American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods

Mark Broadie, Jerome Detemple

Review of Financial Studies, Volume 9, Issue 4 (Winter, 1996), 1211-1250.

American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods

Mark Broadie
Columbia University

Jerome Detemple
McGill University and CIRANO

We develop lower and upper bounds on the prices of American call and put options written on a dividend-paying asset. We provide two option price approximations, one based on the lower bound (termed LBA) and one based on both bounds (termed LUBA). The LUBA approximation has an average accuracy comparable to a 1,000-step binomial tree with a computation speed comparable to a 50-step binomial tree. We introduce a modification of the binomial method (termed BBSR) that is very simple to implement and performs remarkably well. We also conduct a careful large-scale evaluation of many recent methods for computing American option prices.

A wide variety of traded options are American options and therefore may be optimally exercised before

This article is an outgrowth of an earlier article titled "Bounds and Approximations for American Option Values." It has been presented at CERCA (Center for Research on Computation and its Applications), the Workshop on Mathematical Finance at Columbia University, the Financial Mathematics Program at the University of Cambridge, and the 1995 Western Finance Association meetings. We thank Phelim Boyle, Robin Brenner, Peter Carr, George Constantinides, Bob McDonald, Gang Yu, conference participants, and an anonymous referee for their comments. We are grateful to the editor, Kerry Back, for detailed suggestions that have improved this article. We also thank Peter Carr and Dmitri Faguet for providing code for their method of lines. Address correspondence to Mark Broadie, 415 Uris Hall, Columbia University, New York, NY 10027.

The Review of Financial Studies Winter 1996 Vol. 9, No. 4, pp. 1211–1250

the maturity of the contract. Commodity options, commodity futures options, call options on dividend-paying stocks, put options on dividend- or nondividend-paying stocks, foreign exchange options, and index options are examples of contracts for which early exercise may be optimal. The optimality of early exercise presents considerable difficulties from a computational viewpoint. Closed form or analytical solutions are not available to price these American options, so numerical approximation methods are required.

Our article has two aims. First, we propose new methods for computing lower and upper bounds on American option values. Based on the bounds, we provide two option price approximations, termed LBA and LUBA. We also introduce a simple modification of the binomial method, termed BBSR. Second, we conduct a computational study to compare many existing American option price approximation techniques. Methods are compared on the basis of the speed of computation and the accuracy of the approximation.

Our computational results show that our LUBA approximation, which uses both lower and upper bound information, has a root mean squared (RMS) relative error of 0.02% on a sample that represents a wide range of option parameters. This RMS error is slightly better than the RMS error of a 1,000-step binomial tree. Furthermore, the LUBA approximation can be computed as fast as a 50-step binomial tree (or about 500 times faster than a 1,000-step binomial tree). Our LBA and LUBA approximations are not dominated in terms of speed and accuracy by any of the other methods that we tested. Furthermore, these two approximations are sufficiently simple that they can be computed in a spreadsheet.

The valuation of American options on dividend-paying assets is an important problem in financial economics. Early work focused on the case of discrete dividends for which analytical solutions can be derived [Geske (1979), Roll (1977), and Whaley (1981)]. When closed form solutions cannot be derived, numerical methods have been employed to compute the value of the option and the optimal exercise boundary. Brennan and Schwartz (1977, 1978) and Schwartz (1977) introduced finite difference methods and Cox, Ross, and Rubinstein (1979) introduced the binomial method for the valuation of American options. These methods discretize both the time and state spaces in order to approximate the option price. The methods are very easy to implement and are quite flexible in that they can be easily adapted to price many nonstandard or exotic options. Analysis and comparison of these early methods is given in Geske and Shastri (1985). Convergence of the Brennan and Schwartz method is proved in Jaillet, Lamberton, and Lapeyre (1990). Convergence of the binomial method for pricing American options is proved in Amin and Khanna (1994).

Generalizations of the binomial approach include the multinomial methods of Boyle (1988), Omberg (1988), Boyle, Evnine, and Gibbs (1989), and Kamrad and Ritchken (1991). Quasi-analytical solutions were introduced by Barone-Adesi and Whaley (1987), Geske and Johnson (1984), and MacMillan (1986). The Geske and Johnson method gives an exact analytical solution for the American option pricing problem, but their formula is an infinite series that can only be evaluated approximately by numerical methods. The quadratic method of Barone-Adesi and Whaley (1987) and MacMillan (1986) and the method of lines of Carr and Faguet (1994) are based on exact solutions to approximations of the option partial differential equation. In the method of lines the time derivative is replaced by a finite difference approximation. Geske and Johnson (1984) introduced the method of Richardson extrapolation to the option pricing problem. Richardson extrapolation has also been used in Breen (1991), Bunch and Johnson (1992), Carr and Faguet (1994), and Yu (1993). The accelerated binomial method of Breen (1991) can be viewed as a method of approximating the Geske and Johnson (1984) option pricing formula.

Kim (1990) and McKean (1965) provide an integral representation of the option price [see also Carr, Jarrow, and Myneni (1992), Jacka (1991), and Yu (1993)[1]]. Their integral formulas express the value of the American option as the value of the corresponding European option augmented by the present value of the gains from early exercise. The gains from early exercise, in turn, depend parametrically on the optimal exercise boundary, which is the solution of a nonlinear integral equation subject to a boundary condition. While the option price has an explicit representation, the exercise boundary is implicitly defined by the integral equation so that a recursive numerical procedure is required to solve for the exercise boundary and option price.

In the next section of this article we derive a lower bound for the American call option price based on a capped option with an appropriately chosen constant cap. In Section 2 we provide a procedure, based on the same class of capped options, to compute a uniform lower bound, denoted L^*, on the optimal exercise boundary of the American call option. In Section 3 we use the integral representation of the early exercise premium in conjunction with L^* to obtain an upper bound for the theoretical price of the option. Modifications of the procedures for American put options are given at the end of Section 3. Numerical results and comparisons with existing methods are given in Section 4. Concluding remarks are given in Section 5. Proofs

[1] A similar integral representation of the value function which arises in a class of stopping time problems is derived in El Karoui and Karatzas (1991).

are collected in Appendix A. Some details of the implementation of various methods are given in Appendix B.

1. A Lower Bound on the American Call Option Value

We consider an American call option with maturity T and exercise price K that is written on an underlying asset whose price S satisfies

$$dS_t = S_t[(r-\delta)dt + \sigma\, dW_t], \tag{1}$$

where W_t is a standard Brownian motion process. Here r is the rate of interest, δ is the dividend rate, and σ is the volatility coefficient of the asset price, which are all taken to be constant. Throughout the article, we assume $\delta > 0$, unless otherwise noted. The asset price process [Equation (1)] is represented in its risk-neutral form. Let $C_t(S_t)$ denote the value of the American call option, where the parameters r, δ, σ, K, and T are omitted for brevity. The optimal exercise policy can be described by a nonnegative continuous function of time which we denote by B^*. The optimal policy corresponding to B^* is to exercise at the first time $s < T$ such that $S_s = B^*_s$ or at maturity if $S_T \geq K$.

The main tool used in approximating the American call option value is a *capped call option* written on the same asset. If the price of the underlying asset is S, the payoff of a capped call option is $\max(\min(S, L) - K, 0)$, where K is the strike price and L is the cap. The payoff is the same as a standard option, except that the cap L limits the maximum possible payoff. The value of a capped call option with maturity date T, exercise price K, and constant cap L, with automatic exercise when the underlying asset price hits the cap L, is given by

$$\begin{aligned}
C_t(S_t, L) = {} & (L-K)\left[\lambda_t^{2\phi/\sigma^2} N(d_0)) + \lambda_t^{2\alpha/\sigma^2} N(d_0 + 2f\sqrt{T-t}/\sigma)\right] \\
& + S_t e^{-\delta(T-t)}[N(d_1^-(L) - \sigma\sqrt{T-t}) - N(d_1^-(K) - \sigma\sqrt{T-t})] \\
& - \lambda_t^{-2(r-\delta)/\sigma^2} L e^{-\delta(T-t)} \\
& \times [N(d_1^+(L) - \sigma\sqrt{T-t}) - N(d_1^+(K) - \sigma\sqrt{T-t})] \\
& - K e^{-r(T-t)} \Big[N(d_1^-(L)) - N(d_1^-(K)) \\
& \qquad - \lambda_t^{1-2(r-\delta)/\sigma^2}[N(d_1^+(L)) - N(d_1^+(K))]\Big],
\end{aligned} \tag{2}$$

where

$$d_0 = \frac{1}{\sigma\sqrt{T-t}}[\log(\lambda_t) - f(T-t)], \tag{3}$$

$$d_1^{\pm}(x) = \frac{1}{\sigma\sqrt{T-t}}[\pm \log(\lambda_t) - \log(L) + \log(x) + b(T-t)], \quad (4)$$

$$b = \delta - r + \tfrac{1}{2}\sigma^2, \quad f = \sqrt{b^2 + 2r\sigma^2}, \quad \phi = \tfrac{1}{2}(b-f),$$
$$\alpha = \tfrac{1}{2}(b+f), \quad \text{and} \quad \lambda_t = S_t/L. \quad (5)$$

The preceding formula for $C_t(S_t, L)$ holds for $L \geq \max(S_t, K)$. For completeness, we define $C_t(S_t, L) = \max(\min(S_t, L) - K, 0)$ when $L < \max(S_t, K)$.[2] In Equation (2), $N(\cdot)$ denotes the cumulative standard normal distribution function. Although Equation (2) is long, it is nearly as easy to compute as the Black and Scholes formula [Black and Scholes (1973)]. Indeed, Equation (2) is simple enough to implement in a spreadsheet or hand calculator. Note that Equation (2) holds only for constant caps L, *not* for arbitrary exercise boundaries.

The preceding formula for $C_t(S_t, L)$ gives an immediate lower bound on the value of the American call option $C_t(S_t)$. Since the policy of exercising when the asset price reaches the constant cap L is an admissible policy for the American option, $C_t(S_t, L) \leq C_t(S_t)$ for any L. Hence a lower bound is still obtained after optimizing over L. That is, $\max_L C_t(S_t, L) \leq C_t(S_t)$. Note that the maximum is achieved for some $L < \infty$ as long as $\delta > 0$. Define the optimal solution $\hat{L}(S_t)$ by

$$\hat{L}(S_t) = \operatorname*{argmax}_{L \geq S_t} C_t(S_t, L). \quad (6)$$

Thus

$$C_t^l(S_t) \equiv \max_L C_t(S_t, L) \leq C_t(S_t). \quad (7)$$

The lower bound in Equation (7) clearly improves over the European call option value, denoted $c_t(S_t)$.[3] That is, $C_t^l(S_t) > c_t(S_t)$ for $\delta > 0$, since $c_t(S_t) = \lim_{L\uparrow\infty} C_t(S_t, L)$. The lower bound also improves over the immediate exercise value. This follows by taking $L = S_t$, which gives

$$\max(S_t - K, 0) = C_t(S_t, S_t) \leq C_t^l(S_t).$$

The determination of $\hat{L}(S_t)$ is a simple univariate differentiable optimization problem for any given S_t.[4] This problem can be solved

[2] Equation (2) is derived in Broadie and Detemple (1995). Option formulas are also available for capped options with caps that grow at a constant rate. See, for example, Bjerksund and Stensland (1992), Broadie and Detemple (1995), Chesney (1989), or Omberg (1987).

[3] The European call value is $c_t(S_t) = S_t e^{-\delta(T-t)} N(d(K)) - K e^{-r(T-t)} N(d(K) - \sigma\sqrt{T-t})$, where $d(K) = [\log(S_t/K) + (r - \delta + \frac{1}{2}\sigma^2)(T-t)]/(\sigma\sqrt{T-t})$.

[4] A potentially better lower bound could be obtained by optimizing overcaps with a constant

by any number of methods, from a simple line search to more sophisticated algorithms that use derivative information. The derivative $\partial C_t(S_t, L)/\partial L$ is given in Proposition 2 in Appendix A. Derivative information is also used to determine a lower bound for the optimal exercise boundary, as described next.

2. A Lower Bound for the Optimal Exercise Boundary

The procedure relies heavily on the derivative of the capped call option value with respect to the constant cap L, evaluated as S_t approaches L from below:

$$D(L, t) \equiv \lim_{S_t \uparrow L} \frac{\partial C_t(S_t, L)}{\partial L}. \tag{8}$$

Expressions for $\partial C_t(S_t, L)/\partial L$ and $D(L, t)$ are given in Proposition 2 in Appendix A. Denote by L_t^* the solution to the equation

$$D(L, t) = 0. \tag{9}$$

Note that Equation (9) does *not* have to be solved recursively. That is, Equation (9) can be solved for L_t^* without having first solved for L_s^* for $s \in (t, T]$. Equation (9) represents a simple zero-finding problem that can be solved easily, for example, using Newton's method. Derivative information is often useful in these problems, so $\partial D(L, t)/\partial L$ is given in Proposition 2 in Appendix A.

The idea behind the boundary L^* is described next. Suppose one wishes to approximate B_t^* at some fixed time t, without using a recursive procedure. For fixed S_t^1 (which we'll assume is below B_t^*), $\hat{L}(S_t^1)$ is one way to approximate B_t^*. The exercise boundary $\hat{L}(S_t^1)$ can be thought of as the single constant exercise boundary that best approximates B^* in the interval $[t, T]$. Since B_s^* is a decreasing function of s, $B_t^* \geq \hat{L}(S_t^1) \geq B_T^*$, and $B_s^* = \hat{L}(S_t^1)$ for some $t \leq s \leq T$. One difficulty is that $\hat{L}(S_t^1)$ is probably not a good approximation to B^* at time t. However, $\hat{L}(S_t^1)$ is a function of the initial asset price S_t^1. Choosing a new asset price $S_t^2 = \hat{L}(S_t^1)$ leads to a new constant exercise boundary $\hat{L}(S_t^2)$. Note that $\hat{L}(S_t^2) \geq \hat{L}(S_t^1)$ and $B_t^* \geq \hat{L}(S_t^2) \geq B_T^*$. This process can be repeated until the iterates $\hat{L}(S_t^i)$ converge to some L_t^*. Since the iterates form an increasing sequence that is bounded above by

growth rate. In this case, the cap function can be specified by two parameters, for example, the starting point and the growth rate. However, because the cap is convex and the optimal exercise boundary for call options is concave, the improvement in the bound does not appear to be worth the additional effort and complexity of a two-dimensional optimization.

