Localization and Exact Simulation of Brownian Motion Driven Stochastic Differential Equations

Nan Chen and Zhengyu Huang
Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong
email: nchen@se.cuhk.edu.hk and zyhuang@se.cuhk.edu.hk

Generating sample paths of stochastic differential equations (SDE) using the Monte Carlo method finds wide applications in financial engineering. Discretization is a popular approximate approach to generating those paths: it is easy to implement but prone to simulation bias. This article presents a new simulation scheme to exactly generate samples for SDEs. The key observation is that the law of a general SDE can be decomposed into a product of the law of standard Brownian motion and the law of a doubly stochastic Poisson process. An acceptance-rejection algorithm is devised based on the combination of this decomposition and a localization technique. The numerical results corroborates that the mean-square error of the proposed method is in the order of $O(t^{-1/2})$, which is superior to the conventional discretization schemes. Furthermore, the proposed method also can generate exact samples for SDE with boundaries which the discretization schemes usually find difficulty in dealing with.

Key words: Monte Carlo simulation; Stochastic Differential Equations; Option Pricing; Exact Method; Localization

MSC2000 Subject Classification: 65C05, 65C30, 62P05

OR/MS subject classification: Finance: Asset Pricing; Simulation: Efficiency; Probability: Diffusion

---

1. Introduction. Finding an efficient way to implement the Monte Carlo simulation method to generate sample paths for stochastic differential equations (SDE) has triggered a large amount of interest from both the financial engineering and simulation communities. The solutions to SDEs are used in many financial applications to model the evolution of market variables. People are especially interested in evaluating the expected values of functionals on such solutions. For instance, derivative pricing problems are usually reduced down to calculating the expectation of a payoff function over the sample paths of the modeling SDE. However, explicit-form solutions are absent in most circumstances and we have to rely on some numerical methods to obtain them. The Monte Carlo method becomes an attractive option because it can be implemented flexibly and its computational complexity does not depend upon the dimensionality of the problem. In this paper, we focus on constructing an acceptance-rejection method to simulate the sample paths of SDEs. The method is *exact* in the sense that the marginal distribution of the simulated values coincides with the marginal distribution of the continuous-time process on the simulation time grid.

The conventional approach to SDE simulation is via various discretization schemes. The idea is to divide the whole time interval of interest into a discrete time grid and then to simulate a discrete process to approximate the original continuous-time SDE on the time grid according to its finite-difference counterparts. They introduce discretization bias into the simulation. We need to increase the number of time steps considerably to achieve a higher accuracy for the simulation outcomes. Thus, it is computationally expensive to reduce such bias down to an acceptable level. A great deal of research has been devoted to establishing the corresponding convergence rates, or how fast the discretization bias decreases as the number of steps increases, under different schemes; see Kloeden and Platen [37] for a comprehensive overview in that direction.

Given a fixed computational time budget, it is difficult for the discretization schemes to resolve a tradeoff between the discretization bias and the simulation variance. Expending more efforts to increase the number of time steps per path reduces the bias due to discretization, but it also decreases the number of samples that can be completed within the budget, which has a direct impact on the estimator variance. Duffie and Glynn [21] show that the optimal root mean-square error (RMSE), a performance indicator summarizing the effects of both bias and variance, can reach $O(t^{-\beta/(2\beta+1)})$ for a discretization scheme with a convergence rate β for a fixed time budget t. This makes precise the notion that simulating an SDE using a discretization scheme is inferior to simulating it exactly, which can achieve the RMSE with an order of $O(t^{-1/2})$. Meanwhile, the discretization schemes hardly yield valid confidence intervals for estimates because the magnitudes of the simulation bias are often unknown.

The discretization schemes encounter obstacles in generating samples for the SDEs with boundaries. Such SDEs are of particular interest in many financial applications. A well known example is the Cox-Ingersoll-Ross (CIR) model. The solution to this SDE remains nonnegative and it is thus widely used in finance to describe the dynamics of interest rates or equity volatility (see, e.g., Cox, Ingersoll and Ross [18] and Heston [29]). As shown in Section 2, the discretization approximation of this model will produce negative samples with a significant chance. Thus, some kind of truncation is inevitable to avoid these pathological paths, but how to perform it "optimally" is still a challenging issue for the discretization schemes.

In light of all the difficulties of the discretization methods, this paper proposes an innovative simulation scheme to exactly generate samples for SDEs. The key observation underlying our method is that the likelihood ratio between the law of the SDE and a standard Brownian motion $\{W_u, 0 \le u \le T\}$ is proportional to $\exp(-\int_0^T \phi(W_u)du)$, where ϕ is a deterministic function defined explicitly by the SDE's coefficients. Using this exponential function as an acceptance-rejection kernel, we establish an exact simulation algorithm to obtain SDE sample paths from candidates proposed from the Brownian motion. It is easy to see that the exponential is related to the probability that there are no arrivals occurring in $[0, T]$ for a doubly stochastic Poisson process with arrival intensity $\phi(W_u)$ at u, $0 \le u \le T$. If $\phi(\cdot)$ is uniformly bounded above, the stochastic thinning, a technique which dates back to the work of Lewis and Shelder [38], can be applied to generate the Poisson process. Beskos and Roberts [8] propose an exact simulation algorithm based on this observation.

The major contribution of this paper is to introduce an idea of *localization* to generalize the Beskos-Roberts algorithm. The accompanying function ϕ is not bounded at all in many important financial applications. Therefore, a significant obstacle remains if we intend to apply the Beskos-Roberts algorithm to simulate such SDEs. We propose an approach in this paper to break a candidate sample path of the Brownian motion W into small "local" pieces. Within each piece, the part of W is bounded and hence $\phi(W_u)$ is bounded as long as ϕ is continuous. The acceptance-rejection is then applicable again to produce samples for the SDE, piece by piece. Furthermore, our method is readily extensible to simulating the SDEs with boundaries. When we choose the localization width adaptively as the simulation proceeds, we can completely avoid the truncation on the boundary. The numerical results show that the empirical distributions of these models produced by our method perfectly match their theoretical distributions, in contrast to significant distortions in the discretization schemes. The RMSE of the proposed algorithm achieves the order of $O(t^{-1/2})$ because it is exact, which is also confirmed by the numerical experiments in the paper.

This paper is closely related to the literature of exact simulation of SDEs. Beskos et al. [9], [10] suggest a layer Brownian motion to relax the bounded assumption to a condition that either $\liminf_{x\to+\infty} \phi(x)$ or $\liminf_{x\to-\infty} \phi(x)$ is finite. Casella and Roberts [16] discuss how to exactly simulate diffusions killed at a single boundary. In addition, some authors have made breakthroughs in exact simulations for some specific SDEs. Broadie and Kaya [13], [14] develop an exact simulation scheme for the affine stochastic volatility jump diffusions. Glasserman and Kim [26] use gamma expansion in the Broadie-Kaya method to accelerate the simulation for the Heston stochastic volatility model. Our method is also rooted in the literature of diffusion sample path decomposition in the probability theory. As noted by Williams [46], the sample path of a Brownian motion, if the maximum/minimum is given, can be decomposed into two "back-to-back" Brownian meanders. Imhof [31] shows that we can further express the Brownian meander as the square root of a sum of three squared independent Brownian bridges. These two facts play an essential role in the construction of the proposed exact sampling algorithm. To our knowledge, our paper is the first one to apply these theoretical discoveries to the simulation of SDEs.

The rest of this paper is organized as follows. Section 2 briefly reviews some fundamental results about the discretization methods of SDE simulation. Sections 3 and 4 review the Beskos-Roberts exact simulation method and more importantly, our improvement via the localization technique. Section 5 discusses how to extend our method to the SDEs with boundaries and presents some efficiency analysis on the proposed algorithm. Section 6 is devoted to some numerical experiments. All of the proofs are deferred to the appendix.

2. SDEs and Discretization Methods.

Consider the following one-dimensional SDE:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{1}$$

where $\{W_t, 0 \le t \le T\}$ is a standard scalar Brownian motion, and μ and σ are the drift and the volatility of X, respectively. Both of the drift and the volatility can be state-dependent. This SDE model is used widely in financial engineering to describe the evolution of asset prices, interest rates and other market variables. Of interest here is the pricing problem of a derivative security contingent on the dynamic (1). Assume that the derivative payoff function is given by $f(X_T)$, dependent on the values of X at instance T. The no-arbitrage arguments yield that the derivative price should be in the form of the following expectation

$$u(x) = e^{-rT} E[f(X_T)|X_0 = x], \tag{2}$$

if the SDE (1) is assumed to be under the risk neutral probability measure. One minor issue we need to point out is that the drift of the dynamic (1) should be identical with the risk-free interest rate if we intend to use it to model the asset price in a risk-neutral world. This case is just a special case in which the following algorithms works. We consider a broader setting given by a non-constant μ for the interest of simulation.

The formulation (2) paves a way for the application of the Monte Carlo method in derivative pricing. Fix a large integer M and let $\Delta t = T/M$. A simple but crude Euler discretization of the SDE (1) is given by

$$\hat{X}_i = \hat{X}_{i-1} + \mu(\hat{X}_{i-1})\Delta t + \sigma(\hat{X}_{i-1})\Delta W_i, \tag{3}$$

where $\Delta W_i := W_{t_i} - W_{t_{i-1}}$ for all $t_i = i\Delta t,\ 1 \le i \le M$. Note that ΔW_i follows a Gaussian distribution $N(0, \Delta t)$. We can finish the estimation to (2) in two steps: generate a sequence of ΔW_i's from $N(0, \Delta t)$; substitute them into (3) to obtain $\hat{X}_i$ step by step. One straightforward estimate for $E[f(X_T)]$ is then formed by the sample mean $\sum_{k=1}^{N} f(\hat{X}_T^k)/N$ across all N replications.

There are two types of error associated with this Monte Carlo estimator. One error is caused by the discretization bias. The literature has established that the weak convergence rate of the Euler scheme is in the order of 1 under certain proper smooth conditions regarding μ, σ and f. That is, there exists a constant c such that

$$|E[f(\hat{X}_T)] - E[f(X_T)]| \le c\Delta t.$$

See, for example, Bally and Talay [7], Kloeden and Platen [37], Talay and Turbo [44]. The other error comes from the statistical fluctuation of the estimator. The central limit theorem implies that, as the number of replications N tends to $+\infty$, there exists another constant c' such that

$$\text{standard deviation} \approx \frac{c'}{\sqrt{N}}.$$

The simulation literature usually uses the root of mean-square error (RMSE) to measure the overall performance of estimators. It is defined as the square root of the sum of squared bias and variance to take into account the influence of both error sources. From the previous discussion, we know that the RMSE of the Euler scheme should be

$$\sqrt{c^2 \cdot (\Delta t)^2 + \frac{c'}{N}}. \tag{4}$$

Given a fixed computational budget, the RMSE (4) illustrates a tradeoff facing to the discretization scheme: increasing the number of time steps helps to reduce Δt, but it also leads to a lower number of paths that can be completed. To strike the balance, Duffie and Glynn [21] discuss how to allocate computational efforts to achieve the best outcome in terms of RMSE. Their result implies that the Euler scheme can reach $O(t^{-1/3})$ for a total budget time t if we generate $O(t^{2/3})$ paths and $O(t^{1/3})$ steps in each path. One way to improve the performance of the Euler scheme is to pursue higher order approximations. Milstein [40] uses the Taylor expansion to reach a scheme with a convergence rate of order 2. This second-order scheme will improve the optimal RMSE to $O(t^{-2/5})$. In theory, one may be able to repeat the same expansion as in the Milstein scheme to derive even higher-order schemes (Kloeden and Platen (1992), §14.5). As a general conclusion, the optimal RMSE is $O(t^{-\beta/(2\beta+1)})$ for any scheme with a convergence rate of β if we generate $O(t^{1/(2\beta+1)})$ steps within each path and $O(t^{2\beta/(2\beta+1)})$ paths for a fixed time budget t.

The preceding generic discussion reveals several shortcomings of various discretized estimators. To achieve higher-order accuracy, the schemes require high-order smoothness on both SDE coefficients and

the payoff functions, which are not satisfied typically in financial applications. Secondly, regardless of how high β is, the discretization schemes are all inferior to the order of $O(t^{-1/2})$, the rate associated with unbiased estimation. This makes precise the reason why an exact simulation scheme, if possible, is more preferable. Thirdly, it is very difficult to find the exact magnitude of bias, although we know its asymptotic order. Therefore, the discretization schemes can hardly yield valid confidence intervals for their estimates.

As mentioned in the introduction, the discretization methods also meet significant obstacles when applied to simulating some SDEs with boundaries. For example, consider the following CIR model

$$dX_t = (a - bX_t)dt + \sigma\sqrt{X_t}dW_t, \quad X_0 = x, \tag{5}$$

where $a, b > 0$. One of the most important characteristics of the model is its non-negativeness, i.e., $X_t \geq 0$ for all t. When applying the plain Euler scheme, we have

$$\hat{X}_i = \hat{X}_{i-1} + (a - b\hat{X}_{i-1})\Delta t + \sigma\sqrt{\hat{X}_{i-1}}\Delta W_i. \tag{6}$$

There is a considerable chance to obtain a negative $\hat{X}_i$ because $\Delta W_i \sim N(0, \Delta t)$, even if we start from a positive $\hat{X}_{i-1}$. Such $\hat{X}_i$ would be "illegal" when plugged back to the recursion to produce $\hat{X}_{i+1}$. Some researchers have proposed various ways to remedy the pathological behavior of the discretization schemes around the origin. In particular, Deelstra and Delbaen [19], Higham and Mao [30], Berkaoui et al. [11] and Lord et al. [39] suggest a quick "fix" on the Euler scheme (6) by either setting the process equal to zero whenever it attains a negative value, or by reflecting it in the origin and continuing from there on. Alfonsi [3], [4] and Kahl and Jäckel [33] consider an implicit Milstein scheme to preserve the non-negativeness. Anderson [5] uses a combination of approximations to make his method applicable to a variety of parameters. All of these methods are prone to distorting the process distribution at the boundary $X = 0$ significantly.

It is also worthwhile mentioning another recent development in reducing discretization error. If one uses a fine time grid, the bias will be very small but the computational cost will be very large. If working on a coarse grid, the accuracy is much less but the cost is saved. Encouraged by this intuition, Kebaier [36] proposes the Richard-Romberg extrapolation to generate SDE samples on two levels of time grids, using a coarser grid to simulate a crude framework and a finer grid to fine tune the bias. This idea is extended by Giles [24] to multiple grids with different degrees of fineness, instead of just two. The method shows significant bias reduction and computational cost saving.

3. The Beskos-Roberts Exact Simulation. Instead of simulating an approximate process recursively, the exact simulation method offers an alternative approach to generating the sample paths of SDEs according to the process law exactly. The basic idea of the exact simulation is based upon the acceptance-rejection method (ARM), which is among the most applicable approaches to sampling random numbers (see, e.g., Asmussen and Glynn [6] and Glasserman [25]).

