J. Appl. Prob. **55,** 543–558 (2018)
*doi:*10.1017/jpr.2018.34

WEIGHTED LEAST-SQUARES ESTIMATION FOR THE SUBCRITICAL HESTON PROCESS

M. DU ROY DE CHAUMARAY,* *Institut de Mathématiques de Bordeaux*

Abstract

We simultaneously estimate the four parameters of a subcritical Heston process. We do not restrict ourselves to the case where the stochastic volatility process never reaches zero. In order to avoid the use of unmanageable stopping times and a natural but intractable estimator, we use a weighted least-squares estimator. We establish strong consistency and asymptotic normality for this estimator. Numerical simulations are also provided, illustrating the favorable performance of our estimation procedure.

Keywords: Squared radial Ornstein–Uhlenbeck process; Heston process; parameter estimation; weighted least-squares estimate; consistency; asymptotic normality

2010 Mathematics Subject Classification: Primary 62M05
Secondary 60G20; 60G44

1. Introduction

Introduced in 1973 as a hedging tool, the Black–Scholes model uses a geometric Brownian motion to represent asset prices. The implied volatility is supposed to be constant over time, which turned out to be inaccurate when applied to fit real market data, especially during the stock market crash of 1987; see [26]. Several alternative models have been constructed to take into account the so-called smile effect associated to deep in-the-money or out-of-money options. Particular attention has been devoted to the study of stochastic volatility processes in which the volatility is also given as a solution of some stochastic differential equation; see [15], [22], and [25] for financial accuracy. Among them, the Heston process [17] is one of the most popular due to its computational tractability. For example, Lee [21] easily computed call option prices using Fourier inversion techniques. For numerous results on the asymptotic volatility smile, we refer the reader to, for example, [11], [12], and [18].

We denote by Y_t the logarithm of the price of a given asset and by X_t its instantaneous variance, and we consider the following Heston process:

$$\begin{aligned} \mathrm{d}X_t &= (a + bX_t)\,\mathrm{d}t + 2\sqrt{X_t}\,\mathrm{d}B_t, \\ \mathrm{d}Y_t &= (\alpha + \beta X_t)\,\mathrm{d}t + 2\sqrt{X_t}(\rho\,\mathrm{d}B_t + \sqrt{1-\rho^2}\,\mathrm{d}W_t) \end{aligned} \tag{1.1}$$

with $a > 0$, $(b, \alpha, \beta) \in \mathbb{R}^3$, and $\rho \in (-1, 1)$, where (B_t, W_t) is a two-dimensional standard Wiener process and the initial state is $(x_0, y_0) \in \mathbb{R}^+ \times \mathbb{R}$. In this process, the volatility X_t is driven by a generalized squared radial Ornstein–Uhlenbeck process, also known as the Cox–Ingersoll–Ross (CIR) process, first studied by Feller [10] and introduced in a financial context by Cox *et al.* [8] to compute short-term interest rates. The asymptotic behavior of this process has been widely investigated and depends on the values of both coefficients a and b.

Received 31 August 2016; revision received 20 December 2017.
* Current address: ENSAI, Campus de Ker Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz Cedex, France.
Email address: marie.du-roy-de-chaumaray@ensai.fr

Once a model has been chosen for its realistic features, it needs to be calibrated before being used for pricing. Our goal in this paper is to estimate parameters (a, b, α, β) while at the same time using a trajectory of (X_t) and (Y_t) over the time interval $[0, T]$. Azencott and Gadhyan [4] developed an algorithm to estimate some parameters of the Heston process based on discrete-time observations by making use of Euler and Milstein discretization schemes for the maximum likelihood. However, in the special case of a Heston process, the exact likelihood can be computed. It allows us to construct the maximum likelihood estimator (MLE) without using sophisticated approximation methods, which is necessary for many stochastic volatility models; see [1]. The MLE of (a, b, α, β) was recently investigated by Barczy and Pap [5], together with its asymptotic behavior in the special case where $a \geq 2$. Denote by τ_0 the stopping time given by

$$\tau_0 = \inf\left\{T > 0 \;\middle|\; \int_0^T X_t^{-1}\,\mathrm{d}t = \infty\right\}. \tag{1.2}$$

For any $a > 0$, the MLE $\widetilde{\theta}_T = (\widetilde{a}_T, \widetilde{b}_T, \widetilde{\alpha}_T, \widetilde{\beta}_T)$ is given by, for $T < \tau_0$,

$$\widetilde{\theta}_T = \begin{pmatrix} G_T^{-1} & 0 \\ 0 & G_T^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{U}_T \\ \widetilde{V}_T \end{pmatrix},$$

where $\widetilde{U}_T = (\int_0^T X_t^{-1}\,\mathrm{d}X_t, \int_0^T \mathrm{d}X_t)^\top$, $\widetilde{V}_T = (\int_0^T X_t^{-1}\,\mathrm{d}Y_t, \int_0^T \mathrm{d}Y_t)^\top$, and

$$G_T = \begin{pmatrix} \int_0^T X_t^{-1}\,\mathrm{d}t & T \\ T & \int_0^T X_t\,\mathrm{d}t \end{pmatrix}.$$

Observe that $(\widetilde{a}_T, \widetilde{b}_T)$ coincides with the MLE of the parameters (a, b) of the CIR process based on the observation of (X_T) over the time interval $[0, T]$. The asymptotic behavior of this latter estimator is well known; see, for example, [7], [13], and [24]. In the supercritical case of $b > 0$, Overbeck [24] showed that $\widetilde{b}_T$ converges almost surely (a.s.) to b whereas there exists no consistent estimator for a. Hence, we will focus our attention on the geometrically ergodic case of $b < 0$. Furthermore, the value of a governs the behavior at 0 of (X_T): for $a \geq 2$, the process a.s. never reaches 0, whereas for $0 < a < 2$, 0 is quite frequently visited and

$$\mathbb{P}(\tau_0 < \infty) = 1; \tag{1.3}$$

see, for example, [20] or [24]. For $a > 2$, the MLE converges a.s. to $\theta = (a, b, \alpha, \beta)$ and satisfies the following central limit theorem (CLT):

$$\sqrt{T}\left(\widetilde{\theta}_T - \theta\right) \xrightarrow{\mathrm{D}} \mathcal{N}(0, 4R \otimes \Sigma^{-1}),$$

where the matrices R and Σ are given by

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \dfrac{-b}{a-2} & 1 \\ 1 & -\dfrac{a}{b} \end{pmatrix},$$

and '$\otimes$' denotes the Kronecker product.

A large deviation principle for the couple $(\widetilde{a}_T, \widetilde{b}_T)$ was established by Cox *et al.* [9]. In the particular case where one parameter is known and the other one is estimated, large deviations can be found in [28], while moderate deviations can be found in [14].

