Austrian Journal of Statistics
2025, Volume 54, 82–99.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i1.1968

Discretization and Asymptotic Normality of Drift Parameters Estimator in the Cox–Ingersoll–Ross Model

Olha Prykhodko
Taras Shevchenko National University of Kyiv

Kostiantyn Ralchenko
Taras Shevchenko National University of Kyiv
University of Vaasa

Abstract

This paper investigates the simultaneous estimation of two drift parameters of a Cox–Ingersoll–Ross (CIR) model, for which observations can be made either continuously or at discrete time instants. For continuous-time observations, we establish the joint asymptotic normality of the strongly consistent parameter estimators introduced in Dehtiar et al. (Comm. Statist. Theory Methods, 51(19):6818–6833, 2022). Additionally, we study the discrete counterparts of these estimators and prove their strong consistency and joint asymptotic normality.

Keywords: CIR model, continuous observations, discrete observations, strong consistency, joint asymptotic normality.

1. Introduction

To improve the modeling of short-term interest rate evolution, compared to the Vasicek model, Cox, Ingersoll, and Ross (1985) proposed the following equation for the dynamics of the short interest rate:

$$r_t = (a - br_t)\, dt + \sigma\sqrt{r_t}\, dW_t, \quad t \ge 0, \tag{1.1}$$

where $W = \{W_t, t \ge 0\}$ is a Wiener process modeling the random market risk factor, a, b and σ are positive constants. The parameter b corresponds to the speed of adjustment to the mean a/b, and σ represents volatility. It is well-known that this model has several empirically relevant properties. Notably, in this model, r never becomes negative, and the randomly moving interest rate is elastically pulled towards the long-term constant value a/b. Currently, the model (1.1) is widely used in mathematical finance, particularly for valuing interest rate derivatives.

In this paper, we investigate the estimation of the drift parameters (a, b) by observations of a solution r. The diffusion parameter σ is assumed to be known, a typical assumption in parameter estimation for diffusion models, as explained in Remark 2.1 below. It is worth mentioning that the theory of parameter estimation for stochastic differential equations driven by a standard Wiener process, especially homogeneous ones, is well-developed. Comprehensive resources on this topic can be found in the books Bishwal (2008); Iacus (2008); Kutoyants (2004); Liptser and Shiryaev (2001). However, standard approaches for such models usually assume that the diffusion coefficient is Lipschitz continuous. For the coefficient $\sigma\sqrt{x}$ of the CIR model, this assumption does not hold, and moreover, this coefficient is not bounded away from zero. Therefore, the parameter estimation in the CIR model requires a special investigation.

The case of continuous-time statistical inference in the CIR model (when the entire sample path $\{r_t, t \in [0,T]\}$ is observed) was studied by Ben Alaya and Kebaier (2013). They proved the consistency and asymptotic normality of the maximum likelihood estimator (MLE), which has the following form:

$$\hat{a}_T^{\text{mle}} = \frac{\int_0^T r_t dt \int_0^T \frac{dr_t}{r_t} - T(r_T - r_0)}{\int_0^T r_t dt \cdot \int_0^T \frac{dt}{r_t} - T^2}, \qquad \hat{b}_T^{\text{mle}} = \frac{(r_0 - r_T)\int_0^T \frac{dt}{r_t} + T\int_0^T \frac{dr_t}{r_t}}{\int_0^T r_t dt \cdot \int_0^T \frac{dt}{r_t} - T^2}. \tag{1.2}$$

Note that the maximum likelihood estimator is well-defined only if $2a \geq \sigma^2$, as it includes the integral $\int_0^T \frac{1}{r_t} dt$, which exists with probability one if and only if $2a \geq \sigma^2$, see (Ben Alaya and Kebaier 2012, Prop. 4).

To overcome this restriction, Dehtiar, Mishura, and Ralchenko (2022) proposed an alternative estimator that does not require the assumption $2a \geq \sigma^2$. This estimator is based on the ergodic properties of the CIR model and is presented in subsection 2.2 below, see (2.4)–(2.5). Dehtiar *et al.* (2022) proved the strong consistency of this estimator. In this paper, we further study their approach and establish the asymptotic normality.

Furthermore, since in practice the data is discrete, we focus significantly on the estimation of the parameter (a, b) of the CIR model from discrete-time observations. We concentrate on the case of high-frequency sampling, where the sampling step tends to zero, and the observation horizon increases.

A similar problem was addressed in the recent paper Chernova, Dehtiar, Mishura, and Ralchenko (2024), where a discretized version of the continuous-time MLE (1.2) was introduced. The authors found the conditions for weak and strong consistency and asymptotic normality of the estimators and established the rate of convergence of these estimators in probability. A similar approach was studied by (Ben Alaya and Kebaier 2013), where the continuous maximum likelihood estimators were first transformed by the Itô formula and then discretized. Both approaches require the additional condition $a > \sigma^2$, which arises when the integral $\int_0^T \frac{1}{r_t} dt$ is approximated by the corresponding Riemann sum. An alternative method was proposed by Tang and Chen (2009). Rather than discretizing continuous-time MLEs, they considered a discretization of the CIR model and constructed MLEs by maximizing a discrete-time pseudo-likelihood function. Notably, this approach also requires the condition $a > \sigma^2$. Recently, parameter estimation under high-frequency sampling was studied by Cheng, Hufnagel, and Masuda (2022) using a Gaussian quasi-likelihood approach, where an even more restrictive assumption $2a > 5\sigma^2$ was required.

In this paper, we investigate the discrete analog of the strongly consistent estimator proposed in (Dehtiar *et al.* 2022, Thm. 5). As detailed in subsection 2.3, we prove the strong consistency and asymptotic normality of this estimator. We also present a comparison of our approach with alternative approaches for similar sampling schemes. The main advantage of our method is its validity for any positive a and b, in contrast to other known approaches.

The problem of drift parameter estimation for the CIR model in discrete-time settings for different observation schemes was studied in De Rossi (2010); Kladívko (2007); Overbeck and Rydén (1997), where various methods were proposed. Moreover, the case $b = 0$ was explored by Ben Alaya and Kebaier (2013), where they derived the consistency and asymptotic distribution of the maximum likelihood estimator. Additionally, it is important to note there is no consistent estimator for a when $b < 0$, as demonstrated by Overbeck (1998, Thm. 2(v)).

The paper is organized as follows. In Section 2, we describe the statistical problem and formulate the main results. Section 3 contains some simulation experiments. All the proofs are collected in Section 4.

2. Main results

2.1. Model description: the CIR process and its properties

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. We analyze the following stochastic differential equation

$$r_t = r_0 + \int_0^t (a - br_s)\, ds + \sigma \int_0^t \sqrt{r_s}\, dW_s, \quad t \ge 0, \tag{2.1}$$

where $W = \{W_t, t \ge 0\}$ is a Wiener process, and

$$a,\ b,\ \sigma \text{ and } r_0 \text{ are positive constants.} \tag{2.2}$$

Under assumption (2.2) the equation (2.1) possesses a unique non-negative strong solution. This solution is called the *Cox–Ingersoll–Ross* (CIR) process.

If the additional assumption $2a > \sigma^2$ (often referred to as Feller's condition) is satisfied, then the CIR process r remains strictly positive. For $0 < a < 2\sigma^2$, it almost surely hits zero, with the state 0 being instantaneously reflecting.

When $2a > \sigma^2$, the CIR process is ergodic, and its stationary distribution is a Gamma distribution with the following density:

$$p_\infty(x) = \frac{(2b/\sigma^2)^{2a/\sigma^2}}{\Gamma(2a/\sigma^2)}\, x^{2a/\sigma^2-1} e^{-2bx/\sigma^2} \mathbb{1}_{x>0}. \tag{2.3}$$

Ergodicity implies that for any function $f \in L^1(p_\infty)$, the time average $\frac{1}{T}\int_0^T f(r_t)dt$ converges a.s. to the space average $\int_{\mathbb{R}} f(x) p_\infty(dx)$, as $T \to \infty$.

More details on the properties of the CIR process can be found, for example, in Göing-Jaeschke and Yor (2003) or Alfonsi (2015).

