
&\qquad\left. - r_0\int_0^t \frac{A_0}{A_s}\,ds - \int_0^t\left(\int_u^t \frac{A_u}{A_s}\,ds\right)a\bar{r}_u\,du\right\}\\
= \exp\left(-r_0\int_0^t \frac{A_0}{A_s}\,ds\right)&\exp\left\{\int_0^t\left\{\frac{1}{2}\left(\int_u^t \frac{A_u}{A_s}\,ds\right)^2\sigma^2 - \left(\int_u^t \frac{A_u}{A_s}\,ds\right)a\bar{\mathbf{r}}\right\}^{\mathsf{T}}\mathbf{x}_u\,du\right\}
\end{aligned} \tag{2.3}$$

where the first term in equation (2.3) is deterministic. Therefore, we find the zero-coupon bond price by taking the expected value of the second term.

This is similar to the situation described at the end of Subsection 2.2.1. However, in this case the integrand is also a function of t. To deal with this, we fix a maturity time T and define a function

$$g_u = \frac{1}{2}\left(\int_u^T \frac{A_u}{A_s}\,ds\right)^2 \sigma^2 - \left(\int_u^T \frac{A_u}{A_s}\,ds\right) a\bar{\mathbf{r}}.$$

This quantity is deterministic and with a constant rate of mean reversion a we get

$$\int_u^T \frac{A_u}{A_s}\,ds = \frac{1-e^{-a(T-u)}}{a}.$$

Carrying on with the previous notation f_t and $\mathbf{f}_t$, we find the expectation $E[\mathbf{f}_t]$ by solving the homogeneous linear ordinary differential equation

$$\mathbf{y}'(t) = \left\{\mathbf{Q} + \mathrm{diag}[\mathbf{g}_t]\right\}^{\mathsf{T}} \mathbf{y}(t). \tag{2.4}$$

Calling the fundamental matrix in equation (2.4) $\mathbf{\Phi}(t)$ and noting that the initial value is $\mathbf{f}_0 = \mathbf{x}_0$, we write $E[\mathbf{f}_t] = \mathbf{\Phi}(t)\mathbf{x}_0$ and $E[f_t] = \mathbf{1}^{\mathsf{T}}\mathbf{\Phi}(t)\mathbf{x}_0$. Evaluating this at $t = T$ gives the value of a zero-coupon bond maturing at time T

$$B(T) = \exp\left(-r_0 \int_0^T \frac{A_0}{A_s}\,ds\right) \mathbf{1}^{\mathsf{T}}\mathbf{\Phi}(T)\mathbf{x}_0.$$

This is fine when the Markov chain is observable, but in our case we consider the Markov chain hidden. This means that the above bond value is still based on a conditional expectation given $\mathcal{F}_0^x$, and taking expected value requires replacing $\mathbf{x}_0$ with $E[\mathbf{x}_0] = \bar{\mathbf{x}}_0$. In other words, because the Markov chain is hidden, we must base our decisions on the probability distribution of its states. The continuously compounded yield to maturity of such a bond is

$$R(T) = \frac{r_0}{T}\int_0^T \frac{A_0}{A_s}\,ds - \frac{\ln\left(\mathbf{1}^{\mathsf{T}}\mathbf{\Phi}(T)\bar{\mathbf{x}}_0\right)}{T}. \tag{2.5}$$

2.3 Implementation

We implement this model using 7 years of monthly US term structure data from January 1999 to December 2005. The dataset was obtained from the Fama risk-free rate and Fama-Bliss discount structure files of the CRSP database. This data provides continously compounded yield to maturity on 1-month, 3-month, 6-month, and 1-year US T-bills, and it constructs continuously compounded yield to maturity on hypothetical zero-coupon US treasury bonds with maturities ranging annually from 2 to 5 years. This gives eight different maturities observed over 84 months for a total of 672 observations. A quick scan of the data revealed that the July 2003 observation of the six-month yield was erroneously recorded as zero, so we drop this observation leaving a total of 671 remaining observations.

