BIT Numer Math (2011) 51:405–425
DOI 10.1007/s10543-010-0288-y

BIT

Numerical simulation of a strongly nonlinear Ait-Sahalia-type interest rate model

Lukasz Szpruch · Xuerong Mao · Desmond J. Higham · Jiazhu Pan

Received: 31 July 2009 / Accepted: 21 September 2010 / Published online: 6 November 2010

Abstract We are interested in the strong convergence of Euler-Maruyama type approximations to the solution of a class of stochastic differential equations models with highly nonlinear coefficients, arising in mathematical finance. Results in this area can be used to justify Monte Carlo simulations for calibration and valuation. The equations that we study include the Ait-Sahalia type model of the spot interest rate, which has a polynomial drift term that blows up at the origin and a diffusion term with superlinear growth. After establishing existence and uniqueness for the solution, we show that an appropriate implicit numerical method preserves positivity and boundedness of moments, and converges strongly to the true solution.

Keywords Interest rate · Model calibration · Monte Carlo · Moment bound · Implicit · Ito

Mathematics Subject Classification (2000) 65C30

Communicated by Anders Szepessy.

L. Szpruch · X. Mao · D.J. Higham (✉) · J. Pan
Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK
e-mail: d.j.higham@strath.ac.uk

L. Szpruch
e-mail: lukas.szpruch@strath.ac.uk

X. Mao
e-mail: x.mao@strath.ac.uk

J. Pan
e-mail: jiazhu.pan@strath.ac.uk

Present address:
L. Szpruch
Mathematical Institute, 24-29 St. Giles', Oxford, OX1 3LB, UK

1 Introduction

Ait-Sahalia [1] investigated several continuous-time interest rate models empirically. He tested parametric models by comparing their implied densities with the density estimated nonparametrically. In application to Eurodollar interest rates, he rejected all existing univariate linear drift models. This led to the proposal of a new class of highly nonlinear stochastic differential equations (SDEs) to model interest rates.

Subsequent studies supported the observations made by Ait-Sahalia. Stanton [21], using nonparametric kernel regression, also found significant nonlinearities in spot rate data. Some authors [3, 7] pointed out that the Ait-Sahalia test has poor finite sample performance because of persistent dependence in interest rate data and slow convergence of the nonparametric density estimator. However, Hong and Li [15] developed the so called omnibus nonparametric specification test for continuous-time models based on the transition density function, which, unlike the marginal density used by Ait-Sahalia, captures the full dynamics of continuous process. Their test rejected all but the Ait-Sahalia and CKLS [6] model. Along with Ait-Sahalia, Conley et al. [10] and Gallant et al. [8], have used a variety of empirical techniques to estimate model parameters; and all have suggested that the diffusion term in the SDE grows faster than linearly.

In this work we focus on numerical issues arising from SDE models of Ait-Sahalia type. In particular, we are concerned with guaranteeing convergence in numerical simulation; this is clearly a fundamental requirement if the model is to be correctly calibrated and used for valuing financial products in a Monte Carlo setting.

The SDE that we study, which we refer to as the *generalized Ait-Sahalia model*, has the form

$$dx(t) = (\alpha_{-1}x(t)^{-1} - \alpha_0 + \alpha_1 x(t) - \alpha_2 x(t)^r)dt + \sigma x(t)^\rho dw(t), \qquad (1.1)$$

where $\alpha_{-1}, \alpha_0, \alpha, \alpha_1, \alpha_2, \sigma$ are positive constants and $\rho > 1$. In addition, we have introduced a further parameter $r > 1$. Here, $w(t)$ is a scalar Brownian motion defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions, that is to say, it is right continuous and increasing while $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.

The model (1.1) clearly violates the Lipschitz and linear growth conditions which are traditionally imposed in the study of SDEs and their simulation (see e.g. [9, 17, 18]). In particular, although recent work, such as [2, 4, 5, 13, 19], has dealt with numerical simulations of SDEs with diffusion coefficient of the form $\sigma x(t)^\rho$ with $\rho \in [0.5, 1]$, we are not aware of any results that apply to the case of $\rho > 1$, even when the drift is linear. A further complication in (1.1) is that the drift contains a term $\alpha_{-1}x(t)^{-1}$ that does not behave well near the origin.

In Sect. 2 we prove the existence of a unique solution to (1.1). In his original paper Ait-Sahalia applied the Feller test and derived conditions under which the solution will not explode with probability 1. We give an alternative proof that extends the space of admissible parameters. We also show that the pth moment ($p \in (-\infty, \infty)$) of the solution is bounded uniformly in time. In Sect. 3 we propose the backward Euler-Maruyama (BEuM) scheme, which is also known as implicit Euler-Maruyama, to approximate the solution of (1.1). A particularly interesting observation is that, unlike explicit Euler-Maruyama, it preserves positivity. This gives us some intuition that implicit schemes capture more qualitative information about solutions of SDEs (see also the discussion in [20]). In Sect. 4 we show that the pth moment of the BEuM, and a continuous-time extension, can be bounded uniformly in time in pth moment. In Sect. 5 we introduce a new numerical method, which we call forward-backward Euler-Maruyama (FBEuM). We derive moment bounds for FBEuM and a continuous-time extension. Section 6 contains the main results of this work: strong convergence proofs for FBEuM and BEuM applied to (1.1). In Sect. 7 we point out some immediate applications of our results for pricing path-dependent derivatives.

2 Existence and uniqueness of solution

In order to show that the model (1.1) is meaningful, the next theorem guarantees that a unique solution exists, and remains in $\mathbb{R}_+ := (0, \infty)$.

Theorem 2.1 *Given any initial value* $x(0) = x_0 > 0$, *there exists a unique*, *positive global solution* $x(t)$ *to* (1.1) *on* $t \geq 0$.

Proof Define the coefficients of (1.1):

$$f(x) = \alpha_{-1}x^{-1} - \alpha_0 + \alpha_1 x - \alpha_2 x^r \quad \text{and} \quad g(x) = \sigma x^\rho \quad \text{for } x > 0. \tag{2.1}$$

Clearly, they are locally Lipschitz continuous in $(0, \infty)$. Following the standard truncation method (see e.g. [9, 18]), we can show that for any given initial value $x_0 > 0$ there exists a unique maximal local solution $x(t)$, $t \in [0, \tau_e)$, where τ_e is the stopping time of the explosion or first zero time. To prove our theorem, we need to show that $\tau_e = \infty$ a.s.

For every sufficiently large integer $k > 0$, such that $1/k < x(0) < k$, define the stopping time

$$\tau_k = \inf\{t \in [0, \tau_e) : x(t) \notin (1/k, k)\},$$

where throughout this paper we set $\inf(\emptyset) = \infty$. Obviously τ_k is increasing as $k \to \infty$. Set $\tau_\infty = \lim_{k\to\infty} \tau_k$, whence $\tau_\infty \leq \tau_e$ a.s. If we can prove $\tau_k \to \infty$ a.s. as $k \to \infty$, then $\tau_e = \infty$ a.s. and $x(t) \geq 0$ a.s. for all $t \geq 0$. In other words, to complete the proof what we need to show is that $\tau_\infty = \infty$ a.s. To prove this, it is enough to show that $P\{\tau_k \leq T\} \to 0$ as $k \to \infty$ for any given constant $T > 0$, for this immediately implies that $P\{\tau_\infty = \infty\} = 1$ as required.

