Conditioning on One-Step Survival for Barrier Option Simulations
Author(s): Paul Glasserman and Jeremy Staum
Source: *Operations Research*, Vol. 49, No. 6 (Nov. - Dec., 2001), pp. 923-937
Published by: INFORMS
Stable URL: https://www.jstor.org/stable/3088605
Accessed: 20-04-2019 15:28 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/3088605?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Operations Research*

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

CONDITIONING ON ONE-STEP SURVIVAL FOR BARRIER OPTION SIMULATIONS

PAUL GLASSERMAN

Graduate School of Business, Columbia University, New York, New York 10027, pglasser@research.gsb.columbia.edu

JEREMY STAUM

226 Rhodes Hall, Cornell University, Ithaca, New York 14853, staum@orie.cornell.edu

Pricing financial options often requires Monte Carlo methods. One particular case is that of barrier options, whose payoff may be zero depending on whether or not an underlying asset crosses a barrier during the life of the option. This paper develops variance reduction techniques that take advantage of the special structure of barrier options, and are appropriate for general simulation problems with similar structure. We use a change of measure at each step of the simulation to reduce the variance arising from the possibility of a barrier crossing at each monitoring date. The paper details the theoretical underpinnings of this method, and evaluates alternative implementations when exact distributions conditional on one-step survival are available and when not available. When these one-step conditional distributions are unavailable, we introduce algorithms that combine change of measure and estimation of conditional probabilities simultaneously. The methods proposed are more generally applicable to terminal reward problems on Markov processes with absorbing states.

1. INTRODUCTION

Barrier options are derivative securities with the defining characteristic that the payoff may be zero, depending on whether or not an underlying variable crosses a specified barrier during the life of the option. There are two broad types of barrier options: "knock-out" options, which pay zero when there is a barrier crossing, and "knock-in" options, which pay zero unless there is a barrier crossing. A barrier option is cheaper than the equivalent option without a barrier, because it may expire worthless if knocked out (or not knocked in) in the same situation in which the standard option would have paid off. Perhaps because of this property, barrier features are frequently incorporated into option contracts on many different types of underlying assets.

To price an option is to evaluate the integral of its expected discounted payoff under a risk-neutral probability measure. (See, e.g., Duffie 1996 or Hull 1993 for background on option pricing.) In the case of barrier options, this payoff is discontinuous over the space of all paths of the underlying variables. In sufficiently simple cases, there are analytical formulas for the price (e.g., Merton 1973, Kunitomo and Ikeda 1992, Rubinstein and Reiner 1991, and Sidenius 1998). However, there will not be useful formulas if the specification of the barrier or stochastic processes used to model the underlying variables is too complex or high dimensional. Consequently, it is often necessary to price via simulation, which is better suited to high-dimensional and path-dependent problems than other numerical methods, and has the added benefit of providing the estimate's standard error.

A straightforward simulation proceeds by dividing the lifetime of the option into several time steps. Each path begins with the state vector at a specified initial value, and uses an approximate, discretized version of the dynamics to propagate this random vector forward at each time step. A simulation for a knock-out option has the special feature that if at any time the underlying process should cross the barrier, the path may be immediately abandoned, because it is already known that it results in a zero payoff. If a path "survives" by never crossing the barrier, then its payoff is determined at the terminal value of the state vector. This standard Monte Carlo approach to pricing knock-out options suffers from a peculiar defect which creates a possibility for improving the method: Some simulated paths survive, allowing for a positive payoff, while other paths fail to survive and have zero payoff. The lower the probability of survival, the more simulated payoffs are zero. This can make the average payoff among the surviving paths, and hence the variance among all paths, quite large relative to the price.

All the variance due to the possibility of knock-out could be removed by importance sampling if it were possible to use the conditional distribution of the state vector given survival to maturity. However, barrier option simulation is challenging precisely because it is generally impossible to sample conditional on final survival. It becomes necessary to simulate an entire discrete path with observations at each barrier-monitoring point to ascertain whether or not survival occurs. However, there are other possibilities for sampling measures, which may result in more efficient estimators.

Subject classifications: Simulation, efficiency: Variance reduction. Finance, asset pricing: Computational methods.
Area of review: SIMULATION.

0030-364X/01/4906-0923 $05.00
1526-5463 electronic ISSN

Operations Research © 2001 INFORMS
Vol. 49, No. 6, November–December 2001, pp. 923–937

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

One technique presented here is to use the conditional distribution given *one-step* survival, which may be available, even though the conditional distribution given *final* survival is not. By sampling conditional on survival at each step, it is possible to ensure that all simulated paths survive and yield information about potentially positive payoffs. Then it is necessary to incorporate a likelihood ratio to make up for the absence of paths which get knocked out. The result is an unbiased estimator of the option price with reduced variance. The idea of using a conditional distribution at each step in a simulation was investigated by Glynn and Inglehart (1988, §11) in a different context—estimating the steady-state mean of a real-valued Markov chain. Glasserman (1993) analyzes a continuous-time version of this idea. An example of using one-step conditional distributions in barrier option simulations appears in Boyle et al. (1997). Independent of our work, Ross and Shanthikumar (1999) combine two examples from Boyle et al. (1997) to arrive at an estimator similar to the simplest one considered here; they do not consider the case of unknown conditional probabilities for which we propose several alternative estimators. For other simulation methods applied to pricing path-dependent options, see, e.g., Duan and Simonato (1998), Glasserman et al. (1999a), Joy et al. (1996), Lemieux and L'Ecuyer (1998), and Vázquez-Abad and Dufresne (1998).

We propose and analyze a variety of estimators that can be divided into two broad categories defined by whether or not the estimator requires explicit knowledge of the distribution of the underlying process conditional on one-step survival. This conditional distribution enters the problem in two ways—through sampling conditional on survival and through evaluation of a likelihood ratio. The second role turns out to be the more fundamental one; for settings in which the conditional distribution is unknown, we formulate methods which implement a change of measure by estimating the required likelihood ratio.

The techniques we study are most natural for discretely monitored barrier options, which are knocked out only if the barrier is crossed on specified monitoring dates. In this case, there is an obvious way to fix the time steps of the discretization. However, we show that these methods also work in the case of continous monitoring, with an arbitrarily chosen discretization. Other extensions addressed in this paper are knock-in options and rebates, which are paid if the option is knocked out.

For all of these situations where the central idea of conditioning on survival over one step is feasible, we propose algorithms and associated estimators in §2. In §3, we analyze their properties, while deferring any mathematical details which would detract from the flow of the argument until Appendix A. Section 4 reports numerical results for the performance of these algorithms for various knock-out options, while §5 presents conclusions.

2. ESTIMATORS AND ALGORITHMS

2.1. The Mathematical Problem

We treat the following problem in pricing a knock-out option. There is a stochastic model of the evolution of a state vector $S_t, t \in [0, T]$ of underlying financial variables. We assume that S_t has the Markov property. The state vector must be defined so that, if the option has not been knocked out, the terminal value S_T determines the discounted payoff $f(S_T)$. For example, in a model with stochastic interest rates, it will be necessary to include the accumulated discount factor (which is itself stochastic) as part of the underlying process. For an Asian option (whose payoff depends on the average level of the underlying asset), the state vector must include the running sum, as well as the current price of the underlying asset. Also, we deal only with European-style options, which can be exercised only at maturity. To condense notation, from now on we will write the state vector at the monitoring dates as $(S_1, \dots, S_m)$ (abbreviating the more explicit $(S_{t_1}, \dots, S_{t_m})$), and assume that the time between consecutive monitoring dates $t_{i+1} - t_i$ is a constant Δt.

Because of the knock-out feature, the state vector S_i takes a value in $\Re^d \cup \Delta$, where Δ is an absorbing state. If S crosses the barrier at time i, the option is knocked out, and for all $j \geqslant i$, $S_j = \Delta$. Define A_i to be the indicator function $\mathbf{1}\{S_i \neq \Delta\}$; that is, $A_i = 1$ means that the option is alive at time t_i. The structure of this problem is by no means limited to barrier options. It applies to any simulation of expected terminal reward with the feature that there is a known payoff when the state vector exits a specified region of the state space.

We now give three motivating examples to which we will return later as well.

EXAMPLE 1. BLACK-SCHOLES MODEL One example is a one-dimensional Black-Scholes model. The state vector is a single stock price, governed by the dynamics

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t,$$

where W_t is standard Brownian motion, and μ and σ are constants. This has exact discretization

$$S_{i+1} = S_i \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\, Z_i\right), \tag{1}$$

where $Z_1, \dots, Z_{m-1}$ are i.i.d. standard normal. The barrier is a price level $H < S_0$, so $A_i = 1$ if the stock price has not crossed beneath the barrier H by step i. A down-and-out call in this model has discounted payoff $A_m \exp(-rT)(S_m - K)^+$, where K is the strike price (i.e., the price at which the holder of the call option may buy the stock), r the constant interest rate, and $T = t_m$ the maturity. A down-and-out binary call's discounted payoff is $A_m \exp(-rT)\mathbf{1}\{S_m \geqslant K\}$. These examples can be priced in closed form using the results of Merton (1973), but numerical methods become necessary when the barrier is time-varying or the parameters μ and σ are stochastic.

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

EXAMPLE 2. TWO-DIMENSIONAL GEOMETRIC BROWNIAN MOTION In this example, the state vector contains two stock prices, obeying dynamics

$$\frac{dS_t^{(1)}}{S_t^{(1)}} = \mu_1 dt + \sigma_1 dW_t^{(1)},$$

$$\frac{dS_t^{(2)}}{S_t^{(2)}} = \mu_2 dt + \sigma_2 dW_t^{(2)},$$

where $W_t = (W_t^{(1)}, W_t^{(2)})$ is two-dimensional Brownian motion with zero drift and covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The discretized dynamics are

$$\begin{aligned} S_{i+1}^{(1)} &= S_i^{(1)} \exp\left(\left(\mu_1 - \frac{1}{2}\sigma_1^2\right)\Delta t + \sigma^{(1)}\sqrt{\Delta t}\, Z_i^{(1)}\right), \\ S_{i+1}^{(2)} &= S_i^{(2)} \exp\left(\left(\mu_2 - \frac{1}{2}\sigma_2^2\right)\Delta t + \sigma^{(2)}\sqrt{\Delta t}\, Z_i^{(2)}\right), \end{aligned} \tag{2}$$

where

$$Z_i^{(1)} = Y_i^{(1)},$$

$$Z_i^{(2)} = \rho Y_i^{(1)} + \sqrt{1-\rho^2}\, Y_i^{(2)},$$

and all Y are i.i.d. standard normal. In this setting, the barrier crossing may be determined by one asset and the final payoff by the other. For example, the discounted payoff may be

$$e^{-rT}(S_m^{(2)} - K)^+ \mathbf{1}\{\min_{i \leqslant m} S_i^{(1)} \geqslant H\}.$$

Heynen and Kat (1994) have provided closed-form solutions for some such options, but not, for instance, where the barrier is a function of time more complicated than an exponential.