The theoretical yield to maturity (2.5) derived in Subsection 2.2.2 provides a natural non-linear regression model to apply to this data. Writing

$$\alpha(T) = \frac{1}{T}\int_0^T \frac{A_0}{A_s}\,ds = \frac{1-e^{-aT}}{aT}$$

and $R_t(T)$ for the theoretical yield to maturity on a zero-coupon bond at time t that matures T years from then at time $t+T$ we get

$$R_t(T) = \alpha(T)r_t - \frac{\ln\left\{\mathbf{1}^\top \mathbf{\Phi}(T)\bar{\mathbf{x}}_t\right\}}{T}.$$

It is tempting to formulate the second term as a linear function of $\bar{\mathbf{x}}_t$; however, we cannot do this since $\bar{\mathbf{x}}_t$ is not a unit vector as $\mathbf{x}_t$ is. Denoting the observed data as $y_{t,T}$ where T represents the maturity and t represents the date, leads to the regression equation

$$y_{t,T} = R_t(T) + \epsilon_{t,T}.$$

We assume that the residuals $\{\epsilon_{t,T}\}$ are independent with mean 0 and variance η^2. This approach involves minimizing the sum of squared errors or residuals between the predicted theoretical yield and the actual yield observed in the data. Since the theoretical yield is not dynamic in the sense that it does not depend on lagged observations, the parameter estimators are weakly consistent provided the residuals have finite variance $\eta^2 < \infty$, which we will assume to be the case. For details on this see Davidson and MacKinnon [3].

There are three main difficulties we face in implementing the model using non-linear regression. First, in order to solve the differential equation we need the dimension of the Markov chain's state space. Expanding the state space can only reduce the sum of squared residuals because a model with a smaller state space can be considered a nested restriction of a more general model. The restriction could come in the form of requiring both mean-reverting level and volatility values to be equal in two particular states. Because of this, it is impossible to use non-linear regression to estimate the proper dimension of the Markov chain state space. To find an appropriate dimension, an F test could be used to determine when the improvement from increasing the dimension is no longer significant. Therefore, we need to fix the dimension of the Markov chain's state space before running the regression.

The next difficulty involves solving the differential equation. Since the coefficient matrix depends on time t, the fundamental solution matrix does not have a well-known closed form such as an exponential matrix. Therefore we solve the differential equation numerically. We do this by approximating the differential equation with the following difference equation

$$\mathbf{\Phi}\big((n+1)\Delta t\big) = \left\{\mathbf{I} + \big(\mathbf{Q} + \mathrm{diag}[\mathbf{g}_{n\Delta t}]\big)\Delta t\right\}^\top \mathbf{\Phi}(n\Delta t).$$

The solution uses $n = 1000$ intervals for each maturity, so $\Delta t = T/1000$. This provides a degree of accuracy of at least five significant digits for each element of the fundamental matrix $\mathbf{\Phi}(T)$ for all maturities up to $T = 5$ years.

The final problem we face deals with the initial values for the short-term interest rate and the Markov chain, r_t and $\bar{\mathbf{x}}_t$. Neither of these is provided by the data source CRSP. Since the Markov chain is unobservable, there is no hope of finding data elsewhere to use for its initial value at any date. Therefore we must estimate the initial probability distributions for the Markov chain at each date. This is a classic filtering problem and one way to approach it is to use a discrete version of the short rate dynamics

$$\Delta r_t = a(\bar{\mathbf{r}}^\top \mathbf{x}_t - r_t)\Delta t + \sigma^\top \mathbf{x}_t \Delta w_t$$

and monthly observations of the short-term interest rate for the desired period January 1991 to December 2005. An extension of the filtering techniques described in Elliott [5] can be applied to such a problem to get maximum likelihood estimates of the Markov chain state probabilities. Unfortunately, this also requires observation of the short-term interest rate, but more importantly it requires that the Markov chain transition probabilities be the same under the true measure, which is used by the filtering procedure and the risk-neutral measure, which is needed for the term-structure model.

A simpler filtering approach can be devised for our situation. We can simply treat the initial Markov chain state probabilities at each date as unknown parameters in our non-linear regression. Then the parameter estimates produce a filter for the state of the Markov chain and this automatically ensures that from the perspective of minimizing the sum of squared errors for the series of term structures the optimal filter is used. Unfortunately, since the Markov chain state probabilities do not enter the regression equation linearly, this optimal filter cannot be expressed analytically, so the values must be obtained numerically.

