EFFICIENCY IMPROVEMENTS FOR PRICING AMERICAN OPTIONS WITH A STOCHASTIC MESH

Athanassios N. Avramidis
Paul Hyden

School of Operations Research and
Industrial Engineering
Cornell University
Ithaca, NY 14853, U.S.A.

ABSTRACT

We develop and study general-purpose techniques for improving the efficiency of the stochastic mesh method that was recently developed for pricing American options via Monte Carlo simulation. First, we develop a mesh-based, biased-low estimator. By recursively averaging the low and high estimators at each stage, we obtain a significantly more accurate point estimator at each of the mesh points. Second, we adapt the importance sampling ideas for simulation of European path-dependent options in Glasserman, Heidelberger, and Shahabuddin (1998a) to pricing of American options with a stochastic mesh. Third, we sketch generalizations of the mesh method and we discuss links with other techniques for valuing American options. Our empirical results show that the bias-reduced point estimates are much more accurate than the standard mesh-method point estimates. Importance sampling is found to increase accuracy for a smooth option-payoff functions, while variance increases are possible for non-smooth payoffs.

1 INTRODUCTION

In the financial markets, sophisticated, complex products are continuously offered and traded. With the increasing complexity of these products, Monte Carlo simulation is steadily becoming an important tool used in valuing and hedging the products. In this paper, the term American option refers to a discretely, as opposed to coninuously, exercisable option–that is, the option holder can exercise the option at a fixed set of time points (also called exercise opportunities, or stages) up to expiration. When valuing such options through simulation, one is jointly estimating the value of the option and the optimal exercise policy.

An important method developed recently for valuing American options through simulation is the stochastic mesh method (Broadie and Glasserman 1997a), henceforth re-

ferred to as BG1997a. We postpone the description and discussion of the mesh method until Section 2.

In this paper, we are concerned with improving the efficiency of the stochastic mesh method by resorting to general-purpose techniques that can be customized to a given problem instance rather easily. For example, although control- variate-based methods were found to be very effective in BG1997a, we view them as special-purpose, as they must be carefully tailored to the specific option in hand. Without resorting to variance reduction techniques, it is likely that a plain-vanilla mesh-based estimation of the option value will suffer from bias much more than from variance.

Our primary contribution is to develop general-purpose bias-reduced versions of the mesh estimators. A within-mesh biased-low estimator is developed by splitting the set of states in disjoint sets and using one of the two sets for estimation of the optimal exercise policy and the other set for the estimation (if necessary) of the option's continuation value. By recursively averaging the low and high estimators at each stage, we obtain a significantly more accurate point estimator at each of the mesh points.

As a further contribution, we adapt the importance sampling ideas for simulation of European path-dependent options proposed and studied in Glasserman, Heidelberger, and Shahabuddin (1998a) in the pricing of American options hrough a stochastic mesh. We also propose generalizations of the mesh method including variable shapes, and mechanisms for sampling paths conditional on the state of all paths at a given stage; and we discuss links with other techniques for valuing American options.

This paper is organized as follows. In Section 2 we review American-option pricing and the stochastic mesh method. In Section 3 we detail new estimation methods and results, and discuss generalizations of the mesh method. In Section 4 we present Monte Carlo results quantifying the performance of the new estimators.

2 AMERICAN OPTION PRICING: BACKGROUND

Let $S_t = (S_t^1, \ldots, S_t^n)$ denote the vector of securities underlying the option, modeled as a Markov process on R^n with discrete time-parameter $t = 0, 1, \ldots, T$. The argument $t = 0, 1, \ldots, T$ indexes the set of times when the option is exerciseble, also called *exercise opportunities* or simply *stages*.

