

Weighted least squares estimation for the subcritical Heston process

Marie Du Roy de Chaumaray

▶ **To cite this version:**

Marie Du Roy de Chaumaray. Weighted least squares estimation for the subcritical Heston process. 2016. hal-01207617v2

HAL Id: hal-01207617
https://hal.science/hal-01207617v2

Preprint submitted on 20 Jul 2016 (v2), last revised 25 Jan 2018 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WEIGHTED LEAST SQUARES ESTIMATION FOR THE SUBCRITICAL HESTON PROCESS

MARIE DU ROY DE CHAUMARAY

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

1. Introduction

Introduced in 1973, as an 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 to fit real market data, especially during the crash in 1987, see [25]. 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. A particular attention has been drawn to the study of stochastic volatility processes in which the volatility is also given as a solution of some stochastic differential equation, see [24], [21] and [14] for financial accuracy. Among them, Heston process [16] is one of the most popular, due to its computational tractability. For example, [20] easily computes call option prices using Fourier inversion techniques. Numerous results about the asymptotic volatility smile can be found in the very recent literature: see e.g. [10], [11], [17].

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

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

with $a > 0$, $(b, \alpha, \beta) \in \mathbb{R}^3$ and $\rho \in]-1, 1[$, where (B_t, W_t) is a 2-dimensional standard Wiener process and the initial state $(x_0, y_0) \in \mathbb{R}^+ \times \mathbb{R}$. In this process, the volatility X_t is driven by a generalized squared radial Ornstein-Uhlenbeck process, also known as the CIR process, firstly studied by Feller [9] and introduced in a financial context by Cox, Ingersoll and Ross [7] 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.

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) 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 an 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) has been recently

investigated in [5], together with its asymptotic behavior in the special case where $a \geqslant 2$. 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 $\widetilde{\theta}_T = \left(\widetilde{a}_T, \widetilde{b}_T, \widetilde{\alpha}_T, \widetilde{\beta}_T\right)$ is given, for $T < \tau_0$, by:

$$\widetilde{\theta}_T = \begin{pmatrix} G_T^{-1} & 0 \\ 0 & G_T^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{U}_T \\ \widetilde{V}_T \end{pmatrix} \tag{1.3}$$

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

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

One can 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 [12], [23] and [6]. In the supercritical case $b > 0$, Overbeck [23] has shown that $\widetilde{b}_T$ converges a.s. to b whereas there exists no consistent estimator for a. Hence, we will focus our attention on the geometrically ergodic case $b < 0$. Furthermore, the value of a governs the behavior at zero of (X_T): for $a \geqslant 2$, the process almost surely never reaches zero , whereas for $0 < a < 2$, zero is quite frequently visited and

$$\mathbb{P}\left(\tau_0 < \infty\right) = 1, \tag{1.4}$$

see for instance [19] or [23]. For $a > 2$, the MLE converges a.s. to $\theta = (a, b, \alpha, \beta)$ and satisfies the following Central Limit Theorem (CLT)

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

where the block matrix D is given by

$$D = \begin{pmatrix} \Sigma & \rho\Sigma \\ \rho\Sigma & \Sigma \end{pmatrix} \qquad \text{with} \qquad \Sigma = \begin{pmatrix} \frac{-b}{a-2} & 1 \\ 1 & -\frac{a}{b} \end{pmatrix}.$$

A large deviation principle for the couple $(\widetilde{a}_T, \widetilde{b}_T)$ was recently established in [8]. In the particular case where one parameter is known and the other one is estimated, large deviations can be found in [27], while moderate deviations are given in [13].

By contrast, in the case where $0 < a < 2$, (1.4) implies the non-integrability of X_T^{-1} for large values of T so that the MLE does not converge for T going to infinity. Consequently, this case has been less investigated even though it is often of interest in finance, to compute long dated interest rates for instance, as explained in [3], or in FX-markets, see [18]. In the case of the CIR process, Overbeck [23] used accurate stopping times to build a strongly consistent estimator based on the MLE:

$$\mathbb{1}_{T<\tau_0} \begin{pmatrix} \widetilde{a}_T \\ \widetilde{b}_T \end{pmatrix} + \mathbb{1}_{\tau_0 \leqslant T} \begin{pmatrix} \lim\limits_{t\to\tau_0} S_t\Sigma_t^{-1} \\ \left(\int_0^T X_s\,\mathrm{d}s\right)^{-1} \left(X_T - T \lim\limits_{t\to\tau_0} S_t\Sigma_t^{-1}\right) \end{pmatrix} \tag{1.5}$$

where $S_t = \int_0^t X_s^{-1}\,\mathrm{d}X_s$, $\Sigma_t = \int_0^t X_s^{-1}\,\mathrm{d}s$ and τ_0 is given by (1.4). 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 zero and consequently avoid us 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 [26] 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 [6] regarding the CIR process.

