Driven by Brownian motion Cox–Ingersoll–Ross and squared Bessel processes: Interaction and phase transition

Driven by Brownian motion Cox–Ingersoll–Ross and squared Bessel processes: interaction and phase transition

Yuliya Mishura,[1] Kostiantyn Ralchenko,[1, 2, a)] and Svitlana Kushnirenko[3]
[1)]*Department of Probability, Statistics and Actuarial Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrska St., Kyiv, 01601, Ukraine.*
[2)]*School of Technology and Innovations, University of Vaasa, P.O. Box 700, Vaasa, FIN-65101, Finland.*
[3)]*Department of General Mathematics, Taras Shevchenko National University of Kyiv, Volodymyrska St., Kyiv, 01601, Ukraine.*

(*Electronic mail: yuliyamishura@knu.ua, kostiantynralchenko@knu.ua, svilana_kushnirenko@knu.ua)

(Dated: December 9, 2024)

This paper studies two related stochastic processes driven by Brownian motion: the Cox–Ingersoll–Ross (CIR) process and the Bessel process. We investigate their shared and distinct properties, focusing on time-asymptotic growth rates, distance between the processes in integral norms, and parameter estimation. The squared Bessel process is shown to be a phase transition of the CIR process and can be approximated by a sequence of CIR processes. Differences in stochastic stability are also highlighted, with the Bessel process displaying instability, while the CIR process remains ergodic and stable.

I. INTRODUCTION

A. Some historical information

Probability theory (stochastics), which deals with the description and analysis of stochastic objects, is connected with the most diverse phenomena, manifestations of the surrounding reality and with technical phenomena. Stochastics creates theoretical basis of applied applications, that is, adequate models of existing phenomena. At the same time, the created model should allow a simple but flexible image with the help of certain mathematical formulas. Once a good model is created, further development proceeds in two directions: first, the theory of related and more general models itself begins to develop, and second, the model begins to cover more and more applications, since approximately the same regularities are manifested in physics, financial mathematics, biology, economics, climatology, and even in some social sciences. Therefore, the same model or any of its variants can be used in a wide variety of areas. Furthermore, as it is well-known, Brownian motion (Wiener process) is the best possibility to involve randomness into the model.

In 1827, the botanist Robert Brown (1773–1858) first observed the phenomenon, which was later called Brownian motion. A brief description of the observed movement is as follows: imagine a laboratory dish, something like a cup, in which there is a liquid, and in the liquid is poured flower pollen. So, Brown observed the movement of pollen particles in this cup. The phenomenon so impressed him that he described it in detail. Namely, according to observations, the specified movement was chaotic, the trajectories of the particles constantly changed direction, were broken, in fact, at no moment in time did any particle have a fixed direction of movement. Robert Brown could not create a certain simple, non-random description of this system, but only noted that the movement of pollen in a liquid is described by a very chaotic process. From a modern point of view, he actually observed a random process that took on a value of a small part of the plane (the surface of the cup) and that changed in time non-deterministically. At that time, there were no explanations for this phenomenon. They appeared much later.

The next step in the development of the theory of Brownian motion was made by outstanding physicists. Namely, in 1905–1906, the famous works of Albert Einstein (1879–1955) and Marian Smoluchowski (1872–1917),[1] and[2], were published independently of each other; then their original German papers on the specified topic, in which the phenomenon of Brownian motion was explained, were combined into a collection[3]. English translations of Einstein's papers were published in[4]. Einstein and Smoluchowski explained, in particular, the random movement of flower pollen in a liquid by such a phenomenon as the thermal chaotic movement of atoms and molecules.

According to this theory, liquid or gas molecules are in constant thermal motion, and the impulses of different molecules are unequal in magnitude and direction. If the surface of a particle placed in such an environment is small, as is the case for a Brownian particle, in particular, a particle of flower pollen, then the shocks felt by the particle from the surrounding molecules will not be precisely compensated. Therefore, as a result of "bombardment" by molecules, the Brownian particle begins to move randomly, changing the magnitude and direction of its velocity approximately 10^{14} times per second. This is how a seemingly purely biological phenomenon was explained from a physical point of view.

It also turned out that the random process B_t is needed to describe this thermal motion and to create this kinetic theory, i.e. Brownian motion is an adequate model for this kinetic theory. Einstein and Smoluchowski not only described this random process from the point of view of its growth and from the point of view of its dependence on cases, but they found the so-called distribution of process values and formulated a par-

[a)]Corresponding author

tial differential equation that is satisfied by the transition density of the distribution of Brownian motion. It turned out that Brownian motion has a Markov property. This was written using so-called transition probabilities, or rather using partial derivative equations for densities, that is, derivatives of transition probabilities. These equations showed that Brownian motion has a Markov property: independence of the future from the past with a fixed current. This property significantly helps to simplify at least the formulas related to transition probabilities.

Thus, the theory of Brownian motion found a serious reinforcement from physicists and became "overgrown" with interesting properties. However, the real world cannot always be described by a linear model, it is much more complex. Therefore, integration over Brownian motion was constructed first with non-random (N. Wiener), and then with random (K. Itô) integrands, and the theory of stochastic differential equations driven by the Brownian motion was developed. The existence, uniqueness and properties of solutions of such equations, depending on their drift and diffusion coefficients, are thoroughly described and classified in books[5,6]. One very interesting class of equations are those that have a unique nonnegative solution. Such a solution has a natural physical (or financial) meaning. This class includes the Cox–Ingersoll–Ross and Bessel processes, which are solutions of stochastic differential equations with a diffusion coefficient equal to $\sigma\sqrt{x}$ and which differ only in the drift coefficient. However, this drift coefficient, without changing their non-negativity, significantly affects their asymptotic properties. These processes are discussed in more detail in the next sections.

B. Overview of the results presented in the paper

In the present paper we shall consider two closely related stochastic processes, namely, Cox–Ingersoll–Ross and Bessel process, both of them being strictly positive solutions of the respective stochastic differential equations. Strictly positive values make them convenient to model real processes in physics, biology, economics. In finances they are used to forecast interest rates and in bond pricing models, see e.g.[7–10]. Similar models are used to simulate changes in the membrane voltage of a neuron[11]. In our research we combine the methods of stochastic analysis and methods based on the explicit formulas for probability distributions of CIR and Bessel processes.

In a certain sense, the square of the Bessel process can be considered the result of a phase transition in the Cox–Ingersoll–Ross process. We underline their common and distinct properties. More precisely, we begin by presenting several results that provide upper and lower bounds for the time-asymptotic growth rates of both processes. These bounds exhibit notable similarities between the two models.

Next, we explore the approximation of CIR and squared Bessel processes by a sequence of CIR processes. We prove the convergence of this sequence in integral norms, assuming that the corresponding coefficients converge. Additionally, we establish upper bounds on the rate of convergence. It turns out that the CIR and squared Bessel processes are closely related, as the squared Bessel process can be represented as the limit of a sequence of CIR processes. However, as anticipated, the upper bounds involve coefficients that depend on the length of the time interval and tend to infinity as the interval length increases. In this sense, the processes diverge, or, in other words, they move apart. Nevertheless, the coefficients can be sufficiently close such that, over slowly increasing time intervals, the processes remain comparable.

We then apply this approximation to the problem of parameter estimation for the squared Bessel process using the maximum likelihood method. To establish the strong consistency of the constructed drift parameter estimator, we approximate the squared Bessel process by a sequence of CIR processes, for which the necessary convergence can be derived from their ergodic properties. Furthermore, we show how to estimate the diffusion coefficient of the process based on the realized quadratic variations.

Finally, we investigate both processes using the concept of stochastic instability. From this perspective, the properties of the squared Bessel and CIR processes are fundamentally different. We demonstrate that the squared Bessel process exhibits stochastic instability, whereas the CIR process is ergodic and, in this sense, stochastically stable. In addition, we consider an alternative sequence of approximations for the squared Bessel process and show that these approximating processes are also stochastically unstable. Moreover, we prove that, when appropriately normalized, this sequence converges to the (non-squared) Bessel process.

C. Structure of the paper

The remainder of the paper is organized as follows. In Section II, we introduce the CIR and squared Bessel processes as unique solutions to the corresponding stochastic differential equations. We also provide preliminary information on their distributional and pathwise properties that are essential for the subsequent sections. In particular, this section presents formulas for their densities and the first three moments.

Section III contains several results that describe the growth rates of the CIR and squared Bessel processes as functions of time and their coefficients. Section IV focuses on the distance between the CIR and squared Bessel processes in terms of integral norms, expressed through their coefficients. We establish the rate of convergence of CIR processes to either CIR or squared Bessel processes in integral norms over any fixed interval, under the condition that the respective coefficients converge.

Section V addresses the statistical problem of identifying the Bessel process from continuous-time observations of its trajectory. Finally, in Section VI, we demonstrate that the squared Bessel process is stochastically unstable, in contrast to the CIR process. This section also presents some functional limit theorems.

For the reader's convenience, several auxiliary results used in the proofs are provided in the appendix. Specifically, we include definitions and properties of special functions, a limit theorem for stochastic differential equations with non-Lipschitz diffusion terms, as well as additional limit theorems for solutions of stochastic differential equations.

II. PRELIMINARIES

Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space and $W = \{W_t, t \geq 0\}$ be a Wiener process on it. In this paper we study two stochastic differential equations, namely

$$X_t = x_0 + \int_0^t (a - bX_s)\, ds + \sigma \int_0^t \sqrt{X_s}\, dW_s, \tag{1}$$

and

$$Y_t = y_0 + at + \sigma \int_0^t \sqrt{Y_s}\, dW_s, \tag{2}$$

where $x_0 > 0$, $y_0 > 0$, $a > 0$, $b \geq 0$, and $\sigma > 0$.

It is well known that both equations (1) and (2) admit unique non-negative strong solutions, $X = \{X_t, t \geq 0\}$ and $Y = \{Y_t, t \geq 0\}$, respectively.

The process $X = \{X_t, t \geq 0\}$ was introduced in[12] for the purpose of interest rates modeling. It is commonly referred to as the Cox–Ingersoll–Ross (CIR) process. The process $Y = \{Y_t, t \geq 0\}$ is the squared Bessel process, see, e.g.,[13] or[14] (Chapter XI) for details.