Let $h(t, x)$ be the payoff from exercise at time t in state x, discounted to time 0 with the possibly stochastic discount factor recorded in S_t. The option value starting at time t in state x is

$$q(t, x) = \max\left(h(t, x), c(t, x)\right) \quad \text{for } t < T$$

where

$$c(t, x) = E\left[q(t + 1, S_{t+1}) | S_t = x\right] \qquad (1)$$

is the value of the option when not exercised at time t in state x, discounted to time 0, and called the *continuation value* at (t, x); and $q(T, x) = h(x)$ for all states x. Arbitrage-pricing theory suggests that the arbitrage-free price of the option is obtained when the conditional expectation in (1) is with respect to the *risk-neutral measure*, defined as the measure that makes the value of any tradeable security, discounted to time 0, a martingale. The problem is to compute the option value at time 0, $q_o \equiv q(0, s_o)$, where s_o is the known state of underlyings at time 0. *Examples.* In a simple application, S_t is a vector of n stock prices. A *max option* has payoff function

$$h(t, S_t) = \left(\max\left(S_t^1, \ldots, S_t^n\right) - K\right)^+,$$

where K is the *strike price* and x^+ stands for $\max(x, 0)$. A *geometric average option* has

$$h(t, S_t) = \left(\left(\prod_{k=1}^n S_t^k\right)^{(1/n)} - K\right)^+.$$

2.1 Mesh Method Description

In reviewing the method, we follow BG1997a. The method generates a *stochastic mesh* of sample states (also called points) $S_{t,i}, i = 1, \ldots, b$ for each $t = 1, \ldots, T$. Let $S_{0,i} = s_o, i = 1, \ldots, b$. For $t = 1, \ldots, T$, let $g_t(\cdot)$ denote the probability density from which the points $S_{t,i}, i = 1, \ldots, b$ are sampled (to be specified later), and let $f_t(x, \cdot)$ denote the conditional risk-neutral density of S_{t+1} given $S_t = x$. (In accordance with the authors, we assume throughout the paper the existence of such densities.) Finally, let $B = \{1, \ldots, b\}$ be the set of indices of mesh points at each

stage. The mesh estimator of the option value is defined recursively:

$$\widehat{q}_H(T, S_{T,i}) = h(T, S_{T,i})$$

for $i = 1, \ldots, b$; and for $t = T - 1, \ldots, 0$ and for $i = 1, \ldots, b$, the *high mesh estimator* is

$$\widehat{q}_H(t, S_{t,i}) = \max\left(h(t, S_{t,i}), \widehat{c}(t, S_{t,i}, B)\right) \qquad (2)$$

where the continuation value of each point sampled at stage t depends on the previously calculated continuation values of all points sampled at stage $t + 1$:

$$\begin{aligned}
&\widehat{c}(t, S_{t,i}, B) \\
&= \frac{1}{b} \sum_{j \in B} \widehat{q}_H(t + 1, S_{t+1,j}) w(t, S_{t,i}, S_{t+1,j})
\end{aligned}$$

where

$$w(t, S_{t,i}, S_{t+1,j}) = \frac{f_t(S_{t,i}, S_{t+1,j})}{g_{t+1}(S_{t+1,j})}. \qquad (3)$$

The weighing of the combination of points $(S_{t,i}, S_{t+1,j})$ above is necessary in light of the fact that the points at stage $t + 1$ were sampled from the density $g_{t+1}(\cdot)$ instead of the density $f_t(S_{t,i}, \cdot)$ appropriate for point $S_{t,i}$. We make the dependence of $\widehat{c}$ on B explicit for subsequent convenience. Finally, note that $\widehat{c}(t, S_{t,i}, B)$ would be an unbiased estimator of the corresponding continuation values if the estimated values $\widehat{q}_H(t + 1, S_{t+1,j})$ were unbiased for the corresponding option values (which, generally, is not the case). We refer to $\widehat{q}_H$ as the high mesh estimator, in view of the fact that it is biased high as an estimate of the option value at the corresponding time and state.

The choice of densities g_t is crucial. BG1997a make a strong case for using the *average density function*

$$g_t(u) = \frac{1}{b} \sum_{j=1}^b f_{t-1}(S_{t-1,j}, u) \qquad (4)$$

which corresponds to generating b independent paths of S_t and then "forgetting" the path to which each sampled point belongs at each stage $t = 1, \ldots, T$. In agreement with the authors, we call this particular case the *stratified implementation* of the mesh method.

