Pricing American Derivatives using Simulation:

A Biased-Low Approach

Phelim P. Boyle
Adam W. Kolkiewicz
Ken Seng Tan

Department of Statistics and Actuarial Science,
University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada
pboyle@uwaterloo.ca
wakolkiewicz@uwaterloo.ca
kstan@uwaterloo.ca

Abstract

In Boyle et al. (2000) we propose a simulation method for pricing high-dimensional American style derivatives. The method exploits the uniformity property of the low discrepancy sequences so that the resulting biased high estimator can achieve higher rate of convergence of quasi-Monte Carlo method. In this paper, we extend this work by proposing another estimator that is biased low. It has the computational advantage that it can be obtained concurrently with the high-biased estimator using a recursive valuation approach. Some numerical examples are conducted to demonstrate its efficiency. We also show that further enhancement to the proposed estimator is possible by incorporating the standard variance reduction technique such as the use of control variates.

Acknowledgments: Phelim P. Boyle thanks the Social Science and Humanities Research Council of Canada for research support. Adam Kolkiewicz and Ken Seng Tan acknowledge research support from the Natural Sciences and Engineering Research Council of Canada. Phelim P. Boyle and Ken Seng Tan are also grateful to the Committee for Knowledge Extension and Research of the Society of Actuaries for financial support.

1 Introduction

Monte Carlo simulation is an important computational tool for pricing complex derivatives and valuing real options. It widely used in modern risk management since it is well suited for dealing with the large number of variables that are required to analyze the market risk and credit risk of large portfolios. The two most important limitations of the Monte Carlo method are:

- It is sometimes slow.
- The pricing of American options is difficult for high dimensional problems.

In recent years progress has been made along both these fronts. To speed up the method, a wide variety of variance reduction techniques have been proposed. In particular by using specially selected deterministic points, known as quasi-random points, instead of the usual random points, the efficiency of the method can be improved (see e.g. Joy et al. (1996) and Tan and Boyle (2000)). There has also been progress in the application of simulation methods to the valuation of American style derivatives. However even the best methods available are still not very efficient for high dimensional problems and this is exactly where the alternative approaches such as finite difference methods break down.

There are two reasons why the valuation of American style options by simulation holds such fascination. First it is a hard technical problem. Indeed until the publication of Tilley's 1993 paper it was believed[1] that American options could not be valued by Monte Carlo simulation. Since then, this problem has attracted considerable academic attention[2] and despite significant progress has not been adequately solved for high dimensional problems. Second, it is a problem of considerable practical interest. The Monte Carlo method is really the only viable numerical technique for high dimensional problems and such problems are becoming more prevalent.

The difficulty in applying the Monte Carlo method to value an American style contract stems from its early exercise feature. At each point the holder of an American option has to decide whether to exercise the contract or continue to hold on to it. A rational investor will select an optimal exercise strategy that will maximize the value of the contract. This problem can be set up in a dynamic programming framework where we can solve the

optimization problem by working backwards through time. However in the standard Monte Carlo method we generate future price paths of the asset or assets in question starting from the current time. This forward marching approach clashes directly with the backward recursion feature of dynamic programming. The decision to exercise depends on the topography of the early exercise boundary. The standard Monte Carlo method that we use for European options does not furnish this exercise boundary.

A variety of approaches have been used to value the early exercise feature by simulation. Typically they involve some technique for approximating the early exercise boundary or approximating the transitional density function. It is often difficult to disentangle the errors induced by these approximations from the statistical error associated with the Monte Carlo approach. It is not easy to come up with convergence proofs for many of these methods. A common approach is to show that the proposed simulation method produces estimates which converge for large numbers of trials to the results produced by some other numerical method for a range of examples. Broadie and Glasserman have made important contributions to this problem. They proved (1997a) that for a large class of problems the simulation estimator is biased. They also developed a simulated tree approach which produces two estimators: one biased high and the other biased low. These estimators converge asymptotically and furnish confidence intervals for the true value. Unfortunately, this algorithm becomes computational burdensome when the number of exercise points or underlying assets is large. Subsequently, they propose a stochastic mesh method (1997b) which could handle much higher dimensions. The rate of convergence for this method, however, is very slow and significant variance reduction techniques need to be incorporated in order for this method to be practically useful.

As inspired by the stochastic mesh method, Boyle et al. (2000) showed that with an appropriate choice of mesh density, the method can be combined with the quasi-Monte Carlo technique to achieve a significant bias reduction of the high-biased estimator. In this paper we extend the work of Boyle et al. (2000) by studying properties of a low-biased estimator. The motivation for considering these two estimators jointly is so that the correct value of the problem of interest is bounded by these two estimates. Previously, this method has been considered by Avramidis and Hyden (1999) in the context of a stochastic mesh only.

We also investigate two practical issues not resolved in earlier studies: the first one is related to the construction of the biased-low estimator and the sec-

ond to methods of combining high-biased and low-biased estimates together. In the last section, we present some simulation study where we demonstrate their efficiencies. Furthermore, we show that additional enhancement can be achieved by using some standard variance reduction techniques, such as the use of control variates.

2 Low-Biased and Point Estimators

2.1 Introduction and Motivation

We are interested in pricing a derivative security whose payoff depends on s underlying asset prices $\{\vec{S}_t\} = \{(S_t^1, S_t^2, \ldots, S_t^s)\}$. The security can be exercised prior to maturity at $d+1$ time points (including the initial time), which we will denote as $t_0, t_1, \ldots, t_d = T$. If the security is exercised at time τ, its value is equal to $I(\tau, \vec{S}_\tau)$, where $I(\cdot)$ is a known function, which represents the discounted value of the contract.

For now we do not make any special assumptions about the dynamics of the price process $\{\vec{S}_t\}$ except that under the risk-neutral measure Q it is a Markov process with a fixed initial state, whose transition probability densities $f(t_i, t_j, \vec{x}; \cdot)$, defined by

$$P(\vec{S}_{t_j} \in A | \vec{S}_{t_i} = \vec{x}) = \int_A f(t_i, t_j, \vec{x}; \vec{u}) d\vec{u},$$

exist and are known. We denote the time-zero prices by $\vec{S}_0$ and for brevity the transition density functions $f(0, t_i, \vec{S}_0; \cdot)$ by $f_o(t_i; \cdot)$ and $f(t_i, t_{i+1}, \vec{x}; \cdot)$ by $f_i(\vec{x}; \cdot)$.

