APPLICATION OF MAXIMUM LIKELIHOOD ESTIMATION TO STOCHASTIC SHORT RATE MODELS

K. FERGUSSON AND E. PLATEN

ABSTRACT. The application of maximum likelihood estimation is not well studied for stochastic short rate models because of the cumbersome detail of this approach. We investigate the applicability of maximum likelihood estimation to stochastic short rate models. We restrict our consideration to three important short rate models, namely the Vasicek, Cox-Ingersoll-Ross and 3/2 short rate models, each having a closed-form formula for the transition density function. The parameters of the three interest rate models are fitted to US cash rates and are found to be consistent with market assessments.

1. INTRODUCTION

A short rate model is a mathematical model of the instantaneous, continuously compounded deposit rate in a specific currency. The most realistic proxy for the short rate among investible securities is probably the overnight cash deposit rate, expressed as a continuously compounded rate.

Short rates are typically modelled as stochastic processes and coverages of short rate models can be found, for example, in Rebonato [1998] and Brigo and Mercurio [2006]. Previous work on testing models of the short rate has been done by Aït-Sahalia [1996] on seven-day Euro dollar deposit rates from 1st June 1973 to 25th February 1995, where the parameter estimates are obtained by minimising the distance between the parametric density and the nonparametric density. Accurate closed-form approximations to the transition density function of an arbitrary diffusion is described by Aït-Sahalia [1999]. The use of Gaussian estimators has been proposed by Yu and Phillips [2001] as an improvement upon the Euler approximation schemes of Chan et al. [1992]. The use of quasi-maximum likelihood estimation has been described in Treepongkaruna [2003], whereby an approximation to the maximum likelihood estimates is obtained by maximising a function that is related to the logarithm of the likelihood function, but is not equal to it. Other approaches, such as Faff and Gray [2006], employ generalised method of moments to estimate the parameters. Our approach is to find parameters which maximise the likelihood. The short rate models considered in this article are specified by SDEs with one noise source and with constant coefficients. They are chosen because they have explicit formulae for their transition density functions, leading to original proofs of the closed-form formulae of parameter estimates and standard errors for the Vasicek model.

TABLE 1. Stochastic differential equations of short rate models.

Short Rate Model	Dynamics
Vasicek	$dr_t = \kappa(\bar{r} - r_t)dt + \sigma dZ_t$
CIR	$dr_t = \kappa(\bar{r} - r_t)dt + \sigma\sqrt{r_t}dZ_t$
3/2-Model	$dr_t = (pr_t + qr_t^2)dt + \sigma r_t^{3/2}dZ_t$

The three models of the short rate considered in this article are the Vasicek, Cox-Ingersoll-Ross and 3/2 models. The SDE of the short rate of each model is shown in Table 1.

The Vasicek model is a linear mean reverting stochastic model which ensures that interest rates adhere to a long run reference level. It allows for negative interest rates.

The CIR model is a linear mean reverting stochastic model, which avoids the possibility of negative interest rates experienced in the Vasicek model.

Finally we examine the 3/2 model which also prohibits negative interest rates but is not linear mean reverting. As we will see, its inverse is linear mean reverting and, thus, adheres to a long run reference level.

The availability of explicit formulae for the transition density functions makes possible the fitting of each model using maximum likelihood estimation. For estimating the drift parameters the length of the observation window is crucial. Therefore, we fit each model to Shiller's annual data set comprised of US one-year rates from 1871 to 2012, given by Shiller [1989]. The use of the one-year deposit rate as a proxy for the short rate is an assumption that is made here. The magnitude of the biases of a short term deposit rate in lieu of the unobservable short rate was investigated in Chapman et al. [1999], where it was found not to be economically significant.

2. REVIEW OF MODELS

We review each of the three short rate models, showing the transition density function.

2.1. **Vasicek Short Rate Model.** The Vasicek model was proposed in Vasicek [1977], whereby the short rate is described by the SDE

$$(2.1) \qquad dr_t = \kappa(\bar{r} - r_t)dt + \sigma dZ_t$$

for positive constants $\bar{r}$ and σ and κ. The parameter κ denotes the speed of reversion of the short rate r_t to the mean reverting level $\bar{r}$. The parameter $\bar{r}$ denotes the average short rate. The parameter σ is the instantaneous volatility of the short rate. This model is a particular case of the Hull-White model whose drift and diffusion parameters are made time dependent. Since we have long-dated contracts in our focus we concentrate on using constant parameters.

This SDE (2.1) is the Ornstein-Uhlenbeck SDE whose explicit solution is obtained by solving the SDE of $q_t = r_t \exp(\kappa t)$

$$(2.2) \qquad dq_t = d(r_t \exp(\kappa t)) = \kappa \exp(\kappa t)\bar{r}dt + \exp(\kappa t)\sigma dZ_t.$$

Vasicek's model was the first one to capture mean reversion, an essential characteristic of the interest rate that sets it apart from simpler models. Thus, as opposed

to stock prices, for instance, interest rates are not expected to rise indefinitely. This is because at very high levels they would hamper economic activity, prompting a decrease in interest rates. Similarly, interest rates are unlikely to decrease indefinitely. As a result, interest rates move mainly in a range, showing a tendency to revert to a long run value.

The drift factor $\kappa(\bar{r} - r_t)$ represents the expected instantaneous change in the interest rate at time t. The parameter $\bar{r}$ represents the long run reference value towards which the interest rate reverts. Indeed, in the absence of uncertainty, the interest rate would remain constant when it has reached $r_t = \bar{r}$. The parameter κ, governing the speed of adjustment, needs to be positive to ensure stability around the long term value. For example, when r_t is below $\bar{r}$, the drift term $\kappa(\bar{r} - r_t)$ becomes positive for positive κ, generating a tendency for the interest rate to move upwards.

The main disadvantage seems that, under Vasicek's model, it is theoretically possible for the interest rate to become negative. In the academic literature this has been interpreted as an undesirable feature. However, on several occasions the market generated in recent years some slightly negative interest rates, for example in Switzerland and in Europe. The possiblity of negative interest rates is excluded in the Cox-Ingersoll-Ross model (discussed in Section 2.2), the exponential Vasicek model, see Brigo and Mercurio [2001], the model of Black et al. [1990] and the model of Black and Karasinski [1991], among many others, see Brigo and Mercurio [2006] for further discussions. The Vasicek model was further extended in the Hull-White model, see Hull and White [1990], by allowing time dependence in the drift parameters.

Being an Ornstein-Uhlenbeck process, the short rate r_t satisfying the SDE (2.1) has solution

$$(2.3) \quad r_t = r_s \exp(-\kappa(t - s)) + \bar{r}(1 - \exp(-\kappa(t - s))) + \sigma \int_s^t \exp(-\kappa(t - u))dZ_u$$

for times s and t with $0 \leq s < t$ and for positive constants $\bar{r}$, κ and σ. Here Z is the Wiener process in (2.1).

As is the case for the Ho-Lee and Hull-White models, the transition density function of the Vasicek short rate is that of a normal distribution.

For times s and t with $0 \leq s < t \leq T$ the transition density of the short rate r_t in (2.1) is given by

(2.4)

$$p_r(s, r_s, t, r_t) = \frac{1}{\sqrt{2\pi\sigma^2 \frac{1-\exp(-2\kappa(t-s))}{2\kappa}}}$$

$$\times \exp\left(-\frac{1}{2}\left(\frac{r_t - r_s \exp(-\kappa(t - s)) - \bar{r}(1 - \exp(-\kappa(t - s)))}{\sqrt{\sigma^2 \frac{1-\exp(-2\kappa(t-s))}{2\kappa}}}\right)^2\right).$$

As for other Gaussian short rate models such as the Ho-Lee model and the Hull-White model, a potential disadvantage of the Vasicek model is the possibility of negative interest rates.

2.2. **Cox-Ingersoll-Ross Short Rate Model.** The Cox-Ingersoll-Ross (CIR) model was introduced in 1985 by Cox et al. [1985] as an alternative to the Vasicek model. A good explanation of the model is given in Hull [1997]. The short rate is described by the SDE

$$(2.5) \qquad dr_t = \kappa(\bar{r} - r_t)dt + \sigma\sqrt{r_t}dZ_t$$

for positively valued constants $\bar{r}$, σ and κ. A zero valued short rate is avoided because we set $\kappa\bar{r} > \frac{1}{2}\sigma^2$. The parameter κ denotes the speed of reversion of the short rate r_t to the mean reverting level $\bar{r}$. As in the Ho-Lee, Hull-White and Vasicek models, $\bar{r}$ can be thought of as a smoothed average short rate which is targeted by the central bank.

