Chapter 2
Stochastic Volatility or Stochastic Central Tendency: Evidence from a Hidden Markov Model of the Short-Term Interest Rate

Craig A. Wilson and Robert J. Elliott

Abstract We develop a two-factor model for the short-term interest rate that incorporates additional randomness in both the drift and diffusion components. In particular, the model nests stochastic volatility and stochastic central tendency, and therefore provides a medium for testing the overall importance of both factors. The randomness in the drift and diffusion terms is governed by a hidden Markov chain. The likelihood function is determined through an iterative procedure and maximum likelihood estimates are obtained via numerical maximization. This process allows likelihood ratio testing of nested restrictions. These tests show that stochastic volatility is more important than stochastic central tendency for describing the short rate dynamics.

2.1 Introduction

The risk-free interest rate is one of the most vital inputs in financial and economic theory. There is still much debate about the relationship between rates for differing time horizons. Evidence for and against the expectations hypothesis waxes and wanes as additional complexities are incorporated into interest rate models and as more robust empirical analysis is applied to the various models. Most models of the short-term interest rate combine a mean-reverting drift component with a diffusion

C.A. Wilson (✉)
Edwards School of Business, University of Saskatchewan, Saskatoon, SK S7N 5A7, Canada
e-mail: wilson@edwards.usask.ca

R.J. Elliott
Haskayne School of Business, Scurfield Hall, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

School of Mathematical Sciences, North Terrace Campus, University of Adelaide, Adelaide, SA 5005, Australia
e-mail: relliott@ucalgary.ca

R.S. Mamon and R.J. Elliott (eds.), *Hidden Markov Models in Finance*, International Series in Operations Research & Management Science 209, DOI 10.1007/978-1-4899-7442-6_2,

component. The simplest models assume that the drift is a linear function of the short-term interest rate with constant parameters, and that the diffusion is governed by a constant volatility parameter. This model provides a mathematical framework for describing a central bank trying to control interest rates by pushing rates slowly toward a target rate with a force that is proportional to how far the current rate is from the target, but misjudging the effect of its policy independently of (with additive noise models) or proportionally to (with multiplicative noise models) the current rate in a consistent way (with constant volatility). Guthrie and Wright [11] develop an alternative model of central bank behavior that leads to similar observed interest rate behavior.

The main extensions to these simple models involve allowing for non-linearity in the drift function, allowing for randomness in the drift function, and allowing for randomness in the volatility function. Each of these extensions has been found to be empirically significant, when studied individually; however the purpose of this study is to determine which of these extensions is most important for explaining historical short-term interest rates, and we find that stochastic volatility is by far the more important feature.

Current theory about the dynamics of the short-term, default-free interest rate suggests two alternative methods of modeling: an equilibrium approach, and a no arbitrage approach. The later takes the current term structure as an input so as to force an exact fit to longer-term bond prices and other interest rate derivatives. Examples of this approach include the models of Ho and Lee [14], Hull and White [15], and Heath et al. [13]. On the other hand, equilibrium models such as those of Vasicek [22], and Cox et al. [7] generally do not predict values that exactly match current term structures. In this sense, such models are not arbitrage free. However, this shortcoming is often made up for by the model's applicability to future time periods, since they usually lead to a stationary sequence of interest rates. Also, because of the limitations of financial data, the term structures used as inputs for no arbitrage models are finite, so in practice, arbitrage free predictions can often be achieved by equilibrium type models with a sufficiently large number of parameters.

In the particular case of the equilibrium type model used by Chan et al. [5], the interest rate is supposed to follow a mean reverting process described by dynamics of the form

$$dr_t = \alpha(\bar{r} - r_t)dt + \sigma r_t^{\gamma} dW_t. \tag{2.1}$$

In unconstrained estimation by Chan et al. [5] it is found that the variance elasticity, γ, is approximately 1.5, (using GMM estimation on U.S. interest rates based on monthly observations between June 1964 and December 1989), which causes the previous SDE to have a non-stationary solution, (i.e. the variance increases without bound as t gets large), which is undesirable for estimation and testing purposes. The above interest rate model has two important features: the drift term is a linear function of the interest rate and the volatility term is deterministic.

Relaxing one or both of these properties could resolve the problem. For instance letting the drift term be non-linear so that it increased the mean-reverting force as the interest rate became large, could resolve the non-stationarity problem [1]. Or if the volatility were allowed to be stochastic, as in Longstaff and Schwartz [18], the need for randomness implicit in the r_t^γ term could be reduced, requiring a smaller elasticity parameter, γ. These issues are addressed by Sun [21], where he finds that stochastic volatility was significant and non-linear drift was not significant. This finding is consistent with Chapman and Pearson [6], who conclude that non-linear drift is not an essential property of the short-term interest rate.

Although Sun [21] describes a model that nests both stochastic volatility and non-linear drift, he is pitting a two-factor model up against a one-factor model when he tests his restrictions. This approach might bias his study toward finding that the stochastic volatility framework dominates the non-linear drift, since such a framework is able to explain some observed term structure phenomena such as yield curve twists that typically cannot be explained in a one-factor model. It turns out that this concern is unfounded, as we also find that stochastic volatility is most important, even when both full and restricted models have two factors of randomness.

Balduzzi et al. [3] develop a model in which the mean-reverting level, or what they call the central tendency of the short rate process provides a second factor of randomness. In this case the second factor enters through the drift, but they still model the drift as being linear in the short rate. If one compares stochastic central tendency against stochastic volatility, both restricted models have two factors of randomness, and they can be compared on a level playing field, which is the approach we take.

A natural model to nest these two phenomena is the three factor model of Balduzzi et al. [2]. However, this model requires the central tendency factor to be independent of the other factors, so testing restrictions of this model pits a two-factor model against a three-factor model, which may again bias toward rejecting the restrictions of constant central tendency or volatility. One solution to this problem would be to implement a special case of this model's extension by Dai and Singleton [8], where both central tendency and volatility are governed by the same Brownian motion.

Unfortunately, implementing this approach in the Chan et al. [5] framework is difficult because it requires relaxing the affine term structure model assumption. We examine estimation techniques on the Chan et al. [5] model where the central tendency level and volatility parameters themselves are prone to switch in accordance with the same Markov chain. In this way, even the full model with both stochastic drift and stochastic volatility has only two factors of randomness: the Markov chain and an independent Brownian motion. Even so, this framework allows an arbitrary correlation between drift and volatility. Such a feature was found to be important by Dai and Singleton [8].

This regime switching framework has been found helpful in explaining interest rate and term structure characteristics in a number of studies including Hamilton [12], Naik and Lee [19], Bansal and Zhou [4], Smith [20], and Kalimipalli and Susmel [16], to name a few. We assume that the state of the Markov chain cannot be observed directly, and must be estimated through filtering observations of the short-term interest rate. In this way we consider the interest rate to be governed as a hidden Markov model.

We consider the case of a discrete time autoregressive stochastic process. An extension of Hamilton's [12] algorithm provides an iterative construction of the likelihood function. Evaluating this likelihood function at the maximum likelihood estimates obtained for full and restricted models allows the testing of various restrictions using likelihood ratio tests. We do this for observations of Canadian and U.S. nominal interest rates. The null hypothesis of constant volatility can be strongly rejected, whereas that of constant mean reverting level or central tendency cannot.

The remainder of this paper is organized as follows. Section 2.2 discusses the general model, and Sects. 2.3 and 2.4 discus techniques to estimate and test this model in the maximum likelihood framework. Section 2.5 specializes the model to the case of short-term interest rates. Section 2.6 discusses the data, Sect. 2.7 analyzes the results, and Sect. 2.8 concludes.

