The Centre for the Analysis of Risk and Optimisation Modelling Applications

A multi-disciplinary research centre focussed on understanding, modelling, quantification, management and control of RISK

TECHNICAL REPORT

CTR/54/06 November 2006

Parameter estimation of an interest rate model via a HMM filtering method in discrete time

Christina Erlwein, Rogemar Mamon

School of Information Systems,
Computing and Mathematics,
Brunel University,
Uxbridge, Middlesex, UB8 3PH
www.carisma.brunel.ac.uk

Parameter estimation of an interest rate model via an HMM filtering method in discrete time

Christina Erlwein [*] Rogemar Mamon [*] [†]

Abstract

This paper considers the implementation of a mean-reverting interest rate model with Markov-modulated parameters. Hidden Markov model filtering techniques in Elliott [8] and Elliott et al. [9] are employed to obtain optimal estimates of the model parameters via recursive filters of auxiliary quantities of the observation process. Algorithms are developed and implemented on a financial data set of 30-day T-bill yields. We found that within the data set and period studied, a model with three regimes is sufficient to describe the interest rate dynamics on the basis of very small predictions errors.

Keywords: term structure - regime-switching model - change of probability measure technique - optimal parameter estimation - hidden Markov model

1 Introduction

The short term interest rate is a key variable in financial modelling because of its importance in pricing and hedging of fixed income securities and other financial derivatives. Well-known short rate models include the no-arbitrage single-factor model of Vasicek [27], where the interest rate is mean-reverting and its extension developed by Hull and

[*] CARISMA School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom

[†] Department of Statistical and Actuarial Sciences, 2nd Floor Western Science Centre, University of Western Ontario, London, ON Canada N6A 5B7

White [21]. The mean reversion level in the Hull-White model is time varying and the yield curve can be fitted to today's term structure. Cox, Ingersoll and Ross [5] on the other hand proposed a single-factor model, where positive interest rates are guaranteed. Multi-factor models by Duffie and Kan [7] and by Longstaff and Schwartz [23] are able to provide a better fit to the yield curve, but the analysis and parameter estimation for these models are more difficult. Other popular interest rate models are the lognormal short rate model of Black and Karasinski [3] and the Heath-Jarrow-Morton methodology [20], where the entire forward-rate curve is modelled. A detailed discussion of these models can be found in Brigo and Mercurio [4].

Most recent term structure models include the possibility of regime-switching for short-term interest rates. Garcia and Perron [17] analysed the time series behaviour of U.S. real interest rates from 1961 to 1986 and found empirical evidence for jumps caused by important structural events. Hamilton [19] introduced changes in regimes by modelling parameters of an autoregression with a discrete Markov chain applied to a business cycle. Gray [18] used a generalised regime-switching model for short-term interest rates, which allows mean-reversion and captures conditional heteroskedasticity. The model outperforms single-regime models in out-of-sample forecasting. Similar evidence supporting good performance of regime-switching short-term interest-rate models is described by Bansal and Zhou [2]. Here the efficient method of moments is used for the estimation of the model parameters. Amongst others, Naik and Lee [24] and Evans [15] include regime-switching in short-term interest rate models. Driffill, Kenc and Sola [6] found empirical evidence, that regime shifts add more realism to interest rate models. A study by Smith [26] supports Markov switching models over stochastic volatility models, because the volatility seems to depend on the level of the short rate.

Landen [22] developed a hidden Markov model (HMM) for short-term interest rates, where drift and diffusion parameters are modulated by an underlying Markov process. Whilst the paper of Elliott, Fischer and Platen [11] addresses the HMM filtering of a mean-reverting model, the derived filters are in continuous time setting and only a simulation is given. In a model by Elliott, Hunter and Jamieson [12] the short rate process r is a function of a Markov chain. A closed form solution for bond prices, where the underlying short rate is modelled by a mean reversion level governed by a continuous time Markov chain is derived in related work of Elliott and Mamon [13], however the volatility process is constant in their model formulation.

A central concern for these models is parameter estimation. In this paper we provide recursive estimates for the parameters following the approach of Elliott [8] in discrete time. We demonstrate the calculation of exact adaptive filters for Markov chains observed in

Gaussian noise together with the recursive filters for the jump process and occupation times. These filters provide optimal estimates of the parameters of the proposed interest rate model. All calculations are made under an idealised measure $\bar{P}$, equivalently to the real world measure P.

The paper is organised as follows. Section 2 describes the model framework and a discussion on how to incorporate a HMM in the single-factor Hull-White [21] model. In section 3, related processes of the Markov chain are estimated using the filtering technique. The recursive parameter estimations are outlined in Section 4. In Section 5, one-step-ahead forecasts for short-term interest rates are generated on 30-day Treasury-bill rates. The last section concludes with some remarks on the results of the model implementation.

2 Model description

The short term interest rate in the Hull-White model [21] follows the stochastic differential equation (SDE)

$$dr_t = [\theta_t - a_t r_t]\, dt + \xi_t\, dW_t. \tag{1}$$

The parameters a_t, θ_t and ξ_t are deterministic functions and θ_t is chosen so that the model matches the initial term structure. In equation (1), $W = \{W_t : 0 \le t \le T\}$ is a Wiener process independent of θ_t. The volatility of the short rate is described by ξ_t whilst a_t is the mean-reversion rate.

We can re-arrange the SDE in (1) to get

$$dr_t = a_t[\beta_t - r_t]\, dt + \xi_t\, dW_t \tag{2}$$

where $\beta = \frac{\theta}{a}$. Equation (2) is a particular case of the Ornstein-Uhlenbeck process with mean reversion level β. It has the solution

$$r_t \quad = \quad r_0 e^{-at} + (1 - e^{-at})\beta + \xi e^{-at} \int_0^t e^{au}\, dW_u \ \ . \tag{3}$$

Throughout the succeeding discussion, all vectors will be denoted by bold letters in lowercase whilst matrices will be denoted by English or Greek letters in uppercase.