By contrast, in the $0 < a < 2$ case, (1.3) implies the nonintegrability of X_T^{-1} for large values of T so that the MLE does not converge as T goes to ∞. Consequently, this case has received less attention even though it is often of interest in finance in order to compute long-dated interest rates, for example, as explained in [3], or in foreign exchange markets; see [19]. In the case of the CIR process, Overbeck [24] used accurate stopping times to build a strongly consistent estimator based on the MLE:

$$\mathbf{1}_{\{T<\tau_0\}}\begin{pmatrix}\widetilde{a}_T\\ \widetilde{b}_T\end{pmatrix}+\mathbf{1}_{\{\tau_0\le T\}}\begin{pmatrix}\lim_{t\uparrow\tau_0} S_t\Sigma_t^{-1}\\ \left(\int_0^T X_s\,\mathrm{d}s\right)^{-1}\left(X_T-T\lim_{t\uparrow\tau_0}S_t\Sigma_t^{-1}\right)\end{pmatrix},$$

where $S_t=\int_0^t X_s^{-1}\,\mathrm{d}X_s$, $\Sigma_t=\int_0^t X_s^{-1}\,\mathrm{d}s$, τ_0 is given by (1.2), and $\mathbf{1}_A$ is the indicator function on the event A. The aim of this paper is to investigate a new strongly consistent weighted least-squares estimator (WLSE) for the quadruplet of parameters θ (and for (a,b) as a consequence). The weighting allows us to circumvent the explosion for X_T reaching 0 and consequently avoids us having to make use of stopping times, which are not easy to handle in practice. It generalizes to continuous time the original work of Wei and Winnicki [27] for branching processes with immigration, inspired by an analogy with first-order autoregressive processes. Our results answer, by the way, the question of Ben Alaya and Kebaier in the conclusion of [7] regarding the CIR process.

Following the seminal work of Wei and Winnicki [27], denote $C_T=X_T+c$, where c is some positive constant. Our new couple of WLSE is given by

$$\widehat{\theta}_T=\begin{pmatrix}\Gamma_T^{-1} & 0\\ 0 & \Gamma_T^{-1}\end{pmatrix}\begin{pmatrix}U_T\\ V_T\end{pmatrix}, \tag{1.4}$$

where

$$U_T=\left(\int_0^T\frac{1}{C_t}\,\mathrm{d}X_t,\int_0^T\frac{X_t}{C_t}\,\mathrm{d}X_t\right)^{\top},\qquad V_T=\left(\int_0^T\frac{1}{C_t}\,\mathrm{d}Y_t,\int_0^T\frac{X_t}{C_t}\,\mathrm{d}Y_t\right)^{\top},$$

and

$$\Gamma_T=\begin{pmatrix}\int_0^T\frac{1}{C_t}\,\mathrm{d}t & \int_0^T\frac{X_t}{C_t}\,\mathrm{d}t\\ \int_0^T\frac{X_t}{C_t}\,\mathrm{d}t & \int_0^T\frac{X_t^2}{C_t}\,\mathrm{d}t\end{pmatrix}.$$

Further details on the choice of this estimator can be found in Appendix A. We do not restrict ourselves to the case where $c=1$ as it may sometimes lower the variance of the estimators. In the particular case where $c=0$, we observe that the new estimator coincides with the MLE.

The paper is organized as follows. Section 2 contains our main results: the strong consistency of this new couple of estimators as well as its asymptotic normality. In Section 3 we deal with a comparison with the MLE, while the remainder of the paper is devoted to the proofs of our main results, as well as their illustration by some numerical simulations.

2. Main results

Our main results are as follows.

Theorem 2.1. *Assume that $a>0$ and $b<0$. Then the four-dimensional WLSE $\widehat{\theta}_T$ is strongly consistent: as T goes to ∞,*

$$\widehat{\theta}_T\to\theta\quad a.s. \tag{2.1}$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

As T goes to ∞, X_T converges in distribution to a random variable X with gamma distribution $\Gamma(\frac{1}{2}a, -\frac{1}{2}b)$; see, for example, [24, Lemma 3]. Additionally, we denote by C the limiting distribution of $X_T + c$ as T goes to ∞.

Theorem 2.2. *Assume that $a > 0$ and $b < 0$. Then, as T goes to ∞, the estimator $\widehat{\theta}_T$ satisfies the following CLT:*

$$\sqrt{T}(\widehat{\theta}_T - \theta) \xrightarrow{\text{D}} \mathcal{N}(0, 4\Lambda), \tag{2.2}$$

where the asymptotic variance Λ is defined as a block matrix by

$$\Lambda = \begin{pmatrix} ALA & \rho ALA \\ \rho ALA & ALA \end{pmatrix},$$

with the matrices A and L given by

$$A = \bigg(\mathbb{E}[C]\mathbb{E}\bigg[\frac{1}{C}\bigg] - 1\bigg)^{-1} \begin{pmatrix} \mathbb{E}\Big[\dfrac{X^2}{C}\Big] & -\mathbb{E}\Big[\dfrac{X}{C}\Big] \\ -\mathbb{E}\Big[\dfrac{X}{C}\Big] & \mathbb{E}\Big[\dfrac{1}{C}\Big] \end{pmatrix}, \qquad L = \begin{pmatrix} \mathbb{E}\Big[\dfrac{X}{C^2}\Big] & \mathbb{E}\Big[\dfrac{X^2}{C^2}\Big] \\ \mathbb{E}\Big[\dfrac{X^2}{C^2}\Big] & \mathbb{E}\Big[\dfrac{X^3}{C^2}\Big] \end{pmatrix}.$$

We deduce from the previous theorems the following result for the MLE of the two parameters of the CIR process (X_T).

Corollary 2.1. *Assuming that $a > 0$ and $b < 0$, the WLSE $(\widehat{a}_T, \widehat{b}_T)$ of parameters (a, b) is strongly consistent as T goes to ∞; thus,*

$$\begin{pmatrix} \widehat{a}_T \\ \widehat{b}_T \end{pmatrix} \to \begin{pmatrix} a \\ b \end{pmatrix} \quad a.s.$$

and satisfies the following CLT:

$$\sqrt{T}\begin{pmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \end{pmatrix} \xrightarrow{\text{D}} \mathcal{N}(0, 4ALA).$$

Remark 2.1. In the remainder of this paper, we denote

$$\psi_c = \big(-\tfrac{1}{2}bc\big)^{a/2} \mathrm{e}^{-bc/2} \Gamma\big(1 - \tfrac{1}{2}a, -\tfrac{1}{2}bc\big), \tag{2.3}$$

where Γ is the upper incomplete gamma function defined for all $y \in \mathbb{R}$ and $\alpha \in \mathbb{R}_*^+$ by

$$\Gamma(\alpha, y) = \int_y^{+\infty} \mathrm{e}^{-t} t^{\alpha - 1} \, \mathrm{d}t,$$

and extended for $y \neq 0$ to any real α by holomorphy. To simplify the following expressions, we also define

$$\varphi_c = \psi_c\bigg(1 - \frac{a}{bc}\bigg) - 1.$$

In the proof of Theorem 2.2, we evaluate the two matrices A and L involving c to obtain

$$A = \varphi_c^{-1} \begin{pmatrix} c(\psi_c - 1) - \dfrac{a}{b} & \psi_c - 1 \\ \psi_c - 1 & \dfrac{\psi_c}{c} \end{pmatrix}$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

and

$$L = \frac{1}{2}\begin{pmatrix} \dfrac{a}{c}\psi_c + b(1-\psi_c) & (a+2-bc)(1-\psi_c) - a \\ (a+2-bc)(1-\psi_c) - a & c(\psi_c - 1)(a+4-bc) + ac - \dfrac{2a}{b} \end{pmatrix}.$$

