Journal of Statistical Theory and Practice (2019) 13:18
https://doi.org/10.1007/s42519-018-0007-6

ORIGINAL ARTICLE

Least-Squares Estimation for the Subcritical Heston Model Based on Continuous-Time Observations

Mátyás Barczy[1] · Balázs Nyul[2] · Gyula Pap[3]

Published online: 5 November 2018

Abstract

We prove strong consistency and asymptotic normality of least-squares estimators for the subcritical Heston model based on continuous-time observations. We also present some numerical illustrations of our results.

Keywords Heston model · Least-squares estimator · Strong consistency · Asymptotic normality

Mathematics Subject Classification 60H10 · 91G70 · 60F05 · 62F12

1 Introduction

Stochastic processes given by solutions to stochastic differential equations (SDEs) have been frequently applied in financial mathematics. So the theory and practice of stochastic analysis and statistical inference for such processes are important topics. In this note, we consider such a model, namely the Heston model

Mátyás Barczy is supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

✉ Mátyás Barczy
barczy@math.u-szeged.hu

Balázs Nyul
nyul.balazs@inf.unideb.hu

Gyula Pap
papgy@math.u-szeged.hu

[1] MTA-SZTE Analysis and Stochastics Research Group, Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, Szeged 6720, Hungary

[2] Faculty of Informatics, University of Debrecen, Pf. 12, Debrecen 4010, Hungary

[3] Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, Szeged 6720, Hungary

$$\begin{cases} \mathrm{d}Y_t = (a - bY_t)\,\mathrm{d}t + \sigma_1\sqrt{Y_t}\,\mathrm{d}W_t, \\ \mathrm{d}X_t = (\alpha - \beta Y_t)\,\mathrm{d}t + \sigma_2\sqrt{Y_t}\big(\varrho\,\mathrm{d}W_t + \sqrt{1-\varrho^2}\,\mathrm{d}B_t\big), \end{cases} \qquad t \geqslant 0, \tag{1.1}$$

where $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, $\varrho \in (-1, 1)$, and $(W_t, B_t)_{t\geqslant 0}$ is a 2-dimensional standard Wiener process, see Heston [14]. For interpretation of Y and X in financial mathematics, see, e.g., Hurn et al. [20, Section 4], here we only note that X_t is the logarithm of the asset price at time t and Y_t its volatility for each $t \geqslant 0$. The first coordinate process Y is called a Cox–Ingersoll–Ross (CIR) process (see Cox et al. [9]), square-root process or Feller process.

Parameter estimation for the Heston model (1.1) has a long history, for a short survey of the most recent results, see, e.g., the introduction of Barczy and Pap [5]. The importance of the joint estimation of (a, b, α, β) and not only of (a, b) stems from the fact that X_t is the logarithm of the asset price at time t having high importance in finance. In fact, in Barczy and Pap [5], we investigated asymptotic properties of maximum likelihood estimator of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$. In Barczy et al. [6], we studied asymptotic behavior of conditional least-squares estimator of (a, b, α, β) based on discrete-time observations (Y_i, X_i), $i = 1, \ldots, n$, starting the process from some known non-random initial value $(y_0, x_0) \in (0, \infty) \times \mathbb{R}$. In this note, we study least-squares estimator (LSE) of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$, starting the process (Y, X) from some known initial value (Y_0, X_0) satisfying $\mathbb{P}(Y_0 \in (0, \infty)) = 1$. The investigation of the LSE of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$, is motivated by the fact that the LSEs of (a, b, α, β) based on appropriate discrete-time observations converge in probability to the LSE of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$, see Proposition 3.1. We do not suppose that the process $(Y_t)_{t\in[0,T]}$ is observed, since it can be determined using the observations $(X_t)_{t\in[0,T]}$ and the initial value Y_0, which follows by a slight modification of Remark 2.5 in Barczy and Pap [5] (replacing y_0 by Y_0). We do not estimate the parameters σ_1, σ_2 and ϱ, since these parameters could—in principle, at least—be determined (rather than estimated) using the observations $(X_t)_{t\in[0,T]}$ and the initial value Y_0, see Barczy and Pap [5, Remark 2.6]. We investigate only the so-called subcritical case, i.e., when $b > 0$, see Definition 2.3.

In Sect. 2, we recall some properties of the Heston model (1.1) such as the existence and uniqueness of a strong solution of the SDE (1.1), the form of conditional expectation of (Y_t, X_t), $t \geqslant 0$, given the past of the process up to time s with $s \in [0, t]$, a classification of the Heston model and the existence of a unique stationary distribution and ergodicity for the first coordinate process of the SDE (1.1). Section 3 is devoted to derive a LSE of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$, see Proposition 3.1. We note that Overbeck and Rydén [27, Theorems 3.5 and 3.6] have already proved the strong consistency and asymptotic normality of the LSE of (a, b) based on continuous-time observations $(Y_t)_{t\in[0,T]}$, $T > 0$, in case of a subcritical CIR process Y with an initial value having distribution as the unique stationary distribution of the model. Overbeck and Rydén [27, page 433] also noted that (without providing a proof) their results are valid for an arbitrary initial distribution using some coupling argument. In Sect. 4, we prove strong consistency and asymptotic normality of the LSE of (a, b, α, β) introduced in Sect. 3, so our results for the Heston model (1.1) in Sect. 3 can be considered as generalizations of the corresponding ones in Overbeck and Rydén [27, Theorems 3.5 and 3.6] with the advantage that our proof is presented for an arbitrary initial value (Y_0, X_0) satisfying $\mathbb{P}(Y_0 \in (0, \infty)) = 1$, without using any coupling argument. The covariance matrix of the limit normal distribution in question depends on the unknown parameters a and b as well, but somewhat surprisingly not on α and β. We point out that our proof of technique for deriving the asymptotic normality of the LSE in question is completely different from that of Overbeck and Rydén [27]. We use a limit theorem for continuous martingales (see, Theorem 2.6), while Overbeck and Rydén [27] use a limit theorem for ergodic processes due to Jacod and Shiryaev [21, Theorem VIII.3.79] and the so-called Delta method (see, e.g., Theorem 11.2.14 in Lehmann and Romano [24]). We also remark that the approximation in probability of the LSE of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T > 0$, given in Proposition 3.1 is not at all used for proving the asymptotic behavior of the LSE in question as $T \to \infty$ in Theorems 4.1 and 4.2. Further, we mention that the covariance matrix of the limit normal distribution in Theorem 3.6 in Overbeck and Rydén [27] is somewhat complicated, while, as a special case of our Theorem 4.2, it turns out that it can be written in a much simpler form by making a simple reparametrization of the SDE (1) in Overbeck and Rydén [27], estimating $-b$ instead of b (with the notations of Overbeck and Rydén [27]), i.e., considering the SDE (1.1) and estimating b (with our notations), see Corollary 4.3. Section 5 is devoted to present some numerical illustrations of our results in Sect. 4.

2 Preliminaries

Let $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$, $\mathbb{R}_-$ and $\mathbb{R}_{--}$ denote the sets of positive integers, non-negative integers, real numbers, non-negative real numbers, positive real numbers, non-positive real numbers and negative real numbers, respectively. For $x, y \in \mathbb{R}$, we will use the notation $x \wedge y := \min(x, y)$. By $\|x\|$ and $\|A\|$, we denote the Euclidean norm of a vector $x \in \mathbb{R}^d$ and the induced matrix norm of a matrix $A \in \mathbb{R}^{d\times d}$, respectively. By $\boldsymbol{I}_d \in \mathbb{R}^{d\times d}$, we denote the d-dimensional unit matrix.

Let $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ be a probability space equipped with the augmented filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ corresponding to $(W_t, B_t)_{t\in\mathbb{R}_+}$ and a given initial value (η_0, ζ_0) being independent of $(W_t, B_t)_{t\in\mathbb{R}_+}$ such that $\mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1$, constructed as in Karatzas and Shreve [22, Section 5.2]. Note that $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ satisfies the usual conditions, i.e., the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets in $\mathcal{F}$.

By $C^2_c(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$ and $C^\infty_c(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$, we denote the set of twice continuously differentiable real-valued functions on $\mathbb{R}_+ \times \mathbb{R}$ with compact support, and the set of infinitely differentiable real-valued functions on $\mathbb{R}_+ \times \mathbb{R}$ with compact support, respectively.

The next proposition is about the existence and uniqueness of a strong solution of the SDE (1.1), see, e.g., Barczy and Pap [5, Proposition 2.1].

Proposition 2.1 *Let* (η_0, ζ_0) *be a random vector independent of* $(W_t, B_t)_{t\in\mathbb{R}_+}$ *satisfying* $\mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1$. *Then for all* $a \in \mathbb{R}_{++}$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{++}$, *and* $\varrho \in (-1, 1)$, *there is a pathwise unique strong solution* $(Y_t, X_t)_{t\in\mathbb{R}_+}$ *of the SDE* (1.1) such that $\mathbb{P}((Y_0, X_0) = (\eta_0, \zeta_0)) = 1$ *and* $\mathbb{P}(Y_t \in \mathbb{R}_+ \text{ for all } t \in \mathbb{R}_+) = 1$. *Further, for all* $s, t \in \mathbb{R}_+$ *with* $s \leqslant t$,

$$\begin{cases} Y_t = \mathrm{e}^{-b(t-s)} Y_s + a\int_s^t \mathrm{e}^{-b(t-u)}\,\mathrm{d}u + \sigma_1\int_s^t \mathrm{e}^{-b(t-u)}\sqrt{Y_u}\,\mathrm{d}W_u, \\ X_t = X_s + \int_s^t (\alpha - \beta Y_u)\,\mathrm{d}u + \sigma_2\int_s^t \sqrt{Y_u}\,\mathrm{d}(\varrho W_u + \sqrt{1-\varrho^2} B_u). \end{cases} \tag{2.1}$$

Next we present a result about the first moment and the conditional moment of $(Y_t, X_t)_{t\in\mathbb{R}_+}$, see Barczy et al. [6, Proposition 2.2].

Proposition 2.2 *Let* $(Y_t, X_t)_{t\in\mathbb{R}_+}$ *be the unique strong solution of the SDE* (1.1) *satisfying* $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$ *and* $\mathbb{E}(Y_0) < \infty$, $\mathbb{E}(|X_0|) < \infty$. *Then for all* $s, t \in \mathbb{R}_+$ *with* $s \leqslant t$, *we have*

$$\mathbb{E}(Y_t \mid \mathcal{F}_s) = \mathrm{e}^{-b(t-s)} Y_s + a\int_s^t \mathrm{e}^{-b(t-u)}\,\mathrm{d}u, \tag{2.2}$$

$$\begin{aligned} \mathbb{E}(X_t \mid \mathcal{F}_s) &= X_s + \int_s^t (\alpha - \beta\mathbb{E}(Y_u \mid \mathcal{F}_s))\,\mathrm{d}u \\ &= X_s + \alpha(t-s) - \beta Y_s\int_s^t \mathrm{e}^{-b(u-s)}\,\mathrm{d}u - a\beta\int_s^t\left(\int_s^u \mathrm{e}^{-b(u-v)}\,\mathrm{d}v\right)\mathrm{d}u, \end{aligned} \tag{2.3}$$

and hence

$$\begin{bmatrix} \mathbb{E}(Y_t) \\ \mathbb{E}(X_t) \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-bt} & 0 \\ -\beta\int_0^t \mathrm{e}^{-bu}\,\mathrm{d}u & 1 \end{bmatrix}\begin{bmatrix} \mathbb{E}(Y_0) \\ \mathbb{E}(X_0) \end{bmatrix} + \begin{bmatrix} \int_0^t \mathrm{e}^{-bu}\,\mathrm{d}u & 0 \\ -\beta\int_0^t\left(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v\right)\mathrm{d}u & t \end{bmatrix}\begin{bmatrix} a \\ \alpha \end{bmatrix}.$$

Consequently, if $b \in \mathbb{R}_{++}$, *then*

$$\lim_{t\to\infty}\mathbb{E}(Y_t) = \frac{a}{b}, \qquad \lim_{t\to\infty} t^{-1}\mathbb{E}(X_t) = \alpha - \frac{\beta a}{b},$$

if $b = 0$, *then*

$$\lim_{t\to\infty} t^{-1}\mathbb{E}(Y_t) = a, \qquad \lim_{t\to\infty} t^{-2}\mathbb{E}(X_t) = -\frac{1}{2}\beta a,$$

if $b \in \mathbb{R}_{--}$, *then*

$$\lim_{t\to\infty} \mathrm{e}^{bt}\mathbb{E}(Y_t) = \mathbb{E}(Y_0) - \frac{a}{b}, \qquad \lim_{t\to\infty} \mathrm{e}^{bt}\mathbb{E}(X_t) = \frac{\beta}{b}\mathbb{E}(Y_0) - \frac{\beta a}{b^2}.$$

Based on the asymptotic behavior of the expectations $(\mathbb{E}(Y_t), \mathbb{E}(X_t))$ as $t \to \infty$, we recall a classification of the Heston process given by the SDE (1.1), see, Barczy and Pap [5, Definition 2.3].