In general, suppose that we wish to generate samples for a random variable Θ with a density f defined on some set $\mathcal{X}$, which could be a subset of the real line, or a more general set. And we know how to generate samples for another relatively easier random variable θ with a density g. Assume that the likelihood ratio of Θ and θ satisfies a property that

$$\frac{f(s)}{g(s)} \leq c, \ s \in \mathcal{X}$$

for a constant c. Using ARM, Θ can be drawn in the following two steps:

1. Propose a $\theta \sim g$ and a Bernoulli r.v. I with the probability
$$P[I = 1|\theta] = 1 - P[I = 0|\theta] = \frac{f(\theta)}{cg(\theta)};$$
2. If $I = 1$, accept θ as a sample for Θ. Otherwise, reject it and return to Step 1.

The time needed to obtain a qualified candidate is random and its expected waiting time equals to c. In this sense, we prefer to look for a tighter bound c such that fewer candidate samples from g will be wasted.

Now, turn back to simulation of the SDE (1). Our intention is to implement ARM on candidates proposed from a standard Brownian motion W_T to obtain X_T. Simulation of W_T is easy because $W_T \sim N(0,T)$. The key then becomes to figure out the likelihood ratio between the laws of X_T and W_T. We need some technical assumptions to facilitate the search of the likelihood ratio. Let $D_X = (\underline{x}, \bar{x})$ denote the domain of X. We will consider two cases, $D_X = (-\infty, +\infty)$ and $D_X = (0, +\infty)$, in this paper. The latter one is more relevant in finance because many models of asset prices, interest rates or equity volatilities should be always positive. We assume that

Assumption 3.1 (Smoothness of the Coefficients) *The function $\mu(x)$ and $\sigma(x)$ are infinitely differentiable in x for all $x \in D_X$.*

and

Assumption 3.2 (Non-Degeneracy of the Diffusion) *1. If $D_X = (-\infty, \infty)$, there is a constant $c > 0$ such that $\sigma(x) > c$ for all $x \in D_X$;*
2. If $D_X = (0, \infty)$, $\lim_{x \to 0+} \sigma(x) = 0$ and there exist positive constants ξ, k and ρ such that $\sigma(x) \geq kx^\rho$ for all $x \in [0, \xi]$.

Consider the Lamperti transform (see, e.g., Florens [23]) defined as

$$F(y) = \int_x^y \frac{1}{\sigma(u)} du$$

for $y \in D_X$, where x is the initial value of the SDE (1). Assumption 3.2 ensures that the transform is well defined for any $y \in D_X$. Furthermore, F is a strictly increasing function because $\sigma(\cdot) > 0$. Transform X into another process Y defined as $Y_t := F(X_t)$. Ito's lemma implies that Y must satisfy the following SDE:

$$dY_t = \alpha(Y_t)dt + dW_t, \quad Y_0 = 0, \tag{7}$$

where α is given by

$$\alpha(y) = \frac{\mu(F^{-1}(y))}{\sigma(F^{-1}(y))} - \frac{1}{2}\sigma'(F^{-1}(y)).$$

The monotonicity of the transform F makes sure that simulating X_T is equivalent to simulating Y_T. We can recover X_T by the relationship $X_T = F^{-1}(Y_T)$ once we have samples for Y_T, where F^{-1} is the inverse to F.

Denote $\underline{y} = \lim_{x \to \underline{x}+} F(y)$ and $\bar{y} = \lim_{x \to \bar{x}-} F(y)$. The Lamperti transform maps D_X into $D_Y = (\underline{y}, \bar{y})$, the domain of Y. When $D_X = (-\infty, +\infty)$, it is easy to check that $D_Y = (-\infty, +\infty)$ because $\sigma(\cdot) > c > 0$ in this case according to Assumption 3.1. The situation for $D_X = (0, +\infty)$ is more complicated. The following three examples interest us especially:

- D_Y can be $(-\infty, \infty)$. For instance, consider the Black-Scholes model,

  $$dX_t = \mu X_t dt + \sigma X_t dt, \quad X_0 = x.$$

  Its Lamperti transform is given by

  $$F(y) = \frac{1}{\sigma} \ln\left(\frac{y}{x}\right),$$

  which maps D_X into $D_Y = (-\infty, +\infty)$.
- D_Y can be $(\underline{y}, +\infty)$ where $\underline{y} > -\infty$. The aforementioned CIR model (5) falls in this category. Under the transform

  $$F(y) = \frac{2(\sqrt{y} - \sqrt{x})}{\sigma},$$

  the transformed process Y is valued in a range

  $$D_Y = \left(-\frac{2\sqrt{x}}{\sigma}, +\infty\right).$$

- D_Y can be $(-\infty, \bar{y})$ where $\bar{y} < +\infty$. Chan et al. [17] suggest the following linear-drift CEV-type-diffusion model to fit the short-term interest rate in the United States: for $\gamma > 1$,

$$dX_t = (a - bX_t)dt + \sigma X_t^{\gamma} dW_t, \quad X_0 = x \tag{8}$$

The process X is distributed on $(0, +\infty)$ for $a, b > 0$ and $\gamma > 1/2$. With the Lamperti transform

$$F(y) = -\frac{1}{\sigma(\gamma - 1)} \left(\frac{1}{y^{\gamma-1}} - \frac{1}{x^{\gamma-1}} \right),$$

the value range of the transformed process Y for this model is

$$D_Y = \left(-\infty, \frac{1}{\sigma(\gamma - 1)x^{\gamma-1}} \right).$$

To guarantee the existence and uniqueness of the solution to (7), we need an additional assumption to prevent Y from "exploding" (i.e., reaching the boundaries) in finite time. This prevention is necessary from the perspective of financial modeling. As a sensible model for a market variable, X can rarely, if at all, reach $\pm\infty$ or 0 in finite time. The economic literature mentions very little about how the process would behave after the boundaries are hit. Meanwhile, the treatment of the boundary behavior of X will significantly complicate the simulation itself. Different specifications of boundary behavior will lead to different solutions to the SDE, which would totally destroy the uniqueness of the solution (see, e.g. Karlin and Taylor [34, §15.8]). To avoid the complexity of the boundary issue, we introduce

Assumption 3.3 (Boundary Behavior) *1. Left boundary: If $\underline{y} > -\infty$, there are constants ϵ, κ and β such that $\alpha(y) \geq \kappa(y - \underline{y})^{-\beta}$ for $\underline{y} < y < \underline{y} + \epsilon$, where either $\kappa > 0$ and $\beta > 1$ or $\kappa \geq 1$ and $\beta = 1$. If $\underline{y} = -\infty$, there exist positive constants B and K such that $\alpha(y) \geq Ky$ for all $y < -B$.*

2. Right boundary: If $\bar{y} < +\infty$, there exist constants ϵ, κ and β such that $\alpha(y) \leq -\kappa(\bar{y} - y)^{-\beta}$ for $\bar{y} - \epsilon < y < \bar{y}$, where either $\kappa > 0$ and $\beta > 1$ or $\kappa \geq 1/2$ and $\beta = 1$. If $\bar{y} = +\infty$, there exist positive constants B and K such that $\alpha(y) \leq Ky$ for all $y > B$.

The transformed process does not hit the boundary under the above assumptions, which put limits on the rates of the process drift when it approaches the boundaries. For instance, when $D_Y = (-\infty, +\infty)$ and Y is sufficiently large (i.e., it approaches $+\infty$), Assumption 3.3 imposes a requirement that the drift grows sub-linearly, which will make the boundary of $+\infty$ unattainable. When $D_Y = (\underline{y}, +\infty)$ and Y is near the left boundary $\underline{y}$, the drift is assumed to be a positive number larger than $\kappa(y - \underline{y})^{-\beta}$, which pushes Y away from $\underline{y}$. In this way we can rule out the possibility that the process explodes in a finite time horizon. Under Assumptions 3.1-3.3, Aït-Sahalia [2] shows that the SDE (7) admits a weak solution $\{Y_t, t \geq 0\}$, unique in probability law for every initial value Y_0 and furthermore $P[T_Y = \infty] = 1$, where $T_Y = \inf\{t \geq 0 : Y_t \notin (\underline{y}, \bar{y})\}$, the first time that Y moves beyond D_Y. Although all of the assumptions impose restrictions in terms of Y, it is easy to check whether they hold with the process X.

The advantage of introducing Y is that we are able to write down an explicit expression for the likelihood ratio of Y_T with respect to W_T. Beskos and Roberts [8] focus on the case of $D_Y = (-\infty, +\infty)$ and show the following proposition:

Proposition 3.1 *Suppose that α satisfies:*

$$E\left[\exp\left(\frac{1}{2} \int_0^T \alpha^2(W_t)dt \right) \right] < +\infty. \tag{9}$$

Then the density of Y_T is given by

$$f_{Y_T}(y) = \frac{1}{\sqrt{2\pi T}} \exp\left(A(y) - \frac{y^2}{2T} \right) \cdot E\left[\exp\left(-\int_0^T \phi(W_u)du \right) \mid W_T = y \right],$$

where

$$A(y) = \int_0^y \alpha(u)du \text{ and } \phi(y) = \frac{\alpha'(y) + \alpha^2(y)}{2}.$$

To facilitate the application of ARM, they further introduce an assumption that A satisfies

$$C := \int_{-\infty}^{+\infty} \exp\left(A(y) - \frac{y^2}{2T}\right) dy < +\infty. \tag{10}$$

If that is true, a new probability density function can be defined as follows:

$$h(y) := \frac{1}{C} \exp\left(A(y) - \frac{y^2}{2T}\right).$$

By Proposition 3.1, the likelihood ratio of f_{Y_T} with respect to h should be

$$\frac{f_{Y_T}(y)}{h(y)} = \frac{C}{\sqrt{2\pi T}} E\left[\exp\left(-\int_0^T \phi(W_u)du\right) \mid W_T = y\right].$$

To construct the Bernoulli I, note that

$$\exp\left(-\int_0^T \phi(W_u)du\right)$$

equals the probability of the event that there are no arrivals occurring within $[0,T]$ for a doubly stochastic Poisson process with the intensity $\phi(W_u)$ at u, $0 \le u \le T$. Therefore, we accomplish the job if define I as the indicator of such event. More specifically, suppose that there exists a constant K such that

$$0 \le \phi(x) < K \text{ for all } x \in (-\infty, +\infty). \tag{11}$$

Lewis and Shedler [38] propose a technique of stochastic thinning to generate arrivals for doubly stochastic Poisson processes from a homogeneous one with a stronger intensity. Sampling Poisson events with intensity K in $[0,T]$ to obtain a set of arrival times $\{0 < \tau_1 < \cdots < \tau_N < T\}$. For each τ_i, draw a uniform $U_i \sim U(0,1)$. Retain τ_i if $U_i \le \phi(W_{\tau_i})/K$ and erase it otherwise. Let $I = 1$ if there is no arrival left eventually after N retaining-or-erasing tests and $I = 0$ if there are some. It is easy to see that this I is what we desire.

The values of W_{τ_i} must be known to conduct the retaining-or-erasing tests. As W_T is already fixed at y, this is equivalent to sampling from a Brownian bridge. Its simulation is also standard in the literature; we provide a brief overview of the procedure in Appendix A. In summary, the Beskos-Roberts exact algorithm simulates Y_T through the following six steps:

The Beskos-Roberts Exact Simulation

1. Propose a candidate $H \sim h$.
2. Simulate a homogeneous Poisson process with the intensity K and denote $\tau_1, \cdots, \tau_N$ to be all Poisson arrivals in $[0,T]$.
3. If $N = 0$, let $I = 1$ and goto step 6; otherwise, evaluate $W_{\tau_1}, \cdots, W_{\tau_N}$, given $W_T = H$, using the simulation of Brownian bridges.
4. Generate $U_i \sim U(0,1)$ for each τ_i. Retain τ_i if $U_i \le \phi(W_{\tau_i})/K$; otherwise, erase it.
5. Define $I = 1$ if there are no arrivals left in $[0,T]$ and $I = 0$ otherwise.
6. Accept the proposed H if $I = 1$ and reject it if $I = 0$.

Recall that several conditions are indispensable to this algorithm: the Novikov's condition (9), the integrability assumption of $\exp(A(y) - y^2/2T)$, (10), and the uniform bounded assumption of ϕ, (11). However, many SDEs popularly used in financial engineering do not satisfy these conditions. This motivates our work to extend the algorithm.

4. Main Result: A Localization Technique. This section presents our improvement to generalize the above exact simulation method, aiming at the financial applications. The key idea is to use a technique of *localization*, i.e., breaking sample paths of a Brownian motion into small bounded pieces. We illustrate the method using the case of $D_Y = (-\infty, \infty)$ in this section. We also find that this approach is potentially applicable for the cases of $D_Y = (\underline{y}, \infty)$ and $D_Y = (\overline{y}, \infty)$. The relevant discussion is deferred to the next section.

Before proceeding to a detailed technical presentation of our method, we demonstrate the intuition behind the improvement first. Fix a positive constant L and introduce a sequence of first passage times:

$$\zeta_1 = \inf\{t \geq 0 : |Y_t| \geq L\} \text{ and } \zeta_i = \inf\{t \geq 0 : |Y_t - Y_{\zeta_{i-1}}| \geq L\},\ i \geq 1.$$

In words, ζ_1 is the first time that Y leaves the band of $[-L, L]$ and ζ_i is the first time $Y_t - Y_{\zeta_{i-1}}$, the process increment after ζ_{i-1}, leaves $[-L, L]$. Instead of simulating Y_T in one shot as in the last section,

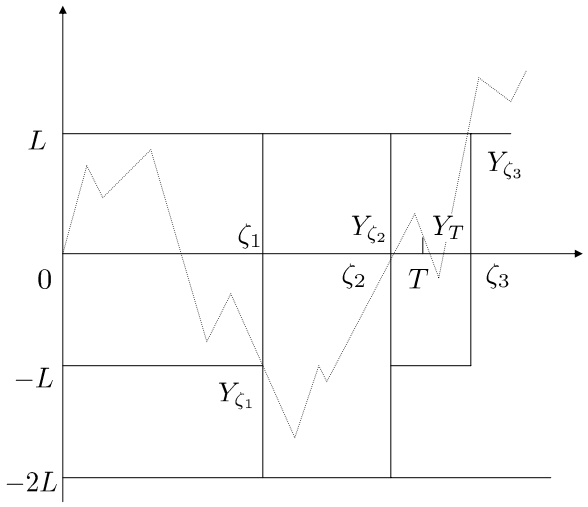

Figure 1: First passage time ζ_i "localizes" Y into bounded pieces $[\zeta_i, \zeta_{i+1}]$. For instance, $\{Y_u, 0 \leq u \leq \zeta_1\}$ is bounded in $[-L, L]$, $\{Y_u, \zeta_1 \leq u \leq \zeta_2\}$ is in $[-2L, 0]$ and so on.

we generate (ζ_i, Y_{ζ_i}) sequentially up to T. More specifically, given $(\zeta_{i-1}, Y_{\zeta_{i-1}})$, when $\zeta_{i-1} < T$, we shall generate two random variables such that

$$\Delta\zeta_i := \zeta_i - \zeta_{i-1} \quad \text{and} \quad \Delta Y_i := Y_{(\zeta_{i-1}+\Delta\zeta_i)\wedge T} - Y_{\zeta_{i-1}}.$$

If $\Delta\zeta_i < T - \zeta_{i-1}$, update $\zeta_i := \zeta_{i-1} + \Delta\zeta_i$, $Y_{\zeta_i} := Y_{\zeta_{i-1}} + \Delta Y_i$, and we then reach the next pair of (ζ_i, Y_{ζ_i}). If $\Delta\zeta_i > T - \zeta_{i-1}$, stop the algorithm by evaluating $Y_T := Y_{\zeta_{i-1}} + \Delta Y_i$. Figure 1 visualizes how the above procedure runs.

