A Method for Simulating Non-Linear Stochastic Differential Equations in $\mathbb{R}^1$

João Nicolau

Instituto Superior de Economia e Gestão
Universidade Técnica de Lisboa

February, 2004

Abstract

Very few specific stochastic differential equations have explicitly known solutions. In order to obtain a simulated path of a solution the most common procedure is based on a discretization of the stochastic differential equations. However, there are some cases where the discrete-time discretization cannot be used. In this article we propose a new method to simulate the solution of a non-linear stochastic differential equation which, in principle, is exempt from error of simulation and can be widely applied including in cases where the discrete-time discretization cannot be used.

Key Words

Diffusion Processes, Statistical Simulation Methods, Simulation-Based Methods, Estimation, Transition Density Estimation.

1E-mail: nicolau@iseg.utl.pt
1 Introduction

Let \( X = \{ X_t, t \geq 0 \} \) be a diffusion process in \( \mathbb{R}^1 \) governed by the Itô stochastic differential equation (SDE)

\[
dX_t = a(X_t) \, dt + b(X_t) \, dW_t, \quad X_0 = x
\]

where \( \{ W_t, t \geq 0 \} \) is a (standard) Wiener process, \( a \) and \( b \) are known coefficients (or functions) with continuous derivatives and \( x \) is either a constant value or a random value \( \mathcal{F}_0 \)-measurable independent of \( W_t \). In appendix 1 we briefly describe some assumptions of the process \( X \) and present some definitions that we will use throughout the paper (such as stationary density, stationary moments and stationary process).

Very few specific SDEs have explicitly known solutions (these being mainly reducible by appropriate transformation to the solution of linear SDEs). For these SDEs the marginal and especially the conditional functionals (e.g. the conditional probabilities) are generally unknown. In order to get insight into the properties of these kinds of processes one usually undertakes a large number of sample simulations of the process. This is done by using algorithms in which the SDE is discretized in time and approximations to its solution at the discretization instants are computed recursively. There is a large body of literature dealing with simulation of stochastic differential equations (see Milstein, 1995, Talay, 1995, Artemiev and Averina, 1997, Kloeden and Platen, 1999 and Schurz, 2002). For instance, to obtain a simulated path of a diffusion process governed by a SDE one considers a finite discretization of the time interval \([0, T]\), i.e. \( 0 = t_0 < t_1 < \ldots < t_n = T \), for example, using the rule \( t_i = \Delta i \), with a step size \( \Delta = T/n \), from which approximate values of the sample
path are generated step by step at the discretization times. There are several discrete-time approximations to the $X$ process at time $t_i$. The simplest and most widely used approximation is the Euler (also known as Euler-Maruyama) scheme which has the form

$$
\tilde{X}_{t_i} = \tilde{X}_{t_{i-1}} + a\left(\tilde{X}_{t_{i-1}}\right) \Delta + b\left(\tilde{X}_{t_{i-1}}\right) \Delta W_{t_i}
$$

for $i = 1, \ldots, n$ with initial value $\tilde{X}_{t_0} = x$, where $\Delta W_{t_i} = W_{t_i} - W_{t_{i-1}}$ are independent normally distributed random variables with zero mean and $\Delta$ variance. This scheme has also been used in the area of parametric estimation (see for example, Chan et al., 1992, Pedersen, 1995, Kloeden et al., 1996 and Gouriéroux and Monfort, 1996).

Naturally, it is important that the simulated paths be close in some sense to those of the Itô process. The approximation $\tilde{X}$ is said to be in the strong sense of order $\gamma$ if $E \left| X_T - \tilde{X}_{t_n} \right| \leq K \delta^\gamma$ for some $K > 0$, where $\delta$ is the maximum step size, $\delta = \max_{i=1,2,\ldots,n} (\Delta_i)$ (note that the approximation is evaluated at the terminal instant $t_n = T$). It is known that the Euler approximation (under some conditions on the infinitesimal coefficients, e.g., the coefficients $a$ and $b$ are Lipschitz continuous and satisfy a linear growth condition) is of strong order $\gamma = 1/2$ for diffusion processes (see Kloeden and Platen, 1999). More general discretization schemes are discussed in Milstein (1995), Talay (1995), Artemiev and Averina (1997), Kloeden and Platen (1999) and Schurz (2002). For example, if we include an extra term in (1) and consider

$$
\tilde{X}_{t_i} = \tilde{X}_{t_{i-1}} + a\left(\tilde{X}_{t_{i-1}}\right) \Delta + b\left(\tilde{X}_{t_{i-1}}\right) \Delta W_{t_i} + \frac{1}{2} b\left(\tilde{X}_{t_{i-1}}\right) b'\left(\tilde{X}_{t_{i-1}}\right) \left(\Delta W_{t_i}\right)^2 - \Delta
$$

then it can be shown that the approximation (denoted by Milstein scheme) is of order $\gamma = 1$. Further additional terms in (2) lead to better approximations.
All these methods based on the approximation values of the sample path, step by step at the discretization times, share the same characteristic: the quantity $E \left| X_T - \tilde{X}_T \right|$ is non null and only vanishes when $\Delta \to 0$ (or $\delta \to 0$).

However, the worst case for applying discretization schemes is when the solution of the discrete-time approximation diverges to $\pm \infty$ with probability one, even if the solution of the SDE has boundaries that are neither attracting, nor attainable (see appendix 1). This is related to stability which has been studied by Hofmann and Platen (1994), Mattingly et al. (2001) and Talay (2002), among others. Mattingly et al. (2001) and Talay (2002) show that the Euler scheme can fail to approximate ergodic processes. They present the example of the SDE $dX_t = -X^3_t dt + dW_t$. The process $X$ is $\beta$-mixing, $\rho$-mixing and geometric ergodic. Nevertheless, they show that the solution of the Euler scheme oscillates with increasing amplitude. Since the diffusion coefficient does not depend on $x$, we can confirm this result in the following way. Consider the Euler deterministic scheme, $y_i = y_{i-1} - y_{i-1}^3 \Delta$. Using the Liapunov function $V(x) = x^2$ it is easy to see that the basin of attraction of the zero equilibrium point is not $\mathbb{R}$. In fact it is $\{ x : x^2 < 2/\Delta \}$. Thus, the solution $y$ goes to $\pm \infty$ if the initial point is outside the set $\{ x : x^2 < 2/\Delta \}$. Mattingly et al. (2001) and Talay (2002) suggest the use of the split-step Backward Euler scheme.

