Expert Systems with Applications 39 (2012) 4679–4689

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An accessible implementation of interest rate models with Markov-switching ☆

Nanxin Zhou [a], Rogemar Mamon [b,*]

[a] Funds Management and Banking Department, Bank of Canada, Ottawa, Ontario, Canada
[b] Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

ARTICLE INFO

Keywords:
Markov chain
Vasicek model
Cox–Ingersoll–Ross model
Black–Karasinski model
Quasi-maximum likelihood method
Parameter estimation
Regime-switching
Model validation

ABSTRACT

We examine the performance of interest rate models with regime-switching feature through a straightforward implementation. In particular, three short-rate models, the Vasicek, CIR and Black–Karasinski models, are extended to capture the switching of economic regimes using a finite-state Markov chain in discrete time. The Markov chain modulates the parameters of the model. We illustrate numerically that the resulting extended models are capable of reproducing various shapes of the yield curve. A quasi-maximum likelihood method based on James and Webber (2000) is employed to estimate the parameters of the regime-switching models. We demonstrate the implementation using actual financial datasets of Canadian yield rates. The numerical results show that under some model validation metrics, the two-state regime-switching models are more flexible, have better forecasting performance and provide better fit than the models without the regime-switching characteristic.

1. Introduction

The modelling of interest rate is a central problem in modern finance. As a key financial variable, interest rate needs to be taken into consideration in almost all financial transactions. It plays an important role in risk management and in the pricing and hedging of fixed income securities as well as other financial derivatives. From an investment perspective, interest rate also provides a benchmark for which other alternative investment opportunities are measured against. In the last three to four decades, researchers have developed different approaches to capture the dynamics of interest rates, including the short-rate approach, the forward rate approach, and the LIBOR market model approach.

The short-rate approach describes the evolution of interest rates by specifying the behaviour of the short rate process, r_t, which is the interest rate applied to an infinitesimally small period of time at time t. Under a short-rate model, fundamental quantities such as bond prices, forward rates and yields are readily defined as a function of the short rate process. Popular short-rate models include those pioneered by Vasicek (1977), Cox, Ingersoll, and Ross (1985), Ho and Lee (1986) and Hull and White (1990), Black and Karasinski (1991) and Black, Derman, and Toy (1990), amongst others. An important class of the short-rate models is the class of exponential affine term structure models. Under an exponential affine term structure model, the price at time t of a zero-coupon bond maturing at time T has the functional form $B(t,T) = e^{D(t,T)-A(t,T)r_t}$ for some deterministic functions A and D. The Vasicek, CIR, Ho–Lee, and Hull–White models all belong to the class of the exponential affine term structure models.

To incorporate more realistic correlation and volatility structures of interest rates, multi-factor models with more than one source of uncertainty were developed. The models previously mentioned are all examples of one-factor models since the short-rate process is driven by only one source of uncertainty. Popular multi-factor models include the Longstaff and Schwartz (1992) two-factor model and the Chen (1996) stochastic mean and stochastic volatility three-factor model, amongst others.

Despite its appealing features, the short-rate approach has its shortcomings as well. For instance, a short-rate model does not automatically provide an analytical description of the yield curve except for a few special cases. Moreover, it does not give the user complete freedom in choosing the volatility structure. To address these problems, Heath, Jarrow, and Morton (1992) (HJM) developed a framework to model interest rates based on the forward rate. They showed that under some no-arbitrage conditions, the drift term of the forward rate process can be expressed as a function of the volatility term. The major advantage of the HJM approach is that it captures the dynamics of the entire yield curve. The problem with the approach is that the short-rate process is generally non-Markovian, which makes it computationally intractable, and Monte Carlo simulation technique has to be used in order to implement the model. A special case of the HJM framework is the LIBOR market model developed by Brace, Gatarek, and

☆ The views expressed in this paper are those of the authors and do not reflect those of (or should not be attributed to) the Bank of Canada.

* Corresponding author. Address: Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada N6A 5B7.
E-mail address: rmamon@stats.uwo.ca (R. Mamon).

0957-4174/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2011.09.053

Musiela (1997). Instead of describing the instantaneous forward rates that are not readily observable in the market, the LIBOR market model considers a set of forward LIBOR rates. This model has the advantage of modelling rates that are directly observable in the market, as well as producing formulae for pricing derivatives that are consistent with market practice.

We focus on three particular short-rate models, namely the Vasicek, CIR, and the Black–Karasinski models. A common and important feature of these models is that they are able to capture the so-called mean-reverting property of interest rate processes. This property suggests that interest rate processes will eventually revert to a long-term average level. This is consistent with empirical studies on interest rates, where monetary authorities set target rates. However, a problem with these models is that the mean-reverting levels are constant. Empirical studies such as those in Mankiw and Miron (1986) have shown that the mean-reverting level of the short rate and the speed of mean reversion are different in periods of extraordinary shocks than in normal periods. Other works such as those in Stanton (1997), Conley, Hansen, Luttmer, and Scheinkman (1997) Buodoukh, Richardson, Stanton, and Whitelaw (1999) found strong nonlinearities in the drift and volatility functions of the short rate.

All of these results lead to the popularity of regime-switching models. Under a regime-switching model, the short-rate process is influenced by a state or regime variable, and exhibits different characteristics depending on which regime the process is in. Hamilton (1994) considered a regime-switching model for which the state variable is described by a finite-state Markov chain. Ang and Bekaert (2002) found that a regime-switching model can better fit empirical data of the US short rate, as well as having better out-of-sample forecasts. Some applications of regime-switching models include Naik (1993), Ang and Bekaert (2002), Buffington and Elliott (2002, 2002) and Siu, Erlwein, and Mamon (2008). Elliott and Mamon (2002, 2003) and Elliott and Siu (2009) derived formulae for bond price valuations when the state variable follows a continuous-time, finite-state Markov chain. Elliott, Fisher, and Platen (1999) developed a filtering technique for parameter estimation of regime-switching models in continuous time whereas Erlwein and Mamon (2009) constructed filters for the estimation of a Markov-switching Vasicek model in discrete time. James and Webber (2000) described the general method of moments approach as well as the method of quasi-maximum likelihood, and Hamilton (1994) described the maximum likelihood estimates and the EM algorithm to obtain numerical solutions.

In this paper, we investigate the Vasicek, CIR and Black–Karasinski models and their associated regime-switching variants. That regime-switching models could retain much of the analytical tractability of the short-rate models they derived from, whilst realistically capturing the movement of interest rates over time, is an attractive feature indeed. However, instead of attempting to derive an analytical solution for the zero-coupon bond price under the different models as in Elliott and Mamon (2002, 2003) and Elliott and Siu (2009), we aim to analyse the term structure of interest rates under regime-switching models vis-a-vis those found in practice. Next, we adopt the quasi-maximum likelihood methodology described by James and Webber (2000) in our parameter estimation using simulated and real-world data. Both the model-fitting ability and forecasting performance of several competing models with varying states are further assessed through some validation criteria.

This paper is organised as follows. In Section 2, we make use of Monte Carlo simulation to numerically solve for the price of zero-coupon bonds of different maturities; this allows us to extract the shape and trend of the entire yield curve. We shall show that regime-switching models exhibit greater flexibility in that it can better replicate different shapes of yield curves observed in real world. In Section 3, we address parameter estimation of the regime-switching models. Numerical estimates and their standard errors will be provided. Model validation is discussed in Section 4. Specifically, we shall fit the regime-switching models to real data and examine their forecasting capabilities. Finally, Section 5 concludes.

2. Yield curve simulation

In the following discussion, the speed of mean reversion, mean-reverting level and volatility of the short-rate process will be denoted by a, b and σ, respectively. When the regime-switching feature is included, these parameters a, b, and σ switch according to a discrete-time, finite-state Markov chain. We further suppose that the the Markov chain is independent from the standard Brownian motion driving the short rate process.

More formally, consider the short-rate process r_t, a finite-state Markov chain $\mathbf{x}_t$ and a standard Brownian motion W_t independent of $\mathbf{x}_t$. Recall that $\mathbf{x}_t$ is an N-state Markov chain if the probability that $\mathbf{x}_t$ equals some particular value $\mathbf{e}_j$ depends on the past only through the most recent observation $\mathbf{x}_{t-1}$. That is,

$$P(\mathbf{x}_t = \mathbf{e}_j | \mathbf{x}_{t-1} = \mathbf{e}_i, \mathbf{x}_{t-2} = \mathbf{e}_h, \ldots) = P(\mathbf{x}_t = \mathbf{e}_j | \mathbf{x}_{t-1} = \mathbf{e}_i) = p_{ij}, \quad (1)$$

and $\mathbf{x}_t$ is governed by an $N \times N$ transition probability matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \vdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}$$

with the property that $p_{i1} + p_{i2} + \cdots + p_{iN} = 1$ for $i = 1,2,\ldots,N$. Here, $\mathbf{e}_j$ represents the jth canonical basis vector of $\mathbb{R}^N$. In other words, $\mathbf{e}_j$ is an $N \times 1$ vector whose jth element equals to unity and is zero elsewhere.