Following the seminal work of [26], denote $C_T = X_T + c$ where c is some positive constant. Our new couple of weighted least squares estimator is given by

$$\widehat{\theta}_T = \begin{pmatrix} \Gamma_T^{-1} & 0 \\ 0 & \Gamma_T^{-1} \end{pmatrix} \begin{pmatrix} U_T \\ V_T \end{pmatrix} \tag{1.6}$$

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

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

We do not restrict ourselves to the case where $c = 1$ as it may lower sometimes the variance of the estimators. In the particular case where $c = 0$, one can observe that the new estimator coincides with the MLE.

The paper is organized as follows. The second section contains our main results: the strong consistency of this new couple of estimators as well as its asymptotic normality. The third section deals with a comparison with the MLE, while the remaining 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: for T going to infinity,*

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

For T going to infinity, X_T converges in distribution to a random variable X with Gamma $\Gamma(a/2, -b/2)$ distribution, see Lemma 3 of [23] for instance. Additionally, we denote by C the limiting distribution of $X_T + c$, as T goes to infinity.

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

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

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

$$\Lambda = \begin{pmatrix} ALA & \rho ALA \\ \rho ALA & ALA \end{pmatrix}, \tag{2.3}$$

with the matrix A and L respectively given by

$$A = \left(\mathbb{E}\left[C\right] \mathbb{E}\left[1/C\right] - 1 \right)^{-1} \begin{pmatrix} \mathbb{E}\left[X^2/C\right] & -\mathbb{E}\left[X/C\right] \\ -\mathbb{E}\left[X/C\right] & \mathbb{E}\left[1/C\right] \end{pmatrix}$$

and

$$L = \begin{pmatrix} \mathbb{E}\left[X/C^2\right] & \mathbb{E}\left[X^2/C^2\right] \\ \mathbb{E}\left[X^2/C^2\right] & \mathbb{E}\left[X^3/C^2\right] \end{pmatrix}.$$

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 for T going to infinity,*

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

and satisfies the following CLT

$$\sqrt{T} \begin{pmatrix} \widehat{a}_T - a \\ \widehat{b}_T - b \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4ALA\right),$$

Remark 2.1. *In the remaining of this paper, we denote*

$$\psi_c = \left(-\frac{bc}{2}\right)^{a/2} e^{-bc/2}\, \Gamma\left(1 - a/2, -bc/2\right), \tag{2.4}$$

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

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

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

$$\varphi_c = \psi_c \left(1 - \frac{a}{bc}\right) - 1. \tag{2.5}$$

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

$$A = \varphi_c^{-1} \begin{pmatrix} c\left(\psi_c - 1\right) - \frac{a}{b} & \psi_c - 1 \\ \psi_c - 1 & \frac{\psi_c}{c} \end{pmatrix} \tag{2.6}$$

and

$$L = \frac{1}{2} \begin{pmatrix} \dfrac{a}{c}\,\psi_c + b\left(1 - \psi_c\right) & \left(a + 2 - bc\right)\left(1 - \psi_c\right) - a \\ \left(a + 2 - bc\right)\left(1 - \psi_c\right) - a & \psi_c c\left(a + 4 - b\right) - 4c - bc^2 - \dfrac{2a}{b} \end{pmatrix}. \tag{2.7}$$

By a straightforward computation, we deduce that $ALA = \left(\varphi_c\right)^{-2} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ where the variances σ_{11} and σ_{22} are respectively given by

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

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

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

Remark 2.2. *For c going to zero (for which we need a to be greater than 2) , we obtain the same covariance matrix than for the MLE. Indeed, using well-known asymptotic results about the incomplete Gamma function Γ, which could be found in* [22]*, we have that, as soon as $a > 2$,*

$$\Gamma\left(1 - a/2, -bc/2\right) \underset{c \to 0}{\longrightarrow} 0 \tag{2.8}$$

and

$$\Gamma\left(1 - a/2, -bc/2\right) \left(-\frac{bc}{2}\right)^{a/2-1} \underset{c \to 0}{\longrightarrow} \frac{1}{1 - a/2} = \frac{2}{a - 2}. \tag{2.9}$$