The Euler scheme can also fail to approximate the solution of some processes with volatility induced stationarity. These processes have recently received attention (see Bibby and Sørensen, 1997, Conley et al., 1997 and Richter, 2002). They have the interesting property that the stationarity is ensured solely by the structure of the diffusion coefficient. Typically, the diffusion coefficient dominates the drift at infinity.
(i.e. \( \lim_{x \to \pm \infty} \frac{|a(x)|}{|b(x)|} = 0 \)). In connection with these kind of processes Richter (2002) has established that under some mild regularity conditions, if

\[
\lim_{x \to \pm \infty} \frac{|x|^p |a(x)|}{|b(x)|} = 0 \quad \text{and} \quad \lim_{x \to \pm \infty} \inf \frac{|a(x)|}{|x|} \geq \varphi
\]

then \( P \left\{ \bigcap_{n \geq 0} \left| \tilde{X}_{t_i} \right| > (1 + \tilde{\varphi}) \left| \tilde{X}_{t_{i-1}} \right| \right\} > 0 \) where \( 0 < \tilde{\varphi} < \varphi \) (see appendix 3). That is, under these conditions, the solution of the Euler scheme oscillates with increasing amplitude and therefore this scheme cannot be used. A particular case of a process with volatility induced stationarity is when the drift is zero (i.e. \( a(x) \equiv 0 \)) (see Bibby and Sørensen, 1997). These are local martingale processes. It can be shown that the Euler approximation of a local martingale process that admits a stationary distribution \( \bar{p} \) and verifies the condition \( \int |x|^p \bar{p}(x) \, dx < \infty \) for some \( p > 1 \), goes to \( \pm \infty \) with probability one as time goes to infinity (see appendix 3). That is, the Euler scheme simply does not work when \( X \) is a stationary local martingale admitting stationary moments of order higher than one. For example, the solution of the SDE

\[
\text{d}X_t = (1 + X_t^2) \, dW_t
\]

is stationary and has finite (stationary) variance. It can be proved that the correspondent Euler approximations, with a fixed step size \( \Delta \), go to \( \pm \infty \) with probability one as time goes to infinity. We study this case in the examples section.

In this article we propose a new method to simulate the \( X \) process which, in principle, is exempt from error of simulation, in the sense that \( \sup_{t=1, \ldots, n} E \left| X_{t_i} - \tilde{X}_{t_i} \right| = 0 \). This method also has the important advantage over the Euler scheme in that it can be applied to those cases where the solution of the SDE cannot be approximated through the Euler scheme as we point out in this section. Our procedure is based on a result of Dacunha-Castelle and Florens-Zmirou (1986) and Nicolau (2002) concerning the
conditional density of a non-linear diffusion process.

The rest of the paper is organized as follows. In Section 2 we discuss the proposed method for simulating a non-linear SDE and in Section 3 we illustrate the proposed method with two examples.

2 Simulating non-linear SDE through conditional density probability estimates

Our goal is to simulate a path $X_t$ from the SDE $dX_t = a(X_t)\, dt + \sigma dW_t$, $X_0 = x$ (the more general case $dX_t = a(X_t)\, dt + b(X_t)\, dW_t$ is discussed below) with a step of discretization of $\Delta$, i.e. we want simulate the sequence $\{X_0, X_{\Delta}, X_{2\Delta}, \ldots, X_{n\Delta}\}$.

To simulate $X_\Delta$ given the initial value $x_0$ we consider the following classic result. Let $X_\Delta$ be a random variable with conditional distribution function $F(\Delta, x_0, x) = P(X_\Delta \leq x \mid X_0 = x_0) = \int_{-\infty}^{x} p(\Delta, x_0, z)\, dz$ ($p$ is the conditional density). If $u$ has uniform distribution on $(0, 1)$ then $X = F^{-1}(\Delta, x_0, u)$ where $F^{-1}(\Delta, x_0, \zeta) = \inf\{x : F(\Delta, x_0, x) \geq \zeta\}$ has conditional distribution function $F(\Delta, x_0, x)$. Thus, a simulated value $X_\Delta$ given $x_0$ is simply the value $z$ such that $F(\Delta, x_0, z) = u_0$ where $u_0$ is the simulated value from the uniform distribution on $(0, 1)$. The next simulated value $X_{2\Delta}$ is obtained using $X_\Delta$ as an initial value. The whole trajectory can then be simulated in this way. For non-linear SDEs the expressions $F(\Delta, x_0, x)$ and $p(\Delta, x_0, x)$ are generally unknown. To overcome this difficulty we consider a result from Dacunha-Castelle and Florens-Zmirou (1986). These authors have proved that the transition (or conditional) density $p(\Delta, x, y)$ associated with the SDE $dX_t = a(X_t)\, dt + \sigma dW_t$, can be written in the following way:
\[ p(\Delta, x, y) = \frac{1}{\sqrt{2\pi\sigma^2\Delta}} \exp\left\{ \frac{(x - y)^2}{2\sigma^2\Delta} + \frac{G(y) - G(x)}{\sigma^2} \right\} E[\psi] \]  

where,

\[ \psi = \exp\left\{ \sigma^2\Delta \int_0^1 g(z_s(x, y) + \sqrt{\sigma^2\Delta}B_s) \, ds \right\}, \]

\[ G(y) = \int_0^y a(s) \, ds, \]

\[ z_s(x, y) = x(1 - s) + ys, \]

\[ B_t = W_t - tW_1, \quad 0 \leq t \leq 1, \quad \text{(Brownian bridge)} \]