The CIR model has the advantage over the Ho-Lee, Hull-White and Vasicek models that for the above conditions on the parameters the interest rates can never be negative.

We can remove the occurrence of r_t in the drift term of (2.5) by means of the integrating factor $\exp(\kappa t)$, giving the SDE for $q_t = r_t\exp(\kappa t)$ in the form

$$(2.6) \qquad dq_t = \kappa\exp(\kappa t)\bar{r}dt + \exp(\kappa t/2)\sigma\sqrt{q_t}dZ_t.$$

We now remove the occurrence of q_t in the diffusion coefficient by making the transformation $\sqrt{q_t}$ for which we have the SDE

$$(2.7) \qquad d\sqrt{q_t} = \frac{1}{2\sqrt{q_t}}dq_t - \frac{1}{8\sqrt{q_t^3}}d[q]_t$$
$$= \frac{\sigma^2\exp(\kappa t)}{8\sqrt{q_t}}\left(\frac{4\kappa\bar{r}}{\sigma^2} - 1\right)dt + \frac{1}{2}\exp(\kappa t/2)\sigma dZ_t.$$

An explicit solution to (2.5) is

$$(2.8) \qquad r_t = \exp(-\kappa t)\sum_{i=1}^{\nu}(\lambda^{(i)} + Z_{\varphi_t}^{(i)})^2$$

where $\lambda^{(1)},\ldots,\lambda^{(\nu)}$ are chosen such that $r_0 = \sum_{i=1}^{\nu}(\lambda^{(i)})^2$, $\varphi_t = \varphi_0 + \frac{1}{4}\sigma^2(\exp(\kappa t) - 1)/\kappa$ and ν is an integer such that $\nu = \frac{4\kappa\bar{r}}{\sigma^2} > 2$.

The transition density function of the short rate process in (2.5) is that of the non-central chi-square distribution, which, for times s and t with $0 \le s < t$, is given by

$$(2.9)$$

$$p_r(s, r_s, t, r_t) = \frac{1}{2(\varphi_t - \varphi_s)\exp(-\kappa t)}\left(\frac{r_t\exp(\kappa t)}{r_s\exp(\kappa s)}\right)^{\frac{1}{2}\left(\frac{\nu}{2} - 1\right)}$$

$$\times \exp\left(-\frac{1}{2}\frac{r_s\exp(\kappa s) + r_t\exp(\kappa t)}{(\varphi_t - \varphi_s)}\right)I_{\frac{\nu}{2} - 1}\left(\frac{\sqrt{r_s r_t\exp(\kappa(s + t))}}{(\varphi_t - \varphi_s)}\right),$$

where $\varphi_t = \varphi_0 + \frac{1}{4}\sigma^2(\exp(\kappa t) - 1)/\kappa$ and $\nu = \frac{4\kappa\bar{r}}{\sigma^2}$ and where

$$(2.10) \qquad I_\nu(x) = \sum_{i=0}^{\infty}\frac{1}{i!\Gamma(i + \nu + 1)}\left(\frac{x}{2}\right)^{2i+\nu}$$

is the power series expansion of the modified Bessel function of the first kind.

Restated, for $t > s$ the conditional random variable

$$(2.11) \qquad \frac{\exp(\kappa t)}{\varphi_t - \varphi_s} r_t$$

given r_s has a non-central chi-squared distribution with $\nu = 4\kappa\bar{r}/\sigma^2$ degrees of freedom and non-centrality parameter $\lambda = r_s \exp(\kappa s)/(\varphi_t - \varphi_s)$, namely

$$(2.12) \qquad \frac{\exp(\kappa t)}{\varphi_t - \varphi_s} r_t \sim \chi^2_{\nu, r_s \exp(\kappa s)/((\varphi_t - \varphi_s))}.$$

2.3. **The 3/2 Short Rate Model.** The 3/2 model was first derived in Platen [1999] and studied by Ahn and Gao [1999], the SDE of which is

$$(2.13) \qquad dr_t = (pr_t + qr_t^2)dt + \sigma r_t^{3/2} dZ_t,$$

where $q < \sigma^2/2$ and $\sigma > 0$ so as to avoid explosive values of r_t. The 3/2 power law for the volatility in this model was shown in Chan et al. [1992] to be the best fitting power law. Also the nonlinear drift term of this model could not be rejected in Aït-Sahalia [1996].

Setting $R_t = 1/r_t$ we obtain the SDE

$$(2.14) \qquad dR_t = (\sigma^2 - q - pR_t)dt - \sigma\sqrt{R_t} dZ_t,$$

which shows that $R_t = 1/r_t$ follows a square root process. This fact makes the derivation of the transition density function of the 3/2 model straightforward.

We remark that it is possible to generalise the 3/2 model to allow for an arbitrary exponent of the short rate in the diffusion term by modelling the short rate as a power transformation of an underlying stochastic process obeying a CIR-type SDE. The maximum likelihood estimation of parameters could then be performed over four parameters and yielding an improved fit. However, we confine our attention to the 3/2 model in this article.

The solution to the SDE (2.13) is

$$(2.15) \qquad r_t = \exp(pt) / \sum_{i=1}^{\nu} (\lambda^{(i)} + Z_{\varphi_t}^{(i)})^2$$

where ν is an integer such that $\nu = \frac{4(\sigma^2 - q)}{\sigma^2}$, $\lambda^{(1)}, \ldots, \lambda^{(\nu)}$ are chosen such that $r_0 = 1/\sum_{i=1}^{\nu} (\lambda^{(i)})^2$ and where $\varphi_t = \varphi_0 + \frac{1}{4}\sigma^2(\exp(pt) - 1)/p$.

The transition density function of the 3/2 short rate model in (2.13) is

$$(2.16) \qquad p_r(s, r_s, t, r_t) = \frac{r_t^{-2}}{2(\varphi_t - \varphi_s)\exp(-pt)} \left(\frac{r_t^{-1}\exp(pt)}{r_s^{-1}\exp(ps)} \right)^{\frac{1}{2}(\frac{\nu}{2} - 1)}$$

$$\times \exp\left(-\frac{1}{2} \frac{r_s^{-1}\exp(ps) + r_t^{-1}\exp(pt)}{(\varphi_t - \varphi_s)} \right)$$

$$\times I_{\frac{\nu}{2} - 1}\left(\frac{\sqrt{r_s^{-1}r_t^{-1}\exp(p(s+t))}}{(\varphi_t - \varphi_s)} \right),$$

where $\varphi_t = \varphi_0 + \frac{1}{4}\sigma^2(\exp(pt) - 1)/p$ and $\nu = 4(1 - q/\sigma^2)$ and where

$$(2.17) \qquad I_\nu(x) = \sum_{i=0}^{\infty} \frac{1}{i!\Gamma(i + \nu + 1)} \left(\frac{x}{2} \right)^{2i + \nu}$$

is the power series expansion of the modified Bessel function of the first kind as in (2.10).

Alternatively stated, the conditional random variable

$$(2.18) \qquad \frac{\exp(pt)}{r_t(\varphi_t - \varphi_s)}$$

given r_s has a non-central chi-squared distribution with $\nu = 4(\sigma^2 - q)/\sigma^2$ degrees of freedom and non-centrality parameter $\lambda = \exp(ps)/(r_s(\varphi_t - \varphi_s))$, namely

$$(2.19) \qquad \frac{\exp(pt)}{r_t(\varphi_t - \varphi_s)} \sim \chi^2_{\nu,\exp(ps)/(r_s(\varphi_t-\varphi_s))},$$

given r_s where $\varphi_t = \varphi_0 + \frac{1}{4}\sigma^2(\exp(pt) - 1)/p$.

3. Maximum Likelihood Estimation Method

The maximum likelihood estimation is demonstrated for each of the three short rate models.