EXAMPLE 3. LIBOR MARKET MODEL Our final motivating example is indicative of a more complex class of models in widespread use for pricing interest rate derivative securities. In this setting, the underlying state vector records interest rates rather than asset prices. We describe a model of London Inter-Bank Offered Rates (LIBOR), a key benchmark for pricing. For further background, see Musiela and Rutkowski (1997, §14.3).

Given a set of maturities $0 < t_1 < t_2 < \cdots < t_{N+1}$, let $F_t^{(k)}$ denote the *forward* interest rate for the period $[t_k, t_{k+1})$ as of time $t < t_k$. This is the interest rate for the period $[t_k, t_{k+1})$ that can be locked in at time t. In a model with proportional volatilities, the dynamics of the vector of forward rates take the form

$$dF_t^{(k)} = \mu^{(k)}(F_t, \Lambda_t)F_t^{(k)}\, dt + F_t^{(k)}\lambda_t^{(k)}\, dW_t, \quad k = 1, \ldots, N.$$

Here, W_t is a d-dimensional standard Brownian motion ($d \leqslant N$), and $\lambda_t^{(k)}$ is the kth row of an $N \times d$ deterministic matrix Λ_t, expressing the instantaneous dependence of $F_t^{(k)}$ on the d components of the Brownian motion. Each $\mu^{(k)}$, $k = 1, \ldots, N$, is a deterministic function of the vector of rates F_t and the volatilities Λ_t. The specification of $\mu^{(k)}$ is fully determined by the requirement that the model be arbitrage-free, as explained by Musiela and Rutkowski (1997, Equation 14.55). The choice of Λ_t determines the instantaneous covariance $\Lambda_t\Lambda_t'$.

Using an Euler discretization of the logarithm of the forward rates, where the time subscript i as usual means t_i, the discretized dynamics are

$$\ln(F_{i+1}^{(k)}) = \ln(F_i^{(k)}) + \left(\mu_i^{(k)}(F_i, \Lambda_i) - \frac{1}{2}\|\lambda_i^{(k)}\|^2\right)\Delta t + C_i^{(k)} Z_i \sqrt{\Delta t}, \tag{3}$$

where Z_i is a vector of d independent standard normals and $C_i^{(k)}$ is the kth row of an $N \times d$ matrix C_i which satisfies $C_i C_i' = \Lambda_i \Lambda_i'$. The choice $C_i = \Lambda_i$ is always available, but we will see that flexibility in the choice of C_i can be useful.

Many types of interest rate options with barrier features are commonly traded. We will consider *barrier swaptions* in particular. These are options that are knocked out if some function of the forward LIBOR rates crosses a barrier, with a payoff that is essentially a call or a put on a swap which is an agreement to exchange interest payments at "floating" (variable) LIBOR rates for payments at a fixed rate κ. The swap rate κ which makes the initial value of the swap zero is itself a complicated function of the forward LIBOR rates. As in Musiela and Rutkowski (1997, Equation 16.5),

$$\kappa = \frac{1 - B_M}{\Delta t \sum_{k=m+1}^{M} B_k} \tag{4}$$

is the swap rate at time t_m for a swap with payment dates t_{m+1} to t_M. Here, B_i is the price (at time t_m) of a bond paying \$1 at time t_i and is given by

$$B_i = \prod_{k=m}^{i-1} \frac{1}{1 + \Delta t F_m^{(k)}}; \tag{5}$$

i.e., by discounting \$1 at the forward LIBOR rates. A swaption on \$1 of notional principal has a time t_m payoff of

$$\left(\Delta t \sum_{k=m+1}^{M} B_k\right) \max\{0, \kappa - K\}, \tag{6}$$

where K is the strike rate, and the summation is the time t_m value of a bond paying \$1 at times t_{m+1} to t_M. The key features of this example for our purposes are that the underlying model dynamics are fairly complex, a typical payoff is a complicated function of the state vector, and a barrier may be imposed on a nonlinear function of the state vector (e.g., the swap rate).

Before analyzing these examples, recall how a standard simulation proceeds, as described in the introduction. In any of the examples above, a standard Monte Carlo algorithm simulates $(S_1, \ldots, S_m)$, where S_i is a vector of

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

underlying prices (or forward interest rates) at the ith discretization time. It proceeds in the usual fashion by generating S_{i+1} from S_i according to the law of the underlying process, or perhaps an approximate discretization thereof. Then it uses this path to evaluate the following estimator. The estimated price is the average of X_m over multiple paths.

Standard Estimator.

$$X_m \triangleq A_m f(S_m). \tag{7}$$

2.2. Known Transition Probabilities

An alternative to the standard estimator generates S_{i+1} from S_i conditional on $A_{i+1}=1$, if this one-step conditional distribution is known. Of course, under this scheme, $A_m=1$ always, and the average of $f(S_m)$ may be terribly biased. For this reason, it is necessary to weight by a likelihood ratio. (See, e.g., Bratley et al. 1987 for background on importance sampling.) Define

$$L_i \triangleq \prod_{j=0}^{i-1} p(S_j), \tag{8}$$

where

$$p(s) \triangleq \mathbf{P}[S_{j+1} \neq \Delta \mid S_j = s]. \tag{9}$$

We adopt the convention that $p(S_j)=0$ if $A_j=0$, because there is no chance of surviving to the next step if the barrier has already been crossed. (This formulation is general enough to allow the probability of one-step survival to depend on j because the time index can be incorporated in the state vector.) So L_i is the "likelihood" of surviving i steps via this path, in a sense to be made precise in §3. Then the new estimator is:

Exact Estimator with Full Importance Sampling.

$$\widehat{X}_m \triangleq L_m f(S_m). \tag{10}$$

Theoretical properties of this and all subsequent estimators are presented in §3. In particular, we show that this estimator is unbiased and has lower variance than the standard estimator.

Example 1 continued. In the example of pricing a down-and-out call under the Black-Scholes model, it is easy to sample conditional on one-step survival. An algorithm to sample S_{i+1} unconditionally implements Equation (1) by

$$S_{i+1} = S_i \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}\,\Phi^{-1}(U)\right), \tag{11}$$

where U is uniformly distributed and Φ is the standard normal cdf. It is easy to evaluate Φ^{-1} numerically, as shown by Marsaglia et al. (1994).

Sampling conditional on one-step survival uses the same equation, except that

$$U = (1-p(S_i)) + Vp(S_i), \tag{12}$$

where V is uniformly distributed and

$$\begin{aligned} p(S_i) &= P(S_{i+1} \geqslant H \mid S_i) \\ &= \Phi\left(\left(\ln\left(\frac{S_i}{H}\right) + \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t\right) \Big/ \left(\sigma\sqrt{\Delta t}\right)\right). \end{aligned} \tag{13}$$

The result is that U is uniformly distributed conditional on being at least as large as necessary to prevent knockout. That is, given S_i and given $S_{i+1} \geqslant H$, $\Phi^{-1}(U)$ has the distribution of $(\ln(S_{i+1}) - \ln(S_i) - (\mu - \sigma^2/2)\Delta t)/(\sigma\sqrt{\Delta t})$. It is also possible (and perhaps faster) to sample from the tail of the normal distribution using acceptance-rejection rather than Φ^{-1}; see Fishman (1996). The logarithmic evaluation in (13) is easily avoided by storing the exponent in (11). Implementation of (13) does entail the overhead of evaluating a cumulative normal probability; fast approximations to Φ are included in many mathematical software libraries.

Example 2 continued. In the setting of Equation (2), sampling conditional on one-step survival works as follows:

$$S_{i+1}^{(1)} = S_i^{(1)} \exp\left(\mu^{(1)}\Delta t + \sigma^{(1)}\sqrt{\Delta t}\,\Phi^{-1}(U^{(1)})\right),$$

$$S_{i+1}^{(2)} = S_i^{(2)} \exp\Big(\mu^{(2)}\Delta t + \sigma^{(2)}\sqrt{\Delta t}\,(\rho\Phi^{-1}(U^{(1)}) + \sqrt{1-\rho^2}\,\Phi^{-1}(U^{(2)}))\Big),$$

where

$$U^{(1)} = (1-p(S_i)) + V^{(1)}p(S_i),$$
$$U^{(2)} = V^{(2)},$$

and $V^{(1)}, V^{(2)}$ are uniformly distributed and independent. The probability of one-step survival $p(S_i)$ is exactly as in Equation (13), but with $S_i^{(1)}$, $\mu^{(1)}$, and $\sigma^{(1)}$ for S_i, μ, and σ, because the barrier condition involves only the first asset price.

Example 3 continued. Suppose that the barrier is a floor beneath the current LIBOR rate, i.e., the option is knocked out at step $i+1$ if $F_{i+1}^{(i+1)}$ is beneath the barrier. It is easiest to implement Equation (3) by choosing the matrix C_i such that $F_{i+1}^{(i+1)}$ depends on a single component of the driving Brownian motion. That is, pick $C_i^{(i+1)}$ such that it has the form $[\sigma, 0, \ldots, 0]$, e.g., through Cholesky factorization. (All rows of C_i with index i or less are zero, because these LIBOR rates refer to maturities already in the past.) Much as in Example 2, first simulate the current LIBOR rate conditional on its being above the floor, and then simulate all forward LIBOR rates conditional on this value of the current LIBOR rate. For the current LIBOR rate,

$$\ln(F_{i+1}^{(i+1)}) = \ln(F_i^{(i+1)}) + \left(\mu_i^{(i+1)}(F_i, \Lambda_i) - \frac{1}{2}\|\lambda_i^{(i+1)}\|^2\right)\Delta t + \sqrt{\Delta t}\,C_i^{(i+1,1)}\Phi^{-1}(U_i^{(1)}),$$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

and the probability of surviving one step is

$$p(F_i) = \Phi\left(\left[\ln(F_i^{(i+1)}/H) + \left(\mu_i^{(i+1)}(F_i, \Lambda_i) - \frac{1}{2}\|\lambda_i^{(i+1)}\|^2\right)\Delta t\right] \Big/ \left(\sigma\sqrt{\Delta t}\right)\right).$$

In general, for $k > i$,

$$\ln(F_{i+1}^{(k)}) = \ln(F_i^{(k)}) + \left(\mu_i^{(k)}(F_i, \Lambda_i) - \frac{1}{2}\|\lambda_i^{(k)}\|^2\right)\Delta t + \sqrt{\Delta t}\sum_{j=1}^{d} C_i^{(k,j)}\Phi^{-1}(U_i^{(j)}),$$

where j indexes the components of the Brownian motion:

$$U_i^{(1)} = (1 - p(S_i)) + V_i^{(1)} p(S_i),$$
$$U_i^{(j)} = V_i^{(j)}, \quad j = 2, \ldots, d,$$

and $V_i^{(1)}, \ldots, V_i^{(d)}$ are uniformly distributed and independent.