Turning our attention back to the initial short-term interest rate at each date, we again have two alternatives. We can use a proxy for the short rate such as the Federal Funds overnight rate, or we can filter values for the initial short rate at each date using our non-linear regression model and the term structure data. Naturally this latter approach uses up many more degrees of freedom by requiring 180 additional estimates. On the other hand, the filtering approach will choose these values optimally. Since the initial interest rate does enter the theoretical yield formula linearly, the optimal value is found to be

$$r_t^* = \frac{\sum_T \left\{ y_{t,T} + \frac{\ln\{\mathbf{1}^\top \mathbf{\Phi}(T)\bar{\mathbf{x}}_t\}}{T} \right\} \alpha(T)}{\sum_T \alpha(T)^2}.$$

These optimal values may differ substantially from the proxy values. One reason for this difference may have to do with the institutional features of the

US banking system that increase demand and thus price for treasury securities beyond an optimal competitive level. In any case, an F test can be used to determine whether the model is significantly hindered by considering the more parsimonious restricted model with the initial short rate proxied by the Federal Funds overnight rate. Next we look at the results from implementing the model.

2.4 Results

In this section we present the results from implementing the term structure model in several situations. Table 2.1 provides the main parameter estimates

$N=1$			$N=3$		
	Fed Fund	Free Est.		Fed Fund	Free Est.
a	0.238872	0.203953	a	0.628730	0.670303
$\bar{r}$	0.059517	0.068345	$\bar{\mathbf{r}}_1$	0.669713	0.355563
σ	0.000114	0.000114	$\bar{\mathbf{r}}_2$	–0.07240	–0.10679
std err	0.004691	0.003095	$\bar{\mathbf{r}}_3$	–0.71369	–0.12026
F_{fed}		11.31533	σ_1	0.024351	0.023138
$N=2$			σ_2	0.024041	0.023241
a	0.404457	0.575902	σ_3	0.000187	0.000187
$\bar{\mathbf{r}}_1$	0.275108	0.105218	$\mathbf{Q}_{12}$	0.040939	0.030791
$\bar{\mathbf{r}}_2$	–0.24453	–0.01730	$\mathbf{Q}_{13}$	5.723191	1.423016
σ_1	0.000185	0.000185	$\mathbf{Q}_{21}$	0.181682	0.229825
σ_2	0.000611	0.000611	$\mathbf{Q}_{23}$	0.287840	0.207573
$\mathbf{Q}_{12}$	0.307727	0.214106	$\mathbf{Q}_{31}$	8.428256	1.270089
$\mathbf{Q}_{21}$	0.680756	0.366236	$\mathbf{Q}_{32}$	0.042292	0.027953
std err	0.002164	0.001370	std err	0.001335	0.000674
F_{fed}		11.33442	F_{fed}		18.02795
F_{mc}	30.46332	29.60259	F_{mc}	12.31672	19.53060

Table 2.1. Parameter Estimates

a, $\bar{\mathbf{r}}$, and σ, the standard error, and F statistics for various restrictions. The standard error is calculated in the usual way as the square root of the sum of squared errors (SSE) divided by the difference between the number of observations and the number of parameters, std err $=$ SSE$/(n-k)$. The F statistic is also calculated in the usual way as

$$F = \frac{\text{SSE(restricted)} - \text{SSE(full)}}{\text{SSE(full)}} \times \frac{n-k}{r},$$

where r is the number of restricted parameters. Of course this statistic only has an F distribution with r and $n-k$ degrees of freedom in the linear case with linear restrictions and independent normally distributed residuals. However, the test is still useful for our situation since even with violations of linearity and normality the F distribution is approached asymptotically provided the parameter estimators are consistent, as they are for our model. In this case it is sometimes called a pseudo-F test. We consider a total of six scenarios: The Markov chain state space has 1, 2, or 3 dimensions and the short-term interest rate is proxied by the Federal Funds overnight rate or it is allowed to be freely estimated at each date by the regression.

The first thing to notice is that all of the F statistics in Table 2.1 are highly significant, having p-values of virtually zero in every case. This implies that all of the restrictions should be rejected, and the fullest model, which has a three dimensional state space and freely estimated initial short rate values at each date, is the best model even when the penalty for its unparsimoniousness is applied in the form of reduced degrees of freedom in the full model. A similar picture evolves when we look at each standard error. This statistic estimates the standard deviation of the residuals and also accounts for the degrees of freedom in the model. We see that the standard error is steadily reduced as more parameters enter the model.

We now turn our attention to the parameter estimates themselves. The rate of mean reversion, a, does behave as our intuition suggests it should. As we allow greater flexibility in the mean-reverting level, the rate at which this level is approached should increase. This is because with a fixed mean-reverting level, the rate of reversion will have to accommodate instances when the data diverges from the average. With a flexible mean-reverting level, these divergences can actually be considered instances of convergence to the more flexible level. From Table 2.1, we see that as we allow our Markov chain to have more states, the rate of mean reversion does increase.