Fix two constants $\gamma_1 \in (0, 1)$ and $\gamma_2 > 1$. Let us define a function $V \in C^2(\mathbb{R}_+, \mathbb{R}_+)$ by

$$V(x) = x^{\gamma_1} + x^{-\gamma_2}. \tag{2.2}$$

It is easy to see that $V(x) \to \infty$ as $x \to \infty$ or $x \to 0$. Compute the diffusion operator

$$\begin{aligned} LV(x) &= V_x(x) f(x) + \frac{1}{2} V_{xx}(x) g(x)^2 \\ &= (\gamma_1 x^{\gamma_1 - 1} - \gamma_2 x^{-(\gamma_2+1)}) f(x) \end{aligned}$$

$$+\frac{1}{2}(\gamma_1(\gamma_1-1)x^{\gamma_1-2}+\gamma_2(\gamma_2+1)x^{-(\gamma_2+2)})g(x)^2$$
$$=\gamma_1\alpha_{-1}x^{\gamma_1-2}-\alpha_0\gamma_1x^{\gamma_1-1}+\alpha_1\gamma_1x^{\gamma_1}-\alpha_2\gamma_1x^{\gamma_1-1+r}$$
$$-\alpha_{-1}\gamma_2x^{-(\gamma_2+2)}+\alpha_0\gamma_2x^{-(\gamma_2+1)}-\alpha_1\gamma_2x^{-\gamma_2}+\gamma_2x^{-(\gamma_2+1)+r}$$
$$+\frac{\sigma^2}{2}(\gamma_1(\gamma_1-1)x^{\gamma_1-2+2\rho}+\gamma_2(\gamma_2+1)x^{-(\gamma_2+2)+2\rho}).$$

Recalling that $\gamma_1\in(0,1)$ and $\gamma_2>1$, we can find a constant K such that

$$LV(x)\le K. \tag{2.3}$$

By the Itô formula,

$$\mathbb{E}V(x(T\wedge\tau_k))\le V(x_0)+KT. \tag{2.4}$$

Therefore

$$\mathbb{P}(\tau_k\le T)[V(1/k)\wedge V(k)]\le\mathbb{E}V(x(T\wedge\tau_k))\le V(x(0))+KT.$$

This implies that $\lim_{k\to\infty}\mathbb{P}(\tau_k\le T)=0$ as desired. The proof is complete. □

In order to proceed with our analysis, we make an assumption about the parameter values. As will become clear from the proofs, this type of assumption allows us to control the potential growth coming from the diffusion term using the dissipative nature of the drift.

Assumption 2.2 *The parameters in* (1.1) *obey*

$$r+1>2\rho.$$

The following lemma gives moment bounds for the solution of the SDE.

Lemma 2.1 *Under Assumption* 2.2, *for any* $p\ge 2$,

$$\sup_{0\le t<\infty}\mathbb{E}|x(t)|^p<\infty \tag{2.5}$$

and

$$\sup_{0\le t<\infty}\mathbb{E}\left(\frac{1}{|x(t)|^p}\right)<\infty. \tag{2.6}$$

Proof For every sufficiently large integer n, define the stopping time

$$\tau_n=\inf\left\{t>0: x(t)\notin\left(\frac{1}{n},n\right)\right\}.$$

Applying the Itô formula to function $V(x,t)=e^t x^p$, we compute the diffusion operator

$$LV(x,t)=e^t\bigg(x^p+px^{p-1}[\alpha_{-1}x^{-1}-\alpha_0+\alpha_1 x-\alpha_2 x^r]+\frac{\sigma^2}{2}p(p-1)x^{p-2+2\rho}\bigg).$$

By Assumption 2.2, there exists a constant $K>0$ such that

$$LV(x,t)\le Ke^t.$$

Therefore

$$\mathbb{E}\Big[e^{t\wedge\tau_n}x(t\wedge\tau_n)^p\Big]\le x_0^p+Ke^t.$$

Letting $n\to\infty$ and applying the Fatou lemma, we have

$$\mathbb{E}|x(t)|^p\le\frac{x_0^p}{e^t}+K,$$

which gives assertion (2.5). In the same way, we can apply the Itô formula to the function $V(x,t)=e^t x^{-p}$ to show (2.6). □

Lemma 2.2 *Under Assumption 2.2, for any* $p\ge 2$,

$$\mathbb{E}(\sup_{0\le t\le T}|x(t)|^p)<\infty,\quad \forall T>0.$$

Proof By the Itô formula, we can show that

$$\begin{aligned}\mathbb{E}[\sup_{0\le t\le T}|x(t)|^p]&\le x_0^p+\mathbb{E}\int_0^T p\big|x(t)^{p-1}(\alpha_{-1}x(t)^{-1}-\alpha_0+\alpha_1 x(t)-\alpha_2 x(t)^r)\\&\quad+0.5(p-1)\sigma^2 x(t)^{2(\rho-1)+p}\big|dt\\&\quad+\mathbb{E}\bigg[\sup_{0\le t\le T}\int_0^t\sigma p x(u)^{\rho+p-1}dw(u)\bigg]\\&\le x(0)^p+KT+\mathbb{E}\bigg[\sup_{0\le t\le T}\int_0^t\sigma p x(u)^{\rho+p-1}dw(u)\bigg],\end{aligned}$$

where K is a constant. By the Hölder and Burkholder-Davis-Gundy (BDG) inequalities, we can show that

$$\begin{aligned}\mathbb{E}\bigg[\sup_{0\le t\le T}\int_0^t\sigma p x(u)^{\rho+p-1}dw(u)\bigg]&\le C\mathbb{E}\bigg(\int_0^T x(t)^{2(\rho+p-1)}dt\bigg)^{\frac{1}{2}}\\&\le C\bigg(\int_0^T\mathbb{E}[x(t)^{2(\rho+p-1)}]dt\bigg)^{\frac{1}{2}},\end{aligned}$$

where C stands for a constant which may vary from line to line. By Lemma 2.1, the conclusion follows. □

3 Backward Euler-Maruyama scheme

In the previous section we showed the existence of a unique global solution to the SDE (1.1), but we are not aware of an explicit expression for the solution or its transition density. We therefore consider computable discrete time approximations that could be used in Monte Carlo simulations. In this initial investigation we focus on the basic property of strong convergence without attempting to establish convergence rates. Two good reasons for considering strong convergence are that

- weak convergence [17] and pathwise convergence [16] follow automatically, and
- efficient multi-level Monte Carlo simulations rely on both weak and strong convergence properties [11, 12].

It was shown in [14] that the classical global Lipschitz assumption, which guarantees strong uniform convergence of Euler-Maruyama, can be significantly relaxed. That work used a local Lipschitz condition plus uniform boundedness of moments of the true solution and its approximation to obtain strong convergence. In this work, we apply and extend those ideas to the problem (1.1), taking care to (a) exploit the controlling influence of the drift and (b) deal with the repulsive nature of the drift at the origin.

Given any step size Δt, we define the partition $\mathcal{P}_{\Delta t} := \{t_k = k\Delta t : k = 0, 1, 2, \dots\}$ of the time interval $[0, \infty)$, and introduce the backward Euler-Maruyama (BEuM) method [14, 17]

$$X_{t_{k+1}} = X_{t_k} + \left(\alpha_{-1} X_{t_{k+1}}^{-1} - \alpha_0 + \alpha_1 X_{t_{k+1}} - \alpha_2 X_{t_{k+1}}^r\right) \Delta t + \sigma X_{t_k}^{\rho} \Delta W_{t_k}, \tag{3.1}$$

where $\Delta W_{t_k} = W_{t_{k+1}} - W_{t_k}$ and $X_{t_0} = x(0)$. The following lemma shows that this implicit method is well defined and preserves positivity of the solution.

Lemma 3.1 *Define, for any given* $\Delta t \le 1/\alpha_1$,

$$F(x) = x - \alpha_{-1} x^{-1} \Delta t + \alpha_0 \Delta t - \alpha_1 x \Delta t + \alpha_2 x^r \Delta t, \quad x \in \mathbb{R}_+.$$

Then for any $b \in \mathbb{R}$ *there exists a unique* $x \in \mathbb{R}_+$ *such that* $F(x) = b$.