Note that in this example, choosing a different square root of the covariance matrix would have made it difficult to find the probability of surviving one step. An inherently difficult case is when the barrier is a floor beneath more than one forward LIBOR rate simultaneously. The distribution for the minimum among those rates is inconvenient, regardless of the square root of the covariance matrix, and it is awkward to find the probability of survival analytically. If the barrier is a floor beneath the forward swap rate itself, this is such a complicated function of forward LIBOR rates that it is effectively impossible to determine the probability of survival. Similar to Equation (4), the forward swap rate from time t_m to time t_M is

$$\kappa_i(m, M) = \frac{B_i^{(m)} - B_i^{(M)}}{\Delta t \sum_{k=m+1}^{M} B_i^{(k)}}. \tag{14}$$

(See Musiela and Rutkowski 1997, 14.64.) This example is featured in the next subsection, which develops methods for use when transition probabilities are unknown.

2.3. Unknown Transition Probabilities

When the conditional probability of one-step survival $p(S_i)$ is not known, the technique of the previous subsection is not applicable. However, as long as $p(S_i) > 0$, it is always possible to sample conditional on one-step survival by generating unconditional successors to S_i, and keeping the first one to survive. This still leaves the problem of evaluating the likelihood ratio when $p(S_i)$ is unknown, a problem of potentially broader scope. To address it, we estimate $p(S_i)$ in the simulation itself. This is made possible by the observation that the number of unconditional successors required to generate a survivor (the *waiting time*) is geometrically distributed with unknown parameter $p(S_i)$. Therefore, one can try to estimate $p(S_i)$ from data observed in the course of simulation, and use the approach of §2.2 with an estimated, instead of an exact, likelihood.

A problem swiftly emerges in the form of an irritating conundrum in elementary statistics: Given a single observation, there is only one unbiased estimator of the parameter of a geometric distribution, and it is perfectly useless for our purposes. If Y_i is the observed geometric waiting time, then the only unbiased estimator is the indicator function which is 1 when $Y_i = 1$ and 0 otherwise (e.g., Cox and Hinkley 1974, p. 253). That is, if the first unconditional successor to S_i survives, estimate that $p(S_i)$ is 1, but if it does not survive, estimate that $p(S_i)$ is 0. Of course, plugging this estimator into the formula for L_m results in A_m, because the only values are 0 and 1, reducing the method to standard simulation.

However, the problem is not intractable, because there is a good unbiased estimator when more than one observation is available. An algorithm incorporating it is this:

1. Produce unconditional successors to S_i until r of them survive;
2. Call the total waiting time Y_i, which has a negative binomial distribution;
3. Let S_{i+1} be any of the r surviving successors;
4. Estimate $p(S_j)$ by

$$\frac{r-1}{Y_i - 1}.$$

Replacing L_m by an estimated likelihood ratio, a new estimator is then:

Negative Binomial Estimator.

$$f(S_m)\prod_{i=0}^{m-1}\frac{r-1}{Y_i - 1}. \tag{15}$$

This estimation algorithm would clearly be less efficient than standard simulation if there were no chance of knock-out, because it would produce $r > 1$ surviving successors at each step, then throw out $r - 1$ of them. As knock-out gets more likely, the standard method grows inefficient in that it does too much of its work in generating successors at early steps, because most paths terminate early. The standard method wastes some of the value of early successors in that early termination of a path means that the random variates already generated provide no information about the final payoff or the probability of surviving later steps. When the probability of knock-out is high enough and the number of steps m is large enough, the negative binomial estimator will outperform the standard estimator.

A potential drawback of this negative binomial method is that there is no upper bound on the amount of time it might take to sample a surviving successor from the unconditional distribution. In order to avoid this difficulty, we can sample a fixed number n of successors from the unconditional distribution, and record the number N_j which survive, which is binomially distributed. Then the estimate of $p(S_j)$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

is N_j/n, but the likelihood ratio need not have the same form as before. It is now possible for a path not to survive, if at one step none of the n successors generated survive. In this sense, the algorithm takes only partial advantage of importance sampling. This requires a new estimator:

Exact Estimator with Partial Importance Sampling.

$$X_m^n \triangleq A_m L_m^n f(S_m), \tag{16}$$

where

$$L_i^n \triangleq \prod_{j=0}^{i-1} \frac{p(S_j)}{1-(1-p(S_j))^n}, \tag{17}$$

because $(1-(1-p(S_j))^n)$ is the probability of finding at least one surviving successor in n attempts; §3 provides a theoretical explanation of this point.

However, the $p(S_j)$ are not known, so L_i must be estimated. One natural possibility is:

Binomial Estimator.

$$A_m f(S_m) \prod_{j=0}^{m-1} \frac{N_j}{n}. \tag{18}$$

The product in (18) is an "empirical" counterpart of the likelihood ratio; as shown in §3, it results in an unbiased estimator.

This binomial estimator is a special case of a more general class of estimators based on right-censored geometric random variables. The insight is that for a large maximum computational budget of n at step i, if $p(S_i)$ is not too small, it might be inefficient (in terms of variance reduction for a fixed amount of work) to spend time estimating $p(S_i)$ extremely precisely rather than to move on, saving the budget for simulating more paths. For instance, each step could involve r trials which end after one successor survives or $N = n/r$ successors fail to survive, whichever comes first. Then the length of the trial is a right-censored geometric random variable, and the binomial estimator is the special case of $N = 1$ and $r = n$. Such a censored geometric estimator is consistent but is not competitive because it is biased at low values of r or N and does not improve efficiency for large r and N. For details, see Staum (2001).

The binomial estimator (18) bears some resemblence to splitting or "RESTART" estimators considered in, e.g., Villén-Altamirano and Villén-Altamirano (1994) and Glasserman et al. (1999b) for estimating rare event probabilities. However, in splitting algorithms all surviving paths are simulated, whereas in (18), multiple survivors are used to estimate the one-step survival probability but simulation continues for just one surviving path.

2.4. Extension: Continuous Monitoring

This framework can also apply to continuously monitored barrier options. In this case, the dates $t_1, \ldots, t_m$ are merely for purposes of discretization, and at step i, the option is knocked out not only if $S_{t_i} = S_i$ is on the wrong side of the barrier, but if S_t crossed the barrier at any time t in the interval $(t_{i-1}, t_i]$. In general, the knock-out condition can be expressed as an inequality $b(S) < H$, where b is a function of the underlying asset price. Then augment the state vector S_i to be (S_i, M_i), where $M_i \triangleq \min_{t\in(t_{i-1},t_i]} b(S_t)$. If the joint distribution of S and M is known, then the discretized simulation can effectively monitor the barrier continuously. This is also possible if b is a vector-valued function, and the minimum is taken coordinatewise. References for simulation of barrier options with continuous monitoring are Andersen and Brotherton-Ratcliffe (1996), Baldi et al. (1999), and Beaglehole et al. (1997). Asmussen et al. (1995) considered the related problem of simulating the maximum of Brownian motion.

If the conditional distribution of M_{i+1} given S_i is not known, then the only choice is to use the methods of the previous subsection where transition probabilities are unknown, and the state vector is now defined to be the pair (S_i, M_i). However, if this conditional distribution is known, it would be desirable to implement conditioning on one-step survival exactly, by simulating M_{i+1} given S_i and the event $A_{i+1} = 1$, then simulating S_{i+1} given S_i and M_{i+1}. Unfortunately, this is often not practical. The difficulty arises in the latter step; often the distribution of S_{i+1} conditional on S_i and M_{i+1} is unknown. For instance, even the case of one-dimensional Brownian motion with nonzero drift is complicated, because it matters at what time the minimum was achieved.

Instead, take advantage once more of the trick of simulating conditional on survival by repeated unconditional sampling. To do this, it is only necessary to know the marginal distributions of S_{i+1} and M_{i+1}, and the distribution of M_{i+1} conditional on S_i and S_{i+1}. Repeat the following process until it succeeds in producing S_{i+1}.

1. Generate S_{i+1} conditional on S_i and $b(S_{i+1}) > H$.
2. Generate M_{i+1} conditional on S_i and S_{i+1}.
3. If $M_{i+1} > H$, accept this value of S_{i+1}.
4. Compute $p(S_i)$ from the marginal distribution of M_{i+1}.

In Example 1, for one-dimensional Brownian motion with drift μ, implement step 1 exactly as specified by Equations (11) and (12) for sampling conditional on survival in the discrete case. Step 2 reduces to sampling the minimum of a Brownian bridge, for which the original drift has become irrelevant. From Karatzas and Shreve (1991, 4.3.40), we get

$$P[M_{i+1} \leqslant x \mid S_i, S_{i+1}] = \exp\left(\frac{2}{\sigma^2 \Delta t} \ln\left(\frac{x}{S_i}\right) \ln\left(\frac{S_{i+1}}{x}\right)\right). \tag{19}$$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

To generate M_{i+1}, invert this cdf, and evaluate at a uniformly distributed random variable U:

$$M_{i+1}=\exp\left(\frac{1}{2}\left(\ln(S_iS_{i+1})\right.\right.$$
$$\left.\left.-\sqrt{\left(\ln\left(\frac{S_i}{S_{i+1}}\right)\right)^2-2\sigma^2\Delta t\ln(U)}\right)\right). \quad (20)$$

This is similar to a result of Asmussen et al. (1995, §4.5). Also, the marginal distribution of M_{i+1} (conditional on S_i, but not on S_{i+1}) is inverse Gaussian, as in Corollary B.3.4 of Musiela and Rutkowski (1997). Evaluating it at H,

$$\begin{aligned}p(S_i)&=P[M_{i+1}>H\mid S_i]\\&=\Phi\left(\frac{\ln(\frac{S_i}{H})-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right)-\exp\left(\frac{2\mu\ln(\frac{S_i}{H})}{\sigma^2}\right)\\&\quad\times\Phi\left(\frac{-\ln(\frac{S_i}{H})-\mu\Delta t}{\sigma\sqrt{\Delta t}}\right).\end{aligned} \quad (21)$$

The procedure functions similarly for Examples 2 and 3, because a single dimension of the underlying Brownian motion determines the barrier crossing. In these cases, use the above method to simulate this single component, then sample the rest of the state vector conditional on it.

2.5. Extension: Rebates

Knock-out options are sometimes written so that the buyer receives a *rebate* if the option is knocked out. Depending on the contract specification, this rebate can be payable either at maturity or at the time of knock-out. The techniques developed in this paper are well suited to handling rebates payable at knock-out, but treatment of the topic is limited to this subsection in order to lighten the burden of notation elsewhere.