Definition 2.3 Let $(Y_t, X_t)_{t\in\mathbb{R}_+}$ be the unique strong solution of the SDE (1.1) satisfying $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. We call $(Y_t, X_t)_{t\in\mathbb{R}_+}$ subcritical, critical or supercritical if $b \in \mathbb{R}_{++}$, $b = 0$ or $b \in \mathbb{R}_{--}$, respectively.

In the sequel $\xrightarrow{\mathbb{P}}$, $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\text{a.s.}}$ will denote convergence in probability, in distribution and almost surely, respectively.

The following result states the existence of a unique stationary distribution and the ergodicity for the process $(Y_t)_{t\in\mathbb{R}_+}$ given by the first equation in (1.1) in the subcritical case, see, e.g., Cox et al. [9, Equation (20)], Li and Ma [25, Theorem 2.6] or Theorem 3.1 with $\alpha = 2$ and Theorem 4.1 in Barczy et al. [4].

Theorem 2.4 *Let* $a, b, \sigma_1 \in \mathbb{R}_{++}$. *Let* $(Y_t)_{t\in\mathbb{R}_+}$ *be the unique strong solution of the first equation of the SDE* (1.1) satisfying $\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1$. *Then*

(1) $Y_t \xrightarrow{\mathcal{L}} Y_\infty$ *as* $t \to \infty$, *and the distribution of* Y_∞ *is given by*

$$\mathbb{E}(\mathrm{e}^{-\lambda Y_\infty}) = \left(1 + \frac{\sigma_1^2}{2b}\lambda\right)^{-2a/\sigma_1^2}, \qquad \lambda \in \mathbb{R}_+, \tag{2.4}$$

i.e., Y_∞ *has Gamma distribution with parameters* $2a/\sigma_1^2$ *and* $2b/\sigma_1^2$, *hence*

$$\mathbb{E}(Y_\infty) = \frac{a}{b}, \qquad \mathbb{E}(Y_\infty^2) = \frac{(2a+\sigma_1^2)a}{2b^2}, \qquad \mathbb{E}(Y_\infty^3) = \frac{(2a+\sigma_1^2)(a+\sigma_1^2)a}{2b^3}.$$

(2) *supposing that the random initial value* Y_0 *has the same distribution as* Y_∞, *the process* $(Y_t)_{t\in\mathbb{R}_+}$ *is strictly stationary.*

(3) *for all Borel measurable functions* $f : \mathbb{R} \to \mathbb{R}$ *such that* $\mathbb{E}(|f(Y_\infty)|) < \infty$, *we have*

$$\frac{1}{T}\int_0^T f(Y_s)\,\mathrm{d}s \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } T \to \infty. \tag{2.5}$$

In what follows we recall some limit theorems for continuous (local) martingales. We will use these limit theorems later on for studying the asymptotic behaviour of least-squares estimators of (a, b, α, β). First we recall a strong law of large numbers for continuous local martingales.

Theorem 2.5 (Liptser and Shiryaev [26, Lemma 17.4]) *Let $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P}\big)$ be a filtered probability space satisfying the usual conditions. Let $(M_t)_{t\in\mathbb{R}_+}$ be a square-integrable continuous local martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ such that $\mathbb{P}(M_0 = 0) = 1$. Let $(\xi_t)_{t\in\mathbb{R}_+}$ be a progressively measurable process such that $\mathbb{P}\big(\int_0^t \xi_u^2 \, \mathrm{d}\langle M\rangle_u < \infty\big) = 1$, $t \in \mathbb{R}_+$, and*

$$\int_0^t \xi_u^2 \, \mathrm{d}\langle M\rangle_u \xrightarrow{\text{a.s.}} \infty \qquad \text{as } t \to \infty, \tag{2.6}$$

where $(\langle M\rangle_t)_{t\in\mathbb{R}_+}$ denotes the quadratic variation process of M. Then

$$\frac{\int_0^t \xi_u \, \mathrm{d}M_u}{\int_0^t \xi_u^2 \, \mathrm{d}\langle M\rangle_u} \xrightarrow{\text{a.s.}} 0 \qquad \text{as } t \to \infty. \tag{2.7}$$

If $(M_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process, the progressive measurability of $(\xi_t)_{t\in\mathbb{R}_+}$ can be relaxed to measurability and adaptedness to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

The next theorem is about the asymptotic behaviour of continuous multivariate local martingales, see van Zanten [28, Theorem 4.1].

Theorem 2.6 (van Zanten [28, Theorem 4.1]) *Let $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P}\big)$ be a filtered probability space satisfying the usual conditions. Let $(\boldsymbol{M}_t)_{t\in\mathbb{R}_+}$ be a d-dimensional square-integrable continuous local martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ such that $\mathbb{P}(\boldsymbol{M}_0 = \boldsymbol{0}) = 1$. Suppose that there exists a function $\boldsymbol{Q} : \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ such that $\boldsymbol{Q}(t)$ is an invertible (non-random) matrix for all $t \in \mathbb{R}_+$, $\lim_{t\to\infty} \|\boldsymbol{Q}(t)\| = 0$ and*

$$\boldsymbol{Q}(t)\langle \boldsymbol{M}\rangle_t \, \boldsymbol{Q}(t)^\top \xrightarrow{\mathbb{P}} \boldsymbol{\eta}\boldsymbol{\eta}^\top \quad \text{as } t \to \infty,$$

where $\boldsymbol{\eta}$ is a $d \times d$ random matrix. Then, for each $\mathbb{R}^k$-valued random vector $\boldsymbol{v}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, we have

$$(\boldsymbol{Q}(t)\boldsymbol{M}_t, \boldsymbol{v}) \xrightarrow{\mathcal{L}} (\boldsymbol{\eta}\boldsymbol{Z}, \boldsymbol{v}) \qquad \text{as } t \to \infty,$$

where $\boldsymbol{Z}$ is a d-dimensional standard normally distributed random vector independent of $(\boldsymbol{\eta}, \boldsymbol{v})$.

We note that Theorem 2.6 remains true if the function $\boldsymbol{Q}$ is defined only on an interval $[t_0, \infty)$ with some $t_0 \in \mathbb{R}_{++}$.

3 Existence of LSE Based on Continuous-Time Observations

First, we define the LSE of (a, b, α, β) based on discrete-time observations $(Y_{\frac{i}{n}}, X_{\frac{i}{n}})_{i\in\{0,1,\ldots,\lfloor nT\rfloor\}}$, $n \in \mathbb{N}$, $T \in \mathbb{R}_{++}$ [see (3.1)] by pointing out that the sum appearing in this definition of LSE can be considered as an approximation of the corresponding sum of the conditional LSE of (a, b, α, β) based on discrete-time observations $(Y_{\frac{i}{n}}, X_{\frac{i}{n}})_{i\in\{0,1,\ldots,\lfloor nT\rfloor\}}$, $n \in \mathbb{N}$, $T \in \mathbb{R}_{++}$ (which was investigated in Barczy et al. [6]). Then, we introduce the LSE of (a, b, α, β) based on continuous-time observations $(X_t)_{t\in[0,T]}$, $T \in \mathbb{R}_{++}$ [see (3.4) and (3.5)] as the limit in probability of the LSE of (a, b, α, β) based on discrete-time observations $(Y_{\frac{i}{n}}, X_{\frac{i}{n}})_{i\in\{0,1,\ldots,\lfloor nT\rfloor\}}$, $n \in \mathbb{N}$, $T \in \mathbb{R}_{++}$ (see Proposition 3.1).

A LSE of (a, b, α, β) based on discrete-time observations $(Y_{\frac{i}{n}}, X_{\frac{i}{n}})_{i\in\{0,1,\ldots,\lfloor nT\rfloor\}}$, $n \in \mathbb{N}$, $T \in \mathbb{R}_{++}$, can be obtained by solving the extremum problem

$$\begin{aligned}&\left(\widehat{a}_{T,n}^{\mathrm{LSE,D}}, \widehat{b}_{T,n}^{\mathrm{LSE,D}}, \widehat{\alpha}_{T,n}^{\mathrm{LSE,D}}, \widehat{\beta}_{T,n}^{\mathrm{LSE,D}}\right)\\ &:= \mathop{\arg\min}_{(a,b,\alpha,\beta)\in\mathbb{R}^4} \sum_{i=1}^{\lfloor nT\rfloor}\left[\left(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} - \frac{1}{n}\left(a - bY_{\frac{i-1}{n}}\right)\right)^2\right.\\ &\qquad\left.+\left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} - \frac{1}{n}\left(\alpha - \beta Y_{\frac{i-1}{n}}\right)\right)^2\right].\end{aligned} \tag{3.1}$$

Here in the notations the letter D refers to discrete-time observations. This definition of LSE can be considered as the corresponding one given in Hu and Long [17, formula (1.2)] for generalized Ornstein-Uhlenbeck processes driven by α-stable motions, see also Hu and Long [18, formula (3.1)]. For a heuristic motivation of the LSE (3.1) based on the discrete observations, see, e.g., Hu and Long [16, page 178] (formulated for Langevin equations), and for a mathematical one, see as follows. By (2.2), for all $i \in \mathbb{N}$,

$$\begin{aligned}Y_{\frac{i}{n}} - \mathbb{E}\left(Y_{\frac{i}{n}} \mid \mathcal{F}_{\frac{i-1}{n}}\right) &= Y_{\frac{i}{n}} - \mathrm{e}^{-\frac{b}{n}} Y_{\frac{i-1}{n}} - a\int_{\frac{i-1}{n}}^{\frac{i}{n}} \mathrm{e}^{-b(\frac{i}{n}-u)}\,\mathrm{d}u\\ &= Y_{\frac{i}{n}} - \mathrm{e}^{-\frac{b}{n}} Y_{\frac{i-1}{n}} - a\int_0^{\frac{1}{n}} \mathrm{e}^{-bv}\,\mathrm{d}v\\ &= \begin{cases} Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} - \frac{a}{n} & \text{if } b = 0,\\ Y_{\frac{i}{n}} - \mathrm{e}^{-\frac{b}{n}} Y_{\frac{i-1}{n}} + \frac{a}{b}(\mathrm{e}^{-\frac{b}{n}} - 1) & \text{if } b \neq 0.\end{cases}\end{aligned}$$