Thus ψ_c goes to zero for c tending to zero and $\frac{\psi_c}{c}$ converges to $\frac{-b}{a-2}$. Hence, we easily obtain that, for c going to zero,

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

3. Asymptotic variance

Even though we considered the weighted least squares estimators in order to investigate the case $0 < a < 2$ for which the MLE is not consistent, it is interesting to compare the asymptotic variances in the CLT of this new estimators and of the MLE, in the case where $a > 2$. This comparison requires a lot of technical calculation as the asymptotic variances depends on the value 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 one of the parameter to be 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

$$\breve{b}_T = \frac{X_T - aT}{\int_0^T X_t \, \mathrm{d}t} \tag{3.1}$$

and satisfies the following CLT

$$\sqrt{T}\left(\breve{b}_T - b\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4/\mathbb{E}\left[X\right])$$

where $\mathbb{E}\left[X\right] = -a/b$, see for instance [23]. On the other hand, if b is known, the MLE of a is given by

$$\breve{a}_T = \frac{\int_0^T 1/X_t \, \mathrm{d}X_t - bT}{\int_0^T 1/X_t \, \mathrm{d}t} \tag{3.2}$$

and satisfies the following CLT

$$\sqrt{T}\left(\breve{a}_T - a\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4/\mathbb{E}\left[X^{-1}\right])$$

with $\mathbb{E}\left[X^{-1}\right] = -b/(a-2)$. Whereas, the weighted least squares estimators are respectively given by

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

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

$$\sqrt{T}\left(\widehat{b}_T - b\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\,\mathbb{E}\left[X^3/C^2\right]\left(\mathbb{E}\left[X^2/C\right]\right)^{-2}\right). \tag{3.3}$$

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

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

where n_T is a martingale given by

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

Using the ergodicity of the process, we obtain for T going to infinity

$$\frac{\langle n \rangle_T}{T} \xrightarrow{a.s.} \mathbb{E}\left[X^3/C^2\right]. \tag{3.4}$$

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

$$\frac{n_T}{\sqrt{T}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{E}\left[X^3/C^2\right]\right). \tag{3.5}$$

Consequently, (3.3) follows from (3.5), Slutsky's lemma and the fact that, by the ergodicity of the process, $\frac{1}{T}\int_0^T X_t^2/C_t \, \mathrm{d}t$ converges a.s. to $\mathbb{E}\left[X^2/C\right]$ for T going to infinity. □

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

$$\sqrt{T}\left(\widehat{a}_T - a\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\,\mathbb{E}\left[X/C^2\right]\left(\mathbb{E}\left[1/C\right]\right)^{-2}\right). \tag{3.6}$$

Proof. It works as in the previous proof. One can observe that

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

where m_T is a martingale term given by

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

Thus, for T going to infinity,

$$\frac{\langle m \rangle_T}{T} \xrightarrow{a.s.} \mathbb{E}\left[X/C^2\right] \tag{3.7}$$

which implies the following convergence in distribution

$$\frac{m_T}{\sqrt{T}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{E}\left[X/C^2\right]\right). \tag{3.8}$$

Finally, (3.8) leads to (3.6) 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 Cauchy-Schwarz Inequality, we notice that

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

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 weighted least squares estimator.*

Proof. Using Cauchy-Schwarz Inequality, we notice that

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

which immediately leads to the result. □

Remark 3.1. *Thus, the weighted least squares estimator is less efficient than the MLE in the case where this later is easily manageable. This could seem to be contradictory to Remark 4.4 of* [26] *which deals with the discrete-time counterpart of the process. In fact, they compare the weighted least squares with the conditional least squares estimator which does not coincide with the MLE.*

Remark 3.2. *One could wonder how to choose which estimator to use, as 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 zero, we know that $a < 2$ and we should use the WLSE instead of the MLE.*

4. Technical Lemmas

In order to prove Theorem 2.1, we need to investigate the almost sure convergence of all the integrals involved in the definition of the estimators. Overbeck recalls in Lemma 3*(i)* of [23] that, for T going to infinity, X_T converges in distribution to X with Gamma $\Gamma(a/2, -b/2)$ distribution, whose probability density function is given by

$$f(x) = \left(\Gamma(a/2)\right)^{-1} (-b/2)^{a/2}\, x^{a/2-1} e^{xb/2} \mathbf{1}_{x>0}. \tag{4.1}$$

