Parameter estimation for the subcritical Heston model based on discrete time observations

Mátyás Barczy[*,◇], Gyula Pap[**] and Tamás T. Szabó[**]

* Faculty of Informatics, University of Debrecen, Pf. 12, H–4010 Debrecen, Hungary.

** Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H–6720 Szeged, Hungary.

e–mails: barczy.matyas@inf.unideb.hu (M. Barczy), papgy@math.u-szeged.hu (G. Pap), tszabo@math.u-szeged.hu (T. T. Szabó).

◇ Corresponding author.

Abstract

We study asymptotic properties of some (essentially conditional least squares) parameter estimators for the subcritical Heston model based on discrete time observations derived from conditional least squares estimators of some modified parameters.

1 Introduction

The Heston model has been extensively used in financial mathematics since one can well-fit them to real financial data set, and they are well-tractable from the point of view of computability as well. Hence parameter estimation for the Heston model is an important task.

In this paper we study the Heston model

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

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 [7]. We investigate only the so-called subcritical case, i.e., when $b > 0$, see Definition 2.3, and we introduce some parameter estimator of (a, b, α, β) based on discrete time observations and derived from conditional least squares estimators (CLSEs) of some modified parameters starting the process (Y, X) from some known non-random initial value $(y_0, x_0) \in (0, \infty) \times \mathbb{R}$. 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 an arbitrarily short continuous time observation $(X_t)_{t \in [0,T]}$ of X, where $T > 0$, see, e.g., Barczy and Pap [1, Remark 2.6]. In Overbeck and Rydén [15, Theorems 3.2 and 3.3] one can find a strongly consistent and asymptotically normal estimator of σ_1 based on discrete time observations for the process Y, and for another estimator of σ_1, see Dokuchaev [5]. Eventually, it turns out that for the calculation of the estimator of (a, b, α, β), one does not need to know the values of the parameters σ_1, σ_2 and ϱ. For interpretations of Y and X in financial mathematics, see, e.g., Hurn et al. [8, Section 4].

CLS estimation has been considered for the Cox-Ingersoll-Ross (CIR) model, which satisfies the first equation of (1.1). For the CIR model, Overbeck and Rydén [15] derived the CLSEs and gave their asymptotic properties, however, they did not investigate the conditions of their existence. Specifically, Theorems 3.1 and 3.3 in Overbeck and Rydén [15] correspond to our Theorem 3.4, but they estimate the volatility coefficient σ_1 as well, which we assume to be known. Li and Ma [14] extended the investigation to so-called stable CIR processes driven by an α-stable process instead of a Brownian motion. For a more complete overview of parameter estimation for the Heston model see, e.g., the introduction in Barczy and Pap [1].

It would be possible to calculate the discretized version of the maximum likelihood estimators derived in Barczy and Pap [1] using the same procedure as in Ben Alaya and Kebaier [3, Section 4] valid for discrete time observations of high frequency. However, this would be basically different from the present line of investigation, therefore we will not discuss it further.

The organization of the paper is the following. In Section 2 we recall some important results about the existence of a unique strong solution to (1.1), and study its asymptotic properties. In the subcritical case, i.e., when $b > 0$, we invoke a result due to Cox et al. [4] on the unique existence of a stationary distribution, and we slightly improve a result due to Li and Ma [14] and Jin et al. [10, Corollary 2.7] and [11, Corollaries 5.9 and 6.4] on the ergodicity of the CIR process $(Y_t)_{t \geqslant 0}$, see Theorem 2.4. We also recall some convergence results for square-integrable martingales. In Section 3 we introduce the CLSE of a transformed parameter vector based on discrete time observations, and derive the asymptotic properties of the estimates – namely, strong consistency and asymptotic normality, see Theorem 3.2. Thereafter, we apply these results together with the so-called delta method to obtain the same asymptotic properties of the estimators for the original parameters, see Theorem 3.4. The point of the parameter transformation is to reduce the minimization in the CLS method to a linear problem, because our objective function depends on the original parameters through complicated functions. The covariance matrices of the limit normal distributions in Theorems 3.2 and 3.4 depend on the unknown parameters a, b and β, as well (but somewhat surprisingly not on α). They also depend on the volatility parameters σ_1, σ_2 and ρ, but, again, we will assume these to be known. Since the considered estimators of a, b and β are proved to be strongly consistent, using random normalization, one may derive counterparts of Theorems 3.2 and 3.4 in a way that the limit distributions are four-dimensional standard normal distributions (having the identity matrix $\boldsymbol{I}_4$ as covariance matrices).

2 Preliminaries

Let $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$, and $\mathbb{R}_{--}$ denote the sets of positive integers, non-negative integers, real numbers, non-negative real numbers, positive real numbers, and negative real numbers, respectively. For $x, y \in \mathbb{R}$, we will use the notation $x \wedge y := \min(x, y)$. By $\|\boldsymbol{x}\|$ and $\|\boldsymbol{A}\|$, we denote the Euclidean norm of a vector $\boldsymbol{x} \in \mathbb{R}^d$ and the induced matrix norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$, respectively. By $\boldsymbol{I}_d \in \mathbb{R}^{d \times d}$, we denote the $d \times d$ unit matrix. The Borel σ-algebra on $\mathbb{R}$ is denoted by $\mathcal{B}(\mathbb{R})$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ 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 [12, 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 $\mathcal{F}$.

The next proposition is about the existence and uniqueness of a strong solution of the SDE (1.1), see, e.g., Barczy and Pap [1, Proposition 2.1].

2.1 Proposition. *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* $s \leqslant t$,

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

Next we present a result about the first moment of $(Y_t, X_t)_{t \in \mathbb{R}_+}$. For a proof, see, e.g., Barczy and Pap [1, Proposition 2.2] together with (2.1) and Proposition 3.2.10 in Karatzas and Shreve [12].

2.2 Proposition. *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*

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

(2.3)
$$\mathbb{E}(X_t \,|\, \mathcal{F}_s) = X_s + \int_s^t (\alpha - \beta\, \mathbb{E}(Y_u \,|\, \mathcal{F}_s))\, \mathrm{d}u$$

$$= X_s + \alpha(t - s) - \beta Y_s \int_s^t \mathrm{e}^{-b(u-s)}\, \mathrm{d}u - a\beta \int_s^t \left(\int_s^u \mathrm{e}^{-b(u-v)}\, \mathrm{d}v \right) \mathrm{d}u,$$

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 introduce a classification of the Heston model given by the SDE (1.1).

2.3 Definition. *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. [4, Equation (20)], Li and Ma [14, Theorem 2.6], Theorem 3.1 with $\alpha = 2$ and Theorem 4.1 in Barczy et al. [2], or Jin et al. [11, Corollaries 5.9 and 6.4]. Only (2.7) of the following Theorem 2.4 can be considered as a slight improvement of the existing results.

2.4 Theorem. *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*

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

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

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

$$(2.5) \qquad \mathbb{E}(Y_\infty) = \frac{a}{b}, \qquad \mathbb{E}(Y_\infty^2) = \frac{(2a + \sigma_1^2)a}{2b^2}, \qquad \mathbb{E}(Y_\infty^3) = \frac{(2a + \sigma_1^2)(a + \sigma_1^2)a}{2b^3}.$$

(ii) *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.*

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

$$(2.6) \qquad \frac{1}{T} \int_0^T f(Y_s) \, \mathrm{d}s \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } T \to \infty,$$

$$(2.7) \qquad \frac{1}{n} \sum_{i=0}^{n-1} f(Y_i) \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } n \to \infty.$$

Proof. Based on the references given before the theorem, we only need to show (2.7). By Corollary 2.7 in Jin et al. [10], the tail σ-field $\bigcap_{t \in \mathbb{R}_+} \sigma(Y_s, s \geqslant t)$ of $(Y_t)_{t \in \mathbb{R}_+}$ is trivial for any initial distribution, i.e., the tail σ-field in question consists of events having probability 0 or 1 for any initial distribution on $\mathbb{R}_+$. But since the tail σ-field of $(Y_t)_{t \in \mathbb{R}_+}$ is richer than that of $(Y_i)_{i \in \mathbb{Z}_+}$, the tail σ-field of $(Y_i)_{i \in \mathbb{Z}_+}$ is also trivial for any initial distribution.

Denoting the distribution of Y_0 and Y_∞ by ν and μ, respectively, let us introduce the distribution $\eta := (\mu + \nu)/2$. Let us introduce the following processes: $(Z_t)_{t \in \mathbb{R}_+}$, which is the pathwise unique strong solution of the first equation in (1.1) with initial condition $Z_0 = \zeta_0$, where ζ_0 has the distribution μ; and $(U_t)_{t \in \mathbb{R}_+}$, which is the pathwise unique strong solution of the same SDE with initial condition $U_0 = \xi_0$, where ξ_0 has the distribution η.

