Kevin Fergusson*

Fast maximum likelihood estimation of parameters for square root and Bessel processes

<https://doi.org/10.1515/snde-2019-0079>

Received July 25, 2019; accepted June 30, 2020; published online August 12, 2020

Abstract: Explicit formulae for maximum likelihood estimates of the parameters of square root processes and Bessel processes and first and second order approximate sufficient statistics are supplied. Applications of the estimation formulae to simulated interest rate and index time series are supplied, demonstrating the accuracy of the approximations and the extreme speed-up in estimation time. This significantly improved run time for parameter estimation has many applications where ex-ante forecasts are required frequently and immediately, such as in hedging interest rate, index and volatility derivatives based on such models, as well as modelling credit risk, mortality rates, population size and voting behaviour.

Keywords: Bessel process; Cox-Ingersoll-Ross model; maximum likelihood estimation; squared Bessel process.

1 Introduction

Fast and accurate statistical methods in finance and other areas of applied probability are becoming increasingly important due to the enormous amounts of data available that have to be analysed. We have extremely fast and accurate methods for modelling Gaussian dynamics, where in a continuous-time setting linearly transformed Brownian motions drive model dynamics, for example, as for the logarithm of an asset price under the Black-Scholes model (see Black and Scholes (1973)) or as for the short term interest rate under the Vasicek model. This seems to be not the case for next-generation models that are based on Bessel processes and square root processes like the CIR model in finance. In population size modelling (see e.g. Feller (1971)) the square root process or the squared Bessel process emerges naturally as a suitable model. Similarly, the voting behaviour seems to be well modelled, similar to population size, by a squared Bessel process; see Karlin and Taylor (1981). Also, in finance the minimal market model (see Platen (1997)) emerges naturally as a suitable model. Interest has evolved also for the $3/2$ stochastic volatility model (see Platen (1997) and Heston (1997)) which appears to be an excellent volatility model; see Goard and Mazur (2013). Furthermore, square root processes have many relevant applications in credit risk modelling (see Bielecki, Jeanblanc, and Rutkowski (2011) for an application to intensity-based models) and in stochastic mortality modelling (see Biffis (2005) and Blackburn and Sherris (2013)).

Various methods have been proposed to estimate the parameters of the model which fit a given time series of observed short rates. For example, in Poulsen (1999) a lognormal approximation is employed. In Fergusson and Platen (2015) the parameters of the CIR short rate model were estimated using the maximum likelihood method using the Newton-Raphson iterative method applied to both the log-likelihood function and an approximate log-likelihood function based on the lognormal distribution. In Fergusson and Platen (2015) the parameters of the $3/2$ short rate model were also estimated using a gradient ascent method. The asymptotic properties of the Heston model are studied in Barczy, Pap, and Szabo (2016), which involves modelling the

*Corresponding author: Kevin Fergusson, Economics, University of Melbourne, Melbourne, Australia,
e-mail: kevin.fergusson@unimelb.edu.au. <https://orcid.org/0000-0002-7584-430X>

volatility process with the square root model in (3). Their estimates, based on discrete-time observations, are derived from conditional least squares estimates of modified parameters, assuming that the diffusion parameter σ is known. On the other hand, in Dokuchaev (2017) estimators of σ are supplied that do not require knowledge of the drift parameters. In Overbeck and Rydén (1997), estimators of parameters of (3), employing observations at equispaced times and based on conditional least squares, are studied. Another approach involving quasi-maximum likelihood estimation via the Wagner-Platen approximation is introduced in Huang (2013). However, in the current paper we supply closed-form approximate formulae for the maximum likelihood estimates of the parameters of the square root process and, consequently, transformations of this, including squared Bessel processes and Bessel processes.

In Section 2 we describe the probabilistic framework for our stochastic processes. In Section 3 we state theorems for closed-form approximate formulae for the parameters $\bar{r}$, κ and σ of (3) based on second order and first order approximations to the log-likelihood function. A corollary is supplied which provides estimates of (5). In Section 4 we demonstrate applications of the theorem to simulated data sets of squared Bessel processes and to an actual short rate data set, exemplifying the accuracy of the formulae. Finally, in Section 5 we conclude with some comments.

2 Probabilistic framework

A Bessel process is specified by the stochastic differential equation (SDE)

$$dR_t = \left(\frac{\alpha}{R_t} + \beta R_t \right) dt + \gamma dZ_t, \quad (1)$$

where $Z = (Z_t)_{t \geq 0}$ is a Wiener process adapted to a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}, P)$, $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ is a family of sigma algebras, $R_0 = x$ is the initial value and α , β and γ are constants. The case when these real-valued functions are Markov chains has been considered in Elliott and Platen (2001), but we confine ourselves to the case when they are constants.

Letting $r_t = R_t^2$ and using the Itô formula, we have the SDE for the squared Bessel process, otherwise termed a square root process, as

$$dr_t = (2\alpha + \gamma^2 + 2\beta r_t) dt + 2\gamma \sqrt{r_t} dZ_t, \quad (2)$$

where $r_0 = x^2 > 0$ is the positive initial value.

A common application of the square root process to finance, is modelling the short rate, which is the continuously compounded rate of interest paid on a cash deposit for an infinitesimally small unit of time. In reality, the short rate is represented by short-term rates such as the overnight cash rate, which is targeted by central banks, or money market rates having terms up to 12 months. The Cox-Ingersoll-Ross (CIR) model of the short rate, given in Cox Ingersoll, and Ross (1985), is specified by the SDE

$$dr_t = \kappa(\bar{r} - r_t) dt + \sigma \sqrt{r_t} dZ_t, \quad (3)$$

where $Z = (Z_t)_{t \geq 0}$ is as before and $\bar{r}$, κ and σ are positive constants. It is readily seen that (3) is the particular case of a squared Bessel process in (2) with

$$\begin{aligned} \alpha &= \frac{1}{2} \bar{r} \kappa - \frac{1}{8} \sigma^2, \\ \beta &= -\frac{1}{2} \kappa, \\ \gamma &= \frac{1}{2} \sigma. \end{aligned} \quad (4)$$

A zero-valued short rate is avoided because we set $\kappa \bar{r} > \frac{1}{2}\sigma^2 > 0$. The parameter κ denotes the speed of reversion of the short rate r_t to the mean reverting level $\bar{r}$. The $\bar{r}$ can be thought of as a smoothed average short rate which is targeted by the central bank.

Related to the CIR short rate model is the 3/2 short rate model, derived in Platen (1997, 1999) and subsequently studied in Ahn and Gao (1999),

$$dr_t = (p r_t + q r_t^2) dt + \sigma(r_t)^{3/2} dZ_t, \quad (5)$$

where $p > 0$, q and $\sigma > 0$ are constants satisfying $q < \sigma^2/2$ to avoid explosive values of r_t . The SDE for the reciprocal of the 3/2 short rate is that for the CIR short rate model and, therefore, maximum likelihood estimation of the parameters of the 3/2 model is equivalent to that for the CIR model.

Both the CIR model and 3/2 model have the property that for the above conditions on the parameters the interest rates can never be negative.

The transition density function of the short rate process in (3) is that of the non-central chi-square distribution, which, for times s and t with $0 \leq s < t$, is given by

$$p_r(s, r_s, t, r_t) = \frac{1}{2(\varphi_t - \varphi_s) e^{-\kappa t}} \left(\frac{r_t e^{\kappa t}}{r_s e^{\kappa s}} \right)^{\frac{1}{2}(\frac{\nu}{2}-1)} \times \exp\left(-\frac{1}{2} \frac{r_s e^{\kappa s} + r_t e^{\kappa t}}{\varphi_t - \varphi_s}\right) I_{\frac{\nu}{2}-1}\left(\frac{\sqrt{r_s r_t} e^{\kappa(s+t)}}{\varphi_t - \varphi_s}\right), \quad (6)$$

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

$$I_\nu(x) = \sum_{i=0}^{\infty} \frac{1}{i! \Gamma(i + \nu + 1)} \left(\frac{x}{2}\right)^{2i+\nu} \quad (7)$$

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

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

$$\frac{e^{\kappa t}}{\varphi_t - \varphi_s} r_t \quad (8)$$

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 $\lambda = r_s e^{\kappa s} / (\varphi_t - \varphi_s)$, namely

$$\frac{e^{\kappa t}}{\varphi_t - \varphi_s} r_t \sim \chi_{\nu, r_s e^{\kappa s} / (\varphi_t - \varphi_s)}^2. \quad (9)$$

Let $t_0, t_1, t_2, \dots, t_n$ be $n+1$ equispaced ordered times at which the short rate is observed, where the spacing is $\Delta = t_i - t_{i-1}$, for $i = 1, 2, \dots, n$. We assume that the observed short rates

$$r_{t_0}, r_{t_1}, \dots, r_{t_n} \quad (10)$$

are all positive. The log-likelihood function which we seek to maximise with respect to the parameters $\bar{r}$, κ and σ is given by

$$\ell(\bar{r}, \kappa, \sigma) = \sum_{i=1}^n \log p_r(t_{i-1}, r_{t_{i-1}}, t_i, r_{t_i}), \quad (11)$$

where the transition density function p_r is as in (6).

3 Main theorem and corollary

Let us first state and prove the main theorem of this paper, which concerns parameter estimates of the square root process based on a second order approximation to the log-likelihood function.

Before doing this, we motivate our definition of an approximately sufficient statistic by recalling the following definition of a sufficient statistic.

Definition 1: In respect of sampled data $r_{t_0}, r_{t_1}, \dots, r_{t_n}$ from the stochastic process $r = (r_t)_{t \geq 0}$, having parameter $\theta \in \mathbb{R}^d$, a statistic $T = T(r_{t_0}, r_{t_1}, \dots, r_{t_n}) \in \mathbb{R}^k$ is said to be sufficient if knowledge of the value of T allows us to specify the likelihood function $L(\theta)$ determined by $r_{t_0}, r_{t_1}, \dots, r_{t_n}$, up to a constant of proportionality which does not depend on the parameter θ .

Remark 1: We see that a sufficient statistic T constitutes a sufficient reduction of the data for the purpose of calculating the maximum likelihood estimates. Indeed, this is a consequence of the Fisher-Neyman factorisation criterion which characterises a sufficient statistic for a probability distribution having density $f_\theta(x)$ as one for which $f_\theta(x)$ factorises as $f_\theta(x) = g_\theta(T)h(x_0, \dots, x_n)$, where g_θ is a function of T depending on θ , and h is a function independent of θ .