2.2 The Model

The first process that we consider is a finite state (N-dimensional), discrete time, homogeneous Markov chain, $X = \{X_t; t \in \mathbf{N} = \{0,1,2,\ldots\}\}$, that takes values in the set of unit (column) vectors, $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$, which is the canonical basis of $\mathbf{R}^N$, (i.e. $X_t = (0,\ldots,0,1,0,\ldots,0)^\top$). Denote by $\mathscr{F}^X = \{\mathscr{F}_t^X\}$ the filtration generated by the Markov chain X, and $\mathbf{P}$ its transition matrix, where $\mathbf{P}_{ij} = \Pr\{X_{t+1} = \mathbf{e}_j | X_t = \mathbf{e}_i, \ldots, X_0 = \mathbf{e}_k\}$ is the probability of going from state i to state j. It follows that $E[X_{t+1}|\mathscr{F}_t^X] = \mathbf{P}^\top X_t$, where the conditional expectation gives the vector of conditional probabilities and the right hand side picks out the appropriate row of $\mathbf{P}$. (Note that the entries of $\mathbf{P}$ must be non-negative and that the rows must sum to 1.) This convenient notation motivates our choice of state space and stochastic matrix notation, which was done without loss of generality.

We presume that this Markov chain is hidden, (i.e. it is not directly observable), so that we do not have access to the information $\mathscr{F}^X$. However, we do observe a stochastic process $\{Y_t; t \in \mathbf{N}\}$, which has the form

$$Y_{t+1} = \propto(X_t) + \zeta(X_t)\varepsilon_{t+1}, \tag{2.2}$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d. standard normal random variables (although other distributions could be used), independent with the Markov chain, X. (For now, we consider $\{Y_t\}$ to be a general observation process, but later we will specialize it to observations of short-term interest rates.) It is clear that $\propto$ and ζ are Markov chains. Also notice that the function $\propto(X_t)$ has the representation $\propto^\top X_t$, where the vector $\propto$ has typical entry $\propto_i = \propto(\mathbf{e}_i)$, and similarly for ζ. In general, the drift and volatility terms $\propto$ and ζ could be functions of other independent and observable variables.

Denote by $\mathscr{F}^Y = \{\mathscr{F}_t^Y\}$ the filtration generated by the observed process, Y, $\mathscr{F}^\varepsilon = \{\mathscr{F}_t^\varepsilon\}$ the filtration generated by the noise, ε, and $\mathscr{G} = \{\mathscr{G}_t\} = \{\mathscr{F}_t^X \vee \mathscr{F}_t^Y\} = \{\mathscr{F}_t^X \vee \mathscr{F}_t^\varepsilon\}$ is the joint (or global) filtration. The filtering problem will therefore involve the optimal use of the available information. We wish to make inferences about $\mathscr{G}$-adapted processes by conditioning on the filtration $\mathscr{F}^Y$. This procedure gives a best, (in mean square error sense), estimate of the unobservable processes, based on information obtained from observing the process Y [10].

2.3 Maximum Likelihood Estimation

The model requires estimates for the transition probabilities, $\mathbf{P}_{ij}$ and the entries of the vectors $\propto$ and ζ. The class of maximum likelihood estimators, (MLE's), has several desirable properties such as consistency, efficiency, and robustness [17]. We therefore attempt to find MLE's for the various parameters. The problem is that MLE's can often be difficult to calculate directly or explicitly. We form the likelihood function iteratively and solve it numerically via the EM algorithm. The procedure we use is a modification of Elliott's [10] filter.

Since the Markov chain is unobservable, we have particular difficulty in estimating the probability matrix. This estimation can be done by using a change of probability technique. Based on the time series of observations and arbitrary starting parameter values, we filter information about the Markov chain's state, which is used to obtain optimal (in the sense of expectation maximization) parameter estimates. The EM algorithm continues by finding new filtered processes using the previous optimal parameter estimates and using the new processes to find new optimal parameter estimates. A fixed point in the parameter space corresponds to maximum likelihood parameter estimates.

The previous algorithm gives maximum likelihood estimates, but not the likelihood function. To perform likelihood ratio (LR) tests, we need the likelihood function evaluated at the optimal parameters for various restrictions. We use a modification of Hamilton's [12] algorithm to obtain the likelihood function and evaluate it at the values found by the EM algorithm. This approach involves manipulating conditional probability mass and density functions at each time and adding their logarithms to get the log-likelihood function. Unfortunately this function cannot be obtained in closed form, which is why we use the EM algorithm to maximize it.

2.4 The Likelihood Function

This section describes an algorithm similar to Hamilton's [12] algorithm. We have a hidden Markov chain $\{X_t\}$ and a sequence of observations $\{Y_t\}$, which are presumed to depend upon the previous state of the Markov chain, and on noise that is independent with the Markov chain. Notice that we are considering probability mass functions and probability density functions in this section and we denote them f and g respectively. (These functions can be thought of as Radon-Nikodym derivatives with respect to counting measure or Lebesgue measure.) Functions of more than one variable refer to joint probability mass or density functions. For ease of notation, the dependence on the parameters is suppressed. It is implied that all of the mass and density functions that follow depend on a common parameterization.

The goal is to obtain the current filter, which is the probability mass function $f(x_t;y_t,\dots,y_0) = \Pr\{X_t = x_t|Y_t = y_t,\dots,Y_0 = y_0\}$ when starting with the previous filter $f(x_{t-1};y_{t-1},\dots,y_0) = \Pr\{X_{t-1} = x_{t-1}|Y_{t-1} = y_{t-1},\dots,Y_0 = y_0\}$. This function provides a filter for the state of the Markov chain. We assume that the conditional density function $g(y_t|x_{t-1},y_{t-1},\dots,y_0)$ is known. In particular, for our general model we use the normal density

$$g(y_t|x_{t-1},y_{t-1},\dots,y_0) = \frac{1}{\sqrt{2\pi}\zeta^{\mathsf{T}}x_{t-1}}\exp\left\{-\frac{(y_t-\propto^{\mathsf{T}}x_{t-1})^2}{2(\zeta^{\mathsf{T}}x_{t-1})^2}\right\}; \qquad (2.3)$$

however, other densities could be used. A consequence of the algorithm is that the conditional density function $g(y_t|y_{t-1},\dots,y_0)$ is obtained, which can be used to construct the likelihood function $L(\theta;y) = g(y_T,\dots,y_1|y_0) = \prod_{t=1}^{T} g(y_t|y_{t-1},\dots,y_0)$. Hamilton advocates maximizing the log-likelihood function numerically to obtain maximum likelihood estimates for the parameters. Knowing the likelihood function explicitly allows likelihood ratio tests to be applied to test equality constraints on the parameters in a straight forward manner: With r distinct equality restrictions, the logarithm of the square of the ratio of the likelihood function evaluated at the unrestricted MLE to that evaluated at the restricted MLE has a central χ^2 distribution with r degrees of freedom (under certain regularity conditions see Lehmann [17] for example).