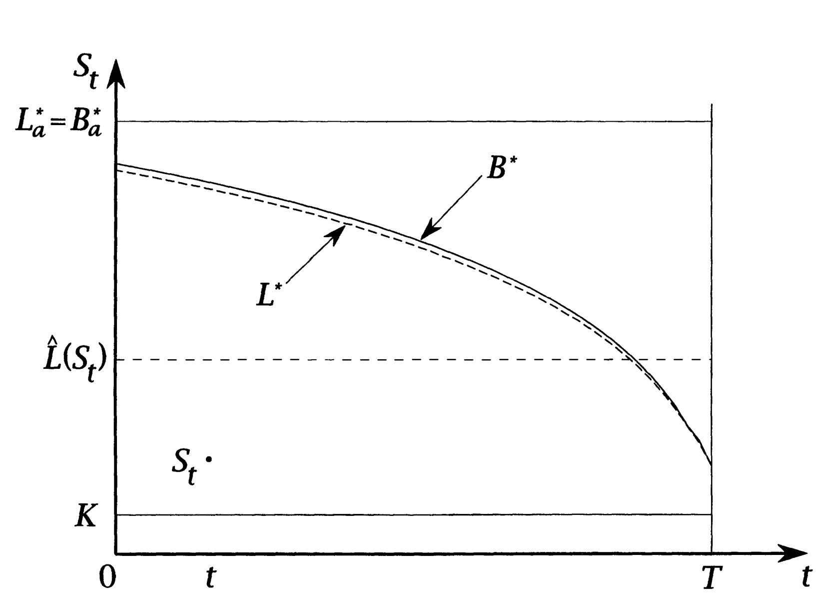

Figure 1
Illustration of B^*, L^*, and $\hat{L}(S_t)$
The solid line B^* is the optimal exercise boundary. The dashed line L^* is the approximate exercise boundary obtained by solving Equation (9). The horizontal dashed line $\hat{L}(S_t) = \operatorname{argmax}_{L \geq S_t} C_t(S_t, L)$ is the optimal constant exercise boundary at time t corresponding to asset price S_t. The asymptotic values of the boundaries, $B_a^* = \lim_{T-t\uparrow\infty} B_t^*$ and $L_a^* = \lim_{T-t\uparrow\infty} L_t^*$, coincide and are equal to the optimal exercise boundary for the corresponding perpetual American option.

B_t^*, each successive iterate is closer to B_t^*. The limiting value L_t^* can be obtained directly by solving Equation (9), that is, the intermediate iterates $\hat{L}(S_t^i)$ never have to be computed.

The relationship between B^*, L^*, and $\hat{L}(S_t)$ is illustrated in Figure 1. At maturity the optimal exercise boundary B^* and the approximate boundary L^* coincide. Let $B_a^* = \lim_{T-t\uparrow\infty} B_t^*$ and $L_a^* = \lim_{T-t\uparrow\infty} L_t^*$. These are the asymptotic values of the boundaries. The boundaries also coincide for very long times to maturity, that is $B_a^* = L_a^*$. Figure 1 illustrates $B_t^* \geq \hat{L}(S_t) \geq B_T^*$ for $S_t \leq B_t^*$. The next theorem summarizes this comparison of the exercise boundary L^* to the optimal boundary B^*.

Theorem 1. *Let B_t^* denote the optimal exercise boundary for the American call option. Let L_t^* denote the exercise boundary given by the solution to Equation (9). Then*

$$(i) \quad L_t^* \leq B_t^*$$

$$(ii) \quad \lim_{T-t\downarrow 0} L_t^* = \max\left(\frac{r}{\delta}K, K\right)$$

$$(iii) \quad \lim_{T-t\uparrow\infty} L_t^* = \frac{b+f}{b+f-\sigma^2}K$$

where $b \equiv \delta - r + \frac{1}{2}\sigma^2$ *and* $f \equiv \sqrt{b^2 + 2r\sigma^2}$ *are defined as before.*

Theorem 1 part (i) says that the approximate exercise boundary L_t^* lies uniformly below the optimal exercise boundary B_t^*. Parts (ii) and (iii) show that $L_t^* \to B_t^*$ in two limiting cases. Since $B_T^* = \max((r/\delta)K, K)$ [see, e.g., Kim (1990)], part (ii) shows that $L_T^* = B_T^*$. Similarly, since $B_t^* \to K(b+f)/(b+f-\sigma^2)$ as $T-t\uparrow\infty$ [again, see Kim (1990)], part (iii) shows that $L_t^* \to B_t^*$ as $T-t\uparrow\infty$.

3. An Upper Bound on the American Call Option Value

Consider the class of contracts consisting of a European call option and a sure flow of payments that are paid at the rate

$$\delta S_t e^{-(\delta-r)(s-t)} N(d_2(S_t, B_s, s-t)) - rK\,N(d_3(S_t, B_s, s-t)) \quad (10)$$

for $s \in [t, T]$, where

$$d_2(S_t, B_s, s-t) = \frac{1}{\sigma\sqrt{s-t}} \times [\log(S_t/B_s) + (r-\delta+\tfrac{1}{2}\sigma^2)(s-t)], \quad (11)$$

$$d_3(S_t, B_s, s-t) = d_2(S_t, B_s, s-t) - \sigma\sqrt{s-t}, \quad (12)$$

and B_s is a nonnegative continuous function of time. Each member of the class of contracts is parametrized by B. The value of the contract at time t is

$$V_t(S_t, B) = c_t(S_t) + \int_{s=t}^{T} [\delta S_t e^{-\delta(s-t)} N(d_2(S_t, B_s, s-t)) - rKe^{-r(s-t)} N(d_3(S_t, B_s, s-t))]ds, \quad (13)$$

where $c_t(S_t)$ denotes the value at time t of a European call option on S with strike price K and maturity T.

The importance of this class of contracts was shown in Carr et al. (1992) and Kim (1990). The optimal exercise boundary for the American call option is obtained by solving

$$V_s(B_s^*, B^*) = B_s^* - K \quad (14)$$

for B_s^* for all $s \in [t, T]$. Equation (14) is often referred to as the *value* *matching* condition. The value of the American call option $C_t(S_t)$ is then given by $V_t(S_t, B^*)$.

Equation (13), subject to the boundary condition of Equation (14), can be numerically approximated by a computationally intensive recursive procedure described in Appendix B. We use Equation (13) in conjunction with L^*, the *lower bound* on the optimal exercise boundary, to obtain an *upper bound* on the theoretical value of an American call option.

Theorem 2. *Let L^* denote the lower bound on the optimal exercise boundary given by the solution to Equation (9). The value of the American call option $C_t(S_t)$ is bounded above by the quantity $C_t^u(S_t) \equiv V_t(S_t, L^*)$. That is,*

$$C_t(S_t) \leq C_t^u(S_t). \tag{15}$$

In practice, the upper bound $C_t^u(S_t)$ is computed by approximating L^* at n discrete points in the time interval $[t, T]$. The points are typically equally spaced throughout the time interval. The intermediate points on the approximate L^* boundary are determined by linear interpolation. Finally, $C_t^u(S_t)$ is computed from Equation (13) taking $B = L^*$. Thus, computing $C_t^u(S_t)$ requires solving Equation (9) n times (to approximate L^*) and performing one numerical integration. Each of the steps can be done very quickly. In practice, small values of n, for example, n between 4 and 10, lead to good upper bounds.

The next proposition characterizes the behavior of the bounds in four limiting cases. It says that the upper and lower bounds become tight for options approaching maturity, for long dated options, for deep out-of-the-money options, and deep in-the-money options. It also says that the bounds become tight for extremely low and high volatilities, for large dividend rates, and for large interest rates.

Proposition 1. *The difference between the upper and lower call option bounds approaches zero, that is,*

$$C_t^u(S_t) - C_t^l(S_t) \downarrow 0,$$

when, holding all other parameters fixed, either (i) $T-t \downarrow 0$, (ii) $T-t \uparrow \infty$, (iii) $S_t \downarrow 0$, (iv) $S_t \uparrow \infty$, (v) $\sigma \downarrow 0$, (vi) $\sigma \uparrow \infty$, (vii) $\delta \uparrow \infty$, or (viii) $r \uparrow \infty$.

3.1 From bounds to approximations

The bounds in Equations (7) and (15) are used to compute two approximations to the American call option value. The approximations are

$$C_t^1(S_t) = \hat{\lambda}_1 C_t^l(S_t) \text{ and}$$
$$C_t^2(S_t) = \hat{\lambda}_2 C_t^l(S_t) + (1 - \hat{\lambda}_2) C_t^u(S_t)$$

for weights $\hat{\lambda}_1 \geq 1$ and $0 \leq \hat{\lambda}_2 \leq 1$. We use the "hat" notation to distinguish the true values of λ_1 and λ_2 defined by $C_t(S_t) = \lambda_1 C_t^l(S_t)$ and $C_t(S_t) = \lambda_2 C_t^l(S_t) + (1 - \lambda_2) C_t^u(S_t)$. For convenience, we refer to the approximation based on the lower bound, $C_t^1(S_t)$, as LBA. Similarly, we sometimes refer to the approximation based on the lower and upper bounds, $C_t^2(S_t)$, as LUBA.

The simple choice of weights $\hat{\lambda}_1 = 1$ and $\hat{\lambda}_2 = 0.5$ usually leads to good approximations. For example, in a large sample of options, we never found a value of λ_1 greater than 1.0133. That is, the lower bound was always within 1.31% of the true option value. However, the original option parameters, together with information obtained in the computation of the lower and upper bounds, can be effectively utilized to quickly compute better weights. We use a weighted regression approach, described in Appendix B, to determine $\hat{\lambda}_1$ and $\hat{\lambda}_2$. Regression techniques have been used in special cases of the American option pricing problem in Johnson (1983) and Kim (1994).

The quality of our bounds and approximations is investigated in Section 4. Next we show the modifications necessary to bound and approximate theoretical American put option values.

3.2 Modifications for American put options

The bounds and approximations for call options can be adapted for put options. Each of the formulas and procedures used for call options could be rederived for put options. For example, corresponding to the capped call option formula is a similar capped put option formula. However, a put-call symmetry result for American options, which holds in the geometric Brownian motion setting, can be used to avoid this additional effort. McDonald and Schroder (1990) show that the value of an American call option with parameters S_t, K, r, δ, and T is related to the value of an American put option by

$$C_t(S_t, K, r, \delta, T) = P_t(K, S_t, \delta, r, T) \tag{16}$$

[see also Chesney and Gibson (1995)]. That is, an American put price equals the American call price with the identification of parameters: $S_t \to K$, $K \to S_t$, $r \to \delta$, and $\delta \to r$.

The intuition for Equation (16) rests on the duality between the underlying asset and cash. A call option gives the right to exchange cash for the asset, while a put option gives the right to exchange the asset for cash. The symmetry result can also be seen as a variation of the international put-call equivalence of Grabbe (1983). The symmetry result means that any American call option pricing routine can be used to price American put options with a simple substitution of parameters.

4. Computational Results

In this section we compare several American option pricing methods on the basis of the speed of computation and the accuracy of the results over a wide range of option parameters. While speed and accuracy are primary concerns of researchers and practitioners, other factors can also be important in an option pricing method. These factors include the economic insights offered by the method, the simplicity of implementation, the ease of adaptability to other types of options, the availability of derivative information, etc.

The speed and accuracy requirements of a pricing method depend on the intended application. A trader wishing to price a single option requires a computation speed on the order of 1 second. However, dealers or large trading desks may need to price thousands of options on an hourly basis. Higher accuracy is always better, but not if economically insignificant price improvements are obtained at an unacceptable cost in terms of computation time. A simple measure of economic significance is the tick size (i.e., minimum price fluctuation) of a contract. For example, some option contracts have tick sizes of 1/8 of a point while others are as little as 1 cent. Generally, option prices are on the order of \$10 (some are less than \$1, but few are over \$100), so accuracy *on the order* of 0.1% (1 cent in \$10) is desirable but clearly not essential in all applications.

In this section we test several existing methods for computing American option prices. We test the binomial method of Cox, Ross, and Rubinstein (1979), the version of the trinomial method described in Kamrad and Ritchken (1991), the quadratic approximation of MacMillan (1986) and Barone-Adesi and Whaley (1987), the two-point Geske and Johnson (1984) method, the modified two-point Geske and Johnson method of Bunch and Johnson (1992), the accelerated binomial method of Breen (1991), the method of lines (ML) from Carr and Faguet (1994), and the integral method of Kim (1990). We test the two approximations proposed in this article, LBA and LUBA, and also two simple modifications of the binomial method. The first modification is the same as the binomial method, except that at the time step just before option maturity, the Black and Scholes formula replaces the usual "continuation value." This method is termed BBS (for *b*inomial with a *B*lack and *S*choles modification). The second modification adds Richardson extrapolation to the BBS method, and we refer to it as the

BBSR method. Unlike the accelerated binomial, which extrapolates based on the number of exercise opportunities, BBSR bases its extrapolation directly on the number of time steps. For algorithms that require the calculation of the cumulative normal distribution function, we use the approximation suggested by Moro (1995). Details of the implementation of several of the methods, including data structures and pseudocode, are given in Appendix B.