In this exposition, it is supposed that the short rate process r can be proxied by the yield rates of T-bills with very short maturity observed in discrete time. A discrete

Markov chain $\mathbf{x}_k$, which represents the state of the economy, is assumed hidden in these observed values. Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space of a homogeneous Markov chain $\mathbf{x}_k$ with finite state in discrete time $(k = 0, 1, ...)$. The distribution of $\mathbf{x}_0$ is known and the state space of $\mathbf{x}_k$ is associated with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$ of $\mathbb{R}^n$. The ith vector $\mathbf{e}_i$ is given by $\mathbf{e}_i = (0, ..., 1, ..., 0)'$, where $'$ denotes the transpose. Let $\mathcal{F}_k^0 = \sigma\{\mathbf{x}_0, \ldots, \mathbf{x}_k\}$ be the $\sigma-$field generated by $\mathbf{x}_0, \ldots, \mathbf{x}_k$, and $\mathcal{F}_k$ be the complete filtration generated by $\mathcal{F}_k^0$. Furthermore let $\mathcal{R}_k$ denote the complete filtration generated by r, so that $\mathcal{H}_k = \mathcal{F}_k \vee \mathcal{R}_k$ is the global filtration generated by $\mathbf{x}$ and r.

Under the real world probability measure P, the Markov chain $\mathbf{x}$ has dynamics

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \mathbf{v}_{k+1} \tag{4}$$

where $\mathbf{v}_{k+1}$ is a martingale increment with $\mathrm{E}[\mathbf{v}_{k+1} \mid \mathcal{F}_k] = 0$, $\mathbf{\Pi} = (\pi_{ji})$ is the transition probability matrix and $\pi_{ji} = P(\mathbf{x}_{k+1} = \mathbf{e}_j | \mathbf{x}_k = \mathbf{e}_i)$.

In our proposed model, the interest rate r follows the stochastic process

$$dr_t = a(\mathbf{x}_t)[\beta(\mathbf{x}_t) - r_t]\, dt + \xi(\mathbf{x}_t)\, dW_t \tag{5}$$

for $r_0 \geq 0$ with $a(\mathbf{x}_t) = \langle \mathbf{a}, \mathbf{x}_t \rangle$, $\beta(\mathbf{x}_t) = \langle \beta, \mathbf{x}_t \rangle$ and $\xi(\mathbf{x}_t) = \langle \xi, \mathbf{x}_t \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual Euclidean scalar product. All three parameters are governed by a Markov chain, which ensures, that the model is switching from one economic regime to another through time.

Consider the interest rate process over the time interval $[s, t]$. Then, if $t - s$ is small and $\mathbf{x}$ is constant over this interval, the solution of equation (5), after invoking (3), is

$$\begin{aligned} r_t &= e^{-a(\mathbf{x}_s)(t-s)} r_s + \beta(\mathbf{x}_s)(1 - e^{-a(\mathbf{x}_s)(t-s)}) \\ &\quad + \xi(\mathbf{x}_s) e^{-a(\mathbf{x}_s)t} \int_s^t e^{a(\mathbf{x}_s)u} dW_u \end{aligned} \tag{6}$$

The stochastic integral $e^{-a(\mathbf{x}_s)t} \int_s^t e^{a(\mathbf{x}_s)u} dW_u$ in (6) is normally distributed with mean zero and variance

$$\int_s^t e^{2a(\mathbf{x}_s)(u-t)} du = \frac{(1 - e^{-2a(\mathbf{x}_s)(t-s)})}{2a(\mathbf{x}_s)}.$$

From equation (6) we can derive the discrete representation of the interest rate as

$$r_{k+1} = \alpha(\mathbf{x}_k) r_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k) z_{k+1} \tag{7}$$

where we set $\alpha(\mathbf{x}_k) = e^{-a(\mathbf{x}_k)\Delta}$, $\gamma(\mathbf{x}_k) = \beta(\mathbf{x}_k)(1 - e^{-a(\mathbf{x}_k)\Delta})$ and $\eta(\mathbf{x}_k) = \xi(\mathbf{x}_k)\sqrt{\frac{1-e^{-2a(\mathbf{x}_k)\Delta}}{2a(\mathbf{x}_k)}}$. Here, $\{\mathbf{x}_k\}$ is a discrete-time Markov chain and $\{z_k\}$ is a sequence of IID standard normal random variables.

Now we assume that we have a series of yield observations $\{y_k : k \in \mathbb{N}\}$ of the form

$$y_{k+1} = \alpha(\mathbf{x}_k)y_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k)z_{k+1} \quad . \tag{8}$$

The filtrations generated by the processes are defined by $\mathcal{F}^y = \sigma(y_1, y_2, ...)$, $\mathcal{F}^{\mathbf{x}} = \sigma(\mathbf{x}_1, \mathbf{x}_2, ...)$ and $\mathcal{G} = \mathcal{F}^y \vee \mathcal{F}^{\mathbf{x}}$. We start with a reference probability measure $\bar{P}$ under which we can find optimal estimates for the unobservable process $\mathbf{x}$ and related quantities. To back out the real world measure P from $\bar{P}$ we follow Elliott, Aggoun and Moore [9], Chapter 8. We define the measure P by $\frac{dP}{d\bar{P}} \mid_{\mathcal{G}_t} = \bar{\Lambda}_t$ with

$$\begin{aligned} \bar{\lambda}_l &= \exp\Big[-\frac{\langle \alpha, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \gamma, \mathbf{x}_{l-1}\rangle}{\langle \eta, \mathbf{x}_{l-1}\rangle} \cdot \frac{y_l}{\langle \eta, \mathbf{x}_{l-1}\rangle} \\ &\qquad - \frac{(\langle \alpha, \mathbf{x}_{l-1}\rangle y_{l-1} + \langle \gamma, \mathbf{x}_{l-1}\rangle)^2}{2\langle \eta, \mathbf{x}_{l-1}\rangle^2}\Big] \qquad (9) \\ \bar{\Lambda}_l &= \prod_{k=1}^{l} \bar{\lambda}_k \qquad (10) \end{aligned}$$

with $\bar{\Lambda}_0 = 1, \{\bar{\lambda}_l : \bar{\lambda} \in \mathbb{N}^+\}$ and $\{\bar{\Lambda}_l : l \in \mathbb{N}\}$. The process $\{\bar{\Lambda}_l\}$ is a $\mathcal{G}$-martingale under $\bar{P}$.