2.1. The three models

Under the Vasicek model, the short rate process is described by the stochastic differential equation (SDE)

$$dr_t = a(\mathbf{x}_t)(b(\mathbf{x}_t) - r_t)dt + \sigma(\mathbf{x}_t)dW_t, \quad (2)$$

and under the CIR model, we have the SDE

$$dr_t = a(\mathbf{x}_t)(b(\mathbf{x}_t) - r_t)dt + \sqrt{r_t}\sigma(\mathbf{x}_t)dW_t. \quad (3)$$

For the Black–Karasinski model, we study the modified version that was given in Cairns (2004), i.e., the short rate dynamics are specified by the SDE

$$d\ln(r_t) = a(\mathbf{x}_t)(\ln b(\mathbf{x}_t) - \ln r_t)dt + \sigma(\mathbf{x}_t)r_t dW_t. \quad (4)$$

For each of the three models in Eqs. (2)–(4), we examine the yield curves obtained from the two-state and three-state regime-switching modelling set-up, and compare them with the yield curves obtained under the non-regime-switching model.

2.2. Vasicek model

Denote by $B(t,T)$ the time t price of a zero-coupon bond maturing at time T. Assuming that the short rate process r_t with no regime-switching is defined on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, Q)$, where Q is the risk-neutral measure, we have

$$B(t,T) = E^Q\left[\exp\left(-\int_t^T r_u du\right)\middle|\mathcal{F}_t\right]. \quad (5)$$

Under the usual Vasicek model, $B(t,T)$ can be expressed as

$$B(t,T) = e^{D(t,T) - A(t,T)r_t},$$

where

$$A(t,T)=\frac{1-e^{-a(T-t)}}{a}$$

and

$$D(t,T)=\left(b-\frac{\sigma^2}{2a^2}\right)[A(t,T)-(T-t)]-\frac{\sigma^2A(t,T)^2}{4a}.$$

Therefore, given the parameters a, b and σ, one can easily obtain the zero-coupon bond yields of different maturities as

$$Y(t,T)=-\frac{\log B(t,T)}{T-t}, \tag{6}$$

and thus, the yield curve is completely determined.

For the regime-switching Vasicek model, we make use of Monte Carlo simulation technique along with the Euler discretisation scheme to obtain zero-coupon bond prices according to formula in (5). The discretised version of the process dynamics in (2) is

$$\begin{aligned} r_{t+\Delta t} &= r_t + a(\mathbf{x}_t)(b(\mathbf{x}_t)-r_t)\Delta t + \sigma(\mathbf{x}_t)\Delta W_t \\ &= r_t + a(\mathbf{x}_t)(b(\mathbf{x}_t)-r_t)\Delta t + \sigma(\mathbf{x}_t)\epsilon_t\sqrt{\Delta t}, \end{aligned} \tag{7}$$

where $\epsilon_t \sim N(0,1)$. Note that the Euler scheme coincides with the Milstein discretisation scheme when the volatility component of the short rate process is deterministic as in the case of the Vasicek model.

2.3. CIR model

Like the Vasicek model, analytical solution exists for the zero-coupon bond price under the CIR model. It is known that

$$B(t,T)=D(t,T)e^{-A(t,T)r_t},$$

where

$$A(t,T)=\frac{2(e^{\gamma(T-t)}-1)}{(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma},$$

$$D(t,T)=\left[\frac{2\gamma e^{(a+\gamma)(T-t)/2}}{(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma}\right]^{\frac{2ab}{\sigma^2}}$$

and

$$\gamma=\sqrt{a^2+2\sigma^2}.$$

Given the parameters a, b and σ, yields of different maturities can be obtained using Eq. (6).

For the regime-switching CIR model, the same procedure used for the Vasicek model can be applied. However, in this case the discretised process under the Milstein scheme based on (3) becomes

$$\begin{aligned} r_{t+\Delta t} &= r_t + a(\mathbf{x}_t)(b(\mathbf{x}_t)-r_t)\Delta t + \sigma(\mathbf{x}_t)\sqrt{r_t}\Delta W_t \\ &= r_t + a(\mathbf{x}_t)(b(\mathbf{x}_t)-r_t)\Delta t + \sigma(\mathbf{x}_t)\sqrt{r_t}\epsilon_t\sqrt{\Delta t} + \frac{1}{4}\sigma^2(\mathbf{x}_t) \\ &\quad \times(\epsilon_t^2\Delta t-\Delta t), \end{aligned} \tag{8}$$

where $\epsilon_t \sim N(0,1)$.

2.4. Black–Karasinski model

Unlike the Vasicek and the CIR model, no analytical solution exists for the zero-coupon bond price for the Black–Karasinski model. As a result, Monte Carlo technique has to be used in the implementation of both the no-switching and the regime-switching models. Under the Milstein scheme, the one-state Black–Karasinski model has the discretised process given by

$$\begin{aligned} \ln r_{t+\Delta t} &= \ln r_t + a(\ln b-\ln r_t)\Delta t + \sigma r_t\Delta W_t = \ln r_t \\ &\quad + a(\ln b-\ln r_t)\Delta t + \sigma r_t\epsilon_t\sqrt{\Delta t} + \frac{1}{2}\sigma^2 r_t(\epsilon_t^2\Delta t-\Delta t), \end{aligned} \tag{9}$$

where $\epsilon_t \sim N(0,1)$.

For the regime-switching model, with the aid of (4), one gets the discretisation

$$\begin{aligned} \ln r_{t+\Delta t} &= \ln r_t + a(\mathbf{x}_t)(\ln b(\mathbf{x}_t)-\ln r_t)\Delta t + \sigma(\mathbf{x}_t)r_t\Delta W_t \\ &= \ln r_t + a(\mathbf{x}_t)(\ln b(\mathbf{x}_t)-\ln r_t)\Delta t + \sigma(\mathbf{x}_t)r_t\epsilon_t\sqrt{\Delta t} \\ &\quad + \frac{1}{2}\sigma^2(\mathbf{x}_t)r_t(\epsilon_t^2\Delta t-\Delta t), \end{aligned} \tag{10}$$

where $\epsilon_t \in N(0,1)$.

We present simulated yield curves under the Vasicek, CIR, and Black–Karasinski models and their two-state and three-state variants. We computed the 1 month, 3 months, 6 months, 1 year, 2 years, 5 years, 10 years, and 30 years zero-coupon bond prices under different models, and used the statistical software R function spline ( ) to fit a curve through the points. For purpose of comparison, we plot in Fig. 1 the Canadian yield curve on 02 January 1986 based on the data compiled by the Bank of Canada. We see that we have a mixed trend, i.e., a mixture of increasing and decreasing patterns across bond maturities.

For the regime-switching models, note that we employed 50,000 sample paths to obtain estimates of bond prices under the Vasicek models whilst we only used 10,000 sample paths for the CIR and Black–Karasinski models. This is because the Milstein discretisation scheme used for the Vasicek model is the same as the less accurate Euler scheme. Thus, more sample paths are needed for the Vasicek model to achieve an approximately equal order of accuracy of those results in the CIR and Black–Karasinski models. Table 1 presents the standard errors for the bond price estimates calculated using Monte Carlo simulation.

From Figs. 2–4, we observe that the one-state models are not adequate as they are only able to produce monotonic yield curves. With the addition of regime-switching into the short-rate models, we are able to observe 'humped' and mixed shapes of the yield curves that are consistent with real-world observations.

3. Parameter estimation

In this paper, we focus on recovering parameter estimates under the one-state and two-state Vasicek, CIR and Black–Karasinski models. We also perform model fitting and forecasting using real financial data set. Simulated data is utilised in this section to ensure that the generating process of the data agrees with the model being studied. This means that any discrepancy of parameter estimates is largely attributed to the model estimation technique and no longer to model misspecification. This makes the judging of the efficiency and accuracy of the parameter estimation method more clear-cut.