Thus, by Lemma 3*(ii)* of [23], for T going to infinity,

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

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

By an integration by part, we easily show the two following properties of the incomplete gamma function, which will be very useful in the following proof:

$$\Gamma(\alpha+1, x) = x^{\alpha} e^{-x} + \alpha\Gamma\left(\alpha, x\right) \tag{4.2}$$

and

$$\Gamma(\alpha+2, x) = x^{\alpha} e^{-x}\left(x+\alpha+1\right) + \alpha\left(\alpha+1\right)\Gamma\left(\alpha, x\right). \tag{4.3}$$

We are now able to prove the following lemmas. The three first points give us the almost sure limit of $T\Gamma_T^{-1}$ as T goes to infinity, while the remaining deals with the increasing process of the right-hand side two-dimensional martingale of (1.6).

Lemma 4.1. *With ψ_c given by (2.4), we have that*

(i) $\mathbb{E}\left[1/C\right] = \frac{\psi_c}{c}$.
(ii) $\mathbb{E}\left[X/C\right] = 1 - \psi_c$.
(iii) $\mathbb{E}\left[X^2/C\right] = c\left(\psi_c - 1\right) - \frac{a}{b}$.
(iv) $\mathbb{E}\left[X/C^2\right] = \frac{a}{2c}\psi_c + \frac{b}{2}\left(1-\psi_c\right)$.
(v) $\mathbb{E}\left[X^2/C^2\right] = \frac{1}{2}\left(\left(a+2-bc\right)\left(1-\psi_c\right) - a\right)$.
(vi) $\mathbb{E}\left[X^3/C^2\right] = \frac{c}{2}\left(a+4-b\right)\psi_c - 2c - \frac{bc^2}{2} - \frac{a}{b}$.

Proof. (i) We have

$$\mathbb{E}\left[1/C\right] = \int_0^{+\infty} \frac{1}{x+c} f(x)\, \mathrm{d}x, \tag{4.4}$$

where f is given by (4.1). Formula 3.38(10) of [15] gives that

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

which leads to

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

and ensures the announced result.

(ii) As in the previous proof, we have

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

By formula 3.38(10) of [15], we know that

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

With formula (4.2), we easily obtain that

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

Combining (4.5), (4.6), (4.7) and the fact that $\Gamma(a/2+1)=a/2\times\Gamma(a/2)$, we deduce the announced result.

(iii) We have

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

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

(iv) By the very definition of f given by (4.1), we have

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

Integrating the right-hand side of (4.8) by part, we obtain that

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

We have already computed both integrals in the proofs of respectively (i) and (ii), which leads to

$$\mathbb{E}\left[X/C^2\right]=\frac{a}{2}\,\mathbb{E}\left[1/C\right]+\frac{b}{2}\,\mathbb{E}\left[X/C\right]=\frac{a}{2c}\,\psi_c+\frac{b}{2}\,\left(1-\psi_c\right). \tag{4.9}$$

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

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

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

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

and we conclude using (iv) and (v).

□

5. Proof of the strong Consistency

We are now in the position to prove Theorem 2.1. We first rewrite (1.6) using (1.1):

$$\widehat{\theta}_T=\theta+\begin{pmatrix}\Gamma_T^{-1} & 0\\ 0 & \Gamma_T^{-1}\end{pmatrix}\begin{pmatrix}M_T\\ N_T\end{pmatrix}, \tag{5.1}$$

where M_T and N_T are martingales respectively given by

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

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 $\left\langle \mathrm{d}B_t, \mathrm{d}\widetilde{B}_t \right\rangle = \rho \mathrm{d}t$, we easily obtain that the increasing process of $\mathcal{M}_T$ is given by

$$\langle \mathcal{M} \rangle_T = \begin{pmatrix} \langle M \rangle_T & \rho \langle M \rangle_T \\ \rho \langle M \rangle_T & \langle M \rangle_T \end{pmatrix}. \tag{5.2}$$