Using first-order Taylor approximation of $\mathrm{e}^{-\frac{b}{n}}$ at $b=0$ by $1-\frac{b}{n}$, and that of $\frac{a}{b}(\mathrm{e}^{-\frac{b}{n}}-1)$ at $(a,b)=(0,0)$ by $-\frac{a}{n}$, the random variable $Y_{\frac{i}{n}}-Y_{\frac{i-1}{n}}-\frac{1}{n}(a-bY_{\frac{i-1}{n}})$ in the definition (3.1) of the LSE of (a,b,α,β) can be considered as a first-order Taylor approximation of

$$Y_{\frac{i}{n}}-\mathbb{E}\Big(Y_{\frac{i}{n}}\mid Y_0,X_0,Y_{\frac{1}{n}},X_{\frac{1}{n}},\dots,Y_{\frac{i-1}{n}},X_{\frac{i-1}{n}}\Big)=Y_{\frac{i}{n}}-\mathbb{E}\Big(Y_{\frac{i}{n}}\mid\mathcal{F}_{\frac{i-1}{n}}\Big),$$

which appears in the definition of the conditional LSE of (a,b,α,β) based on discrete-time observations $(Y_{\frac{i}{n}},X_{\frac{i}{n}})_{i\in\{0,1,\dots,\lfloor nT\rfloor\}}$, $n\in\mathbb{N}$, $T\in\mathbb{R}_{++}$. Similarly, by (2.3), for all $i\in\mathbb{N}$,

$$\begin{aligned}
X_{\frac{i}{n}}-\mathbb{E}(X_{\frac{i}{n}}\mid\mathcal{F}_{\frac{i-1}{n}})&=X_{\frac{i}{n}}-X_{\frac{i-1}{n}}-\frac{\alpha}{n}\\
&\quad+\beta Y_{\frac{i-1}{n}}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\mathrm{e}^{-b(u-\frac{i-1}{n})}\,\mathrm{d}u+a\beta\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(\int_{\frac{i-1}{n}}^{u}\mathrm{e}^{-b(u-v)}\,\mathrm{d}v\right)\mathrm{d}u\\
&=X_{\frac{i}{n}}-X_{\frac{i-1}{n}}-\frac{\alpha}{n}+\beta Y_{\frac{i-1}{n}}\int_0^{\frac{1}{n}}\mathrm{e}^{-bu}\,\mathrm{d}u+a\beta\int_0^{\frac{1}{n}}\left(\int_0^{u}\mathrm{e}^{-bv}\,\mathrm{d}v\right)\mathrm{d}u\\
&=\begin{cases}X_{\frac{i}{n}}-X_{\frac{i-1}{n}}-\frac{\alpha}{n}+\frac{\beta}{n}Y_{\frac{i-1}{n}}+\frac{a\beta}{2n^2} & \text{if } b=0,\\ X_{\frac{i}{n}}-X_{\frac{i-1}{n}}-\frac{\alpha}{n}+\frac{\beta}{b}(1-\mathrm{e}^{-\frac{b}{n}})Y_{\frac{i-1}{n}}+\frac{a\beta}{b}\Big(\frac{1}{n}-\frac{1-\mathrm{e}^{-\frac{b}{n}}}{b}\Big) & \text{if } b\neq 0.\end{cases}
\end{aligned}$$

Using first-order Taylor approximation of $\frac{a\beta}{2n^2}$ at $(a,\beta)=(0,0)$ by 0, that of $\frac{\beta}{b}(1-\mathrm{e}^{-\frac{b}{n}})$ at $(b,\beta)=(0,0)$ by $\frac{\beta}{n}$, and that of $\frac{a\beta}{b}\Big(\frac{1}{n}-\frac{1-\mathrm{e}^{-\frac{b}{n}}}{b}\Big)=\frac{a\beta}{n^2}\sum_{k=0}^{\infty}(-1)^k\frac{(b/n)^k}{(k+2)!}$ at $(a,b,\beta)=(0,0,0)$ by 0, the random variable $X_{\frac{i}{n}}-X_{\frac{i-1}{n}}-\frac{1}{n}(\alpha-\beta Y_{\frac{i-1}{n}})$ in the definition (3.1) of the LSE of (a,b,α,β) can be considered as a first-order Taylor approximation of

$$X_{\frac{i}{n}}-\mathbb{E}(X_{\frac{i}{n}}\mid Y_0,X_0,Y_{\frac{1}{n}},X_{\frac{1}{n}},\dots,Y_{\frac{i-1}{n}},X_{\frac{i-1}{n}})=X_{\frac{i}{n}}-\mathbb{E}(X_{\frac{i}{n}}\mid\mathcal{F}_{\frac{i-1}{n}}),$$

which appears in the definition of the conditional LSE of (a,b,α,β) based on discrete-time observations $(Y_{\frac{i}{n}},X_{\frac{i}{n}})_{i\in\{0,1,\dots,\lfloor nT\rfloor\}}$, $n\in\mathbb{N}$, $T\in\mathbb{R}_{++}$.

We note that in Barczy et al. [6] we proved strong consistency and asymptotic normality of conditional LSE of (a,b,α,β) based on discrete-time observations $(Y_i,X_i)_{i\in\{1,\dots,n\}}$, $n\in\mathbb{N}$, starting the process from some known non-random initial value $(y_0,x_0)\in\mathbb{R}_{++}\times\mathbb{R}$, as the sample size n tends to infinity in the subcritical case.

Solving the extremum problem (3.1), we have

$$\big(\widehat{a}_{T,n}^{\mathrm{LSE,D}}, \widehat{b}_{T,n}^{\mathrm{LSE,D}}\big) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{\lfloor nT\rfloor} \Big(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} - \frac{1}{n}\Big(a - bY_{\frac{i-1}{n}}\Big)\Big)^2,$$

$$\big(\widehat{\alpha}_{T,n}^{\mathrm{LSE,D}}, \widehat{\beta}_{T,n}^{\mathrm{LSE,D}}\big) = \underset{(\alpha,\beta)\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{\lfloor nT\rfloor} \Big(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} - \frac{1}{n}\Big(\alpha - \beta Y_{\frac{i-1}{n}}\Big)\Big)^2,$$

hence, similarly as on page 675 in Barczy et al. [3], we get

$$\begin{bmatrix} \hat{a}_{T,n}^{LSE,D} \\ \hat{b}_{T,n}^{LSE,D} \end{bmatrix} = n \begin{bmatrix} \lfloor nT\rfloor & -\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}} \\ -\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}} & \sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}^2 \end{bmatrix}^{-1} \begin{bmatrix} Y_{\frac{\lfloor nT\rfloor}{n}} - Y_0 \\ -\sum_{i=1}^{\lfloor nT\rfloor} (Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}})Y_{\frac{i-1}{n}} \end{bmatrix}, \tag{3.2}$$

and

$$\begin{bmatrix} \hat{\alpha}_{T,n}^{\mathrm{LSE,D}} \\ \hat{\beta}_{T,n}^{\mathrm{LSE,D}} \end{bmatrix} = n \begin{bmatrix} \lfloor nT\rfloor & -\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}} \\ -\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}} & \sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}^2 \end{bmatrix}^{-1} \begin{bmatrix} X_{\frac{\lfloor nT\rfloor}{n}} - X_0 \\ -\sum_{i=1}^{\lfloor nT\rfloor} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})Y_{\frac{i-1}{n}} \end{bmatrix}, \tag{3.3}$$

provided that the inverse exists, i.e., $\lfloor nT\rfloor \sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}^2 > \Big(\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}\Big)^2$. By Lemma 3.1 in Barczy et al. [6], for all $n \in \mathbb{N}$ and $T \in \mathbb{R}_{++}$ with $\lfloor nT\rfloor \geqslant 2$, we have $\mathbb{P}\Big(\lfloor nT\rfloor \sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}^2 > \Big(\sum_{i=1}^{\lfloor nT\rfloor} Y_{\frac{i-1}{n}}\Big)^2\Big) = 1$.

Proposition 3.1 *If* $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, *and* $\mathbb{P}(Y_0 \in \mathbb{R}_{++}) = 1$, *then for any* $T \in \mathbb{R}_{++}$, *we have*

$$\begin{bmatrix} \hat{a}_{T,n}^{\mathrm{LSE,D}} \\ \hat{b}_{T,n}^{\mathrm{LSE,D}} \\ \hat{\alpha}_{T,n}^{\mathrm{LSE,D}} \\ \hat{\beta}_{T,n}^{\mathrm{LSE,D}} \end{bmatrix} \xrightarrow{\mathbb{P}} \begin{bmatrix} \hat{a}_{T}^{\mathrm{LSE}} \\ \hat{b}_{T}^{\mathrm{LSE}} \\ \hat{\alpha}_{T}^{\mathrm{LSE}} \\ \hat{\beta}_{T}^{\mathrm{LSE}} \end{bmatrix} \qquad \text{as } n \to \infty,$$

where

$$\begin{bmatrix} \hat{a}_{T}^{\mathrm{LSE}} \\ \hat{b}_{T}^{\mathrm{LSE}} \end{bmatrix} := \begin{bmatrix} T & -\int_0^T Y_s\,\mathrm{d}s \\ -\int_0^T Y_s\,\mathrm{d}s & \int_0^T Y_s^2\,\mathrm{d}s \end{bmatrix}^{-1} \begin{bmatrix} Y_T - Y_0 \\ -\int_0^T Y_s\,\mathrm{d}Y_s \end{bmatrix}$$
$$= \frac{1}{T\int_0^T Y_s^2\,\mathrm{d}s - \Big(\int_0^T Y_s\,\mathrm{d}s\Big)^2} \begin{bmatrix} (Y_T - Y_0)\int_0^T Y_s^2\,\mathrm{d}s - \int_0^T Y_s\,\mathrm{d}s \int_0^T Y_s\,\mathrm{d}Y_s \\ (Y_T - Y_0)\int_0^T Y_s\,\mathrm{d}s - T\int_0^T Y_s\,\mathrm{d}Y_s \end{bmatrix}, \tag{3.4}$$

and

$$\begin{bmatrix} \hat{\alpha}_T^{\mathrm{LSE}} \\ \hat{\beta}_T^{\mathrm{LSE}} \end{bmatrix} := \begin{bmatrix} T & -\int_0^T Y_s \mathrm{d}s \\ -\int_0^T Y_s \mathrm{d}s & \int_0^T Y_s^2 \mathrm{d}s \end{bmatrix}^{-1} \begin{bmatrix} X_T - X_0 \\ -\int_0^T Y_s \mathrm{d}X_s \end{bmatrix}$$
$$= \frac{1}{T\int_0^T Y_s^2 \mathrm{d}s - \left(\int_0^T Y_s \mathrm{d}s\right)^2} \begin{bmatrix} (X_T - X_0)\int_0^T Y_s^2 \mathrm{d}s - \int_0^T Y_s \mathrm{d}s \int_0^T Y_s \mathrm{d}X_s \\ (X_T - X_0)\int_0^T Y_s \mathrm{d}s - T\int_0^T Y_s \mathrm{d}X_s \end{bmatrix}, \tag{3.5}$$

which exist almost surely, since

$$\mathbb{P}\left(T\int_0^T Y_s^2 \, \mathrm{d}s > \left(\int_0^T Y_s \, \mathrm{d}s\right)^2\right) = 1 \qquad \text{for all } T \in \mathbb{R}_{++}. \tag{3.6}$$

By definition, we call $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\right)$ *the LSE of* (a, b, α, β) *based on continuous-time observations* $(X_t)_{t\in[0,T]}$, $T \in \mathbb{R}_{++}$.