Since true American option values are unknown, how can numerical approximation methods be compared? We solve this problem by taking a convergent method and computing option values to an error that is an order of magnitude less than the error in the methods we are trying to compare. For our results, we use the convergent binomial method [see Amin and Khanna (1994)] with $n = 15{,}000$ as the basis for comparison. That is, we take values generated by this method to be the "true" option values. Hence, the "errors" that we report would not change significantly if we knew the exact option values.

In order to get a preliminary flavor of the results, Tables 1 and 2 give American option values for several methods. The results are given for call options, but the American put-call symmetry of McDonald and Schroder (1990) implies identical results for puts after a renaming of parameters. In particular, the call option results for $r = 0$ and $\delta = 0.07$ can be more naturally thought of as put option results for $\delta = 0$ and $r = 0.07$. The results in Tables 1 and 2 suggest that the lower bound approximation (LBA), the lower and upper bound approximation (LUBA), and the binomial method with $n = 300$ give fairly accurate results. The accuracy of the quadratic approximation degrades for longer maturity options, consistent with the finding in Barone-Adesi and Whaley (1987). The modified Geske and Johnson two-point method appears to be more accurate than the original Geske and Johnson two-point method. This finding is consistent with Bunch and Johnson (1992).

The 40 options in Tables 1 and 2 do not represent a large enough sample to draw any firm conclusions about the methods. The tables do not give summary information about errors, nor information about computational speed. More thorough and systematic results concerning the speed-accuracy trade-off of various American option pricing methods are given in Figures 2 through 8. These figures are based on average results from nearly 2,500 options determined from a random distribution of parameters. The probability distribution of call option parameters is described next.

We chose a distribution of parameters that is a reasonable reflection of options that are of interest to academics and practitioners. Volatility, denoted σ, is distributed uniformly between 0.1 and 0.6. Time to maturity is, with probability 0.75, uniform between 0.1 and 1.0 years

Table 1
American option value bounds and approximations (maturity $T = 0.5$ years)

Option param	Asset price	Lower bound	Upper bound	LBA	LUBA	Binom	Accel binom	GJ 2-pt	GJ 2-pt modif	Quad approx	Method of lines	True value
$r = 0.03$,	80	0.218	0.220	0.219	0.220	0.220	0.218	0.215	0.219	0.230*	0.224	0.219
$\sigma = 0.20$,	90	1.376	1.389	1.382	1.386	1.389	1.389	1.363*	1.382	1.405*	1.373*	1.386
$\delta = 0.07$	100	4.750	4.792	4.771*	4.782	4.780	4.763*	4.769*	4.786	4.782	4.779	4.783
	110	11.049	11.125	11.090	11.098	11.098	11.125*	11.212*	11.255*	11.041*	11.105	11.098
	120	20.000	20.061	20.000	20.000	20.000	20.003	20.029	20.000	20.000	20.072*	20.000
$r = 0.03$,	80	2.676	2.691	2.689	2.689	2.690	2.689	2.661*	2.683	2.711*	2.668*	2.689
$\sigma = 0.40$,	90	5.694	5.727	5.721	5.723	5.725	5.727	5.676*	5.716	5.742*	5.715	5.722
$\delta = 0.07$	100	10.190	10.250	10.235	10.240	10.233	10.225	10.198*	10.235	10.242	10.241	10.239
	110	16.110	16.201	16.176	16.182	16.180	16.151	16.195	16.211	16.152	16.187	16.181
	120	23.271	23.392	23.356	23.357	23.366	23.343	23.477*	23.477*	23.288*	23.370	23.360
$r = 0.00$,	80	1.029	1.039	1.036	1.037	1.036	1.034	1.015*	1.032	1.062*	1.026*	1.037
$\sigma = 0.30$,	90	3.098	3.129	3.118	3.123	3.127	3.123	3.077*	3.115	3.147*	3.107*	3.123
$\delta = 0.07$	100	6.985	7.051	7.027	7.035	7.032	6.999*	7.018*	7.042	7.028	7.029	7.035
	110	12.882	12.988	12.952	12.953	12.955	12.929*	13.064*	13.064*	12.886*	12.957	12.955
	120	20.650	20.779	20.743	20.721	20.719	20.826*	20.934*	20.438*	20.607*	20.750	20.717
$r = 0.07$,	80	1.664	1.664	1.664	1.664	1.662	1.662	1.664	1.664	1.665	1.663	1.664
$\sigma = 0.30$,	90	4.495	4.495	4.495	4.495	4.499	4.499	4.495	4.495	4.495	4.500	4.495
$\delta = 0.03$	100	9.251	9.251	9.251	9.251	9.244	9.244	9.251	9.251	9.251	9.284*	9.251
	110	15.798	15.798	15.798	15.798	15.796	15.796	15.798	15.798	15.799	15.845*	15.798
	120	23.706	23.706	23.706	23.706	23.710	23.710	23.706	23.706	23.709	23.774*	23.706

All options have $K = 100$.
Binomial and accelerated binomial methods are based on $n = 300$ time steps.
The upper bound is based on a discretization with $n = 200$.
The method of lines is based on $n = 3$ time steps.
The "true value" column is based on the binomial method with $n = 15{,}000$ time steps.
*Relative error > 0.2% and absolute error > 0.01.

Table 2
American option value bounds and approximations (maturity T = 3 years)

Option param	Asset price	Lower bound	Upper bound	LBA	LUBA	Binom	Accel binom	GJ 2-pt	GJ 2-pt modif	Quad approx	Method of lines	True value
$r = 0.03$,	80	2.553	2.589	2.572	2.580	2.584	2.560*	2.515*	2.560*	2.711*	2.547*	2.580
$\sigma = 0.20$,	90	5.121	5.187	5.155*	5.168	5.172	5.123*	5.164	5.164	5.301*	5.149*	5.167
$\delta = 0.07$	100	9.002	9.103	9.053	9.065	9.062	9.052	9.216*	9.216*	9.154*	9.056	9.066
	110	14.371	14.504	14.432	14.444	14.447	14.636*	14.741*	15.077*	14.444	14.449	14.443
	120	21.354	21.506	21.409	21.412	21.415	21.762*	21.621*	20.000*	21.336*	21.422	21.414
$r = 0.03$,	80	11.238	11.354	11.315	11.327	11.334	11.267*	11.275*	11.353*	11.625*	11.278*	11.326
$\sigma = 0.40$,	90	15.609	15.763	15.706	15.724	15.734	15.631*	15.787*	15.797*	16.028*	15.683*	15.722
$\delta = 0.07$	100	20.656	20.850	20.770	20.793	20.787	20.670*	21.029*	21.029*	21.084*	20.752	20.793
	110	26.337	26.569	26.459	26.489	26.503	26.465	26.939*	26.939*	26.749*	26.464	26.495
	120	32.607	32.876	32.730	32.772	32.792	32.841	33.448*	33.569*	32.982*	32.756	32.781
$r = 0.00$,	80	5.463	5.540	5.510	5.520	5.523	5.457*	5.459*	5.495*	5.658*	5.483*	5.518
$\sigma = 0.30$,	90	8.766	8.879	8.834	8.843	8.846	8.785*	8.889*	8.889*	8.947*	8.824*	8.842
$\delta = 0.07$	100	13.048	13.199	13.135	13.142	13.137	13.168	13.346*	13.346*	13.177*	13.130	13.142
	110	18.347	18.535	18.447	18.453	18.458	18.610*	18.800*	19.126*	18.394*	18.460	18.453
	120	24.685	24.903	24.788	24.797	24.786	25.206*	25.173*	25.266*	24.638*	24.806	24.791
$r = 0.07$,	80	12.145	12.145	12.167	12.145	12.158	12.162	12.137	12.146	12.282*	12.137	12.145
$\sigma = 0.30$,	90	17.367	17.368	17.397	17.368	17.381	17.390	17.355	17.372	17.553*	17.391	17.369
$\delta = 0.03$	100	23.347	23.349	23.383	23.349	23.341	23.355	23.331	23.356	23.586*	23.380	23.348
	110	29.961	29.964	30.001	29.964	29.983	29.999	29.946	29.976	30.259*	30.009	29.964
	120	37.099	37.104	37.142	37.104	37.100	37.113	37.091	37.126	37.459*	37.146	37.104

All options have $K = 100$.
Binomial and accelerated binomial methods are based on $n = 300$ time steps.
The upper bound is based on a discretization with $n = 200$.
The method of lines is based on $n = 3$ time steps.
The "true value" column is based on the binomial method with $n = 15{,}000$ time steps.
*Relative error > 0.2% and absolute error > 0.01.

and, with probability 0.25, uniform between 1.0 and 5.0 years. We fix the strike price at $K = 100$ and take the initial asset price $S \equiv S_0$ to be uniform between 70 and 130. Relative errors do not change if S and K are scaled by the same factor, that is, only the ratio S/K is of interest. The dividend rate δ is uniform between 0.0 and 0.10. The riskless rate r is, with probability 0.8, uniform between 0.0 and 0.10, and with probability 0.2, equal to 0.0. By American put-call symmetry, the roles of r and δ and the roles of S and K are reversed between puts and calls. Hence, when we price call options with this distribution of parameters, we are also pricing put options with a similar distribution. In particular, the put option dividends are, with probability 0.8, uniform between 0.0 and 0.10, and with probability 0.2, equal to 0.0. Each parameter is selected independently of the others.

The main error measure that we report is root mean squared (RMS) relative error. RMS error is defined by

$$\text{RMS} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} e_i^2}\,, \quad \text{where} \quad e_i = \frac{\hat{C}_i - C_i}{C_i}$$

is the relative error, C_i is the "true" option value (estimated by a 15,000-step binomial tree), and $\hat{C}_i$ is the estimated option value.[5] To make relative error meaningful, the summation is taken over options in the data set satisfying $C_i \geq 0.50$. Out of the 2,500 options, 2,271 satisfied this criterion. For option values less than 50 cents, the RMS absolute error measure yielded qualitatively similar results.

Computation speed is measured in option prices calculated per second. The exact hardware is inconsequential, since only relative speeds matter.[6] Care was taken to "tune" the methods as best as possible. That is, many methods have several choices that affect the speed-accuracy trade-off. For example, to implement a method that requires the solution of a nonlinear equation, the programmer must select a solution algorithm and must set iteration and/or tolerance parameters. Similar choices are required if the method requires one or more numerical integrations. Even in the simpler methods, significant computation time can be saved by eliminating redundant or unnecessary computations. Some methods take advantage of the computation of a critical stock

[5] The RMS error criterion seems to be very reasonable. Mean absolute error does not penalize large errors enough. Maximum absolute error penalizes large errors too much. Even so, we obtained similar qualitative results when we used the mean absolute relative error and maximum relative error measures.

[6] The results were computed on a 25-MHz 68040 NeXTstation. The methods were all compared using the same compiler settings.

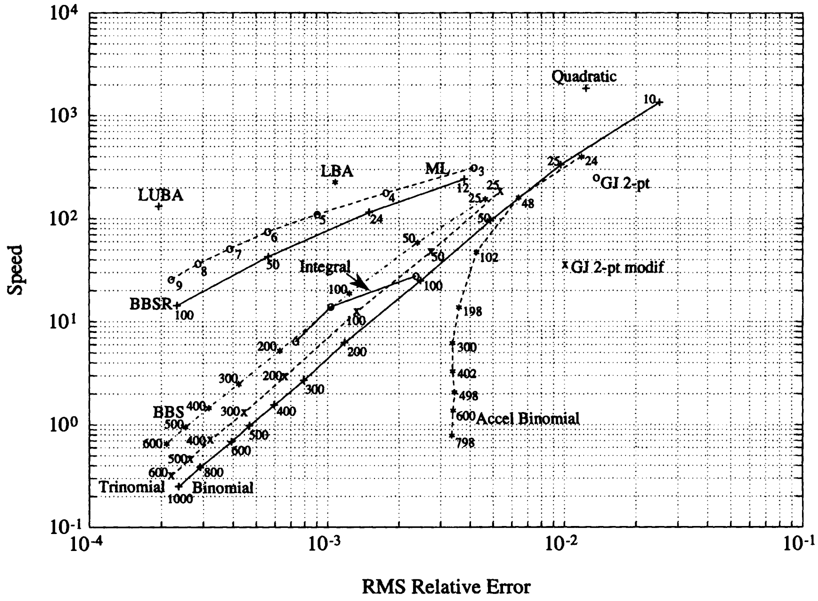

Figure 2
Speed-accuracy trade-off for all observations with true price ≥ 0.50

RMS relative error is defined by $\sqrt{\frac{1}{m}\sum_{i=1}^{m}((\hat{C}_i - C_i)/C_i)^2}$, where C_i is the "true" option value (estimated by a 15,000 step binomial tree), $\hat{C}_i$ is the approximate option value estimated by the corresponding numerical method, and $m = 2,271$ is the number of options satisfying $C_i \geq 0.50$. Speed is measured in option prices calculated per second (on a 25-MHz 68040 NeXTstation). Numbers next to some of the methods indicate the number of time steps. Preferred methods are in the upper-left corner.

price or boundary. We priced options at five stock values for a given set of other parameters.

The overall results are given in Figure 2. Because of the extreme differences in speed and accuracy, the results are plotted on a log-log scale. Numbers next to the methods indicate the number of time steps. (These numbers are identical in the later graphs, but are not repeated for clarity of presentation.) The integral method results are based on the discretizations 4, 8, and 16, in order of decreasing error and speed. Figures 3 and 4 break the results down by option maturity, Figures 5 and 6 by S/K (the "moneyness" of the option), and Figures 7 and 8 by option volatility.