We outline the algorithm as follows: Assume we know $f(x_0|y_0)$, we have iterated through the algorithm to the tth observation to get $f(x_t|y_t,\dots,y_0)$, and $g(y_{t+1}|x_t,y_t,\dots,y_0)$ is given as in Eq. 2.3. Then

1. $g(y_{t+1},x_t|y_t,\dots,y_0) = g(y_{t+1}|x_t,y_t,\dots,y_0)f(x_t|y_t,\dots,y_0)$
2. $g(y_{t+1}|y_t,\dots,y_0) = \sum_i g(y_{t+1},\mathbf{e}_i|y_t,\dots,y_0)$
3. $f(x_t|y_{t+1},\dots,y_0) = \frac{g(y_{t+1},x_t|y_t,\dots,y_0)}{g(y_{t+1}|y_t,\dots,y_0)}$
4. $f(x_{t+1},x_t|y_{t+1},\dots,y_0) = f(x_{t+1}|x_t,y_{t+1},\dots,y_0)f(x_t|y_{t+1},\dots,y_0)$
5. $f(x_{t+1}|y_{t+1},\dots,y_0) = \sum_i f(x_{t+1},\mathbf{e}_i|y_{t+1},\dots,y_0)$

Each step follows from the definition of conditional probability or a straightforward application of Bayes' theorem. In Steps 2 and 5, the term $\mathbf{e}_i$ in the joint density or mass function refers to the case when $x_t = \mathbf{e}_i$. In Step 4, the conditional mass is $f(x_{t+1}|x_t,y_{t+1},\dots,y_0) = f(x_{t+1}|x_t)$, (by the Markov property and independence between the noise and the Markov chain), which is simply the transition probability. We obtain the likelihood function as the product of conditional density functions found in Step 2.

2.5 The Interest Rate Model

We assume that the interest rate follows a discrete version of the continuous time stochastic process defined by the SDE,

$$dr_t = \alpha_t(\bar{r}_t - r_t)\,dt + \sigma_t r_t^{\gamma}\,dW_t, \tag{2.4}$$

where $\alpha_t = \alpha$, $\bar{r}_t = \bar{r}(X_t)$, $\sigma_t = \sigma(X_t)$, and W is a standard Brownian motion independent with the continuous time Markov chain, X. The continuous time Markov chain is characterized by its transition rate matrix, $\mathbf{Q}$, which is related to the probability transition matrix through the forward and backward Kolmogorov equations. In particular, for a homogeneous Markov chain whose rate matrix doesn't depend on time, we have $\mathbf{P} = e^{\mathbf{Q}\Delta t}$, obtained by the matrix exponential. A well-behaved Markov chain has a rate matrix that is a so-called conservative Q-matrix, which means that $\mathbf{Q}$ has non-negative off-diagonal entries and its rows sum to zero, so the probability transition matrix does turn out to be a stochastic matrix with entries $\mathbf{P}_{ij}$ representing the probability of going from state i at time s to state j at time $s + \Delta t$.

If Δt is small, then an Euler approximation of the above SDE provides the following discrete representation of the interest rate:

$$\Delta r_{t+1} = \alpha\{\bar{r}(X_t) - r_t\}\Delta t + \sigma(X_t) r_t^{\gamma}\sqrt{\Delta t}\,\varepsilon_{t+1}, \tag{2.5}$$

where $\{X_t\}$ is a discrete time Markov chain with transition matrix $\mathbf{P} = e^{\mathbf{Q}\Delta t}$ and $\{\varepsilon_t\}$ are i.i.d. standard normal. Here α is the rate of mean reversion, $\bar{r}$ is the mean reverting level or central tendency, σ is the volatility of the interest rate process, and γ is the variance elasticity. We estimate the following equation

$$\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \varsigma(X_t) r_t^{\gamma}\varepsilon_{t+1}, \tag{2.6}$$

and then transform the coefficients to the more meaningful term structure coefficients.

We now turn our attention to what this model implies about the behavior of interest rates. First of all the short-term rates, following this model will be positively auto-correlated through time. The auto-correlation is $\rho = 1 - \alpha\Delta t$, which is less than 1 whenever the mean reversion rate, α, is positive, where α measures the rate at which r is expected to approach the mean reverting level, $\bar{r}$. If $\rho > 1$, then the process will drift away from the mean. If $\rho = 1$, then $\propto$ must equal zero and thus r follows a random walk.

Allowing the parameters to depend on a Markov chain means that the central tendency and interest rate volatility will change from time to time. When it does change, the interest rate will begin to converge toward the new central tendency level, when it changes back, the interest rate will turn around and begin to converge back. It seems intuitive that such a data generating process would describe a cyclical pattern, but with a random cycle length. In particular, such a data generating process would be able to create large cycles with a relatively small volatility parameter, as is typically seen in a series of interest rates.

Since short-term interest rates are essentially controlled by the central bank, it might seem unreasonable that they switch so violently. However, whatever the underlying variable that the bank is primarily controlling through its choice of interest rates, be it inflation, exchange rates, or unemployment, etc, it might not be unreasonable to model these, exogenously, as having the impact of a Markov chain switching the converging level. Furthermore, the continuous time rate process has continuous sample paths, which provides a certain smoothness in interest changes that is often desired by central banks. This intuition suggests that randomness in the central tendency or drift term of the short rate process will be the more natural. Surprisingly, it is randomness in the volatility that seems to be more important.

2.6 Data

For the Canadian data we use monthly and weekly observations of the short term interest rate implied by Government of Canada 3-month Treasury bills. The monthly rates were obtained from the Bank of Canada website, excerpted from *Selected Canadian and International Interest Rates Including Bond Yields and Interest Arbitrage*. The data set includes bills from March 1934, when the first public tender occurred, until December 2004. The rates quoted in this data set are measured in units of percent and quoted as a discount style of interest rate. To be consistent with our modeling, we convert these rates to unitless annual continuously compounded values. Monthly rates are based on the last Wednesday of the month. Data with weekly observations starts Wednesday January 3, 1962. The weekly Canadian data was taken from the CANSIM website (series V121778).

For the U.S. data, we also use monthly and weekly observations of the U.S. 3-month Treasury bill returns provided by the St. Louis Federal Reserve website. Monthly returns are provided from January 1934 to December 2004 and weekly returns are provided from January 1954; however, to be consistent with our Canadian data, we restrict attention to the post 1962 period. These returns are based on averages over the week or month, so we use daily data and choose data from each Wednesday or the last Wednesday of the month for our observations. For those Wednesdays that land on a holiday, we use data from the following Thursday. The returns in these data sets are discrete discount returns, so we convert them into continuously compounded annual returns.

Interest rates obtained from 3-month T-bills are used because these products have much longer time series of data available than 1-month T-bills. The interest rate model in the previous section is a discrete time approximation of a continuous time model of the infinitesimally short term risk-free rate. As such it would be better to use a shorter term product with more frequent observations; however, the institutional features of the interest rate market and data availability leave us with the current compromise.[1]

[1] Moving to daily observations may also introduce too much serial correlation, which may lead to inconsistent estimators.

Before we proceed to the general case, it is informative to first consider the simpler case where the parameters are constant and do not depend on the Markov chain. First we rewrite Eq. 2.6 as $\Delta\mathbf{r} = \mathbf{X}\beta + \mathbf{u}$, where $\mathbf{u} \sim N(\mathbf{0}, \Omega)$, $\Omega = \varsigma^2 \mathrm{diag}[r_0^{2\gamma}, \ldots, r_{T-1}^{2\gamma}]$, $\mathbf{X} = [\mathbf{1} \quad \ell_1(\mathbf{r})]$, $\mathbf{1}$ is a column vector of 1's, $\ell_1(\mathbf{r}) = (r_0, \ldots, r_{T-1})^\mathsf{T}$, and $\beta = (\alpha\bar{r}\Delta t, 1-\alpha\Delta t)^\mathsf{T}$. This form makes it clear that for any known γ, β can be estimated by generalized (or weighted) least squares, $\hat{\beta}(\gamma) = (\mathbf{X}^\mathsf{T}\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\Omega^{-1}\Delta\mathbf{r}$, which is the output of the OLS regression $\eta\Delta\mathbf{r} = \eta\mathbf{X}\beta + \eta\mathbf{u}$, for $\eta = \mathrm{diag}[r_0^{-\gamma}, \ldots, r_{T-1}^{-\gamma}]$.