For actually pricing options, BG1997a suggest obtaining a second estimator by simulating paths of the process S_t independent of the mesh points $S_{t,i}$ until the exercise region implied by the mesh is reached. Specifically, the approximate optimal policy implied by the mesh exercises at $\widehat{\tau} \equiv \min\{t : h(t, S_t) \geq \widehat{q}_H(t, S_t)\}$, with $\widehat{q}_H(t, S_t)$ as in (2). The *path estimator* is then $\widehat{q}_P = h(\widehat{\tau}, S_{\widehat{\tau}})$.

See BG1997a for other properties of the estimators, considerable computational enhancements (mainly through control variates), and an extensive numerical study of the method's performance.

The main alternative for pricing high-dimensional American options using provably consistent estimators is a tree-based simulation where simulated paths branch out at each sampled point for each exercise opportunity (Broadie and Glasserman 1997b). Compared to this alternative, the mesh method has two important advantages: (*a*) it alleviates the exponential growth of the number of points to be sampled with the number of exercise opportunities; and (*b*) sampled paths help each other in the estimation: all states sampled at stage $j + 1$ are used in the estimation of option values at stage j. On the downside, the applicability and ease of use of the mesh method in application might be restricted by the requirement to calculate the conditional risk-neutral densities $f_t(S_{t,i}, \cdot)$.

3 ENHANCING MESH EFFICIENCY

3.1 Bias Reduction for Mesh Estimation

The idea behind the construction of a biased-low estimate is to use disjoint sets of points for estimation of the optimal exercise policy and the estimation of continuation values (in case the estimated optimal policy is to continue).

Assume the mesh points are sampled from the average density function (4). Let $I \subseteq B$ denote an arbitrary subset of all indices, and $I' = B - I$ its complement with respect to B. To simplify notation, we occasionally drop the explicit dependence of estimators on t, writing, for example, $\widehat{c}(S_{t,i})$ for $\widehat{c}(t, S_{t,i})$.

To calculate the low estimator at each stage $t = T - 1, \ldots, 0$, assume the low estimator of the values at all sampled points at stage $t + 1$ has been calculated. Define the estimate of the continuation values at t using *only* the points in I from stage $t + 1$:

$$\widehat{c}(t, S_{t,i}, I)$$
$$= \frac{1}{|I|} \sum_{j \in I} \widehat{q}_{\mathrm{L}}(t + 1, S_{t+1,j}) w(t, S_{t,i}, S_{t+1,j}),$$

where $w(t, S_{t,i}, S_{t+1,j})$ are as in (3) and $|I|$ is the number of elements in I. Define the estimate of the option value at point $S_{t,i}$

$$\widehat{q}_{\mathrm{L}}(t, S_{t,i}, I) = \begin{cases} h(S_{t,i}), & \text{if } h(S_{t,i}) \geq \widehat{c}(t, S_{t,i}, I) \\ \widehat{c}(t, S_{t,i}, I'), & \text{otherwise.} \end{cases}$$

Note that the dependence of $\widehat{q}_{\mathrm{L}}$ on I, the set of points used to estimate the optimal exercise policy, was made explicit; and note the implicit dependence on I'. To maximize usage

of known information at stage $t + 1$ on the estimation at stage t, we form the overall low estimator by averaging b copies of $\widehat{q}_{\mathrm{L}}$ where the j-th copy uses $B_{-j} \equiv B - \{j\}$ in place of I. The *low mesh estimator* of the option value is defined recursively:

$$\widehat{q}_{\mathrm{L}}(T, S_{T,i}) = h(T, S_{T,i})$$

for $i = 1, \ldots, b$; and for $t = T - 1, \ldots, 0$ and for $i = 1, \ldots, b$, the low mesh estimator is

$$\widehat{q}_{\mathrm{L}}(t, S_{t,i}) = \frac{1}{b} \sum_{j=1}^{b} \widehat{q}_{\mathrm{L}}(t, S_{t,i}, B_{-j})$$

Theorem 1. *The estimator $\widehat{q}_{\mathrm{L}}$ is biased low, i.e.,*

$$\mathrm{E}[\widehat{q}_{\mathrm{L}}(t, x)] \leq q(t, x)$$

for all t, x.