In our formulation of the main theorem we require a definition of an approximate sufficient statistic, where we are interested in the degree to which a stochastic model should be modified to make the statistic sufficient; see McCullagh (1984) for a discussion of local sufficiency or Chapter 5 of Le Cam (1986) for a discussion of approximate sufficiency, for example.

Definition 2: In respect of equispaced sampled data $r_{t_0}, r_{t_1}, \dots, r_{t_n}$, with time step $\Delta = t_1 - t_0$, from the stochastic process $r = (r_t)_{t \geq 0}$, having parameter $\theta \in \mathbb{R}^d$, a statistic $T = T(r_{t_0}, r_{t_1}, \dots, r_{t_n}) \in \mathbb{R}^m$, for some $m \in \mathbb{Z}^+$, is said to be approximately sufficient of order k if knowledge of the value of T allows us to specify the θ -dependent part of the likelihood function determined by $r_{t_0}, r_{t_1}, \dots, r_{t_n}$ up to a tolerance $O(n\Delta^{k+1})$, as $\Delta \rightarrow 0$. Mathematically speaking, there exist a function $g_\theta(T)$, which depends on the statistic T and parameter θ , a function $h(r_{t_0}, r_{t_1}, \dots, r_{t_n})$, which is independent of θ , and positive constants K_0 and K_1 such that for all $\Delta \in [0, K_1]$,

$$|\log L(r_{t_0}, r_{t_1}, \dots, r_{t_n}; \theta) - \log g_\theta(T(r_{t_0}, r_{t_1}, \dots, r_{t_n})) - \log h(r_{t_0}, r_{t_1}, \dots, r_{t_n})| \leq K_0 n \Delta^{k+1}.$$

Theorem 1: Given the observations of the square root process in (10), the maximum likelihood estimates of the parameters $\bar{r}, \kappa, \sigma$ of the model specified in (3) are approximately given by, as $\Delta \rightarrow 0$,

$$\begin{aligned} \hat{\bar{r}} &= \frac{2(v+1)a}{e^k - e^{-k}}, \\ \hat{\kappa} &= 2k\frac{1}{\Delta}, \\ \hat{\sigma} &= \sqrt{\frac{8ak}{(e^k - e^{-k})\Delta}}, \end{aligned} \quad (12)$$

respectively, subject to the condition

$$(q'_0)^2 - 2q_0 q''_0 > 0, \quad (\text{Condition A})$$

where

$$\begin{aligned} k &= \frac{-q'_0 + \text{sign}(q'_0) \sqrt{(q'_0)^2 - 2q_0 q''_0}}{q'_0}, \\ q_0 &= \frac{1}{2} R_5 b_0^2 - \left(\frac{1}{4} R_5 f_0 - R_3 \right) b_0 c_0 - \left(\frac{1}{2} R_3 f_0 - h_0 \right) c_0^2, \\ q'_0 &= R_5 b_0 b'_0 - \left\{ \frac{1}{4} R_5 f'_0 b_0 c_0 + \left(\frac{1}{4} R_5 f_0 - R_3 \right) (b'_0 c_0 + b_0 c'_0) \right\} \\ &\quad - \left\{ \left(\frac{1}{2} R_3 f'_0 - h'_0 \right) c_0^2 + 2 \left(\frac{1}{2} R_3 f_0 - h_0 \right) c_0 c'_0 \right\}, \end{aligned} \quad (13)$$

$$\begin{aligned}
q'' &= R_5 \{ b_0 b_0'' + (b_0')^2 \} - \frac{1}{4} R_5 (f_0' b_0 c_0) (f_0'' / f_0' + b_0' / b_0 + c_0' / c_0) \\
&\quad - \left(\frac{1}{4} R_5 - R_3 \right) (b_0'' c_0 + 2 b_0' c_0' + b_0 c_0'') - \left(-\frac{1}{2} R_3 f_0'' - h_0' \right) c_0^2 \\
&\quad - 4 \left(-\frac{1}{2} R_3 f_0' - h_0' \right) c_0 c_0' - 2 \left(-\frac{1}{2} R_3 f_0 - h_0 \right) (c_0 c_0'' + (c_0')^2), \\
a &= \frac{b_k}{c_k}, \\
v &= \frac{k + \frac{1}{2}L}{a R_3 + \frac{1}{2}a^2 R_5}, \\
b_k &= \frac{3}{8} R_3 f_k^2 + \left(\frac{3}{4} \frac{R_3^2}{R_5} - \frac{5}{4} h_k \right) f_k - \frac{3}{2} \frac{R_3}{R_5} h_k + g_k, \\
c_k &= -\frac{1}{16} R_5 f_k^2 + \frac{5}{8} R_3 f_k + 1 - \frac{3}{2} h_k + \frac{3}{2} \frac{R_3^2}{R_5}, \\
f_k &= -R_0 e^{-k} + R_1 e^k, \\
g_k &= R_0 e^{-k} + R_1 e^k - 2R_2, \\
h_k &= k + \frac{1}{2}L
\end{aligned}$$

and where we have the second order approximate sufficient statistics L, R_i for $i \in \{1, 2, 3, 5\}$

$$L = \frac{1}{n} \log(r_{t_n}/r_{t_0}), \quad (14)$$

$$R_0 = \frac{1}{n} \sum_{i=1}^n r_{t_{i-1}}, \quad (15)$$

$$R_1 = \frac{1}{n} \sum_{i=1}^n r_{t_i}, \quad (16)$$

$$R_2 = \frac{1}{n} \sum_{i=1}^n \sqrt{r_{t_{i-1}} r_{t_i}}, \quad (17)$$

$$R_3 = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{r_{t_{i-1}} r_{t_i}}}, \quad (18)$$

$$R_4 = \frac{1}{n} \sum_{i=1}^n \log(r_{t_{i-1}} r_{t_i}), \quad (19)$$

$$R_5 = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_{t_{i-1}} r_{t_i}}, \quad (20)$$

$$R_6 = \frac{1}{n} \sum_{i=1}^n \frac{1}{(r_{t_{i-1}} r_{t_i})^{3/2}}. \quad (21)$$

Also, the Fisher information matrix is approximately given by

$$\mathcal{J}(\bar{r}, \kappa, \sigma) \approx B^T \mathcal{J}(a, k, v) B, \quad (22)$$

where

$$a = \frac{\sigma^2}{4\kappa} \{e^{\kappa\Delta/2} - e^{-\kappa\Delta/2}\},$$

$$k = \frac{1}{2} \kappa \Delta, \quad (23)$$

$$\nu = \frac{2\bar{r}\kappa}{\sigma^2} - 1,$$

$$B = \begin{pmatrix} 0 & \sigma^2 \kappa \Delta^3 / 48 & \sigma \Delta / 2 \\ 0 & \Delta / 2 & 0 \\ 2\kappa / \sigma^2 & 2\bar{r} / \sigma^2 & -4\bar{r}\kappa / \sigma^3 \end{pmatrix}, \quad (24)$$

and

$$\mathcal{J}(a, k, \nu) = -n \begin{pmatrix} \frac{1}{2a^2} - \frac{1}{a^3} \{R_0 e^{-k} + R_1 e^k - 2R_2\} & \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -R_3 \nu \\ \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -\frac{1}{2a} \{R_0 e^{-k} + R_1 e^k\} & 1 \\ -R_3 \nu & 1 & -a R_3 \end{pmatrix}. \quad (25)$$

Finally, we have that as $n \rightarrow \infty$,

$$\begin{pmatrix} \hat{\bar{r}} \\ \hat{\kappa} \\ \hat{\sigma} \end{pmatrix} - \begin{pmatrix} \bar{r} \\ \kappa \\ \sigma \end{pmatrix} \rightarrow N(0, \mathcal{J}(\bar{r}, \kappa, \sigma)^{-1}). \quad (26)$$

Proof: See Appendix A for a proof.

The following theorem supplies parameter estimates of the square root process using a first order approximation to the log-likelihood function.

Theorem 2: Given the observations of the short rate in (10), the maximum likelihood estimates of the parameters $\bar{r}$, κ , σ of the CIR model specified in (3) are approximately given by, as $\Delta \rightarrow 0$,

$$\hat{\bar{r}} = \frac{2(\nu+1)a}{e^k - e^{-k}},$$

$$\hat{\kappa} = 2k_{\Delta}, \quad (27)$$

$$\hat{\sigma} = \sqrt{\frac{8ak}{(e^k - e^{-k})\Delta}},$$

respectively, subject to the condition

$$(p_0')^2 - 2p_0 p_0'' > 0, \quad (\text{Condition A'})$$

where

$$k = \frac{-p_0' + \text{sign}(p_0') \sqrt{(p_0')^2 - 2p_0 p_0''}}{p_0''}, \quad (28)$$

$$p_0 = \frac{1}{4} \left(\frac{1}{2} R_3 f_0 - h_0 \right)^2 - \left(\frac{1}{2} R_3 f_0 - h_0 \right) + R_3 g_0 - h_0^2$$

$$p_0' = \frac{1}{2} \left(\frac{1}{2} R_3 f_0' - h_0' \right) \left(\frac{1}{2} R_3 f_0 - h_0 \right) - \left(\frac{1}{2} R_3 f_0' - h_0' \right) + R_3 g_0' - 2h_0' h_0$$

$$p''_0 = \frac{1}{2} \left(\frac{1}{2} R_3 f''_0 - h''_0 \right) \left(\frac{1}{2} R_3 f_0 - h_0 \right) + \frac{1}{2} \left(\frac{1}{2} R_3 f'_0 - h'_0 \right)^2 - \left(\frac{1}{2} R_3 f''_0 - h''_0 \right) + R_3 g''_0 - 2(h''_0 h_0 + (h'_0)^2), \quad (29)$$

$$a = \frac{1}{2} f_k - \frac{h_k}{R_3},$$

$$v = \frac{h_k}{a R_3},$$

$$p_k = \frac{1}{4} \left(\frac{1}{2} R_3 f_k - h_k \right)^2 - \left(\frac{1}{2} R_3 f_k - h_k \right) + R_3 g_k - h_k^2,$$

$$f_k = -R_0 e^{-k} + R_1 e^k,$$

$$g_k = R_0 e^{-k} + R_1 e^k - 2R_2,$$

$$h_k = k + \frac{1}{2} L \quad (30)$$

and where we have the first order approximate sufficient statistics L, R_i for $i \in \{1, 2, 3\}$ given in (14)–(18).

Also, the Fisher information matrix is given approximately in (22). Finally, asymptotic normality, as $n \rightarrow \infty$, is as in (26).

Proof: See Appendix A for a proof.

The following corollary to Theorem 1 is straightforwardly established.