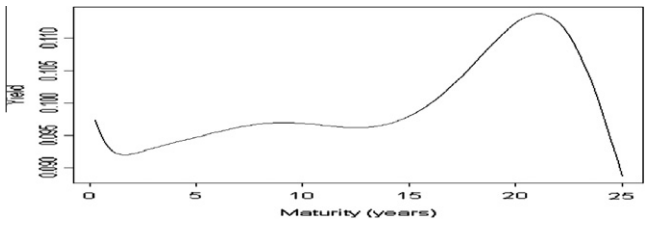


Fig. 1. Canadian yield curve on 02 January 1986 as provided by the Bank of Canada.

Table 1
Standard errors of bond price estimates.

Bond Maturity	Models						
	Black-Karasinski	Two-state Vasicek	Two-state CIR	Two-state Black–Karasinski	Three-state Vasicek	Three-state CIR	Three-state Black–Karasinski
1 month	3.94e−08	2.51e−06	3036e−06	5.73e−07	8.27e−06	3.36e−06	1.22e−06
3 month	1.46e−07	1.04e−05	1.63e−05	7.65e−06	3.58e−05	1.54e−05	1.56e−05
6 month	2.74e−07	2.49e−05	4.44e−05	3.02e−05	7.83e−05	4.05e−05	6.34e−05
1 yr	4.37e−07	5.94e−05	0.0001	9.43e−05	0.0001	0.0001	0.0203
2 yr	6.15e−07	0.0001	0.0003	0.0002	0.0002	0.0003	0.0005
5 yr	8.63e−07	0.0002	0.0005	0.0004	0.0004	0.0006	0.0007
10 yr	9.55e−07	0.0003	0.0005	0.0005	0.0004	0.0006	0.0007
30 yr	6.06e−07	0.0001	0.0002	0.0002	0.0001	0.0002	0.0002

3.1. One-state models

Exact maximum likelihood estimates of the parameters can be obtained under both the Vasicek and CIR models. For a detailed description of the maximum likelihood method, see James and Webber (2000). In this paper, we focus on the quasi-maximum likelihood method, owing to its ease of implementation. As outlined in James and Webber (2000), the quasi-maximum likelihood method makes the approximation that, if a process $dr_t = \mu(t,r_t)dt + \sigma(t,r_t)dW_t$ is discretised using the Euler approximation as

$$r_{t+\Delta t} = r_t + \mu(t,r_t)\Delta t + \sigma(t,r_t)\epsilon_t\sqrt{\Delta t}, \tag{11}$$

where $\epsilon_t \sim N(0,1)$, then $r_{t+\Delta t}$ is normal with mean $r_t + \mu(t,r_t)\Delta t$ and variance $\sigma^2(t,r_t)\Delta t$. Collect all the parameters into a vector θ. Based on series of observations $\{r_{t_i}\}$, $i = 1,\ldots,N$, where $t_{i+1} - t_i = \Delta t$ is constant, the maximum likelihood estimate $\hat{\theta}$ can be obtained by maximising the quasi-log-likelihood function

$$QL(\theta) = -\frac{N-1}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{N-1}\ln\left(\sigma^2(t_i,r_{t_i})\Delta t\right) - \frac{1}{2} \times \sum_{i=1}^{N-1}\left(\frac{r_{t_{i+1}} - r_{t_i} - \mu(t_i,r_{t_i})\Delta t}{\sigma(t_i,r_{t_i})\sqrt{\Delta t}}\right)^2. \tag{12}$$

The Hessian matrix is defined as $\mathbf{H}(\hat{\theta}) = \left(\frac{\partial^2 QL}{\partial\theta_i\theta_j}\right)$. Consequently, the standard error (SE) of $\hat{\theta}_i$ can be obtained as the square root of the (i,i)th entry of the inverse of $\mathbf{H}(\hat{\theta})$, i.e., by evaluating $\sqrt{\mathbf{H}(\hat{\theta})^{-1}_{ii}}$. This formula is employed to provide the SEs of parameter estimates in the succeeding discussion.

3.2. Vasicek model

For the Vasicek model, Eq. (11) gives $\mu(t,r_t) = a(b - r_t)$ and $\sigma(t,r_t) = \sigma$. Therefore, $r_{t+\Delta t}$ is normally distributed with mean $r_t + a(b - r_t)\Delta t$ and variance $\sigma^2\Delta t$. Thus, the quasi-log-likelihood is

$$QL_{Vasicek} = -\frac{N-1}{2}\ln 2\pi - (N-1)\ln\sigma - \frac{N-1}{2}\ln\Delta t - \frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{N-1}\left(r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t\right)^2. \tag{13}$$

To obtain the estimates $\hat{a}$, $\hat{b}$ and $\hat{\sigma}$, we take partial derivatives of the function QL with respect to a, b, and σ and set them equal to zero:

$$\frac{\partial QL_{Vasicek}}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{N-1}\left(r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t\right)(b - r_{t_i}) = 0, \tag{14}$$

$$\frac{\partial QL_{Vasicek}}{\partial b} = \frac{a}{\sigma^2}\sum_{i=1}^{N-1}\left(r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t\right) = 0, \tag{15}$$

$$\frac{\partial QL_{Vasicek}}{\partial \sigma} = -\frac{N-1}{\sigma} + \frac{1}{\sigma^3\Delta t}\sum_{i=1}^{N-1}\left(r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t\right)^2 = 0. \tag{16}$$

Solving the set of Eqs. (14)–(16) yields the Vasicek parameters

$$\hat{a} = -\frac{\sum_{i=1}^{N-1} r_{t_i}\left(\sum_{i=1}^{N-1} r_{t_i} - \sum_{i=1}^{N-1} r_{t_{i+1}}\right) + (N-1)\left(\sum_{i=1}^{N-1} r_{t_{i+1}} r_{t_i} - \sum_{i=1}^{N-1} r_{t_i}^2\right)}{\Delta t\left[(N-1)\sum_{i=1}^{N-1} r_{t_i}^2 - \left(\sum_{i=1}^{N-1} r_{t_i}\right)^2\right]} \tag{17}$$

$$\hat{b} = -\frac{-\sum_{i=1}^{N-1} r_{t_{i+1}} r_{t_i}\sum_{i=1}^{N-1} r_{t_i} + \sum_{i=1}^{N-1} r_{t_{i+1}}\sum_{i=1}^{N-1} r_{t_i}^2}{\sum_{i=1}^{N-1} r_{t_i}\left(\sum_{i=1}^{N-1} r_{t_i} - \sum_{i=1}^{N-1} r_{t_{i+1}}\right) + (N-1)\left(\sum_{i=1}^{N-1} r_{t_{i+1}} r_{t_i} - \sum_{i=1}^{N-1} r_{t_i}^2\right)} \tag{18}$$

$$\hat{\sigma} = \sqrt{\frac{1}{(N-1)\Delta t}\sum_{i=1}^{N-1}\left(r_{t_{i+1}} - r_{t_i} - \hat{a}(\hat{b} - r_{t_i})\Delta t\right)^2}. \tag{19}$$

3.3. CIR model

For the CIR model, $\mu(t,r_t) = a(b - r_t)$ and $\sigma(t,r_t) = \sigma\sqrt{r_t}$ in Eq. (11), and $r_{t+\Delta t}$ is normally distributed with mean $r_t + a(b - r_t)\Delta t$ and variance $\sigma^2 r_t\Delta t$. The quasi-log-likelihood is

$$QL_{CIR} = -\frac{N-1}{2}\ln 2\pi - (N-1)\ln\sigma - \frac{N-1}{2}\ln\Delta t - \frac{1}{2}\sum_{i=1}^{N-1}\ln r_{t_i} - \frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{N-1}\left(\frac{r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t}{\sqrt{r_{t_i}}}\right)^2. \tag{20}$$

Taking the partial derivatives, we have

$$\frac{\partial QL_{CIR}}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{N-1}\frac{r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t}{r_{t_i}}(b - r_{t_i}) \tag{21}$$

$$\frac{\partial QL_{CIR}}{\partial b} = \frac{a}{\sigma^2}\sum_{i=1}^{N-1}\frac{r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t}{r_{t_i}} \tag{22}$$

$$\frac{\partial QL_{CIR}}{\partial \sigma} = \frac{1}{\sigma^3\Delta t}\sum_{i=1}^{N-1}\left(\frac{r_{t_{i+1}} - r_{t_i} - a(b - r_{t_i})\Delta t}{\sqrt{r_{t_i}}}\right)^2 - \frac{N-1}{\sigma}. \tag{23}$$

Equating to 0 each partial derivative from (21)–(23) and solving the resulting set of equations produce the CIR parameter estimates