3 Calculation of filters

We wish to determine the expectation of any $\mathcal{G}-$adapted stochastic process H given the filtration $\mathcal{F}_k^y$. Using Bayes-theorem, a filter for any adapted process H is given by

$$E\left[H_k \mid \mathcal{F}_k^y\right] = \frac{\overline{E}\left[H_k\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]}{\overline{E}\left[\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]} \quad .$$

We define $\sigma(H_k) := \overline{E}\left[H_k\bar{\Lambda}_k \mid \mathcal{F}_k^y\right]$, so that $E\left[H_k \mid \mathcal{F}_k^y\right] = \frac{\sigma(H_k)}{\sigma(1)}$. Clearly, $\sigma(H_0) = E[H_0]$. We shall find a recursive formula for $\sigma(H_{k-1}\mathbf{x}_{k-1})$. Since H is a scalar, $\sigma(H_{k-1}\mathbf{x}_{k-1})$ is a vector. To relate $\sigma(H_k)$ and $\sigma(H_k\mathbf{x}_k)$ we note that $\langle \mathbf{1}, \mathbf{x}_k\rangle = 1$. Hence,

$$\langle \mathbf{1}, \sigma(H_k\mathbf{x}_k)\rangle = \sigma(H_k\langle \mathbf{1}, \mathbf{x}_k\rangle) = \sigma(H_k). \tag{11}$$

Therefore $E\big[H_k \mid \mathcal{F}_k^y\big] = \frac{\langle \mathbf{1}, \sigma(H_k \mathbf{x}_k)\rangle}{\langle \mathbf{1}, \sigma(\mathbf{x}_k)\rangle}$. Suppose H_l is a scalar $\mathcal{G}-$adapted process, H_0 is $\mathcal{F}_0^{\mathbf{x}}$ measurable and

$$H_l = H_{l-1} + a_l + \langle \mathbf{b}_l, \mathbf{v}_l \rangle + g_l f(y_l) \tag{12}$$

where a, b and g are $\mathcal{G}$-predictable, f is a scalar-valued function and $\mathbf{v}_l = \mathbf{x}_l + \mathbf{\Pi}\mathbf{x}_{l-1}$. From theorem 5.3 of Elliott [8], a recursive relation for $\sigma_k(H_k\mathbf{x}_k)$ is given by

$$\begin{aligned} \sigma_k(H_k\mathbf{x}_k) \;\; = \;\; & \sum_{i=1}^{n} \Gamma^i(y_k)\big[\langle \mathbf{e}_i, \sigma_{k-1}(H_{k-1}\mathbf{x}_{k-1})\rangle \mathbf{\Pi}\mathbf{e}_i \\ & + \langle \mathbf{e}_i, \sigma_{k-1}(a_k \mathbf{x}_{k-1})\rangle \mathbf{\Pi}\mathbf{e}_i \\ & + (\text{diag}(\mathbf{\Pi}\mathbf{e}_i) - \mathbf{\Pi}\mathbf{e}_i \otimes \mathbf{\Pi}\mathbf{e}_i)\sigma_{k-1}(b_k\langle \mathbf{e}_i, \mathbf{x}_{k-1}\rangle) \\ & + \sigma_{k-1}(g_k\langle \mathbf{e}_i, \mathbf{x}_{k-1}\rangle) f(y_k)\mathbf{\Pi}\mathbf{e}_i\big] \end{aligned} \tag{13}$$

with

$$\Gamma^i(y_l) = \exp\left[-\frac{\alpha_i y_{l-1} + \gamma_i}{\eta_i} \cdot \frac{y_l}{\eta_i} - \frac{(\alpha_i y_{l-1} + \gamma_i)^2}{2\eta_i^2}\right] \tag{14}$$

where $\otimes$ denotes the tensor product of vectors in $\mathbb{R}^n$ and diag(B) is a diagonal matrix B with $(b_1, b_2, ..., b_N)'$ in the diagonal.

4 Recursive estimation of processes

Now, for the estimation of the unknown parameters we need an estimator for the state of the Markov chain as well as for the three related processes. These processes are special cases of the general form $H_l = H_{l-1} + a_l + \langle \mathbf{b}_l, \mathbf{v}_l\rangle + g_l f(y_l)$, where H_0 is $\mathcal{F}_0^{\mathbf{x}}$ measurable.

The estimator $\sigma(\mathbf{x}_k)$ can be derived from $\sigma(H_k\mathbf{x}_k)$ by setting $H_k = 1$, $a_k = 0$, $b_k = 0$ and $g_k = 0$. From (13), this implies

$$\sigma(\mathbf{x}_k) = \sum_{i=1}^{N} \Gamma^i(y_k)\langle \mathbf{e}_i, \sigma_{k-1}(\mathbf{x}_{k-1})\rangle \mathbf{\Pi}\mathbf{e}_i \ \ . \tag{15}$$

Let J_k^{sr} represent the number of jumps of $\mathbf{x}_k$ from state $\mathbf{e}_r$ to state $\mathbf{e}_s$ in time k. So,

$$\begin{aligned} J_k^{sr} \;\; &= \;\; \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r\rangle\langle \mathbf{x}_l, \mathbf{e}_s\rangle \\ &= \;\; J_{k-1}^{sr} + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \pi_{sr} + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle\langle \mathbf{v}_k, \mathbf{e}_s\rangle \ \ . \end{aligned} \tag{16}$$

Setting $H_k = J_k^{(sr)}$,$H_0 = 0$, $a_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \pi_{sr}$, $b_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \mathbf{e}_s^{'}$ and $g_k = 0$ in equation (13) we get

$$\begin{aligned} \sigma_k(J_k^{sr}\mathbf{x}_k) &= \sum_{i=1}^{N} \Gamma^i(y_k)\overline{E}\big[\sigma_{k-1}\langle \mathbf{x}_{k-1}, \mathbf{e}_i\rangle \{J_{k-1}^{sr}\mathbf{\Pi}\mathbf{e}_i + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \pi_{sr}\mathbf{\Pi}\mathbf{e}_i \\ &\quad + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \mathbf{e}_s'(\text{diag}\mathbf{\Pi}\mathbf{e}_i - \mathbf{\Pi}\mathbf{e}_i \otimes \mathbf{\Pi}\mathbf{e}_i)\}\big] \\ &= \sum_{i=1}^{N} \Gamma(y_k)\langle \sigma_{k-1}(J_{k-1}^{sr}\mathbf{x}_{k-1}), \mathbf{e}_i\rangle \mathbf{\Pi}\mathbf{e}_i \\ &\quad + \Gamma^r(y_k)\sigma_{k-1}(\langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle)\pi_{sr}\mathbf{e}_s \ . \end{aligned} \tag{17}$$

Let O_k^r denote the occupation time of the Markov process $\mathbf{x}$, that is the length of time $\mathbf{x}$ spent in state r up to time k. Then,

$$O_k^r = \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r\rangle = O_{k-1}^r + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \tag{18}$$