It follows from the comparison theorem[15] (Proposition 2.18, p. 293) that if $x_0 \leq y_0$, then

$$\mathbf{P}\left(X_t \leq Y_t \text{ for all } t \geq 0\right) = 1.$$

In what follows we additionally assume that $2a \geq \sigma^2$. In this case, the trajectories of both processes X and Y with probability 1 remain strictly positive, while in the case $0 < a < \sigma^2/2$, they almost surely hit zero, where the state 0 is instantaneously reflecting (see, e.g., classical paper[13] and the more recent ones[16,17] for more details). For the sake of technical simplicity, we assume throughout the paper that

$$2a > \sigma^2.$$

A. Distributional and path-wise properties of CIR process

It is well known[12] that X_t follows a non-central chi-squared distribution with the following probability density function:

$$p_t(x) = \frac{1}{c(t)} \left(\frac{x}{x_0 e^{-bt}}\right)^{\nu/2} \times \exp\left\{-\frac{x + x_0 e^{-bt}}{c(t)}\right\} I_\nu \left(\frac{2e^{-bt/2}\sqrt{x x_0}}{c(t)}\right) \mathbb{1}_{x>0},$$

where

$$c(t) = \frac{\sigma^2}{2b}\left(1 - e^{-bt}\right), \quad \nu = \frac{2a}{\sigma^2} - 1,$$

and I_ν is the modified Bessel function of the first kind. For $\nu > -1$ and $x \in \mathbb{R}$ this function is defined by the following power series[18] (Formula 50:6:1):

$$I_\nu(x) = \sum_{j=0}^{\infty} \frac{(x/2)^{2j+\nu}}{j!\,\Gamma(j+1+\nu)},$$

where Γ stands for the Gamma function. Obviously, $I_\nu(0) = 0$ for all $\nu > 0$, more precisely I_ν has the following behavior as $x \to 0$:

$$I_\nu(x) \sim \frac{(x/2)^\nu}{\Gamma(\nu+1)}.$$

Using this relation, one can show that, as $t \to \infty$,

$$p_t(x) \to \frac{(2b/\sigma^2)^{2a/\sigma^2}}{\Gamma(2a/\sigma^2)} x^{2a/\sigma^2 - 1} e^{-2bx/\sigma^2} \mathbb{1}_{x>0} =: p_\infty(x) \tag{3}$$

Note that the limiting distribution is a Gamma distribution.

Moreover, the CIR process X is ergodic[12] (see also[19] (Section 1.2) and[20]). Ergodicity implies that for any function $f \in L^1(\mathbb{R}, p_\infty(x)dx)$, the time average $\frac{1}{T}\int_0^T f(X_t)dt$ converges a.s. to the space average $\int_{\mathbb{R}} f(x) p_\infty(x) dx$, as $T \to \infty$. In particular, for $a > \frac{\sigma^2}{2}$,

$$\frac{1}{T}\int_0^T \frac{dt}{X_t} \to \int_{\mathbb{R}} \frac{p_\infty(x)}{x} dx = \frac{b}{a - \sigma^2/2}, \quad \text{a. s., when } T \to \infty. \tag{4}$$

The first two moments of X_t are equal to

$$\mathbf{E}X_t = x_0 e^{-bt} + \frac{a}{b}\left(1 - e^{-bt}\right), \tag{5}$$

and

$$\mathbf{E}X_t^2 = \frac{x_0(\sigma^2 + 2a)}{b}\left(e^{-bt} - e^{-2bt}\right) + \frac{a(\sigma^2 + 2a)}{2b^2}\left(1 - e^{-bt}\right)^2 + x_0^2 e^{-2bt}. \tag{6}$$

The formula for higher moments of the CIR processes is presented in[21] (Proposition 1). In particular,

$$\mathbf{E}X_t^3 = x_0^3 e^{-3bt} + \left(1 + \frac{3\sigma^2}{2a} + \frac{\sigma^4}{2a^2}\right) \times \left(\frac{a^3}{b^3}\left(1 - e^{-bt}\right)^3 + \frac{3x_0 a^2}{b^2}\left(e^{-bt} - 2e^{-2bt} + e^{-3bt}\right)\right) + \frac{3x_0^2 a}{b}\left(1 + \frac{\sigma^2}{a}\right)\left(e^{-2bt} - e^{-3bt}\right).$$

B. Distributional and path-wise properties of squared Bessel process

The probability density function of the squared Bessel process Y_t is given by

$$g_t(x) = \frac{1}{2}\left(\frac{x}{y_0}\right)^{\nu/2} \exp\left\{-\frac{2(x + y_0)}{\sigma^2 t}\right\} I_\nu\left(\frac{4\sqrt{x y_0}}{\sigma^2 t}\right) \mathbb{1}_{x>0},$$

where, as before, $\nu = \frac{2a}{\sigma^2} - 1$ (see, e.g.,[14] (Chapter XI, Corollary (1.4))).

Unlike the CIR process X, the squared Bessel process Y is non-ergodic. For a detailed discussion on the properties of squared Bessel processes, we refer the reader to[14] (Chapter XI). A comparison of the properties of both ergodic and non-ergodic processes, X and Y, can be found in[22].

Since $I_\nu(0) = 0$, we see that

$$g_t(x) \to 0, \quad t \to \infty. \tag{7}$$

Therefore, for the squared Bessel process, the limiting distribution does not exist.

The first and second moments of Y_t are equal to:

$$\mathbf{E}Y_t = y_0 + at, \quad \mathbf{E}Y_t^2 = y_0^2 + \left(\frac{\sigma^2}{2} + a\right)\left(2y_0t + at^2\right). \tag{8}$$

Both equalities can be derived directly from the equation (2), taking into account that the stochastic integral $\int_0^t \sqrt{Y_s}\, dW_s$ is a square-integrable martingale with zero mean whose second moment equals $\int_0^t \mathbf{E}Y_s\, ds$.

Remark II.1. We see from (5) and (6) that the first two moments of the CIR process exist for all t. Moreover, they are totally bounded. Indeed, as it was established in[22] (Proposition 3), $\sup_{t\geq 0} \mathbf{E}X_t^p < \infty$ for any $p > -2a/\sigma^2$. In contrast, the first and second moments of the squared Bessel process exhibit linear and quadratic growth with respect to t, respectively.

The general formula for the moments of the Bessel process has the following form: for any $p \geq -\frac{2a}{\sigma^2}$

$$\mathbf{E}Y_t^p = \left(\frac{\sigma^2 t}{2}\right)^p \frac{\Gamma\left(\frac{2a}{\sigma^2}+p\right)}{\Gamma\left(\frac{2a}{\sigma^2}\right)} \exp\left\{-\frac{2y_0}{\sigma^2 t}\right\} {}_1F_1\left(\frac{2a}{\sigma^2}+p, \frac{2a}{\sigma^2}, \frac{2y_0}{\sigma^2 t}\right),$$

see the proof of Proposition 3 in[22]. Here ${}_1F_1$ is the confluent hypergeometric function, see Appendix A. Using formula (A1) in Appendix, we can derive for $p = 3$

$$\begin{aligned}\mathbf{E}Y_t^3 &= \left(\frac{a\sigma^4}{2} + \frac{3a^2\sigma^2}{2} + a^3\right)t^3 \\ &+ 3\left(\frac{y_0\sigma^4}{2} + \frac{3ay_0\sigma^2}{2} + a^2y_0\right)t^2 + 3y_0^2\left(\sigma^2 + a\right)t + y_0^3. \end{aligned}\tag{9}$$

III. TIME-ASYMPTOTIC GROWTH RATE FOR CIR AND QUADRATIC BESSEL PROCESSES

Now we establish several results that provide a growth rate for the solution to equations (2) and (1), as the function of time and coefficients. As usual, time is included in the constants, since the time interval is fixed in many situations, but for us it is the asymptotic behaviour of functionals of solutions that is most important. We demonstrate what growth rates can be obtained by different methods, and compare the results. The first result follows from the Grönwall inequality therefore gives an exponential growth rate. This growth rate is determined by coefficients a, x_0 and σ and is valid both for solution to (1) or (2).

Proposition III.1. *Let $Z = \{Z_t, t \geq 0\}$ be a unique solution to (1) or (2) (i.e., $Z = X$ or $Z = Y$). Then, for all $t \geq 0$,*

$$\mathbf{E}\left(\sup_{s\leq t} Z_s\right)^2 \leq 2\left((x_0 + at)^2 + 2\sigma^2 t\right)e^{4\sigma^2 t}. \tag{10}$$

Proof. Consider the equation (1), i.e., $Z = X$. Define

$$\tau_N := \inf\{t \geq 0 : X_t \geq N\}, \quad N \geq 1.$$

Then

$$\begin{aligned} 0 \leq X_{t\wedge\tau_N} &= x_0 + a(t\wedge\tau_N) - b\int_0^{t\wedge\tau_N} X_u\, du \qquad (11)\\ &+ \sigma\int_0^{t\wedge\tau_N}\sqrt{X_u}\, dW_u \\ &\leq x_0 + a(t\wedge\tau_N) + \sigma\int_0^{t\wedge\tau_N}\sqrt{X_u}\, dW_u, \qquad (12)\end{aligned}$$

which implies

$$\begin{aligned}&\mathbf{E}\left(\sup_{s\leq t} X_{s\wedge\tau_N}\right)^2 \\ &\leq 2\left((x_0+at)^2 + \sigma^2 \mathbf{E}\sup_{s\leq t}\left(\int_0^{s\wedge\tau_N}\sqrt{X_u}\, dW_u\right)^2\right). \end{aligned}\tag{13}$$

Since

$$\begin{aligned}\mathbf{E}\left(\int_0^{t\wedge\tau_N}\sqrt{X_s}\, dW_s\right)^2 &= \mathbf{E}\left(\int_0^t \mathbb{1}_{s\leq\tau_N}\sqrt{X_s}\, dW_s\right)^2 \\ &= \int_0^t \mathbf{E}\left(\mathbb{1}_{s\leq\tau_N}\sqrt{X_s}\right)^2 ds \leq Nt,\end{aligned}$$

we see that $\int_0^{t\wedge\tau_N}\sqrt{X_s}\, dW_s$ is a square-integrable martingale. Therefore, by the Doob maximal quadratic inequality,

$$\begin{aligned}\mathbf{E}\sup_{s\leq t}\left(\int_0^{s\wedge\tau_N}\sqrt{X_u}\, dW_u\right)^2 &\leq 4\mathbf{E}\int_0^t X_{s\wedge\tau_N}\, ds \\ &\leq 2\mathbf{E}\int_0^t \left(1 + X_{s\wedge\tau_N}^2\right) ds \\ &\leq 2t + 2\int_0^t \mathbf{E}\left(\sup_{u\leq s} X_{u\wedge\tau_N}\right)^2 ds. \end{aligned}\tag{14}$$

Combining (13)–(14), we derive the inequality

$$\begin{aligned}&\mathbf{E}\left(\sup_{s\leq t} X_{s\wedge\tau_N}\right)^2 \\ &\leq 2\left((x_0+at)^2 + 2\sigma^2 t + 2\sigma^2\int_0^t \mathbf{E}\left(\sup_{u\leq s} X_{u\wedge\tau_N}\right)^2 ds\right).\end{aligned}$$

By the Grönwall inequality,

$$\mathbf{E}\left(\sup_{s\leq t} X_{s\wedge\tau_N}\right)^2 \leq 2\left((x_0+at)^2 + 2\sigma^2 t\right)e^{4\sigma^2 t}.$$

Letting $N \to \infty$ concludes the proof for X. The equation (2) is considered similarly. $\square$

Remark III.2. Similarly to the above proof, one can establish that for the solution $X = \{X_t, t \geq 0\}$ of the equation (1), the following inequality holds:

$$\mathbf{E}\sup_{s\leq t}\left(X_s + b\int_0^s X_u\,du\right)^2 \leq 2\left((x_0+at)^2 + 2\sigma^2 t\right)e^{4\sigma^2 t}.$$

The disadvantage of these upper bounds is that being exponential, they grow quickly in time t, but their advantage is that they do not depend on coefficient b.