The sequence of (ζ_i, Y_{ζ_i}) "localizes" the sample paths of Y into bounded pieces. Consider the piece of Y in $[0, \zeta_1]$. The value of increment ΔY_1 should be bounded between $[-L, L]$ by the definition of ζ_1. Denote $\varsigma = \inf\{t \geq 0 : |W_t| \geq L\}$. Theorem 4.1 states that the likelihood ratio of $(\Delta\zeta_1, \Delta Y_1)$ with respect to (ς, W_ς) is proportional to

$$E\left[\exp\left(-\int_0^\varsigma \phi(W_u)du\right)\Big|\varsigma, W_\varsigma\right].$$

Note that, prior to ς, the path of W should be bounded in $[-L, L]$. The continuity of the function ϕ is good enough to make sure both of $\min_{y\in[-L,L]}\phi(y)$ and $\max_{y\in[-L,L]}\phi(y)$ are finite. Hence $\phi(W_u)$ will be sandwiched between these two:

$$\min_{y\in[-L,L]}\phi(y) \leq \phi(W_u) \leq \max_{y\in[-L,L]}\phi(y),\ \text{for } u \leq \varsigma.$$

This observation allows us to find the intensity for the homogeneous Poisson process used in ARM to generate $(\Delta\zeta_1, \Delta Y_1)$ from the standard Brownian motion candidates (ς, W_ς). It is clear to see the advantage of localization from the preceding discussion: we need only the continuity of ϕ to make the algorithm feasible, which is much weaker than the conditions imposed by the Beskos-Roberts algorithm and satisfied by many SDEs with financial applications. Given (ζ_1, Y_{ζ_1}), the law of the increment $\{Y_t - Y_{\zeta_1}, t \geq \zeta_1\}$ is independent of the part before ζ_1, thanks to the Markov property of Y. We then repeat the same approach to generate $(\Delta\zeta_2, \Delta Y_2)$. The whole procedure stops until Y_T is reached.

The following theorem lays out the theoretical foundation of our localization algorithm. It represents the likelihood ratio (more rigorously, the Radon-Nikodym derivative) of the joint distribution function of $(\Delta\zeta_i, \Delta Y_i)$ in terms of the standard Brownian motion. Note that we successfully drop off the Novikov's condition (9) from it.

THEOREM 4.1 *Suppose Assumptions 3.1-3.3 are satisfied. Given $Y_{\zeta_{i-1}} = x, \zeta_{i-1} = t$, the Radon-Nikodym derivative of the law of $(\Delta\zeta_i, \Delta Y_i)$ with respect to that of $(\varsigma, W_{\varsigma\wedge(T-t)})$ is given by*

$$\begin{aligned} & \frac{P[\Delta\zeta_i \in ds, \Delta Y_i \in dy | Y_{\zeta_{i-1}} = x, \zeta_{i-1} = t]}{P[\varsigma \in ds, W_{\varsigma\wedge(T-t)} \in dy]} \\ = \quad & \exp(-A(x)) \cdot \exp(A(x+y)) \cdot E\left[\exp\left(-\int_0^{s\wedge(T-t)} \phi(x+W_u)du\right) \Big| \varsigma = s, W_{\varsigma\wedge(T-t)} = y\right]. \end{aligned}$$

Based on this theorem, we present the following algorithm to generate (ζ_i, Y_{ζ_i}) from a given pair of $(\zeta_{i-1}, Y_{\zeta_{i-1}}) = (t, x)$.

Localization Algorithm in a Nutshell

1. Propose $(\varsigma, W_{\varsigma\wedge(T-t)})$ from a standard Brownian motion.
2. Simulate an indicator binomial r.v. I whose distribution satisfies
$$P[I=1] = 1 - P[I=0] \propto \exp(A(x + W_{\varsigma\wedge(T-t)})) \cdot E\left[\exp\left(-\int_0^{\varsigma\wedge(T-t)} \phi(x+W_u)du\right) \Big| \varsigma, W_{\varsigma\wedge(T-t)}\right]$$
3. If $I = 1$, let $\Delta\zeta_i = \varsigma$ and $\Delta Y_i = W_{\varsigma\wedge(T-t)}$. Otherwise, go to Step 1.
4. If $\Delta\zeta_i < T - t$, update
$$\zeta_i := \zeta_{i-1} + \Delta\zeta_i, \quad Y_{\zeta_i} := Y_{\zeta_{i-1}} + \Delta Y_i.$$
Go to step 1 to restart the algorithm.
5. If $\Delta\zeta_i > T - t$, stop the algorithm and output
$$Y_T := Y_{\zeta_{i-1}} + \Delta Y_i.$$

Now we shall specify the implementation details for each step of the preceding localization algorithm. Subsections 4.1 and 4.2 are devoted to Step 1 and Subsection 4.3 focuses how to generate the Bernoulli I as required by Step 2. In all of these three steps, we need to compare a number with an infinite sum. A special structure of the sums lends an efficient comparison method. We present the related discussion in Subsection 4.4.

4.1 Sampling of ς. Denote $\varsigma_1 = \inf\{t \geq 0 : |W_t| = 1\}$. The self-similarity property of the standard Brownian motion implies that $\{LW_{t/L^2}, t \geq 0\}$ is identical with $\{W_t, t \geq 0\}$ in distribution (Karatzas and Shreve [35, Lemma 2.9.4]). Therefore, the sampling of ς can be accomplished by letting $\varsigma := L^2\varsigma_1$ if we know how to obtain ς_1.

The probability density function of ς_1 is known explicitly through the following infinite series:

$$g_{\varsigma_1}(t) = \frac{1}{\sqrt{2\pi t^3}} e^{-\frac{1}{2t}} + \sum_{j=1}^{\infty} \frac{(-1)^j}{\sqrt{2\pi t^3}} \left[(2j+1)\exp\left(-\frac{(2j+1)^2}{2t}\right) - (2j-1)\exp\left(-\frac{(2j-1)^2}{2t}\right)\right] \tag{12}$$

(Karatzas and Shreve [35, Exercise 2.8.11]). Burq and Jones [15] obeserve that the right hand side of (12) is bounded above by a Gamma distribution density. More specifically, there must exist constants a, b, λ such that

$$g_{\varsigma_1}(t) \leq a g(t; b, \lambda) := a \cdot \frac{\lambda^b t^{b-1} e^{-\lambda t}}{\Gamma(b)}$$

for all $t > 0$, where

$$\Gamma(b) = \int_0^{+\infty} t^{b-1} e^{-t} dt.$$

They also recommend that the optimal choices for the parameters should be $a = 1.243707$, $b = 1.088870$, and $\lambda = 1.233701$ via some numerical experiments.

Based on this fact, we can use ARM to obtain samples of ς_1. Keep generating $U \sim U(0,1)$ and $V \sim Gamma(b,\lambda)$; accept $\varsigma_1 = V$ until

$$U < \frac{g_{\varsigma_1}(V)}{ag(V;b,\lambda)}. \tag{13}$$

The exact evaluation of $g_{\varsigma_1}(V)$ is impossible because it involves with an infinite sum. However, we still can conduct the comparison in (13) exactly due to a special structure of the summation. This will be illustrated in Section 4.4. Burq and Jones [15] prove that the expected number of operations to simulate one sample of ς_1 is finite.

4.2 Sampling of $W_{\varsigma\wedge(T-t)}$. Given $\varsigma < T-t$, sampling $W_{\varsigma\wedge(T-t)}$ is equivalent to sampling W_ς. Note that

$$P[W_\varsigma = L] = P[W_\varsigma = -L] = \frac{1}{2} \tag{14}$$

by the symmetric property of Brownian motion. We can easily generate $W_{\varsigma\wedge(T-t)}$ in this case from the observation 14.

Now, consider the simulation of $W_{\varsigma\wedge(T-t)}$ if we get $T-t < \varsigma$, or more generally, generating the values of W at a sequence of instances: $0 < t_1 < \cdots < t_n < \varsigma$. Thanks to the self-similarity property of the standard Brownian motion again, we focus our attention on the case when $L = 1$ only. It is easy to obtain the samples for a general L if we note that the following relationship holds:

$$(W_{t_1}, \cdots W_{t_n} | \varsigma, W_\varsigma) \stackrel{d}{=} L \cdot (W_{t_1/L^2}, \cdots W_{t_n/L^2} | \varsigma_1, W_{\varsigma_1}).$$

Given ς_1 and W_{ς_1}, the law of $(W_t, 0 \le t < \varsigma_1)$ turns out to be related to Brownian bridges too. Without loss of generality, assume that $W_{\varsigma_1} = -1$ first. The Brownian motion W achieves its minimum in $[0,\varsigma_1]$ at ς_1. A well-known result in probability theory — the Williams path decomposition, named after the work of Williams [46] — states that the part of a standard Brownian motion prior to its minimum should behave like a time-reverse Brownian meander. Moreover, as noted by Imhof [31], a Brownian meander can be further decomposed into three independent Brownian bridges. A symmetric observation applies for the case when $W_{\varsigma_1} = 1$. More formally, we introduce a new process such that, for $t \le \varsigma_1$,

$$\widetilde{W}_t := \begin{cases} 1 - W_{\varsigma_1 - t}, & \text{if } W_{\varsigma_1} = 1; \\ 1 + W_{\varsigma_1 - t}, & \text{if } W_{\varsigma_1} = -1. \end{cases} \tag{15}$$

Denote B^i_t, $i = 1,2,3$, to be three independent Brownian bridges from 0 to 0 on $[0,\varsigma_1]$. Let

$$\widetilde{B}_t := \sqrt{(t/\varsigma_1 + B^1_{\varsigma_1 - t})^2 + (B^2_{\varsigma_1 - t})^2 + (B^3_{\varsigma_1 - t})^2}. \tag{16}$$

The following theorem shows the explicit form of the likelihood ratio of $\widetilde{W}$ with respect to $\widetilde{B}$.

THEOREM 4.2 *Let $y_{n+1} = 1$. For any $0 = t_0 < t_1 < \cdots < t_n < t_{n+1} = \varsigma_1$ and $y_1, \cdots, y_n > 0$, the joint conditional distribution of $(\widetilde{W}_{t_1}, \cdots \widetilde{W}_{t_n})$ has the following likelihood ratio with respect to $(\widetilde{B}_{t_1}, \cdots \widetilde{B}_{t_n})$:*

$$\frac{P[\widetilde{W}_{t_1} \in dy_1, \cdots, \widetilde{W}_{t_n} \in dy_n | \varsigma_1, W_{\varsigma_1}]}{P[\widetilde{B}_{t_1} \in dy_1, \cdots, \widetilde{B}_{t_n} \in dy_n]} = c \prod_{i=1}^{n} p(t_i, y_i; t_{i+1}, y_{i+1}) \cdot q(t_1, y_1) \cdot 1_{\{y_i \in (0,2)\}}.$$

where the constant $c = 1/P[\max_{0 \le t \le \varsigma_1} \tilde{B}_t \le 2]$. Both functions p and q are nonnegative and are bounded above by 1. They both admit closed-form expressions. For any $0 < s < t < \varsigma_1$ and $x, y \in (0,2)$,

$$p(s,x;t,y) \quad = \quad \frac{1 - \sum_{j=1}^{+\infty} (\theta_j - \vartheta_j)}{1 - \exp(-2xy/(t-s))},$$

where

$$\begin{aligned} \theta_j(s,x;t,y) &= \exp\left(-\frac{2(2j-x)(2j-y)}{t-s}\right) + \exp\left(-\frac{2(2(j-1)+x)(2(j-1)+y)}{t-s}\right), \\ \vartheta_j(s,x;t,y) &= \exp\left(-\frac{2j(4j+2(x-y))}{t-s}\right) + \exp\left(-\frac{2j(4j-2(x-y))}{t-s}\right). \end{aligned}$$

And

$$q(s,x) \quad = \quad 1 - \frac{1}{x}\sum_{j=1}^{+\infty}(\rho_j - \varrho_j),$$

where

$$\rho_j(s,x) = (4j-x)\exp\left(\frac{-4j(2j-x)}{s}\right), \quad \varrho_j(s,x) = (4j+x)\exp\left(\frac{-4j(2j+x)}{s}\right).$$

We can use Theorem 4.2 to produce samples of $(W_{t_1},\cdots,W_{t_n})$, conditioning on ς_1 and W_{ς_1}, in the following steps:

Generating $(W_{t_1},\cdots,W_{t_n})$ When ς_1 and W_{ς_1} Are Given

1. Generate Brownian bridges $(B^i_{t_1},\cdots,B^i_{t_n})$ for $i=1,2,3$.
2. Use (16) to construct $(\widetilde{B}_{\varsigma_1-t_n},\cdots,\widetilde{B}_{\varsigma_1-t_1})$.
3. Generate $n+1$ uniforms $U_j \sim U(0,1)$, $0\le j\le n$.
4. Accept $(\widetilde{B}_{\varsigma_1-t_n},\cdots,\widetilde{B}_{\varsigma_1-t_1})$ as a sample of $(\widetilde{W}_{\varsigma_1-t_n},\cdots,\widetilde{W}_{\varsigma_1-t_1})$ when $\widetilde{B}_{\varsigma_1-t_j}<2$,
$$U_0 < q(\varsigma_1-t_n,\widetilde{B}_{\varsigma_1-t_n}) \text{ and } U_j < p(\varsigma_1-t_j,\widetilde{B}_{\varsigma_1-t_j};\varsigma_1-t_{j+1},\widetilde{B}_{\varsigma_1-t_{j+1}}) \tag{17}$$
for $1\le j\le n$.
5. Obtain $(W_{t_1},\cdots,W_{t_n})$ from $(\widetilde{W}_{\varsigma_1-t_n},\cdots,\widetilde{W}_{\varsigma_1-t_1})$ by (15).

Note that we also have to deal with comparisons involving infinite sums in (17). Like (13), exact evaluation of the infinite sums in functions p and q can be avoided if we take advantage of a special structure of these sums. The detailed discussion is in Section 4.4.

