

Fast Greeks in Forward Libor Models

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

Xiaoliang Zhao
Department of Statistics
Columbia University
New York, NY 10027

January 1999

Abstract

This paper develops methods for fast estimation of option price sensitivities in Monte Carlo simulation of term structure models. The models considered are based on discretely compounded forward rates with proportional volatilities. We investigate the efficient estimation of option deltas, gammas, and vegas in this setting. Various general methods are available in the Monte Carlo literature for computing such estimates; we tailor these methods to the term structure models and develop approximations specific to this setting in order to either accelerate the methods or expand their applicability. We provide some theoretical support for the application of the basic methods and evaluate the approximations through numerical experiments. The results indicate that the proposed algorithms can substantially improve over standard finite difference estimates of sensitivities.

1 Introduction

This paper develops methods for fast estimation of option price sensitivities based on Monte Carlo simulation of forward Libor models of the type developed by Brace, Gatarek, and Musiela [3], Jamshidian [12], Miltersen, Sandmann, and Sondermann [14], and Musiela and Rutkowski [15]. These models are similar in spirit to the general framework of Heath, Jarrow, and Morton [10], but differ in that they model the dynamics of discretely compounded forward rates (directly observable in the market) rather than instantaneous continuously compounded forward rates. As in the HJM setting, arbitrage restrictions determine the dynamics of the forward curve (now represented by a vector of discrete rates) once the volatility structure and numeraire have been chosen. The resulting dynamics are typically complex enough to make Monte Carlo simulation the primary computational tool for use with these models.

Price sensitivities are, of course, of central importance in any model for pricing derivative securities because the sensitivities determine the trading strategy that hedges the derivative security. A common criticism of Monte Carlo simulation is that it produces poor estimates of Greeks. Indeed, using straightforward simulation, estimating deltas with respect to N underlying assets or rates requires simulating a minimum of $N + 1$ times as many paths as estimating a price alone, and in

spite of this the delta estimates obtained will often be much less accurate than the estimated price. There are, however, Monte Carlo methods specifically designed for the estimation of sensitivities. Some of these are treated in, e.g., Glasserman [7], Glynn and L’Ecuyer [9], Ho and Cao [11], Reiman and Weiss [17], Rubinstein and Shapiro [18], and the application of these and related methods to option pricing has been considered in Broadie and Glasserman [2], Fournié, Lasry, Lebuchoux, Lions, and Touzi [5], Fu and Hu [6] and Pikovsky [4]; see also the overview in Boyle et al. [1]. But the class of models for in which we implement the methods here is somewhat more complex than previous financial applications of these methods and it raises both practical and theoretical issues.

The problem of estimating sensitivities by simulation may be formulated quite generally as one of estimating the derivative of an expectation with respect to a parameter. In the case of estimating a delta, for example, the relevant parameter is the initial value of a price or rate. Methods for estimating sensitivities may be broadly classified by whether they put the dependence on the parameter in the underlying stochastic process or in the probability measure. Both perspectives are generally possible, and this flexibility is analogous to two ways of adding a drift μ to Brownian motion: we may add μt at time t to each Brownian path, or we can leave the paths unchanged and use Girsanov’s Theorem to add a drift through a change of probability measure. Putting the dependence on the parameter in the sample paths of the stochastic process leads to estimators that differentiate the paths of the process — we call these *pathwise* derivatives. Putting the dependence in the measure leads to estimators based on differentiating probability densities; this is often referred to as the *likelihood ratio method* (LRM).

We investigate the use of both pathwise derivatives and LRM in estimating deltas and gammas and the use of pathwise estimators for “vega” (sensitivity to changes in volatility). Our primary contribution to the literature on forward Libor models lies in deriving and comparing a variety of methods and identifying which are most practical and effective in this context. In this regard our conclusions are as follows:

- For estimating deltas when the option payoff is a (Lipschitz) continuous function of the forward rates, use the pathwise method with a *forward-drift approximation*.
- For estimating deltas when the payoff is discontinuous (e.g., a digital or knockout payoff), use LRM with the forward-drift approximation.
- No method is entirely satisfactory for estimating gammas. Conventional central difference approximations are very sensitive to the size of the perturbation introduced. A mixed pathwise-LRM method appears preferable.
- A pathwise estimator using a forward-drift approximation is fast and effective in estimating vega when the payoff is continuous.

Relative to the general literature on estimating sensitivities through simulation, this paper makes three principal contributions:

- (i) it proposes and evaluates fast approximations to an exact pathwise algorithm specific to the forward Libor setting;
- (ii) it analyzes the convergence to the continuous-time limit of pathwise estimators based on discrete-time simulation;
- (iii) it uses an approximate LRM estimator in a setting where the relevant probability density is unknown and develops a method for applying LRM in a singular setting where no density exists.

We comment briefly on each of these points. (i) In its exact version, the pathwise method entails simulating a stochastic process of derivatives of state variables in addition to the original state variables. In a model with the complexity of the forward Libor models, the effort involved in simulating the derivatives process can be comparable to that required to simulate a perturbed copy of the original process, so the pathwise method may not offer a large advantage over a standard finite difference approximation to a derivative based on resimulating the original process. The approximations we develop address this issue. (ii) The pathwise method can be formulated in continuous time (differentiating a diffusion process with respect to a parameter) or in discrete time (differentiating the discretized process in the simulation). We give conditions under which the discrete-time estimator gives unbiased derivative estimates for the simulated process and also under which it converges to the correct continuous-time limit. (iii) The application of LRM to estimating delta entails knowledge of the transition density of the underlying state variables. No such density is available in forward Libor models, so we use a Gaussian approximation. This does not entirely resolve the problem because in a model with fewer factors than state variables (i.e., a model in which the dimension of the driving Brownian motion is smaller than the dimension of the state vector) the distribution of the increments of the state variables over one simulated time step is singular and fails to have a density — even in the Gaussian case. The increment over multiple time steps may nevertheless have a density and we use this observation to apply LRM.

The rest of this paper is organized as follows. Section 2 reviews the dynamics of forward Libor models. Section 3 develops pathwise delta estimators, first deriving an exact method and then proposing and evaluating approximations. Section 4 develops LRM delta estimators, first reviewing the method in a purely Gaussian setting then tailoring its application to forward Libor models. Section 5 addresses the somewhat harder problem of estimating gamma and Section 6 deals with vega. Theoretical analysis of the pathwise estimators is given in Section 7.

2 Preliminaries on the Model

We begin with a brief review of Libor market models based on a finite set of maturities, as developed by Jamshidian [12]. The *tenor structure* is a finite set of dates

$$0 = T_0 < T_1 < \dots < T_N < T_{N+1}$$

representing maturities spaced, e.g., three months or six months apart. For simplicity, we assume that the day-count fractions $\delta_i \triangleq T_{i+1} - T_i$, $i = 0, \dots, N$, are all equal to a fixed δ (e.g., $\delta = 0.25$ years). In practice, day-count conventions would make the lengths of these intervals slightly different. The left-continuous function $\eta : (0, T_{N+1}] \rightarrow \{1, \dots, N+1\}$ defined by taking $\eta(t)$ to be the unique integer satisfying

$$T_{\eta(t)-1} < t \leq T_{\eta(t)}$$

gives the index of the next tenor date at time t . Associated with each tenor date T_i is a zero-coupon bond maturing at that date; $B_i(t)$ is the price of that bond at time $t \in [0, T_i]$ and $B_i(T_i) = 1$.

The *forward Libor rate* at time t for the accrual period $[T_i, T_{i+1}]$, $t \leq T_i$ is

$$L_i(t) = \frac{1}{\delta} \left(\frac{B_i(t)}{B_{i+1}(t)} - 1 \right), \quad i = 1, \dots, N. \quad (1)$$

It is at times notationally convenient to extend the definition of L_i beyond the i th tenor date; we do so by setting $L_i(t) = L_i(T_i)$ for $t > T_i$. At a tenor date T_i the price of any bond B_n , $n > i$, that has not yet matured is given by

$$B_n(T_i) = \prod_{j=i}^{n-1} \frac{1}{1 + \delta L_j(T_i)};$$

more generally, at an arbitrary time $t < T_n$ we have

$$B_n(t) = B_{\eta(t)}(t) \prod_{j=\eta(t)}^{n-1} \frac{1}{1 + \delta L_j(t)}. \quad (2)$$

The dynamics of the forward Libor rates depend on the form assumed for their volatilities and on the measure under which the model is specified. Throughout, we assume the Libor rates have deterministic volatilities (so that caplets are priced by Black's formula, as in [3]) and we work in the *spot Libor measure* introduced by Jamshidian [12]. This is the equivalent martingale measure associated with the numeraire

$$B^*(t) = \frac{B_{\eta(t)}(t)}{B_1(0)} \prod_{j=1}^{\eta(t)-1} \frac{B_j(T_j)}{B_{j+1}(T_j)},$$

which may be interpreted as the result of buying $1/B_1(0)$ bonds at time 0 maturing at T_1 , and then at each tenor date selling the bonds that matured and investing the proceeds in the bond

that matures next. This is thus a discretely compounded analog of the money market account that gives rise to the usual risk-neutral measure. A particular case of Jamshidian's construction is the specification

$$\frac{dL_n(t)}{L_n(t)} = \sum_{i=\eta(t)}^n \frac{\delta \lambda_n(t) \lambda_i(t)' L_i(t)}{1 + \delta L_i(t)} dt + \lambda_n(t) dW_t, \quad n = 1, \dots, N, \quad (3)$$

in which W_t is a standard d -dimensional Brownian motion under the spot Libor measure and each λ_n is deterministic, bounded, and possibly time-varying, with $\lambda_n(t)$ a d -dimensional row vector. (Take $\lambda_n(t) \equiv 0$ for $t \geq T_n$ to keep $L_n(t) = L_n(T_n)$ on $[T_n, T_{N+1}]$.) The form of the drift in (3) is necessitated by the absence of arbitrage once the volatilities (and the numeraire) are specified. In particular, with this choice of drift deflated asset prices (ratios of asset prices to $B^*(t)$) are martingales. It follows that the time- t value $C(t)$ and time- T value $C(T)$ of a derivative security (that can be replicated by trading in the basic bonds) are related by the pricing rule

$$C(t) = B^*(t) E \left[\frac{C(T)}{B^*(T)} \middle| \mathcal{F}_t \right], \quad (4)$$

where $\{\mathcal{F}, t \geq 0\}$ is the filtration generated by the Brownian motion.

In order to delta hedge a derivative with positions in the underlying bonds, we need to calculate, e.g.,

$$\frac{\partial C(0)}{\partial B_n(0)} = \frac{\partial}{\partial B_n(0)} \left(B^*(0) E \left[\frac{C(T)}{B^*(T)} \right] \right), \quad n = 1, \dots, N + 1.$$

In light of the deterministic relations (1) and (2), this is equivalent (through the chain rule of ordinary calculus) to computing sensitivities

$$\frac{\partial C(0)}{\partial L_k(0)} = \frac{\partial}{\partial L_k(0)} \left(B^*(0) E \left[\frac{C(T)}{B^*(T)} \right] \right), \quad k = 0, \dots, N,$$

setting $L_0 = ([1/B_1(0)] - 1)/\delta$. This may be viewed as a type of *bucket hedging* in which a separate delta is computed with respect to each component of the forward rate vector. Given either bucket deltas or deltas with respect to zero-coupon bonds one can in turn compute deltas with respect to the basic instruments used to build a forward curve again using just the ordinary chain rule, because the bond prices and forward rates are deterministically related to the basic instruments — the deterministic relation being embodied in the curve building algorithm.

3 Pathwise Deltas

3.1 Preliminaries

To motivate the first method we develop, consider a caplet with strike K paying $\delta(L_n(T_n) - K)^+$ at T_{n+1} . From (4) and the definition of B^* it follows that its price at time 0 is

$$C_n(0) = B_1(0) \delta E \left[(L_n(T_n) - K)^+ \prod_{i=1}^n \frac{1}{1 + \delta L_i(T_n)} \right]. \quad (5)$$

Much as in Brace, Gatarek, and Musiela [3], this expectation is evaluated by the Black formula

$$C_n(0) = C_{\text{Black}}(\bar{\lambda}_n, K, L_n(0), B_{n+1}(0), T_n)$$

where

$$C_{\text{Black}}(\sigma, K, r, b, T) = \delta b \left[r \Phi \left(\frac{\log(r/K) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \right) - K \Phi \left(\frac{\log(r/K) - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}} \right) \right], \quad (6)$$

where Φ is the cumulative normal distribution and

$$\bar{\lambda}_n = \sqrt{\int_0^{T_n} \|\lambda_n(t)\|^2 dt}.$$

Of course, since we want to develop a general method we will not use the fact that a caplet can be priced in closed form except to compare numerical results.