Proof The lemma follows if we can show that the function F is continuous, coercive and strictly monotone (see [22]). Clearly, $F(x)$ is continuous on $\mathbb{R}_+$ with $\lim_{x\to\infty} F(x) = \infty$ and $\lim_{x\to 0^+} F(x) = -\infty$, so the function F is coercive on $\mathbb{R}_+$. Since $F'(x) = 1 + (\alpha_{-1} x^{-2} - \alpha_1 + r\alpha_2 x^{r-1})\Delta t > 1 - \alpha_1 \Delta t$, we see that $\dot{F}(x) > 0$ whenever $\Delta t \le 1/\alpha_1$, showing strict monotonicity. □

From now on we always let $\Delta t \le 1/\alpha_1$ so that the BEuM is well defined and preserves positivity.

In contrast, let us point out that the (standard) Euler-Maruyama scheme does not preserve the positivity of the solution to (1.1). In fact, recall that the Euler-Maruyama scheme applied to (1.1) has the form

$$X_{t_{k+1}} = X_{t_k} + \left(\alpha_{-1} X_{t_k}^{-1} - \alpha_0 + \alpha_1 X_{t_k} - \alpha_2 X_{t_k}^r\right) \Delta t + \sigma X_{t_k}^{\rho} \Delta W_{t_k}.$$

Without loss of generality, we assume that $X_{t_k} > 0$ is given. Note that $X_{t_{k+1}} < 0$ is equivalent to $\Delta W_{t_k} < -(X_{t_k} + (\alpha_{-1} X_{t_k}^{-1} - \alpha_0 + \alpha_1 X_{t_k} - \alpha_2 X_{t_k}^r)\Delta t)/\sigma X_{t_k}^\rho := K(X_{t_k})$, but clearly $\mathbb{P}(\Delta W_{t_k} < K(X_{t_k})) > 0$.

4 Moment properties of BEuM

4.1 Boundedness of moments

We will work on the discrete filtration $\{\mathcal{F}_{t_k}\}_{k\geq 0}$. By Lemma 3.1, X_{t_k} is $\mathcal{F}_{t_k}$-measurable.

Lemma 4.1 *Let $r > 1$, then for any $p \geq 2$ and sufficiently large integer n, there exists a constant $K(p, n)$, such that*

$$\sup_{\Delta t \leq 1/2\alpha_1} \mathbb{E}\left|X_{t_k}\right|^p \mathbf{1}_{[0,\lambda_n]}(k) < K(p,n) \quad \textit{for any } k \geq 0,$$

where

$$\lambda_n = \inf\left\{k : X_{t_k} \notin \left(\frac{1}{n}, n\right)\right\}. \tag{4.1}$$

Proof We observe that when $k \in [0, \lambda_n]$, $X_{t_{k-1}} \in (\frac{1}{n}, n)$, but X_{t_k} may not stay in $(\frac{1}{n}, n)$, so the lemma is not obvious. Because

$$\left(X_{t_{k+1}} - f(X_{t_{k+1}})\Delta t\right)^2 = \left(X_{t_k} + g(X_{t_k})\Delta W_{t_k}\right)^2,$$

we have

$$\left|X_{t_{k+1}}\right|^2 - \left|X_{t_k}\right|^2 \leq 2X_{t_{k+1}} f(X_{t_{k+1}})\Delta t + 2X_{t_k} g(X_{t_k})\Delta W_{t_k} + g(X_{t_k})^2(\Delta W_{t_k})^2. \tag{4.2}$$

Recalling that $\Delta t < \frac{1}{2\alpha_1}$, we see from (4.2) that

$$\begin{aligned} 0 &< \left|X_{t_{k+1}}\right|^2 - 2X_{t_{k+1}} f(X_{t_{k+1}})\Delta t + 2\alpha_{-1}\Delta t \\ &\leq \left|X_{t_k}\right|^2 + 2X_{t_k} g(X_{t_k})\Delta W_{t_k} + g^2(X_{t_k})(\Delta W_{t_k})^2 + 2\alpha_{-1}\Delta t. \end{aligned} \tag{4.3}$$

Without any loss of generality, we may let $p \geq 2$ be an even number. Raising both sides of (4.3) to the power p gives

$$\begin{aligned} &\sum_{j=0}^{p}\sum_{s=0}^{p-j}(-1)^s A\left(\left|X_{t_{k+1}}\right|^2\right)^{p-j-s}\left(2X_{t_{k+1}} f(X_{t_{k+1}})\Delta t\right)^s (2\alpha_{-1}\Delta t)^j \\ &\quad\leq \sum_{j=0}^{p}\sum_{s=0}^{p-j}\sum_{i=0}^{j} B\left(\left|X_{t_k}\right|^2\right)^{p-j-s}\left(2X_{t_k} g(X_{t_k})\Delta W_{t_k}\right)^s\left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i} \\ &\quad\quad\times (2\alpha_{-1}\Delta t)^i, \end{aligned} \tag{4.4}$$

where $A = A(p, j, s)$ and $B = B(p, j, s, i)$ are all positive constants. Rearranging this inequality we get

$$\begin{aligned}
\left|X_{t_{k+1}}\right|^{2p} \le \left|X_{t_k}\right|^{2p} &+ \sum_{s=1}^{p}(-1)^{s+1}A\left(\left|X_{t_{k+1}}\right|^2\right)^{p-s}\left(2X_{t_{k+1}}f(X_{t_{k+1}})\Delta t\right)^s \\
&+ \sum_{j=1}^{p}\sum_{s=0}^{p-j}(-1)^{s+1}A\left(\left|X_{t_{k+1}}\right|^2\right)^{p-j-s}\left(2X_{t_{k+1}}f(X_{t_{k+1}})\Delta t\right)^s(2\alpha_{-1}\Delta t)^j \\
&+ \sum_{s=1}^{p}B\left(\left|X_{t_k}\right|^2\right)^{p-s}\left(2X_{t_k}g(X_{t_k})\Delta W_{t_k}\right)^s \\
&+ \sum_{j=1}^{p}\sum_{s=0}^{p-j}\sum_{i=0}^{j}B\left(\left|X_{t_k}\right|^2\right)^{p-j-s}\left(2X_{t_k}g(X_{t_k})\Delta W_{t_k}\right)^s \\
&\times\left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i}(2\alpha_{-1}\Delta t)^i.
\end{aligned}$$

By Assumption 2.2, there exists a constant $K = K(p) > 0$, independent of k, and Δt, such that

$$\begin{aligned}
&\sum_{s=1}^{p}(-1)^{s+1}A\left(\left|X_{t_{k+1}}\right|^2\right)^{p-s}\left(2X_{t_{k+1}}f(X_{t_{k+1}})\Delta t\right)^s \\
&\quad+ \sum_{j=1}^{p}\sum_{s=0}^{p-j}(-1)^{s+1}A\left(\left|X_{t_{k+1}}\right|^2\right)^{p-j-s}\left(2X_{t_{k+1}}f(X_{t_{k+1}})\Delta t\right)^s(2\alpha_{-1}\Delta t)^j < K(p).
\end{aligned}$$

For sufficiently large integer n , such that $1/n < X_{t_0} < n$, by Hölder's inequality and properties of ΔW_{t_k}, there exists a constant $K(p, n) > 0$, such that

$$\begin{aligned}
&\mathbb{E}\left|X_{t_{k+1}}\right|^{2p}\mathbf{1}_{[0,\lambda_n]}(k+1) \\
&\le K(p) + \mathbb{E}\left|X_{t_k}\right|^{2p}\mathbf{1}_{[0,\lambda_n]}(k+1) \\
&\quad+ \mathbb{E}\sum_{s=1}^{p}B\left(\left|X_{t_k}\right|^2\right)^{p-s}\left(2X_{t_k}g(X_{t_k})\Delta W_{t_k}\right)^s\mathbf{1}_{[0,\lambda_n]}(k+1) \\
&\quad+ \mathbb{E}\sum_{j=1}^{p}\sum_{s=0}^{p-j}\sum_{i=0}^{j}B\left(\left|X_{t_k}\right|^2\right)^{p-j-s}\left(2X_{t_k}g(X_{t_k})\Delta W_{t_k}\right)^s\left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i} \\
&\quad\times(2\alpha_{-1}\Delta t)^i\mathbf{1}_{[0,\lambda_n]}(k+1) < K(p, n),
\end{aligned}$$

which is the required assertion. □

Theorem 4.1 *Let Assumption 2.2 hold. Let $\Delta t^* \in (0, 1/2\alpha_1)$ be sufficiently small so that whenever $\Delta t \le \Delta t^*$,*

$$0 < \frac{[e^{\Delta t} - 1]}{\Delta t} \le 2 \quad and \quad e^{\Delta t} \le 2. \tag{4.5}$$