We use Birkhoff's ergodic theorem (see, e.g., Theorem 8.4.1 in Dudley [6]) in the usual setting: the probability space is $(\mathbb{R}^{\mathbb{Z}_+}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}_+}), \mathcal{L}((Z_i)_{i \in \mathbb{Z}_+}))$, where $\mathcal{L}((Z_i)_{i \in \mathbb{Z}_+})$ denotes the distribution of $(Z_i)_{i \in \mathbb{Z}_+}$, and the measure-preserving transformation T is the shift operator, i.e., $T((x_i)_{i \in \mathbb{Z}_+}) :=$

4

$(x_{i+1})_{i\in\mathbb{Z}_+}$ for $(x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+}$ (the measure preservability follows from (ii)). All invariant sets of T are included in the tail σ-field of the coordinate mappings π_i, $i \in \mathbb{Z}_+$, on $\mathbb{R}^{\mathbb{Z}_+}$, since for any invariant set A we have $A \in \sigma(\pi_0, \pi_1, \ldots)$, but as $T^k(A) = A$ for all $k \in \mathbb{N}$, it is also true that $A \in \sigma(\pi_k, \pi_{k+1}, \ldots)$ for all $k \in \mathbb{N}$. This implies that T is ergodic, since the tail σ-field is trivial. Hence we can apply the ergodic theorem for the function

$$g : \mathbb{R}^{\mathbb{Z}_+} \to \mathbb{R}, \qquad g((x_i)_{i\in\mathbb{Z}_+}) := f(x_0), \qquad (x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+},$$

where f is given in (iii), to obtain

$$\frac{1}{n} \sum_{i=0}^{n-1} f(x_i) \to \int_{\mathbb{R}_+} f(x_0)\, \mu(\mathrm{d}x_0) \qquad \text{as } n \to \infty$$

for almost every $(x_i)_{i\in\mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+}$ with respect to the measure $\mathcal{L}((Z_i)_{i\in\mathbb{Z}_+})$, and consequently

(2.8)
$$\frac{1}{n} \sum_{i=0}^{n-1} f(Z_i) \xrightarrow{\text{a.s.}} \mathbb{E}(f(Y_\infty)) \qquad \text{as } n \to \infty,$$

because, clearly, the distribution of Y_∞ does not depend on the initial distribution. We introduce the following event, which is clearly a tail event of $(Z_i)_{i\in\mathbb{Z}_+}$ and has probability 1 by (2.8):

$$C_Z := \left\{ \omega \in \Omega : \frac{1}{n} \sum_{i=0}^{n-1} f(Z_i(\omega)) \to \mathbb{E}(f(Y_\infty)) \ \text{as } n \to \infty \right\}.$$

The events C_Y and C_U are defined in a similar way and are clearly tail events of $(Y_i)_{i\in\mathbb{Z}_+}$ and $(U_i)_{i\in\mathbb{Z}_+}$, respectively. Clearly,

$$\mathbb{P}(C_U) = \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\eta(x) = \frac{1}{2} \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\mu(x) + \frac{1}{2} \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\nu(x)$$

$$\geqslant \frac{1}{2} \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\mu(x) = \frac{1}{2} \int_0^\infty \mathbb{P}(C_Z \,|\, Z_0 = x)\, \mathrm{d}\mu(x)$$

$$= \frac{1}{2} \mathbb{P}(C_Z) = \frac{1}{2}.$$

Here we used that $\mathbb{P}(C_U \,|\, U_0 = x) = \mathbb{P}(C_Z \,|\, Z_0 = x)$ μ-a.e. $x \in \mathbb{R}_+$, since the conditional probabilities on both sides depend only on the transition probability kernel of the CIR process given by the first SDE of (1.1) irrespective of the initial distribution. Further, we note that $\mathbb{P}(C_U \,|\, U_0 = x)$ is defined uniquely only η-a.e. $x \in \mathbb{R}_+$, but, by the definition of η, this means both μ-a.e. $x \in \mathbb{R}_+$, and ν-a.e. $x \in \mathbb{R}_+$, and similarly $\mathbb{P}(C_Z \,|\, Z_0 = x)$ is defined μ-a.e. $x \in \mathbb{R}_+$, so our equalities are valid. Thus, we have $\mathbb{P}(C_U) \geqslant \frac{1}{2}$. But since C_U is a tail event of $(U_i)_{i\in\mathbb{Z}_+}$, its probability must be either 0 or 1 (since the tail σ-field is trivial), hence $\mathbb{P}(C_U) = 1$. Hence

$$2 = \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\mu(x) + \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\nu(x) \leqslant \mu([0, \infty)) + \nu([0, \infty)) = 2,$$

yielding that

$$\int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\mu(x) = \int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\, \mathrm{d}\nu(x) = 1,$$

and the second equality is exactly (2.7) after we note that, by the same argument as above,

$$\int_0^\infty \mathbb{P}(C_U \,|\, U_0 = x)\,\mathrm{d}\nu(x) = \int_0^\infty \mathbb{P}(C_Y \,|\, Y_0 = x)\,\mathrm{d}\nu(x) = \mathbb{P}(C_Y).$$

With this our proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

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

First, we recall a strong law of large numbers for discrete time square-integrable martingales.

2.5 Theorem. (Shiryaev [16, Chapter VII, Section 5, Theorem 4]) *Let* $\big(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{N}}, \mathbb{P}\big)$ *be a filtered probability space. Let* $(M_n)_{n\in\mathbb{N}}$ *be a square-integrable martingale with respect to the filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *such that* $\mathbb{P}(M_0 = 0) = 1$ *and* $\mathbb{P}(\lim_{n\to\infty}\langle M\rangle_n = \infty) = 1$, *where* $(\langle M\rangle_n)_{n\in\mathbb{N}}$ *denotes the predictable quadratic variation process of* M. *Then*

$$\frac{M_n}{\langle M\rangle_n} \xrightarrow{\text{a.s.}} 0 \qquad as \ \ n \to \infty.$$

Next, we recall a martingale central limit theorem in discrete time.

2.6 Theorem. (Jacod and Shiryaev [9, Chapter VIII, Theorem 3.33]) *Let* $\{(\boldsymbol{M}_{n,k}, \mathcal{F}_{n,k}) : k = 0, 1, \ldots, k_n\}_{n\in\mathbb{N}}$ *be a sequence of* d-*dimensional square-integrable martingales with* $\boldsymbol{M}_{n,0} = \boldsymbol{0}$ *such that there exists some symmetric, positive semi-definite non-random matrix* $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ *such that*

$$\sum_{k=1}^{k_n} \mathbb{E}\big((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^\top \,|\, \mathcal{F}_{n,k-1}\big) \xrightarrow{\mathbb{P}} \boldsymbol{D} \qquad as \ \ n \to \infty,$$

and for all $\varepsilon \in \mathbb{R}_{++}$,

$$(2.9) \qquad \sum_{k=1}^{k_n} \mathbb{E}\big(\|\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1}\|^2 \mathbb{1}_{\{\|\boldsymbol{M}_{n,k}-\boldsymbol{M}_{n,k-1}\|\geqslant\varepsilon\}} \,|\, \mathcal{F}_{n,k-1}\big) \xrightarrow{\mathbb{P}} 0 \qquad as \ \ n \to \infty.$$

Then

$$\sum_{k=1}^{k_n} (\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1}) = \boldsymbol{M}_{n,k_n} \xrightarrow{\mathcal{L}} \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{D}) \qquad as \ \ n \to \infty,$$

where $\mathcal{N}_d(\boldsymbol{0}, \boldsymbol{D})$ *denotes a* d-*dimensional normal distribution with mean vector* $\boldsymbol{0}$ *and covariance matrix* $\boldsymbol{D}$.

In all the remaining sections, we will consider the subcritical Heston model (1.1) with a non-random initial value $(y_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}$. Note that 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).

3 CLSE based on discrete time observations

Using (2.2) and (2.3), by an easy calculation, for all $i \in \mathbb{N}$,

$$(3.1) \qquad \mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \,\middle|\, \mathcal{F}_{i-1}\right) = \begin{bmatrix} \mathrm{e}^{-b} & 0 \\ -\beta \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} Y_{i-1} \\ X_{i-1} \end{bmatrix} + \begin{bmatrix} \int_0^1 \mathrm{e}^{-bu}\,\mathrm{d}u & 0 \\ -\beta \int_0^1 \big(\int_0^u \mathrm{e}^{-bv}\,\mathrm{d}v\big)\,\mathrm{d}u & 1 \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix}.$$

6

Using that $\sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1}) \subseteq \mathcal{F}_{i-1}$, $i \in \mathbb{N}$, by tower rule for conditional expectations, we have

$$\mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \middle| \sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1})\right) = \mathbb{E}\left(\mathbb{E}\left(\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \middle| \mathcal{F}_{i-1}\right) \middle| \sigma(X_1, Y_1, \ldots, X_{i-1}, Y_{i-1})\right)$$

$$= \begin{bmatrix} e^{-b} & 0 \\ -\beta \int_0^1 e^{-bu}\,du & 1 \end{bmatrix} \begin{bmatrix} Y_{i-1} \\ X_{i-1} \end{bmatrix} + \begin{bmatrix} \int_0^1 e^{-bu}\,du & 0 \\ -\beta \int_0^1 \left(\int_0^u e^{-bv}\,dv\right) du & 1 \end{bmatrix} \begin{bmatrix} a \\ \alpha \end{bmatrix}, \qquad i \in \mathbb{N},$$

and hence a CLSE of (a, b, α, β) based on discrete time observations $(Y_i, X_i)_{i \in \{1, \ldots, n\}}$ could be obtained by solving the extremum problem

$$(3.2) \qquad \underset{(a,b,\alpha,\beta) \in \mathbb{R}^4}{\arg\min} \sum_{i=1}^{n} \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2\right],$$

where

$$d := d(b) := e^{-b}, \qquad c := c(a, b) := a \int_0^1 e^{-bu}\,du,$$

$$(3.3)$$

$$\delta := \delta(b, \beta) := -\beta \int_0^1 e^{-bu}\,du, \qquad \gamma := \gamma(a, b, \alpha, \beta) := \alpha - a\beta \int_0^1 \left(\int_0^u e^{-bv}\,dv\right) du.$$

First, we determine the CLSE of (c, d, γ, δ) by minimizing the sum on the right hand side of (3.2) with respect to $(c, d, \gamma, \delta) \in \mathbb{R}^4$.