\[ g(x) = -\frac{1}{2} \left( \frac{a(x)^2}{\sigma^4} + \frac{da(x)}{dx} \frac{1}{\sigma^2} \right) \]

\[ g \text{ being continuous and such that } |g(x)| = O(x^2). \quad \text{Dacunha-Castelle and Florens-Zmirou (1986) give (somewhat restrictive) conditions under which } E[\psi] \text{ can be written as (for simplicity we take } \sigma = 1), \]

\[ E[\psi] = E \left[ \sum_{i=0}^k \frac{\Delta^i}{i!} \left( \sum_{j=0}^{2k} \frac{\Delta^{j/2}}{j!} [B^j g^{(j)}] \right)^i \right] + R \]

where \( R \) is a “small” quantity depending on \( k \) and \( \Delta, \) \( g^{(j)} \) is the derivative \( d^j g(x)/dx^j, \) \( g \) is given in (4) and \( [B^j g^{(j)}] = \int_0^1 g^{(j)}(z_s(x, y)) B_s^j \, ds. \) Nicolau (2002) proposes a simulated maximum likelihood method based on equation (3), where the quantity \( E[\psi] \) is evaluated through Monte Carlo experiments:

\[ \hat{\psi} = \frac{1}{S} \sum_{j=1}^S \psi_N(\omega_j) \]

where

\[ \psi_N(\omega) = \exp\left\{ \frac{\sigma^2\Delta}{N} \sum_{i=0}^{N-1} g(z_t(x, y) + \sqrt{\sigma^2\Delta}B_t(\omega)) \right\} \]

with \( t_i = i/N \ (i = 0, 1, ..., N) \) and \( \{\omega_j; j \geq 1\} \) are i.i.d..
Let
\[ \hat{p}(\Delta, x, y) = \frac{1}{\sqrt{2\pi \sigma^2 \Delta}} \exp \left\{ \frac{-(x - y)^2}{2\sigma^2 \Delta} + \frac{G(y) - G(x)}{\sigma^2} \right\} \hat{\psi}. \] (7)

We now present some extensions of Nicolau (2002).

**Proposition 1** Under the result (3), with \( g \) continuously differentiable in \( I \), where \( I \) is the state space of the \( X \) process, \( \hat{p}(\Delta, x, y) \) converges uniformly in probability on

\[ \Xi = \{(x, y) : x, y \in I, p(\Delta, x, y) \text{ is finite}\} \subset \mathbb{R}^2 \] to \( p(\Delta, x, y) \) when \( N \to +\infty \) and \( S \to +\infty \).

**Proof.** See appendix 2.

The \( \hat{p}(\Delta, x, y) \) estimator is highly computational efficient and it converges rapidly to the \( p(\Delta, x, y) \) estimator (see Nicolau, 2002).

The assumption in the previous proposition, that \( g \) is continuously differentiable, is not necessary for pointwise convergence in probability (and even with probability one). In fact, the assumption that \( g \) is continuous is sufficient for pointwise convergence. However, we note that uniform convergence in probability is more interesting in our case since we are dealing with functional convergence. We require

\[ \hat{p}(\Delta, x, y) \xrightarrow{p} p(\Delta, x, y) \] for each \( y \) of an uncountable set of points in some interval of the state space.

We need now to study the estimator

\[ \hat{F}(\Delta, x_0, x) = \sum_{y_i \leq x} \hat{p}(\Delta, x_0, y_i) \Delta y_i \] (8)

for \( F(\Delta, x_0, x) = \int_{-\infty}^{x} p(\Delta, x_0, y) \, dy \), which forms the basis for the simulated proposed method.
Proposition 2 Under the conditions of proposition 1, one has

\[
\sum_{i=1}^{M} \hat{p}(\Delta, x, y_i) \Delta y_i \xrightarrow{pr} \int_{y_0}^{y_1} p(\Delta, x, y) \, dy,
\]

uniformly on \( \Xi \) as \( M \to +\infty \), where \( y_i = y_0 + i \frac{y_1 - y_0}{M} \) and \( \Delta y_i = \frac{y_1 - y_0}{M} \).

Proof. See appendix 2.

Uniform convergence is the right criterion to use, since \( \hat{F}(\Delta, x_0, x) \) requires convergence at each \( x \) of an uncountable set of points in the interval \([y_0, y_1] \).

Proposition 2 deals with finite limits of integration. To estimate the integral \( \int_{-\infty}^{x} p(\Delta, x_0, y) \, dy \) with negative infinity lower limit, we simply truncate the lower limit point by a suitable constant \( y_0 \) such that \( \int_{y_0}^{x} p(\Delta, x_0, y) \, dy \approx \int_{-\infty}^{x} p(\Delta, x_0, y) \, dy \).

This procedure should not involve errors since \( \lim_{|y| \to +\infty} p(\Delta, x, y) = 0 \) (the probability mass concentrates mainly in the neighborhood of the conditional mean and the tails should tend quickly to zero as \( y \) moves away from the conditional mean).

Proposition 2 establishes convergence in terms of Riemann sums. Nevertheless, other approximations, such as the Simpson rule and the Gauss-Legendre quadrature can also be used.

We have the following procedure to simulate \( X \):

1. Set \( i = 0 \), set \( S \) and \( N \) (large enough)

2. Set \( i = i + 1 \)

3. Fix \( x = X_{(i-1)\Delta} \)

4. Simulate independently \( u \) with distribution uniform \( U(0, 1) \)
5. Determine \( z \) such that

\[
\hat{F}(\Delta, x, z) = u
\]  

(9)

where \( \hat{F}(\Delta, x, z) \) is defined in equation 8.

