Exercise boundaries and efficient approximations to American option prices and hedge parameters

Farid AitSahlia

Financial Engines, 1804 Embarcadero Road, Palo Alto, California 94303, USA

Tze Leung Lai

Department of Statistics, Stanford University, Stanford, California 94305, USA

This paper presents a new numerical method to solve the integral equation defining the early exercise boundary of an American option. It is shown that the early exercise boundaries of standard American options are well approximated by linear splines with a few knots, implying that the new solution method can actually be carried out on a coarse grid of time points with reasonable accuracy. This leads to a fast and reasonably accurate method to compute the early exercise boundaries, values, and hedge parameters of American options. In this connection, a brief survey of recent developments in approximations to American option prices and hedge parameters are also given.

1. INTRODUCTION

The explosive growth in the use of derivatives by investors and institutions (financial and otherwise) has fueled the need for their efficient and accurate valuation. The problem of efficient and accurate valuation of American options has a large literature, which can be broadly characterized by two directions. The first involves a discretization scheme to approximate the continuous-time pricing problem. This approach is used in the finite-difference method of Brennan and Schwartz (1977), Courtadon (1982), and Wu and Kwok (1997) to solve the corresponding free-boundary PDE, and in the binomial tree method of Cox, Ross, and Rubinstein (1979), who approximate the underlying price process by a discrete process on a tree to which a dynamic programming algorithm can be applied.

To improve computational efficiency, a second direction has emerged whereby analytical characterizations or approximations are sought to circumvent the fine resolution required by direct discretization for accurate valuation. It can be traced back to Geske and Johnson (1984), who first characterized American options as compound European options and then used the Richardson extrapolation (with three or four points, typically) for the numerical approximation phase. The past few years have witnessed the emergence of a new approach to approximate American option prices and hedge parameters, based on the integral representation formula of the difference between American and European option prices due to Kim (1990), Jacka (1991), and Carr, Jarrow, and Myneni (1992). Although the representation formula is exact, it requires the determination of the early exercise boundary for its implementation. This integral representation formula also leads to an integral equation for the early exercise boundary. Huang, Subrahmanyam, and Yu (1996) proposed to approximate the integrands in both the integral representation formula and in the integral equation by piecewise constant functions, with $n = 1$, 2, and 3 pieces, yielding three crude approximations P_1, P_2, and P_3 to the option price, and then to combine them via a three-point Richardson extrapolation scheme so that greater accuracy can be achieved. Instead of using piecewise constant approximations to the integrands, Ju (1998) subsequently proposed to approximate the boundary by a piecewise exponential function, for which the integral in the representation formula can be evaluated in closed form. This approach leads to greater accuracy than that of Huang, Subrahmanyam, and Yu (1996), while maintaining the latter's computational simplicity. Ju (1998) reported numerical studies showing that his method with $n = 3$ pieces substantially improves those of Geske and Johnson (1984) and of Huang, Subrahmanyam, and Yu (1996) and other approximations in the literature. By using the finite-difference method to solve for the free-boundary PDE in one example, he computed an approximation to the early exercise boundary and found it to differ substantially from his piecewise exponential approximations with $n = 1$, 2, and 3 pieces. This led him to conclude that "an accurate estimate of the early exercise boundary is not required for pricing an American option accurately".

AitSahlia and Lai (1999) recently carried out extensive computations of early exercise boundaries for a wide range of maturities, interest rates, dividend rates, volatilities, and strike prices via reparametrization to reduce American option valuation to a single optimal stopping problem for standard Brownian motion, indexed by one parameter in the absence of dividends, and by two parameters otherwise. Their results show, however, that the early exercise boundary is actually *very well approximated* by a piecewise exponential boundary which uses a small number of pieces. This explains Ju's finding that his method has superior performance over other approximation approaches. In this paper, we make use of Ju's closed-form integration to develop (i) a new numerical method to solve the integral equation defining the early exercise boundary and (ii) an improvement of Ju's approximation to the boundary. It involves three basic ideas. The first is a change of variables under which the exercise boundary appears in the integrand in a simpler form. Moreover, under this change of variables, an exponential function is transformed into a linear function. The second idea uses a numerically stable root-finding algorithm in solving for the boundary points at successive times t_i. The third idea is to use, instead of the usual step function approximation to the boundary as in Kim (1990), a linear interpolation for the boundary between two adjacent time points and to apply Ju's (1998) closed-form expression for the premium integral when the boundary is piecewise linear in the transformed coordinates (or, equivalently, piecewise exponential in the original coordinates).

The last idea above is of particular importance in developing a good approximation to the boundary that involves only a few time points. Specifically, if the actual boundary for the transformed coordinates should be nearly linear over a wide interval with endpoints t_i and t_{i-1}, then interpolating the boundary linearly between t_i and t_{i-1} and using Ju's analytic formula for the premium integral between t_i and t_{i-1} would yield an almost exact value of the premium integral even though t_i and t_{i-1} may be quite far apart. Solving the integral equation for the early exercise boundary at narrowly spaced time points provides a benchmark to compare approximations that solve the integral equation on a coarse grid consisting of only a few points. This is done in Section 3 and confirms that the boundary is indeed well approximated by a linear spline having only a few knots. Hence, the three key ideas mentioned above also lead to an efficient approximation to the early exercise boundary that involves finding its values at a few time points and linearly interpolating elsewhere (in the transformed coordinate system). This fast and accurate approximation to the early exercise boundary yields an efficient method for the valuation of American options, thanks to Ju's (1998) closed-form expression for the premium integral. Moreover, we also develop here a closed-form expression for the hedge parameters using this piecewise linear approximation (in the transformed coordinate system) for the early exercise boundary.

The paper is organized as follows. Section 2 describes our method to solve the integral equation defining the early exercise boundary. Numerical results and efficient linear spline approximation to the boundary are given in Section 3. Section 4 gives closed-form expressions for the approximate computation of hedge parameters. Section 5 compares our approximation of American option prices and exercise boundaries with those of Ju (1998) and Huang, Subrahmanyam, and Yu (1996) and provides a deeper understanding of their methods and results via the benchmark developed in Section 2. In this connection, we also give a brief review of other approximations in the literature. Section 6 summarizes and concludes the paper.

2. A NUMERICAL METHOD TO SOLVE THE INTEGRAL EQUATION FOR THE EARLY EXERCISE BOUNDARY

In the standard Black–Scholes environment with a riskless interest rate r and an underlying asset having volatility σ and paying dividend at rate μ, the price of an American option at time t is the optimum value in the optimal stopping problem

$$U(t, P) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathsf{E}\big[e^{-r(\tau-t)} f(P_\tau) \mid P_t = P\big], \tag{1}$$

where $P_t = P_0 e^{(r-\mu-\frac{1}{2}\sigma^2)t+\sigma W_t}$ with initial security price P_0, $\{W_t\}$ is a standard Brownian motion (so that the stochastic process $\{P_t\}$ is a geometric Brownian motion), and $\mathcal{T}_{a,b}$ is the set of stopping times taking values between a and b with $b > a$. Given a strike price K, the payoff $f(P)$ in (1) is $(K-P)^+$ for a put, and $(P-K)^+$ for a call. We shall focus on American puts, as American calls on dividend-paying securities can be evaluated similarly.