The only further assumption necessary is that the present value of the rebate payable if knock-out occurs at time t_i be a function of the state vector at time t_{i-1}. For rebates paid at knock-out, this assumption is not very restrictive. The present value of the rebate is the product of the nominal value (the amount paid) and a discount factor. The nominal value of the rebate is generally constant, and at step $i-1$ both the discount factor up to time t_{i-1} and the interest rate r_{i-1} for the interval $(t_{i-1},t_i]$ are known. The discrete dynamics of the discount factor D are $D_i=D_{i-1}\exp(-r_{i-1}\Delta t)$, so D_i is known at t_{i-1}.

Write the present value of the rebate earned at time t_i as $g(S_{i-1})$. Then the standard estimator, the realized value of the option on a simulated path, is

$$X_m=A_mf(S_m)+\sum_{i=1}^{m}(A_{i-1}-A_i)g(S_{i-1}). \quad (22)$$

The expression $A_{i-1}-A_i$ is an indicator function which is one when the option is knocked out at step i. This formula also holds if $g(S_{i-1})$ is the expected present value of the rebate at step i. The exact estimator with full importance sampling defined in Equation (10) is now

$$\widehat{X}_m=L_mf(S_m)+\sum_{i=1}^{m}L_{i-1}(1-p(S_{i-1}))g(S_{i-1}). \quad (23)$$

Rebates payable at maturity do not in general fit this framework because the discounting involves interest rates conditional on knock-out. (If discounting is not stochastic or is independent of the rest of the process, this is not an objection.) To handle a knock-out option with rebate payable at maturity, decompose it into the sum of an ordinary knock-out option and a binary knock-in option which knocks in and pays the rebate at maturity precisely when the other knocks out. The next subsection treats knock-in options.

2.6. Extension: Knock-In Options

Dealing with knock-in options is not so simple, but is possible if there is a known expression $f_i(S_i)$ for the present value of a barrierless option, received at time t_i when the state vector is S_i, whose payoff will be $f(S_m)$ at time t_m. This is the case for sufficiently simple knock-in options, and in particular for valuing rebates payable at maturity as discussed in the previous subsection. In that case, $f_i(S_i)=rD_iB_i(t_m)$, where r is the nominal rebate amount and $B_i(t_m)$ the price at time t_i of \$1 paid at time t_m.

We continue to use A_i to mean the indicator function which is one if the barrier has not been crossed by time t_i, which in this situation means that the option has *not* yet been knocked in. As before, $p(S_i)$ is the probability of not crossing the barrier over this step. Then, a different standard-type estimator taking advantage of the principle that a knocked-in barrier option is effectively transformed into a barrierless option is

$$X_m=\sum_{i=1}^{m}(A_{i-1}-A_i)f_i(S_i). \quad (24)$$

To take advantage of full importance sampling, it is necessary to complicate the sampling scheme. At each step there must be two successors to S_i. Simulating conditional on no knock-in produces S_{i+1}, while simulating conditional on knock-in produces S^*_{i+1}. The S^*_{i+1} are not part of the path, but are used to estimate the value of the option should it be knocked in at step $i+1$, since in this case it is not realistic to expect that this value should be known at time i. The estimator is:

$$\widehat{X}_m=\sum_{i=1}^{m}L_{i-1}(1-p(S_{i-1}))f_i(S^*_i). \quad (25)$$

3. PROPERTIES OF THE ESTIMATORS

This section will explore statistical properties, such as bias, variance, and consistency of the estimators which the previous section discussed. We defer proofs to an appendix.

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

Precise statements of some of the properties of the estimators will require a discussion of probability measures as they relate to various simulation algorithms.

Let $\mathbf{P}$ be the measure under which the underlying state vector process $(S_1, \ldots, S_m)$ has its usual joint distribution on a universe called Ω. In a standard Monte Carlo simulation, the simulated price vectors obey the law of $\mathbf{P}$. Let $\mathcal{A}_i$ be the subset of Ω on which the previously defined indicator function A_i equals 1, i.e., where the option is "alive" at time i, and let $\mathcal{F}_i$ be the sigma-algebra generated by $(S_1, \ldots, S_i)$. Then $\widehat{\mathbf{P}}$ is defined on $\mathcal{A}_m$ relative to $\mathcal{F}_m$ through the conditional distributions

$$\widehat{\mathbf{P}}[S_{i+1} \in Q \mid S_i] \triangleq \mathbf{P}[S_{i+1} \in Q \mid S_i, A_{i+1} = 1], \tag{26}$$

where $S_0 = s_0$ is fixed.

Simulating under $\widehat{\mathbf{P}}$ means sampling the next state vector conditional on survival at the next step. Of course, this implies that $A_m = 1$ with probability 1 under $\widehat{\mathbf{P}}$; i.e., all paths simulated under $\widehat{\mathbf{P}}$ survive until the end. Our first result shows that it is possible to compensate with a likelihood ratio. Let $\widehat{\mathbf{E}}$ denote expectation with respect to $\widehat{\mathbf{P}}$.

Lemma 1. *L_i, defined in (8), is the likelihood ratio relating the measures $\widehat{\mathbf{P}}$ and the restriction of $\mathbf{P}$ to $\mathcal{A}_i$ as follows:*

$$\widehat{\mathbf{E}}[L_i Y] = \mathbf{E}[A_i Y] \tag{27}$$

for any $\mathcal{F}_i$-measurable function Y such that the expectation $\mathbf{E}[A_i Y]$ exists and is finite.

This is useful because we can define the expected payoff at step i in the simulation, which is an $\mathcal{F}_i$-measurable function:

$$X_i \triangleq \mathbf{E}[X_m \mid \mathcal{F}_i],$$

$$\widehat{X}_i \triangleq \widehat{\mathbf{E}}[\widehat{X}_m \mid \mathcal{F}_i].$$

These are not observable during the simulation, except when $i = m$, which corresponds to the final estimator. (Recall that X_m and $\widehat{X}_m$ are estimators defined in (7) and (10), respectively.) Throughout this section, we also require the nonrestrictive technical condition

$$\mathbf{E}[X_m^2] < \infty \tag{28}$$

in order to allow the use of Lemma 1 and ensure that variances exist.

Lemma 2. *The Condition (28) implies $\widehat{\mathbf{E}}[\widehat{X}_m^2] < \infty$ and $\forall\, i \leqslant m,\ |\mathbf{E}[A_i X_i]| < \infty$.*

Under this assumption, a direct consequence of Lemma 1 is

Theorem 1. *The estimator $\widehat{X}_m$ defined in Equation (10) is unbiased. That is,*

$$\widehat{\mathbf{E}}[\widehat{X}_m] = \widehat{\mathbf{E}}[L_m f(S_m)] = \mathbf{E}[A_m f(S_m)] = \mathbf{E}[X_m],$$

which equals the price of the option.

Note that while $\widehat{X}_m$ is defined as an "exact estimator with full importance sampling," we are not in the usual setting of importance sampling, because in nondegenerate examples, it is not the probability measure $\mathbf{P}$ but only its restriction to $\mathcal{A}_m$ which is absolutely continuous with respect to $\widehat{\mathbf{P}}$. (A similar lack of absolute continuity arises in other applications of importance sampling; see, e.g., Asmussen 1987, §14.7, and Glynn and Iglehart 1989.)

A simple extension of Theorem 1 is:

Theorem 2. *The estimator for knock-out options with rebates in (23) is unbiased.*

$$\widehat{\mathbf{E}}\left[L_m f(S_m) + \sum_{i=1}^{m} L_{i-1}(1 - p(S_{i-1}))g(S_{i-1})\right]$$
$$= \mathbf{E}\left[A_m f(S_m) + \sum_{i=1}^{m} (A_{i-1} - A_i) g(S_{i-1})\right].$$

This result relies on linearity of expectation and $\widehat{\mathbf{E}}[L_i(1 - p(S_i))Y_i] = \mathbf{E}[(A_i - A_{i+1})Y_i]$, which is much like Lemma 1.

The following lemma is useful for analyzing the reduction in variance which $\widehat{X}_m$ provides, which may be expressed as a sum of expected one-step conditional variances.

Lemma 3. *The variance of X_m can be expressed as*

$$\mathbf{Var}[X_m] = \sum_{i=1}^{m} \mathbf{E}[\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}]]. \tag{29}$$

We are interested in algorithms which improve upon the one-step conditional variance in the standard estimator, $\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}]$, which can be expressed as follows.

Lemma 4. *For the standard estimator, the one-step conditional variance is*

$$\begin{aligned}\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}] = {} & p(S_{i-1})\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}, A_i = 1] \\ & + p(S_{i-1})(1 - p(S_{i-1})) \\ & \times (\mathbf{E}[X_i \mid \mathcal{F}_{i-1}, A_i = 1])^2,\end{aligned} \tag{30}$$

and its expectation is

$$\mathbf{E}[\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}]] = \mathbf{E}[X_i^2] - \mathbf{E}[X_{i-1}^2]. \tag{31}$$

Finding the expected one-step conditional variance leads to the following theorem, which summarizes the variance reduction results. In the theorem and throughout, $\widehat{\mathbf{Var}}$ denotes variance with respect to $\widehat{\mathbf{P}}$.

Theorem 3. *The estimator $\widehat{X}_m$ has reduced variance*

$$\widehat{\mathbf{Var}}[\widehat{X}_m] = \mathbf{E}[L_m X_m^2] - X_0^2 \leqslant \mathbf{E}[X_m^2] - X_0^2 = \mathbf{Var}[X_m]. \tag{32}$$

The inequality is strict if $\mathbf{E}[A_m] < 1$ and $\mathbf{E}[f(S_m)] > 0$, i.e., if there is any chance of knock-out and positive payoff.

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

The variance reduction relies on the comparison of $\mathbf{E}[L_m X_m^2]$ and $\mathbf{E}[X_m^2]$. The greater the probability of knock-out, the smaller L_m tends to be, so the greater the reduction in variance.

To analyze estimation schemes with a maximum computational budget of n forward simulations at each step, define the measure $\mathbf{P}^n$, which governs the process where the path survives a step when at least one of n potential successors survives. Both $\mathbf{P}$ and $\widehat{\mathbf{P}}$ are special cases, when $n=1$ and $n=\infty$, respectively. The interpretation of $\mathbf{P}^n$ is that it takes partial advantage of importance sampling, where $\widehat{\mathbf{P}}$ takes full advantage of it and $\mathbf{P}$ takes none. $\mathbf{P}^n$ is defined on Ω relative to $\mathcal{F}_m$ by

$$\mathbf{P}^n[S_{i+1}=\Delta \mid S_i] \triangleq (1-p(S_i))^n, \tag{33}$$

$$\mathbf{P}^n[S_{i+1}\in Q \mid S_i] \triangleq (1-(1-p(S_i))^n) \times \mathbf{P}[S_{i+1}\in Q \mid S_i, A_{i+1}=1] \quad \text{for } Q\subseteq \mathcal{A}_{i+1}, \tag{34}$$

where $S_0=s_0$ is fixed, because $(1-(1-p(S_i))^n)$ is the $\mathbf{P}$-probability that at least one of n successors of S_i survives.