The valuation of the American derivative can be formulated as the maximization of its expected exercise value taken over all stopping times (see Duffie (1996)). Thus we are interested in finding V such that

$$V := \max_\tau E[I(\tau, \vec{S}_\tau)],$$

where τ is a stopping time taking values in the set $\{t_0, t_1, ..., T\}$. We can solve this problem using the principle of dynamic programming: We find the functions $V(t_i, \cdot)$ through the backward recursion:

$$\begin{aligned} V(T, \vec{x}) &= I(T, \vec{x}), \\ V(t_i, \vec{x}) &= \max[I(t_i, \vec{x}),\ C(t_i, \vec{x})], \quad i = d-1, ..., 0, \end{aligned}$$

where $C(t_i, \vec{x})$ is the continuation value at point $\vec{x}$

$$C(t_i, \vec{x}) = E[V(t_{i+1}, \vec{S}_{t_{i+1}})|\vec{S}_{t_i} = \vec{x}]), \tag{1}$$

and finally set $V = V(0, \vec{S}_0)$.

To use this method in practice we must be able to calculate or approximate efficiently the continuations values $C(t_i, \vec{x})$ for all $i = 0, ..., d-1$ and some selected set of points $\vec{x}$ from the state space. For $t = t_0$ this can be accomplished using the Monte Carlo method:

$$E[V(t_1, \vec{S}_{t_1})|\vec{S}_{t_0}] \simeq \frac{1}{n}\sum_{j=1}^{n} \hat{V}(t_1, \vec{X}_{t_1}(j)),$$

where $\hat{V}(t_1, \cdot)$ is an approximation of $V(t_1, \cdot)$ obtained from the backward recursion and $\vec{X}_{t_1}(1), ..., \vec{X}_{t_1}(n)$ is a random sample drawn from a distribution with density function $f_0(t_1; \cdot)$. We can repeat the same procedure at $t = t_1$, now, however, instead of one integral we have to evaluate n integrals. Proceeding as before, we generate n random points for each conditional expectation $E[\hat{V}(t_2, \vec{S}_{t_2})|\vec{S}_{t_1} = \vec{X}_{t_1}(j)],\ j = 1, ..., n$. We could, in principle, proceed in a similar way for the remaining time points t. In practice, however, this method will quickly become unworkable because the number of points will grow exponentially.

The key to the stochastic mesh method (Broadie and Glasserman, 1997b) and low discrepancy mesh (LDM) method (Boyle et. al., 2000) is the observation that points that we generate to calculate one conditional expectation can also be used to calculate other expectations. By doing this, we may reduce the rate at which the number of points growths, or even keep it constant. In both methods, first we generate a mesh of state points $\{\vec{X}_{t_i}(j), i = 1, \ldots, d, j = 1, \ldots, n\}$, from $\mathcal{R}^s$. In the stochastic mesh method this is achieved by generating random points from a distribution specified by a certain density function g_{t_i}, which is referred to as a mesh density. Similar methods are used for the low discrepancy mesh method except that the mesh points are generated by applying the inverse method to a low discrepancy sequence. Once a mesh is constructed, the values $V(t_i, \cdot)$ are calculated using the backward recursion method, where the continuation values are approximated by

$$\hat{C}(t_i, \vec{X}_{t_i}(j)) = \sum_{l=1}^{n} \hat{V}(t_{i+1}, \vec{X}_{t_{i+1}}(l)) w(t_i, \vec{X}_{t_i}(j), \vec{X}_{t_{i+1}}(l)),$$

for $j = 1, .., n,\ i = 0, 1, \ldots, d$, where

$$w(t_i, \vec{X}_{t_i}(j), \vec{X}_{t_{i+1}}(l)) := \frac{f_i(\vec{X}_{t_i}(j); \vec{X}_{t_{i+1}}(l))}{g_{t_i}(\vec{X}_{t_{i+1}}(l))}$$

is the Radon-Nikodym derivative. We denote the resulting estimate $\hat{V}(0, \vec{S}_0)$ as $\hat{V}_H$. The introduction of the weights w is necessary as the points in the mesh are sampled from the density g_{t_i}. This procedure, applicable to both stochastic and deterministic mesh points, generates an estimate that is biased high.

A good choice of the densities g_{t_i}, $i = 1, \ldots, d$, is crucial for the success of the mesh method. For random sampling, Broadie and Glasserman (1997b) suggest to use the average density

$$g_M(t_i, \vec{x}; \vec{u}) := \frac{1}{n} \sum_{i=1}^{n} f_i(\vec{x}; \vec{u}), \tag{2}$$

which can be interpreted as a mixture distribution of transition densities with the initial points equally likely to be selected.

The authors justify this selection by demonstrating that for European style contracts other choices may lead to a build-up of the error. For American options, a more formal argument in favor of this distribution is presented by Boyle et al. (2000). In the same paper it has been also shown that a simpler choice of the mesh density g equal to the marginal distribution $f_o(t_i; \cdot)$ will give a very similar rate of convergence. Results presented by Boyle et al. (2000) indicate that an application of low discrepancy sequences can improve significantly the accuracy of the mesh method. In this paper we consider a method that yields biased low estimates.

2.2 Low-Biased Estimator

Let us partition the set of indices $\mathcal{I} = \{1, \ldots, n\}$ into two subsets, $\mathcal{A}$ and its complement $\mathcal{A}^c = \mathcal{I} - \mathcal{A}$, and denote by $|\cdot|$ the number of elements in a set. A biased low estimator is defined in a recursive way similar to the high-biased $\hat{V}_H$ but at each time step we use the following formula:

$$\hat{V}_L(t_i, \vec{x}) = \begin{cases} I(t_i, \vec{x}) & \text{if } I(t_i, \vec{x}) \geq C_{\mathcal{A}}(t_i, \vec{x}) \\ C_{\mathcal{A}^c}(t_i, \vec{x}) & \text{otherwise,} \end{cases} \tag{3}$$

where for an arbitrary subset of $\mathcal{I}$, say $\mathcal{B}$, we define

$$C_{\mathcal{B}}(t_i, \vec{x}) = \frac{1}{|\mathcal{B}|} \sum_{l \in \mathcal{B}} \hat{V}_L(t_{i+1}, \vec{X}_{t_{i+1}}(l)) w(t_i, \vec{x}, \vec{X}_{t_{i+1}}(l)). \tag{4}$$

The above estimator is considered by Avramidis and Hyden (1999), who in the case when g_{t_i} is equal to the average density (2) state that it is a biased low estimator of V. Since for low discrepancy sequences we are also interested in using the marginal distributions as mesh densities, we need a result that would cover this case.