3.1. **Vasicek Short Rate Model.** Estimating the parameters of the Vasicek model is achieved using maximum likelihood estimation. To avoid any potential confusion we derive the estimators showing all steps. Because the transition density function of the Vasicek short rate is normal it suffices to obtain formulae for the conditional mean and variance, which are given as

$$(3.1) \qquad m_s(t) = \bar{r}\kappa B(s,t) + r_s(1 - \kappa B(s,t))$$

$$v_s(t) = \sigma^2\left(B(s,t) - \frac{1}{2}\kappa B(s,t)^2\right),$$

where

$$(3.2) \qquad B(s,t) = (1 - \exp(-\kappa(t-s)))/\kappa.$$

Therefore our log-likelihood function under the Vasicek model on the set of observed short rates r_{t_i}, for $i = 0, 2, \ldots, n$ is

$$(3.3) \qquad \ell(\bar{r}, \kappa, \sigma) = -\frac{1}{2}\sum_{i=1}^{n}\left(\log(2\pi v_{t_{i-1}}(t_i)) + \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)}\right),$$

where

$$(3.4) \qquad m_{t_{i-1}}(t_i) = \bar{r}\kappa B(t_{i-1}, t_i) + r_{t_{i-1}}(1 - \kappa B(t_{i-1}, t_i)),$$

$$v_{t_{i-1}}(t_i) = \sigma^2\left(B(t_{i-1}, t_i) - \frac{1}{2}\kappa B(t_{i-1}, t_i)^2\right),$$

and where $B(s,t)$ is as in (3.2).

The following theorem provides explicit maximum likelihood estimates (MLEs) of $\bar{r}$, κ and σ for a fixed value of $\bar{r}$. Such formulae are supplied in Liptser and Shiryaev [2001] but we supply them explicitly here.

Theorem 1. *Assume that the times $t_0 < t_1 < \ldots < t_n$ are equidistant with spacing δt. Then the MLEs of $\bar{r}$, κ and σ are given by*

$$(3.5) \qquad \bar{r} = \frac{S_1 S_{00} - S_0 S_{01}}{S_0 S_1 - S_0^2 - S_{01} + S_{00}},$$

$$\kappa = \frac{1}{\delta t} \log \frac{S_0 - \bar{r}}{S_1 - \bar{r}},$$

$$\sigma^2 = \frac{1}{n\beta(1 - \frac{1}{2}\kappa\beta)} \sum_{i=1}^{n} (r_{t_i} - m_{t_{i-1}}(t_i))^2,$$

where

$$(3.6) \qquad S_0 = \frac{1}{n}\sum_{i=1}^{n} r_{t_{i-1}}, \qquad S_1 = \frac{1}{n}\sum_{i=1}^{n} r_{t_i},$$

$$S_{00} = \frac{1}{n}\sum_{i=1}^{n} r_{t_{i-1}} r_{t_{i-1}}, \qquad S_{01} = \frac{1}{n}\sum_{i=1}^{n} r_{t_{i-1}} r_{t_i},$$

and $\beta = \frac{1}{\kappa}(1 - \exp(-\kappa\delta t))$.

Proof. See Appendix. ☐

As a result, Theorem 1 supplies the explicit MLEs $(\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma})$, denoted by the three-vector

$$(3.7) \qquad \hat{\underset{\sim}{\theta}}_{Vasicek} = \begin{pmatrix} \hat{\bar{r}} \\ \hat{\kappa} \\ \hat{\sigma} \end{pmatrix}.$$

To provide standard errors of these MLEs, we note that their variances satisfy the Cramér-Rao inequality

$$(3.8) \qquad VAR((\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma})) \geq \frac{1}{I(\bar{r}, \kappa, \sigma)},$$

where $I(\bar{r}, \kappa, \sigma)$ is the Fisher information matrix. As the number of observations approaches infinity the variance is asymptotic to the lower bound. Also the Fisher information matrix is approximated by the observed Fisher information matrix

$$(3.9) \qquad I(\bar{r}, \kappa, \sigma) \approx I(\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma}) = -\nabla^2 \ell(\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma}).$$

The following theorem supplies the observed Fisher information matrix in respect of MLEs of the Vasicek model. As far as we can ascertain, there is no explicit statement of the Fisher Information Matrix for the Vasicek model in the literature.

Theorem 2. *The observed Fisher information matrix in respect of the MLEs in Theorem 1 is given by*

$$(3.10) \qquad \begin{pmatrix} \frac{n\kappa^2\beta^2}{\sigma^2(\hat{\beta} + \frac{1}{2}\kappa\beta^2)} & 0 & 0 \\ 0 & e & 0 \\ 0 & 0 & \frac{2n}{\sigma^2} \end{pmatrix},$$

where $\beta = \frac{1}{\kappa}(1 - \exp(-\kappa\delta t))$ and we have assumed that the times $t_0 < t_1 < \ldots < t_n$ are equidistant with spacing δt.

Proof. See Appendix. ☐

3.2. **Cox-Ingersoll-Ross Short Rate Model.** Estimating the parameters of the CIR model is achieved using the maximum likelihood method. The log-likelihood function is given by

$$(3.11) \qquad \ell(\bar{r}, \kappa, \sigma) = \sum_{i=1}^{n} \log p_r(t_{i-1}, r_{t_{i-1}}, t_i, r_{t_i}),$$

where the transition density function p_r is as in (2.9). Employing a gradient ascent algorithm in log parameter space ensures positivity of the estimates. The initial estimate $\underset{\sim}{u}_0$ of the log parameter triple is computed as

$$(3.12) \qquad \underset{\sim}{u}_0 = \log \begin{pmatrix} \hat{\underset{\sim}{\theta}}^{(1)}_{Vasicek} \\ \hat{\underset{\sim}{\theta}}^{(2)}_{Vasicek} \\ \hat{\underset{\sim}{\theta}}^{(3)}_{Vasicek} / \sqrt{\hat{\underset{\sim}{\theta}}^{(1)}_{Vasicek}} \end{pmatrix},$$

derived from the $\hat{\underset{\sim}{\theta}}_{Vasicek}$ triple in (3.7). Subsequent approximations are iteratively obtained using the formula

$$(3.13) \qquad \underset{\sim}{u}_k = \underset{\sim}{u}_{k-1} + \frac{1}{\|\nabla \ell(\underset{\sim}{u}_{k-1})\|^2} \nabla \ell(\underset{\sim}{u}_{k-1}),$$

for $k \geq 1$ and where ℓ is the log likelihood function in terms of the log parameters.

3.3. **Cox-Ingersoll-Ross Model with a Log-Normal Approximation.** Poulsen [1999] shows that good estimates can be obtained by approximating the transition density of the CIR process with a Gaussian distribution having the same mean and variance as the transition density function. However, our fitting of the CIR model to the data suggests that that the Gaussian distribution is inadequate as an approximation and further, we find that similar estimates to those derived using the exact transition density can be obtained with the lognormal approximation by matching first and second moments. This is not surprising given that the lognormal distribution is positively skewed and is defined for positive values of the random variable, as is the case for the non-central chi-squared distribution.

For the CIR process in (2.5) and times s, t with $s \leq t$ the mean and variance of r_t given r_s are given by

$$(3.14) \qquad m_s(t) = \bar{r}\kappa B(s,t) + r_s(1 - \kappa B(s,t))$$

$$v_s(t) = \sigma^2 \left(\frac{1}{2}\kappa B(s,t)^2 \bar{r} + (B(s,t) - \kappa B(s,t)^2)r_s \right),$$

where

$$(3.15) \qquad B(s,t) = (1 - \exp(-\kappa(t-s)))/\kappa.$$

We approximate the transition density of the CIR process from time s to time t by a lognormal distribution that matches the mean and variance. It is straightforward to show that the approximating lognormal distribution has parameters $\mu_s^{(LN)}(t)$,

$\sigma_s^{(LN)}(t)$ given by

$$(3.16) \qquad \mu_s^{(LN)}(t) = \log m_s(t) - \frac{1}{2}(\sigma_s^{(LN)}(t))^2$$

$$(\sigma_s^{(LN)}(t))^2 = \log\left(1 + \frac{v_s(t)}{m_s(t)^2}\right)$$

where

$$(3.17) \qquad m_{t_{i-1}}(t_i) = \bar{r}\kappa B(t_{i-1}, t_i) + r_{t_{i-1}}(1 - \kappa B(t_{i-1}, t_i))$$

$$v_{t_{i-1}}(t_i) = \sigma^2\left(\frac{1}{2}\kappa B(t_{i-1}, t_i)^2 \bar{r} + (B(t_{i-1}, t_i) - \kappa B(t_{i-1}, t_i)^2)r_{t_{i-1}}\right)$$

and $B(s,t)$ is as in (3.15).