By a straightforward computation, we deduce that

$$ALA = (\varphi_c)^{-2}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

where the variances σ_{11} and σ_{22} are given by

$$\sigma_{11} = \frac{a}{b}(\psi_c - 1)^2 - \frac{a^2}{2b}\varphi_c, \qquad \sigma_{22} = \varphi_c\left(\frac{\psi_c}{c} - \frac{b}{2}\right) + \frac{\psi_c}{c}(\psi_c - 1),$$

and the covariance σ_{12} is given by $\sigma_{12} = (\psi_c - 1)^2 - (\frac{1}{2}a)\varphi_c$.

Remark 2.2. As c goes to 0 (for which we need a to be greater than 2), we obtain the same covariance matrix as for the MLE. Indeed, using well-known asymptotic results about the incomplete gamma function Γ, which can be found in [23], it follows that as soon as $a > 2$,

$$\Gamma\left(1 - \frac{a}{2}, -\frac{bc}{2}\right)\left(-\frac{bc}{2}\right)^{a/2-1} \to \frac{-1}{1-a/2} = \frac{2}{a-2}, \qquad c \to 0.$$

Thus, ψ_c goes to 0 for c tending to 0, ψ_c/c converges to $-b/(a-2)$, and φ_c tends to $2/(a-2)$. Hence, we easily obtain, as c goes to 0,

$$\sigma_{11} \to -\frac{2a}{b(a-2)}, \qquad \sigma_{22} \to -\frac{2b}{(a-2)^2}, \qquad \sigma_{12} \to -\frac{2}{a-2}, \qquad c \to 0,$$

where σ_{11}, σ_{22}, and σ_{12} are defined in Remark 2.1. This leads to

$$ALA \to \Sigma^{-1}, \qquad c \to 0, \quad \text{where } \Sigma = \begin{pmatrix} \dfrac{-b}{a-2} & 1 \\ 1 & -\dfrac{a}{b} \end{pmatrix}.$$

3. Asymptotic variance

Even though we considered the WLSEs in order to investigate the $0 < a < 2$ case for which the MLE is not consistent, it is interesting to compare the asymptotic variances in the CLT of this new estimator, and of the MLE in the case where $a > 2$. This comparison is technically involved as the asymptotic variances depend on the values of a, b, and c. However, it is quite easy to compare variances for the MLE of the parameters of the CIR process in the case where we suppose that one of the parameters is known and we estimate the other one as it simplifies substantially the expression of the estimators. On the one hand, if a is known, the MLE for b is given by

$$\check{b}_T = \frac{X_T - x_0 - aT}{\int_0^T X_t \, \mathrm{d}t}$$

and satisfies the following CLT:

$$\sqrt{T}(\check{b}_T - b) \xrightarrow{\mathrm{D}} \mathcal{N}\left(0, \frac{4}{\mathbb{E}[X]}\right),$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

where $\mathbb{E}[X] = -a/b$; see, for example, [6]. On the other hand, if b is known, the MLE of a is given by

$$\check{a}_T = \frac{\int_0^T 1/X_t \, \mathrm{d}X_t - bT}{\int_0^T 1/X_t \, \mathrm{d}t}$$

and satisfies the following CLT:

$$\sqrt{T}(\check{a}_T - a) \xrightarrow{\mathrm{D}} \mathcal{N}\bigg(0, \frac{4}{\mathbb{E}[X^{-1}]}\bigg)$$

with $\mathbb{E}[X^{-1}] = -b/(a-2)$. Whereas the WLSEs are given by

$$\widehat{b}_T = \frac{\int_0^T (X_t/C_t) \, \mathrm{d}X_t - a \int_0^T (X_t/C_t) \, \mathrm{d}t}{\int_0^T (X_t^2/C_t) \, \mathrm{d}t}, \qquad \widehat{a}_T = \frac{\int_0^T (1/C_t) \, \mathrm{d}X_t - b \int_0^T (X_t/C_t) \, \mathrm{d}t}{\int_0^T (1/C_t) \, \mathrm{d}t}.$$

Proposition 3.1. *Assume that $a > 0$ and $b < 0$. As T goes to ∞, $\widehat{b}_T$ satisfies the following CLT:*

$$\sqrt{T}(\widehat{b}_T - b) \xrightarrow{\mathrm{D}} \mathcal{N}\bigg(0, 4\mathbb{E}\bigg[\frac{X^3}{C^2}\bigg]\bigg(\mathbb{E}\bigg[\frac{X^2}{C}\bigg]\bigg)^{-2}\bigg). \tag{3.1}$$

Proof. Replacing $\mathrm{d}X_t$ by its expression (1.1), we easily obtain

$$\sqrt{T}(\widehat{b}_T - b) = 2\bigg(\frac{1}{T}\int_0^T \frac{X_t^2}{C_t} \, \mathrm{d}t\bigg)^{-1} \frac{n_T}{\sqrt{T}},$$

where n_T is a martingale given by

$$n_T = \int_0^T \frac{X_t\sqrt{X_t}}{C_t} \, \mathrm{d}B_t, \qquad \langle n \rangle_T = \int_0^T \frac{X_t^3}{C_t^2} \, \mathrm{d}t.$$

Using the ergodicity of the process, we obtain, as T goes to ∞,

$$\frac{\langle n \rangle_T}{T} \to \mathbb{E}\bigg[\frac{X^3}{C^2}\bigg] \quad \text{a.s.} \tag{3.2}$$

Thus, by the CLT for martingales, we obtain the following convergence in distribution:

$$\frac{n_T}{\sqrt{T}} \xrightarrow{\mathrm{D}} \mathcal{N}\bigg(0, \mathbb{E}\bigg[\frac{X^3}{C^2}\bigg]\bigg). \tag{3.3}$$

Consequently, (3.1) follows from (3.3), Slutsky's lemma, and the fact that, by the ergodicity of the process, $(1/T)\int_0^T X_t^2/C_t \, \mathrm{d}t$ converges a.s. to $\mathbb{E}[X^2/C]$ as T goes to ∞. □