The primary focus of this paper is the estimation of the unknown drift parameter (a, b) through observations of the trajectory of the CIR process r. We investigate two scenarios for the observations:

1. *Continuous observations:* The process r is observed continuously over the interval $[0, T]$ (see Subsection 2.2).
2. *Discrete observations:* The process r is observed at deterministic and equidistant time points, under the conditions of high frequency and infinite horizon (see Subsection 2.3).

Further details are provided in the respective subsections below.

Remark 2.1. In both continuous and discrete frameworks, we assume that the diffusion parameter σ is known. This assumption is standard in the context of parameter estimation for diffusion models (see, e.g., Kutoyants (2004); Sørensen (2002)) because there are straightforward estimators, such as those utilizing L^p-variation of r_t, which converge much more quickly than the estimators for the parameters a and b.

For example, the parameter σ can be estimated independently of the drift parameters as follows. For any fixed $t > 0$, let $\{t_k, k = 0, \dots, n\}$, $n \ge 1$, be a sequence of partitions of the interval $[0, t]$, with diameter $\delta_n \to 0$ as $n \to \infty$. It is well known that as $n \to \infty$,

$$\sum_{t_k \le t} (r_{t_k} - r_{t_{k-1}})^2 \to \sigma^2 \int_0^t r_t\, dt \quad \text{a.s.}$$

(see, e.g., (Jacod and Shiryaev 1987, p. 52)). Therefore, by substituting the integral with a Riemann sum, we derive the following strongly consistent estimator for σ^2:

$$\hat{\sigma}_n^2 = \frac{\sum_{t_k \le t} \left(r_{t_k} - r_{t_{k-1}}\right)^2}{\delta_n \sum_{t_k \le t} r_{t_k}}.$$

Additional methods for estimating the diffusion coefficient of the CIR process are available in Cheng *et al.* (2022); Dokuchaev (2017); Mishura, Ralchenko, and Dehtiar (2022); Tang and Chen (2009).

2.2. Estimation by continuous observations

Assume that we observe a sample path $\{r_t, t \in [0,T]\}$ continuously. The following strongly consistent estimator of the parameter (a, b) was constructed in (Dehtiar *et al.* 2022, Thm. 5):

$$\tilde{a}_T = \frac{\sigma^2}{2} \cdot \frac{\left(\int_0^T r_t\,dt\right)^2}{T\int_0^T r_t^2\,dt - \left(\int_0^T r_t\,dt\right)^2}, \tag{2.4}$$

$$\tilde{b}_T = \frac{\sigma^2}{2} \cdot \frac{T\int_0^T r_t\,dt}{T\int_0^T r_t^2\,dt - \left(\int_0^T r_t\,dt\right)^2}. \tag{2.5}$$

In the following theorem, we establish its asymptotic normality. This theorem represents the first main result of our paper.

Theorem 2.2. *The estimator* $(\tilde{a}_T, \tilde{b}_T)$ *is asymptotically normal:*

$$\sqrt{T}\begin{pmatrix}\tilde{a}_T - a\\ \tilde{b}_T - b\end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma), \quad \text{as } T \to \infty,$$

where

$$\Sigma = \begin{pmatrix}\frac{a}{b}\left(2a+\sigma^2\right) & 2a+\sigma^2\\ 2a+\sigma^2 & \frac{2b}{a}\left(a+\sigma^2\right)\end{pmatrix}.$$

Remark 2.3 (Comparison with MLE). Let us compare the estimator $(\tilde{a}_T, \tilde{b}_T)$, as defined by (2.4)–(2.5), with the maximum likelihood estimator, studied in Ben Alaya and Kebaier (2013). On one hand, the maximum likelihood estimator is well-defined only under the additional condition $2a > \sigma^2$. In contrast, the estimator $(\tilde{a}_T, \tilde{b}_T)$ is suitable for all $a > 0$, which is a crucial advantage. On the other hand, according to (Ben Alaya and Kebaier 2013, Thm. 5), the maximum likelihood estimator is asymptotically normal with the following covariance matrix:

$$\Sigma_{\text{mle}} = \begin{pmatrix}\frac{a}{b}\left(2a-\sigma^2\right) & 2a-\sigma^2\\ 2a-\sigma^2 & 2b\end{pmatrix}. \tag{2.6}$$

Comparing Σ and Σ_{mle}, we see that when the maximum likelihood estimator is well-defined, it is always more efficient than the estimator $(\tilde{a}_T, \tilde{b}_T)$.

2.3. Estimation by discrete observations

Now let us introduce a discrete counterpart of the estimator $(\tilde{a}_T, \tilde{b}_T)$. Following Chernova *et al.* (2024), we adopt the following observation scheme, which enables the construction of a consistent estimator. Let n be a positive integer. We consider an equidistant partition of the interval $[0, n]$ with a step size of $\delta_n = n^{-\beta}$, where $\beta > 0$. Observations are made at the times $t_k = k\delta_n = kn^{-\beta}$. Consequently, the total number of observations is $m_n = n^{\beta+1}$.

We define the estimator of (a,b) based on the discrete observations $\{r_{t_k}, k = 0,\dots m_n - 1\}$ as follows:

$$\hat{a}_n = \frac{\sigma^2}{2} \cdot \frac{\left(\sum_{k=0}^{m_n-1} r_{t_k}\right)^2}{m_n \sum_{k=0}^{m_n-1} r_{t_k}^2 - \left(\sum_{k=0}^{m_n-1} r_{t_k}\right)^2}, \tag{2.7}$$

$$\hat{b}_n = \frac{\sigma^2}{2} \cdot \frac{m_n \sum_{k=0}^{m_n-1} r_{t_k}}{m_n \sum_{k=0}^{m_n-1} r_{t_k}^2 - \left(\sum_{k=0}^{m_n-1} r_{t_k}\right)^2}. \tag{2.8}$$

Our second main result states that $(\hat{a}_n, \hat{b}_n)$ is a strongly consistent and asymptotically normal estimator of the parameter (a,b).

Theorem 2.4. *1. For any $\beta > 0$,*

$$\hat{a}_n \to a, \quad \hat{b}_n \to b, \quad \text{a.s. as } n \to \infty,$$

i.e. $(\hat{a}_n, \hat{b}_n)$ is a strongly consistent estimator of the parameter (a,b) respectively.
2. Let $\beta > 1$. Then the estimator $(\hat{a}_n, \hat{b}_n)$ is asymptotically normal:

$$\sqrt{n} \begin{pmatrix} \hat{a}_n - a \\ \hat{b}_n - b \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma), \quad \text{as } n \to \infty, \tag{2.9}$$

where Σ is defined in Theorem 2.2.

Remark 2.5 (Comparison with discretized MLE). Recently, Chernova *et al.* (2024) studied a discretized version of MLE and established the conditions for its (strong) consistency and asymptotic normality. Notably, compared to our estimator $(\hat{a}_n, \hat{b}_n)$, the discretized MLE requires the additional condition $a > \sigma^2$, which is even more restrictive than the condition $2a > \sigma^2$ needed for the continuous-time MLE, see (Chernova *et al.* 2024, Thm. 2.2).

Furthermore, the discretized MLE is strongly consistent for $\beta \ge 2$, whereas our estimator $(\hat{a}_n, \hat{b}_n)$ is strongly consistent for all $\beta > 0$. However, if the condition $a > \sigma^2$ is satisfied, then the discretized MLE is asymptotically normal with the asymptotic covariance matrix Σ_{mle}, see (2.6). This implies that, under these condition, the discretized MLE is more efficient than $(\hat{a}_n, \hat{b}_n)$.

Remark 2.6 (Comparison with pseudo-MLE). Tang and Chen (2009) proposed an alternative method for parameter estimation in the CIR model using discrete observations. Their approach is based on a discrete-time approximation of the CIR model, followed by the maximization of the discrete-time pseudo-log-likelihood function. They utilized a different parameterization of the drift coefficient, specifically $\kappa(\alpha - r_t)$ instead of $a - br_t$. Within this framework, they also examined the parameter estimation problem using an observation scheme where the observation horizon increases while the interval between observations tends to zero.

