Discretization of Deflated Bond Prices

Paul Glasserman
Graduate School of Business
Columbia University
New York, NY 10027

Hui Wang
Department of Statistics
Columbia University
New York, NY 10027

March 1999

Abstract

This paper proposes and analyzes discrete-time approximations to a class of diffusions, with an emphasis on preserving certain important features of the continuous-time processes in the approximations. We start with multivariate diffusions having three features in particular: they are martingales, each of their components evolves within the unit interval, and the components are almost surely ordered. In the models of the term structure of interest rates that motivate our investigation, these properties have the important implications that the model is arbitrage-free and that interest rates remain positive. In practice, numerical work with such models often requires Monte Carlo simulation and thus entails replacing the original continuous-time model with a discrete-time approximation. It is desirable that the approximating processes preserve the three features of the original model just noted, though standard discretization methods do not. We introduce new discretizations based on first applying nonlinear transformations from the unit interval to the real line (in particular, the inverse normal and inverse logit), then using an Euler discretization, and finally applying a small adjustment to the drift in the Euler scheme. We verify that these methods enforce important features in the discretization with no loss in the order of convergence (weak or strong). Numerical results suggest that these methods can also yield a better approximation to the law of the continuous-time process than does a more standard discretization.

Key words: Interest rate models, Monte Carlo simulation, martingales

1 Introduction

A central principle of financial economics, sometimes referred to as the fundamental theorem of asset pricing, is that in the absence of arbitrage (opportunities for riskless profits) the prices of traded assets are martingales when properly normalized. The normalization consists of dividing by the price of another asset, called the numeraire, which may be thought of as the asset relative to which all other assets are priced. A *deflated* asset price is the ratio of the asset price to the price of the numeraire. A more accurate paraphrasing of the fundamental theorem is then that for each choice of numeraire (which can be essentially any positive price process) there exists a measure under which deflated asset prices are martingales. For precise formulations of these ideas see Harrison and Kreps [7], Harrison and Pliska [8], Delbaen and Schachermayer [2], Duffie [3]; for a discussion of the relation between changing numeraires and changing measure see Geman, El Karoui, and Rochet [5].

In a stochastic model of interest rates, the most basic assets are pure-discount bonds — bonds making a sure payment of 1 unit of account at a fixed date in the future. Clearly, bond prices should always be positive (buying a bond at a negative price would be an arbitrage). In a model with positive interest rates bond prices should never exceed 1. Indeed, if we write $B_T(t)$ for the time-t price of a bond maturing at $T \geq t$, then $[B_{T_1}(t) - B_{T_2}(t)]/[B_{T_2}(t)(T_2 - T_1)]$ defines the *forward* rate of interest for the interval $[T_1, T_2]$, $t \leq T_1 < T_2$. For all forward rates to be positive, bond prices must decrease with maturity, so $B_T(t) < B_t(t) \equiv 1$ whenever $T > t$. If, therefore, we choose as numeraire an asset whose price never falls below 1, then under the measure associated with this numeraire deflated bond prices have the following properties:

(i) they are martingales;

(ii) they are bounded between 0 and 1;

(iii) they are almost surely ordered, decreasing with maturity.

That the numeraire should be bounded from below (by 1) occurs naturally if the numeraire corresponds to the value of depositing 1 unit at time 0 in an interest bearing account and reinvesting all interest payments. Indeed, this type of "money market account" is probably the most common choice of numeraire; see, for example, the general class of models identified by Heath, Jarrow, and Morton [9].

The three characteristics above lead us to consider processes of the form

$$dX_i(t) = X_i(t) \sum_{j=1}^{d} \sigma_{ij}(X(t))\, dW_j(t), \quad i = 1, \ldots, N, \tag{1}$$

$$0 < X_N(t) < \cdots < X_1(t) < 1, \tag{2}$$

where $X(t) = (X_1(t), \ldots, X_N(t))$ and $W(t) = (W_1(t), \ldots, W_d(t))$ is a d-dimensional standard Brownian motion. We have in mind a model with finitely many maturities $T_N > T_{N-1} > \cdots > T_1 > T_0 = 0$ in which $X_i(t)$ represents the time-t deflated price of the bond maturing at $T_i \geq t$. (How X_i evolves after T_i will be immaterial.) Models in which deflated bond prices admit precisely such a representation have recently been put forth in an important stream of research that includes Brace, Gatarek, and Musiela [1], Jamshidian [11], Miltersen, Sandmann, and Sondermann [15], and Musiela and Rutkowski [16]. These models actually take forward interest rates rather than bond prices as model primitives (as do Heath et al. [9]), but it is possible to solve for the bond price dynamics and to put them in the form (1). Using the numeraire proposed in Jamshidian [11] (a discretely compounded money market account) the property in (2) holds automatically. Using instead the numeraire in Brace, Gatarek, and Musiela [1] and Musiela and Rutkowski [16] (a bond maturing later than T_N), (2) should be replaced with $1 < X_N(t) < \cdots < X_1(t)$. This is similar to but simpler than the restriction in (2), so it suffices to consider (2).

Our interest is in the time-discretization of (1) consistent with properties (i)-(iii) above. Pricing interest rate derivative securities in a model of the whole term structure (of the Heath-Jarrow-Morton [9] type or the others cited above) generally requires a numerical method and for instruments that depend on the *path* of interest rates the only viable method is often Monte Carlo simulation. Simulation requires approximating diffusions with discrete-time processes. Typically, the forward rate processes through which the models are defined are themselves discretized; but as argued in Glasserman and Zhao [6] (in the specific setting of [1, 11, 16]) there can be advantages to simulating instead the deflated bond prices or their increments. Hence, here we take (1) (rather than forward rate processes leading to it) as our starting point.

Each of the properties (i)-(iii) is important in the continuous-time formulation of an interest rate model. Care should therefore be taken to preserve these properties in the discretization, since it is ultimately the discretized model from which prices are computed. To further motivate our line of investigation, consider the one-dimensional case

$$dX(t) = X(t)\sigma(X(t))\,dW(t), \quad 0 < X(t) < 1. \tag{3}$$

A standard Euler discretization of this process with step size h is given recursively by

$$\hat{X}((i+1)h) = \hat{X}(ih) + \hat{X}(ih)\sigma(\hat{X}(ih))\sqrt{h}Z_{i+1}, \quad \hat{X}(0) = X(0) \equiv x_0,$$

where $Z_1, Z_2, \ldots$ are independent standard normal random variables. The discretized process $\hat{X}$ is clearly a martingale, but it is not restricted to the unit interval. We can keep $\hat{X}$ between 0 and 1 by projecting values outside this range back to $(0,1)$, but in so doing we destroy the martingale property. We are left in the position of having to choose between two important features of the original model. More accurate approximations to the law of the continuous-time process can sometimes be obtained using a higher-order discretization (of the type detailed in Kloeden and Platen [12], Milstein [14], and Talay [18]) but these do not directly address the issue of preserving properties (i)-(iii).

The methods we investigate begin by applying a transformation $Y(t) = g(X(t))$ where g is an increasing, twice continuously differentiable mapping from $[0,1]$ onto the real line. We discretize Y and then apply the inverse transformation $f = g^{-1}$. This produces a discretization of the original process that is automatically restricted to the unit interval. In the univariate case we show how to carry out the discretization of Y so that the resulting discretization of X will furthermore be a martingale. This approach (developed in Section 2) is particularly explicit when g is the inverse of the cumulative normal distribution. In Section 3 we turn to the multivariate case and show that under a general class of transformations the martingale property holds "to first order" in a sense to be made precise. In Section 4, we develop a second-order adjustment. We show that all methods proposed have the same strong and weak convergence orders as the standard Euler scheme and are therefore no worse than an Euler scheme in this respect. Section 5 gives some numerical illustrations.

2 One-Dimensional Setting

We proceed with the approach sketched towards the end of the previous section. For now, we simply assume that the stochastic differential equation (SDE) in (3) has a unique strong solution that remains in the unit interval whenever $0 < X(0) < 1$. Let $g : (0,1) \to R$ be increasing, surjective, and twice continuously differentiable. By Ito's rule, the process $Y(t) = g(X(t))$ satisfies

$$\begin{aligned} dY(t) &= \tfrac{1}{2}g''(X_t)\sigma^2(X_t)X_t^2\,dt + g'(X_t)\sigma(X_t)X_t\,dW_t \\ &\equiv \tilde{\mu}(Y_t)\,dt + \tilde{\sigma}(Y_t)\,dW_t, \end{aligned}$$

with

$$\begin{aligned} \tilde{\mu}(y) &= \tfrac{1}{2}g''(f(y))\sigma^2(f(y))f^2(y) \\ \tilde{\sigma}(Y_t) &= g'(f(y))\sigma(f(y))f(y), \end{aligned}$$

and $f = g^{-1}$. A standard Euler discretization gives

$$\hat{Y}((i+1)h) = \hat{Y}(ih) + \tilde{\mu}(\hat{Y}(ih))h + \tilde{\sigma}(\hat{Y}(ih))\sqrt{h}Z_{i+1}, \quad \hat{Y}(0) = g(x_0),$$

with $Z_1, Z_2, \ldots,$ independent standard normals. The process $\hat{X}(ih) = f(\hat{Y}(ih))$ never leaves the unit interval, but it is not a martingale.