On the other hand, estimates of the mean-reverting level, $\bar{r}$, are less economically intuitive. In the degenerate case, the level is quite reasonable at around 6.0 or 6.8%. However, when the Markov chain is allowed to switch between distinct states, the mean-reverting level tends to switch between unreasonably high and low values. When the short-term interest rate is restricted to be the Federal Funds rate, the mean-reverting level ranges between –24.4 and 27.5% for a two-state chain and –71.4 and 67.0% for a three-state chain. This is especially troubling when the high rate of mean reversion is also considered. We can see that restricting the initial short rate causes some of this problem, since when this constraint is relaxed, the mean-reverting levels become more reasonable. In particular, the two-state case switches between –1.7 and 10.5%, but the three-state case is still between –12.0 and 35.6%.

The volatility parameter, σ, turns out to be quite unimportant when the model is applied to the term structure data. For the degenerate case when the Markov chain has only one state, a 1% confidence interval is $0 \le \sigma \le 0.030851$. In fact, the sum of squared errors does not change perceptibly when the volatility is restricted to be $\sigma = 0$, and the p-value is virtually 1. A similar comment applies to the other cases, even in states one and two of the three state case, where the volatility is estimated to be somewhat larger, it is still not significantly different from zero. An explanation for this can be seen quite clearly in the degenerate case, which is Vasiček's [10] model. In this case the term structure equation can be written as

$$R_t(T) = \alpha(T)r_t + R_\infty\{1 - \alpha(T)\} + \frac{\sigma^2 T}{4a}\alpha(T)^2,$$

where $\alpha(T)$ is given previously and $R_\infty = \bar{r} - 0.5(\sigma/a)^2$, (see also Elliott and Kopp [6]). From this we can see that σ enters the formula through $(\sigma/a)^2$ and σ^2/a and with a small optimal volatility relative to the mean reversion rate, both of these quantities are small and likely to have little impact on the predicted yield to maturity. Although we do not have a closed-form solution for the more general cases, a similar reasoning may apply.

The entries of the transition rate matrix $\mathbf{Q}_{ij}$ can be interpreted as the rate at which the Markov chain is switching from state i to state j. It is perhaps easier to interpret these values if we convert them to monthly transition probabilities. We do this by calculating the exponential matrices $\mathbf{P} = e^{\mathbf{Q}t}$, with t taken to be one month (i.e. 1/12 of a year). The four cases with the number of states being two or three and the short-term interest rate being restricted or not are given as follows:

$$\begin{array}{ccc}
 & N=2 & N=3 \\
\text{Fed Fund} & \begin{bmatrix} 0.975384 & 0.024616 \\ 0.054456 & 0.945544 \end{bmatrix} & \begin{bmatrix} 0.717509 & 0.003359 & 0.279132 \\ 0.018137 & 0.961695 & 0.020168 \\ 0.411041 & 0.003423 & 0.585536 \end{bmatrix} \\
 & & \\
\text{Free Est.} & \begin{bmatrix} 0.982582 & 0.017418 \\ 0.029793 & 0.970207 \end{bmatrix} & \begin{bmatrix} 0.891514 & 0.002504 & 0.105982 \\ 0.018582 & 0.964250 & 0.017168 \\ 0.094595 & 0.002296 & 0.903109 \end{bmatrix}
\end{array}$$

The i, j elements of these matrices represent the transition probabilities of going from state i to state j next month. From this, we can see that there is a fairly low probability of switching for either of the $N = 2$ cases, and a low probability of switching out of state 2 for the $N = 3$ cases, but there is a fairly high probability of switching from state 1 to 3 and from 3 to 1, especially for the restricted case when these probabilities are 27.9 and 41.1% respectively. Another informative quantity we can find is the steady state or limiting probabilities for each state. These limiting probabilities can be interpreted as the proportion of time spent in each state in the limit as time becomes large. It is easy to show that these transition matrices are associated with irreducible and ergodic Markov chains, so the limiting probabilities correspond to stationary probabilities, which satisfy the equations $\mathbf{P}^{\mathsf{T}}\pi = \pi$ and $\mathbf{1}^{\mathsf{T}}\pi = 1$. These steady state probabilities are given as follows (transposed as row vectors):

$$\begin{array}{lcc} & N=2 & N=3 \\ \text{Fed Fund} & \begin{bmatrix}0.688688 & 0.311312\end{bmatrix} & \begin{bmatrix}0.546683 & 0.081187 & 0.372130\end{bmatrix} \\ \\ \text{Free Est.} & \begin{bmatrix}0.631069 & 0.368931\end{bmatrix} & \begin{bmatrix}0.442306 & 0.062766 & 0.494928\end{bmatrix} \end{array}$$