6. Set \( X_{i\Delta} = z \) and go to step 2.

Remarks:

(i) The sequence \( \{\omega_j; j \geq 1\} \) defined in equation (6) has to be kept fixed when evaluating \( \{p(\Delta, x, y_i); i \geq 0\} \);

(ii) If the diffusion coefficient depends on \( x \), i.e. if the process is governed by a more general SDE \( dX_t = a(X_t)dt + b(X_t)dW_t \) it is necessary to standardize the diffusion coefficient of \( X \), i.e. transforming \( X \) into \( Y \) defined as

\[
Y = U(X), \quad U(x) = \int^x b^{-1}(s)ds.
\]

Since \( b > 0 \) on \( \mathbb{R} \), the function \( U(x) \) is increasing and invertible. By applying Itô’s lemma, \( Y \) has unit diffusion coefficient

\[
dY_t = \alpha(Y_t) dt + dW_t
\]

where

\[
\alpha(y) = \frac{a(U^{-1}(y))}{b(U^{-1}(y))} - \frac{1}{2}b'(U^{-1}(y)).
\]

In the next section we provide an example where the diffusion coefficient depends on \( x \). Once a simulation of \( \{Y_t\} \) is available the simulation of the process \( X \) can be recovered using the formula \( X_t = U^{-1}(Y_t) \).
3 Examples

To illustrate the proposed method to simulate the solution of SDEs we consider two examples. In the first one we study the simulation problem of the Ornstein-Uhlenbeck process $X$,

$$dX_t = -X_t dt + dW_t, \quad X_0 = 0.$$ 

The solution of this SDE can be easily simulated without simulation errors as the conditional probabilities are known (equivalently, the simulations can also be carried out using the closed form of the solution). With this example we intend to show that the proposed method is highly computational efficient (in the sense that for moderate $N$, $S$ and $M$ values it gives almost the same results as the exact method, based on the true conditional probabilities). To verify this, we will show that the estimated conditional distribution function $\hat{F}(\Delta, x_0, x)$ (equation (8)) is, for some select values $x_0$, (almost) identical to the true conditional distribution function

$$F(\Delta, x_0, x) = \Phi \left( \frac{x - x_0 e^{-\Delta}}{\sqrt{\frac{1}{2} (1 - e^{-2\Delta})}} \right),$$ 

where $\Phi$ is the cumulative distribution function of the Normal distribution. To estimate $p(\Delta, x, y)$ we set $N = 100$ and $S = 20$ and to integrate the density in (8) we used the Gauss-Legendre quadrature.

It is also interesting to compare these two previous methods with the Euler approximation. Once again, we consider the conditional distribution function associated with the Euler scheme, which is

$$F_e(\Delta, x_0, x) = \Phi \left( \frac{x - (x_0 + x_0 \Delta)}{\sqrt{\Delta}} \right).$$ 

11
It is well known that the quality of the Euler scheme depends heavily on the magnitude of the step of discretization $\Delta$ (the Euler approximation is of strong order $\gamma = 1/2$).

We select $\Delta = 0.5$ which, as we will see, can be considered a high value for the Euler approximation of the SDE $dX_t = -X_t dt + dW_t$.

In figure 1 we observe that the differences between the proposed estimated conditional distribution function (8) and the true conditional distribution (10) are, for practical purposes, negligible across the interval $(-4, 4)$ (note that the distributions are conditioned on the values $x_0 = -1$, $x_0 = 0$ and $x_0 = 1$). We confirm, as in Nicolau (2002), that the proposed estimators for the conditional density function and the conditional distribution function are highly computational efficient.

However, substantial differences appear between the conditional distribution function associated with the Euler scheme and the true conditional distribution function (obviously, the high value of the discretization step $\Delta = 0.5$ is responsible for these differences).

**FIGURE 1**

In the second example we consider the SDE

$$dX_t = \left(1 + X_t^3\right) dW_t, \quad X_0 = 0.$$  

Since the conditional probabilities are unknown, the best method to simulate the solution (based on the conditional distribution) cannot be used. Despite the fact that $X$ is a stationary process, with stationary density $\bar{p}(x) = 2/\left(\pi \left(1 + x^2\right)^2\right)$ and
stationary moments $\int x\hat{p}(x)\,dx = 0$ and $\int x^2\hat{p}(x)\,dx = 1$, the Euler scheme (and other schemes, such as the Milstein and Platen) simply cannot be applied here. As we have pointed out, this is due to the fact that the solution $X_t = x + \int_0^t (1 + X_s^2)\,dW_s$ is a stationary local martingale with finite variance, which cannot be approximated by the Euler scheme (with $\Delta$ fixed and choose independently of $X$). Some simulations carried out by the author have shown that these convergence problems also occur with the Milstein and Platen schemes.

The method proposed in the previous section solves the problem of simulating $X$. According to remark (ii) in section 2, we first standardize the diffusion coefficient of $X$, by considering the transformation

$$U(x) = \int_x^x \frac{1}{1 + z^2}dz = \arctan x.$$  

By applying Itô’s lemma we have $dY_t = \alpha(Y_t)\,dt + dW_t$ where

$$\alpha(y) = \frac{a(U^{-1}(y))}{b(U^{-1}(y))} - \frac{1}{2}b'(U^{-1}(y)) = -\tan y, \quad -\frac{\pi}{2} < y < \frac{\pi}{2}. $$

Thus

$$G(x) = \int_x^x -\tan y = \log(\cos x), \quad -\frac{\pi}{2} < x < \frac{\pi}{2},$$

$$g(x) = -\frac{1}{2} \left( \frac{\alpha(x)^2}{\sigma^4} + d\alpha(x) \cdot \frac{1}{\sigma^2} \right) = \frac{1}{2}$$

and

$$p(\Delta, x, y) = \frac{1}{\sqrt{2\pi\Delta}} \exp \left\{ -\frac{(x-y)^2}{2\Delta} + \log(\cos y) - \log(\cos x) \right\} E[\psi].$$

We estimate $p(\Delta, x, y)$ using equation (7), setting $N = 100$ and $S = 10$ (we have found similar results for $N \in \{20, 21, \ldots\}$ and $S \in \{10, 11, \ldots\}$). We are now ready
to simulate the solution of the SDE $dY_t = -\tan Y_t \, dt + dW_t$ following steps 1 to 6 defined in section 2. Using the inverse function of $U$ we can obtain a simulation of $X$, i.e. $X_{i\Delta} = \tan(Y_{i\Delta})$, $i = 1, 2, ...$. We simulated 10,000 observations with a discretization step $\Delta = 0.05$ (note that $X_t$ is simulated in the interval $t \in [0, 500]$, since $10,000 \times 0.05 = 500$). The simulated path is presented in figure 2.