To compute, e.g., $\partial C_n(0)/\partial L_k(0)$ we need to compute the sensitivity of the expectation in (5) with respect to $L_k(0)$. Provided derivative and expectation can be interchanged, we have

$$\frac{\partial}{\partial L_k(0)} E \left[(L_n(T_n) - K)^+ \prod_{i=1}^n \frac{1}{1 + \delta L_i(T_n)} \right] = E \left[\frac{\partial}{\partial L_k(0)} \left\{ (L_n(T_n) - K)^+ \prod_{i=1}^n \frac{1}{1 + \delta L_i(T_n)} \right\} \right]. \quad (7)$$

If we can evaluate the derivative inside the expectation on the right for each simulated path, then by averaging over paths we obtain an estimate of the expectation and thus of the derivative on the left. The chain rule suggests

$$\frac{\partial}{\partial L_k(0)} \left\{ (L_n(T_n) - K)^+ \prod_{i=1}^n \frac{1}{1 + \delta L_i(T_n)} \right\} = \sum_{i=1}^n \frac{\partial}{\partial L_i(T_n)} \left\{ (L_n(T_n) - K)^+ \prod_{i=1}^n \frac{1}{1 + \delta L_i(T_n)} \right\} \frac{\partial L_i(T_n)}{\partial L_k(0)}.$$

One may question whether the expression in curly braces can be differentiated as indicated in light of the presence of the positive-part operator. However, the mapping $x \mapsto (x - K)^+$ is Lipschitz continuous and thus differentiable almost everywhere and equal to the indefinite integral of its a.e.-defined derivative. With probability one, we have

$$\frac{\partial}{\partial L_k(0)} (L_n(T_n) - K)^+ = \mathbf{1}\{L_n(T_n) > K\} \frac{\partial L_n(T_n)}{\partial L_k(0)}.$$

(The expression $\mathbf{1}\{\cdot\}$ takes the value 1 when the event in braces occurs and 0 otherwise.)

Generalizing the setting, to estimate

$$\frac{\partial}{\partial L_k(0)} E[g(L_1(t_1), \dots, L_N(t_N))]$$

for some Lipschitz continuous g and arbitrary dates t_i , we bring the derivative inside the expectation to arrive at the (continuous-time) pathwise delta estimator

$$\sum_{n=1}^N \left\{ \frac{\partial}{\partial L_n(t_n)} g(L_1(t_1), \dots, L_N(t_N)) \right\} \Delta_{nk}(t_n),$$

with

$$\Delta_{nk}(t) = \frac{\partial L_n(t)}{\partial L_k(0)}, \quad n, k = 1, \dots, N.$$

In practice, we can at best simulate discrete-time approximations \hat{L}_n and $\hat{\Delta}_{nk}$ to these continuous-time variables. We thus arrive at

Pathwise Delta Estimator

$$\sum_{n=1}^N \left\{ \frac{\partial}{\partial L_n(t_n)} g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_N)) \right\} \hat{\Delta}_{nk}(t_n). \quad (8)$$

It should now be clear that the key to this method is the evaluation of the Libor sensitivities Δ_{nk} and their discretized counterparts $\hat{\Delta}_{nk}$.

3.2 Exact Pathwise Method

Recall that the evolution of the forward Libor rates is determined by

$$dL_n(t) = \mu_n(t)L_n(t) dt + \lambda_n(t)L_n(t) dW_t, \quad n = 1, \dots, N, \quad (9)$$

where

$$\mu_n(t) = \sum_{i=\eta(t)}^n \frac{\delta \lambda_n(t) \lambda_i(t)' L_i(t)}{1 + \delta L_i(t)}.$$

Heuristically differentiating both sides with respect to $L_k(0)$ suggests

$$\begin{aligned} d\Delta_{nk}(t) &= \Delta_{nk}(t) [\mu_n(t)dt + \lambda_n(t)dW_t] + L_n(t) \sum_{j=1}^N \frac{\partial \mu_n(t)}{\partial L_j(t)} \Delta_{jk}(t) dt \\ &= \Delta_{nk}(t) \frac{dL_n(t)}{L_n(t)} + L_n(t) \sum_{j=1}^N \frac{\partial \mu_n(t)}{\partial L_j(t)} \Delta_{jk}(t) dt, \\ n &= 1, \dots, N, \quad k = 1, \dots, N. \end{aligned} \quad (10)$$

In Section 7, we justify this equation by showing that the system of SDEs (9)-(10) has a solution for which indeed $\Delta_{nk}(t) = \partial L_n(t) / \partial L_k(0)$.

From the perspective of simulation, the problem has now been reduced to one of simulating discrete-time approximations to the system of SDEs (9) and (10). The question of discretization of (9) is investigated in Glasserman and Zhao [8], where we show that there are advantages to discretizing SDEs for deflated bond prices (or their increments) rather than the forward Libor rates themselves. Differentiating these SDEs leads to a set of derivative SDEs analogous to (10) and it is possible to simulate a discrete-time approximation to those. Indeed, one could even choose

to simulate the deflated bond price SDEs together with the derivative SDEs (10). In continuous-time, all such variations are ultimately equivalent. The possible discrete-time approximations are limitless. To make the general method as transparent as possible, we restrict attention to (9)–(10).

Among the methods for simulating (9) considered in [8] is the recursion

$$\hat{L}_n((i+1)h) = \hat{L}_n(ih) \exp \left[(\hat{\mu}_n(ih) - \frac{1}{2} \lambda_n(ih) \lambda_n(ih)') h + \lambda_n(ih) \sqrt{h} Z_{i+1} \right] \quad (11)$$

in which h is the time increment, Z_1, Z_2, \dots are independent d -dimensional standard normal vectors,

$$\hat{\mu}_n(ih) = \sum_{j=\eta(ih)}^n \frac{\delta \lambda_n(ih) \lambda_j(ih)' \hat{L}_j(ih)}{1 + \delta \hat{L}_j(ih)}, \quad (12)$$

and $\hat{L}_n(0) = L_n(0)$. (A hat indicates a discrete-time approximation to a continuous-time variable.) Equation (11) may be interpreted as an Euler scheme for $\log L_n$. Among all ways of discretizing (10), the one that differentiates (11) seems most natural and yields

Exact Pathwise Algorithm

$$\hat{\Delta}_{nk}((i+1)h) = \hat{\Delta}_{nk}(ih) \frac{\hat{L}_n((i+1)h)}{\hat{L}_n(ih)} + \hat{L}_n((i+1)h) \sum_{j=1}^N \frac{\partial \hat{\mu}_n(ih)}{\partial \hat{L}_j(ih)} \hat{\Delta}_{jk}(ih) h, \quad (13)$$

with initial condition $\hat{\Delta}_{nk}(0) = \mathbf{1}\{n = k\}$.

It is easy to see that, indeed,

$$\hat{\Delta}_{nk}(ih) = \frac{\partial \hat{L}_n(ih)}{\partial L_k(0)}, \quad (14)$$

for every outcome of Z_1, \dots, Z_i , and this is the sense in which the algorithm is exact. Moreover, the same algorithm evaluates $\partial \hat{L}_n(ih) / \partial B_k(0)$ if we change the initial condition $\hat{\Delta}_{nk}(0)$ to $\partial L_n(0) / \partial B_k(0)$.

More significant than the sample path property (14) is the fact that for any Lipschitz continuous $g : \mathbf{R}^N \rightarrow \mathbf{R}$ and any fixed times $i_1 h, \dots, i_N h$,

$$\frac{\partial}{\partial L_k(0)} E[g(\hat{L}_1(i_1 h), \dots, \hat{L}_N(i_N h))] = E \left[\sum_{j=k}^N \frac{\partial g}{\partial \hat{L}_j(i_j h)} \hat{\Delta}_{jk}(i_j h) \right], \quad (15)$$

as shown in Section 7. (The range of summation starts at k because $\hat{\Delta}_{jk} \equiv 0$ if $j < k$.) For example, to use this to estimate the sensitivity of a caplet price to an initial forward rate, let g be the discounted caplet payoff (the right side of (5) but without the expectation) and evaluate

$$\sum_{j=k}^N \frac{\partial g}{\partial \hat{L}_j(T_n)} \hat{\Delta}_{jk}(T_n)$$

with

$$\frac{\partial g}{\partial \hat{L}_j(T_n)} = B_1(0)\delta \prod_{i=1}^n \frac{1}{1 + \delta \hat{L}_i(T_n)} \left(\mathbf{1}\{\hat{L}_n(T_n) > K\} - (\hat{L}_n(T_n) - K)^+ \frac{\delta}{1 + \delta \hat{L}_j(T_n)} \right).$$

Equation (15) indicates that this gives an unbiased estimate of the delta in the discrete-time model (assuming the time grid $\{h, 2h, \dots\}$ includes the tenor dates T_n).

3.3 Approximations

For options with Lipschitz continuous payoffs, the pathwise method makes it possible to estimate deltas from a single simulation path — i.e., without actually changing any initial values in the model. However, implementing the exact pathwise method simulating (13) together with (11). The computational effort required by (11) (for all $k = 1, \dots, N$ and all $n = k, \dots, N$) is comparable to the effort involved in resimulating all $(\hat{L}_1, \dots, \hat{L}_N)$ and additional N times, slightly changing the value $L_k(0)$ on the k th of these. Hence, the exact pathwise method may not offer an overwhelming advantage compared with a standard finite difference estimator. We propose approximations to the exact algorithm that are much faster to simulate and appear to give very good accuracy.

One of the most time-consuming steps in (13) is the recomputation of all $\partial \hat{\mu}_n / \partial \hat{L}_j$ at every time step. For typical parameter values, each μ_n will be quite small (they differ from 0 just enough to keep the forward rate dynamics arbitrage-free) so our first approximation simply sets $\partial \hat{\mu}_n / \partial \hat{L}_j \equiv 0$ in the derivative recursions. Clearly, (13) then collapses to

Zero Drift Pathwise Approximation

$$\hat{\Delta}_{nk}((i+1)h) = \frac{\hat{L}_n((i+1)h)}{L_k(0)} \mathbf{1}\{n = k\}. \quad (16)$$

This would give the exact pathwise derivative if the forward rates were driftless multivariate geometric Brownian motion — i.e., if all the μ_n were indeed 0. It must be emphasized that although we make this approximation in the algorithm for $\hat{\Delta}$, we continue to use the original $\hat{\mu}_n$ for the simulation of the \hat{L}_n , as in (11).

Evaluating the $\hat{\Delta}_{nk}$ under the zero-drift approximation requires virtually no effort beyond that involved in simulating the forward Libor rates themselves. However, the approximation seems rather crude. Our next approximation lies between the exact and zero-drift methods both in terms of computing time and the accuracy with which it estimates $\partial \hat{L}_n / \partial L_k$. In this approximation, we differentiate \hat{L}_n as through the ordinary discretized drift $\hat{\mu}_n(ih)$ in (12) were instead

$$\hat{\mu}_n^o(ih) = \sum_{j=\eta(ih)}^n \frac{\delta \lambda_n(ih) \lambda_j(ih)' L_j(0)}{1 + \delta L_j(0)}.$$

In other words, we replace the $\hat{L}_j(ih)$ with their time-0 forward values $L_j(0)$, in the spirit of the approximations introduced by Brace et al. [3] to derive pricing formulas. The sensitivity of the approximate drift to $L_k(0)$ simplifies to

$$\frac{\partial \hat{\mu}_n^o(ih)}{\partial L_k(0)} = \frac{\delta \lambda_n(ih) \lambda'_k(ih)}{(1 + \delta L_k(0))^2} \mathbf{1}\{\eta(ih) \leq k \leq n\}.$$

Observe that these values are time-varying but deterministic. If $\hat{\mu}^o$ were the true drift, we would be able to solve the SDE for the forward Libor rates and differentiate this solution with respect its initial condition. Doing so yields

Forward Drift Approximation

$$\hat{\Delta}_{nk}(ih) = \frac{\hat{L}_n(ih)}{L_k(0)} \mathbf{1}\{n = k\} + \hat{L}_n(ih) \sum_{r=0}^{i-1} \frac{\partial \hat{\mu}_n^o(rh)}{\partial L_k(0)}. \quad (17)$$

The derivatives of $\hat{\mu}^o$ used in this expression can be precomputed so this approximation is only slightly more effort to implement than the zero-drift approximation. In particular, unlike the exact algorithm, it does not entail simulation of an additional recursion.