Consider a modified Euler scheme

$$\hat{Y}((i+1)h) = \hat{Y}(ih) + \mu(\hat{Y}(ih))h + \tilde{\sigma}(\hat{Y}(ih))\sqrt{h}Z_{i+1},$$

in which μ is to be chosen close enough to $\tilde{\mu}$ to preserve convergence of the scheme but perturbed sufficiently to make $\hat{X} = f(\hat{Y})$ a martingale. Imposing the martingale condition is equivalent to requiring

$$\mathsf{E}[f(\hat{Y}(ih) + \mu(\hat{Y}(ih))h + \tilde{\sigma}(\hat{Y}(ih))\sqrt{h}Z_{i+1}|\hat{Y}(ih)] = f(\hat{Y}(ih)) \tag{4}$$

We now focus on the particular case $g = \Phi^{-1}$ and

$$f(y) = \Phi(y) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2}\,du.$$

This transformation is made particularly convenient by the fact that we can explicitly evaluate, for any constants a, b,

$$\mathsf{E}[\Phi(a + bZ)] = P(Z' \le a + bZ) = P(Z' - bZ \le a) = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right),$$

with Z, Z' independent standard normals. With $y = \hat{Y}(ih)$, (4) becomes

$$\Phi\left(\frac{y + \mu(y)h}{\sqrt{1 + \tilde{\sigma}^2(y)h}}\right) = \Phi(y).$$

We therefore require

$$y = \frac{y + \mu(y)h}{\sqrt{1 + \tilde{\sigma}^2(y)h}},$$

which is to say

$$\mu(y) = \frac{1}{h}\left(\sqrt{1 + \tilde{\sigma}^2(y)h} - 1\right) y. \tag{5}$$

A simple calculation verifies that (for general g and $f = g^{-1}$)

$$\tilde{\sigma}^2(y) = \frac{-2f'(y)}{f''(y)}\tilde{\mu}(y).$$

With $f = \Phi$ the expression on the right becomes $2\tilde{\mu}(y)/y$ and (5) becomes

$$\mu(y) = \frac{1}{h}\left(\sqrt{1 + \frac{2h}{y}\tilde{\mu}(y)} - 1\right) y. \tag{6}$$

Substituting this expression for μ results in the discretized process

$$\hat{Y}((i+1)h) = \hat{Y}(ih) + \left(\sqrt{1 + \frac{2h}{\hat{Y}(ih)}\tilde{\mu}(\hat{Y}(ih))} - 1\right)\hat{Y}(ih) + \tilde{\sigma}(\hat{Y}(ih))\sqrt{h}Z_{i+1}. \tag{7}$$

We summarize the derivation leading to this scheme in the following:

Proposition 1 *Let $\hat{Y}$ be as in (7). Then the process $\hat{X} = \Phi(\hat{Y})$ is a martingale on the unit interval.*

Of course, for this scheme to be of any interest $\hat{X}$ must converge to X, so we now address this issue. We recall some notions of convergence of discrete-time approximations to SDEs, as presented in Kloeden and Platen [12]. A discretization $\hat{X}$ based on time-step h converges to X with *strong order* γ if

$$\mathsf{E}[|\hat{X}(nh) - X(T)|] \le Ch^{\gamma}, \quad n = T/h,$$

for some constant C and all sufficiently small h. It converges with *weak* order γ if

$$|\mathsf{E}[\psi(\hat{X}(nh)] - \mathsf{E}[\psi(X(T))]| \leq Ch^{\gamma},$$

for all functions ψ having $2(\gamma+1)$ continuous and polynomially bounded derivatives. The Euler scheme typically has strong order 1/2 (see p.327 of Kloeden and Platen [12]), but it also has weak order 1 under additional smoothness conditions on the coefficients of the SDE. We now have

Theorem 1 *For the SDE in (3), suppose that $0 < X(0) \equiv x_0 < 1$, that $\sigma(\cdot)$ is Lipschitz continuous on $[0,1]$, and that $\sigma(1) = 0$. Then for any $T > 0$ we have $0 < X(t) < 1$, $0 \leq t \leq T$, almost surely, and $\hat{X} = \Phi(\hat{Y})$ has strong convergence order $1/2$.*

Theorem 2 *Suppose the conditions in Theorem 1 hold and that $\sigma(\cdot)$ is in fact $C^4([0,1])$. Then $\hat{X}$ converges to X with weak order 1.*

Proofs of these and all other results are given in an appendix.

3 Multidimensional Setting

We now turn to the general model in (1). To keep each coordinate between 0 and 1 we could apply a transformation g to each coordinate separately, and to preserve the martingale property we could make the drift adjustment introduced in the previous section. In this way, the univariate results of the previous section extend to vector processes. However, these transformations do not by themselves ensure that the coordinates remain ordered as indicated in (2). In the interest rate setting that motivates our investigation, when some increment $X_i(t) - X_{i+1}(t)$ hits zero, the forward rate

$$\frac{1}{T_{i+1} - T_i}\left(\frac{X_i}{X_{i+1}} - 1\right) \tag{8}$$

also hits zero, and if the increment were to become negative, so would the forward rate. Since negative interest rates can create serious anomalies, we would like to ensure that even after discretization the increments $\hat{X}_i - \hat{X}_{i+1}$ remain positive.

3.1 Discretization

To this end, we define

$$Y_1 = g(X_1), \quad Y_i = g\left(X_i / X_{i-1}\right), \quad i = 2, \ldots, N,$$

where $g : (0,1) \to \mathbf{R}$, as before, is increasing, surjective, and twice continuously differentiable. Inverting this transformation gives

$$X_i = \prod_{j=1}^{i} f(Y_j),$$

with f the inverse of g.

To make the dynamics of $Y = (Y_1, \ldots, Y_N)$ explicit, let

$$\sigma_i = \begin{pmatrix} \sigma_{i1} \\ \sigma_{i2} \\ \vdots \\ \sigma_{id} \end{pmatrix} \quad \text{and} \quad W(t) = \begin{pmatrix} W_1(t) \\ W_2(t) \\ \vdots \\ W_d(t) \end{pmatrix},$$

with the σ_{ij} as in (1). Using angle brackets to denote quadratic variation, Ito's rule gives

$$\begin{aligned} d\left(\frac{X_{i+1}}{X_i}\right) &= \frac{dX_{i+1}}{X_i} - \frac{X_{i+1}dX_i}{X_i^2} - \frac{d < X_i, X_{i+1} >}{X_i^2} + \frac{X_{i+1} d < X_i >}{X_i^3} \\ &= \frac{X_{i+1}}{X_i}\sigma_{i+1}^\top \, dW - \frac{X_{i+1}}{X_i}\sigma_i^\top \, dW - \frac{X_{i+1}}{X_i}\sigma_{i+1}^\top \sigma_i \, dt + \frac{X_{i+1}}{X_i}\sigma_i^\top \sigma_i \, dt \\ &= -\frac{X_{i+1}}{X_i}(\sigma_{i+1} - \sigma_i)^\top \sigma_i \, dt + \frac{X_{i+1}}{X_i}(\sigma_{i+1} - \sigma_i)^\top \, dW, \end{aligned} \tag{9}$$

each σ_j evaluated at $X(t)$. It now follows from a further application of Ito's rule that

$$\begin{aligned} dY_1 &= \tfrac{1}{2} g''(X_1) X_1^2 \sigma_1^\top \sigma_1 \, dt + g'(X_1) X_1 \sigma_1^\top \, dW \\ &\triangleq a_1(Y_1)\, dt + b_1(Y_1)^\top \, dW, \end{aligned} \tag{10}$$

and

$$\begin{aligned} dY_{i+1} &= \left[\tfrac{1}{2} g''(\frac{X_{i+1}}{X_i})(\frac{X_{i+1}}{X_i})^2 (\sigma_{i+1} - \sigma_i)^\top (\sigma_{i+1} - \sigma_i) - g'(\frac{X_{i+1}}{X_i})(\frac{X_{i+1}}{X_i})(\sigma_{i+1} - \sigma_i)^\top \sigma_i\right] dt \\ &\quad + g'(\frac{X_{i+1}}{X_i})(\frac{X_{i+1}}{X_i})(\sigma_{i+1} - \sigma_i)^\top \, dW \\ &\triangleq a_{i+1}(Y_{i+1})\, dt + b_{i+1}^\top (Y_{i+1})\, dW, \quad i = 1, \ldots, N-1. \end{aligned} \tag{11}$$

An Euler scheme for Y with step size h has the form

$$\hat{Y}((k+1)h) = \hat{Y}(kh) + a_k(\hat{Y}(kh))h + b_k(\hat{Y}(kh))^\top Z(k+1)\sqrt{h}, \quad \hat{Y}(0) = Y(0), \tag{12}$$

where

$$Z(k) = \begin{pmatrix} Z_1(k) \\ Z_2(k) \\ \vdots \\ Z_d(k) \end{pmatrix}, \quad k = 1, 2, \ldots$$

are independent standard normal vectors. This in turn defines a discretization of the original process X if we define

$$\hat{X}_i(kh) = \prod_{j=1}^{i} f(\hat{Y}_j(kh)).$$

Moreover, for all $k = 0, 1, 2, \ldots$ we have

$$1 > \hat{X}_1(kh) > \hat{X}_2(kh) > \cdots > \hat{X}_N(kh) > 0,$$

almost surely, in view of the fact that f maps the real line into $(0, 1)$. We have therefore met conditions (ii) and (iii) of Section 1.

We proceed to verify that the proposed discretization scheme has the same orders of strong and weak convergence as a standard Euler scheme. To state these results we need to introduce two classes of functions. As before, we consider $f : \mathbf{R} \to (0, 1)$ with $f(-\infty) = 0$ and $f(\infty) = 1$. We say that $f \in \mathcal{D}$ if

(i) f' is strictly positive and bounded on $\mathbf{R}$.

(ii) $f(1 - f)/f'$ is bounded.

(iii) $f(1 - f)f''/(f')^2$ is bounded.

(iv) $f^2(1 - f)^2 f'''/(f')^3$ is bounded.

We say that $f \in \mathcal{D}'$ if in addition

(v) $f^3(1 - f)^3 f''''/(f')^4$ is bounded.