Two independent samples of option parameters, both drawn from the distribution described earlier, were used in the computational study. The first sample was used to estimate the $\hat{\lambda}_1$ and $\hat{\lambda}_2$ functions, while the results in Figures 2 through 8 were computed with the second sample. This eliminates the bias that would result from using the same sample to estimate and test the approximations. However,

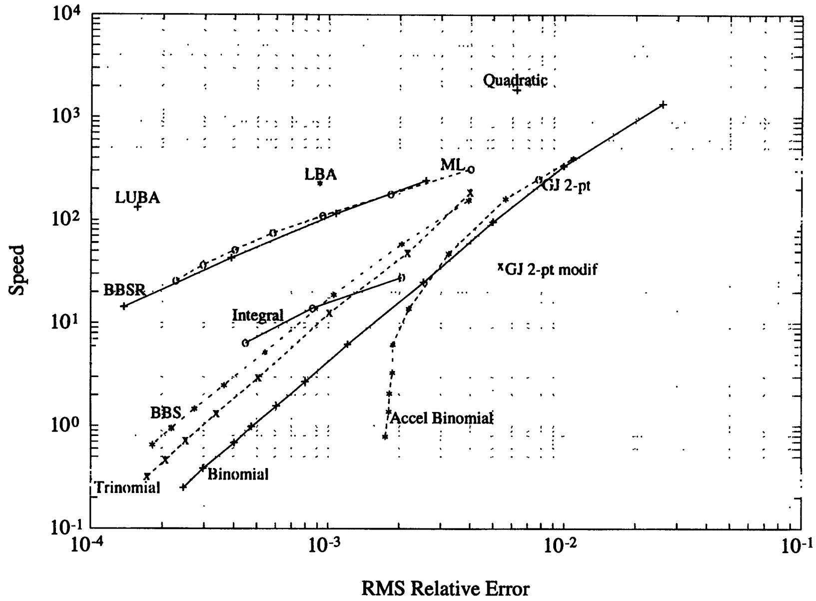

Figure 3
Speed-accuracy trade-off for short maturity options ($T < 1.0$) and true price ≥ 0.50

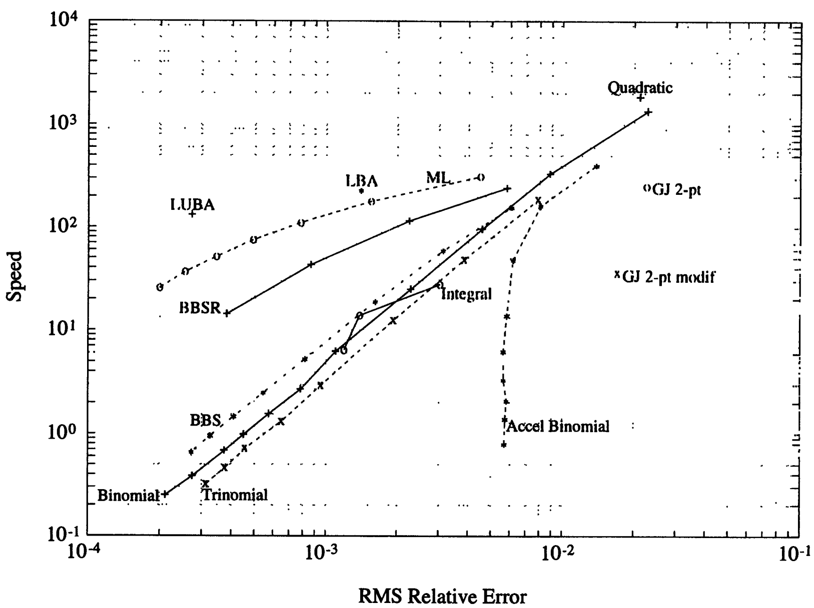

Figure 4
Speed-accuracy trade-off for long maturity options ($T \geq 1.0$) and true price ≥ 0.50

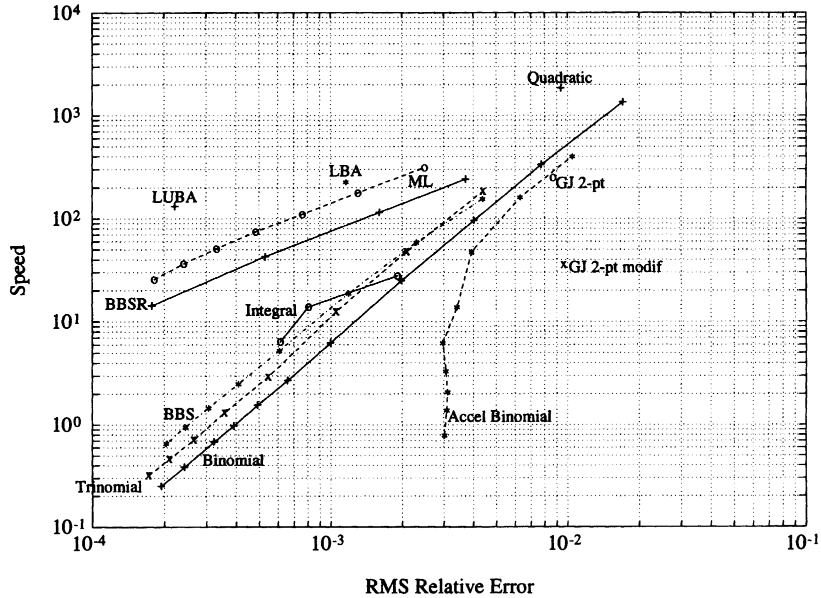

Figure 5
Speed-accuracy trade-off for at-the-money options ($0.9 < S/K < 1.1$) and true price ≥ 0.50

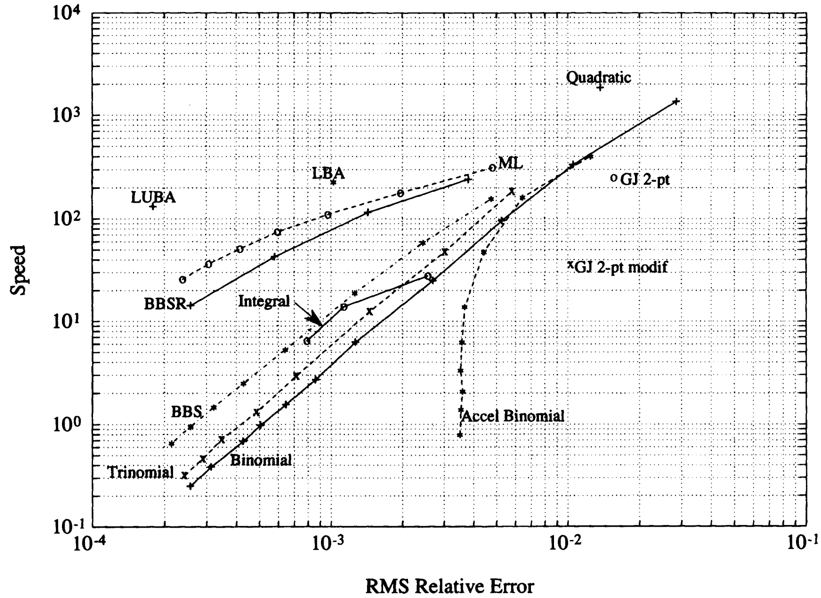

Figure 6
Speed-accuracy trade-off for in- and out-of-the-money options ($S/K \leq 0.9$ or $S/K \geq 1.1$) and true price ≥ 0.50

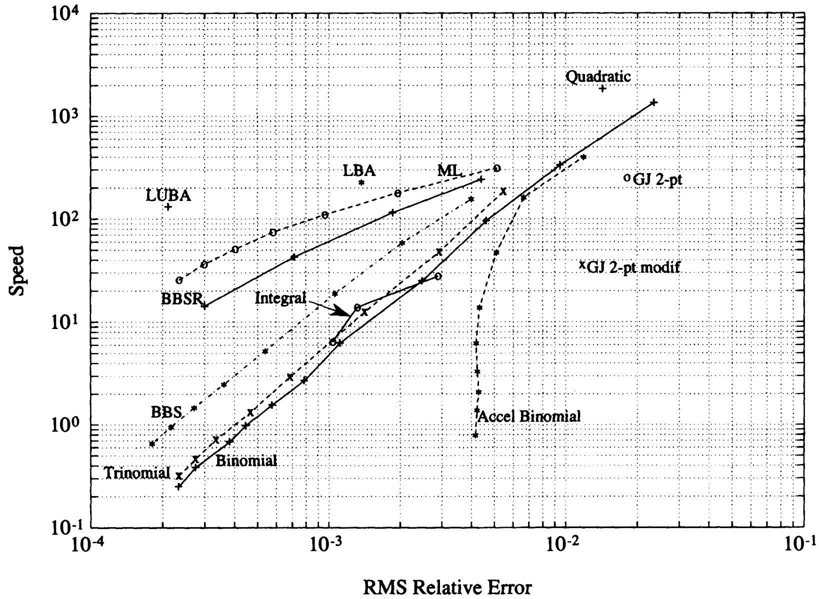

Figure 7
Speed-accuracy trade-off for low volatility options (σ < 0.35) and true price $\geq$ 0.50

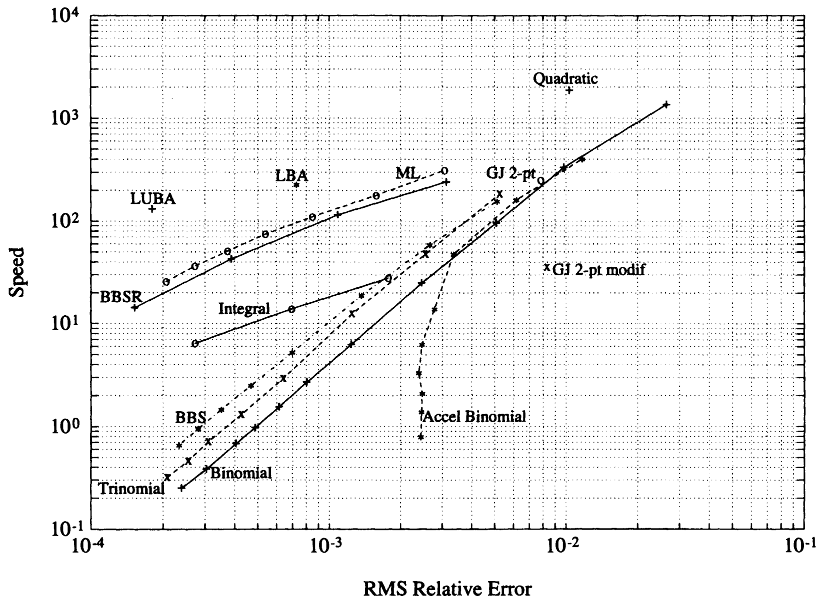

Figure 8
Speed-accuracy trade-off for high volatility options ($\sigma \geq$ 0.35) and true price $\geq$ 0.50

the results may still be influenced by the choice of the distribution of option parameters. A comparison of the results for various subsets of option parameters provides an indication of the sensitivity of the results to this choice. Figures 3 through 8 show that the performance of the LBA and LUBA approximations is fairly consistent across these subsets. Also, Proposition 1 shows that the LBA and LUBA approximations converge to the true American option value for extreme parameter values. These observations suggest that the LBA and LUBA results are not extremely sensitive to the particular distribution of option parameters. In reality, practitioners could face a different distribution of option parameters. In this case, the functions $\hat{\lambda}_1$ and $\hat{\lambda}_2$ could be reestimated to achieve the best performance.

4.1 Discussion of results

The binomial method is striking in its elegance and simplicity and is very useful because it is a convergent method. Computation time with the binomial method is quadratic in the number of time steps. The binomial method plots as a nearly straight line in Figures 2 through 8, that is, the binomial error decreases linearly with the number of time steps. As a result, to get one extra decimal point in RMS accuracy using the binomial method requires increasing the number of time steps by a factor of 10, which results in computation time increasing by a factor of 100.

The quadratic method is by far the fastest method, with an RMS error of about 0.6% for options with less than 1 year maturity. The two-point Geske and Johnson methods are dominated by the binomial method. The American option formula given in Geske and Johnson (1984) is an *exact* representation of the option value in terms of an infinite series. Evaluation of n^{th} order terms requires the computation of n-dimensional cumulative normals. The two-point Geske and Johnson methods require only the evaluation of bivariate cumulative normals, which is very reasonable in terms of speed. However, two exercise points do not capture enough of the early exercise opportunities of American options to give high accuracy.

The accelerated binomial curve in Figure 2 requires explanation. As the number of time steps increases, the accelerated binomial converges to the three-point Geske and Johnson approximation, not to the American option value. To have convergence to the American option value, both the number of time steps (n) *and* the number of exercise points (m) must increase to infinity. In Figure 2, large values of n, with m fixed at 3, lead to an RMS error of about 0.3%. The accelerated binomial approximation is faster to compute but less accurate than the binomial for each n (with $m = 3$). Surprisingly, the binomial method dominates the accelerated binomial in the overall speed-accuracy trade-off.

We tested a variation of the binomial method that is common among practitioners. In this variation, the result of the binomial method with n time steps is averaged with the $n+1$ time step result. The idea is to take advantage of the well-known oscillatory convergence of the binomial. In the speed-accuracy figures, this variation plots almost directly on top of the binomial method. That is, this variation used with a given value of n and $n+1$ has the same speed and accuracy as the original binomial method with a larger value of n. In other words, this binomial variation has little to recommend. We also tested several other variations of the binomial method, including Amin (1991), Trigeorgis (1991), and the binomial method with the parameter choices described in Hull and White (1988), and the original parameters in Cox, Ross, and Rubinstein (1979). *These methods were all essentially identical in the speed-accuracy trade-off.* Convergence properties and additional variations of the binomial method are investigated in Leisen and Reimer (1995).