Now our goal is to obtain the upper bounds for $\mathbf{E}\sup_{s\leq t}(X_s + b\int_0^s X_u\,du)^2$ that will grow not so quickly. We will not apply the Grönwall inequality that always gives exponential growth, but apply directly the values of the moments of CIR-process. Oppositely to previous bounds, the next ones will depend on b and the next goal will be to analyze this dependence when $b \downarrow 0$.

Proposition III.3. *Let $X = \{X_t, t \geq 0\}$ be a solution to* (1). *Then, for all $t \geq 0$,*

$$\mathbf{E}\sup_{s\leq t}\left(X_s + b\int_0^s X_u\,du\right)^2 \leq 2(x_0+at)^2 + \frac{8\sigma^2}{b}\left(x_0 - \frac{a}{b}\right)\left(1-e^{-bt}\right) + \frac{8\sigma^2 a}{b}t. \quad (15)$$

Proof. Similarly to the proof of Proposition III.1, we get

$$\begin{aligned}\mathbf{E}\sup_{s\leq t}\left(X_s + b\int_0^s X_u\,du\right)^2 &\leq 2(x_0+at)^2 + 2\sigma^2\mathbf{E}\sup_{s\leq t}\left(\int_0^s \sqrt{X_u}\,dW_u\right)^2\\ &\leq 2(x_0+at)^2 + 8\sigma^2\int_0^t \mathbf{E}X_s\,ds.\end{aligned}$$

Integrating (5), we obtain

$$\int_0^t \mathbf{E}X_s\,ds = \frac{1}{b}\left(x_0 - \frac{a}{b}\right)\left(1-e^{-bt}\right) + \frac{a}{b}t. \quad (16)$$

Combining these two formulas we conclude the proof. □

Remark III.4. It follows from (6) and (16) that

$$\begin{aligned}\mathbf{E}\sup_{s\leq t}\left(X_s + b\int_0^s X_u\,du\right)^2 &\geq \mathbf{E}X_t^2 + b^2\mathbf{E}\left(\int_0^t X_u\,du\right)^2\\ &\geq \mathbf{E}X_t^2 + b^2\left(\int_0^t \mathbf{E}X_u\,du\right)^2\\ &= \frac{x_0(\sigma^2+2a)}{b}\left(e^{-bt} - e^{-2bt}\right) + \frac{a(\sigma^2+2a)}{2b^2}\left(1-e^{-bt}\right)^2\\ &\quad + x_0^2 e^{-2bt} + \left(\left(x_0 - \frac{a}{b}\right)\left(1-e^{-bt}\right) + at\right)^2\end{aligned}$$

Comparing this lower bound with the upper bound in (15), we observe that both the upper and lower bounds for $\mathbf{E}\sup_{s\leq t}(X_s + b\int_0^s X_u, du)^2$ exhibit a quadratic rate of growth, as $t \to \infty$. The upper bound follows the asymptotic behavior $2a^2t^2 + 4a(x_0 + \frac{2\sigma^2}{b})t + O(1)$, while the lower bound behaves as $a^2t^2 + 2a(x_0 - \frac{a}{b})t + O(1)$.

Remark III.5. (i) Arguing as in the proof of Proposition III.3 and using (8) instead of (5), we get the bound

$$\mathbf{E}\sup_{s\leq t} Y_s^2 \leq 2(y_0+at)^2 + 4\sigma^2\left(2y_0 t + at^2\right). \quad (17)$$

Note that the right-hand side of (17) is a limit, as $b \downarrow 0$, of the right-hand side of (15). This can be easily seen from the relation $1 - e^{-bt} = bt - \frac{b^2t^2}{2} + o(b^3)$, $b \downarrow 0$. However, if to fix b and consider the asymptotic behaviour, as $t \to \infty$, of the right-hand sides of (15) and (17), we see that the main part of the right-hand side of (15) equals $2a^2t^2$ while the main part of the right-hand side of (17) equals $(2a^2 + 4a\sigma^2)t^2$. It means (a bit unexpectedly) that the difference between asymptotic behaviour of $\mathbf{E}\sup_{s\leq t} Y_s^2$ and $\mathbf{E}\sup_{s\leq t} X_s^2$ is very significantly determined by the diffusion coefficient σ as well as (more expected) of the drift coefficient a. Of course, this difference in some latent way depends on b because the value $\frac{8\sigma^2}{b}\left(x_0 - \frac{a}{b}\right)\left(1-e^{-bt}\right)$ in the right-hand side of (15) is bounded in t for any $b > 0$, however, it is growing as t^2 if we come to the limit, as $b \to 0$.

(ii) The second formula in (8) implies the following lower bound:

$$\mathbf{E}\sup_{s\leq t} Y_s^2 \geq \mathbf{E}Y_t^2 = (y_0+at)^2 + \frac{\sigma^2}{2}\left(2y_0 t + at^2\right). \quad (18)$$

It can be observed that, compared to the upper bound in (17), this lower bound contains the same terms, but with smaller coefficients. Therefore, we conclude that $\mathbf{E}\sup_{s\leq t} Y_s^2$ grows quadratically as a function of t.

(iii) Let us compare two upper bounds for the squared Bessel process, specifically (10) and (17). On the one hand, the bound given by (17) is clearly more advantageous for large values of t. On the other hand, when t is near zero, the bound (10) may provide greater accuracy. More precisely, the bound (10) is superior to (17) if and only if $e^{4\sigma^2 t} \leq 2x_0 + at$. If $x_0 > 1/2$, this condition is satisfied for sufficiently small values of t.

IV. DISTANCE BETWEEN CIR AND SQUARED BESSEL PROCESSES IN INTEGRAL NORMS

In this section, we establish the rate of convergence of a sequence of CIR processes to a limiting CIR process in two integral norms, $L_1([0,T],\mathbf{P})$ and $L_2([0,T],\mathbf{P})$, over any fixed interval $[0,T]$, under the assumption that the corresponding coefficients converge. Additionally, we analyze the rate of convergence of the CIR processes to a squared Bessel process in both norms, demonstrating a close relationship between these two classes of processes in this context.

However, as expected, the upper bounds of the distance between them contain coefficients depending on the length of the interval and tending to ∞, as the length tends to ∞. In this sense, the processes disperse, or, in other words, move away. Despite this fact, the coefficients can be so close that, under slow growth of time interval, the processes can be still close.

So, consider the following sequence of stochastic differential equations:

$$X_n(t) = x_0 + \int_0^t (a_n - b_n X_n(s))\,ds + \sigma_n \int_0^t \sqrt{X_n(s)}\,dW_s, \quad (19)$$

$n \ge 0$, where $x_0 > 0$, $a_n > 0$, $b_n \ge 0$, and $\sigma_n > 0$ for all $n \ge 0$.

Assume that

$$a_n \to a_0, \quad b_n \to b_0, \quad \sigma_n \to \sigma_0, \quad \text{as } n \to \infty.$$

Note that the equations in (19) satisfy conditions (Y1$_n$)–(Y4$_n$) from Section 4 of[23], see Appendix B. According to Theorem B.1, we have that for any $T > 0$

$$\sup_{t\in[0,T]} \mathbf{E}\,|X_n(t) - X_0(t)| \to 0, \quad \text{as } n \to \infty.$$

In the following theorem, we establish an upper bound for rate of convergence.

Theorem IV.1. (*i*) *Let $b_0 > 0$. Then for any $T > 0$, the following upper bound holds:*

$$\sup_{t\in[0,T]} \mathbf{E}\,|X_n(t) - X_0(t)| \le e^{b_n T}\Big(|a_n - a_0|\,T + |b_n - b_0|\,A_0^2(T) + |\sigma_n - \sigma_0|\,A_0(T)\Big), \quad (20)$$

where

$$A_0^2(T) := \frac{1}{b_0}\left(x_0 - \frac{a_0}{b_0}\right)\left(1 - e^{-b_0 T}\right) + \frac{a_0}{b_0} T. \quad (21)$$

(*ii*) *Let $b_0 = 0$. Then for any $T > 0$, the following upper bound holds:*

$$\sup_{t\in[0,T]} \mathbf{E}\,|X_n(t) - X_0(t)| \le e^{b_n T}\Big(|a_n - a_0|\,T + b_n B_0^2(T) + |\sigma_n - \sigma_0|\,B_0(T)\Big), \quad (22)$$

where

$$B_0^2(T) = x_0 T + \frac{a_0}{2} T^2.$$

Remark IV.2. The function $A_0^2(T)$, defined by (21), is positive, since $A_0^2(T) = \int_0^T \mathbf{E}X_0(s)\,ds$, according to (16). Moreover, it exhibits linear growth and satisfies the following bounds:

$$\min\left\{x_0, \frac{a_0}{b_0}\right\} T \le A_0^2(T) \le \max\left\{x_0, \frac{a_0}{b_0}\right\} T.$$

To verify these inequalities, we consider two cases.

Case 1: $x_0 - \frac{a_0}{b_0} \ge 0$. Using the inequality $0 \le 1 - e^{-b_0 T} \le b_0 T$, we obtain that

$$\frac{a_0}{b_0} T \le A_0^2(T) \le \frac{1}{b_0}\left(x_0 - \frac{a_0}{b_0}\right) b_0 T + \frac{a_0}{b_0} T = x_0 T. \quad (23)$$

Case 2: $x_0 - \frac{a_0}{b_0} < 0$. In this case, both inequalities in (23) are reversed.