Proposition 1 *Suppose that the mesh points are generated using* $g_{t_i}(\cdot) = f_o(t_i, \cdot)$, $i = 0, \ldots, d$ *and that the payoff function* $I(\cdot)$ *is bounded. Then for any positive number* α *there exist* N *such that for* $n > N$ *the estimator* $\hat{V}_L$, *as defined above, satisfies*

$$E[\hat{V}_L(t_i, \vec{x})] \leq V(t_i, \vec{x}) + \alpha$$

for all $\vec{x} \in \mathcal{R}^s$ *and* $i = 0, \ldots, d$. *In the case when* $I(0, \vec{S}_0) \neq C(0, \vec{S}_0)$, *the number* α *can be taken equal to zero.*

This result can be extended to the case when the mesh densities g_{t_i} depend on the previously generated points in the mesh, as considered by Boyle et al. (2000). However, for low discrepancy sequences there is very little difference between mesh methods based on such a density and the marginal densities f_o. The assumption of the payoff function being bounded is a technical one and can be relaxed at a cost of adding more complexity to the proof.

Proposition 1 will hold for any selection of the subset $\mathcal{A}$ and we need an additional criterion to determine a desirable partition. In previous studies, two different partitions of $\mathcal{I}$ have been used. Avramidis and Hyden (1999) propose to include in the set $\mathcal{A}$ only one element but then they take an average of the resulting estimators over all possible choices of this element. Avramidis et al. (2000) in their studies use equal number of points in $\mathcal{A}$ and $\mathcal{A}^c$. In both cases no formal justification has been provided, however.

To get more insight into the problem, now we consider the variance of the low-biased estimators as a function of the number of points in $\mathcal{A}$. To simplify the exposition, we present an analysis assuming that at the next time step, t_{i+1}, the value of the contract is known.

Let us fix time step, t_i, and a point in the state space, $\vec{x}$, and introduce the following notation:

$$\begin{aligned}
\mu(\vec{x}) &:= E[C_{\mathcal{A}^c}(t_i,\vec{x})] = E[C_{\mathcal{A}}(t_i,\vec{x})] \\
p_L(m) &:= P\{C_{\mathcal{A}}(t_i,\vec{x}) \le I(t_i,\vec{x})\}, \quad q_L(m) := 1 - p_L(m), \\
\sigma^2(t_i,\vec{x}) &:= Var[V(t_{i+1},\vec{U})w(t_i,\vec{x},\vec{U})],
\end{aligned}$$

where $\vec{U}$ is a random variable with the density function g_{t_i}, and $m = |\mathcal{A}|$. In the expression for $\sigma^2(t_i,\vec{x})$ we have used V to calculate the variance. For the analysis to be valid this is not an essential assumption; what really matters is that V is non-random.

Using basic probability rules, it is easy to find the variance of $\hat{V}_L$:

$$\begin{aligned}
Var[\hat{V}_L(t_i,\vec{x})] &= q_L(m)Var[C_{\mathcal{A}^c}] + p_L(m)q_L(m)[I(t_i,\vec{x}) - \mu(\vec{x})]^2 \quad (5) \\
&= q_L(m)\frac{\sigma^2(t_i,\vec{x})}{n-m} + p_L(m)q_L(m)[I(t_i,\vec{x}) - \mu(\vec{x})]^2. \quad (6)
\end{aligned}$$

The last expression suggests ways of partitioning $\mathcal{I}$ with the objective of reducing the variance. If $\vec{x}$ belongs to the continuation region, $\{\vec{x} : \ I(t_i,\vec{x}) < C(t_i,\vec{x})\}$, the probability $p_L(m)$ will be small for large values of m. This will reduce the size of the second component in the variance but for a fixed n will increase the first term, which corresponds to the variance of estimation of the continuation value. This suggests that for $\vec{x}$ from the continuation region we should increase the number of points in $\mathcal{A}^c$. On the other hand, in the exercise region with $I(t_i,\vec{x}) \ge C(t_i,\vec{x})$, we can reduce the variance by increasing the accuracy of estimation $p_L(m)$ which suggests increasing the number of points in $\mathcal{A}$.

The above analysis is consistent with intuition and can be useful for an adaptive approach, when a partition of the index set $\mathcal{I}$ is allowed to depend on the state point. For a simpler method where n and m are fixed in advance, we may select a good partition by a closer examination of the above formula. For this purpose, we can use the Central Limit Theorem to find the following approximation to $p_L(m)$:

$$p_L(m) \cong \Phi(\sqrt{m}M(t_i,\vec{x})), \quad (7)$$

where Φ is the standard normal cumulative distribution function and $M \equiv M(t_i,\vec{x}) := [I(t_i,\vec{x}) - \mu(\vec{x})]/\sigma(t_i,\vec{x})$. From this we can get the following

approximation to the variance as a function of the fraction of points, δ, from $\mathcal{I}$ included in set $\mathcal{A}$:

$$\begin{aligned} W(\delta) &:= Var[\hat{V}_L(t_i, \vec{x})](\delta) \\ &\propto (1 - \Phi(M\sqrt{\delta n})) \left[\frac{1}{(1-\delta)n} + M^2 \Phi(M\sqrt{\delta n}) \right]. \end{aligned}$$

In this formula, the variance depends on the point $\vec{x}$ only through a scalar parameter M. Otherwise it is quite general, as it does not depend on the payoff function I nor the dimension of the state space. Fig. 1 depicts typical shapes of this function: in the middle of the range of possible values of its argument, it is almost flat and close to the minimal value. This suggests that any partitioning that corresponds to δ near 0.5 should be close to the optimal one, regardless of M. This has been verified numerically by constructing a dense set of points for δ and M.

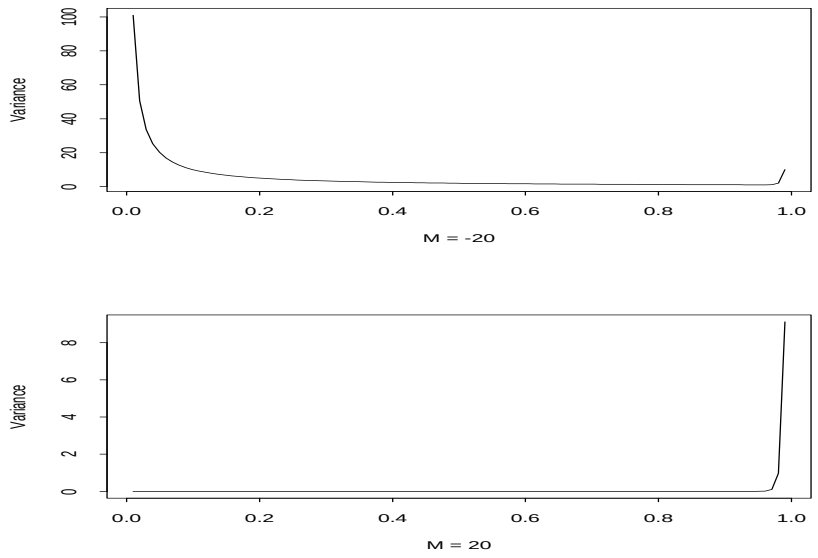

Figure 1: Typical shapes of the variance as a function of δ.