The trinomial method slightly edges out the binomial method, except for long-maturity options. Likewise, the BBS method described earlier is slightly better than the trinomial method. The binomial Black and Scholes method with Richardson extrapolation (BBSR) is significantly better than the other binomial-type methods. Indeed, the BBSR method for $n = 100$ (which extrapolates prices from the $n = 50$ and $n = 100$ BBS results) is about as accurate as the binomial method with $n = 1,000$ and is about 55 times faster. The oscillatory convergence of the binomial is dampened by the use of the Black and Scholes formula at the penultimate time step. (See Figures 14 and 15 in Appendix B for an illustration of this point.) Two-point Richardson extrapolation is then able to significantly improve the option price estimates. Higher order Richardson extrapolation does worse than two-point extrapolation in this application.

Our implementation of the integral method appears to be competitive with the binomial method. Because the integral method requires equation solving and numerical integration, there are many choices that affect the speed-accuracy trade-off. Yu (1993) implemented the integral method with a less accurate but quicker step function approximation for the integrals. Our implementation is slower and more accurate, and it is not clear which is the better choice.

The method of lines (ML) is a slight improvement over the BBSR method. The method of lines also uses Richardson extrapolation to improve its estimates, and this technique is responsible in a large part for the good results obtained by this method. Although not shown in the figures, the binomial method dominates the method of lines when

Richardson extrapolation is not used. The convergence rates of the ML and BBSR methods appear to be faster than the binomial method, as shown by the slopes of the lines in Figure 2.

The two approximations developed in this article, LBA and LUBA, are undominated in the overall speed-accuracy trade-off in Figure 2. The LUBA method has an accuracy comparable to a 1,000 time step binomial tree and a speed comparable to a 50 time step tree. This represents an average error of about 0.02% and a computation speed on the order of 100 options per second (on a 25-MHz 68040 CPU or comparable 486-based PC).

Although these results focused on option prices, a similar analysis could be done for the "Greeks," that is, the partial derivatives of the option prices with respect to the option parameters. For all of the methods tested, there are straightforward ways to quickly compute derivatives once the prices have been computed.

5. Conclusion

The theoretical values of many European options can be computed by evaluating a simple formula. The computation of theoretical American option values is considerably more difficult because of the possibility of optimal early exercise. In this article, lower and upper bounds on the theoretical American option value were developed. These bounds were shown to become tight for extreme values of the parameters.

Based on the bounds, we developed two option value approximations. LBA, the approximation based on the lower bound, has an RMS error of about 0.1% on a large range of option parameters, which is comparable to a 200-step binomial tree. LUBA, the approximation based on the lower and upper bound, has an RMS error of 0.02%, which is comparable to a 1,000-step binomial tree. Both methods are more complicated to implement than the binomial method. However, they are simple enough that they can be directly implemented in today's spreadsheets. One drawback of the methods is that they are not convergent, that is, there is no parameter than can be increased to give arbitrarily high accuracy. The bounds could be improved, but the resulting algorithm would likely resemble the integral equation approach.

We compared many existing American option approximation techniques based on speed and accuracy. The binomial method has stood the test of time for its combinations of speed and accuracy. In addition, the binomial method is valuable for its simplicity, elegance, and adaptability to other options. Among the other methods tested, *the main results are that the LBA, LUBA, ML, and BBSR approximations are all significant improvements over existing methods*. Among these four methods, the BBSR method introduced in this article is the simplest to program. The ML method slightly improves over the BBSR method and is a quite promising approach. The LUBA method introduces a new idea to option price approximation and it is the only method that also gives tight bounds on the option price.

In principle, the LBA and LUBA methodology developed in this article based on capped option values can be used to obtain bounds for other American-style contracts. However, the quality of the bounds and approximations for other contracts remains to be investigated.

Appendix A

Proof of Theorem 1. (i) Fix time t. Without loss of generality, consider the case where $\delta < r$. In this case, $B_T^* = (r/\delta)K$. (The case $\delta \geq r$ with $B_T^* = K$ is similar.) Consider some arbitrary asset price $S_t^1 \leq (r/\delta)K$. The value of the capped option with constant cap L is $C_t(S_t^1, L)$. Maximizing the value of the option with the constraint $L \geq S_t^1$ yields the first-order condition

$$\frac{\partial C_t(S_t^1, L)}{\partial L} = 0$$

for $L > S_t^1$ or $\partial C_t(S_t^1, L)/\partial L \leq 0$ for $L = S_t^1$. The first-order condition admits a solution $\hat{L}_t(S_t^1)$ such that $S_t^1 \leq \hat{L}_t(S_t^1) \leq B_t^*$. The fact that $\hat{L}_t(S_t^1)$ is bounded above by B_t^* follows from Lemma 1 below. Indeed Lemma 1 implies that for constant boundaries L^1 and L^2 such that $L^2 > L^1 \equiv B_t^*$ we have $C_t(S_t, L^2) < C_t(S_t, L^1)$. The optimal strategy, if one is restricted to a constant exercise barrier, will necessarily lie below B_t^*. Now set $S_t^2 = \hat{L}_t(S_t^1)$ and repeat the procedure, that is, select the cap $\hat{L}_t(S_t^2)$ that maximizes the capped option value when the asset price is S_t^2. Clearly, $S_t^2 \leq \hat{L}_t(S_t^2)$, for otherwise value is lost. (The exercise value in the case $S_t^2 > \hat{L}_t(S_t^2)$ would be $\hat{L}_t(S_t^2) - K$ which is less than $S_t^2 - K$.) By Lemma 1, $\hat{L}_t(S_t^2) \leq B_t^*$. Following this procedure we construct an increasing sequence $\hat{L}_t(S^n)$ which is bounded above and therefore converges to a limit L_t^*. Hence inequality (i) follows.

(ii) As $t \uparrow T$ clearly $L_t^* \to \max((r/\delta)K, K)$.

(iii) Using the analytic expression for $D(L, t)$ given in Proposition 2, it can be shown that

$$D(L, t) \to 1 - \frac{(b+f)}{\sigma^2}(1 - K/L)$$

as $T - t \uparrow \infty$. The result follows by solving Equation (9) for that case. ■

Lemma 1. *Suppose that L^1 and L^2 are any continuous time-dependent boundaries satisfying $L_s^2 > L_s^1 \geq B_s^*$ for all $s \in [t, T]$. Then $C_t(S_t, L^2) < C_t(S_t, L^1)$.*

Proof of Lemma 1. Let E_t denote the expectation at time t under the risk-neutral probability measure. Denote the first time that S hits L^i by τ_i, for $i = 1, 2$. Let the operator x^+ denote $\max(x, 0)$. Then

$$\begin{aligned}
C_t(S_t, L^2) &= E_t[e^{-r(\tau_2-t)}(L_{\tau_2}^2 - K)1_{\{\tau_2<T\}}] \\
&\quad + E_t[e^{-r(T-t)}(S_T - K)^+ 1_{\{\tau_2 \geq T\}}] \\
&= E_t\Big\{e^{-r(\tau_1-t)}1_{\{\tau_1<T\}}E_{\tau_1}[e^{-r(\tau_2-\tau_1)}(L_{\tau_2}^2 - K)1_{\{\tau_2<T\}} \\
&\qquad + e^{-r(T-\tau_1)}(S_T - K)^+ 1_{\{\tau_2 \geq T\}}]\Big\} \\
&\quad + E_t[e^{-r(T-t)}(S_T - K)^+ 1_{\{\tau_1 \geq T\}}1_{\{\tau_2 \geq T\}}] \\
&< E_t[e^{-r(\tau_1-t)}(L_{\tau_1}^1 - K)1_{\{\tau_1<T\}}] \\
&\quad + E_t[e^{-r(T-t)}(S_T - K)^+ 1_{\{\tau_1 \geq T\}}] \\
&= C_t(S_t, L^1).
\end{aligned}$$

The first equality follows from the risk-neutral representation of the option value with deterministic cap L^2. The second equality follows from the law of iterated expectations and from the fact that $1_{\{\tau_1<T\}}$ is measurable relative to information at time τ_1. The next inequality follows from the fact that at time τ_1 for $S_{\tau_1} = L_{\tau_1}^1 \geq B_{\tau_1}^*$ immediate exercise dominates any waiting strategy. Thus, $L_{\tau_1}^1 - K > C_{\tau_1}(S_{\tau_1}, L^2)$. The second term on the right-hand side of the inequality also makes use of the relationship $1_{\{\tau_1 \geq T\}}1_{\{\tau_2 \geq T\}} = 1_{\{\tau_1 \geq T\}}$. The last equality follows from the risk-neutral representation of the option value with deterministic cap L^1. ■

Proof of Theorem 2. Consider the class of contracts whose value at time t is given by

$$V_t(S_t, B) = c_t(S_t) + \int_{s=t}^{T} \Phi_t(B_s, S_t, s)ds$$

where

$$\begin{aligned}
\Phi_t(B_s, S_t, s) &= \delta S_t e^{-\delta(s-t)}N(d_2(S_t, B_s, s-t)) \\
&\quad - rKe^{-r(s-t)}N(d_3(S_t, B_s, s-t)).
\end{aligned}$$

The functions d_2 and d_3 are defined in Equations (11) and (12), respectively, and B_s is a continuous function. For each s, consider $\Phi_t(x_s, S_t, s) : \mathbb{R}^+ \mapsto \mathbb{R}$ as a function of x_s. It can be verified that $\Phi_t(x_s, S_t, s)$ is single peaked with global maximum at $x_s = (r/\delta)K$, strictly decreasing for $x_s \in [(r/\delta)K, \infty)$, and satisfies $\lim_{x_s \uparrow \infty} \Phi_t(x_s, S_t, s) = 0$. Recall that the theoretical value of the American option is $V_t(S_t, B^*)$, where B^* solves Equation (14). Since $B_s^* \geq L_s^* \geq (r/\delta)K$ an upper bound for the American option value is obtained by pointwise maximization of the function $\Phi_t(x_s, S_t, s)$ over the set $x_s \in [L_s^*, \infty)$:

$$V_t(S_t, B^*) \leq c_t(S_t) + \int_{s=t}^{T} \max_{x_s \geq L_s^*} \Phi_t(x_s, S_t, s) ds.$$

By the monotonicity property of the function $\Phi_t(x_s, S_t, s)$ for $x_s \geq (r/\delta)K$, the solution to the pointwise maximization problem is $x_s = L_s^*$. It follows that

$$\begin{aligned} V_t(S_t, B^*) &\leq c_t(S_t) + \int_{s=t}^{T} \Phi_t(L_s^*, S_t, s) ds \\ &\equiv C_t^u(S_t). \end{aligned}$$

■

Proof of Proposition 1. A sketch of the proof for each of the cases is provided. Details of each step can be checked directly using the functional forms for each quantity.

(i) As $T - t \downarrow 0$, $L_t^* \to B_T^*$. Also $C_t^u(S_t) \to c_t(S_t)$. Note that $c_t(S_t) \to \max(S_t - K, 0)$ as $T - t \downarrow 0$. Similarly, $C_t^l(S_t) \to c_t(S_t)$ as $T - t \downarrow 0$. Combining these straightforward results gives $C_t^u(S_t) - C_t^l(S_t) \downarrow 0$ as $T - t \downarrow 0$.

(ii) For a perpetual call option, the optimal exercise boundary is $B_t^* = K(b+f)/(b+f-\sigma^2)$. For any t and any S_t, the optimal solution to $\max_L C_t(S_t, L)$ is achieved at $\hat{L}(S_t) = K(b+f)/(b+f-\sigma^2)$. Since $\hat{L}(S_t)$ is independent of S_t, it follows that $L^* = \hat{L}(S_t)$ and therefore $L^* = B^*$ and $C_t^u(S_t) = C_t(S_t, L^*) = C_t(S_t) = C_t^l(S_t)$. Hence, $C_t^u(S_t) - C_t^l(S_t) \downarrow 0$ as $T - t \uparrow \infty$.

(iii) As $S_t \downarrow 0$, both $C_t^u(S_t) \to c_t(S_t)$ and $C_t^l(S_t) \to c_t(S_t)$. (Also $c_t(S_t) \to 0$ as $S_t \downarrow 0$.) Hence, $C_t^u(S_t) - C_t^l(S_t) \downarrow 0$ as $S_t \downarrow 0$.

(iv) As $S_t \uparrow \infty$, both $C_t^u(S_t) \to S_t - K$ and $C_t^l(S_t) \to S_t - K$. Hence, $C_t^u(S_t) - C_t^l(S_t) \downarrow 0$ as $S_t \uparrow \infty$.

(v) For $\sigma \downarrow 0$, consider two cases: (a) $\delta > r$ and (b) $\delta \leq r$. For case (a), the boundaries B^* and L^* approach the constant K as $\sigma \downarrow 0$. For $S_t \leq K$, $\hat{L} \to K$ and for $S_t > K$, $\hat{L} \to S_t$. Thus, $C_t^l(S_t) \to 0$ or $C_t^l(S_t) \to S_t - K$, respectively. Also, for $S_t \leq K$, $V(S_t, L^*) \to c_t(S_t) \to 0$. For $S_t > K$, $V(S_t, L^*) \to S_t - K$.

For case (b), the boundaries B^* and L^* approach the constant $(r/\delta)K$ as $\sigma \downarrow 0$. For $S_t \leq (r/\delta)K$, $\hat{L} \to (r/\delta)K$ and for $S_t > (r/\delta)K$, $\hat{L} \to S_t$. Now there are several subcases to consider, depending on the direction of the inequality between $rK/(\delta S_t)$ and $e^{-(\delta-r)(T-t)}$ and between K/S_t and $e^{-(\delta-r)(T-t)}$. For example, if $S_t < (r/\delta)K$ and $K/S_t > e^{-(\delta-r)(T-t)}$ (and hence $rK/(\delta S_t) > e^{-(\delta-r)(T-t)}$), then $C_t^l(S_t) \to 0$ and $V(S_t, L^*) \to 0$. In another subcase, if $S_t < (r/\delta)K$, $rK/(\delta S_t) > e^{-(\delta-r)(T-t)}$, and $K/S_t < e^{-(\delta-r)(T-t)}$, then $C_t^l(S_t)$ and $V(S_t, L^*)$ both approach $e^{-\delta(T-t)}S_t - Ke^{-r(T-t)}$. The other subcases follow similarly.