We get

$$(3.4) \qquad \begin{bmatrix} \widehat{c}_n^{\mathrm{CLSE}} \\ \widehat{d}_n^{\mathrm{CLSE}} \\ \widehat{\gamma}_n^{\mathrm{CLSE}} \\ \widehat{\delta}_n^{\mathrm{CLSE}} \end{bmatrix} = \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} n & \sum_{i=1}^{n} Y_{i-1} \\ \sum_{i=1}^{n} Y_{i-1} & \sum_{i=1}^{n} Y_{i-1}^2 \end{bmatrix}^{-1}\right) \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} Y_i Y_{i-1} \\ X_n - x_0 \\ \sum_{i=1}^{n} (X_i - X_{i-1}) Y_{i-1} \end{bmatrix}$$

provided that $n \sum_{i=1}^{n} Y_{i-1}^2 > \left(\sum_{i=1}^{n} Y_{i-1}\right)^2$, where $\otimes$ denotes Kronecker product of matrices. Indeed, with the notation

$$f(c, d, \gamma, \delta) := \sum_{i=1}^{n} \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2\right], \qquad (c, d, \gamma, \delta) \in \mathbb{R}^4,$$

we have

$$\frac{\partial f}{\partial c}(c, d, \gamma, \delta) = -2 \sum_{i=1}^{n} (Y_i - dY_{i-1} - c),$$

$$\frac{\partial f}{\partial d}(c, d, \gamma, \delta) = -2 \sum_{i=1}^{n} Y_{i-1}(Y_i - dY_{i-1} - c),$$

$$\frac{\partial f}{\partial \gamma}(c, d, \gamma, \delta) = -2 \sum_{i=1}^{n} (X_i - X_{i-1} - \gamma - \delta Y_{i-1}),$$

$$\frac{\partial f}{\partial \delta}(c, d, \gamma, \delta) = -2 \sum_{i=1}^{n} Y_{i-1}(X_i - X_{i-1} - \gamma - \delta Y_{i-1}).$$

Hence the system of equations consisting of the first order partial derivates of f being equal to 0 takes the form

$$\left(I_2 \otimes \begin{bmatrix} n & \sum_{i=1}^n Y_{i-1} \\ \sum_{i=1}^n Y_{i-1} & \sum_{i=1}^n Y_{i-1}^2 \end{bmatrix}\right) \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_{i-1}Y_i \\ X_n - x_0 \\ \sum_{i=1}^n (X_i - X_{i-1})Y_{i-1} \end{bmatrix}.$$

This implies (3.4), since the 4×4-matrix consisting of the second order partial derivatives of f having the form

$$2I_2 \otimes \begin{bmatrix} n & \sum_{i=1}^n Y_{i-1} \\ \sum_{i=1}^n Y_{i-1} & \sum_{i=1}^n Y_{i-1}^2 \end{bmatrix}$$

is positive definite provided that $n \sum_{i=1}^n Y_{i-1}^2 > (\sum_{i=1}^n Y_{i-1})^2$. In fact, it turned out that for the calculation of the CLSE of (c, d, γ, δ), one does not need to know the values of the parameters σ_1, σ_2 and ϱ.

The next lemma assures the unique existence of the CLSE of (c, d, γ, δ) based on discrete time observations.

3.1 Lemma. *If $a \in \mathbb{R}_{++}$, $b \in \mathbb{R}$, $\sigma_1 \in \mathbb{R}_{++}$, and $Y_0 = y_0 \in \mathbb{R}_+$, then for all $n \geqslant 2$, $n \in \mathbb{N}$, we have*

$$\mathbb{P}\left(n \sum_{i=1}^n Y_{i-1}^2 > \left(\sum_{i=1}^n Y_{i-1}\right)^2\right) = 1,$$

and hence, supposing also that $\alpha, \beta \in \mathbb{R}$, $\sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$, there exists a unique CLSE $(\widehat{c}_n^{\mathrm{CLSE}}, \widehat{d}_n^{\mathrm{CLSE}}, \widehat{\gamma}_n^{\mathrm{CLSE}}, \widehat{\delta}_n^{\mathrm{CLSE}})$ of (c, d, γ, δ) which has the form given in (3.4).

Proof. By an easy calculation,

$$n \sum_{i=1}^n Y_{i-1}^2 - \left(\sum_{i=1}^n Y_{i-1}\right)^2 = n \sum_{i=1}^n \left(Y_{i-1} - \frac{1}{n} \sum_{j=1}^n Y_{j-1}\right)^2 \geqslant 0,$$

and equality holds if and only if

$$Y_{i-1} = \frac{1}{n} \sum_{j=1}^n Y_{j-1}, \qquad i = 1, \ldots, n \qquad \Longleftrightarrow \qquad Y_0 = Y_1 = \cdots = Y_{n-1}.$$

Then, for all $n \geqslant 2$,

$$\mathbb{P}(Y_0 = Y_1 = \cdots = Y_{n-1}) \leqslant \mathbb{P}(Y_0 = Y_1) = \mathbb{P}(Y_1 = y_0) = 0,$$

since the law of Y_1 is absolutely continuous, see, e.g., Cox et al. [4, formula 18]. $\square$

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

Next, we describe the asymptotic behaviour of the CLSE of (c, d, γ, δ).

3.2 Theorem. *If* $a, b \in \mathbb{R}_{++}$, $\alpha, \beta \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{++}$, $\varrho \in (-1, 1)$ *and* $(Y_0, X_0) = (y_0, x_0) \in \mathbb{R}_{++} \times \mathbb{R}$, *then the CLSE* $(\widehat{c}_n^{\mathrm{CLSE}}, \widehat{d}_n^{\mathrm{CLSE}}, \widehat{\gamma}_n^{\mathrm{CLSE}}, \widehat{\delta}_n^{\mathrm{CLSE}})$ *of* (c, d, γ, δ) *given in* (3.4) *is strongly consistent and asymptotically normal, i.e.,*

$$(\widehat{c}_n^{\mathrm{CLSE}}, \widehat{d}_n^{\mathrm{CLSE}}, \widehat{\gamma}_n^{\mathrm{CLSE}}, \widehat{\delta}_n^{\mathrm{CLSE}}) \xrightarrow{\text{a.s.}} (c, d, \gamma, \delta) \qquad as \ \ n \to \infty,$$

and

$$\sqrt{n} \begin{bmatrix} \widehat{c}_n^{\mathrm{CLSE}} - c \\ \widehat{d}_n^{\mathrm{CLSE}} - d \\ \widehat{\gamma}_n^{\mathrm{CLSE}} - \gamma \\ \widehat{\delta}_n^{\mathrm{CLSE}} - \delta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4(\mathbf{0}, \boldsymbol{E}) \qquad as \ \ n \to \infty,$$

with some explicitly given symmetric, positive definite matrix $\boldsymbol{E} \in \mathbb{R}^{2 \times 2}$ *given in* (3.14).

Proof. By (3.4), we get

$$
\begin{aligned}
\begin{bmatrix} \widehat{c}_n^{\mathrm{CLSE}} \\ \widehat{d}_n^{\mathrm{CLSE}} \end{bmatrix} &= \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} Y_i \right) \\[2mm]
&= \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right) \begin{bmatrix} c \\ d \end{bmatrix} \\[2mm]
&\quad + \left(\sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} (Y_i - c - d Y_{i-1}) \\[2mm]
&= \begin{bmatrix} c \\ d \end{bmatrix} + \left(\frac{1}{n} \sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \varepsilon_i,
\end{aligned}
$$

(3.5)

where $\varepsilon_i := Y_i - c - d Y_{i-1}$, $i \in \mathbb{N}$, provided that $n \sum_{i=1}^n Y_{i-1}^2 > \left(\sum_{i=1}^n Y_{i-1} \right)^2$. By (3.1) and (3.3), $\mathbb{E}(Y_i \mid \mathcal{F}_{i-1}) = d Y_{i-1} + c$, $i \in \mathbb{N}$, and hence $(\varepsilon_i)_{i \in \mathbb{N}}$ is a sequence of martingale differences with respect to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}_+}$. By (2.1), we have

$$Y_i = \mathrm{e}^{-b} Y_{i-1} + a \int_{i-1}^i \mathrm{e}^{-b(i-u)} \, \mathrm{d}u + \sigma_1 \int_{i-1}^i \mathrm{e}^{-b(i-u)} \sqrt{Y_u} \, \mathrm{d}W_u$$

$$= d Y_{i-1} + c + \sigma_1 \int_{i-1}^i \mathrm{e}^{-b(i-u)} \sqrt{Y_u} \, \mathrm{d}W_u, \qquad i \in \mathbb{N},$$

hence, by Proposition 3.2.10 in Karatzas and Shreve [12] and (2.2), we have

$$
\begin{aligned}
\mathbb{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1}) &= \sigma_1^2 \, \mathbb{E}\left(\left(\int_{i-1}^i \mathrm{e}^{-b(i-u)} \sqrt{Y_u} \, \mathrm{d}W_u \right)^2 \, \Big| \, \mathcal{F}_{i-1} \right) = \sigma_1^2 \int_{i-1}^i \mathrm{e}^{-2b(i-u)} \, \mathbb{E}(Y_u \mid \mathcal{F}_{i-1}) \, \mathrm{d}u \\[2mm]
&= \sigma_1^2 \int_{i-1}^i \mathrm{e}^{-2b(i-u)} \mathrm{e}^{-b(u-i+1)} Y_{i-1} \, \mathrm{d}u + \sigma_1^2 \int_{i-1}^i \mathrm{e}^{-2b(i-u)} a \int_{i-1}^u \mathrm{e}^{-b(u-v)} \, \mathrm{d}v \, \mathrm{d}u \\[2mm]
&= \sigma_1^2 Y_{i-1} \int_0^1 \mathrm{e}^{-b(2-v)} \, \mathrm{d}v + \sigma_1^2 a \int_0^1 \int_0^u \mathrm{e}^{-b(2-v-u)} \, \mathrm{d}v \, \mathrm{d}u =: C_1 Y_{i-1} + C_2.
\end{aligned}
$$

9

Now we apply Theorem 2.5 to the square-integrable martingale $M_n^{(c)} := \sum_{i=1}^{n} \varepsilon_i$, $n \in \mathbb{N}$, which has predictable quadratic variation process $\langle M^{(c)} \rangle_n = \sum_{i=1}^{n} \mathbb{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1}) = C_1 \sum_{i=1}^{n} Y_{i-1} + C_2 n$, $n \in \mathbb{N}$, see, e.g., Shiryaev [16, Chapter VII, Section 1, formula (15)]. By (2.5) and (2.7),

$$\frac{\langle M^{(c)} \rangle_n}{n} \xrightarrow{\text{a.s.}} C_1 \mathbb{E}(Y_\infty) + C_2 \qquad \text{as } n \to \infty,$$

and since $C_1, C_2 \in \mathbb{R}_{++}$, $\langle M^{(c)} \rangle_n \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$. Hence, by Theorem 2.5,

$$(3.6) \qquad \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i = \frac{M_n^{(c)}}{\langle M^{(c)} \rangle_n} \frac{\langle M^{(c)} \rangle_n}{n} \xrightarrow{\text{a.s.}} 0 \cdot (C_1 \mathbb{E}(Y_\infty) + C_2) = 0 \qquad \text{as } n \to \infty.$$