Proof First, we check (3.6). Note that $\mathbb{P}(\int_0^T Y_s \, \mathrm{d}s < \infty) = 1$ and $\mathbb{P}(\int_0^T Y_s^2 \, \mathrm{d}s < \infty) = 1$ for all $T \in \mathbb{R}_+$, since Y has continuous trajectories almost surely. For each $T \in \mathbb{R}_{++}$, put

$$A_T := \{\omega \in \Omega : t \mapsto Y_t(\omega) \text{ is continuous and non-negative on } [0, T]\}.$$

Then $A_T \in \mathcal{F}$, $\mathbb{P}(A_T) = 1$, and for all $\omega \in A_T$, by the Cauchy–Schwarz's inequality, we have

$$T\int_0^T Y_s(\omega)^2 \, \mathrm{d}s \geqslant \left(\int_0^T Y_s(\omega) \, \mathrm{d}s\right)^2,$$

and $T\int_0^T Y_s(\omega)^2 \, \mathrm{d}s - \left(\int_0^T Y_s(\omega) \, \mathrm{d}s\right)^2 = 0$ if and only if $Y_s(\omega) = K_T(\omega)$ for almost every $s \in [0, T]$ with some $K_T(\omega) \in \mathbb{R}_+$. Hence $Y_s(\omega) = Y_0(\omega)$ for all $s \in [0, T]$ if $\omega \in A_T$ and $T\int_0^T Y_s^2(\omega) \, \mathrm{d}s - \left(\int_0^T Y_s(\omega) \, \mathrm{d}s\right)^2 = 0$. Consequently, using that $\mathbb{P}(A_T) = 1$, we have

$$\mathbb{P}\left(T\int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2 = 0\right) = \mathbb{P}\left(\left\{T\int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2 = 0\right\} \cap A_T\right)$$
$$\leqslant \mathbb{P}(Y_s = Y_0, \ \forall s \in [0, T]) \leqslant \mathbb{P}(Y_T = Y_0) = 0,$$

where the last equality follows by the fact that Y_T is absolutely continuous (see, e.g., Alfonsi [2, Proposition 1.2.11]) together with the law of total probability. Hence $\mathbb{P}\left(T\int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2 = 0\right) = 0$, yielding (3.6).

Further, we have

$$\frac{1}{n}\begin{bmatrix} \lfloor nT \rfloor & -\sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}} \\ -\sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}} & \sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}}^2 \end{bmatrix} \xrightarrow{a.s.} \begin{bmatrix} T & -\int_0^T Y_s ds \\ -\int_0^T Y_s ds & \int_0^T Y_s^2 ds \end{bmatrix} \quad as\ n \to \infty,$$

since $(Y_t)_{t\in\mathbb{R}_+}$ is almost surely continuous. By Proposition I.4.44 in Jacod and Shiryaev [21] with the Riemann sequence of deterministic subdivisions $\left(\frac{i}{n} \wedge T\right)_{i\in\mathbb{N}}$, $n \in \mathbb{N}$, and using the almost sure continuity of $(Y_t, X_t)_{t\in\mathbb{R}_+}$, we obtain

$$\begin{bmatrix} Y_{\frac{\lfloor nT \rfloor}{n}} - Y_0 \\ -\sum_{i=1}^{\lfloor nT \rfloor} (Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}})Y_{\frac{i-1}{n}} \end{bmatrix} \xrightarrow{\mathbb{P}} \begin{bmatrix} Y_T - Y_0 \\ -\int_0^T Y_s \mathrm{d}Y_s \end{bmatrix} \quad \text{as } n \to \infty,$$

$$\begin{bmatrix} X_{\frac{\lfloor nT \rfloor}{n}} - X_0 \\ -\sum_{i=1}^{\lfloor nT \rfloor} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})Y_{\frac{i-1}{n}} \end{bmatrix} \xrightarrow{\mathbb{P}} \begin{bmatrix} X_T - X_0 \\ -\int_0^T Y_s \mathrm{d}X_s \end{bmatrix} \quad \text{as } n \to \infty.$$

By Slutsky's lemma, using also (3.2), (3.3) and (3.6), we obtain the assertion.□

Note that Proposition 3.1 is valid for all $b \in \mathbb{R}$, i.e., not only for subcritical Heston models.

We call the attention that $(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}})$ can be considered to be based only on $(X_t)_{t\in[0,T]}$, since the process $(Y_t)_{t\in[0,T]}$ can be determined using the observations $(X_t)_{t\in[0,T]}$ and the initial value Y_0, see Barczy and Pap [5, Remark 2.5]. We also point out that Overbeck and Rydén [27, formulae (22) and (23)] have already come up with the definition of LSE $(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}})$ of (a, b) based on continuous-time observations $(Y_t)_{t\in[0,T]}$, $T \in \mathbb{R}_{++}$, for the CIR process Y. They investigated only the CIR process Y, so our definitions (3.4) and (3.5) can be considered as generalizations of formulae (22) and (23) in Overbeck and Rydén [27] for the Heston model (1.1). Overbeck and Rydén [27, Theorem 3.4] also proved that the LSE of (a, b) based on continuous-time observations can be approximated in probability by conditional LSEs of (a, b) based on appropriate discrete-time observations.

In the next remark, we point out that the LSE of (a, b, α, β) given in (3.4) and (3.5) can be approximated using discrete-time observations for X, which can be reassuring for practical applications, where data in continuous record is not available.

Remark 3.2 The stochastic integral $\int_0^T Y_s \, \mathrm{d}Y_s$ in (3.4) is a measurable function of $(X_s)_{s\in[0,T]}$ and Y_0. Indeed, for all $t \in [0, T]$, Y_t and $\int_0^t Y_s \, \mathrm{d}s$ are measurable functions of $(X_s)_{s\in[0,T]}$ and Y_0, i.e., they can be determined from a sample $(X_s)_{s\in[0,T]}$ and Y_0 following from a slight modification of Remark 2.5 in Barczy and Pap [5] (replacing y_0 by Y_0), and, by Itô's formula, we have $\mathrm{d}(Y_t^2) = 2Y_t \, \mathrm{d}Y_t + \sigma_1^2 Y_t \, \mathrm{d}t$, $t \in \mathbb{R}_+$, implying that $\int_0^T Y_s \, \mathrm{d}Y_s = \frac{1}{2}\left(Y_T^2 - Y_0^2 - \sigma_1^2 \int_0^T Y_s \, \mathrm{d}s\right)$, $T \in \mathbb{R}_+$. For the stochastic integral $\int_0^T Y_s \, \mathrm{d}X_s$ in (3.5), we have

$$\sum_{i=1}^{\lfloor nT \rfloor} Y_{\frac{i-1}{n}}(X_{\frac{i}{n}} - X_{\frac{i-1}{n}}) \xrightarrow{\mathbb{P}} \int_0^T Y_s \, \mathrm{d}X_s \qquad \text{as } n \to \infty, \tag{3.7}$$

following from Proposition I.4.44 in Jacod and Shiryaev [21] with the Riemann sequence of deterministic subdivisions $\left(\frac{i}{n} \wedge T\right)_{i \in \mathbb{N}}$, $n \in \mathbb{N}$. Thus, there exists a measurable function $\Phi : C([0,T], \mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ such that $\int_0^T Y_s \, \mathrm{d}X_s = \Phi((X_s)_{s\in[0,T]}, Y_0)$, since the convergence in (3.7) holds almost surely along a suitable subsequence, for each $n \in \mathbb{N}$, the members of the sequence in (3.7) are measurable functions of $(X_s)_{s\in[0,T]}$ and Y_0, and one can use Theorems 4.2.2 and 4.2.8 in Dudley [13]. Hence, the right-hand sides of (3.4) and (3.5) are measurable functions of $(X_s)_{s\in[0,T]}$ and Y_0, i.e., they are statistics. □

Using the SDE (1.1) and Corollary 3.2.20 in Karatzas and Shreve [22], one can check that

$$\begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \end{bmatrix} = \begin{bmatrix} T & -\int_0^T Y_s \mathrm{d}s \\ -\int_0^T Y_s \mathrm{d}s & \int_0^T Y_s^2 \mathrm{d}s \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1 \int_0^T Y_s^{1/2} \mathrm{d}W_s \\ -\sigma_1 \int_0^T Y_s^{3/2} \mathrm{d}W_s \end{bmatrix},$$

$$\begin{bmatrix} \hat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \hat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} = \begin{bmatrix} T & -\int_0^T Y_s \mathrm{d}s \\ -\int_0^T Y_s \mathrm{d}s & \int_0^T Y_s^2 \mathrm{d}s \end{bmatrix}^{-1} \begin{bmatrix} \sigma_2 \int_0^T Y_s^{1/2} \mathrm{d}\widetilde{W}_s \\ -\sigma_2 \int_0^T Y_s^{3/2} \mathrm{d}\widetilde{W}_s \end{bmatrix},$$

provided that $T \int_0^T Y_s^2 \, \mathrm{d}s > \left(\int_0^T Y_s \, \mathrm{d}s\right)^2$, where $\widetilde{W}_t := \varrho W_t + \sqrt{1-\varrho^2} B_t$, $t \in \mathbb{R}_+$, and hence

$$\begin{aligned}
\widehat{a}_T^{\mathrm{LSE}} - a &= \frac{\sigma_1\left(\int_0^T Y_s^{1/2} \, \mathrm{d}W_s\right)\left(\int_0^T Y_s^2 \, \mathrm{d}s\right) - \sigma_1\left(\int_0^T Y_s \, \mathrm{d}s\right)\left(\int_0^T Y_s^{3/2} \, \mathrm{d}W_s\right)}{T \int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2}, \\
\widehat{b}_T^{\mathrm{LSE}} - b &= \frac{\sigma_1\left(\int_0^T Y_s^{1/2} \, \mathrm{d}W_s\right)\left(\int_0^T Y_s \, \mathrm{d}s\right) - \sigma_1 T \int_0^T Y_s^{3/2} \, \mathrm{d}W_s}{T \int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2}, \\
\widehat{\alpha}_T^{\mathrm{LSE}} - \alpha &= \frac{\sigma_2\left(\int_0^T Y_s^{1/2} \, \mathrm{d}\widetilde{W}_s\right)\left(\int_0^T Y_s^2 \, \mathrm{d}s\right) - \sigma_2\left(\int_0^T Y_s \, \mathrm{d}s\right)\left(\int_0^T Y_s^{3/2} \, \mathrm{d}\widetilde{W}_s\right)}{T \int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2}, \\
\widehat{\beta}_T^{\mathrm{LSE}} - \beta &= \frac{\sigma_2\left(\int_0^T Y_s^{1/2} \, \mathrm{d}\widetilde{W}_s\right)\left(\int_0^T Y_s \, \mathrm{d}s\right) - \sigma_2 T \int_0^T Y_s^{3/2} \, \mathrm{d}\widetilde{W}_s}{T \int_0^T Y_s^2 \, \mathrm{d}s - \left(\int_0^T Y_s \, \mathrm{d}s\right)^2},
\end{aligned} \tag{3.8}$$

provided that $T \int_0^T Y_s^2 \,\mathrm{d}s > \left(\int_0^T Y_s \,\mathrm{d}s\right)^2$.

4 Consistency and Asymptotic Normality of LSE

Our first result is about the consistency of LSE in case of subcritical Heston models.

Theorem 4.1 *If* $a, b, \sigma_1, \sigma_2 \in \mathbb{R}_{++}$, $\alpha, \beta \in \mathbb{R}$, $\varrho \in (-1, 1)$, *and* $\mathbb{P}((Y_0, X_0) \in \mathbb{R}_{++} \times \mathbb{R}) = 1$, *then the LSE of* (a, b, α, β) *is strongly consistent, i.e.,* $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\right) \xrightarrow{\mathrm{a.s.}} (a, b, \alpha, \beta)$ *as* $T \to \infty$.