** FIGURE 2 **

We present a brief description on the behavior of the simulated process. Despite the zero drift (and thus the tendency of the process to behave like a random walk), the path exhibits reversion effects to zero, which are assured solely by the structure of the diffusion coefficient. It is the instantaneous volatility (or simply volatility), i.e. the value of the diffusion coefficient, that induces stationarity. In the neighborhood of the stationary mean the volatility is low so the process tends to spend more time in this interval. If there is a (large) random shock (i.e. large increment of the Brownian motion), the process moves away from the stationary mean (zero) and the volatility increases (since the diffusion is $1 + x^2$) which, in turn, increases the probability of the $X$ crossing zero again. The process can reach extreme peaks in a very short time but quickly return to the neighborhood of the stationary mean. We conjecture that volatility induced stationary processes can be applied to financial time series.

We now study the quality of the approximation. The question is to verify whether the simulated process really reproduces the probabilist structure of the $X$ process. In order to assess the approximation we consider two general measures. In the first
one we follow Talay (1995) who has proposed a criterion for assessing discrete-time approximation \( \tilde{X} \) (say). Let \( F \) be the functional

\[
F = \int_{\mathbb{R}} f(x) \tilde{p}(x) \, dx < \infty
\]  

(12)

where \( \tilde{p}(x) \) is the stationary density. According to Talay (1995), a discrete-time approximation based on a step of discretization of \( \Delta \), \( \tilde{X}^\Delta \), is said to converge with respect to the ergodic criterion with order \( \beta > 0 \) to an ergodic process \( X \) as \( \Delta \to 0 \) if for each function \( f \) satisfying (12) there exists a positive constant \( C_f \) which does not depend on \( \Delta \) such that

\[
|F^\Delta - F| \leq C_f \Delta^\beta
\]

where

\[
F^\Delta = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} f\left( \tilde{X}^\Delta_{t_i} \right), \quad t_i = t_0 + i\Delta.
\]

With slight modifications we can adapt this criterion to build a test statistic for assessing the proposed simulated process, which we denote by \( \hat{X} \). Firstly, we require that the differences \( F_n - F \), where

\[
F_n = \frac{1}{n} \sum_{i=1}^{n} f\left( \hat{X}_{t_i} \right),
\]

are not statistically significant, for various choices of \( f \). We can select several functions \( f \) satisfying (12) (in fact there are infinite functions \( f \) that can be selected). However, \( X \) does not possess moments higher than 3; in fact, \( \int |x|^\alpha \tilde{p}(x) \, dx < \infty \) only if \( \alpha < 3 \).
To assess the simulation we evaluate the hypothesis

\[
H_0: \begin{bmatrix}
E \left[ \hat{X} \right] &= \int_{\mathbb{R}} x \bar{p}(x) \, dx = 0 \\
E \left[ \hat{X}^2 \right] &= \int_{\mathbb{R}} x^2 \bar{p}(x) \, dx = 1 \\
E \left[ |\hat{X}| \right] &= \int_{\mathbb{R}} |x| \bar{p}(x) \, dx = \frac{2}{\pi} \\
E \left[ \frac{1}{1 + X^2} \right] &= \int_{\mathbb{R}} \frac{1}{1 + x^2} \bar{p}(x) \, dx = \frac{3}{\pi}
\end{bmatrix}
\]

which explores some measures of the distribution such as the location and scale. Let

\[
G_n = \sum_{i=1}^{n} h_i, \quad h_i = \begin{bmatrix}
\hat{X}_{i\Delta} \\
\hat{X}_{i\Delta}^2 \\
|\hat{X}_{i\Delta}| \\
\frac{1}{1 + \hat{X}_{i\Delta}^2}
\end{bmatrix}, \quad \theta = \begin{bmatrix}
0 \\
1 \\
\frac{2}{\pi} \\
\frac{3}{\pi}
\end{bmatrix}
\]

Under \( H_0 \) and assuming that the components of \( h_i \) are i.i.d., we have by the central limit theorem that

\[
\frac{1}{\sqrt{n}} (G_n - \theta) \xrightarrow{d} N(0, S), \quad S = \lim_{n \to +\infty} Var\left( \frac{1}{\sqrt{n}} G_n \right).
\]  

Equation (13) implies \( Q = \frac{1}{n} (G_n - \theta)' S^{-1} (G_n - \theta) \xrightarrow{d} \chi^2_4 \) (the \( S \) matrix can be consistently estimated through the estimator \( \hat{h} h' \) where \( h = \begin{bmatrix} h_1 & \ldots & h_n \end{bmatrix} \) is a \( 4 \times n \) matrix). Thus, \( Q \) is the test statistic for the null hypothesis. Let \( q \) be the observed statistics. We reject \( H_0 \) if \( P(Q > q) < 0.05 \).

Our simulated path, based on the step \( \Delta = 0.05 \), displays strong correlation. This is natural since \( \Delta \) is a small value and \( X \) is a local martingale (thus \( \hat{X} \) should behave like a random walk - the correspondent Ljung-Box p-value (10 lags) is zero). Since result (13) involves independence we break temporal dependence by using only the simulated values multiple of 40, that is, we consider, \( \{ \hat{X}_{d\iota}; \delta = 40\Delta; i = 1, \ldots, 250 \} \).
This procedure is always possible since (it can be shown) \( X \) is a \( \rho \)-mixing process (see, Chen et al., 1998). Now, the associated Ljung-Box p-value (10 lags) is higher than 0.35 (thus, the joint null hypothesis \( \rho_1 = \rho_2 = \ldots = \rho_{10} \) cannot be rejected at any standard levels). Considering \( \{ \hat{X}_{\delta i}, \delta = 40 \Delta; i = 1, \ldots, 250 \} \) we find \( Q = 2.03 \) and the associated p-value is higher than 0.84. Therefore the null hypothesis cannot be rejected at any standard levels and we conclude that the simulated value reproduces, to some extent, the structure of the ergodic moments.