THEOREM 4. *The estimator X_m^n of Equation* (16) *is unbiased, i.e.,* $\mathbf{E}^n[X_m^n]=\mathbf{E}[X_m]$.

This follows from $\mathbf{E}^n[A_i Y_i L_i^n]=\mathbf{E}[A_i Y_i]$, in the spirit of Lemma 1.

THEOREM 5. *The estimator X_m^n achieves an intermediate reduction in variance*: $\mathbf{Var}^n[X_m^n]=\mathbf{E}[L_m^n X_m^2]-X_0^2$, *hence*

$$\widehat{\mathbf{Var}}[\widehat{X}_m] \leqslant \mathbf{Var}^n[X_m^n] \leqslant \mathbf{Var}[X_m], \tag{35}$$

where the inequalities are strict for $1<n<\infty$ if $\mathbf{E}[A_m]<1$ and $\mathbf{E}[f(S_m)]>0$, i.e., if there is any chance of knock-out and positive payoff.

Also interesting is the behavior of estimators when $p(S_i)$ are not known. It is highly desirable that these estimators be consistent as the number of sample paths goes to infinity, in order that barrier options may be priced to arbitrary accuracy by increasing the number of paths. Because the simulated price is an average of the values of an estimator realized on each path, it would be best if this estimator were unbiased. This is true of both the negative binomial estimator (15) under $\widehat{\mathbf{P}}$ and the binomial estimator (18) under $\mathbf{P}^n$. Both involve products of factors which are individually conditionally unbiased for what they seek to estimate, namely $p(S_i)$, and whose errors are uncorrelated.

THEOREM 6. *The negative binomial estimator is unbiased.*

$$\widehat{\mathbf{E}}\left[f(S_m)\prod_{i=0}^{m-1}\frac{r-1}{Y_i-1}\right]=\widehat{\mathbf{E}}[\widehat{X}_m].$$

THEOREM 7. *The binomial estimator is unbiased.*

$$\mathbf{E}^n\left[f(S_m)\prod_{i=0}^{m-1}\frac{N_i}{n}\right]=\mathbf{E}^n[X_m^n].$$

The proofs are based on conditional independence of Y_i and S_{i+1} in conjunction with the unbiasedness of the individual estimates of $p(S_i)$; see the appendix.

It is difficult to produce variance comparisons for estimators where the one-step survival probabilities are unknown. Instead, we rely on numerical comparisons.

4. NUMERICAL RESULTS

Section 3 contained theorems formalizing the idea that conditioning on one-step survival produces a reduction in variance, compared to standard Monte Carlo simulation. This suggests that at least the exact estimator $\widehat{X}_m$ defined in (10) should be superior to the standard estimator X_m of (7). On the other hand, $\widehat{X}_m$ requires a greater average number of transitions simulated per path than does X_m. This is because an efficient implementation of the standard simulation scheme will cease work on a path as soon as knock-out occurs, whereas an algorithm which conditions on surviving each step necessarily simulates every path for m steps. Algorithms which generate a maximum of n potential successors at each step have an intermediate number of expected transitions simulated. It is necessary to test the performance of the various estimators numerically. As the following results show, the effectiveness of the new methods depends on the specific problem to a great extent.

We examine the performance of the proposed estimators relative to the standard technique for specific benchmark options. For each of the three examples used in §2, we analyze one option with a moderate knock-out probability and one with a high knock-out probability. We do not consider options with low knock-out probability because our methods do not produce substantial variance reduction for them, as noted after the statement of Theorem 3.

EXAMPLE 1A. The underlying process is a single stock price which, under the risk-neutral measure, obeys geometric Brownian motion with annual drift $\mu=5\%$ and volatility $\sigma=60\%$. The stock's initial price $S_0=100$, and there are barriers at $H_l=95$ and $H_u=105$, so that the option is knocked out if it crosses either of these. The option has $K=S_0$ and its maturity $T=0.25$, with three monitoring dates.

EXAMPLE 1B. The option's specification is the same, except that the lower barrier $H_l=100$.

EXAMPLE 2A. The state vector contains a stock price and an index level which obey geometric Brownian motion. They both have annual drift $\mu=5\%$, and while the index has volatility $\sigma_1=40\%$, the stock has volatility $\sigma_2=60\%$. Their correlation is $\rho=0.5$. The initial value of the index is $S_0^{(1)}=1{,}000$, and the stock's initial price is $S_0^{(2)}=100$. The barriers on the index are at $H_l=950$ and $H_u=1{,}050$, and the strike for the stock price is $K=100$. The option still has maturity $T=0.25$ and three monitoring dates.

EXAMPLE 2B. This example is like the previous, but the maturity of the option is $T=3$ years, with quarterly

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

monitoring. The index has volatility $\sigma_1 = 15\%$ and the barriers are $H_l = 900$ and $H_u = 1{,}050$. The stock has volatility $\sigma_2 = 25\%$, and correlation $\rho = -0.5$ with the index.

EXAMPLE 3A. This example uses the LIBOR market model based on bonds with maturities 0.5 years apart, and this equals the simulation time step Δt. All forward rates are initially 5%. The driving Brownian motion has dimension $d = 1$, and each $\lambda^{(k)}$ is a constant 0.3. The contract is a "2 into 2" payer swaption with strike $K = 5\%$. That is, in two years the owner has the option to enter into a swap to pay 5% interest and receive the floating rate for two years. The owner will exercise the option if the swap rate for years 2 to 4 (i.e., Steps 4 to 8) at maturity $\kappa_4(4, 8)$ is above K—see Equation (4). At $T = 2$ years, the payoff is the maximum of zero, and the present value of receiving interest at rate $\kappa_4(4, 8)$—5% for two years—see Equation (6). There are barriers of $H_l = 5\%$ and $H_u = 7\%$ on the current LIBOR rate.

EXAMPLE 3B. The setting is similar, but the swaption is now 4 into 2. The maturity is $T = 4$, and again we consider a payoff based on a two-year swap rate which is now $\kappa_8(8, 12)$. The barriers are $H_l = 5\%$ and $H_u = 6\%$, but they only take effect in the second two years of the option's life; that is, there is only monitoring at steps 4 through 8.

We compare the performance of the estimators based on the product of the average number of transitions simulated per path (N) and the variance per path (V). This figure of merit appropriately penalizes the new estimators for simulating every path to maturity while the standard method abandons a path as soon as a barrier is crossed. For the estimators of §2.3 (i.e., with unknown transition probabilities), N counts all candidate transitions generated at every step, not just the number of survivors. The computational effort per transition is the same in these methods as in the standard estimator, and the total computational effort in both cases is essentially proportional to the number of such transitions. Thus, comparing the performance of the estimators of §2.3 with the standard estimator based on N*V is essentially equivalent to comparing them based on the product of average computer time per path and variance per path.

For the "exact" estimator of §2.2, a comparison based on N*V does not reflect the difference in time required to generate a conditional transition and an unconditional transition. Ordinarily, generating a conditional transition will take longer, but not much longer. An exact comparison turns out to be extremely sensitive to the precise implementation of the method (e.g., how one generates normal random variables and how one computes normal probabilities). In contrast, the product N*V should be nearly independent of the implementation.

The overhead involved in generating conditional transitions is illustrated by the comparison of (11) and (13). By simulating $\ln S_i$, and exponentiating only to get S_m (to compute the terminal payoff), we can accelerate the basic simulation (11) and avoid the evaluation of the logarithm in (13) after a one-time computation of $\ln H$. The overhead in (11) primarily consists of generating a conditional normal rather than an unconditional normal, and of one evaluation of Φ. For this example, we find that generating a conditional transition takes approximately 50% longer than an unconditional survivor. We view this as close to a worst case precisely because the basic model is so simple. For more complicated models, the time per transition in a standard simulation is greater; the additional effort to generate a conditional transition should be similar to that for (13) in absolute terms, and thus smaller as a percent of the time to generate an unconditional transition.

Table 1.

Example	Binary	Standard
1a	1.5%	1.9%
2a	4.5%	6.1%
3a	0.4%	3.9%

Tables 1–4 report numerical results for the examples listed above. In each case, we report the product N*V by normalizing the corresponding product for the standard method to be 100%.

In Table 1, we present the performance of the "$\widehat{\mathbf{P}}$ Estimator" of Equation (10) relative to the standard estimator of Equation (7) for the typical options. For each benchmark option, we report results for two versions: the standard version and a binary version, in which the payoff is 1 when the payoff of the standard version is positive, and 0 otherwise. The binary payoffs have less variance conditional on final survival. The methods proposed in this paper reduce only the variance associated with knock-out, and have no effect on variance conditional on final survival. However, it should be possible to combine them with other methods which do reduce that part of the variance.

Next, we test values of the estimator parameter r (the number of trials per step) in order to give guidelines for choosing the parameter in practice. Table 2 gives the results for Examples 1a, 2a, and 3a, while Table 3 contains the results for the higher knock-out probability options of Examples 1b, 2b, and 3b. We focus on binary options, for which the methods are most effective.

Increasing r improves the accuracy of the likelihood ratio estimation, but this accuracy comes at the price of increased

Table 2.

Method	r	Example 1a	Example 2a	Example 3a
Binomial	2	52%	72%	83%
	3	46%	70%	82%
	4	43%	72%	81%
	6	43%	80%	79%
	8	46%	88%	77%
Negative Binomial	2	173%	205%	162%
	3	104%	152%	118%
	4	85%	142%	102%
	6	82%	152%	89%
	8	90%	173%	84%

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

Table 3.

Method	r	Example 1b	Example 2b	Example 3b
Binomial	2	37%	35%	27%
	3	24%	32%	15%
	4	19%	29%	12%
	6	15%	24%	11%
	8	14%	22%	10%
Negative Binomial	2	11%	0.29%	
	3	3.7%	0.14%	
	4	2.2%	0.08%	n. a.
	6	1.6%	0.06%	
	8	1.3%	0.06%	

computational effort. The marginal benefit of large r is decreasing in r; that is, the more effort has been expended on estimating the likelihood ratio accurately, the less value there is to expending further effort on this task. This is reflected in several examples in which increasing r too much eventually increases the computational expense.