Proof By Proposition 3.1, there exists a unique LSE $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\right)$ of (a, b, α, β) for all $T \in \mathbb{R}_{++}$. By (3.8), we have

$$\widehat{a}_T^{\mathrm{LSE}} - a = \frac{\sigma_1 \cdot \frac{1}{T}\int_0^T Y_s \,\mathrm{d}s \cdot \frac{1}{T}\int_0^T Y_s^2 \,\mathrm{d}s \cdot \frac{\int_0^T Y_s^{1/2} \,\mathrm{d}W_s}{\int_0^T Y_s \,\mathrm{d}s} - \sigma_1 \cdot \frac{1}{T}\int_0^T Y_s \,\mathrm{d}s \cdot \frac{1}{T}\int_0^T Y_s^3 \,\mathrm{d}s \cdot \frac{\int_0^T Y_s^{3/2} \,\mathrm{d}W_s}{\int_0^T Y_s^3 \,\mathrm{d}s}}{\frac{1}{T}\int_0^T Y_s^2 \,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s \,\mathrm{d}s\right)^2}$$

provided that $\int_0^T Y_s \,\mathrm{d}s \in \mathbb{R}_{++}$, which holds almost surely, see the proof of Proposition 3.1. Since, by part (1) of Theorem 2.4, $\mathbb{E}(Y_\infty), \mathbb{E}(Y_\infty^2), \mathbb{E}(Y_\infty^3) \in \mathbb{R}_{++}$, part (3) of Theorem 2.4 yields

$$\frac{1}{T}\int_0^T Y_s \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \mathbb{E}(Y_\infty), \qquad \frac{1}{T}\int_0^T Y_s^2 \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \mathbb{E}(Y_\infty^2), \qquad \frac{1}{T}\int_0^T Y_s^3 \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \mathbb{E}(Y_\infty^3)$$

as $T \to \infty$, and then

$$\int_0^T Y_s \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \infty, \qquad \int_0^T Y_s^2 \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \infty, \qquad \int_0^T Y_s^3 \,\mathrm{d}s \xrightarrow{\mathrm{a.s.}} \infty$$

as $T \to \infty$. Hence, by a strong law of large numbers for continuous local martingales (see, e.g., Theorem 2.5), we obtain

$$\widehat{a}_T^{\mathrm{LSE}} - a \xrightarrow{\mathrm{a.s.}} \frac{\sigma_1 \cdot \mathbb{E}(Y_\infty) \cdot \mathbb{E}(Y_\infty^2) \cdot 0 - \sigma_1 \cdot \mathbb{E}(Y_\infty) \cdot \mathbb{E}(Y_\infty^3) \cdot 0}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} = 0 \qquad \text{as } T \to \infty,$$

where for the last step we also used that $\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2 = \frac{a\sigma_1^2}{2b^2} \in \mathbb{R}_{++}$.

Similarly, by (3.8),

$$\widehat{b}_T^{\mathrm{LSE}} - b = \frac{\sigma_1 \cdot \left(\frac{1}{T}\int_0^T Y_s \,\mathrm{d}s\right)^2 \cdot \frac{\int_0^T Y_s^{1/2}\,\mathrm{d}W_s}{\int_0^T Y_s\,\mathrm{d}s} - \sigma_1 \cdot \frac{1}{T}\int_0^T Y_s^3\,\mathrm{d}s \cdot \frac{\int_0^T Y_s^{3/2}\,\mathrm{d}W_s}{\int_0^T Y_s^3\,\mathrm{d}s}}{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\right)^2}$$

$$\xrightarrow{\text{a.s.}} \frac{\sigma_1 \cdot (\mathbb{E}(Y_\infty))^2 \cdot 0 - \sigma_1 \cdot \mathbb{E}(Y_\infty^3)\cdot 0}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} = 0 \qquad \text{as } T \to \infty.$$

One can prove

$$\widehat{\alpha}_T^{\mathrm{LSE}} - \alpha \xrightarrow{\text{a.s.}} 0 \qquad \text{and} \qquad \widehat{\beta}_T^{\mathrm{LSE}} - \beta \xrightarrow{\text{a.s.}} 0 \qquad \text{as } T \to \infty$$

in a similar way. □

Our next result is about the asymptotic normality of LSE in case of subcritical Heston models.

Theorem 4.2 *If* $a, b, \sigma_1, \sigma_2 \in \mathbb{R}_{++}$, $\alpha, \beta \in \mathbb{R}$, $\varrho \in (-1, 1)$ *and* $\mathbb{P}((Y_0, X_0) \in \mathbb{R}_{++} \times \mathbb{R}) = 1$, *then the LSE of* (a, b, α, β) *is asymptotically normal, i.e.,*

$$T^{\frac{1}{2}} \begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \\ \hat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \hat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4\left(\mathbf{0}, \boldsymbol{S} \otimes \begin{bmatrix} \frac{(2a+\sigma_1^2)a}{\sigma_1^2 b} & \frac{2a+\sigma_1^2}{\sigma_1^2} \\ \frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{\sigma_1^2 a} \end{bmatrix}\right) \qquad \text{as } T \to \infty, \tag{4.1}$$

where $\otimes$ *denotes the tensor product of matrices, and*

$$\boldsymbol{S} := \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

With a random scaling, we have

$$E_{1,T}^{-\frac{1}{2}} I_2 \otimes \begin{bmatrix} (TE_{2,T} - E_{1,T}^2)(E_{1,T}E_{3,T} - E_{2,T}^2)^{-\frac{1}{2}} & 0 \\ -T & E_{1,T} \end{bmatrix} \begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \\ \hat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \hat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4\left(\mathbf{0}, \boldsymbol{S} \otimes \boldsymbol{I}_2\right) \tag{4.2}$$

as $T \to \infty$, *where* $E_{i,T} := \int_0^T Y_s^i \,\mathrm{d}s$, $T \in \mathbb{R}_{++}$, $i = 1, 2, 3$.

Proof By Proposition 3.1, there exists a unique LSE $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\right)$ of (a, b, α, β). By (3.8), we have

$$\sqrt{T}(\widehat{a}_T^{\mathrm{LSE}} - a) = \frac{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s \cdot \frac{\sigma_1}{\sqrt{T}}\int_0^T Y_s^{1/2}\,\mathrm{d}W_s - \frac{1}{T}\int_0^T Y_s\,\mathrm{d}s \cdot \frac{\sigma_1}{\sqrt{T}}\int_0^T Y_s^{3/2}\,\mathrm{d}W_s}{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\right)^2},$$

$$\sqrt{T}(\widehat{b}_T^{\mathrm{LSE}} - b) = \frac{\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s \cdot \frac{\sigma_1}{\sqrt{T}}\int_0^T Y_s^{1/2}\,\mathrm{d}W_s - \frac{\sigma_1}{\sqrt{T}}\int_0^T Y_s^{3/2}\,\mathrm{d}W_s}{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\right)^2},$$

$$\sqrt{T}(\widehat{\alpha}_T^{\mathrm{LSE}} - \alpha) = \frac{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s \cdot \frac{\sigma_2}{\sqrt{T}}\int_0^T Y_s^{1/2}\,\mathrm{d}\widetilde{W}_s - \frac{1}{T}\int_0^T Y_s\,\mathrm{d}s \cdot \frac{\sigma_2}{\sqrt{T}}\int_0^T Y_s^{3/2}\,\mathrm{d}\widetilde{W}_s}{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\right)^2},$$

$$\sqrt{T}(\widehat{\beta}_T^{\mathrm{LSE}} - \beta) = \frac{\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s \cdot \frac{\sigma_2}{\sqrt{T}}\int_0^T Y_s^{1/2}\,\mathrm{d}\widetilde{W}_s - \frac{\sigma_2}{\sqrt{T}}\int_0^T Y_s^{3/2}\,\mathrm{d}\widetilde{W}_s}{\frac{1}{T}\int_0^T Y_s^2\,\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\,\mathrm{d}s\right)^2},$$

provided that $T\int_0^T Y_s^2\,\mathrm{d}s > \left(\int_0^T Y_s\,\mathrm{d}s\right)^2$, which holds almost surely. Consequently,

$$\begin{aligned}\sqrt{T}\begin{bmatrix}\widehat{a}_T^{\mathrm{LSE}} - a\\ \widehat{b}_T^{\mathrm{LSE}} - b\\ \widehat{\alpha}_T^{\mathrm{LSE}} - \alpha\\ \widehat{\beta}_T^{\mathrm{LSE}} - \beta\end{bmatrix} &= \frac{1}{\frac{1}{T}\int_0^T Y_s^2\mathrm{d}s - \left(\frac{1}{T}\int_0^T Y_s\mathrm{d}s\right)^2}\left(\boldsymbol{I}_2 \otimes \begin{bmatrix}\frac{1}{T}\int_0^T Y_s^2\mathrm{d}s & \frac{1}{T}\int_0^T Y_s\mathrm{d}s\\ \frac{1}{T}\int_0^T Y_s\mathrm{d}s & 1\end{bmatrix}\right)\frac{1}{\sqrt{T}}\boldsymbol{M}_T\\ &= \left(\boldsymbol{I}_2 \otimes \begin{bmatrix}1 & -\frac{1}{T}\int_0^T Y_s\mathrm{d}s\\ -\frac{1}{T}\int_0^T Y_s\mathrm{d}s & \frac{1}{T}\int_0^T Y_s^2\mathrm{d}s\end{bmatrix}^{-1}\right)\frac{1}{\sqrt{T}}\boldsymbol{M}_T,\end{aligned} \tag{4.3}$$

provided that $T\int_0^T Y_s^2\,\mathrm{d}s > \left(\int_0^T Y_s\,\mathrm{d}s\right)^2$, which holds almost surely, where

$$\boldsymbol{M}_t := \begin{bmatrix}\sigma_1\int_0^t Y_s^{1/2}\mathrm{d}W_s\\ -\sigma_1\int_0^t Y_s^{3/2}\mathrm{d}W_s\\ \sigma_2\int_0^t Y_s^{1/2}\mathrm{d}\tilde{W}_s\\ -\sigma_2\int_0^t Y_s^{3/2}\mathrm{d}\tilde{W}_s\end{bmatrix}, \qquad t \in \mathbb{R}_+,$$

is a 4-dimensional square-integrable continuous local martingale due to $\int_0^t \mathbb{E}(Y_s)\,\mathrm{d}s < \infty$ and $\int_0^t \mathbb{E}(Y_s^3)\,\mathrm{d}s < \infty$, $t \in \mathbb{R}_+$. Next, we show that

$$\frac{1}{\sqrt{T}} \boldsymbol{M}_T \xrightarrow{\mathcal{L}} \boldsymbol{\eta} \boldsymbol{Z} \qquad \text{as } T \to \infty, \tag{4.4}$$

where $\boldsymbol{Z}$ is a 4-dimensional standard normally distributed random vector and $\boldsymbol{\eta} \in \mathbb{R}^{4\times4}$ such that

$$\boldsymbol{\eta}\boldsymbol{\eta}^\top = \boldsymbol{S} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix}.$$

Here, the two symmetric matrices on the right-hand side are positive definite, since $\sigma_1, \sigma_2 \in \mathbb{R}_{++}, \varrho \in (-1, 1), \mathbb{E}(Y_\infty) = \frac{a}{b} \in \mathbb{R}_{++}$ and

$$\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - \left(-\mathbb{E}\left(Y_\infty^2\right)\right)^2 = \frac{a^2\sigma_1^2}{4b^4}(2a + \sigma_1^2) \in \mathbb{R}_{++},$$

and, so is their Kronecker product. Hence $\boldsymbol{\eta}$ can be chosen, for instance, as the uniquely defined symmetric positive definite square root of the Kronecker product of the two matrices in question. We have

$$\langle \boldsymbol{M} \rangle_t = \boldsymbol{S} \otimes \begin{bmatrix} \int_0^t Y_s \, \mathrm{d}s & -\int_0^t Y_s^2 \, \mathrm{d}s \\ -\int_0^t Y_s^2 \, \mathrm{d}s & \int_0^t Y_s^3 \, \mathrm{d}s \end{bmatrix}, \qquad t \in \mathbb{R}_+.$$