4.3 Sampling of the Bernoulli *I*. Once we have $(\varsigma, W_{\varsigma\wedge(T-t)})$, we need a Bernoulli I to decide whether or not to accept this proposed candidate pair. The indicator I should satisfy

$$P[I=1|\varsigma,W_{\varsigma\wedge(T-t)}] \propto \exp(A(x+W_{\varsigma\wedge(T-t)}))\cdot E\left[\exp\left(-\int_0^{\varsigma\wedge(T-t)}\phi(x+W_u)du\right)\Big|\varsigma,W_{\varsigma\wedge(T-t)}\right].$$

As illustrated at the beginning of this section, introduction of the stopping time ς enables us to use the idea of stochastic thinning to accomplish this task. By the definition of ς, we know $W_u\in[-L,L]$ for all $0\le u\le\varsigma$. Define

$$m := \min_{y\in[-L,L]}\phi(x+y), \quad M := \max_{y\in[-L,L]}\phi(x+y) \quad \text{and} \quad \Gamma = \max_{y\in[-L,L]}\exp(A(x+y)).$$

Theorem 4.3 shows that we can obtain such I if we go through the following steps:

Sampling of the Bernoulli I

1. Simulate a homogeneous Poisson process with intensity $M-m$ to obtain arrivals
$$0\le\tau_1<\tau_2<\cdots<\tau_N\le(T-t)\wedge\varsigma.$$
2. Use the method presented in Section 4.2 to generate $(W_{\tau_1},\cdots,W_{\tau_N})$ and meanwhile simulate $N+1$ uniforms $U_j\sim U(0,1)$, $0\le j\le N$ and another independent uniform $V\sim U(0,1)$.
3. Let $I=1$ if and only if the following three inequalities hold simultaneously:
$$U_0\le\frac{\exp(A(x+W_{(T-t)\wedge\varsigma}))}{\Gamma},\ U_j<\frac{M-\phi(x+W_{\tau_j})}{M-m},\ \text{and } V\le\frac{\exp(-m((T-t)\wedge\varsigma))}{\max(1,\exp(-m(T-t)))}$$
for all $1\le j\le N$.

THEOREM 4.3 *The Bernoulli I obtained through the above three-step procedure satisfies*

$$P[I=1|\varsigma,W_{\varsigma\wedge(T-t)}] = \frac{\exp(A(x+W_{\varsigma\wedge(T-t)}))}{\Gamma\max(1,\exp(-m(T-t)))}\cdot E\left[\exp\left(-\int_0^{\varsigma\wedge(T-t)}\phi(x+W_u)du\right)\mid\varsigma,W_{\varsigma\wedge(T-t)}\right].$$

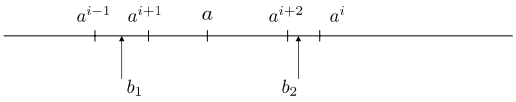

Figure 2: The sequence $\{a_i\}$ oscillates around its limit a. $b_1 < a$ iff $b_1 < a_{i+1} < a_i$ for some i; $b_2 > a$ iff $b_2 > a_{i+1} > a_i$ for some i.

From this theorem, we can argue for the validity of the localization method as follows. By Theorems 4.1 and 4.3, we have

$$P[I = 1|\varsigma = s, W_{\varsigma\wedge(T-t)} = y] = \frac{P[\Delta\zeta_i \in ds, \Delta Y_i \in dy|Y_{\zeta_{i-1}} = x, \zeta_{i-1} = t]}{cP[\varsigma \in ds, W_{\varsigma\wedge(T-t)} \in dy]},$$

where c is a constant given by

$$c = \exp(\max_{y\in[-L,L]} A(x+y) - A(x)) \cdot \max(1, \exp(-m(T-t))).$$

The general principle of ARM in Section 3 implies that our method will generate exact samples of $(\Delta\zeta_i, \Delta Y_i)$.

4.4 Comparison Involving Infinite Sums. At several points in the previous subsections, we have to compare a number and an infinite sum (cf. (13) and (17)). One special structure of the sums turns out to be very useful for constructing an efficient method without exactly valuing these infinite sums. They are all *oscillating* in the sense of the following definition:

Definition 4.1 *A sequence of numbers $\{a_i, i \geq 1\}$ is called to be oscillating if there exists a positive integer N such that for $i \geq N$,*

$$0 < -\frac{a_{i+1} - a_i}{a_i - a_{i-1}} < 1.$$

Suppose that the limit of an oscillating sequence $\{a_i, i \geq 1\}$ exists and is equal to a. We are willing to compare a with another number b. As illustrated in Figure 2, we have

$$a_i < a_{i+1} < b \Rightarrow a < b, \quad \text{and} \quad a_i > a_{i+1} > b \Rightarrow a > b \tag{18}$$

for some i. From this fact, we can easily see that, as long as $a \neq b$, the comparison of a and b can be finished in finite steps of calculation and without knowing the exact value of a. Both comparisons (13) and (17) involve summations of oscillating series. Let

$$g_J(t) := \frac{1}{\sqrt{2\pi t^3}} e^{-\frac{1}{2t}} + \sum_{j=1}^{J} \frac{(-1)^j}{\sqrt{2\pi t^3}} \left[(2j+1)\exp\left(-\frac{(2j+1)^2}{2t}\right) - (2j-1)\exp\left(-\frac{(2j-1)^2}{2t}\right)\right].$$

Define

$$p_{2J}(s,x;t,y) = \sum_{j=1}^{J}(\theta_j - \vartheta_j), \quad p_{2J+1}(s,x;t,y) = p_{2J}(s,x;t,y) + \theta_{J+1}$$

and

$$q_{2J}(s,x) = \sum_{j=1}^{J}(\rho_j - \varrho_j), \quad q_{2J+1}(s,x) = q_{2J}(s,x) + \rho_{J+1}$$

for all $J \geq 1$. Appendix C verifies that all of these three sequences are actually oscillating and their limits exist such that

$$\lim_{J\to+\infty} g_J(t) = g_{\varsigma_1}(t), \quad \lim_{J\to+\infty} p_J(s,x;t,y) = p_\infty(s,x;t,y), \quad \text{and} \quad \lim_{J\to+\infty} q_J(s,x) = q_\infty(s,x).$$

Exploiting the oscillation property, the following proposition shows that we can finish the comparison between a number and $q(s,x)$ in finite steps. A similar result holds for the other two sequences. We skip the detailed discussion for the interest of space.

PROPOSITION 4.1 *Given $U \in U(0,1)$, $s > 0$, and $x \in (0,2)$, suppose that N is a minimal integer such that*

$$N > \frac{(\log 4)s}{8}\frac{1}{x} + 2 \quad \textit{and} \quad 6(N+1)\exp(-8N^2/s) < |x(1-U) - q_\infty(s,x)|.$$

Then, at most $2N$ terms, $\{q_1(s,x), \cdots, q_{2N}(s,x)\}$, need to be calculated before finishing the comparison between U and $q(s,x)$.

For any given ς_1, $\widetilde{B}_{\varsigma_1 - t_n}$, and U, the preceding proposition indicates apparently that we do not need to evaluate the exact value of $q_\infty(\varsigma_1 - t_n, \widetilde{B}_{\varsigma_1 - t_n})$ in order to perform the comparison in (17). The partial sum of the first $2N$ terms of the sequence of $\{q_J\}$ is sufficient for us to make a judgement. Clearly, Proposition 4.1 also implies that N tends to $+\infty$ when ς_1 is extremely large or $\widetilde{B}_{\varsigma_1 - t_n}$ is very close to 0. However, we can prove that the probabilities of both events are negligible. In conjunction with this observation, the expected number of operations to finish the comparisons in (13) and (17) is shown to be finite. We skip the detailed proof of this conclusion in this paper.

The above method can be regarded as an application of the alternating series algorithm in the simulation literature, a general method to simulate non-uniform random variables (see, e.g., Devroye [22, §IV.5]). The method is to deal with simulation of a random number with a density function f that can be approximated from above and below by sequences of functions f_n and g_n. In particular, assume that (i) $\lim_{n\to+\infty} f_n = f$ and $\lim_{n\to+\infty} g_n = f$; (ii) $f_n \le f \le g_n$; (iii) $f \le ch$ for a constant $c \ge 1$ and an easy density h. The alternating series algorithm suggests that we can generate such random variable X through ARM. In our setting, the oscillating property helps us easily find the squeezing sequences to bound g_{ς_1}, p_∞, and q_∞ from above and from below. For instance, q_∞ is sandwiched by the sequences of $\{q_{2J}\}$ and $\{q_{2J+1}\}$.

5. Discussions. We discuss the efficiency of our exact simulation and its possible extension to those SDEs with boundaries in this section. Our aim is not to present here a complete analysis, which is left for future exploration, but to raise some general advices on the algorithm implementation.

5.1 One Extension: SDEs with Boundaries In Section 4, we fix the localization parameter L for the whole simulation procedure. This fixation is not critical at all for the feasibility of our algorithm. Thanks to the Markovian property of the SDE (7), we may use different L_i in simulating different $(\Delta\zeta_i, \Delta Y_i)$. Such adaptive selection of L is very appealing in simulating SDEs with boundaries such as the CIR model (1). Consider a process Y with the range $D_Y = (\underline{y}, +\infty)$ for a finite $\underline{y}$. As shown by the numerical examples in Section 6, the function $\phi(y)$ will diverge to ∞ as $y \downarrow \underline{y}$ for many SDEs with such kind of D_Y. This makes it difficult to find a uniformly stronger intensity for the homogeneous Poisson process if we run the Besko-Roberts exact simulation scheme.

However, we can still use the localization idea to circumvent the obstacle. Rather than sticking to a fixed constant L to do the job over time, we change it adaptively as the generation of (ζ_i, Y_{ζ_i}) proceeds. Figure 3 illustrates the procedure with an adaptive localization scheme. Start with an L_1 such that $-L_1 > \underline{y}$. Use the same technique presented in Section 4 to obtain (ζ_1, Y_{ζ_1}). When $Y_{\zeta_1} = -L_1$, the localization width L_1 is not suitable any more for generating (ζ_2, Y_{ζ_2}) because such choice would lead to a possibility that $Y_{\zeta_2} = Y_{\zeta_1} - L_1 = -2L_1 < \underline{y}$, which is shown to be impossible by the unattainability of $\underline{y}$. Applying the idea of adaptive localization, we may replace L_1 by a smaller L_2 to ensure that $-L_1 - L_2 > \underline{y}$ to keep the simulated process away from the boundary. Continue the procedure until we reach time T. How to find an optimal rule of change-of-L to achieve the best performance of such extension is left for future investigation. In the numerical experiments in Section 6, we use a very simple rule such that update $L_{i+1} = (Y_{\zeta_i} - \underline{y})/2$ for the generation of $Y_{\zeta_{i+1}}$, after simulating Y_{ζ_i}. This rule of thumb performs very well in the experiments.

5.2 Simulation Efficiency. Our localization algorithm demonstrates a nested structure: the inner loop generates candidate pairs of $(\varsigma, W_{\varsigma\wedge(T-t)})$ and the Bernoulli I using the ARM schemes introduced in Subsections 4.1 and 4.2; the outer loop decides whether or not accept the proposed candidates of $(\varsigma, W_{\varsigma\wedge(T-t)})$ as an output of $(\Delta\zeta, \Delta Y)$. The efficiency of this algorithm is determined by how fast this nested ARM scheme can be accomplished.

Let us focus on the inner loop first. By Burq and Jones [15], we know that we can generate one sample

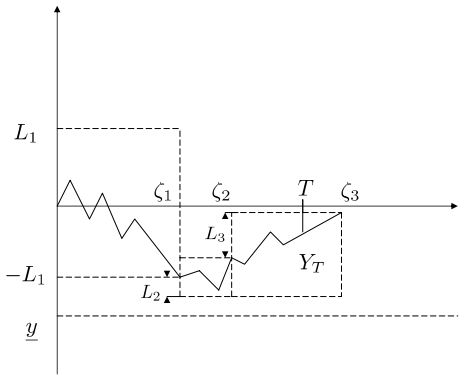

Figure 3: Barrier L can be adjusted adaptively in the simulation of an SDE with $D_Y = (\underline{y}, +\infty)$. This figure shows that we change to a smaller L_2 when Y approaches to the boundary $\underline{y}$.

of ς in finite steps. To analyze the computational cost for generating $W_{\varsigma\wedge(T-t)}$ and I, observe that the bottleneck is the ARM scheme that we use to produce the intermediate variables $\{\widetilde{W}_{\tau_1}, \cdots, \widetilde{W}_{\tau_N}\}$. Theorem 4.2 indicates that the expected number of trials needed to generate a qualified set of $\widetilde{W}_{\tau_i}$'s equals $1/P[\max_{0\le t\le \varsigma_1} \widetilde{B}_t \le 2]$. Within each trial, the required computational cost is proportional to how many τ_i's are simulated by the homogeneous Poisson process in the algorithm of sampling I in Subsection 4.3. Therefore, the expectation of the total computational cost to simulate $W_{\varsigma\wedge(T-t)}$ and I should be proportional to

$$(T-t)(M-m)\cdot \frac{1}{P[\max_{0\le t\le \varsigma_1} \widetilde{B}_t \le 2]},$$

where $M-m$ is the intensity of the homogeneous Poisson process and $(T-t)(M-m)$ is the average number of τ_i's. The proof of Theorem 4.2 shows that $P[\max_{0\le t\le \varsigma_1} \widetilde{B}_t \le 2] = q(\varsigma_1, 1)$. Some numerical experiments we conduct point out that $E[1/q(\varsigma_1, 1)] = 1.5425$. Therefore, the efficiency of the inner loop is determined mainly by the value of $M-m$, or in other words, the fluctuation magnitude of ϕ in the interval $[x-L, x+L]$. Assumptions 3.1-3.3 ensure that the function ϕ is locally Lipschitz and thus we expect a finite computational cost for the inner loop.

We then proceed to discuss the efficiency of the outer loop. According to Theorem 4.1, on average, we need to repeat $1/P[I=1]$ inner loops to generate a qualified pair of $(\Delta\zeta, \Delta Y)$. Note that

$$P[I=1|W_{\varsigma\wedge(T-t)}, \varsigma] \ge \exp\{-[\max_{y\in[-L,L]} A(x+y) - \min_{y\in[-L,L]} A(x+y)]\}\cdot \frac{\exp\{-M(T-t)\}}{\max(1, \exp\{-m(T-t)\})}. \quad (19)$$

Therefore, as long as functions A and ϕ do not fluctuate too much in the interval $[x-L, x+L]$, the right hand side of (19) will have a uniform lower bound and it implies that the number of inner loops needed is expected to be finite. Combining with the fact that the average computational cost in each inner loop is finite, we conclude that our localization algorithm will finish the simulation of one pair of $(\Delta\zeta, \Delta Y)$ in finite operations. The preceding discussion also points out that two quantities,

$$\max_{y\in[-L,L]} A(x+y) - \min_{y\in[-L,L]} A(x+y) \quad \text{and} \quad \max_{y\in[-L,L]} \phi(x+y) - \min_{y\in[-L,L]} \phi(x+y)$$

play an important role in determining how fast the whole procedure will be: smaller values for both differences lead to faster execution of the algorithm.

Finally, we show that, for any SDE satisfying Assumptions 3.1-3.3, the localization algorithm presented in Section 4 needs only finite $(\Delta\zeta, \Delta Y)$ in Step 4 to reach Y_T.

THEOREM 5.1 *Assume that Assumptions 3.1-3.3 are satisfied. Let* $J = \inf\{j : \sum_{i=1}^{j} \Delta\zeta_i > T\}$. *Then,* $J < +\infty$ *almost surely.*

Note that the discussion before Theorem 5.1 indicates that the computational effort is finite for generating each update of $(\Delta\zeta, \Delta Y)$. In conjunction with this theorem, we finish the efficiency analysis to claim that the localization algorithm will generate samples of Y_T with finite number of operations.