Then

$$\sup_{\Delta t \le \Delta t^*} \sup_{k \ge 0} \mathbb{E}\left|X_{t_k}\right|^2 < \infty.$$

Proof Let n be a sufficiently large integer. Clearly λ_n in (4.1) is a stopping time with respect to $\{\mathcal{F}_{t_k}\}_{k\ge 0}$. By (4.2), writing $\Delta W_{t_k} = \sqrt{\Delta t}\xi_{t_{k+1}}$, where the $\xi_{t_{k+1}}$ are $i.i.d.$ with $\xi_{t_{k+1}} \sim N(0, 1)$ and independent of $\mathcal{F}_{t_k}$, we have

$$\begin{aligned}\left|X_{t_{k+1}}\right|^2 - \left|X_{t_k}\right|^2 &\le 2(\alpha_{-1} - \alpha_0 X_{t_{k+1}} + \alpha_1 X_{t_{k+1}}^2 - \alpha_2 X_{t_{k+1}}^{r+1})\Delta t \\ &\quad + 2X_{t_k}^{\rho+1}\sqrt{\Delta t}\xi_{t_{k+1}} + \sigma^2 X_{t_k}^{2\rho}\Delta t \xi_{t_{k+1}}^2.\end{aligned}$$

Note that

$$\begin{aligned}e^{t_{k+1}} X_{t_{k+1}}^2 - e^{t_k} X_{t_k}^2 &= e^{t_k}\left[X_{t_{k+1}}^2 - X_{t_k}^2\right] + \left[e^{t_{k+1}} - e^{t_k}\right] X_{t_{k+1}}^2 \\ &= e^{t_k}\left[X_{t_{k+1}}^2 - X_{t_k}^2\right] + e^{t_k}\frac{[e^{\Delta t} - 1]}{\Delta t}\Delta t X_{t_{k+1}}^2.\end{aligned}$$

Recalling (4.5), we hence have

$$\begin{aligned}&e^{t_{k+1}}\left|X_{t_{k+1}}\right|^2 - e^{t_k}\left|X_{t_k}\right|^2 \\ &\le e^{t_k}\Big[2(\alpha_{-1} - \alpha_0 X_{t_{k+1}} + (\alpha_1 + 2)X_{t_{k+1}}^2 - \alpha_2 X_{t_{k+1}}^{r+1})\Delta t \\ &\quad + 2X_{t_k}^{\rho+1}\sqrt{\Delta t}\xi_{t_{k+1}} + \sigma^2 X_{t_k}^{2\rho}\Delta t \xi_{t_{k+1}}^2\Big].\end{aligned}$$

Let N be any nonnegative integer. Summing both sides of the above inequality from $k = 0$ to $N \wedge \lambda_n$, we get

$$\begin{aligned}&e^{t_{N\wedge\lambda_n+1}}\left|X_{t_{N\wedge\lambda_n+1}}\right|^2 \\ &\le e^{t_0}\left(\left|X_{t_0}\right|^2 + \sigma^2 X_{t_0}^{2\rho}\Delta t\xi_{t_1}^2 + 2X_{t_0}^{\rho+1}\sqrt{\Delta t}\xi_{t_1}\right) \\ &\quad + \sum_{k=1}^{N\wedge\lambda_n+1} 2e^{t_{k-1}}(\alpha_{-1} - \alpha_0 X_{t_k} + (\alpha_1 + 2)X_{t_k}^2 - \alpha_2 X_{t_k}^{r+1})\Delta t \\ &\quad + \sum_{k=1}^{N\wedge\lambda_n} e^{t_k}\sigma^2 X_{t_k}^{2\rho}\Delta t\xi_{t_{k+1}}^2 + \sum_{k=1}^{N\wedge\lambda_n} 2e^{t_k}X_{t_k}^{\rho+1}\sqrt{\Delta t}\xi_{t_{k+1}} \\ &= e^{t_0}\left(\left|X_{t_0}\right|^2 + \sigma^2 X_{t_0}^{2\rho}\Delta t\xi_{t_1}^2 + 2X_{t_0}^{\rho+1}\sqrt{\Delta t}\xi_{t_1}\right)\end{aligned}$$

$$+\sum_{k=1}^{N+1} 2e^{t_{k-1}}(\alpha_{-1}-\alpha_0 X_{t_k}+(\alpha_1+2)X_{t_k}^2-\alpha_2 X_{t_k}^{r+1})\mathbf{1}_{[0,\lambda_n+1]}(k)\Delta t$$

$$+\sum_{k=1}^{N} e^{t_k}\sigma^2 X_{t_k}^{2\rho}\mathbf{1}_{[0,\lambda_n]}(k)\Delta t\xi_{t_{k+1}}^2+\sum_{k=1}^{N} 2e^{t_k}X_{t_k}^{\rho+1}\mathbf{1}_{[0,\lambda_n]}(k)\sqrt{\Delta t}\xi_{t_{k+1}}.$$

Applying Lemma 4.1 and noting that X_{t_k} and $\mathbf{1}_{[0,\lambda_n]}(k)$ are $\mathcal{F}_{t_k}$-measurable while $\xi_{t_{k+1}}$ is independent of $\mathcal{F}_{t_k}$, we can take the expectation on both sides of (4.6) to get

$$\begin{aligned}
&\mathbb{E}e^{t_{N\wedge\lambda_n+1}}\left|X_{t_{N\wedge\lambda_n+1}}\right|^2\\
&\quad\le K_1+\mathbb{E}\left[\sum_{k=1}^{N+1} 2e^{t_{k-1}}(\alpha_{-1}-\alpha_0 X_{t_k}+(\alpha_1+2)X_{t_k}^2-\alpha_2 X_{t_k}^{r+1})\mathbf{1}_{[0,\lambda_n+1]}(k)\Delta t\right]\\
&\qquad+\mathbb{E}\left[\sum_{k=1}^{N} e^{t_k}\sigma^2 X_{t_k}^{2\rho}\mathbf{1}_{[0,\lambda_n]}(k)\Delta t\right],\\
&\quad\le K_1+\mathbb{E}\Bigg[\sum_{k=1}^{N+1} 2e^{t_{k-1}}(\alpha_{-1}-\alpha_0 X_{t_k}+(\alpha_1+2)X_{t_k}^2-\alpha_2 X_{t_k}^{r+1}+\sigma^2 X_{t_k}^{2\rho})\\
&\qquad\times\mathbf{1}_{[0,\lambda_n+1]}(k)\Delta t\Bigg],
\end{aligned}$$

where $K_1=e^{t_0}|X_{t_0}|^2+e^{t_0}\sigma^2 X_{t_0}^{2\rho}\Delta t^*$. By Assumption 2.2, there exists a constant $K>0$, such that

$$\alpha_{-1}-\alpha_0 x+(\alpha_1+2)x^2-\alpha_2 x^{r+1}+\sigma^2 x^{2\rho}\le K,\quad x>0.$$

We hence obtain that

$$\mathbb{E}e^{t_{N\wedge\lambda_n+1}}\left|X_{t_{N\wedge\lambda_n+1}}\right|^2\le K_1+2K\left[\sum_{k=1}^{N+1} e^{t_{k-1}}\Delta t\right]. \tag{4.6}$$

But

$$\sum_{k=1}^{N+1} e^{t_{k-1}}\Delta t\le\int_0^{t_{N+1}} e^t\,dt\le e^{t_{N+1}}.$$

Thus, letting $n\to\infty$ in (4.6) and applying the Fatou lemma, we get

$$e^{t_{N+1}}\mathbb{E}\left|X_{t_{N+1}}\right|^2\le K_1+2Ke^{t_{N+1}},$$

which implies

$$\mathbb{E}\left|X_{t_{N+1}}\right|^2\le K_1+2K,\quad\forall N\ge 0,$$

as required. The proof is complete. □

In the next section we will need a stronger version of Theorem 4.1, namely the boundedness of pth moment for $p > 2$. The following theorem is of course a generalization of Theorem 4.1. The reason why we present Theorem 4.1 and give a detailed proof above rather than present the following more general theorem directly is to make our proofs more understandable.