In particular, their estimator for the drift parameters (κ, α) is asymptotically normal with the following asymptotic covariance matrix:

$$\begin{pmatrix} 2\kappa & 2 \\ 2 & \frac{\alpha\sigma^2}{\kappa^2} \end{pmatrix},$$

as presented in (Tang and Chen 2009, Thm. 3.2.4).

By setting $a = \kappa\alpha$ and $b = \kappa$, defining the estimator for (a,b) accordingly, and applying the delta method, it can be shown that their estimator for (a,b) is asymptotically normal with the following asymptotic covariance matrix:

$$\Sigma_{\text{pseudo}} = \begin{pmatrix} \frac{a}{b}(2a + \sigma^2 + 4b) & 2(a+b) \\ 2(a+b) & 2b \end{pmatrix}.$$

Comparing Σ_{pseudo} with Σ, we observe that our estimator (2.7) for the parameter a is more efficient than that proposed by Tang and Chen (2009). Conversely, their estimator for b has the same asymptotic variance as the discretized MLE and is more efficient than our estimator (2.8).

It should also be noted that Tang and Chen (2009) assume the condition $a \geq \sigma^2$ is satisfied. This restriction is similar to the one described for the discretized MLE in Remark 2.5. We emphasize that our approach does not require such a restriction.

Remark 2.7 (Comparison with Gaussian quasi-MLE). Cheng *et al.* (2022) proposed an alternative approach for parameter estimation in the CIR model, based on Gaussian quasi-likelihood estimation. Similar to our approach, they considered the case of high-frequency sampling, where the observation horizon approaches infinity while the observation step size approaches zero. As a result, Cheng *et al.* (2022) constructed an asymptotically efficient estimator of the drift parameters, characterized by an asymptotic covariance matrix Σ_{mle} (see (2.6)). Additionally, they formulated a practical two-stage method that avoids numerical optimization. However, their approach imposes a rather restrictive assumption on the parameters, specifically $2a > 5\sigma^2$ (see condition (1.2) in Cheng *et al.* (2022)).

3. Simulations

To demonstrate the quality of the constructed discretized estimators, we conducted a simulation experiment using the R programming language. We selected the initial value $r_0 = 1$ and, for certain fixed parameter values, generated 1000 trajectories for the process r as a solution to the stochastic differential equation (2.1) through Euler–Maruyama approximations.

To investigate the influence of a, b, and σ on the behavior of the estimator $(\hat{a}_n, \hat{b}_n)$, we considered 12 different sets of model parameters, namely $(a, b, \sigma) \in \{1, 2\} \times \{1, 2\} \times \{1, 2, 3\}$. We set the parameter β equal to 1.5. For each set, we computed the estimates for various horizons of observations, namely $n = 2^4, 2^6, 2^8$. The biases of the estimators are reported in Table 1. The theoretical and empirical standard deviations of $\hat{a}_n$ and $\hat{b}_n$ are reported in Tables 2 and 3, respectively.

Table 1: Biases of $\hat{a}_n$ and $\hat{b}_n$

	$\mu(\hat{a}_n) - a$				$\mu(\hat{b}_n) - b$			
	$a=1$, $b=1$	$a=1$, $b=2$	$a=2$, $b=1$	$a=2$, $b=2$	$a=1$, $b=1$	$a=1$, $b=2$	$a=2$, $b=1$	$a=2$, $b=2$
$\sigma = 1$								
$n = 2^4$	0.3679	0.1924	0.5203	0.3644	0.4219	0.3593	0.3298	0.3908
$n = 2^6$	0.1087	0.0569	0.1596	0.1005	0.1286	0.1175	0.0990	0.1116
$n = 2^8$	0.0254	0.0121	0.0365	0.0211	0.0301	0.0249	0.0226	0.0235
$\sigma = 2$								
$n = 2^4$	0.4927	0.3115	0.7356	0.4717	0.7497	0.7945	0.5377	0.5981
$n = 2^6$	0.1670	0.1008	0.2412	0.1442	0.2438	0.2635	0.1686	0.1843
$n = 2^8$	0.0450	0.0247	0.0596	0.0320	0.0623	0.0627	0.0407	0.0409
$\sigma = 3$								
$n = 2^4$	0.6350	0.4189	0.9386	0.6109	1.2819	1.3760	0.8269	0.9225
$n = 2^6$	0.2322	0.1424	0.3320	0.2014	0.4110	0.4447	0.2632	0.2928
$n = 2^8$	0.0709	0.0415	0.0912	0.0503	0.1107	0.1161	0.0673	0.0697

We observed that the numerical results confirm our theoretical results. Namely, the biases and the standard deviations of the estimators approach zero as n increases. Moreover, the empirical standard deviations are close enough to the corresponding theoretical values. This fact serves as confirmation for the asymptotic normality stated in Theorem 2.4.

Table 2: Empirical and theoretical standard deviations of $\hat{a}_n$

	$a=1$, $b=1$		$a=1$, $b=2$		$a=2$, $b=1$		$a=2$, $b=2$	
	emp	theor	emp	theor	emp	theor	emp	theor
$\sigma=1$								
$n=2^4$	0.4427	0.4330	0.2916	0.3062	0.7988	0.7906	0.5696	0.5590
$n=2^6$	0.2039	0.2165	0.1461	0.1531	0.3819	0.3953	0.2740	0.2795
$n=2^8$	0.1047	0.1083	0.0763	0.0765	0.1937	0.1976	0.1410	0.1398
$\sigma=2$								
$n=2^4$	0.4965	0.6124	0.3445	0.4330	0.8797	1.0000	0.6232	0.7071
$n=2^6$	0.2481	0.3062	0.1856	0.2165	0.4375	0.5000	0.3239	0.3536
$n=2^8$	0.1384	0.1531	0.1028	0.1083	0.2358	0.2500	0.1732	0.1768
$\sigma=3$								
$n=2^4$	0.5780	0.8292	0.3947	0.5863	0.9795	1.2748	0.6907	0.9014
$n=2^6$	0.2909	0.4146	0.2244	0.2932	0.5047	0.6374	0.3801	0.4507
$n=2^8$	0.1728	0.2073	0.1317	0.1466	0.2846	0.3187	0.2124	0.2253

Moreover, we observed that among the considered parameter values, the estimator $\hat{a}_n$ demonstrates the best performance when $a=1$ and $b=2$, while the estimator $\hat{b}_n$ performs best when $a=2$ and $b=1$. This observation holds true from both bias and standard deviation perspectives. More generally, we notice the following pattern: a decrease in the parameter a and an increase in b result in improved performance of the estimator $\hat{a}_n$. Conversely, the estimator $\hat{b}_n$ exhibits the opposite behavior.

The increase of the volatility parameter σ negatively affects the performance of both estimators.

The mentioned facts align perfectly with the formulas for asymptotic variances in Theorem 2.4.

Moreover, our estimators indeed work regardless of whether Feller's condition $2a > \sigma^2$ is satisfied or not. In the case when it is satisfied, we can compare our results with the corresponding results for discretized MLEs from the paper Chernova *et al.* (2024). The discretized MLEs slightly outperform our estimators $\hat{a}_n$ and $\hat{b}_n$. However, our estimators still seem quite reasonable.

4. Proofs

In the following proofs, let C denote a generic constant whose value may change from one occurrence to another.

4.1. Asymptotic behavior of the integrals

In this subsection we investigate the properties of the integrals $\int_0^T r_t dt$ and $\int_0^T r_t^2 dt$, which are involved into the estimator $(\tilde{a}, \tilde{b})$. In particular, we express the above-mentioned integrals through the Itô martingales $\int_0^T r_t^{1/2} dW_t$ and $\int_0^T r_t^{3/2} dW_t$. The joint asymptotic normality of these martingales is established as well.

For further reference, let us recall some well-known facts about the CIR process from the papers Ben Alaya and Kebaier (2013); Chernova *et al.* (2024); Deelstra and Delbaen (1995); Dehtiar *et al.* (2022).