3.4 Numerical Comparisons

We compare the speed and accuracy of the exact and approximate pathwise algorithms through numerical results. All our results are based on $\delta = 0.25$ (quarterly rates), $h = \delta$ (simulation time step equal to length of accrual intervals) and $N + 1 = 20$ (a five-year horizon). The initial term structure takes the form $L_n(0) = \log(a + bn)$, with a and b chosen so that $L_0(0) = .05$ and $L_{19}(0) = .07$. The volatilities are constant over the intervals $[T_i, T_i + 1)$, with

$$\lambda_n(T_i) = \lambda(n - i), \quad i = 0, \dots, n - 1, \quad n = 1, \dots, 19,$$

and each $\lambda(j)$ drawn randomly from the uniform distribution on $[0.15, 0.25]$. The specific values of the $\lambda(j)$ used are

$$0.2216, 0.1919, 0.1631, 0.1751, 0.1993, 0.2444, 0.1894, 0.2286, 0.1539, 0.2147 \\ 0.1741, 0.2441, 0.2414, 0.1820, 0.1866, 0.2423, 0.2169, 0.1917, 0.1520, 0.2128 \quad .$$

We compare the performance of the exact pathwise algorithm, the zero drift approximation and the forward drift approximation in estimating $\partial C_n(0)/\partial L_k(0)$ with C_n a caplet price as in (5). Although in practice one would be interested in deltas for more complicated instruments or portfolios of instruments, using caplets allows us to compare with exact (continuous-time) values from Black's formula.

In principle, there are N^2 values of $\partial C_n(0)/\partial L_k(0)$ to be estimated corresponding to the possible combinations of n and k , though the delta is clearly 0 for $k > n$ and the most interesting case is $n = k$. Estimating all these deltas using finite differences (i.e., changing each $L_k(0)$ and resimulating) requires $N + 1$ simulated paths per observation — one for the original scenario and additional path for each perturbed $L_k(0)$. Using central differences the number increases to $2N + 1$. At the expense of some overhead per path, all the deltas can be estimated from the same simulated paths using any of the pathwise algorithms. In our experiments, estimating all deltas using the exact pathwise algorithm is about four times as fast as estimating all deltas using finite differences, the forward drift approximation is about three times as fast as the exact pathwise algorithm, and the zero drift approximation is faster by another factor of two.

Rather than attempt to report numerical results for all N^2 deltas, we focus on the most interesting and most difficult cases. The most interesting deltas are the diagonal cases, $n = k$. The most computationally demanding cases fix $n = N$ and let k range from 1 to N . In comparing methods, there are two standards one might reasonably apply in gauging accuracy: proximity to the discrete-time delta obtained by differentiating with respect to $L_k(0)$ while keeping the time step h fixed, or proximity to the continuous-time delta. The exact pathwise algorithm is unbiased for the former but — like any simulation method — is subject to discretization error in estimating the latter. In order to give as complete a picture as possible, we include information on both types of error.

Figure 1 shows estimated biases (in percent) for the diagonal deltas $\partial C_k/\partial L_k$ compared with the deltas obtained from Black’s formula. The exact values range from 0.10 to 0.13. We can see that all three methods are close to each other and perform well. The standard errors of these estimates are about 0.1%, so most of the estimated biases for the exact and forward drift methods fail to be statistically significant. It should be stressed that the results for the exact pathwise estimate represent the best one could hope to achieve using ordinary finite difference estimates. If we wanted to compare with the discrete-time delta rather than the continuous-time limit, we could use the exact pathwise estimate as the standard, since it is unbiased for the discrete-time delta. The forward drift approximation does a particularly effective job of approximating the exact method, with some gradual degradation at longer maturities.

Figure 2 shows the relative bias in deltas of the last caplet, $\partial C_{19}/\partial L_k(0)$, $k = 1, \dots, 19$. The absolute values of these deltas for $k = 1, \dots, 18$ are around 0.0005, which is about 200 times smaller than $\partial C_{19}/\partial L_{19}(0)$. In relative terms, the zero drift approximation method produces a large bias (up to 60%) for the off-diagonal deltas, the forward drift approximation produces a substantially smaller bias, and the exact method produces no discernible bias at all. It should be emphasized that in all cases in Figure 2 the absolute errors are very small and the large relative errors are due

to the fact that we estimating values so close to 0. Based on these and other consistent numerical results, taking into account both accuracy and computing time, the forward drift approximation appears to be the most effective method.

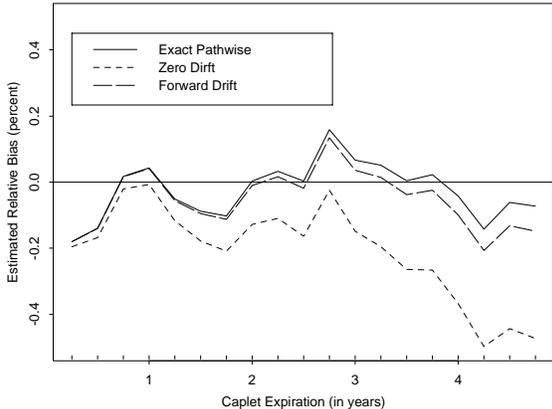


Figure 1: Estimated bias of $\partial C_k / \partial L_k(0)$

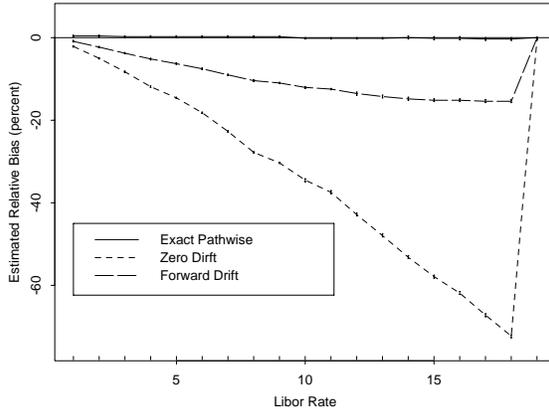


Figure 2: Estimated bias of $\partial C_{19} / \partial L_k(0)$

4 Likelihood Ratio Deltas

The only significant limitation of the method developed in the previous section is that it is restricted to payoffs that are at least continuous. This precludes application of the method to, e.g., a caplet with a digital payoff

$$\mathbf{1}\{L_N(T_N) > K\}$$

or a knockout caplet with payoff

$$(L_N(T_N) - K)^+ \mathbf{1}\left\{\min_{i=1, \dots, N} L_i(T_i) > b\right\}. \quad (18)$$

In both cases, the pathwise derivative with respect to some $L_k(0)$ actually exists with probability 1, but fails to reflect the discontinuity in the indicator function and thus provides an uninformative estimate. (For the digital caplet, the pathwise derivative is identically zero wherever it exists.) Put more precisely, these are examples in which the interchange of derivative and expectation required in (7) does not hold. We now present an alternative method for estimating deltas based on moving the dependence on $L_k(0)$ from the sample paths to the measure, thereby eliminating the need for smoothness in the option payoff. As noted in Section 1, the distinction is analogous to two ways of adding a drift to Brownian motion: we can add μt at time t to each Brownian path, or we can leave the paths unchanged and use Girsanov's Theorem to add a drift through a change of probability measure.

4.1 LRM in the Gaussian Setting

We begin our discussion of the *likelihood ratio method* (LRM) by considering the somewhat simpler setting of estimating sensitivities with respect to a parameter of the mean of a Gaussian vector. We then extend this to assets described by geometric Brownian motion and ultimately show how the method can be applied (with some necessary modifications) to Libor models.

Suppose, then, that the random n -vector X is multivariate normal with mean vector $\mathbf{m}(\theta)$ and covariance matrix Σ . Here, θ is a scalar parameter and we are interested in sensitivities with respect to θ . We suppose Σ has full rank and denote by

$$\phi(x; \mathbf{m}(\theta), \Sigma) = \frac{\exp(-\frac{1}{2}(x - \mathbf{m}(\theta))' \Sigma^{-1} (x - \mathbf{m}(\theta)))}{(2\pi)^{n/2} |\Sigma|^{1/2}}$$

the density of X . For any $g : \mathbf{R}^n \rightarrow \mathbf{R}$,

$$E_\theta[g(X)] = \int_{\mathbf{R}^n} g(x) \phi(x; \mathbf{m}(\theta), \Sigma) dx, \quad (19)$$

where we have subscripted the expectation to emphasize the dependence of the measure on θ . Differentiating and then interchanging derivative and integral yields

$$\begin{aligned} \frac{d}{d\theta} E_\theta[g(X)] &= \int g(x) \dot{\phi}(x; \mathbf{m}(\theta), \Sigma) dx \\ &= \int g(x) \frac{\dot{\phi}(x; \mathbf{m}(\theta), \Sigma)}{\phi(x; \mathbf{m}(\theta), \Sigma)} \phi(x; \mathbf{m}(\theta), \Sigma) dx, \end{aligned} \quad (20)$$

the dot on ϕ indicating differentiation with respect to θ . Some algebra shows that

$$\frac{\dot{\phi}(x; \mathbf{m}(\theta), \Sigma)}{\phi(x; \mathbf{m}(\theta), \Sigma)} = (x - \mathbf{m}(\theta))' \Sigma^{-1} \dot{\mathbf{m}}(\theta).$$

Making this substitution in (20) and interpreting the integral there as an expectation, we arrive at

$$\frac{d}{d\theta} E_\theta[g(X)] = E_\theta[g(X) (X - \mathbf{m}(\theta))' \Sigma^{-1} \dot{\mathbf{m}}(\theta)]. \quad (21)$$

Hence, the expression inside the expectation on the right provides an unbiased estimator of the derivative on the left. Moreover, this derivation requires smoothness in the dependence of ϕ on θ , but no smoothness at all in g . The key quantity $\dot{\phi}/\phi$ is the derivative with respect to ϵ of the likelihood ratio $\phi(x; \mathbf{m}(\theta + \epsilon), \Sigma)/\phi(x; \mathbf{m}(\theta), \Sigma)$, hence the name *likelihood ratio method*.

In a simulation, we would typically sample X by setting $X = \mathbf{m}(\theta) + AZ$ where A is an $n \times n$ matrix satisfying $AA' = \Sigma$ and Z is a vector of independent standard normal random variables. Making this substitution, we get

$$\frac{d}{d\theta} E_\theta[g(X)] = E[g(\mathbf{m}(\theta) + AZ) Z' A^{-1} \dot{\mathbf{m}}(\theta)]. \quad (22)$$

The expectation on the right is with respect to the n -dimensional standard normal distribution, hence not subscripted by θ .

This derivation applies directly to the estimation of delta for path-dependent options on geometric Brownian motion. Let

$$S_t = S_0 \exp(\mu t + \sigma W_t), \quad t \geq 0,$$

with W_t a one-dimensional Brownian motion and μ and σ constants. Suppose we want to estimate

$$\frac{d}{dS_0} E[f(S_{t_1}, \dots, S_{t_n})]$$

for some dates $0 < t_1 < \dots < t_n$ and some f to be interpreted as the discounted payoff of a path-dependent option. With g chosen appropriately, we can re-express the payoff in the following form:

$$f(S_{t_1}, \dots, S_{t_n}) = g(\log S_{t_1}, \log S_{t_2} - \log S_{t_1}, \dots, \log S_{t_n} - \log S_{t_{n-1}}).$$

Now make the correspondences $\theta \leftarrow S_0$,

$$X \leftarrow \begin{pmatrix} \log S_{t_1} \\ \log S_{t_2} - \log S_{t_1} \\ \vdots \\ \log S_{t_n} - \log S_{t_{n-1}} \end{pmatrix}, \quad \mathbf{m} \leftarrow \begin{pmatrix} \log S_0 + \mu t_1 \\ \mu(t_2 - t_1) \\ \vdots \\ \mu(t_n - t_{n-1}) \end{pmatrix},$$

and

$$\Sigma \leftarrow \begin{pmatrix} \sigma^2 t_1 & 0 & \dots & 0 \\ 0 & \sigma^2(t_2 - t_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2(t_n - t_{n-1}) \end{pmatrix}.$$

We may clearly take A to be diagonal in solving $AA' = \Sigma$. Let $Z' = (Z_1, \dots, Z_n)$ be independent standard normals used to simulate the process, in the sense that

$$S_{t_{i+1}} = S_{t_i} \exp(\mu(t_{i+1} - t_i) + \sigma \sqrt{t_{i+1} - t_i} Z_{i+1}), \quad i = 0, 1, \dots, n-1,$$

(so that $X = \mathbf{m} + AZ$). We now find that

$$Z' A^{-1} \mathbf{m} = \frac{Z_1}{S_0 \sigma \sqrt{t_1}}$$

and (22) becomes

$$\frac{d}{dS_0} E[f(S_{t_1}, \dots, S_{t_n})] = E \left[f(S_{t_1}, \dots, S_{t_n}) \frac{Z_1}{S_0 \sigma \sqrt{t_1}} \right].$$

We may therefore use

$$f(S_{t_1}, \dots, S_{t_n}) \frac{Z_1}{S_0 \sigma \sqrt{t_1}}$$

to estimate delta. A similar expression was derived in Broadie and Glasserman [2]. Notice that we used the function g to make a direct correspondence with the previous example but g plays no role in the final estimator. Moreover, f could be generalized to any function of the path of the underlying asset that depends only on values of the underlying after some time $t_1 > 0$.