Proof of Theorem IV.1. (*i*) Define an auxiliary process

$$\widetilde{X}_n(t) = x_0 + \int_0^t (a_n - b_n X_0(s))\,ds + \sigma_n \int_0^t \sqrt{X_0(s)}\,dW_s.$$

Then using the Cauchy–Schwarz inequality and the Itô isometry, we get

$$\begin{aligned}\mathbf{E}\left|\widetilde{X}_n(t) - X_0(t)\right| &= |a_n - a_0|\,t + |b_n - b_0| \int_0^t \mathbf{E}X_0(s)\,ds \\ &\quad + |\sigma_n - \sigma_0|\,\mathbf{E}\left|\int_0^t \sqrt{X_0(s)}\,dW_s\right| \\ &\le |a_n - a_0|\,t + |b_n - b_0| \int_0^t \mathbf{E}X_0(s)\,ds \\ &\quad + |\sigma_n - \sigma_0| \left(\int_0^t \mathbf{E}X_0(s)\,ds\right)^{\frac{1}{2}}. \end{aligned} \quad (24)$$

Note that, by (16),

$$\int_0^t \mathbf{E}X_0(s)\,ds \le \int_0^T \mathbf{E}X_0(s)\,ds = A_0^2(T), \quad (25)$$

therefore, (24) becomes

$$\mathbf{E}\left|\widetilde{X}_n(t) - X_0(t)\right| \le |a_n - a_0|\,T + |b_n - b_0|\,A_0^2(T) + |\sigma_n - \sigma_0|\,A_0(T). \quad (26)$$

Now let us follow the Yamada method from[24]. Define

$$\alpha_m = \exp\left\{-\frac{m(m+1)}{2}\right\}, \quad m \ge 0.$$

It follows that $1 = \alpha_0 > \alpha_1 > \alpha_2 > \cdots > 0$ and

$$\int_{\alpha_m}^{\alpha_{m-1}} \frac{1}{x}\,dx = m, \quad m \ge 1.$$

Thus, for every $m \ge 1$, there exists a continuous function $\psi_m \colon \mathbb{R} \to \mathbb{R}$ with compact support in (α_m, α_{m-1}) such that

$$0 \le \psi_m(x) \le \frac{2}{mx},\ x \in (\alpha_m, \alpha_{m-1}), \quad \text{and} \quad \int_0^\infty \psi_m(x)dx = 1.$$

Next, we define

$$\varphi_m(x) = \int_0^{|x|}\int_0^y \psi_m(u)\,du\,dy, \quad x \in \mathbb{R}.$$

The function φ_m satisfies the following properties:

$$0 \le \varphi_m(x) \le |x| \quad \text{and} \quad \left|\varphi_m'(x)\right| \le 1,$$

and it is twice continuously differentiable since $\varphi_m''(x) = \psi_m(|x|)$.

Moreover, by the Lebesgue dominated convergence theorem, $\lim_{m\to\infty} \varphi_m(x) = |x|$, since $\lim_{m\to\infty} \int_0^y \psi_m(u)\,du = 1$.

By applying the Itô formula, we obtain

$$\mathbf{E}\varphi_m\left(X_n(t) - \widetilde{X}_n(t)\right)$$

$$
\begin{aligned}
&= -b_n \int_0^t \mathbf{E}\left[(X_n(s)-X_0(s))\,\varphi_m'(X_n(s)-X_0(s))\right] ds \\
&\quad + \frac{\sigma_n^2}{2}\int_0^t \mathbf{E}\left[\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)^2 \varphi_m''(X_n(s)-X_0(s))\right] ds \\
&\le b_n \int_0^t \mathbf{E}\left|X_n(s)-X_0(s)\right| ds \\
&\quad + \frac{\sigma_n^2}{2}\int_0^t \mathbf{E}\left[\left|X_n(s)-X_0(s)\right| \frac{2}{m\left|X_n(s)-X_0(s)\right|}\right] ds \\
&= b_n \int_0^t \mathbf{E}\left|X_n(s)-X_0(s)\right| ds + \frac{\sigma_n^2 t}{m},
\end{aligned}
$$

Here, we have used the facts that $|\varphi_m'(x)| \le 1$, $\varphi_m''(x) = \psi_m(|x|) \le \frac{2}{m|x|}$, and $|\sqrt{x}-\sqrt{y}| \le \sqrt{|x-y|}$. By taking the limit, as $m\to\infty$, and applying Fatou's lemma, we obtain

$$\mathbf{E}\left|X_n(t)-\widetilde{X}_n(t)\right| \le b_n \int_0^t \mathbf{E}\left|X_n(s)-X_0(s)\right| ds.$$

Then

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathbf{E}\left|X_n(t)-X_0(t)\right| \\
&\quad \le \sup_{t\in[0,T]} \mathbf{E}\left|\widetilde{X}_n(t)-X_0(t)\right| + \sup_{t\in[0,T]} \mathbf{E}\left|X_n(t)-\widetilde{X}_n(t)\right| \\
&\quad \le \sup_{t\in[0,T]} \mathbf{E}\left|\widetilde{X}_n(t)-X_0(t)\right| \\
&\qquad + b_n \int_0^T \sup_{u\in[0,s]} \mathbf{E}\left|X_n(u)-X_0(u)\right| ds.
\end{aligned}
$$

Applying the Grönwall inequality yields

$$\sup_{t\in[0,T]} \mathbf{E}\left|X_n(t)-X_0(t)\right| \le e^{b_n T} \sup_{t\in[0,T]} \mathbf{E}\left|\widetilde{X}_n(t)-X_0(t)\right|.$$

Finally, utilizing the bound from (26), we arrive at (20).

(ii) The proof of (22) follows similar steps as the proof of (20). In this case, instead of (25), we have

$$\int_0^T \mathbf{E}X_0(s)\,ds = x_0 T + \frac{a_0}{2}T^2 = B_0^2(T),$$

which is derived from (8). □

Remark IV.3. Note that, as $T\to\infty$,

$$A_0^2(T) \sim \frac{a_0}{b_0}T, \quad B_0^2(T) \sim \frac{a_0}{2}T^2.$$

Hence, as $T\to\infty$, the right-hand sides of (20) and (22) are asymptotically equivalent to

$$\frac{1}{b_0}\left(b_0\,|a_n-a_0| + a_0\,|b_n-b_0|\right) T e^{b_n T} \quad \text{and} \quad \frac{1}{2}a_0 b_n T^2 e^{b_n T}$$

respectively.

If, in the above theorem, we take $a_n \equiv a$, $\sigma_n \equiv \sigma$, $b_0 = 0$, the resulting bound simplifies significantly. Specifically, we obtain the following result concerning the approximation of the squared Bessel process by a sequence of CIR processes.

Corollary IV.4. *Let Y be a solution to* (2)*. Consider a sequence of the stochastic differential equations:*

$$Y_n(t) = y_0 + \int_0^t (a - b_n Y_n(s))\,ds + \sigma \int_0^t \sqrt{Y_n(s)}\,dW_s, \quad n \ge 1,$$

where $b_n \downarrow 0$, $n\to\infty$. Then, for any $T>0$, the following bound holds:

$$\sup_{t\in[0,T]} \mathbf{E}\left|Y_n(t)-Y(t)\right| \le e^{b_n T} b_n T\left(x_0 + \tfrac{1}{2}aT\right).$$

Corollary IV.5. *Let $b_n \downarrow 0$, $n\to\infty$, and let the sequence $\{T_n, n\ge 1\}$ satisfy*

$$T_n \to \infty, \quad e^{b_n T_n} b_n T_n^2 \to 0, \quad \text{as } n\to\infty. \tag{27}$$

Then

$$\sup_{t\in[0,T_n]} \mathbf{E}\left|Y_n(t)-Y(t)\right| \to 0, \quad n\to\infty.$$

Remark IV.6. For example, the condition (27) is fulfilled for $b_n = 1/n$, $T_n = \log(\log n)$, $n\ge 2$.

Remark IV.7. Note that the technique used in the proof of Theorem IV.1 does not allow us to establish an upper bound for the second moment $\mathbf{E}\left(X_n(t)-X_0(t)\right)^2$. The proof of Theorem IV.1 consists of two main steps: deriving inequality (24) and applying the Yamada method. While an inequality of the form (24) can be established for the second moment $\mathbf{E}(\widetilde{X}_n(t)-X_0(t))^2$, the Yamada method is more involved and provides only an upper bound for the first moment $\mathbf{E}|\widetilde{X}_n(t)-X_0(t)|$. Nevertheless, as stated in the next result, it is possible to derive an upper bound for $\mathbf{E}\left(X_n(t)-X_0(t)\right)^2$ as well.

Theorem IV.8. (i) *Let $b_n>0$ for all $n\ge 0$. Then for any $T>0$, the following upper bound holds:*

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathbf{E}\left(X_n(t)-X_0(t)\right)^2 \le 2\left(R_n(T)+R_0(T)\right)^{\frac{1}{2}} e^{\frac{b_n T}{2}} \\
&\times \Big(|a_n-a_0|\,T + |b_n-b_0|\,A_0^2(T) + |\sigma_n-\sigma_0|\,A_0(T)\Big)^{\frac{1}{2}},
\end{aligned} \tag{28}
$$

where $A_0(T)$ is defined in Theorem IV.1 and

$$
\begin{aligned}
R_n(T) &= x_0^3 + \left(1 + \frac{3\sigma_n^2}{2a_n} + \frac{\sigma_n^4}{2a_n^2}\right) \\
&\quad \times \left(\frac{a_n^3}{b_n^3}\left(1-e^{-b_n T}\right)^3 + \frac{3x_0 a_n^2}{b_n^2}\left(1-e^{-b_n T}\right)^2\right) \\
&\quad + \frac{3x_0^2 a_n}{b_n}\left(1+\frac{\sigma_n^2}{a_n}\right)\left(1-e^{-b_n T}\right).
\end{aligned}
$$

(ii) *Let $b_0 = 0$ and $b_n>0$ for $n\ge 1$. Then for any $T>0$, the following upper bound holds:*

$$\sup_{t\in[0,T]} \mathbf{E}\left(X_n(t)-X_0(t)\right)^2 \le 2\left(R_n(T)+\widetilde{R}_0(T)\right)^{\frac{1}{2}} e^{\frac{b_n T}{2}}$$

$$\times\left(|a_n-a_0|\,T+b_nB_0^2(T)+|\sigma_n-\sigma_0|\,B_0(T)\right)^{\frac{1}{2}},$$

where $B_0(T)$ is defined in Theorem IV.1 and

$$\widetilde{R}_0(T)=\left(\frac{a_0\sigma_0^4}{2}+\frac{3a_0^2\sigma_0^2}{2}+a_0^3\right)T^3$$
$$+3\left(\frac{x_0\sigma_0^4}{2}+\frac{3a_0x_0\sigma_0^2}{2}+a_0^2x_0\right)T^2+3x_0^2\left(\sigma_0^2+a_0\right)T+x_0^3.$$

Proof. (i) By the Cauchy–Schwarz inequality,

$$\begin{aligned}\mathbf{E}\left(X_n(t)-X_0(t)\right)^2&\le\mathbf{E}\,|X_n(t)-X_0(t)|^{\frac{1}{2}}\,|X_n(t)-X_0(t)|^{\frac{3}{2}}\\&\le\left(\mathbf{E}\,|X_n(t)-X_0(t)|\right)^{\frac{1}{2}}\left(\mathbf{E}\,|X_n(t)-X_0(t)|^3\right)^{\frac{1}{2}}.\\&\le 2\left(\mathbf{E}\,|X_n(t)-X_0(t)|\right)^{\frac{1}{2}}\left(\mathbf{E}X_n^3(t)+\mathbf{E}X_0^3(t)\right)^{\frac{1}{2}},\end{aligned}$$

where the elementary inequality $(x+y)^3\le 4(x^3+y^3)$ has been used.