Proposition 3.2. *Assume that $a > 0$ and $b < 0$. As T goes to ∞, $\widehat{a}_T$ satisfies the following CLT:*

$$\sqrt{T}(\widehat{a}_T - a) \xrightarrow{\mathrm{D}} \mathcal{N}\bigg(0, 4\mathbb{E}\bigg[\frac{X}{C^2}\bigg]\bigg(\mathbb{E}\bigg[\frac{1}{C}\bigg]\bigg)^{-2}\bigg). \tag{3.4}$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

Proof. This follows the same lines as the previous proof. Observe that

$$\sqrt{T}(\widehat{a}_T - a) = 2\left(\frac{1}{T}\int_0^T \frac{1}{C_t}\,\mathrm{d}t\right)^{-1}\frac{m_T}{\sqrt{T}},$$

where m_T is a martingale term given by

$$m_T = \int_0^T \frac{\sqrt{X_t}}{C_t}\,\mathrm{d}B_t, \qquad \langle m\rangle_T = \int_0^T \frac{X_t}{C_t^2}\,\mathrm{d}t.$$

Thus, as T goes to ∞,

$$\frac{\langle m\rangle_T}{T} \to \mathbb{E}\left[\frac{X}{C^2}\right] \quad \text{a.s.,} \tag{3.5}$$

which implies the following convergence in distribution:

$$\frac{m_T}{\sqrt{T}} \xrightarrow{\mathrm{D}} \mathcal{N}\left(0, \mathbb{E}\left[\frac{X}{C^2}\right]\right). \tag{3.6}$$

Finally, (3.6) leads to (3.4) thanks to the ergodicity of the process and Slutsky's lemma. □

Proposition 3.3. *Assume that $a > 2$ is known and $b < 0$. Then the MLE of b satisfies a CLT with a smaller asymptotic variance than the weighted least-squares estimator.*

Proof. Using the Cauchy–Schwarz inequality, we note that

$$\left(\mathbb{E}\left[\frac{X^2}{C}\right]\right)^2 = \left(\mathbb{E}\left[\sqrt{X}\frac{X^{3/2}}{C}\right]\right)^2 \le \mathbb{E}[X]\mathbb{E}\left[\frac{X^3}{C^2}\right],$$

which immediately leads to the result. □

Proposition 3.4. *Assume that $a > 2$ and $b < 0$ is known. Then the MLE of a satisfies a CLT with a smaller asymptotic variance than the WLSE.*

Proof. Using the Cauchy–Schwarz inequality, we note that

$$\left(\mathbb{E}\left[\frac{1}{C}\right]\right)^2 = \left(\mathbb{E}\left[X^{-1/2}\frac{X^{1/2}}{C}\right]\right)^2 \le \mathbb{E}\left[\frac{1}{X}\right]\mathbb{E}\left[\frac{X}{C^2}\right],$$

which immediately leads to the result. □

Remark 3.1. Thus, the WLSE is less efficient than the MLE in the case where the latter is easily manageable. This may appear to be contradictory to [27, Remark 4.4], in which the authors dealt with the discrete-time counterpart of the process. In fact, they compared the weighted least-squares with the conditional least-squares estimator which does not coincide with the MLE.

Remark 3.2. We could speculate on our choice of estimator, since the parameter a is unknown. However, we suppose that we observe the whole trajectory of the process over the time interval $[0, T]$. Thus, if we are able to detect some local time at level 0 then we know that $a < 2$ and we should use the WLSE instead of the MLE.

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

4. Technical lemmas

First, using (1.1), (1.4 can be written as

$$\widehat{\theta}_T = \theta + \begin{pmatrix} \Gamma_T^{-1} & 0 \\ 0 & \Gamma_T^{-1} \end{pmatrix} \begin{pmatrix} M_T \\ N_T \end{pmatrix}, \tag{4.1}$$

where M_T and N_T are martingales given by

$$M_T = \begin{pmatrix} \displaystyle\int_0^T \frac{2\sqrt{X_t}}{C_t}\,\mathrm{d}B_t \\ \displaystyle\int_0^T \frac{2\sqrt{X_t}X_t}{C_t}\,\mathrm{d}B_t \end{pmatrix}, \qquad N_T = \begin{pmatrix} \displaystyle\int_0^T \frac{2\sqrt{X_t}}{C_t}\,\mathrm{d}\widetilde{B}_t \\ \displaystyle\int_0^T \frac{2\sqrt{X_t}X_t}{C_t}\,\mathrm{d}\widetilde{B}_t \end{pmatrix}$$

with $\mathrm{d}\widetilde{B}_t = \rho\,\mathrm{d}B_t + \sqrt{1-\rho^2}\,\mathrm{d}W_t$.

We denote by $\mathcal{M}_T$ the martingale $\mathcal{M}_T = (M_T, N_T)$. As $\langle \mathrm{d}B_t, \mathrm{d}\widetilde{B}_t\rangle = \rho\,\mathrm{d}t$, we can easily see that the increasing process of $\mathcal{M}_T$ is given by

$$\langle \mathcal{M}\rangle_T = \begin{pmatrix} \langle M\rangle_T & \rho\langle M\rangle_T \\ \rho\langle M\rangle_T & \langle M\rangle_T \end{pmatrix}, \tag{4.2}$$

where the increasing process $\langle M\rangle_T$ of M_T is given by

$$\langle M\rangle_T = 4\begin{pmatrix} \langle m\rangle_T & \displaystyle\int_0^T \frac{X_t^2}{C_t^2}\,\mathrm{d}t \\ \displaystyle\int_0^T \frac{X_t^2}{C_t^2}\,\mathrm{d}t & \langle n\rangle_T \end{pmatrix}$$

with $\langle m\rangle_T$ and $\langle n\rangle_T$ given by (3.5) and (3.2), respectively.

In order to prove Theorems 2.1 and 2.2, we need to investigate the almost sure convergence of all the integrals involved in the definition of the estimators. Applying the result of Overbeck [24, Lemma 3(i)], we see that as T goes to ∞, X_T converges in distribution to X with gamma distribution $\Gamma(\frac{1}{2}a, -\frac{1}{2}b)$, whose probability density function is given by

$$f(x) = \big(\Gamma\big(\tfrac{1}{2}a\big)\big)^{-1}\big(-\tfrac{1}{2}b\big)^{a/2}x^{a/2-1}\mathrm{e}^{xb/2}\mathbf{1}_{\{x>0\}}. \tag{4.3}$$

Thus, by [24, Lemma 3(ii)], as T goes to ∞,

$$\frac{1}{T}\int_0^T g(X_t)\,\mathrm{d}t \to \mathbb{E}[g(X)] = \int_0^{+\infty} g(x)f(x)\,\mathrm{d}x \quad \text{a.s.}$$

for any function g such that the right-hand side exists.