The above discussion does not exclude the possibility that we may need to wait for a very long time to obtain one sample of Y_T, although the time is shown to be finite by Theorem 5.1. Recall that the required computational time in our rejection-based sampling method is correlated with the values of proposed candidates. Particularly, we change the localization width L in the extension presented in Subsection 5.1 as the simulation proceeds. When the simulated process Y_{ζ_i} approaches the lower boundary $\underline{y}$ in Figure 3, a small L will be used and that may "trap" our simulation around $\underline{y}$ for a long time before it reaches T. Sometimes, such long waiting time would lead to another potential source of bias: users may have no patience to wait until the final result comes out and abort the algorithm at an early stage. It is referred to as the "user-impatience bias" commonly in the literature of perfect simulation; see Propp and Wilson [43]. The numerical experiments in the next section illustrate that the impact of such bias is limited. We leave further investigation in this direction for future research.

Another possible direction for future research is to explore the effect of L on the efficiency. On the one hand, a larger L will leads to a larger first-passage time ς, and in turn, smaller amount of intermediate steps to reach the target time T. On the other hand, we need to simulate more W_{τ_i}'s to preform ARM if L is larger. In light of this trade-off, an appropriate choice of L will yield the best performance in terms of computational time, although the sampling under any L is exact according to Section 4. One may run small-scale pilot experiments to obtain the optimal L empirically.

6. Numerical Examples. In the section, we present some numerical results of the exact simulation on four models: the Ornstein-Uhlenbeck (OU) mean-reverting process, a double-well potential model, the Cox-Ingersoll-Ross square root process, and the linear-drift CEV-type-diffusion model. The former two models have ranges $D_X = (-\infty, \infty)$ and the last two $D_X = (0, +\infty)$.

6.1 Ornstein-Uhlenbeck Mean-Reverting Process Consider an Ornstein-Uhlenbeck process

$$dS_t = -bS_t dt + dW_t,\ 0 < t < T$$

with b is a positive constant. Note that the drift is positive when $S_t < 0$ and negative when $S_t > 0$. Thus, S is pulled toward level 0, a property generally referred to as mean reversion. A variation of this model is used by Vasicek [45] to model short rates in the interest market. The marginal distribution of S_T is normal with mean $S_0 \exp(-bT)$ and variance $(1 - \exp(-2bT))/2b$. It is easy to see that we do not need perform the Lamperti transform for this process because its diffusion coefficient is already a unit. In addition, $\alpha(S) = -bS$, $\phi(S) = ((bS)^2 - b)/2$, and $A(S) = -bS^2/2$. The function ϕ has a lower bound $-b/2$ but does not have a uniform upper bound.

Table 1 demonstrates a comparison of the exact simulation and the Euler discretization scheme for a particular case when $T = 1$, $b = 1$, and $S_0 = 0$. Here, we estimate $E[S_T^2]$ using the sample averages and

Euler Scheme					Exact Simulation			
Sample	Step	Estimation	RMSE	RV Simulated	Sample	Estimation	RMSE	RV Simulated
400	16	0.4571	0.0408	6400	65	0.4594	0.0632	3018
1600	32	0.4434	0.0173	51200	520	0.4334	0.0274	24905
6400	64	0.4359	0.0095	409600	4200	0.4349	0.0097	201974
25600	128	0.4353	0.0040	3276800	33600	0.4324	0.0031	1607682
Avg. reduction			0.475				0.363	

Table 1: A numerical comparison of the Euler scheme and exact simulation in an OU Process. The RMSE is calculated on the basis of 100 trials. The reduction factor is defined as a ratio of the RMSEs of two consecutive rows. The columns of *RV Simulated* records how many random variables are generated to obtain the corresponding RMSE.

calculate its error with respect to the true value $(1 - \exp(-2))/2 = 0.4323$. The first 5 columns show the results of the Euler scheme. To improve the precision of the estimation, more computational effort is needed to increase the number of time steps within each path and the total number of sample paths simultaneously. Following the Duffie-Glynn rule, we need to double the number of time steps (Column

2) and quadruple the total number of simulated paths (Column 1) at the same time. The total work load therefore increases by eight times from one row to the next, which is reflected in Column 5, the total number of random variables needed to simulate. The level of RMSE (Column 4) shrinks by a factor of almost 1/2. This is consistent with the conclusion of Duffie and Glynn [21] that the optimal RMSE for the Euler schemes is in the order of $O(t^{-1/3})$ if we generate $O(t^{1/3})$ time steps in each sample path and $O(t^{2/3})$ paths. The last four columns display the results from our exact simulation method. It is obvious that the RMSE decreases by a factor of almost $8^{-1/2} = 0.354$ as the number of generated samples increases by eight times since our method is unbiased. The proposed algorithm achieves the order of $O(t^{-1/2})$, the optimal convergence rate of RMSE for Monte Carlo simulation.

Figure 4 is the empirical density of S_T obtained from our exact simulation. It coincides with the theoretical density of S_T perfectly after 1 million sampling. The Kolmogorov-Smirnov goodness-of-fit test also suggests strongly that there is no significant difference between the empirical probability density of the samples generated by the exact simulation and the true density of S_T. On the other hand, we need a large number of time steps for the Euler scheme to generate samples with a comparable quality.

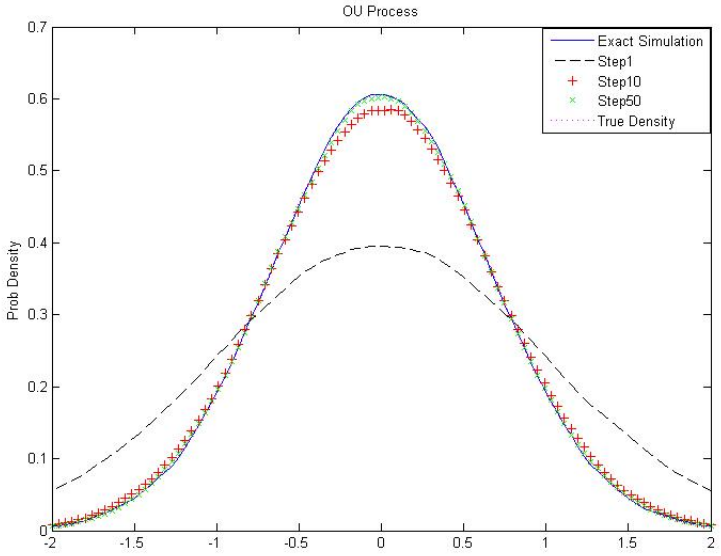

Figure 4: The empirical probability densities of S_T generated from the exact simulation and the Euler scheme against its theoretical density. $S_0 = 0$ and $b = T = 1$. One million accepted samples were simulated in our exact simulation. The Kolmogorov-Smirnov goodness-of-fit test justifies the outcomes from the exact simulation. The KS statistic is 0.7314 and its corresponding p-value equals 0.6486. For the purpose of comparison, we simulate the Euler scheme with the number of time steps $N = 1$, $N = 10$, and $N = 50$. The computation time for 1 million samples under our exact simulation is 3,624 seconds. The configuration of the PC we used is Intel Core 2, 2.66GHz, and 2.87GB of Ram. The version of Matlab is Matlab 7.12.0(R2011a).

6.2 A Double-Well Potential Model The following model is artificially made to illustrate the ability of the exact simulation to deal with transitional densities with strong non-normality. Suppose that X follows

$$dX_t = (X_t - X_t^3)dt + dW_t.$$

The value range of this model is $D_X = (-\infty, +\infty)$. We have $\phi(y) = (1 - 2y^2 - 2y^4 + y^6)/2$ and $A(y) = y^2/2 - y^4/4$. It is easy to show there are two modes: $x = \pm 1$ in the long-run stationary distribution of this process. Hence its transition distribution significantly differs from normal distributions. Figure 5 plots the empirical transition densities of X_T given $X_0 = 0.5$ and $X_0 = 0$, respectively, with $T = 1/2$. Both the pictures exhibit a strong non-normality of the transition density of the process. No matter which value the process starts from initially, the exact simulation method can capture the non-normality immediately. However, the Euler schemes are not able to deliver accurate outcomes unless we considerably increase the number of discretization steps.

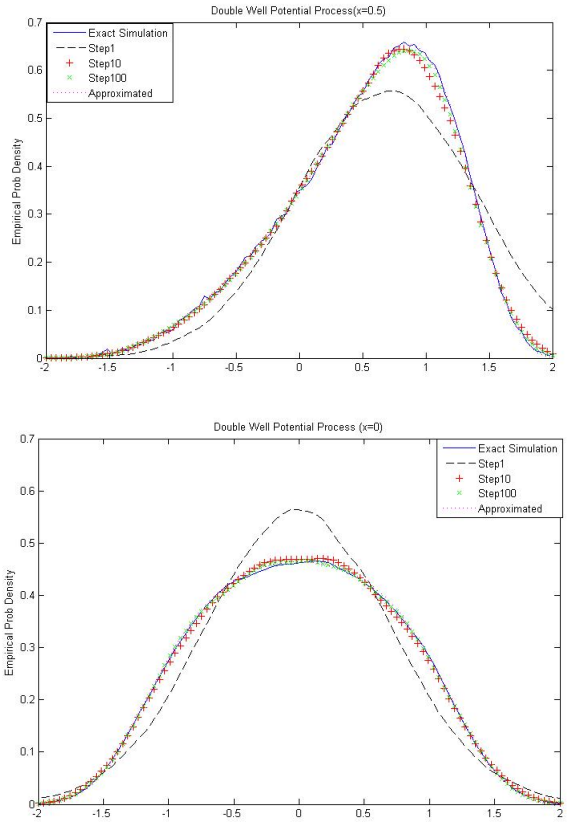

Figure 5: The empirical transitional densities for a double-well potential process under the exact simulation and the Euler scheme. The initial values are $x = 0.5$ and $x = 0$ for the top and bottom panels, respectively. For the purpose of comparison, we use an approximation proposed by Aït-Shahalia [5] as the true density of X_T. It is apparent that the outputs from the Euler schemes tend to what is obtained from the exact simulation as the discretization is increasingly fine. The computation times for 1 million samples under our exact simulation are 4,259 and 4,385 seconds, respectively. The configuration of the PC we used is Intel Core 2, 2.66GHz, and 2.87GB of Ram. The version of Matlab is Matlab 7.12.0(R2011a).

6.3 Cox-Ingersoll-Ross Square-Root Process Consider the sampling of the CIR process (5) in this subsection. Given $X_0 = x$, the transition distribution of X_T is known as (Feller [23] and Cox, Ingersoll and Ross [18]):

$$X_T \stackrel{d}{=} \frac{\sigma^2(1-e^{-bT})}{4b} \cdot \chi^2_d\left(\frac{4be^{-bT}}{\sigma^2(1-e^{-bT})}x\right),$$

where $\chi^2_d(\lambda)$ is a noncentral chi-square random number with $d = 4a/\sigma^2$ degrees of freedom and noncentral parameter λ. Under the corresponding Lamperti transform $F(y) = 2(\sqrt{y} - \sqrt{x})/\sigma$, the transformed process $Y_t = F(X_t)$ then follows the SDE:

$$dY_t = \alpha(Y_t)dt + dW_t = \left[\left(\frac{4a-\sigma^2}{2\sigma^2}\right)\frac{1}{(Y_t + 2\sqrt{x}/\sigma)} - \frac{b}{2}\left(Y_t + \frac{2\sqrt{x}}{\sigma}\right)\right]dt + dW_t,\ Y_0 = 0.$$

Because $X_t \geq 0$, Y_t should be larger than $\underline{y} = -2\sqrt{x}/\sigma$ for $t \geq 0$. Assumption 3.3 imposes a constraint on the process parameters that $2a/\sigma^2 > 3/2$ to ensure that $\alpha(y) \geq 1/(y - \underline{y})$. Figure 6 illustrates the results of our algorithm in the simulation of a case when $a = 0.07$, $b = 0.2\overline{5}$, and $\sigma = 0.3$. The exact simulation yields a perfect consistence with the true probability density of the CIR model.

In Subsection 5.2, we point out that the efficiency of our exact simulation may be under the influence of the user-impatience bias, in particular, in the simulation of SDEs with boundaries. The following experiment is designed to investigate this issue. Let $2a/\sigma^2 = 1.50001$, $X_0 = 0.001$, and $T = 10$. Apparently, these values are very extreme to model the real financial market. The purpose of such parameter choices is to make the process have a larger chance to stay near the lower boundary so that we have

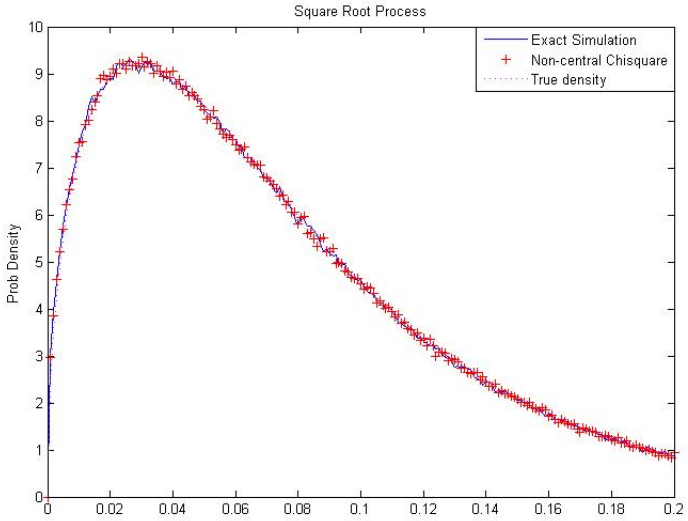

Figure 6: The exact simulation of the CIR model. The picture shows the empirical density function generated by our exact simulation against the theoretical density of the CIR model. $X_0 = 0.02$, $a = 0.07$, $b = 0.25$, $\sigma = 0.3$, and $T = 1$. For the purpose of comparison, we also generate samples from the true density of the CIR model. The three density curves are indistinguishable. We employ the Kolmogorov-Smirnov goodness-of-fit test for validation. The KS statistic is 0.6755 and the corresponding p-value is 0.7516, respectively. The computation times for 1 million samples under our exact simulation is 1,745 seconds. The configuration of the PC we used is Intel Core 2, 2.66GHz, and 2.87GB of Ram. The version of Matlab is Matlab 7.12.0(R2011a).

to use many intermediate steps to reach $T = 10$, which might bring us a tremendous computational burden potentially. However, Table 2 demonstrates that even such extreme parameters have only a very limited impact on the efficiency of the exact simulation. As in Subsection 6.1, we use the number of

$1000 \leq RVS < 2000$	$2000 \leq RVS < 3000$	$3000 \leq RVS < 5000$	$RVS \geq 5000$
97801	533	279	307

Table 2: The distribution of RVS to produce 1 M samples of X_{10} when $X_0 = 0.001$, $a = 0.00750005$, $b = 0.2$, $\sigma = 0.1$, and $T = 10$. The average of RVS is 500.58 and the maximum of RVS is 79755.

random variable simulated (RVS) for each sample path as a rough measure of the computational effort. The table records the distribution of RVS to produce 1 million samples of X_{10}. We can see that about 99.9 percents of the samples need less than 2000 RVS. Furthermore, we find in the experiments that the actual computational time is very small even for those samples which needs large RVS. Both observations strongly suggests that the user-impatience will not constitute a serious concern for our method.