2.3 Point Estimator

An advantage of using the low-biased estimator described in previous section is that it can be constructed using the same mesh points as the high-biased

estimator. An obvious question now is how these two can be combined to define a good point estimate. Below we provide a discussion of a possible solution of this issue and return to it in the next section where we present numerical examples.

As before, we analyze the problem assuming that the value of the contract at the next time step, say t_{i+1}, is known. Following our findings about optimal choices of the partition of $\mathcal{I}$, here we assume that the number of points n is even and that each subset of $\mathcal{I}$ contains the same number of points. For ease of the exposition, let us rewrite the definitions of the biased high and biased low estimators, respectively, as follows:

$$\hat{V}_H = \begin{cases} I(t_i, \vec{x}) & \text{if } I(t_i, \vec{x}) \geq C \\ C & \text{otherwise,} \end{cases}$$

and

$$\hat{V}_L = \begin{cases} I(t_i, \vec{x}) & \text{if } I(t_i, \vec{x}) \geq C_1 \\ C_2 & \text{otherwise,} \end{cases}$$

where C, C_1, and C_2 are three estimators of the continuation value at mesh point $\vec{x}$, defined from formula (4). The estimator C encompasses the entire set $\mathcal{I} = \{1, \ldots, n\}$ while both C_1 and C_2 correspond to equal partitioning of the set $\mathcal{I}$ with the points being randomly chosen.

Our objective is to find a point estimator, say $\hat{V}_P$, of the value of the contract $V_{\vec{x}} \equiv V(t_i, \vec{x})$ as a linear combination of $\hat{V}_H$ and $\hat{V}_L$:

$$\hat{V}_P = \gamma \hat{V}_H + (1 - \gamma)\hat{V}_L, \quad \text{for } \gamma \in [0, 1]. \tag{8}$$

We are interested in finding γ such that V_P has good statistical properties. As a criterion, we shall use the mean square error of V_P, and for this we need formulas for the variances of $\hat{V}_H$ and $\hat{V}_L$. As the variance for the former is given by (6), here we only need to consider the variance of the later. Let us denote by $\aleph_A$ the indicator function of a set A. Using conditioning argument we have

$$Var[\hat{V}_H] = I^2 p_H(n) q_H(n) + Var[C\aleph_{\{C>I\}}] - 2I p_H(n) E[C\aleph_{\{C>I\}}],$$

where $I \equiv I(t_i, \vec{x})$, $p_H(n) = P\{C \leq I\}$ and $q_H(n) = 1 - p_H(n)$. From this, it can be shown that there exists a constant, say D, such that

$$Var[\hat{V}_H] \leq D \cdot I^2 p_H(n) q_H(n) + Var[C]. \tag{9}$$

To identify the leading term in the right-hand-side of this inequality, we use (7) to find that $p_H(n) = \Phi(M\sqrt{n})$, where as before, $M = (I(t_i, \vec{x}) - \mu(\vec{x}))/\sigma(t_i, \vec{x})$. The cumulative distribution function Φ for large values of its arguments admits the following approximation (Abramowitz and Stegun, 1972):

$$1 - \Phi(z) \sim \frac{1}{2\sqrt{\pi}} z^{-1} e^{-z^2} \left(1 + \frac{1}{z^2} + R_1(z)\right), \tag{10}$$

where $|R_1(z)| < 1/(2z^2)$, as $z \to \infty$. Therefore, either $p_H(n)$ or $q_H(n)$, depending whether $\mu(\vec{x}) > I$ or not, converges to zero at a higher rate than $1/n$. This implies that for large n the term that will contribute more to $Var[\hat{V}_H]$ is $Var[C]$.

Using representation (6), an analogous analysis can be carried out for the variance of $\hat{V}_L$. In this case the leading term is $Var[C_2]$.

These results can be used to find γ in (8) that minimizes the mean square error of $\hat{V}_P$. We have,

$$\begin{aligned} MSE[\hat{V}_P] &:= E[(\gamma\hat{V}_H + (1-\gamma)\hat{V}_L - V_{\vec{x}})^2] \\ &= Var[\gamma\hat{V}_H + (1-\gamma)\hat{V}_L] + [E(\hat{V}) - V_{\vec{x}}]^2. \end{aligned} \tag{11}$$

Our previous analysis suggests that

$$Var[\hat{V}_H] \simeq \frac{1}{n}\sigma^2 \qquad \text{and} \qquad Var[\hat{V}_L] \simeq \frac{2}{n}\sigma^2, \tag{12}$$

where $\sigma^2 \equiv \sigma^2(t_i, \vec{x})$. In order to asses the bias of these estimators, we need to find their expectations:

$$E[\hat{V}_L] = Ip_L(n/2) + V_{\vec{x}} q_L(n/2) \tag{13}$$

and

$$\begin{aligned} E[\hat{V}_H] &= Ip_H(n) + E[C\aleph_{\{C>I\}}] \\ &= Ip_H(n) + V_{\vec{x}} q_L(n) + Cov(C, \aleph_{\{C>I\}}). \end{aligned} \tag{14}$$

By comparing (13) and (14), we conclude that $\hat{V}_H$ will have a larger bias than $\hat{V}_L$. This is consistent with our numerical results. These formulae, however, suggest also that both estimators converge to the true value at a rate that corresponds to representation (10) of the c.d.f Φ. This high rate would make the bias term in the mean square error (11) negligible when

compared with the variances. This, however, is due to the fact that in our analysis we assume that $\hat{V}(t_{i+1},\cdot)$ is equal to the true value $V(t_{i+1},\cdot)$. This assumption does not play an important role for estimates of relative variances but it is crucial for assessment of the bias. Therefore, to proceed with our analysis we have to rely on our empirical experience, which suggests that both estimators converge at a rate that is close to $1/\sqrt{n}$. If we apply this to the bias term in (11), then we get

$$MSE[\hat{V}_P] = \gamma^2\frac{\sigma^2}{n}+(1-\gamma)^2\frac{2\sigma^2}{n}+2\gamma(1-\gamma)Cov(\hat{V}_L,\hat{V}_H)+\left[\frac{\gamma}{\sqrt{n}}-(1-\gamma)\sqrt{\frac{2}{n}}\right]^2. \tag{15}$$

The covariance $Cov(\hat{V}_L,\hat{V}_H)$ will have the same order as $Cov(C,C_2)$, which is equal to σ^2/n. Thus, representation (15) can be used to minimize the mean square error with respect to γ and the optimal value is $\gamma_o = 0.65$. In practice, especially for smaller values of n, we expect this number to be lower due to a lower bias of $\hat{V}_L$ compared with $\hat{V}_H$.