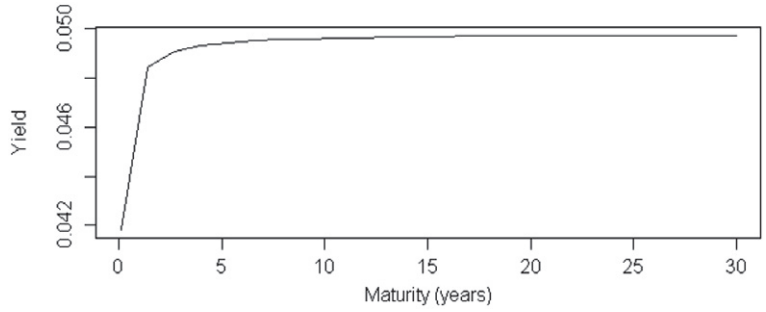
$$\hat{a} = \frac{\tau}{\Delta t\left(-(N-1)^2 + \sum_{i=1}^{N-1}\frac{1}{r_{t_i}}\sum_{i=1}^{N-1} r_{t_i}\right)}, \tag{24}$$

$$\hat{b} = \frac{\sum_{i=1}^{N-1}\frac{r_{t_{i+1}}}{r_{t_i}}\sum_{i=1}^{N-1} r_{t_i} - (N-1)\sum_{i=1}^{N-1} r_{t_{i+1}}}{\tau} \tag{25}$$

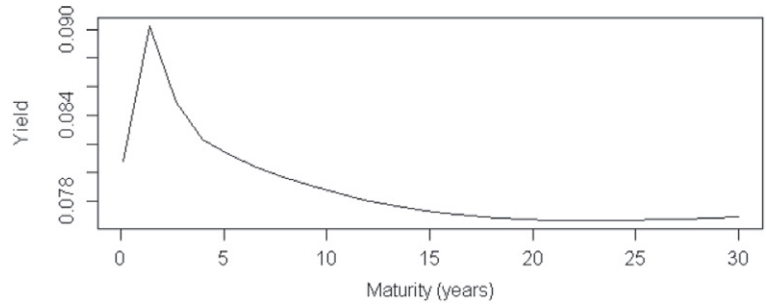
$$\hat{\sigma} = \sqrt{\frac{1}{(N-1)\Delta t}\sum_{i=1}^{N-1}\left(\frac{r_{t_{i+1}} - r_{t_i} - \hat{a}(\hat{b} - r_{t_i})\Delta t}{\sqrt{r_{t_i}}}\right)^2}, \tag{26}$$

where

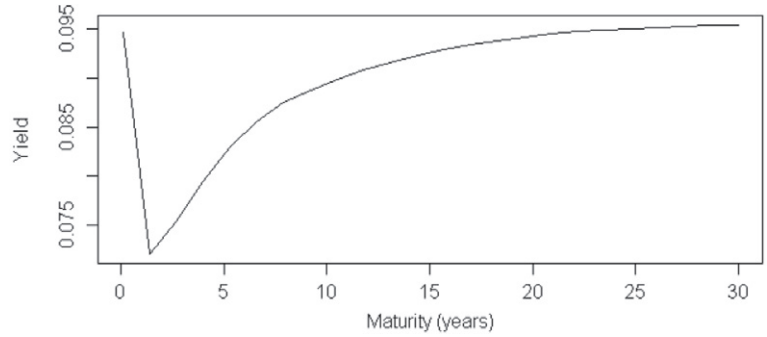
$$\tau = -(N-1)^2 + (N-1)\sum_{i=1}^{N-1}\frac{r_{t_{i+1}}}{r_{t_i}} + \sum_{i=1}^{N-1}\frac{1}{r_{t_i}}\left(\sum_{i=1}^{N-1} r_{t_i} - \sum_{i=1}^{N-1} r_{t_{i+1}}\right).$$



Vasicek yield curve: a=5, b=0.05, sigma=0.1 and r_0=0.04



Two-state Vasicek yield curve: a_1=7, a_2=3, b_1=0.1, b_2=0.05, sigma$_1$=0.05, sigma$_2$=0.10, r_0=0.075, p_{11}=0.998, p_{22}=0.998, initial state=1, r_0=0.075



Three-state Vasicek yield curve: a_1=7, a_2=5, a_3=3, b_1=0.15, b_2=0.10, b_3=0.05, sigma$_1$=0.05, sigma$_2$=0.10 and sigma$_3$=0.15, initial state=1 and r_0=0.10

The transition probability matrix for the 3-state case is $\mathbf{P} = \begin{pmatrix} 0.998 & 0.002 & 0 \\ 0.001 & 0.998 & 0.001 \\ 0 & 0.002 & 0.998 \end{pmatrix}$.

Fig. 2. Yield curves for a Vasicek model with varying number of states.

3.4. Black–Karasinski model

The Euler approximation scheme for the Black–Karasinski model is

$$\ln r_{t+\Delta t} = \ln r_t + a(\ln b - \ln r_t)\Delta t + \sigma r_t \sqrt{\Delta t}\,\epsilon_t. \tag{27}$$

This implies that $\ln r_{t+\Delta t}$ is normally distributed with mean $\ln r_t + a(\ln b - \ln r_t)\Delta t$ and variance $\sigma^2 r_t^2 \Delta t$, i.e., $r_{t+\Delta t}$ is lognormally distributed. The quasi-log-likelihood function is

$$QL_{BK} = -\frac{N-1}{2}\ln 2\pi - (N-1)\ln\sigma - \frac{N-1}{2}\ln\Delta t - \sum_{i=1}^{N-1}\ln r_{t_i} - \frac{1}{2\sigma^2\Delta t}\sum_{i=1}^{N-1}\left(\frac{\ln r_{t_{i+1}} - \ln r_{t_i} - a(\ln b - \ln r_{t_i})\Delta t}{r_{t_i}}\right)^2. \tag{28}$$

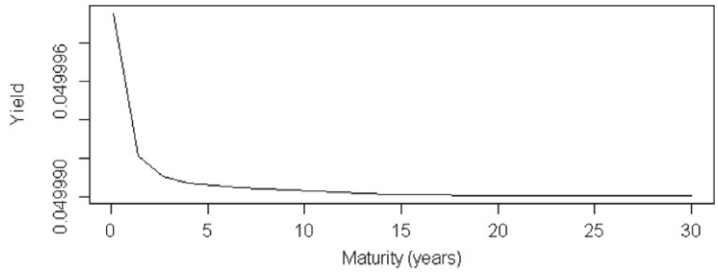
The corresponding partial derivatives are

$$\frac{\partial QL_{BK}}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{N-1}\left(\frac{\ln r_{t_{i+1}} - \ln r_{t_i} - a(\ln b - \ln r_{t_i})\Delta t}{r_{t_i}}\right) \times \left(\frac{b - \ln r_{t_i}}{r_{t_i}}\right), \tag{29}$$

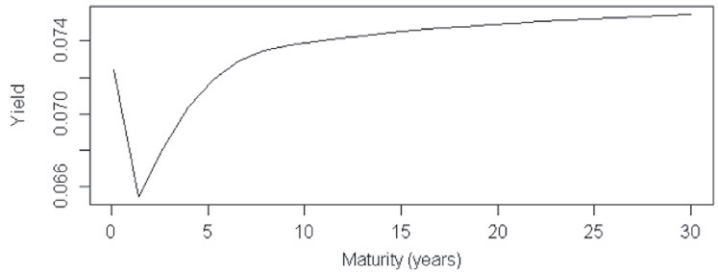
$$\frac{\partial QL_{BK}}{\partial b} = \frac{a}{b\sigma^2}\sum_{i=1}^{N-1}\frac{\ln r_{t_{i+1}} - \ln r_{t_i} - a(\ln b - \ln r_{t_i})\Delta t}{r_{t_i}^2}, \tag{30}$$

$$\frac{\partial QL_{BK}}{\partial \sigma} = \frac{1}{\sigma^3\Delta t}\sum_{i=1}^{N-1}\left(\frac{\ln r_{t_{i+1}} - \ln r_{t_i} - a(\ln b - \ln r_{t_i})\Delta t}{r_{t_i}}\right)^2 - \frac{N-1}{\sigma}. \tag{31}$$