Corollary 1: Given the observations of the short rate in (10), the maximum likelihood estimates of the parameters p, q and σ of the 3/2 short rate model given in (5) are

$$\begin{aligned} \hat{p} &= \hat{\kappa}, \\ \hat{q} &= \hat{\sigma}^2 - \hat{r}\hat{\kappa}, \end{aligned} \quad (31)$$

where formulae for $\hat{r}, \hat{\kappa}$ and $\hat{\sigma}$, based upon the reciprocals of the short rates, are supplied in Theorem 1.

4 Applications

In this section we supply various aspects of the estimation methods. Firstly, we check the ability of the estimated parameters from simulated data to recover the actual parameter values, for varying lengths of the series. Secondly, we fit the models to US interest rate data. Thirdly, we compare the run times with the Newton-Raphson method.

4.1 Estimation from simulated data

Here we simulate data sets from each of the processes, squared-Bessel, 3/2 and Bessel processes, and estimate its parameters over varying window lengths and time steps Δ .

4.1.1 CIR model

Commencing with $r_0 = 0.05$, we simulate with a daily time step, $\Delta = 1/365$, the CIR model having parameter values $\bar{r} = 0.0427, \kappa = 0.095$ and $\sigma = 0.06439$, which correspond closely to those estimated for the CIR model in Table 1. We compare estimation of parameters over varying window lengths, from one quarter to 150 years, and

time steps $\Delta \in \left\{\frac{1}{4}, \frac{1}{12}, \frac{1}{52}, \frac{1}{365}\right\}$. The evolutions of estimates of the parameters $\bar{r}$, κ , σ and $\nu = 4\bar{r}\kappa/\sigma^2$ are depicted respectively in Figures 1–4, varying the time steps. We maintain shorter time horizons in these figures to allow the differences in estimated parameter values among the various time steps to be shown clearly. Because the effects of varying the time steps on estimated values become less pronounced over the long term, making them almost indistinguishable, in Figure 5 we show only the estimates of the parameters using $\Delta = 1/365$. This allows for a comparison of the behaviour of estimated values of different parameters over various window lengths up to 150 years.

In Figure 1 we show the effect of Δ on the estimate of $\bar{r}$. For $\Delta = 1/52$ and $\Delta = 1/365$ the estimates are almost indistinguishable. For $\Delta = 1/4$ the estimates with window lengths of less than 3 years are wildly unstable, which is to be expected when using so few data points. It is not until around 3 years, corresponding to 12 data

Table 1: Parameter estimates of the CIR, 3/2 and Bessel process short rate models fitted to monthly US short rates via several methods.

Model	Method	Log-likelihood	Parameter estimates
Vasicek	Exact formula	7109.33145040	$\hat{r} = 0.043648$ (0.008476) $\hat{\kappa} = 0.149909$ (0.045259) $\hat{\sigma} = 0.015437$ (0.000261)
CIR	Newton-Raphson	7559.08057481	$\hat{r} = 0.042712$ (0.011234) $\hat{\kappa} = 0.094537$ (0.034902) $\hat{\sigma} = 0.064243$ (0.001083) $\hat{\nu} = 3.913439$ (1.067639)
CIR	1st order approx	7559.08042386	$\hat{r} = 0.042724$ (0.011232) $\hat{\kappa} = 0.094962$ (0.035044) $\hat{\sigma} = 0.064244$ (0.001083) $\hat{\nu} = 3.931961$ (1.067294)
CIR	2nd order approx	7559.08057479	$\hat{r} = 0.042712$ (0.011234) $\hat{\kappa} = 0.094543$ (0.034907) $\hat{\sigma} = 0.064243$ (0.001083) $\hat{\nu} = 3.913676$ (1.067673)
3/2	Newton-Raphson	7213.56087570	$\hat{p} = 0.110247$ (0.038861) $\hat{q} = -0.019455$ (1.216960) $\hat{\sigma} = 2.251059$ (0.037980) $\hat{\nu} = 4.015358$ (0.960646)
3/2	1st order approx	7213.56078067	$\hat{p} = 0.110624$ (0.038916) $\hat{q} = -0.036201$ (1.220414) $\hat{\sigma} = 2.251093$ (0.037982) $\hat{\nu} = 4.028575$ (0.963344)
3/2	2nd order approx	7213.56087569	$\hat{p} = 0.110250$ (0.038894) $\hat{q} = -0.019558$ (1.219089) $\hat{\sigma} = 2.251059$ (0.037980) $\hat{\nu} = 4.015439$ (0.962326)
Bessel	Newton-Raphson	7168.45470719	$\hat{\alpha} = 0.000151$ (0.000015) $\hat{\beta} = -0.106942$ (0.026841) $\hat{\gamma} = 0.016822$ (0.000403) $\hat{\nu} = 2.068122$ (0.092544)
Bessel	1st order approx	7162.69765360	$\hat{\alpha} = 0.000140$ (0.000014) $\hat{\beta} = -0.096360$ (0.024824) $\hat{\gamma} = 0.015654$ (0.000314) $\hat{\nu} = 2.146483$ (0.101064)
Bessel	2nd order approx	7164.59619582	$\hat{\alpha} = 0.000138$ (0.000014) $\hat{\beta} = -0.096364$ (0.025122) $\hat{\gamma} = 0.015836$ (0.000324) $\hat{\nu} = 2.100565$ (0.100157)

Figure 1: In respect of the CIR model, a graph of the evolution of the estimate of $\bar{r}$ from 1/4 to 20 years of simulated daily data.

points, that we see some stability. For $\Delta = 1/12$ the estimates of $\bar{r}$ are unstable over window lengths less than one year, corresponding to fewer than 12 data points. We see from Figure 5a that window lengths of many years are required to estimate the true value of $\bar{r}$. In fact, after 25 years the estimates remain within about 50% of the true value and accurate estimates of $\bar{r}$ are achieved for windows around 100 years in length.

In Figure 2 we show the effect of Δ on the estimate of κ . Beyond window lengths of about 10 years all four time steps give rise to similar estimated parameter values. However, for window lengths of less than 10 years all time steps fail to give good estimates of κ . We see from Figure 5b that the 95% confidence intervals contain the true value of κ for window lengths of at least 10 years. Also, it is not until around 25 years when we see estimates of κ being within about 80% of the true value and until around 85 years for estimates to be within roughly 50% of the true value.

In Figure 3 we show the effect of Δ on the estimate of σ . For $\Delta = 1/4$, the estimates stabilise after two and a half years, corresponding to 10 or more data points. For $\Delta = 1/12$, estimates stabilise after one year, namely for 12 or more data points. For $\Delta = 1/52$, estimates stabilise after half a year, which is for 26 or more data points. Finally, for $\Delta = 1/365$, the estimates are stable over windows at least as long as one quarter of a year. We see in Figure 5c that estimates of σ are within 0.001 of the true value over horizons beyond 10 years and are almost exact to four decimal places after 50 years.

In Figure 4 we show the effect of Δ on the estimate of $\nu = 4\bar{r}\kappa/\sigma^2$, which is the degrees of freedom parameter and which is crucial in ascertaining whether the process avoids hitting zero, namely when $\nu \geq 2$. For all $\Delta \in \left\{\frac{1}{4}, \frac{1}{12}, \frac{1}{52}, \frac{1}{365}\right\}$, the estimates are unstable for windows shorter than two and a half to three years, after which the estimates provide similar values. In Figure 5d we see that the estimates do not appear to converge to the true value until around 30 years. Yet, even for windows longer than 30 years the estimates hover above the true value by 1 to 1.25, which is roughly 25% of the true value.

In conclusion, σ and $\bar{r}$ are the only parameters which can be accurately estimated over windows shorter than 150 years. With daily sampling of the process, good estimates of σ can be obtained in about half a year and with weekly sampling, good estimates are obtainable in about two years. For windows longer than 10 years, quarterly, monthly, weekly and daily sampling give rise to estimates of similar accuracy. With such accurate estimates of σ , we see from the relation $\sigma^2 = 4\bar{r}\kappa/\nu$ that inaccuracies in any one of the parameters $\bar{r}$, κ and ν will result to some extent in inaccuracies in the other two. This is exemplified in Figures 1, 2 and 4 and, over the longer term, in Figure 5. In practical applications, such as when modelling interest rates, the choice of Δ will depend on the window length, that is the length in years of the data series, all other variables being equal. Assuming interest rate cycles of around 20–40 years, having 50–100 years of data generally allows sufficient time for interest rates to revert, meaning that $\bar{r}$ and κ , in particular, can be estimated fairly reliably, and meaning that using $\Delta = 1/4$, or even $\Delta = 1$, will give good estimates. On the other hand, having only five years of data, using $\Delta \leq 1/52$ will give reliable estimates only for σ , with $\bar{r}$ and κ having large standard errors.

Figure 2: In respect of the CIR model, a graph of the evolution of the estimate of κ from 1/4 to 20 years of simulated daily data.

Figure 3: In respect of the CIR model, a graph of the evolution of the estimate of σ from 1/4 to 20 years of simulated daily data.

Figure 4: In respect of the CIR model, a graph of the evolution of the estimate of $v = 4\bar{\kappa}\kappa/\sigma^2$ from 1/4 to 20 years of simulated daily data.

4.1.2 3/2 model

Commencing with $r_0 = 0.05$, we simulate with a daily time step, $\Delta = 1/365$, the 3/2 model having parameter values $p = 0.11$, $q = -0.019$ and $\sigma = 2.25093$, which correspond closely to those estimated for the 3/2 model in Table 1. We compare estimation of parameters over the same varying window lengths and time steps examined

(a) Estimates of $\bar{r}$. (b) Estimates of κ . (c) Estimates of σ . (d) Estimates of ν .

Figure 5: In respect of the CIR model, graphs of the evolutions of the estimates of $\bar{r}$, κ , σ and ν from 1/4 to 150 years of simulated daily data. The dashed lines surrounding each of the evolutions of estimates provide a 95% confidence interval for the parameter value. The dashed horizontal red line shows the true value of the parameter.

in the CIR model. The evolutions of estimates of the parameters p , q , σ and $\nu = 4(1 - q/\sigma^2)$ are depicted respectively in Figures 6–9, varying the time steps. For the same reasons given for the CIR model, we maintain shorter time horizons in these figures, whereas in Figure 10 we show only the estimates of the parameters using $\Delta = 1/365$ over window lengths up to 150 years.