For low discrepancy sequences the corresponding rates of convergence of variances will be higher. In particular, for randomized low discrepancy sequences that we have used in our numerical examples, the rates in (12) will become:

$$Var[\hat{V}_H] \simeq \frac{\log(n)^{s-1}}{n^3} \qquad \text{and} \qquad Var[\hat{V}_L] \simeq \frac{2^3\log(n/2)^{s-1}}{n^3}.$$

Assuming that now estimators converge at the rate $\log(n)^{(s-1)/2}n^{-3/2}$, which is slightly higher than our empirical studies show, we can find the value γ that minimizes the mean square error (15). For LDM method, we do not know the correlation coefficient between C and C_2. If we denote it by ρ, then the optimal value can be expressed as a function of ρ in the following way

$$\gamma_o = \frac{A_nB_n(1-\rho)+2B_n^2}{2(A_n+B_n)^2-2A_nB_n(1+\rho)},$$

where

$$A_n = \frac{[\log(n)]^{(s-1)/2}}{n^{3/2}}, \qquad \text{and} \qquad B_n = \frac{2^{3/2}[\log(n/2)]^{(s-1)/2}}{n^{3/2}}.$$

Unlike the stochastic case, now the results depend on the dimension of the state space. For example, for $s = 2$ and ρ between 0.5 and 1, the optimal γ is

in the interval $[0.83, 0.88]$. When $s = 5$, the optimal values are in the interval $[0.79, 0.84]$. These particular cases illustrate well the general dependency of γ on some of the parameters. We make the following remarks:

(i) the optimal combination of the high- and low-biased estimators is not sensitive to the choice of the value of the correlation parameter,

(ii) for LDM method, the high-biased estimator has a larger weight assigned to it than in the random sequences case,

The first observation above makes applications of this analysis easy, as we do not have to estimate ρ. Remark (ii) is consistent with the numerical examples presented in the next section.

3 Numerical Results

In this section we discuss some applications of the low discrepancy mesh method and the stochastic mesh method to the valuation of American derivative contracts. We assume that (under the risk-neutral measure Q) the asset prices follow a multivariate geometric Brownian motion

$$dS^i(t) = S^i(t) \left[rdt + \sum_{j=1}^{s} M_{ij} dB^j(t) \right], \quad i = 1, \ldots, s,$$

where r is the risk-free rate of return, $B^1(t), \ldots, B^s(t)$ are independent standard Brownian motions, and $\vec{M}$ is the volatility matrix. We assume that the instantaneous covariance matrix $\vec{\Sigma} = \vec{M}\vec{M}'$ is nonsingular. This model implies that conditional on $\vec{S}_{t_{i-1}}$, $\vec{S}_{t_i}$ has a multivariate lognormal distribution. Hence, the mesh density for the stochastic method becomes the mixture of lognormals (see (2)). In the following simulation studies, the stochastic mesh is generated using the "ran2" pseudo-random number generator of Press et al. (1992). For the LDM, we use the marginal (lognormal) density $f_o(t_{i+1}; \vec{u})$ as the mesh density together with the scrambled low discrepancy sequences of Owen (1995). This is motivated by Proposition 1 as well as the results in Boyle et al. (2000).

3.1 Low Discrepancy Mesh vs Stochastic Mesh Estimators

The goal in this subsection is to evaluate the efficiency of the various mesh estimators. For each generated mesh, either based on average density (ADM) or LDM, three estimators are computed:

1. the high-biased mesh estimator,
2. the low-biased mesh estimator,
3. the point mesh estimator (8). For the ADM method, we set $\gamma = 0.65$. For the LDM, we fix $\gamma = 0.85$ for the first example and $\gamma = 0.8$ for the second example, as justified in the previous section.

For each simulated mesh estimator, we compute the root mean square error (RMSE) as defined by

$$RMSE_n = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{\hat{C}_i - C_i}{C_i}\right)^2},$$

where M is the number of option contracts, $\hat{C}_i$ is the simulated estimator for the i-th American option contract based on n mesh points and C_i is the true price of the contract.

In our numerical examples, we consider two types of option contracts with the following features:

- American call options on the maximum of two assets and with 21 exercise dates,
- American call options on the geometric average of three assets and with 21 exercise dates.

We have chosen These two types of option contracts mostly because the correct option premiums can be approximated accurately using the bivariate lattice method of Boyle, Evnine and Gibbs (1989) and the Cox, Ross and Rubinstein (1979) lattice model, respectively. The option values from these procedures are then used as a benchmark for computing the RMSE.

For the American call option on the maximum of two assets, we consider $M = 50$ option contracts with mesh sizes $n \in \{128, 256, 512, 1024, 2048, 4096\}$.

The initial asset prices are assumed to be fixed at 100 while the remaining parameter values are randomly generated as follows: the strike price $K \sim U[80, 120]$, the time until maturity of the contract $T \sim U[1\ month, 4\ years]$, the correlation between assets is $\rho \sim U[-0.4, 0.7]$, the annual interest rate $r \sim U[1\%, 5\%]$, the annual volatilities and the annual dividend rates for each asset are, respectively, $\sigma_i \sim U[10\%, 50\%]$ and $d_i \sim U[5\%, 12\%], i = 1, 2$.

For American call options on the geometric average of three assets, we generate 25 option contracts using the same randomization procedure as in the previous example except for the correlations. In this case, we assume that all the pairwise correlations are identical and their values are randomly generated from the range $(0, 0.5)$. For the mesh sizes, we consider $n \in \{243, 729, 2187, 3645, 5832\}$.