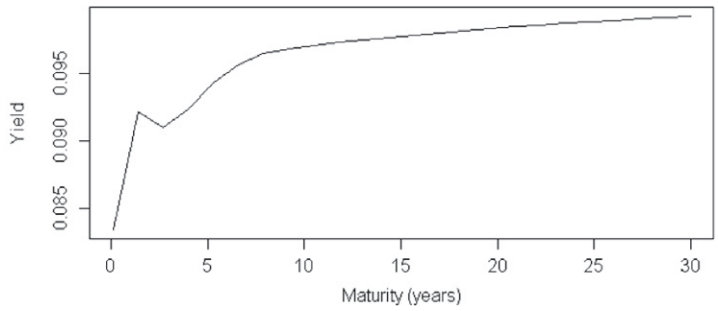
Equating Eqs. (29)–(31) each to 0 and solving yield the following solutions for the Black–Karasinski parameters:



CIR yield curve: a=5, b=0.05, sigma=0.1 and r_0=0.05



Two-state CIR yield curve: a_1=7, a_2=3, b_1=0.1, b_2=0.05, sigma$_1$=0.05, sigma$_2$=0.1, p_{11}=0.998, p_{22}=0.998, initial state =2 and r_0=0.075



Three-state CIR yield curve: a_1=7, a_2=5, a_3=3, b_1=0.15, b_2=0.10, b_3=0.05, sigma$_1$=0.05, sigma$_2$=0.10, sigma$_3$=0.15, initial state=2 and r_0=0.08

The transition probability matrix for the 3-state case is $\mathbf{P} = \begin{pmatrix} 0.998 & 0.002 & 0 \\ 0 & 0.998 & 0.002 \\ 0.002 & 0 & 0.998 \end{pmatrix}$.

Fig. 3. Yield curves for a CIR model with varying number of states.

$$\hat{a} = -\frac{\beta}{\alpha}, \tag{32}$$

$$\hat{b} = e^{\gamma/\beta}, \tag{33}$$

$$\hat{\sigma} = \sqrt{\frac{1}{(N-1)\Delta t}\sum_{i=1}^{N-1}\left(\frac{\ln r_{t_{i+1}} - \ln r_{t_i} - \hat{a}(\ln \hat{b} - r_{t_i})\Delta t}{r_{t_i}}\right)^2}, \tag{34}$$

where

$$\alpha = \Delta t\left[-\sum_{i=1}^{N-1}\frac{1}{r_{t_i}^2}\sum_{i=1}^{N-1}\frac{\ln r_{t_i}^2}{r_{t_i}^2} + \left(\sum_{i=1}^{N-1}\frac{\ln r_{t_i}}{r_{t_i}^2}\right)^2\right],$$

$$\beta = \sum_{i=1}^{N-1}\frac{\ln r_{t_i}}{r_{t_i}^2}\left(-\sum_{i=1}^{N-1}\frac{\ln r_{t_i}}{r_{t_i}^2} + \sum_{i=1}^{N-1}\frac{\ln r_{t_{i+1}}}{r_{t_i}^2}\right) - \sum_{i=1}^{N-1}\frac{1}{r_{t_i}^2}\left(\sum_{i=1}^{N-1}\frac{\ln r_{t_i}\ln r_{t_{i+1}}}{r_{t_i}^2} - \sum_{i=1}^{N-1}\frac{\ln r_{t_i}^2}{r_{t_i}^2}\right),$$

and

$$\gamma = -\sum_{i=1}^{N-1}\frac{\ln r_{t_i}}{r_{t_i}^2}\sum_{i=1}^{N-1}\frac{\ln r_{t_i}\ln r_{t_{i+1}}}{r_{t_i}^2} + \sum_{i=1}^{N-1}\frac{\ln r_{t_{i+1}}}{r_{t_i}^2}\sum_{i=1}^{N-1}\frac{\ln r_{t_i}^2}{r_{t_i}^2}.$$

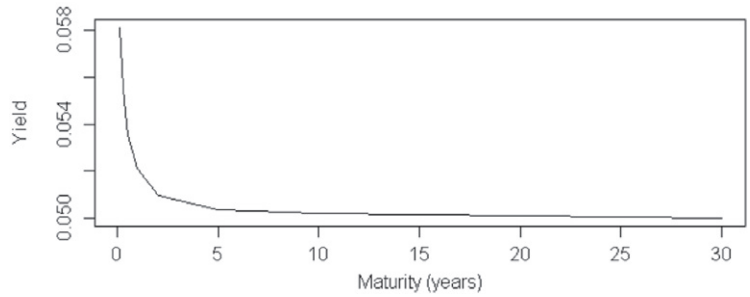
3.5. Two-state regime-switching models

For a two-state regime-switching model, the distribution of $r_{t+\Delta_t}$ given r_t is a mixture of two Gaussian distributions. Collect all the parameters into a vector $\boldsymbol{\theta}$ and let $\mathcal{F}_t$ denote the σ-algebra generated by the observations $r_t, r_{t-1}, \ldots, r_0$. Define

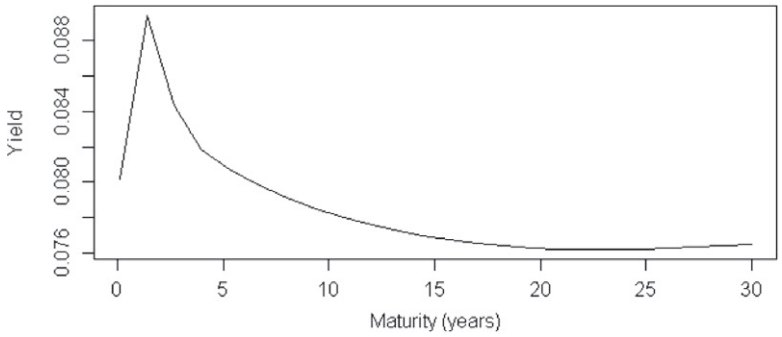
$$\boldsymbol{\eta}_t = \begin{bmatrix} f(r_{t+\Delta t}|\mathbf{x}_t = \mathbf{e}_1, \mathcal{F}_t; \boldsymbol{\theta}) \\ f(r_{t+\Delta t}|\mathbf{x}_t = \mathbf{e}_2, \mathcal{F}_t; \boldsymbol{\theta}) \end{bmatrix} \tag{35}$$

and write $\boldsymbol{\pi} := \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix}$ as the unconditional probabilities of the underlying Markov process. Then, the distribution of $r_{t+\Delta t}$ is simply

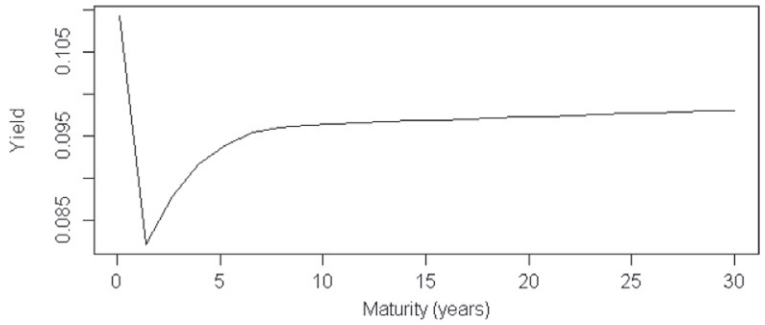
$$f(r_{t+\Delta t}|\mathcal{F}_t; \boldsymbol{\theta}) = \boldsymbol{\pi}^\top \cdot \boldsymbol{\eta}_t, \tag{36}$$



Black-Karasinski yield curve: a=5, b=0.05, sigma=0.1 and r_0=0.06



Two-state Black-Karasinski yield curve: a_1=7, a_2=3, $b1$=0.1, b_2=0.05, sigma$_1$=0.05, sigma2=0.1, $p_{11} = p_{22} = 0.998$, initial state = 1 and $r_0 = 0.075$



Three-state Black-Karasinski yield curve: a_1=7, a_2=5, a_3=3, b_1=0.15, b_2=0.10, b_3=0.05, sigma$_1$=0.05, sigma$_2$=0.10, sigma$_3$=0.15, initial state = 3 and r_0=0.08

$$\text{The transition probability matrix for the 3-state case is } \mathbf{P} = \begin{pmatrix} 0.998 & 0.002 & 0 \\ 0 & 0.998 & 0.002 \\ 0.002 & 0 & 0.998 \end{pmatrix}.$$

Fig. 4. Yield curves for a Black–Karasinski model with varying number of states.

where $\top$ denotes the transpose of a vector or matrix.

Recall that for a two-state Markov chain with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix} \tag{37}$$

and the unconditional probabilities can be obtained as

$$\boldsymbol{\pi} = \begin{bmatrix} (1-p_{22})/(2-p_{11}-p_{22}) \\ (1-p_{11})/(2-p_{11}-p_{22}) \end{bmatrix}. \tag{38}$$

Thus, there are eight parameters of interest for a two-state model, namely, a_1, a_2, b_1, b_2, σ_1, σ_2, p_{11} and p_{22}.