Table 3 highlights one of the shortcomings of the negative binomial estimator. It is impractical to use it for Example 3b, in which the barrier only takes effect after two years. If the forward rate is too far outside the barrier, there is a very low probability that an unconditional successor will survive the step in which the barrier first takes effect. This makes the expected waiting time unreasonably long.

This difficulty of "painting oneself into a corner" illustrates a potential shortcoming of our methodology. What one would really wish to do to reduce variance is to condition on final survival. As this is seldom possible, conditioning on one-step survival can be an effective substitute, but is not precisely the same. In general, the barrier at time i has no effect on simulation of steps $j < i-1$, so for certain problems, the state vector S_i can have a very large probability of being in a region in which there is minuscule chance of surviving step i. This situation weakens the effectiveness of all of the methods, but especially the negative binomial.

Table 4 examines the characteristics of a barrier option which make conditioning on one-step survival an effective technique. We modify Example 1 so that it is a down-and-out call, for which there is only a lower barrier.

EXAMPLE 1C. The stock's initial price is $S_0 = 100$, and it has drift $\mu = 0$ and volatility $\sigma = 30\%$. The option's maturity is $T = 0.25$. There is a single barrier at $H = 94.30$, and monthly monitoring ($m = 3$). The strike is $K = H$, so that the binary option pays 1 unless it is knocked out.

Table 4.

σ	T	H	m	Binary	Standard
30.0%	0.25	\$94.30	3	12%	58%
73.7%	0.25	\$94.30	3	9%	48%
30.0%	1.5	\$94.30	3	9%	48%
30.0%	0.25	\$98.62	3	8%	42%
30.0%	0.25	\$94.30	63	126%	155%

This means that the price of the binary option equals the probability of final survival, because there is no discounting. The price of the option is 50 cents. Then we construct four new scenarios in each of which one parameter differs:

- The volatility $\sigma = 73.7\%$.
- The maturity $T = 1.5$ years, but there are still only three monitoring dates.
- The barrier $H = 98.62$.
- The monitoring is daily, so $m = 63$.

All of these changes produce the same new, lower price of 33.9 cents.

All else being equal, a lower probability of final survival is associated with a greater benefit to these methods, because there is more variance due to survival to be eliminated. However, different factors affect the results in slightly different ways. For instance, the scenario with the tighter barrier produces superior results for the standard call. Tightening the barrier produces the greatest improvement in variance reduction because it neither increases the variance conditional on survival nor causes wild sample paths with highly variable likelihoods.

Most significantly, we see that with this knock-out probability and a large number of steps, the exact method can underperform the standard method. With the total knock-out probability held constant, increasing the number of steps weakens the performance of the exact method for two reasons. For one, the ratio of the expected number of steps simulated by the standard and exact methods is 1 when $m = 1$ and decreases to some limit as m increases. Also, from Theorem 3, the variance of the exact estimator is $\widehat{\mathbf{E}}[L_m^2(f(S_m))^2] - X_0^2$, which increases with the variance of L_m. When $m = 1$, the likelihood ratio L_m is constant and has zero variance, and its variance grows with m. This is another illustration that successive one-step survival is not the same as final survival; the more steps there are, the less conditioning on one-step survival resembles conditioning on final survival.

Table 5 illustrates this point using Example 1d with a binary payoff. The rows labeled N and N*V contain, as usual, the ratio of the estimate for the exact $\widehat{\mathbf{P}}$ estimator divided by that for the standard estimator. The option price approaches a limit which is the price with continuous monitoring. Likewise, the ratio of N approaches a limit which is T divided by the expected time until knock-out or maturity with continuous monitoring. For $m = 1$, the exact estimator has zero variance for this binary option. Despite the increasing probability of knock-out as m increases, the efficiency of the exact method decreases.

EXAMPLE 1D. This is the same as Example 1c, except that there are two barriers, $H_l = 94.40$ and $H_u = 105.60$. The frequency of monitoring m varies.

5. CONCLUSIONS

Conditioning on one-step survival at each step of a barrier option simulation is a natural extension of importance sampling which takes advantage of the particular structure of

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

Table 5.

m	1	3	6	12	24	36	48	60
Price	0.290	0.101	0.042	0.018	0.007	0.005	0.004	0.003
N	100%	176%	248%	321%	402%	441%	461%	487%
N*V	0%	0.1%	0.5%	1.3%	3.2%	5.5%	7.0%	9.2%

barrier options. We have proposed an estimator (10), which implements this technique when the state vector's distribution conditional on one-step survival is explicitly known. The estimator is unbiased and has less variance than a standard estimator, but also requires more computational effort on average. For typical barrier options, it produces a substantial improvement in efficiency, as measured by the product of work and variance. When the probability of knock-out is high, this estimator can be far more efficient than a standard Monte Carlo estimate.

It is possible to implement the same concept even when the one-step conditional distribution is unknown. Using a properly estimated likelihood ratio results in consistent estimators. Standard Monte Carlo simulation is seen to be a special case of this type of algorithm, where each step's contribution to the likelihood ratio is estimated by a single Bernoulli trial. The computational expense of using more simulations per step to estimate the likelihood ratio is justified by a sufficient reduction in variance only when the probability of knock-out is high. Consequently, the structure of the estimated likelihood ratio can have a significant effect on the estimator's performance, making it potentially difficult to choose a suitable estimator before having analyzed the problem already.

Nonetheless, it is possible to give some guidelines about when conditioning on one-step survival will be most effective. In general, the probability of knock-out is high when the underlying asset is "close" to the barrier, relative to its volatility and the maturity of the option. A fixed barrier is effectively closer when volatility is high or maturity is long. However, large volatilities and maturities are associated with higher payoff variance, which can reduce the effectiveness of the method.

An important point is that conditioning on survival reduces only the variance associated with knock-out, not the variance of the payoff, which remains conditional on final survival. When the former type of variance is small compared with the latter, this variance reduction method will not prove very effective if applied alone. If there is significant variance conditional on final survival, conditioning on one-step survival may be used in conjuction with other variance reduction methods such as control variates or antithetic variates.

Another situation in which conditioning on survival is effective is when knock-out is disproportionately likely to occur late in the simulation. In this case, standard Monte Carlo simulation will waste a lot of time in computing paths that get knocked out, making it attractive to simulate under a scheme where paths have a higher probability of surviving each step. This situation can arise if a combination of the drift of the process or time dependence of the barrier or volatility make the barrier initially distant from the underlying asset, but very likely to be close after the elapse of some time.

For typical barrier options, conditioning on one-step survival is an effective variance reduction technique when the one-step conditional distribution is known. If it is unknown, the lower the probability of final survival, the more computational effort should be spent at each step in estimating the contribution to the likelihood ratio. However, algorithms which expend too much effort in this direction may underperform standard Monte Carlo simulation.

APPENDIX A. ANALYSIS OF THE ESTIMATORS

Although Lemma 1 is in some respects standard (see, e.g., Glynn and Iglehart 1989 for related results), we detail the proof because the fact that $\mathbf{P}$ and $\widehat{\mathbf{P}}$ do not have common support requires some care.

Proof of Lemma 1. Let ψ be a path up to time i: $(S_1 = s_1, \ldots, S_i = s_i)$, and let Ψ_i be the set of all such paths. The space Ψ_i is the Cartesian product $\prod_{j=1}^{i} \Omega_j$, where each Ω_j, the outcome space for S_j, is a copy of $\Re^d \cup \{\Delta\}$, for some d. The subsets of Ψ_i may be naturally identified with the elements of $\mathcal{F}_i$ by mapping each point $(s_1, \ldots, s_i) \in \Psi_i$ to the set $\{(S_1, \ldots, S_m) \mid S_1 = s_1, \ldots, S_i = s_i\} \subseteq \Omega$. Then we may stretch notation by treating $\mathbf{P}$ and $\widehat{\mathbf{P}}$ as measures on Ψ_i, and using $\mathcal{A}_j$ to refer to the projection of $\mathcal{A}_j \subset \Omega$ onto Ω_j. With this machinery, the proof of the lemma is as follows:

$$\begin{aligned}
\mathbf{E}[A_i Y] &= \int_{\Psi_i} A_i Y \, d\mathbf{P}(\psi) \\
&= \int_{\Omega_1} \cdots \int_{\Omega_i} A_i Y \, d\mathbf{P}(S_i \mid S_{i-1} = s_{i-1}) \cdots d\mathbf{P}(S_1 \mid S_0 = s_0) \\
&= \int_{\mathcal{A}_1} \cdots \int_{\mathcal{A}_i} Y \, d\mathbf{P}(S_i \mid S_{i-1} = s_{i-1}) \cdots d\mathbf{P}(S_1 \mid S_0 = s_0) \\
&= \int_{\mathcal{A}_1} \cdots \int_{\mathcal{A}_i} Y L_i \prod_{j=0}^{i-1} \frac{1}{p(S_j)} d\mathbf{P}(S_i \mid S_{i-1} = s_{i-1}) \\
&\qquad \cdots d\mathbf{P}(S_1 \mid S_0 = s_0) \\
&= \int_{\mathcal{A}_1} \cdots \int_{\mathcal{A}_i} Y L_i \frac{d\mathbf{P}(S_i \mid S_{i-1} = s_{i-1})}{p(S_{i-1})} \cdots \frac{d\mathbf{P}(S_1 \mid S_0 = s_0)}{p(S_0)} \\
&= \int_{\mathcal{A}_1} \cdots \int_{\mathcal{A}_i} Y L_i d\mathbf{P}(S_i \mid S_{i-1} = s_{i-1}, A_i = 1) \\
&\qquad \cdots d\mathbf{P}(S_1 \mid S_0 = s_0, A_1 = 1) \\
&= \int_{\mathcal{A}_1} \cdots \int_{\mathcal{A}_i} Y L_i d\widehat{\mathbf{P}}(S_i \mid S_{i-1} = s_{i-1}) \cdots d\widehat{\mathbf{P}}(S_1 \mid S_0 = s_0)
\end{aligned}$$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

$$= \int_{\Psi_i} A_i Y L_i d\widehat{\mathbf{P}}(\psi)$$
$$= \int_{\Psi_i} Y L_i d\widehat{\mathbf{P}}(\psi) = \widehat{\mathbf{E}}[L_i Y].$$

Thus, L_i is the Radon-Nikodym derivative of the restriction of $\mathbf{P}$ to $\mathscr{A}_m$ (which is a measure, but not a probability measure) with respect to $\widehat{\mathbf{P}}$ relative to $\mathscr{F}_i$. See, e.g., Billingsley (1995, pp. 422–423). As mentioned previously, $\mathbf{P}$ is not absolutely continuous with respect to $\widehat{\mathbf{P}}$ because $\mathbf{P}[A_m = 0] > 0$ while $\widehat{\mathbf{P}}[A_m = 0] = 0$. □