Similarly,

$$\mathbb{E}(Y_{i-1}^2 \varepsilon_i^2 \mid \mathcal{F}_{i-1}) = Y_{i-1}^2 \mathbb{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1}) = C_1 Y_{i-1}^3 + C_2 Y_{i-1}^2, \qquad i \in \mathbb{N},$$

and, by essentially the same reasoning as before, $\frac{1}{n} \sum_{i=1}^{n} Y_{i-1} \varepsilon_i \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. By (2.5) and (2.7),

$$(3.7) \qquad \left(\frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} = \begin{bmatrix} 1 & \frac{1}{n} \sum_{i=1}^{n} Y_{i-1} \\ \frac{1}{n} \sum_{i=1}^{n} Y_{i-1} & \frac{1}{n} \sum_{i=1}^{n} Y_{i-1}^2 \end{bmatrix}^{-1} \xrightarrow{\text{a.s.}} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}$$

as $n \to \infty$, where we used that $\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2 = \frac{a\sigma_1^2}{2b^2} \in \mathbb{R}_{++}$, and consequently, the limit is indeed non-singular. Thus, by (3.5), $(\widehat{c}_n^{\mathrm{CLSE}}, \widehat{d}_n^{\mathrm{CLSE}}) \xrightarrow{\text{a.s.}} (c, d)$ as $n \to \infty$.

Further, by (3.4),

$$(3.8)$$
$$\begin{bmatrix} \widehat{\gamma}_n^{\mathrm{CLSE}} \\ \widehat{\delta}_n^{\mathrm{CLSE}} \end{bmatrix} = \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} (X_i - X_{i-1}) \right)$$

$$= \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right) \begin{bmatrix} \gamma \\ \delta \end{bmatrix}$$

$$+ \left(\sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} (X_i - X_{i-1} - \gamma - \delta Y_{i-1})$$

$$= \begin{bmatrix} \gamma \\ \delta \end{bmatrix} + \left(\frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^\top \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \eta_i,$$

where $\eta_i := X_i - X_{i-1} - \gamma - \delta Y_{i-1}$, $i \in \mathbb{N}$, provided that $n \sum_{i=1}^{n} Y_{i-1}^2 > (\sum_{i=1}^{n} Y_{i-1})^2$. By (3.1) and (3.3), $\mathbb{E}(X_i \mid \mathcal{F}_{i-1}) = X_{i-1} + \delta Y_{i-1} + \gamma$, $i \in \mathbb{N}$, and hence $(\eta_i)_{i \in \mathbb{N}}$ is a sequence of martingale differences with respect to the filtration $(\mathcal{F}_i)_{i \in \mathbb{Z}_+}$. By (2.1) and (2.2), with the notation

10

$\widetilde{W}_t := \varrho W_t + \sqrt{1-\varrho^2} B_t, \ t \in \mathbb{R}_+, \ $ we compute

$$X_i - X_{i-1} = \int_{i-1}^i (\alpha - \beta Y_u)\, \mathrm{d}u + \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u = \alpha - \beta \int_{i-1}^i Y_u\, \mathrm{d}u + \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u$$

$$= \alpha - \beta \int_{i-1}^i \left(\mathrm{e}^{-b(u-(i-1))} Y_{i-1} + a \int_{i-1}^u \mathrm{e}^{-b(u-v)}\, \mathrm{d}v + \sigma_1 \int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v \right) \mathrm{d}u$$

$$+ \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u$$

$$= \alpha - \beta Y_{i-1} \int_{i-1}^i \mathrm{e}^{-b(u-i+1)}\, \mathrm{d}u - a\beta \int_{i-1}^i \left(\int_{i-1}^u \mathrm{e}^{-b(u-v)}\, \mathrm{d}v \right) \mathrm{d}u$$

$$- \beta\sigma_1 \int_{i-1}^i \left(\int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v \right) \mathrm{d}u + \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u$$

$$= \alpha - \beta Y_{i-1} \int_0^1 \mathrm{e}^{-bv}\, \mathrm{d}v - a\beta \int_0^1 \left(\int_0^u \mathrm{e}^{-bv}\, \mathrm{d}v \right) \mathrm{d}u$$

$$- \beta\sigma_1 \int_{i-1}^i \left(\int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v \right) \mathrm{d}u + \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u$$

$$= \delta Y_{i-1} + \gamma - \beta\sigma_1 \int_{i-1}^i \left(\int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v \right) \mathrm{d}u + \sigma_2 \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u,$$

and consequently,

$$\mathbb{E}(\eta_i^2 \,|\, \mathcal{F}_{i-1}) = \beta^2 \sigma_1^2\, \mathbb{E}\left[\left(\int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v\, \mathrm{d}u \right)^2 \Bigg| \mathcal{F}_{i-1} \right] + \sigma_2^2\, \mathbb{E}\left[\left(\int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}\widetilde{W}_u \right)^2 \Bigg| \mathcal{F}_{i-1} \right]$$

$$- 2\beta\sigma_1\sigma_2\, \mathbb{E}\left[\left(\int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v\, \mathrm{d}u \right) \left(\varrho \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}W_u \right) \Bigg| \mathcal{F}_{i-1} \right]$$

$$- 2\beta\sigma_1\sigma_2\, \mathbb{E}\left[\left(\int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-v)} \sqrt{Y_v}\, \mathrm{d}W_v\, \mathrm{d}u \right) \left(\sqrt{1-\varrho^2} \int_{i-1}^i \sqrt{Y_u}\, \mathrm{d}B_u \right) \Bigg| \mathcal{F}_{i-1} \right].$$

We use Equation (3.2.23) from Karatzas and Shreve [12] to the first, second and third terms, and Proposition 3.2.17 from Karatzas and Shreve [12] to the fourth term (together with the independence of W and B):

$$\mathbb{E}(\eta_i^2 | \mathcal{F}_{i-1}) = \beta^2 \sigma_1^2 \int_{i-1}^i \int_{i-1}^i \mathbb{E}\left(\int_{i-1}^u \mathrm{e}^{-b(u-w)} \sqrt{Y_w}\, \mathrm{d}W_w \int_{i-1}^v \mathrm{e}^{-b(v-w)} \sqrt{Y_w}\, \mathrm{d}W_w \,\Big|\, \mathcal{F}_{i-1} \right) \mathrm{d}v\, \mathrm{d}u$$

$$+ \sigma_2^2 \int_{i-1}^i \mathbb{E}(Y_u \,|\, \mathcal{F}_{i-1})\, \mathrm{d}u$$

$$- 2\beta\sigma_1\sigma_2\varrho \int_{i-1}^i \mathbb{E}\left(\int_{i-1}^u \mathrm{e}^{-b(u-w)} \sqrt{Y_w}\, \mathrm{d}W_w \int_{i-1}^i \sqrt{Y_w}\, \mathrm{d}W_w \,\Big|\, \mathcal{F}_{i-1} \right) \mathrm{d}u - 0$$

$$= \beta^2 \sigma_1^2 \int_{i-1}^i \int_{i-1}^i \int_{i-1}^{u \wedge v} \mathrm{e}^{-b(u+v-2w)} \mathbb{E}(Y_w \,|\, \mathcal{F}_{i-1})\, \mathrm{d}w\, \mathrm{d}u\, \mathrm{d}v + \sigma_2^2 \int_{i-1}^i \mathbb{E}(Y_u \,|\, \mathcal{F}_{i-1})\, \mathrm{d}u$$

$$- 2\beta\sigma_1\sigma_2\varrho \int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-v)} \mathbb{E}(Y_v \,|\, \mathcal{F}_{i-1})\, \mathrm{d}v\, \mathrm{d}u.$$

11

Using again (2.2), we get

$$\mathbb{E}(\eta_i^2|\mathcal{F}_{i-1}) = \beta^2 \sigma_1^2 Y_{i-1} \int_{i-1}^i \int_{i-1}^i \int_{i-1}^{u \wedge v} \mathrm{e}^{-b(u+v-w-(i-1))} \, \mathrm{d}w \, \mathrm{d}v \, \mathrm{d}u$$

$$+ a\beta^2 \sigma_1^2 \int_{i-1}^i \int_{i-1}^i \int_{i-1}^{u \wedge v} \int_{i-1}^w \mathrm{e}^{-b(u+v-w-z)} \, \mathrm{d}z \, \mathrm{d}w \, \mathrm{d}v \, \mathrm{d}u + \sigma_2^2 Y_{i-1} \int_{i-1}^i \mathrm{e}^{-b(u-(i-1))} \, \mathrm{d}u$$

$$+ a\sigma_2^2 \int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-v)} \, \mathrm{d}v \, \mathrm{d}u - 2\beta\sigma_1\sigma_2\varrho Y_{i-1} \int_{i-1}^i \int_{i-1}^u \mathrm{e}^{-b(u-(i-1))} \, \mathrm{d}v \, \mathrm{d}u$$

$$- 2a\beta\sigma_1\sigma_2\varrho \int_{i-1}^i \int_{i-1}^u \int_{i-1}^v \mathrm{e}^{-b(u-w)} \, \mathrm{d}w \, \mathrm{d}v \, \mathrm{d}u$$

$$= \left(\beta^2 \sigma_1^2 \int_0^1 \int_0^1 \int_0^{u' \wedge v'} \mathrm{e}^{-b(u'+v'-w')} \, \mathrm{d}w' \, \mathrm{d}v' \, \mathrm{d}u' - 2\beta\sigma_1\sigma_2\varrho \int_0^1 \int_0^{u'} \mathrm{e}^{-bu'} \, \mathrm{d}v' \, \mathrm{d}u' + \sigma_2^2 \int_0^1 \mathrm{e}^{-bu'} \, \mathrm{d}u' \right) Y_{i-1}$$

$$+ a\beta^2 \sigma_1^2 \int_0^1 \int_0^1 \int_0^{u' \wedge v'} \int_0^{w'} \mathrm{e}^{-b(u'+v'-w'-z')} \, \mathrm{d}z' \, \mathrm{d}w' \, \mathrm{d}v' \, \mathrm{d}u'$$

$$+ a\sigma_2^2 \int_0^1 \int_0^{u'} \mathrm{e}^{-b(u'-v')} \, \mathrm{d}v' \, \mathrm{d}u' - 2a\beta\sigma_1\sigma_2\varrho \int_0^1 \int_0^{u'} \int_0^{v'} \mathrm{e}^{-b(u'-w')} \, \mathrm{d}w' \, \mathrm{d}v' \, \mathrm{d}u' =: C_3 Y_{i-1} + C_4.$$