For a two-state Vasicek model,

$$\boldsymbol{\eta}_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\Delta t}\sigma_1} \exp\frac{\left(r_{t+\Delta_t} - r_t - a_1(b_1 - r_t)\Delta t\right)^2}{2\sigma_1^2\Delta t} \\ \frac{1}{\sqrt{2\pi\Delta t}\sigma_2} \exp\frac{\left(r_{t+\Delta_t} - r_t - a_2(b_2 - r_t)\Delta t\right)^2}{2\sigma_2^2\Delta t} \end{bmatrix}. \tag{39}$$

For a two-state CIR model, we have

$$\boldsymbol{\eta}_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\Delta t r_t}\sigma_1} \exp\frac{\left(r_{t+\Delta_t} - r_t - a_1(b_1 - r_t)\Delta t\right)^2}{2\sigma_1^2 r_t\Delta t} \\ \frac{1}{\sqrt{2\pi\Delta t r_t}\sigma_2} \exp\frac{\left(r_{t+\Delta_t} - r_t - a_2(b_2 - r_t)\Delta t\right)^2}{2\sigma_2^2 r_t\Delta t} \end{bmatrix} \tag{40}$$

and for the Black–Karasinski model, one obtains

$$\boldsymbol{\eta}_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\Delta t} r_t \sigma_1} \exp \frac{\left(\ln r_{t+\Delta_t} - \ln r_t - a_1(\ln b_1 - \ln r_t)\Delta t\right)^2}{2\sigma_1^2 r_t^2 \Delta t} \\ \frac{1}{\sqrt{2\pi\Delta t} r_t \sigma_2} \exp \frac{\left(\ln r_{t+\Delta_t} - \ln r_t - a_2(\ln b_2 - \ln r_t)\Delta t\right)^2}{2\sigma_2^2 r_t^2 \Delta t} \end{bmatrix}. \quad (41)$$

The maximum likelihood estimate can be obtained by maximising

$$l(\boldsymbol{\theta}) = \sum_{t=0}^{n-1} \log f(r_{t+\Delta t} | \mathcal{F}_t; \boldsymbol{\theta}), \quad (42)$$

where $f(r_{t+\Delta t}|\mathcal{F}_t;\boldsymbol{\theta})$ is specified by Eq. (36).

3.6. Estimation using simulated data

We wish to emphasise again that the use of simulated data to estimate parameters is motivated by the fact that we wish to analyse the extent of the discrepancy between true and estimated parameter values coming only from the performance of the estimation approach, and not from the generating process of the data.

3.7. One-state models

Here, we generate five years of simulated data (1260 data points assuming 252 days in a year) under each of the three one-state models. We update the parameter estimates every month (using a window of 21 data points) by applying the estimation formulae provided in subSection 3.1. The running estimates of the parameters for each of the three models are depicted in Figs. 5–7.

As we observe from Figs. 5–7, the estimates for the speed of mean reversion a vary widely and are the most unstable compared to the estimated values of the two other parameters. The estimates for the mean reversion level b and the volatility σ exhibit more steady patterns and they vary within a much tighter range.

3.8. Two-state models

We also generate five years of data under each of the three two-state models. However, due to complexity caused by the increased number of parameters involved, we only estimated the parameters using the entire dataset and provided their standard errors. Whilst no running estimates are calculated in this setting, the parameter update methodology in the one-state setting can be extended definitely to the two- or three-state settings in a straightforward manner. The function optim ( ) in the statistical software R was used to carry out the computations.

From Tables 2–4, we can see that the method of quasi-maximum likelihood applied to regime-switching framework produces reasonably accurate parameter estimates and can be used as a much simpler alternative to other estimation methods such as the filtering technique proposed in Erlwein and Mamon (2009) or Elliott et al. (1999), which often employ the heavy machinery of measure-theoretic probability. Note that the estimates for a_1 and a_2 for all three models are the least accurate compared to the estimated values of other parameters and have much larger standard errors. This result is consistent with the result obtained under the one-state models, where running windows of 21 data points are employed.

When performing the numerical optimisation in the two-state parameter estimation, we observed that the choice of initial values can have a great impact on the results obtained. This is due to the fact that if the initial values are far away from the true values, the numerical algorithm may not be able to find the maximiser. To choose, our initial value, we perform an evaluation of some 100 starting random values and choose the initial value that yields the highest value for the log-likelihood function.

4. Model validation

4.1. Model fitting and forecasting using real data

We fit the one-state and two-state models using the three-month zero-coupon yields from January 1986 to December 1995 compiled by the Bank of Canada. Fig. 8 shows a plot of the data. Taking a quick look at the data, it is apparent that switching of regimes occur during the time period considered.

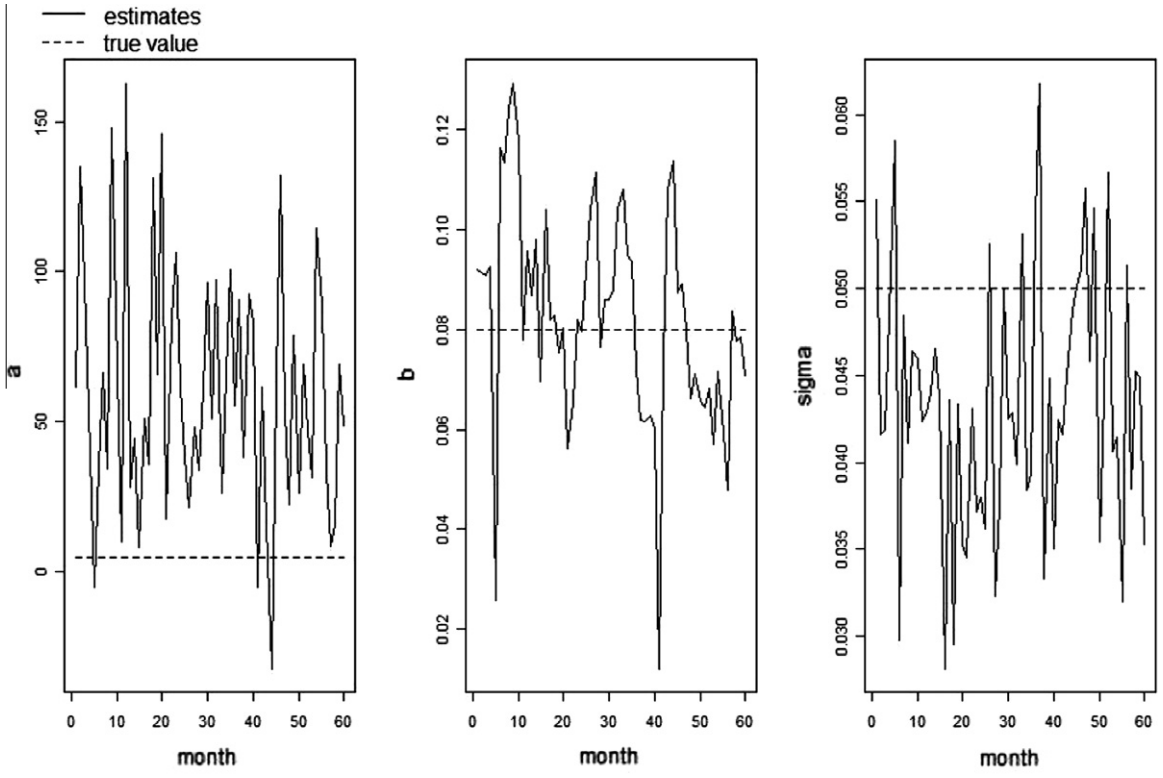


Fig. 5. Evolution of parameter estimates under the Vasicek model with $a = 5$, $b = 0.08$, $\sigma = 0.05$, and $r_0 = 0.10$.