We recall two properties of the incomplete gamma function that will be very useful in the proof of the next lemma:

$$\Gamma(\alpha+1, x) = x^{\alpha}\mathrm{e}^{-x} + \alpha\Gamma(\alpha, x) \tag{4.4}$$

and

$$\Gamma(\alpha+2, x) = x^{\alpha}\mathrm{e}^{-x}(x+\alpha+1) + \alpha(\alpha+1)\Gamma(\alpha, x).$$

We are now able to prove the following lemma. The first three points yield the almost sure limit of the matrix $T\Gamma_T^{-1}$ as T goes to ∞, while the remaining deals with the increasing process of the four-dimensional martingale $\mathcal{M}_T$ given by (4.2).

Lemma 4.1. *With ψ_c given by (2.3), we have*

(i) $\mathbb{E}[1/C] = \psi_c/c$;

(ii) $\mathbb{E}[X/C] = 1 - \psi_c$;

(iii) $\mathbb{E}[X^2/C] = c(\psi_c - 1) - a/b$;

(iv) $\mathbb{E}[X/C^2] = (a/2c)\psi_c + \frac{1}{2}b(1 - \psi_c)$;

(v) $\mathbb{E}[X^2/C^2] = \frac{1}{2}((a + 2 - bc)(1 - \psi_c) - a)$;

(vi) $\mathbb{E}[X^3/C^2] = \frac{1}{2}c(a + 4 - bc)(\psi_c - 1) + (\frac{1}{2}ac) - a/b$.

Proof. (i) We have

$$\mathbb{E}\bigg[\frac{1}{C}\bigg] = \int_0^{+\infty} \frac{1}{x + c} f(x)\,\mathrm{d}x,$$

where f is given by (4.3). Using [16, Equation 3.383(10)], we have

$$\int_0^{+\infty} \frac{1}{x + c} x^{a/2-1} \mathrm{e}^{xb/2}\,\mathrm{d}x = c^{a/2-1} \mathrm{e}^{-bc/2} \Gamma\bigg(\frac{a}{2}\bigg) \Gamma\bigg(1 - \frac{a}{2}, -\frac{bc}{2}\bigg),$$

which leads to

$$\mathbb{E}\bigg[\frac{1}{C}\bigg] = \frac{1}{c}\bigg(-\frac{bc}{2}\bigg)^{a/2} \mathrm{e}^{-bc/2} \Gamma\bigg(1 - \frac{a}{2}, -\frac{bc}{2}\bigg)$$

and ensures the announced result.

(ii) As before, we have

$$\mathbb{E}\bigg[\frac{X}{C}\bigg] = \int_0^{+\infty} \frac{x}{x + c} f(x)\,\mathrm{d}x. \tag{4.5}$$

Using [16, Equation 3.383(10)], we have

$$\int_0^{+\infty} \frac{1}{x + c} x^{a/2} \mathrm{e}^{xb/2}\,\mathrm{d}x = c^{a/2} \mathrm{e}^{-bc/2} \Gamma\bigg(\frac{a}{2} + 1\bigg) \Gamma\bigg(-\frac{a}{2}, -\frac{bc}{2}\bigg). \tag{4.6}$$

Together with (4.4), we easily obtain

$$\Gamma\bigg(-\frac{a}{2}, -\frac{bc}{2}\bigg) = \bigg(-\frac{2}{a}\bigg)\bigg(\Gamma\bigg(1 - \frac{a}{2}, -\frac{bc}{2}\bigg) - \bigg(-\frac{bc}{2}\bigg)^{-a/2} \mathrm{e}^{bc/2}\bigg). \tag{4.7}$$

Combining (4.5)–(4.7) and the fact that $\Gamma(\frac{1}{2}a + 1) = \frac{1}{2}a \times \Gamma(\frac{1}{2}a)$, we deduce the announced result.

(iii) We have

$$\mathbb{E}\bigg[\frac{X^2}{C}\bigg] = \mathbb{E}\bigg[\frac{(X + c - c)^2}{X + c}\bigg] = \mathbb{E}[X] - c + c^2 \mathbb{E}\bigg[\frac{1}{C}\bigg],$$

and we conclude using (i) and the fact that $\mathbb{E}[X] = -a/b$.

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

(iv) By the definition of f in (4.3), we have

$$\mathbb{E}\left[\frac{X}{C^2}\right] = \int_0^{+\infty} \frac{x}{(x+c)^2} f(x)\,\mathrm{d}x = \frac{(-b/2)^{a/2}}{\Gamma(a/2)} \int_0^{+\infty} \frac{x^{a/2}}{(x+c)^2} \mathrm{e}^{xb/2}\,\mathrm{d}x. \tag{4.8}$$

Using integration by parts on the right-hand side of (4.8), we obtain

$$\mathbb{E}\left[\frac{X}{C^2}\right] = \frac{(-b/2)^{a/2}}{\Gamma(a/2)} \left[\frac{a}{2} \int_0^{+\infty} \frac{x^{a/2-1}}{x+c} \mathrm{e}^{xb/2}\,\mathrm{d}x + \frac{b}{2} \int_0^{+\infty} \frac{x^{a/2}}{x+c} \mathrm{e}^{xb/2}\,\mathrm{d}x\right].$$

Since both integrals were computed in the proofs of (i) and (ii), we obtain

$$\mathbb{E}\left[\frac{X}{C^2}\right] = \frac{a}{2}\mathbb{E}\left[\frac{1}{C}\right] + \frac{b}{2}\mathbb{E}\left[\frac{X}{C}\right] = \frac{a}{2c}\psi_c + \frac{b}{2}(1-\psi_c).$$

(v) Integrating by parts and using (iii) and (iv),

$$\begin{aligned}
\mathbb{E}\left[\frac{X^2}{C^2}\right] &= \frac{(-b/2)^{a/2}}{\Gamma(a/2)} \int_0^{+\infty} \frac{x^{a/2+1}}{(x+c)^2} \mathrm{e}^{xb/2}\,\mathrm{d}x \\
&= \frac{(-b/2)^{a/2}}{\Gamma(a/2)} \left[\frac{a+2}{2} \int_0^{+\infty} \frac{x^{a/2}}{x+c} \mathrm{e}^{xb/2}\,\mathrm{d}x + \frac{b}{2} \int_0^{+\infty} \frac{x^{a/2+1}}{x+c} \mathrm{e}^{xb/2}\,\mathrm{d}x\right] \\
&= \frac{1}{2}\left((a+2)\mathbb{E}\left[\frac{X}{C}\right] + b\mathbb{E}\left[\frac{X^2}{C}\right]\right) \\
&= \frac{1}{2}\left((a+2)(1-\psi_c) + b\left(c(\psi_c - 1) - \frac{a}{b}\right)\right).
\end{aligned}$$

(vi) Noting that $X^3 = X(X+c)^2 - 2cX^2 - c^2X$, we obtain

$$\mathbb{E}\left[\frac{X^3}{C^2}\right] = \mathbb{E}[X] - 2c\mathbb{E}\left[\frac{X^2}{C^2}\right] - c^2\mathbb{E}\left[\frac{X}{C^2}\right]$$

and we conclude using (iv) and (v). □

5. Proof of the strong consistency

We are now in a position to prove Theorem 2.1.