Kim (1990), Jacka (1991), and Carr, Jarrow, and Myneni (1992) have obtained the following representation for an American put:

$$U(t, P) = U_{\mathrm{E}}(t, P) + \int_t^T \big[rK\mathrm{e}^{-r(\tau-t)}N\big(-d_2(P, B_\tau, \tau-t)\big) - \mu P\mathrm{e}^{-\mu(\tau-t)}N\big(-d_1(P, B_\tau, \tau-t)\big)\big]\,\mathrm{d}\tau, \tag{2}$$

in which $N(\cdot)$ denotes the standard normal distribution function, B_t is the early exercise boundary, U_{E} the corresponding European put price given by

$$U_{\mathrm{E}}(t, P) = K\mathrm{e}^{-r(T-t)}N\big(-d_2(P, K, T-t)\big) - P\mathrm{e}^{-\mu(T-t)}N\big(-d_1(P, K, T-t)\big),$$

where

$$d_1(x, y, \tau) = \frac{ln(x/y) + (r - \mu + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \quad \text{and} \quad d_2(x, y, \tau) = d_1(x, y, \tau) - \sigma\sqrt{\tau}.$$

Since $U(t, B_t) = K - B_t$, it follows from (2) that B_t satisfies the integral equation

$$K - B_t = U_{\mathrm{E}}(t, B_t) + \int_t^T \big[rK\mathrm{e}^{-r(\tau-t)}N\big(-d_2(B_t, B_\tau, \tau-t)\big) - \mu B_t\mathrm{e}^{-\mu(\tau-t)}N\big(-d_1(B_t, B_\tau, \tau-t)\big)\big]\,\mathrm{d}\tau. \tag{3}$$

Kim (1990) proposed to solve (3) numerically by dividing the interval $[0, T]$ into n subintervals $[t_{i-1}, t_i]$, with $t_0 = 0$, $t_n = T$, and $t_i - t_{i-1} = T/n$, and replacing (3) by the system of nonlinear equations

$$K - B_{t_i} = U_{\mathrm{E}}(t_i, B_{t_i}) + \sum_{j=i+1}^{n} \big[rK\mathrm{e}^{-r(t_j-t_i)}N\big(-d_2(B_{t_i}, B_{t_j}, t_j - t_i)\big) - \mu B_{t_i}\mathrm{e}^{-\mu(t_j-t_i)}N\big(-d_1(B_{t_i}, B_{t_j}, t_j - t_i)\big)\big]T/n, \tag{4}$$

which is solved recursively backwards with $B_{t_n} = K$ if $\mu \leqslant r$ and $B_{t_n} = rK/\mu$ if $\mu > r$. The basic idea behind (4) is to approximate the integrand in (3) by a step function, which gives good accuracy when n is large.

2.1 *Change of Variables*

The boundary B_t appears in (3) (and therefore (4) also) in a complicated manner. Letting $\rho = r/\sigma^2$ and $\alpha = \mu/r$, we introduce the change of variables

$$s = \sigma^2(t - T), \qquad z = \log(P/K) - (\rho - \alpha\rho - \tfrac{1}{2})s, \tag{5}$$

so that the boundary B_t becomes $\bar{z}(s)$ in the new coordinate system, with $B_t = K\mathrm{e}^{\bar{z}(s)+(\rho-\alpha\rho-\frac{1}{2})s}$. Note that a piecewise exponential boundary B_t is transformed into a piecewise linear function of s. Moreover, the integral equation (3) can be expressed in the simpler form

$$
\begin{aligned}
1-\mathrm{e}^{\bar{z}(s)+(\rho-\alpha\rho-\frac{1}{2})s} = \\
&\mathrm{e}^{\rho s}\big[N\big(-\bar{z}(s)/\sqrt{-s}\big) - \mathrm{e}^{\bar{z}(s)-\frac{1}{2}s}N\big(-\bar{z}(s)/\sqrt{-s}-\sqrt{-s}\big)\big] \\
&+\rho\mathrm{e}^{\rho s}\int_s^0\left[\mathrm{e}^{-\rho u}N\left(\frac{\bar{z}(u)-\bar{z}(s)}{\sqrt{u-s}}\right) - \alpha\mathrm{e}^{-(\alpha\rho u+\frac{1}{2}s)+\bar{z}(s)}N\left(\frac{\bar{z}(u)-\bar{z}(s)}{\sqrt{u-s}}-\sqrt{u-s}\right)\right]\mathrm{d}u.
\end{aligned}
\tag{6}
$$

The transformation (5) was introduced by AitSahlia and Lai (1999) to reduce the optimal stopping problem (1) to a canonical form that involves only one parameter, ρ, in the absence of dividends and two parameters, ρ and α, when dividends are paid. An important advantage of this transformation is that for values of σ (between 0.1 and 0.4) that are of practical interest, the time horizon $\sigma^2 T$ in the canonical scale is only a small fraction of T.

2.2 *Evaluation of Integrals*

To solve (6), instead of using a step function approximation to the entire integrand as in (4), we use a piecewise linear approximation to $\bar{z}(\cdot)$. This is clearly more accurate but requires evaluation of the integral, which has a closed-form expression when $\bar{z}(\cdot)$ is piecewise linear. First note that

$$
\begin{aligned}
&\rho\mathrm{e}^{\rho s}\int_s^0\left[\mathrm{e}^{-\rho u}N\left(\frac{\bar{z}(u)-z}{\sqrt{u-s}}\right) - \alpha\mathrm{e}^{-(\alpha\rho u+\frac{1}{2}s)+z}N\left(\frac{\bar{z}(u)-z}{\sqrt{u-s}}-\sqrt{u-s}\right)\right]\mathrm{d}u \\
&\quad = 1-\mathrm{e}^{\rho s}-\mathrm{e}^{z+(\rho-\alpha\rho-\frac{1}{2})s}(1-\mathrm{e}^{\alpha\rho s}) - \int_0^{-s}\rho\mathrm{e}^{-\rho t}N\left(\frac{z-\bar{z}(s+t)}{\sqrt{t}}\right)\mathrm{d}t \\
&\qquad + \mathrm{e}^{z+(\rho-\alpha\rho-\frac{1}{2})s}\int_0^{-s}\alpha\rho\mathrm{e}^{-\alpha\rho t}N\left(\frac{z-\bar{z}(s+t)+t}{\sqrt{t}}\right)\mathrm{d}t.
\end{aligned}
\tag{7}
$$

To evaluate the last two integrals in (7), suppose that $s = s_m < \cdots < s_0 = 0$ divide the interval $[s, 0]$ into m subintervals such that

$$\bar{z}(u) = \beta_i u + \gamma_i \quad \text{for } s_i \leqslant u \leqslant s_{i-1} \quad (1 \leqslant i \leqslant m). \tag{8}$$

Let $\tau_i = s_i - s_m$. Then $\bar{z}(t+s_m) - z = -(b_i t + c_i)$ for $\tau_i \leqslant t \leqslant \tau_{i-1}$, where $b_i = -\beta_i$, $c_i = z - \gamma_i - \beta_i s_m$, noting that $\tau_i + s_m = s_i$. Let $a_i = \sqrt{b_i^2 + 2\rho}$. Then, for $1 \leqslant i \leqslant m$, we have

$$
\begin{aligned}
&\int_{\tau_i}^{\tau_{i-1}}\rho\mathrm{e}^{-\rho t}N\left(\frac{z-\bar{z}(t+s_m)}{\sqrt{t}}\right)\mathrm{d}t \\
&\quad = \mathrm{e}^{-\rho\tau_i}N(b_i\tau_i^{1/2}+c_i\tau_i^{-1/2}) - \mathrm{e}^{-\rho\tau_{i-1}}N(b_i\tau_{i-1}^{1/2}+c_i\tau_{i-1}^{-1/2}) \\
&\qquad + \tfrac{1}{2}\left(\frac{b_i}{a_i}+1\right)\mathrm{e}^{(a_i-b_i)c_i}[N(a_i\tau_{i-1}^{1/2}+c_i\tau_{i-1}^{-1/2}) - N(a_i\tau_i^{1/2}+c_i\tau_i^{-1/2})] \\
&\qquad + \tfrac{1}{2}\left(\frac{b_i}{a_i}-1\right)\mathrm{e}^{-(a_i+b_i)c_i}[N(a_i\tau_{i-1}^{1/2}-c_i\tau_{i-1}^{-1/2}) - N(a_i\tau_i^{1/2}-c_i\tau_i^{-1/2})],
\end{aligned}
\tag{9}
$$

which corresponds to equation (5) of Ju (1998). In the case where $\tau_i = 0$, we replace τ_i by $t > 0$ and take the limit as $t \to 0$, which amounts to setting $c_i \tau_i^{-1/2} = 0$, or ∞, or $-\infty$ according as $c_i = 0$, or $c_i > 0$, or $c_i < 0$. Similarly, letting $\tilde{b}_i = b_i + 1$ and $\tilde{a}_i = (\tilde{b}_i^2 + 2\alpha\rho)^{1/2}$, we have