We then introduce the effect of the user-impatience bias to the above experiment intentionally by abandoning one sample path in which we need more than 3,000 RVS. We choose 3,000 as the threshold because it is almost equivalent to the workload of an Euler scheme with daily discretization (i.e., $\Delta T = 1/252$) for $T = 10$. We assume that the user will be impatient if the workload from our exact simulation algorithm for one sample path exceeds the workload of a corresponding Euler scheme. Figure 7 shows the outcomes of the probability density estimation changes very little before and after such operation. This is consistent with the observations we made in Table 2. Since we only have a negligible portion of samples in which large simulation effort is needed, whether or not abort those samples will not generate too much difference between both estimations.

Finally, we point out one advantage of the exact simulation over the discretization schemes: they do not distort the distribution around the boundary, while we need to increase the number of time step very significantly for the discretization schemes in avoidance of such distortions even for a very short time horizon. As discussed in Section 2, we cannot apply the vanilla Euler scheme directly on the model because it does not preserve the non-negativeness of the solution. Thus, a variety of truncations are

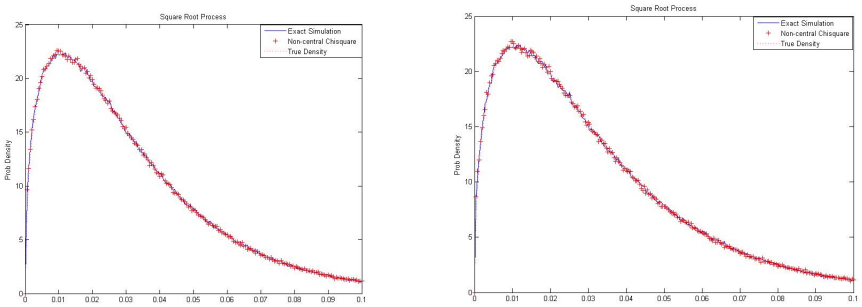

Figure 7: The exact simulation of the CIR model with and without the user-impatience. $X_0 = 0.001$, $a = 0.00750005$, $b = 0.2$, $\sigma = 0.1$, and $T = 10$. Thus, $2a/\sigma^2 \approx 1.5$. For the purpose of comparison, we also generate samples from the true density of the CIR model. The left and right plot shows the outcomes without and with abandonments, respectively. The empirical density estimations in both plots are indistinguishable from the true density of the CIR model.

suggested in the literature to correct this. We use two of them to compare with the result of the proposed exact simulation algorithm. The first is a full truncation method, as suggested by Lord et al. [39]. The scheme can be written in the following form:

$$\widetilde{X}_{t_i} = \widetilde{X}_{t_{i-1}} + (a - b\widetilde{X}^+_{t_{i-1}})\Delta t + \sigma\sqrt{\widetilde{X}^+_{t_{i-1}}}\Delta W_i,$$

where $x^+ = \max(x, 0)$. This is a modified first-order Euler scheme by truncating the generated samples on the origin and it appears to produce the smallest error in all of the schemes reviewed by Lord et al. [39]. The second method is an implicit second-order Milstein scheme in Kahl and Jäckel [33]:

$$\widetilde{X}_{t_i} = \frac{\widetilde{X}_{t_{i-1}} + a\Delta t + \sigma\sqrt{\widetilde{X}_{t_{i-1}}}\Delta W_i + \sigma^2((\Delta W_i)^2 - \Delta t)/4}{1 + b\Delta t}.$$

This discretization scheme is shown to preserve the positivity of the sample paths when $4a > \sigma^2$.

Figure 8 illustrates the results of our algorithm, the full truncation and the implicit Milstein scheme for a case when $T = 1/52$ (one week). The full truncation method produces a non-zero probability mass at 0. The implicit Milstein scheme tends to overestimate the density at its peak and right tail and underestimate the left tail.

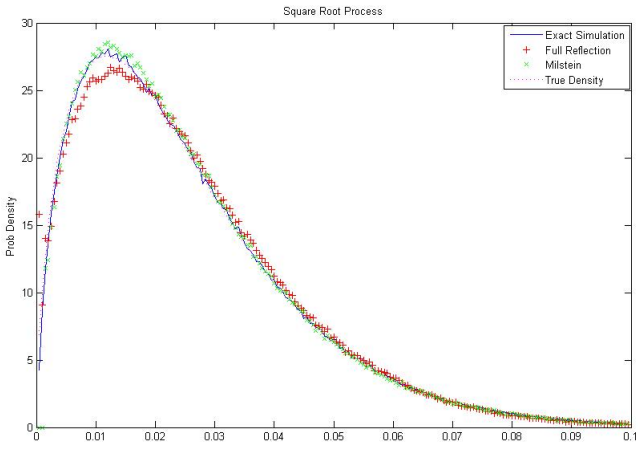

Figure 8: The exact simulation of the CIR model under the exact simulation, the full truncation method and the Milstein scheme. $X_0 = 0.01$, $a = 0.8$, $b = 0.01$, $\sigma = 1$, and $T = 1/52$. We use the daily discretization in the latter two schemes, i.e., $\Delta T = 1/252$. The computation times for 1 million samples under our exact simulation is 2,317 seconds. The configuration of the PC we used is Intel Core 2, 2.66GHz, and 2.87GB of Ram. The version of Matlab is Matlab 7.12.0(R2011a).

6.4 Linear-Drift CEV-Type-Diffusion Model The final example of this section uses the model (8). The Lamperti transform maps X into Y with the dynamic as follows:

$$dY_t = \left[-\frac{\gamma}{2(\gamma-1)} \cdot \frac{1}{f(Y_t)} - b(\gamma-1)f(Y_t) + a\sigma^{1/(\gamma-1)}(\gamma-1)^{\gamma/(\gamma-1)} f^{\gamma/(\gamma-1)}(Y_t) \right] dt + dW_t$$

where

$$f(Y_t) = \frac{x^{1-\gamma}}{\sigma(\gamma-1)} - Y_t.$$

As Y_t approaches $x^{1-\gamma}/(\sigma(\gamma-1))$, the right boundary of D_Y, the drift of Y will be influenced dominatingly by its first term

$$-\frac{\gamma}{2(\gamma-1)} \cdot \frac{1}{f(Y_t)}.$$

The condition that $\gamma \geq 1$ ensures that the model satisfies Assumption 3. Figure 7 displays the empirical transition densities of the process for $T = 1/12$ and $T = 1$, respectively. We observe a similar conclusion that we have to increase the number of time steps significantly for the Euler scheme to catch up with the performance of the exact method.

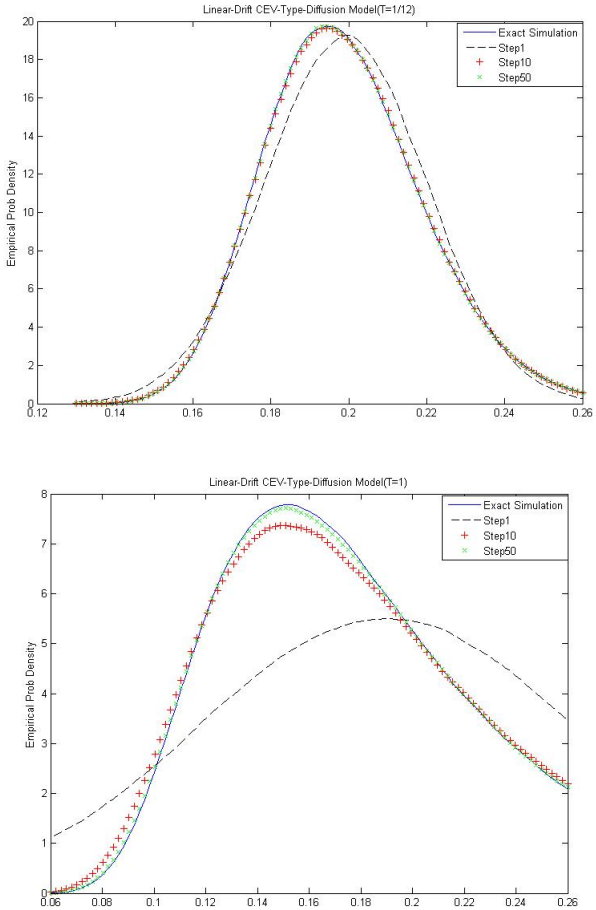

Figure 9: The exact simulation for Linear-Drift CEV-type. The two panels show the density functions of X_T when $T = 1/12$ and $T = 1$, respectively. The parameter set is cited from Ait-Sahalia [1]: $a = 0.0074, b = 0.0876, \sigma = 0.7791, \gamma = 1.48$, which are the MLE estimations fitting Fed fund data from January 1963 to December 1998. For all the schemes displayed in the plots, we ran 1 million paths. The empirical density obtained by the exact simulation matches very well with the density function produced by Ait-Sahalia's closed-form approximation. The computation times for 1 million samples under our exact simulation are 1,325 and 1,399 seconds, respectively. The configuration of the PC we used is Intel Core 2, 2.66GHz, and 2.87GB of Ram. The version of Matlab is Matlab 7.12.0(R2011a).

Appendix A. Simulation of Brownian Bridges. A Brownian bridge are defined as a solution to a particular SDE:

$$dB_t = \frac{b - B_t}{T - t}dt + dW_t,\ B_0 = a. \tag{20}$$

for $t \in [0, T)$ and given numbers a, b and $T > 0$. It has a more appealing representation for simulating purposes: a standard Brownian motion conditioning on that it starts from a and ends at b. From this, we can derive that B_t follows a normal distribution

$$B_t \sim N\left(\frac{(T-t)a + tb}{T}, \frac{t(T-t)}{T}\right).$$

To sample it, simply let

$$B_t = \frac{(T-t)a + tb}{T} + \sqrt{\frac{t(T-t)}{T}}Z,$$

with $Z \sim N(0, 1)$.

To simulate a vector of $(B_{t_1}, \cdots, B_{t_n})$ from (20) for $0 < t_1 < \cdots < t_n < T$, we can obtain them one by one. The Markov property of Brownian motion implies that for a given $B_{t_{i-1}}$, B_{t_i} is independent of all B_t with $t < t_{i-1}$, and given $B_{t_{i+1}}$, B_{t_i} is independent of all B_t with $t > t_{i+1}$. Thus,

$$\begin{aligned}(B_{t_i}|B_{t_1}, \cdots, B_{t_{i-1}}, B_{t_{i+1}}, \cdots, B_{t_n}) &\stackrel{d}{=} (B_{t_i}|B_{t_{i-1}}, B_{t_{i+1}}) \\ &\sim N\left(\frac{(t_{i+1} - t_i)B_{t_{i-1}} + (t_i - t_{i-1})B_{t_{i+1}}}{t_{i+1} - t_{i-1}}, \frac{(t_{i+1} - t_i)(t_i - t_{i-1})}{t_{i+1} - t_{i-1}}\right).\end{aligned}$$

Then, the simulation of B_{t_i} is accomplished through

$$B_{t_i} = \frac{(t_{i+1} - t_i)B_{t_{i-1}} + (t_i - t_{i-1})B_{t_{i+1}}}{t_{i+1} - t_{i-1}} + \sqrt{\frac{(t_{i+1} - t_i)(t_i - t_{i-1})}{t_{i+1} - t_{i-1}}}Z,$$

where $Z \sim N(0, 1)$, independent of all $B_{t_1}, \cdots, B_{t_{i-1}}, B_{t_{i+1}}, \cdots, B_{t_n}$. By repeatedly using the above observations, we can sample $(B_{t_1}, \cdots, B_{t_n})$ in any order. In particular, start by generating B_{t_n} from $N(((T - t_n)a + t_n b)/T, t_n(T - t_n)/T)$ and proceed backward to sample the other intermediate values, at each step conditioning on the two closet time points.

Appendix B. Proofs of Main Results. We prove the main results of the paper in this appendix.

Proof of Theorem 4.1. Given $Y_{\zeta_{i-1}} = x, \zeta_{i-1} = t$, denote $\widetilde{Y}_u := Y_{u+\zeta_{i-1}} - Y_{\zeta_{i-1}}$ for $u \geq 0$. Then

$$d\widetilde{Y}_u = \alpha(x + \widetilde{Y}_u)du + dW_u$$

and $\widetilde{Y}_0 = 0$. Therefore, $\Delta\zeta_i = \inf\{u \geq 0 : |\widetilde{Y}_u| = L\}$ and $\Delta Y_i = \widetilde{Y}_{(T-t)\wedge\zeta}$, respectively. Let $S_n = \inf\{u \geq 0 : |\widetilde{Y}_u| = n\}$ and $S = \lim_{n\to+\infty} S_n$. Assumptions 3.1-3.3 allow us to use a generalized Girsanov theorem (see, e.g., Karatzas and Shreve [35, Exercise 5.5.28]) to conclude that for every finite $T > 0$, any Borel set $\Gamma \subset \mathbf{R}$ and $s > 0$,

$$\begin{aligned}&P[\widetilde{Y}_{(T-t)\wedge\Delta\zeta_i} \in \Gamma, \zeta \in ds, S > T - t] \\ =\ & E\left[\exp\left(\int_0^{T-t} \alpha(x + W_u)dW_u - \frac{1}{2}\int_0^{T-t} \alpha^2(x + W_u)du\right) \cdot \mathbf{1}_{\{W_{(T-t)\wedge\varsigma} \in \Gamma, \varsigma \in ds\}}\right].\end{aligned} \tag{21}$$

Note that Aït-Sahalia [2] proves that $P[S = +\infty] = 1$ when Assumptions 3.1-3.3 hold. Part (iii) of Exercise 5.5.28 in Karatzas and Shreve [35] further implies that the nonnegative supermartingale

$$Z_t = \exp\left(\int_0^t \alpha(x + W_u)dW_u - \frac{1}{2}\int_0^t \alpha^2(x + W_u)du\right),\ 0 \leq t < +\infty,$$

is a true martingale.