Here we set $H_k = O_k^r$, $H_0 = 0$, $a_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle$, $b_k = 0$ and $g_k = 0$ in equation (13) to obtain

$$\begin{aligned} \sigma_k(O_k^r\mathbf{x}_k) &= \sum_{i=1}^{N} \Gamma^i(y_k)\{\langle \sigma_{k-1}(O_{k-1}^r\mathbf{x}_{k-1}), \mathbf{e}_i\rangle \mathbf{\Pi}\mathbf{e}_i \\ &\quad + \sigma_{k-1}(\langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_i\rangle)\mathbf{\Pi}\mathbf{e}_i\} \\ &= \sum_{i=1}^{N} \Gamma^i(y_k)\langle \sigma_{k-1}(O_{k-1}^r\mathbf{x}_{k-1}, \mathbf{e}_i\rangle \mathbf{\Pi}\mathbf{e}_i \\ &\quad + \Gamma^r(y_k)\langle \sigma_{k-1}(\mathbf{x}_{k-1}), \mathbf{e}_r\rangle \mathbf{\Pi}\mathbf{e}_r \ . \end{aligned} \tag{19}$$

Finally, define the process $T_k^r(f)$ as

$$\begin{aligned} T_k^r(f) &:= \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r\rangle f(y_l) \\ &= T_{k-1}^r(f) + \langle \mathbf{x}_{k-1}, \mathbf{e}_r\rangle f(y_k) \end{aligned} \tag{20}$$

where f is a function of the form $f(y) = y$, $f(y) = y^2$ or $f(y) = y_{l+1}y_l, 1 \leq l \leq k$. We apply formula (13) with substitution $H_k = T_k^r(g)$, $H_0 = 0$, $a_k = 0$, $b_k = 0$ and

$g_k = \langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle$ and get

$$\begin{aligned}
\sigma_k(T_k^r(f)\mathbf{x}_k) &= \sum_{i=1}^{N} \Gamma^i(y_k)\{\langle \sigma_{k-1}(T_{k-1}^r(f)\mathbf{x}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi} \mathbf{e}_i \\
&\quad + \sigma_{k-1}(\langle \mathbf{x}_{k-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_i \rangle) f(y_k) \mathbf{\Pi} \mathbf{e}_i\} \\
&= \sum_{i=1}^{N} \Gamma^i(y_k)\{\langle \sigma_{k-1}(T_{k-1}^r(f)\mathbf{x}_{k-1}), \mathbf{e}_i \rangle \mathbf{\Pi} \mathbf{e}_i \\
&\quad + \Gamma^r(y_k) \langle \sigma_{k-1}(\mathbf{x}_{k-1}), \mathbf{e}_r \rangle f(y_k) \mathbf{\Pi} \mathbf{e}_r
\end{aligned} \tag{21}$$

The recursive optimal estimates of J, O and T can be calculated using equation (11).

5 Parameter estimation

In this section we discuss the estimation of the model parameters for the observation process

$$y_{k+1} = \alpha(\mathbf{x}_k) y_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k) z_{k+1} \tag{22}$$

and the transition probability matrix $\mathbf{\Pi}$ of the Markov chain. The set of parameters ρ, which determines our model is

$$\rho = \{\pi_{ji}, \alpha_i, \gamma_i, \eta_i, 1 \leq i, j \leq n\}. \tag{23}$$

Initial values of these parameters are assumed to be given. With the EM algorithm (see [10] for details) we wish to find a new set of parameters $\hat{\rho}$, which maximises the conditional expectation of the log-likelihoods. Write $\hat{H}_l = E[H_l | \mathcal{Y}_k]$. To find an estimate for the transition probability matrix $\mathbf{\Pi} = \pi_{ji}$, where $\sum_i^n \pi_{ji} = 1$ we consider the Radon-Nikodym derivative

$$\begin{aligned}
\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{Y}_k} = \Lambda_k &= \prod_{l=1}^{k} \left(\sum_{r,s=1}^{n} \left(\frac{\hat{\pi}_{sr}}{\pi_{sr}} \right)^{\langle \mathbf{x}_l, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle} \right) \\
\text{with} \quad \Lambda_0 &= 1
\end{aligned}$$

The log-likelihood is therefore

$$\log \Lambda_k = \sum_{s,r}^{n} J_k^{sr} \log \hat{\pi}_{sr} + Remainder$$

where the remainder does not involve $\hat{\pi}_{sr}$. Subject to the constraint $\sum_i^n \hat{\pi}_{ji} = 1$ we maximise the log-likelihood. Consequently, the optimal estimates for the parameters $\hat{\pi}, \hat{\alpha}, \hat{\gamma}, \hat{\eta}$ are

$$\hat{\pi}_{ji} = \frac{\hat{J}_l^{ji}}{\hat{O}_l^i} \tag{24}$$

$$\hat{\alpha}_i = \frac{\hat{T}_l^i(y_{l+1}, y_l) - \hat{T}_l^i(y)\gamma_i}{\hat{T}_l^i(y^2)} \tag{25}$$

$$\hat{\gamma}_i = \frac{\hat{T}_{l+1}^i(y) - \hat{T}_l^i(y)\hat{\alpha}_i}{\hat{O}_l^i} \tag{26}$$

$$\begin{aligned}\hat{\eta}_i = &\frac{\hat{T}_{l+1}^i(y^2) + \hat{\alpha}_i^2\hat{T}_l^i(y^2) + \hat{\gamma}_i^2\hat{O}_l^i - 2\hat{\alpha}_i\hat{T}_l^i(y_{l+1}, y_l)}{\hat{O}_l^i} \\ &- \frac{2\hat{\gamma}_i\hat{T}_{l+1}^i(y) + 2\hat{\alpha}_i\hat{\gamma}_i\hat{T}_l^i(y)}{\hat{O}_l^i} \quad .\end{aligned} \tag{27}$$

The sketch of the proofs of equations $(25) - (27)$ are provided in the Appendix. As in Elliott, Sick and Stein [14] the first parameter α is updated using the parameter estimate γ from the previous calculated optimal parameter set. However, once α is updated, this new optimal parameter estimate is used for updating the remaining parameters. Evidently, the above parameter estimates rely on the processes $\sigma(J_k^{ji}), \sigma(O_k^i)$ and $\sigma(T_k^i(f))$, which were explicitly specified by recursive equations $(17), (19)$ and (21), respectively. Since these optimal estimates are updated upon the arrival of new information throughout the implementation the filtering procedure produces a self-calibrating model.