By Theorem 2.4, we have

$$\boldsymbol{Q}(t)\langle \boldsymbol{M} \rangle_t \boldsymbol{Q}(t)^\top \xrightarrow{\text{a.s.}} \boldsymbol{S} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \qquad \text{as } t \to \infty$$

with $\boldsymbol{Q}(t) := t^{-1/2}\boldsymbol{I}_4$, $t \in \mathbb{R}_{++}$. Hence, Theorem 2.6 yields (4.4). Then, by (4.3), Slutsky's lemma yields

$$\sqrt{T} \begin{bmatrix} \widehat{a}_T^{\mathrm{LSE}} - a \\ \widehat{b}_T^{\mathrm{LSE}} - b \\ \widehat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \widehat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} \xrightarrow{\mathcal{L}} \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \right) \boldsymbol{\eta} \boldsymbol{Z} \stackrel{\mathcal{L}}{=} \mathcal{N}_4(\boldsymbol{0}, \boldsymbol{\Sigma}) \quad \text{as } T \to \infty,$$

where (applying the identities $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$ and $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A}\boldsymbol{C}) \otimes (\boldsymbol{B}\boldsymbol{D})$)

$$
\begin{aligned}
\boldsymbol{\Sigma} := &\left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right) \boldsymbol{\eta}\mathbb{E}(\boldsymbol{Z}\boldsymbol{Z}^\top)\boldsymbol{\eta}^\top \left(\mathbf{I}_2 \otimes \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right)^\top \\
= &\left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right) \left(\boldsymbol{S} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix}\right) \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right) \\
= &\left(\boldsymbol{I}_2\boldsymbol{S}\boldsymbol{I}_2\right) \otimes \left(\begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}(Y_\infty) & -\mathbb{E}(Y_\infty^2) \\ -\mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \begin{bmatrix} 1 & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right) \\
= &\frac{1}{(\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)^2} \\
&\times \boldsymbol{S} \otimes \begin{bmatrix} (\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2 \\ \mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2 & \mathbb{E}(Y_\infty^3) - 2\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^2) + (\mathbb{E}(Y_\infty))^3 \end{bmatrix},
\end{aligned}
$$

which yields (4.1). Indeed, by Theorem 2.4, an easy calculation shows that

$$
\begin{aligned}
\big(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2\big)\mathbb{E}(Y_\infty) &= \frac{a^3\sigma_1^2}{4b^5}(2a + \sigma_1^2), \\
\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2 &= \frac{a^2\sigma_1^2}{4b^4}(2a + \sigma_1^2), \\
\mathbb{E}(Y_\infty^3) - 2\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^2) + (\mathbb{E}(Y_\infty))^3 &= \frac{a\sigma_1^2}{2b^3}(a + \sigma_1^2), \\
\mathbb{E}(Y_\infty^2) - \big(\mathbb{E}(Y_\infty)\big)^2 &= \frac{a\sigma_1^2}{2b^2}.
\end{aligned}
\tag{4.5}
$$

Now we turn to prove (4.2). Slutsky's lemma, (4.1) and (4.5) yield

$$
\begin{aligned}
&E_{1,T}^{-\frac{1}{2}}\boldsymbol{I}_2 \otimes \begin{bmatrix} (TE_{2,T} - E_{1,T}^2)(E_{1,T}E_{3,T} - E_{2,T}^2)^{-\frac{1}{2}} & 0 \\ -T & E_{1,T} \end{bmatrix} \begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \\ \hat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \hat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} \\
&= \bar{E}_{1,T}^{-\frac{1}{2}}\boldsymbol{I}_2 \otimes \begin{bmatrix} (\bar{E}_{2,T} - \bar{E}_{1,T}^2)(\bar{E}_{1,T}\bar{E}_{3,T} - \bar{E}_{2,T}^2)^{-\frac{1}{2}} & 0 \\ -1 & \bar{E}_{1,T} \end{bmatrix} \sqrt{T}\begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \\ \hat{\alpha}_T^{\mathrm{LSE}} - \alpha \\ \hat{\beta}_T^{\mathrm{LSE}} - \beta \end{bmatrix} \\
&\xrightarrow{\mathcal{L}} (\mathbb{E}(Y_\infty))^{-\frac{1}{2}}\boldsymbol{I}_2 \otimes \begin{bmatrix} (\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)^{-\frac{1}{2}} & 0 \\ -1 & \mathbb{E}(Y_\infty) \end{bmatrix} \\
&\quad \times \mathcal{N}_4\left(\boldsymbol{0}, \boldsymbol{S} \otimes \begin{bmatrix} \frac{(2a+\sigma_1^2)a}{\sigma_1^2 b} & \frac{2a+\sigma_1^2}{\sigma_1^2} \\ \frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{\sigma_1^2 a} \end{bmatrix}\right) \\
&\stackrel{\mathcal{L}}{=} \mathcal{N}_4(0, \boldsymbol{\Xi}) \qquad \text{as } T \to \infty,
\end{aligned}
$$

where $\overline{E}_{i,T} := \frac{1}{T}\int_0^T Y_s^i \, ds$, $T \in \mathbb{R}_{++}$, $i = 1,2,3$, and, applying the identities $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$, $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$, and using (4.5),

$$
\begin{aligned}
\boldsymbol{\Xi} :=& \frac{1}{\mathbb{E}(Y_\infty)}\left(\boldsymbol{I}_2 \otimes \begin{bmatrix} (\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)^{-\frac{1}{2}} & 0 \\ -1 & \mathbb{E}(Y_\infty) \end{bmatrix}\right) \\
&\times \left(\boldsymbol{S} \otimes \begin{bmatrix} \frac{(2a+\sigma_1^2)a}{\sigma_1^2 b} & \frac{2a+\sigma_1^2}{\sigma_1^2} \\ \frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{\sigma_1^2 a} \end{bmatrix}\right) \\
&\times \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} (\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)^{-\frac{1}{2}} & ;0 \\ -1 & \mathbb{E}(Y_\infty) \end{bmatrix}\right)^\top \\
=& \frac{1}{\mathbb{E}(Y_\infty)}(\boldsymbol{I}_2 \boldsymbol{S} \boldsymbol{I}_2) \otimes \left(\begin{bmatrix} (\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)^{-\frac{1}{2}} & 0 \\ -1 & \mathbb{E}(Y_\infty) \end{bmatrix}\right. \\
&\times \begin{bmatrix} \frac{(2a+\sigma_1^2)a}{\sigma_1^2 b} & \frac{2a+\sigma_1^2}{\sigma_1^2} \\ \frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{\sigma_1^2 a} \end{bmatrix} \\
&\times \left.\begin{bmatrix} (\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2)^{-\frac{1}{2}} & -1 \\ 0 & \mathbb{E}(Y_\infty) \end{bmatrix}\right) \\
=& \frac{b}{a}\boldsymbol{S} \otimes \left(\begin{bmatrix} \sigma_1(2a+\sigma_1^2)^{-\frac{1}{2}} & 0 \\ -1 & \frac{a}{b} \end{bmatrix}\begin{bmatrix} \frac{(2a+\sigma_1^2)a}{\sigma_1^2 b} & \frac{2a+\sigma_1^2}{\sigma_1^2} \\ \frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{\sigma_1^2 a} \end{bmatrix}\begin{bmatrix} \sigma_1(2a+\sigma_1^2)^{-\frac{1}{2}} & -1 \\ 0 & \frac{a}{b} \end{bmatrix}\right) \\
=& \boldsymbol{S} \otimes \boldsymbol{I}_2.
\end{aligned}
$$

Thus we obtain (4.2). □

Next, we formulate a corollary of Theorem 4.2 presenting separately the asymptotic behavior of the LSE of (a, b) based on continuous-time observations $(Y_t)_{t\in[0,T]}$, $T > 0$. We call the attention that Overbeck and Rydén [27, Theorem 3.6] already derived this asymptotic behavior (for more details on the role of the initial distribution, see the Introduction); however, the covariance matrix of the limit normal distribution in their Theorem 3.6 is somewhat complicated. It turns out that it can be written in a much simpler form by making a simple reparametrization of the SDE (1) in Overbeck and Rydén [27], estimating $-b$ instead of b (with the notations of Overbeck and Rydén [27]), i.e., considering the SDE (1.1) and estimating b (with our notations).

Corollary 4.3 If $a, b, \sigma_1 \in \mathbb{R}_{++}$, and $\mathbb{P}(Y_0 \in \mathbb{R}_{++}) = 1$, then the LSE of (a, b) given in (3.4) based on continuous-time observations $(Y_t)_{t\in[0,T]}$, $T > 0$, is strongly consistent and asymptotically normal, i.e., $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}\right) \xrightarrow{\text{a.s.}} (a, b)$ as $T \to \infty$, and

$$
T^{\frac{1}{2}}\begin{bmatrix} \hat{a}_T^{\mathrm{LSE}} - a \\ \hat{b}_T^{\mathrm{LSE}} - b \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2\left(\boldsymbol{0}, \begin{bmatrix} \frac{(2a+\sigma_1^2)a}{b} & 2a+\sigma_1^2 \\ 2a+\sigma_1^2 & \frac{2b(a+\sigma_1^2)}{a} \end{bmatrix}\right) \qquad \text{as } T \to \infty.
$$

5 Numerical Illustrations

In this section, first, we demonstrate some methods for the simulation of the Heston model (1.1), and then we illustrate Theorem 4.1 and convergence (4.1) in Theorem 4.2 using generated sample paths of the Heston model (1.1). We will consider a subcritical Heston model (1.1) (i.e., $b \in \mathbb{R}_{++}$) with a known non-random initial value $(y_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}$. Note that in this case, the augmented filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ corresponding to $(W_t, B_t)_{t\in\mathbb{R}_+}$ and the initial value $(y_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}$, in fact, does not depend on (y_0, x_0). We recall five simulation methods which differ from each other in how the CIR process in the Heston model (1.1) is simulated.

In what follows, let η_k, $k \in \{1, \dots, N\}$, be independent standard normally distributed random variables with some $N \in \mathbb{N}$, and put $t_k := k\frac{T}{N}$, $k \in \{0, 1, \dots, N\}$, with some $T \in \mathbb{R}_{++}$.

Higham and Mao [15] introduced the Absolute Value Euler (AVE) method

$$Y_{t_k}^{(N)} = Y_{t_{k-1}}^{(N)} + (a - bY_{t_{k-1}}^{(N)})(t_k - t_{k-1}) + \sigma_1 \sqrt{|Y_{t_{k-1}}^{(N)}|}\sqrt{t_k - t_{k-1}}\,\eta_k, \qquad k \in \{1, \dots, N\},$$

with $Y_0^{(N)} = y_0$ for the approximation of the CIR process, where $a, b, \sigma_1 \in \mathbb{R}_{++}$. This scheme does not preserve non-negativity of the CIR process.

The Truncated Euler (TE) scheme uses the discretization

$$Y_{t_k}^{(N)} = Y_{t_{k-1}}^{(N)} + (a - bY_{t_{k-1}}^{(N)})(t_k - t_{k-1}) + \sigma_1 \sqrt{\max(Y_{t_{k-1}}^{(N)}, 0)}\sqrt{t_k - t_{k-1}}\,\eta_k, \qquad k \in \{1, \dots, N\},$$

with $Y_0^{(N)} = y_0$, where $a, b, \sigma_1 \in \mathbb{R}_{++}$, for approximation of the CIR process Y, see, e.g., Deelstra and Delbaen [10]. This scheme does not preserve non-negativity of the CIR process.

The Symmetrized Euler (SE) method gives an approximation of the CIR process Y via the recursion

$$Y_{t_k}^{(N)} = \left| Y_{t_{k-1}}^{(N)} + \left(a - bY_{t_{k-1}}^{(N)}\right)(t_k - t_{k-1}) + \sigma_1 \sqrt{Y_{t_{k-1}}^{(N)}}\sqrt{t_k - t_{k-1}}\,\eta_k \right|, \qquad k \in \{1, \dots, N\},$$

with $Y_0^{(N)} = y_0$, where $a, b, \sigma_1 \in \mathbb{R}_{++}$, see, Diop [12] or Berkaoui et al. [7] (where the method is analyzed for more general SDEs including so-called alpha-root processes as well with diffusion coefficient $\sqrt[\alpha]{x}$ with $\alpha \in (1, 2]$ instead of $\sqrt{x}$). This scheme gives a non-negative approximation of the CIR process Y.