Table 3: Empirical and theoretical standard deviations of $\hat{b}_n$

	$a=1$, $b=1$		$a=1$, $b=2$		$a=2$, $b=1$		$a=2$, $b=2$	
	emp	theor	emp	theor	emp	theor	emp	theor
$\sigma=1$								
$n=2^4$	0.5152	0.5000	0.6408	0.7071	0.4608	0.4330	0.6230	0.6124
$n=2^6$	0.2455	0.2500	0.3422	0.3536	0.2173	0.2165	0.3057	0.3062
$n=2^8$	0.1245	0.1250	0.1807	0.1768	0.1084	0.1083	0.1573	0.1531
$\sigma=2$								
$n=2^4$	0.8386	0.7906	1.0524	1.1180	0.6360	0.6124	0.8448	0.8660
$n=2^6$	0.3827	0.3953	0.5312	0.5590	0.2968	0.3062	0.4213	0.4330
$n=2^8$	0.1923	0.1976	0.2832	0.2795	0.1513	0.1531	0.2213	0.2165
$\sigma=3$								
$n=2^4$	1.3543	1.1180	1.6015	1.5811	0.9300	0.8292	1.1789	1.1726
$n=2^6$	0.5647	0.5590	0.7581	0.7906	0.4064	0.4146	0.5631	0.5863
$n=2^8$	0.2680	0.2795	0.3944	0.3953	0.2016	0.2073	0.2969	0.2932

1. The following convergences hold a.s. as $T \to \infty$:

$$\frac{1}{T}\int_0^T r_t dt \to \frac{a}{b}, \tag{4.1}$$

$$\frac{1}{T}\int_0^T r_t^2 dt \to \frac{a(a+\sigma^2/2)}{b^2}. \tag{4.2}$$

The convergences (4.1) and (4.2) were proved in (Deelstra and Delbaen 1995, Thm. 1) and (Dehtiar *et al.* 2022, Thm. 4) respectively. Note that under additional condition $2a > \sigma^2$ (when the process r is ergodic with stationary density (2.3)), (4.1) and (4.2) follow from the ergodic theorem, see, e.g., (Dehtiar *et al.* 2022, Cor. 2).

2. For any $q \geq 1$ there exists a constant $C_q > 0$ such that

$$\mathbf{E}\left|\frac{1}{T}\int_0^T r_t dt - \frac{a}{b}\right|^q \leq C_q T^{-q/2}. \tag{4.3}$$

This inequality was proved in (Chernova *et al.* 2024, Lemma 4.3 (1)). It is worth noting that the statement of that lemma includes an additional condition, $2a > \sigma^2$. However, a detailed analysis of the proof indicates that this condition is necessary only for the second part of the lemma, and not for the bound (4.3).

3. For any $p > -2a/\sigma^2$

$$\sup_{t\geq 0} \mathbf{E} r_t^p < \infty, \tag{4.4}$$

see (Ben Alaya and Kebaier 2013, Prop. 3).

Lemma 4.1. *For any $p > 0$ and $q \geq 1$ there exists a constant $C = C_{p,q} > 0$ such that*

$$\mathbf{E}\left|\frac{1}{T}\int_0^T r_s^p \, dW_s\right|^q \leq C T^{-\frac{q}{2}}. \tag{4.5}$$

Proof. In view of Lyapunov's inequality between L_q-norms for different q, it suffices to prove the lemma for $q \geq 2$. Using successively the Burkholder–Davis–Gundy and Hölder inequalities we derive that for any $q \geq 2$,

$$\mathbf{E}\left|\int_0^T r_s^p\, dW_s\right|^q \leq C\mathbf{E}\left(\int_0^T r_s^{2p}\, ds\right)^{\frac{q}{2}} \leq CT^{\frac{q}{2}-1}\int_0^T \mathbf{E} r_t^{pq}\, dt \leq CT^{\frac{q}{2}}, \tag{4.6}$$

where the last inequality follows from (4.4). Dividing both sides of (4.6) by T^q, we get (4.5). ☐

Remark 4.2. For the particular case $p = \frac{1}{2}$, Lemma 4.1 was proved in (Chernova *et al.* 2024, Lemma 4.2).

In the following lemma we express the integrals $\int_0^T r_t dt$ and $\int_0^T r_t^2 dt$ via $\int_0^T r_t^{1/2}\, dW_t$ and $\int_0^T r_t^{3/2}\, dW_t$. As usual, we denote by $\xrightarrow{L_q}$ the convergence in the space $L_q = L_q(\Omega, \mathcal{F}, \mathbf{P})$.

Lemma 4.3. *Let $q \geq 1$.*

(i) For any $T > 0$,

$$T^{-\frac{1}{2}}\int_0^T r_t dt = \frac{a}{b}T^{\frac{1}{2}} + \frac{\sigma}{b}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t + \rho_1(T), \tag{4.7}$$

where

$$\rho_1(T) := \frac{r_0 - r_T}{b\sqrt{T}} \xrightarrow{L_q} 0, \quad \text{as } T \to \infty. \tag{4.8}$$

(ii) For any $T > 0$,

$$T^{-\frac{3}{2}}\left(\int_0^T r_t dt\right)^2 = \frac{a^2}{b^2}T^{\frac{1}{2}} + \frac{2a\sigma}{b^2}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t + \rho_2(T), \tag{4.9}$$

where

$$\rho_2(T) := \frac{\sigma^2}{b^2}T^{-\frac{3}{2}}\left(\int_0^T r_t^{1/2}\, dW_t\right)^2 + \frac{\rho_1^2(T)}{\sqrt{T}} + \frac{2a}{b}\rho_1(T) + \frac{2\sigma}{b}T^{-1}\rho_1(T)\int_0^T r_t^{1/2}\, dW_t \xrightarrow{L_q} 0, \quad \text{as } T \to \infty. \tag{4.10}$$

(iii) For any $T > 0$,

$$T^{-\frac{1}{2}}\int_0^T r_t^2 dt = \frac{a(a+\sigma^2/2)}{b^2}T^{\frac{1}{2}} + \frac{\sigma(a+\sigma^2/2)}{b^2}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t + \frac{\sigma}{b}T^{-\frac{1}{2}}\int_0^T r_t^{3/2}\, dW_t + \rho_3(T), \tag{4.11}$$

where

$$\rho_3(T) := \frac{r_0^2 - r_T^2}{2b\sqrt{T}} + \frac{(a+\sigma^2/2)}{b}\rho_1(T) \xrightarrow{L_q} 0, \quad \text{as } T \to \infty. \tag{4.12}$$

Proof. (i) The representation (4.7) is a direct consequence of the equation (2.1); the convergence (4.8) follows immediately from (4.4).

(ii) By squaring (4.7), we get the representation (4.9). The convergence (4.10) is derived from the bound (4.5) together with (4.8).

(iii) By application of the Itô formula with the function $H(x) = x^2$, we derive from (2.1) that

$$r_T^2 - r_0^2 = \int_0^T \left((a - br_t)\cdot H'(r_t) + \frac{\sigma^2 r_t}{2}H''(r_t)\right) dt + \int_0^T \sigma\sqrt{r_s}H'(r_t)dW_t$$

$$= \int_0^T \left((a - br_t) \cdot 2r_t + \sigma^2 r_t \right) dt + 2\sigma \int_0^T r_t^{\frac{3}{2}} dW_t$$
$$= (2a + \sigma^2) \left(\int_0^T r_t dt - \frac{aT}{b} \right) - 2b \int_0^T r_t^2 dt + \frac{a(2a + \sigma^2)T}{b} + 2\sigma \int_0^T r_t^{\frac{3}{2}} dW_t.$$

We express the integral $\int_0^T r_t^2$ from this equality and get

$$\int_0^T r_t^2 dt = \frac{a(a + \sigma^2/2)}{b^2} T + \frac{r_0^2 - r_T^2}{2b} + \frac{2a + \sigma^2}{2b} \left(\int_0^T r_t dt - \frac{aT}{b} \right) + \frac{\sigma}{b} \int_0^T r_t^{\frac{3}{2}} dW_t. \quad (4.13)$$

Now in order to establish (4.11), it suffices to substitute the representation (4.7) for $\int_0^T r_t \, dt$ into the right-hand side of (4.13). The convergence (4.12) follows from (4.4) and (4.8). □

In the next two lemmas we estimate the rate of convergence of the normalized integrals $\frac{1}{T} \int_0^T r_s^2 \, ds$ and $\frac{1}{T} \int_0^T r_s^3 \, ds$ to their limits. As a corollary of (4.11) we get the following result.