For a proof, see Avramidis and Hyden (1999).

A key component in the development and proofs of properies of $\widehat{q}_{\mathrm{H}}$ and $\widehat{q}_{\mathrm{L}}$ is that the points at stage $t + 1$ are sampled from the average density function (4). In particular, this implies that, conditional on $S_{t,i}$, the points $\{S_{t+1,j} : j \in B\}$ are independent and identically distributed (i.i.d.). We claim that in most applications, better estimation can be achieved by "remembering" the path to which each point belongs. This changes the densities $g_t(\cdot)$ and thus the weights of all point combinations, and the i.i.d. property of points stated above is lost. For more details and results on this new view of the mesh, see Avramidis and Hyden (1999).

Both $\widehat{q}_{\mathrm{H}}$ and $\widehat{q}_{\mathrm{L}}$ suffer from recursive bias: assuming that $\widehat{q}_{\mathrm{H}}(t + 1, x)$ and $\widehat{q}_{\mathrm{L}}(t + 1, x)$ are unbiased for all x, one can show that the estimates $\widehat{q}_{\mathrm{H}}(t, x)$ and $\widehat{q}_{\mathrm{L}}(t, x)$ are biased high and low as estimates of $q(t, x)$, respectively. A simple glance at the derivation of these properties shows that bias is accumulating from stage $T - 1$ down to stage 0 for both estimators; see BG1997a for the high mesh estimator and Avramidis and Hyden (1999) for the low mesh estimator. As an intuitive bias reduction scheme, we propose the *average mesh estimator* which is defined recursively:

$$\widehat{q}_{\mathrm{A}}(T, S_{T,i}) = h(T, S_{T,i})$$

for $i = 1, \ldots, b$; and for $t = T - 1, \ldots, 0$ and for $i = 1, \ldots, b$, the average mesh estimator is

$$\widehat{q}_{\mathrm{A}}(t, S_{t,i}) = \frac{1}{2} \left(\widehat{q}_{\mathrm{H}}(t, S_{t,i}) + \widehat{q}_{\mathrm{L}}(t, S_{t,i}) \right),$$

where the values $\widehat{q}_A(t+1, \cdot)$ are used as substitutes for the values $\widehat{q}_H(t+1, \cdot)$ and $\widehat{q}_L(t+1, \cdot)$ in the calculation of $\widehat{q}_H(t, \cdot)$ and $\widehat{q}_L(t, \cdot)$, respectively.

3.2 Importance Sampling

Our discussion in this section is restricted to option-pricing problems that can be cast in a way such that all input random variates are normally distributed. To our knowledge, the majority of option-pricing models fit this framework. Our approach is an adaption of the importance sampling techniques proposed and studied in Glasserman, Heidelberger, and Shahabuddin (1998) in the context of pricing path-dependent European-style options via simulation.

Let Z be the vector of k independent standard normal random variables necessary for sampling an entire path of the underlying securities, $\{S_1, S_2, \ldots, S_T\}$. We denote this as $Z \sim N(0, I_k)$, where I_k is the $k \times k$ identity matrix. Note that any multivariate Normal distribution can be generated as a deterministic function of Z. We obtain a new sampling density for paths of S_t as follows. Treating the option as if it were European, we form the product of option payoff at expiration and likelihood under the risk-neutral measure as a function of Z. The value of Z that maximizes the function above, say μ_o suggests a new sampling density for Z, namely $N(\mu_o, I_k)$. Noting that the latter normal density has its maximum at μ_o, we have effectively chosen the new sampling density so that its maximum is attained at the point where the product of payoff at expiration and likelihood under the risk-neutral measure is maximized. For further motivation on this choice of change of measure, see Glasserman, Heidelberger, and Shahabuddin (1998a).