The following two methods do not directly simulate the CIR process Y, but its square root $Z = (Z_t := \sqrt{Y_t})_{t\in\mathbb{R}_+}$. If $a > \frac{\sigma_1^2}{2}$, then $\mathbb{P}(Y_t \in \mathbb{R}_{++},\ \forall\, t \in \mathbb{R}_+) = 1$, and, by Itô's formula,

$$\mathrm{d}Z_t = \left(\left(\frac{a}{2} - \frac{\sigma_1^2}{8}\right)\frac{1}{Z_t} - \frac{b}{2}Z_t\right)\mathrm{d}t + \frac{\sigma_1}{2}\,\mathrm{d}W_t, \qquad t \in \mathbb{R}_+.$$

The Drift Explicit Square Root Euler (DESRE) method (see, e.g., Kloeden and Platen [23, Section 10.2] or Hutzenthaler et al. [19, equation (4)] for general SDEs) simulates Z by

$$Z_{t_k}^{(N)} = Z_{t_{k-1}}^{(N)} + \left(\left(\frac{a}{2} - \frac{\sigma_1^2}{8} \right) \frac{1}{Z_{t_{k-1}}^{(N)}} - \frac{b}{2} Z_{t_{k-1}}^{(N)} \right)(t_k - t_{k-1}) + \frac{\sigma_1}{2} \sqrt{t_k - t_{k-1}}\, \eta_k, \qquad k \in \{1, \dots, N\},$$

with $Z_0^{(N)} = \sqrt{y_0}$, where $a > \frac{\sigma_1^2}{2}$ and $b, \sigma_1 \in \mathbb{R}_{++}$. Here note that $\mathbb{P}(Z_{t_k}^{(N)} = 0) = 0$, $k \in \{1, \dots, N\}$, since $Z_{t_k}^{(N)}$ is absolutely continuous. Transforming back, i.e., $Y_{t_k}^{(N)} = (Z_{t_k}^{(N)})^2$, $k \in \{0, 1, \dots, N\}$, gives a non-negative approximation of the CIR process Y.

The Drift Implicit Square Root Euler (DISRE) method (see, Alfonsi [1] or Dereich et al. [11]) simulates Z by

$$Z_{t_k}^{(N)} = Z_{t_{k-1}}^{(N)} + \left(\left(\frac{a}{2} - \frac{\sigma_1^2}{8} \right) \frac{1}{Z_{t_k}^{(N)}} - \frac{b}{2} Z_{t_k}^{(N)} \right)(t_k - t_{k-1}) + \frac{\sigma_1}{2} \sqrt{t_k - t_{k-1}}\, \eta_k, \qquad k \in \{1, \dots, N\},$$

with $Z_0^{(N)} = \sqrt{y_0}$, where $a > \frac{\sigma_1^2}{2}$ and $b, \sigma_1 \in \mathbb{R}_{++}$. This recursion has a unique positive solution given by

$$Z_{t_k}^{(N)} = \frac{Z_{t_{k-1}}^{(N)} + \frac{\sigma_1}{2}\sqrt{t_k - t_{k-1}}\,\eta_k}{2 + b(t_k - t_{k-1})} + \sqrt{\frac{\left(Z_{t_{k-1}}^{(N)} + \frac{\sigma_1}{2}\sqrt{t_k - t_{k-1}}\,\eta_k\right)^2}{(2 + b(t_k - t_{k-1}))^2} + \frac{\left(a - \frac{\sigma_1^2}{4}\right)(t_k - t_{k-1})}{2 + b(t_k - t_{k-1})}}$$

for $k \in \{1, \dots, N\}$ with $Z_0^{(N)} = \sqrt{y_0}$. Transforming again back, i.e., $Y_{t_k}^{(N)} = (Z_{t_k}^{(N)})^2$, $k \in \{0, 1, \dots, N\}$, gives a strictly positive approximation of the CIR process Y.

We mention that there exist so-called exact simulation methods for the CIR process, see, e.g., Alfonsi [2, Section 3.1]. In our simulations, we will use the SE, DESRE and DISRE methods for approximating the CIR process which preserve non-negativity of the CIR process.

The second coordinate process X of the Heston process (1.1) will be approximated via the usual Euler–Maruyama scheme given by

$$X_{t_k}^{(N)} = X_{t_{k-1}}^{(N)} + (\alpha - \beta Y_{t_{k-1}}^{(N)})(t_k - t_{k-1}) + \sigma_2 \sqrt{Y_{t_{k-1}}^{(N)}} \sqrt{t_k - t_{k-1}} \left(\varrho\, \eta_k + \sqrt{1 - \varrho^2}\, \zeta_k \right) \tag{5.1}$$

for $k \in \{1, \dots, N\}$ with $X_0^{(N)} = x_0$, where $\alpha, \beta \in \mathbb{R}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, and ζ_k, $k \in \{1, \dots, N\}$, be independent standard normally distributed random variables independent of η_k, $k \in \{1, \dots, N\}$. Note that in (5.1) the factor $\sqrt{Y_{t_{k-1}}^{(N)}}$ appears, which is well-defined in case of the CIR process Y is approximated by the SE, DESRE or DISRE methods, that we will consider.

We also mention that there exist exact simulation methods for the Heston process (1.1), see, e.g., Broadie and Kaya [8] or Alfonsi [2, Section 4.2.6].

We will approximate the estimator $\big(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\big)$ given in (3.4) and (3.5) using the generated sample paths of (Y, X). For this, we need to simulate, for a large time $T \in \mathbb{R}_{++}$, the random variables

$$Y_T, \quad X_T, I_{1,T} := \int_0^T Y_s \,\mathrm{d}s, I_{2,T} := \int_0^T Y_s^2 \,\mathrm{d}s, I_{3,T} := \int_0^T Y_s \,\mathrm{d}Y_s, I_{4,T} := \int_0^T Y_s \,\mathrm{d}X_s.$$

We can easily approximate the $I_{i,T}$, $i \in \{1, 2, 3, 4\}$, respectively, by

$$I_{1,T}^N := \sum_{k=1}^N Y_{t_{k-1}}^{(N)}(t_k - t_{k-1}) = \frac{T}{N}\sum_{k=1}^N Y_{t_{k-1}}^{(N)}, \qquad I_{2,T}^N := \sum_{k=1}^N (Y_{t_{k-1}}^{(N)})^2(t_k - t_{k-1}) = \frac{T}{N}\sum_{k=1}^N (Y_{t_{k-1}}^{(N)})^2,$$
$$I_{3,T}^N := \sum_{k=1}^N Y_{t_{k-1}}^{(N)}(Y_{t_k}^{(N)} - Y_{t_{k-1}}^{(N)}), \qquad I_{4,T}^N := \sum_{k=1}^N Y_{t_{k-1}}^{(N)}(X_{t_k}^{(N)} - X_{t_{k-1}}^{(N)}).$$

Hence, we can approximate $\widehat{a}_T^{\mathrm{LSE}}$, $\widehat{b}_T^{\mathrm{LSE}}$, $\widehat{\alpha}_T^{\mathrm{LSE}}$, and $\widehat{\beta}_T^{\mathrm{LSE}}$ by

$$\widehat{a}_T^{(N)} := \frac{(Y_T^{(N)} - y_0)I_{2,T}^N - I_{1,T}^N I_{3,T}^N}{TI_{2,T}^N - (I_{1,T}^N)^2}, \qquad \widehat{b}_T^{(N)} := \frac{(Y_T^{(N)} - y_0)I_{1,T}^N - TI_{3,T}^N}{TI_{2,T}^N - (I_{1,T}^N)^2},$$
$$\widehat{\alpha}_T^{(N)} := \frac{(X_T^{(N)} - x_0)I_{2,T}^N - I_{1,T}^N I_{4,T}^N}{TI_{2,T}^N - (I_{1,T}^N)^2}, \qquad \widehat{\beta}_T^{(N)} := \frac{(X_T^{(N)} - x_0)I_{1,T}^N - TI_{4,T}^N}{TI_{2,T}^N - (I_{1,T}^N)^2}.$$

We point out that $\widehat{a}_T^{(N)}$, $\widehat{b}_T^{(N)}$, $\widehat{\alpha}_T^{(N)}$ and $\widehat{\beta}_T^{(N)}$ are well-defined, since

$$TI_{2,T}^N - (I_{1,T}^N)^2 = \frac{T^2}{N}\sum_{k=1}^N \left(Y_{t_k}^{(N)} - \frac{1}{N}\sum_{k=1}^N Y_{t_{k-1}}^{(N)}\right)^2 \geqslant 0,$$

and

$$TI_{2,T}^N - (I_{1,T}^N)^2 = 0 \qquad \iff \qquad Y_{t_k}^{(N)} = \frac{1}{N}\sum_{\ell=1}^N Y_{t_{\ell-1}}^{(N)}, \qquad k \in \{1, \ldots, N\}$$
$$\iff \qquad Y_0^{(N)} = Y_{t_1}^{(N)} = \cdots = Y_{t_{N-1}}^{(N)}.$$

Consequently, using that $Y_{t_1}^{(N)}$ is absolutely continuous together with the law of total probability, we have $\mathbb{P}(TI_{2,T}^N - (I_{1,T}^N)^2 \in \mathbb{R}_{++}) = 1$.

For the numerical implementation, we take $y_0 = 0.2$, $x_0 = 0.1$, $a = 0.4$, $b = 0.3$, $\alpha = 0.1$, $\beta = 0.15$, $\sigma_1 = 0.4$, $\sigma_2 = 0.3$, $\varrho = 0.2$, $T = 3000$, and $N = 30000$

(consequently, $t_k - t_{k-1} = 0.1$, $k \in \{1, \dots, N\}$). Note that $a > \frac{\sigma_1^2}{2}$ with this choice of parameters. We simulate 10,000 independent trajectories of (Y_T, X_T) and the normalized error $T^{\frac{1}{2}}\left(\widehat{a}_T^{\mathrm{LSE}} - a, \widehat{b}_T^{\mathrm{LSE}} - b, \widehat{\alpha}_T^{\mathrm{LSE}} - \alpha, \widehat{\beta}_T^{\mathrm{LSE}} - \beta\right)$. Table 1 contains the empirical mean of $Y_T^{(N)}$ and $\frac{1}{T}X_T^{(N)}$, based on 10,000 independent trajectories of (Y_T, X_T), and the (theoretical) limit $\lim_{t\to\infty} \mathbb{E}(Y_t) = \frac{a}{b}$ and $\lim_{t\to\infty} t^{-1}\mathbb{E}(X_t) = \alpha - \frac{\beta a}{b}$, respectively (following from Proposition 2.2), using the schemes SE, DESRE and DISRE for simulating the CIR process.

Henceforth, we will use the above choice of parameters except that $T = 5000$ and $N = 50,000$ (yielding $t_k - t_{k-1} = 0.1$, $k \in \{1, \dots, N\}$).

In Table 2, we calculate the expected bias $(\mathbb{E}(\widehat{\theta}_T^{\mathrm{LSE}} - \theta))$, the L_1-norm of error $(\mathbb{E}|\widehat{\theta}_T^{\mathrm{LSE}} - \theta|)$ and the L_2-norm of error $\left(\left(\mathbb{E}(\widehat{\theta}_T^{\mathrm{LSE}} - \theta)^2\right)^{1/2}\right)$, where $\theta \in \{a, b, \alpha, \beta\}$, using the scheme DISRE for simulating the CIR process.

In Table 3, we give the relative errors $(\widehat{\theta}_T^{(N)} - \theta)/\theta$, where $\theta \in \{a, b, \alpha, \beta\}$, for $T = 5000$ using the scheme DISRE for simulating the CIR process.

In Fig. 1, we illustrate the limit law of each coordinate of the LSE $\left(\widehat{a}_T^{\mathrm{LSE}}, \widehat{b}_T^{\mathrm{LSE}}, \widehat{\alpha}_T^{\mathrm{LSE}}, \widehat{\beta}_T^{\mathrm{LSE}}\right)$ given in (4.1). To do so, we plot the obtained density histograms of each of its coordinates based on 10,000 independently generated trajectories using the scheme DISRE for simulating the CIR process, we also plotted the density functions of the corresponding normal limit distributions in red.With the above choice of parameters, as a consequence of (4.1), we have

$$
\begin{aligned}
&T^{\frac{1}{2}}(\hat{a}_T^{\mathrm{LSE}} - a) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{a}{b}(2a + \sigma_1^2)\right) = \mathcal{N}(0, 1.28) \qquad \text{as } T \to \infty,\\
&T^{\frac{1}{2}}(\hat{b}_T^{\mathrm{LSE}} - b) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2b}{a}(a + \sigma_1^2)\right) = \mathcal{N}(0, 0.84) \qquad \text{as } T \to \infty,\\
&T^{\frac{1}{2}}(\hat{\alpha}_T^{\mathrm{LSE}} - \alpha) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{a\sigma_2^2}{b\sigma_1^2}(2a + \sigma_1^2)\right) = \mathcal{N}(0, 0.72) \qquad \text{as } T \to \infty,\\
&T^{\frac{1}{2}}(\hat{\beta}_T^{\mathrm{LSE}} - \beta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{2b\sigma_2^2}{a\sigma_1^2}(a + \sigma_1^2)\right) = \mathcal{N}(0, 0.4725) \qquad \text{as } T \to \infty.
\end{aligned}
$$

In case of the parameters a and b, one can see a bias in Fig. 1, which, in our opinion, may be related with the different speeds of weak convergence for the LSE of (a, b) and that of (α, β), and with the bad performance of the applied discretization scheme for Y.