Lemma 4.4. *For any $q \geq 1$ there exists a constant $C_q > 0$ such that*

$$\mathbf{E} \left| \frac{1}{T} \int_0^T r_s^2 \, ds - \frac{a(a + \sigma^2/2)}{b^2} \right|^q \leq C_q T^{-\frac{q}{2}}. \quad (4.14)$$

Proof. By (4.11), we have

$$\mathbf{E} \left| \frac{1}{T} \int_0^T r_t^2 dt - \frac{a(a + \sigma^2/2)}{b^2} \right|^q$$
$$\leq C \left(\frac{\sigma^q (a + \sigma^2/2)^q}{b^{2q}} \mathbf{E} \left| \frac{1}{T} \int_0^T r_t^{1/2} \, dW_t \right|^q + \frac{\sigma^q}{b^q} \cdot \mathbf{E} \left| \frac{1}{T} \int_0^T r_t^{\frac{3}{2}} dW_t \right|^q + T^{-\frac{q}{2}} \mathbf{E} \left| \rho_3(T) \right|^q \right).$$

Now the proof follows from (4.5) and (4.12). □

The following lemma deals with the integral $\int_0^T r_t^3 dt$. Although it is not directly included in the estimators, it represents the quadratic variation of the martingale $int_0^T r_t^{3/2} dW_t$. Therefore, it is essential for subsequent proofs, which rely on the central limit theorem for martingales.

Lemma 4.5. *(i) The integral $\int_0^T r_t^3 dt$ admits the following representation:*

$$\frac{1}{T} \int_0^T r_t^3 dt = \frac{a \left(a + \sigma^2\right) \left(a + \sigma^2/2\right)}{b^3} + \frac{r_0^3 - r_T^3}{3bT}$$
$$+ \frac{a + \sigma^2}{b} \left(\frac{1}{T} \int_0^T r_t^2 dt - \frac{a(a + \sigma^2/2)}{b^2} \right) + \frac{\sigma}{bT} \int_0^T r_t^{\frac{5}{2}} dW_t. \quad (4.15)$$

(ii) Moreover, for any $q \geq 1$ there exists a constant $C = C(q) > 0$ such that

$$\mathbf{E} \left| \frac{1}{T} \int_0^T r_t^3 \, dt - \frac{a \left(a + \sigma^2\right) \left(a + \sigma^2/2\right)}{b^3} \right|^q \leq C T^{-\frac{q}{2}}. \quad (4.16)$$

Proof. (i) Similarly to the proof of Lemma 4.3(iii), we apply the Itô formula with the function $H(x) = x^3$ to the stochastic differential equation (2.1). We obtain

$$r_T^3 - r_0^3 = \int_0^T \left((a - br_t) \cdot H'(r_t) + \frac{\sigma^2 r_t}{2} H''(r_t) \right) dt + \int_0^T \sigma \sqrt{r_s} H'(r_t) dW_t$$
$$= 3 \int_0^T \left(\left(a + \sigma^2\right) r_t^2 - br_t^3 \right) dt + 3\sigma \int_0^T r_t^{\frac{5}{2}} dW_t$$

$$
\begin{aligned}
&= 3\left(a+\sigma^2\right)\left(\int_0^T r_t^2 dt - \frac{a(a+\sigma^2/2)}{b^2}T\right) - 3b\int_0^T r_t^3 dt \\
&\quad + 3\sigma\int_0^T r_t^{\frac{5}{2}} dW_t + \frac{3a(a+\sigma^2)(a+\sigma^2/2)}{b^2}T,
\end{aligned}
$$

whence

$$
\begin{aligned}
\int_0^T r_t^3 dt &= \frac{a(a+\sigma^2)(a+\sigma^2/2)}{b^3}T + \frac{r_0^3 - r_T^3}{3b} \\
&\quad + \frac{a+\sigma^2}{b}\left(\int_0^T r_t^2 dt - \frac{a(a+\sigma^2/2)}{b^2}T\right) + \frac{\sigma}{b}\int_0^T r_t^{\frac{5}{2}} dW_t.
\end{aligned}
$$

Thus (4.15) is proved.

(*ii*) The bound (4.16) is derived from (4.15) similarly to the proof of Lemma 4.4. The equality (4.15) implies

$$
\begin{aligned}
&\mathbf{E}\left|\frac{1}{T}\int_0^T r_t^3\, dt - \frac{a\left(a+\sigma^2\right)\left(a+\sigma^2/2\right)}{b^3}\right|^q \\
&\le C\left(\frac{r_0^{3q} + \mathbf{E} r_T^{3q}}{(3bT)^q} + \left(\frac{a+\sigma^2}{b}\right)^q \mathbf{E}\left|\frac{1}{T}\int_0^T r_t^2 dt - \frac{a(a+\sigma^2/2)}{b^2}\right|^q \right. \\
&\qquad \left. + \left(\frac{\sigma}{b}\right)^q \mathbf{E}\left|\frac{1}{T}\int_0^T r_t^{\frac{5}{2}} dW_t\right|^q\right).
\end{aligned} \tag{4.17}
$$

The first term in the right-hand side of (4.17) can be bounded using (4.4). For the second term, we use the bound from Lemma 4.4, and for the third term, we apply Lemma 4.1. This completes the proof of (4.16). □

Remark 4.6. It is possible to establish the convergence

$$
\frac{1}{T}\int_0^T r_t^3\, dt \to \frac{a\left(a+\sigma^2\right)\left(a+\sigma^2/2\right)}{b^3} \quad \text{a.s., as } T\to\infty, \tag{4.18}
$$

in addition to L_q-convergence, which follows from (4.16) (and is sufficient for further proofs). The proof of (4.18) is derived from (4.16) by a standard application of the Borel–Cantelli lemma, similarly to the proof of the convergence (4.32) in Lemma 4.12 below. Note that in the case $2a > \sigma^2$ (when the process r has ergodic properties), (4.18) follows from the ergodic theorem.

Lemma 4.7. *As* $T \to \infty$

$$
\frac{1}{\sqrt{T}}\begin{pmatrix} \int_0^T r_t^{1/2} dW_t \\ -\int_0^T r_t^{3/2} dW_t \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Gamma),
$$

where

$$
\Gamma = \begin{pmatrix} \frac{a}{b} & -\frac{a}{b^2}\left(a+\frac{\sigma^2}{2}\right) \\ -\frac{a}{b^2}\left(a+\frac{\sigma^2}{2}\right) & \frac{a}{b^3}\left(a+\sigma^2\right)\left(a+\frac{\sigma^2}{2}\right) \end{pmatrix}.
$$

Proof. Note that the vector process

$$
M_T := \begin{pmatrix} \int_0^T r_t^{1/2} dW_t \\ -\int_0^T r_t^{3/2} dW_t \end{pmatrix}, \quad T \ge 0,
$$

is a two-dimensional Brownian martingale with the quadratic variation matrix

$$
\langle M\rangle_T = \begin{pmatrix} \int_0^T r_t dW_t & -\int_0^T r_t^2 dW_t \\ -\int_0^T r_t^2 dW_t & \int_0^T r_t^3 dW_t \end{pmatrix}.
$$

According to (4.1), (4.2) and Lemma 4.5, we have

$$\frac{1}{T}\langle M\rangle_T \xrightarrow{\mathbf{P}} \Gamma, \quad \text{as } T \to \infty.$$

Hence, the required convergence

$$\frac{1}{\sqrt{T}} M_T \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Gamma), \quad \text{as } T \to \infty,$$

follows from the central limit theorem for multidimensional martingales, see, e.g., Heyde (1997, Thm. 12.6). □

4.2. Proof of Theorem 2.2

We split the proof into several steps. Denote

$$D_T := \frac{1}{T}\int_0^T r_t^2\, dt - \left(\frac{1}{T}\int_0^T r_t\, dt\right)^2. \tag{4.19}$$

It follows from (4.1) and (4.2) that

$$D_T \to \frac{a\sigma^2}{2b^2} \quad \text{a.s. as } T \to \infty. \tag{4.20}$$

Further, in the following two lemmas we express $\sqrt{T}(\tilde{a}_T - a)$ and $\sqrt{T}(\tilde{b}_T - b)$ via $\int_0^T r_t^{1/2} dW_t$ and $\int_0^T r_t^{3/2} dW_t$.