In Figure 6 we show the effect of Δ on the estimate of p . In (31) we see that p is equal to κ in the corresponding CIR model for the reciprocal of the $3/2$ process. Therefore the effects of Δ on $\hat{p}$ correspond well with those on $\hat{\kappa}$ shown in Figure 2. Beyond window lengths of about 10 years all four time steps give rise to similar estimated parameter values. However, for windows shorter than 10 years all time steps fail to give good estimates of p . We see from Figure 10a that the 95% confidence intervals contain the true value of p for window lengths of at least 10 years. Also, it is not until after around 25 years when we see estimates of p remaining within about 100% of the true value and until around 85 years for estimates to be within roughly 50% of the true value.

In Figure 7 we show the effect of Δ on the estimate of q . In (31) we see that q is equal to $\sigma^2 - \bar{r}\kappa$ in the corresponding CIR model, and this can be rewritten in terms of σ and ν as $\sigma^2(1 - \nu/4)$. Therefore, if $\hat{\sigma}$ is accurate then we would expect the effect of Δ on q to be similar to that on ν . Indeed Figure 10b appears to be a mirror image of Figure 5d. Beyond window lengths of about 10 years all four time steps give rise to similar estimated parameter values. We see from Figure 10b that windows at least around 25 years in length give estimates of q within 2 of the true value and windows at least around 85 years gives estimates within 1 of the true value.

In Figure 8 we show the effect of Δ on the estimate of σ , which are similar to those on σ in the corresponding CIR model. For windows longer than 14 years, the effects of Δ wear off. For $\Delta = 1/4$ and $\Delta = 12$, the estimates stabilise after 12 years or so, corresponding to at least 48, respectively 144, data points. For $\Delta = 1/52$, estimates stabilise after 2 years, corresponding to 104 or more data points. Finally, for $\Delta = 1/365$, the estimates are stable over windows at least as long as one quarter of a year. We see in Figure 10c that estimates of σ are within 0.05 of the true value after 5 years and are within 0.01 over horizons beyond 45 years.

In Figure 9 we show the effect of Δ on the estimate of $\nu = 4(1 - q/\sigma^2)$, which is the degrees of freedom parameter and which is crucial in ascertaining whether the process avoids hitting zero, namely when $\nu \geq 2$. For

Figure 6: In respect of the 3/2 model, a graph of the evolution of the estimate of p from 1/4 to 20 years of simulated daily data.

Figure 7: In respect of the 3/2 model, a graph of the evolution of the estimate of q from 1/4 to 20 years of simulated daily data.

Figure 8: In respect of the 3/2 model, a graph of the evolution of the estimate of σ from 1/4 to 20 years of simulated daily data.

all $\Delta \in \left\{ \frac{1}{4}, \frac{1}{12}, \frac{1}{52}, \frac{1}{365} \right\}$, the estimates are unstable for windows shorter than two and a half to three years, after which the estimates provide similar values. In Figure 10d we see that the estimates do not appear to converge to the true value until around 30 years. Yet, even for windows longer than 30 years the estimates hover above the true value by 1–1.25, which is roughly 25% of the true value.

Figure 9: In respect of the 3/2 model, a graph of the evolution of the estimate of $v = 4(1 - q/\sigma^2)$ from 1/4 to 20 years of simulated daily data.

In conclusion, σ is the most accurately estimated parameter over all window lengths up to 150 years, followed by p and lastly q . With daily sampling of the process, good estimates of σ can be obtained in about half a year and with weekly sampling, good estimates are obtainable in about 2 years. For windows longer than 10 years, quarterly, monthly, weekly and daily sampling give rise to estimates of similar accuracy. With such accurate estimates of σ , we see from the relation $q = \sigma^2(1 - v/4)$ that inaccuracies in q and v are linearly related. This is exemplified in Figures 7 and 9 and, over the longer term, in Figures 10b and 10d. In practical applications, the choice of Δ is similar to the choice for the CIR model.

4.1.3 Bessel model

Commencing with $r_0 = 0.05$, we simulate with a daily time step, $\Delta = 1/365$, the Bessel model having parameter values $\alpha = 0.000151$, $\beta = -0.107$ and $\gamma = 0.01682$, which correspond closely to those estimated for the Bessel model in Table 1. We compare estimation of parameters over the same varying window lengths and time steps examined in the CIR model. The evolutions of estimates of the parameters α , β , γ and $v = 1 + 2\alpha/\gamma^2$ are depicted respectively in Figures 11–14, varying the time steps. For the same reasons given for the CIR model, we maintain shorter time horizons in these figures, whereas in Figure 15 we show only the estimates of the parameters using $\Delta = 1/365$ over window lengths up to 150 years.

In Figure 11 we show the effect of Δ on the estimate of α , where stability in estimates is achieved in windows of around 5.7 years in length and longer, regardless of the time step. For shorter windows, the estimates all diverge from the true value by about 0.005 for windows longer than one year, and diverge in varying amounts for windows shorter than one year. We see from Figure 15a that estimates are within 0.00015 for window lengths between 7.5 and 130 years, and converging very close to the true value beyond 130 years.

In Figure 12 we show the effect of Δ on the estimate of β . For windows shorter than 5.7 years the estimates vary with Δ but fail to achieve any value consistently close to the true value. Beyond this, all values of Δ provide similarly valued estimates, within 0.1 of the true value of -0.107 . We see from Figure 15b that windows between 60 and 100 years in length provide estimates within 0.05 of the true value, and after 100 years the estimates are almost exact.

In Figure 13 we show the effect of Δ on the estimate of γ . For $\Delta = 1/365$ the estimates are very close to the true value, with estimates being almost indistinguishable from the true value after 7.5 years. The larger the value of Δ the worse the estimates are, although most of estimates improve beyond 7.5 years, except for $\Delta = 1/4$. We see in Figure 15c that estimates of γ , based on $\Delta = 1/365$, are within 0.0001 of the true value after 10 years and are within about 0.00005 beyond 50 years.

In Figure 14 we show the effect of Δ on the estimate of $v = 1 + 2\alpha/\gamma^2$, for the same reasons given for the CIR and 3/2 models. For all $\Delta \in \left\{\frac{1}{4}, \frac{1}{12}, \frac{1}{52}, \frac{1}{365}\right\}$, the estimates are unstable for windows shorter than 5.5 years, after

(a) Estimates of p . (b) Estimates of q . (c) Estimates of σ . (d) Estimates of v .

Figure 10: In respect of the 3/2 model, graphs of the evolutions of the estimates of p , q , σ and $v = 4(1 - q/\sigma^2)$ from 1/4 to 150 years of simulated daily data. The dashed lines surrounding each of the evolutions of estimates provide a 95% confidence interval for the parameter value. The dashed horizontal red line shows the true value of the parameter.

which the estimates provide similar values and achieving stability after 7.5 years. In Figure 15d we see that the estimates are within 1 of the true value beyond 10 years and within 0.1 beyond 133 years.

In conclusion, γ is the most accurately estimated parameter over all windows, particularly for $\Delta = 1/365$, which gives good estimates over windows about half a year in length. The other parameters can be estimated fairly reliably over longer windows. The choice of Δ is largely determined by the data window. For shorter windows using $\Delta = 1/365$ is imperative. However, for windows over several decades, quarterly sampling provides reliable estimates.

4.2 Estimation from historical interest rates

In this section we demonstrate the parameter estimation for each of the processes when used to model interest rates. For the set of monthly US short rates over the period 1871–2019, as given by the one-year rates in Shiller (2019) and supplemented by six-month LIBOR rates in Federal Reserve Bank of St. Louis (2019), we obtain the maximum likelihood estimates in Table 1. The Newton-Raphson approximation algorithm was used to compute the exact parameter estimates for each of the models, whereas the estimates derived using the first and second term approximations employ the formulae given in Theorem 1 and Corollary 1. The parameter estimates of the Vasicek model were used in determining an initial set of parameter estimates for the Newton-Raphson method applied to the CIR model, as described in Fergusson and Platen (2015).

We can draw several conclusions from Table 1, where the data series is monthly and nearly 150 years long. Firstly, simulations performed in Section 4.1.1 for the CIR model, having the same parameters used in Table 1, indicate that the minimum window length for reasonable estimates of $\bar{r}$ is around 30 years, as shown in Figure 5a. For κ , this is around 90 years, as shown in Figure 5b and for σ , this is around 14 years for monthly or less frequent observations, as shown in Figure 3. So based on the monthly frequency and 150 year series length, we would expect the parameter estimates for the CIR model in Table 1 to be reliable. Secondly, analogous simulations performed in Section 4.1.2 for the 3/2 model indicate that the minimum window length for estimates of p within

Figure 11: In respect of the Bessel model, a graph of the evolution of the estimate of α from 1/4 to 20 years of simulated daily data.

Figure 12: In respect of the Bessel model, a graph of the evolution of the estimate of β from 1/4 to 20 years of simulated daily data.

Figure 13: In respect of the Bessel model, a graph of the evolution of the estimate of γ from 1/4 to 20 years of simulated daily data.

0.1 of the true value is around 20 years, as shown in Figure 10a. For q being within 1 of its true value, this is around 90 years, as shown in Figure 10b and for σ , this is around 14 years, as shown in Figure 8. As for the CIR model, we would expect the parameter estimates for the 3/2 model in Table 1 to be reliable. Thirdly, analogous simulations performed in Section 4.1.3 for the Bessel model indicate that the minimum window length for estimates of α within 0.00006 of the true value is around 133 years, as shown in Figure 15a. For β being within 0.01 of its true

Figure 14: In respect of the Bessel model, a graph of the evolution of the estimate of $v = 1 + 2\alpha/\gamma^2$ from 1/4 to 20 years of simulated daily data.

(a) Estimates of α .

(b) Estimates of β .

(c) Estimates of γ .

(d) Estimates of v .

Figure 15: In respect of the Bessel model, graphs of the evolutions of the estimates of α , β , γ and v from 1/4 to 150 years of simulated daily data. The dashed lines surrounding each of the evolutions of estimates provide a 95% confidence interval for the parameter value. The dashed horizontal red line shows the true value of the parameter.

value, this is around 101 years, as shown in Figure 15b and for γ being within 0.001 of its true value, this is around 8 years, as shown in Figure 13. As for the two previous models, we would expect the parameter estimates for the Bessel model in Table 1 to be reliable. In general, the minimum series length is around 100 years for estimating parameters of any of our models, for all choices of $\Delta \leq 1/4$. Finally, using either the log-likelihood, AIC or BIC as the criterion, the CIR model is the best of the three models.

To illustrate the estimation methods for more frequent observations, for the set of daily US Effective Federal Funds rates over the period July 1954 to January 2020, as given in Federal Reserve Bank of St. Louis (2020), we obtain the maximum likelihood estimates in Table 2. We note that the second order approximation in respect of the Bessel model fails due to Condition A not being met.