$$\int_{\tau_i}^{\tau_{i-1}} \alpha\rho e^{-\alpha\rho t} N\left(\frac{z - \bar{z}(t + s_m) + t}{\sqrt{t}}\right) dt$$
$$= \text{RHS of (9) with } (b_i, a_i) \text{ replaced by } (\tilde{b}_i, \tilde{a}_i) \text{ and } \rho \text{ by } \alpha\rho, \quad (10)$$

which corresponds to equation (6) of Ju (1998). When either $(a_i - b_i)c_i$ or $-(a_i + b_i)c_i$ exceeds some large number (say 400), its exponential in (9) becomes exceedingly large, causing numerical instability. In these situations, which typically do not occur unless $\tau_{i-1} - \tau_i$ is very small, we evaluate the LHS of (9) by the midpoint rule:

$$\int_{\tau_i}^{\tau_{i-1}} \rho e^{-\rho t} N\left(\frac{z - \bar{z}(t + s_m)}{\sqrt{t}}\right) dt \doteq (e^{-\rho\tau_i} - e^{-\rho\tau_{i-1}}) N\left(\frac{z - \bar{z}(\tau_i^* + s_m)}{\sqrt{\tau_i^*}}\right), \quad (11)$$

where $\tau_i^* = \frac{1}{2}(\tau_i + \tau_{i-1})$. We also evaluate the LHS of (10) by the midpoint rule if $(\tilde{a}_i - \tilde{b}_i)c_i$ or $-(\tilde{a}_i + \tilde{b}_i)c_i$ is large. Since $\int_0^{-s_m} = \sum_{i=1}^m \int_{\tau_i}^{\tau_{i-1}}$, the integral on the RHS of (7) can be expressed as a sum of terms in (9) and (10).

2.3 *Solving for the Boundary at Successive Time Points*

To solve the integral equation (6) recursively on a grid of time points $s_0 = 0 > s_1 > \cdots > s_n = -\sigma^2 T$, we linearly interpolate the boundary between s_j and s_{j+1} and use (7) together with (9), (10), or (11) to evaluate the integrals in (6) in terms of simple sums. The recursion is initialized at $s_0 = 0$ by $\bar{z}(0) = 0$ if $0 \leqslant \alpha \leqslant 1$ and $\bar{z}(0) = -\ln\alpha$ if $\alpha > 1$. For $1 \leqslant m \leqslant n$ and $1 \leqslant j \leqslant m$, let $\bar{z}_j = \bar{z}(s_j)$ and $\tau_j = s_j - s_m$. Suppose that $\bar{z}_0, \ldots, \bar{z}_{m-1}$ have been determined. Then $b_i = -\beta_i = (\bar{z}_i - \bar{z}_{i-1})/(s_{i-1} - s_i)$, $\gamma_i = \bar{z}_{i-1} - \beta_i s_{i-1}$, and $c_i = \bar{z}_i - \bar{z}_{i-1} + \beta_i \tau_{i-1}$ for $1 \leqslant i \leqslant m-1$. To determine $\bar{z}_m$, let z be a candidate value and let $b(z) = (z - \bar{z}_{m-1})/(s_{m-1} - s_m)$, $a(z) = [b^2(z) + 2\rho]^{1/2}$, $\tilde{b}(z) = b(z) + 1$, and $\tilde{a}(z) = [\tilde{b}^2(z) + 2\alpha\rho]^{1/2}$. Noting that $c(z) := z - \bar{z}_{m-1} - (s_{m-1} - s_m)b(z) = 0$ and $\tau_m = 0$, we obtain from (6)–(10) the following equation (in the variable z) defining $\bar{z}_m$:

$$1 - e^{z + (\rho - \alpha\rho - \frac{1}{2})s_m}$$
$$= e^{\rho s_m}\left[N(-z/\sqrt{-s_m}) - e^{z - \frac{1}{2}s_m} N(-z/\sqrt{-s_m} - \sqrt{-s_m})\right]$$
$$+ 1 - e^{\rho s_m} - e^{z + (\rho - \alpha\rho - \frac{1}{2})s_m}(1 - e^{\alpha\rho s_m}) + e^{-\rho\tau_{m-1}} N\big(b(z)\tau_{m-1}^{1/2}\big) - \tfrac{1}{2}$$
$$- \frac{b(z)}{a(z)} N\big(a(z)\tau_{m-1}^{1/2} - \tfrac{1}{2}\big) - \sum_{i=1}^{m-1} A_i(z) + e^{z + (\rho - \alpha\rho - \frac{1}{2})s_m}\left[\frac{\tilde{b}(z)}{\tilde{a}(z)}\left[N\big(\tilde{a}(z)\tau_{m-1}^{1/2}\big) - \tfrac{1}{2}\right]\right.$$
$$\left. + \tfrac{1}{2} - e^{-\alpha\rho\tau_{m-1}} N\big(\tilde{b}(z)\tau_{m-1}^{1/2}\big) + \textstyle\sum_{i=1}^{m-1} \tilde{A}_i(z)\right], \quad (12)$$

where $A_i(z)$ is given by the RHS of (9) and $\tilde{A}_i(z)$ by the RHS of (10).

The nonlinear equation (12) can be solved by a Van Wijngaarden–Dekker–Brent-type method (see Press *et al.* 1992). First, as shown by AitSahlia and Lai (1999), the early exercise boundary $\bar{z}(s)$ is bounded above by $\bar{z}_{\mathrm{u}}(s)$ and bounded below by $\bar{z}_{\ell}(s)$, where

$$\bar{z}_{\mathrm{u}}(s) = -[\rho(1-\alpha) - \tfrac{1}{2}]s - (\ln \alpha)^{+},$$
$$\bar{z}_{\ell}(s) = -[\rho(1-\alpha) - \tfrac{1}{2}]s - \ln[\theta/(\theta-1)],$$
$$\theta = -[\rho(1-\alpha) - \tfrac{1}{2}] - \{[\rho(1-\alpha) - \tfrac{1}{2}]^2 + 2\rho\}^{1/2}.$$

Let $D(z)$ denote the difference between the two sides of (12); specifically, $D(z) = \mathrm{LHS} - \mathrm{RHS}$. If $D(\bar{z}_{m-1}) = 0$, we have found the solution and can set $\bar{z}_m = \bar{z}_{m-1}$. If $D(\bar{z}_{m-1}) < 0$, set $z' = \bar{z}_{m-1}$ and $z'' = \bar{z}_{\ell}(s_m)$ after checking that $D(z'') > 0$. In the unlikely event that the latter condition is violated, set $\bar{z}_m = \bar{z}_{m-1}$. If $D(\bar{z}_{m-1}) > 0$, set $z'' = \bar{z}_{m-1}$ and $z' = \bar{z}_u(s_m)$ after checking that $D(z') < 0$, whose unlikely violation results in setting $\bar{z}_m = \bar{z}_{m-1}$. After bracketing the solution in this way between z' and z'' with $D(z') < 0$ and $D(z'') > 0$, we can use successive linear approximations that replace $D(z)$ by the linear function $D_1(z)$ with $D_1(z') = D(z')$ and $D_1(z'') = D(z'')$. Let

$$z^* = z'' - D(z'')(z' - z'')/[D(z') - D(z'')]$$

be the solution of $D_1(z) = 0$. Note that z^* lies between z' and z''. If $D(z^*) < 0$, reset z' at z^*. If $D(z^*) > 0$, reset z'' at z^*. Proceeding inductively in this way, the procedure terminates when a solution is reached or when $|z' - z''|$ falls below some prescribed tolerance level.