Lemma 4.8. *For any $T > 0$,*

$$\sqrt{T}\,(\tilde{a}_T - a) = \frac{a\sigma}{bD_T\sqrt{T}}\left(\frac{a+\sigma^2/2}{b}\int_0^T r_t^{1/2} dW_t - \int_0^T r_t^{3/2} dW_t\right) + R_1(T), \tag{4.21}$$

where

$$R_1(T) \xrightarrow{\mathbf{P}} 0, \quad \text{as } T \to \infty. \tag{4.22}$$

Proof. By the definition (2.4) of the estimator $\tilde{a}_T$, taking into account the notation (4.19), we have

$$\begin{aligned}\sqrt{T}\,(\tilde{a}_T - a) &= \sqrt{T}\left(\frac{\sigma^2}{2}\cdot\frac{\left(\int_0^T r_t\, dt\right)^2}{T\int_0^T r_t^2\, dt - \left(\int_0^T r_t\, dt\right)^2} - a\right) \\ &= \frac{\left(\frac{\sigma^2}{2} + a\right)T^{-3/2}\left(\int_0^T r_t\, dt\right)^2 - aT^{-\frac{1}{2}}\int_0^T r_t^2\, dt}{D_T}. \end{aligned} \tag{4.23}$$

Next, we transform (4.23) by expressing the integrals in the numerator according to Lemma 4.3 (ii)–(iii). We get

$$\begin{aligned}\sqrt{T}\,(\tilde{a}_T - a) &= \frac{1}{D_T}\Bigg[\frac{a^2\,(a+\sigma^2/2)}{b^2}T^{\frac{1}{2}} + \frac{2a\sigma\,(a+\sigma^2/2)}{b^2}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t \\ &\quad + \left(a + \frac{\sigma^2}{2}\right)\rho_2(T) - \frac{a^2(a+\sigma^2/2)}{b^2}T^{\frac{1}{2}} \\ &\quad - \frac{a\sigma(a+\sigma^2/2)}{b^2}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t - \frac{a\sigma}{b}T^{-\frac{1}{2}}\int_0^T r_t^{3/2}\, dW_t - a\rho_3(T)\Bigg] \\ &= \frac{1}{D_T}\Bigg[\frac{a\sigma\,(a+\sigma^2/2)}{b^2}T^{-\frac{1}{2}}\int_0^T r_t^{1/2}\, dW_t - \frac{a\sigma}{b}T^{-\frac{1}{2}}\int_0^T r_t^{3/2}\, dW_t\end{aligned}$$

$$+ \left(a + \frac{\sigma^2}{2}\right) \rho_2(T) - a\rho_3(T)\Big].$$

Hence, (4.21) holds with the following remainder

$$R_1(T) := \frac{1}{D_T}\left[\left(a + \frac{\sigma^2}{2}\right) \rho_2(T) - a\rho_3(T)\right],$$

for which the convergence (4.22) is valid due to (4.10), (4.12) and (4.20). □

Lemma 4.9. *For any* $T > 0$,

$$\sqrt{T}\left(\tilde{b}_T - b\right) = \frac{\sigma}{D_T\sqrt{T}}\left(\frac{a}{b}\int_0^T r_t^{1/2} dW_t - \int_0^T r_t^{3/2} dW_t\right) + R_2(T), \tag{4.24}$$

where

$$R_2(T) \xrightarrow{\mathbf{P}} 0, \quad as\ T \to \infty. \tag{4.25}$$

Proof. From (2.5), (4.19) and Lemma (4.3), we derive the following representation

$$\sqrt{T}\left(\tilde{b}_T - b\right) = \frac{\frac{\sigma^2}{2}T^{-\frac12}\int_0^T r_t\,dt - bT^{-\frac12}\int_0^T r_t^2\,dt + bT^{-\frac32}\left(\int_0^T r_t\,dt\right)^2}{\frac1T\int_0^T r_t^2\,dt - \left(\frac1T\int_0^T r_t\,dt\right)^2}$$

$$= \frac{1}{D_T}\left[\frac{a\sigma}{b}T^{-\frac12}\int_0^T r_t^{1/2}\,dW_t - \sigma T^{-\frac12}\int_0^T r_t^{3/2}\,dW_t + \frac{\sigma^2}{2}\rho_1(T) - b\rho_3(T) + b\rho_2(T)\right],$$

whence we obtain the formula (4.24) with

$$R_2(T) := \frac{1}{D_T}\left[\frac{\sigma^2}{2}\rho_1(T) - b\rho_3(T) + b\rho_2(T)\right].$$

The convergence (4.25) follows from (4.20) and Lemma 4.3. □

Proof of Theorem 2.2. According to Lemmas 4.8 and 4.9, we have the following representation

$$\sqrt{T}\begin{pmatrix}\tilde{a}_T - a\\ \tilde{b}_T - b\end{pmatrix} = \frac{\sigma}{D_T}\begin{pmatrix}\frac{a}{b^2}\left(a + \frac{\sigma^2}{2}\right) & \frac{a}{b}\\ \frac{a}{b} & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{T}}\int_0^T r_t^{1/2}dW_t\\ -\frac{1}{\sqrt{T}}\int_0^T r_t^{3/2}dW_t\end{pmatrix} + \begin{pmatrix}R_1(T)\\ R_2(T)\end{pmatrix}. \tag{4.26}$$

Taking into account the convergence (4.20), Lemma 4.7, and the Slutsky theorem, we derive that the right-hand side of (4.26) converges to the bi-variate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma)$ with

$$\Sigma = \left(\frac{2b^2}{a\sigma}\right)^2\begin{pmatrix}\frac{a}{b^2}\left(a + \frac{\sigma^2}{2}\right) & \frac{a}{b}\\ \frac{a}{b} & 1\end{pmatrix}\begin{pmatrix}\frac{a}{b} & -\frac{a}{b^2}\left(a + \frac{\sigma^2}{2}\right)\\ -\frac{a}{b^2}\left(a + \frac{\sigma^2}{2}\right) & \frac{a}{b^3}\left(a + \sigma^2\right)\left(a + \frac{\sigma^2}{2}\right)\end{pmatrix}\begin{pmatrix}\frac{a}{b^2}\left(a + \frac{\sigma^2}{2}\right) & \frac{a}{b}\\ \frac{a}{b} & 1\end{pmatrix}$$

$$= \begin{pmatrix}\frac{2a}{b}\left(a + \frac{\sigma^2}{2}\right) & 2\left(a + \frac{\sigma^2}{2}\right)\\ 2\left(a + \frac{\sigma^2}{2}\right) & \frac{2b}{a}\left(a + \sigma^2\right)\end{pmatrix}.$$

This completes the proof. □

4.3. Approximation of integrals by sums. Asymptotic behavior of sums

In this subsection, we focus on the sums $\sum_{k=0}^{m_n-1} r_{t_k}$ and $\sum_{k=0}^{m_n-1} r_{t_k}^2$, which are part of the discrete-time estimators (2.7)–(2.8). Our objective is to estimate the difference between these sums and the corresponding integrals $\int_0^T r_t dt$ and $\int_0^T r_t^2 dt$ discussed in subsection 4.1. Subsequently, we will utilize the asymptotic properties of these integrals and continuous-time estimators to establish the corresponding properties of the discrete-time estimators.