Now we apply Theorem 2.5 to the square-integrable martingale $M_n^{(\gamma)} := \sum_{i=1}^n \eta_i, \; n \in \mathbb{N}$, which has predictable quadratic variation process $\langle M^{(\gamma)} \rangle_n = \sum_{i=1}^n \mathbb{E}(\eta_i^2 \mid \mathcal{F}_{i-1}) = C_3 \sum_{i=1}^n Y_{i-1} + C_4 n, \; n \in \mathbb{N}$. By (2.7),

$$(3.9) \qquad \qquad \frac{\langle M^{(\gamma)} \rangle_n}{n} \xrightarrow{\text{a.s.}} C_3 \, \mathbb{E}(Y_\infty) + C_4 \qquad \text{as } \; n \to \infty.$$

Note that $C_3 \geqslant 0$ and $C_4 \geqslant 0$, since $\mathbb{E}(\eta_1^2 \mid \mathcal{F}_0) = C_3 y_0 + C_4 \geqslant 0$ for all $y_0 \in \mathbb{R}_+$. By setting $y_0 = 0$, we can see that $C_4 \geqslant 0$, and then, by taking the limit $y_0 \to \infty$ on the right-hand side of the inequality $C_3 \geqslant -\frac{C_4}{y_0}, \; y_0 > 0$, we get $C_3 \geqslant 0$ as well. Note also that $\langle M^{(\gamma)} \rangle_n \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$ provided that $C_3 + C_4 > 0$. If $C_3 = 0$ and $C_4 = 0$, then $\mathbb{E}(\eta_i^2 \mid \mathcal{F}_{i-1}) = 0, \; i \in \mathbb{N}$, and consequently $\mathbb{E}(\eta_i^2) = 0, \; i \in \mathbb{N}$, and, since $\mathbb{E}(\eta_i) = 0, \; i \in \mathbb{N}$, we have $\mathbb{P}(\eta_i = 0) = 1, \; i \in \mathbb{N}$, implying that $\mathbb{P}(\sum_{i=1}^n \eta_i = 0) = 1$ and $\mathbb{P}(\sum_{i=1}^n Y_{i-1}\eta_i = 0) = 1, \; n \in \mathbb{N}$, i.e., in this case, by (3.8), $(\widehat{\gamma}_n^{\text{CLSE}}, \widehat{\delta}_n^{\text{CLSE}}) = (\gamma, \delta), \; n \in \mathbb{N}$, almost surely. If $C_3 + C_4 > 0$, then, by Theorem 2.5,

$$(3.10) \qquad \frac{1}{n} \sum_{i=1}^n \eta_i = \frac{M_n^{(\gamma)}}{\langle M^{(\gamma)} \rangle_n} \frac{\langle M^{(\gamma)} \rangle_n}{n} \xrightarrow{\text{a.s.}} 0 \cdot (C_3 \, \mathbb{E}(Y_\infty) + C_4) = 0 \qquad \text{as } \; n \to \infty.$$

Similarly,

$$\mathbb{E}(Y_{i-1}^2 \eta_i^2 \mid \mathcal{F}_{i-1}) = Y_{i-1}^2 \, \mathbb{E}(\eta_i^2 \mid \mathcal{F}_{i-1}) = C_3 Y_{i-1}^3 + C_4 Y_{i-1}^2, \qquad i \in \mathbb{N},$$

and, by essentially the same reasoning as before, $\frac{1}{n} \sum_{i=1}^n Y_{i-1}\eta_i \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$ (in the case $C_3 + C_4 > 0$). Using (3.7) and (3.8), we have $(\widehat{\gamma}_n^{\text{CLSE}}, \widehat{\delta}_n^{\text{CLSE}}) \xrightarrow{\text{a.s.}} (\gamma, \delta)$ as $n \to \infty$.

Since the intersection of two events having probability 1 is an event having probability 1, we get $(\widehat{c}_n^{\text{CLSE}}, \widehat{d}_n^{\text{CLSE}}, \widehat{\gamma}_n^{\text{CLSE}}, \widehat{\delta}_n^{\text{CLSE}}) \xrightarrow{\text{a.s.}} (c, d, \gamma, \delta)$ as $n \to \infty$, as desired.

Next, we turn to prove that the CLSE of (c, d, γ, δ) is asymptotically normal. First, using (3.5) and (3.8), we can write

$$
(3.11) \qquad \sqrt{n}
\begin{bmatrix}
\widehat{c}_n^{\mathrm{CLSE}} - c \\
\widehat{d}_n^{\mathrm{CLSE}} - d \\
\widehat{\gamma}_n^{\mathrm{CLSE}} - \gamma \\
\widehat{\delta}_n^{\mathrm{CLSE}} - \delta
\end{bmatrix}
= \left(\boldsymbol{I}_2 \otimes \left(n^{-1} \sum_{i=1}^{n} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}^{\top} \right)^{-1} \right) n^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix},
$$

provided that $n \sum_{i=1}^{n} Y_{i-1}^2 > \left(\sum_{i=1}^{n} Y_{i-1} \right)^2$. By (3.7), the first factor converges almost surely to

$$
\boldsymbol{I}_2 \otimes
\begin{bmatrix}
1 & \mathbb{E}(Y_\infty) \\
\mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2)
\end{bmatrix}^{-1}
\qquad \text{as } n \to \infty.
$$

For the second factor, we are going to apply the martingale central limit theorem (see Theorem 2.6) with the following choices: $d = 4$, $k_n = n$, $n \in \mathbb{N}$, $\mathcal{F}_{n,k} = \mathcal{F}_k$, $n \in \mathbb{N}$, $k \in \{1, \ldots, n\}$, and

$$
\boldsymbol{M}_{n,k} = n^{-\frac{1}{2}} \sum_{i=1}^{k} \begin{bmatrix} \varepsilon_i \\ \eta_i \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{i-1} \end{bmatrix}, \qquad n \in \mathbb{N}, \quad k \in \{1, \ldots, n\}.
$$

Then, applying the identities $(\boldsymbol{A}_1 \otimes \boldsymbol{A}_2)^{\top} = \boldsymbol{A}_1^{\top} \otimes \boldsymbol{A}_2^{\top}$ and $(\boldsymbol{A}_1 \otimes \boldsymbol{A}_2)(\boldsymbol{A}_3 \otimes \boldsymbol{A}_4) = (\boldsymbol{A}_1 \boldsymbol{A}_3) \otimes (\boldsymbol{A}_2 \boldsymbol{A}_4)$,

$$
\mathbb{E}\left((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^{\top} \,\big|\, \mathcal{F}_{n,k-1} \right)
$$

$$
= \frac{1}{n} \mathbb{E}\left(\left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \right) \left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \right)^{\top} \,\bigg|\, \mathcal{F}_{k-1} \right)
$$

$$
= \frac{1}{n} \mathbb{E}\left(\left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix}^{\top} \right) \otimes \left(\begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix}^{\top} \right) \,\bigg|\, \mathcal{F}_{k-1} \right)
$$

$$
= \frac{1}{n} \mathbb{E}\left(\begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix}^{\top} \,\bigg|\, \mathcal{F}_{k-1} \right) \otimes \left(\begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix}^{\top} \right), \qquad n \in \mathbb{N}, \quad k \in \{1, \ldots, n\}.
$$

Since $\mathbb{E}(\varepsilon_k^2 \,|\, \mathcal{F}_{k-1}) = C_1 Y_{k-1} + C_2$, $k \in \mathbb{N}$, and $\mathbb{E}(\eta_k^2 \,|\, \mathcal{F}_{k-1}) = C_3 Y_{k-1} + C_4$, $k \in \mathbb{N}$, it remains to calculate

$$
\mathbb{E}(\varepsilon_k \eta_k \,|\, \mathcal{F}_{k-1}) = \mathbb{E}\left((Y_k - c - d Y_{k-1})(X_k - X_{k-1} - \gamma - \delta Y_{k-1}) \,\big|\, \mathcal{F}_{k-1} \right)
$$

$$
= \mathbb{E}\left(\sigma_1 \int_{k-1}^{k} \mathrm{e}^{-b(k-s)} \sqrt{Y_s} \, \mathrm{d}W_s \left(-\beta \sigma_1 \int_{k-1}^{k} \int_{k-1}^{u} \mathrm{e}^{-b(u-v)} \sqrt{Y_v} \, \mathrm{d}W_v \, \mathrm{d}u + \sigma_2 \int_{k-1}^{k} \sqrt{Y_u} \, \mathrm{d}\widetilde{W}_u \right) \,\bigg|\, \mathcal{F}_{k-1} \right)
$$

$$
= -\beta \sigma_1^2 \int_{k-1}^{k} \mathbb{E}\left(\int_{k-1}^{k} \mathrm{e}^{-b(k-s)} \sqrt{Y_s} \, \mathrm{d}W_s \int_{k-1}^{u} \mathrm{e}^{-b(u-v)} \sqrt{Y_v} \, \mathrm{d}W_v \,\bigg|\, \mathcal{F}_{k-1} \right) \mathrm{d}u
$$

$$
+ \sigma_1 \sigma_2 \, \mathbb{E}\left(\int_{k-1}^{k} \mathrm{e}^{-b(k-s)} \sqrt{Y_s} \, \mathrm{d}W_s \int_{k-1}^{k} \sqrt{Y_u} \, \mathrm{d}\widetilde{W}_u \,\bigg|\, \mathcal{F}_{k-1} \right).
$$