6 Implementation

We use 30-day Treasury-bill rates as a proxy for the short-term interest rates. The data set is compiled by the Bank of Canada and consists of daily 30-day Treasury-bill yields between 1996 and 2005 and there are 2500 data points in total. The data is processed in batches of 20 data points, therefore the parameters are roughly updated monthly. For this data set, the parameters were updated 125 times.

A preliminary analysis of the actual data reveals that evolution of the T-bill rates undergoes several distinct regimes characterised by states with high and low means as well as high and low standard deviations. The regime-switching model is developed to capture this particular behaviour. Tables 1 and 2 present the segregation of actual data into either two or three states.

1st state	2nd state
Jan 1996 - Oct 1996 *mean*: 4.49 *std*: 0.625	Nov 1996 - Dec 1997 *mean*: 2.86 *std*: 0.286
Jan 1998 - Jun 2001 *mean*: 4.74 *std*: 0.42	Jul 2001 - Dec 2005 *mean*: 2.52 *std*: 0.47

Table 1: **Segregation of the period of actual data into 2 states**

1st state	2nd state	3rd state
Jan 1996 - June 1997 *mean:* 3.723 *std:* 0.9924	July 1997 - Dec 2000 *mean:* 4.50 *std:* 0.7343	Jan 2002 - Dec 2005 *mean:* 2.452 *std:* 0.3447
Jan 2001 - Dec 2001 *mean:* 3.930 *std:* 1.0929		

Table 2: **Segregation of the period of actual data into 3 states**

The calculated means and standard deviations are used as rough guides for the initial values in estimating the parameters α, γ, η and the transition probability matrix $\mathbf{\Pi}$. The values of the parameters after 125 passes are also included in Tables 3 and 4.

The one-step ahead predicted yields of the T-bill rates are calculated by

$$\begin{aligned} E\big[y_{k+1} \mid \mathcal{G}_k\big] &= E\big[\alpha(\mathbf{x}_k)y_k + \gamma(\mathbf{x}_k) + \eta(\mathbf{x}_k)z_{k+1} \mid \mathcal{G}_k\big] \\ &= \langle \alpha, \mathbf{\Pi}\hat{\mathbf{x}}_k\rangle y_k + \langle \gamma, \mathbf{\Pi}\hat{\mathbf{x}}_k\rangle \end{aligned} \tag{28}$$

parameter	initial value	value after 125 passes
α	(0.8187 0.7204)	(1.4625 2.0332)
γ	(0.6363 0.9816)	(-1.3726 -3.0352)
η	(3.2762 5.0681)	(6.9096 8.4849)
Π	$\begin{bmatrix} 0.3333 & 0.6667 \\ 0.6667 & 0.3333 \end{bmatrix}$	$\begin{bmatrix} 0.0107 & 0.9004 \\ 0.9893 & 0.0996 \end{bmatrix}$

Table 3: **Results of parameter estimation for a 2-state HMM-based interest rate model**

parameter	initial value	value after 125 passes
α	(0.9048 0.8597 0.7977)	(0.8206 1.6322 1.7706)
γ	(0.3340 0.4926 0.7100)	(0.4973 -1.8747 -2.2781)
η	(1.7178 2.5343 3.6576)	(4.6131 6.4416 6.8690)
Π	$\begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.0359 & 0.0736 & 0.0193 \\ 0.9641 & 0.9264 & 0.9807 \end{bmatrix}$

Table 4: **Results of parameter estimation for a 3-state HMM-based interest rate model**

where $\hat{\mathbf{x}}_k = E\big[\mathbf{x}_k \mid \mathcal{G}_k\big]$. Figure 1 shows the actual time series and the resulting one-step ahead forecasts for a 3-state HMM-based interest rate model between 1996 and 2005.

Apparently, we impose the number of states in the implementation. Of course, in doing this, we are guided by the realistic features of the actual data. We put them into different categories according to mean and standard deviation. Following the date segregation displayed in tables 1 and 2 we generate a 1-step ahead forecast with a 2- and 3-state HMM. It is evident that all forecasts and the actual data are very close to each other. We did generate forecasts in a 4-state HMM-based interest rate model to see if any further improvement can be gained. However, we did not find evidence of this. The evolution of the parameters after 125 passes for the 3-state HMM-based interest rate model can be seen in figure 2.

We adopt the criterion of Armstrong and Collopy [1] in assessing the goodness of fit of the one-step ahead forecasts. We evaluate the Median Relative Absolute Error (MdRAE) and the Median Absolute Percentage Error (MdAPE) for the 2- and 3-state HMM respectively. Furthermore we calculate the Mean Square Error (MSE), which can be used for a comparison of the models. The results of this error analysis are given in Table 5.

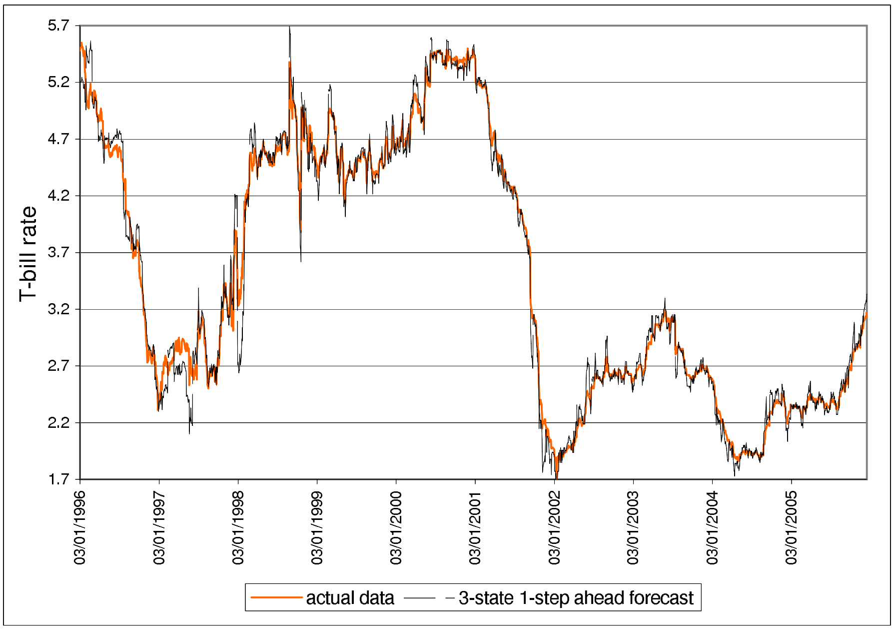

Figure 1: **Plot of actual and one-step ahead forecasts generated by a 3-state HMM-based interest rate model**

A comparison of the MdRAE, MdAPE and MSE gives the 3-state HMM-based interest rate model a better fit than the 2-state HMM-based model, although the error differences between the 2-state and 3-state models is not that significant.