In view of the results and suggestions in BG1997a, we maintain a pathwise sampling of mesh points, followed by "forgetting" the path to which each point belongs. For each $t = 0, \ldots, T-1$, let $f_{t,\mu_o}(x, \cdot)$ denote the new conditional density of S_{t+1} given $S_t = x$, in view of the fact that transitions were simulated by sampling Z from $N(\mu_o, I_k)$ instead of the original (risk-neutral) measure $N(0, I_k)$. The new average density function becomes

$$g_{t,\mu_o}(u) = \frac{1}{b} \sum_{j=1}^{b} f_{t-1,\mu_o}(S_{t-1,j}, u)$$

for $t = 1, \ldots, T$. This in turn implies new weights with g_{t,μ_o} replacing g_t in (3). Except for the weight adjustment, all other estimators remain as in Sections 2.1 for the standard mesh method and as in Section 3 for the enhanced estimators. For a specific example of the choice of new mesaure, see Section 4.

Since the optimal exercise policy for many options is to hold until expiration, our treatment of an American option as if it were European for the purposes of determining a change

of measure is likely to be a good approximation. In light of the successful results with this change of measure reported in Glasserman, Heidelberger, and Shahabuddin (1998a), we expect similar to slightly less successful results in improving estimation efficiency for American options.

3.3 Mesh Generalizations and Links to Other Methods

We propose generalizations of the mesh method and sketch ideas for dynamically growing the mesh, aiming at more efficient sampling of the entire mesh points. We introduce the concept of *parent* and *child*. When a mesh point B is generated by extending the path of mesh point A, mesh point A is called the parent relative to the child mesh point B. In this terminology, the stratified implementation in BG1997a generates exactly one child per parent. Also, define the *budget* at any stage t as the number of points to sample.

First note that the current choice of a fixed budget per stage is arbitrary. Intuitively, we can say that the density of the sampled mesh points is spread more thinly in later stages. More concretely, our experience with realistic sample sizes, say $b \leq 1000$, suggests that the importance of non-children to any parent is orders of magnitude lower than the importance of its child. This effect is progressively stronger in higher dimensions and, more important to our discussion, later stages. In other words, the effective sample size for the estimation at each parent is generally small, certainly much less than b, and decreasing with dimensionality and stage. Although there appears to be no cure for the effect of dimensionality, this suggests that an increasing budget over stages should enhance efficiency. Preliminary experimentation (not reported here) has confirmed this conjecture.

Thus, our first generalization is to allow the budget to vary with stage, so we have b_t points at stage t. At stage t, and assuming no preference to giving more children to any parent, we can afford on average b_{t+1}/b_t children per parent, called the *growth rate* at stage t. One choice is to allocate to each parent a number of children equal at least to the integral part of this ratio, and then choose randomly the parents to receive the additional children up to the stage's budget b_{t+1}. Interestingly, this method leaves the computation of the actual sampling densities $g_t(\cdot)$ unchanged with respect to the stratified implementation.

In tree-based simulations for american options as in Broadie and Glasserman (1997b), the minimum growth rate for meaningful estimation is two (every parent gets at least two children), requiring 2^T points to be sampled at stage T. Instead, the generalized mesh allows any non-integer growth rate, with rates greater than one but less than two, thus allowing a less explosive growth than the tree-based case. But for now, we have no formal procedure for optimally allocating the b_t's.

A second important generalization is to allocate a varying number of children to the points at a given stage. Define

the *cum-weight* (short for cumulative weight) of a point at stage t as

$$w_{\mathrm{C}}(S_{t,i}) = \sum_{j=1}^{b_{t-1}} w_{\mathrm{C}}(S_{t-1,j}) w(t-1, S_{t-1,j}, S_{t,i}),$$

where $w_{\mathrm{C}}(s_o) = 1$. It is easy to see that $w_{\mathrm{C}}(S_{t,i})$ is also equal to the sum of the *path weights* of all paths from s_o to $S_{t,i}$, where a path weight is equal to the product of weights of all 1-step transitions of the path. The cumweight $w_{\mathrm{C}}(S_{t,i})$ measures the importance of point $S_{t,i}$ on the estimation of the option value at s_o. Asumming no a priori importance sampling of the paths, the higher the cumweight of a point at stage T, the higher its importance in the overall estimation. In early experiments, we have observed strong 1-step cumweight correlation, defined as the correlation of the cumweight of parents to their children. This suggests that more efficient growth of the mesh can be dynamically achieved by allocating more children to higher-cumweight points. More complex rules that also take into account surrogates to the option value (as the immediate-exercise value of a point) might prove even more effective in growing the mesh efficiently. For example, if a path on a call option is way out of the money, it is highly likely that the optimal decision is to delay exercise of the option. Increasing the number of children for that kind of parent is unlikely to be an efficient choice.