The case of multidimensional geometric Brownian motion works similarly and will bring us one step closer to the Libor model. Suppose we have d assets $S_t^{(i)}$, $i = 1, \dots, d$, satisfying

$$S_t^{(i)} = S_0^{(i)} \exp(\mu_i t + \sigma_i W_t^{(i)}),$$

with $E[W_t^{(i)} W_t^{(j)}] = \rho_{ij} t$. Suppose the $d \times d$ matrix Σ with entries $\Sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$ has full rank and let A satisfy $AA' = \Sigma$. Write $f(S)$ for the value of some function of the d assets that depends on their values only after some time $t_1 > 0$. Suppose we simulate the d assets by setting

$$S_{t_1}^{(i)} = S_0^{(i)} \exp(\mu_i t_1 + \sqrt{t_1} (AZ)_i),$$

with Z a vector of d independent standard normals. Proceeding as before, we arrive at

$$\frac{d}{dS_0^{(k)}} E[f(S)] = E \left[f(S) \frac{(Z' A^{-1})_k}{S_0^{(k)} \sqrt{t_1}} \right].$$

The expression inside the expectation on the right thus provides an unbiased estimator of the delta with respect to the k th asset.

4.2 LRM in Libor Models

In order to see both the possibilities and difficulties in applying LRM in the Libor model, it is convenient to take logarithms in (11) to get

$$\log \hat{L}_n((i+1)h) = \log \hat{L}_n(ih) + [\hat{\mu}_n(ih) - \frac{1}{2} \|\lambda_n(ih)\|^2]h + \sqrt{h} \lambda_n(ih) Z_{i+1}, \quad n = 1, \dots, N. \quad (23)$$

Two issues now arise. The first is that $\hat{\mu}_n$ is a function of the forward Libor rates themselves and hence implicitly of the $L_k(0)$ s. This makes it difficult to move all the dependence on the $L_k(0)$ s out of the sample paths and into the probability measure. We address this issue as we did in Section 3.3 by differentiating *as though* the drift were deterministic (while simulating the forward Libor rates with the original drift). If we use the zero-drift approximation, the problem reduces to applying LRM to constant-drift, multidimensional geometric Brownian motion, just as in Section 4.1. But we work primarily with the forward-drift approximation, which is only slightly more complicated.

Under the forward-drift approximation, (23) describes the evolution of a Gaussian process so the development of the previous section potentially applies — if only as an approximation. But we still face a second issue not dealt with previously: equation (23) describes the evolution of a vector

of N rates driven by (say) d -dimensional vectors of normal random variables, where d is simply the number of factors in the original formulation of the model. Over a single time step, the covariance matrix of the increments in (23) has rank d . If $d < N$ (and we usually have $d \ll N$), the matrix is singular so the development in Section 4.1 — which includes inverting the covariance matrix — is not applicable. Indeed, even the starting point of the derivation (19) is problematic because the N -vector of increments fails to have a density in \mathbf{R}^N . This issue is not specific to the Libor setting. Had we not assumed that the covariance matrix Σ of the multidimensional Brownian motion in Section 4.1 is nonsingular, precisely the same issue would have arisen there.

To address this issue, we consider the distribution of the increments over multiple time steps rather than just one. Unraveling (23) yields

$$\log \hat{L}_n(ih) = \log L_n(0) + h \sum_{j=0}^{i-1} [\hat{\mu}_n(jh) - \frac{1}{2} \|\lambda_n(jh)\|^2] + \sqrt{h} [\lambda_n(0) | \lambda_n(h) | \cdots | \lambda_n((i-1)h)] \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_i \end{bmatrix},$$

where the row vectors $\lambda_n(jh)$ have been concatenated into a single vector of length $i \cdot d$ and the column vectors Z_j have been stacked into a column vector of the same length. For sufficiently large i^* , the $N \times i^*d$ matrix

$$\mathbf{\Lambda}_h(i^*) = \begin{pmatrix} \lambda_1(0) & | & \lambda_1(h) & | & \cdots & | & \lambda_1((i^*-1)h) \\ \lambda_2(0) & | & \lambda_2(h) & | & \cdots & | & \lambda_2((i^*-1)h) \\ \vdots & & \vdots & & \vdots & & \vdots \\ \lambda_N(0) & | & \lambda_N(h) & | & \cdots & | & \lambda_N((i^*-1)h) \end{pmatrix}$$

may have rank N , even if $d < N$. This means that the covariance matrix $\mathbf{\Lambda}_h(i^*)\mathbf{\Lambda}_h(i^*)'$ of the $\log L_n(i^*h)$, $n = 1, \dots, N$, is invertible and — using a deterministic approximation to the drift — the derivation of the previous section applies.

Suppose, then, that $\mathbf{\Lambda}_h(i^*)$ has full rank. To apply the method of Section 4.1 in the form given in (21), make the following correspondences: $\theta \leftarrow \hat{L}_k(0)$,

$$X \leftarrow (\log \hat{L}_1(i^*h), \dots, \log \hat{L}_N(i^*h))', \quad (24)$$

$$\mathbf{m}_n(\theta) \leftarrow \log L_n(0) + h \sum_{r=0}^{i^*-1} [\hat{\mu}_n^o(rh) - \frac{1}{2} \|\lambda_n(rh)\|^2], \quad n = 1, \dots, N, \quad (25)$$

($\hat{\mu}^o$ is the forward-drift approximation of Section 3.3), $\Sigma \leftarrow \sqrt{h}\Sigma_h(i^*) \equiv \sqrt{h}\mathbf{\Lambda}_h(i^*)\mathbf{\Lambda}_h(i^*)'$, and $A \leftarrow \sqrt{h}A_h(i^*)$ for any $i^*d \times i^*d$ matrix $A_h(i^*)$ satisfying $A_h(i^*)A_h(i^*)' = \mathbf{\Lambda}_h(i^*)\mathbf{\Lambda}_h(i^*)'$, and

$$\mathbf{m}_n \leftarrow \frac{\mathbf{1}\{n = k\}}{L_k(0)} + h \sum_{r=0}^{i^*-1} \frac{\partial \hat{\mu}_n^o(rh)}{\partial L_k(0)}, \quad n = 1, \dots, N. \quad (26)$$

With these substitutions we arrive at (see (21)) the following delta estimator for an arbitrary discounted payoff $g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_N))$:

LRM Delta Estimator

$$g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_N))(X - \mathbf{m})' \boldsymbol{\Sigma}^{-1} \dot{\mathbf{m}}, \quad (27)$$

with X , \mathbf{m} , $\boldsymbol{\Sigma}$, and $\dot{\mathbf{m}}$ as in (24)-(26).

Precomputing the vector $\boldsymbol{\Sigma}^{-1} \dot{\mathbf{m}}$ reduces the computational effort per simulated path to evaluate the quadratic form in (27) from $O(N^2)$ to $O(N)$. If we can take $i^* = N/d$ so that $\boldsymbol{\Lambda}_h(i^*)$ is square, then we can rewrite the quadratic form to get the estimator

$$g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_N)) h^{-1/2} Z' \boldsymbol{\Lambda}_h(i^*)^{-1} \dot{\mathbf{m}},$$

where Z is the column vector obtained by stacking the i^* d -vectors of independent normals used to simulate the d -factor model for i^* steps. This puts the estimator in the form of (22). Much as in (27), we can precompute $\boldsymbol{\Lambda}_h(i^*)^{-1} \dot{\mathbf{m}}$.

To illustrate the use of this method, we return to the examples with which we began this section. Consider the estimation of

$$B_1(0) \cdot \frac{d}{dL_k(0)} E \left[\mathbf{1}\{L_N(T_N) > K\} \prod_{i=1}^N \frac{1}{1 + \delta L_i(T_i)} \right],$$

the delta of the digital caplet with respect to the k th forward rate. For the LRM method to be applicable, we need the quantity inside the expectation to be a function of the $\hat{L}_i(t)$ for $t \geq i^*h$ but not $t < i^*h$. One way to achieve this chooses the time step h small enough that $T_1 \geq i^*h$. But the method can actually be applied with any $h \leq \delta$ if we recall that $L_i(t) \equiv L_i(T_i)$ for all $t \geq T_i$. The discounted payoff on the digital caplet can thus be re-expressed as

$$B_1(0) \mathbf{1}\{L_N(T_N) > K\} \prod_{i=1}^N \frac{1}{1 + \delta L_i(T_N)}$$

and the delta estimated using (with the notation in (24)-(26))

$$\left(B_1(0) \mathbf{1}\{L_N(T_N) > K\} \prod_{i=1}^N \frac{1}{1 + \delta L_i(T_N)} \right) h^{-1/2} (X - \mathbf{m})' \boldsymbol{\Sigma}_h(i^*)^{-1} \dot{\mathbf{m}}.$$

A similar rewriting of the discounted payoff on the knockout caplet leads to the estimator

$$\left(B_1(0) \delta(L_N(T_N) - K)^+ \mathbf{1}\left\{ \min_{i=1, \dots, N} L_i(T_N) > b \right\} \prod_{i=1}^N \frac{1}{1 + \delta L_i(T_N)} \right) h^{-1/2} (X - \mathbf{m})' \boldsymbol{\Sigma}_h(i^*)^{-1} \dot{\mathbf{m}},$$

for its delta with respect to $L_k(0)$.

The derivation above simplifies somewhat in the important special case that $h = \delta$ — i.e., when the simulation time step coincides with the spacing between tenor dates. If, in particular, we have $d = 1$ (a single-factor model), then the key matrix to check for nonsingularity is

$$\mathbf{\Lambda}_\delta(N) = \begin{pmatrix} \lambda_1(0) & | & \lambda_1(T_1) & | & \cdots & | & \lambda_1(T_{N-1}) \\ \lambda_2(0) & | & \lambda_2(T_1) & | & \cdots & | & \lambda_2(T_{N-1}) \\ \vdots & & \vdots & & \vdots & & \vdots \\ \lambda_N(0) & | & \lambda_N(T_1) & | & \cdots & | & \lambda_N(T_{N-1}) \end{pmatrix}.$$

Under our convention that $\lambda_n(t) = 0$ for $t \geq T_n$ (so that $L_n(t) \equiv L_n(T_n)$ for $t \geq T_n$), this matrix is block lower triangular. Suppose the forward rate volatilities depend solely on time-to-maturity in the sense that

$$\lambda_n(t) = \lambda(T_n - t)$$

for some function $\lambda(\cdot)$ and all n and t . (We require that λ assign a value of 0 to negative arguments.)