3. NUMERICAL RESULTS AND AN EFFICIENT APPROXIMATION TO THE EARLY EXERCISE BOUNDARY

Using even spacing in the choice of grid points $s_0 = 0 > s_1 > \cdots > s_n = -\sigma^2 T$ in Section 2, i.e., $s_i - s_{i+1} = \delta := \sigma^2 T/n$ for $0 \leqslant i \leqslant n-1$, Table 1 gives values of the early exercise boundary $\bar{z}(\cdot)$ obtained by using the method of Section 2 to solve the Volterra integral equation (6) with $\delta = 10^{-2}$, 10^{-3}, or 10^{-4}. These results show convergence of the method with diminishing δ; moreover, they show that the integral equation approach is already quite accurate for $\delta = 10^{-2}$ and $-s \geqslant 0.05$. Note that, for $\delta = 10^{-2}$, $s = -0.01$ corresponds to the first time point at which the integral equation method computes a boundary value to start the recursion and $s = -0.05$ is the fifth recursive stage.

For small δ, a much faster method to compute $\bar{z}(\cdot)$ is the corrected Bernoulli walk method introduced by Chernoff and Petkau (1986). This method has recently been used to compute the benchmark values of early exercise boundaries of American options by AitSahlia and Lai (1999). Table 1 also gives for comparison the values of $\bar{z}(\cdot)$ computed by the corrected Bernoulli walk method with time step size $\delta = 10^{-4}$, 10^{-5}, or 10^{-6}. The results show that,

TABLE 1. Early exercise boundary, in canonical scale (5), using the Volterra integral equation (6) or the corrected Bernoulli walk method for the American put with $\rho = 0.5$ and $\alpha = 0$.

	Volterra			Bernoulli		
$-s$	$\delta = 10^{-2}$	$\delta = 10^{-3}$	$\delta = 10^{-4}$	$\delta = 10^{-4}$	$\delta = 10^{-5}$	$\delta = 10^{-6}$
.0001			−0.03111	−0.01000	−0.02530	−0.02800
.0002			−0.03656	−0.02000	−0.03795	−0.03700
.0003			−0.04489	−0.03000	−0.04427	−0.04400
.0004			−0.05022	−0.04000	−0.05060	−0.05000
.0005			−0.05532	−0.05000	−0.05376	−0.05500
.0006			−0.05967	−0.06000	−0.06008	−0.06000
.0007			−0.06366	−0.06000	−0.06325	−0.06400
.0008			−0.06729	−0.06000	−0.06641	−0.06700
.0009			−0.07067	−0.07000	−0.06957	−0.07100
.0010		−0.08537	−0.07382	−0.07000	−0.07273	−0.07400
.0020		−0.09522	−0.09801	−0.10000	−0.09803	−0.09800
.0030		−0.11680	−0.11531	−0.12000	−0.11384	−0.11500
.0040		−0.12909	−0.12917	−0.13000	−0.12965	−0.12900
.0050		−0.14126	−0.14098	−0.14161	−0.14085	−0.14086
.0100	−0.22116	−0.18350	−0.18342	−0.18456	−0.18337	−0.18350
.0500	−0.32180	−0.32095	−0.32095	−0.32107	−0.32096	−0.32094
.1000	−0.39610	−0.39587	−0.39570	−0.39586	−0.39578	−0.39586
.2000	−0.47572	−0.47594	−0.47567	−0.47564	−0.47561	−0.47525
.2400	−0.49669	−0.49669	−0.49663	−0.49656	−0.49653	−0.49669
.2800	−0.51420	−0.51418	−0.51411	−0.51413	−0.51412	−0.51430

for $.002 \leqslant -s < .005$, the corrected Bernoulli walk method needs $\delta = 10^{-6}$ to give boundary values comparable to those obtained by using the method of Section 2 to solve the integral equation (6). For this range of $-s$, using $\delta = 10^{-4}$ in the corrected Bernoulli walk method produces a substantial relative error for the boundary. For $-s \leqslant .001$, Table 1 shows that even smaller values of δ than 10^{-6} are needed for the corrected Bernoulli walk method.

Although the numerical procedure of Section 2 is computationally expensive for small values of δ because of the large number of computations and memory needed in evaluating $A_i(z)$ and $\tilde{A}_i(z)$ for $1 \leqslant i \leqslant m-1$ and different values of z in (12) when m is large, it is very fast when m ($\leqslant n$) is small. For $T = 1$ and $\sigma = .1$, $n = \sigma^2 T/\delta = 10$ if we choose $\delta = 10^{-2}$, which gives quite accurate results for the boundary values at $-s \geqslant 0.05$. On the other hand, the Bernoulli method with 10 time steps ($\delta = 10^{-2}$) is grossly inaccurate.

As noted by AitSahlia and Lai (1999), the Chernoff–Petkau continuity correction to compute the optimal stopping boundary $\bar{z}(\cdot)$ at s via the Bernoulli walk approximation to Brownian motion requires boundedness of the derivative

$\bar{z}'(\cdot)$ in some neighborhood of s and is not applicable around $s = 0$. Since the corrected Bernoulli walk method can be applied for $s \leqslant -.005$, we can restrict ourselves to $0 > s > -.005$ in applying the preceding numerical procedure to solve the integral equation that defines the optimal stopping boundary. In particular, with $\delta = 10^{-4}$, the range $0 > s > -.005$ only involves $n = 50$ time steps. Let $s_0 = -.005$ and $t_0 = T + s_0/\sigma^2$. Having computed $\bar{z}(\cdot)$ for $0 > s \geqslant s_0$ by using the preceding procedure to solve the integral equation defining $\bar{z}(\cdot)$, we can use (13) below together with (9) and (10) to evaluate the value function at s_0 as the terminal payoff of the corrected Bernoulli walk method that computes $\bar{z}(\cdot)$ by backward induction for $s \leqslant -s_0$. As shown by Lai, Yao, and AitSahlia (1999), this hybrid approach that combines the corrected Bernoulli walk method of Chernoff and Petkau (for $s \leqslant s_0$) with the integral equation method (for $s_0 \leqslant s < 0$) has $o(\sqrt{\delta})$ error in computing the boundary $\bar{z}(\cdot)$. Note the computational efficiency of this hybrid approach, which exploits the recursive nature of the corrected Bernoulli walk method and only involves $-s_0/\delta$ time steps for the integral equation method applied to the initial time segment $[-s_0, 0)$.

Figure 1 plots the graphs of $\bar{z}(s)$ for $-0.3 \leqslant s \leqslant 0$, with $\rho = 0.5$, and for different values of α, computed by the preceding approach with $\delta = 10^{-4}$. It shows that $\bar{z}(\cdot)$ is well approximated by a linear spline with knots at $s = 0$, $-.005$, $-.025$, $-.05$, $-.1$, $-.15$, $-.3$. For $\sigma = .1$, $s = -.3$ corresponds to a very long maturity T of 30 years. The approximate piecewise linearity of $\bar{z}(\cdot)$ shown in Figure 1 is also confirmed by extensive numerical computations over a wide range of ρ and α values for puts and calls in AitSahlia and Lai (1999). Hence, using a small number of time steps in the procedure of Section 2 results in a reasonably accurate and fast approximation to the early exercise boundary.

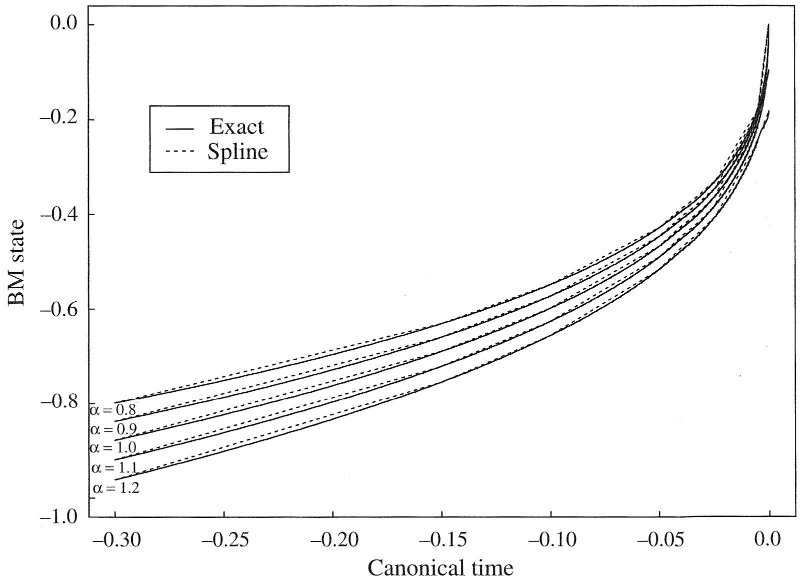

FIGURE 1. Optimal stopping boundaries: standard put with $\rho = 0.5$.