Proof of Theorem 2.1. First, we have

$$\frac{1}{T^2}\det\Gamma_T = \frac{1}{T}\int_0^T \frac{1}{C_t}\,\mathrm{d}t \frac{1}{T}\int_0^T \frac{X_t^2}{C_t}\,\mathrm{d}t - \left(\frac{1}{T}\int_0^T \frac{X_t}{C_t}\,\mathrm{d}t\right)^2.$$

Thus, as the process is ergodic, we obtain, as T goes to ∞,

$$\frac{1}{T^2}\det\Gamma_T \to \mathbb{E}\left[\frac{1}{C}\right]\mathbb{E}\left[\frac{X^2}{C}\right] - \left(\mathbb{E}\left[\frac{X}{C}\right]\right)^2 \quad \text{a.s.} \tag{5.1}$$

Using the fact that $X = C - c$, we obtain

$$\mathbb{E}\left[\frac{X^2}{C}\right] = \mathbb{E}[C] - 2c + c^2\mathbb{E}\left[\frac{1}{C}\right] \quad \text{and} \quad \mathbb{E}\left[\frac{X}{C}\right] = 1 - c\mathbb{E}\left[\frac{1}{C}\right].$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

We can easily derive

$$\mathbb{E}\Big[\frac{1}{C}\Big]\mathbb{E}\Big[\frac{X^2}{C}\Big]-\Big(\mathbb{E}\Big[\frac{X}{C}\Big]\Big)^2=\mathbb{E}\Big[\frac{1}{C}\Big]\mathbb{E}[C]-1,$$

which leads to the limit as T goes to ∞,

$$\frac{1}{T^2}\det\Gamma_T\to\mathbb{E}\Big[\frac{1}{C}\Big]\mathbb{E}[C]-1\quad\text{a.s.}$$

Thus, as T goes to ∞, we obtain

$$T\Gamma_T^{-1}\to A\quad\text{a.s.,}\tag{5.2}$$

where A is given by

$$A=\Big(\mathbb{E}[C]\mathbb{E}\Big[\frac{1}{C}\Big]-1\Big)^{-1}\begin{pmatrix}\mathbb{E}\Big[\frac{X^2}{C}\Big] & -\mathbb{E}\Big[\frac{X}{C}\Big]\\ -\mathbb{E}\Big[\frac{X}{C}\Big] & \mathbb{E}\Big[\frac{1}{C}\Big]\end{pmatrix}.\tag{5.3}$$

A straightforward application of Lemmas 4.1(i)–(iii) yields

$$A=\frac{1}{\psi_c(1-a/bc)-1}\begin{pmatrix}c(\psi_c-1)-\dfrac{a}{b} & \psi_c-1\\ \psi_c-1 & \dfrac{\psi_c}{c}\end{pmatrix}.$$

Besides, using the strong law of large numbers for the martingale, we see that the martingale M_T satisfies, as T goes to ∞,

$$\frac{M_T}{T}\to 0\quad\text{a.s.}\tag{5.4}$$

As a matter of fact, by convergences (3.2) and (3.5), we know that a.s. $\langle n\rangle_T=\mathcal{O}(T)$ and $\langle m\rangle_T=\mathcal{O}(T)$. This ensures that, as T goes to ∞, $n_T/T\to 0$ a.s. and $m_T/T\to 0$ a.s. As N_T and M_T share the same increasing process, this result holds by replacing M_T by N_T. Finally, the almost sure convergence (2.1) follows from (4.1), (5.2), and (5.4). □

6. Proof of the asymptotic normality

Proof of Theorem 2.2. First, from (4.1), we deduce that

$$\sqrt{T}(\widehat{\theta}_T-\theta)=\begin{pmatrix}T\Gamma_T^{-1} & 0\\ 0 & T\Gamma_T^{-1}\end{pmatrix}\begin{pmatrix}\dfrac{M_T}{\sqrt{T}}\\ \dfrac{N_T}{\sqrt{T}}\end{pmatrix}.\tag{6.1}$$

We already know that $T\Gamma_T^{-1}$ converges a.s. as T goes to ∞ and its limit A is given by (5.3). We now have to establish the asymptotic normality of $M_T/\sqrt{T}$. By the ergodicity of the process, we obtain

$$\frac{\langle M\rangle_T}{T}\to 4L\quad\text{a.s.,}\quad\text{where } L=\begin{pmatrix}\mathbb{E}\Big[\frac{X}{C^2}\Big] & \mathbb{E}\Big[\frac{X^2}{C^2}\Big]\\ \mathbb{E}\Big[\frac{X^2}{C^2}\Big] & \mathbb{E}\Big[\frac{X^3}{C^2}\Big]\end{pmatrix}.$$

As a straightforward consequence of Lemmas 4.1(iv)–(vi), we obtain

$$L = \frac{1}{2}\begin{pmatrix} \dfrac{a}{c}\psi_c + b(1-\psi_c) & (a+2-bc)(1-\psi_c) - a \\ (a+2-bc)(1-\psi_c) - a & c(a+4-bc)(\psi_c - 1) + ac - \dfrac{2a}{b} \end{pmatrix}.$$

We easily obtain the following almost sure convergence:

$$\frac{\langle \mathcal{M} \rangle_T}{T} \to 4\mathcal{L} \quad \text{a.s.},$$

where $\mathcal{L}$ is a block matrix given by

$$\mathcal{L} = \begin{pmatrix} L & \rho L \\ \rho L & L \end{pmatrix}.$$

From the CLT for martingales, we deduce that

$$\frac{\mathcal{M}_T}{\sqrt{T}} \xrightarrow{\text{D}} \mathcal{N}(0, 4\mathcal{L}). \tag{6.2}$$

Finally, the asymptotic normality (2.2) follows from (6.1) and (6.2) together with Slutsky's lemma. □

7. Numerical simulations

The efficient discretization of the CIR process is a challenging problem; see, for example, [2] and [3]. We choose to implement the quadratic-exponential algorithm based on quadratic-exponential approximations proposed by Andersen [3]. Andersen introduced this algorithm to deal with the $a < 2$ case, for which common discretization schemes are not accurate.

7.1. Asymptotic behavior for $c = 1$

In Figures 1 and 2 we illustrate our main results: strong consistency and asymptotic normality, respectively, in the case where $a = 1$ and $b = -2$ with the weighting parameter $c = 1$. The curves in Figure 2 correspond to the standard normal distribution. We denote by v_a (respectively, v_b) the variance of $\widehat{a}_T$ (respectively, $\widehat{b}_T$).