In this case, $\mathbf{\Lambda}_\delta(N)$ and its inverse have the general (Toeplitz) form

$$\mathbf{\Lambda}_\delta(N) = \begin{pmatrix} a_1 & & & & \\ a_2 & a_1 & & & \\ a_3 & a_2 & a_1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ a_N & a_{N-1} & a_{N-2} & \cdots & a_1 \end{pmatrix}, \quad \mathbf{\Lambda}_\delta(N)^{-1} = \begin{pmatrix} b_1 & & & & \\ b_2 & b_1 & & & \\ b_3 & b_2 & b_1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ b_N & b_{N-1} & b_{N-2} & \cdots & b_1 \end{pmatrix}.$$

The inverse is particularly easy to compute because $b_1 = 1/a_1$,

$$b_2 = -(a_2 b_1) b_1, \quad b_3 = -(a_3 b_1 + a_2 b_2) b_1, \quad \dots, \quad b_N = -(a_N b_1 + \cdots + a_2 b_{N-1}) b_1.$$

Fast computation of $\mathbf{\Lambda}_\delta^{-1}(N)$ may be especially important when the simulation is embedded in an iterative procedure to calibrate a model through choice of $\lambda(\cdot)$.

4.3 Numerical Results

Consistent with observations in [2] and the broader literature on sensitivity estimation, we find that when the pathwise method is applicable — in the present context meaning that the option payoff is Lipschitz continuous — it provides more precise estimates than LRM. We therefore evaluate the LRM estimator in estimating deltas for two discontinuous payoffs: a caplet with a digital payoff $\mathbf{1}\{L_n(T_n) > K\}$, and a knockout caplet with payoff

$$(L_n(T_n) - K)^+ \prod_{i=1}^n \mathbf{1}\{L_i(T_i) < b_i\}; \quad b_i = 1.2L_i(0), \quad i = 1, \dots, N.$$

The model parameters are as in Section 3.4. Because the pathwise method is inapplicable, the alternative against which we compare is a finite difference estimator. For the digital caplets we can

Method		Delta	Estimator	Std. Err.	RMSE
Exact Value		$\partial C_9/\partial B_{10}(0)$	-22.665	—	—
		$\partial C_{10}/\partial B_{10}(0)$	20.731	—	—
Likelihood Ratio Method		$\partial C_9/\partial B_{10}(0)$	-22.731	0.073	0.098
		$\partial C_{10}/\partial B_{10}(0)$	20.777	0.072	0.085
Finite Difference Method	$\Delta b =$	$\partial C_9/\partial B_{10}(0)$	-22.868	0.294	0.357
	0.0005	$\partial C_{10}/\partial B_{10}(0)$	20.798	0.278	0.286
	$\Delta b =$	$\partial C_9/\partial B_{10}(0)$	-22.793	0.201	0.238
	0.001	$\partial C_{10}/\partial B_{10}(0)$	20.233	0.190	0.533
	$\Delta b =$	$\partial C_9/\partial B_{10}(0)$	-22.432	0.132	0.268
	0.002	$\partial C_{10}/\partial B_{10}(0)$	19.316	0.125	1.421

Table 1: Deltas for digital caplets using LRM and finite differences. To balance the computing time required to estimate all deltas, we use 1,000,000 replications for LRM and 110,000 for the finite difference estimators. Δb is the increment in $B_{10}(0)$ used in the finite difference estimation.

find the exact (continuous-time) delta from a straightforward variant of the Black formula; for the knock-out caplet our “accurate” value is obtained from a large number of simulations of a finite difference estimator using a small increment.

As in Section 3.4 the large number of deltas one could consider makes it necessary to focus the numerical comparison on informative cases. Each caplet C_n (whether digital or standard) can be perfectly hedged using the bonds B_n and B_{n+1} , so hedge ratios with respect to these underlying assets are particularly interesting. Our numerical results focus on these deltas.

Applying the likelihood ratio method, we can compute the estimators of all deltas $\partial C_n/\partial B_i(0)$ for all $n = 1, \dots, 19$, $i = 1, \dots, 20$, from each simulation path. However, using a finite difference method, each pair of paths yields an estimate of $\partial C_n/\partial B_i(0)$ for all $n = 1, \dots, 19$ but with i fixed. It follows that the computing effort required to estimate all deltas using finite difference estimation is approximately 20 times greater than using LRM. In our numerical experiments, we balance the number of paths simulated using each method so that the computing time used to estimate *all* $\partial C_n/\partial B_i(0)$ is the same across methods.

Table 1 shows the numerical comparison of selected deltas for the digital option. (We choose $i = 10$ as a typical case.) Because $E[Z'\Lambda_h(i^*)^{-1}\mathbf{m}] = 0$, we use $Z'\Lambda_h(i^*)^{-1}\mathbf{m}$ as control variate in the implementation of LRM, as is often done. This reduces the standard error by about 25%. The finite difference methods are based on simulating from the initial values $B_{10}(0)$ and $B_{10}(0) + \Delta b$ for three values of Δb . Smaller values tend to reduce bias but increase variance; the two effects are captured by the root mean square error (RMSE). The results indicate that the LRM estimator outperforms the finite difference estimators.

Table 2 summarizes a similar numerical comparison for knockout caplets. The “accurate” values are estimates calculated using the finite difference method with $\Delta b = 0.00001$ and 100 million

Method		Delta	Estimator	Std. Err.	RMSE
Accurate Value		$\partial C_9 / \partial B_{10}(0)$	-0.05934	0.00041	–
		$\partial C_{10} / \partial B_{10}(0)$	0.06965	0.00042	–
Likelihood Ratio Method		$\partial C_9 / \partial B_{10}(0)$	-0.05889	0.00032	0.00047
		$\partial C_{10} / \partial B_{10}(0)$	0.06959	0.00058	0.00059
Finite Difference Method	$\Delta b = 0.0002$	$\partial C_9 / \partial B_{10}(0)$	-0.06218	0.00283	0.00401
		$\partial C_{10} / \partial B_{10}(0)$	0.06704	0.00307	0.00403
	$\Delta b = 0.0005$	$\partial C_9 / \partial B_{10}(0)$	-0.05579	0.00185	0.00400
		$\partial C_{10} / \partial B_{10}(0)$	0.06285	0.00205	0.00701
	$\Delta b = 0.001$	$\partial C_9 / \partial B_{10}(0)$	-0.05464	0.00124	0.00486
		$\partial C_{10} / \partial B_{10}(0)$	0.05793	0.00151	0.01182

Table 2: Deltas for knockout caplets using LRM and finite differences. To balance the computing time required to estimate all deltas, we use 1,000,000 replications for LRM and 100,000 for the finite difference estimators. Δb is the increment in $B_{10}(0)$ used in the finite difference estimation.

replications. Root mean square errors (RMSE) are estimated relative to the accurate values. The LRM method substantially outperforms the finite difference estimators when the computing effort required to estimate all deltas is held fixed.

5 Gamma

Second derivatives are typically somewhat harder to estimate than first derivatives. In this section, we present and compare three methods for estimating second derivatives of option prices with respect to initial values of forward rates (or equivalently of initial bond prices): the standard central difference estimator, a combination of the pathwise and likelihood ratio methods, and a pure likelihood ratio method.

As in Section 3, denote by $g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n))$ the discounted payoff of some derivative security and consider the generic problem of estimating

$$\frac{\partial^2}{\partial L_k(0)^2} E[g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n))] \quad (28)$$

or

$$\frac{\partial^2}{\partial B_k(0)^2} E[g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n))]. \quad (29)$$

Using the deterministic relation between the initial forward rates and the initial bond prices we can convert an estimator of either gamma into an estimator of the other. To emphasize the dependence of the expected value on the initial term structure we write

$$G(L_0(0), L_1(0), \dots, L_N(0)) = E[g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n))]$$

or $G(B_1(0), \dots, B_{N+1}(0))$ for the same quantity. To emphasize the role of a single forward rate $L_k(0)$ with all others held fixed, we write $G(L_k(0))$.

For a central difference estimator, we choose an $\epsilon > 0$ and make the approximation

$$\frac{\partial^2}{\partial L_k(0)^2} E[g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n))] \approx \frac{1}{\epsilon^2} [G(L_k(0) + \epsilon) - 2G(L_k(0)) + G(L_k(0) - \epsilon)]. \quad (30)$$

The terms $G(L_k(0) \pm \epsilon)$ are estimated in separate simulations in which the initial value of the k th forward Libor rate is set to $L_k(0) \pm \epsilon$. The accuracy of this method can be very sensitive to the choice of ϵ : smaller values will lead to larger variance in the difference estimator because of the ϵ in the denominator of (30); larger values will lead to larger bias due to the approximation in (30). Ideally, one would like to choose ϵ to balance these considerations by, e.g., minimizing mean square error, but the optimal ϵ may be quite sensitive to the form of the discounted payoff g and to model parameters.

A pure extension of the pathwise method in Section 3 to gamma is rarely possible because option payoffs are seldom twice differentiable. The example of a caplet should make this clear. We argued in Section 3 that the mapping $L_n \mapsto (L_n - K)^+$ could be differentiated almost everywhere to yield

$$\frac{d}{dL_n} (L_n - K)^+ = \mathbf{1}\{L_n > K\}. \quad (31)$$

When we try to differentiate a second time to produce a gamma estimator we face exactly the same obstacle as in estimating delta for a digital payoff. The indicator can be differentiated almost everywhere, but its derivative is zero wherever it exists and is thus completely uninformative.

It is, however, possible to combine a pathwise estimate of delta with an LRM term to arrive at a mixed estimate of gamma: the LRM term has the effect of “differentiating” the pathwise delta estimate. Consider, then, a pathwise delta estimate of the form

$$\sum_{i=1}^N \frac{\partial g}{\partial \hat{L}_i(t_i)} \hat{\Delta}_{ik}(t_i). \quad (32)$$

We restrict attention to the case in which $\hat{\Delta}$ is calculated based on the forward-drift approximation (17). If we now want to apply the LRM method, we must note that the initial values $L_k(0)$ affect the expected value of (32) in two ways: implicitly through the distribution of the $\hat{L}_i(t_i)$ (just as in Section 4.2) but also explicitly through the functional dependence of the $\hat{\mu}_i^o$ on these values. The latter dependence enters (32) through the dynamics of $\hat{\Delta}_{ik}$. This dependence *is* differentiable a second time (even though the derivative of g may not be) so we use a second application of the

pathwise method for this term together with LRM. The resulting estimator is

Mixed Pathwise-LRM Gamma Estimator

$$\left(\sum_{n=1}^N \frac{\partial g}{\partial \hat{L}_n(t_n)} \hat{\Delta}_{nk}(t_n) \right) (X - m) \Sigma^{-1} \hat{\mathbf{m}} + \sum_{n=1}^N \frac{\partial g}{\partial \hat{L}_n(t_n)} \hat{\Delta}_{nkk}(t_n) \quad (33)$$

with

$$\begin{aligned} \hat{\Delta}_{nkj}(ih) &= \frac{\partial \hat{\Delta}_{nk}(ih)}{\partial L_j(0)} = \\ &= \frac{\hat{L}_n(ih)}{L_k(0)} \sum_{r=1}^i \frac{\partial \hat{\mu}_n^o(rh)}{\partial L_j(0)} + \hat{\Delta}_{nj}(ih) \sum_{r=1}^i \frac{\partial \hat{\mu}_n^o(rh)}{\partial L_k(0)} + \hat{L}_n(ih) \sum_{r=1}^i \frac{\partial^2 \hat{\mu}_n^o(rh)}{\partial L_k(0) \partial L_j(0)} \end{aligned} \quad (34)$$

and X , \mathbf{m} , $\hat{\mathbf{m}}$, and Σ as in (24)-(26).

A few remarks on this estimator are in order:

- Equation (34), though a bit more complex than those we encountered in Section 3, is easily evaluated because (unlike the exact pathwise expression (13)) it is not recursive — it can be evaluated at time ih directly from the simulated forward Libor rates at that time and from derivatives of $\hat{\mu}^o$ that can be precomputed.
- The estimator in (33) applies to the “diagonal” gamma $\partial^2/\partial L_k^2(0)$ and uses $\hat{\Delta}_{nkj}$ only with $j = k$. We have included the more general case in (34) to include the possibility of estimating an “off-diagonal” gamma of the form $\partial^2/(\partial L_j(0)\partial L_k(0))$. For this case, replace $\hat{\Delta}_{nkk}$ with $\hat{\Delta}_{nkj}$ in (33), and in the definition (26) of $\hat{\mathbf{m}}$ replace k with j .
- The recursion in (34) determines the values of the $\hat{\Delta}_{nkj}$ on the time grid $\{0, h, 2h, \dots\}$ whereas in (33) we have implicitly allowed evaluation of these variables at arbitrary times. In practice, one can either arrange to have all relevant dates lie on the simulation time grid or else interpolate linearly between grid points.
- The estimator in (33) would in fact be unbiased (for the discretized process with time step h) if the forward-drift approximation held exactly (i.e., if the Libor rates were multivariate geometric Brownian motion with time-varying drift) provided the full-rank condition necessary to define Σ^{-1} holds. Hence, (33) does not entail any approximations beyond the forward-drift approximation and the time-discretization inherent to simulation.