Theorem 4.2 *Under Assumption 2.2, for any $p > 2$, there is a $\Delta t^* \in (0, 1/2\alpha_1)$ such that*

$$\sup_{\Delta t \le \Delta t^*} \sup_{k \ge 0} \mathbb{E}\left|X_{t_k}\right|^p < \infty.$$

Proof The proof is analogous to that of Theorem 4.1 so we only highlight the main steps. Without loss of generality we assume that p is an even number. Rearranging inequality (4.4) we get

$$\begin{aligned}
&\left|X_{t_{k+1}}\right|^{2p} - \left|X_{t_k}\right|^{2p} \\
&\le \sum_{s=1}^{p} (-1)^{s+1} \binom{p}{s} \left(\left|X_{t_{k+1}}\right|^2\right)^{p-s} \left(2X_{t_{k+1}} f(X_{t_{k+1}})\right)^s \Delta t^s \\
&\quad + \sum_{j=1}^{p} \sum_{s=0}^{p-j} (-1)^{s+1} \binom{p}{j} \binom{p-j}{s} \left(\left|X_{t_{k+1}}\right|^2\right)^{p-j-s} \\
&\quad \times \left(2X_{t_{k+1}} f(X_{t_{k+1}})\right)^s (2\alpha_{-1})^j \Delta t^{j+s} \\
&\quad + \sum_{s=1}^{p/2} \binom{p}{2s} \left(\left|X_{t_k}\right|^2\right)^{p-2s} \left(2X_{t_k} g(X_{t_k}) \Delta W_{t_k}\right)^{2s} \\
&\quad + \sum_{\substack{s=1 \\ s=odd}}^{p} \binom{p}{s} \left(\left|X_{t_k}\right|^2\right)^{p-s} \left(2X_{t_k} g(X_{t_k}) \Delta W_{t_k}\right)^s \\
&\quad + \sum_{j=1}^{p} \sum_{i=0}^{j} \binom{p}{j} \binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j} \left[g(X_{t_k})^2\right]^{j-i} [(\Delta W_{t_k})^2]^{j-i} (2\alpha_{-1} \Delta t)^i \\
&\quad + \sum_{\substack{j=1 \\ j=odd}}^{p} \sum_{s=1}^{(p-j)/2} \sum_{i=0}^{j} \binom{p}{j} \binom{p-j}{2s} \binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j-2s} \\
&\quad \times \left(2X_{t_k} g(X_{t_k})\right)^{2s} \left(g(X_{t_k})^2 (\Delta W_{t_k})^2\right)^{j-i} \left(\Delta W_{t_k}\right)^{2s} (2\alpha_{-1} \Delta t)^i \\
&\quad + \sum_{\substack{j=1 \\ j=even}}^{p} \sum_{s=1}^{(p-j)/2} \sum_{i=0}^{j} \binom{p}{j} \binom{p-j}{2s} \binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j-2s}
\end{aligned}$$

$$
\begin{aligned}
&\times \left(2X_{t_k} g(X_{t_k})\Delta W_{t_k}\right)^{2s} \left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i} (2\alpha_{-1}\Delta t)^i \\
&+ \sum_{\substack{j=1\\ j=odd}}^{p} \sum_{\substack{s=1\\ s=odd}}^{p-j} \sum_{i=0}^{j} \binom{p}{j}\binom{p-j}{s}\binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j-s} \\
&\times \left(2X_{t_k} g(X_{t_k})\Delta W_{t_k}\right)^{s} \left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i} (2\alpha_{-1}\Delta t)^i \\
&+ \sum_{\substack{j=1\\ j=even}}^{p} \sum_{\substack{s=1\\ s=odd}}^{p-j} \sum_{i=0}^{j} \binom{p}{j}\binom{p-j}{s}\binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j-s} \\
&\times \left(2X_{t_k} g(X_{t_k})\Delta W_{t_k}\right)^{s} \left(g(X_{t_k})^2(\Delta W_{t_k})^2\right)^{j-i} (2\alpha_{-1}\Delta t)^i .
\end{aligned}
$$

The rest of the proof is similar to the proof of Theorem 4.1. For every sufficiently large integer n, we use the stopping time λ_n in (4.1). Noting that

$$
e^{t_{k+1}} X_{t_{k+1}}^p - e^{t_k} X_{t_k}^p \le e^{t_k}\left[X_{t_{k+1}}^p - X_{t_k}^p\right] + 2e^{t_k}\Delta t X_{t_{k+1}}^p,
$$

we then have, for any nonnegative integer N,

$$
\begin{aligned}
&\mathbb{E}e^{t_{N+1}} \left|X_{t_{N\wedge\lambda_n}+1}\right|^{2p} \\
&\quad \le K_1 + \sum_{k=1}^{N+1} \mathbb{E}e^{t_{k-1}} \left(\mathbf{1}_{[0,\lambda_n+1]} p \left(\left|X_{t_k}\right|^2\right)^{p-1} 2X_{t_k} f(X_{t_k})\Delta t + 2X_{t_k}^p\right) \\
&\qquad + \sum_{k=1}^{N+1} \mathbb{E}e^{t_{k-1}} \Bigg(\mathbf{1}_{[0,\lambda_n+1]} \sum_{s=2}^{p} (-1)^{s+1}\binom{p}{s} \left(\left|X_{t_{k+1}}\right|^2\right)^{p-s} \\
&\qquad \times \left(2X_{t_{k+1}} f(X_{t_{k+1}})\right)^s \Delta t^s\Bigg) \\
&\qquad + \sum_{k=1}^{N+1} \mathbb{E}e^{t_{k-1}} \Bigg(\mathbf{1}_{[0,\lambda_n+1]} \sum_{j=1}^{p}\sum_{s=0}^{p-j} (-1)^{s+1}\binom{p}{j}\binom{p-j}{s} \left(\left|X_{t_{k+1}}\right|^2\right)^{p-j-s} \\
&\qquad \times \left(2X_{t_{k+1}} f(X_{t_{k+1}})\right)^s (2\alpha_{-1})^j \Delta t^{j+s}\Bigg) \\
&\qquad + \sum_{k=1}^{N} \mathbb{E}e^{t_k} \left(\mathbf{1}_{[0,\lambda_n]} \sum_{s=1}^{p/2} \binom{p}{2s} \left(\left|X_{t_k}\right|^2\right)^{p-2s} \left(2X_{t_k} g(X_{t_k})\Delta W_{t_k}\right)^{2s}\right) \\
&\qquad + \sum_{k=1}^{N} \mathbb{E}e^{t_k} \Bigg(\mathbf{1}_{[0,\lambda_n]} \sum_{j=1}^{p}\sum_{i=0}^{j} \binom{p}{j}\binom{j}{i} \left(\left|X_{t_k}\right|^2\right)^{p-j} \left[g(X_{t_k})^2\right]^{j-i} [(\Delta W_{t_k})^2]^{j-i}
\end{aligned}
$$