PROOF OF LEMMA 2. By assumption (28), $\mathbf{E}[A_m X_m^2]$ exists and is finite, since $A_m X_m^2 = X_m^2$ because $X_m = A_m f(S_m)$ as defined in (7). As X_m^2 is $\mathscr{F}_m$-measurable, Lemma 1 applies, so

$$\infty > \mathbf{E}[A_m X_m^2] = \widehat{\mathbf{E}}[L_m X_m^2] \geqslant \widehat{\mathbf{E}}[L_m^2 X_m^2] = \widehat{\mathbf{E}}[\widehat{X}_m^2],$$

since $L_m \leqslant 1$. Consequently, $\widehat{\mathbf{E}}[\widehat{X}_m^2]$ is finite. For the second part,

$$\begin{aligned} |\mathbf{E}[A_i X_i]| &= \left| \int A_i X_i d\mathbf{P} \right| \\ &\leqslant \int A_i |X_i| d\mathbf{P} \\ &\leqslant \mathbf{P}[|X_i| \leqslant 1] + \int_{|X_i|>1} A_i |X_i| d\mathbf{P} \\ &\leqslant \mathbf{P}[|X_i| \leqslant 1] + \int_{|X_i|>1} A_i X_i^2 d\mathbf{P} \\ &\leqslant 1 + \mathbf{E}[X_i^2] \leqslant 1 + \mathbf{E}[X_m^2] < \infty, \end{aligned}$$

where $\mathbf{E}[X_i^2] \leqslant \mathbf{E}[X_m^2]$ is justified by Jensen's inequality. □

PROOF OF THEOREM 2. In light of Theorem 1 and linearity of expectation, the conclusion would follow directly if it were proven that

$$\widehat{\mathbf{E}}[L_i(1 - p(S_i))Y_i] = \mathbf{E}[(A_i - A_{i+1})Y_i]$$

where Y_i is an $\mathscr{F}_i$-measurable random variable. First observe that

$$\widehat{\mathbf{E}}[L_i(1 - p(S_i))Y_i] = \mathbf{E}[A_i(1 - p(S_i))Y_i]$$

by Lemma 1, since $(1 - p(S_i))Y_i$ is $\mathscr{F}_i$-measurable. Next, noting that on $\{A_i = 1\}$ we have $p(S_i) = \mathbf{E}[A_{i+1} \mid \mathscr{F}_i]$, we get

$$\begin{aligned} \mathbf{E}[A_i p(S_i) Y_i] &= \mathbf{E}[A_i \mathbf{E}[A_{i+1} \mid \mathscr{F}_i] Y_i] \\ &= \mathbf{E}[\mathbf{E}[A_{i+1} Y_i \mid \mathscr{F}_i]] = \mathbf{E}[A_{i+1} Y_i]. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{E}[A_i(1 - p(S_i))Y_i] &= \mathbf{E}[A_i Y_i] - \mathbf{E}[A_i p(S_i) Y_i] \\ &= \mathbf{E}[A_i Y_i] - \mathbf{E}[A_{i+1} Y_i] \\ &= \mathbf{E}[(A_i - A_{i+1})Y_i] \end{aligned}$$

as desired, proving the result. □

PROOF OF LEMMA 3. First,

$$\mathbf{Var}[X_m] = \mathbf{Var}\left[\sum_{i=1}^{m} (X_i - X_{i-1})\right].$$

Next, show that (for $i \neq j$) $X_i - X_{i-1}$ and $X_j - X_{j-1}$ are uncorrelated, although not necessarily independent. This is so because $X_i = \mathbf{E}[X_{i+1} \mid \mathscr{F}_i]$, so if Y is $\mathscr{F}_{i-1}$-measurable,

$$\begin{aligned} \mathbf{E}[Y(X_i - X_{i-1})] &= \mathbf{E}[YX_i] - \mathbf{E}[Y\mathbf{E}(X_i \mid \mathscr{F}_{i-1})] \\ &= 0 = \mathbf{E}[X_i - X_{i-1}]\mathbf{E}[Y]. \end{aligned}$$

Take $j < i$, then $X_j - X_{j-1}$ is $\mathscr{F}_i$-measurable, so substituting for Y, $X_i - X_{i-1}$, and $X_j - X_{j-1}$ are uncorrelated. Consequently,

$$\mathbf{Var}[X_m] = \sum_{i=1}^{m} \mathbf{Var}[X_i - X_{i-1}].$$

By a standard decomposition of variance,

$$\begin{aligned} \mathbf{Var}[X_i - X_{i-1}] &= \mathbf{E}[\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}]] \\ &\quad + \mathbf{Var}[\mathbf{E}[X_i - X_{i-1} \mid \mathscr{F}_{i-1}]]. \end{aligned}$$

Observe from its definition that X_i is a martingale (see, e.g., Karlin and Taylor 1975, p. 246). It follows that the second term in this variance decomposition is zero. Therefore,

$$\mathbf{Var}[X_m] = \sum_{i=1}^{m} \mathbf{E}[\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}]]. \qquad \square$$

PROOF OF LEMMA 4. First write, using the standard variance decomposition again,

$$\begin{aligned} \mathbf{Var}[X_i \mid \mathscr{F}_{i-1}] &= \mathbf{E}[\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}, A_i] \mid \mathscr{F}_{i-1}] \\ &\quad + \mathbf{Var}[\mathbf{E}[X_i \mid \mathscr{F}_{i-1}, A_i] \mid \mathscr{F}_{i-1}]. \end{aligned}$$

The conditional expectation and variance of X_i given $A_i = 0$ are both zero. Therefore, given $\mathscr{F}_{i-1}$,

$$\mathbf{E}[X_i \mid \mathscr{F}_{i-1}, A_i] = A_i \mathbf{E}[X_i \mid \mathscr{F}_{i-1}, A_i = 1],$$

and the distribution of A_i conditioned on $\mathscr{F}_{i-1}$ is Bernoulli with parameter $p(S_{i-1})$. Consequently,

$$\begin{aligned} &\mathbf{Var}[\mathbf{E}[X_i \mid \mathscr{F}_{i-1}, A_i] \mid \mathscr{F}_{i-1}] \\ &\quad = p(S_{i-1})(1 - p(S_{i-1}))(\mathbf{E}[X_i \mid \mathscr{F}_{i-1}, A_i = 1])^2. \end{aligned}$$

Likewise, given $\mathscr{F}_{i-1}$,

$$\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}, A_i] = A_i \mathbf{Var}[X_i \mid \mathscr{F}_{i-1}, A_i = 1]$$

and

$$\begin{aligned} &\mathbf{E}[\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}, A_i] \mid \mathscr{F}_{i-1}] \\ &\quad = p(S_{i-1})\mathbf{Var}[X_i \mid \mathscr{F}_{i-1}, A_i = 1]. \end{aligned}$$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

The expectation of the one-step conditional variance is, since X_i is a martingale,

$$\begin{aligned}\mathbf{E}[\mathbf{Var}[X_i \mid \mathcal{F}_{i-1}]] &= \mathbf{E}[\mathbf{E}[X_i^2 \mid \mathcal{F}_{i-1}] - (\mathbf{E}[X_i \mid \mathcal{F}_{i-1}])^2] \\ &= \mathbf{E}[\mathbf{E}[X_i^2 \mid \mathcal{F}_{i-1}]] - \mathbf{E}[X_{i-1}^2] \\ &= \mathbf{E}[X_i^2] - \mathbf{E}[X_{i-1}^2]. \qquad \square\end{aligned}$$

Proof of Theorem 3. Variance reduction follows from a bound on the likelihood ratio, a property frequently exploited in analyzing importance sampling methods:

$$\begin{aligned}\widehat{\mathbf{Var}}[\widehat{X}_m] &= \widehat{\mathbf{E}}[\widehat{X}_m^2] - (\widehat{\mathbf{E}}[\widehat{X}_m])^2 \\ &= \widehat{\mathbf{E}}[L_m^2 X_m^2] - X_0^2 \\ &= \mathbf{E}[L_m X_m^2] - X_0^2 \\ &\leqslant \mathbf{E}[X_m^2] - X_0^2 \\ &= \mathbf{Var}[X_m],\end{aligned}$$

where the inequality follows because $L_m \leqslant 1$, and $L_m < 1$ if there is any chance of knock-out. $X_0 = \mathbf{E}[X_m] = \widehat{\mathbf{E}}[\widehat{X}_m]$ is the nonrandom (but unknown) price, the unconditional expectation of these unbiased estimators (see Theorem 1). Similarly, the expected one-step conditional variance is:

$$\begin{aligned}\widehat{\mathbf{E}}[\widehat{\mathbf{Var}}[\widehat{X}_i \mid \mathcal{F}_{i-1}]] &= \widehat{\mathbf{E}}[\widehat{\mathbf{E}}[(\widehat{X}_i - \mathbf{E}[\widehat{X}_i \mid \mathcal{F}_{i-1}])^2 \mid \mathcal{F}_{i-1}]] \\ &= \widehat{\mathbf{E}}[\widehat{\mathbf{E}}[(\widehat{X}_i - \widehat{X}_{i-1})^2 \mid \mathcal{F}_{i-1}]] \\ &= \widehat{\mathbf{E}}[\widehat{\mathbf{E}}[\widehat{X}_i^2] - 2\widehat{X}_{i-1}\widehat{\mathbf{E}}[\widehat{X}_i \mid \mathcal{F}_{i-1}] + \widehat{X}_{i-1}^2] \\ &= \widehat{\mathbf{E}}[\widehat{X}_i^2] - \widehat{\mathbf{E}}[\widehat{X}_{i-1}^2] \\ &= \widehat{\mathbf{E}}[L_i^2 X_i^2] - \widehat{\mathbf{E}}[L_{i-1}^2 X_{i-1}^2] \\ &= \mathbf{E}[L_i X_i^2] - \mathbf{E}[L_{i-1} X_{i-1}^2],\end{aligned}$$

where the second equality holds because X_i is a martingale. $\square$

Proof of Theorems 4 and 5. Theorem 4 is an immediate consequence of the more general result

$$\mathbf{E}^n[A_i Y L_i^n] = \mathbf{E}[A_i Y],$$

where Y is $\mathcal{F}_i$-measurable. This result is very similar to Lemma 1, and the proof is exactly parallel. Here the crucial observation is

$$\begin{aligned}&\int_{\mathcal{A}_j} \bullet \left(\frac{1-(1-p(S_{j-1}))^n}{p(S_{j-1})}\right) d\mathbf{P}(S_j \mid S_{j-1} = s_{j-1}) \\ &\qquad = \int_{\mathcal{A}_j} \bullet\, d\mathbf{P}^n(S_j \mid S_{j-1} = s_{j-1}),\end{aligned}$$

because $\mathbf{P}^n[A_j = 1 \mid S_{j-1}] = 1-(1-p(S_{j-1}))^n$, while $\mathbf{P}[A_j = 1 \mid S_{j-1}] = p(S_{j-1})$. The proof of Theorem 5 is exactly the same as that of Theorem 3, with the measure $\mathbf{P}^n$ substituted for $\widehat{\mathbf{P}}$. $\square$