Therefore, our approximating log-likelihood function on the set of observed short rates r_{t_i}, for $i = 0, 2, \ldots, n$ is

$$(3.18) \quad \ell(\bar{r}, \kappa, \sigma) = -\frac{1}{2}\sum_{i=1}^{n}\left(\log(2\pi v_{t_{i-1}}(t_i)) + \log r_{t_i} + \frac{(\log r_{t_i} - \mu_{t_{i-1}}^{(LN)}(t_i))^2}{(\sigma_{t_{i-1}}^{(LN)}(t_i))^2}\right).$$

3.4. 3/2 **Short Rate Model.** Estimating the parameters of the 3/2 model is achieved using the maximum likelihood method. The log-likelihood function is given by

$$(3.19) \qquad \ell(p, q, \sigma) = \sum_{i=1}^{n}\log p_r(t_{i-1}, r_{t_{i-1}}, t_i, r_{t_i}),$$

where the transition density function p_r is as in (2.16). Exploiting the fact that the 3/2 short rate model is a CIR model of the reciprocal of the short rate, we compute initial estimate $\underset{\sim}{u}_0$ under the CIR short rate model using the reciprocal of the elements in the data set. The corresponding $(\hat{p}, \hat{q}, \hat{\sigma})$ triple for the 3/2 short rate model is computed as

$$(3.20) \qquad \begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{\sigma} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{u}_0^{(2)} \\ (\underset{\sim}{u}_0^{(3)})^2 - \underset{\sim}{u}_0^{(1)}\underset{\sim}{u}_0^{(2)} \\ \underset{\sim}{u}_0^{(3)} \end{pmatrix}.$$

In the same vein as that done for the CIR model, a gradient ascent algorithm in log parameter space converges to the solution.

4. Numerical Results of Method

Maximum likelihood estimation is applied to the annual series of one-year deposit rates from 1871 to 2012, given by Shiller [1989]. The numerical results are shown in this section for each model and comparisons of goodness-of-fits are made.

FIGURE 1. Actual short rate and fitted Vasicek mean reverting level for US cash rates.
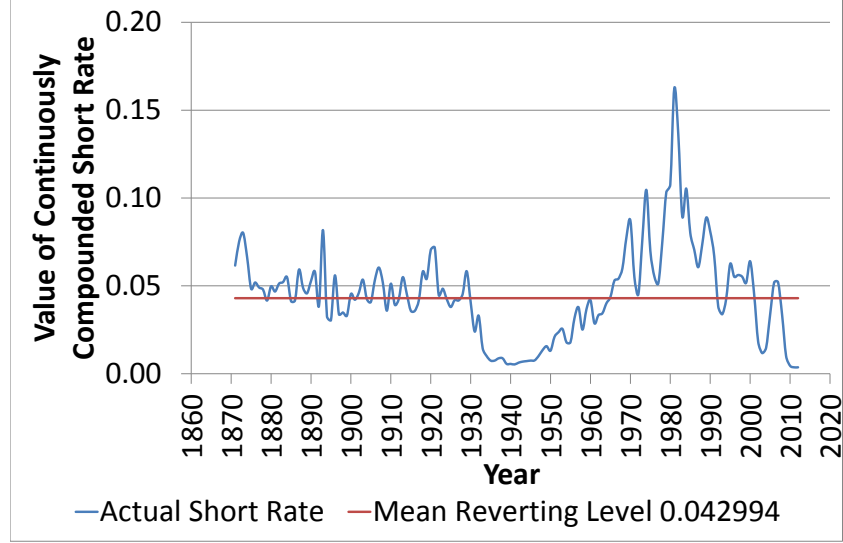
4.1. **Vasicek Model.** We fit the Vasicek model to the data set. We obtain the MLEs, with standard errors shown in brackets,

$$
\begin{aligned}
\bar{r} &= 0.042994 \, (0.0080023) \\
\kappa &= 0.162953 \, (0.053703) \\
\sigma &= 0.015384 \, (0.00099592).
\end{aligned}
\tag{4.1}
$$

We show the parameter estimate for the mean reverting level $\bar{r}$ alongside the historical short rates in Figure 1. We note that for the periods after 1930 a time dependent reference level may be appropriate but we deliberately keep constant parameters in our consideration.

One way of assessing the goodness of fit of the parameter estimates for σ and κ in a graphical fashion is to compare the theoretical quadratic variation of $q_t = r_t \exp(\kappa t)$ with the observed quadratic variation.

From the SDE (2.2) the theoretical quadratic variation of $q_t = r_t \exp(\kappa t)$ is

$$
[q]_t = \sigma^2 \int_0^t \exp(2\kappa s) ds = \sigma^2 \frac{\exp(2\kappa t) - 1}{2\kappa}
\tag{4.2}
$$

and the observed quadratic variation of q is computed using the formula

$$
[q]_t \approx \sum_{j:t_j \leq t} (r_{t_j} \exp(\kappa t_j) - r_{t_{j-1}} \exp(\kappa t_{j-1}))^2.
\tag{4.3}
$$

FIGURE 2. Logarithm of quadratic variation of $\exp(\kappa t)r_t$.
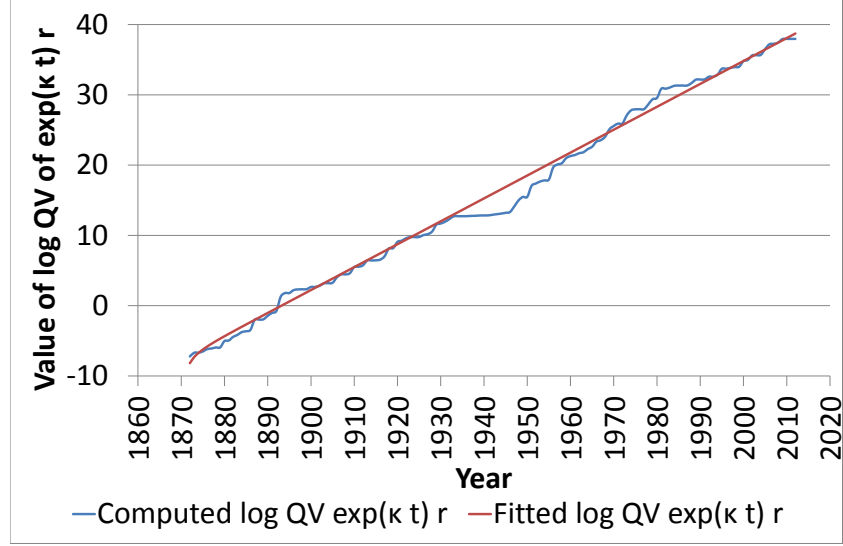
The logarithm of the observed quadratic variation of $q_t = r_t \exp(\kappa t)$ in (4.3) is shown alongside the logarithm of the fitted quadratic variation function (4.2) in Figure 2. We note that we have visually a good fit.

4.2. **Cox-Ingersoll-Ross Model.** We fit the CIR model to the data set, giving the maximum likelihood estimates

$$(4.4) \qquad\qquad \bar{r} = 0.041078\,(0.011421)$$
$$\kappa = 0.092540\,(0.038668)$$
$$\sigma = 0.064670\,(0.0040761),$$

where the standard errors are shown in brackets.

In Figure 3 we plot the actual short rate versus the fitted mean reversion level. We obtain a similar reference level as for the Vasicek model, see (4.1).

The logarithm of the empirically calculated quadratic variation of $\sqrt{r_t \exp(\kappa t)}$ is shown alongside the logarithm of the theoretically computed quadratic variation function, $[\sqrt{q}]_t = \frac{1}{4}\sigma^2(\exp(\kappa t) - 1)/\kappa$, in Figure 4. We note visually a reasonable long term fit.

We remark that substituting the values for $\bar{r}$, κ and σ given in (4.4) into the equation $\nu = \frac{4\kappa\bar{r}}{\sigma^2}$ gives $\nu = 4 \times 0.041078 \times 0.092540/(0.064670)^2 = 3.6357$, which could be approximated by $\nu \approx 4$.

Unlike the Vasicek model, under the CIR model there is no possibility of negative interest rates, as demonstrated by (2.8).