$$\times (2\alpha_{-1}\Delta t)^i \Bigg)$$

$$+\sum_{k=1}^{N} \mathbb{E} e^{t_k} (\mathbf{1}_{[0,\lambda_n]} \sum_{j=1}^{p} \sum_{s=1}^{(p-j)/2} \sum_{i=0}^{j} \binom{p}{j} \binom{p-j}{2s} \binom{j}{i} \left(|X_{t_k}|^2\right)^{p-j-2s}$$

$$\times \left(2X_{t_k} g(X_{t_k})\right)^{2s} \left(g(X_{t_k})^2 (\Delta W_{t_k})^2\right)^{j-i} \left(\Delta W_{t_k}\right)^{2s} (2\alpha_{-1}\Delta t)^i ,$$

where $K_1 = K_1(X_{t_0})$ is a constant. To this end, for any given z , $1 < z < p$, we need to find the highest power q_1 of $-X_{t_k}^{q_1} \Delta t^z$, from the drift terms (not involving ΔW_{t_k}). Note that the highest power among the drift terms arises when $j = 0$. Next we need to obtain the highest power q_2 of $X_{t_k}^{q_2} \Delta t^z$ from the diffusion terms (involving ΔW_{t_k}). Notice that we get the highest power of the diffusion terms when $i = 0$. We can then easily verify that the highest power q_1 of $-X_{t_k}^{q_1} \Delta t^z$ is $q_1 = 2(p - z) + z(1 + j)$. To find the highest power q_2 of $X_{t_k}^{q_2} \Delta t^z$, we note $z = s + j$ and consider three possible cases:

Case 1 $s = z$, $j = 0$. Because $q_2 = 2(p - z) + 2z\rho$, by Assumption 2.2, $q_1 > q_2$.
Case 2 $s = 0$, $j = z$. Because $q_2 = 2(p - z) + 2z\rho$, by Assumption 2.2, $q_1 > q_2$.
Case 3 $1 \le s, j \le z - 1$. By Assumption 2.2, we still have $q_1 > q_2$.

We can therefore bound the right-hand-side terms of the inequality to obtain

$$\mathbb{E} e^{t_{N+1}} \left|X_{t_{N\wedge\lambda_n}+1}\right|^{2p} \le K_1 + K \sum_{k=1}^{N+1} e^{t_{k-1}} \Delta t,$$

which implies the assertion. □

5 Backward-Forward Euler-Maruyama scheme

For the purpose of analysis we find it convenient to introduce another numerical method. In terms of the general drift and diffusion coefficients (2.1), once we compute the value X_{t_k} from BEuM, that is

$$X_{t_k} = X_{t_{k-1}} + f(X_{t_k})\Delta t + g(X_{t_{k-1}})\Delta W_{t_{k-1}},$$

we define the Backward-Forward Euler-Maruyama (BFEuM) scheme as follows

$$\hat{X}_{t_{k+1}} = \hat{X}_{t_k} + f(X_{t_k})\Delta t + g(X_{t_k})\Delta W_{t_k},$$

where $\hat{X}_{t_0} = X_{t_0} = x(t_0)$. The two methods are, of course, closely related. Our approach will be to establish strong convergence of BEuM by first showing strong convergence of BFEuM. In this section we derive some basic results for BFEuM.

Theorem 5.1 *Under Assumption 2.2, for any $p > 2$, there is a $\Delta t^* \in (0, 1/2\alpha_1)$ such that*

$$\sup_{\Delta t \le \Delta t^*} \sup_{k \ge 0} \mathbb{E} \left| \hat{X}_{t_k} \right|^p < \infty.$$

Proof Let N be any nonnegative integer. Summing backward-forward and backward schemes we have

$$\hat{X}_{t_N+1} = \hat{X}_{t_0} + \sum_{k=0}^{N} f(X_{t_k})\Delta t + \sum_{k=0}^{N} g(X_{t_k})\Delta W_{t_k},$$

$$X_{t_N} = X_{t_0} + \sum_{k=1}^{N} f(X_{t_k})\Delta t + \sum_{k=0}^{N-1} g(X_{t_k})\Delta W_{t_k}.$$

By the Hölder inequality

$$\left| \hat{X}_{t_N+1} \right|^p \le 2^{p-1} \left(\left| \hat{X}_{t_N+1} - X_{t_N} \right|^p + \left| X_{t_N} \right|^p \right), \tag{5.1}$$

where

$$\left| \hat{X}_{t_N+1} - X_{t_N} \right|^p = \left| f(X_{t_0})\Delta t + g(X_{t_N})\Delta W_{t_N} \right|^p. \tag{5.2}$$

By Theorem 4.2, taking the expectation of both sides of inequality (5.1), we get

$$\mathbb{E} \left| \hat{X}_{t_N+1} \right|^p \le K(1 + \Delta t^{\frac{p}{2}}),$$

as required. □

5.1 Continuous BFEuM

Let us now define our continuous version of the BFEuM. We introduce the following notation

$$\eta(t) := t_k, \quad \text{for } t \in [t_k, t_{k+1}), k \ge 0.$$

The continuous version of the BFEuM is then given by

$$\hat{X}(t) = \hat{X}_{t_0} + \int_0^t f(X_{\eta(s)})ds + \int_0^t g(X_{\eta(s)})dw(s), \quad t \ge 0. \tag{5.3}$$

Note that the continuous and discrete BFEuM coincide at the gridpoints; that is, $\hat{X}(t_k) = \hat{X}_{t_k}$. Having bounded the moments for the discrete BFEuM, we can bound the continuous BFEuM in the following sense.

Lemma 5.1 *Under Assumption 2.2, there is a $\Delta t^* > 0$ such that for any $p \ge 2$,*

$$\sup_{\Delta t \le \Delta t^*} \mathbb{E} \left[\int_0^t \frac{1}{X_{\eta(s)}} ds \right]^p < \infty, \quad \forall t > 0.$$

Proof We only need to prove the lemma for $t = t_N$ for any $N \geq 1$. It follows from (5.3) that

$$\alpha_{-1} \int_0^{t_N} \frac{1}{X_{\eta(s)}} ds = \hat{X}_{t_N} - X_{t_0} + \alpha_0 t_N - \alpha_1 \int_0^{t_N} X_{\eta(s)} ds + \alpha_2 \int_0^t X^r_{\eta(s)} ds - \sigma \int_0^{t_N} X^\rho_{\eta(s)} ds.$$

It is then straightforward to show the assertion by Theorem 4.2. □

Lemma 5.2 *Under Assumption 2.2, there is a $\Delta t^* > 0$ such that for any $p \geq 2$,*

$$\sup_{\Delta t \leq \Delta t^*} \mathbb{E}\Big(\sup_{0 \leq t \leq T} \left|\hat{X}(t)\right|^p \Big) < \infty, \quad \forall T > 0.$$

Proof It follows from (5.3) that

$$\mathbb{E}\Big(\sup_{0 \leq t \leq T} \left|\hat{X}(t)\right|^p \Big) \leq 3^{p-1} \mathbb{E}\left(\hat{X}(0)^p\right) + \mathbb{E}\left(\int_0^T (\alpha_{-1} X^{-1}_{\eta(s)} + \alpha_0 + \alpha_1 X_{\eta(s)} + \alpha_2 X^r_{\eta(s)}) ds \right)^p + \mathbb{E}\left(\sup_{0 \leq t \leq T} \left| \int_0^t \sigma X^\rho_{\eta(s)} dw(s) \right|^p \right).$$

This, together with Theorem 4.2 and Lemma 5.1, implies the assertion. □

Theorem 5.2 *Let Assumption 2.2 hold and $T > 0$ be fixed. Then, for any given $\epsilon > 0$, there exists an N_0 such that for every $n \geq N_0$, we can find a $\Delta t_0 = \Delta t_0(n)$ so that whenever $\Delta t \leq \Delta t_0$,*

$$\mathbb{P}(\vartheta_n < T) \leq \epsilon,$$

where $\vartheta_n = \inf\{t > 0 : \hat{X}(t) \notin (\frac{1}{n}, n) \text{ or } X_{\eta(s)} \notin (\frac{1}{n}, n)\}$.