Table 2: Parameter estimates of the CIR, 3/2 and Bessel process short rate models fitted to daily Effective Federal Funds rates via several methods.

Model	Method	Log-likelihood	Parameter estimates
Vasicek	Exact formula	102991.65	$\hat{r} = 0.048432$ (0.005261) $\hat{\kappa} = 1.476785$ (0.510347) $\hat{\sigma} = 0.062912$ (0.000291)
CIR	Newton-Raphson	108524.04	$\hat{r} = 0.048446$ (0.005710) $\hat{\kappa} = 1.061058$ (0.126698) $\hat{\sigma} = 0.292883$ (0.001342) $\hat{v} = 2.397021$ (0.157883)
CIR	1st order approx	108524.03	$\hat{r} = 0.048451$ (0.010082) $\hat{\kappa} = 1.064971$ (0.252687) $\hat{\sigma} = 0.292780$ (0.001343) $\hat{v} = 2.407777$ (0.176680)
CIR	2nd order approx	108524.04	$\hat{r} = 0.048450$ (0.007161) $\hat{\kappa} = 1.060475$ (0.170004) $\hat{\sigma} = 0.292790$ (0.001341) $\hat{v} = 2.397399$ (0.163386)
3/2	Newton-Raphson	81391.08	$\hat{p} = 4.854270$ (0.371083) $\hat{q} = 400.427422$ (4.604012) $\hat{\sigma} = 31.562020$ (0.144021) $\hat{v} = 2.392118$ (0.023049)
3/2	1st order approx	81391.06	$\hat{p} = 4.876184$ (0.350792) $\hat{q} = 396.963983$ (0.000000) $\hat{\sigma} = 31.549896$ (0.143300) $\hat{v} = 2.404800$ (0.000000)
3/2	2nd order approx	81391.07	$\hat{p} = 4.851194$ (0.374552) $\hat{q} = 400.090431$ (4.202087) $\hat{\sigma} = 31.552349$ (0.143901) $\hat{v} = 2.392486$ (0.022038)
Bessel	Newton-Raphson	127474.83	$\hat{\alpha} = 0.003649$ (0.000059) $\hat{\beta} = -1.976072$ (0.175107) $\hat{\gamma} = 0.085175$ (0.000454) $\hat{v} = 2.006025$ (0.012245)
Bessel	1st order approx	127474.83	$\hat{\alpha} = 0.003649$ (0.000059) $\hat{\beta} = -1.976072$ (0.175107) $\hat{\gamma} = 0.085175$ (0.000454) $\hat{v} = 2.006025$ (0.012245)
Bessel	2nd order approx	Undefined (Condition A is not met)	$\hat{\alpha}$ undefined $\hat{\beta}$ undefined $\hat{\gamma}$ undefined $\hat{v}$ undefined

We can draw several conclusions from Table 2, where the data series is daily and roughly 65 1/2 years long. Firstly, simulations performed in Section 4.1.1 for the CIR model, despite being based on different parameters, indicate that the respective minimum window lengths for reasonable estimates of $\bar{r}$, κ and σ are around 30, 90 and 14 years for daily observations. So based on the daily frequency and 65 1/2 year series length, we would expect the parameter estimates for $\bar{r}$ and σ of the CIR model in Table 2 to be reliable, while the estimate of κ to be unreliable. Secondly, analogous simulations performed in Section 4.1.2 for the 3/2 model indicate that the respective minimum window lengths for estimates of p , q and σ to be around 20, 90 and 14 years for daily observations. So, the parameter estimates of p and σ are reliable while the estimate of q

Table 3: Average run time of CIR estimation method per simulated data set using CIR parameter estimates and 1000 simulations.

Method	Average run time	Average parameter estimates
Newton-Raphson	1.162925	$\hat{r} = 0.039402$ (0.015420) $\hat{\kappa} = 0.121061$ (0.056607) $\hat{\sigma} = 0.072203$ (0.008800) $\hat{\nu} = 4.648468$ (2.035282)
1st order approx	0.000157	$\hat{r} = 0.042323$ (0.011295) $\hat{\kappa} = 0.127420$ (0.048523) $\hat{\sigma} = 0.064637$ (0.001141) $\hat{\nu} = 4.894986$ (1.662848)
2nd order approx	0.000244	$\hat{r} = 0.042316$ (0.011302) $\hat{\kappa} = 0.126511$ (0.048135) $\hat{\sigma} = 0.064636$ (0.001141) $\hat{\nu} = 4.861427$ (1.657352)

is not. Thirdly, analogous simulations performed in Section 4.1.3 for the Bessel model indicate that the respective minimum window lengths for estimates of α , β and γ are around 133, 101 and 8 years. So in Table 2, only the estimate of γ is reliable, while estimates of α and β are not. As mentioned previously, the minimum series length is around 130 to 140 years for estimating parameters for any of our models, for all choices of $\Delta \leq 1/4$. With this caveat in mind, based on either the log-likelihood, AIC or BIC, the Bessel model is the best of the three models.

4.3 Comparison of run times

To give an idea of the substantial improvement in run times of the first order and second order approximation methods over the standard Newton-Raphson estimation method, we simulated in MATLAB 1000 data sets of CIR short rates, based on the CIR parameter values $\bar{r} = 0.041954$, $\kappa = 0.093950$ and $\sigma = 0.064619$ (i.e. $\nu = 3.775778$), and then computed the average estimation run time per simulation, along with the averages of the parameter estimates. The results are shown in Table 3, where it is evident that the approximation methods take roughly two or three ten-thousandths of a second, far less than the 1 s or so taken on average by the Newton-Raphson method. This represents a speed-up of roughly 5,000 times with almost the same level of accuracy.

5 Conclusion

The closed-form formulae presented in this paper permit accurate parameter estimates of the CIR and 3/2 short rate model to be performed in spreadsheets and other high-level software packages, such as MATLAB and R, and permit exceedingly fast computations in computer algorithms. This significantly improved run time for parameter estimation has many applications where ex-ante forecasts are required frequently and immediately, such as in hedging interest rate derivatives and volatility based on such models.

Appendix A Proofs of theorems 1 and 2

In this technical appendix we supply proofs of Theorems 1 and 2. Our proof of Theorem 1 is as follows.

Proof: For convenience, we make the change of variables given in (23) and we rewrite the log-likelihood function in (11) in terms of these new variables as

$$\ell(a, k, v) = n \left\{ -\log(2) - \log a + k + \frac{1}{2}vL + vk - \frac{1}{2a}(R_0 e^{-k} + R_1 e^k) + \frac{1}{n} \sum_{i=1}^n \log I_v \left(\frac{\sqrt{r_{t_{i-1}} r_{t_i}}}{a} \right) \right\}, \quad (32)$$

where L , R_0 and R_1 are as in (14)–(16).

Our proof of the theorem involves solving for the new variables using an approximation to the first order derivatives of the log-likelihood function in (32).

The modified Bessel function of the first kind given in (7) can also be written as

$$I_v(z) = \frac{1}{\pi} \int_0^\pi e^{z \cos \theta} \cos(v\theta) d\theta - \frac{\sin(v\pi)}{\pi} \int_0^\infty e^{-z \cosh t - vt} dt, \quad (33)$$

from which we can derive the asymptotic expansion, given in §7.23 of Watson (1966),

$$I_v(z) \sim \frac{e^z}{\sqrt{2\pi z}} \sum_{j=0}^{\infty} \frac{(-1)^j (v, j)}{(2z)^j} + \frac{e^{-z + (v+\frac{1}{2})\pi i}}{\sqrt{2\pi z}} \sum_{j=0}^{\infty} \frac{(v, m)}{(2z)^m} \quad (34)$$

as $z \rightarrow \infty$, where

$$(v, j) = \frac{1}{j!} \prod_{m=1}^j \left(v^2 - \frac{1}{4}(2m-1)^2 \right) \quad (35)$$

and $\arg z \in (-\pi/2, 3\pi/2)$. Employing the first few terms of the expansion gives

$$I_v(z) = \frac{e^z}{\sqrt{2\pi z}} \left\{ 1 - \frac{v^2 - \frac{1}{4}}{2z} + \frac{(v^2 - \frac{1}{4})(v^2 - \frac{9}{4})}{2! (2z)^2} + O_v \left(\frac{1}{z^3} \right) \right\}, \quad (36)$$

where the subscript v in the big-O notation indicates dependence of the implied constant upon v . Taking logarithms of both sides gives

$$\log I_v(z) = z - \frac{1}{2} \log(2\pi z) - \frac{v^2 - \frac{1}{4}}{2z} - \frac{v^2 - \frac{1}{4}}{4z^2} + O_v \left(\frac{1}{z^3} \right). \quad (37)$$

In terms of our new variables given in (23) we rewrite the log-likelihood function in (32) as

$$\ell(a, k, v) = \ell_1(a, k, v) + n O_v(a^3 R_6), \quad (38)$$

where

$$\begin{aligned} \ell_1(a, k, v) = n \left\{ \log(2\sqrt{2\pi}) - \frac{1}{2} \log a + (v+1)k + \frac{1}{2}vL \right. \\ \left. - \frac{1}{2a}(R_0 e^{-k} + R_1 e^k - 2R_2) - \frac{1}{4}R_4 \right. \\ \left. - \frac{1}{2}aR_3 \left(v^2 - \frac{1}{4} \right) - \frac{1}{4}a^2 R_5 \left(v^2 - \frac{1}{4} \right) \right\}, \end{aligned} \quad (39)$$

and where L , R_0 , R_1 , R_2 , R_3 , R_4 , R_5 and R_6 are as in (14)–(21). According to (23), $a \approx \frac{1}{4}\sigma^2\Delta$, for small Δ , so that in (38), $\frac{1}{n}\ell(a, k, v) - \frac{1}{n}\ell_1(a, k, v) = O_{v, \sigma, R_6}(\Delta^3)$ as $\Delta \rightarrow 0$, where the subscripted variables in the big-O notation indicate dependence of the implied constant on these variables. Therefore, knowledge of the values of L , R_0 , R_1 , R_2 , R_3 and R_5 is sufficient to determine the $(\bar{r}, \kappa, \sigma)$ -dependent part of ℓ_1 and therefore the statistic $T = (L, R_0, R_1, R_2, R_3, R_5)$ is approximately sufficient of the second order.