What about allocating zero children to a point? For such a point i at stage t, the estimation of the option value will depend on only non-children at stage $t+1$. Of these, the ones that "look" more than i's children will get the greatest weight in the estimation of i's value. In this way, the estimation at i is "tied" to random neighbors of i, namely the parents of the stage-$(t+1)$ points with the highest weight in the estimation of i. This is reminiscent of the bundling methods of Tilley (1993), where paths are bundled at each stage according to a *predefined* similarity (the resulting estimators are not consistent, in general). Yet, the mesh method and its generalizations do not preconceive a path similarity criterion, likely preserving most desirable theoretical properties.

4 NUMERICAL RESULTS AND RECOMMENDATIONS

We report limited results to demonstrate the degree of efficiency improvement achieved over a standard implementation of the mesh method with no variance reduction techniques. As test cases, we use a subset of the test cases in BG1997a. Under the risk-neutral measure, the n assets

are independent, and each follows a geometric Brownian motion process:

$$dS_t^k = S_t^k[(r-\delta)dt + \sigma dW_t^k], \quad k = 1, \ldots, n,$$

where W_t^k is standard Brownian motion, r is the riskless interest rate, δ is the divident rate, and σ is the stock volatility parameter. Exercise opportunities occur at the set of calendar times $t_i = iT/d, i = 0, \ldots, d$, where T is the calendar option expiration time (note that t and T denoted indices of the calendar times t_i and T in the earlier sections). Under the risk-neutral measure, $\ln(S_{t_i}/S_{t_{i-1}})$ is normally distributed with mean $(r - \delta - \sigma^2/2)(t_i - t_{i-1})$ and variance $\sigma^2(t_i - t_{i-1})$.

Case 1 is a max option with $n = 5$, $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $T = 3$, $d = 9$, $s_o = 90$, and $K = 100$. Case 2 is a geometric average option with $n = 7$, $r = 0.03$, $\delta = 0.05$, $\sigma = 0.4$, $T = 1$, $d = 10$, $s_o = 90$ and $K = 100$. We remark that these cases were selected as the most difficult cases (based on our own computations) from the set in BG1997a.

No variance reduction techniques are applied in Cases 1 and 2. Case 3 is the same option as Case 2, where the mesh is sampled through importance sampling as in Section 3.2. Specifically, we let Z_T denote a $1 \times n$ row vector of n independent standard normal variates that drive the sampling of S_T. Let α be a $1 \times n$ row vector with $(r - \delta - \sigma^2/2)T$ at each entry. The new mean of Z_T, μ_o, is found as the solution to

$$\max_z h(T, s_o \exp(\alpha + \sigma\sqrt{T}z) \exp(-\frac{1}{2}zz')$$

where z is a $1 \times n$ row vector and z' is the transpose of z. In view of the new mean for S_T, the paths of S_t are sampled with a new drift implied by μ_o.

We report point-estimator performance (as opposed to confidence-interval performance) for two estimators: (a) $\widehat{q}_{\mathrm{BG}}$ is the point estimator suggested in BG1997a, defined as the average of the high mesh estimator and the path estimator based on a number of simulated paths equal to the parameter b of the mesh; and (b) $\widehat{q}_{\mathrm{A}}$ is the average mesh estimator defined in Section 3.1.