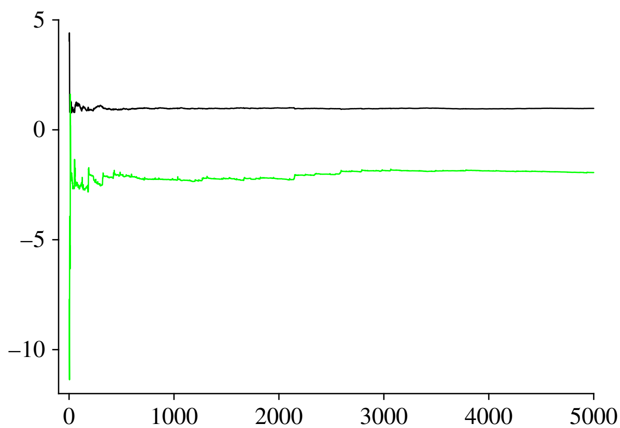

FIGURE 1: Strong consistency of $(\widehat{a}_T)$ (*upper plot*) and $(\widehat{b}_T)$ (*lower plot*).

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

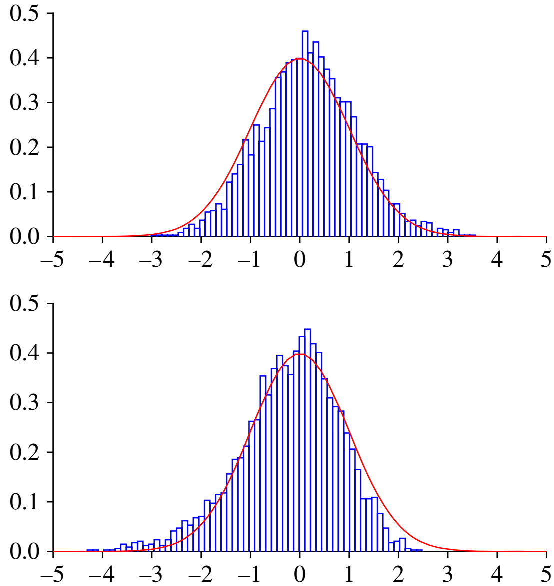

FIGURE 2: Histograms of 3000 outcomes of $\sqrt{T/v_a}(\widehat{a}_T - a)$ (*upper*) and $\sqrt{T/v_b}(\widehat{b}_T - b)$ (*lower*) at time $T = 70$.

7.2. Choice of the constant c

We choose to introduce a constant c in our weighting, instead of considering only the $c = 1$ case (as carried out for the discrete-time case in [27]) with the aim of lowering the variance of the estimators. However, this raises the question as to the optimal choice of the constant c, which depends on the values of parameters a and b. We set $a = 1$ and $b = -4$ and simulate 500 trajectories of the process over the time interval $[0, 50]$. We compute the empirical variance of the estimators given by each trajectory for c varying between 10^{-10} and 1; see Figure 3. It appears that we should choose a small value of c. The value should not be too small to avoid the growth seen in the right-hand panels of Figure 3, which might, however, be a consequence of the discretized version of the CIR process we used. For $\widehat{a}_T$, there is a significant difference (a five-fold factor) between the empirical variances obtained with $c = 0.01$ and $c = 1$. However, for $\widehat{b}_T$ both empirical variances do not significantly differ.

Appendix A. Motivation of the chosen estimator

We rewrite (1.1) by making use of the weighting $C_t^{-1/2}$:

$$C_t^{-1/2}\,\mathrm{d}X_t = (a + bX_t)C_t^{-1/2}\,\mathrm{d}t + 2C_t^{-1/2}\sqrt{X_t}\,\mathrm{d}B_t,$$
$$C_t^{-1/2}\,\mathrm{d}Y_t = (\alpha + \beta X_t)C_t^{-1/2}\,\mathrm{d}t + 2C_t^{-1/2}\sqrt{X_t}\,\mathrm{d}\widetilde{B}_t.$$

Using the fact that $X_t = C_t - c$, we obtain

$$C_t^{-1/2}\,\mathrm{d}X_t = (\ell C_t^{-1/2} + bC_t^{1/2})\,\mathrm{d}t + 2C_t^{-1/2}\sqrt{X_t}\,\mathrm{d}B_t,$$
$$C_t^{-1/2}\,\mathrm{d}Y_t = (\lambda C_t^{-1/2} + \beta C_t^{1/2})\,\mathrm{d}t + 2C_t^{-1/2}\sqrt{X_t}\,\mathrm{d}\widetilde{B}_t,$$

where $\ell = a - cb$ and $\lambda = \alpha - c\beta$.

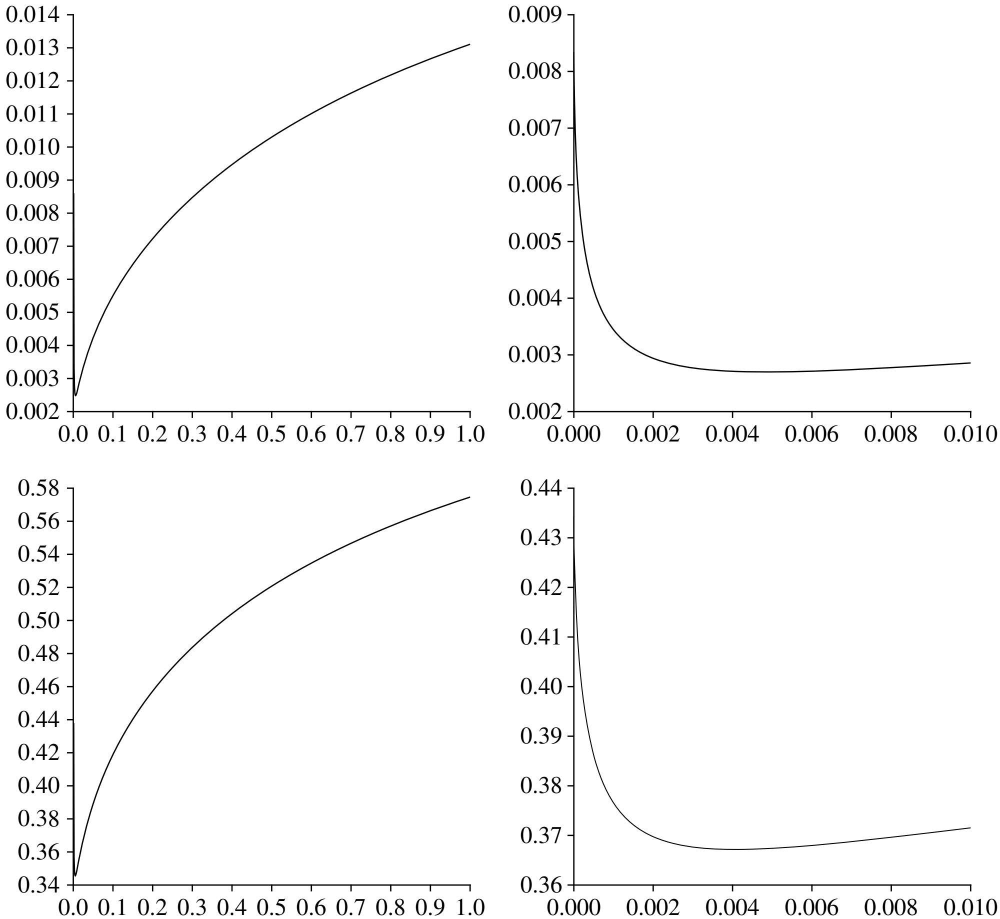

FIGURE 3: Variance of $\widehat{a}_{50}$ (*upper*) and $\widehat{b}_{50}$ (*lower*) as functions of c. Note that the left- and right-hand plots are from two different simulations.