Our plan is firstly to show that the solution $x_1 = (a_1, k_1, v_1)^\top$ to

$$f_1 \begin{pmatrix} a \\ k \\ v \end{pmatrix} \equiv \begin{pmatrix} \frac{\partial \ell_1}{\partial a} \\ \frac{\partial \ell_1}{\partial k} \\ \frac{\partial \ell_1}{\partial v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (40)$$

is close to the solution $x = (a, k, v)^\top$ to

$$f \begin{pmatrix} a \\ k \\ v \end{pmatrix} \equiv \begin{pmatrix} \frac{\partial \ell}{\partial a} \\ \frac{\partial \ell}{\partial k} \\ \frac{\partial \ell}{\partial v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (41)$$

and secondly to solve (40).

Write $h = x - x_1$ and by Taylor's Theorem

$$f(x_1 + h) = f(x_1) + Df(x_1 + \xi_1 h)h, \quad (42)$$

for some $\xi_1 \in (0, 1)$, where

$$Df(x) = \begin{pmatrix} \frac{\partial^2 \ell}{\partial a^2} & \frac{\partial^2 \ell}{\partial a \partial k} & \frac{\partial^2 \ell}{\partial a \partial v} \\ \frac{\partial^2 \ell}{\partial a \partial k} & \frac{\partial^2 \ell}{\partial k^2} & \frac{\partial^2 \ell}{\partial k \partial v} \\ \frac{\partial^2 \ell}{\partial a \partial v} & \frac{\partial^2 \ell}{\partial k \partial v} & \frac{\partial^2 \ell}{\partial v^2} \end{pmatrix}. \quad (43)$$

Therefore,

$$h = \{Df(x_1 + \xi_1 h)\}^{-1} f(x_1). \quad (44)$$

We know that

$$f(x_1) - f_1(x_1) = \begin{pmatrix} n O_{v_1}(a_1^2 R_6) \\ 0 \\ n O_{v_1}(a_1^3 R_6) \end{pmatrix} \quad (45)$$

and that

$$Df(x_1 + \xi_1 h) - Df_1(x_1 + \xi_1 h) = \begin{pmatrix} n O_{v_1}(a_1 R_6) & 0 & n O_{v_1}(a_1^2 R_6) \\ 0 & 0 & 0 \\ n O_{v_1}(a_1^2 R_6) & 0 & n O_{v_1}(a_1^3 R_6) \end{pmatrix}. \quad (46)$$

It is apparent from (44) and (45) that h is close to zero, that is, x_1 is close to x . Specifically,

$$h = O_{v_1} \left[\{Df_1(x_1 + \xi_1 h)\}^{-1} \begin{pmatrix} n a_1^2 R_6 \\ 0 \\ n a_1^3 R_6 \end{pmatrix} \right]. \quad (47)$$

To solve (40), we rewrite $\ell_1(a, k, v)$ as a quadratic in v , with the consequent inequality

$$\begin{aligned}
\ell_1(a, k, v) &= n \left\{ -\log(2\sqrt{2\pi}) - \frac{1}{2} \log a + k - \frac{1}{2a} (R_0 e^{-k} + R_1 e^k - 2R_2) - \frac{1}{4} R_4 \right. \\
&\quad \left. + \frac{1}{8} a R_3 + \frac{1}{16} a^2 R_5 + \left(k + \frac{1}{2} L\right) v - \left(\frac{1}{2} a R_3 + \frac{1}{4} a^2 R_5\right) v^2 \right\} \\
&= n \left\{ -\log(2\sqrt{2\pi}) - \frac{1}{2} \log a + k - \frac{1}{2a} (R_0 e^{-k} + R_1 e^k - 2R_2) - \frac{1}{4} R_4 \right. \\
&\quad \left. + \frac{1}{8} a R_3 + \frac{1}{16} a^2 R_5 + \frac{\left(k + \frac{1}{2} L\right)^2}{2a R_3 + a^2 R_5} \right. \\
&\quad \left. - \left(\frac{1}{2} a R_3 + \frac{1}{4} a^2 R_5\right) \left(v - \frac{k + \frac{1}{2} L}{a R_3 + \frac{1}{2} a^2 R_5}\right)^2 \right\} \\
&\leq \ell_2(a, k),
\end{aligned} \tag{48}$$

where

$$\begin{aligned}
\ell_2(a, k) &= n \left\{ -\log(2\sqrt{2\pi}) - \frac{1}{2} \log a + k - \frac{1}{2a} (R_0 e^{-k} + R_1 e^k - 2R_2) - \frac{1}{4} R_4 \right. \\
&\quad \left. + \frac{1}{8} a R_3 + \frac{1}{16} a^2 R_5 + \frac{\left(k + \frac{1}{2} L\right)^2}{2a R_3 + a^2 R_5} \right\}
\end{aligned} \tag{49}$$

and where, for ease of legibility in our algebra, the functions f , g and h are defined as

$$f_k = -R_0 e^{-k} + R_1 e^k, g_k = R_0 e^{-k} + R_1 e^k - 2R_2, h_k = k + \frac{1}{2} L. \tag{50}$$

Equality in (48) is achieved when

$$v = \frac{k + \frac{1}{2} L}{a R_3 + \frac{1}{2} a^2 R_5}. \tag{51}$$

We now focus on determining the optimum parameter values of the new log-likelihood function $\ell_2(a, k)$.

The first order partial derivatives of ℓ_2 with respect to our new set of variables a and k are readily deduced to be

$$\begin{aligned}
\frac{\partial \ell_2}{\partial a} &= -\frac{n}{2a} + \frac{n}{2a^2} g_k + \frac{n}{8} R_3 + \frac{n}{8} a R_5 - n \frac{h_k^2}{(2a R_3 + a^2 R_5)^2} (2R_3 + 2a R_5), \\
\frac{\partial \ell_2}{\partial k} &= n - \frac{n}{2a} f_k + 2n \frac{h_k}{2a R_3 + a^2 R_5}.
\end{aligned} \tag{52}$$

The maximum likelihood estimates are found by equating these first order partial derivatives to zero and solving for a and k .

Thus, from (52) we have the two equations

$$(R_3 + a R_5) h_k^2 = \left(\frac{1}{4} a^3 R_5 + \frac{1}{4} a^2 R_3 - a + g_k\right) \left(R_3 + \frac{1}{2} a R_5\right)^2, \tag{53}$$

$$h_k = \left(\frac{1}{2} f_k - a\right) \left(R_3 + \frac{1}{2} a R_5\right). \tag{54}$$

Rearranging (54) gives the quadratic equation in a

$$\frac{1}{2} R_5 a^2 + \left(R_3 - \frac{1}{4} R_5 f_k\right) a + h_k - \frac{1}{2} R_3 f_k = 0. \tag{55}$$

In the light of (54), we can reduce the degree of the polynomial in a in (53) as follows

$$\begin{aligned} & \frac{1}{4}a^3R_5 + \frac{1}{4}a^2R_3 - a + g_k \\ &= (R_3 + aR_5)h_k^2\left(R_3 + \frac{1}{2}aR_5\right)^{-2} \\ &= (R_3 + aR_5)\left(\frac{1}{2}f_k - a\right)^2 \\ &= \frac{1}{4}R_3f_k^2 + \left(\frac{1}{4}R_5f_k^2 - R_3f_k\right)a + (-R_5f_k + R_3)a^2 + R_5a^3. \end{aligned} \quad (56)$$

This is a cubic equation in a that can be rearranged as

$$\frac{3}{4}R_5a^3 + \left(-R_5f_k + \frac{3}{4}R_3\right)a^2 + \left(\frac{1}{4}R_5f_k^2 - R_3f_k + 1\right)a + \frac{1}{4}R_3f_k^2 - g_k = 0. \quad (57)$$

Euclid's greatest common divisor algorithm applied to the polynomial equations (55) and (57) gives a monic polynomial in a . The first step of this algorithm is (57) minus $\frac{3}{2}a$ of (55), giving the quadratic polynomial equation in a

$$\begin{aligned} & \left(-R_5f_k + \frac{3}{4}R_3\right)a^2 + \frac{3}{2}\left(\frac{1}{4}R_5f_k - R_3\right)a^2 \\ &+ \left(\frac{1}{4}R_5f_k^2 - R_3f_k + 1\right)a + \frac{3}{4}R_3f_ka - \frac{3}{2}h_ka + \frac{1}{4}R_3f_k^2 - g_k \\ &= \left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)a^2 + \left(\frac{1}{4}R_5f_k^2 - \frac{1}{4}R_3f_k + 1 - \frac{3}{2}h_k\right)a + \frac{1}{4}R_3f_k^2 - g_k \\ &= 0. \end{aligned} \quad (58)$$

The second step of this algorithm is (58) minus $\frac{2}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)$ of (55), giving the monic polynomial equation in a

$$\begin{aligned} & \left(\frac{1}{4}R_5f_k^2 - \frac{1}{4}R_3f_k + 1 - \frac{3}{2}h_k\right)a + \frac{1}{4}R_3f_k^2 - g_k + \frac{2}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)\left(\frac{1}{4}R_5f_k - R_3\right)a \\ &+ \frac{1}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)R_3f_k - \frac{2}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)h_k \\ &= \left[\left(\frac{1}{4}R_5f_k^2 - \frac{1}{4}R_3f_k + 1 - \frac{3}{2}h_k\right) + \frac{2}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)\left(\frac{1}{4}R_5f_k - R_3\right)\right]a \\ &+ \left[\frac{1}{4}R_3f_k^2 - g_k + \frac{1}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)R_3f_k - \frac{2}{R_5}\left(-\frac{5}{8}R_5f_k - \frac{3}{4}R_3\right)h_k\right] \\ &= 0. \end{aligned} \quad (59)$$