On the other hand, if β is known, then the log-likelihood function is

$$\ell(\mathbf{r};\beta,\gamma,\hat{\sigma}) = \{-T\ln(2\pi) - T - T\ln(\hat{\sigma}^2) - 2\gamma\sum_{t=1}^{T}\ln(r_{t-1})\}/2, \qquad (2.7)$$

where $\hat{\sigma}^2 = \frac{1}{T}\sum\left(\frac{\Delta r_t - \mathbf{X}\beta}{r_{t-1}^{\gamma}}\right)^2$. Given β, this equation can be maximized numerically over the one unknown variable γ.

With normally distributed errors, $\hat{\beta}(\gamma)$ is a maximum likelihood estimate conditional on γ. Since γ is independent of β terms, we can iterate back and forth maximizing conditional on γ, then conditional on β, etc. In fact, thinking of $\hat{\beta}(\gamma)$ as a function of γ allows us to maximize over γ in one step. Details of this approach can be found in Davidson and MacKinnon [9].

The output of this estimation is provided in Table 2.1. All of the parameters have the expected signs; however, none of them are significantly different from zero. Furthermore, none of the series are expected to differ from zero as is seen by the F statistics. This observation foreshadows our more general finding that once the proper diffusion parameters are employed, the drift parameters are not very important. None of the four series differs significantly from a unit root as tested by the Dickey-Fuller τ statistics, all four series have residuals that exhibit significant serial correlation as seen by the Durbin-Watson statistics, and the residuals of all four series have a significantly non-normal distribution according to the Jarque-Bera statistics. These negative results, together with the very low R^2 statistics demonstrate that this type of linear drift one-factor model does not adequately describe 3-month T-bill rates in either country, at least not for such a long time period.[2] Particularly troubling is the combination of serial correlation in the residuals and regressing on lagged dependent variables, which can cause the maximum likelihood estimators to be inconsistent. Combining this observation with the non-normality of the residuals implies that the parameter estimates may not be very accurate.

Nevertheless, to compare the parameters and to allow them to be more easily understood, it is helpful to convert them to the form in Eq. 2.5. This conversion is done in the second panel of Table 2.1. The parameters should be approximately

[2] For robustness we repeated the analysis for the shorter time period from May 1990 to December 2004 using 3-month T-bill rates, 3-month LIBOR rates, and 1-month LIBOR rates, and although the parameters differed substantially from the longer period, the findings were generally similar. The only noteworthy differences were that β_1 was significantly less than zero, ruling out a unit root in each case, and for the Canadian short rates, serial correlation in the residuals was no longer present. Data for the LIBOR rates was obtained at the British Bankers' Association website.

CEV	US-T-3-m	US-T-3-w	Can-T-3-m	Can-T-3-w
β_0	0.000123	2.59×10^{-5}	0.000552	6.81×10^{-5}
t stat	1.137823	0.83111	1.498601	1.065669
β_1	−0.00037	−0.00013	−0.00784	−0.00098
DF τ stat	−0.09936	−0.13147	−1.15416	−0.82891
std err	0.012858	0.002313	0.001547	0.000208
F stat	1.542211	0.668327	1.183339	0.594016
γ	1.156092	1.134993	0.770751	0.772217
r^2	0.009546	0.000596	0.004592	0.00053
DW stat	1.621623	1.831274	1.584126	1.564956
JB stat	241.351	4030.834	2596.504	88514.14
# obs	515	2,243	515	2,243
Δt	0.083333	0.019231	0.083333	0.019231
$\bar{r}$	0.330697	0.200768	0.070411	0.06966
α	0.004479	0.006698	0.094047	0.050839
σ	0.392037	0.34664	0.135987	0.103896
γ	1.156092	1.134993	0.770751	0.772217

Table 2.1 Quasi-maximum likelihood estimation of the regression equation $\Delta r_{t+1} = \beta_0 + \beta_1 r_t + \varsigma r_t^{\gamma} \varepsilon_{t+1}$, where ε_{t+1} are independent standard normal random variables. For a given γ this equation can be estimated independently of the distribution of ε by GLS. The normal distribution is used to obtain the maximum likelihood estimate for γ. Here $\Delta r_{t+1} = r_{t+1} - r_t$, and r_t is the unitless, annual, continuously-compounded yield to maturity at date t on a US or Canadian T-bill maturing 3 months later. These interest rate observations occur monthly or weekly from January 1962 to December 2004, with weekly observations occurring each Wednesday (or the next available date), and monthly observations occurring on the last Wednesday of each month (or the last weekly observation of the month). The column notation is used to indicate the country, interest rate type (T-bill v. LIBOR), maturity (3 month v. 1 month), and observation frequency (monthly v. weekly) of the short rate process. (Although only 3-month T-bill data was available from 1962, other short rates became available from May 1990 and were considered for robustness.) The regression coefficients have Student t statistics reported, which in the case of β_1 is actually a Dickey-Fuller τ statistic used to test against a unit root. The F statistic tests against the fully restricted model with both coefficients being zero. The Durbin-Watson statistic is used to test for serial correlation in the residuals, and the Jarque-Bera statistic is used to test whether the residuals are normally distributed. The values for Δt are $1/12$ for monthly and $1/52$ for weekly observations, which are used to convert the regression coefficient estimates into interest rate model coefficients for the equation $\Delta r_{t+1} = \alpha(\bar{r} - r_t)\Delta t + \sigma r_t^{\gamma} \sqrt{\Delta t}\varepsilon_{t+1}$

equal when comparing the monthly to the weekly observed time series, which is the case for all parameters except the rate of mean reversion parameter α, being difficult to estimate anyway [5, 2, 21].

Comparing parameter estimates for both countries yields a further contrast. The Canadian rates have a central tendency ($\bar{r}$) around 0.07 and a rate of mean reversion (α) between 0.05 and 0.09, whereas the U.S. rates have a much higher central tendency (greater than 0.2) and a much slower rate of mean reversion between 0.0045 and 0.0067. The data suggests that U.S. rates tend toward a fairly high interest rate at a fairly slow rate and Canadian rates tend more quickly toward a modest interest rate. This high US-low Canadian central tendency is even more puzzling when

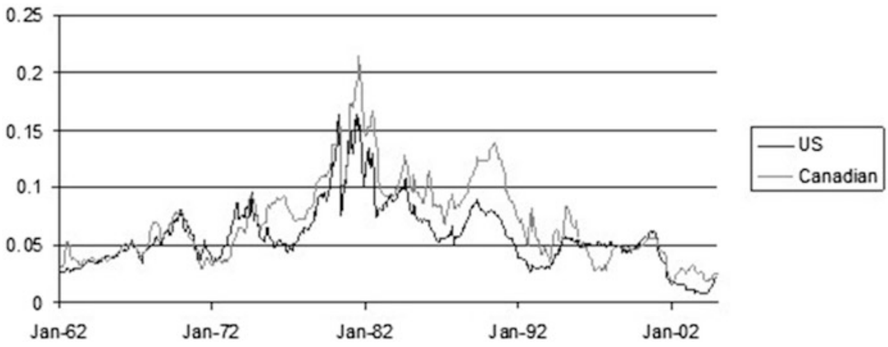

Fig. 2.1 The graph depicts US and Canadian 3-month T-bill rates from January 1962 to December 2004. The rates are in unitless continuously compounded form and they are based on monthly observations on the last Wednesday in each month

considering Fig. 2.1, which shows that US rates never once reach 20 %, and Canadian rates peak higher and typically are higher for most of this period.