The important special cases $f = \Phi$ and the inverse logit transformation

$$f(y) = \frac{e^y}{1 + e^y} \qquad \left(g(x) = \log\left(\frac{x}{1 - x}\right)\right)$$

both belong to $\mathcal{D}'$. Write $\Delta\sigma_i$ for $\sigma_i - \sigma_{i-1}$, $i = 2, \ldots, N$, and $\Delta\sigma_1 \equiv \sigma_1$.

Theorem 3 *Suppose that*

(i) σ_1 *depends only on* X_1 *and* $\Delta\sigma_i$ *depends only on the ratio* X_i/X_{i-1}, $i = 2, \ldots, N$;

(ii) $\Delta\sigma_i$ *is Lipschitz continuous on* $[0, 1]$ *with* $\Delta\sigma_i(1) = 0$, $i = 1, \ldots, N$.

Then the SDE (1) admits a unique strong solution for every initial condition $1 > X_1(0) > X_2(0) > \cdots > X_N(0) > 0$ *such that* $1 > X_1(t) > X_2(t) > \cdots > X_N(t) > 0$ *for all* t*. Moreover, if we define* $\hat{Y}$ *as in (12) for any* $f \in \mathcal{D}$ *and* $g = f^{-1}$*, then*

$$\hat{X}_i(kh) = \prod_{j=1}^{i} f(\hat{Y}_j(kh)), \quad i = 1, \ldots, N,$$

converge to X *with strong order* $1/2$*.*

Through the expression in (8) for the forward rate and from (9) (but with i and $i+1$ reversed) we may interpret condition (ii) of this theorem as stating that the diffusion coefficient of each forward rate is a function of that forward rate only.

For the weak convergence order we need an additional definition: a function is in C_p^k if the function and its derivatives up to order k are continuous and polynomially bounded. We now have

Theorem 4 *Suppose* $f \in \mathcal{D}$ *and in addition, for* $i = 1, \ldots, N$*,*

(i) $\Delta\sigma_i \in C^4([0,1])$*,*

(ii) $a_i, b_i \in C_p^4$*.*

Then $\hat{X}$ *converges to* X *with weak order 1. Condition (ii) is automatically satisfied if* f *is* Φ *or the inverse logit transformation.*

Theorem 3 is proved in an appendix; the existence and uniqueness of the solution to the SDE is a prerequisite to everything else in the paper. The proof of Theorem 4 amounts to a verification of the conditions in Theorem 14.5.2 of Kloeden and Platen [12] under the stated hypotheses. As the calculations involved are lengthy but routine, we omit them.

3.2 First-Order Property and Second-Order Adjustment

We have thus far verified that discretizing the X_i by transforming first to the Y_i, applying an Euler discretization, and then transforming back results in a convergent scheme that preservers the ordering of the variables on $(0,1)$. We turn next to the martingale property. Using the scheme above, the $\hat{X}$ will not in general be martingales, and in this multidimensional setting explicitly identifying a drift adjustment similar to the one in Section 2 appears to be infeasible. However, the following result indicates that, in a precise sense, the martingale property automatically holds to first order:

Theorem 5 *Define $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_N)$ by setting $\hat{Y}(0) = Y(0)$*

$$\hat{Y}_i((k+1)h) = \hat{Y}_i(kh) + a_i(\hat{Y}(kh))h + b_i(\hat{Y}(kh))^\top Z(k+1)\sqrt{h}, \quad k = 0, 1, \ldots,$$

with a_i and b_i as in (10)-(11). Set

$$\hat{X}_i = \prod_{j=1}^{i} f(\hat{Y}_j), \quad i = 1, \ldots, N.$$

Then

$$\frac{d}{dh} E\left[\prod_{j=1}^{i} f(\hat{Y}_j(h))\right]\Bigg|_{h=0} = 0, \quad i = 1, \ldots, N.$$

We interpret this result as stating that

$$\mathsf{E}[\hat{X}((k+1)h)|\hat{X}(kh)] = \hat{X}(kh) + o(h),$$

so that in transforming variables before discretizing we have maintained the ordering of the components of X while partially preserving the martingale property.

This result indicates a positive feature of the transformation method, but it also suggests a strategy for further improvement of the method. We cannot entirely eliminate the $o(h)$ term (as we did in the one-dimensional setting) but we can perhaps reduce it to $o(h^2)$ through a modification of the drift term in the Euler scheme for $\hat{Y}$.

Consider, then, a scheme of the form

$$\tilde{Y}_i((k+1)h) = \tilde{Y}_i(kh) + [a_i(\tilde{Y}(kh)) + e_i(\tilde{Y}(kh))h]h + b_i(\tilde{Y}(kh))^\top Z(k+1)\sqrt{h}, \quad i = 1, \ldots, N,$$

with the a_i as before and with the $e_i(y)$ chosen to satisfy

$$\frac{d^2}{dh^2} E\left[\prod_{j=1}^{i} f(\tilde{Y}_j(h))|\tilde{Y}(0) = y\right]\Bigg|_{h=0} = 0, \quad i = 1, \ldots, N.$$

The e_i do not admit a simple closed-form expression; however, they can be fairly easily evaluated at each step of a simulation. Details of the evaluation of the e_i's are given at the beginning of the proof of Theorem 6 in the Appendix. As before, we then set

$$\tilde{X}_i = \prod_{j=1}^{i} f(\tilde{Y}_j), \quad i = 1, \ldots, N.$$

Clearly, this method continues to ensure

$$1 > \tilde{X}_1(kh) > \tilde{X}_2(kh) > \cdots > \tilde{X}_N(kh) > 0,$$

and by construction it enforces the martingale property to terms of order h^2. We now verify that under appropriate conditions its convergence orders are no worse than those of a standard Euler scheme. The only assumption required beyond those in Theorems 3 and 4 is that f belong to $\mathcal{D}'$ rather than merely $\mathcal{D}$.

Theorem 6 *Under the conditions of Theorem 3 but with $f \in \mathcal{D}'$, the $\tilde{X}_i$ converge to the X_i with strong order* $1/2$. *Under the conditions of Theorem 4 but with $f \in \mathcal{D}'$ we have convergence with weak order 1.*

4 Numerical Examples

Our analysis has focused on ensuring that the martingale property and bounds on variables are preserved after discretization, and further ensuring that this is achieved without a decrease in the convergence order in comparison with a standard Euler scheme. In practice, it is also important that the discretization provide a good approximation to the law of the continuous-time process, even if the time step h is not very small. To gauge the quality of the approximation, we present a few numerical examples. These suggest that the transformation methods we have proposed provide a better approximation than a standard Euler scheme.

We present three types of examples. We start with a one-dimensional model, then consider a ten-dimensional example, and finally compare the performance of the methods in pricing interest rate *caps*, which are simply options on forward rates. In comparing multiple methods on a single problem, we keep the computer time fixed for all methods by varying the time-step h. For example, if implementing the second-order adjustment takes twice the time of an ordinary Euler scheme, we use a time-step half as small for the Euler scheme as for the adjusted scheme. This puts the various methods on level ground in the comparison.

4.1 One-Dimensional Case

With $f = \Phi$, the exact martingale adjustment can be established as in (7) for a one-dimensional model. We consider the following example:

$$dX(t) = -\lambda X(t)(1 - X(t))dW(t), \quad 0 < X(0) = x_0 < 1,$$

where $\lambda > 0$ is a constant. For all t, we have $0 < X(t) < 1$.

To evaluate the quality of any discretization scheme, we need to have exact results against which to compare. To this end, we will examine the distribution not of $X(t)$ itself (which is rather complex) but rather of the "forward rate" implied by this deflated bond price, namely

$$L(t) \stackrel{\triangle}{=} \frac{1}{X(t)} - 1.$$

Its dynamics are given by

$$\frac{dL(t)}{L(t)} = \frac{\lambda^2 L(t)}{1+L(t)}\,dt + \lambda\,dW(t), \quad L(0) = \frac{1}{x_0} - 1.$$

Define a new measure by

$$\left.\frac{d\hat{\mathbf{P}}}{d\mathbf{P}}\right|_{\mathcal{F}_t} \stackrel{\triangle}{=} \exp\left\{-\int_0^T \lambda(1-X(t))dW(t) - \frac{1}{2}\lambda^2\int_0^T (1-X(t))^2\,dt\right\} = \frac{X(t)}{x_0};$$

then by Girsanov's Theorem (see, e.g., Karatzas and Shreve [13], Theorem 3.5.1) $\hat{W}(t) \stackrel{\triangle}{=} W(t) + \int_0^t \lambda(1-X(s))\,ds$ is a standard Brownian motion under this new measure. Moreover, under $\hat{\mathbf{P}}$ we have

$$\frac{dL(t)}{L(t)} = \lambda\,d\hat{W}(t)$$

since $1 - X(t) = L(t)/(1+L(t))$. This implies that $L(t)$ is lognormally distributed under $\hat{\mathbf{P}}$ and thus

$$\mathsf{E}\left[\frac{X(t)}{x_0}\mathbf{1}_{\left\{\frac{1}{X(t)}-1\le x\right\}}\right] = \hat{\mathbf{P}}(L(t)\le x) = \Phi\left(\frac{\log\left(\frac{x}{L(0)}\right) + \frac{1}{2}\lambda^2 t}{\lambda\sqrt{t}}\right).$$

We will use this formula to check the quality of the simulated distribution. Figure 1 compares the estimated bias for $\hat{\mathbf{P}}(L(T)\le x)$ for three different methods: a standard Euler scheme for the original process $X(t)$, a standard Euler scheme for the transformed process $Y(t) \stackrel{\triangle}{=} \Phi^{-1}(X(t))$, and a scheme with the martingale adjustment applied to $Y(t)$ as in (7). We use the following parameter values: time horizon $T = 10$, "volatility" $\lambda = 0.5$, and initial value $x_0 = 0.95$ (corresponding to an interest rate $L(0)$ of about 5.26%).