The performance measures are all relative to the true option value. We measure RBIAS, RSTDE, and RRMSE, which are acronyms for relative bias, relative standard error, and relative root mean square error of an estimator, respectively. Figures 1 and 2 show performance for $\widehat{q}_{\mathrm{BG}}$ and $\widehat{q}_{\mathrm{A}}$, respectively, as a function of estimator work, measured in CPU time in a SUN Ultra 1 workstation. Figures 3-4 and 5-6 are analogs of Figures 1-2 for Cases 2 and 3, respectively.
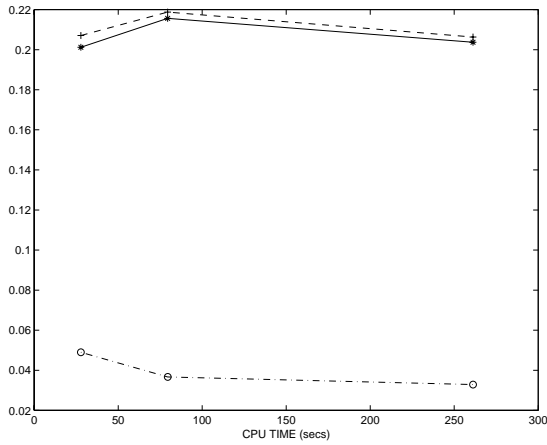
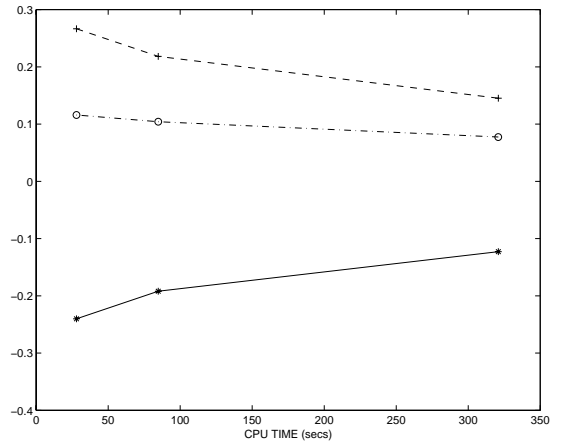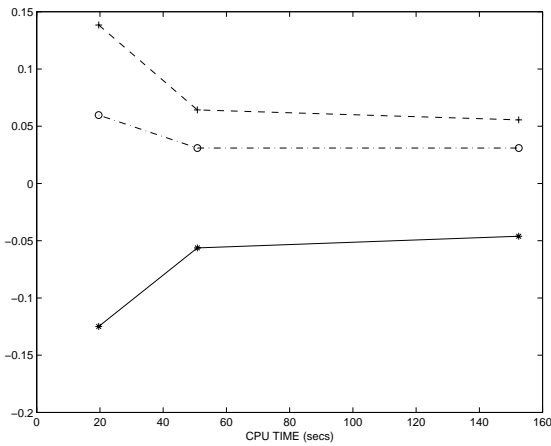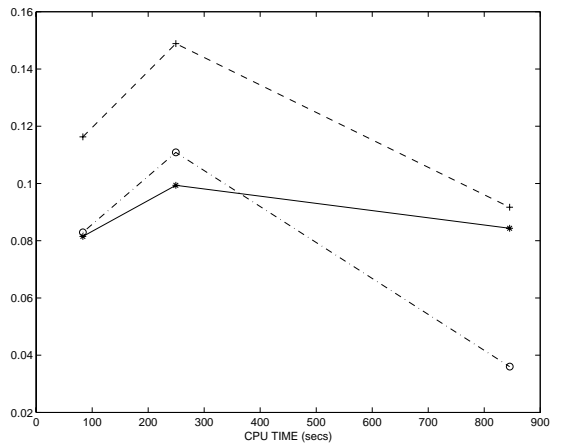Figure 1: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{BG}}$ for Case 1