Finally, we notice, like Chan et al. [5], that the elasticity parameter for this period is greater than 1 for the U.S. data (between 1.135 and 1.156), which provides a good opportunity to see if incorporating stochastic volatility will reduce the elasticity to an acceptable level as suggested by Sun [21]. It turns out that we find that elasticity is not affected in a predictable way by incorporating stochastic volatility. (In particular, we find that elasticity is decreased in the Canadian case, which was already sufficiently low, but it is not changed significantly in US rates, which could be considered too high for purposes of term structure modeling.)

Yong [23] provides some conditions on the variance elasticity parameter γ for a continuous time model similar to Eq. 2.4, in which the parameter functions α_t, $\bar{r}_t$, and σ_t are bounded, deterministic functions of time. In that case, there does not exist a continuous associated wealth process bounded in expectation when $\gamma > 1/2$. In particular, the condition $E[\exp(\lambda \int_0^T r_t dt)] < \infty$ does not hold for any $T > 0$ and $\lambda > 0$. Yong [23] also shows that Novikov's condition fails when $\gamma \geq 1/2$, which causes problems for determining the existence of an equivalent martingale measure, when using this interest model in conjunction with a Black-Scholes type model for the risky asset(s). Failure of the existence of an equivalent martingale measure implies that the model permits arbitrage.

In empirical investigation, we find $\gamma > 1/2$ for all of our samples except for weekly observations of Canadian T-bill rates. This indicates that a linear interest rate model of the type described by Eq. 2.4 may not adequately fit with historical interest rate data. One approach to deal with this problem could be to apply a non-linear model, such as that put forward by Aït-Sahalia [1]. However, two main distinctions between our model and Yong's [23] model could help explain these problematic empirical results: First, we actually estimate the discrete model given by Eqs. 2.5 and 2.6, instead of the continuous model given by Eq. 2.4, so the parameters may not exactly coincide with those of a continuous model. Second, we do not permit the parameter functions to vary directly and deterministically with time. By permitting the parameters to vary directly with time, information from the term structure could be used to estimate the parameter functions, which might decrease estimates of the elasticity parameter. However, since our main objective in this paper is to compare the relative importance of stochastic volatility and stochastic central tendency, we leave a detailed investigation of this issue for future research.

Figure 2.2 plots the monthly observed US 3-month T-bill rates and the residuals to the estimation equation used in this section. Note that the residuals seem to cluster into high and low volatility regimes, which provides further motivation for our method.

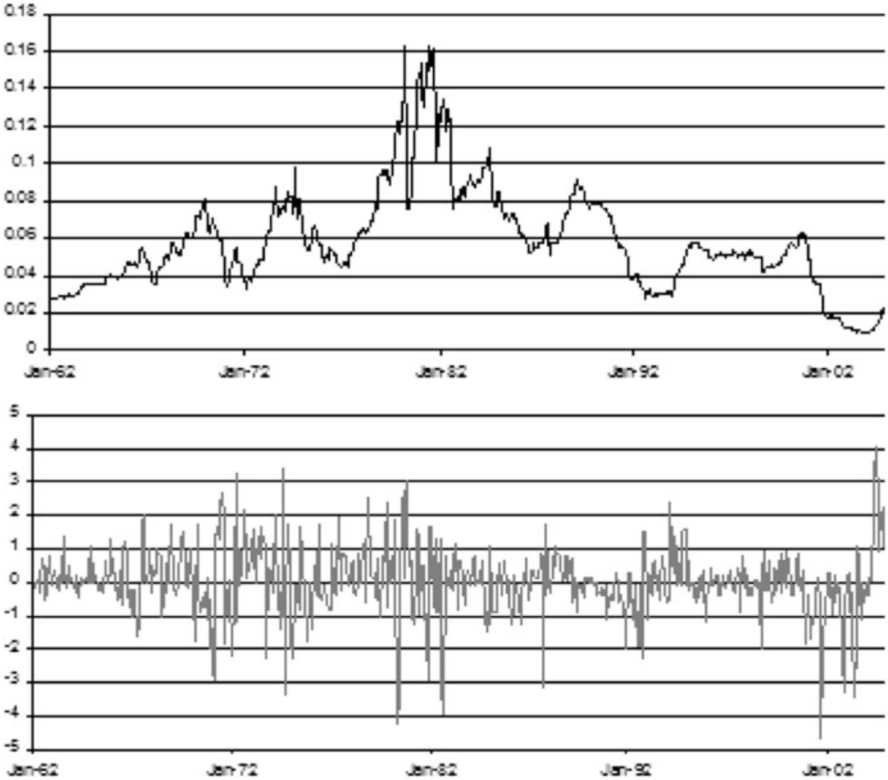

Fig. 2.2 The *upper panel* depicts the US 3-month T-bill rate between January 1962 and December 2004. The *lower panel* plots the standardized residuals from the constant elasticity of volatility model $\Delta r_{t+1} = \beta_0 + \beta_1 r_t + \sigma r_t^{\gamma} \varepsilon_{t+1}$ estimated via maximum likelihood

2.7 Results

We implement the model for a 2-state Markov chain.[3] Because the columns of the Markov matrix must sum to 1, the matrix is associated with two free parameters, also the central tendency and volatility components of the model are associated with two free parameters each. Together with the mean reversion rate and volatility elasticity parameters the model has a total of eight parameters. The estimates obtained by the algorithm are presented in Tables 2.2–2.5. Tables 2.2 and 2.3 use monthly and weekly observed US data respectively and Tables 2.4 and 2.5 use the equivalent Canadian data.

These tables provide parameter estimates for hidden Markov interest rate models. Each table is associated with a separate time series. The first column reports values for the fully restricted model, in which neither parameter is permitted to switch according to the Markov chain. The remaining columns relax various parameter restrictions. We consider restrictions on the stochastic nature of the mean-reverting level or volatility parameters. These restrictions force the level or volatility to be constant by requiring them to take the same values in each state, (although they allow the constant to be arbitrary). For each restricted model, we report maximum likelihood parameter estimates and the value of the log-likelihood function evaluated at the MLE. We also report likelihood ratio statistics, and Durbin-Watson and Jarque-Bera statistics, which are constructed from each model's residuals. The partially restricted models are associated with only one fewer degrees of freedom, (seven free parameters instead of eight); however, the fully restricted model has four fewer degrees of freedom since the stochastic matrix parameters are no longer relevant in that case.

By looking at the likelihood ratio statistics, we see that for all short rate series, the constant volatility restriction can be rejected and the constant central tendency restriction cannot be rejected. The small improvement in likelihood from relaxing the constant level restriction suggests that the parsimonious model is best. A caveat to this finding is that it depends heavily on the second factor of randomness being present. When comparing the fully restricted model with constant central tendency and constant volatility to a stochastic central tendency model, the likelihood does improve substantially. We don't report statistics in the tables, but the smallest likelihood ratio statistic would be about 40, which is highly significant.

This result shows that on its own, stochastic central tendency seems to be very important, which is consistent with Balduzzi et al. [3]. However, when compared on an equal footing with an alternative 2-factor model having stochastic volatility, it is found to be almost completely unimportant in explaining historical interest rates, which suggests that it is the second factor of randomness in general that is important rather than the stochastic central tendency in particular.

[3] For robustness, the complete analysis was repeated for a 3-state Markov chain and the results were qualitatively the same: Stochastic central tendency was important only when compared to the 1-factor model. It was unimportant when compared with a stochastic volatility model.