	MdRAE	MdAPE	MSE
2-state MC	0.0041	0.0164	0.0252
3-state MC	0.004	0.0169	0.0178

Table 5: **An error analysis of the 2-state and 3-state HMM-based interest rate model.**

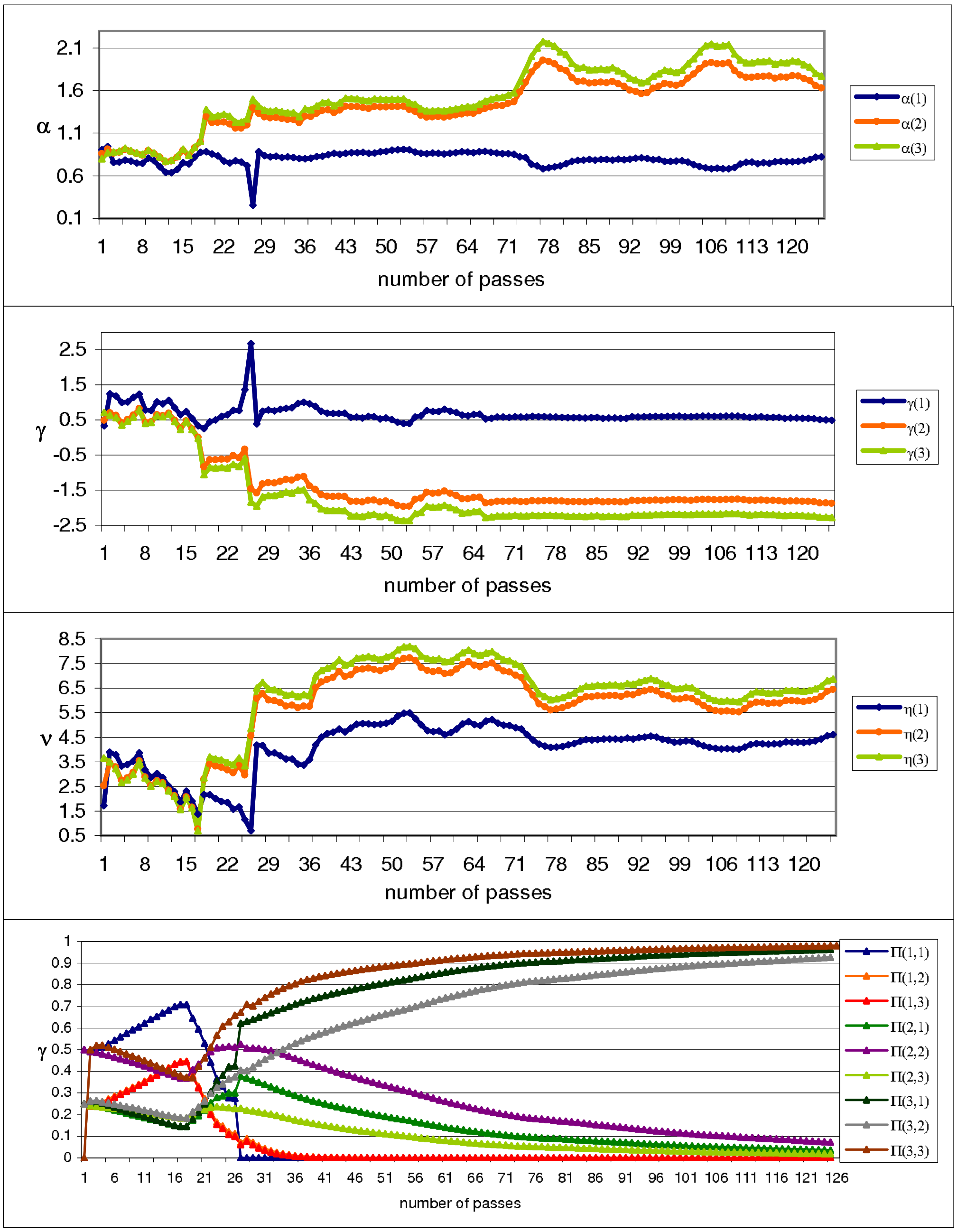

Figure 2: **Evolution of estimates for the parameters α, γ, η and the transition probabilities π_{ij} for a 3-state HMM-based interest rate model**

7 Concluding remarks

By utilising HMM filtering techniques, we implemented successfully an interest rate model in which the volatility, and both the level and speed of mean-reversion are governed by a Markov chain in discrete time. In particular, the proposed model is tested on a financial time series of T-bill yields with very short maturity. The model possesses self-tuning characteristics since the parameters can be updated every time a new set of information arrives. This is made possible through the use of recursive filters developed for the optimal estimate of the state of the Markov chain and other related quantities of the observation process. Our empirical results show that an interest rate model with a Hull-White specification and which allows switching between two regimes is capable of capturing the dynamics of the short-term interest rate process. However, within the algorithms and procedures outlined in this paper the systematic choice of initial values and an approach to selecting the optimal number of states still largely remain unexplored areas which warrant further analysis and research.

8 Acknowledgement

Both authors would like to thank the financial support provided by a Marie Curie Fellowship for Early Stage Researchers Training.