4. APPLICATION TO VALUATION AND HEDGING

As in Section 2, we again focus on American puts since the treatment of American calls on dividend-paying securities is similar. After the change of variables (5), the value of (2) can be written as

$$U(t, P) = K\mathrm{e}^{\rho s}\big[N\big(-z/\sqrt{-s}\big) - \mathrm{e}^{z-\frac{1}{2}s}N\big(-z/\sqrt{-s} - \sqrt{-s}\big)\big] + \rho K\mathrm{e}^{\rho s}\int_s^0 \bigg[\mathrm{e}^{-\rho u}N\bigg(\frac{\bar{z}(u) - z}{\sqrt{u-s}}\bigg) - \alpha\mathrm{e}^{-\alpha\rho u - \frac{1}{2}s + z}N\bigg(\frac{\bar{z}(u) - z}{\sqrt{u-s}} - \sqrt{u-s}\bigg)\bigg]\,\mathrm{d}u. \quad (13)$$

Suppose that $\bar{z}(\cdot)$ can be approximated by a piecewise linear function of the form (8). Then, in view of (7) and $\int_0^{-s_m} = \sum_{i=1}^m \int_{\tau_i}^{\tau_{i-1}}$, we can express the integral in (13) as a sum of the terms in (9) and (10) (in which $c_i\tau_i^{1/2}$ is set to be equal to 0, ∞, or $-\infty$ according as $c_i = 0$, or $c_i > 0$, or $c_i < 0$ when $\tau_i = 0$) or in (11) when $(a_i - b_i)c_i$ or $-(a_i + b_i)c_i$ is too large.

Formula (13) also leads to explicit expressions for the hedge parameters through differentiation. In particular, the parameter delta can be expressed as

$$\frac{\partial U}{\partial P}(t, P) = -\mathrm{e}^{\alpha\rho s}N\bigg(-\frac{z}{\sqrt{-s}} - \sqrt{-s}\bigg) - \rho\mathrm{e}^{-z-(\rho-\alpha\rho-\frac{1}{2})s}\int_0^{-s}\frac{\mathrm{e}^{-\rho\tau}}{\sqrt{\tau}}n\bigg(\frac{\bar{z}(\tau+s) - z}{\sqrt{\tau}}\bigg)\,\mathrm{d}\tau - \int_0^{-s}\alpha\rho\mathrm{e}^{-\alpha\rho\tau}N\bigg(\frac{\bar{z}(\tau+s) - z - \tau}{\sqrt{\tau}}\bigg) + \alpha\rho\int_0^{-s}\frac{\mathrm{e}^{-\alpha\rho\tau}}{\sqrt{\tau}}n\bigg(\frac{\bar{z}(\tau+s) - z - \tau}{\sqrt{\tau}}\bigg)\,\mathrm{d}\tau, \quad (14)$$

where $n(x) = \mathrm{e}^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ is the standard normal density function. Suppose that $\bar{z}(s)$ can be approximated by a piecewise linear function of the form (8). Then the second integral in (14) can be written as

$$\int_0^{-s}\alpha\rho\mathrm{e}^{-\alpha\rho\tau}N\bigg(\frac{\bar{z}(\tau+s) - z - \tau}{\sqrt{\tau}}\bigg)\,\mathrm{d}\tau = 1 - \mathrm{e}^{\alpha\rho s} - \int_0^{-s}\alpha\rho\mathrm{e}^{-\alpha\rho\tau}N\bigg(\frac{z - \bar{z}(\tau+s) + \tau}{\sqrt{\tau}}\bigg)\,\mathrm{d}\tau,$$

which can therefore be expressed as a sum of terms of the form (10), with $s = s_m$. Since $a_i^2 = b_i^2 + 2\rho$, the first integral can also be written as a sum of the terms

$$\begin{aligned}\int_{\tau_i}^{\tau_{i-1}}\frac{\mathrm{e}^{-\rho\tau}}{\sqrt{\tau}}n\bigg(\frac{\bar{z}(\tau+s_m) - z}{\sqrt{\tau}}\bigg)\,\mathrm{d}\tau &= \int_{\tau_i}^{\tau_{i-1}}\frac{\mathrm{e}^{-\rho\tau}}{\sqrt{\tau}}n(b_i\tau^{1/2} + c_i\tau^{1/2})\,\mathrm{d}\tau \\ &= (2\pi)^{-1/2}\mathrm{e}^{-b_ic_i}\int_{\tau_i}^{\tau_{i-1}}\tau^{-1/2}\mathrm{e}^{-\frac{1}{2}(a_i^2\tau + c_i^2\tau^{-1})}\,\mathrm{d}\tau \\ &= a_i^{-1}\mathrm{e}^{(a_i-b_i)c_i}\big[N\big(a_i\tau_{i-1}^{1/2} + c_i\tau_{i-1}^{-1/2}\big) - N\big(a_i\tau_i^{1/2} + c_i\tau_i^{-1/2}\big)\big] \\ &\quad + a_i^{-1}\mathrm{e}^{-(a_i+b_i)c_i}\big[N\big(a_i\tau_{i-1}^{1/2} - c_i\tau_{i-1}^{-1/2}\big) - N\big(a_i\tau_i^{1/2} - c_i\tau_i^{-1/2}\big)\big]. \quad (15)\end{aligned}$$

TABLE 2. Accuracy of three-piece linear spline approximation (LSP3) relative to binomial and hybrid benchmarks. Case I: $T = 3$, $\sigma = 0.1$, $r = 0.06$; Case II: $T = 0.5$, $\sigma = 0.2$, $r = 0.06$.

			Price			Delta		
	μ	P	Binomial	Hybrid	LSP3	Binomial	Hybrid	LSP3
Case I	.09	80	22.787	22.787	22.787	−0.754	−0.754	−0.754
		90	15.706	15.706	15.707	−0.655	−0.655	−0.655
		100	9.843	9.843	9.844	−0.511	−0.511	−0.511
		110	5.561	5.561	5.562	−0.346	−0.346	−0.346
		120	2.840	2.840	2.840	−0.204	−0.204	−0.204
	.06	80	20.000	20.000	19.982	−1.000	−1.000	−1.000
		90	11.593	11.598	11.603	−0.686	−0.682	−0.683
		100	6.087	6.082	6.101	−0.425	−0.428	−0.423
		110	2.870	2.870	2.883	−0.231	−0.231	−0.230
		120	1.221	1.221	1.230	−0.110	−0.110	−0.110
	.03	80	20.000	19.998	19.994	−1.000	−1.000	−0.997
		90	10.057	10.051	10.064	−0.905	−0.904	−0.899
		100	3.964	3.962	3.965	−0.383	−0.383	−0.386
		110	1.449	1.448	1.445	−0.152	−0.152	−0.153
		120	0.489	0.488	0.487	−0.055	−0.055	−0.056
	.00	80	20.000	19.992	19.971	−1.000	−1.000	−0.994
		90	10.000	9.984	10.017	−1.000	−1.000	−1.000
		100	2.723	2.727	2.734	−0.366	−0.365	−0.368
		110	0.706	0.705	0.706	−0.096	−0.096	−0.096
		120	0.178	0.178	0.187	−0.025	−0.025	−0.023
Case II	.09	80	20.802	20.803	20.803	−0.906	−0.906	−0.906
		90	12.422	12.422	12.422	−0.748	−0.748	−0.748
		100	6.183	6.183	6.183	−0.491	−0.491	−0.491
		110	2.535	2.534	2.535	−0.250	−0.250	−0.250
		120	0.865	0.865	0.865	−0.100	−0.100	−0.100
	.06	80	20.093	20.094	20.055	−0.949	−0.949	−0.945
		90	11.545	11.545	11.542	−0.742	−0.742	−0.740
		100	5.504	5.504	5.510	−0.462	−0.462	−0.462
		110	2.154	2.154	2.158	−0.223	−0.223	−0.223
		120	0.701	0.701	0.702	−0.085	−0.085	−0.085
	.03	80	20.000	20.000	19.968	−1.000	−1.000	−0.995
		90	10.953	10.953	10.906	−0.760	−0.760	−0.759
		100	4.961	4.961	4.940	−0.442	−0.443	−0.440
		110	1.843	1.843	1.841	−0.200	−0.200	−0.200
		120	0.570	0.570	0.571	−0.072	−0.072	−0.072
	.00	80	20.000	20.000	19.996	−1.000	−1.000	−0.996
		90	10.522	10.523	10.524	−0.794	−0.794	−0.794
		100	4.493	4.492	4.493	−0.427	−0.427	−0.427
		110	1.578	1.577	1.582	−0.180	−0.180	−0.181
		120	0.462	0.462	0.465	−0.061	−0.061	−0.062