Let us recall the following facts from Ben Alaya and Kebaier (2013); Chernova *et al.* (2024):

1. Let $0 \le s < t$ such that $0 < t - s < 1$. Then for all $q \ge 1$, there exists a constant $C_q > 0$ such that
$$\mathbf{E}\,|r_t - r_s|^q \le C_q(t-s)^{\frac{q}{2}}, \tag{4.27}$$
see (Ben Alaya and Kebaier 2013, Prop. 4).

2. According to (Chernova *et al.* 2024, formula (4.17)), for any $q \ge 1$,
$$\mathbf{E}\left|\frac{1}{n}\int_0^T r_t dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right|^q \le Cn^{-\frac{\beta q}{2}} \tag{4.28}$$
for some constant $C > 0$.

3. By (Chernova *et al.* 2024, Lemma 4.6),
$$\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n \to \frac{a}{b} \text{ a.s., as } n \to \infty. \tag{4.29}$$

To prove Theorem 2.4, in addition to the cited results for the process r, we need their analogs for the process r^2. These analogs are stated in the following three lemmas.

Lemma 4.10. *Assume that $0 \le s < t$, such that $0 < t - s < 1$. Then for any $q \ge 1$ there exists a constant $C = C(q) > 0$ such that*
$$\mathbf{E}\left|r_t^2 - r_s^2\right|^q \le C(t-s)^{q/2} \tag{4.30}$$

Proof. Applying the Cauchy–Schwarz inequality, we obtain
$$\mathbf{E}\left|r_t^2 - r_s^2\right|^q = \mathbf{E}\left[(r_t + r_s)^q\,|r_t - r_s|^q\right] \le \sqrt{\mathbf{E}(r_t + r_s)^{2q}\cdot\mathbf{E}\,|r_t - r_s|^{2q}}.$$
The first term under the square root is bounded due to (4.4), and the second term, according to (4.27), admits the upper bound $\mathbf{E}\,|r_t - r_s|^{2q} \le C(t-s)^q$, which completes the proof. □

Lemma 4.11. *For any $q \ge 1$,*
$$\mathbf{E}\left|\frac{1}{n}\int_0^T r_t^2 dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n\right|^q \le Cn^{-\frac{\beta q}{2}}. \tag{4.31}$$

Proof. For $q = 1$ using (4.30) we get
$$\begin{aligned}\mathbf{E}\left|\frac{1}{n}\int_0^n r_t^2 dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n\right| &= \frac{1}{n}\mathbf{E}\left|\int_0^n \sum_{k=0}^{m_n-1}(r_t^2 - r_{t_k}^2)\mathbb{1}_{t\in(t_k,t_{k+1})}dt\right| \\ &\le \frac{1}{n}\sum_{k=0}^{m_n-1}\int_{t_k}^{t_{k+1}}\mathbf{E}\left|r_t^2 - r_{t_k}^2\right|dt \le n^{-1}\cdot C\cdot n^{\beta/2}\cdot n = Cn^{-\beta/2}.\end{aligned}$$
For $q > 1$ we apply Hölder's inequality and obtain
$$\begin{aligned}\mathbf{E}\left|\frac{1}{n}\int_0^n r_t^2 dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n\right|^q &= \frac{1}{n^q}\mathbf{E}\left|\int_0^n \sum_{k=0}^{m_n-1}(r_t^2 - r_{t_k}^2)\mathbb{1}_{t\in(t_k,t_{k+1})}dt\right|^q \\ &\le \frac{1}{n^q}\mathbf{E}\int_0^n\left|\sum_{k=0}^{m_n-1}\left(r_t^2 - r_{t_k}^2\right)\mathbb{1}_{t\in(t_k,t_{k+1})}\right|^q dt\left(\int_0^n dt\right)^{q-1} \\ &\le \frac{1}{n^q}n^{q-1}\sum_{k=0}^{m_n-1}\int_{t_k}^{t_{k+1}}\mathbf{E}\left|r_t^2 - r_{t_k}^2\right|^q \le n^{-1}\cdot C\cdot n^{-\beta q/2}\cdot n = Cn^{-\beta q/2}.\end{aligned}$$
Thus, (4.31) is proved. □

Lemma 4.12. *One has the following convergence:*

$$\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n \to \frac{a(a+\sigma^2/2)}{b^2} \quad a.s., \text{ as } n\to\infty. \tag{4.32}$$

Moreover, for any $q \geq 1$

$$\mathbf{E}\left|\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n - \frac{a(a+\sigma^2/2)}{b^2}\right|^q \leq Cn^{-\frac{(\beta\wedge 1)q}{2}}. \tag{4.33}$$

Proof. Combining Lemmas 4.4 and 4.11, we establish the bound (4.33). Moreover, Lemma 4.11 and Markov's inequality reveal that for any $\varepsilon > 0$ and any $q \geq 1$ the following inequality holds:

$$\mathbf{P}(A_n) := \mathbf{P}\left\{\left|\frac{1}{n}\int_0^n r_t^2 dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n\right| \geq \varepsilon\right\} \leq \frac{C}{\varepsilon^q} n^{-\frac{\beta q}{2}}.$$

Choosing sufficiently large q, we can guarantee that the condition $\sum_{n\geq 1}\mathbf{P}(A_n) < \infty$ is satisfied. Then the Borel–Cantelli lemma yields the convergence

$$\frac{1}{n}\int_0^n r_t^2 dt - \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n \to 0, \quad \text{a.s., as } n\to\infty.$$

Combining this convergence with (4.2) we obtain (4.32). □

Lemma 4.13. *Let* $\beta > 1$*. Then*

$$n^{-\frac{1}{2}}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n = n^{-\frac{1}{2}}\int_0^n r_t dt + \tau_1(n), \tag{4.34}$$

$$n^{-\frac{3}{2}}\left(\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2 = n^{-\frac{3}{2}}\left(\int_0^n r_t dt\right)^2 + \tau_2(n), \tag{4.35}$$

$$n^{-\frac{1}{2}}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n = n^{-\frac{1}{2}}\int_0^n r_t^2 dt + \tau_3(n), \tag{4.36}$$

where

$$\tau_i(n) \xrightarrow{\mathbf{P}} 0 \quad \text{as } n\to\infty, \quad i = 1,2,3. \tag{4.37}$$

Proof. The asymptotic relations (4.34) and (4.36) follow from the bounds (4.28) and (4.31) respectively.

The formula (4.35) is derived from (4.34) as follows:

$$n^{-\frac{3}{2}}\left(\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2 = n^{-\frac{1}{2}}\left(n^{-\frac{1}{2}}\int_0^n r_t dt + \tau_1(n)\right)^2 = n^{-\frac{3}{2}}\left(\int_0^n r_t dt\right)^2 + \tau_2(n),$$

where

$$\tau_2(n) = 2\tau_1(n)\,\frac{1}{n}\int_0^n r_t dt + n^{-\frac{1}{2}}\tau_1^2(n) \xrightarrow{\mathbf{P}} 0 \quad \text{as } n\to\infty,$$

because $\tau_1(n) \xrightarrow{\mathbf{P}} 0$ and $\frac{1}{n}\int_0^n r_t dt \xrightarrow{\mathbf{P}} \frac{a}{b}$ by (4.1). □

4.4. Proof of Theorem 2.4

Proof of strong consistency. In order to prove the strong consistency of $\hat{a}_n$, we divide both the numerator and the denominator of (2.7) by n^2, and use (4.29) and (4.32):

$$\hat{a}_n = \frac{\sigma^2}{2} \cdot \frac{\left(\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2}{\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n - \left(\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2} \to \frac{\sigma^2}{2} \cdot \frac{\left(\frac{a}{b}\right)^2}{\frac{a^2}{b^2} + \frac{a\sigma^2}{2b^2} - \left(\frac{a}{b}\right)^2} = a \quad \text{a.s. as } n \to \infty.$$

The strong consistency of $\hat{b}_n$ is derived from (2.8), (4.29) and (4.32) in the same way. □

Proof of asymptotic normality. Let us denote

$$\widehat{D}_n \coloneqq \frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n - \left(\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2.$$