In what follows, we consider only the first equation as the results for the second one can be derived the same way. Suppose that we want to compute the usual least-squares estimator $\widetilde{\mu}_n$ of the couple (ℓ, b) based on discrete-time observations $(X_0, X_1, \dots, X_n)$, which would, in fact, be a WLSE due to the presence of C_t in the above equations. It is the solution of the following minimization problem:

$$\widetilde{\mu}_n = (\widetilde{\ell}_n, \widetilde{b}_n) = \arg\min_{\ell>0, b<0} \sum_{i=1}^{n} \biggl(\frac{X_i - X_{i-1}}{\sqrt{X_{i-1}+c}} - \frac{\ell}{\sqrt{X_{i-1}+c}} - b\sqrt{X_{i-1}+c} \biggr)^2 .$$

We investigate the critical points and easily obtain

$$\widetilde{\ell}_n = \frac{n \sum_{i=1}^{n} (X_i - X_{i-1}) - \sum_{i=1}^{n} (X_{i-1}+c) \sum_{i=1}^{n} (X_i - X_{i-1})/(X_{i-1+c})}{n^2 - \sum_{i=1}^{n} (X_{i-1}+c) \sum_{i=1}^{n} 1/(X_{i-1+c})}$$

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

and

$$\widetilde{b}_n = \frac{n \sum_{i=1}^{n} (X_i - X_{i-1})/(X_{i-1+c}) - \sum_{i=1}^{n} (X_i - X_{i-1}) \sum_{i=1}^{n} 1/(X_{i-1+c})}{n^2 - \sum_{i=1}^{n} (X_{i-1} + c) \sum_{i=1}^{n} 1/(X_{i-1+c})}.$$

Besides, as $\ell = a - cb$, we obtain an estimator $\widetilde{a}_n$ for a given by $\widetilde{a}_n = \widetilde{\ell}_n + c\widetilde{b}_n$.

Motivated by those equations, we introduce a WLSE based on the continuous-time observations $(X_t)_{t \leq T}$ given by

$$\widehat{b}_T = \frac{T \int_0^T \mathrm{d}X_t / C_t - (X_T - X_0) \int_0^T (1/C_t)\,\mathrm{d}t}{T^2 - \int_0^T C_t \,\mathrm{d}t \int_0^T (1/C_t)\,\mathrm{d}t}$$

$$\widehat{a}_T = \frac{(T - c \int_0^T (1/C_t)\,\mathrm{d}t)(X_T - X_0) - \int_0^T X_t \,\mathrm{d}t \int_0^T \mathrm{d}X_t / C_t}{T^2 - \int_0^T C_t \,\mathrm{d}t \int_0^T 1/C_t \,\mathrm{d}t}.$$

Using the fact that $X_t = C_t - c$, these

Acknowledgements

The author would like to thank Bernard Bercu and Adrien Richou as well as the anonymous referee for their careful reading and their fruitful comments.

References

[1] Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *J. Financial Econom.* **83,** 413–452.

[2] Alfonsi, A. (2010). High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Math. Comput.* **79,** 209–237.

[3] Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *J. Comput. Finance* **11,** 1–42.

[4] Azencott, R. and Gadhyan, Y. (2009). Accurate parameter estimation for coupled stochastic dynamics. *Discrete Contin. Dyn. Syst.* **2009,** 44–53.

[5] Barczy, M. and Pap, G. (2016). Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations. *Statistics* **50,** 389–417.

[6] Ben Alaya, M. and Kebaier, A. (2012). Parameter estimation for the square-root diffusions: ergodic and nonergodic cases. *Stoch. Models* **28,** 609–634.

[7] Ben Alaya, M. and Kebaier, A. (2013). Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stoch. Anal. Appl.* **31,** 552–573.

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

[9] Du Roy de Chaumaray, M. (2017). Large deviations for the squared radial Ornstein–Uhlenbeck process. *Theory Prob. Appl.* **61,** 408–441.

[10] Feller, W. (1951). Two singular diffusion problems. *Ann. Math. (2)* **54,** 173–182.

[11] Forde, M. and Jacquier, A. (2011). The large-maturity smile for the Heston model. *Finance Stoch.* **15,** 755–780.

[12] Forde, M., Jacquier, A. and Lee, R. (2012). The small-time smile and term structure of implied volatility under the Heston model. *SIAM J. Financial Math.* **3,** 690–708.

[13] Fournié, E. and Talay, D. (1991). Application de la statistique des diffusions à un modèle de taux d'intérêt. *Finance* **12,** 79–111.

[14] Gao, F. and Jiang, H. (2009). Moderate deviations for squared Ornstein–Uhlenbeck process. *Statist. Prob. Lett.* **79,** 1378–1386.

[15] Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. John Wiley, Hoboken, NJ.

[16] Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products*. Academic Press, New York.

[17] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Studies* **6,** 327–343.

https://doi.org/10.1017/jpr.2018.34 Published online by Cambridge University Press

[18] Jacquier, A. and Roome, P. (2016). Large-maturity regimes of the Heston forward smile. *Stoch. Process. Appl.* **126,** 1087–1123.
[19] Janek, A., Kluge, T., Weron, R. and Wystup, U. (2011). FX smile in the Heston model. In *Statistical Tools for Finance and Insurance*, Springer, Heidelberg, pp. 133–162.
[20] Lamberton, D. and Lapeyre, B. (1997). *Introduction au Calcul Stochastique Appliqué à la Finance*, 2nd edn. Ellipses Édition Marketing, Paris.
[21] Lee, R. W. (2004). Option pricing by transform methods: extensions, unification and error control. *J. Comput. Finance* **7,** 51–86.
[22] Lewis, A. L. (2000). *Option Valuation Under Stochastic Volatility*. Finance Press, Newport Beach, CA.
[23] Luke, Y. L. (1969). *The Special Functions and Their Approximations*, Vol. II. Academic Press, New York.
[24] Overbeck, L. (1998). Estimation for continuous branching processes. *Scand. J. Statist.* **25,** 111–126.
[25] Stein, E. M. and Stein, J. C. (1991). Stock price distributions with stochastic volatility: an analytic approach. *Rev. Financial Studies* **4,** 727–752.
[26] Stein, J. (1989). Overreactions in the options market. *J. Finance* **44,** 1011–1023.
[27] Wei, C. Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.* **18,** 1757–1773.
[28] Zani, M. (2002). Large deviations for squared radial Ornstein–Uhlenbeck processes. *Stoch. Process. Appl.* **102,** 25–42.