13

Again, by Equation (3.2.23) and Proposition 3.2.17 from Karatzas and Shreve [12], we have

$$\mathbb{E}(\varepsilon_k \eta_k \mid \mathcal{F}_{k-1}) = -\beta\sigma_1^2 \int_{k-1}^{k} \int_{k-1}^{u} \mathrm{e}^{-b(k+u-2v)} \mathbb{E}(Y_v \mid \mathcal{F}_{k-1}) \, \mathrm{d}v \, \mathrm{d}u$$

$$+ \sigma_1\sigma_2\varrho \int_{k-1}^{k} \mathrm{e}^{-b(k-v)} \mathbb{E}(Y_v \mid \mathcal{F}_{k-1}) \, \mathrm{d}v.$$

Using (2.2), by an easy calculation,

$$\mathbb{E}(\varepsilon_k \eta_k \mid \mathcal{F}_{k-1})$$

$$= -\beta\sigma_1^2 \int_{k-1}^{k} \int_{k-1}^{u} \mathrm{e}^{-b(k+u-2v)} \left(\mathrm{e}^{-b(v-k+1)} Y_{k-1} + a \int_{k-1}^{v} \mathrm{e}^{-b(v-s)} \, \mathrm{d}s \right) \mathrm{d}v \, \mathrm{d}u$$

$$+ \sigma_1\sigma_2\varrho \int_{k-1}^{k} \mathrm{e}^{-b(k-v)} \left(\mathrm{e}^{-b(v-k+1)} Y_{k-1} + a \int_{k-1}^{v} \mathrm{e}^{-b(v-s)} \, \mathrm{d}s \right) \mathrm{d}v$$

$$= \left(-\beta\sigma_1^2 \int_0^1 \int_0^{u'} \mathrm{e}^{-b(u'-v'+1)} \, \mathrm{d}v' \, \mathrm{d}u' + \sigma_1\sigma_2\varrho\mathrm{e}^{-b} \right) Y_{k-1} - a\beta\sigma_1^2 \int_0^1 \int_0^{u'} \int_0^{v'} \mathrm{e}^{-b(u'-v'-s'+1)} \, \mathrm{d}s' \, \mathrm{d}v' \, \mathrm{d}u'$$

$$+ a\sigma_1\sigma_2\varrho \int_0^1 \int_0^{v'} \mathrm{e}^{-b(1-s')} \, \mathrm{d}s' \, \mathrm{d}v' =: C_5 Y_{k-1} + C_6, \qquad k \in \mathbb{N}.$$

Hence, by (2.5) and (2.7),

$$\sum_{k=1}^{n} \mathbb{E}\left((\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})(\boldsymbol{M}_{n,k} - \boldsymbol{M}_{n,k-1})^\top \mid \mathcal{F}_{n,k-1} \right)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \begin{bmatrix} C_1 Y_{k-1} + C_2 & C_5 Y_{k-1} + C_6 \\ C_5 Y_{k-1} + C_6 & C_3 Y_{k-1} + C_4 \end{bmatrix} \otimes \begin{bmatrix} 1 & Y_{k-1} \\ Y_{k-1} & Y_{k-1}^2 \end{bmatrix}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \begin{bmatrix} Y_{k-1} & Y_{k-1}^2 \\ Y_{k-1}^2 & Y_{k-1}^3 \end{bmatrix} + \frac{1}{n} \sum_{k=1}^{n} \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} 1 & Y_{k-1} \\ Y_{k-1} & Y_{k-1}^2 \end{bmatrix}$$

$$\xrightarrow{\text{a.s.}} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} + \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} =: \boldsymbol{D} \qquad \text{as} \quad n \to \infty,$$

where the 4×4 limit matrix $\boldsymbol{D}$ is necessarily symmetric and positive semi-definite (indeed, the limit of positive semi-definite matrices is positive semi-definite).

Next, we check Lindeberg condition (2.9). Since

$$\|\boldsymbol{x}\|^2 \mathbb{1}_{\{\|\boldsymbol{x}\| \geqslant \varepsilon\}} \leqslant \frac{\|\boldsymbol{x}\|^4}{\varepsilon^2} \mathbb{1}_{\{\|\boldsymbol{x}\| \geqslant \varepsilon\}} \leqslant \frac{\|\boldsymbol{x}\|^4}{\varepsilon^2}, \qquad \boldsymbol{x} \in \mathbb{R}^4, \quad \varepsilon \in \mathbb{R}_{++},$$

and $\|\boldsymbol{x}\|^4 = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^2 \leqslant 4(x_1^4 + x_2^4 + x_3^4 + x_4^4)$, $x_1, x_2, x_3, x_4 \in \mathbb{R}$, it is enough to check that

$$\frac{1}{n^2} \sum_{k=1}^{n} \left(\mathbb{E}(\varepsilon_k^4 \mid \mathcal{F}_{k-1}) + Y_{k-1}^4 \mathbb{E}(\varepsilon_k^4 \mid \mathcal{F}_{k-1}) + \mathbb{E}(\eta_k^4 \mid \mathcal{F}_{k-1}) + Y_{k-1}^4 \mathbb{E}(\eta_k^4 \mid \mathcal{F}_{k-1}) \right)$$

$$= \frac{1}{n^2} \sum_{k=1}^{n} \mathbb{E}((1 + Y_{k-1}^4)(\varepsilon_k^4 + \eta_k^4) \mid \mathcal{F}_{k-1}) \xrightarrow{\mathbb{P}} 0 \qquad \text{as} \quad n \to \infty.$$

14

Instead of convergence in probability, we show convergence in L^1, i.e., we check that

$$\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}((1+Y_{k-1}^4)(\varepsilon_k^4+\eta_k^4)) \to 0 \qquad \text{as } n\to\infty.$$

Clearly, it is enough to show that

$$\sup_{k\in\mathbb{N}}\mathbb{E}((1+Y_{k-1}^4)(\varepsilon_k^4+\eta_k^4)) < \infty.$$

By Cauchy–Schwarz inequality,

$$\mathbb{E}((1+Y_{k-1}^4)(\varepsilon_k^4+\eta_k^4)) \leqslant \sqrt{\mathbb{E}((1+Y_{k-1}^4)^2)\,\mathbb{E}((\varepsilon_k^4+\eta_k^4)^2)} \leqslant \sqrt{2}\sqrt{\mathbb{E}((1+Y_{k-1}^4)^2)\,\mathbb{E}(\varepsilon_k^8+\eta_k^8)}$$

for all $k\in\mathbb{N}$. Since, by Proposition 3 in Ben Alaya and Kebaier [3],

$$(3.12) \qquad\qquad \sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^\kappa) < \infty, \qquad \kappa\in\mathbb{R}_+,$$

it remains to check that $\sup_{k\in\mathbb{N}}\mathbb{E}(\varepsilon_k^8+\eta_k^8) < \infty$. Since, by the power mean inequality,

$$\mathbb{E}(\varepsilon_k^8) = \mathbb{E}(|Y_k - dY_{k-1} - c|^8) \leqslant \mathbb{E}((Y_k + dY_{k-1} + c)^8) \leqslant 3^7\,\mathbb{E}(Y_k^8 + d^8 Y_{k-1}^8 + c^8), \qquad k\in\mathbb{N},$$

using (3.12), we have $\sup_{k\in\mathbb{N}}\mathbb{E}(\varepsilon_k^8) < \infty$. Using (2.1) and again the power mean inequality, we have

$$\mathbb{E}(\eta_k^8) = \mathbb{E}((X_k - X_{k-1} - \gamma - \delta Y_{k-1})^8)$$

$$= \mathbb{E}\left(\left(\alpha - \beta\int_{k-1}^{k} Y_u\,\mathrm{d}u + \sigma_2\varrho\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}W_u + \sigma_2\sqrt{1-\varrho^2}\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}B_u - \gamma - \delta Y_{k-1}\right)^8\right)$$

$$\leqslant 6^7\,\mathbb{E}\left(\alpha^8 + \beta^8\left(\int_{k-1}^{k} Y_u\,\mathrm{d}u\right)^8 + \sigma_2^8\varrho^8\left(\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}W_u\right)^8 + \sigma_2^8(1-\varrho^2)^4\left(\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}B_u\right)^8\right.$$

$$\left. + \delta^8 Y_{k-1}^8 + \gamma^8\right), \qquad k\in\mathbb{N}.$$

By Jensen's inequality and (3.12),

$$(3.13)$$

$$\sup_{k\in\mathbb{N}}\mathbb{E}\left(\left(\int_{k-1}^{k} Y_u\,\mathrm{d}u\right)^8\right) \leqslant \sup_{k\in\mathbb{N}}\mathbb{E}\left(\int_{k-1}^{k} Y_u^8\,\mathrm{d}u\right) = \sup_{k\in\mathbb{N}}\int_{k-1}^{k}\mathbb{E}(Y_u^8)\,\mathrm{d}u$$

$$\leqslant \left(\sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8)\right)\left(\sup_{k\in\mathbb{N}}\int_{k-1}^{k} 1\,\mathrm{d}u\right) = \sup_{t\in\mathbb{R}_+}\mathbb{E}(Y_t^8) < \infty.$$

By the SDE (1.1) and the power mean inequality,

$$\mathbb{E}\left(\left(\int_{k-1}^{k}\sqrt{Y_u}\,\mathrm{d}W_u\right)^8\right) \leqslant \frac{1}{\sigma_1^8}\mathbb{E}\left(\left(Y_k - Y_{k-1} - a - b\int_{k-1}^{k} Y_u\,\mathrm{d}u\right)^8\right)$$

$$\leqslant \frac{4^7}{\sigma_1^8}\mathbb{E}\left(Y_k^8 + Y_{k-1}^8 + a^8 + b^8\left(\int_{k-1}^{k} Y_u\,\mathrm{d}u\right)^8\right), \qquad k\in\mathbb{N},$$

15

and hence, by (3.13),

$$\sup_{k\in\mathbb{N}} \mathbb{E}\left(\left(\int_{k-1}^{k} \sqrt{Y_u}\, dW_u\right)^8\right) \leqslant \frac{4^7}{\sigma_1^8}\left(2\sup_{t\in\mathbb{R}_+} \mathbb{E}(Y_t^8) + a^8 + b^8 \sup_{t\in\mathbb{R}_+} \mathbb{E}(Y_t^8)\right) < \infty.$$