US-T-3-m	Constant	Constant ς	Constant β_0	Full
$\beta_0(\mathbf{e}_1)$	0.000123	0.00184	8.32×10^{-5}	6.68×10^{-5}
$\beta_0(\mathbf{e}_2)$		0.00011		0.000173
β_1	−0.000373	−0.029164	-2.12×10^{-8}	-2.1×10^{-8}
$\varsigma(\mathbf{e}_1)$	0.113171	0.156445	0.053643	0.053674
$\varsigma(\mathbf{e}_2)$			0.167192	0.167746
γ	1.156092	1.286568	1.159406	1.160018
$\mathbf{P}_{11}$		0.990704	0.919711	0.918096
$\mathbf{P}_{12}$		0.009296	0.080289	0.081904
$\mathbf{P}_{21}$		0.044543	0.120808	0.123999
$\mathbf{P}_{22}$		0.955457	0.879192	0.876001
loglikelihood	2161.35	2181.783	2245.361	2245.467
LR-stat	168.2335	127.3688	0.211469	
DW-stat	1.621623	1.723194	1.677849	1.67412
JB-stat	241.351	328.0739	166.5115	183.8003
# obs	515	515	515	515
Δt	0.083333	0.083333	0.083333	0.083333
$\bar{r}(\mathbf{e}_1)$	0.330697	0.0631	3930.426	3157.727
$\bar{r}(\mathbf{e}_2)$		0.003787		8157.085
α	0.004479	0.349969	2.54×10^{-7}	2.54×10^{-7}
$\sigma(\mathbf{e}_1)$	0.392037	0.541941	0.185826	0.185933
$\sigma(\mathbf{e}_2)$			0.579172	0.581091
γ	1.156092	1.286568	1.159406	1.160018
$\mathbf{Q}_{11}$		−0.114663	−1.075674	−1.1005
$\mathbf{Q}_{12}$		0.114663	1.075674	1.100502
$\mathbf{Q}_{21}$		0.549448	1.618516	1.666099
$\mathbf{Q}_{22}$		−0.549448	−1.618516	−1.6661

Table 2.2 Quasi-maximum likelihood estimation of the regression equation $\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \varsigma(X_t) r_t^{\gamma} \varepsilon_{t+1}$, where ε_{t+1} are independent standard normal random variables and X_t is a hidden Markov chain. Here $\Delta r_{t+1} = r_{t+1} - r_t$, and r_t is the unitless, annual, continuously-compounded yield to maturity at date t on a US T-bill maturing 3 months later based on monthly observations from January 1962 to December 2004 on the last Wednesday of each month. The column notation is used to indicate various model restrictions from the full model with both β_0 and ς allowed to take different values in different states of the Markov chain. The columns "constant ς" "constant β_0" and "constant" imply that ς, β_0 or both respectively take the same value in each state of the Markov chain. The matrix $\mathbf{P}$ is the transition probability matrix with $\mathbf{P}_{ij}$ being the probability of switching from state i to state j. The loglikelihood is used to calculate the likelihood ratio statistics, which are based on the difference in loglikelihood between the full model and each particular restriction. The Durbin-Watson statistic is used to test for serial correlation in the residuals, and the Jarque-Bera statistic is used to test whether the residuals are normally distributed. The value for Δt is used to convert the regression coefficient estimates into interest rate model coefficients for the equation $\Delta r_{t+1} = \alpha(\bar{r}(X_t) - r_t)\Delta t + \sigma(X_t) r_t^{\gamma} \sqrt{\Delta t}\varepsilon_{t+1}$. The matrix $\mathbf{Q}$ is the transition rate matrix $\mathbf{P} = e^{\mathbf{Q}\Delta t}$

US-T-3-w	Constant	Constant ς	Constant β_0	Full
$\beta_0(\mathbf{e}_1)$	2.59×10^{-5}	4.76×10^{-5}	2.01×10^{-5}	1.86×10^{-5}
$\beta_0(\mathbf{e}_2)$		-0.004438		2.65×10^{-5}
β_1	-0.000129	-3.79×10^{-8}	-1.75×10^{-11}	-1.75×10^{-11}
$\varsigma(\mathbf{e}_1)$	0.04807	0.045367	0.026938	0.026941
$\varsigma(\mathbf{e}_2)$			0.077429	0.077445
γ	1.134993	1.137806	1.151832	1.151875
$\mathbf{P}_{11}$		0.993943	0.967915	0.9679
$\mathbf{P}_{12}$		0.006057	0.032085	0.0321
$\mathbf{P}_{21}$		0.343071	0.060586	0.060636
$\mathbf{P}_{22}$		0.656929	0.939414	0.939364
loglikelihood	11195.18	11269.19	11612.03	11612.04
LR-stat	833.7179	685.7047	0.028266	
DW-stat	1.831274	0.858285	1.853169	1.853177
JB-stat	4030.834	12186.02	5194.324	5194.194
# obs	2,243	2,243	2,243	2,243
Δt	0.019231	0.019231	0.019231	0.019231
$\bar{r}(\mathbf{e}_1)$	0.200768	1255.337	1,150,858	1,067,508
$\bar{r}(\mathbf{e}_2)$		-116999.2		1,514,390
α	0.006698	1.97×10^{-6}	9.08×10^{-10}	9.08×10^{-10}
$\sigma(\mathbf{e}_1)$	0.34664	0.327146	0.194254	0.194272
$\sigma(\mathbf{e}_2)$			0.558349	0.558466
γ	1.134993	1.137806	1.151832	1.151875
$\mathbf{Q}_{11}$		-0.387444	-1.750856	-1.751751
$\mathbf{Q}_{12}$		0.387444	1.750856	1.751751
$\mathbf{Q}_{21}$		21.94357	3.306166	3.308974
$\mathbf{Q}_{22}$		-21.94357	-3.306166	-3.308974

Table 2.3 Quasi-maximum likelihood estimation of the regression equation $\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \varsigma(X_t) r_t^{\gamma} \varepsilon_{t+1}$, where ε_{t+1} are independent standard normal random variables and X_t is a hidden Markov chain. Here $\Delta r_{t+1} = r_{t+1} - r_t$, and r_t is the unitless, annual, continuously-compounded yield to maturity at date t on a US T-bill maturing 3 months later based on weekly observations from January 1962 to December 2004 on the last Wednesday of each month. The column notation is used to indicate various model restrictions from the full model with both β_0 and ς allowed to take different values in different states of the Markov chain. The columns "constant ς" "constant β_0" and "constant" imply that ς, β_0 or both respectively take the same value in each state of the Markov chain. The matrix $\mathbf{P}$ is the transition probability matrix with $\mathbf{P}_{ij}$ being the probability of switching from state i to state j. The loglikelihood is used to calculate the likelihood ratio statistics, which are based on the difference in loglikelihood between the full model and each particular restriction. The Durbin-Watson statistic is used to test for serial correlation in the residuals, and the Jarque-Bera statistic is used to test whether the residuals are normally distributed. The value for Δt is used to convert the regression coefficient estimates into interest rate model coefficients for the equation $\Delta r_{t+1} = \alpha(\bar{r}(X_t) - r_t)\Delta t + \sigma(X_t) r_t^{\gamma} \sqrt{\Delta t}\varepsilon_{t+1}$. The matrix $\mathbf{Q}$ is the transition rate matrix $\mathbf{P} = e^{\mathbf{Q}\Delta t}$