Proof Let $s \in [0, T \wedge \vartheta_n)$. Then for the function $V(x)$ defined by (2.2) we have

$$V_x(\hat{X}(s)) \left(f(X_{\eta(s)}) - f(\hat{X}(s))\right) + V_{xx}(\hat{X}(s)) \left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s))\right) \leq L(n) \left|X_{\eta(s)} - \hat{X}(s)\right|,$$

where $L(n)$ are (local Lipschitz) constant. By the Itô formula, we can show that

$$dV(\hat{X}(s)) = \Big[LV(\hat{X}(s)) + V_x(\hat{X}(s)) \left(f(X_{\eta(s)}) - f(\hat{X}(s))\right) + \frac{1}{2} V_{xx}(\hat{X}(s)) \left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s))\right) \Big] ds + V_x(\hat{X}(s)) g(X_{\eta(s)}) dw(s),$$

where LV has been defined in the proof of Theorem 2.1. Recalling (2.3), we then have

$$
\begin{aligned}
&\mathbb{E}V(\hat{X}(T\wedge\vartheta_n)) \\
&\quad\le V(\hat{X}(0)) + KT + \mathbb{E}\int_0^{T\wedge\vartheta_n} V_x(\hat{X}(s))\left(f(X_{\eta(s)}) - f(\hat{X}(s))\right)ds \\
&\qquad + \mathbb{E}\int_0^{T\wedge\vartheta_n} V_{xx}(\hat{X}(s))\left(g^2(X_{\eta(s)}) - g^2(\hat{X}(s))\right)ds \\
&\quad\le V(\hat{X}(0)) + KT + L(n)\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|X_{\eta(s)} - \hat{X}(s)\right|ds \\
&\quad\le V(\hat{X}(0)) + KT + L(n)\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|\hat{X}_{\eta(s)+1} - \hat{X}(s)\right|ds \\
&\qquad + L(n)\int_0^{T}\mathbb{E}\left|X_{\eta(s)} - \hat{X}_{\eta(s)+1}\right|ds.
\end{aligned}
$$

By (5.2)

$$
\mathbb{E}\left|X_{\eta(s)} - \hat{X}_{\eta(s)+1}\right| < K\Delta t^{\frac{1}{2}}. \tag{5.4}
$$

To bound the term $\mathbb{E}\int_0^{T\wedge\vartheta_n}|\hat{X}_{\eta(s)+1} - \hat{X}(s)|ds$, given $s\in[0, T\wedge\vartheta_n)$, let k be an integer for which $s\in[t_k, t_{k+1})$. Then

$$
\left|\hat{X}_{\eta(s)+1} - \hat{X}(s)\right| = \int_s^{t_{k+1}} f(X_{t_k})ds + \int_s^{t_{k+1}} g(X_{t_k})dw(s).
$$

By Hölder's inequality

$$
\mathbb{E}\int_0^{T\wedge\vartheta_n}\left|\hat{X}_{\eta(s)+1} - \hat{X}(s)\right|ds \le C(n,T)\Delta t^{\frac{1}{2}}, \tag{5.5}
$$

where $C(n,T) > 0$ is constant. Therefore

$$
\mathbb{E}V(\hat{X}(t\wedge\vartheta_n)) \le V(\hat{X}(0)) + KT + (L(n))(K + C(T,n))\Delta t^{\frac{1}{2}},
$$

which implies that

$$
\mathbb{P}(\vartheta_n < T) \le \frac{V(\hat{X}(0)) + KT + (L(n))(K + C(T,n))\Delta t^{\frac{1}{2}}}{V(1/n)\wedge V(n)}.
$$

Now for any given $\epsilon > 0$ we choose N_0 such that for any $n \ge N_0$

$$
\frac{V(\hat{X}(0)) + KT}{V(1/n)\wedge V(n)} \le \frac{\epsilon}{2}.
$$

Then we can choose $\Delta t_0 = \Delta t_0(n)$, such that for any $\Delta t \leq \Delta t_0$

$$\frac{(L(n))(K + C(T,n))\Delta t^{\frac{1}{2}}}{V(1/n) \wedge V(n)} \leq \frac{\epsilon}{2},$$

whence $\mathbb{P}(\vartheta_n < T) \leq \epsilon$ as required. □

6 Convergence of BFEuM scheme

In this section, we use the previous results to establish the main conclusion of this work: strong convergence of BEuM in Theorem 6.2, via strong convergence of BFEuM in Theorem 6.1. First, we define the stopping time θ_n

$$\theta_n = \tau_n \wedge \vartheta_n,$$

and prove a final lemma.

Lemma 6.1 *For any $p \geq 2$, $T > 0$ and sufficiently large n, there exists a constant $C = C(p, T, n)$, such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t \wedge \theta_n) - x(t \wedge \theta_n)\right|^p\right] \leq C\Delta t^{\frac{p}{2}}.$$

Proof For any $T_1 \in [0, T]$, by the Hölder and BDG inequalities

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq t\leq T_1}\left|\hat{X}(t \wedge \theta_n) - x(t \wedge \theta_n)\right|^p\right] \\
&\leq 2^{p-1}\Bigg(T^{p-1}\mathbb{E}\int_0^{T_1\wedge\theta_n}\left[f(X(\eta(s))) - f(x(s))\right]^p ds \\
&\quad + C(p)\mathbb{E}\int_0^{T_1\wedge\theta_n}\left[g(X(\eta(s))) - g(x(s))\right]^p ds\Bigg),
\end{aligned}$$

where $C(p)$ is a constant. Let $s \in [0, T_1 \wedge \vartheta_n)$. Then local Lipschitz conditions on f and g imply that

$$\left|f(X_{\eta(s)}) - f(x(s))\right|^p + \left|g(X_{\eta(s)}) - g(x(s))\right|^p \leq L_1(n)\left|X_{\eta(s)} - x(s)\right|^p$$

and

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq t\leq T_1}\left|\hat{X}(t \wedge \theta_n) - x(t \wedge \theta_n)\right|^p\right] \\
&\leq 2^{p-1}L_1(n)(T^{p-1} + C(p))\mathbb{E}\int_0^{T_1\wedge\theta_n}\left|X_{\eta(s)} - x(s)\right|^p ds
\end{aligned}$$

$$\leq 6^{p-1} L_1(n)(T^{p-1} + C(p))\mathbb{E}\int_0^{T_1\wedge\theta_n} \left|\hat{X}(s) - x(s)\right|^p + \left|\hat{X}_{\eta(s)+1} - \hat{X}(s)\right|^p + \left|X_{\eta(s)} - \hat{X}_{\eta(s)+1}\right|^p ds.$$

Now by (5.4) and (5.5)

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t\wedge\theta_n) - x(t\wedge\theta_n)\right|^p\right] \leq 6^{(p-1)} L_1(n)(T^{p-1} + C(p))\Bigg[(K + C(T,n))\Delta t^{\frac{p}{2}} + \int_0^{T_1}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t\wedge\theta_n) - x(t\wedge\theta_n)\right|^p\right]ds\Bigg].$$

By Gronwall's inequality

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t\wedge\theta_n) - x(t\wedge\theta_n)\right|^p\right] \leq C_2\Delta t^{\frac{p}{2}} e^{C_1},$$

where $C_1 = 6^{(p-1)} L_1(n)(T^{p-1} + C(p))$ and $C_2 = C_1(K + C(T,n))$. □

Theorem 6.1 *Under Assumption* (2.2), *we have*

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\hat{X}(t) - x(t)\right|^p\right] = 0.$$

Proof Let

$$e(t) = \hat{X}(t) - x(t).$$

Using the notation of Lemma 6.1 and applying Young's inequality

$$ab \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2, \quad \forall a, b, \delta > 0,$$

we have

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p\right] = \mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p \mathbf{1}_{\{\tau_n > T, \vartheta_n > T\}}\right] + \mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p \mathbf{1}_{\{\tau_n\leq T \text{ or } \vartheta_n\leq T\}}\right]$$
$$\leq \mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^p \mathbf{1}_{\{\theta_n > T\}}\right] + \frac{\delta}{2}\mathbb{E}\left[\sup_{0\leq t\leq T}|e(t)|^{2p}\right] + \frac{1}{2\delta}\mathbb{P}(\tau_n\leq T \text{ or } \vartheta_n\leq T).$$

To finish the proof we need to estimate the expressions on the right hand side of this inequality.