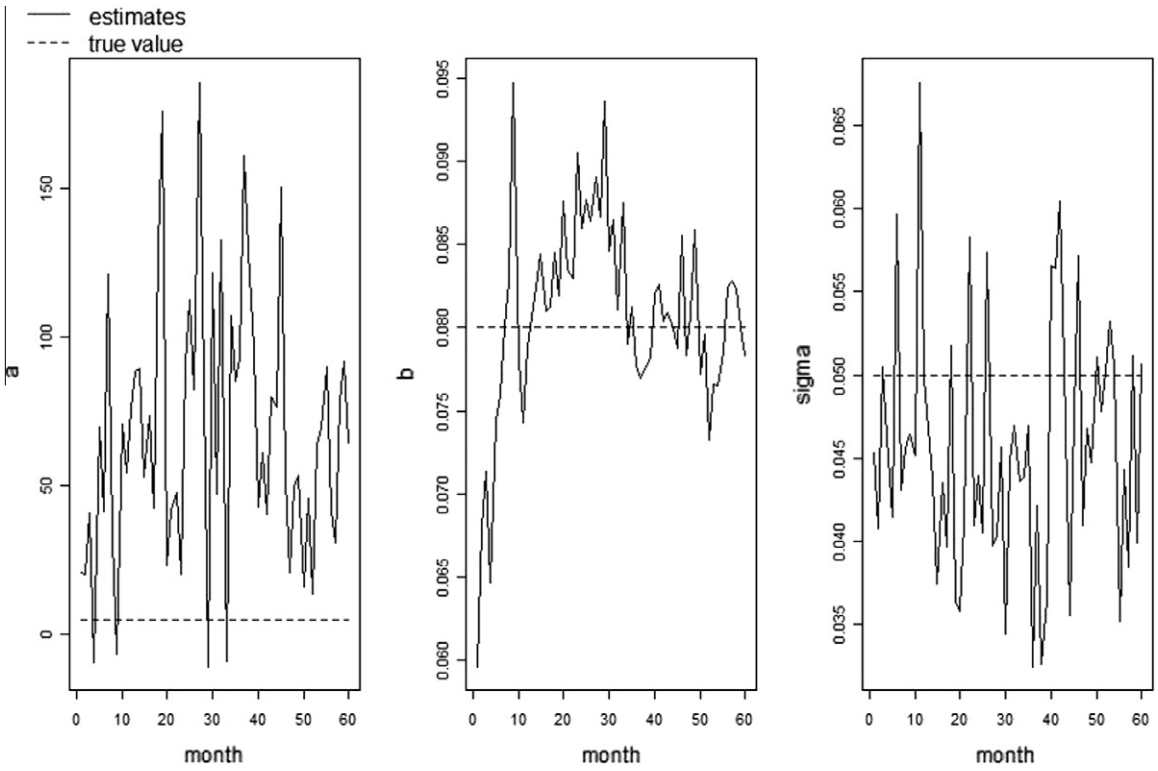


Fig. 6. Evolution of parameter estimates under the CIR model with $a = 5$, $b = 0.08$, $\sigma = 0.05$, and $r_0 = 0.05$.

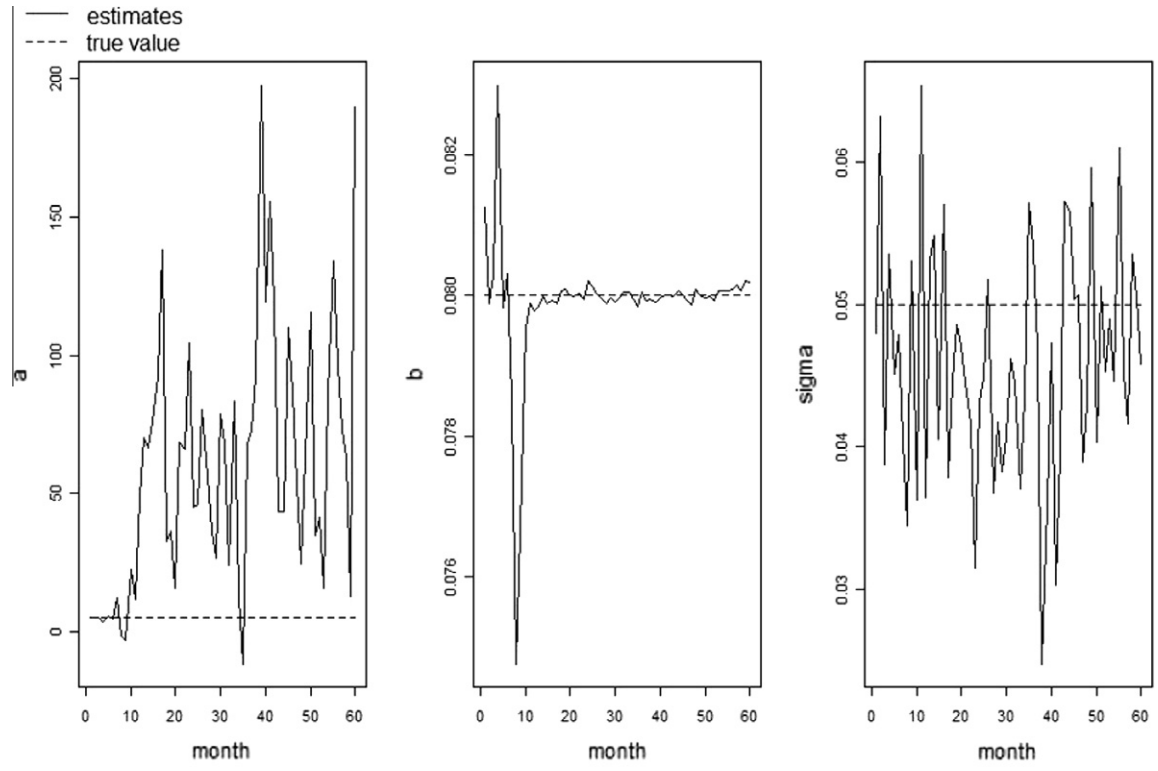


Fig. 7. Evolution of parameter estimates under the Black–Karasinski model with $a = 5$, $b = 0.08$, $\sigma = 0.05$, and $r_0 = 0.05$.

Hence, we expect that regime-switching models to perform better in capturing the dynamics of the data. To test our hypothesis, we compared the fit of the regime-switching models by calculating the Akaike Information Criterion (AIC). We also tested their forecasting performance by computing the mean squared error (MSE) between the one-step ahead predictions and actual observed values.

Table 2
Parameter estimation for a two-state Vasicek model.

Parameter	True value	Estimate	Standard error
a_1	6.00	4.9204	1.9368
a_2	2.00	1.4123	3.2822
b_1	0.10	0.0933	0.0111
b_2	0.05	0.0320	0.1400
σ_1	0.05	0.0526	0.0049
σ_2	0.10	0.1026	0.0071
p_{11}	0.95	0.9513	0.1523
p_{22}	0.95	0.9365	0.1980

Table 3
Parameter estimation for a two-state CIR model.

Parameter	True value	Estimate	Standard error
a_1	6.00	6.0321	3.7708
a_2	2.00	2.0766	4.3387
b_1	0.10	0.1247	0.0681
b_2	0.05	0.0742	0.0055
σ_1	0.05	0.0555	0.0085
σ_2	0.10	0.1029	0.0129
p_{11}	0.95	0.9472	0.2383
p_{22}	0.95	0.9406	0.2327

Table 4
Parameter estimation for a two-state Black–Karasinski model.

Parameter	True value	Estimate	Standard error
a_1	6.00	6.0556	0.0300
a_2	2.00	1.9891	0.0451
b_1	0.10	0.0998	0.0002
b_2	0.05	0.0500	0.0005
σ_1	0.05	0.0508	0.0016
σ_2	0.10	0.0994	0.0026
p_{11}	0.95	0.9387	0.0728
p_{22}	0.95	0.9561	0.0520

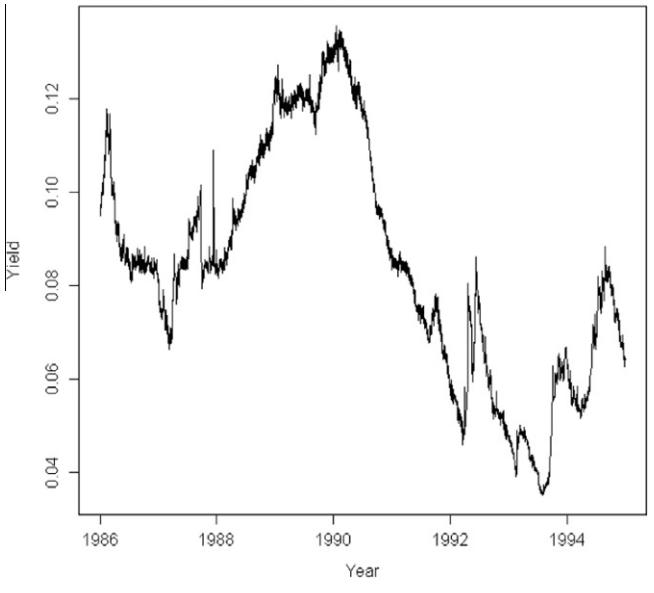


Fig. 8. Bank of Canada's three-month zero-coupon yields from January 1986 to December 1995.