Conditioning on $\mathcal{F}_{(T-t)\wedge\varsigma}$ and invoking the towering rule of conditional expectations, the right hand side of (21) equals

$$\begin{aligned}&E\left[E\left[\exp\left(\int_0^{T-t} \alpha(x + W_u)dW_u - \frac{1}{2}\int_0^{T-t} \alpha^2(x + W_u)du\right) \cdot \mathbf{1}_{\{W_{(T-t)\wedge\varsigma} \in \Gamma, \varsigma \in ds\}}\Big|\mathcal{F}_{(T-t)\wedge\varsigma}\right]\right] \\ =\ & E\left[\exp\left(\int_0^{(T-t)\wedge\varsigma} \alpha(x + W_u)dW_u - \frac{1}{2}\int_0^{(T-t)\wedge\varsigma} \alpha^2(x + W_u)du\right) \cdot \mathbf{1}_{\{W_{(T-t)\wedge\varsigma} \in \Gamma, \varsigma \in ds\}}\right]\end{aligned} \tag{22}$$

On the other hand, applying Ito's lemma on $A(x+W_t)$ for any t leads to

$$A(x+W_t) = A(x) + \int_0^t \alpha(x+W_u)dW_u + \frac{1}{2}\int_0^t \alpha'(x+W_u)du. \tag{23}$$

Combining (22) and (23) with (21) yields

$$\begin{aligned} & P[\widetilde{Y}_{(T-t)\wedge\zeta} \in \Gamma, \zeta \in ds, S > T-t] \\ = \; & E\left[\exp\left(A(x+W_{(T-t)\wedge\varsigma}) - A(x) - \int_0^{(T-t)\wedge\varsigma} \phi(x+W_u)du\right)\cdot \mathbf{1}_{\{W_{(T-t)\wedge\varsigma}\in\Gamma,\varsigma\in ds\}}\right], \end{aligned} \tag{24}$$

where $\phi = (\alpha^2+\alpha')/2$. The left hand side of (24) equals to $P[\widetilde{Y}_{(T-t)\wedge\zeta} \in \Gamma, \zeta \in ds]$ because $P[S > T-t] = 1$ and thus,

$$\begin{aligned} & P[\widetilde{Y}_{(T-t)\wedge\zeta} \in \Gamma, \zeta \in ds] \\ = \; & \exp(-A(x))\cdot E\left[\exp\left(A(x+W_{(T-t)\wedge\varsigma}) - \int_0^{(T-t)\wedge\varsigma} \phi(x+W_u)du\right)\mathbf{1}_{\{W_{(T-t)\wedge\varsigma}\in\Gamma,\varsigma\in ds\}}\right]. \end{aligned}$$

From the above equality, we know

$$\begin{aligned} & \frac{P[\Delta\zeta_i \in ds, \Delta Y_i \in dy | Y_{\zeta_{i-1}} = x, \zeta_{i-1} = t]}{P[\varsigma \in ds, W_{\varsigma\wedge(T-t)} \in dy]} \\ = \; & \exp(-A(x)+A(x+y))\cdot E\left[\exp\left(-\int_0^{s\wedge(T-t)} \phi(x+W_u)du\right)\Big|\varsigma = s, W_{\varsigma\wedge(T-t)} = y\right]. \end{aligned}$$

□

Proof of Theorem 4.2. Consider the case of $\varsigma_1 = s$ and $W_{\varsigma_1} = -1$ only. The arguments for the case of $W_{\varsigma_1} = 1$ are totally same by the symmetrical property of the law of standard Brownian motion. Denote $m_s = \min\{W_t : 0 \le t \le s\}$ and $T_s = \inf\{0 \le t \le s : W_t = m_s\}$. Using these new notations, the event $\{\varsigma_1 = s, W_{\varsigma_1} = -1\}$ occurs if and only if $T_s = s$, $W_{T_s} = -1$, and $\max_{0\le t\le s} W_t < 1$. Therefore, for all $y_i \in (0,2)$, $1 \le i \le n$,

$$\begin{aligned} & P[\widetilde{W}_{t_1} \in dy_1, \cdots, \widetilde{W}_{t_n} \in dy_n | \varsigma_1 = s, W_{\varsigma_1} = -1] \\ = \; & P[\widetilde{W}_{t_1} \in dy_1, \cdots, \widetilde{W}_{t_n} \in dy_n | T_s = s, W_{T_s} = -1, \max_{0\le t\le s} \widetilde{W}_t < 2]. \end{aligned}$$

Applying the Bayes' rule on the right hand side of the above equality, we have that it must be equal to

$$\frac{P[\widetilde{W}_{t_1} \in dy_1, \cdots, \widetilde{W}_{t_n} \in dy_n, \max_{0\le t\le s} \widetilde{W}_t < 2 | T_s = s, W_{T_s} = -1]}{P[\max_{0\le t\le s} \widetilde{W}_t < 2 | T_s = s, W_{T_s} = -1]}. \tag{25}$$

Williams [46] shows that the path of W decomposes at its time of minimum into two back-to-back path fragments, which are two independent Brownian meanders respectively. Imhof [31] further proves that a Brownian meander can be represented in terms of three independent Brownian bridges. Applying these two classical results here,

$$(\widetilde{W}_t, 0 \le t \le s | T_s = s, W_{T_s} = -1) \stackrel{d}{=} (\widetilde{B}_t, 0 \le t \le s),$$

where $\widetilde{B}$ is defined in (16). In particular, the numerator of (25) should equal

$$\begin{aligned} & P[\widetilde{W}_{t_1} \in dy_1, \cdots, \widetilde{W}_{t_n} \in dy_n, \max_{0\le t\le s} \widetilde{W}_t < 2 | T_s = s, W_{T_s} = -1] \\ = \; & P[\widetilde{B}_{t_1} \in dy_1, \cdots, \widetilde{B}_{t_n} \in dy_n, \max_{0\le t\le s} \widetilde{B}_t < 2] \\ = \; & P[\max_{0\le t\le s} \widetilde{B}_t < 2 | \widetilde{B}_{t_1} = y_1, \cdots, \widetilde{B}_{t_n} = y_n]\cdot P[\widetilde{B}_{t_1} \in dy_1, \cdots, \widetilde{B}_{t_n} \in dy_n]. \end{aligned} \tag{26}$$

By the Markov property of $\widetilde{B}$, we can further decompose the right hand side of (26) into

$$\prod_{i=1}^{n} P[\max_{t_i\le t\le t_{i+1}} \widetilde{B}_t < 2 | \widetilde{B}_{t_i} = y_i, \widetilde{B}_{t_{i+1}} = y_{i+1}]\cdot P[\max_{0\le t\le t_1} \widetilde{B}_t < 2 | \widetilde{B}_{t_1} = y_1].$$

In addition, the denominator of (25) should equal to

$$P[\max_{0\le t\le s}\widetilde{W}_t<2|T_s=s,W_{T_s}=-1]=P[\max_{0\le t\le s}\widetilde{B}_t<2].$$

Introduce two quantities such that

$$p(s,x;t,y):=P[\max_{s\le u\le t}\widetilde{B}_u<2|\widetilde{B}_s=x,\widetilde{B}_t=y]\quad\text{and}\quad q(s,x)=P[\max_{0\le u\le s}\widetilde{B}_u<2|\widetilde{B}_s=x].$$

Then,

$$\frac{P[\widetilde{W}_{t_1}\in dy_1,\cdots,\widetilde{W}_{t_n}\in dy_n|\varsigma_1=s,W_{\varsigma_1}=-1]}{P[\widetilde{B}_{t_1}\in dy_1,\cdots,\widetilde{B}_{t_n}\in dy_n]}=\frac{\prod_{i=1}^n p(t_i,y_i;t_{i+1},y_{i+1})\cdot q(t_1,y_1)}{P[\max_{0\le t\le s}\widetilde{B}_t<2]}.$$

The first half of the theorem statements is thus proved.

The closed-form expressions of p and q can be derived from the results obtained by Pötzelberger and Wang [42]. According to Imhof [31], the probability law of $\widetilde{B}$ is identical with the law of a Brownian bridge conditioned to stay positive. Denote $BB^{x\to y}$ to be the Brownian bridge from x to y on $[s,t]$. So,

$$p(s,x;t,y)=\frac{P[0<BB_u^{x\to y}<2,\text{ for all }u\in[s,t]]}{\lim_{M\to+\infty}P[0<BB_u^{x\to y}<M,\text{ for all }u\in[s,t]]}$$

and $q(s,x)=\lim_{u\to 0}p(0,u;s,x)$. Pötzelberger and Wang [42] provide an explicit expression for $P[0<BB_u^{x\to y}<M,\text{ for all }t\in[s,t]]$. That is,

$$P[0<BB_u^{x\to y}<M,\text{ for all }u\in[s,t]]=1-\sum_{j=1}^{+\infty}q(j)$$

where

$$\begin{aligned}q(j)&=\left(-\frac{2(Mj-x)(Mj-y)}{t-s}\right)+\exp\left(-\frac{2(M(j-1)+x)(M(j-1)+y)}{t-s}\right)\\&-\left[\exp\left(-\frac{2j(M^2j+M(x-y))}{t-s}\right)+\exp\left(-\frac{2j(M^2j-M(x-y))}{t-s}\right)\right].\end{aligned}$$

Substituting it into the expression of p and some routine calculation will lead to the second half of the theorem statements. □

Proof of Theorem (4.3). Conditional on that the number of homogenous Poisson arrivals N, the arrival times $\tau_1,\cdots,\tau_N$ and the sample path of Brownian motions $\{W_t\}$, we have

$$P[I=1|N,\tau_1,\cdots,\tau_N,W]=\frac{\exp(A(x+W_{(T-t)\wedge\varsigma}))}{\Gamma}\cdot\prod_{j=1}^N\frac{\phi(x+W_{\tau_j})-m}{M-m}\cdot\frac{\exp(-m((T-t)\wedge\varsigma))}{\max(1,\exp(-m(T-t)))}$$

because $U_i,0\le i\le N$ and V follow uniform distribution. By the tower rule of conditional expectation,

$$\begin{aligned}&P[I=1|N,\varsigma,W_{(T-t)\wedge\varsigma}]=E[P[I=1|N,\tau_1,\cdots,\tau_N,W]|N,\varsigma,W_{(T-t)\wedge\varsigma}]]\\=\;&\frac{\exp(A(x+W_{(T-t)\wedge\varsigma}))}{\Gamma}\cdot E\left[\prod_{j=1}^N\frac{M-\phi(x+W_{\tau_j})}{M-m}\Big|N,\varsigma,W_{(T-t)\wedge\varsigma}\right]\cdot\frac{\exp(-m((T-t)\wedge\varsigma))}{\max(1,\exp(-m(T-t)))}.\end{aligned}$$

We know that $(\tau_1,\cdots,\tau_N)$ have the same joint distribution as the order statistics of N independent uniformly distributed random variables on $[0,(T-t)\wedge\varsigma]$. Thus,

$$\begin{aligned}&E\left[\prod_{j=1}^N\frac{M-\phi(x+W_{\tau_j})}{M-m}\Big|N,\varsigma,W_{(T-t)\wedge\varsigma}\right]\\=\;&E\left[\left(\frac{1}{(T-t)\wedge\varsigma}\int_0^{(T-t)\wedge\varsigma}\left[\frac{M-\phi(x+W_u)}{M-m}\right]du\right)^N\Big|N,\varsigma,W_{(T-t)\wedge\varsigma}\right].\end{aligned}\tag{27}$$

Note that $N \sim Poisson((M-m)((T-t)\wedge\varsigma))$. Then, the right hand side of (27) equals

$$\sum_{n=0}^{+\infty} E\left[\left(\frac{1}{(T-t)\wedge\varsigma}\int_0^{(T-t)\wedge\varsigma}\left[\frac{M-\phi(x+W_u)}{M-m}\right]du\right)^n \frac{e^{-(M-m)((T-t)\wedge\varsigma)}((M-m)((T-t)\wedge\varsigma))^n}{n!}\Big|\varsigma, W_{\varsigma\wedge(T-t)}\right]$$

Some straightforward algebra manipulations on the above quantity lead to a conclusion that

$$E\left[\prod_{j=1}^{N}\frac{M-\phi(x+W_{\tau_j})}{M-m}\Big|N,\varsigma,W_{(T-t)\wedge\varsigma}\right] = E\left[\exp\left(-\int_0^{\varsigma\wedge(T-t)}(\phi(x+W_u)-m)du\right)|\varsigma, W_{\varsigma\wedge(T-t)}\right].$$

In light of all the results ahead, we know

$$P[I=1|\varsigma,W_{(T-t)\wedge\varsigma}] = \frac{\exp(A(x+W_{(T-t)\wedge\varsigma}))}{\Gamma\max(1,\exp(-m(T-t)))}\cdot E\left[\exp\left(-\int_0^{\varsigma\wedge(T-t)}\phi(x+W_u)du\right)|\varsigma, W_{\varsigma\wedge(T-t)}\right].$$

□

Proof of Theorem (5.1). Here we only present a proof to the case of $D_Y = (-\infty,\infty)$. The treatment of SDEs with boundaries is similar but involves with much more details and we skip it for the interest of space.

Suppose that the theorem statement is not true. That is, for some sample paths, we have $J=\infty$. There are two possibilities: either one subsequence of $\{Y_{\zeta_1}, Y_{\zeta_2},\cdots,Y_{\zeta_i},\cdots\}$ tends to $\pm\infty$ or the whole sequence is bounded in an interval, say, $[-NL, NL]$ for a large positive integer N. The first possibility is ruled out by Assumption 3.3 since Y does not explode in any finite time horizon $[0,T]$. We consider the second possibility only. Note that all Y_ζ's can only take values in a finite set $\{-NL, -(N-1)L,\cdots,(N-1)L, NL\}$. Therefore, there must exist some integer $k\in[-N,N]$ such that $Y_{\zeta_{n_j}} = kL$ for all j, where $\{Y_{\zeta_{n_j}}, j\geq 1\}$ is a subsequence of $\{Y_{\zeta_1}, Y_{\zeta_2},\cdots,Y_{\zeta_i},\cdots\}$.

On the other hand, the assumption that $J=\infty$ implies that

$$T > \sum_{i=1}^{+\infty}\Delta\zeta_i \geq \sum_j \Delta\zeta_{n_j}. \tag{28}$$

In our simulation scheme, the distribution of $\Delta\zeta_{n_j}$ depends only on $Y_{\zeta_{n_j}}$, which is the same across all j. Therefore, by the Markovian property of the process Y, all $\Delta\zeta_{n_j}$, $j\geq 1$, should be distributed independently and identically. The law of large number implies that $\sum_j \Delta\zeta_{n_j} = +\infty$. Contradiction to (28). Then we know that J must be finite almost surely.