Further, for any $n\ge 0$ and any $t\in[0,T]$,

$$\begin{aligned}\mathbf{E}X_n^3(t)&=x_0^3e^{-3b_nt}+\left(1+\frac{3\sigma_n^2}{2a_n}+\frac{\sigma_n^4}{2a_n^2}\right)\\&\times\left(\frac{a_n^3}{b_n^3}\left(1-e^{-b_nt}\right)^3+\frac{3x_0a_n^2}{b_n^2}e^{-b_nt}\left(1-e^{-b_nt}\right)^2\right)\\&+\frac{3x_0^2a_n}{b_n}\left(1+\frac{\sigma_n^2}{a_n}\right)e^{-2b_nt}\left(1-e^{-b_nt}\right)\le R_n(T).\end{aligned}$$

Then using the bound (20), we obtain

$$\begin{aligned}&\sup_{t\in[0,T]}\mathbf{E}\left(X_n(t)-X_0(t)\right)^2\\&\le 2\left(R_n(T)+R_0(T)\right)^{\frac{1}{2}}\left(\sup_{t\in[0,T]}\mathbf{E}\,|X_n(t)-X_0(t)|\right)^{1/2}\\&\le 2\left(R_n(T)+R_0(T)\right)^{\frac{1}{2}}e^{b_nT/2}\\&\quad\times\left(|a_n-a_0|\,T+|b_n-b_0|A_0^2(T)+|\sigma_n-\sigma_0|A_0(T)\right)^{\frac{1}{2}}.\end{aligned}$$

The statement (ii) is derived from the second statement of Theorem IV.1 in a similar manner, taking into account that $\sup_{t\in[0,T]}\mathbf{E}X_0^3\le\widetilde{R}(T)$ for $b_0=0$, according to the formula (9) for the third moment of the squared Bessel process. □

Now we can establish the upper bound for the same value, $\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^2$, as in the previous theorem, but using not knowledge of probability distributions, but methods of stochastic analysis.

Theorem IV.9. (i) *Let $b_n>0$ for all $n\ge 0$. Then for any $T>0$, the following upper bound holds:*

$$\begin{aligned}&\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^2\\&\le e^{(b_n+b_0)T}\left(2\,|a_n-a_0|+\sigma_0^2+2\sigma_0\,|\sigma_n-\sigma_0|\right)\\&\quad\times\left(|a_n-a_0|\,T+|b_n-b_0|A_0^2(T)+|\sigma_n-\sigma_0|A_0(T)\right)T\\&+e^{b_0T}2e^{b_0T}\,|b_n-b_0|\,T\left(D_n^2(T)+D_n(T)D_0(T)\right)\\&+e^{b_0T}\,|\sigma_n-\sigma_0|^2A_n^2(T),\end{aligned}\tag{29}$$

where

$$\begin{aligned}A_n^2(T)&=\int_0^T\mathbf{E}X_n(s)\,ds\\&=\frac{1}{b_n}\left(x_n-\frac{a_n}{b_n}\right)\left(1-e^{-b_nT}\right)+\frac{a_n}{b_n}T,\\D_n^2(T)&=\frac{x_0(\sigma_n^2+2a_n)}{b_n}\left(1-e^{-bT}\right)\\&\quad+\frac{a(\sigma_n^2+2a_n)}{2b_n^2}\left(1-e^{-bT}\right)^2+x_0^2.\end{aligned}$$

(ii) *Let $b_0=0$ and $b_n>0$ for $n\ge 1$. Then for any $T>0$, the following upper bound holds:*

$$\begin{aligned}&\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^2\\&\le\left(2\,|a_n-a_0|+\sigma_0^2+2\sigma_0\,|\sigma_n-\sigma_0|\right)\\&\quad\times\left(|a_n-a_0|\,T+b_nB_0^2(T)+|\sigma_n-\sigma_0|\,B_0(T)\right)Te^{b_nT}\\&+2b_nT\left(D_n^2(T)+D_n(T)E_0(T)\right)+|\sigma_n-\sigma_0|^2A_n^2(T).\end{aligned}$$

where $B_0(T)$ is defined in Theorem IV.1 and

$$E_0^2(T)=x_0^2+\left(\frac{\sigma_0^2}{2}+a_0\right)\left(2x_0T+a_0T^2\right).$$

Proof. (i) Let us apply the Itô formula to the function $F(x)=x^2$ and the process

$$\begin{aligned}X_n(t)-X_0(t)&=(a_n-a_0)t+(b_n-b_0)\int_0^tX_n(s)\,ds\\&+b_0\int_0^t\left(X_n(s)-X_0(s)\right)ds+(\sigma_n-\sigma_0)\int_0^t\sqrt{X_n(s)}\,dW_s\\&+\sigma_0\int_0^t\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)dW_s.\end{aligned}$$

We get that

$$\begin{aligned}&(X_n(t)-X_0(t))^2=2\int_0^t(X_n(s)-X_0(s))\,d(X_n(s)-X_0(s))\\&+\sigma_0^2\int_0^t\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)^2ds+(\sigma_n-\sigma_0)^2\int_0^tX_n(s)\,ds\\&+2\sigma_0(\sigma_n-\sigma_0)\int_0^t\sqrt{X_n(s)}\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)ds.\end{aligned}\tag{30}$$

Taking expectation, which is zero for stochastic integrals, we obtain from (30) that

$$\begin{aligned}&\mathbf{E}(X_n(t)-X_0(t))^2=2(a_n-a_0)\int_0^t\mathbf{E}(X_n(s)-X_0(s))\,ds\\&+2(b_n-b_0)\int_0^t\mathbf{E}\left[(X_n(s)-X_0(s))X_n(s)\right]ds\\&+b_0\int_0^t\mathbf{E}(X_n(s)-X_0(s))^2\,ds\end{aligned}$$

$$
\begin{aligned}
&+\sigma_0^2\int_0^t \mathbf{E}\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)^2 ds\\
&+(\sigma_n-\sigma_0)^2\int_0^t \mathbf{E}X_n(s)\,ds\\
&+2\sigma_0(\sigma_n-\sigma_0)\int_0^t \mathbf{E}\left[\sqrt{X_n(s)}\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)\right]ds.
\end{aligned}
\tag{31}
$$

The expectations in the right-hand side of (31) can be bounded as follows:

$$
\begin{aligned}
\mathbf{E}\left|(X_n(s)-X_0(s))X_n(s)\right| &\le \mathbf{E}X_n^2(s)+\mathbf{E}X_0(s)X_n(s)\\
&\le \mathbf{E}X_n^2(s)+\sqrt{\mathbf{E}X_0^2(s)\mathbf{E}X_n^2(s)},\\
\mathbf{E}\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)^2 &\le \mathbf{E}\left|X_n(s)-X_0(s)\right|,
\end{aligned}
$$

and

$$
\mathbf{E}\left[\sqrt{X_n(s)}\left(\sqrt{X_n(s)}-\sqrt{X_0(s)}\right)\right] \le \mathbf{E}\left|X_n(s)-X_0(s)\right|,
$$

where we have used inequalities

$$
\begin{aligned}
\left|\sqrt{x}-\sqrt{y}\right| &\le \sqrt{|x-y|},\\
\sqrt{x}\left|\sqrt{x}-\sqrt{y}\right| &= \frac{\sqrt{x}}{\sqrt{x}+\sqrt{y}}\,|x-y| \le |x-y|.
\end{aligned}
$$

Hence, (31) implies

$$
\begin{aligned}
&\mathbf{E}(X_n(t)-X_0(t))^2\\
&\le \left(2|a_n-a_0|+\sigma_0^2+2\sigma_0|\sigma_n-\sigma_0|\right)\int_0^t \mathbf{E}\left|X_n(s)-X_0(s)\right| ds\\
&+2|b_n-b_0|\int_0^t\left(\mathbf{E}X_n^2(s)+\sqrt{\mathbf{E}X_0^2(s)\mathbf{E}X_n^2(s)}\right)ds\\
&+b_0\int_0^t \mathbf{E}(X_n(s)-X_0(s))^2\,ds+|\sigma_n-\sigma_0|^2\int_0^t \mathbf{E}X_n(s)\,ds.
\end{aligned}
\tag{32}
$$

Now, the expectation $\mathbf{E}|X_n(s)-X_0(s)|$ is the first term in the right-hand side of (32) can be bounded with the help of Theorem IV.1, the expectations $\mathbf{E}X_n^2(s)$, $\mathbf{E}X_0^2(s)$ and $\mathbf{E}X_n(s)$ can be computed by explicit formulas (5) and (6). Thus taking supremum over all $t\in[0,T]$ and denoting

$$
F_n(T)=\int_0^T\left(\mathbf{E}X_n^2(s)+\sqrt{\mathbf{E}X_0^2(s)\mathbf{E}X_n^2(s)}\right)ds,
$$

we arrive at

$$
\begin{aligned}
&\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^2\\
&\le \left(2|a_n-a_0|+\sigma_0^2+2\sigma_0|\sigma_n-\sigma_0|\right)\\
&\times\left(|a_n-a_0|T+|b_n-b_0|A_0^2(T)+|\sigma_n-\sigma_0|A_0(T)\right)Te^{b_nT}\\
&+2|b_n-b_0|F_n(T)+|\sigma_n-\sigma_0|^2A_n^2(T)\\
&+b_0\int_0^T\sup_{u\in[0,s]}\mathbf{E}(X_n(u)-X_0(u))^2\,ds.
\end{aligned}
\tag{33}
$$

Finally, the Grönwall inequality yields

$$
\begin{aligned}
&\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^2\\
&\le e^{(b_n+b_0)T}\left(2|a_n-a_0|+\sigma_0^2+2\sigma_0|\sigma_n-\sigma_0|\right)\\
&\quad\times\left(|a_n-a_0|T+|b_n-b_0|A_0^2(T)+|\sigma_n-\sigma_0|A_0(T)\right)T\\
&\quad+e^{b_0T}\left(2|b_n-b_0|F_n(T)+|\sigma_n-\sigma_0|^2A_n^2(T)\right).
\end{aligned}
$$

It remains to note that (6) implies

$$
\begin{aligned}
\sup_{t\in[0,T]}\mathbf{E}X_n(s)^2 &\le \frac{x_0(\sigma_n^2+2a_n)}{b_n}\left(1-e^{-bT}\right)\\
&\quad+\frac{a(\sigma_n^2+2a_n)}{2b_n^2}\left(1-e^{-bT}\right)^2+x_0^2=D_n^2(T),
\end{aligned}
$$

consequently,

$$
F_n(T)\le T\left(D_n^2(T)+D_n(T)D_0(T)\right).
$$

The proof of claim (ii) follows in a manner similar to that of claim (i). However, in this case, the last term in (33) vanishes, which simplifies the final steps of the proof, as the application of the Grönwall inequality is no longer required. □

Remark IV.10. Let us now discuss and compare the upper bounds established in Theorems IV.1, IV.8, and IV.9.