Further, using that the conditional distribution of $\int_{k-1}^{k} \sqrt{Y_u}\, dB_u$ given $(Y_u)_{u\in[0,k]}$ is normal with mean 0 and variance $\int_{k-1}^{k} Y_u\, du$ for all $k \in \mathbb{N}$, we have

$$\mathbb{E}\left(\left(\int_{k-1}^{k} \sqrt{Y_u}\, dB_u\right)^8 \bigg| (Y_u)_{u\in[0,k]}\right) = 105\left(\int_{k-1}^{k} Y_u\, du\right)^4, \qquad k \in \mathbb{N},$$

and consequently

$$\mathbb{E}\left(\left(\int_{k-1}^{k} \sqrt{Y_u}\, dB_u\right)^8\right) = 105\, \mathbb{E}\left(\left(\int_{k-1}^{k} Y_u\, du\right)^4\right), \qquad k \in \mathbb{N}.$$

Hence, similarly to (3.13), we have

$$\sup_{k\in\mathbb{N}} \mathbb{E}\left(\left(\int_{k-1}^{k} \sqrt{Y_u}\, dB_u\right)^8\right) \leqslant 105 \sup_{t\in\mathbb{R}_+} \mathbb{E}(Y_t^4) < \infty,$$

which yields that $\sup_{k\in\mathbb{N}} \mathbb{E}(\eta_k^8) < \infty$. All in all, by the martingale central limit theorem (see, Theorem 2.6),

$$\boldsymbol{M}_{n,n} = n^{-1/2} \sum_{k=1}^{n} \begin{bmatrix} \varepsilon_k \\ \eta_k \end{bmatrix} \otimes \begin{bmatrix} 1 \\ Y_{k-1} \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4(\boldsymbol{0}, \boldsymbol{D}) \qquad \text{as } n \to \infty.$$

Consequently, by (3.11) and Slutsky's lemma,

$$\sqrt{n} \begin{bmatrix} \widehat{c}_n^{\text{CLSE}} - c \\ \widehat{d}_n^{\text{CLSE}} - d \\ \widehat{\gamma}_n^{\text{CLSE}} - \gamma \\ \widehat{\delta}_n^{\text{CLSE}} - \delta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4\left(\boldsymbol{0}, \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)^{-1} \boldsymbol{D} \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)^{-1}\right)$$

as $n \to \infty$, where the covariance matrix of the limit distribution takes the form

$$\left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)^{-1} \boldsymbol{D} \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)^{-1}$$

$$= \left(\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix}\right)\right) \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right)$$

$$+ \left(\begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}\right)\right) \left(\boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1}\right)$$

16

$$= \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \right)$$

$$+ \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \left(\begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix} \begin{bmatrix} 1 & \mathbb{E}(Y_\infty) \\ \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \end{bmatrix}^{-1} \right)$$

$$= \frac{1}{(\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)^2} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}$$

$$\otimes \left(\begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y_\infty) & \mathbb{E}(Y_\infty^2) \\ \mathbb{E}(Y_\infty^2) & \mathbb{E}(Y_\infty^3) \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix} \right)$$

$$+ \frac{1}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix}$$

$$= \frac{1}{(\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2)^2} \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}$$

$$\otimes \begin{bmatrix} -\mathbb{E}(Y_\infty)((\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3)) & (\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) \\ (\mathbb{E}(Y_\infty^2))^2 - \mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) & \mathbb{E}(Y_\infty^3) - 2\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^2) + (\mathbb{E}(Y_\infty))^3 \end{bmatrix}$$

$$+ \frac{1}{\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2} \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \mathbb{E}(Y_\infty^2) & -\mathbb{E}(Y_\infty) \\ -\mathbb{E}(Y_\infty) & 1 \end{bmatrix}$$

(3.14)
$$= \begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix} \otimes \begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix} + \begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix} \otimes \begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix} := \boldsymbol{E}.$$

Indeed, by (2.5), an easy calculation shows that

$$\left(\mathbb{E}(Y_\infty)\mathbb{E}(Y_\infty^3) - (\mathbb{E}(Y_\infty^2))^2 \right) \mathbb{E}(Y_\infty) = \frac{a^3\sigma_1^2}{4b^5}(2a + \sigma_1^2),$$

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

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

$$\mathbb{E}(Y_\infty^2) - (\mathbb{E}(Y_\infty))^2 = \frac{a\sigma_1^2}{2b^2}.$$

Finally, we show that $\boldsymbol{E}$ is positive definite. To show this, it is enough to check that

(i) the matrix

$$\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}$$

is positive definite,

17

(ii) the matrices

$$
\begin{bmatrix} C_2 & C_6 \\ C_6 & C_4 \end{bmatrix}, \qquad
\begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix}
\qquad \text{and} \qquad
\begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix}
$$

are positive semi-definite.

Indeed, the sum of a positive definite and a positive semi-definite square matrix is positive definite, the Kronecker product of positive semi-definite matrices is positive semi-definite and the Kronecker product of positive definite matrices is positive definite (as a consequence of the fact that the eigenvalues of the Kronecker product of two square matrices are the product of the eigenvalues of the two square matrices in question including multiplicities). The positive semi-definiteness of the matrices

$$
\begin{bmatrix} \frac{a(2a+\sigma_1^2)}{b\sigma_1^2} & -\frac{2a+\sigma_1^2}{\sigma_1^2} \\ -\frac{2a+\sigma_1^2}{\sigma_1^2} & \frac{2b(a+\sigma_1^2)}{a\sigma_1^2} \end{bmatrix}
\qquad \text{and} \qquad
\begin{bmatrix} \frac{2a+\sigma_1^2}{\sigma_1^2} & -\frac{2b}{\sigma_1^2} \\ -\frac{2b}{\sigma_1^2} & \frac{2b^2}{a\sigma_1^2} \end{bmatrix}
$$

readily follows, since $\frac{a(2a+\sigma_1^2)}{b\sigma_1^2} > 0$, $\frac{2a+\sigma_1^2}{\sigma_1^2} > 0$, and the determinant of the matrices in question are $\frac{2a+\sigma_1^2}{\sigma_1^2} > 0$ and $\frac{2b}{a\sigma_1^2} > 0$, respectively. Next, we prove that the matrices

$$
\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}
\qquad \text{and} \qquad
\begin{bmatrix} C_2 & C_4 \\ C_4 & C_6 \end{bmatrix}
$$

are positive semi-definite. Since $\mathbb{P}(Y_0 = y_0) = 1$, we have $\mathbb{E}(\varepsilon_1^2 \,|\, \mathcal{F}_0) = C_1 y_0 + C_2$, $\mathbb{E}(\eta_1^2 \,|\, \mathcal{F}_0) = C_3 y_0 + C_4$, and $\mathbb{E}(\varepsilon_1 \eta_1 \,|\, \mathcal{F}_0) = C_5 y_0 + C_6$ $\mathbb{P}$-almost surely, hence

$$
\mathbb{E}(\varepsilon_1^2)\,\mathbb{E}(\eta_1^2) - \big(\mathbb{E}(\varepsilon_1 \eta_1)\big)^2 = (C_1 C_3 - C_5^2)y_0^2 + (C_1 C_4 + C_2 C_3 - 2C_5 C_6)y_0 + C_2 C_4 - C_6^2.
$$

Clearly, by Cauchy–Schwarz's inequality,

$$
\mathbb{E}(\varepsilon_1^2)\,\mathbb{E}(\eta_1^2) - \big(\mathbb{E}(\varepsilon_1 \eta_1)\big)^2 \geqslant 0,
$$

hence, by setting an arbitrary initial value $Y_0 = y_0 \in \mathbb{R}_+$, we obtain $C_1 C_3 - C_5^2 \geqslant 0$ and $C_2 C_4 - C_6^2 \geqslant 0$. Thus, both matrices

$$
\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}
\qquad \text{and} \qquad
\begin{bmatrix} C_2 & C_4 \\ C_4 & C_6 \end{bmatrix}
$$

are positive semi-definite, since $C_1 > 0$ and $C_2 > 0$. Now we turn to check that

$$
\begin{bmatrix} C_1 & C_5 \\ C_5 & C_3 \end{bmatrix}
$$

is positive definite. Since $C_1 > 0$, this is equivalent to showing that $C_1 C_3 - C_5^2 > 0$. Recalling the

definition of the constants, we have

$$C_1 = \sigma_1^2 \int_0^1 e^{-b(2-v)}\, dv = \sigma_1^2 e^{-2b}\frac{e^b - 1}{b},$$

$$C_3 = \beta^2\sigma_1^2 \int_0^1 \int_0^1 \int_0^{u'\wedge v'} e^{-b(u'+v'-w')}\, dw'\, dv'\, du' - 2\beta\sigma_1\sigma_2\varrho \int_0^1 \int_0^{u'} e^{-bu'}\, dv'\, du' + \sigma_2^2 \int_0^1 e^{-bu'}\, du'$$

$$= b^{-3}\left(2e^{-b}\beta^2\sigma_1^2(\sinh b - b) + 2b\beta\varrho\sigma_1\sigma_2((1+b)e^{-b} - 1) + b^2\sigma_2^2(1 - e^{-b})\right),$$

$$C_5 = -\beta\sigma_1^2 \int_0^1 \int_0^{u'} e^{-b(u'-v'+1)}\, dv'\, du' + \sigma_1\sigma_2\varrho e^{-b} = b^{-2}\sigma_1 e^{-b}\left(-e^{-b}\beta\sigma_1(1 + (b-1)e^b) + \varrho\sigma_2 b^2\right),$$

thus we have

$$C_1 C_3 - C_5^2 = b^{-4}e^{-2b}\sigma_1^2\Big(2b(2+b^2)\beta\varrho\sigma_1\sigma_2 + 2(\beta^2\sigma_1^2 - 2b\beta\varrho\sigma_1\sigma_2 + b^2\sigma_2^2)\cosh b - (2+b^2)\beta^2\sigma_1^2$$