FIGURE 3. Actual short rate and fitted CIR mean reversion level for US cash rates.
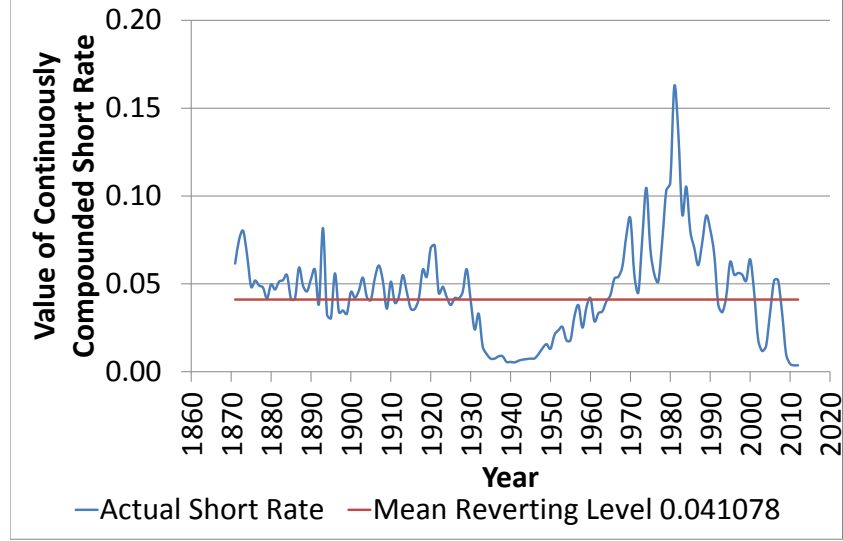
4.3. Cox-Ingersoll-Ross Model with a Log-Normal Approximation.
We fit the CIR model to the data set, giving the maximum likelihood estimates

$$(4.5) \qquad \bar{r} = 0.042470 \, (0.010935)$$

$$\kappa = 0.102677 \, (0.041105)$$

$$\sigma = 0.070948 \, (0.0041198).$$

We note that these estimates are close to those in (4.4).

4.4. The $3/2$ Short Rate Model.
We fit the $3/2$ model to the data set using the maximum likelihood method giving the parameter values

$$(4.6) \qquad p = 0.038506 \, (0.042284)$$

$$q = 0.877908 \, (1.177853)$$
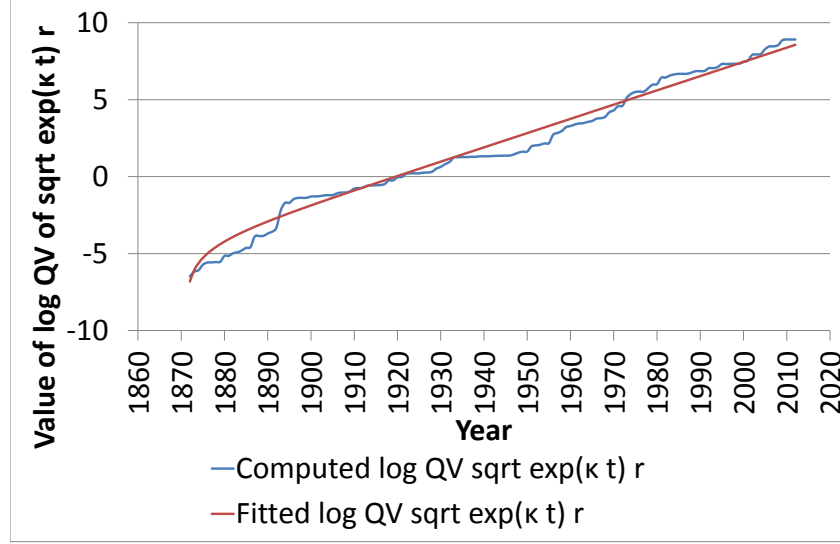
$$\sigma = 2.0681 \, (0.13241),$$

where the standard errors are shown in brackets.

The mean reverting level of $R_t = 1/r_t$ in (2.14) is given by

$$(4.7) \qquad \frac{\sigma^2 - q}{p}.$$

The inverse of this level is not the mean reverting level of r_t. The limiting distribution of r_t/p as $t \to \infty$ is an inverse chi-squared distribution with ν degrees of

FIGURE 4. Logarithm of quadratic variation of $\sqrt{\exp(\kappa t)r_t}$ for US cash rates.

freedom and the mean of r_t as $t \to \infty$ is deduced to be

$$(4.8) \qquad \frac{p/\sigma^2}{\nu - 2} = \frac{p}{2\sigma^2 - 4q}.$$

In Figure 5 the graph of the short rate is shown along with the implied reverting level, the mean, of 0.03054 of the 3/2 model. The dimension of the square root process $1/r_t$ is here estimated as $\nu = 4(\sigma^2 - q)/\sigma^2 \approx 3.1790$, which is reasonably close to three.

4.5. **Comparing Fits.** The three models considered in this article have explicit formulae for their transition density functions and this has allowed the fitting of parameters using maximum likelihood estimation. The Vasicek model is most easily fitted to the data because it has closed form expressions for the parameter estimates. In contrast, the CIR and 3/2 models each require iterative algorithms to find the best fitting parameters.

In fitting the three models to the US cash rates we can identify which model provides the best fit to the data by looking at the value of the maximum likelihood function. Since each of the models has the same number of parameters it is not necessary to use the Akaike Information Criterion, as given in Burnham and Anderson [2004] for example. The log likelihood value of each model is shown in Table 2, where the CIR model appears to be the best fitting model.

To establish further whether the CIR model is a good fitting model we consider Pearson's goodness-of-fit chi-squared statistic, described in Kendall and Stuart

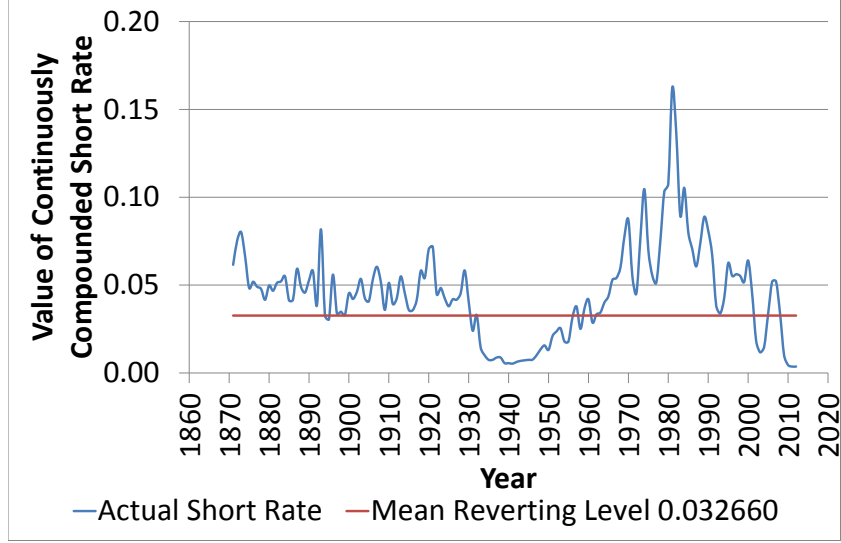FIGURE 5. The fitted reverting level under the 3/2 model.

TABLE 2. Values of the log likelihood in respect of various short
rate models (US cash rates).

| Model	Parameters	Log Likelihood
Vasicek	3	399.7019
CIR	3	427.8116
3/2	3	406.2713

[1961], given by

$$(4.9) \qquad S = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i,$$

where O_i is the number of observations in category i and E_i is the corresponding expected number of observations according to the hypothesised model. The test statistic S is asymptotically distributed as χ^2_ν, where ν equals the number of categories k less the number of constraints and estimated parameters of the model.

Given a time series of interest rates $\{r_{t_j} : j = 1, 2, \ldots, n\}$ and given a hypothesised transition density function with corresponding cumulative distribution function F we compute the $n-1$ quantiles $q_j = F(t_{j-1}, r_{t_{j-1}}, t_j, r_{t_j})$ for $j = 2, 3, \ldots, n$. Under the hypothesised model the quantiles q_j are independent and uniformly distributed. These quantiles are graphed against those of the uniform distribution in Figure 6. One notes that the CIR model remains in some sense visually closest over the $[0, 1]$ interval.