1. A key advantage of all three theorems is that they provide explicit rates of convergence in terms of the coefficients of the corresponding equations. This makes them particularly valuable for practical analysis.
2. Theorems IV.8 and IV.9 present upper bounds for the second moments, which are often crucial for practical applications. These bounds cannot be directly obtained from the results for the first moments, such as those provided by Theorem IV.1. Furthermore, it is worth noting that, in a similar manner, one can derive upper bounds for $\sup_{t\in[0,T]}\mathbf{E}(X_n(t)-X_0(t))^{2p}$ for any $p\ge 1$.
3. For a fixed $T>0$, the convergence rates established in Theorems IV.1, IV.8, and IV.9 can be compared as follows. Assume that

$$
|a_n-a_0|\le\delta_n,\quad |b_n-b_0|\le\delta_n,\quad |\sigma_n-\sigma_0|\le\delta_n
$$

for some sequence $\delta_n\downarrow 0$ as $n\to\infty$. Then for the quantity $\sup_{t\in[0,T]}\mathbf{E}|X_n(t)-X_0(t)|$ Theorems IV.1, IV.8, and IV.9 yield rates of convergence of orders $O(\delta_n)$, $O(\delta_n^{1/4})$, and $O(\delta_n^{1/2})$, respectively. Hence, from the perspective of convergence rates, Theorem IV.1 offers the fastest rate. Similarly, Theorem IV.9 demonstrates a superior rate of convergence compared to Theorem IV.8.
4. We can also compare Theorems IV.8 and IV.9 in the asymptotic case as $T\to\infty$. Note that the functions $R_n(T)$ are bounded, while $A_0^2(T)$ grows linearly with T. Consequently, the right-hand side of (28) behaves as $O(T^{1/2}e^{b_nT/2})$, as $T\to\infty$. In contrast, the right-hand side of (29) grows significantly faster, at a rate of $O(T^2e^{(b_n+b_0)T})$. From this comparison, for large T, Theorem IV.8 provides a tighter bound than Theorem IV.9.

V. PARAMETER ESTIMATION

We now address the problem of identifying the squared Bessel process, i.e., the estimation of its parameters. Suppose we have continuous-time observations of a trajectory $\{Y_t, t \in [0,T]\}$ of the squared Bessel process (2) over some interval $[0,T]$. Note that these parameters are also defining for the CIR process, and the corresponding estimates can be easily modified for it. Moreover, we can assume that the parameter σ is known and focus on estimating the parameter a. For continuous-time observations, this assumption is natural, because σ can be determined almost surely from the observations on any fixed interval, as explained in the following remark.

Remark V.1. (Estimation of σ) Let $T > 0$ be fixed, $\delta = \frac{T}{n}$, and $t_k = k\delta$, for $0 \le k \le n$. Then

$$\sum_{k=1}^{n} \left(Y_{t_k} - Y_{t_{k-1}}\right)^2 \to \sigma^2 \int_0^T Y_s\, ds \quad \text{a.s., as } n \to \infty.$$

Indeed, as it is clear, quadratic variation of the linear function tends to zero with the diameter of the partition of the interval. This implies that the parameter σ can be evaluated using the following identity:

$$\sigma^2 = \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \left(Y_{t_k} - Y_{t_{k-1}}\right)^2}{\int_0^T Y_s\, ds} \quad \text{a.s.}$$

To estimate the parameter a, we apply the maximum likelihood method, see, e.g.,[25] (Section 10.7). First, we transform equation (2) using the Itô formula as follows:

$$\begin{aligned}\sqrt{Y_t} &= \sqrt{y_0} + \frac{1}{2}\int_0^t \frac{1}{\sqrt{Y_s}}\, dY_s - \frac{1}{8}\sigma^2 \int_0^t \frac{1}{\sqrt{Y_s}}\, ds \\ &= \sqrt{y_0} + \frac{a}{2}\int_0^t \frac{ds}{\sqrt{Y_s}} + \frac{\sigma}{2} W_t - \frac{1}{8}\sigma^2 \int_0^t \frac{1}{\sqrt{Y_s}}\, ds \\ &= \sqrt{y_0} + \frac{1}{2}\int_0^t \frac{a - \frac{\sigma^2}{4}}{\sqrt{Y_s}}\, ds + \frac{\sigma}{2} W_t.\end{aligned}$$

It follows from the Girsanov theorem that the likelihood function is then given by

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{[0,T]} = \exp\left\{ \frac{\frac{\sigma^2}{4} - a}{\sigma} \int_0^T \frac{1}{\sqrt{Y_s}}\, dW_s - \frac{1}{2}\left(\frac{\frac{\sigma^2}{4} - a}{\sigma}\right)^2 \int_0^T \frac{ds}{Y_s} \right\}.$$

We aim to find the value of a that maximizes this likelihood function.

To proceed, we set

$$-\theta = \frac{\sigma}{4} - \frac{a}{\sigma}$$

and minimize the likelihood function with respect to θ, replacing

$$-\theta \int_0^T \frac{1}{\sqrt{Y_s}}\, dW_s = -\frac{2\theta}{\sigma}\int_0^T \frac{1}{\sqrt{Y_s}}\, d\sqrt{Y_s} + \theta^2 \int_0^T \frac{ds}{Y_s}.$$

Letting $Z_t = \sqrt{Y_t}$, the expression simplifies to minimizing the following:

$$-\frac{2\theta}{\sigma}\int_0^T \frac{dZ_s}{Z_s} + \frac{1}{2}\theta^2 \int_0^T \frac{ds}{Z_s^2}.$$

This minimization leads to the following maximum likelihood estimator:

$$\hat{\theta}_T = \frac{2\int_0^T \frac{dZ_s}{Z_s}}{\sigma \int_0^T \frac{ds}{Z_s^2}}.$$

Proposition V.2. *Let $2a > \sigma^2$. Then $\hat{\theta}_T$ is a strongly consistent estimator of θ, i.e.,*

$$\hat{\theta}_T \to \theta \quad \textit{a.s., when } T \to \infty.$$

Proof. Using the equation $dZ_t = \frac{\sigma\theta dt}{2Z_t} + \frac{\sigma}{2} dW_t$, we represent the estimator $\hat{\theta}_T$ in the following form

$$\hat{\theta}_T = \theta + \frac{\int_0^T \frac{dW_s}{Z_s}}{\int_0^T \frac{ds}{Z_s^2}} = \theta + \frac{M_T}{\langle M \rangle_T},$$

where $M_T = \int_0^T \frac{dW_s}{Z_s}$ is a locally square-integrable martingale with the quadratic variation $\langle M \rangle_T = \int_0^T \frac{ds}{Z_s^2}$. According to the strong law of large numbers for martingales[26] (Ch. 2, §6, Thm. 10, Cor. 1), if $\langle M \rangle_T \to \infty$ a.s., as $T \to \infty$, then $\frac{M_T}{\langle M \rangle_T} \to 0$ a.s., as $T \to \infty$. Thus, to establish strong consistency, we need to show that

$$\int_0^T \frac{ds}{Z_s^2} \to \infty \quad \text{a.s., as } T \to \infty. \tag{34}$$

Return to the equation

$$Z_t^2 = Z_0^2 + at + \sigma \int_0^t Z_s\, dW_s.$$

We divide both sides by $\int_0^t Z_s^2\, ds$:

$$\frac{Z_t^2}{\int_0^t Z_s^2\, ds} = \frac{Z_0^2}{\int_0^t Z_s^2\, ds} + \frac{at}{\int_0^t Z_s^2\, ds} + \sigma \frac{\int_0^t Z_s\, dW_s}{\int_0^t Z_s^2\, ds}. \tag{35}$$

Let $b_n \downarrow 0$, as $n \to \infty$, and consider the approximating CIR processes $\{Z_{t,b_n}^2, t \ge 0\}$, $n \ge 1$, which solve the equations

$$Z_{t,b_n}^2 = y_0 + \int_0^t \left(a - b_n Z_{s,b_n}^2\right) ds + \sigma \int_0^t Z_{s,b_n}\, dW_s, \quad n \ge 1.$$

It follows from (4) that for any $n > 1$ and any $\varepsilon > 0$ there exists $T_n > 0$ such that for all $T \ge T_n$,

$$\left| \frac{1}{T}\int_0^T \frac{ds}{Z_{s,b_n}^2} - \frac{b_n}{a - \frac{\sigma^2}{2}} \right| < \frac{\varepsilon}{2} \quad \text{a.s.}$$

Choose $n_0 > 0$ such that for $n \ge n_0$

$$\frac{b_n}{a - \frac{\sigma^2}{2}} < \frac{\varepsilon}{2}.$$

Then for all $T \geq T_{n_0}$

$$\frac{1}{T}\int_0^T \frac{ds}{Z^2_{s,b_{n_0}}} < \varepsilon \quad \text{a.s.} \tag{36}$$

By the comparison theorem[15] (Proposition 2.18, p. 293), we have

$$\mathbf{P}\left(Z^2_{t,b_{n_0}} \leq Z^2_t \text{ for all } t \geq 0\right) = 1.$$

Hence, (36) implies that for all $T \geq T_{n_0}$

$$\frac{1}{T}\int_0^T \frac{ds}{Z^2_s} < \varepsilon \quad \text{a.s.}$$

Consequently,

$$\frac{1}{T}\int_0^T \frac{ds}{Z^2_s} \to 0, \quad \text{a.s., as } T \to \infty. \tag{37}$$

This implies that

$$\frac{1}{T}\int_0^T Z^2_s\,ds \to \infty, \quad \text{a.s., as } t \to \infty,$$

since

$$\frac{1}{T^2}\int_0^T Z^2_s\,ds \int_0^T \frac{ds}{Z^2_s} \geq 1$$

by the Cauchy–Schwarz inequality.

Thus, we obtain

$$\frac{t}{\int_0^t Z^2_s\,ds} \to 0 \quad \text{and} \quad \frac{\int_0^t Z_s\,dW_s}{\int_0^t Z^2_s\,ds} \to 0 \quad \text{a.s., as } t \to \infty.$$

From (35), it follows that

$$\frac{Z^2_t}{\int_0^t Z^2_s\,ds} \to 0 \quad \text{a.s., as } t \to \infty,$$

or

$$\frac{\int_0^t Z^2_s\,ds}{Z^2_t} \to \infty \quad \text{a.s., as } t \to \infty.$$

Therefore, for almost all ω and for all $C > 0$ there exists t_0 such that for all $t \geq t_0$

$$\frac{1}{Z^2_t} \geq \frac{C}{\int_0^t Z^2_s\,ds},$$

which implies that

$$\int_{t_0}^{\infty} \frac{ds}{Z^2_s} \geq C\int_{t_0}^{\infty} \frac{ds}{\int_0^s Z^2_u\,du},$$

Letting $C \to \infty$, we obtain

$$\int_{t_0}^{\infty} \frac{ds}{Z^2_s} = \infty \quad \text{a.s.}$$

Thus, condition (34) is satisfied, and we conclude that strong consistency holds. □

Corollary V.3. *Let $2a > \sigma^2$. The maximum likelihood estimator of the parameter a of the squared Bessel process* (2) *is given by*

$$\hat{a}_T = \sigma\hat{\theta}_T + \frac{\sigma^2}{4} = \frac{2\int_0^T \frac{d\sqrt{Y_s}}{\sqrt{Y_s}}}{\int_0^T \frac{ds}{Y_s}} + \frac{\sigma^2}{4}$$

and it is strongly consistent, as $T \to \infty$.