Can-T-3-m	Constant	Constant ς	Constant β_0	Full
$\beta_0(\mathbf{e}_1)$	0.000552	0.014384	0.000277	0.000273
$\beta_0(\mathbf{e}_2)$		0.000433		−0.000328
β_1	−0.007837	−0.012064	−0.001139	-6.23×10^{-5}
$\varsigma(\mathbf{e}_1)$	0.039256	0.034997	0.011153	0.010552
$\varsigma(\mathbf{e}_2)$			0.038387	0.036238
γ	0.770751	0.788342	0.553748	0.539309
$\mathbf{P}_{11}$		0.269987	0.932768	0.933597
$\mathbf{P}_{12}$		0.730013	0.067232	0.066403
$\mathbf{P}_{21}$		0.02281	0.191737	0.177576
$\mathbf{P}_{22}$		0.97719	0.808263	0.822424
loglikelihood	2038.84	2073.68	2143.932	2144.174
LR-stat	210.6678	140.9884	0.485521	
DW-stat	1.584126	0.647096	1.528304	1.528396
JB-stat	2596.504	31575.5	1117.866	1076.651
# obs	515	515	515	515
Δt	0.083333	0.083333	0.083333	0.083333
$\bar{r}(\mathbf{e}_1)$	0.070411	1.19231	0.243123	4.38532
$\bar{r}(\mathbf{e}_2)$		0.035868		−5.27452
α	0.094047	0.144767	0.013662	0.000747
$\sigma(\mathbf{e}_1)$	0.135987	0.121232	0.038637	0.036552
$\sigma(\mathbf{e}_2)$			0.132975	0.125534
γ	0.770751	0.788342	0.553748	0.539309
$\mathbf{Q}_{11}$		−16.26365	−0.933712	−0.913454
$\mathbf{Q}_{12}$		16.26365	0.933712	0.913454
$\mathbf{Q}_{21}$		0.508179	2.662839	2.442781
$\mathbf{Q}_{22}$		−0.508179	−2.662839	−2.442781

Table 2.4 Quasi-maximum likelihood estimation of the regression equation $\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \varsigma(X_t) r_t^{\gamma} \varepsilon_{t+1}$, where ε_{t+1} are independent standard normal random variables and X_t is a hidden Markov chain. Here $\Delta r_{t+1} = r_{t+1} - r_t$, and r_t is the unitless, annual, continuously-compounded yield to maturity at date t on a Canadian T-bill maturing 3 months later based on monthly observations from January 1962 to December 2004 on the last Wednesday of each month. The column notation is used to indicate various model restrictions from the full model with both β_0 and ς allowed to take different values in different states of the Markov chain. The columns "constant ς" "constant β_0" and "constant" imply that ς, β_0 or both respectively take the same value in each state of the Markov chain. The matrix $\mathbf{P}$ is the transition probability matrix with $\mathbf{P}_{ij}$ being the probability of switching from state i to state j. The loglikelihood is used to calculate the likelihood ratio statistics, which are based on the difference in loglikelihood between the full model and each particular restriction. The Durbin-Watson statistic is used to test for serial correlation in the residuals, and the Jarque-Bera statistic is used to test whether the residuals are normally distributed. The value for Δt is used to convert the regression coefficient estimates into interest rate model coefficients for the equation $\Delta r_{t+1} = \alpha(\bar{r}(X_t) - r_t)\Delta t + \sigma(X_t) r_t^{\gamma} \sqrt{\Delta t}\varepsilon_{t+1}$. The matrix $\mathbf{Q}$ is the transition rate matrix $\mathbf{P} = e^{\mathbf{Q}\Delta t}$

Can-T-3-w	Constant	Constant ς	Constant β_0	Full
$\beta_0(\mathbf{e}_1)$	6.81×10^{-5}	0.008376	2.45×10^{-5}	2.54×10^{-5}
$\beta_0(\mathbf{e}_2)$		7.37×10^{-5}		-1.17×10^{-5}
β_1	−0.000978	−0.00254	-7.84×10^{-5}	-6.09×10^{-5}
$\varsigma(\mathbf{e}_1)$	0.014408	0.012304	0.002357	0.002356
$\varsigma(\mathbf{e}_2)$			0.010228	0.010221
γ	0.772217	0.771539	0.431682	0.431884
$\mathbf{P}_{11}$		0.296443	0.933413	0.933518
$\mathbf{P}_{12}$		0.703557	0.066587	0.066482
$\mathbf{P}_{21}$		0.008998	0.172207	0.170958
$\mathbf{P}_{22}$		0.991002	0.827793	0.829042
loglikelihood	11140.16	11373.37	11825.4	11825.44
LR-stat	1370.562	904.1384	0.074158	
DW-stat	1.564956	0.322833	1.59396	1.594426
JB-stat	88514.14	358305.9	12302.88	12244.49
# obs	2,243	2,243	2,243	2,243
Δt	0.019231	0.019231	0.019231	0.019231
$\bar{r}(\mathbf{e}_1)$	0.06966	3.297269	0.312126	0.416996
$\bar{r}(\mathbf{e}_2)$		0.029008		−0.192659
α	0.050839	0.13209	0.004077	0.003166
$\sigma(\mathbf{e}_1)$	0.103896	0.088725	0.016999	0.016991
$\sigma(\mathbf{e}_2)$			0.073757	0.073705
γ	0.772217	0.771539	0.431682	0.431884
$\mathbf{Q}_{11}$		−64.01087	−3.95636	−3.946761
$\mathbf{Q}_{12}$		64.01087	3.95636	3.946761
$\mathbf{Q}_{21}$		0.818628	10.23188	10.14906
$\mathbf{Q}_{22}$		−0.818628	−10.23188	−10.14906

Table 2.5 Quasi-maximum likelihood estimation of the regression equation $\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \varsigma(X_t) r_t^{\gamma} \varepsilon_{t+1}$, where ε_{t+1} are independent standard normal random variables and X_t is a hidden Markov chain. Here $\Delta r_{t+1} = r_{t+1} - r_t$, and r_t is the unitless, annual, continuously-compounded yield to maturity at date t on a Canadian T-bill maturing 3 months later based on weekly observations from January 1962 to December 2004 on the last Wednesday of each month. The column notation is used to indicate various model restrictions from the full model with both β_0 and ς allowed to take different values in different states of the Markov chain. The columns "constant ς" "constant β_0" and "constant" imply that ς, β_0 or both respectively take the same value in each state of the Markov chain. The matrix $\mathbf{P}$ is the transition probability matrix with $\mathbf{P}_{ij}$ being the probability of switching from state i to state j. The loglikelihood is used to calculate the likelihood ratio statistics, which are based on the difference in loglikelihood between the full model and each particular restriction. The Durbin-Watson statistic is used to test for serial correlation in the residuals, and the Jarque-Bera statistic is used to test whether the residuals are normally distributed. The value for Δt is used to convert the regression coefficient estimates into interest rate model coefficients for the equation $\Delta r_{t+1} = \alpha(\bar{r}(X_t) - r_t)\Delta t + \sigma(X_t) r_t^{\gamma} \sqrt{\Delta t}\varepsilon_{t+1}$. The matrix $\mathbf{Q}$ is the transition rate matrix $\mathbf{P} = e^{\mathbf{Q}\Delta t}$

The remaining rows report estimates for the equivalent parameterization of Eq. 2.5. These estimates are more meaningful and they are comparable with each other among the different observation frequencies. The entries of the matrix $\mathbf{Q}$ represent the rate that a continuous Markov chain switches between states. They are obtained by solving $\mathbf{P} = e^{\mathbf{Q}\Delta t}$. The values of these observation frequency-robust estimates generally do seem to be quite close to each other; however the LR statistics indicate a significant difference (e.g. comparing monthly to weekly observations for US 3-month T-bills yields a LR statistic of 7.28).

As in Table 2.1, most models predict that β_1 is close to 0, which is associated with a rate of mean reversion α close to 0. The exception is the case with stochastic drift and constant volatility. In that case the mean reversion rate is quite high, which is not too surprising since, for the central tendency to be important, the mean reversion rate can't be too close to zero in our models.

One problem that occurs in some cases is a very high central tendency, (around 151 million percent for the US weekly observed rate). On the other hand this very high central tendency was also associated with a very slow rate of mean reversion in all but the Canadian stochastic central tendency and constant volatility cases, which had high values for P_{12} suggesting that the process switches out of state 1 soon after it enters, but the potential for unreasonably high interest rates to be generated by this process does still exist.[4]

Finally, for the Canadian data, the elasticity parameter does seem to decrease with the introduction of stochastic volatility. On the other hand it increases slightly for the U.S. data. There doesn't seem to be a conclusive empirical finding on this result.