FIGURE 6. Comparison of quantile-quantile plots of short rate models (Shiller US data).
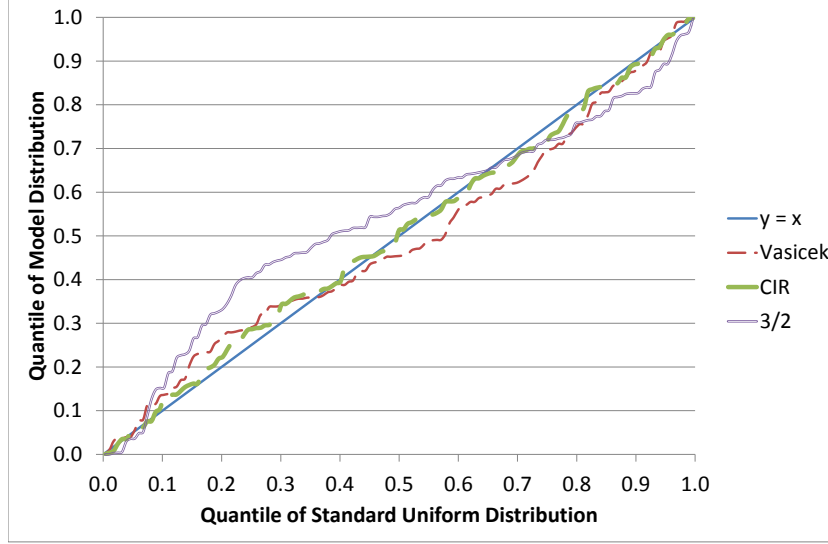
A similar comparison is shown in Figure 7 for the monthly data series of one-year US Treasury bond yields from January 1962 to June 2014, sourced from the US Federal Reserve Bank website, where a similar conclusion follows.

For a fixed integer k satisfying $2 \leq k \leq (n-1)/5$ we partition the unit interval into k equally sized subintervals. Hence we compute the number of observations O_i in the i-th subinterval $((i-1)/k, i/k]$ for $i = 1, 2, \ldots, k$. The corresponding expected number of observations E_i in the i-th subinterval is $(n-1)/k$. Our test statistic is thus computed as

$$(4.10) \qquad S = k \sum_{i=1}^{k} (O_i - (n-1)/k)^2/(n-1),$$

which is approximately chi-squared distributed with $\nu = k - 1 - n_{parameters}$ degrees of freedom.

The value of the Pearson's chi-squared statistic and corresponding p-value for each model and for a range of partition sizes is shown in Table 3. It is evident that the 3/2 model and, for some partition sizes, the Vasicek model can be rejected at the 1% level of significance and that the CIR model cannot be rejected at this level of significance. We conclude that the CIR model cannot be rejected as a valid model whereas we can reject the validity of the Vasicek model and the 3/2 model.

Another test of goodness-of-fit is the Kolmogorov-Smirnov test, as described by Stephens [1974]. Under the null hypothesis that the set of n observations $u_1, u_2, \ldots, u_n$ emanate from a uniform distribution, the Kolmogorov test statistic

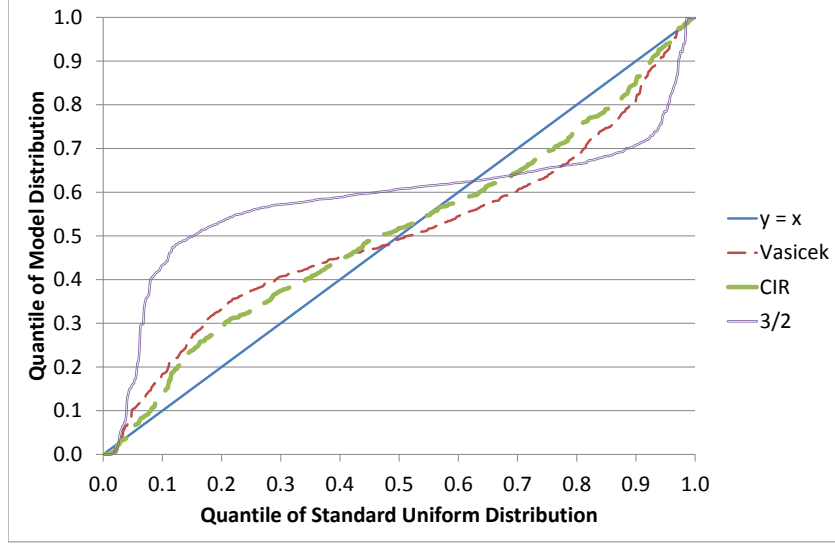FIGURE 7. Comparison of quantile-quantile plots of short rate models (US Federal Reserve monthly data).

TABLE 3. Pearson's chi-squared statistic with p-values shown in brackets in respect of various short rate models (US cash rates).

k	ν	Vasicek	CIR	3/2
5	1	8.7143 (0.3157%)	0.5714 (44.9703%)	30.6429 (0.0000%)
10	6	13.7143 (3.2996%)	1.8571 (93.2361%)	35.5714 (0.0003%)
15	11	21.7857 (2.6087%)	10.2143 (51.1220%)	37.6429 (0.0090%)
20	16	34.0000 (0.5433%)	11.1429 (80.0580%)	42.5714 (0.0324%)
25	21	30.3571 (8.5044%)	18.9286 (58.9720%)	41.7857 (0.4476%)

is

$$(4.11) \qquad D_n = \sup_{x \in \{u_1, u_2, \dots, u_n\}} \max\left(F^{(n)}(x) - x, x - F^{(n)}(x) - \frac{1}{n}\right)$$

and the modified test statistic $K_n = \sqrt{n} D_n$ has the limiting distribution function, as $n \to \infty$,

$$(4.12) \qquad F(x) = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{\infty} \exp(-(2k-1)^2 \pi^2/(8x^2)),$$

where

$$(4.13) \qquad F^{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{u_i \leq x}$$

TABLE 4. Kolmogorov-Smirnov test statistics in respect of various short rate models (US cash rates).

	Vasicek	CIR	3/2
D_n	0.08468	0.04526	0.17092
n	141	141	141
K_n	1.00556	0.53749	2.02958
$F(K_n)$	0.73592	0.06518	0.99947
p-value	0.26408	0.93482	0.00053

TABLE 5. Anderson-Darling test statistics in respect of various short rate models (US cash rates).

	Vasicek	CIR	3/2
S	-142.50533	-141.28374	-146.21838
n	141	141	141
$A^2 = -n - S$	1.50533	0.28374	5.21838
A	1.22692	0.53267	2.28438
p-value	0.1898	0.9546	0.0015

is the empirical cumulative distribution function. We compute the test statistics in Table 4 where we see that only the 3/2 short rate model can be rejected at the 1% level of significance.

Finally, another test of goodness-of-fit is the Anderson-Darling test, as described in Stephens [1974]. Under the null hypothesis that the set of n observations $u_1, u_2, \ldots, u_n$ emanate from a uniform distribution, the test statistic A is given by

$$(4.14) \qquad A = \sqrt{-n - S},$$

where

$$(4.15) \qquad S = \sum_{i=1}^{n} \frac{2i-1}{n} \big(\log u_i + \log(1 - u_{n+1-i}) \big).$$

We compute the test statistics in Table 5 where, as for the Kolmogorov-Smirnov test, we see that only the 3/2 short rate model can be rejected at the 1% level of significance. The p-values of the test statistic A in Table 5 were estimated using sample Anderson-Darling statistics of 1000 simulations of sets of 141 uniformly distributed observations.

5. CONCLUSIONS

In this article we have demonstrated applicability of the maximum likelihood estimation of parameters of short rate models. The fitted parameters fitted to the data set have values consistent with market assessments. Clearly the possibility of negative interest rates under the Vasicek model makes it less preferable to the other two models which preclude such a possibility. An example of a modification to the Vasicek model which precludes negative interest rates is the exponential Vasicek

model. Also, we have demonstrated several ways of assessing the goodness-of-fits of short rate models.

APPENDIX: PROOFS OF RESULTS ON THE VASICEK MODEL

Here we present proofs of results in Section 2.1.