The last equality in (15) can be derived by using the change of variables

$$x = a_i\tau^{1/2} + c_i\tau^{-1/2}, \qquad y = a_i\tau^{1/2} - c_i\tau^{-1/2},$$

so that

$$\mathrm{d}x = \tfrac{1}{2}(a_i\tau^{-1/2} - c_i\tau^{-3/2})\,\mathrm{d}\tau, \qquad \mathrm{d}y = \tfrac{1}{2}(a_i\tau^{-1/2} + c_i\tau^{-3/2})\,\mathrm{d}\tau,$$

and, by using the identity

$$\tau^{-1/2} = \frac{(a_i\tau^{-1/2} - c_i\tau^{-3/2}) + (a_i\tau^{-1/2} + c_i\tau^{-3/2})}{2a_i}.$$

Similarly, letting $\tilde{b}_i = b_i + 1$ and $\tilde{a}_i = (\tilde{b}_i^2 + 2\alpha\rho)^{1/2}$ as before, we can express the third integral as a sum of the terms

$$\begin{aligned}
&\int_{\tau_i}^{\tau_{i-1}} \frac{\mathrm{e}^{-\alpha\rho\tau}}{\sqrt{\tau}} n\left(\frac{\bar{z}(\tau + s_m) - z - \tau}{\sqrt{\tau}}\right) \mathrm{d}\tau \\
&\quad = \int_{\tau_i}^{\tau_{i-1}} \frac{\mathrm{e}^{-\alpha\rho\tau}}{\sqrt{\tau}} n\big(\tilde{b}_i\tau^{1/2} + c_i\tau^{-1/2}\big)\,\mathrm{d}\tau \\
&\quad = \tilde{a}_i^{-1}\mathrm{e}^{(\tilde{a}_i - \tilde{b}_i)c_i}\big[N\big(\tilde{a}_i\tau_{i-1}^{1/2} + c_i\tau_{i-1}^{-1/2}\big) - N\big(\tilde{a}_i\tau_i^{1/2} + c_i\tau_i^{-1/2}\big)\big] \\
&\quad\quad + \tilde{a}_i^{-1}\mathrm{e}^{-(\tilde{a}_i + \tilde{b}_i)c_i}\big[N\big(\tilde{a}_i\tau_{i-1}^{1/2} - c_i\tau_{i-1}^{-1/2}\big) - N\big(\tilde{a}_i\tau_i^{1/2} - c_i\tau_i^{-1/2}\big)\big]. \qquad (16)
\end{aligned}$$

When $\tau_i = 0$ in (10), (15), or (16), we take $c_i\tau_i^{-1/2} = 0$, or ∞, or $-\infty$ according as $c_i = 0$, or $c_i > 0$, or $c_i < 0$.

The closed-form valuation and hedge parameter formulas given by (13)–(16), together with (9) and (10), are based on the piecewise linear approximation (8) to the early exercise boundary after the change of variables (5). We have seen in Section 3 that $\bar{z}(\cdot)$ can be well approximated by a linear spline with a few knots that are determined by using the numerical method of Section 2. Table 2 gives the results for the option prices and deltas computed by using a three-piece linear approximation (denoted by LSP3) of $\bar{z}(\cdot)$, which places the first knot at $s = -.005$ and evenly spaces the remaining two knots and which limits the solution of (12) to only five iterations to achieve greater speed than previously proposed approximations. These results are overwhelmingly within 0.1% of the corresponding benchmark values computed by (i) using (13) and $\bar{z}(\cdot)$ determined by the hybrid method in Section 3 with $\delta = 10^{-4}$, and (ii) using the binomial tree method with 10 000 steps.

5. COMPARISON WITH OTHER APPROXIMATIONS

Shortly after the Geske–Johnson approximation which is reviewed in Section 1, McMillan (1986) and Barone-Adesi and Whaley (1987) developed an alternative analytic approximation of the option price. The basic idea is to approximate the PDE for the difference between American and European option prices by an ODE. Applying the boundary conditions to the ODE gives a nonlinear equation for a constant (time-invariant) approximation P^* to the exercise boundary, and a closed-form expression involving P^* yields an approximation to the American option price. Although the method is fast and gives an adequate approximation to the option price when the maturity is short or very long, it suffers from lack of accuracy for intermediate lifespans. Moreover, P^* fails to capture the shape of the time-varying exercise boundary.

Bunch and Johnson (1992) introduced a modification of the Geske–Johnson method. Broadie and Detemple (1996) developed upper and lower bounds for an American option and used a convex combination of these bounds with empirically determined weights to approximate the option price. Carr (1998) discretized the time dimension of the PDE with a few points and used a randomization method to approximate the option price. Ju (1998) has carried out an extensive numerical study, comparing these approximations with his own and that of Huang, Subrahmanyam, and Yu (1996), and has found that his piecewise exponential approximation to the early exercise boundary using three evenly spaced pieces gives the best performance in terms of both speed and accuracy. We shall therefore focus mainly on Ju's method but also consider briefly the closely related method of Huang, Subrahmanyam, and Yu (1996).

Since a piecewise exponential exercise boundary B_t is equivalent to a piecewise linear $\bar{z}(s)$ after the change of variables (5), our method and Ju's basically use the same functional form to approximate the early exercise boundary. However, there are important differences. First, like the actual boundary, our approximation is continuous; it is a linear spline. On the other hand, Ju chooses the coefficients Q_m and q_m of the exponential function $Q_m e^{q_m t}$ separately for each of three pieces, resulting in discontinuities at the time points between successive intervals. Secondly, instead of even spacing between successive time points, we allow unevenly spaced knots to better fit the shape of the early exercise boundary as explained in Section 3, and the number of pieces in our piecewise linear approximation depends on $\sigma^2 T$ and may vary between 1 and 6. Thirdly, our algorithm in Section 2 to solve (10) for $\bar{z}(s_m)$ is simpler than Ju's method, which has to determine the two (instead of one) parameters Q_m and q_m recursively.

The differences between our method and Ju's are primarily due to major differences in the rationale behind the two approaches. The motivation behind our use of the linear spline approximation to $\bar{z}(\cdot)$ with a few knots comes from an extensive numerical study in which we computed $\bar{z}(\cdot)$ almost exactly for different values of ρ and α and which shows that $\bar{z}(\cdot)$ can indeed be well approximated by such linear splines. In contrast, Ju (1998) was motivated by his attempt to use the closed-form expression for the integral in (2) when B_t is piecewise exponential to improve on the much cruder approximation of Huang, Subrahmanyam, and Yu (1996) that treats the entire integrand as piecewise constant even though the interval $[0, T]$ is coarsely partitioned into three subintervals of equal width. Since he felt that the actual exercise boundary might differ substantially from a piecewise exponential curve with three or fewer pieces, he used two adjustable parameters q_m and Q_m to fit an exponential function in each piece. The following example illustrates the adequacy of Ju's three-piece approximation (with jump discontinuities) to the early exercise boundary and compares it with ours and that of Huang, Subrahmanyam, and Yu (1996).