Furthermore, it seems that the potential for our more general models to alleviate the observed serial correlation and non-normality of the residuals for the constant models is not achieved. While the stochastic volatility models tend to increase the Durbin-Watson statistics and reduce the Jarque-Bera statistics, these statistics are nowhere near their optimal values of 2 and 0 respectively. A caveat to this finding is that maximum likelihood estimation does not necessarily minimize the sum of squared residuals, so the reported statistics are not as meaningful as they are for regression models.

Another quantity of interest is the filter for the Markov chain. Figure 2.3 plots the conditional probability of being in the high-volatility state over time as estimated by the full model on monthly observations US 3-month T-bill rates. Several features to notice are that the state probability seems to traverse quickly between high and low values of approximately 0.88 and 0.17. Furthermore, it seems to switch out of each state quite frequently, (and more frequently down than up, consistent with the transition probability estimates of 0.082 for switching up and 0.124 for switching down). Also, the volatility seems to be more likely in the high state between the late 1960s and 1982 and more likely in the low volatility state in the post 1982 period. However, the frequent switching of the Markov chain suggests that it is picking up

[4] For robustness, we further restricted the stochastic volatility models by requiring all drift parameters to be 0. Such restrictions had little effect on the log-likelihoods and the LR statistics were all less than 0.4.

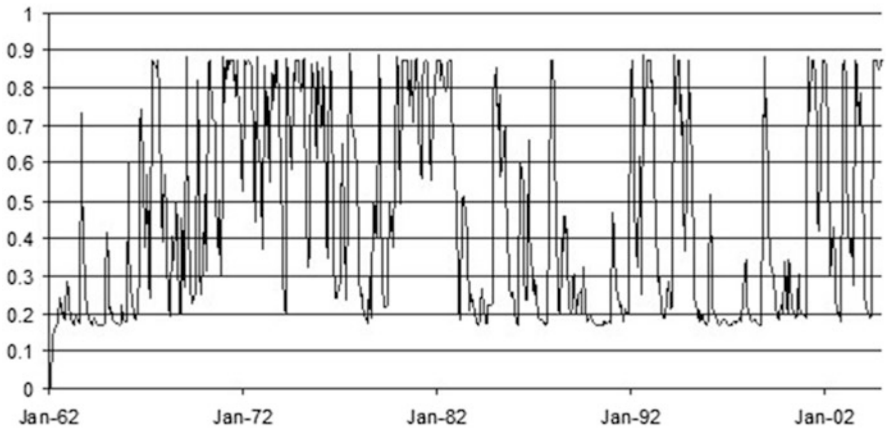

Fig. 2.3 The filter for the Markov chain. The conditional probability (given the observations up to date) of the high volatility (and high central tendency) state as estimated from the hidden Markov model $\Delta r_{t+1} = \beta_0(X_t) + \beta_1 r_t + \sigma(X_t) r_t^\gamma \varepsilon_{t+1}$ via maximum likelihood using monthly observations of the US 3-month T-bill rate, where $\{X_t\}$ is an unobserved 2-state Markov chain

more than just a structural break at 1982. It is more likely that the split is due to the rising and falling long-term trends apparent in Fig. 2.2 together with the fact that the high volatility state is also the high central tendency state in this case.

2.8 Conclusion

We develop a 2-factor model for the short-term interest rate where the second factor enters through both the volatility and the drift, (and in particular, the central tendency). We develop a method for estimating and testing the model. Since the decrease in the likelihood is much greater for the constant volatility restriction, it seems that a stochastic volatility is very important, and far more important than stochastic drift components for explaining nominal interest rate movements. This conclusion is true for both Canadian and U.S. interest rates using both monthly and weekly observations.

A potential weakness of this finding is the limitation of a linear stochastic drift. The combination of non-linear drift and stochastic central tendency may affect our conclusions. This issue is particularly relevant in light of our estimates of the volatility elasticity parameter γ being greater than 1/2 for most samples, which is inappropriate for term structure modeling. We leave this problem to future research.

References

1. Aï t-Sahalia, Y.: Testing continuous-time models of the spot interest rate. Rev. Financ. Stud. **9**, 385–426 (1996)
2. Balduzzi, P., Das, S.R., Foresi, S., Sundaram, R.: A simple approach to three-factor term structure models. J. Fixed Income **6**, 43–53 (1996)
3. Balduzzi, P., Das, S.R., Foresi, S.: The central tendency: a second factor in bond yields. Rev. Econ. Stat. **80**, 62–72 (1998)
4. Bansal, R., Zhou, H.: Term structure of interest rates with regime shifts. J. Finance **57**, 1997–2043 (2002)
5. Chan, K., Karolyi, G., Longstaff, F., Sanders, A.: An empirical comparison of alternative models of the short-term interest rate. J. Finance **47**, 1209–1227 (1992)
6. Chapman, D.A., Pearson, N.D.: Is the short rate drift actually nonlinear? J. Finance **55**, 355–388 (2000)
7. Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385–408 (1985)
8. Dai, Q., Singleton, K.J.: Specification analysis of affine term structure models. J. Finance **55**, 1943–1978 (2000)
9. Davidson, R., MacKinnon, J.G.: Estimation and Inference in Econometrics. Oxford University Press, New York (1993)
10. Elliott, R.J.: Exact adaptive filters for Markov chains observed in Gaussian noise. Automatica **30**, 1399–1408 (1994)
11. Guthrie, G., Wright, J.: The optimal design of interest rate target changes. J. Money Credit Bank. **36**(1), 115–137 (2004)
12. Hamilton, J.D.: Rational expectations economic analysis of changes in regime. J. Econ. Dyn. Control **12**, 385–423 (1988)
13. Heath, D., Jarrow, R., Morton, A.: Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econometrica **60**, 77–105 (1992)
14. Ho, T.S.Y., Lee, S.B.: Term structure movements and pricing interest rate contingent claims. J. Finance **41**, 1011–1029 (1986)
15. Hull, J., White, A.: Pricing interest-rate derivative securities. Rev. Financ. Stud. **3**, 573–592 (1990)
16. Kalimipalli, M., Susmel, R.: Regime-switching stochastic volatility and short-term interest rates. J. Empir. Finance **11**, 309–329 (2004)
17. Lehmann, E.L.: Theory of Point Estimation. Wiley, New York (1983)
18. Longstaff, F., Schwartz, E.: Interest rate volatility and the term structure: a two-factor general equilibrium model. J. Finance **47**, 1259–1282 (1992)
19. Naik, V., Lee, M.H.: Yield curve dynamics with discrete shifts in economic regimes: theory and estimation. Working paper, University of British Columbia (1997)
20. Smith, D.R.: Markov switching and stochastic volatility diffusion models of short-term interest rates. J. Bus. Econ. Stat. **20**(2), 183–197 (2002)

21. Sun, L.: Nonlinear drift and stochastic volatility: an empirical investigation of short-term interest rate models. J. Financ. Res. **26**, 389–404 (2003)
22. Vasicek, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**, 177–188 (1977)
23. Yong, J.: Some estimates on exponentials of solutions to stochastic differential equations. J. Appl. Math. Stoch. Anal. **4**, 287–316 (2004)

Springer

http://www.springer.com/978-1-4899-7441-9

Hidden Markov Models in Finance
Further Developments and Applications, Volume II
Mamon, R.S.; Elliott, R.J. (Eds.)
2014, XXII, 261 p. 47 illus., 39 illus. in color., Hardcover
ISBN: 978-1-4899-7441-9