Letting

$$\begin{aligned} b_k &= \frac{3}{8}R_3f_k^2 + \left(\frac{3}{4}\frac{R_3^2}{R_5} - \frac{5}{4}h_k\right)f_k - \frac{3}{2}\frac{R_3}{R_5}h_k + g_k, \\ c_k &= -\frac{1}{16}R_5f_k^2 + \frac{5}{8}R_3f_k + 1 - \frac{3}{2}h_k + \frac{3}{2}\frac{R_3^2}{R_5}, \end{aligned} \quad (60)$$

we have from (59) the solution for a

$$a = \frac{b_k}{c_k}. \quad (61)$$

Note that replacing a in (55) by the RHS expression in (61) gives

$$\frac{1}{2}R_5b_k^2 + \left(-\frac{1}{4}R_5f_k + R_3\right)b_kc_k + \left(-\frac{1}{2}R_3f_k + h_k\right)c_k^2 = 0, \quad (62)$$

which only involves the variable k . Defining the function q_k as

$$q_k = \frac{1}{2}R_5b_k^2 + \left(-\frac{1}{4}R_5f_k + R_3\right)b_kc_k + \left(-\frac{1}{2}R_3f_k + h_k\right)c_k^2, \quad (63)$$

and making the assumption that k is close to zero, an approximate solution can be found by solving

$$q_0 + q'_0k + \frac{1}{2}q''_0k^2 = 0, \quad (64)$$

where

$$\begin{aligned} q_0 &= \frac{1}{2}R_5b_0^2 - \left(\frac{1}{4}R_5f_0 - R_3\right)b_0c_0 - \left(\frac{1}{2}R_3f_0 - h_0\right)c_0^2, \\ q'_0 &= R_5b_0b'_0 - \left\{\frac{1}{4}R_5f'_0b_0c_0 + \left(\frac{1}{4}R_5f_0 - R_3\right)(b'_0c_0 + b_0c'_0)\right\} - \left\{\left(\frac{1}{2}R_3f'_0 - h'_0\right)c_0^2 + 2\left(\frac{1}{2}R_3f_0 - h_0\right)c_0c'_0\right\}, \\ q''_0 &= R_5(b_0b''_0 + (b'_0)^2 - \frac{1}{4}R_5(f'_0b_0c_0)(f''_0/f'_0 + b'_0/b_0 + c'_0/c_0) - \left(\frac{1}{4}R_5 - R_3\right)(b''_0c_0 + 2b'_0c'_0 + b_0c''_0) \\ &\quad - \left(-\frac{1}{2}R_3f''_0 - h''_0\right)c_0^2 - 4\left(-\frac{1}{2}R_3f'_0 - h'_0\right)c_0c'_0 - 2\left(-\frac{1}{2}R_3f_0 - h_0\right)(c_0c''_0 + (c'_0)^2)). \end{aligned} \quad (65)$$

Therefore, from (64) we have

$$k = \frac{-q'_0 \pm \sqrt{(q'_0)^2 - 2q_0q''_0}}{q''_0}, \quad (66)$$

where the choice of sign of the square root is the same as that of q'_0 , thus being the root closest to zero. Then we can determine a from (61) and, finally, determine v from (51).

The original parameters are estimated as

$$\begin{aligned} \kappa &= \frac{2k}{\Delta}, \\ \sigma &= \sqrt{\frac{4\kappa a}{e^k - e^{-k}}}, \\ \bar{r} &= \frac{(v+1)\sigma^2}{2\kappa}. \end{aligned} \quad (67)$$

The second order partial derivatives are

$$\begin{aligned} \frac{\partial^2 \ell}{\partial a^2} &= \frac{n}{2a^2} - \frac{n}{a^3} \{R_0 e^{-k} + R_1 e^k - 2R_2\}, \\ \frac{\partial^2 \ell}{\partial a \partial k} &= \frac{n}{2a^2} \{-R_0 e^{-k} + R_1 e^k\}, \\ \frac{\partial^2 \ell}{\partial a \partial v} &= -nR_3v, \\ \frac{\partial^2 \ell}{\partial k^2} &= -\frac{n}{2a} \{R_0 e^{-k} + R_1 e^k\}, \\ \frac{\partial^2 \ell}{\partial k \partial v} &= n, \\ \frac{\partial^2 \ell}{\partial v^2} &= -naR_3, \end{aligned} \quad (68)$$

and the Fisher information matrix with respect to the variables a , k and v is simply

$$\mathcal{J}(a, k, v) = -n \begin{pmatrix} \frac{1}{2a^2} - \frac{1}{a^3} \{R_0 e^{-k} + R_1 e^k - 2R_2\} & \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -R_3 v \\ \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -\frac{1}{2a} \{R_0 e^{-k} + R_1 e^k\} & 1 \\ -R_3 v & 1 & -a R_3 \end{pmatrix}. \quad (69)$$

From (23),

$$\begin{pmatrix} \frac{\partial a}{\partial \bar{r}} & \frac{\partial a}{\partial \kappa} & \frac{\partial a}{\partial \sigma} \\ \frac{\partial k}{\partial \bar{r}} & \frac{\partial k}{\partial \kappa} & \frac{\partial k}{\partial \sigma} \\ \frac{\partial v}{\partial \bar{r}} & \frac{\partial v}{\partial \kappa} & \frac{\partial v}{\partial \sigma} \end{pmatrix} \approx \begin{pmatrix} 0 & \sigma^2 \kappa \Delta^3 / 48 & \sigma \Delta / 2 \\ 0 & \Delta / 2 & 0 \\ 2\kappa / \sigma^2 & 2\bar{r} / \sigma^2 & -4\bar{r}\kappa / \sigma^3 \end{pmatrix} \equiv B, \quad (70)$$

for small values of $\kappa\Delta/2$, and we have

$$\mathcal{J}(\bar{r}, \kappa, \sigma) \approx B^T \mathcal{J}(a, k, v) B. \quad (71)$$

Finally, asymptotic normality in (26) of the maximum likelihood estimates follows from (12) and asymptotic theory of MLEs; see for example Theorem 3.10 in Lehmann and Casella (1998).

Our proof of Theorem 2 is as follows.

Proof: We continue in a similar vein as in the proof of Theorem 1 up to the equations formed by equating the first order partial derivatives in (52) to zero and solving for a and k , where instead the log-likelihood function in (32) is rewritten as

$$\ell(a, k, v) = \ell_1(a, k, v) + n O_v(a^2 R_5), \quad (72)$$

where

$$\begin{aligned} \ell_1(a, k, v) &= n \left\{ \log(2\sqrt{2\pi}) - \frac{1}{2} \log a + (v+1)k + \frac{1}{2} vL \right. \\ &\quad \left. - \frac{1}{2a} (R_0 e^{-k} + R_1 e^k - 2R_2) - \frac{1}{4} R_4 - \frac{1}{2} a R_3 \left(v^2 - \frac{1}{4} \right) \right\} \\ &\leq \ell_2(a, k) \\ &\equiv n \left\{ -\log(2\sqrt{2\pi}) - \frac{1}{2} \log a + k - \frac{1}{2a} g_k - \frac{1}{4} R_4 + \frac{1}{8} a R_3 + \frac{h_k^2}{2a R_3} \right\} \end{aligned} \quad (73)$$

with equality achieved when

$$v = \frac{h_k}{a R_3}, \quad (74)$$

and where the first order partial derivatives of ℓ_2 are

$$\begin{aligned} \frac{\partial \ell_2}{\partial a} &= -\frac{n}{2a} + \frac{n}{2a^2} g_k + \frac{n}{8} R_3 - n \frac{h_k^2}{(2a R_3)^2} (2R_3), \\ \frac{\partial \ell_2}{\partial k} &= n - \frac{n}{2a} f_k + 2n \frac{h_k}{2a R_3}. \end{aligned} \quad (75)$$

Thus, setting (75) to zero, we have the two equations

$$R_3 h_k^2 = \left(\frac{1}{4} a^2 R_3 - a + g_k \right) R_3^2, \quad (76)$$

$$h_k = \left(\frac{1}{2} f_k - a \right) R_3. \quad (77)$$

Rearranging (77) gives the equation for a

$$a = \frac{1}{2} f_k - \frac{h_k}{R_3}. \quad (78)$$

Eliminating a from (76) gives

$$p_k \equiv \frac{1}{4} \left(\frac{1}{2} R_3 f_k - h_k \right)^2 - \left(\frac{1}{2} R_3 f_k - h_k \right) + R_3 g_k - h_k^2 = 0, \quad (79)$$

which only involves the variable k and where we have defined the function p_k as indicated. Making the assumption that k is close to zero, an approximate solution can be found by solving

$$p_0 + p'_0 k + \frac{1}{2} p''_0 k^2 = 0, \quad (80)$$

where

$$\begin{aligned} p_0 &= \frac{1}{4} \left(\frac{1}{2} R_3 f_0 - h_0 \right)^2 - \left(\frac{1}{2} R_3 f_0 - h_0 \right) + R_3 g_0 - h_0^2 \\ p'_0 &= \frac{1}{2} \left(\frac{1}{2} R_3 f'_0 - h'_0 \right) \left(\frac{1}{2} R_3 f_0 - h_0 \right) - \left(\frac{1}{2} R_3 f'_0 - h'_0 \right) + R_3 g'_0 - 2h'_0 h_0 \\ p''_0 &= \frac{1}{2} \left(\frac{1}{2} R_3 f''_0 - h''_0 \right) \left(\frac{1}{2} R_3 f_0 - h_0 \right) + \frac{1}{2} \left(\frac{1}{2} R_3 f'_0 - h'_0 \right)^2 \\ &\quad - \left(\frac{1}{2} R_3 f''_0 - h''_0 \right) + R_3 g''_0 - 2(h'_0 h_0 + (h'_0)^2). \end{aligned} \quad (81)$$

We remark that the expressions for p_0 , p'_0 and p''_0 in (81) can be obtained by placing $R_5 = 0$ in the expressions for q_0 , q'_0 and q''_0 in (65).

Therefore, from (80) we have

$$k = \frac{-p'_0 \pm \sqrt{(p'_0)^2 - 2p_0 p''_0}}{p''_0}, \quad (82)$$

where the choice of sign of the square root is the same as that of p'_0 , thus being the root closest to zero. Then we can determine a from (78) and, finally, determine v from (74).

The original parameters are estimated as

$$\begin{aligned} \kappa &= \frac{2k}{\Delta}, \\ \sigma &= \sqrt{\frac{4\kappa a}{e^k - e^{-k}}}, \\ \bar{r} &= \frac{(v+1)\sigma^2}{2\kappa}. \end{aligned} \quad (83)$$

The second order partial derivatives are

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial a^2} &= \frac{n}{2a^2} - \frac{n}{a^3} \{R_0 e^{-k} + R_1 e^k - 2R_2\}, \\
\frac{\partial^2 \ell}{\partial a \partial k} &= \frac{n}{2a^2} \{-R_0 e^{-k} + R_1 e^k\}, \\
\frac{\partial^2 \ell}{\partial a \partial v} &= -nR_3v, \\
\frac{\partial^2 \ell}{\partial k^2} &= -\frac{n}{2a} \{R_0 e^{-k} + R_1 e^k\}, \\
\frac{\partial^2 \ell}{\partial k \partial v} &= n, \\
\frac{\partial^2 \ell}{\partial v^2} &= -naR_3
\end{aligned} \tag{84}$$

and the Fisher information matrix with respect to the variables a , k and v is simply

$$\mathcal{J} = -n \begin{pmatrix} \frac{1}{2a^2} - \frac{1}{a^3} \{R_0 e^{-k} + R_1 e^k - 2R_2\} & \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -R_3v \\ \frac{1}{2a^2} \{-R_0 e^{-k} + R_1 e^k\} & -\frac{1}{2a} \{R_0 e^{-k} + R_1 e^k\} & 1 \\ -R_3v & 1 & -aR_3 \end{pmatrix}. \tag{85}$$

The Fisher information matrix $\mathcal{J}$ with respect to the variables $\bar{r}$, κ and σ is derived as in the proof of Theorem 1.