Proof of Theorem 2.1. First of all, we have

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

Thus, as the process is ergodic, we obtain for T going to infinity,

$$\frac{1}{T^2} \det \Gamma_T \xrightarrow{a.s.} \mathbb{E}\left[1/C\right] \mathbb{E}\left[X^2/C\right] - \left(\mathbb{E}\left[X/C\right]\right)^2 \tag{5.3}$$

and

$$T\Gamma_T^{-1} \xrightarrow{a.s.} A \tag{5.4}$$

where A is given by

$$A = \left(\mathbb{E}\left[C\right] \mathbb{E}\left[1/C\right] - 1\right)^{-1} \begin{pmatrix} \mathbb{E}\left[X^2/C\right] & -\mathbb{E}\left[X/C\right] \\ -\mathbb{E}\left[X/C\right] & \mathbb{E}\left[1/C\right] \end{pmatrix}. \tag{5.5}$$

A straightforward application of Lemmas 4.1 (i) to (iii) gives that

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

Besides, the martingale M_T satisfies for T going to infinity

$$\frac{M_T}{T} \xrightarrow{a.s.} 0. \tag{5.6}$$

As a matter of fact, by convergences (3.4) and (3.7), we know that a.s. $\langle n \rangle_T = \mathcal{O}\left(T\right)$ and $\langle m \rangle_T = \mathcal{O}\left(T\right)$. It ensures that for T going to infinity,

$$\frac{n_T}{T} \xrightarrow{a.s.} 0 \qquad \text{and} \qquad \frac{m_T}{T} \xrightarrow{a.s.} 0.$$

As N_T and M_T share the same increasing process, this result remains true by replacing M_T by N_T. Finally, the almost sure convergence (2.1) follows from (5.1), (5.4) and (5.6). □

6. Proof of the asymptotic normality

Proof of Theorem 2.2. First of all, we deduce from (5.1) that

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

We already saw that $T\Gamma_T^{-1}$ converges almost surely as T goes to infinity and its limit A is given by (5.5). We now have to establish the asymptotic normality of $\frac{M_T}{\sqrt{T}}$. The increasing process of M_T is given by

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

where $\langle m \rangle_T$ and $\langle n \rangle_T$ are respectively given by (3.7) and (3.4). Thus, by the ergodicity of the process, we obtain that

$$\frac{\langle M \rangle_T}{T} \xrightarrow{a.s.} 4L \qquad \text{where} \qquad L = \begin{pmatrix} \mathbb{E}\left[X/C^2\right] & \mathbb{E}\left[X^2/C^2\right] \\ \mathbb{E}\left[X^2/C^2\right] & \mathbb{E}\left[X^3/C^2\right] \end{pmatrix}.$$

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

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

We easily obtain the following a.s. convergence

$$\frac{\langle \mathcal{M} \rangle_T}{T} \xrightarrow{a.s.} 4\mathcal{L}$$

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

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

and we deduce from the CLT for martingales that

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

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 question, see for example [3] and [2]. We choose to implement the QE-algorithm based on quadratic-exponential approximations proposed in [3]. Andersen introduced this algorithm to deal with the case $a < 2$, for which common discretization schemes are not accurate.

7.1. **Asymptotic behavior for $c = 1$.** The two following figures illustrate our main results (strong consistency and asymptotic normality) in the case $a = 1$ and $b = -2$, with the weighting parameter $c = 1$. The red curves in the second figure displays the standard normal distribution.

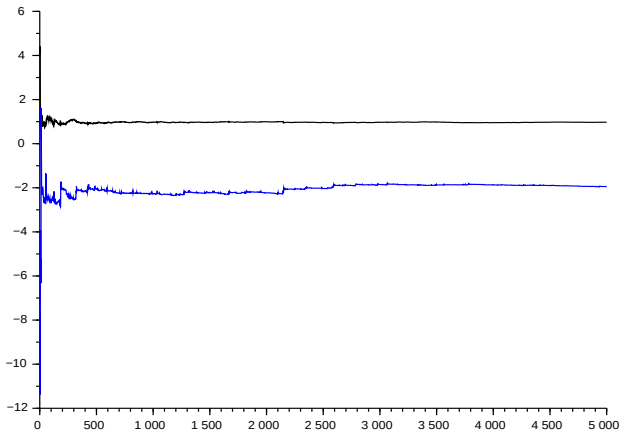


Figure 1. strong consistency: $(\widehat{a}_T)$ in black and $(\widehat{b}_T)$ in blue.