Just as we derived the mixed gamma estimator by applying LRM to a pathwise delta estimate, we can derive an alternative estimator by applying LRM to an LRM delta estimate. For example,

in the Gaussian setting surrounding (20) we could differentiate twice to get

$$\begin{aligned} \frac{d^2}{d\theta^2} E_\theta[g(X)] &= \int g(x) \ddot{\phi}(x; \mathbf{m}(\theta), \Sigma) dx \\ &= \int g(x) \frac{\ddot{\phi}(x; \mathbf{m}(\theta), \Sigma)}{\phi(x; \mathbf{m}(\theta), \Sigma)} \phi(x; \mathbf{m}(\theta), \Sigma) dx. \end{aligned} \quad (35)$$

Simple calculations show that

$$\frac{\ddot{\phi}(x; \mathbf{m}(\theta), \Sigma)}{\phi(x; \mathbf{m}(\theta), \Sigma)} = [(x - \mathbf{m}(\theta))' \Sigma^{-1} \dot{\mathbf{m}}]^2 - \dot{\mathbf{m}}' \Sigma^{-1} \dot{\mathbf{m}} + (x - \mathbf{m}(\theta))' \Sigma^{-1} \ddot{\mathbf{m}}. \quad (36)$$

Evaluating this expression at $x = X$ and multiplying it by $g(X)$ yields the LRM estimator of the second derivative with respect to θ . In the Libor model, we make the correspondences (24)-(26) and

$$\ddot{\mathbf{m}}_n \leftarrow -\frac{\mathbf{1}\{n = k\}}{L_k^2(0)} + h \sum_{r=0}^{i^*-1} \frac{\partial^2 \hat{\mu}_n^o(rh)}{\partial L_k^2(0)}, \quad n = 1, \dots, N, \quad (37)$$

with

$$\frac{\partial^2 \hat{\mu}_n^o(rh)}{\partial L_k^2(0)} = \frac{-2\delta^2 \lambda_n(ih) \lambda_k'(ih)}{(1 + \delta L_k(0))^3} \mathbf{1}\{\eta(ih) \leq k \leq n\}.$$

We thus arrive at

LRM Gamma Estimator

$$g(\hat{L}_1(t_1), \dots, \hat{L}_N(t_n)) \left\{ [(X - \mathbf{m})' \Sigma^{-1} \dot{\mathbf{m}}]^2 - \dot{\mathbf{m}}' \Sigma^{-1} \dot{\mathbf{m}} + (X - \mathbf{m})' \Sigma^{-1} \ddot{\mathbf{m}} \right\}, \quad (38)$$

with X , \mathbf{m} , Σ , $\dot{\mathbf{m}}$ as in (24)-(26) and $\ddot{\mathbf{m}}$ as in (37).

An LRM estimator for $\partial^2/(\partial L_j(0)L_k(0))$ can be derived similarly: replace the scalar parameter θ of ϕ with a vector and replace (36) with the calculation of $\partial^2/(\partial\theta_j\partial\theta_k)$; then make the usual correspondences to convert to the Libor setting.

5.1 Numerical Results

As mentioned in Section 4.3, a standard caplet C_n can be hedged using just the bonds B_n and B_{n+1} . Conversely, B_n is useful in hedging C_{n-1} and C_n . Second derivatives of the form $\partial^2 C_{n-1}/\partial B_n^2$ and $\partial^2 C_n/\partial B_n^2$ are thus relevant in checking gamma estimates.

Table 3 presents numerical results for caplet gammas of this form. The pure LRM gamma estimator produces very large standard errors in this application and is therefore omitted from further comparison. Exact values are calculated from Black's formula. Much as in Section 4.3, the computing time required to estimate all gammas using finite differences is roughly 20 times as great as the time required to estimate all gammas using the mixed pathwise-LRM method. The

Method		Delta	Estimator	Std. Err.	RMSE
Exact Value		$\partial^2 C_9 / \partial B_{10}(0)^2$	105.851	—	—
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	96.374	—	—
Pathwise-LRM Method		$\partial^2 C_9 / \partial B_{10}(0)^2$	105.377	0.639	0.796
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	96.872	1.365	1.452
Finite Difference Method	$\Delta b = 0.0005$	$\partial^2 C_9 / \partial B_{10}(0)^2$	105.420	1.146	1.225
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	97.549	1.099	1.609
Finite Difference Method	$\Delta b = 0.001$	$\partial^2 C_9 / \partial B_{10}(0)^2$	105.566	0.778	0.829
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	97.007	0.745	0.977
Finite Difference Method	$\Delta b = 0.002$	$\partial^2 C_9 / \partial B_{10}(0)^2$	104.295	0.499	1.634
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	95.620	0.481	0.895
Finite Difference Method	$\Delta b = 0.003$	$\partial^2 C_9 / \partial B_{10}(0)^2$	102.071	0.365	3.326
		$\partial^2 C_{10} / \partial B_{10}(0)^2$	93.853	0.355	2.546

Table 3: Gamma estimates for caplets. To balance the computing time required to estimate all gammas, we use 500,000 replications for the mixed pathwise-LRM estimator and 100,000 for the finite difference estimators. Δb is the increment in $B_{10}(0)$ used in the finite difference estimation.

results in the table are based on balancing the number of simulated paths for each method so that the total computing time to estimate all gammas would be equal across methods. The results for the mixed method also use the control variate $Z' \mathbf{\Lambda}_h(i^*)^{-1} \mathbf{m}$, which reduces the standard error by about 20%. The increment in B_n used for the finite difference estimates is indicated in the table by Δb .

The results in the table are consistent with the view that estimating gammas is more difficult than estimating deltas; the accuracy we get for the gamma of these Lipschitz continuous payoffs is similar to what we get for the deltas of discontinuous payoffs, consistent with the discussion surrounding (31). The mixed pathwise-LRM method is sometimes outperformed by the best finite difference estimate. However, the finite difference method is very sensitive to the choice of Δb , and a good Δb may not be known in advance. Unless Δb can be chosen carefully, the mixed method appears preferable.

6 Vega

We now turn to estimating sensitivities with respect to changes in volatility. We frame the problem by introducing a parameter θ in the volatilities $\lambda_n(t)$ and considering derivatives with respect to θ . Setting $\partial \lambda_n(\theta, t) / \partial \theta \equiv 1$ for some n (and all t) but $\partial \lambda_i(\theta, t) / \partial \theta \equiv 0$ for all $i \neq n$ corresponds to a parallel shift in the volatilities of L_n ; setting $\partial \lambda_n(\theta, t) / \partial \theta \equiv 1$ for all n corresponds to a parallel shift in all volatilities. Sensitivities to volatility buckets can be modeled by restricting nonzero values of $\partial \lambda_n(\theta, t) / \partial \theta$ to t in some interval.

Write $\hat{\Delta}_n(t) \equiv \hat{\Delta}_n(\theta, t)$ for $\partial \hat{L}_n(t) / \partial \theta$. Once we have computed the $\hat{\Delta}_n$, the general pathwise

estimator takes the form in (8) but with $\hat{\Delta}_{nk}$ replaced by $\hat{\Delta}_n$. Differentiating (11) yields

Exact Pathwise Algorithm: Vega

$$\hat{\Delta}_n((i+1)h) = \hat{\Delta}_n(ih) \frac{\hat{L}_n((i+1)h)}{\hat{L}_n(ih)} + \hat{L}_n((i+1)h) \left(\left\{ \frac{\partial \hat{\mu}_n(ih)}{\partial \theta} - \frac{\partial \lambda_n(ih)}{\partial \theta} \lambda'_n(ih) \right\} h + \frac{\partial \lambda_n}{\partial \theta} Z_{i+1} \sqrt{h} \right) \quad (39)$$

with initial condition $\hat{\Delta}_n(0) \equiv 0$.

In (39), $\partial \lambda_n / \partial \theta$ and λ_n are row vectors and Z_{i+1} and λ'_n are column vectors. The differentiated drift appearing in (39) abbreviates the full expression

$$\frac{\partial \hat{\mu}_n}{\partial \theta} = \sum_{j=\eta(ih)}^n \left\{ \frac{\partial \hat{\mu}_n}{\partial \hat{L}_j} \hat{\Delta}_j + \sum_{k=1}^d \frac{\partial \hat{\mu}_n}{\partial \lambda_{jk}} \frac{\partial \lambda_{jk}}{\partial \theta} \right\},$$

with λ_{jk} denoting the k th component of λ_j . The presence of the $\hat{\Delta}_j$ in this expression makes simulation of (39) somewhat time-consuming, requiring effort comparable to simulating another copy of the Libor rates with a perturbed value of θ .

We therefore consider the forward-drift approximation. Differentiating $\hat{\mu}_n^o$ with respect to θ yields

$$\frac{\partial \hat{\mu}_n^o(ih)}{\partial \theta} = \sum_{j=\eta(ih)}^n \frac{\delta L_j(0)}{1 + \delta L_j(0)} \left(\frac{\partial \lambda_n(ih)}{\partial \theta} \lambda'_j(ih) + \frac{\partial \lambda_j(ih)}{\partial \theta} \lambda'_n(ih) \right). \quad (40)$$

This expression is independent of the simulated path and can thus be precomputed. Replacing $\hat{\mu}$ with $\hat{\mu}^o$ in (11), differentiating and then simplifying the resulting expression yields

Forward Drift Approximation: Vega

$$\hat{\Delta}_n(ih) = \hat{L}_n(ih) \left(\sum_{j=1}^i \left\{ \frac{\partial \hat{\mu}_n^o(jh)}{\partial \theta} - \frac{\partial \lambda_n(jh)}{\partial \theta} \lambda'_n(jh) \right\} h + \sum_{j=1}^i \frac{\partial \lambda_n(jh)}{\partial \theta} Z_j \sqrt{h} \right). \quad (41)$$

Besides \hat{L}_n , the only term on the right side of (41) that cannot be precomputed is

$$\sum_{j=1}^i \frac{\partial \lambda_n(jh)}{\partial \theta} Z_j \sqrt{h}.$$

But evaluating this expression along each simulated path requires very little effort, making (41) much faster than the exact pathwise method or simulation of a second copy of the Libor rates.

Table 4 presents numerical results for caplet vegas $\partial C_n / \partial \theta$ with all $\partial \lambda_i / \partial \theta \equiv 1$, corresponding to a parallel shift in the term structure of volatility. Exact values are calculated from Black's

n	Exact Value	Exact Pathwise		Pathwise Approximation	
		Estimator	Std. Err.	Estimator	Std. Err.
1	24.77	24.62	0.13	24.62	0.13
2	35.13	35.18	0.19	35.17	0.19
3	43.03	43.06	0.24	43.03	0.24
4	49.95	50.08	0.29	50.03	0.29
5	56.11	56.01	0.33	55.91	0.33
6	61.38	61.20	0.38	61.01	0.37
7	66.53	66.05	0.41	65.76	0.41
8	71.23	70.61	0.45	70.17	0.44
9	75.50	74.80	0.48	74.23	0.47
10	79.69	78.91	0.51	78.12	0.50
11	83.62	82.82	0.54	81.81	0.53
12	87.16	86.42	0.57	85.06	0.56
13	90.55	89.54	0.59	87.77	0.58
14	93.88	92.85	0.61	90.71	0.60
15	97.04	95.83	0.63	93.24	0.62
16	99.89	98.35	0.65	95.16	0.63
17	102.69	100.80	0.67	96.98	0.65
18	105.36	103.36	0.68	98.92	0.66
19	107.74	106.00	0.70	101.00	0.67

Table 4: Estimated caplet vegas $\partial C_n / \partial \theta$ and standard errors (all in basis points). Estimates are based on 100,000 replications.

formula. Both the exact pathwise and the forward drift approximation methods perform well. The approximation method produces larger bias at long maturities, but saves nearly one third of the computing time compared to the exact method.

7 Convergence of Pathwise Estimators

In this section, we give a theoretical analysis of the (exact) pathwise delta estimators of Section 3. We show that the discrete-time algorithm produces unbiased estimators for deltas of the discrete-time forward Libor process, the continuous-time estimators are unbiased for the continuous-time forward Libor process, and the discrete-time estimators converge to the continuous-time deltas as the simulation time step decreases to zero. Theoretical support for the (discrete-time) LRM estimators in a Gaussian setting follows fairly well-established lines (e.g., [9]) and is therefore omitted.