4.2. Model fitting

Denote by $\hat{\theta} = \text{argmax}_\theta l(\theta)$ the quasi-maximum-likelihood estimates. We compute the AIC as $-2l(\hat{\theta}) + 2k$, where k is the number of parameters in the model. Based on the AIC numbers shown in Table 5, we see that the regime-switching models provide better fit than their one-state counterparts, and we conclude that the two-state CIR model provides the best fit to the data.

Table 5
AIC comparison.

Model	Log-likelihood	k	AIC
Vasicek	10966.88	3	−21297.76
CIR	11006.70	3	−22007.41
Black–Karasinski	4791.98	3	−9577.95
Two-state Vasicek	11227.31	8	−22438.62
Two-state CIR	11287.96	8	−22559.92
Two-state Black–Karasinski	5316.38	8	−10616.76

4.3. Forecasting

Given observations $r_0, r_1, \ldots, r_n$, the mean-squared error (MSE) for a given model can be computed as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{r}_i - r_i)^2, \tag{43}$$

where $\hat{r}_t$ is the predicted value of r_t given all the information available up to time $t-1$.

Recall that $\hat{r}_{t+\Delta t} = E[r_{t+\Delta t}|\mathcal{F}_t]$ is a function of r_t and the current estimates of the parameters. For both the Vasicek and CIR model, the distribution of $r_{t+\Delta t}$ is normal with mean $r_t + a(b - r_t)\Delta t$. Therefore,

$$E[r_{t+\Delta t}|\mathcal{F}_t] = r_t + \hat{a}(\hat{b} - r_t)\Delta t. \tag{44}$$

For the Black–Karasinski model, the distribution of $r_{t+\Delta t}$ is lognormally distributed, and so

$$E[r_{t+\Delta t}|\mathcal{F}_t] = \exp\left(\ln r_t + \hat{a}(\ln \hat{b} - \ln r_t)\Delta t + \hat{\sigma}^2 r_t^2 \Delta t/2\right). \tag{45}$$

For the two-state models, the predicted value can be obtained as

$$E[r_{t+\Delta t}|\mathcal{F}_t] = \pi_1 \cdot E[r_{t+\Delta t}|\mathbf{x}_t = \mathbf{e}_1, \mathcal{F}_t] + \pi_2 \cdot E[r_{t+\Delta t}|\mathbf{x}_t = \mathbf{e}_2, \mathcal{F}_t]. \tag{46}$$

In Table 6, we display the MSE statistics for different models. For simplicity and illustrative purpose only, we updated the parameter estimates every year in the calculation. The parameter updates can certainly be set to other frequency (e.g., semi-annually, monthly, etc.). We use the data during the first year as a training dataset to learn the parameters of the model. So, no predictions were made for the first year of data as no previous estimates were available and we require a full year of data to obtain our first set of estimates. Therefore, starting from the second year, we performed the one-step prediction using the estimates obtained using previous year's data. For the two-state models, previous estimates are used as the initial values for the new estimate inputs to the statistical software R function optim ( ). As we can see from Table 6, the MSEs for the two-state models are smaller than those obtained from their one-state counterpart, indicating better forecasting performance for the regime-switching model. On the basis of MSE criterion, the

Table 6
MSE comparison.

Model	MSE
Vasicek	4.4499e−06
CIR	4.4040e−06
Black–Karasinski	4.2755e−06
Two-state Vasicek	3.7133e−06
Two-state CIR	3.7432e−06
Two-state Black–Karasinski	3.8338e−06

two-state Vasicek model best fits the dataset. But, we note that the MSEs for all the models are too close to each other and therefore, this result may not be conclusive. Other models examined here could outdo the two-state Vasicek if a different period of the dataset is examined.

5. Conclusion

It is apparent that regime-switching models allow greater modelling flexibility whilst realistically capturing the switching of economic regimes. For instance, we were able to reproduce 'humped' and other irregularly shaped yield curves that are much more consistent with real world observations under regime-switching models. The one-factor short rate models were only able to produce yield curves that are monotone in nature.

We also illustrated that the method of quasi-maximum likelihood along with the running window parameter update approach can be utilised as a simpler alternative to parameter estimation by filtering technique when implementing the regime-switching models. With the proper choice of initial values when performing the numerical optimisation, the method produces accurate results and is fairly robust. When implemented to real financial data set, we have evidence to suggest that regime-switching models provided improved forecasting performance and better fit than those models without regime-switching.

Some further directions can be explored on the basis of this work. For instance, variance reduction techniques can be employed to accelerate the yield curve simulation process. Undoubtedly, significant progress, in terms of speed and efficiency of the proposed estimation approach, can be achieved if one can come up with a systematic way of specifying the appropriate initial values given a set of data for the optimisation of the likelihood function.

Acknowledgements

Rogemar Mamon thanks the Natural Sciences and Engineering Research Council of Canada for the support of his research.

References

Ang, A., & Bekaert, G. (2002). Short rate nonlinearities and regime switches. *Journal of Economic Dynamics and Control, 26*(7-8), 1243–1274.

Ang, A., & Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies, 15*(4), 1137–1187.

Black, F., Derman, E., & Toy, W. (1990). A one-factor model of interest rates and its applications to treasury bond options. *Financial Analysts Journal, 46*(1), 33–39.

Black, F., & Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal, 47*(4), 52–59.

Brace, A., Gatarek, D., & Musiela, M. (1997). The Market model of interest rate dynamics. *Mathematical Finance, 7*(2), 127–154.

Buffington, J., & Elliott, R. J. (2002). Regime switching and European options. In K. S. Lawrence (Ed.), *Stochastic theory and control. Proceedings of a workshop* (pp. 73–81). Berlin: Springer.

Buffington, J., & Elliott, R. J. (2002). American options with regime switching. *International Journal of Theoretical and Applied Finance, 5*(5), 497–514.

Buodoukh, J., Richardson, M., Stanton, R. & Whitelaw, R. (1999). A multifactor, nonlinear, continuous-time model of interest rate volatility. NBER Working Paper 7213.

Cairns, A. (2004). *Interest rate models: An introduction*. Princeton: Princeton University Press.

Chen, L. (1996). Stochastic mean and stochastic volatility – A three-factor model of the term structure of interest rates and its applications to the pricing of interest rate derivatives. *Financial Markets, Institutions and Instruments, 5*, 1–88.

Conley, T., Hansen, L., Luttmer, E., & Scheinkman, J. (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies, 10*(3), 525–577.

Cox, J., Ingersoll, J., & Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica, 53*(2), 385–407.

Elliott, R., Fisher, P., & Platen, E. (1999). Filtering and parameter estimation for a mean reverting interest rate model. *Canadian Applied Math Quarterly, 7*, 381–400.

Elliott, R., & Mamon, R. (2002). An interest rate model with a Markovian mean-reverting level. *Quantitative Finance, 2*(6), 454–458.

Elliott, R., & Mamon, R. (2003). A complete yield curve description of a Markov interest rate model. *International Journal of Theoretical and Applied Finance, 6*(4), 317–326.

Elliott, R., & Siu, T. (2009). On Markov-modulated exponential-affine bond price formulae. *Applied Mathematical Finance, 16*(1), 1–15.

Erlwein, C., & Mamon, R. (2009). An on-line estimation scheme for a Hull–White model with HMM-driven parameters. *Statistical Methods and Applications, 18*(1), 87–107.

Hamilton, J. (1994). *Time series analysis*. Princeton: Princeton University Press.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica, 60*(1), 77–105.

Ho, T., & Lee, S. (1986). Term structure movements and the pricing of interest rate contingent claims. *The Journal of Finance, 41*(5), 1011–1029.

Hull, J., & White, A. (1990). Pricing interest rate derivatives. *Review of Financial Studies, 3*(4), 573–592.

James, J., & Webber, N. (2000). *Interest rate modeling*. Chichester: John Wiley & Sons.

Longstaff, F., & Schwartz, E. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance, 47*(4), 1259–1282.

Mankiw, N., & Miron, J. (1986). The changing behavior of the term structure of interest rates. *Quarterly Journal of Economics, 101*(2), 211–228.

Naik, V. (1993). Option valuation and hedging strategies with jumps in the volatility of asset returns. *Journal of Finance, 48*(2), 1969–1984.

Siu, T., Erlwein, C., & Mamon, R. (2008). The pricing of credit default swaps under a Markov-modulated Merton's structural model. *North American Actuarial Journal, 12*(1), 19–46.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance, 52*(5), 1973–2002.

Vasicek, O. A. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics, 5*(2), 177–188.