(vi) As $\sigma \uparrow \infty$, $C_t^l(S_t) \to (\hat{L} - K)S_t/\hat{L}$. Since $\hat{L} \to \infty$, $C_t^l(S_t) \to S_t$. Similarly, as $\sigma \uparrow \infty$,

$$\begin{aligned} V(S_t, L^*) &\to c_t(S_t) + \int_{s=t}^{T} \delta S_t e^{-\delta(s-t)} ds \\ &\to S_t e^{-\delta(T-t)} - S_t e^{-\delta(s-t)} \Big|_t^T = S_t. \end{aligned}$$

(vii) and (viii) As $\delta \uparrow \infty$ or $r \uparrow \infty$, both $C_t^u(S_t) \to 0$ and $C_t^l(S_t) \to 0$. ■

Proposition 2 gives explicit expressions for various partial derivatives of $C_t(S_t, L)$.

Proposition 2. *Let $\tau = T - t$ and $\lambda_t = S_t/L$. Suppose $L \geq \max(S_t, K)$. Let $b = \delta - r + \frac{1}{2}\sigma^2$ and $f = \sqrt{b^2 + 2r\sigma^2}$, as before. Then $\partial C_t(S_t, L)/\partial L$ can be written as*

$$\begin{aligned} \frac{\partial C_t(S_t, L)}{\partial L} &= \left[1 - \left(\frac{L-K}{L}\right)(2\phi/\sigma^2)\right]\lambda_t^{2\phi/\sigma^2} N(d_0) \\ &\quad + \left[1 - \left(\frac{L-K}{L}\right)(2\alpha/\sigma^2)\right]\lambda_t^{2\alpha/\sigma^2} N(d_0 + 2f\sqrt{\tau}/\sigma) \\ &\quad + e^{-\delta\tau}\frac{2(b-\sigma^2)}{\sigma^2}\lambda_t^{-2(r-\delta)/\sigma^2} \\ &\quad \times [N(d_1^+(L) - \sigma\sqrt{\tau}) - N(d_1^+(K) - \sigma\sqrt{\tau})] \\ &\quad - e^{-r\tau}\frac{2bK}{\sigma^2 L}\lambda_t^{2b/\sigma^2}[N(d_1^+(L)) - N(d_1^+(K))], \end{aligned}$$

$\partial C_t(S_t, L)/\partial S$ can be written as

$$\frac{\partial C_t(S_t, L)}{\partial S} = \frac{L-K}{L}[(2\phi/\sigma^2)\lambda_t^{2\phi/\sigma^2 - 1} N(d_0)$$

$$
\begin{aligned}
&+ (2\alpha/\sigma^2)\lambda_t^{2\alpha/\sigma^2-1} N(d_0 + 2f\sqrt{\tau}/\sigma)] \\
&+ e^{-\delta\tau}[N(d_1^-(L) - \sigma\sqrt{\tau}) - N(d_1^-(K) - \sigma\sqrt{\tau})] \\
&+ 2(r-\delta)/\sigma^2 e^{-\delta\tau}\lambda_t^{-1-2(r-\delta)/\sigma^2} \\
&\times [N(d_1^+(L) - \sigma\sqrt{\tau}) - N(d_1^+(K) - \sigma\sqrt{\tau})] \\
&+ \frac{K}{L} e^{-r\tau}\lambda_t^{-2(r-\delta)/\sigma^2}(1 - 2(r-\delta)/\sigma^2) \\
&\times [N(d_1^+(L)) - N(d_1^+(K))],
\end{aligned}
$$

$D(L, t)$ can be written as

$$
\begin{aligned}
D(L,t) = &\left[1 - \left(\frac{L-K}{L}\right)(2\phi/\sigma^2)\right] N(-f\sqrt{\tau}/\sigma) \\
&+ \left[1 - \left(\frac{L-K}{L}\right)(2\alpha/\sigma^2)\right] N(f\sqrt{\tau}/\sigma) \\
&+ e^{-\delta\tau}\frac{2(b-\sigma^2)}{\sigma^2}[N(d_1^+(L) - \sigma\sqrt{\tau}) - N(d_1^+(K) - \sigma\sqrt{\tau})] \\
&- e^{-r\tau}\frac{2bK}{\sigma^2 L}[N(d_1^+(L)) - N(d_1^+(K))],
\end{aligned}
$$

and $\partial D(L, t)/\partial L$ can be written as

$$
\begin{aligned}
\frac{\partial D(L,t)}{\partial L} = &-\frac{K}{L^2}[2\phi/\sigma^2 + (2f/\sigma^2)N(f\sqrt{\tau}/\sigma)] \\
&- 2e^{-\delta\tau} n(d_1^+(K) - \sigma\sqrt{\tau})/(L\sigma\sqrt{\tau}) \\
&+ e^{-r\tau}\frac{2bK}{\sigma^2 L^2}[N(d_1^+(L)) - N(d_1^+(K))].
\end{aligned}
$$

Proof of Proposition 2. The expression for $\partial C_t(S_t, L)/\partial L$ follows by taking the partial derivative of Equation (2) and simplifying. Identities used in the simplification include $\lambda_t^{2b/\sigma^2} n(d_1^+(L)) = n(d_1^-(L))$, $\lambda_t^{2(b-\sigma\sqrt{\tau})/\sigma^2} n(d_1^+(L) - \sigma\sqrt{\tau}) = n(d_1^-(L) - \sigma\sqrt{\tau})$, $n(d_1^-(L) - \sigma\sqrt{\tau})\lambda_t = n(d_1^-(L))e^{-(r-\delta)\tau}$, $n(d_1^+(K))e^{-(r-\delta)\tau}\lambda_t K/L = n(d_1^+(K) - \sigma\sqrt{\tau})$, $e^{-r\tau} n(d_1^-(L)) = n(d_0)\lambda_t^{2\phi/\sigma^2}$, and $e^{-r\tau} n(d_1^-(L)) = n(d_0 + 2f\sqrt{\tau}/\sigma)$ $\lambda_t^{2\alpha/\sigma^2}$. In the previous identities, $n(\cdot)$ denotes the density function of a standard normal random variable. The expression for $D(L, t)$ follows by taking the limit as S_t increases to L. The other expressions follow similarly using standard calculus. ■

Appendix B

In this appendix we provide details of the implementation of various American option pricing methods. Although the binomial method is easy to program, we begin with this method in order to present a particular implementation that is easily adapted to the accelerated binomial method and the trinomial method.

B.1 Binomial method

The binomial method was proposed in Cox, Ross, and Rubinstein (1979) [see also Rendleman and Bartter (1979)]. The parameters that we use for the binomial procedure are modified from Hull and White (1988, footnote 4) to account for dividends.

Because a binomial tree with n time steps has $O(n^2)$ nodes, the computation time increases as $O(n^2)$. Our implementation of the binomial method uses only $O(n)$ storage. It is not necessary to store the entire tree in memory; only information related to the current time step is required. Our implementation computes the stock price values $Su^j d^{n-j}$ recursively. This approach uses only multiplications and hence avoids the use of the more time-consuming power function. In addition, these $2n$ stock price values need only be computed *once*. The parameters p' and q' are the binomial up and down probabilities, respectively, deflated by the discount factor. Adjusting the probabilities initially means that discounting is done automatically at each node as part of the present value computation. This saves one multiplication at each node.

A pseudocode expression of our implementation is given in Figure 9. The inputs to the routine are the option parameters S, K, T, r, and δ, and the binomial time step parameter n. The output of the routine is the American call option value C. In order to clarify our routine, a small binomial tree indicating the indexing of time and stock price states is given in Figure 10. Our indexing scheme avoids the need for a separate temporary storage vector. Finally, note that this code will not be correct for extremely large values of δ (which lead to negative probabilities in the tree).

B.2 Accelerated binomial method

The accelerated binomial method was proposed in Breen (1991). The main "trick" to an efficient implementation of this method involves the computation of the binomial formula. We use a simple recursion to avoid redundant computations. As before, the tree parameters for this implementation are modified from Hull and White (1988, footnote 4) to account for dividends.

```
/* allocate space */
vectors v[j], s[j], for j = −n to n by 1;

/* initialize parameters */
Δt   = T/n; r_inv = e^{−rΔt}; a = e^{(r−δ)Δt}; b² = a²(e^{σ²Δt} − 1);
tmp  = a² + b² + 1; u = (tmp + √(tmp² − 4a²))/(2a); d = 1/u;
p    = (a − d)/(u − d); q = 1 − p; p′ = r_inv ∗ p;
q′   = r_inv ∗ q; s[0] = S;
for j = 1 to n by 1;
    s[j] = s[j − 1] ∗ u;
    s[−j] = s[−j + 1] ∗ d;
    end;

/* store option values at time index i = n */
v[j] = max(s[j] − K, 0), for j = −n to n by 2;

/* work backwards in time */
for i = n − 1 to 0 by −1;
    v[j] = max(p′ ∗ v[j + 1] + q′ ∗ v[j − 1], s[j] − K), for j = −i to i by 2;
    end;

/* return binomial option value */
C = v[0];
```

Figure 9
Binomial routine pseudocode
This routine estimates the price of an American call option with input parameters S, K, T, σ, r, and δ using n time steps.

The binomial formula involves terms of the form $b_j \equiv \binom{n}{j} p^{n-j} q^j$. If the term b_j has already been computed, then the next term b_{j+1} can be computed using the recursion

$$b_{j+1} \equiv \binom{n}{j+1} p^{n-j-1} q^{j+1} = b_j \frac{n-j+1}{j} \frac{q}{p}. \tag{17}$$

These binomial terms only need to be computed once.

Using the notation of Breen (1991), the accelerated binomial requires the computation of $P(1)$, $P(2)$, and $P(3)$. For brevity, we illustrate the computation of $P(3)$ only. Recall $P(3)$ is the option value allowing exercise at T, $2T/3$, and $T/3$ only. The pseudocode for our computation of $P(3)$ is given in Figure 11. The inputs to the routine are the option parameters S, K, T, r, and δ, and the time step parameter n. We assume that the routine is called with an integer n that is divisible by 6. The output of the routine is the value $P(3)$. The accelerated binomial value is given by the Richardson extrapolation

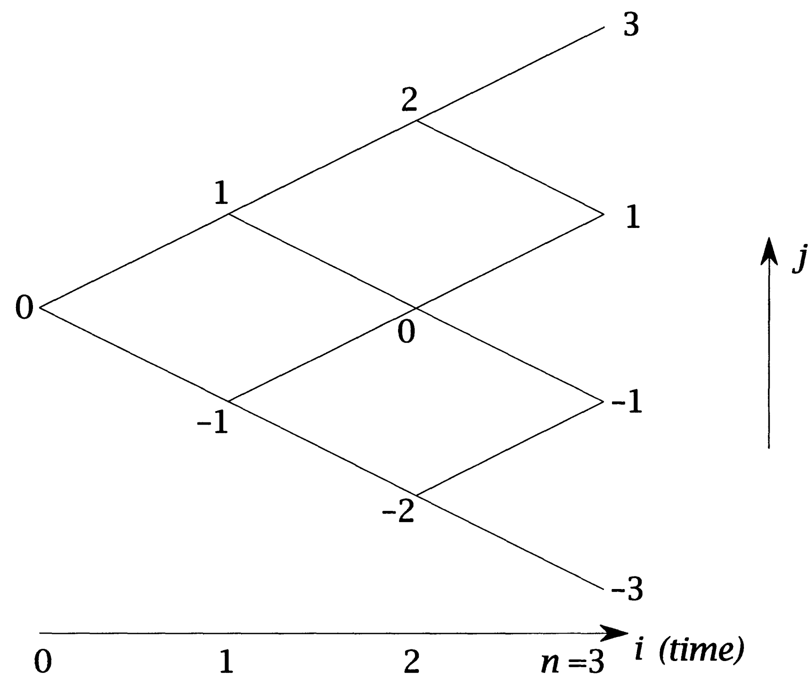

Figure 10
Illustration of the binomial tree for $n = 3$
Time is indexed $i = 0, \ldots, n$. At time i, asset prices are indexed by $j = -i, -i + 2, \ldots, i - 2, i$.

formula $C = P(3) + 3.5(P(3) - P(2)) - 0.5(P(2) - P(1))$. As before, the nodes of the tree are indexed as in Figure 10. [This leads to a slightly different indexing of the binomial terms in our routine below than indicated in Equation (17)].

Since the main computational effort in this routine involves multiplication, the work is easily shown to be $\sim 7/12n^2$.[7] The work in the binomial routine is $\sim n^2$ (2 multiplications at $n^2/2$ nodes). Thus the accelerated binomial is faster than the binomial routine for the same n.