7.1 Unbiasedness: Discrete Time

Fix a time increment h and consider the processes defined by (11) and (13).

Theorem 1 Suppose $g : \mathbf{R}^N \rightarrow \mathbf{R}$ is Lipschitz continuous. Then

$$E \left[\sum_{n=1}^N \left\{ \frac{\partial}{\hat{L}_n(i_n h)} g(\hat{L}_1(i_1 h), \dots, \hat{L}_N(i_N h)) \right\} \hat{\Delta}_{nk}(i_n h) \right] = \frac{\partial}{\partial L_k(0)} E[g(\hat{L}_1(i_1 h), \dots, \hat{L}_N(i_N h))],$$

for any i_1, \dots, i_N ; i.e., the discrete-time pathwise estimator is unbiased.

Proof. Because g is Lipschitz it is differentiable almost everywhere so the partial derivatives evaluated at the \hat{L}_n exist with probability one. To emphasize the dependence on initial conditions, write $\hat{L}_n(ih) = \hat{L}_n(ih, L_k(0))$, and $\hat{L}_n^\epsilon(ih) = \hat{L}_n(ih, L_k(0) + \epsilon)$. Then

$$|g(\hat{L}_1^\epsilon(i_1 h), \dots, \hat{L}_N^\epsilon(i_N h)) - g(\hat{L}_1(i_1 h), \dots, \hat{L}_N(i_N h))| \leq K_g \sum_{n=1}^N |\hat{L}_n^\epsilon(i_n h) - \hat{L}_n(i_n h)|$$

for some constant K_g . The theorem now follows from the dominated convergence theorem once we show that

$$E[\hat{\Delta}_{nk}(ih)] = \frac{\partial}{\partial \hat{L}_k(0)} E[\hat{L}_n(ih)] \quad (42)$$

for all n and i .

Since n is arbitrary, we lighten the notation by writing simply $\hat{L}(ih) = \hat{L}_n(ih, L_k(0))$ and $\hat{L}^\epsilon(ih) = \hat{L}_n(ih, L_k(0) + \epsilon)$. We will use induction to prove that

$$\frac{1}{\epsilon} \left| \hat{L}^\epsilon(ih) - \hat{L}(ih) \right| \leq K_i, \quad (43)$$

where K_i is a random variable measurable with respect to $\{Z_1, \dots, Z_i\}$ with $E[K_i] < \infty$. It is easy to see that (43) holds for $i = 0$. Assuming that (43) holds for some i , we have

$$\begin{aligned} & \frac{1}{\epsilon} \left| \hat{L}^\epsilon((i+1)h) - \hat{L}((i+1)h) \right| \\ &= \frac{1}{\epsilon} \exp \left(-\frac{1}{2} \lambda_n(ih) \lambda_n(ih)' h + \lambda_n(ih) \sqrt{h} Z_{i+1} \right) \left| \hat{L}^\epsilon(ih) \exp(\hat{\mu}(\hat{L}^\epsilon(ih))) - \hat{L}(ih) \exp(\hat{\mu}(\hat{L}(ih))) \right|. \\ &\leq \frac{\xi}{\epsilon} \left[\left| \exp(\hat{\mu}(\hat{L}^\epsilon(ih))) (\hat{L}^\epsilon(ih) - \hat{L}(ih)) \right| + \left| \hat{L}(ih) \left(\exp(\hat{\mu}(\hat{L}^\epsilon(ih))) - \exp(\hat{\mu}(\hat{L}(ih))) \right) \right| \right] \\ &\leq \xi C_1 K_i + \frac{\xi}{\epsilon} \left| \hat{L}(ih) C_2 \left(\hat{\mu}(\hat{L}^\epsilon(ih)) - \hat{\mu}(\hat{L}(ih)) \right) \right| \\ &\leq \xi C_1 K_i + \frac{\xi}{\epsilon} C_2 \left[\left| \left(\hat{L}(ih) - \hat{L}^\epsilon(ih) \right) \hat{\mu}(\hat{L}^\epsilon(ih)) \right| + \left| \hat{L}^\epsilon(ih) \hat{\mu}(\hat{L}^\epsilon(ih)) - \hat{L}(ih) \hat{\mu}(\hat{L}(ih)) \right| \right] \\ &\leq \xi C_1 K_i + \xi C_3 K_i + \xi C_4 K_i. \end{aligned}$$

We have defined $\xi = \exp \left(-\frac{1}{2} \lambda_n(ih) \lambda_n(ih)' h + \lambda_n(ih) \sqrt{h} Z_{i+1} \right)$, which is a random variable independent of K_i . We have also used the fact that $\hat{\mu}(\cdot)$ is bounded, $\hat{\mu}(\hat{L}(\cdot)) \hat{L}_n(\cdot)$ is Lipschitz (as in (44) below) and the inequality $|e^x - e^y| \leq C|x - y|$ for bounded x, y . Now we can define $K_{i+1} = \xi C_1 K_i + \xi C_3 K_i + \xi C_4 K_i$, and clearly, we have $E[K_{i+1}] < \infty$. Equation (42) now follows from the dominated convergence theorem. \square

7.2 Unbiasedness: Continuous Time

We first justify equation (10) for the dynamics of the Δ_{nk} using a result from the theory of stochastic flows (as in Chapter 5 of Protter [16]). Pikovsky [4] has also used the stochastic-flow formulation for the Monte Carlo estimation of sensitivities. Write L for the vector (L_1, \dots, L_N) and Δ for $(\Delta_{1k}, \dots, \Delta_{Nk})$ with k fixed but arbitrary. Insert the initial conditions $L(0)$ and $\Delta(0)$ as arguments of these processes. Write $\mathbf{1}_k$ for the N -vector whose n th coordinate is $\mathbf{1}\{n = k\}$.

Lemma 1 *There exists a unique process $(L(t, L(0)), \Delta(t, \mathbf{1}_k))$ satisfying the stochastic differential equation (9)–(10). For each t , $L(t, L(0))$ is continuously differentiable with respect to $L(0)$, and $\frac{\partial}{\partial L_k(0)} L_n(t, L(0)) = \Delta_{nk}(t, \mathbf{1}_k)$.*

Proof. Set $\lambda_{max} = \sup_{n,t} \|\lambda_n(t)\|$. It is straightforward to see that $\mu_n(t)L_n(t)$ is globally Lipschitz, because

$$\|\mu_n(t, x)x_n - \mu_n(t, y)y_n\| \leq 2N\lambda_{max}^2\|x - y\|.$$

The first derivatives

$$\frac{\partial}{\partial x_k} (\mu_n(t, x)x_n) = \begin{cases} 0 & ; \quad k < \eta(t) \\ \frac{\delta\lambda_n\lambda'_k x_n}{(1+\delta x_k)^2} & ; \quad \eta(t) \leq k < n \\ \frac{\delta\lambda_n\lambda'_n x_n}{(1+\delta x_n)^2} + \frac{\delta\lambda_n\lambda'_n x_n}{1+\delta x_n} & ; \quad k = n \\ 0 & ; \quad k > n \end{cases} \quad (44)$$

also satisfy a Lipschitz condition. Thus, the lemma follows from Theorem V.39 of Protter [16]. In particular, we can write the solution of (9), (10) as

$$L_n(t) = L_n(0) \exp \left[\int_0^t \mu_n(s) ds - \frac{1}{2} \int_0^t \lambda_n(s)\lambda'_n(s) ds + \int_0^t \lambda_n(s) dW_s \right] \quad (45)$$

$$\Delta_{nk}(t) = \mathbf{1}\{n = k\} \frac{L_n(t)}{L_n(0)} + L_n(t) \int_0^t \sum_j \frac{\partial \mu_n(s)}{\partial L_j(s)} \Delta_{jk}(s) ds. \quad (46)$$

□

We now show that the continuous-time pathwise estimator is unbiased.

Theorem 2 *Let $(L(\cdot), \Delta(\cdot))$ be the solution of equation (9),(10). Suppose $g : \mathbf{R}^N \rightarrow \mathbf{R}$ is Lipschitz continuous. Then*

$$E \left[\sum_{n=1}^N \left\{ \frac{\partial}{\partial L_n(t_n)} g(L_1(t_1), \dots, L_N(t_N)) \right\} \Delta_{nk}(t_n) \right] = \frac{\partial}{\partial L_k(0)} E[g(L_1(t_1), \dots, L_N(t_N))],$$

for any t_1, \dots, t_N ; i.e., the continuous-time pathwise estimator is unbiased.

Proof. As in Theorem 1, it suffices to prove

$$\frac{\partial}{\partial L_k(0)} E[L_n(t)] = E[\Delta_{nk}(t)]. \quad (47)$$

Define $X_n(t) = \log L_n(t)$ so that

$$dX_n(t) = (\mu_n(t) - \frac{1}{2}\lambda_n(t)\lambda'_n(t))dt + \lambda_n(t)dW(t). \quad (48)$$

Differentiate to get

$$dY_{nk}(t) = \sum_j \frac{\partial \mu_n(t)}{\partial X_j(t)} Y_{jk}(t) dt = \sum_j \frac{\partial \mu_n(t)}{\partial L_k(t)} L_k(t) Y_{jk}(t) dt. \quad (49)$$

From Theorem V.39 of Protter [16], we know that there exists $(X(t), Y(t))$ solving the equations (48)-(49) with $Y_{nk}(t) = \partial X_n(t) / \partial X_k(0) = L_k(0) \partial X_n(t) / \partial L_k(0)$. In particular, we have

$$X_n(t) = \log L_n(t) = \log L_n(0) + \int_0^t \mu_n(s) ds - \frac{1}{2} \int_0^t \lambda_n(s) \lambda'_n(s) ds + \int_0^t \lambda_n(s) dW(s), \quad (50)$$

$$Y_{nk}(t) = \mathbf{1}\{n = k\} + \int_0^t \sum_j \frac{\partial \mu_n(s)}{\partial L_j(s)} \frac{\partial L_j(s)}{\partial X_j(s)} Y_{jk}(s) ds = \mathbf{1}\{n = k\} + \int_0^t \sum_j \frac{\partial \mu_n(s)}{\partial L_j(s)} L_j(s) Y_{jk}(s) ds, \quad (51)$$

and $\Delta_{nk}(t)L_k(0) = L_n(t)Y_{nk}(t)$. Together, (45) and (51) solve the SDEs (9) and (49). Moreover, $\mu_n(\cdot)L_n(\cdot)$, $\lambda(\cdot)L_n(\cdot)$ and $\sum_j (\partial \mu_n(\cdot) / \partial L_j(\cdot)) L_j(\cdot) Y_{jk}(\cdot)$ satisfy a Lipschitz condition, so from Theorem V.9 of Protter [16] we have

$$\lim_{\epsilon \rightarrow 0} E \|(L_n, Y_{nk})(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - (L_n, Y_{nk})(t, (L_k(0), \mathbf{1}_k))\|^2 = 0. \quad (52)$$

This gives us

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} E \|\Delta_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - \Delta_{nk}(t, (L_k(0), \mathbf{1}_k))\| \\ &= \lim_{\epsilon \rightarrow 0} E \left[\frac{1}{L_k(0)} \|(L_k(0) + \epsilon) \Delta_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - L_k(0) \Delta_{nk}(t, (L_k(0), \mathbf{1}_k))\| \right. \\ & \quad \left. - \epsilon \Delta_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k))\| \right] \\ &\leq \lim_{\epsilon \rightarrow 0} C_1 E [\|L_n(t, L_k(0) + \epsilon) Y_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - L_n(t, L_k(0)) Y_{nk}(t, (L_k(0), \mathbf{1}_k))\|] \\ &\leq \lim_{\epsilon \rightarrow 0} C_1 E [\|L_n(t, L_k(0) + \epsilon)\| \|Y_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - Y_{nk}(t, (L_k(0), \mathbf{1}_k))\|] \\ & \quad + \lim_{\epsilon \rightarrow 0} C_1 E [\|Y_{nk}(t, (L_k(0), \mathbf{1}_k))\| \|L_n(t, L_k(0) + \epsilon) - L_n(t, L_k(0))\|] \\ &\leq \lim_{\epsilon \rightarrow 0} C_1 \left[E \|L_n(t, L_k(0) + \epsilon)\|^2 E \|Y_{nk}(t, (L_k(0) + \epsilon, \mathbf{1}_k)) - Y_{nk}(t, (L_k(0), \mathbf{1}_k))\|^2 \right]^{1/2} \\ & \quad + \lim_{\epsilon \rightarrow 0} C_1 \left[E \|Y_{nk}(t, (L_k(0), \mathbf{1}_k))\|^2 E \|L_n(t, L_k(0) + \epsilon) - L_n(t, L_k(0))\|^2 \right]^{1/2} \\ &= 0, \end{aligned}$$

because $E\|L_n(t)\|^2$ and $E\|Y_{nk}(t)\|^2$ are finite. This establishes the continuity of $E\Delta(t, L_k(0))$. On the other hand, from Lemma 1, we have

$$L_n(t, L_k(0) + \epsilon) = L_n(t, L_k(0)) + \int_0^\epsilon \Delta_{nk}(t, L(0) + \delta) d\delta.$$