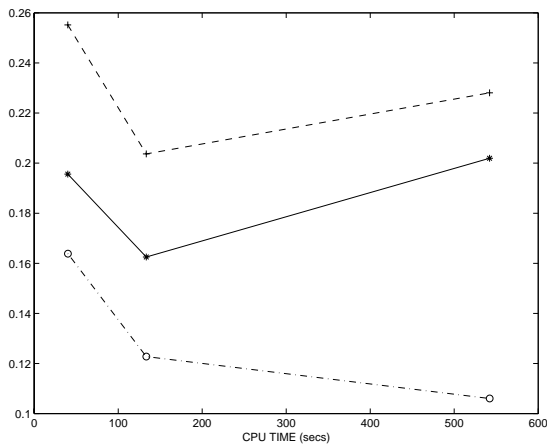Figure 2: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{A}}$ for Case 1

Figure 3: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{BG}}$ for Case 2

Figure 4: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{A}}$ for Case 2

Figure 5: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{BG}}$ for Case 3
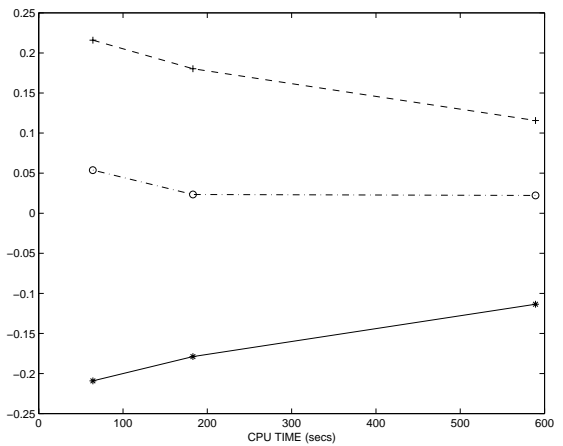

Figure 6: Estimated RBIAS(*), RSTDE(o), and RRMSE(+) of $\widehat{q}_{\mathrm{A}}$ for Case 3

Overall, these results and additional experience not reported here suggest the following:

- The estimator $\widehat{q}_{\mathrm{BG}}$ in a standard implementation of the stochastic mesh method with no variance reduction techniques as in BG1997a suffers seriously from bias. This bias is mainly due to the very strong bias of $\widehat{q}_{\mathrm{H}}$, the mesh high estimator.

- The recursively averaged mesh estimator $\widehat{q}_{\mathrm{A}}$ has small to moderate bias without resorting to variance reduction techniques and appears to be the best point estimator for a general-purpose mesh implementation.

- Even the better $\widehat{q}_{\mathrm{A}}$ estimator will generally yield only moderate accuracy with realistic sample sizes. Although the cases we studied here are the hardest from the set of cases in BG1997a, they underline that general-purpose mesh estimation needs deeper study and that there is substantial room for accuracy improvement.

- Importance sampling for the mesh method is fairly straightfowrward to implement. For smooth payoff functions such as the geometric average option, it yields substantial efficiency improvements. For non-smooth payoffs such as the max option, results not reported here show that importance sampling may backfire.

ACKNOWLEDGMENTS

REFERENCES

Avramidis, A., and P. Hyden. 1999. Efficient Simulation Techniques for Pricing American Options. Unpublished manuscript.

Broadie, M., and P. Glasserman. 1997a. A stochastic mesh method for Pricing High-Dimensional American Options. Unpublished manuscript.

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

Glasserman. P., P. Heidelberger, and P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* 9: (2) 117–152.

Tilley, J.A. 1993. Valuing American Options in a Path Simulation Model. *Transactions of the Society of Actuaries* 45: 83–104.

AUTHOR BIOGRAPHIES

ATHANASSIOS (THANOS) N. AVRAMIDIS joined the faculty of the School of Operations Research and Industrial Engineering at Cornell University in 1997. Prior to that, he was a consultant with SABRE Decision Technologies in Dallas, TX and Paris, France. His recent research interests are in Monte Carlo simulation methods with applications to derivative valuation, hedging, and risk management. He holds M.S. and Ph.D. degrees from the School of Industrial Engineering at Purdue University and a diploma in Mechanical Engineering from the University of Thessaloniki in Greece.

PAUL HYDEN is a Ph.D. candidate at Cornell University's School of Operations Research and Industrial Engineering. His dissertation work focuses on designing efficient and practical simultaneous simulation experiments. He is also interested in semiconductor manufacturing, service operations, and finance.