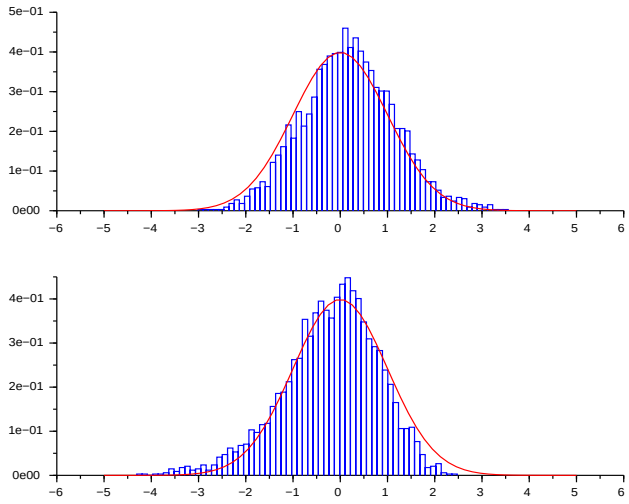


FIGURE 2. Histograms of 3000 outcomes of $\sqrt{T/4\sigma_{11}}\,(\widehat{a}_T - a)$ and $\sqrt{T/4\sigma_{22}}(\widehat{b}_T - b)$ at time $T = 70$.

7.2. **Choice of the constant** c**.** We have chosen to introduce a constant c in our weighting, instead of only considering the case $c = 1$ (as done in the discrete-time case in [26]) with the aim of lowering the variance of the estimators. However, this raises the question of 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. It appears that one should choose a small value of c. The value should not be to small to avoid the growth illustrated by the second figure, 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 (factor 5) between the empirical variances obtained with $c = 0.01$ and $c = 1$. However, for $\widehat{b}_T$ both empirical variances do not significantly differ.

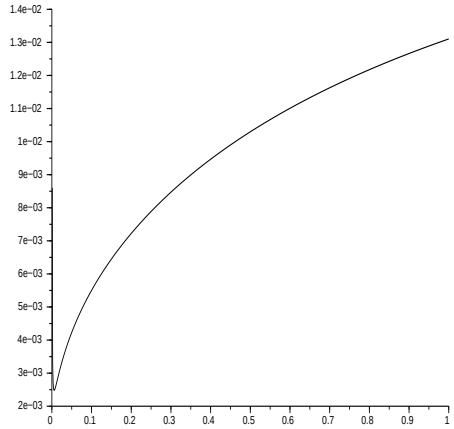


FIGURE 3. Variance of $\widehat{a}_{50}$

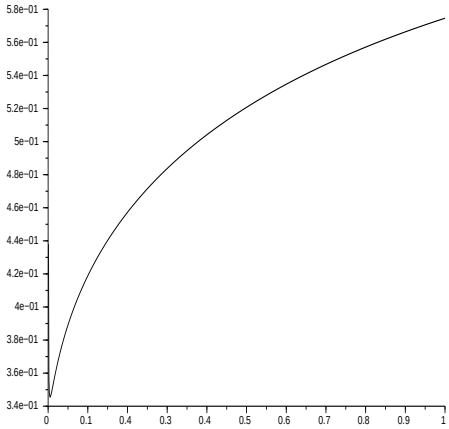


FIGURE 4. Variance of $\widehat{b}_{50}$

Empirical variances of the estimators

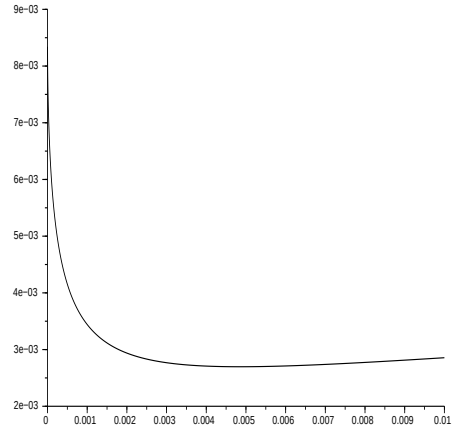


FIGURE 5. Variance of $\widehat{a}_{50}$

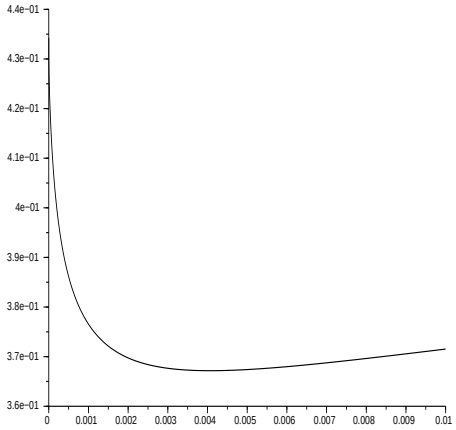


FIGURE 6. Variance of $\widehat{b}_{50}$

Empirical variances for very small values of c

REFERENCES

[1] AÏT-SAHALIA, Y., AND KIMMEL, R. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*, 2 (feb 2007), 413–452.