We also can see how to prove the theorem for the SDEs with boundaries. Take $D_Y = (\underline{y}, +\infty)$ as an example. Using Assumption 3.3, Y cannot reach $\underline{y}$ and $+\infty$ before T. If $J=+\infty$, this implies that it will be "recurrent" in D_Y. In other words, there exists some level y such that the process $\{Y_t, 0\leq t\leq T\}$ hits y infinite times. This will lead to a contradiction like (28). We skip the detailed reasoning here. □

Appendix C. The Oscillating Property of the Series This part of the Appendix is devoted to the discussion on the oscillating property of $\{g_J(t)\}$, $\{p_J(s,x;t,y)\}$ and $\{q_J(s,x)\}$. Burq and Jones [15, Lemma 4] have already established the oscillating property of $\{g_J(t)\}$. Based on this, they also prove the expected number of operations needed to simulate ς_1 is finite. Thus, we skip sequence $\{g_J\}$ and only concentrate on the latter two sequences. The main result is the following proposition:

Proposition C.1 *(i). When $x,y\in(0,2)$ and $0<s<t$, for any integer*

$$J \geq \frac{(\log 3)(t-s)}{8}\cdot\max\left(\frac{1}{x},\frac{1}{y}\right)+1, \tag{29}$$

the sequence $\{p_J(s,x;t,y)\}$ satisfies

$$0 < p_{2J+1} - p_{2J} \leq \frac{2}{3}(p_{2J-1}-p_{2J}) \leq \left(\frac{2}{3}\right)^2(p_{2J-1}-p_{2J-2}).$$

(ii). When $x\in(0,2)$ and $0<s$, for any integer

$$J \geq \frac{(\log 4)s}{8}\cdot\frac{1}{x}+2, \tag{30}$$

the sequence $\{q_J(s,x)\}$ *satisfies*

$$0 < q_{2J+1} - q_{2J} \le \frac{1}{2}(q_{2J-1} - q_{2J}) \le \frac{1}{4}(q_{2J-1} - q_{2J-2}).$$

PROOF. (i). For any integer J satisfying (29) and $x, y \in (0,2)$,

$$\exp\left(\frac{2x(4J-y)}{t-s}\right) \ge 2 \text{ and } \exp\left(\frac{2y(4J-x)}{t-s}\right) \ge 2. \tag{31}$$

Rearranging the terms in (31), it is easy to see

$$\exp\left(-\frac{2(2J-x)(2J-y)}{t-s}\right) \ge 2\max\left\{\exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right), \exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right)\right\}. \tag{32}$$

On the other hand, the fact $0 < x, y < 2$ also implies

$$\min\left\{\exp\left(\frac{2(2-x)(4J-(2-y))}{t-s}\right), \exp\left(\frac{2(2-y)(4J-(2-x))}{t-s}\right)\right\} \ge 1.$$

From it, we have

$$\begin{aligned} & \exp\left(-\frac{2(2(J-1)+x)(2(J-1)+y)}{s_2-s_1}\right) \\ \ge\ & \max\left\{\exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right), \exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right)\right\}. \end{aligned} \tag{33}$$

Summing up (32) and (33) will yield

$$\begin{aligned} & \exp\left(-\frac{2(2J-x)(2J-y)}{s_2-s_1}\right) + \exp\left(-\frac{2(2(J-1)+x)(2(J-1)+y)}{s_2-s_1}\right) \\ \ge\ & \frac{3}{2}\exp\left(-\frac{2J(4J+2(x-y))}{s_2-s_1}\right) + \frac{3}{2}\exp\left(-\frac{2J(4J-2(x-y))}{s_2-s_1}\right), \end{aligned}$$

i.e., $2\theta_J/3 \ge \vartheta_J$. Therefore, we have

$$0 < p_{2J-1} - p_{2J} < \frac{2}{3}(p_{2J-1} - p_{2J-2}).$$

The inequality (29) also leads to

$$\exp\left(\frac{8Jx}{t-s}\right) \ge 3 \quad \text{and} \quad \exp\left(\frac{8Jy}{t-s}\right) \ge 3.$$

After some algebraic manipulations, we know that

$$\min\left\{\exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right), \exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right)\right\} \ge 3\exp\left(-\frac{2(2J+x)(2J+y)}{t-s}\right).$$

Thus,

$$\frac{1}{4}\exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right) + \frac{1}{4}\exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right) \ge \frac{3}{2}\exp\left(-\frac{2(2J+x)(2J+y)}{t-s}\right). \tag{34}$$

Meanwhile, it is easy to show that, for any integer J and $0 < x, y < 2$,

$$\min\left\{\exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right), \exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right)\right\} \ge \exp\left(-\frac{2(2(J+1)-x)(2(J+1)-y)}{t-s}\right).$$

Then,

$$\frac{3}{4}\exp\left(-\frac{2J(4J+2(x-y))}{t-s}\right) + \frac{3}{4}\exp\left(-\frac{2J(4J-2(x-y))}{t-s}\right) \ge \frac{3}{2}\exp\left(-\frac{2(2(J+1)-x)(2(J+1)-y)}{t-s}\right). \tag{35}$$

Summing up the inequalities (34) and (35), we have $2\vartheta_J/3 \ge \theta_{J+1}$, which implies

$$0 < p_{2J+1} - p_{2J} < \frac{2}{3}(p_{2J-1} - p_{2J}).$$

(ii). When J satisfies (30) and $x \in (0,2)$,

$$\exp\left(\frac{8Jx}{s}\right) \geq 4 \geq 2 \cdot \frac{4J+x}{4J-x}, \tag{36}$$

where the first inequality is due to $J \geq (\log 4)s/(8x)$ and the second inequality is because $J \geq 2$ and $x < 2$. By (36),

$$\frac{1}{2}(4J-x)\exp\left(-\frac{4J(2J-x)}{s}\right) \geq (4J+x)\exp\left(-\frac{4J(2J+x)}{s}\right),$$

i.e., $\rho_J/2 \geq \varrho_J$. This implies $(q_{2J-1} - q_{2J-2})/2 \geq (q_{2J-1} - q_{2J})$. We also are able to establish

$$\exp\left(\frac{4(2-x)(2J+1)}{s}\right) \geq 1 \geq \frac{1}{2} \cdot \frac{4(J+1)-x}{4J+x}$$

for those J satisfying (30). Thus, $\varrho_J/2 \geq \rho_{J+1}$ and we can obtain $(q_{2J-1} - q_{2J})/2 \geq (q_{2J+1} - q_{2J})$. □

Proposition C.1 reveals that both sequences of $\{p_J\}$ and $\{q_J\}$ converge to their respective limits exponentially fast. This observation helps us to establish Theorem 4.1.

Proof of Proposition 4.1. Note that $q(s,x) = 1 - q_\infty(s,x)/x$. Given $U \in (0,1)$, $s > 0$, and $x \in (0,2)$, comparing U and $q(s,x)$ is the same thing as comparing $(1-U)x$ and $q_\infty(s,x)$. we intend to show that we can finish the comparison between $(1-U)x$ and $q_\infty(s,x)$ within at most $2N$ steps, where N is specified in the proposition statement. Suppose that it is not true, we must have the following inequality:

$$|(1-U)x - q_\infty(s,x)| \leq \max\{|q_{2N}(s,x) - q_\infty(s,x)|, |q_{2N+1}(s,x) - q_\infty(s,x)|\}.$$

Note that

$$N \geq \frac{\log(4)s}{8}\frac{1}{x} + 2.$$

Following part (ii) of Proposition C.1, for any $k \geq 2N$, we know that $|q_k(s,x) - q_{k+1}(s,x)| \leq (1/2)^{k-2N}|q_{2N}(s,x) - q_{2N+1}(s,x)|$. Therefore,

$$|q_{2N}(s,x) - q_\infty(s,x)| \leq \sum_{k=2N}^{+\infty} |q_k(s,x) - q_{k+1}(s,x)| \leq |q_{2N}(s,x) - q_{2N+1}(s,x)| \cdot \sum_{k=2N}^{+\infty} \left(\frac{1}{2}\right)^{k-2N} = 2\rho_{N+1}.$$

Similarly, we obtain that

$$|q_{2N+1}(s,x) - q_\infty(s,x)| \leq 2\varrho_{N+1}.$$

Consequently,

$$\begin{aligned} & |(1-U)x - q_\infty(s,x)| \leq 2\max\{\rho_{N+1}, \varrho_{N+1}\} \\ = \;& 2\max\left\{(4N+4-x)\exp\left(-\frac{(4N+4)(2N+2-x)}{s}\right), (4N+4+x)\exp\left(-\frac{(4N+4)(2N+2+x)}{s}\right)\right\}. \end{aligned} \tag{37}$$

Since $x \in (0,2)$, we know that the right hand side of (37) must be less than $6(N+1)\exp(-8N^2/s)$. From the assumption of the proposition, N satisfies that

$$6(N+1)\exp(-8N^2/s) < |(1-U)x - q_\infty(s,x)|.$$

Contradicting to (37). Then, we know that $(1-U)x$ should not be inside the interval $(q_{2N}(s,x), q_{2N+1}(s,x))$. If $(1-U)x < q_{2N}(s,x)$, $(1-U)x$ must be less than $q_\infty(s,x)$; otherwise, $(1-U)x$ must be larger than $q_\infty(s,x)$. From this observation, we can finish the comparison between U and $q(s,x)$ just within $2N$ steps. □

Acknowledgment. The author is indebted to Paul Glasserman, Jeremy Staum and David Yao for their helpful comments on the draft of this paper. An earlier version was presented at the 2008 Monte Carlo and Quasi Monte Carlo conference, Montreal, Canada and the 2008 conference of Quantitative Method in Finance, Sydney, Australia. This research is supported under Grant No. CUHK411108 and CUHK411309 by the Hong Kong Research Council.

References

[1] Y. Aït-Sahalia, *Transition densities for interest rate and other nonlinear diffusions*. Journal of Finance **70** (1999), 1361–1395.

[2] Y. Aït-Sahalia, *Maximum likelihood estimation of discretely sampled diffusions: a closed-from approximation approach*. Econometrica **70** (2002), 223–262.

[3] A. Alfonis, *On the discretization schemes for the cir (and bessel squared) processes*. Monte Carlo Methods and Applications **11** (2005), 355–384.

[4] A. Alfonis, *High order discretization schemes for the cir process: application to affine term structure and heston model*. Mathematics of Computation (2008). Forthcoming.

[5] L. Andersen, *Simple and efficient simulation of the heston stochastic volatility model*. Journal of Computational Finance **11** (2008), 1–42.

[6] S. Asmussen and P. W. Glynn, *Stochastic simulation: algorithms and analysis*. Springer-Verlag, New York, 2007.

[7] V. Bally and D. Talay, *The law of the euler scheme for stochastic differential equations I: convergence rate of the distribution function*. Probability Theory and Related Fields **104** (1996), 43–60.

[8] A. Beskos and G. O. Roberts, *Exact simulation of diffusions*, Annals of Applied Probability **15** (2005), 2422–2444.

[9] A. Beskos, O. Papaspiliopoulos, and G. O. Roberts, *Retrospective exact simulation of diffusion sample paths with applications*. Bernoulli **12** (2006), 1077–1098.

[10] A. Beskos, O. Papaspiliopoulos, and G. O. Roberts, *A factorisation of diffusion measure and finite sample path constructions*. Metholodogy and Computing in Applied Probability, **10** (2008), pp. 85–104.

[11] A. Berkaoui, M. Bossy and A. Diop, *Euler scheme for SDEs with non-lipschitz diffusion coefficient: strong convergence*. ESAIM: Probability and Statistics **12** (2008), 1–11.

[12] M. Bossy and A. Diop, *An efficient discretization scheme for one dimensional SDEs with a diffusion coefficient function of the form* $|x|^\alpha, \alpha \in [1/2, 1)$. Research report, Institut Nationale de Recherche en Informatique et en Automatique (INRIA), No. 5396, December 2004.

[13] M. Broadie, and O. Kaya, *Exact simulation of option Greeks under stochastic volatility and jump diffusion models*. Proceedings of the 2004 Winter Simulation Conference. Ingalls et al. eds. IEEE, Piscataway, NJ, 1607-1615.

[14] M. Broadie and O. Kaya, *Exact simulation of stochastic volatility and other affine jump diffusion processes*. Operations Research **54** (2006), 217–231.

[15] Z. A. Burq and O. D. Jones, *Simulation of Brownian motion at first-passage times*. Mathematics and Computers in Simulation **77** (2008), 64–81.

[16] B. Casella and G. O. Roberts, *Exact Monte Carlo simulation of killed diffusions*. Advances in Applied Probability **40** (2007), 273–291.

[17] K. C. Chan, G. A. Karolyi, F. A. Longstaff and A. B. Sanders, *An empirical comparison of alternative models of the short-term interest rate*. Journal of Finance **47** (1992), 1209–1227

[18] J. C. Cox, J. E. Ingersoll and S. A. Ross, *A theory of the term structure of interest rates*. Econometrica **53** (1985), 129–151.

[19] G. Deelstra and F. Delbaen, *Convergence of discretized stochastic (interest rate) processes with stochastic drift term*. Applied Stochastic Models and Data Analysis **14** (1998), 77–84.

[20] A. Diop, *Sur la Discrétisation et le Comportement à Petit Bruit d'EDS Unidimensionelles dont les Coefficients sont à Derives Singulières*. PhD thesis, INRIA, 2003.

[21] D. Duffie and P. Glynn, *Effiecient monte carlo simulation of security prices*. Annals of Applied Probability **5** (1995), 897–905.

[22] L. Devroye, *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

[23] D. Florens, *Estimation of the diffusion coefficient from crossing*. Statistical Inference for Stochastic Processes **1** (1999), 175–195.

[24] M. B. Giles, *Multi-level Monte Carlo path simulation*. Operations Research **56** (2008), 607–617.

[25] P. Glasserman, *Monte Carlo methods in financial engineering*. Springer-Verlag, New York, 2004.

[26] P. Glasserman and K. Kim, *Gamma expansion of the Heston stochastic volatility model*. To appear in *Finance and Stochastics* (2008).

[27] P. Glasserman, and J. Staum, *Conditioning on one-step survival in barrier option simulation*. Operations Research **49** (2001), 923–937.

[28] P. Glynn and D. Inglehart, *Simulation methods for queues: an overview*. Queueing Systems **3** (1988), 221–256.

[29] S. Heston, *A closed-form solution of options with stochastic volatility with applications to bond and currency options*. Review of Financial Studies **6** (1993), 327–343.

[30] D. J. Higham and X. Mao, *Convergence of the Monte Carlo simulations involving the meanreverting square root process*, Journal of Computational Finance **8** (2005), 35–62.

[31] J. P. Imhof, *Density factorizations for Brownian motion, meander and the three dimensional Bessel process, and applications*. Journal of Applied Probability **21** (1984), 500–510.

[32] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*. Vol. 2. John Wiley & Sons, New York, 1995.

[33] C. Kahl and P. Jäckel, *Fast strong approximation Monte-Carlo schemes for stochastic models*. Quantitative Finance **6** (2006), 513–536.

[34] S. Karlin and H. M. Talor, *A second course in stochastic processes*. Academic Press, San Diego, 1981.

[35] I. Karatzas and S. E. Shreve, *Browinian motion and stochastic calculus*. 2nd Edition. Springer, New York, 1991.

[36] A. Kebaier, *Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing*. Annals of Applied Probability **15** (2005), 2681–2705.

[37] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin, 1992.

[38] P. A. W. Lewis and G. S. Shedler, *Simulation of nonhomogeneous poisson processes by thinning*. Naval Research Logistics Quarterly **26** (1979), 403–413.

[39] R. Lord, R. Koekkoek and D. van Dijk, *A Comparison of biased simulation schemes for stochastic volatility models*. To appear in Quantitative Finance (2009).

[40] G. N. Milstein, *Approximate integration of stochastic differential equations*. Theory of Probability and its Application **19** (1974), 557–562.

[41] B. Moro, *The Full Monte*. Risk **8** (Feb) (1995), 57–58.

[42] K. Pötzelberger and L. Wang, *Boundary crossing probability for Brownian motion*. Journal of Applied Probability **38** (2001), 152–164.

[43] J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*. Random Structures and Algorithms **9** (1996), 223–252.

[44] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*. Stochastic Analysis and Applications **8** (1990), 94–120.

[45] O. A. Vasicek, *An equilibrium characterization of the term structrue*. Journal of Financial Economics **5** (1977), 177–188.

[46] D. Williams, *Decomposing the Brownian Path*. Bulletin of the American Mathematical Society **76** (1970), 871–873.