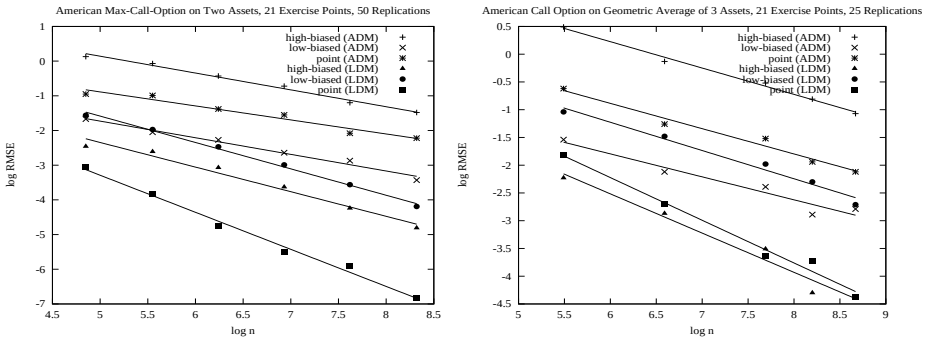

Figure 2: Comparison of high-biased, low-biased and point estimators of LDM and ADM methods

The results (plotted in log n vs. log RMSE) are depicted in Fig. 2. We make the following remarks:

- For both types of option contracts, the high-biased estimators of ADM are extremely biased. The low-biased estimators of ADM, on the other hand, are significantly better relative to the corresponding biased high estimators.

- The proposed high-biased LDM estimator is extremely effective at reducing the bias. The low-biased estimator, on the other hand, is not as efficient. The LDM low-biased estimator compares favorably to the low-biased estimator of ADM for the low-dimensional option contracts

while compares inferiorly for the high-dimensional examples. The deterioration of the LDM low-biased estimator as dimension increases can be attributed to the lack of independence of the mesh points but to get a better understanding of this issue more work is necessary.

- For the point estimator, we take the proportion γ as 0.65 for the ADM estimator. This proportion is optimum asymptotically as argued in Section 2.3. The point estimators compare favorably to the high-biased of ADM but inferiorly to the corresponding low-biased estimators. This is a consequence of the slow rate of convergence of the ADM methods and from the mesh sizes considered in these examples, the asymptotic rate has not been attained. In view of the significant biased of the high estimators relative to the low-biased estimators, a possible enhancement to the point estimators is to allow greater weight in the low estimator and a smaller weight to the high estimator. In fact if we had taken γ to be around 10%, the resulting point estimators would compare favorably to both high and low estimators of ADM.

- For the LDM methods, we have used γ equal to 0.85 and 0.8 for the options that depended on two and five assets, respectively. The point estimators are very effective for the low-dimensional problems while not as efficient for the higher-dimensional examples. The larger bias in the underlying low estimator of LDM implies that higher proportion should be assigned to the high LDM estimator. We have again confirmed this observation that using $\gamma = 0.9$ would produce point estimators that outperform both high and low estimators.

3.2 Mesh Estimators with Control Variates

In the numerical results reported in the last subsection, we show that by taking an appropriate proportion between the low-biased and high-biased estimators, the efficiency of the point estimator can further be enhanced. The experiment, however, also indicates that if an "inappropriate" proportion has been taken, the resulting point estimator can be worst than either the bias-high and bias-high estimates, depending which one is closer to the true value. This can be attributed to the different rates of convergence exhibited by the low- and high-biased estimators. In practice, it is desirable to report the low- and high-biased estimators since this provides a range for which the

true value must lie. More specifically, let $\bar{V}_L$ and $\bar{V}_H$ be the low- and high-biased estimators based on M independent meshes and $\hat{\sigma}(\hat{V}_L)$ and $\hat{\sigma}(\hat{V}_H)$ are respectively, the sample standard deviations of $\hat{V}_L$ and $\hat{V}_H$. Then the $100(1-\alpha)\%$ confidence interval can be constructed from these two estimators as

$$\left[\bar{V}_L - z_{\alpha/2}\frac{\hat{\sigma}(\hat{V}_L)}{\sqrt{M}},\ \bar{V}_H + z_{\alpha/2}\frac{\hat{\sigma}(\hat{V}_H)}{\sqrt{M}} \right], \tag{16}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. To ensure that this interval is as tight as possible, one source of enhancement is to reduce the magnitude of both $\hat{\sigma}(\hat{V}_L)$ and $\hat{\sigma}(\hat{V}_H)$. This can be achieved through the variance reduction techniques. In this subsection, we consider several examples and discuss the impact of incorporating a commonly used variance reduction tool known as the control variate technique.

In our study, we have considered an arithmetic average options with 5 assets and 11 different exercise dates. This is a classical example where there exists no competitive approach for obtaining an accurate estimate of the option price. The parameter values are: $K = 100, r = 5\%, T = 3$ years, $S_i = \{90, 100, 110\}$, $d_i = 10\%, \sigma_i = 30\%, \rho_{ij} = 0.5$ for $i, j = 1, \ldots, 5, i \neq j$. A natural candidate for the control variate in this case is the corresponding geometric average option since the option premium of the geometric average option can be approximated accurately from the CRR model. The results for both ADM and LDM methods based on 10 independent replications are shown in Table 1. Based on these simulation results, we make the following remarks:

- Incorporating control variates is effective at reducing the variance for both ADM and LDM methods. The impact is marginally more pronounced for the ADM methods (see columns (i) and (ii)).

- The confidence intervals constructed from LDM are narrower than the corresponding intervals from ADM. The order of improvement, as measured by the ratio of the width of the 90% confidence intervals of ADM to LDM, ranging from 3.5 to 5.6 (see column (iv)) without the control variate and 3.3 to 4.5 (see column (v)) in the presence of control variate.

Table 1: Comparison of ADM and LDM methods with and without control variates. The values in brackets denote the respective standard errors based on 10 independent replications