The results in Figure 1 show that using the transformation (with or without martingale adjustment) substantially improves the fit of the distribution. The bias curve for the ordinary Euler scheme jumps at 0 showing that a certain proportion of $\hat{X}(T)$s go beyond 1. It also has a large spike indicating that too many simulated values are close to 1. Without the martingale adjustment, the Euler scheme for $Y(t)$ introduces a small bias in $\mathsf{E}[X(T)]$ (visible by examining the limit as $x\to\infty$ in the figure), while the martingale adjustment (7) erases this bias successfully.

4.2 Multidimensional Case

The diffusion model we consider is

$$X_0(t) \;\equiv\; 1;$$

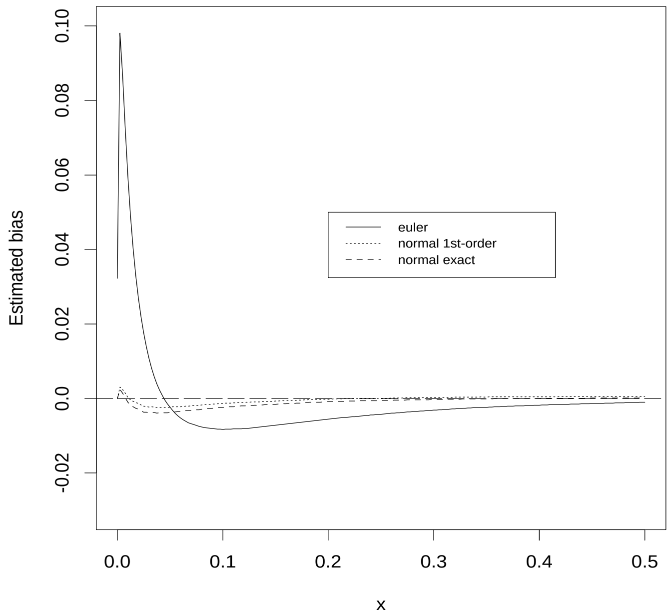

Figure 1: One-Dimensional Model

$$\begin{aligned} dX_i(t) &= -\lambda X_i(t) \cdot \left(\sum_{j=1}^{i} \left(1 - \frac{X_j(t)}{X_{j-1}(t)} \right) \right) dW(t), \quad i = 1, \ldots, N; \\ & 1 > X_1(0) > X_2(0) > \cdots > X_N(0) > 0, \end{aligned}$$

with $W(t)$ a standard one-dimensional Brownian motion (i.e. $d = 1$) and $\lambda > 0$ is a constant. These processes will always be ordered and bounded, i.e., $1 > X_1(t) > X_2(t) > \cdots > X_N(t) > 0$ at every $t \geq 0$.

As in the one-dimensional case, we derive closed-form expressions against which to compare by introducing forward rates and changing measures: let

$$\begin{aligned} L_i(t) &\triangleq \frac{X_{i-1}(t)}{X_i(t)} - 1, \qquad i = 1, 2, \cdots, N, \\ \left.\frac{d\hat{\mathbf{P}}_i}{d\mathbf{P}}\right|_{\mathcal{F}_t} &\triangleq \frac{X_i(t)}{X_i(0)}; \\ \hat{W}^{(i)}(t) &\triangleq W(t) + \lambda \int_0^t \sum_{j=1}^{i} \left(1 - \frac{X_j(s)}{X_{j-1}(s)} \right) ds. \end{aligned}$$

It follows readily that $dL_i(t) = \lambda L_i(t) d\hat{W}^{(i)}(t)$ and, much as in the one-dimensional case,

$$\mathsf{E}\left[\frac{X_i(t)}{X_i(0)} \mathbf{1}_{\left\{ \frac{X_{i-1}(t)}{X_i(t)} - 1 \leq x \right\}} \right] = \hat{\mathbf{P}}_i(L_i(t) \leq x) = \Phi\left(\frac{\log\left(\frac{x}{L_i(0)}\right) + \frac{1}{2}\lambda^2 t}{\lambda\sqrt{t}} \right).$$

We compare simulated estimates of this expectation with the formula. The simulation uses a time horizon $T = 10$, dimension $N = 10$, constant coefficient $\lambda = 0.5$, and initial value $X_i(0) = 0.95^i$, $i = 1, \cdots, 10$.

No exact martingale adjustment is available in this multidimensional setting. Instead, we carry out the second-order adjustments. As discussed in Section 3.2, there is no simple closed-form expression for the adjustment parameters e_i, but they can be evaluated fairly easily at each step of a simulation. We consider two transformations in this example — namely, $f = \Phi$ and the inverse logit transformation $f(x) = e^x/(1+e^x)$. Figures 2, 3, and 4 show estimated bias curves for $X_1(T)$, $X_5(T)$, and $X_{10}(T)$ and thus give an indication of the performance across coordinates as well as across methods. In each case, we compare five schemes: a standard Euler scheme for the original process, a standard Euler scheme for the transformed process $Y_i(t) \stackrel{\triangle}{=} f^{-1}(X_i(t))$ (with $f = \Phi$ or the inverse logit function), and discretizations with the second-order adjustment for $Y(t)$.

The bias curves from the standard Euler scheme display a jump at 0, indicating that $\hat{X}_{i+1}(T)$ exceeds $\hat{X}_i(T)$ with positive probability. The large spike shows that a disproportionate fraction of the $\hat{X}_{i+1}(T)$ and $\hat{X}_i(T)$ are very close (roughly corresponding to forward rates that are too close to zero). We get a better fit to the target distribution using either of the transformations. As in the one-dimensional case, the Euler scheme for $Y_i(t)$ without adjustment introduces bias (more pronounced for larger i). Examination of the bias curves for large values of x indicates that with the second-order adjustment the martingale property is well preserved.

Figures 2, 3, and 4 are consistent with a broader pattern we have observed in other numerical examples as well: the transformation methods are most effective for the lower-indexed components of the vector X, while the benefit of the second-order adjustment is most pronounced for the higher-indexed components.

4.3 Pricing Caps

Interest rate *caps* are among the most actively traded of all interest rate options, so the ability to price caps accurately is a prerequisite for any computational procedure in this context. To assess the performance of our discretization schemes in pricing caps, we need to enrich the setting slightly.

Fix a set of maturity dates $0 = T_0 < T_1 < \cdots < T_N < T_{N+1}$ and for simplicity suppose they are evenly spaced with an increment of $\delta \equiv T_{i+1} - T_i$ $i = 0, \cdots, N$. We take $\delta = 0.25$, corresponding to a quarter of a year. Define $\eta(t)$ to be the unique integer for which $T_{\eta(t)-1} < t \leq T_{\eta(t)}$. The model takes the form

$$\begin{aligned} \frac{dX_{n+1}(t)}{X_{n+1}(t)} &= -\lambda \sum_{i=\eta(t)}^{n} \left(1 - \frac{X_{i+1}(t)}{X_i(t)}\right) dW(t), \qquad n = 0, 1, \ldots, N, \\ & \quad 1 > X_1(0) > X_2(0) > \cdots > X_N(0) > 0. \end{aligned}$$