VI. INSTABILITY AND SOME FUNCTIONAL LIMIT THEOREMS FOR THE SQUARED BESSEL PROCESS

As we have already mentioned, the process X is ergodic, while Y is non-ergodic. Now our goal is to establish how these properties reflect in the notion of stochastic instability. There are several approaches to stochastic instability of the processes. We shall consider the following definition, see book[27].

Definition VI.1. *A stochastic process ξ is called stochastically unstable if for any constant $N > 0$*

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t \mathbf{P}\{|\xi_s| < N\}\,ds = 0.$$

Proposition VI.2. *Let $N > 0$ be an arbitrary constant.*

(i) *The CIR process X has the following property:*

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t \mathbf{P}\{X_s < N\}\,ds = \frac{\gamma\left(\frac{2a}{\sigma^2}, \frac{2bN}{\sigma^2}\right)}{\Gamma\left(\frac{2a}{\sigma^2}\right)},$$

where $\gamma(a,x) = \int_0^x u^{a-1}e^{-u}\,du$ is the lower incomplete Gamma function.

(ii) *The squared Bessel process Y is stochastically unstable, i.e.,*

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t \mathbf{P}\{Y_s < N\}\,ds = 0.$$

Proof. (i) Using the convergence (3), we obtain

$$\begin{aligned}\mathbf{P}\{X_t < N\} &= \int_0^N p_t(x)\,dx \to \int_0^N p_\infty(x)\,dx \\ &= \frac{(2b/\sigma^2)^{2a/\sigma^2}}{\Gamma(2a/\sigma^2)}\int_0^N x^{2a/\sigma^2-1}e^{-2bx/\sigma^2}\,dx \\ &= \frac{1}{\Gamma(2a/\sigma^2)}\int_0^{2bN/\sigma^2} u^{2a/\sigma^2-1}e^{-u}\,du = \frac{\gamma\left(\frac{2a}{\sigma^2}, \frac{2bN}{\sigma^2}\right)}{\Gamma\left(\frac{2a}{\sigma^2}\right)},\end{aligned}$$

$t \to +\infty$. Taking into account continuity of the integrand and applying the l'Hôpital's rule, we get

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t \mathbf{P}\{X_s < N\}\,ds = \lim_{t\to+\infty} \mathbf{P}\{X_t < N\} = \frac{\gamma\left(\frac{2a}{\sigma^2}, \frac{2bN}{\sigma^2}\right)}{\Gamma\left(\frac{2a}{\sigma^2}\right)}.$$

(ii) The proof is similar to that of the statement (i). However, we get zero limit, since the probability density $g_t(x)$ of the process Y converges to zero, as $t \to \infty$, according to (7). □

Remark VI.3. Proposition VI.2 justifies why we say that we have phase transition when coefficient b of CIR process tends to zero and we get squared Bessel process. Indeed, squared Bessel process is stochastically unstable while, oppositely, CIR process is ergodic, and in this sense, stochastically stable. Therefore, we use here the term "phase transition" in order to describe principally different long-term behaviour of the respective dynamical system. In more detail the difference of the asymptotic behaviour is described for example in the paper[28] where the authors analyze the transition from an ergodic model to a non-ergodic one, for which standard analytical and statistical methods are insufficient for fitting the data, necessitating the development and application of alternative tools.

Now, note that in some sense, we were lucky because knowledge of the distribution allowed us to establish instability of the squared Bessel process directly, without application of the tools of stochastic analysis. However, we can establish instability for the approximations of the squared Bessel process whose distribution is unknown. So, again, consider the squared Bessel process determined as the unique solution of the equation (2). Let for simplicity $\sigma = 2$. General condition $a \geq \frac{\sigma^2}{2}$ leads in our case to $a \geq 2$. So, we assume that $a \geq 2$, then the trajectories of the solution are strictly positive with probability 1. Therefore, we can consider function $F(x) = \sqrt{x}$ and apply Itô formula to Y_t, obtaining equation

$$\begin{aligned}\sqrt{Y_t} &= \sqrt{y_0} + \frac{1}{2}\int_0^t \frac{a - \frac{\sigma^2}{4}}{\sqrt{Y_s}}\,ds + \frac{\sigma}{2}W_t\\ &= \sqrt{y_0} + \frac{1}{2}\int_0^t \frac{a-1}{\sqrt{Y_s}}\,ds + W_t,\end{aligned}$$

or, that is the same,

$$V_t = V_0 + \frac{1}{2}\int_0^t \frac{a-1}{V_s}\,ds + W_t,$$

where $V_t = \sqrt{Y_t}$, $V_0 = \sqrt{Y_0}$. Note that $a - 1 \geq 1$. Now our goal is to consider a smooth version of Bessel process, i.e., a solution of the stochastic differential equation

$$V_t^{\varepsilon} = \int_0^t \frac{c\,ds}{\sqrt{(V_s^{\varepsilon})^2 + \varepsilon^2}} + W_t, \tag{38}$$

where $\varepsilon \neq 0$, $c > 0$, $V_0^{\varepsilon} = \sqrt{y}_0 > 0$. The coefficient $\frac{c}{\sqrt{x^2+\varepsilon^2}}$ is Lipschitz because it has a bounded derivative:

$$\begin{aligned}\left|\left(\frac{c}{\sqrt{x^2+\varepsilon^2}}\right)'\right| &= \frac{c|x|}{(x^2+\varepsilon^2)^{3/2}}\\ &\leq c\frac{|x|}{(x^2+\varepsilon^2)^{1/2}} \cdot \frac{1}{x^2+\varepsilon^2} \leq \frac{c}{\varepsilon^2}.\end{aligned}$$

Also it is bounded, therefore, due the standard existence-uniqueness theorem for stochastic differential equations, equation (38) has a unique strong solution.

We want to achieve three goals. First, we establish the convergence of V^{ε} to the (non-squared) Bessel process $V = \sqrt{Y}$, as $\varepsilon \to 0$, that is more or less the expected result.

Proposition VI.4. *Let $c = \frac{1}{2}(a-1)$, $\varepsilon \to 0$. Then for any $t > 0$*

$$V_t^{\varepsilon} \to V_t \quad a.s.$$

Proof. Consider the difference

$$V_t^{\varepsilon} - V_t = c\left(\int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} - \int_0^t \frac{ds}{V_s}\right). \tag{39}$$

According to the comparison theorem, $V_t^{\varepsilon} \leq V_t$, $t \geq 0$, $\varepsilon > 0$, a.s. Therefore, (39) implies that

$$\int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} \leq \int_0^t \frac{ds}{V_s}, \quad t \geq 0,\ \varepsilon > 0, \text{ a.s.}$$

Moreover, V_t^{ε} increases in ε when ε decreases, therefore $\int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}}$ also increases and has a limit.

Since

$$\int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} \geq \int_0^t \frac{ds}{\sqrt{V_s^2+\varepsilon^2}} \uparrow \int_0^t \frac{ds}{V_s}, \quad \text{as } \varepsilon \downarrow 0,$$

we get from the double inequality

$$\int_0^t \frac{ds}{\sqrt{V_s^2+\varepsilon^2}} \leq \int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} \leq \int_0^t \frac{ds}{V_s},$$

that

$$\int_0^t \frac{ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} \to \int_0^t \frac{ds}{V_s}, \quad \text{a.s., as } \varepsilon \downarrow 0,$$

whence

$$V_t^{\varepsilon} = V_0 + \int_0^t \frac{c\,ds}{\sqrt{(V_s^{\varepsilon})^2+\varepsilon^2}} + W_t \to V_t, \quad \text{a.s., as } \varepsilon \downarrow 0,$$

for $t \geq 0$. □

Second, we wish to establish stochastic instability of V^{ε}.

Proposition VI.5. *Let $c > 0$. Then the processes V^{ε} are stochastically unstable for any $\varepsilon \neq 0$, i.e.,*

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t \mathbf{P}\{V_s^{\varepsilon} < N\}\,ds = 0$$

for any constant $N > 0$.

Remark VI.6. Stochastic instability of V^{ε} does not follow from Proposition VI.2 because $V^{\varepsilon} \leq V$. Oppositely, stochastic instability of V follows from Proposition VI.5. However, having explicit formulas for the distributions of X and $Y = (V)^2$, we preferred to give the direct proof to Proposition VI.2.

Proof. We shall apply Theorem 3.1, item 2. from[27], see Theorem C.1 in Appendix. So, in terms of Section 3.1 from[27], the process V^ε is the unique solution of the equation

$$dV_t^\varepsilon = a(V_t^\varepsilon)\,dt + dW_t, \quad t>0,\, V_0>0,$$

with the drift $a(x) = \frac{c}{\sqrt{x^2+\varepsilon^2}}$. Applying the l'Hôpital's rule, let us check the value of the limit

$$\lim_{|x|\to+\infty} \frac{1}{\log|x|}\int_0^x a(v)dv = \lim_{x\to+\infty}|x|a(x) = c>0,$$

and the proof immediately follows from Theorem C.1 ($c_0 = c>0$, $2c_0>-1$). □

And finally, we establish a bit unexpected statement: if $\varepsilon \neq 0$ is fixed, then the properly normalised processess V^ε weakly converge to the (non-squared) Bessel process as time tends to infinity. The definition of weak convergence is given in Appendix, Definition C.2.

Theorem VI.7. *Normalised stochastic process $Z_\varepsilon(t) = \frac{Y_{tT}^\varepsilon}{\sqrt{T}}$ converges weakly, as $T\to\infty$, to the Bessel process Y_t that is the solution of the equation*

$$Y_t^2 = 3t + 2\int_0^t Y_s\,dW_s, \quad t\ge 0. \tag{40}$$

Proof. We shall apply Theorem C.4 from Appendix. In our case

$$\frac{1}{x}\int_0^x va_\varepsilon(v)\,dv = \frac{1}{x}\int_0^x \frac{v}{\sqrt{v^2+\varepsilon^2}}\,dv \to \begin{cases}1, & x\to+\infty,\\ -1, & x\to-\infty,\end{cases}$$

for any $\varepsilon>0$. Therefore we have that $c_1=1$, $c_2=-1$ in Theorem C.4 and

$$2c_1 = 2>1 \quad \text{and} \quad 2c_2 = -2<1.$$

Then the stochastic process $\frac{Y_{tT}^\varepsilon}{\sqrt{T}}$ weakly converges, as $T\to\infty$, to the Bessel process satisfying equation (40). □

ACKNOWLEDGMENTS

The first author is supported by The Swedish Foundation for Strategic Research, grant Nr. UKR24-0004, by Japan Science and Technology Agency CREST, project reference number 811JPMJCR2115 and the ToppForsk project no. 274410 of the Research Council of Norway with the title STORM: Stochastics for Time-Space Risk Models. The second author is supported by the Research Council of Finland, decision number 359815.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Appendix A: Special functions

In this subsection we provide definitions and selected properties of special functions, used in the paper. We refer to the handbooks and[29] (Chapter 13) or[18] (Chapter 47) for more details.