Proof of Theorem 1. Differentiating (3.3) with respect to $\bar{r}$ we have

$$(A1) \qquad \frac{\partial}{\partial \bar{r}} \ell(\bar{r}, \kappa, \sigma) = -\frac{1}{2} \sum_{i=1}^{n} \frac{2(r_{t_i} - m_{t_{i-1}}(t_i))}{v_{t_{i-1}}(t_i)} \times -\frac{\partial m_{t_i}}{\partial \bar{r}}$$

$$= \kappa \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)},$$

where we have used the fact that $\frac{\partial m_{t_{i-1}}(t_i)}{\partial \bar{r}} = \kappa B(t_{i-1}, t_i)$. Equating (A1) to zero gives the equation

$$(A2) \qquad \sum_{i=1}^{n} (r_{t_i} - m_{t_{i-1}}(t_i)) = 0$$

which, using (3.1), is equivalent to

$$(A3) \qquad \sum_{i=1}^{n} (r_{t_i} - \bar{r}) = \sum_{i=1}^{n} (r_{t_{i-1}} - \bar{r})(1 - \kappa B(t_{i-1}, t_i)).$$

Since the sampling times t_i are equidistant we can solve for κ, giving the solution

$$(A4) \qquad \kappa = \frac{1}{\delta t} \log \frac{\sum_{i=1}^{n} (r_{t_{i-1}} - \bar{r})}{\sum_{i=1}^{n} (r_{t_i} - \bar{r})}.$$

Differentiating (3.3) with respect to σ we have

$$(A5) \qquad \frac{\partial}{\partial \sigma} \ell(\bar{r}, \kappa, \sigma) = -\frac{1}{2} \sum_{i=1}^{n} \left(\frac{1}{v_{t_{i-1}}(t_i)} - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)^2} \right) \frac{\partial v_{t_{i-1}}(t_i)}{\partial \sigma}$$

$$= -\sum_{i=1}^{n} \left(\frac{1}{v_{t_{i-1}}(t_i)} - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)^2} \right) \frac{v_{t_{i-1}}(t_i)}{\sigma}$$

$$= -\frac{1}{\sigma} \sum_{i=1}^{n} \left(1 - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right),$$

where we have used the fact that $\frac{\partial v_{t_{i-1}}(t_i)}{\partial \sigma} = 2\frac{v_{t_{i-1}}(t_i)}{\sigma}$. Equating (A5) to zero gives the equation

$$(A6) \qquad \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} = \sum_{i=1}^{n} 1$$

which, using the equation for $v_{t_{i-1}}(t_i)$ in (3.1), is equivalent to

$$(A7) \qquad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{B(t_{i-1}, t_i)(1 - \frac{1}{2}\kappa B(t_{i-1}, t_i))}.$$

Since the sampling times t_i are equidistant we can simplify the equation for σ^2 to

$$(A8) \qquad \sigma^2 = \frac{1}{n\beta(1 - \frac{1}{2}\kappa\beta)} \sum_{i=1}^{n}(r_{t_i} - m_{t_{i-1}}(t_i))^2.$$

Differentiating (3.3) with respect to κ we have

$$(A9) \qquad \frac{\partial}{\partial\kappa}\ell(\bar{r},\kappa,\sigma) = -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{v_{t_i}}\frac{\partial v_{t_i}}{\partial\kappa} + \frac{\partial}{\partial\kappa}\left(\frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)}\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{1}{v_{t_{i-1}}(t_i)}\frac{\partial v_{t_{i-1}}(t_i)}{\partial\kappa}\right.$$

$$+ \frac{2(r_{t_i} - m_{t_{i-1}}(t_i))}{v_{t_{i-1}}(t_i)}\times\left(-\frac{\partial m_{t_{i-1}}(t_i)}{\partial\kappa}\right)$$

$$\left. - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)^2}\times\frac{\partial v_{t_{i-1}}(t_i)}{\partial\kappa}\right\}$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{v_{t_{i-1}}(t_i)}\frac{\partial v_{t_{i-1}}(t_i)}{\partial\kappa}\left(1 - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)}\right)$$

$$- \frac{2(r_{t_i} - m_{t_{i-1}}(t_i))}{v_{t_{i-1}}(t_i)}\frac{\partial m_{t_{i-1}}(t_i)}{\partial\kappa}.$$

To simplify (A9) we determine expressions for $\frac{\partial v_{t_{i-1}}(t_i)}{\partial\kappa}$ and $\frac{\partial m_{t_{i-1}}(t_i)}{\partial\kappa}$. Firstly,

$$(A10) \qquad \frac{\partial v_{t_i}}{\partial\kappa} = \frac{\partial}{\partial\kappa}\left\{\sigma^2\left(B(t_{i-1},t_i) - \frac{1}{2}\kappa B(t_{i-1},t_i)^2\right)\right\}$$

$$= \sigma^2\left(\frac{\partial B(t_{i-1},t_i)}{\partial\kappa} - \frac{1}{2}B(t_{i-1},t_i)^2 - \kappa B(t_{i-1},t_i)\frac{\partial B(t_{i-1},t_i)}{\partial\kappa}\right).$$

We note that

$$(A11) \qquad \frac{\partial(\kappa B(t_{i-1},t_i))}{\partial\kappa} = \frac{\partial(1 - \exp(-\kappa(t_i - t_{i-1})))}{\partial\kappa}$$

$$= (t_i - t_{i-1})\exp(-\kappa(t_i - t_{i-1}))$$

$$= (t_i - t_{i-1})(1 - \kappa B(t_{i-1},t_i))$$

and, therefore,

$$(A12) \qquad \frac{\partial B(t_{i-1},t_i)}{\partial\kappa} = \frac{1}{\kappa}\left(\frac{\partial(\kappa B(t_{i-1},t_i))}{\partial\kappa} - B(t_{i-1},t_i)\right)$$

$$= \frac{1}{\kappa}\left((t_i - t_{i-1})(1 - \kappa B(t_{i-1},t_i)) - B(t_{i-1},t_i)\right).$$

Hence (A10) becomes

(A13) $\quad \dfrac{\partial v_{t_i}}{\partial \kappa} = \sigma^2\left(-\dfrac{1}{2}B(t_{i-1},t_i)^2 + (1 - \kappa B(t_{i-1},t_i))\dfrac{\partial B(t_{i-1},t_i)}{\partial \kappa}\right)$

$$= \sigma^2\Bigg\{-\frac{1}{2}B(t_{i-1},t_i)^2$$
$$+ (1 - \kappa B(t_{i-1},t_i))\frac{1}{\kappa}\left((t_i - t_{i-1})(1 - \kappa B(t_{i-1},t_i)) - B(t_{i-1},t_i)\right)\Bigg\}$$
$$= \sigma^2\Bigg(-\frac{1}{2}B(t_{i-1},t_i)^2$$
$$+ \frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{1}{\kappa}B(t_{i-1},t_i)(1 - \kappa B(t_{i-1},t_i))\Bigg)$$
$$= \sigma^2\Bigg(-\frac{1}{2}B(t_{i-1},t_i)^2$$
$$+ \frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{1}{\kappa}B(t_{i-1},t_i) + B(t_{i-1},t_i)^2\Bigg)$$
$$= \sigma^2\left(\frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{1}{\kappa}B(t_{i-1},t_i) + \frac{1}{2}B(t_{i-1},t_i)^2\right)$$
$$= \sigma^2\frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{\sigma^2}{\kappa}\left(B(t_{i-1},t_i) - \frac{1}{2}\kappa B(t_{i-1},t_i)^2\right)$$
$$= \sigma^2\frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{1}{\kappa}v_{t_i}.$$

Secondly,

(A14)
$$\frac{\partial m_{t_{i-1}}(t_i)}{\partial \kappa} = \frac{\partial}{\partial \kappa}\left(r_s + (\bar{r} - r_s)\kappa B(t_{i-1},t_i)\right)$$
$$= (\bar{r} - r_s)(t_i - t_{i-1})(1 - \kappa B(t_{i-1},t_i))$$
$$= -(t_i - t_{i-1})(m_{t_{i-1}}(t_i) - \bar{r}).$$

Substituting (A13) and (A14) into (A9) gives

(A15)
$$\frac{\partial}{\partial \kappa}\ell(\bar{r},\kappa,\sigma) = -\frac{1}{2}\sum_{i=1}^{n}\Bigg\{\frac{1}{v_{t_{i-1}}(t_i)}\left(\sigma^2\frac{t_i - t_{i-1}}{\kappa}(1 - \kappa B(t_{i-1},t_i))^2 - \frac{1}{\kappa}v_{t_{i-1}}(t_i)\right)$$
$$\times\left(1 - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)}\right)$$
$$+ \frac{2(r_{t_i} - m_{t_{i-1}}(t_i))}{v_{t_{i-1}}(t_i)}(t_i - t_{i-1})(m_{t_{i-1}}(t_i) - \bar{r})\Bigg\}.$$