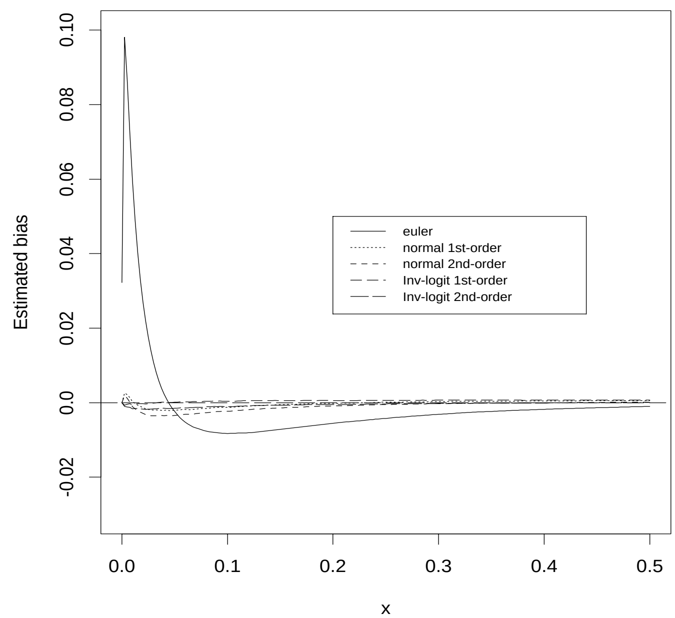

Figure 2: Multidimensional Model, X_1

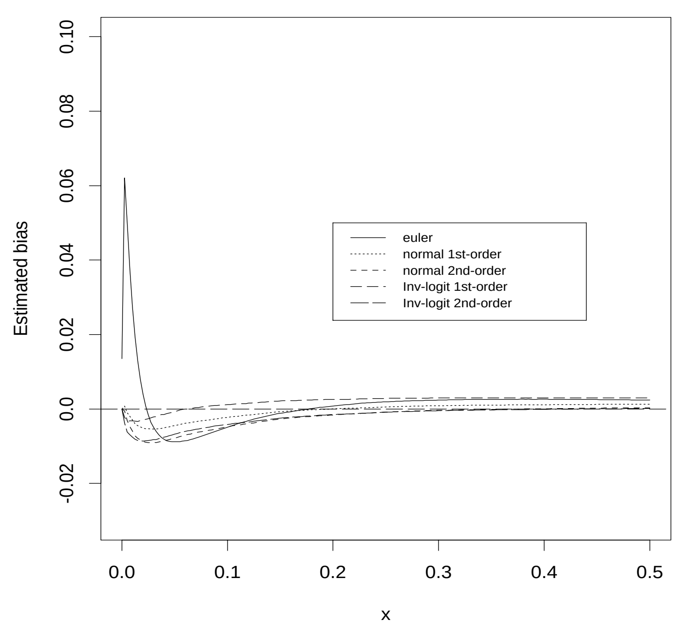

Figure 3: Multidimensional Model, X_5

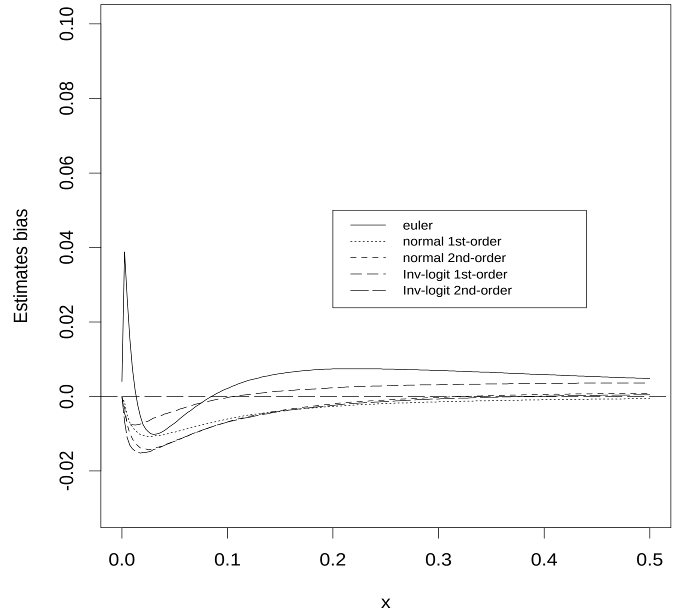

Figure 4: Multidimensional Model, X_{10}

Each $X_i(t)$ should be interpreted as the deflated price of a bond maturing at T_i. This model differs from the previous one only in that the range of summation in the diffusion coefficient varies over time, in a manner consistent with the formulation in Jamshidian [11]. Although, for simplicity, we have not explicitly considered time-varying coefficients in our analysis, the proposed transformations can still be applied in this setting.

From the deflated bond prices we can define forward interest rates. The forward rate at time t for the accrual period $[T_i, T_{i+1}]$, $t \le T_i$ is

$$L_i(t) = \frac{1}{\delta}\left(\frac{X_i(t)}{X_{i+1}(t)} - 1\right), \quad i = 1, \ldots, N,$$

which evolves according to

$$\frac{dL_n(t)}{L_n(t)} = \sum_{i=\eta(t)}^{n} \frac{\delta\lambda^2 L_i(t)}{1+\delta L_i(t)} dt + \lambda\, dW(t), \quad n = 1, \ldots, N.$$

For background and a more detailed discussion, see Jamshidian [11].

An interest rate cap for the period $[T_n, T_{n+1}]$ (a *caplet*) pays the holder $\delta(L_n(T_n) - K)^+$ at time T_{n+1}, where the constant K is the *strike*. As shown in Jamshidian [11], the value of this option is given by $C(\lambda, K, L_n(0), B_{n+1}(0), T_n)$, where

$$C(\sigma, K, r, b, T) \triangleq \delta b \left[r\Phi\left(\frac{\log\left(\frac{r}{K}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right) - K\Phi\left(\frac{\log\left(\frac{r}{K}\right) - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right)\right] \tag{13}$$

and $B_{n+1} = X_{n+1}(0)(1 + \delta L_0(0))$. (The factor $1 + \delta L_0(0)$ is the initial value of the numeraire and is arbitrary; we take $L_0(0) = .04$.) This expression evaluates

$$B_{n+1}(0)\hat{\mathsf{E}}_{n+1}[\delta(L_n(T_n) - K)^+] = (1 + \delta L_0(0))\mathsf{E}\left[X_{n+1}(T_{n+1}) \cdot \delta(L_n(T_n) - K)^+\right].$$

The expectation on the right can be estimated by simulation and then compared with the formula in (13) to estimate the discretization bias.

We use the following model parameters: the time horizon is $T = 10$ years with a total of $N = 40$ bonds (since $\delta = 1/4$), the volatility coefficient is $\lambda = 0.5$, initial values are $X_i(0) = 0.99^i$. We consider at-the-money caplets, meaning that for the n-th caplet we use a strike of $K = L_n(0)$.

Figure 5 shows the estimated bias for the five schemes used previously. The results are based on four million replications (using a control variate for variance reduction) which suffices to make the standard error a small fraction of the estimated bias. The apparent biases in the graphs are thus statistically significant. The time-increments for the various schemes were chosen to balance the computer times per path; specifically, the second-order (resp. first-order) adjustment will usually take three (resp. two) times as long per step as an ordinary Euler scheme so we use a time increment one third (resp. half) as small for the Euler scheme as for the second-order (resp. first-order) adjusted scheme. We use the same number of paths for all methods, so the faster time per step associated with the Euler scheme is used exclusively to reduce its bias through a smaller time increment. (For an analysis of the tradeoff between bias and variance in simulations of SDE, see Duffie and Glynn [4].)

All the transformation schemes appear to be biased low across all maturities whereas the standard Euler scheme appears to be biased high. The smaller bias observed in the previous examples does not translate to a markedly smaller bias in caplet prices in this example. The bias in the nth caplet price is closely related to the integral over x from K to ∞ of the bias in estimating $\hat{\mathbf{P}}_{n+1}(L_n(t) > x)$. The large positive and negative spikes typical of the bias in the Euler scheme in estimating such probabilities makes the caplet bias using this method very sensitive to the choice of K, with larger K typically resulting in a more positive caplet bias over the relevant range of strike prices. Similarly, the shape of the bias curves in Figures 1-4 for the transformation methods is consistent with the negative bias observed in the caplet prices. It is worth emphasizing, however, that the transformation methods have other desirable properties and that Figure 5 suggests that these properties have not come at the expense of inferior performance in pricing caplets.

5 Concluding Remarks

We have investigated and analyzed the use of nonlinear transformations in discretizing a class of continuous-time martingales restricted to a wedge-shaped region of $\mathbf{R}^N$. The martingale condition and the inequalities defining the domain of the processes have important implications in the financial applications that motivate this investigation; care should therefore be taken to try to preserve these properties in numerical work. Our results indicate that by applying nonlinear transformations

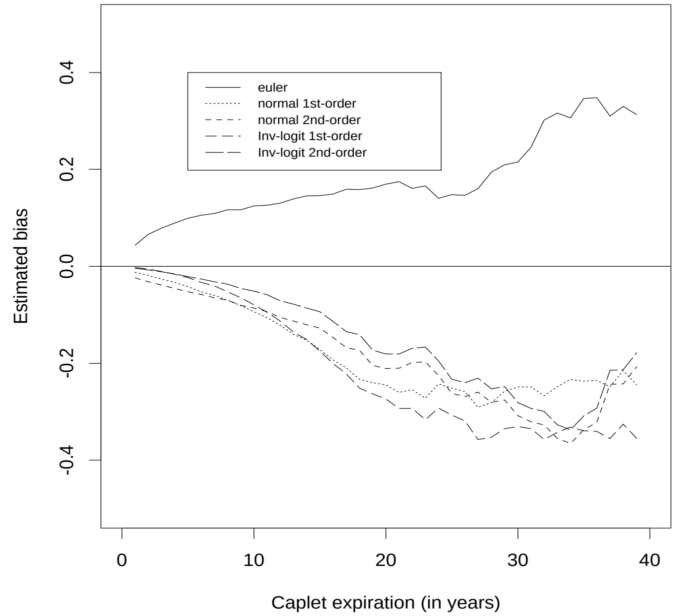

Figure 5: Caplet prices

before time-discretization we can better preserve these properties with no loss in the convergence order of the discretization (compared with a standard Euler scheme) and in some cases with a much better fit to the ideal continuous-time distribution.

Appendix: Proofs

Proof of Theorem 1

We need to prove that $\mathsf{E}[|X(T) - \Phi(\hat{Y}(T))|] \leq c\sqrt{h}$. (Here and throughout the proof, $c, c_1, c_2, \ldots$ represent unspecified constants.) Because Φ is Lipschitz, it suffices to prove

$$\mathsf{E}[|Y(T) - \hat{Y}(T)|] \leq c\sqrt{h}. \tag{14}$$

We will show that

(I) $\tilde{\mu}$ and $\tilde{\sigma}$ are Lipschitz continuous;

(II) $\tilde{\mu}^2(y) \leq K$ and $\tilde{\sigma}^2(y) \leq K$, for some $K > 0$.

We leave verification of these two properties until the end, and proceed to show that they imply (14).

With h the time increment, let $n_t = \lfloor t/h \rfloor$, $t_n = nh$ and write $\hat{Y}_i$ for $\hat{Y}(ih)$. In the construction of $\hat{Y}$, let the normal random variables $\sqrt{h}Z_i$ be the increments $\Delta W_i = W_{ih} - W_{(i-1)h}$ of the Brownian motion defining X and let $\{\mathcal{F}_t, t \geq 0\}$ be the standard filtration generated by the Brownian motion.