Proof of Theorem 6. The proof is by induction on m. The crucial observation is that the waiting time Y_i and the successor S_{i+1} are independent given $\mathcal{F}_i$. Then for $m = 1$, $f(S_1)$ and Y_0 are independent, so

$$\begin{aligned}\widehat{\mathbf{E}}\left[f(S_1)\frac{r-1}{Y_0-1}\right] &= \widehat{\mathbf{E}}[f(S_1)]\widehat{\mathbf{E}}\left[\frac{r-1}{Y_0-1}\right] \\ &= \widehat{\mathbf{E}}[f(S_1)]p(S_0) = \widehat{\mathbf{E}}[f(S_1)L_1]\end{aligned}$$

as desired. Recall that where Y_i has negative binomial distribution with parameters $(r, p(S_i))$, the estimator $(r-1)/(Y_i-1)$ is unbiased. Next assume that the property holds for some m, and prove it holds for $m+1$:

$$\begin{aligned}&\widehat{\mathbf{E}}\left[f(S_{m+1})\prod_{i=0}^{m}\frac{r-1}{Y_i-1}\right] \\ &= \widehat{\mathbf{E}}\left[\prod_{i=0}^{m-1}\frac{r-1}{Y_i-1} f(S_{m+1})\frac{r-1}{Y_m-1}\right] \\ &= \widehat{\mathbf{E}}\left[\widehat{\mathbf{E}}\left[\prod_{i=0}^{m-1}\frac{r-1}{Y_i-1} f(S_{m+1})\frac{r-1}{Y_m-1} \mid \mathcal{F}_m\right]\right] \\ &= \widehat{\mathbf{E}}\left[\prod_{i=0}^{m-1}\frac{r-1}{Y_i-1}\widehat{\mathbf{E}}[f(S_{m+1}) \mid \mathcal{F}_m]\widehat{\mathbf{E}}\left[\frac{r-1}{Y_m-1} \mid \mathcal{F}_m\right]\right] \\ &= \widehat{\mathbf{E}}[L_m \widehat{\mathbf{E}}[f(S_{m+1}) \mid \mathcal{F}_m]p(S_m)] \\ &= \widehat{\mathbf{E}}[\widehat{\mathbf{E}}[L_{m+1} f(S_{m+1}) \mid \mathcal{F}_m]] \\ &= \widehat{\mathbf{E}}[L_{m+1} f(S_{m+1})].\end{aligned}$$

The fourth equality follows from the inductive assumption, because $\widehat{\mathbf{E}}[f(S_{m+1}) \mid \mathcal{F}_m]\widehat{\mathbf{E}}[\frac{r-1}{Y_m-1} \mid \mathcal{F}_m]$ is just an $\mathcal{F}_m$-measurable random variable. $\square$

Proof of Theorem 7. The proof is very similar to that of Theorem 6, proceeding by induction on m. For $m = 1$,

$$\begin{aligned}\mathbf{E}^n\left[f(S_1)\frac{N_0}{n}\right] &= \mathbf{E}^n(f(S_1))\mathbf{E}^n\left[\frac{N_0}{n}\right] \\ &= \mathbf{E}^n[f(S_1)]p(S_0) = \mathbf{E}^n[f(S_1)L_1],\end{aligned}$$

because N_i/n is unbiased for $p(S_i)$. The inductive step is justified by:

$$\begin{aligned}\mathbf{E}^n\left[f(S_{m+1})\prod_{i=0}^{m}\frac{N_i}{n}\right] &= \mathbf{E}^n\left[\mathbf{E}^n\left[\prod_{i=0}^{m-1}\frac{N_i}{n} f(S_{m+1})\frac{N_m}{n} \mid \mathcal{F}_m\right]\right] \\ &= \mathbf{E}^n\left[\prod_{i=0}^{m-1}\frac{N_i}{n}\mathbf{E}^n(f(S_{m+1}) \mid \mathcal{F}_m) \right.\\ &\qquad \left. \times \mathbf{E}^n\left[\frac{N_m}{n} \mid \mathcal{F}_m\right]\right].\end{aligned}$$

Next, note that

$$\begin{aligned}\mathbf{E}^n\left[\frac{N_m}{n} \mid \mathcal{F}_m\right] &= \mathbf{E}^n\left[A_{m+1}\frac{N_m}{n} \mid \mathcal{F}_m\right] \\ &= \mathbf{E}^n\left[A_{m+1}\mathbf{E}^n\left[\frac{N_m}{n} \mid \mathcal{F}_m, A_{m+1} = 1\right] \mid \mathcal{F}_m\right] \\ &= \mathbf{E}^n\left[A_{m+1}\frac{p(S_m)}{1-(1-p(S_m))^n} \mid \mathcal{F}_m\right],\end{aligned}$$

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms

because

$$A_{m+1} = 0 \iff N_m = 0,$$
$$\mathbf{P}^n[A_{m+1} = 1 \mid \mathcal{F}_m] = 1 - (1 - p(S_m))^n,$$
$$\mathbf{E}^n\left[\frac{N_m}{n} \mid \mathcal{F}_m, A_{m+1} = 0\right] = 0.$$

Returning to the main argument,

$$\begin{aligned}
&\mathbf{E}^n\left[f(S_{m+1})\prod_{i=0}^{m}\frac{N_i}{n}\right] \\
&= \mathbf{E}^n\left[A_m L_m^n \mathbf{E}^n\left[f(S_{m+1})A_{m+1}\frac{p(S_m)}{1-(1-p(S_m))^n} \mid \mathcal{F}_m\right]\right] \\
&= \mathbf{E}^n[\mathbf{E}^n[A_{m+1}f(S_{m+1})L_{m+1}^n \mid \mathcal{F}_m]] \\
&= \mathbf{E}^n[A_{m+1}f(S_{m+1})L_{m+1}^n]. \qquad \square
\end{aligned}$$

ACKNOWLEDGMENT

The authors thank the referees for many constructive comments. This research is supported by NSF grant DMI9457189, a grant from the IBM Corporation, and an NSF graduate research fellowship.

REFERENCES

Andersen, L., R. Brotherton-Ratcliffe. 1996. Exact exotics. *Risk* **9** 85–89.

Asmussen, S. 1987. *Applied Probability and Queues.* Wiley, Chichester, U.K.

——, P. Glynn, J. Pitman. 1995. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. App. Probab.* **5** 875–896.

Baldi, P., L. Caramellino, M. Iovino. 1999. Pricing general barrier options: A numerical approach using sharp large deviations. *Math. Finance* **9** 239–321.

Beaglehole, D., P. Dybvig, G. Zhou. 1997. Going to extremes: Correcting simulation bias in exotic option valuation. *Financial Analysts J.* **53** 62–68.

Billingsley, P. 1995. *Probability and Measure*, 3rd ed. Wiley-Interscience, New York.

Boyle, P., M. Broadie, P. Glasserman. 1997. Monte Carlo methods for security pricing. *J. Econom. Dynam. Control* **21** 1267–1321.

Bratley, P., B. L. Fox, L. Schrage. 1987. *A Guide to Simulation*, 2nd ed. Springer-Verlag, New York.

Cox, D., D. Hinkley. 1974. *Theoretical Statistics.* Chapman and Hall, New York.

Duan, J.-C., J.-G. Simonato. 1998. Empirical martingale simulation for asset prices. *Management Sci.* **9** 1218–1233.

Duffie, D. 1996. *Dynamic Asset Pricing Theory*, 2nd ed. Princeton University Press, Princeton, NJ.

Fishman, G. 1996. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer-Verlag, New York.

Glasserman, P. 1993. Filtered Monte Carlo. *Math. Oper. Res.* **18** 610–634.

——, P. Heidelberger, P. Shahabuddin. 1999a. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* **9** 117–152.

——, ——, ——, T. Zajic. 1999b. Multilevel splitting for estimating rare event probabilities. *Oper. Res.* **47** 585–600.

Glynn, P., D. Iglehart. 1988. Simulation methods for queues: An overview. *Queueing Sys.* **3** 221–256.

——, ——. 1989. Importance sampling for stochastic simulations. *Management Sci.* **35** 1367–1392.

Heynen, R., H. Kat. 1994. Crossing barriers. *Risk* **8** 46–49.

Hull, J. 1993. *Options, Futures, and Other Derivative Securities*, 2nd ed. Prentice-Hall, Englewood Cliffs, NJ.

Joy, C., P. Boyle, K. Tan. 1996. Quasi-Monte Carlo methods in numerical finance. *Management Sci.* **42** 926–938.

Karatzas, I., S. Shreve. 1991. *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer-Verlag, New York.

Karlin, S., H. Taylor. 1975. *A First Course in Stochastic Processes.* Academic Press, New York.

Kunitomo, N., M. Ikeda. 1992. Pricing options with curved boundaries. *Math. Finance* **2** 275–298.

Lemieux, C., P. L'Ecuyer. 1998. Efficiency improvement by lattice rules for pricing Asian options. *Proc. 1998 Winter Simulation Conf.* IEEE Press, 579–586.

Marsaglia, G., A. Zaman, J. Marsaglia. 1994. Rapid evaluation of the inverse of the normal distribution function. *Statist. Probab. Lett.* **19** 259–266.

Merton, R. 1973. Theory of rational option pricing. *Bell J. Econom. Management Sci.* **4** 141–183.

Musiela, M., M. Rutkowski. 1997. *Martingale Methods in Financial Modelling.* Springer-Verlag, New York.

Ross, S., J. G. Shanthikumar. 1999. Pricing exotic options: Monotonicity in volatility and efficient simulation. Working paper, IEOR Department, University of California, Berkeley, CA.

Rubinstein, M., E. Reiner. 1991. Breaking down the barriers. *Risk* **4** 28–35.

Sidenius, J. 1998. Double barrier options: Valuation by path counting. *J. Comput. Finance* **1** 63–79.

Staum, J. 2001. Ph. D. Dissertation, Graduate School of Business, Columbia University, New York.

Vázquez-Abad, F., D. Dufresne. 1998. Accelerated simulation for pricing asian options. *Proc. 1998 Winter Simulation Conf.* IEEE Press, 1493–1500.

Villén-Altamirano, M., J. Villén-Altamirano. 1994. RESTART: A straightforward method for fast simulation of rare events. *Proc. 1994 Winter Simulation Conf.* IEEE Press, 282–289.

This content downloaded from 62.99.145.238 on Sat, 20 Apr 2019 15:28:38 UTC
All use subject to https://about.jstor.org/terms