Then (see the proof of strong consistency)

$$\widehat{D}_n \to \frac{a\sigma^2}{2b^2} \quad \text{a.s. as } T \to \infty, \tag{4.38}$$

and $\sqrt{n}\,(\hat{a}_n - a)$ can be represented as follows

$$\begin{aligned}\sqrt{n}\,(\hat{a}_n - a) &= \sqrt{n}\left(\frac{\frac{\sigma^2}{2}\left(\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2}{\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n - \left(\frac{1}{n}\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2} - a\right)\\ &= \frac{1}{\widehat{D}_n}\left[\left(a + \frac{\sigma^2}{2}\right) n^{-3/2}\left(\sum_{k=0}^{m_n-1} r_{t_k}\delta_n\right)^2 - an^{-1/2}\sum_{k=0}^{m_n-1} r_{t_k}^2\delta_n\right].\end{aligned}$$

Then using Lemma 4.13 and the representation (4.23), we get

$$\begin{aligned}\sqrt{n}\,(\hat{a}_n - a) &= \frac{1}{\widehat{D}_n}\Bigg[\left(a + \frac{\sigma^2}{2}\right) n^{-\frac{3}{2}}\left(\int_0^n r_t dt\right)^2 - an^{-1/2}\int_0^n r_t^2 dt\\ &\qquad + \left(a + \frac{\sigma^2}{2}\right)\tau_2(n) - a\tau_3(n)\Bigg]\\ &= \frac{D_n}{\widehat{D}_n}\sqrt{n}\,(\tilde{a}_n - a) + Q_1(n),\end{aligned} \tag{4.39}$$

where

$$Q_1(n) \coloneqq \frac{\left(a + \frac{\sigma^2}{2}\right)\tau_2(n) - a\tau_3(n)}{\widehat{D}_n} \xrightarrow{\mathbf{P}} 0 \quad \text{as } n \to \infty,$$

by (4.37) and (4.38). Similarly, we can prove that

$$\sqrt{n}\left(\hat{b}_n - b\right) = \frac{D_n}{\widehat{D}_n}\sqrt{n}\left(\tilde{b}_n - b\right) + Q_2(n) \tag{4.40}$$

with

$$Q_2(n) \coloneqq \frac{\frac{\sigma^2}{2}\tau_1(n) - b\tau_3(n) + b\tau_2(n)}{\widehat{D}_n} \xrightarrow{\mathbf{P}} 0 \quad \text{as } n \to \infty.$$

Note that $D_n/\widehat{D}_n \xrightarrow{\mathbf{P}} 1$ as $n \to \infty$, by (4.20) and (4.38). Thus, now the convergence (2.9) follows from representations (4.39)–(4.40) and Theorem 2.2 by the Slutsky theorem. □

Acknowledgment

The second author is supported by the Research Council of Finland, decision number 359815.

References

Alfonsi A (2015). *Affine Diffusions and Related Processes: Simulation, Theory and Applications*, volume 6 of *Bocconi & Springer Series*. Springer and Bocconi University Press.

Ben Alaya M, Kebaier A (2012). "Parameter Estimation for the Square-root Diffusions: Ergodic and Nonergodic Cases." *Stochastic Models*, **28**(4), 609–634.

Ben Alaya M, Kebaier A (2013). "Asymptotic Behavior of the Maximum Likelihood Estimator for Ergodic and Nonergodic Square-root Diffusions." *Stochastic Analysis and Applications*, **31**(4), 552–573. doi:10.1080/07362994.2013.798175.

Bishwal JPN (2008). *Parameter Estimation in Stochastic Differential Equations*, volume 1923 of *Lecture Notes in Mathematics*. Springer, Berlin. doi:10.1007/978-3-540-74448-1.

Cheng Y, Hufnagel N, Masuda H (2022). "Estimation of Ergodic Square-root Diffusion under High-frequency Sampling." *Econometrics and Statistics*. doi:10.1016/j.ecosta.2022.05.003.

Chernova O, Dehtiar O, Mishura Y, Ralchenko K (2024). "Rate of Convergence of Discretized Drift Parameters Estimators in the Cox–Ingersoll–Ross Model." *Communications in Statistics. Theory and Methods*, **53**(13), 4857–4879. doi:10.1080/03610926.2023.2196591.

Cox JC, Ingersoll Jr JE, Ross SA (1985). "A Theory of the Term Structure of Interest Rates." *Econometrica*, **53**(2), 385–407. doi:10.2307/1911242.

De Rossi G (2010). "Maximum Likelihood Estimation of the Cox–Ingersoll–Ross Model Using Particle Filters." *Computational Economics*, **36**(1), 1–16.

Deelstra G, Delbaen F (1995). "Long-term Returns in Stochastic Interest Rate Models." *Insurance: Mathematics & Economics*, **17**(2), 163–169. doi:10.1016/0167-6687(95)00018-N.

Dehtiar O, Mishura Y, Ralchenko K (2022). "Two Methods of Estimation of the Drift Parameters of the Cox-Ingersoll-Ross process: Continuous Observations." *Communications in Statistics. Theory and Methods*, **51**(19), 6818–6833. doi:10.1080/03610926.2020.1866611.

Dokuchaev N (2017). "A Pathwise Inference Method for the Parameters of Diffusion Terms." *Journal of Nonparametric Statistics*, **29**(4), 731–743.

Göing-Jaeschke A, Yor M (2003). "A Survey and Some Generalizations of Bessel Processes." *Bernoulli*, **9**(2), 313–349. doi:10.3150/bj/1068128980.

Heyde CC (1997). *Quasi-likelihood and Its Application. A General Approach to Optimal Parameter Estimation.* Springer Series in Statistics. Springer-Verlag, New York. doi:10.1007/b98823.

Iacus SM (2008). *Simulation and Inference for Stochastic Differential Equations. With R Examples.* Springer Series in Statistics. Springer, New York. doi:10.1007/978-0-387-75839-8.

Jacod J, Shiryaev AN (1987). *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin.

Kladívko K (2007). "Maximum Likelihood Estimation of the Cox–Ingersoll–Ross Process: the Matlab Implementation." *Technical Computing Prague*, **7**.

Kutoyants YA (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer Series in Statistics. Springer-Verlag London, Ltd., London. doi:10.1007/978-1-4471-3866-2.

Liptser RS, Shiryaev AN (2001). *Statistics of Random Processes. II. Applications*, volume 6 of *Applications of Mathematics*. Springer-Verlag, Berlin.

Mishura Y, Ralchenko K, Dehtiar O (2022). "Parameter Estimation in CKLS Model by Continuous Observations." *Statistics & Probability Letters*, **184**, Paper No. 109391, 10. doi:10.1016/j.spl.2022.109391.

Overbeck L (1998). "Estimation for Continuous Branching Processes." *Scandinavian Journal of Statistics. Theory and Applications*, **25**(1), 111–126. doi:10.1111/1467-9469.00092.

Overbeck L, Rydén T (1997). "Estimation in the Cox–Ingersoll–Ross Model." *Econometric Theory*, **13**(3), 430–461.

Sørensen H (2002). "Estimation of Diffusion Parameters for Discretely Observed Diffusion Processes." *Bernoulli*, **8**(4), 491–508.

Tang CY, Chen SX (2009). "Parameter Estimation and Bias Correction for Diffusion Processes." *Journal of Econometrics*, **149**(1), 65–81. doi:10.1016/j.jeconom.2008.11.001.

Affiliation:

Olha Prykhodko
Department of Probability, Statistics and Actuarial Mathematics
Taras Shevchenko National University of Kyiv
64/13, Volodymyrs'ka St., 01601 Kyiv, Ukraine

Kostiantyn Ralchenko
Department of Probability, Statistics and Actuarial Mathematics
Taras Shevchenko National University of Kyiv
64/13, Volodymyrs'ka St., 01601 Kyiv, Ukraine
E-mail: kostiantynralchenko@knu.ua

School of Technology and Innovations
University of Vaasa
P.O. Box 700, Wolffintie 34, FIN-65101 Vaasa, Finland
E-mail: kostiantyn.ralchenko@uwasa.fi

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 54 *Submitted:* 2024-06-08
2025 *Accepted:* 2024-06-20