B.3 Trinomial method

Trinomial methods have been proposed in Boyle (1988), Kamrad and Ritchken (1991), Omberg (1988), and Parkinson (1977). We test the

[7] The symbol $\sim$ means asymptotic to. The computation of $P(1)$ requires n multiplications, $P(2)$ requires $\sim n^2/4$ multiplications ($n/2$ multiplications at $n/2$ nodes), and $P(3)$ requires $\sim n^2/3$ multiplications ($n/3$ multiplications at n nodes). Thus computing C in the accelerated binomial requires $\sim n^2/4 + n^2/3$ multiplications.

```
/* allocate space */
vectors v[j], s[j], b[j], vtmp[j], for j = −n to n by 2;

/* initialize parameters */
Δt   = T/n;  r_inv = e^{−rT/3};  a = e^{(r−δ)Δt};  b^2 = a^2(e^{σ^2 Δt} − 1);
tmp  = a^2 + b^2 + 1;  u = (tmp + sqrt(tmp^2 − 4a^2))/(2a);  d = 1/u;
p    = (a − d)/(u − d);  q = 1 − p;  s[0] = S;
for  j = 2 to n by 2;
     s[j] = s[j − 2] * u^2;
     s[−j] = s[−j + 2] * d^2;
     end;

/* store option values at time index i = n */
v[j] = max(s[j] − K, 0), for j = −n to n by 2;

/* store binomial terms */
m = n/3; b[m] = p^m;
for  j = 1 to m by 1;
     k = m − 2j;
     b[k] = b[k + 2] * ((m − j + 1)/j) * (q/p);
     end;

/* evaluate at time index i = 2n/3 = 2m */
for  j = −2m to 2m by 2;
     vtmp[j] = sumproduct(b[2k − m], v[2k − m + j], k = 0 to m by 1);
     vtmp[j] = max(r_inv * vtmp[j], s[j] − K);
     end;
v[j] = vtmp[j], for j = −2m to 2m by 2;

/* evaluate at time index i = n/3 = m */
for  j = −m to m by 2;
     vtmp[j] = sumproduct(b[2k − m], v[2k − m + j], k = 0 to m by 1);
     vtmp[j] = max(r_inv * vtmp[j], s[j] − K);
     end;
v[j] = vtmp[j], for j = −m to m by 2;

/* evaluate at time i = 0 */
vtmp[0] = r_inv * sumproduct(b[k], v[k], k = −m to m by 2);

/* return P(3) value */
P(3) = vtmp[0];
```

Figure 11
Portion of the accelerated binomial pseudocode

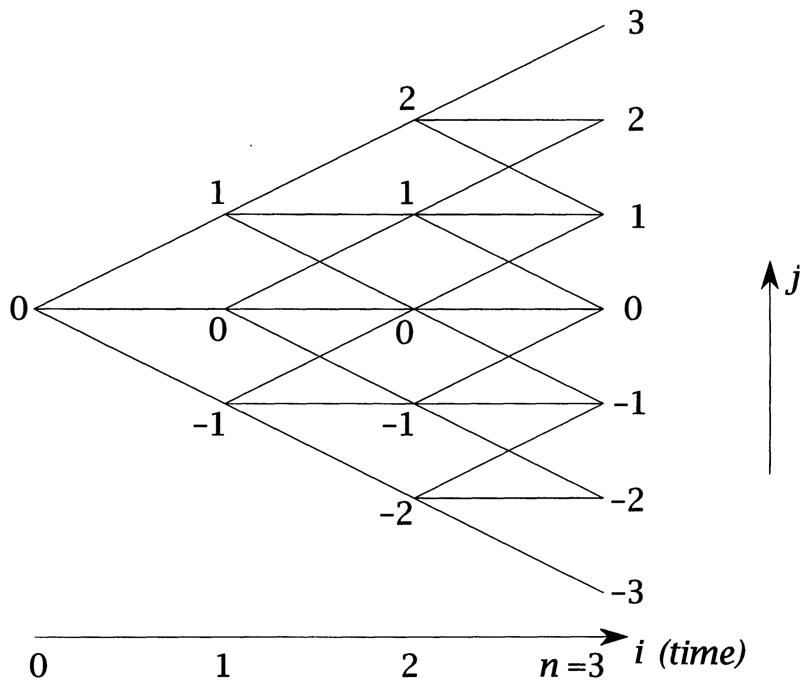

Figure 12
Illustration of the trinomial tree for $n = 3$
Time is indexed $i = 0, \ldots, n$. At time i, asset prices are indexed by $j = -i, -i+1, \ldots, i-1, i$.

Kamrad and Ritchken (1991) version. Our trinomial implementation follows easily from our binomial implementation. A small trinomial tree indicating the indexing of time and stock price states is given in Figure 12. The parameters p'_u, p'_m, and p'_d are the trinomial up, middle, and down probabilities, respectively, deflated by the discount factor.

A pseudocode expression of our implementation is given in Figure 13. The inputs to the routine are the option parameters S, K, T, r, and δ, the time step parameter n, and the trinomial parameter λ (which we set to $\sqrt{3/2}$). The output of the routine is the American call option value C. As before, note that this code will not be correct for extremely large values of δ (which lead to negative probabilities in the tree).

Since the main computational effort in this routine involves multiplication, the work is seen to be $\sim 3/2n^2$ (3 multiplications at $n^2/2$ nodes). This compares with $\sim n^2$ work for the binomial. So the computational work in the trinomial for large n should be comparable to the work in the binomial for $3/2n$. In Figure 2, the speed of the

```
/* allocate space */
vectors v[j], s[j], vtmp[j], for j = -n to n by 1;

/* initialize parameters */
λ    = √(3/2);  Δt = T/n;  r_inv = e^(-rΔt);  u = e^(λσ√Δt);
d    = 1/u;  μ = r - δ - ½σ²;  p_u = 1/(2λ²) + μ√Δt/(2λσ);
p_m  = 1 - 1/λ²;  p_d = 1 - p_u - p_d;  p'_u = r_inv * p_u;
p'_m = r_inv * p_m;  p'_d = r_inv * p_d;  s[0] = S;
for  j = 1 to n by 1;
     s[j] = s[j - 1] * u;
     s[-j] = s[-j + 1] * d;
     end;

/* store option values at time index i = n */
v[j] = max(s[j] - K, 0), for j = -n to n by 2;

/* work backwards in time */
for  i = n - 1 to 0 by -1;
     vtmp[j] = max(p'_u * v[j + 1] + p'_m * v[j] + p'_d * v[j - 1], s[j] - K),
                              for j = -i to i by 1;
     v[j] = vtmp[j], for j = -i to i by 1;
     end;

/* return trinomial option value */
C = v[0];
```

Figure 13
Trinomial routine pseudocode

trinomial for $n = 400$ is close to the binomial speed corresponding to $n = 600$.

B.4 Binomial Black and Scholes (BBS) method

The BBS method is identical to the binomial method, except that at the time step just before option maturity the Black and Scholes formula replaces the usual "continuation value." Evaluating the Black and Scholes formula involves more work than computing the continuation value (which involves two multiplications). However, this additional work is done only at n nodes, so the work of the BBS method is still $\sim n^2$. Figure 2 is consistent with this observation. For example, the speed of the binomial and BBS methods are nearly identical for $n = 600$.

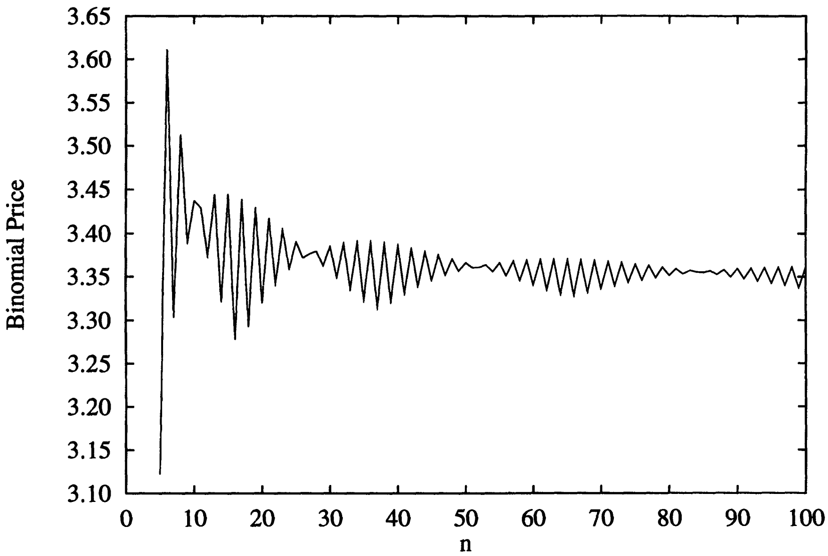

Figure 14
Binomial price versus number of time steps
The put option parameters are $S = 100$, $K = 90$, $r = 0.05$, $\delta = 0$, $\sigma = 0.30$, and $T = 0.5$. The true price is 3.345. The oscillatory convergence of the binomial is quite evident.

B.5 Binomial Black and Scholes method with Richardson extrapolation (BBSR)

The BBSR method adds two-point Richardson extrapolation to the BBS method. For example, the BBSR method computes the BBS price corresponding to $n = 25$ (say C_1) and $n = 50$ (say C_2) and then sets the approximate price to $C = 2C_2 - C_1$. This result is called the BBSR estimate for $n = 50$. For an extensive treatment of Richardson extrapolation see Marchuk and Shaidurov (1983).

To gain some intuition about the advantages of this approach, consider pricing a put option with parameters $S = 100$, $K = 90$, $r = 0.05$, $\delta = 0$, $\sigma = 0.30$, and $T = 0.5$. The true value of this option is 3.345. Figure 14 shows the binomial approximation as a function of the number of time steps n. Likewise, Figure 15 shows how the BBS approximation varies with the number of time steps. The oscillatory convergence of the binomial is quite evident in Figure 14. This effect is considerably dampened in Figure 15.

The effect of the oscillations on Richardson extrapolation is significant. For example, the binomial values corresponding to $n = 6$ and $n = 12$ are 3.611 and 3.374, respectively. Applying two-point Richardson extrapolation gives the estimate 3.136 (an error of -0.209) which is much worse than the binomial estimate at $n = 12$ (with an error of

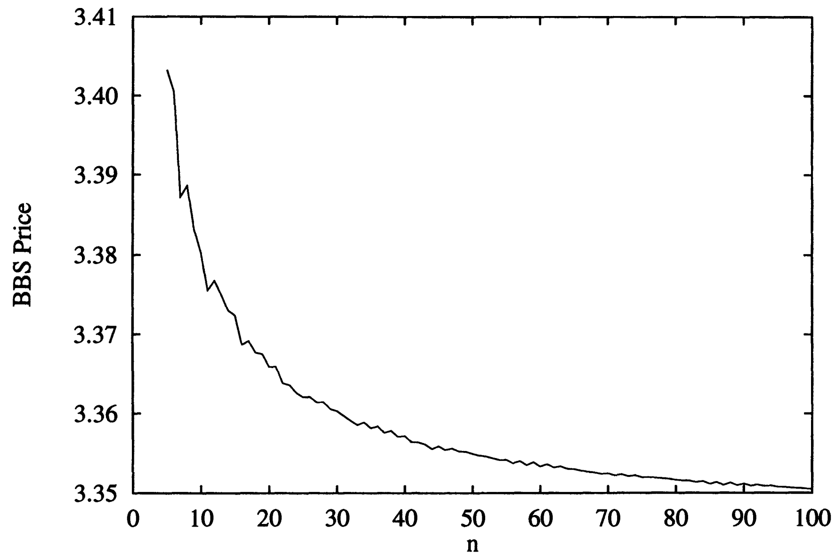

Figure 15
BBS price versus number of time steps
The put option parameters are $S = 100$, $K = 90$, $r = 0.05$, $\delta = 0$, $\sigma = 0.30$, and $T = 0.5$. The true price is 3.345. The convergence of the BBS method is considerably smoother compared to the binomial method.

0.028). The BBS estimates are 3.400 and 3.377 at $n = 6$ and $n = 12$, respectively. Two-point Richardson extrapolation gives the estimate 3.353 (an error of 0.008) which is much better than the BBS estimate at $n = 12$ (with an error of 0.031). Of course this example is merely illustrative. The results in Figures 2 through 8 indicate the improvement of the BBSR method over the BBS method.

B.6 LBA: Approximation based on the lower bound

Next we detail the approach that is used to convert the lower bound $C^l(S)$ to the option value approximation $C^1(S)$. The relationship between the bound and approximation is

$$C^1(S) = \hat{\lambda}_1 C^l(S),$$

where $\hat{\lambda}_1 \geq 1$ is a function of the option parameters S, K, T, r, and δ.

In order to define $\hat{\lambda}_1 = \hat{\lambda}_1(S, K, T, r, \delta)$, we first define some intermediate variables. Let $a \vee b \equiv \max(a, b)$ and $a \wedge b \equiv \min(a, b)$. Define $x_1 = T$, $x_2 = \sqrt{T}$, $x_3 = S/K$, $x_4 = r$, $x_5 = \delta$, $x_6 = \min(r/(\delta \vee 10^{-5}), 5)$, $x_7 = x_6^2$, $x_8 = (C^l(S) - c(S))/K$, $x_9 = x_8^2$, $x_{10} = C^l(S)/c(S)$. Recall

$c(S)$ denotes the European call option value. Then define y_1 by

$$\begin{aligned} y_1 = {} & 1.002 \times 10^{+0} - 1.485 \times 10^{-3} x_1 + 6.693 \times 10^{-3} x_2 \\ & - 1.451 \times 10^{-3} x_3 - 3.430 \times 10^{-2} x_4 + 6.301 \times 10^{-2} x_5 \\ & - 1.954 \times 10^{-3} x_6 + 2.740 \times 10^{-4} x_7 - 1.043 \times 10^{-1} x_8 \\ & + 5.077 \times 10^{-1} x_9 - 2.509 \times 10^{-3} x_{10}. \end{aligned}$$

Finally, define $\hat{\lambda}_1$ by

$$\hat{\lambda}_1 = \begin{cases} 1 & \text{if } C^l(S) = c(S) \text{ or } C^l(S) \leq S - K \\ \max(y_1 \wedge 1.0133, 1) & \text{otherwise.} \end{cases}$$

The coefficients in the formula for y_1 were determined from a regression on approximately 2,500 option values, where the option parameters were sampled from the distribution described in Section 4.