Extend $\hat{Y}$ to arbitrary times by setting $\hat{Y}(t) = \hat{Y}(n_t h)$. Set $Z(t) = \sup_{0\le s\le t} \mathsf{E}[|\hat{Y}_{n_s} - Y_s|^2]$ so

$$\begin{aligned}
Z(t) &= \sup_{0\le s\le t} E\left[\left|\sum_{n=0}^{n_s-1}(\hat{Y}_{n+1}-\hat{Y}_n) - \int_0^s \tilde{\mu}(Y_u)\,du - \int_0^s \tilde{\sigma}(Y_u)\,dW_u\right|^2\right] \\
&\le c_1 \sup_{0\le s\le t}\left\{E\left|\sum_{n=0}^{n_s-1}(\mathsf{E}[\hat{Y}_{n+1}-\hat{Y}_n|\mathcal{F}_{nh}) - \tilde{\mu}(\hat{Y}_n)h)\right|^2\right. \\
&\quad +E\left|\sum_{n=0}^{n_s-1}(\hat{Y}_{n+1}-\hat{Y}_n) - \mathsf{E}[\hat{Y}_{n+1}-\hat{Y}_n|\mathcal{F}_{nh}) - \tilde{\sigma}(\hat{Y}_n)\Delta W_n)\right|^2 \\
&\quad +E\left|\int_0^{t_{n_s}}[\tilde{\mu}(\hat{Y}_{n_u}) - \tilde{\mu}(Y_u)]\,du\right|^2 + E\left|\int_0^{t_{n_s}}[\tilde{\sigma}(\hat{Y}_{n_u}) - \tilde{\sigma}(Y_u)]\,dW_u\right|^2 \\
&\quad \left. +E\left|\int_{t_{n_s}}^s \tilde{\mu}(Y_u)\,du\right|^2 + E\left|\int_{t_{n_s}}^s \tilde{\sigma}(Y_u)\,dW_u\right|^2\right\} \\
&\equiv A_1 + A_2 + A_3 + A_4 + A_5 + A_6.
\end{aligned}$$

By construction, $A_2 = 0$. By property (I) above

$$A_3 + A_4 \le c\int_0^t E(\hat{Y}_{n_s} - Y_s)^2\,ds \le c\int_0^t Z(s)\,ds,$$

and by property (II)

$$A_5 + A_6 \le c_2 h.$$

Next we analyze A_1, which can be rewritten as

$$A_1 = E\left|\sum_{n=0}^{n_s-1}[\mu(\hat{Y}_n) - \tilde{\mu}(\hat{Y}_n)]^2\right|.$$

In light of (5), $\tilde{\mu}(y)/y > 0$ for all y. From (6) and the fact that $|\sqrt{1+2a} - 1 - a| \le a^2/2$ for $a \ge 0$, we may therefore conclude that

$$|\mu(y) - \tilde{\mu}(y)| \le \frac{1}{h}\cdot\tfrac{1}{2}h^2\frac{\tilde{\mu}^2(y)}{|y|} = \tfrac{1}{2}\frac{\tilde{\mu}^2(y)}{|y|}h.$$

If we can establish that

(III) $\sup_{y\in\mathbf{R}} \frac{\tilde{\mu}^2(y)}{|y|} < \infty$,

it will follow that $A_1 \le c_3 h^2$ and

$$Z(t) \le c_1\left(c_3 h^2 + K\int_0^t Z(s)\,ds + c_2 h\right) \le c_4\int_0^t Z(s)\,ds + c_5 h.$$

By Gronwall's inequality, this implies $Z(t) \le c_6 h$ and finally

$$E|Y_T - \hat{Y}_T| \le \sqrt{Z(T)} \le c_7\sqrt{h}.$$

It remains to establish properties (I)-(III). For (III) observe that

$$|\sigma(x)| = |\sigma(x) - \sigma(1)| \le c(1-x),$$

since $\sigma(\cdot)$ is Lipschitz with $\sigma(1) = 0$. It is not difficult to verify that

$$\frac{\tilde{\mu}^2(y)}{|y|} = \frac{|y|}{4\phi^4(y)}\sigma^4(\Phi(y))\Phi^4(y) = O(1/|y|^3), \quad \text{as} \quad |y| \to \infty,$$

using $\Phi(y) \sim \phi(y)/|y|$ as $y \to -\infty$.

From the proof of (III) we have $\tilde{\mu}^2(y) = O(\frac{1}{y^2})$ as $|y| \to \infty$. Similarly,

$$\tilde{\sigma}^2(y) = \frac{1}{\phi^2(y)}\sigma^2(\Phi(y))\Phi^2(y) = O(\frac{1}{y^2}), \quad |y| \to \infty,$$

Therefore, $\tilde{\mu}^2(y) \le K$ and $\tilde{\sigma}^2(y) \le K$ for some positive constant K and (II) holds.

To establish (I) it suffices (through, e.g., exercise 17.23 of Hewitt and Stromberg [10]) to show that

$$\limsup_{h\downarrow 0} \frac{1}{h}[\tilde{\mu}(y+h) - \tilde{\mu}(y)] \quad \text{and} \quad \limsup_{h\downarrow 0} \frac{1}{h}[\tilde{\sigma}(y+h) - \tilde{\sigma}(y)]$$

are bounded independent of y. Explicit calculation verifies that the limsups are indeed bounded. We omit the details. □

Proof of Theorem 2

It suffices to show that $\hat{Y}$ converges to Y with weak order 1 by checking that conditions (14.5.7)–(14.5.12) in Theorem 14.5.2 of Kloeden and Platen [12] are satisfied. This is to verify, in our context, that for any $p = 1, 2, \ldots$, there exists $C < \infty$, which does not depend on h, such that for any $q = 1, 2, \ldots, p$, we have

$$\mathsf{E}\left(\max_{0\le n\le n_T} |\hat{Y}_n|^{2q}\right) \le C;$$

$$\mathsf{E}\left(|\hat{Y}_{n+1} - \hat{Y}_n|^{2q}\Big|\, \mathcal{F}_{nh}\right) \le Ch^q;$$

$$\left|\mathsf{E}\left((\hat{Y}_{n+1} - \hat{Y}_n)^l - (\tilde{\mu}(\hat{Y}_n)h + \tilde{\sigma}(\hat{Y}_n)\sqrt{h}Z_{n+1})^l)|\mathcal{F}_{nh}\right)\right| \le Ch^2; \quad \text{for } l = 1, 2, 3$$

From the proof of Theorem 1 it follows that $\tilde{\mu}, \tilde{\sigma}$ are both bounded and Lipschitz continuous, and $|\mu(y) - \tilde{\mu}(y)| \le Kh$ holds for any real number y (in particular, μ is bounded too). The first inequality is immediate from the boundedness of $\tilde{\sigma}$ and μ (cf. Exercise 14.5.3 of Kloeden and Platen [12]). As to the second inequality, we have

$$\begin{aligned}
\mathsf{E}\left(|\hat{Y}_{n+1} - \hat{Y}_n|^{2q}\Big|\, \mathcal{F}_{nh}\right) &= \mathsf{E}\left(|\mu(\hat{Y}_n)h + \tilde{\sigma}(\hat{Y}_n)\sqrt{h}Z_{n+1}|^{2q}\Big|\, \mathcal{F}_{nh}\right) \\
&\le c_1\left(|\mu(\hat{Y}_n)h|^{2q} + |\tilde{\sigma}(\hat{Y}_n)\sqrt{h}Z_{n+1}|^{2q}\Big|\, \mathcal{F}_{nh}\right) \\
&\le c_1(c_2h^{2q} + c_3h^q) \\
&\le c_4h^q,
\end{aligned}$$

where the last inequality is from the boundedness of $\tilde{\mu}(\cdot)$ and $\tilde{\sigma}(\cdot)$. It is not difficult to check the remaining inequality via similar explicit calculation; we omit the details. $\square$

Proof of Theorem 3

It is sufficient to prove the boundedness and Lipschitz continuity of all the coefficients $\{a_i(\cdot)\}$ and $\{b_i(\cdot)\}$ in SDE (10)–(11). Actually, these properties imply the existence and uniqueness of the strong solution to SDE (10)–(11) (Theorem 5.2.9 of Karatzas and Shreve [13]). Moreover, $\hat{Y}_i$ converges to Y_i with strong order 1/2 (Theorem 9.6.2 of Kloeden and Platen [12]), so $\mathsf{E}|\hat{Y}_i - Y_i| \le c\sqrt{h}$, which yields

$$\begin{aligned}
\mathsf{E}|\hat{X}_i - X_i| &= \mathsf{E}\left|\prod_{j=1}^{i} f(\hat{Y}_j) - \prod_{j=1}^{i} f(Y_j)\right| \\
&\le \sum_{j=1}^{i} \mathsf{E}\left|f(\hat{Y}_j) - f(Y_j)\right| \\
&\le c_1 \sum_{j=1}^{i} \mathsf{E}\left|\hat{Y}_j - Y_j\right| \\
&\le c_2\sqrt{h},
\end{aligned}$$

where the first and second inequalities are from $0 \le f \le 1$ and the Lipschitz continuity of f. The rest of this subsection is devoted to the proof of the properties of boundedness and Lipschitz continuity.

We recall that,

$$\begin{aligned}
a_i(Y) &= \frac{1}{2} g''(f(Y_i)) \cdot f^2(Y_i) \cdot (\triangle\sigma_i^\top \triangle\sigma_i)(f(Y_i)) - g'(f(Y_i)) \cdot f(Y_i) \cdot \triangle\sigma_i(f(Y_i))^\top \sigma_{i-1}(Y), \\
b_i(Y) &= g'(f(Y_i)) \cdot f(Y_i) \cdot \triangle\sigma_i(f(Y_i)),
\end{aligned}$$

since $\triangle\sigma_i$ only depends on $X_i/X_{i-1} = f(Y_i)$.

Boundedness: Note that $g' \circ f = \frac{1}{f'}$, $g'' \circ f = -\frac{f''}{(f')^3}$, $\triangle\sigma_i$ is bounded, σ_{i-1} is bounded and $\| \triangle\sigma_i(f) \| \le c_1(1-f)$ (by Lipschitz continuity of $\triangle\sigma_i$ and $\triangle\sigma_i(1) = 0$). It follows that

$$\sup \| b_i \| \le c \cdot \sup \left| \frac{f(1-f)}{f'} \right| < \infty,$$

and

$$\begin{aligned}
\sup \| a_i \| &\le \sup\left|(g'' \circ f) \cdot f^2 \cdot \| \triangle\sigma_i \circ f \|^2\right| + \sup\left|(g' \circ f) \cdot f \cdot (\triangle\sigma_i \circ f)^\top \sigma_{i-1}\right| \\
&\le \sup\left|\frac{f^2 f''}{(f')^3} \cdot \| \triangle\sigma_i \circ f \|^2\right| + c\sup\left|\frac{f}{f'} \| \triangle\sigma_i \circ f \|\right| \\
&\le c_1 \cdot \sup\left|\frac{f^2(1-f)^2 f''}{(f')^3}\right| + c_2 \sup\left|\frac{f(1-f)}{f'}\right| < \infty,
\end{aligned}$$

where the last inequality follows readily from the fact that $f \in \mathcal{D}$.