S	n	No Control Variate		With Control Variate				(iii)[c]	(iv)[d]	(v)[e]
		Low-Biased	High-Biased	Low-Biased	(i)[a]	High-Biased	(ii)[b]			
ADM: Stochastic Mesh Method										
90	1250	5.29(0.09)	10.58(0.30)	5.78(0.02)	4.1	7.07(0.05)	5.4	4.2		
	2500	5.36(0.07)	9.65(0.26)	5.81(0.02)	3.8	6.89(0.04)	5.8	4.1		
	5000	5.55(0.04)	9.01(0.07)	5.85(0.01)	5.3	6.80(0.02)	3.5	3.7		
100	1250	9.10(0.10)	16.75(0.39)	9.61(0.01)	7.8	11.50(0.07)	5.4	4.2		
	2500	9.21(0.07)	15.17(0.19)	9.67(0.02)	4.3	11.22(0.04)	5.1	3.9		
	5000	9.34(0.04)	14.04(0.09)	9.70(0.01)	4.9	11.09(0.02)	3.9	3.4		
110	1250	14.30(0.10)	24.36(0.47)	14.80(0.02)	3.9	17.42(0.08)	5.7	3.9		
	2500	14.34(0.11)	21.93(0.39)	14.84(0.02)	5.0	17.06(0.07)	5.8	3.6		
	5000	14.48(0.04)	20.23(0.12)	14.88(0.02)	2.4	16.81(0.03)	3.9	3.0		
LDM: Low Discrepancy Mesh Method										
90	1250	4.64(0.02)	5.95(0.07)	5.62(0.01)	2.3	5.91(0.02)	3.2	4.2	4.1	4.1
	2500	4.90(0.02)	6.12(0.05)	5.67(0.00)	4.6	5.97(0.01)	4.7	4.1	3.6	3.7
	5000	5.11(0.01)	6.11(0.02)	5.71(0.00)	3.0	5.99(0.01)	3.1	3.5	3.5	3.3
100	1250	8.44(0.02)	10.12(0.05)	9.45(0.01)	2.1	9.87(0.01)	3.5	3.8	4.8	4.4
	2500	8.66(0.02)	10.21(0.05)	9.49(0.01)	3.7	9.95(0.01)	4.2	3.4	3.8	3.4
	5000	8.93(0.01)	10.09(0.02)	9.54(0.00)	2.5	9.94(0.01)	3.0	3.0	4.1	3.5
110	1250	13.70(0.02)	15.54(0.06)	14.60(0.01)	2.7	15.22(0.02)	2.6	2.9	5.6	4.2
	2500	13.95(0.01)	15.48(0.05)	14.66(0.00)	2.8	15.24(0.02)	2.8	2.7	5.2	3.8
	5000	14.20(0.01)	15.25(0.02)	14.71(0.00)	2.4	15.14(0.01)	2.9	2.5	5.5	4.5

[a] Ratio of low-biased standard errors no control variate to control variates.
[b] Ratio of high-biased standard errors of no control variate to control variates.
[c] Ratio of the width of 90% confidence interval without control variate to with control variate.
[d] Ratio of the width of 90% confidence interval of ADM to LDM (no control variate)
[e] Ratio of the width of 90% confidence interval of ADM to LDM (control variate)

4 Conclusion

The current challenge in using simulation to price high dimensional American derivatives is to obtain a procedure that is computationally efficient. In this paper we have made progress on this problem. We have been able to combine two useful labour saving devices in our implementation. The first is the stochastic mesh method of Broadie and Glasserman which solves the problem of the exploding tree by restricting the expansion of the state space as we move through time. The second is the use of low discrepancy sequences, which have already proved useful in the case of European derivatives. We were able to combine both these features in our approach and this is the source of its computational efficiency. In this paper, we have studied two estimators: biased low and biased high estimators. The efficiencies of these estimators are analyzed. The LDM high-biased estimator is in particular very promising.

Furthermore, we have also shown that the proposed estimators can be enhanced by incorporating the standard variance reduction techniques such as the control variates. Other techniques are certainly possible and similar order of improvement can be expected. As many of them will be problem-dependent, we have not implemented them in this study since our goal was to provide a more general framework for pricing American options.

Appendix

Here we present a proof of Proposition 1. By the assumption, there exists a constant, say A, such that $I(\cdot) \leq A$. From this it is easy to show that

$$Var[\hat{V}_L(t_{i+1}, \vec{S}_{t_{i+1}})|\vec{S}_{t_i} = \vec{x}] = O\left(\frac{1}{n}\right)$$

uniformly for all states $\vec{x}$ and $i = 0, \ldots, d-1$.

The proof is carried out using the induction method. At the last stage, t_d, the estimate is equal to I and the result is satisfied trivially. Suppose that at stage t_{i+1} for arbitrary $\epsilon_{i+1} > 0$ there exists a number, say N, such that for all $n > N$ we have

$$E[\hat{V}_L(t_{i+1}, \vec{x})] \leq V(t_{i+1}, \vec{x}) + \epsilon_{i+1} \text{ for } \vec{x} \in \mathcal{R}^s.$$

Let us denote by $\aleph_A$ the indicator function of a set A. Then, using basic probability rules, we have

$E[\hat{V}_L(t_i,\vec{x})]$

$$
\begin{aligned}
&= E[I(t_i,\vec{x})\aleph_{\{I(t_i,\vec{x})\geq C_{\mathcal{A}}(t_i,\vec{x})\}}] + E[C_{\mathcal{A}^c}(t_i,\vec{x})\aleph_{\{I(t_i,\vec{x})<C_{\mathcal{A}}(t_i,\vec{x})\}}] \\
&= I(t_i,\vec{x})p + (1-p)E[C_{\mathcal{A}^c}(t_i,\vec{x})] + Cov(C_{\mathcal{A}^c}(t_i,\vec{x}),\aleph_{\{I(t_i,\vec{x})<C_{\mathcal{A}}(t_i,\vec{x})\}}) \\
&\leq I(t_i,\vec{x})p + (1-p)E[C_{\mathcal{A}^c}(t_i,\vec{x})] + Var[C_{\mathcal{A}^c}(t_i,\vec{x})]Var[\aleph_{\{I(t_i,\vec{x})<C_{\mathcal{A}}(t_i,\vec{x})\}}] \\
&= I(t_i,\vec{x})p + (1-p)E[C_{\mathcal{A}^c}(t_i,\vec{x})] + p(1-p)Var[C_{\mathcal{A}^c}(t_i,\vec{x})], \qquad (17)
\end{aligned}
$$

where $p = P\{I(t_i,\vec{x}) \geq C_{\mathcal{A}}(t_i,\vec{x})\}$. Let us observe that by the induction assumption we have

$$
\begin{aligned}
E[C_{\mathcal{A}^c}(t_i,\vec{x})] &= E\left[\frac{1}{|\mathcal{A}^c|}\sum_{l\in\mathcal{A}^c} E[\hat{V}_L(t_{i+1},\vec{X}_{t_{i+1}}(l))w(t_i,\vec{x},\vec{X}_{t_{i+1}}(l)) \mid \vec{X}_{t_{i+1}}(l)]\right] \\
&\leq E\left[\frac{1}{|\mathcal{A}^c|}\sum_{l\in\mathcal{A}^c} V(t_{i+1},U)w(t_i,\vec{x},U)\right] + A\epsilon_{i+1} \\
&= E[V(t_{i+1},\vec{S}_{t_{i+1}})|\vec{S}_{t_i}=\vec{x}] + A\epsilon_{i+1}, \qquad (18)
\end{aligned}
$$

where the random variable U has a distribution with the density function $g_{t_{i+1}}$.