Example 1. Ju (1998, p. 635) has considered the following American put with $K = 100$, $T = 3$ (years), $\sigma = .2$, $r = .08$, and $\mu = .12$, and has reported values of the coefficients $q_m^{(3)}, Q_m^{(3)}$ $(m = 1, 2, 3)$ in the three-piece exponential curve (denoted by EXP3) to approximate the early exercise boundary. Using these values, Figure 2 plots Ju's approximation and the "exact" (benchmark) boundary, which is computed by the hybrid approach described in Section 3 with $\delta = 10^{-4}$. Also shown in Figure 2 is our linear spline approximation with four unevenly spaced pieces (denoted by LSP4), with the first knot at $s = -.005$ followed by three evenly spaced knots. The figure shows that our approximation is close to the early exercise boundary. Although the first piece of Ju's approximation is not very close, it has little effect on the integrand in (13), as shown in Figure 3 which plots the integrands in (13) for $P = 100$ using the exact, Ju's, and our exercise boundaries, showing close agreement between them. On the other hand, if we approximate the integrand by a piecewise constant function using four pieces as in Huang, Subrahmanyam, and Yu (1996, p. 287), then its graph (denoted by HSY4), also shown in Figure 3, differs substantially from that of the integrand associated with the exact boundary, explaining why their method performs worse than Ju's as shown in Ju's Table 2. Also shown in Figure 2 is the step function approximation of Huang, Subrahmanyam, and Yu (HSY4) that differs substantially from the exact boundary.

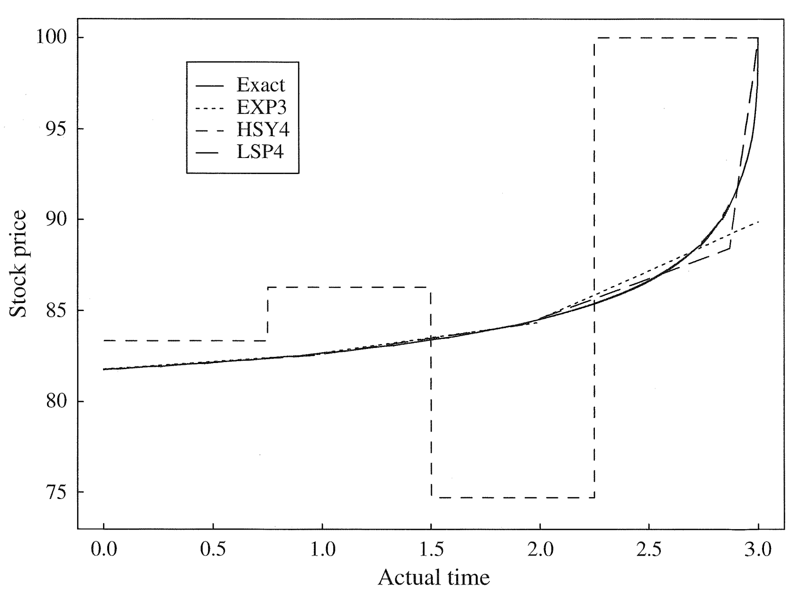

FIGURE 2. Approximations to early exercise boundary.

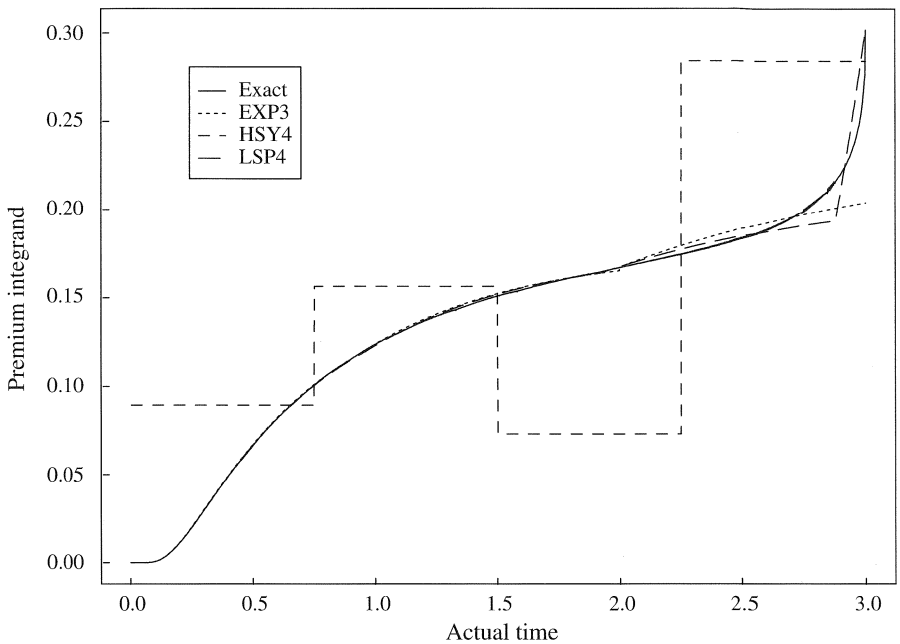

FIGURE 3. Approximations to premium integrand.

To solve for $(q_m^{(n)}, Q_m^{(n)})$ recursively in an n-piece exponential approximation to B_t, Ju needs two nonlinear equations $f_1(q_m^{(n)}, Q_m^{(n)}) = 0$ and $f_2(q_m^{(n)}, Q_m^{(n)}) = 0$, where f_1 comes directly from (3) by writing the integral on the RHS as $\sum_{i=1}^{m} \int_{\tau_i}^{\tau_{i-1}}$, and f_2 comes from the integral equation obtained by differentiating the RHS of (2) with respect to P and setting it equal to -1 (= derivative of $K - P$) when $P = B_t$. These two nonlinear equations are solved by Newton's method, which requires good starting values. In contrast, the method in Section 2 to solve (12) for $\bar{z}(s_m)$ only involves a *one-dimensional* search, which can be carried out by the Van Wijngaarden–Dekker–Brent method, a method that is numerically much more stable than Newton's method.

As indicated by Ju (1998), his n-piece exponential approximation to the early exercise boundary involves besides the $2n$ parameters $q_1^{(n)}, Q_1^{(n)}; \ldots; q_n^{(n)}, Q_n^{(n)}$ of the n pieces also the $2(n-1)$ parameters $q_1^{(1)}, Q_1^{(1)}; \ldots; q_1^{(n-1)}, Q_1^{(n-1)}$ of the 1-piece, $\ldots$, $(n-1)$-piece approximations. The reason is that he uses $(q_1^{(j-1)}, Q_1^{(j-1)})$ as the starting value for solving the two nonlinear equations defining the parameters of the function $Q_1^{(j)} \exp(q_1^{(j)} t)$ for $(j-1)T/j < t \leqslant T$. In particular, to initialize the one-piece approximation, he can rely on the good time-invariant approximation $(0, P^*)$ to the early exercise boundary provided by MacMillan (1986) and Barone-Adesi and Whaley (1987) whose method is described at the beginning of this section. For $i \geqslant 2$, $(q_{i-1}^{(n)}, Q_{i-1}^{(n)})$ is used as the starting value for solving the nonlinear equations defining $(q_i^{(n)}, Q_i^{(n)})$. This "bottom-up" approach requires the determination of P^* and $2(n-1)$ additional parameters simply to initialize $q_1^{(n)}, Q_1^{(n)}$. In contrast, our method is more direct and involves n equations, one for each $\bar{z}(s_i)$, in our n-piece linear approximation to $\bar{z}(\cdot)$.

It is natural to try to bypass Ju's bottom-up approach by initializing the solution for $(q_1^{(n)}, Q_1^{(n)})$ more directly without using $(q_1^{(n-1)}, Q_1^{(n-1)})$, which in turn involves $(q_1^{(n-2)}, Q_1^{(n-2)}), \ldots, (q_1^{(1)}, Q_1^{(1)})$, as was done by Gao, Huang, and Subrahmanyam (2000) in their extension of Ju's approximation to American barrier options. An obvious way is to initialize with the $(0, P^*)$ of MacMillan (1986) and Barone-Adesi and Whaley (1987). Another way is to initialize with $(0, B_T)$, as suggested by Gao, Huang, and Subrahmanyam (2000), in which $B_T = K\min(1, r/\mu)$. A third way is to initialize with the first piece of our piecewise linear approximation for the transformed coordinates (5), which yields the starting value $(\bar{q}, \bar{Q})$ for $(q_1^{(n)}, Q_1^{(n)})$, where

$$\bar{q} = [\bar{z}(0) - \bar{z}(s_1)]n/T + r - \mu - \tfrac{1}{2}\sigma^2, \qquad \bar{Q} = B_T e^{-\bar{q}T}, \tag{17}$$

noting that T/n is the width of each subinterval when $[0, T]$ is partitioned into n subintervals of equal width.