Table 4 contains the skewness and excess kurtosis of $T^{\frac{1}{2}}(\widehat{\theta}_T^{(N)} - \theta)$, where $\theta \in \{a, b, \alpha, \beta\}$, using the scheme DISRE for simulating the CIR process. This confirms our results in (4.1) as well.

Using the Anderson–Darling and Jarque–Bera tests, we test whether each of the coordinates of $T^{\frac{1}{2}}\left(\widehat{a}_T^{\mathrm{LSE}} - a, \widehat{b}_T^{\mathrm{LSE}} - b, \widehat{\alpha}_T^{\mathrm{LSE}} - \alpha, \widehat{\beta}_T^{\mathrm{LSE}} - \beta\right)$ follows a normal distribution or not for $T = 5000$. In Table 5 we give the test values and (in parenthesis) the p-values of the Anderson–Darling and Jarque–Bera tests using the scheme DISRE

Table 1 Empirical mean of $Y_T^{(N)}$ (first row) and $\frac{1}{T}X_T^{(N)}$ (second row)

Empirical mean of $Y_T^{(N)}$ and $\frac{1}{T}X_T^{(N)}$	SE	DESRE	DISRE
$\lim_{t\to\infty} E(Y_t) = \frac{a}{b} = 1.3333$	1.321025	1.325539	1.331852
$\lim_{t\to\infty} t^{-1}E(X_t) = \alpha - \frac{\beta a}{b} = -0.1$	− 0.09978663	− 0.100054	− 0.09941841

Table 2 Expected bias, L_1- and L_2-norm of error using DISRE scheme

Errors	Expected bias	L_1-norm of error	L_2-norm of error
a	− 0.01089369	0.0153848	0.0190123
b	− 0.007639168	0.01189344	0.01474495
α	0.0001779072	0.00957648	0.0120646
β	0.0001402452	0.007776999	0.009771835

Table 3 Relative errors using DISRE scheme

Relative errors	$T = 5000$
$(\widehat{a}_T^{(N)} - a)/a$	− 0.02723421
$(\widehat{b}_T^{(N)} - b)/b$	− 0.02546389
$(\widehat{\alpha}_T^{(N)} - \alpha)/\alpha$	0.001779072
$(\widehat{\beta}_T^{(N)} - \beta)/\beta$	0.0009349683

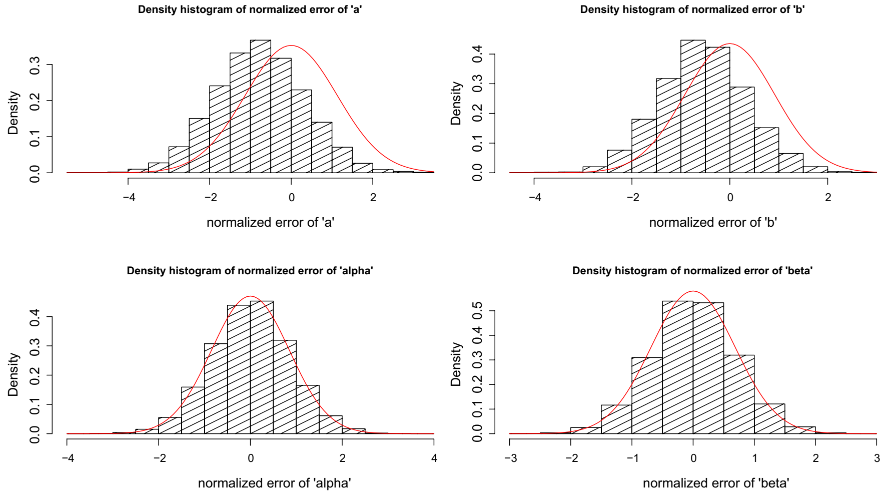

Fig. 1 In the first line from left to right, the density histograms of the normalized errors of $T^{1/2}(\widehat{a}_T^{(N)} - a)$ and $T^{1/2}(\widehat{b}_T^{(N)} - b)$, in the second line from left to right, the density histograms of the normalized errors of $T^{1/2}(\widehat{\alpha}_T^{(N)} - \alpha)$ and $T^{1/2}(\widehat{\beta}_T^{(N)} - \beta)$. In each case, the red line denotes the density function of the corresponding normal limit distribution

Table 4 Skewness and excess kurtosis using the scheme DISRE for simulating the CIR process

Skewness and excess kurtosis	$T^{\frac{1}{2}}(\widehat{a}_T^{(N)} - a)$	$T^{\frac{1}{2}}(\widehat{b}_T^{(N)} - b)$	$T^{\frac{1}{2}}(\widehat{\alpha}_T^{(N)} - \alpha)$	$T^{\frac{1}{2}}(\widehat{\beta}_T^{(N)} - \beta)$
Skewness	0.04915124	0.04544189	− 0.02317407	− 0.01399869
Excess kurtosis	0.07666643	0.05226811	0.09994108	0.07877347

Table 5 Test of normality in case of $y_0 = 0.2$, $x_0 = 0.1$, $a = 0.4$, $b = 0.3$, $\alpha = 0.1$, $\beta = 0.15$, $\sigma_1 = 0.4$, $\sigma_2 = 0.3$, $\varrho = 0.2$, $T = 5000$, and $N = 50,000$ generating 10,000 independent sample paths using the scheme DISRE for simulating the CIR process

Test of normality	$T^{\frac{1}{2}}(\widehat{a}_T^{(N)} - a)$	$T^{\frac{1}{2}}(\widehat{b}_T^{(N)} - b)$	$T^{\frac{1}{2}}(\widehat{\alpha}_T^{(N)} - \alpha)$	$T^{\frac{1}{2}}(\widehat{\beta}_T^{(N)} - \beta)$
Anderson–Darling	0.34486 (0.4857*)	0.62481 (0.1037*)	0.34078 (0.4962*)	0.35232 (0.467*)
Jarque–Bera	6.5162 (0.03846)	4.6077 (0.09987*)	5.1089 (0.07774*)	2.9528 (0.2285*)

for simulating the CIR process (the $*$ after a p-value denotes that the p-value in question is greater than any reasonable significance level). It turns out that, with this choice of parameters, at any reasonable significance level the Anderson–Darling test accepts that $T^{\frac{1}{2}}(\widehat{a}_T^{\mathrm{LSE}} - a)$, $T^{\frac{1}{2}}(\widehat{b}_T^{\mathrm{LSE}} - b)$, $T^{\frac{1}{2}}(\widehat{\alpha}_T^{\mathrm{LSE}} - \alpha)$, and $T^{\frac{1}{2}}(\widehat{\beta}_T^{\mathrm{LSE}} - \beta)$ follow normal laws. The Jarque–Bera test also accepts that $T^{\frac{1}{2}}(\widehat{b}_T^{\mathrm{LSE}} - b)$, $T^{\frac{1}{2}}(\widehat{\alpha}_T^{\mathrm{LSE}} - \alpha)$, and $T^{\frac{1}{2}}(\widehat{\beta}_T^{\mathrm{LSE}} - \beta)$ follow normal laws, but rejects that $T^{\frac{1}{2}}(\widehat{a}_T^{\mathrm{LSE}} - a)$ follows a normal law.

All in all, our numerical illustrations are more or less in accordance with our theoretical results in (4.1). Finally, we note that we used the open source software R for making the simulations.

References

1. Alfonsi A (2005) On the discretization schemes for the CIR (and Bessel squared) processes. Monte Carlo Methods Appl 11(4):355–384
2. Alfonsi A (2015) Affine Diffusions and related processes: simulation, theory and applications. Springer, Cham
3. Barczy M, Döring L, Li Z, Pap G (2013) On parameter estimation for critical affine processes. Electron J Stat 7:647–696
4. Barczy M, Döring L, Li Z, Pap G (2014) Stationarity and ergodicity for an affine two factor model. Adv Appl Probab 46(3):878–898
5. Barczy M, Pap G (2016) Asymptotic properties of maximum likelihood estimators for Heston models based on continuous time observations. Statistics 50(2):389–417
6. Barczy M, Pap G, Szabó TT (2016) Parameter estimation for the subcritical Heston model based on discrete time observations. Acta Sci Math (Szeged) 82:313–338
7. Berkaoui A, Bossy M, Diop A (2008) Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. ESAIM Probab Stat 12:1–11
8. Broadie M, Kaya Ö (2006) Exact simulation of stochastic volatility and other affine jump diffusion processes. Oper Res 54(2):217–231
9. Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. Econometrica 53(2):385–407

10. Deelstra G, Delbaen F (1998) Convergence of discretized stochastic (interest rate) processes with stochastic drift term. Appl Stoch Models Data Anal 14(1):77–84
11. Dereich S, Neuenkirch A, Szpruch L (2012) An Euler-type method for the strong approximation of the Cox–Ingersoll–Ross process. Proc R Soc Lond Ser A Math Phys Eng Sci 468(2140):1105–1115
12. Diop A (2003) Sur la discretisation et le comportement a petit bruit d'EDS multidimensionnelles dont les coefficients sont a derivees singulieres. Ph.D thesis, INRIA, France
13. Dudley RM (1989) Real analysis and probability. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove
14. Heston S (1993) A closed-form solution for options with stochastic volatilities with applications to bond and currency options. Rev Financ Stud 6:327–343
15. Higham D, Mao X (2005) Convergence of Monte Carlo simulations involving the mean-reverting square root process. J Comput Finance 8(3):35–61
16. Hu Y, Long H (2007) Parameter estimation for Ornstein–Uhlenbeck processes driven by α-stable Lévy motions. Commun Stoch Anal 1(2):175–192
17. Hu Y, Long H (2009) Least squares estimator for Ornstein–Uhlenbeck processes driven by α-stable motions. Stoch Process Appl 119(8):2465–2480
18. Hu Y, Long H (2009) On the singularity of least squares estimator for mean-reverting α-stable motions. Acta Math Sci 29B(3):599–608
19. Hutzenthaler M, Jentzen A, Kloeden PE (2012) Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. Ann Appl Probab 22(4):1611–1641
20. Hurn AS, Lindsay KA, McClelland AJ (2013) A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. J Econom 172:106–126
21. Jacod J, Shiryaev AN (2003) Limit theorems for stochastic processes, 2nd edn. Springer, Berlin
22. Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, Berlin
23. Kloeden PE, Platen E (1992) Numerical solution of stochastic differential equations. Applications of mathematics (New York), vol 23. Springer, Berlin
24. Lehmann EL, Romano JP (2009) Testing statistical hypotheses, 3rd edn. Springer, Berlin
25. Li Z, Ma C (2015) Asymptotic properties of estimators in a stable Cox–Ingersoll–Ross model. Stoch Process Appl 125(8):3196–3233
26. Liptser RS, Shiryaev AN (2001) Statistics of random processes II. Applications, 2nd edn. Springer, Berlin
27. Overbeck L, Rydén T (1997) Estimation in the Cox–Ingersoll–Ross model. Econometric theory 13(3):430–461
28. van Zanten H (2000) A multivariate central limit theorem for continuous local martingales. Stat Probab Lett 50(3):229–235