B Calculations of parameter standard errors in Table 1

We supply here the full details of the calculations of the standard errors of the parameters for the CIR model, given in Table 1. The calculations of the standard errors of the parameters for the 3/2 and Bessel models are calculated analogously. Once the maximum likelihood estimates of the parameters of the CIR model have been obtained, namely

$$\hat{\bar{r}} = 0.042712, \hat{\kappa} = 0.094537, \hat{\sigma} = 0.064243,$$

we compute numerically the Fisher Information Matrix, approximated by the observed Fisher Information Matrix,

$$\mathcal{J}(\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma}) = - \begin{pmatrix} \frac{\partial^2 \ell}{\partial \bar{r}^2} & \frac{\partial^2 \ell}{\partial \bar{r} \partial \kappa} & \frac{\partial^2 \ell}{\partial \bar{r} \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \kappa \partial \bar{r}} & \frac{\partial^2 \ell}{\partial \kappa^2} & \frac{\partial^2 \ell}{\partial \kappa \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \bar{r}} & \frac{\partial^2 \ell}{\partial \sigma \partial \kappa} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{pmatrix}$$

With $h = 0.00001$, the value of the log-likelihood function at the maximum likelihood estimate vector

$$\ell(\hat{\bar{r}}, \hat{\kappa}, \hat{\sigma}) = 7559.08057481, \tag{86}$$

and values of the log-likelihood function at points neighbouring the maximum likelihood estimate vector

$$\begin{aligned}
\ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma}) &= 7559.08057408832 \\
\ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma}) &= 7559.08057408832 \\
\ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma}) &= 7559.08057473559 \\
\ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma}) &= 7559.0805747356 \\
\ell(\hat{r}, \hat{\kappa}, \hat{\sigma} + h) &= 7559.08053039953 \\
\ell(\hat{r}, \hat{\kappa}, \hat{\sigma} - h) &= 7559.08053261873,
\end{aligned} \tag{87}$$

$$\begin{aligned}
\ell(\hat{r} + h, \hat{\kappa} + h, \hat{\sigma}) &= 7559.08057369888 \\
\ell(\hat{r} + h, \hat{\kappa} - h, \hat{\sigma}) &= 7559.08057432685 \\
\ell(\hat{r} - h, \hat{\kappa} + h, \hat{\sigma}) &= 7559.08057432661 \\
\ell(\hat{r} - h, \hat{\kappa} - h, \hat{\sigma}) &= 7559.08057369926,
\end{aligned} \tag{88}$$

$$\begin{aligned}
\ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma} + h) &= 7559.08053065437 \\
\ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma} - h) &= 7559.08053091805 \\
\ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma} + h) &= 7559.08052869978 \\
\ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma} - h) &= 7559.08053287355
\end{aligned} \tag{89}$$

and

$$\begin{aligned}
\ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma} + h) &= 7559.08053076554 \\
\ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma} - h) &= 7559.08053210167 \\
\ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma} + h) &= 7559.08052988274 \\
\ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma} - h) &= 7559.08053298488
\end{aligned} \tag{90}$$

we compute numerically the second order partial derivatives as

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \bar{r}^2} &= \frac{1}{h^2} \{ \ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma}) - 2\ell(\hat{r}, \hat{\kappa}, \hat{\sigma}) + \ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma}) \} = -14453.7989399396 \\
\frac{\partial^2 \ell}{\partial \bar{r} \partial \kappa} &= \frac{1}{4h^2} \{ \ell(\hat{r} + h, \hat{\kappa} + h, \hat{\sigma}) - \ell(\hat{r} + h, \hat{\kappa} - h, \hat{\sigma}) - \ell(\hat{r} - h, \hat{\kappa} + h, \hat{\sigma}) + \ell(\hat{r} - h, \hat{\kappa} - h, \hat{\sigma}) \} \\
&= -3138.30241793767 \\
\frac{\partial^2 \ell}{\partial \bar{r} \partial \sigma} &= \frac{1}{4h^2} \{ \ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma} + h) - \ell(\hat{r} + h, \hat{\kappa}, \hat{\sigma} - h) - \ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma} + h) + \ell(\hat{r} - h, \hat{\kappa}, \hat{\sigma} - h) \} \\
&= 9775.22404355113 \\
\frac{\partial^2 \ell}{\partial \kappa^2} &= \frac{1}{h^2} \{ \ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma}) - 2\ell(\hat{r}, \hat{\kappa}, \hat{\sigma}) + \ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma}) \} = -1508.28782352619 \\
\frac{\partial^2 \ell}{\partial \kappa \partial \sigma} &= \frac{1}{4h^2} \{ \ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma} + h) - \ell(\hat{r}, \hat{\kappa} + h, \hat{\sigma} - h) - \ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma} + h) + \ell(\hat{r}, \hat{\kappa} - h, \hat{\sigma} - h) \} \\
&= 4415.0260691822 \\
\frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{1}{h^2} \{ \ell(\hat{r}, \hat{\kappa}, \hat{\sigma} + h) - 2\ell(\hat{r}, \hat{\kappa}, \hat{\sigma}) + \ell(\hat{r}, \hat{\kappa}, \hat{\sigma} - h) \} = -866037.589730695
\end{aligned} \tag{91}$$

so that the Fisher Information Matrix becomes

$$\mathcal{J} \approx - \begin{pmatrix} -14453.7989399396 & -3138.30241793767 & 9775.22404355113 \\ -3138.30241793767 & -1508.28782352619 & 4415.0260691822 \\ 9775.22404355113 & 4415.0260691822 & -866037.589730695 \end{pmatrix}.$$

The covariance matrix $\text{COV}(\hat{r}, \hat{\kappa}, \hat{\sigma})$ of parameter estimates is the inverse of this matrix, namely

$$\mathcal{J}^{-1} \approx \begin{pmatrix} 0.000126206691887228 & -0.000262343932416607 & 8.7113295284852E-08 \\ -0.000262343932416607 & 0.00121837741220894 & 3.25008677976887E-06 \\ 8.71132952848433E-08 & 3.25008677976891E-06 & 1.17223638081813E-06 \end{pmatrix}$$

and the standard errors emerge as square roots of the diagonal elements of this matrix

$$SE(\hat{r}) = 0.011234, SE(\hat{\kappa}) = 0.034905, SE(\hat{\sigma}) = 0.001083,$$

which are very close to those given for the CIR model in Table 1.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This research is supported by an Australian Government Research Training Program Scholarship.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Ahn, D., and B. Gao. 1999. "A Parametric Nonlinear Model of Term Structure Dynamics." *Review of Financial Studies* 12: 721–62.
- Barczy, M., G. Pap, and T. T. Szabo. 2016. "Parameter Estimation for the Subcritical Heston Model Based on Discrete Time Observations." *Acta Scientiarum Mathematicarum (Szeged)* 82: 313–38.
- Bielecki, T., M. Jeanblanc, and M. Rutkowski. 2011. "Hedging of a Credit Default Swaption in the CIR Default Intensity Model." *Finance and Stochastics* 15: 541–72.
- Biffis, E. 2005. "Affine Processes for Dynamic Mortality and Actuarial Valuations." *Insurance: Mathematics and Economics* 37: 443–68.
- Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81: 637–54.
- Blackburn, F., and M. Sherris. 2013. "Consistent Dynamic Affine Mortality Models for Longevity Risk Applications." *Insurance: Mathematics and Economics* 53: 64–73.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53: 385–407.
- Dokuchaev, N. 2017. "A Pathwise Inference Method for the Parameters of Diffusion Terms." *Journal of Nonparametric Statistics* 29: 731–43.
- Elliott, R., and E. Platen. 2001. *Hidden Markov Chain Filtering for Generalised Bessel Processes*, Trends in Mathematics, 123–43. Boston, MA: Birkhäuser. chapter 7.
- Federal Reserve Bank of St. Louis. 2019. *6-Month London Interbank Offered Rate (LIBOR), Based on U.S. Dollar (USD6MTD156N)*, <https://fred.stlouisfed.org/series/USD6MTD156N>, [Online; accessed April 16 2019].
- Federal Reserve Bank of St. Louis. 2020. *Effective Federal Funds Rate [DFF]*. <https://fred.stlouisfed.org/series/DFF>, [Online; accessed February 3 2020].
- Feller, W. 1971. *An Introduction to Probability Theory and its Applications*, Vol 2, 2nd ed. New York: Wiley.
- Fergusson, K., and E. Platen. 2015. "Application of Maximum Likelihood Estimation to Stochastic Short Rate Models." *Annals of Financial Economics* 10: 1–26.
- Goard, J., and M. Mazur. 2013. "Stochastic Volatility Models and the Pricing of VIX Options." *Mathematical Finance* 23: 439–58.
- Heston, S. L. 1997. *A Simple New Formula for Options with Stochastic Volatility*, Technical report: Washington University of St. Louis.
- Huang, X. 2013. "Quasi-maximum Likelihood Estimation of Multivariate Diffusions." *Studies in Nonlinear Dynamics and Econometrics* 17: 179–97.
- Karlin, S., and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. New York: Academic Press.
- Le Cam, L. 1986. *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.
- Lehmann, E. L., and G. Casella. 1998. *Theory of Point Estimation*. New York: Springer.
- McCullagh, P. 1984. "Local Sufficiency." *Biometrika* 71: 233–44.
- Overbeck, L., and T. Rydén. 1997. "Estimation in the Cox-Ingersoll-Ross Model." *Econometric Theory* 13: 430–61.
- Platen, E. 1997. *A Non-linear Stochastic Volatility Model*, Working Paper: Australian National University, Financial Mathematics Research Reports. FMRR 005-97.
- Platen, E. 1999. "A Short-Term Interest Rate Model." *Finance and Stochastics* 3: 215–25.
- Poulsen, R. 1999. *Approximate Maximum Likelihood Estimation of Discretely Observed Diffusion Processes*, Working Paper No. 29: University of Aarhus.
- Shiller, R. 2019. *One-Year Interest Rate*, <http://www.econ.yale.edu/shiller/data.htm>, [Online; accessed April 16 2019].
- Watson, G. N. 1966. *A Treatise on the Theory of Bessel Functions*, 2nd ed.: Cambridge University Press.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/snde-2019-0079>).