By Hölder's inequality and Lemmas 5.2 and 2.2, we choose δ such that

$$\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^{2p}\right]\le 2^{2p-1}\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}|x(t)|^{2p}+\sup_{0\le t\le T}\left|\hat{X}(t)\right|^{2p}\right]\le\frac{\epsilon}{3}$$

Next, by Theorem 2.1 there exists N_0 such that for $n\ge N_0$

$$\frac{1}{2\delta}\mathbb{P}(\tau_n\le T)\le\frac{\epsilon}{3},$$

and finally by Theorem 5.2 and Lemma 6.1 we may choose Δt sufficiently small, such that

$$\mathbb{E}\left[\sup_{0\le t\le T}|e(t)|^p\,\mathbf{1}_{\{\theta_n>T\}}\right]+\mathbb{P}(\vartheta_n\le T)\le\frac{\epsilon}{3}. \qquad \square$$

Now we are ready to formulate the main theorem of this paper. We will show that the Backward-Euler scheme (3.1) strongly converges to the solution of (1.1).

Theorem 6.2 *Under Assumption* (2.2), *we have*

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-x(\eta(t))\right|^p\right]=0.$$

Proof By Hölder's inequality

$$\begin{aligned}&\mathbb{E}\left|X_{\eta(t)}-x(\eta(t))\right|^p\\ &\quad\le 3^{p-1}\left[\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-\hat{X}_{\eta(t)+1}\right|\right]^p+\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)}-x(\eta(t))\right|\right]^p\right.\\ &\qquad\left.+\,\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\right]\right].\end{aligned}$$

Now from (5.2) and Theorem 4.2

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|X_{\eta(t)}-\hat{X}_{\eta(t)+1}\right|\right]^p\le K\,\Delta t^{\frac{p}{2}}.$$

By Theorem 6.1

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)}-x(\eta(t))\right|^p\right]=0.$$

To finish the proof it is enough to show that

$$\lim_{\Delta t\to 0}\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\right]=0.$$

By analogy to the proof of Theorem 6.1, we have

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\right]&=\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\mathbf{1}_{\{\vartheta_n>T\}}\right]\\&\quad+\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\mathbf{1}_{\{\vartheta_n\le T\}}\right]\\&\le\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^p\mathbf{1}_{\{\vartheta_n>T\}}\right]\\&\quad+\frac{\delta}{2}\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{X}_{\eta(t)+1}-\hat{X}_{\eta(t)}\right|^{2p}\right]\\&\quad+\frac{1}{2\delta}\mathbb{P}(\vartheta_n\le T).\end{aligned}$$

By Lemma 5.2 and Theorem 5.2 it is straightforward to finish the proof. □

7 Corollary on option valuation

Theorem 6.2 is relevant in any context where the SDE (1.1) is to be simulated numerically. For example, sample paths may be needed within a model calibration exercise. Furthermore, the SDE may represent an asset on which an option is to be valued. It is shown in [12, 13] that for many path dependent options, strong convergence of the SDE asset simulation guarantees convergent Monte Carlo simulations for the option value. For example, an up-and-out call gives a European payoff if the asset never exceeds the fixed barrier, B, where $B>K$ and K is the exercise price; otherwise it pays zero. The payoff at the expiry date T thus has the form

$$P=\mathbb{E}\left[(x(T)-K)^+\mathbf{1}_{\{\sup_{0\le t\le T}x(t)<B\}}\right].$$

Accordingly, we may define the approximate payoff based on the numerical method (3.1)

$$P_{\Delta t}=\mathbb{E}\left[(X_{\eta(T)}-K)^+\mathbf{1}_{\{\sup_{0\le t\le T}X_{\eta(t)}<B\}}\right].$$

It then follows from Theorem 6.2 that

$$\lim_{\Delta t\to 0}|P-P_{\Delta t}|=0.$$

We refer to [12, 13] for details and further examples.

References

1. Ait-Sahalia, Y.: Testing continuous-time models of the spot interest rate. Rev. Financ. Stud. **9**(2), 385–426 (1996)
2. Alfonsi, A.: On the discretization schemes for the CIR (and Bessel squared) processes. Monte Carlo Methods Appl. **11**(4), 355–384 (2005)
3. Ang, A., Bekaert, G.: Short rate nonlinearities and regime switches. J. Econ. Dyn. Control **26**, 1243–1274 (2002)
4. Berkaoui, A., Bossy, M., Diop, A.: Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. ESAIM **12**, 1–11 (2008)
5. Bossy, M., Diop, A.: An efficient discretisation scheme for one dimensional SDEs with a diffusion coefficient function of the form x^{α}, $\alpha \in [1/2, 1)$. INRIA Rapport de recherche (2007)
6. Chan, K.C., Karolyi, A.G., Longstaff, F.A., Sanders, A.B.: An empirical comparison of alternative models of the short-term interest rates. J. Finance **XLVII**(3) (1992)
7. Chapman, D.A., Pearson, N.D.: Is the short rate drift actually nonlinear? J. Finance **55**, 355–388 (2000)
8. Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A.: Short-term interest rates as subordinated diffusions. Rev. Financ. Stud. **10**, 525–578 (1997)
9. Friedman, A.: Stochastic Differential Equations and Applications, vol. 1–2. Academic Press, New York (1975)
10. Gallant, A.R., Tauchen, G.: Estimation of continuous-time models for stock returns and interest rates. In: Macroeconomic Dynamics. Cambridge University Press, Cambridge (1997)
11. Giles, M.B.: Multilevel Monte Carlo path simulation. Oper. Res. **56**, 607–617 (2008)
12. Giles, M.B., Higham, D.J., Mao, X.: Analyzing multi-level Monte Carlo for options with non-globally Lipschitz payoff. Finance Stoch. **13**, 403–413 (2009)
13. Higham, D.J., Mao, X.: Convergence of Monte Carlo simulations involving the mean-reverting square root process. J. Comput. Finance **8**(3), 35–62 (2005)
14. Higham, D.J., Mao, X., Stuart, A.M.: Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM J. Numer. Anal. **40**(3), 1041–1063 (2002)
15. Hong, Y., Haitao, L.: Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev. Finance Stud. **18**(1) (2005)
16. Kloeden, P.E., Neuenkirch, A.: The pathwise convergence of approximation schemes for stochastic differential equations. LMS J. Comput. Math. **10**, 235–253 (2007)
17. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations, 3rd edn. Springer, Berlin (1992)
18. Mao, X.: Stochastic Differential Equations with Applications, 2nd edn. Horwood Publishing, Chichester (2007)
19. Mao, X., Truman, A., Yuan, C.: Euler-Maruyama approximations in mean-reverting stochastic volatility model under regime-switching. J. Appl. Math. Stoch. Anal. **2006**, 80967 (2006)
20. Milstein, G.N., Platen, E., Schurz, H.: Balanced implicit methods for stiff stochastic systems. SIAM J. Numer. Anal. (1998)
21. Stanton, R.: A nonparametric model of term structure dynamics and the market price of interest rate risk. J. Finance **52**, 1973–2002 (1997)
22. Zeidler, E.: Nonlinear Monotone Operators. Springer, Berlin (1989)