Finally, we consider another measure to assess the simulations. We now want to verify if the (unconditional) probability density of the simulated path equals the stationary density \( \bar{p}(x) = \frac{2}{\pi \left(1 + x^2\right)^2} \). We consider the estimator

\[
\hat{p}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left( \frac{x - X_{ti}}{h_n} \right)
\]  

(14)

where \( K \) is a Gaussian kernel and \( h_n \) is a bandwidth. Under some conditions, as \( X \) is ergodic and \( h_n \to 0 \) and \( nh_n \to +\infty \) as \( n \to +\infty \), the estimator \( \hat{p}(x) \) converges in MSE to \( \bar{p}(x) \) (see for example Rao, 1983). Using \( \hat{X}_{ti} \) instead of \( X_{ti} \) in equation (14) (and all 10,000 observations) we can estimate the unconditional probability density of the simulated process. From figure (3) we conclude that the probabilist structure of the simulated path is almost indistinguishable from that of the \( X \) process. A Gaussian density is also shown for comparison purposes.

** FIGURE 3 **
References


Talay, D. (2002), Stochastic Hamiltonian dissipative systems with non globally Lipschitz coefficients: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme, To appear in Markov Processes and
Related Fields.
Appendix 1

We assume that $X = \{X_t, t \geq 0\}$ is a diffusion process in $\mathbb{R}^1$, with state space $I = (l, r)$, governed by the Itô stochastic differential equation

$$dX_t = a(X_t) \, dt + b(X_t) \, dW_t, \quad X_0 = x$$

where $\{W_t, t \geq 0\}$ is a (standard) Wiener process, $a$ and $b$ are unknown coefficients or functions with continuous derivatives and $x$ is either a constant value or a random value $\mathcal{F}_0$-measurable independent of $W_t$ (a standard treatise on stochastic differential equations is Gikhman and Skorokhod, 1972; see also Arnold, 1974 and Ikeda and Watanable, 1981).

Let $S[x, x_0] = \int_{x_0}^x s(u) \, du$ be the scale function where $s(z) = \exp \left\{ - \int_{z_0}^z 2a(u)/b^2(u) \, du \right\}$ ($x_0$ and $z_0$ are arbitrary fixed points inside $I$).

Our main assumption is $S(l, x) = S[x, r] = +\infty$ for $x \in I$.

According to Arnold (1974, page 114), if the infinitesimal coefficients $a$ and $b$ have continuous derivatives with respect to $x$, then there exists a unique continuous process that is defined up to a random explosion time $\eta$ in the interval $t_0 < \eta \leq +\infty$. Our assumption assures that the boundaries $l$ and $r$ are neither attracting, nor attainable (see Karlin and Taylor, 1981, chapter 15) and the process is recurrent, i.e. $P[T_y < \infty | X_0 = x] = 1$ for every $x, y \in I$ where $T_y = \inf \{t \geq 0, X_t = y\}$ (Ikeda and Watanable, 1981, Theorem 3.1, Chapter VI). We note that $P[T_\alpha < T_\beta | X_0 = x] = S[x, \beta]/S[\alpha, \beta]$ for $\alpha < x < \beta$. Thus, if $S(l, x) = +\infty$ then $\lim_{\alpha \to l} P[T_\alpha < T_\beta | X_0 = x] = 0$, i.e. the probability of the process reaching the boundary $l$ first rather than an arbitrary state $\beta$ is zero from any interior starting point $x < \beta$. In consequence, when both conditions $S(l, x) = S[x, r] = +\infty$ for $x \in I$
hold, it can be proved that $P[\eta = +\infty | X_0 = x] = 1$ (Ikeda and Watanable, 1981, pp. 362-363).

Roughly speaking, the boundaries $l$ and $r$ are never attained although every finite point can be reached with probability one in finite time. Global Lipschitz and growth conditions, which fail to be satisfied for many interesting models in economics and finance (Aït-Sahalia, 1996), are not needed in the presence of the previous assumptions. The above condition is not very strong: for example, the standard Brownian motion satisfies $S(l, x) = S(x, r) = +\infty$ for $x \in I$ (observe that in this case $s(x) = 1$).

If $S(l, x) = S(x, r) = +\infty$ and \( \int_{l}^{r} m(x) \, dx < \infty \) where \( m(u) = \left( b^2(u) s(u) \right)^{-1} \) is the speed density, then $X$ is ergodic and the invariant distribution $P_0$ has (stationary) density \( \bar{p}(x) = m(x) / \int_{l}^{r} m(u) \, du \) with respect to the Lebesgue measure [see Skorokhod (1989), theorem 16]. We say that $X$ has stationary moments of order $r$ if $E[|X|^r] = \int_{l}^{r} |x|^r \, \bar{p}(x) \, dx < \infty$. Note that if $X$ is ergodic and $X_0 = x$ has distribution $P_0$ then $X$ is a stationary process (Arnold, 1974).
Appendix 2

Proof of proposition 1 To guarantee that \( \hat{p}(\Delta, x, y) \) converges uniformly in probability on \( \Xi = \{(x, y) : x, y \in I \} \subset \mathbb{R}^2 \) to \( p(\Delta, x, y) \) it is sufficient to assure that

(i) \( \hat{p}(\Delta, x, y) \) converges in probability for each \( x, y \in I \) (pointwise convergence) and

(ii) \( \hat{p}(\Delta, x, y) \) is stochastically equicontinuous (see Davidson, 1994, theorem 21.9).