TABLE 3. Comparison of benchmark values (binomial) of American option prices ($K = 100$, $T = 3$, $\sigma = 0.2$, $r = 0.08$) with five fast approximations: LSP4, EXP3, Ju$(0, P^*)$, Ju$(0, B_T)$, Ju$(\bar{q}, \bar{Q})$). The last three represent initializing Ju's method at three alternative starting values.

μ	P	Binomial	LSP4	EXP3	Ju$(0, P^*)$	Ju$(\bar{q}, \bar{Q})$	Ju$(0, B_T)$
.12	80	25.658	25.658	25.657	25.656	25.656	*
	90	20.083	20.075	20.082	20.082	20.082	*
	100	15.498	15.505	15.497	15.498	15.497	*
	110	11.803	11.809	11.802	11.803	11.802	*
	120	8.886	8.890	8.885	8.885	8.885	*
.08	80	22.205	22.193	22.208	22.198	*	*
	90	16.207	16.201	16.211	16.198	*	*
	100	11.704	11.700	11.707	11.694	*	*
	110	8.367	8.365	8.369	8.357	*	*
	120	5.930	5.929	5.932	5.921	*	*
.04	80	20.350	20.345	20.351	20.347	20.347	*
	90	13.499	13.491	13.500	13.487	13.486	*
	100	8.944	8.939	8.947	8.931	8.930	*
	110	5.912	5.907	5.915	5.898	5.898	*
	120	3.897	3.893	3.900	3.885	3.885	*
.00	80	20.000	20.000	20.000	20.000	*	*
	90	11.697	11.698	11.699	11.692	*	*
	100	6.932	6.932	6.935	6.922	*	*
	110	4.155	4.154	4.157	4.145	*	*
	120	2.510	2.510	2.512	2.501	*	*

In his extensive numerical study comparing his method with other approximations in the literature, Ju (1998) uses three pieces and no more than three iterations in applying Newton's procedure to solve the two nonlinear equations for the (q, Q) of each piece, after initializing the procedure in the way described above. This small number of pieces and of iterations makes his method faster than many other approximations in the literature, as shown in Tables 1–4 of his paper. Since he builds the three-piece approximation "bottom-up" from the one- and two-piece approximations, it is natural for him to combine the option values obtained from the one-, two- and three-piece approximations by using the Richardson extrapolation to obtain a slightly more accurate approximation than that involving only the three-piece approximation.

Table 3 compares Ju's method involving a three-piece exponential approximation (EXP3), whose results are taken from Table 2 of Ju (1998), with our linear spline approximation LSP4. To be comparable with Ju in speed, we have limited the number of iterations in our solution of (12) to 5 in Table 3 and Figure 2. LSP4 and EXP3 are in close agreement with the benchmark values computed by the binomial tree method with 10 000 steps. Table 3 also considers the three simpler ways to initialize $(q_1^{(3)}, Q_1^{(3)})$ described above, namely, starting with $(0, P^*)$ or $(0, B_T)$, or with the $(\bar{q}, \bar{Q})$ defined in (17). It shows the superiority of Ju's elaborate choice of the starting value over other equally plausible choices, which produce either less accurate or numerically unstable results. In particular, the entries marked by $*$ in Table 3 denote results where the Newton's iterative procedure to solve Ju's simultaneous nonlinear equations is terminated because of a singular (or nearly singular) Jacobian matrix.

6. CONCLUSIONS

AitSahlia and Lai (1999) used the corrected Bernoulli walk approach introduced by Chernoff and Petkau (1986) to compute the early exercise boundary of a standard American option. In this paper, we use a different approach based on numerical solution of the integral equation defining the early exercise boundary. More importantly, by combining both approaches, we obtain a hybrid method that is an improvement over either approach in accuracy and speed. A new method is also developed to solve the integral equation numerically.

The approximately piecewise linear shape, with a few unevenly spaced pieces, of the early exercise boundary in the canonical scale, already noted by AitSahlia and Lai (1999), suggests that our new method to solve the integral equation defining the early exercise boundary can be applied to a coarse grid with a few time points to yield a fast and reasonably accurate approximation. It also explains why Ju's (1998) method that involves solving simultaneous nonlinear equations for the parameters of piecewise exponential approximations to the early exercise boundary (in the original coordinates) is both faster and more accurate than previous approximations in the literature. We have noted in Section 5 that another ingredient for the success of Ju's method is his elaborate starting value for the iterative solution of these simultaneous nonlinear equations, and that other equally plausible starting values can result in singular or ill-conditioned Jacobian matrices in the Newton-type iterations used by Ju to solve the simultaneous nonlinear equations. In contrast, our method to solve for $\bar{z}(t_i)$ involves a numerically stable one-dimensional search and has superior convergence properties.

Acknowledgements

We would like to thank Nengjiu Ju for providing us with the code he wrote to implement his method, and Wenlong Weng for helpful discussions. We also acknowledge stimulating comments from participants at the Second Conference on Quantitative Methods in Finance (Sydney, December 1998), at which some of these results were presented.

REFERENCES

AitSahlia, F., and Lai, T. L. (1999). A canonical optimal stopping problem for American options and its numerical solution. *Journal of Computational Finance*, **3**(2), 33–52.

Barone-Adesi, G., and Whaley, R. (1987). Efficient analytic approximations of American option values. *Journal of Finance*, **42**, 301–320.

Brennan, M. J., and Schwartz, E. S. (1977). The valuation of the American put option. *Journal of Finance*, **32**, 449–462.

Broadie, M., and Detemple, J. (1996). American option valuation: New bounds, approximations and a comparison of existing methods. *Review of Financial Studies*, **9**, 1211–1250.

Bunch, D., and Johnson, H. (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. (1998). Randomization and the American put. *Review of Financial Studies*, **11**, 597–626.

Carr, P., Jarrow, R., and Myneni, R. (1992). Alternative characterizations of American put options. *Mathematical Finance*, **2**, 87–106.

Chernoff, H., and Petkau, A. J. (1986). Numerical solutions for Bayes sequential decision problems. *SIAM Journal of Scientific and Statistical Computing*, **7**, 46–59.

Cox, J., Ross, S., and Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, **7**, 229–263.

Courtadon, G. R. (1982). A more accurate finite difference approximation for the valuation of options. *Journal of Financial and Quantitative Analysis*, **17**, 697–703.

Gao, B., Huang, J., and Subrahmanyam, M. G. (2000). An analytical approach to the valuation of American path-dependent options. *Journal of Economic Dynamics and Control*, **24**, 1783–1827.

Geske, R., and Johnson, H. (1984). The American put options valued analytically. *Journal of Finance*, **39**, 1511–1524.

Huang, J., Subrahmanyam, M. G., and Yu, G. G. (1996). Pricing and hedging American options: A recursive integration method. *Review of Financial Studies*, **9**, 277–300.

Jacka, S. D. (1991). Optimal stopping and the American put. *Mathematical Finance*, **1**, 1–14.

Ju, N. (1998). Pricing an American option by approximating its early exercise boundary as a multipiece exponential function. *Review of Financial Studies*, **11**, 627–646.

Kim, I. J. (1990). The analytic valuation of the American options. *Review of Financial Studies*, **3**, 547–572.

Lai, T. L., Yao, Y. C., and AitSahlia, F. (1999). Corrected random walk approximations to free boundary problems in optimal stopping. Working Paper, Stanford University.

MacMillan, L. (1986). Analytic approximation for the American put option. *Advances in Futures and Options Research*, **1**, 119–139.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. Cambridge University Press, New York.

Wu, L., and Kwok, Y.-K. (1997). A front-fixing finite difference method for the valuation of American options. *Journal of Financial Engineering*, **6**, 83–97.