B.7 LUBA: Approximation based on the lower and upper bound

Next we detail the regression approach that is used to convert the lower bound $C^l(S)$ and upper bound $C^u(S)$ to the option value approximation $C^2(S)$. The relationship between the bounds and approximation is

$$C^2(S) = \hat{\lambda}_2 C^l(S) + (1 - \hat{\lambda}_2) C^u(S),$$

where $0 \leq \hat{\lambda}_2 \leq 1$ is a function of the option parameters S, K, T, r, and δ.

In order to define $\hat{\lambda}_2 = \hat{\lambda}_2(S, K, T, r, \delta)$, we first define some intermediate variables. Let $x_1 = T$, $x_2 = \sqrt{T}$, $x_3 = r$, $x_4 = \delta$, $x_5 = \min(r/(\delta \vee 10^{-5}), 5)$, $x_6 = x_5^2$, $x_7 = dC^l(S)/dS$, $x_8 = x_7^2$, $x_9 = (C^l(S) - c(S))/K$, $x_{10} = x_9^2$, $x_{11} = C^l(S)/c(S)$, $x_{12} = (C^u(S) - C^l(S))/K$, $x_{13} = C^u(S)/C^l(S)$, $x_{14} = S/L_0^*$, and $x_{15} = x_{14}^2$. Recall $dC^l(S)/dS$ is defined in Proposition 2. Then define y_2 by

$$\begin{aligned} y_2 = {} & 8.664 \times 10^{-1} - 7.668 \times 10^{-2} x_1 + 3.092 \times 10^{-1} x_2 \\ & - 3.356 \times 10^{-1} x_3 + 1.200 \times 10^{+0} x_4 - 3.507 \times 10^{-2} x_5 \\ & - 9.755 \times 10^{-2} x_6 - 7.208 \times 10^{-1} x_7 + 6.071 \times 10^{-1} x_8 \\ & + 7.379 \times 10^{+0} x_9 - 4.999 \times 10^{+1} x_{10} + 1.148 \times 10^{-1} x_{11} \\ & - 5.037 \times 10^{+1} x_{12} - 6.629 \times 10^{-1} x_{13} \\ & - 4.745 \times 10^{-1} x_{14} + 5.995 \times 10^{-1} x_{15}. \end{aligned}$$

Finally, define $\hat{\lambda}_2$ by

$$\hat{\lambda}_2 = \begin{cases} 1 & \text{if } C^l(S) = c(S) \text{ or } C^l(S) \leq S - K \\ \max(y_2 \wedge 1, 0) & \text{otherwise.} \end{cases}$$

The computation of the upper bound is computed approximating L^* at n discrete points in the time interval $[0, T]$. To compute $C^2(S)$, we use $n = 8$ in the computation of L^*. To compute the upper bound $C^u(S)$, we need to evaluate the integral in Equation (13). We do this using Simpson's rule with $n = 8$.[8] In this way, the evaluations of the function in the integral in Equation (13) coincide with the computed values of L^*.

The coefficients in the formula for y_2 were determined from a *weighted* regression, which is described next. Suppose we want to solve the optimization problem $\min \sum_i [(C_i^2 - C_i)/C_i]^2$. Applying the definitions gives $C_i^2 - C_i = (\lambda - \hat{\lambda})(C_i^u - C_i^l)$. So instead of a simple regression of λ on the x-variables, we weight each observation by $(C_i^u - C_i^l)/C_i$. Intuitively this makes a great deal of sense. If the lower and upper bounds are close, the value of $\hat{\lambda}$ does not matter in the prediction C_i^2. The larger the difference between the bounds, the more important it is to have an accurate estimate $\hat{\lambda}$ of λ.

The coefficients in the formula for y_2 were determined from a weighted regression on approximately 2,500 options, where the parameters were sampled from the distribution described in Section 4. As noted earlier, practitioners facing a different distribution of option parameters could reestimate the function $\hat{\lambda}_2$ to achieve the best performance.

B.8 Integral equation method

Equation (13), subject to the boundary condition $V_s(B_s^*, B^*) = B_s^* - K$, can be numerically approximated by discretizing the time interval $[t, T]$. Denote the time intervals by $t_0 < t_1 < \cdots < t_n$ with $t_0 = t$ and $t_n = T$. We take equally spaced intervals: $t_i = t + (T - t)i/n$. To solve for the boundary with n equally spaced increments, denoted B^n, first set $B_T^n = \max((r/\delta)K, K)$. Next solve for $B_{t_{n-1}}^n$ by setting the left-hand side of Equation (13) to $B_{t_{n-1}}^n - K$ and use numerical integration to evaluate the right-hand side of Equation (13). This nonlinear integral

[8] For $n = 8$, Simpson's rule approximates the integral of f over $[t_0, t_8]$ by

$$\int_{t=t_0}^{t_8} f(t)dt = h/3(f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + 4f_5 + 2f_6 + 4f_7 + f_8),$$

where $h = (t_8 - t_0)/8$.

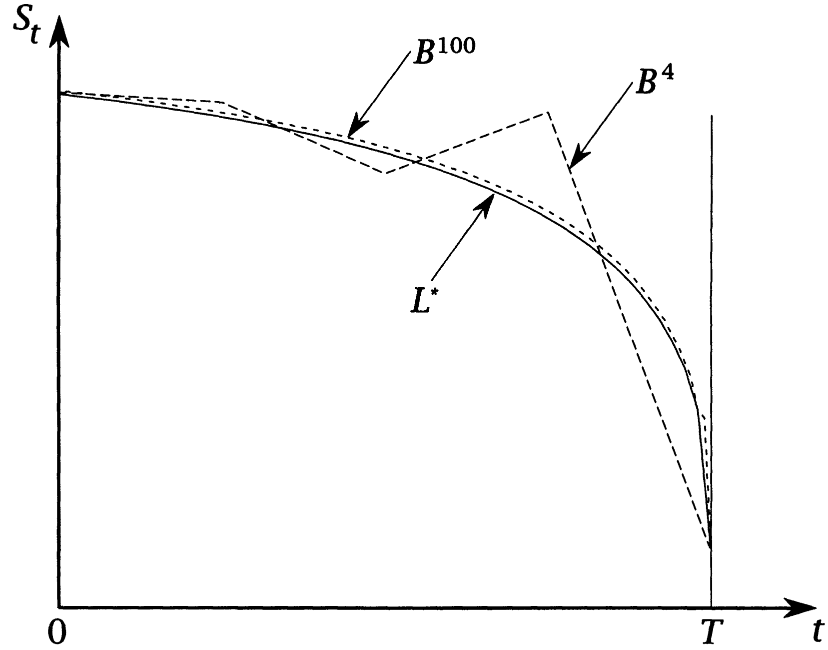

Figure 16
Illustration of B^4, B^{100}, and L^*
The solid line L^* is the approximate exercise boundary obtained by solving Equation (9). The dashed lines B^4 and B^{100} are approximations to the exercise boundary obtained with the integral equation method using 4 and 100 time steps, respectively. Note that B^4 is not monotonic in the time to maturity.

equation can be solved for the single unknown $B^n_{t_{n-1}}$. The boundary between adjacent points of B^n is taken to be linear. Continue this procedure for $i = n-2, \dots, 0$. This procedure is based on Kim (1990).

This method requires solving n integral equations, where n is the number of time steps. Like the binomial procedure, this procedure converges to the American option value as n increases to infinity.

Even though the optimal exercise boundary B^* is monotonic in the time to maturity, the discrete implementation of the integral method need not produce monotonic approximations to the boundary. This situation is illustrated in Figure 16 for a call option. The parameters used in Figure 16 are $\sigma = 0.2$, $r = 0.08$, $\delta = 0.12$, $K = 100$, and $T = 3$. For $n = 4$, B^n is not monotonic in the time to maturity.

References

Amin, K., 1991, "On the Computation of Continuous Time Option Prices Using Discrete Approximations," *Journal of Financial and Quantitative Analysis*, 26, 477–496.

Amin, K., and A. Khanna, 1994, "Convergence of American Option Values from Discrete- to Continuous-Time Financial Models," *Mathematical Finance*, 4, 289–304.

Barone-Adesi, G., and R. Whaley, 1987, "Efficient Analytical Approximation of American Option Values," *Journal of Finance*, 42, 301–320.

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Bjerksund, P., and G. Stensland, 1992, "Closed Form Approximation of American Options," *Scandinavian Journal of Management*.

Boyle, P., 1988, "A Lattice Framework for Option Pricing with Two State Variables," *Journal of Financial and Quantitative Analysis*, 23, 1–12.

Boyle, P., J. Evnine, and S. Gibbs, 1989, "Numerical Evaluation of Multivariate Contingent Claims," *Review of Financial Studies*, 2, 241–250.

Breen, R., 1991, "The Accelerated Binomial Option Pricing Model," *Journal of Financial and Quantitative Analysis*, 26, 153–164.

Brennan, M., and E. Schwartz, 1977, "The Valuation of American Put Options," *Journal of Finance*, 32, 449–462.

Brennan, M., and E. Schwartz, 1978, "Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis," *Journal of Financial and Quantitative Analysis*, 13, 461–474.

Broadie, M., and J. B. Detemple, 1995, "American Capped Call Options on Dividend Paying Assets," *Review of Financial Studies*, 8, 161–191.

Bunch, D., and H. Johnson, 1992, "A Simple and Numerically Efficient Valuation Method for American Puts Using a Modified Geske-Johnson Approach," *Journal of Finance*, 47, 809–816.

Carr, P., and D. Faguet, 1994, "Fast Accurate Valuation of American Options," working paper, Cornell University.

Carr, P., R. Jarrow, and R. Myneni, 1992, "Alternative Characterizations of American Put Options," *Mathematical Finance*, 2, 87–106.

Chesney, M., 1989, "Pricing American Currency Options: An Analytical Approach," working paper, Université de Lausanne.

Chesney, M., and R. Gibson, 1995, "State Space Symmetry and Two-Factor Option Pricing Models," *Advances in Futures and Options Research*, 8, 85–112.

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

El Karoui, N., and I. Karatzas, 1991, "A New Approach to the Skorohod Problem and its Applications," *Stochastics and Stochastics Reports*, 34, 57–82.

Geske, R., 1979, "A Note on an Analytical Valuation Formula for Unprotected American Options on Stocks with Known Dividends," *Journal of Financial Economics*, 7, 375–380.

Geske, R., and H. E. Johnson, 1984, "The American Put Options Valued Analytically," *Journal of Finance*, 39, 1511–1524.

Geske, R., and K. Shastri, 1985, "Valuation by Approximation: A Comparison of Alternative Option Valuation Techniques," *Journal of Financial and Quantitative Analysis*, 20, 45–71.

Grabbe, O., 1983, "The Pricing of Call and Put Options on Foreign Exchange," *Journal of International Money and Finance*, 2, 239–253.

Hull, J. C., and A. White, 1988, "The Use of the Control Variate Technique in Option Pricing," *Journal of Financial and Quantitative Analysis*, 23, 237–251.

Jacka, S. D., 1991, "Optimal Stopping and the American Put," *Mathematical Finance*, 1, 1–14.

Jaillet, P., D. Lamberton, and B. Lapeyre, 1990, "Variational Inequalities and the Pricing of American Options," *Acta Applicandae Mathematicae*, 21, 263–289.

Johnson, H., 1983, "An Analytic Approximation for the American Put Price," *Journal of Financial and Quantitative Analysis*, 18, 141–148.

Kamrad, B., and P. Ritchken, 1991, "Multinomial Approximating Models for Options with *k* State Variables," *Management Science*, 37, 1640–1652.

Kim, I. J., 1990, "The Analytic Valuation of American Options," *Review of Financial Studies*, 3, 547–572.

Kim, I. J., 1994, "Analytical Approximation of the Optimal Exercise Boundaries for American Futures Option," *Journal of Futures Markets*, 14, 1–24.

Leisen, D., and M. Reimer, 1995, "Binomial Models for Option Valuation—Examining and Improving Convergence," Working Paper B–309, University of Bonn.

MacMillan, L. W., 1986, "An Analytic Approximation for the American Put Price," *Advances in Futures and Options Research*, 1, 119–139.

Marchuk, G., and V. Shaidurov, 1983, *Difference Methods and Their Extrapolations*, Springer Verlag, New York.

McDonald, R., and M. Schroder, 1990, "A Parity Result for American Options," working paper, Northwestern University.

McKean, H. P., Jr., 1965, "Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics," *Industrial Management Review*, 6, 32–39.

Moro, B., 1995, "Fast Computation of Cumulative Normal Distribution Function," working paper, TMG Financial Products, Greenwich, Connecticut.

Omberg, E., 1987, "The Valuation of American Puts with Exponential Exercise Policies," *Advances in Futures and Options Research*, 2, 117–142.

Omberg, E., 1988, "Efficient Discrete Time Jump Process Models in Option Pricing," *Journal of Financial and Quantitative Analysis*, 23, 161–174.

Parkinson, M., 1977, "Option Pricing: The American Put," *Journal of Business*, 50, 21–36.

Rendleman, R., and B. Bartter, 1979, "Two-State Option Pricing," *Journal of Finance*, 34, 1093–1110.

Roll, R., 1977, "An Analytic Valuation Formula for Unprotected American Call Options on Stocks with Known Dividends," *Journal of Financial Economics*, 5, 251–258.

Schwartz, E. S., 1977, "The Valuation of Warrants: Implementing a New Approach," *Journal of Financial Economics*, 4, 79–93.

Trigeorgis, L., 1991, "A Log-Transformed Binomial Numerical Analysis Method for Valuing Complex Multi-Option Investments," *Journal of Financial and Quantitative Analysis*, 26, 309–326.

Whaley, R. E., 1981, "On the Valuation of American Call Options on Stocks with Known Dividends," *Journal of Financial Economics*, 9, 207–211.

Yu, G., 1993, "Essays on the Valuation of American Options," Ph. D. dissertation, New York University.