The last set of estimates for us to consider are the estimated Markov chain state probabilities and the estimated initial short-term interest rates. Rather than reporting these values in a tabular format, we present them graphically in Figure 2.1. We present the estimates for the three-state Markov chain. The top panel shows a graph of the probabilities for states 1, 2, and 3 through time, the middle panel shows a graph of the estimated short rate and the Federal Funds rate through time, and the lower panel shows a graph of the yields to maturity of the various zero-coupon bonds through time. The time scale of the graphs is matched to help draw inferences regarding how these three components are related to each other.

First we notice that the Markov chain state probabilities seem to behave as expected. Since according to Table 2.1, state 1 and state 3 are associated with a high and low mean-reverting level respectively, we expect the probability of being in state 1 to be higher and the probability of being in state 3 to be lower when rates are rising (and vice versa when rates are falling). This is generally observed when we compare the first and second panels. During 2002, we see the probability of state 2 rising and the probability of state 3 falling. Both of these states are associated with low mean-reverting levels, and therefore they should both correspond to falling interest rates as they do. However, state 3 has a much lower mean-reverting level than state 2, so state 2 should be associated with interest rates falling at a slower rate, which is consistent with what we observe between 2002 and 2004. As rates begin to rise again in 2004 and 2005, state 1 reasserts itself as the most likely state.

The second panel also allows us to compare the Federal Funds interest rate with the short rate filtered by the model. In general, they seem to agree quite well, although the filtered rate is usually slightly lower. Recall that this was also observed about the short maturity T-bills. There are a few substantial departures in 1999, 2000, and 2001, which likely account for most of the increase in sum of squared errors when we restrict the short-term interest rate. Indeed, when allowed to be freely estimated, the short rate is seen to follow the 1-month T-bill very closely.

The third panel graphs the data. We can see that this period is associated with a number of interesting phenomena. There are times when the interest rates are rising and the spread or slope of the yield curve is also increasing. The yield curve becomes very flat at the end of 2000 and beginning of 2001. The more common situation occurs when rates are falling but the spread is widening in 2001 and also the other common situation with rising rates and decreasing spread also occurs in 2004 and 2005. The most interesting feature