(i) Firstly, we study the pointwise converge of \( \hat{\psi} \). Given the almost sure (a.s.)

continuity of \( g(z_u(x, y) + \sqrt{\sigma^2 \Delta} B_u(\omega)) \) in \( u \) it is known that

\[
\lim_{S \to +\infty} \frac{1}{S} \sum_{j=1}^{S} \psi_N(\omega_j) = \mathbb{E}[\psi]
\]

by the law of large numbers, since the \( \psi(\omega_j) \) are i.i.d. and \( \mathbb{E}[|\psi|] = \mathbb{E}[\psi] < +\infty \). Thus, for each \( x \) and \( y \), the estimator \( \hat{p}(\Delta, x, y) \) converges in probability to \( p(\Delta, x, y) \) as \( N \to +\infty \) and \( S \to +\infty \).

(ii) By theorem 21.10 of Davidson (1994), a sufficient condition for \( \hat{p}(\Delta, x, y) \) to
be stochastically equicontinuous is that $\beta = O_p(1)$ where

$$\beta = \sup_{(x^*, y^*) \in \Xi^*} \left\| \begin{bmatrix} \frac{\partial \hat{p}(\Delta, x^*, y^*)}{\partial x} \\ \frac{\partial \hat{p}(\Delta, x^*, y^*)}{\partial y} \end{bmatrix} \right\|$$

and $\|\cdot\|$ is a norm, $\Xi^*$ is an open convex set containing $\Xi$ and $(x^*, y^*) \in \Xi^*$ is a point on the line segment joining two arbitrary points $z$ and $z'$ in $\Xi^*$.

We now prove that $\beta_n = O_p(1)$. Let $\hat{p}(\Delta, x, y) = \phi(x, y) \hat{\psi}(x, y)$ where

$$\phi(x, y) = \frac{1}{\sqrt{2\pi \sigma^2 \Delta}} \exp \left\{ -\frac{(x - y)^2}{2\sigma^2 \Delta} + \frac{G(y) - G(x)}{\sigma^2} \right\}.$$}

We have

$$\frac{\partial \hat{p}(\Delta, x, y)}{\partial y} = \frac{\partial \phi(x, y)}{\partial y} \hat{\psi}(x, y) + \phi(x, y) \frac{\partial \hat{\psi}(x, y)}{\partial y} = \phi(x, y) \left( -\frac{(x - y)}{\Delta \sigma^2} + \frac{a(y)}{\sigma^2} \right) \hat{\psi}(x, y) + \phi(x, y) \frac{1}{S} \sum_{j=1}^{s} \frac{\partial \psi_N(\omega_j)}{\partial y}$$

The term $\phi(x, y) \left( -\frac{(x - y)}{\Delta \sigma^2} + \frac{a(y)}{\sigma^2} \right)$ is a continuous deterministic function. Therefore, it is bounded on all points on the line segment joining the points $(x, y)$ and $(x', y')$ in $\Xi^*$ such that $p(\Delta, x, y)$ and $p(\Delta, x', y')$ are finite. Thus, by part (i) the term $\phi(x, y) \left( -\frac{(x - y)}{\Delta \sigma^2} + \frac{a(y)}{\sigma^2} \right) \hat{\psi}(x, y)$ is bounded in probability (observe that $E[|\psi|] < \infty$).

On the other hand, given the almost sure (a.s.) continuity of $g \left( z_u(x, y) + \sqrt{\sigma^2 \Delta} B_u(\omega) \right)$ and $g' \left( z_u(x, y) + \sqrt{\sigma^2 \Delta} B_u(\omega) \right)$ in $u$ it follows that

$$\frac{\partial \psi_N(\omega)}{\partial y} = \psi_N(\omega) \left\{ \frac{\sigma^2 \Delta}{N} \sum_{i=0}^{N-1} g' \left( z_{t_i}(x, y) + \sqrt{\sigma^2 \Delta} B_{t_i}(\omega) \right) \right\}$$

converges in probability to $\psi(\omega) \xi(\omega)$ as $N \to +\infty$ where

$$\xi(\omega) = \sigma^2 \Delta \int_0^1 g' \left( z_u(x, y) + \sqrt{\sigma^2 \Delta} B_u(\omega) \right) u \, du.$$
\( \xi(\omega) \) is a random variable that is bounded in probability since the limits of integration are finite, the integrand is continuous and \( B_t, \ t \in [0, 1] \) is a bounded process (a.s.). Using the same arguments in part (i), one can conclude that the random quantity
\[
\frac{1}{S} \sum_{j=1}^{s} \frac{\partial \psi_N(\omega_j)}{\partial y} \text{ converges to } E \left[ \frac{\partial \psi_N(\omega)}{\partial y} \right].
\]

Similar arguments can be used to show that \( \frac{\partial \hat{p}(\Delta, x, y)}{\partial x} = O_p(1) \).

\textbf{Proof of proposition 2} By proposition 1 we can write
\[
\hat{p}(\Delta, x, y_i) = p(\Delta, x, y_i) + o_p(1)
\]
where \( \lim o_p(1) = 0 \) uniformly in probability on \( \Xi \) as \( N \to +\infty \) and \( S \to +\infty \).

Without loss of generality consider \( y_0 = 0 \) and \( y_1 = 1 \).

\[
\left| \sum_{i=1}^{M} \hat{p}(\Delta, x, y_i) \Delta y_i - \int_{y_0}^{y_1} p(\Delta, x, y) \, dy \right|
= \left| \frac{1}{M} \sum_{i=1}^{M} \hat{p}(\Delta, x, y_i) - \int_{0}^{1} p(\Delta, x, y) \, dy \right|
\leq \left| \frac{1}{M} \sum_{i=1}^{M} p\left(\Delta, x, \frac{i}{M}\right) - \int_{0}^{1} p(\Delta, x, y) \, dy \right| + \frac{1}{M} \sum_{i=1}^{M} o_p(1)
\leq \left| \frac{1}{M} \sum_{i=1}^{M} p\left(\Delta, x, \frac{i}{M}\right) - \int_{0}^{1} p(\Delta, x, y) \, dy \right| + \frac{1}{M} \sum_{i=1}^{M} o_p(1)
\]

Due to the continuity of \( p(\Delta, x, y) \), \( \frac{1}{M} \sum_{i=1}^{M} p\left(\Delta, x, \frac{i}{M}\right) \) is a Riemann sum. Therefore, the first term goes to zero as \( M \to +\infty \). The second term goes to zero uniformly in probability on \( \Xi \) as \( N \to +\infty \) and \( S \to +\infty \).
Appendix 3

Since Richter (2002) is currently unpublished, we present in this appendix two propositions from Richter’s paper that we used in our paper. A draft of the proofs are also presented.