Lipschitz continuity of b_i: Note that $b_i = (b_{i1}, b_{i2}, \cdots, b_{id})$ is a $\mathbf{R}^d$–valued function that only depends on Y_i. The proof is therefore similar to the one-dimensional case. We only need to prove for any $j = 1, 2, \ldots, d$, the boundedness of

$$\limsup_{h \downarrow 0} \frac{1}{h}[b_{ij}(y+h) - b_{ij}(y)]$$

Since $b_{ij}(y) = g'(f(y)) \cdot f(y) \cdot \triangle\sigma_{ij}(f(y)) = \frac{f(y)}{f'(y)} \triangle\sigma_{ij}(f(y))$, it follows that

$$\begin{aligned}
\frac{1}{h}\left|b_{ij}(y+h) - b_{ij}(y)\right| &\leq \frac{1}{h}\left|\frac{f(y+h)}{f'(y+h)}\left(\triangle\sigma_{ij}(f(y+h)) - \triangle\sigma_{ij}(f(y))\right)\right| \\
&\quad + \frac{1}{h}\left|\triangle\sigma_{ij}(f(y))\left(\frac{f(y+h)}{f'(y+h)} - \frac{f(y)}{f'(y)}\right)\right| \\
&\triangleq A + B.
\end{aligned}$$

However, by Lipschitz continuity of $\triangle\sigma_{ij}$, we have

$$\limsup_{h \downarrow 0} A \leq c \limsup_{h \downarrow 0} \frac{1}{h}\left|\frac{f(y+h)}{f'(y+h)}(f(y+h) - f(y))\right| = cf(y) \leq c,$$

and from the fact that $|\triangle\sigma_{ij}(x) - 1| \leq cx$ we obtain

$$\begin{aligned}
\limsup_{h \downarrow 0} B &\leq c(1 - f(y)) \limsup_{h \downarrow 0} \frac{1}{h}\left|\frac{f(y+h)}{f'(y+h)} - \frac{f(y)}{f'(y)}\right| \\
&= c(1 - f(y))\left|\left(\frac{f}{f'}\right)'(y)\right| = c(1 - f(y))\left|1 - \frac{f(y)f''(y)}{(f'(y))^2}\right| \\
&\leq c(1 + \sup\left|\frac{f(1-f)f''}{(f')^2}\right|) < \infty,
\end{aligned}$$

where the last inequality holds since $f \in \mathcal{D}$.

Lipschitz continuity of a_i: It is immediate that

$$g'(f(Y_i)) \cdot f(Y_i) \cdot \triangle\sigma_i(f(Y_i))^\top \sigma_{i-1}(Y) = b_i(Y)^\top \sigma_{i-1}(Y)$$

is Lipschitz continuous since both b_i and σ_{i-1} are bounded and Lipschitz continuous. Hence it suffices to show the Lipschitz continuity of $\phi_i \triangleq \frac{1}{2}(g'' \circ f) \cdot f^2 \cdot (\triangle\sigma_i^\top \triangle\sigma_i) \circ f$. As before, we need to prove that

$$\limsup_{h \downarrow 0} \frac{1}{h}[\phi_i(y+h) - \phi_i(y)]$$

is bounded. The proof is similar to that of Lipschitz continuity of b_i and is omitted. □

Proof of Theorem 5

Let $y_j \triangleq Y_j(0), A_j \triangleq a_j(Y(0))$ and $B_j \triangleq b_j(Y(0))$ for $j = 1, \cdots, N$. We have

$$f(\tilde{Y}_j(h)) = f(y_j) + f'(y_j)B_j^\top Z\sqrt{h} + \left(f'(y_j)A_j + \frac{1}{2}f''(y_j)(B_j^\top Z)^2\right)h + o(h),$$

which implies, with $\mathsf{E}Z = 0,\ \mathsf{E}(Z_iZ_j) = \delta_{ij}$,

$$\begin{aligned}
\frac{d}{dh}E\left[\prod_{j=1}^{i} f(\tilde{Y}_j(h))\right]\Bigg|_{h=0} &= \sum_{j=1}^{i}\left(u_j\prod_{m\neq j} f(y_m)\right) + \sum_{1\le m<j\le i}\left(f'(y_m)f'(y_j)B_m^\top B_j \cdot \prod_{l\neq m,j} f(y_l)\right) \\
&= \prod_{m=1}^{i} f(y_m)\cdot\left(\sum_{j=1}^{i}\frac{u_j}{f(y_j)} + \sum_{1\le m<j\le i}\frac{f'(y_m)f'(y_j)}{f(y_m)f(y_j)}B_m^\top B_j\right).
\end{aligned}$$

Here $u_j \triangleq A_jf'(y_j) + \frac{1}{2}f''(y_j)B_j^\top B_j$. We only need to prove that, for every $1 \le i \le N$,

$$\frac{u_i}{f(y_i)} + \frac{f'(y_i)}{f(y_i)}\sum_{0\le m<i}\frac{f'(y_m)}{f(y_m)}\cdot(B_m^\top B_i) = 0,$$

or equivalently,

$$u_i = -f'(y_i)\sum_{0\le m<i}\frac{f'(y_m)}{f(y_m)}\cdot(B_m^\top B_i) = -f'(y_i)\left(\sum_{0\le m<i}\frac{f'(y_m)}{f(y_m)}B_m\right)^\top B_i.$$

However, by definition,

$$B_m = b_m(Y(0)) = g'(f(y_m))f(y_m)\triangle\sigma_m = \frac{f(y_m)}{f'(y_m)}\triangle\sigma_m.$$

Hence

$$\sum_{0\le m<i}\frac{f'(y_m)}{f(y_m)}B_m = \sum_{0\le m<i}\frac{f'(y_m)}{f(y_m)}\frac{f(y_m)}{f'(y_m)}\triangle\sigma_m = \sum_{0\le m<i}\triangle\sigma_m = \sigma_{i-1}$$

and we need only to prove $u_i = -f'(y_i)\sigma_{i-1}^\top B_i$. But by definition,

$$\begin{aligned}
A_i = a_i(Y(0)) &= \frac{1}{2}g''(f(y_i))f^2(y_i)(\triangle\sigma_i^\top\triangle\sigma_i)(f(y_i)) - g'(f(y_i))f(y_i)\triangle\sigma_i(f(y_i))^\top\sigma_{i-1} \\
&= \frac{1}{2}\frac{g''(f(y_i))}{(g'(f(y_i)))^2}(B_i^\top B_i) - \sigma_{i-1}^\top B_i \\
&= -\frac{1}{2}\frac{f''(y_i)}{f'(y_i)}(B_i^\top B_i) - \sigma_{i-1}^\top B_i.
\end{aligned}$$

Therefore,

$$\begin{aligned}
u_i &= \left(-\frac{1}{2}\frac{f''(y_i)}{f'(y_i)}(B_i^\top B_i) - \sigma_{i-1}^\top B_i\right)f'(y_i) + \frac{1}{2}f''(y_i)B_i^\top B_i \\
&= -f'(y_i)\sigma_{i-1}^\top B_i,
\end{aligned}$$

which concludes the proof. □

Proof of Theorem 6

Here we present the proof for the case of a one-dimensional driving Brownian motion, i.e. $d = 1$. The proof for $d > 1$ is essentially the same except for more burdensome notation. We shall prove that e_i is actually *bounded* for every $i = 1, \cdots, N$. The strong and weak convergence order then follow from Theorems 9.6.2 and 14.5.2 of Kloeden and Platen [12].

Before we prove the boundedness of e_i, let us briefly describe how e_i can be evaluated at each step of simulation. As before, Let $y_j \triangleq Y_j(0), A_j \triangleq a_j(\tilde{Y}(0))$ and $B_j \triangleq b_j(\tilde{Y}(0))$ for $j = 1, \cdots, N$. Explicit calculation yields that

$$f(\tilde{Y}_j(h)) = d_j^{(0)} + d_j^{(1)} Z\sqrt{h} + d_j^{(2)} h + d_j^{(3)} Z^2 h + d_j^{(4)} Zh\sqrt{h} + d_j^{(5)} Z^3 h\sqrt{h} + d_j^{(6)} h^2 + d_j^{(7)} Z^2 h^2 + d_j^{(8)} Z^4 h^2 + o(h^2),$$

with

$$\begin{aligned} d_j^{(0)} &= f(y_j), \quad d_j^{(1)} = f'(y_j)B_j, \quad d_j^{(2)} = f'(y_j)A_j, \\ d_j^{(3)} &= \frac{1}{2}f''(y_j)B_j, \quad d_j^{(4)} = f''(y_j)A_jB_j, \quad d_j^{(5)} = \frac{1}{6}f'''(y_j)B_j^3, \\ d_j^{(6)} &= f'(y_j)e_j + \frac{1}{2}f''(y_j)A_j^2, \quad d_j^{(7)} = \frac{1}{2}f'''(y_j)A_jB_j^2, \quad d_j^{(8)} = \frac{1}{24}f''''(y_j)B_j^4. \end{aligned}$$