Appendices

A Optimal estimate for α

To derive an optimal estimate for α we consider a new measure $\hat{P}$, which is defined by

$$\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{Y}_k} = \Lambda_k^{\alpha} \;=\; \prod_{l=1}^{k} \lambda_l^{\alpha}$$

where

$$
\begin{aligned}
\lambda_l^\alpha &= \frac{\exp\left(-\frac{1}{2\eta^2(\mathbf{x}_l)}(y_{l+1}^2 + (\hat{\alpha}(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l))\right)}{\exp\left(-\frac{1}{2\eta^2(\mathbf{x}_l)}(y_{l+1}^2 + (\alpha(\mathbf{x}_l)y_l)^2 + \gamma(\mathbf{x}_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l - 2y_{l+1}\gamma(\mathbf{x}_l) + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l))\right)} \\
&= \exp\big(\frac{1}{2\eta^2(\mathbf{x}_l)}(\alpha(\mathbf{x}_l)y_l)^2 - (\hat{\alpha}(\mathbf{x}_l)y_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l \\
&\qquad + 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)\big).
\end{aligned}
$$

This means that

$$
\begin{aligned}
\log\Lambda_k^\alpha &= \sum_{l=1}^{k}\big[(\alpha(\mathbf{x}_l)y_l)^2 - (\hat{\alpha}(\mathbf{x}_l)y_l)^2 - 2y_{l+1}\alpha(\mathbf{x}_l)y_l + 2y_{l+1}\hat{\alpha}(\mathbf{x}_l)y_l \\
&\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\hat{\alpha}(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l)\big]/2\eta^2(\mathbf{x}_l) \\
&= \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle\mathbf{x}_l, e_i\rangle\big(\alpha_i^2y_l^2 - \hat{\alpha}_i^2y_l^2 - 2y_{l+1}y_l\alpha_i + 2y_{l+1}y_l\hat{\alpha}_i + 2y_l\alpha_i\gamma_i - 2\hat{\alpha}_iy_l\gamma_i\big)/2\eta_i^2\Big) \\
&= \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle\mathbf{x}_l, e_i\rangle\big(-\hat{\alpha}_i^2y_l^2 + 2y_{l+1}y_l\hat{\alpha}_i - 2\hat{\alpha}_iy_l\gamma_i\big)/2\eta_i^2\Big) + R(\alpha_i) \\
&= \sum_{i=1}^{n}\big(-T_l^i(y^2)\hat{\alpha}_i^2 + 2T_l^i(y_{l+1}, y_l)\hat{\alpha}_i - 2T_l^i(y)\gamma_i\hat{\alpha}_i\big)/2\eta_i^2 + R(\alpha_i)
\end{aligned}
$$

where $R(\alpha_i)$ is a remainder and does not contain $\hat{\alpha}$. The expectation of the log-likelihood conditional on $\mathcal{Y}_k$ is

$$
L(\hat{\alpha}) = E\Big[\log\Lambda_k^\alpha|\,\mathcal{Y}_k\Big] = \sum_{i=1}^{n}\big(-\hat{T}_l^i(y^2)\hat{\alpha}_i^2 + 2\hat{T}_l^i(y_{l+1}, y_l)\hat{\alpha}_i - 2\hat{T}_l^i(y)\gamma_i\hat{\alpha}_i\big)/2\eta_i^2 + R(\alpha_i)
$$

where $\hat{H}_l = E[H_l|\mathcal{Y}_k]$. We differentiate $L(\hat{\alpha})$ in $\hat{\alpha}_i$ and equate the result to 0. This gives us the optimal choice of the parameter $\hat{\alpha}$. In particular,

$$
(-2\hat{\alpha}_i\hat{T}_l^i(y^2) + 2\hat{T}_l^i(y_{l+1}, y_l) - 2\hat{T}_l^i(y)\gamma_i)/2\eta_i^2 = 0,
$$

and solving for $\hat{\alpha}_i$, we get

$$
\hat{\alpha}_i = \frac{\hat{T}_l^i(y_{l+1}, y_l) - \hat{T}_l^i(y)\gamma_i}{\hat{T}_l^i(y^2)}.
$$

B Optimal estimate for γ

To calculate the optimal estimate $\hat{\gamma}_i$ we consider again the following Radon-Nikodym derivative

$$\begin{aligned}\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{Y}_k} &= \Lambda_k^\gamma = \prod_{l=1}^{k} \exp\Big(\frac{1}{2\eta^2(\mathbf{x}_l)}\big(\gamma(\mathbf{x}_l)^2 - \hat{\gamma}(\mathbf{x}_l)^2 - 2y_{l+1}\gamma(\mathbf{x}_l) + 2y_{k+1}\hat{\gamma}(\mathbf{x}_l) \\ &\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\alpha(\mathbf{x}_l)y_l\hat{\gamma}(\mathbf{x}_l)\big)\Big) \ .\end{aligned}$$

Now

$$\begin{aligned}\log(\Lambda_k^\gamma) &= \sum_{l=1}^{k}\big(\gamma(\mathbf{x}_l)^2 - \hat{\gamma}(\mathbf{x}_l)^2 - 2y_{l+1}\gamma(\mathbf{x}_l) + 2y_{l+1}\hat{\gamma}(\mathbf{x}_l) \\ &\qquad + 2\alpha(\mathbf{x}_l)y_l\gamma(\mathbf{x}_l) - 2\alpha(\mathbf{x}_l)y_l\hat{\gamma}(\mathbf{x}_l)\big)/2\eta^2(\mathbf{x}_l) \\ &= \sum_{l=1}^{k}\Big(\sum_{i=1}^{n}\langle \mathbf{x}_l, e_i\rangle\big(-\hat{\gamma}_i^2 + 2y_{k+1}\hat{\gamma}_i - 2y_k\alpha_i\hat{\gamma}_i\big)/2\eta_i\Big) + R(\gamma_i)\end{aligned}$$

where $R(\gamma_i)$ is independent of $\hat{\gamma}_i$. Thus

$$L(\hat{\gamma}_i) = \sum_{i=1}^{n}\big(-\hat{O}_l^i\hat{\gamma}_i^2 + 2\hat{T}_{l+1}^i(y)\hat{\gamma}_i - 2\hat{T}_l^i(y)\alpha_i\hat{\gamma}_i\big)/2\eta_i + R(\gamma_i).$$

We differentiate $L(\hat{\gamma}_i)$ and set the derivative to 0. The optimal choice of $\hat{\gamma}_i$ is given by

$$\hat{\gamma}_i = \frac{\hat{T}_{l+1}^i(y) - \hat{T}_l^i(y)\alpha_i}{\hat{O}_l^i} \ .$$

C Optimal estimate for η

To find the optimal estimate $\hat{\eta}$, we start with the Radon-Nikodym derivative

$$\begin{aligned}\left.\frac{d\hat{P}}{dP}\right|_{\mathcal{Y}_k} = \Lambda_k^\eta &= \prod_{l=1}^{k}\frac{\eta(\mathbf{x}_l)}{\hat{\eta}(\mathbf{x}_l)}\exp\Big(\frac{1}{2\eta^2(\mathbf{x}_l)}\big(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \gamma(\mathbf{x}_l)\big)^2 \\ &\qquad - \frac{1}{2\hat{\eta}^2(\mathbf{x}_l)}\big(y_{l+1} - \alpha(\mathbf{x}_l)y_l - \gamma(\mathbf{x}_l)\big)^2\Big) \ .\end{aligned}$$