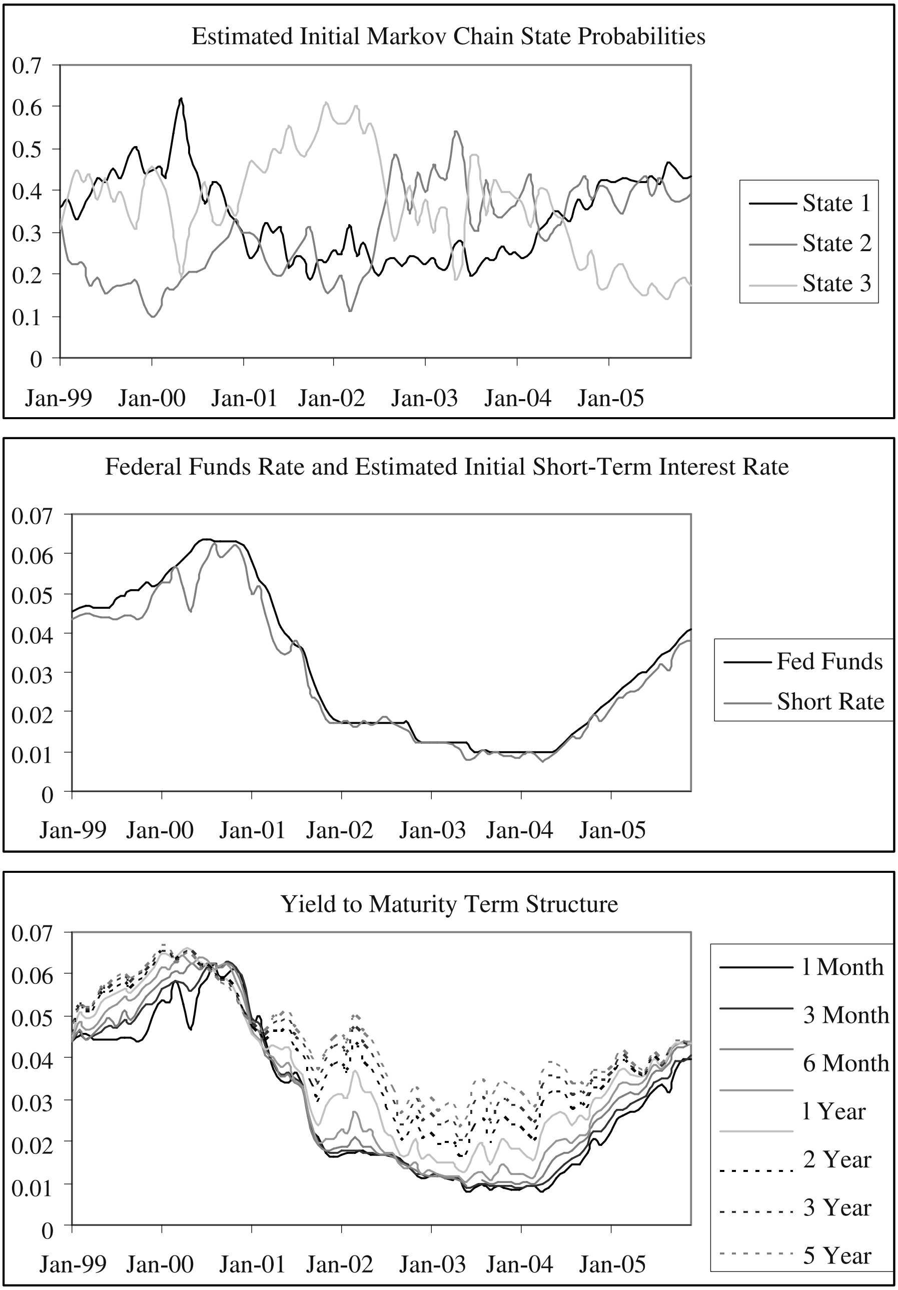

Fig. 2.1. State Probabilities and Initial Short-Term Interest Rate Estimates

occurs at the end of 2001 when the yield curve begins to twist with longer-term rates rising and short-term rates still falling. This is the time when a model with only one risk factor such as our degenerate case has the most difficulty in describing term structure dynamics, and this might explain why allowing the Markov chain as second factor of randomness to enter makes such a vast improvement in the models ability to explain this data.

2.5 Conclusion

We outline a methodology to incorporate a stochastic volatility and mean-reverting level into the short-term interest rate dynamics by using a Markov chain. We then show how to calculate the value of a zero-coupon bond. Using recent yield to maturity data, we estimate the model using a non-linear regression technique, and we find that the model makes a significant improvement in explaining the data over the basic model that excludes the Markov chain. Furthermore, we find that a three-state Markov chain makes a significant improvement over the two-state case. These improvements remain significant even when the initial short-rate is chosen optimally at each date, rather than being constrained to take values that proxy this rate such as the Federal Funds overnight rate. This suggests that models based on two-state regime switching may benefit from our more general N-state model construction.

References

1. Balduzzi, P., Das, S. R. and Foresi S. (1998). "The central tendency: A second factor in bond yields". *Review of Economics and Statistics*, 80(1): 62–72.
2. Cox, J. C., Ingersoll, J. E. and S. A. Ross (1985). "A theory of the term structure of interest rates". *Econometrica*, 53(2): 385–408.
3. Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford University Press.
4. Elliott, R. J. (1993). "New finite dimensional filters and smoothers for noisily observed Markov chains". *IEEE Transactions on Information Theory*, 39(1): 265–271.
5. Elliott, R. J. (1994). "Exact adaptive filters for Markov chains observed in Gaussian noise". *Automatica*, 30(9): 1399–1408.
6. Elliott, R. J. and P. E. Kopp. (1999). *Mathematics of Financial Markets.* Springer-Verlag.
7. Hull, J. and A. White (1990). "Pricing interest-rate derivative securities". *Review of Financial Studies*, 3(4): 573–592.
8. Naik, V. and M. H. Lee. (1997). "Yield curve dynamics with discrete shifts in economic regimes: Theory and estimation". Working Paper, University of British Columbia.
9. Roma, A. and W. Torous (1997). "The cyclical behavior of interest rates". *Journal of Finance*, 52(4): 1519–1542.
10. Vasiček, O. (1977). "An equilibrium characterization of the term structure". *Journal of Financial Economics*, 5(2): 177–188.

http://www.springer.com/978-0-387-71081-5

Hidden Markov Models in Finance
Mamon, R.S.; Elliott, R.J. (Eds.)
2007, XX, 186 p., Hardcover
ISBN: 978-0-387-71081-5