Using conditioning argument, the assumption that $I(\cdot)$ is bounded, and the fact that $\hat{V}_L(t_{i+1},\vec{X}_{t_{i+1}}(l))w(t_i,\vec{x},\vec{X}_{t_{i+1}}(l))$, $l=1,\ldots,n$, are independent conditionally on points $\vec{X}_{t_k}$, $k>i+1$, it can be shown that

$$
Var[C_{\mathcal{A}^c}(t_i,\vec{x})] = O\left(\frac{1}{n}\right), \qquad (19)
$$

uniformly over the state space. Combining (17), (18), and (19), we get

$E[\hat{V}_L(t_i,\vec{x})]$

$$
\begin{aligned}
&\leq pI(t_i,\vec{x}) + (1-p)E[V(t_{i+1},\vec{S}_{t_{i+1}})|\vec{S}_{t_i}=\vec{x}] + (1-p)\left[\epsilon_{i+1} + p\frac{M}{n}\right] \qquad (20) \\
&\leq \max[I(t_i,\vec{x}), E[V(t_{i+1},\vec{S}_{t_{i+1}})|\vec{S}_{t_i}=\vec{x}]] + (1-p)\left[\epsilon_{i+1} + p\frac{M}{n}\right], \qquad (21)
\end{aligned}
$$

where M is a constant. From (21) the result follows. In the case when $I(0,\vec{S}_0) \neq C(0,\vec{S}_0)$, it is easy to see that for sufficiently large n the maximum of (20) over $p\in[0,1]$, is less than $\max[I(t_i,\vec{x}), E[V(t_{i+1},\vec{S}_{t_{i+1}})|\vec{S}_{t_i}=\vec{x}]]$, which proves the second statement in the proposition.

Notes

[1]Bossaerts (1989) also made an important contribution to the American option problem by solving for the early exercise strategy but his work is still in working paper form and this probably explains why it has often been overlooked.

[2] Boyle, Broadie and Glasserman (1997) give an overview of the research in this area in their review paper. Subsequent contributions, which deal with the estimation of American derivatives by Monte Carlo simulation include Barraquand and Martineau (1995), Carrière (1996), Grant, Vora and Weeks (1997), Broadie and Glasserman (1997a, 1997b), Broadie, Glasserman and Jain (1997), Raymar and Zwecher (1997), Carr and Yang (1997, 1998), Andersen (1999), Garcia (1999), Ibanez and Zapatero (1999), Longstaff and Schwartz (2000).

References

[1] M. Abramowitz and I.A. Stegun *Handbook of mathematical functions, with formulas, graphs, and mathematical tables.* Washington, D.C. U.S. Dept. of Commerce, 1972.

[2] L.B.G. Andersen. A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor market model. Working paper, General Re Financial Products, NY, 1999.

[3] A.N. Avramidis and P. Hyden. Efficiency improvement for pricing American options with a stochastic mesh. In P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, editors, *Proceedings of the 1999 Winter Conference*, pages 344–350. 1999.

[4] T. Avramidis, Y. Zinchenko, T.F. Coleman and A. Verma. Efficiency improvement for pricing American options with a stochastic mesh: Parallel implementation. Working paper, Cornell University, 2000.

[5] J. Barraquand and D. Martineau. Numerical valuation of high dimensional multivariate American securities. *Journal of Financial and Quantitative Analysis*, 30(3):383–405, 1995.

[6] P. Bossaerts. Simulation Estimators of Optimal Early Exercise. Working paper, Carnegie Mellon University, 1989.

[7] P.P. Boyle, M. Broadie, and P. Glasserman. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

[8] P.P. Boyle, J. Evnine, and S. Gibbs. Numerical evaluation of multivariate contingent claims. *Review of Financial Studies*, 2(2):241–250, 1989.

[9] P.P. Boyle, A. Kolkiewicz, and K.S. Tan. Pricing American style options using low discrepancy mesh methods. Technical report, IIPR 00-07, University of Waterloo, 2000.

[10] M. Broadie and P. Glasserman. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control*, 21(8-9):1323–1352, 1997a.

[11] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional American options, 1997b. Working paper, Columbia University.

[12] M. Broadie, P. Glasserman, and G. Jain. Enhanced Monte Carlo estimates for American option prices. *Journal of Derivatives*, 5(1):25–44, 1997.

[13] P. Carr and G. Yang. Simulating Bermudan interest rate derivatives. Working paper, Morgan Stanley/Numerix, 1997.

[14] P. Carr and G. Yang. Simulating American bond options in HJM framework. Working paper, Morgan Stanley/Numerix, 1998.

[15] J.F. Carrière. Valuation of the early-exercise price for options using simulation and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

[16] Cox, J.C. and S.A. Ross and M. Rubinstein. Option Pricing: A Simplified Approach. *Journal of Financial Economics*, 7:229–263, 1979.

[17] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, 2nd edition, 1996.

[18] D. Garcia. A Monte Carlo procedure for pricing American options. Working Paper, University of California, Berkeley, 1999.

[19] D. Grant, G. Vora, and D. Weeks. Path-dependent options: Extending the Monte Carlo simulation approach. *Management Science*, 43(11):1589–1602, 1997.

[20] A. Ibanez and F. Zapatero. Monte Carlo valuation of American options through computation of the optimal exercise frontier. Working Paper, Instituto Tecnologico Autonomo de Mexico, 1999.

[21] C. Joy, P.P. Boyle, and K.S. Tan. Quasi-Monte Carlo methods in numerical finance. *Management Science*, 42(6):926–938, 1996.

[22] F.A. Longstaff and E.S. Schwartz. Valuing American options by simulation: A simple least-square approach. to appear in *Review in Financial Studies*.

[23] A.B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P.J-S. Shiue, editors, *Monte and quasi-Monte-Carlo Methods in Scientific Computing: proceedings of a conference at the University of Nevada, Las Vegas, Nevada, USA, June 23-25, 1994*, volume 106 of *Lecture Notes in Statistics*, pages 299–317. Springer-Verlag, New York, 1995.

[24] W. Press, S. Teukolsky, T.V. William, and P.F. Brian. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.

[25] S. Raymar and M. Zwecher. A Monte Carlo valuation of American options on the maximum of several stocks. *Journal of Derivatives*, 5(1):7–23, 1997.

[26] J. A. Tilley. Valuing American options in a path simulation model. *Transactions of the Society of Actuaries*, XLV:499–520, 1993.