Therefore the log-likelihood is

$$\log \Lambda_k^\eta \;=\; \sum_{l=1}^{k} \Big(-\log(\hat{\eta}(\mathbf{x}_k)) - \frac{1}{2\hat{\eta}^2(\mathbf{x}_l)} (y_{l+1} - \alpha(\mathbf{x}_l) y_l - \gamma(\mathbf{x}_l))^2 \Big) + R(\eta)$$

where $R(\eta)$ is the remainder independent of $\hat{\eta}$. Hence,

$$\begin{aligned} L(\hat{\eta}) &= E\Bigg[\sum_{l=1}^{k} \sum_{i=1}^{n} \Big(-\langle \mathbf{x}_l, e_i \rangle \log \hat{\eta}_i - \frac{\langle \mathbf{x}_l, e_i \rangle}{2\hat{\eta}_i^2} \\ &\qquad * \big(y_{l+1}^2 + (\alpha_i y_l)^2 + \gamma_i^2 - 2y_{l+1}\alpha_i y_l - 2y_{l+1}\gamma_i - 2\alpha_i y_l \gamma_i \big) \Big) \mid \mathcal{Y}_k \Bigg] \\ &= \sum_{i=1}^{n} \Bigg(-\log \hat{\eta}_i \hat{(O)}_l^i - \frac{1}{2\hat{\eta}_i^2} \hat{T}_{l+1}^i(y^2) - \frac{\alpha_i^2}{2\hat{\eta}_i^2} \hat{T}_l^i(y^2) - \frac{\gamma_i^2}{2\hat{\eta}_i^2} \hat{O}_l^i \\ &\qquad + \frac{\alpha_i}{\hat{\eta}_i^2} \hat{T}_l^i(y_{l+1}, y_l) + \frac{\gamma_i}{\hat{\eta}_i^2} \hat{T}_{l+1}^i(y) + \frac{\alpha_i \gamma_i}{\hat{\eta}_i^2} \hat{T}_l^i(y) \Bigg) + R(\eta) \;. \end{aligned}$$

After differentiating $L(\hat{\eta})$ with respect to $\hat{\eta}$ and maximising $L(\hat{\eta})$ the optimal estimate for η may be shown as

$$\hat{\eta}_i = \sqrt{\frac{\hat{T}_{l+1}^i(y^2) + \alpha_i^2 \hat{T}_l^i(y^2) + \gamma_i^2 \hat{O}_l^i - 2\alpha_i \hat{T}_l^i(y_{l+1}, y_l) - 2\gamma_i \hat{T}_{l+1}^i(y) - 2\alpha_i\gamma_i \hat{T}_l^i(y)}{\hat{O}_l^i}} \quad .$$

References

[1] Armstrong J and Collopy F (1992) Error measures for generalizing about forecasting methods: empirical comparisons. International Journal of Forecasting 8: 69-80

[2] Bansal R and Zhou H (2002) Term structure of interest rate with regime shifts. Journal of Finance 57: 1997-2043

[3] Black F and Karasinski P (1991) Bond and Option Pricing when Short Rates are Log Normal. Financial Analysts Journal 47: 52-9

[4] Brigo D and Mercurio F (2001) Interest Rate Models: Theory and Practice. Springer-Verlag, Heidelberg

[5] Cox J, Ingersoll J and Ross S (1985) A theory of the term structure of interest rates. Econometrica 53: 363-384

[6] Driffill J, Kenc T and Sola M (2003) An Empirical Examination of Term Structure Models with Regime Shifts. Computing in Economics and Finance 65. Society for Computational Economics

[7] Duffie D and Kan R (1996) A Yield-Factor Model of Interest Rates. Mathematical Finance 64: 379-406

[8] Elliott R (1994) Exact adaptive filters for Markov chains observed in Gaussian noise. Automatica 30: 1399-1408

[9] Elliott R, Aggoun L and Moore J (1995) Hidden Markov Models: Estimation and Control. Springer-Verlag, New York

[10] Elliott R and Krishnamurthy V (1999) New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models. IEEE Transactions on Automatic Control 44: 938-951

[11] Elliott R, Fischer P and Platen E (1999) Filtering ans parameter estimation for a mean reverting interest rate model. Canadian Applied Mathematics Quarterly 7: 381-400

[12] Elliott R, Hunter W and Jamieson B (2001) Financial signal processing: a self calibrating model. International Journal of Theoretical and Applied Finance 4: 567-584

[13] Elliott R and Mamon R (2002) An interst rate model with a Markovian mean reverting level. Quantitative Finance 2: 454-458

[14] Elliott R, Sick G and Stein M (2003) Modelling Electricity Price Risk. Working paper. University of Calgary

[15] Evans M (2003) Real risk, inflation risk and the term structure. The Economic Journal 113: 345-389

[16] Fama E and Gibbons M (1984) A comparison of inflation forecasts. Journal of Monetary Economics 13: 327-348

[17] Garcia R and Perron P (1996) An analysis of the real interest rate under regime shifts. Review of Economics and Statistics 78: 111-125

[18] Gray S (1996) Modelling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42: 27-62

[19] Hamilton J (1989) A new approach to the economic analysis of nonstationary time series and business cycle. Econometrica 57: 357-384

[20] Heath D, Jarrow R and Morton A (1992) Bond pricing and the term structure of interest rates: a new methodology. Econometrica 60: 77-105

[21] Hull J and White A (1990) Pricing interest rate derivative securities. Review of Financial Studies 3: 573-592

[22] Landén C (2000) Bond pricing in a hidden Markov model of the short rate. Finance and Stochastics 4: 371-389

[23] Longstaff F and Schwartz E (1992) Interest volatility and the term structure: a two-factor general equilibrium model. Journal of Finance 47: 1259-1282

[24] Naik V and Lee M (1997) Yield curve dynamics with discrete shifts in economic regimes: theory and estimation. Working paper. University of British Columbia

[25] Pelsser A (2000) Efficient methods for valuing interest rate derivatives. Springer-Verlag, London

[26] Smith D (2002) Markov-switching and stochastic volatility diffusion models of short-term interest rates. Journal of Business and Economic Statistics 20: 183-197

[27] Vasicek O (1977) An equilibrium characterization of the term structure. Journal of Financial Economics 5: 177-188