Moreover, we may write

$$\begin{aligned} \prod_{i=1}^{j} f(\tilde{Y}_i(h)) &= w_j^{(0)} + w_j^{(1)} Z\sqrt{h} + w_j^{(2)} h + w_j^{(3)} Z^2 h + w_j^{(4)} Zh\sqrt{h} + w_j^{(5)} Z^3 h\sqrt{h} \\ &\quad + w_j^{(6)} h^2 + w_j^{(7)} Z^2 h^2 + w_j^{(8)} Z^4 h^2 + o(h^2), \end{aligned}$$

where $w_j^{(\cdot)}$'s can be expressed in terms of $w_{j-1}^{(\cdot)}$'s and $d_j^{(\cdot)}$'s. Actually, $w_1^{(\cdot)} = d_1^{(\cdot)}$ and for $i \geq 2$,

$$\begin{aligned} w_i^{(0)} &= w_{i-1}^{(0)} d_i^{(0)}; \quad w_i^{(1)} = w_{i-1}^{(1)} d_i^{(0)} + w_{i-1}^{(0)} d_i^{(1)}; \quad w_i^{(2)} = w_{i-1}^{(2)} d_i^{(0)} + w_{i-1}^{(0)} d_i^{(2)}; \\ w_i^{(3)} &= w_{i-1}^{(3)} d_i^{(0)} + w_{i-1}^{(1)} d_i^{(1)} + w_{i-1}^{(0)} d_i^{(3)}; \quad w_i^{(4)} = w_{i-1}^{(4)} d_i^{(0)} + w_{i-1}^{(2)} d_i^{(1)} + w_{i-1}^{(1)} d_i^{(2)} + w_{i-1}^{(0)} d_i^{(4)}; \\ w_i^{(5)} &= w_{i-1}^{(4)} d_i^{(0)} + w_{i-1}^{(2)} d_i^{(1)} + w_{i-1}^{(1)} d_i^{(2)} + w_{i-1}^{(0)} d_i^{(4)}; \quad w_i^{(6)} = w_{i-1}^{(6)} d_i^{(0)} + w_{i-1}^{(2)} d_i^{(2)} + w_{i-1}^{(0)} d_i^{(6)}; \\ w_i^{(7)} &= w_{i-1}^{(7)} d_i^{(0)} + w_{i-1}^{(4)} d_i^{(1)} + w_{i-1}^{(3)} d_i^{(2)} + w_{i-1}^{(2)} d_i^{(3)} + w_{i-1}^{(1)} d_i^{(4)} + w_{i-1}^{(0)} d_i^{(7)}; \\ w_i^{(8)} &= w_{i-1}^{(8)} d_i^{(0)} + w_{i-1}^{(5)} d_i^{(1)} + w_{i-1}^{(3)} d_i^{(3)} + w_{i-1}^{(1)} d_i^{(5)} + w_{i-1}^{(0)} d_i^{(8)}; \end{aligned}$$

The value of e_i can be updated at each step through the equation

$$w_i^{(6)} + w_i^{(7)} + 3w_i^{(8)} = 0,$$

which makes the term of second-order have expectation zero since $\mathsf{E}Z = \mathsf{E}Z^3 = 0, \mathsf{E}Z^2 = 1$ and $\mathsf{E}Z^4 = 3$. Solving this equation, we have

$$d_i^{(6)} = -\frac{w_i^{(7)} + 3w_i^{(8)} + w_{i-1}^{(6)} d_i^{(0)} + w_{i-1}^{(2)} d_i^{(2)}}{w_{i-1}^{(0)}} \quad \text{and} \quad e_i = \frac{d_i^{(6)} - \frac{1}{2} f''(y_i) A_i^2}{f'(y_i)}. \tag{15}$$

More specifically, when $i = 1$, the above equation gives $d_1^{(6)} + d_1^{(7)} + 3d_1^{(8)} = 0$, which in turn yields

$$e_1 = \frac{-1}{f'(y_1)} \left(\frac{1}{2} f''(y_1) A_1^2 + \frac{1}{2} f'''(y_1) A_1 B_1^2 + \frac{1}{8} f''''(y_1) B_1^4 \right).$$

With this e_1, we can determine all the $d_1^{(\cdot)}$'s. As $i = 2$, all the $w_2^{(\cdot)}$'s can updated immediately except $w_i^{(6)}$ since we do not know $d_2^{(6)}$. However, e_2 can be evaluated from (15), which in turn determine the value of $d_2^{(6)}$ and therefore $w_2^{(6)}$; and so on.

To prove the boundedness of e_j, we employ induction on j.

$j = 1$: It follows from $f \in \mathcal{D}'$ that

$$\sup \left| \frac{f'' A_1^2}{f'} \right| = \sup \left| \frac{f'' f^2 (\sigma_1 \circ f)^2}{(f')^3} \right| \le c \sup \left| \frac{f'' f^2 (1-f)^2}{(f')^3} \right| < \infty,$$

and similarly,

$$\sup \left| \frac{f''' A_1 B_1^2}{f'} \right| \le c \sup \left| \frac{f''' (f'')^2 f^5 (1-f)^5}{(f')^8} \right| < \infty,$$

$$\sup \left| \frac{f'''' B_1^4}{f'} \right| \le c \sup \left| \frac{f'''' (f'')^4 f^8 (1-f)^8}{(f')^{13}} \right| < \infty.$$

The boundedness of e_1 follows readily.

Suppose for $1 \le i \le j-1$, e_i is bounded. We need to prove that e_j is bounded too.

First we prove that all $d_i^{(p)}$s with $1 \le i \le j-1, 0 \le p \le 8$ are bounded. Actually, by assumption, e_i is bounded. Hence $d_i^{(6)}$ is bounded too. For $p \neq 6$, explicit calculation will do. For example, when $p = 5$,

$$\sup \left| d_i^{(5)} \right| = \sup \left| \frac{1}{6} f''' B_i^3 \right| \le c \sup \left| \frac{f''' (f'')^3 f^6 (1-f)^6}{(f')^9} \right| < \infty$$

since $f \in \mathcal{D}'$.

e_j is determined by $w_j^{(6)} + w_j^{(7)} + 3w_j^{(8)} = 0$. However,

$$\begin{aligned} w_j^{(6)} h^2 + w_j^{(7)} Z^2 h^2 + w_j^{(8)} Z^4 h^2 \quad = \quad & d_j^{(0)} \cdot \left(\text{terms of order } h^2 \text{ from } \prod_{i=1}^{j-1} f(\tilde{Y}_i(h)) \right) \\ & + d_j^{(1)} \cdot \left(\text{terms of order } h\sqrt{h} \text{ from } \prod_{i=1}^{j-1} f(\tilde{Y}_i(h)) \right) + \cdots \\ & + \cdots + d_j^{(8)} \cdot \left(\text{constant term from } \prod_{i=1}^{j-1} f(\tilde{Y}_i(h)) \right). \end{aligned}$$

Because $\left(\text{terms of order } h^2 \text{ from } \prod_{i=1}^{j-1} f(\tilde{Y}_i(h))\right)$ already has expectation 0, it suffices to prove that $\frac{d_j^{(p)}}{f'}$ *for all* $1 \leq p \leq 8$ *and* $\frac{f'' a_j^2}{f'}$ *are bounded* since all the $d_i^{(p)}$s with $1 \leq i \leq j-1, 0 \leq p \leq 8$ are bounded. Using $f \in \mathcal{D}'$, we can carry out the proof by explicit calculation. We omit the details. □

References

[1] Brace, A., Gatarek, D., and Musiela, M., The market model of interest rate dynamics, *Math. Finance*, 7, 127-155 (1997).

[2] Delbaen, F., and Schachermayer, W., A general version of the fundamental theorem of asset pricing, *Math. Ann.*, 300, 463–520, (1997).

[3] Duffie, D., *Dynamic Asset Pricing Theory*, second edition, Princeton University Press, Princeton, New Jersey (1995).

[4] Duffie, D., and Glynn, P.W. (1995), Efficient Monte Carlo simulation of security prices, *Annals Appl. Probab.*, 5, 897–905.

[5] El Karoui, N., Geman, H., and Rochet, J.C., Changes of numeraire, changes of probability measure, and option pricing, *J. Appl. Probab.* 32, 443-458 (1995).

[6] Glasserman, P., and Zhao, X., Arbitrage-free discretization of lognormal forward Libor and swap rate models, *Finance and Stochastics*, to appear.

[7] Harrison, J.M., and Kreps, D., Martingales and arbitrage in multiperiod securities markets, *J. Economic Theory*, 20, 381-408 (1979).

[8] Harrison, J.M., and Pliska, S., Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and Their Applications* 11, 215-260 (1981).

[9] Heath, D., Jarrow, R., and Morton, A., Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation, *Econometrica*, 60, 77-105 (1992).

[10] Hewitt, E., and Stromberg, K., *Real and Abstract Analysis,* Springer, New York (1975).

[11] Jamshidian, F., LIBOR and swap market models and measures, *Finance and Stochastics*, 1, 293–330 (1997).

[12] Kloeden, P., and Platen, E., *Numerical Solution of Stochastic Differential Equations*, Springer, New York (1992).

[13] Karatzas, I. and Shreve, S. *Brownian Motion and Stochastic Calculus.* Second Edition, Springer, New York (1992).

[14] Milstein, G.N., A method of second-order accuracy integration of stochastic differential equations, *Theor. Probab. Appl.*, 19, 557-562 (1978).

[15] Miltersen, K.R., Sandmann, K, and Sondermann, D., Closed-form solutions for term structure derivatives with lognormal interest rates, *J. Finance*, 409-430 (1997).

[16] Musiela, M., and Rutkowski, M., Continuous-time term structure models: Forward measure approach, *Finance and Stochastics*, 1, 261-292 (1997).

[17] Musiela, M., and Rutkowski, M., *Martingale methods in financial modeling*, Springer, New York, 1997.

[18] Talay, D., Efficient numerical schemes for the approximation of expectations of functionals of the solution of a S.D.E, and applications, in *Filtering and Control of Random Processes*, H. Korezlioglu, G. Mazziotto, and J. Szpirglas, eds., Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin (1984).