If the right hand side of (A5) is zero, then (A15) simplifies to

(A16)
$$\frac{\partial}{\partial \kappa}\ell(\bar{r},\kappa,\sigma) = -\frac{1}{2}\sum_{i=1}^{n}\frac{2(r_{t_i} - _{t_{i-1}}(t_i))}{v_{t_{i-1}}(t_i)}(t_i - t_{i-1})(m_{t_{i-1}}(t_i) - \bar{r})$$
$$= \delta t\frac{1}{\sigma^2(\beta + \frac{1}{2}\kappa\beta^2)}\sum_{i=1}^{n}(r_{t_i} - m_{t_{i-1}}(t_i))(m_{t_{i-1}}(t_i) - \bar{r}).$$

Hence $\frac{\partial}{\partial \kappa} \ell(\bar{r}, \kappa, \sigma) = 0$ is equivalent to

$$
\text{(A17)} \qquad 0 = \sum_{i=1}^{n} (r_{t_i} - m_{t_{i-1}}(t_i))(m_{t_{i-1}}(t_i) - \bar{r})
$$

$$
= \sum_{i=1}^{n} (r_{t_i} - \bar{r} - (1 - \kappa\beta)(r_{t_{i-1}} - \bar{r}))(r_{t_{i-1}} - \bar{r})(1 - \kappa\beta)
$$

from which we have

$$
\text{(A18)} \qquad 1 - \kappa\beta = \frac{\sum_{i=1}^{n}(r_{t_i} - \bar{r})(r_{t_{i-1}} - \bar{r})}{\sum_{i=1}^{n}(r_{t_{i-1}} - \bar{r})^2}.
$$

Thus in addition to (A3), we have an expression for $1 - \kappa\beta$ in (A18) and this allows us to solve explicitly for $\bar{r}$. $\qquad\square$

Proof of Theorem 2. We compute the second order partial derivatives of the log-likelihood function. Differentiating (A1) with respect to $\bar{r}$ gives

$$
\text{(A19)} \qquad \frac{\partial^2 \ell}{\partial \bar{r}^2} = \frac{\partial}{\partial \bar{r}}\left\{ \kappa \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)} \right\}
$$

$$
= -\kappa^2 \sum_{i=1}^{n} \frac{B(t_{i-1}, t_i)^2}{v_{t_{i-1}}(t_i)}.
$$

Differentiating (A1) with respect to σ gives

$$
\text{(A20)} \qquad \frac{\partial^2 \ell}{\partial \sigma \partial \bar{r}} = \frac{\partial}{\partial \sigma}\left\{ \kappa \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)} \right\}
$$

$$
= -\frac{2\kappa}{\sigma} \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)}.
$$

At the point of maximum likelihood the right hand side of (A1) vanishes and therefore so does the right hand side of (A20).

Differentiating (A1) with respect to κ gives

$$
\text{(A21)} \qquad \frac{\partial^2 \ell}{\partial \kappa \partial \bar{r}} = \frac{\partial}{\partial \kappa}\left\{ \kappa \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)} \right\}
$$

$$
= -\frac{2\kappa}{\sigma} \sum_{i=1}^{n} \frac{(r_{t_i} - m_{t_{i-1}}(t_i))B(t_{i-1}, t_i)}{v_{t_{i-1}}(t_i)}.
$$

At the point of maximum likelihood the right hand side of (A1) vanishes and therefore so does the right hand side of (A21).

Differentiating (A5) with respect to σ gives

$$
\text{(A22)} \qquad \frac{\partial^2 \ell}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left\{ -\frac{1}{\sigma} \sum_{i=1}^{n} \left(1 - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right) \right\}
$$

$$
= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(1 - 3\frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right).
$$

At the point of maximum likelihood the right hand side of (A5) vanishes and therefore at this point of maximum likelihood the right hand side of (A22) becomes $-\frac{2n}{\sigma^2}$.

Differentiating (A5) with respect to κ gives

$$\text{(A23)} \qquad \frac{\partial^2 \ell}{\partial \kappa \partial \sigma} = \frac{\partial}{\partial \kappa} \left\{ -\frac{1}{\sigma} \sum_{i=1}^{n} \left(1 - \frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right) \right\}$$
$$= \frac{1}{\sigma} \sum_{i=1}^{n} \frac{\partial}{\partial \kappa} \left(\frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right).$$

At the point of maximum likelihood the right hand side of (A9) vanishes and therefore at this point of maximum likelihood we have

$$\text{(A24)} \qquad \sum_{i=1}^{n} \frac{\partial}{\partial \kappa} \left(\frac{(r_{t_i} - m_{t_{i-1}}(t_i))^2}{v_{t_{i-1}}(t_i)} \right) = \sum_{i=1}^{n} \frac{1}{v_{t_{i-1}}(t_i)} \frac{\partial v_{t_{i-1}}(t_i)}{\partial \kappa}.$$

Thus (A23) becomes

$$\text{(A25)} \qquad \frac{\partial^2 \ell}{\partial \kappa \partial \sigma} = -\frac{1}{\sigma} \sum_{i=1}^{n} \frac{1}{v_{t_{i-1}}(t_i)} \frac{\partial v_{t_{i-1}}(t_i)}{\partial \kappa}.$$

From (A13) we have

$$\text{(A26)} \qquad \frac{1}{v_{t_{i-1}}(t_i)} \frac{\partial v_{t_{i-1}}(t_i)}{\partial \kappa} = -\frac{1}{\kappa} + \frac{t_i - t_{i-1}}{\kappa} \frac{(1 - \kappa B(t_{i-1}, t_i))^2}{B(t_{i-1}, t_i) - \frac{1}{2}\kappa B(t_{i-1}, t_i)^2}$$

$\square$

References

D.H. Ahn and B. Gao. A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12:721–762, 1999.

Y. Aït-Sahalia. Testing continuous time models of the spot interest rate. *Review of Financial Studies*, 9(2):385–426, 1996.

Y. Aït-Sahalia. Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance*, 54:136195, 1999.

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

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

D. Brigo and F. Mercurio. *Interest Rate Models – Theory and Practice*. Springer Finance, 2001.

D. Brigo and F. Mercurio. *Interest Rate Models – Theory and Practice*. Springer Finance, second edition, 2006.

K. P. Burnham and D. R. Anderson. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33:261–304, 2004.

K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3):1209–1227, 1992.

D. Chapman, J. Long, and Pearson N. Using proxies for the short-rate: When are three months like an instant? *Review of Financial Studies*, 12(4):763–806, 1999.

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

R. Faff and P. Gray. On the estimation and comparison of short-rate models using the generalised method of moments. *Journal of banking and finance*, 30:3131–3146, 2006.

J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice-Hall, third edition, 1997.

J. C. Hull and A. White. Pricing interest rate derivative securities. *The Review of Financial Studies*, 3:573–592, 1990.

M. Kendall and A. Stuart. Inference and relationship. In *Vol. 2 of The Advanced Theory of Statistics*. Charles Griffin, London, 1961.

R. S. Liptser and A. Shiryaev. *Statistics of Random Processes II*. Springer Verlag, Berlin Heidelberg, second edition, 2001.

E. Platen. A short term interest rate model. *Finance and Stochastics*, 3:215–225, 1999.

R. Poulsen. Approximate maximum likelihood estimationg of discretely observed diffusion processes. Working Paper No. 29, University of Aarhus, 1999.

R. Rebonato. *Interest-rate option models*. Wiley Series in Financial Engineering. John Wiley and Sons, second edition, 1998.

R. Shiller. *Market Volatility*. The MIT Press, Cambridge, Massachusetts, 1989.

M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

S. Treepongkaruna. Quasi-maximum likelihood estimates of Kiwi short-term interest rate. *Applied Economics Letters*, 10:937942, 2003.

O. A. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

J. Yu and P. Phillips. A Gaussian approach for continuous time models of the short-term interest rate. *Econometrics Journal*, 4:210–224, 2001.

(K. Fergusson) Curtin University, Kent Street, Bentley WA 6102, Australia
E-mail address, K. Fergusson: `kevin.fergusson@curtin.edu.au`

(E. Platen) University of Technology, Sydney, PO Box 123, Broadway NSW 2007, Australia
E-mail address, E. Platen: `eckhard.platen@uts.edu.au`