[2] ALFONSI, A. High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Mathematics of Computation 79*, 269 (2010), 209–237.

[3] ANDERSON, L. Simple and efficient simulation of the heston stochastic volatility model. *Journal of Computational Finance 11* (2008).

[4] AZENCOTT, R., AND GADHYAN, Y. Accurate parameter estimation for coupled stochastic dynamics. *Discrete Contin. Dyn. Syst.*, Dynamical systems, differential equations and applications. 7th AIMS Conference, suppl. (2009), 44–53.

[5] BARCZY, M., AND PAP, G. Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations. *Statistics 50*, 2 (2016), 389–417.

[6] BEN ALAYA, M., AND KEBAIER, A. Asymptotic Behavior of The Maximum Likelihood Estimator For Ergodic and Nonergodic Square-Root Diffusions. *Stochastic Analysis and Applications* (2013).

[7] COX, J. C., INGERSOLL, JR., J. E., AND ROSS, S. A. A theory of the term structure of interest rates. *Econometrica 53*, 2 (1985), 385–407.

[8] DU ROY DE CHAUMARAY, M. Large deviations for the squared radial Ornstein-Uhlenbeck process. *Theory of Probability and its Applications* (2015), To appear.

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

[10] FORDE, M., AND JACQUIER, A. The large-maturity smile for the Heston model. *Finance Stoch. 15*, 4 (2011), 755–780.

[11] FORDE, M., JACQUIER, A., AND LEE, R. The small-time smile and term structure of implied volatility under the Heston model. *SIAM J. Financial Math. 3*, 1 (2012), 690–708.

[12] FOURNIÉ, E., AND TALAY, D. Application de la statistique des diffusions à un modèle de taux d'intérêt. *Finance 12* (1991).

[13] GAO, F., AND JIANG, H. Moderate deviations for squared Ornstein-Uhlenbeck process. *Statistics & Probability Letters 79*, 11 (2009), 1378–1386.

[14] GATHERAL, J. *The Volatility Surface: A Practitioner's Guide*. Wiley, 2006.

[15] GRADSHTEYN, I. S., AND RYZHIK, I. M. *Table of integrals, series, and products*. Academic Press, 1980.

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

[17] JACQUIER, A., AND ROOME, P. Large-maturity regimes of the Heston forward smile. *Stochastic Process. Appl. 126*, 4 (2016), 1087–1123.

[18] Janek, A., Kluge, T., Weron, R., and Wystup, U. FX smile in the Heston model. In *Statistical tools for finance and insurance*. Springer, Heidelberg, 2011, pp. 133–162.
[19] Lamberton, D., and Lapeyre, B. *Introduction au calcul stochastique appliqué à la finance*, second ed. Ellipses Édition Marketing, Paris, 1997.
[20] Lee, R. Option pricing by transform methods: Extensions, unification, and error control. *Journal of Computational Finance 7*, 51–86.
[21] Lewis, A. L. *Option valuation under stochastic volatility*. Finance Press, Newport Beach, CA, 2000.
[22] Luke, Y. L. *The Special Functions and their Approximations. Vol. 2*. Academic Press, 1969.
[23] Overbeck, L. Estimation for continuous branching processes. *Scandinavian Journal of Statistics. Theory and Applications 25* (1998).
[24] Stein, E., and Stein, J. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies 4* (1991), 727–752.
[25] Stein, J. Overreactions in the options market. *Journal of finance 44* (1989), 1011–1024.
[26] Wei, C., and Winnicki, J. Estimation of the means in the branching process with immigration. *Annals of Statistics 18* (1990).
[27] Zani, M. Large deviations for squared radial OrnsteinUhlenbeck processes. *Stochastic Processes and their Applications 102*, 1 (2002), 25 – 42.

Université de Bordeaux, Institut de Mathématiques de Bordeaux, UMR 5251, 351 Cours de la Libération, 33405 Talence cedex, France.