Proposition 3 of Richter (2002) reads:

Let \( \tilde{X}_t; i \geq 0 \) be the Euler approximation of the SDE \( dX_t = a(X_t)\, dt + b(X_t)\, dW_t, \) \( X_0 = \zeta \) (where \( \zeta \) is independent of the filtration generated by the Brownian motion. Assume that \( P[b(X_0) = 0] < 1 \) and that the diffusion coefficient does not degenerate outside a compact interval \( K \), i.e. \( b(x) \neq 0 \) for \( x \notin K \). If there exists \( p > 0 \) and \( \phi > 0 \) such that

\[
\lim_{x \to \pm \infty} \frac{|x|^p \left| a(x) \right|}{|b(x)|} = 0 \quad \text{and} \quad \lim_{x \to \pm \infty} \inf \frac{|a(x)|}{|x|} \geq \phi
\]

then \( P \left[ \bigcap_{n \geq 0} \left| \bar{X}_t \right| > (1 + \tilde{\phi}) \left| \bar{X}_{t-1} \right| \right] > 0 \) where \( 0 < \tilde{\phi} < \phi \).

The proof is based on a preliminary lemma: let \( \{\alpha_n; n \geq 0\} \) be a positive sequence such that \( \lim_{n \to +\infty} \alpha_n = 0 \) and define the set \( A = \cap_{n \geq 0} (|\Delta W_n| \geq \alpha_n) \). Then \( P(A) > 0 \) if and only if \( \sum_{n \geq 0} \alpha_n < \infty \).

The main steps in the proof of the proposition consist in defining the sequence

\[
\alpha_n = c \sup_{|x| \geq (1+\varphi)^n y_0} \frac{|a(x)|}{|b(x)|}, \quad n \geq 0
\]

and the set \( A = \cap_{n \geq 0} (|\Delta W_n| \geq \alpha_n) \) and in proving that \( \sum_{n \geq 0} \alpha_n < \infty \) under the conditions of the proposition. Then it is shown that \( \left| \bar{X}_t \right| > (1 + \tilde{\phi}) \left| \bar{X}_{t-1} \right| \) on the set \( A \). It follows that \( P(A) > 0 \), i.e. \( P \left[ \bigcap_{n \geq 0} \left| \bar{X}_t \right| > (1 + \tilde{\phi}) \left| \bar{X}_{t-1} \right| \right] > 0 \) where \( 0 < \tilde{\phi} < \phi \).

Proposition 4 of Richter (2002) reads
Let \( \tilde{X}_t; i \geq 0 \) be the Euler approximation of the SDE \( dX_t = b(X_t) \, dW_t, X_0 = \zeta \) (where \( \zeta \) is independent of the filtration generated by the Brownian motion. Assume that \( b > 0 \). If there exists \( p > 1 \) such that

\[
\lim_{x \to \pm \infty} \frac{|x|^p}{|b(x)|} = 0
\]

then for any \( \varphi > 0 \), \( P \left[ \bigcap_{n \geq 0} \left| \tilde{X}_{t_i} \right| > (1 + \tilde{\varphi}) \left| \tilde{X}_{t_{i-1}} \right| \right] > 0. \)

The proof of this proposition is similar to the previous one (the sequence \( \{\alpha_n\} \) is now defined as \( \alpha_n = (2 + \varphi) \sup_{|x| \geq (1 + \varphi)^n \zeta} |x| / |b(x)|. \)

Since every local martingale that admits a stationary distribution \( \bar{p} \) and has a finite moment \( p > 1 \) satisfies the conditions of proposition 4, one has the following corollary:

Corollary 1 (Richter) Let \( X \) be a local martingale defined by the SDE \( dX_t = b(X_t) \, dW_t \) where \( b(x) > 0 \) is a function with continuous derivative admitting a stationary distribution \( \bar{p}(x) = b^{-2}(x) / \int_{-\infty}^{\infty} b^{-2}(u) \, du \), such that \( \int |x|^p \, \bar{p}(x) \, dx < \infty \) for some \( p > 1 \). Then, the corresponding Euler scheme \( \tilde{X} \), i.e. \( \tilde{X}_{t_i} = \tilde{X}_{t_{i-1}} + b\left(\tilde{X}_{t_{i-1}}\right) \Delta W_{t_i} \) (\( X_0 = x \)), verifies \( P \left[ \bigcap_{n \geq 0} \left| \tilde{X}_{t_i} \right| > (1 + \tilde{\varphi}) \left| \tilde{X}_{t_{i-1}} \right| \right] > 0 \) where \( \tilde{\varphi} > 0. \)

Acknowledgments

I would like to thank a referee for the comments and suggestions which led to the improvement of the paper. I am also grateful to Tom Kundert for helpful comments.

This research was supported by the Fundação para a Ciência e a Tecnologia (FCT) and by POCTI.

28
List of Figure Captions

Figure 1: Conditional Density and Conditional Distribution Function Estimations: Comparison of Three Methods

Figure 2: Simulated Path of the process $dX_t = (1 + X_t^2)\,dW_t$, $X_0 = 0$ (10,000 observations)

Figure 3: True Density $\bar{p}(x)$, Estimated Density $\hat{p}(x)$ and Gaussian Density
Figure 1: Conditional Density and Conditional Distribution Function Estimations: Comparison of Three Methods
Figure 2: Simulated Path of the process \( dX_t = (1 + X_t^2) \, dW_t, \, X_0 = 0 \) (10,000 observations)

\[
dX_t = 1 + X_t^2 \, dW_t, \quad X_0 = 0
\]

Figure 3: True Density \( \rho(x) \), Estimated Density \( \hat{\rho}(x) \) and Gaussian Density