The confluent hypergeometric function ${}_1F_1$ (Kummer's function) is defined by the series:

$${}_1F_1(a,c,x) = \sum_{j=0}^\infty \frac{(a)_j}{(c)_j}\frac{x^j}{j!},$$

where $(a)_n = a(a+1)\dots(a+n-1)$, $(a)_0=1$. When a is a negative integer, say $a=-n$, this function becomes a polynomial of order n. Specifically, it is expressed via the generalized Laguerre polynomial $L_n^{(c-1)}$:

$${}_1F_1(-n,c,x) = \frac{n!}{(c)_n}L_n^{(c-1)}(x) = \sum_{j=0}^n \binom{n}{j}\frac{(-x)^j}{(c)_j}.$$

The following identity is known as Kummer's transformation:

$${}_1F_1(a,c,-x) = e^{-x}{}_1F_1(c-a,c,x).$$

Therefore, combining two above formulas, we derive that for $n\in\mathbb{N}$,

$$e^{-x}{}_1F_1(c+n,c,x) = {}_1F_1(-n,c,-x) = \sum_{j=0}^n \binom{n}{j}\frac{x^j}{(c)_j}.$$

In particular,

$$\begin{aligned} e^{-x}{}_1F_1(c+1,c,x) &= 1+\frac{x}{c},\\ e^{-x}{}_1F_1(c+2,c,x) &= 1+\frac{2x}{c}+\frac{x^2}{c(c+1)},\\ e^{-x}{}_1F_1(c+3,c,x) &= 1+\frac{3x}{c}+\frac{3x^2}{c(c+1)}+\frac{x^3}{c(c+1)(c+2)}.\end{aligned} \tag{A1}$$

Appendix B: A limit theorem for equations with nonhomogeneous coefficients and non-Lipschitz diffusion

Here we present a limit theorem from[23]. Consider the following sequence of stochastic differential equations:

$$X_n(t) = X_n(0) + \int_0^t b_n\big(s,X_n(s)\big)ds + \int_0^t \sigma_n\big(s,X_n(s)\big)dW(s),$$

$n\ge 0$, where the initial conditions $X_n(0)$ are nonrandom. Assume that the coefficients of these equations satisfy the following conditions:

($Y1_n$) b_n and σ_n are continuous with respect to all arguments;

($Y2_n$) linear growth:

$$|\sigma_n(t,x)| + |b_n(t,x)| \le L(1+|x|), \quad t\ge 0,\ x\in\mathbb{R};$$

($Y3_n$) Lipschitz condition for b_n:

$$|b_n(t,x)-b_n(t,y)| \le L|x-y|, \quad t \ge 0,\ x,y \in \mathbb{R};$$

($Y4_n$) there exists an increasing function $\rho_n\colon \mathbb{R}^+ \to \mathbb{R}^+$ such that $\int_{0+} \rho_n^{-2}(u)du = \infty$ and

$$|\sigma_n(t,x)-\sigma_n(t,y)| \le \rho_n(|x-y|), \quad t \ge 0, \quad x,y \in \mathbb{R}.$$

Additionally, assume that as $n \to \infty$, the following convergences hold:

$$X_n(0) \to X_0(0), \quad b_n(t,x) \to b_0(t,x), \quad \sigma_n(t,x) \to \sigma_0(t,x) \tag{B1}$$

for $t \ge 0$ and $x \in \mathbb{R}$.

Theorem B.1 ([23] (Theorem 4.1)). *If conditions* ($Y1_n$)–($Y4_n$) *and* (B1) *hold, then*

$$\mathbf{E}\left[|X_n(t)-X_0(t)|\right] \to 0, \quad n \to \infty,$$

uniformly in any finite interval.

Appendix C: Functional and other limit theorems for the solutions of stochastic differential equations

Let us consider an equation of the form

$$d\xi_t = a(\xi_t)dt + dW_t, \quad t > 0, \quad \xi_0 = x_0, \tag{C1}$$

with real measurable drift coefficient satisfying additional assumption: $|xa(x)| \le M$ for a certain constant M and for all $x \in \mathbb{R}$. We use the following notation:

$$\psi(x,c) = \frac{1}{\log|x|}\int_0^x a(v)dv - c.$$

Theorem C.1 ([27] (Theorem 3.1, item 2.)). *Let ξ be a solution to equation (C1) and let*

$$\lim_{|x|\to+\infty} \psi(x,c_0) = 0.$$

Then for $2c_0 > -1$ the solution ξ is stochastically unstable.

For investigating the behavior when $T \to +\infty$ of the distribution of the normalized random process $\xi_T(t) = \frac{\xi_{tT}}{\sqrt{T}}$, $t > 0$, where T is parameter, we study the weak convergence to some limit process ζ in the following sense.

Definition C.2. *A family $\xi_T = \{\xi_T(t), t \ge 0\}$ of stochastic processes is said to converge weakly, as $T \to +\infty$, to a process $\zeta = \{\zeta(t), t \ge 0\}$ if, for any $L > 0$, the measures $\mu_T[0,L]$, generated by the processes $\xi_T(\cdot)$ on the interval $[0,L]$ converge weakly to the measure $\mu[0,L]$ generated by the process $\zeta(\cdot)$.*

Remark C.3. Since the processes ξ_T are continuous with probability 1 as the solutions to Itô's stochastic differential equations, Definition C.2 is a definition of the weak convergence of the processes ξ_T to the continuous process ζ in a uniform topology of the space of continuous functions.

Theorem C.4 ([27] (Theorem 3.3(2))). *Let ξ be a solution to equation (C1), and let there exist the constants c_1 and c_2 such that*

$$\lim_{|x|\to+\infty}\left[\frac{1}{x}\int_0^x va(v)dv - \overline{c}(x)\right] = 0, \quad \overline{c}(x) = \begin{cases} c_1, & x \ge 0, \\ c_2, & x < 0. \end{cases}$$

If $2c_1 > 1$ and $2c_2 < 1$, then the stochastic process $|\xi_{tT}|T^{-\frac{1}{2}}$ converges weakly, as $T \to +\infty$, to the process Y, which is the solution of Itô's stochastic differential equation

$$Y_t^2 = (2c+1)t + 2\int_0^t Y_s\, dW_s$$

for $c = c_1$.

REFERENCES

[1] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," Annalen der Physik **322**, 549–560 (1905).
[2] M. Von Smoluchowski, "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen," Annalen der physik **326**, 756–780 (1906).
[3] A. Einstein and M. von Smoluchowski, *Untersuchungen über die Theorie der Brownschen Bewegung. Abhandlung über die Brownsche Bewegung und verwandte Erscheinungen*, Ostwalds Klassiker der exakten Wissenschaften, Band 199 (Verlag Harri Deutsch, 1997).
[4] A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover publications, New-York, 1956) edited with notes by R. Fürth.
[5] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, 2nd ed., North-Holland Mathematical Library, Vol. 24 (North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989) pp. xvi+555.
[6] A. S. Cherny and H.-J. Engelbert, *Singular stochastic differential equations*, Lecture Notes in Mathematics, Vol. 1858 (Springer-Verlag, Berlin, 2005) pp. viii+128.
[7] D. Brigo and F. Mercurio, "A deterministic-shift extension of analytically-tractable and time-homogeneous short-rate models," Finance Stoch. **5**, 369–387 (2001).
[8] M. Di Francesco and K. Kamm, "How to handle negative interest rates in a CIR framework," SeMA J. **79**, 593–618 (2022).
[9] Y. Maghsoodi, "Solution of the extended CIR term structure and bond option valuation," Math. Finance **6**, 89–109 (1996).
[10] G. Orlando, R. M. Mininni, and M. Bufalo, "Interest rates calibration with a CIR model," The Journal of Risk Finance **20**, 370–387 (2019).
[11] C. Kelly and G. J. Lord, "An adaptive splitting method for the Cox-Ingersoll-Ross process," Appl. Numer. Math. **186**, 252–273 (2023).
[12] J. C. Cox, J. E. Ingersoll, and S. A. Ross, "A theory of the term structure of interest rates," Econometrica **53**, 385–407 (1985).
[13] A. Göing-Jaeschke and M. Yor, "A survey and some generalizations of Bessel processes," Bernoulli **9**, 313–349 (2003).
[14] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3rd ed., Grundlehren der mathematischen Wissenschaften, Vol. 293 (Springer-Verlag, Berlin, 1999) pp. xiv+602.
[15] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics, Vol. 113 (Springer-Verlag, New York, 1991) pp. xxiv+470.
[16] Y. Mishura, A. Pilipenko, and A. Yurchenko-Tytarenko, "Low-dimensional Cox-Ingersoll-Ross process," Stochastics (2024), dOI:10.1080/17442508.2023.2300291.

[17] Y. Mishura and A. Yurchenko-Tytarenko, "Standard and fractional reflected Ornstein-Uhlenbeck processes as the limits of square roots of Cox-Ingersoll-Ross processes," Stochastics **95**, 99–117 (2023).
[18] K. Oldham, J. Myland, and J. Spanier, *An atlas of functions*, 2nd ed. (Springer, New York, 2009).
[19] A. Alfonsi, *Affine diffusions and related processes: simulation, theory and applications*, Bocconi & Springer Series, Vol. 6 (Springer, Cham; Bocconi University Press, Milan, 2015) pp. xiv+252.
[20] O. Dehtiar, Y. Mishura, and K. Ralchenko, "Two methods of estimation of the drift parameters of the Cox–Ingersoll–Ross process: continuous observations," Comm. Statist. Theory Methods **51**, 6818–6833 (2022).
[21] O. Okhrin, M. Rockinger, and M. Schmid, "Distributional properties of continuous time processes: from CIR to Bates," AStA Adv. Stat. Anal. **107**, 397–419 (2023).
[22] M. Ben Alaya and A. Kebaier, "Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions," Stoch. Anal. Appl. **31**, 552–573 (2013).
[23] Y. Mishura, S. Posashkova, and G. Shevchenko, "Properties of solutions of stochastic differential equations with nonhomogeneous coefficients and non-Lipschitz diffusion," Theory Probab. Math. Statist. **79**, 117–126 (2009).
[24] T. Yamada and S. Watanabe, "On the uniqueness of solutions of stochastic differential equations," J. Math. Kyoto Univ. **11**, 155–167 (1971).
[25] Y. Mishura and G. Shevchenko, *Theory and Statistical Applications of Stochastic Processes* (ISTE Ltd and John Wiley & Sons, 2017).
[26] R. S. Liptser and A. N. Shiryayev, *Theory of martingales*, Mathematics and its Applications (Soviet Series), Vol. 49 (Kluwer Academic Publishers Group, Dordrecht, 1989) pp. xiv+792.
[27] G. Kulinich, S. Kushnirenko, and Y. Mishura, *Asymptotic analysis of unstable solutions of stochastic differential equations*, Bocconi & Springer Series, Vol. 9 (Bocconi University Press; Springer, 2020) p. 248.
[28] A. G. Cherstvy, A. V. Chechkin, and R. Metzler, "Anomalous diffusion and ergodicity breaking in heterogeneous diffusion processes," New J. Phys. **15**, 083039, 13 (2013).
[29] M. Abramowitz and I. A. Stegun, eds., *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (Dover Publications, Inc., New York, 1992).