Taking expectation on both sides, we get

$$\begin{aligned} E[L_n(t, L_k(0) + \epsilon)] &= E[L_n(t, L_k(0))] + E\left[\int_0^\epsilon \Delta_{nk}(t, L(0) + \delta) d\delta\right] \\ &= E[L_n(t, L_k(0))] + \int_0^\epsilon E[\Delta_{nk}(t, L(0) + \delta)] d\delta \end{aligned}$$

with the interchange of integral and expectation justified by Tonelli's Theorem (Fubini for non-negative integrands). This representation of $E[L_n(t, L_k(0))]$ implies (47) through the continuity of $E[\Delta_{nk}(t, L(0))]$. \square

7.3 Convergence from Discrete to Continuous Time

We now show that the discrete-time $\hat{\Delta}_{nk}$ are asymptotically unbiased estimators of the continuous-time derivatives $\partial E[L_n]/\partial L_k(0)$ as the time increment h decreases to 0.

Theorem 3 *Suppose that $(L_n(t), \Delta_{nk}(t))$ solve (9)–(10) and $(\hat{L}_n(t), \hat{\Delta}_{nk}(t))$ solve (11)–(13). Then we have*

$$\lim_{h \rightarrow 0} E[\hat{\Delta}_{nk}(t)] = E[\Delta_{nk}(t)]. \quad (53)$$

Proof. Discretize $Y_{nk}(t)$ using

$$\begin{aligned} \hat{Y}_{nk}((i+1)h) &= \hat{Y}_{nk}(ih) + \sum_{j=\eta(ih)}^n \frac{\partial \mu_n(ih, \hat{L}(ih))}{\partial \hat{L}_j(ih)} \hat{L}_j(ih) \hat{Y}_{jk}(ih) h \\ &= \hat{Y}_{nk}(ih) + \sum_{j=\eta(ih)}^n \frac{\delta \hat{L}_j(ih)}{(1 + \delta \hat{L}_j(ih))^2} \hat{Y}_{jk}(ih) h. \end{aligned} \quad (54)$$

Clearly, we have $L_k(0)\hat{\Delta}_{nk}(ih) = \hat{Y}_{nk}(ih)\hat{L}_n(ih)$. From the fact that $\mu(\cdot)$ is positive and bounded, and the inequality $e^x \leq 1 + x + x^2 e^x, \forall x > 0$, we get (writing \mathcal{A}_i for $\{Z_1, \dots, Z_i\}$)

$$\begin{aligned} &E \left[\left| E \left[\frac{\hat{L}_n((i+1)h) - \hat{L}_n(ih)}{h} \middle| \mathcal{A}_i \right] - \mu_n(\hat{L}(ih))\hat{L}_n(ih) \right|^2 \right] \\ &= E \left[\left| E \left[\frac{\hat{L}_n(ih)}{h} \left(\exp(\mu(\hat{L}(ih))h) - \frac{1}{2}\lambda_n\lambda'_n + \lambda_n\sqrt{h}Z_i \right) - 1 \right] - \mu_n(\hat{L}(ih))\hat{L}_n(ih) \right|^2 \right] \\ &= E \left[\left| \hat{L}_n(ih) \right|^2 \left| \frac{1}{h} \left(\exp(\mu(\hat{L}(ih))h) - 1 \right) - \mu_n(\hat{L}(ih)) \right|^2 \right] \\ &\leq E \left[\left| \hat{L}_n(ih) \right|^2 \left| \frac{1}{h} \left(\mu(\hat{L}(ih))h + \exp(\mu(\hat{L}(ih))h)(\mu(\hat{L}(ih))h)^2 \right) - \mu(\hat{L}(ih)) \right|^2 \right] \\ &\leq C_1 h^2, \end{aligned}$$

and

$$\begin{aligned}
& E \left[\frac{1}{h} \left| \hat{L}_n((i+1)h) - \hat{L}_n(ih) - \mathbb{E} \left(\hat{L}_n((i+1)h) - \hat{L}_n(ih) \mid \mathcal{A}_i \right) - \lambda_n(ih) \hat{L}_n(ih) Z_{i+1} \sqrt{h} \right|^2 \right] \\
&= E \left[\frac{|\hat{L}_n(ih)|^2}{h} \left| e^{(\mu_n(\hat{L}_n(ih))h - \frac{1}{2} \lambda_n \lambda'_n h + \lambda_n Z_{i+1} \sqrt{h})} - e^{(\mu_n(\hat{L}_n(ih))h) - \lambda_n(ih) Z_{i+1} \sqrt{h}} \right|^2 \right] \\
&= E \left\{ \frac{|\hat{L}_n(ih)|^2}{h} \mathbb{E} \left[\left(e^{(\mu_n(\hat{L}_n(ih))h - \frac{1}{2} \lambda_n \lambda'_n h + \lambda_n Z_{i+1} \sqrt{h})} - e^{(\mu_n(\hat{L}_n(ih))h) - \lambda_n(ih) Z_{i+1} \sqrt{h}} \right)^2 \mid \mathcal{A}_i \right] \right\} \\
&= E \left[\frac{|\hat{L}_n(ih)|^2}{h} \left(e^{(2\mu_n(\hat{L}_n(ih))h + \lambda_n \lambda'_n h)} + e^{(2\mu_n(\hat{L}_n(ih))h)} + \lambda_n \lambda'_n h - 2e^{(2\mu_n(\hat{L}_n(ih))h)} - 2\lambda_n \lambda'_n h e^{(\mu_n(\hat{L}_n(ih))h)} \right) \right] \\
&= E \left[\frac{|\hat{L}_n(ih)|^2}{h} \left(e^{(2\mu_n(\hat{L}_n(ih))h)} (e^{\lambda_n \lambda'_n h} - 1) - 2\lambda_n \lambda'_n h e^{(\mu_n(\hat{L}_n(ih))h)} + \lambda_n \lambda'_n h \right) \right] \\
&\leq E \left[\frac{|\hat{L}_n(ih)|^2}{h} \left(e^{(2\mu_n(\hat{L}_n(ih))h)} (\lambda_n \lambda'_n h + e^{\lambda_n \lambda'_n h} (\lambda_n \lambda'_n)^2 h^2) - 2\lambda_n \lambda'_n h e^{(\mu_n(\hat{L}_n(ih))h)} + \lambda_n \lambda'_n h \right) \right] \\
&\leq E \left[\lambda_n \lambda_n |\hat{L}_n(ih)|^2 \left(e^{(\mu_n(\hat{L}_n(ih))h)} - 1 \right)^2 \right] + C_2 h \\
&\leq C_3 h^2 + C_2 h.
\end{aligned}$$

A similar argument shows that $\hat{Y}_{nk}(\cdot)$ also satisfies these conditions. From Theorem 9.6.2 of Kloeden and Platen [13], we get

$$\lim_{h \rightarrow 0} E \left| \hat{L}_n(t) - L_n(t) \right|^2 = 0, \quad \lim_{h \rightarrow 0} E \left| \hat{Y}_{nk}(t) - Y_{nk}(t) \right|^2 = 0.$$

Thus,

$$\begin{aligned}
& \lim_{h \rightarrow 0} E \left| \hat{\Delta}_{nk}(t) - \Delta_{nk}(t) \right| \\
&= \frac{1}{L_k(0)} \lim_{h \rightarrow 0} E \left| L_k(0) \hat{\Delta}_{nk}(t) - L_k(0) \Delta_{nk}(t) \right| \\
&= \frac{1}{L_k(0)} \lim_{\epsilon \rightarrow 0} E \left| \hat{Y}_{nk}(t) \hat{L}_n(t) - Y_{nk}(t) L_n(t) \right| \\
&\leq \frac{1}{L_k(0)} \lim_{h \rightarrow 0} \left[E \left| \hat{Y}_{nk}(t) \left(\hat{L}_n(t) - L_n(t) \right) \right| + E \left| L_n(t) \left(\hat{Y}_{nk}(t) - Y_{nk}(t) \right) \right| \right] \\
&\leq \frac{1}{L_k(0)} \lim_{h \rightarrow 0} \left[\left(E \left| \hat{Y}_{nk}(t) \right|^2 E \left| \hat{L}_n(t) - L_n(t) \right|^2 \right)^{1/2} + \left(E \left| L_n(t) \right|^2 E \left| \hat{Y}_{nk}(t) - Y_{nk}(t) \right|^2 \right)^{1/2} \right] \\
&= 0.
\end{aligned}$$

□

In (53), we could replace $\hat{\Delta}_{nk}$ with any Lipschitz continuous function of all $(\hat{L}_n, \hat{\Delta}_{nk})$ and obtain the corresponding result; however, the Lipschitz requirement rules out even the indicator function appearing in the pathwise estimator of caplet deltas.

Acknowledgements. We thank Leif Andersen of General Re Financial Products for several valuable discussions. This work is supported in part by NSF grant DMI-9457189 and an IBM University Partnership Award.

References

- [1] Boyle, P., Broadie, M., and Glasserman, P., Simulation methods for security pricing, *J. Economic Dynamics and Control*, 21, 1267-1321 (1997).
- [2] Broadie, M., and Glasserman, P., Estimating security price derivatives using simulation, *Management Science*, 42, 269-285, 1996.
- [3] Brace, A., Gatarek, D., and Musiela, M., The market model of interest rate dynamics, *Math. Finance*, 7, 127-155 (1997).
- [4] Pikovsky, I., Methods of stochastic flows in contingent claims valuation, working paper, (1998)
- [5] Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., and Touzi, N., Applications of Malliavin calculus to Monte Carlo methods in finance, *Finance and Stochastics*, 3, (1999).
- [6] Fu, M., and Hu, J.Q., Sensitivity analysis for Monte Carlo simulation of option pricing, *Probability in the Engineering and Information Sciences*, 9, 417-446, (1995).
- [7] Glasserman, P., *Gradient Estimation via Perturbation Analysis*, Kluwer Publishers, Norwell, Mass., (1991).
- [8] Glasserman, P., and Zhao, X., Arbitrage-free discretization of lognormal forward Libor and swap rate models, *Finance and Stochastics*, to appear.
- [9] Glynn, P., and L'Ecuyer, P., Likelihood ratio gradient estimation for stochastic recursions, *Advances in Applied Probability*, 27, 1019-1053, (1995).
- [10] 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).
- [11] Ho, Y.C., and Cao, X., *Perturbation Analysis of Discrete Event Dynamic Systems*, Kluwer Publishers, Norwell, Mass., (1991).
- [12] Jamshidian, F., LIBOR and swap market models and measures, *Finance and Stochastics*, 1, 293-330 (1997).
- [13] Kloeden, P., and Platen, E., *Numerical Solution of Stochastic Differential Equations*, Springer, New York (1992).
- [14] Miltersen, K.R., Sandmann, K., and Sondermann, D., Closed-form solutions for term structure derivatives with lognormal interest rates, *J. Finance*, 409-430 (1997).
- [15] Musiela, M., and Rutkowski, M., Continuous-time term structure models: Forward measure approach, *Finance and Stochastics*, 1, 261-292 (1997).
- [16] Protter, P., *Stochastic Integration and Differential Equations*, Springer-Verlag, New York, (1990).
- [17] Reiman, M.I. and Weiss, A., Sensitivity analysis for simulations via likelihood ratios, *Oper. Res.*, 37, 830-844, (1989).
- [18] Rubinstein, R., and Shapiro, A., *Discrete Event Systems*, Wiley, New York, (1993).