$$- b^2(2 + b^2\varrho^2)\sigma_2^2\Big).$$

Consequently, using that $\cosh b = \sum_{k=0}^\infty \frac{b^{2k}}{(2k)!} > 1 + \frac{b^2}{2}$ and that

$$\beta^2\sigma_1^2 - 2b\beta\varrho\sigma_1\sigma_2 + b^2\sigma_2^2 = (\beta\sigma_1 - b\varrho\sigma_2)^2 + b^2(1 - \varrho^2)\sigma_2^2 > 0,$$

we have

$$C_1 C_3 - C_5^2 > b^{-4}e^{-2b}\sigma_1^2\Big(4b\beta\varrho\sigma_1\sigma_2 + 2b^3\beta\varrho\sigma_1\sigma_2 + 2\beta^2\sigma_1^2 + b^2\beta^2\sigma_1^2 - 4b\beta\varrho\sigma_1\sigma_2 - 2b^3\beta\varrho\sigma_1\sigma_2$$

$$+ 2b^2\sigma_2^2 + b^4\sigma_2^2 - 2\beta^2\sigma_1^2 - b^2\beta^2\sigma_1^2 - 2b^2\sigma_2^2 - b^4\varrho^2\sigma_2^2\Big)$$

$$= b^{-4}e^{-2b}\sigma_1^2(b^4(1 - \varrho^2)\sigma_2^2) > 0.$$

With this our proof is finished. □

So far we have obtained the limit distribution of the CLSE of the transformed parameters (c, d, γ, δ). A natural estimator of (a, b, α, β) can be obtained from (3.2) using relation (3.3) detailed as follows. Calculating the integrals in (3.3) in the subcritical case, let us introduce the function $g : \mathbb{R}_{++}^2 \times \mathbb{R}^2 \to \mathbb{R}_{++} \times (0,1) \times \mathbb{R}^2$,

(3.15) $\qquad g(a, b, \alpha, \beta) := \begin{bmatrix} ab^{-1}(1 - e^{-b}) \\ e^{-b} \\ \alpha - a\beta b^{-2}(e^{-b} - 1 + b) \\ -\beta b^{-1}(1 - e^{-b}) \end{bmatrix} = \begin{bmatrix} c \\ d \\ \gamma \\ \delta \end{bmatrix}, \qquad (a, b, \alpha, \beta) \in \mathbb{R}_{++}^2 \times \mathbb{R}^2.$

Note that g is bijective having inverse

(3.16) $\qquad g^{-1}(c, d, \gamma, \delta) = \begin{bmatrix} -c\frac{\log d}{1-d} \\ -\log d \\ \gamma - c\delta\frac{d-1-\log d}{(1-d)^2} \\ \delta\frac{\log d}{1-d} \end{bmatrix} = \begin{bmatrix} a \\ b \\ \alpha \\ \beta \end{bmatrix}, \qquad (c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0,1) \times \mathbb{R}^2.$

19

Indeed, for all $(c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$, we have

$$\alpha = \gamma + a\beta b^{-2}(\mathrm{e}^{-b} - 1 + b) = \gamma + (-c)\frac{\log d}{1 - d}\delta\frac{\log d}{1 - d}(-\log d)^{-2}(d - 1 - \log d)$$

$$= \gamma - c\delta\frac{d - 1 - \log d}{(1 - d)^2}.$$

Under the conditions of Theorem 3.2 the CLSE $(\widehat{c}_n^{\,\mathrm{CLSE}}, \widehat{d}_n^{\,\mathrm{CLSE}}, \widehat{\gamma}_n^{\,\mathrm{CLSE}}, \widehat{\delta}_n^{\,\mathrm{CLSE}})$ of (c, d, γ, δ) is strongly consistent, hence in the subcritical case $(\widehat{c}_n^{\,\mathrm{CLSE}}, \widehat{d}_n^{\,\mathrm{CLSE}}, \widehat{\gamma}_n^{\,\mathrm{CLSE}}, \widehat{\delta}_n^{\,\mathrm{CLSE}})$ fall into the set $\mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$ for sufficiently large $n \in \mathbb{N}$ with probability one. Hence, in the subcritical case, one can introduce a natural estimator of (a, b, α, β) based on discrete time observations $(Y_i, X_i)_{i \in \{1, \ldots, n\}}$ by applying the inverse of g to the CLSE of (c, d, γ, δ), i.e.,

$$(3.17) \qquad (\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) := g^{-1}(\widehat{c}_n^{\,\mathrm{CLSE}}, \widehat{d}_n^{\,\mathrm{CLSE}}, \widehat{\gamma}_n^{\,\mathrm{CLSE}}, \widehat{\delta}_n^{\,\mathrm{CLSE}})$$

for sufficiently large $n \in \mathbb{N}$ with probability one.

3.3 Remark. We would like to stress the point that the estimator of (a, b, α, β) introduced in (3.17) exists only for sufficiently large $n \in \mathbb{N}$ with probability of 1. However, as all our results are asymptotic, this will not cause a problem. From the considerations before this remark, we obtain

$$(3.18) \qquad (\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) = \underset{(a,b,\alpha,\beta) \in \mathbb{R}_{++}^2 \times \mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} \left[(Y_i - dY_{i-1} - c)^2 + (X_i - X_{i-1} - \gamma - \delta Y_{i-1})^2\right]$$

for sufficiently large $n \in \mathbb{N}$ with probability one. We call the attention that $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ does not necessarily provides a CLSE of (a, b, α, β), since in (3.18) one takes the infimum only on the set $\mathbb{R}_{++}^2 \times \mathbb{R}^2$ instead of $\mathbb{R}^4$. Formula (3.18) serves as a motivation for calling $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ essentially conditional least squares estimator in the Abstract. $\qquad\square$

3.4 Theorem. *Under the conditions of Theorem 3.2 the sequence* $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$, $n \in \mathbb{N}$, *is strongly consistent and asymptotically normal, i.e.,*

$$(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n) \xrightarrow{\mathrm{a.s.}} (a, b, \alpha, \beta) \qquad \textit{as } n \to \infty,$$

and

$$\sqrt{n}\begin{bmatrix} \widehat{a}_n - a \\ \widehat{b}_n - b \\ \widehat{\alpha}_n - \alpha \\ \widehat{\beta}_n - \beta \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_4\left(0, \boldsymbol{J}\boldsymbol{E}\boldsymbol{J}^\top\right) \qquad \textit{as } n \to \infty,$$

where $\boldsymbol{E} \in \mathbb{R}^{2 \times 2}$ *is a symmetric, positive definite matrix given in (3.14) and*

$$\boldsymbol{J} := \begin{bmatrix} -\frac{\log d}{1 - d} & -c\frac{\log d - 1 + d^{-1}}{(1 - d)^2} & 0 & 0 \\ 0 & -\frac{1}{d} & 0 & 0 \\ \delta\frac{\log d + 1 - d}{(1 - d)^2} & c\delta\frac{2\log d - d + d^{-1}}{(1 - d)^3} & 1 & c\frac{\log d + 1 - d}{(1 - d)^2} \\ 0 & \delta\frac{\log d - 1 + d^{-1}}{(1 - d)^2} & 0 & \frac{\log d}{1 - d} \end{bmatrix}$$

with c, d *and* δ *given in (3.3).*

Proof. The strong consistency of $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$, $n \in \mathbb{N}$, follows from the strong consistency of the CLSE of (c, d, γ, δ) proved in Theorem 3.2 using also that the inverse function g^{-1} given in (3.16) is continuous on $\mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$. For the second part of the theorem we use Theorem 3.2, and the so-called delta method (see, e.g., Theorem 11.2.14 in Lehmann and Romano [13]). Indeed, one can extend the function g^{-1} to be defined on $\mathbb{R}^4$ not only on $\mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$ (e.g., let it be zero on the complement of $\mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$), $(\widehat{a}_n, \widehat{b}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ takes the form given in (3.17) with this extension of g^{-1} as well, and the Jacobian of g^{-1} at $(c, d, \gamma, \delta) \in \mathbb{R}_{++} \times (0, 1) \times \mathbb{R}^2$ is clearly $\boldsymbol{J}$. $\square$

Acknowledgements

References

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

[2] BARCZY, M., DÖRING, L., LI, Z. and PAP, G. (2014). Stationarity and ergodicity for an affine two-factor model. *Advances in Applied Probability* **46(3)** 878–898.

[3] BEN ALAYA, M. and KEBAIER, M. (2013). Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stochastic Analysis and Applications* **31(4)** 552–573.

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

[5] DOKUCHAEV, N. (2015). Pathwise estimation of the diffusion term for Cox-Ingersoll-Ross and similar processes. Available on the ArXiv: `http://arxiv.org/abs/1506.05627`

[6] DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California.

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

[8] HURN, A. S., LINDSAY, K. A. and MCCLELLAND, A. J. (2013). A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions. *Journal of Econometrics* **172** 106–126.

[9] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. Springer-Verlag, Berlin.

[10] JIN, P., MANDREKAR, V., RÜDIGER, B. and TRABELSI, C. (2013). Positive Harris recurrence of the CIR process and its applications. *Communications on Stochastic Analysis* **7(3)** 409–424.

[11] JIN, P., RÜDIGER, B. and TRABELSI, C. (2016). Positive Harris recurrence and exponential ergodicity of the basic affine jump-diffusion. *Stochastic Analysis and Applications* **34(1)** 75–95.

[12] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus,* 2nd ed. Springer-Verlag.

[13] LEHMANN, E.L. and ROMANO, J. P. (2009). *Testing Statistical Hypotheses,* 3rd ed. Springer-Verlag Berlin Heidelberg.

[14] LI, Z. and MA, C. (2015). Asymptotic properties of estimators in a stable Cox-Ingersoll-Ross model. *Stochastic Processes and their Applications* **125(8)** 3196–3233.

[15] OVERBECK, L. and RYDÉN, T. (1997). Estimation in the Cox-Ingersoll-Ross model. *Econometric Theory* **13(3)** 430–461.

[16] SHIRYAEV, A. N. (1989). *Probability,* 2nd ed. Springer, New York.