ELSEVIER

Journal of Economic Dynamics and Control
22 (1998) 1419–1444

JOURNAL OF
Economic
Dynamics
& Control

Pricing the American put option: A detailed convergence analysis for binomial models

Dietmar P.J. Leisen[a,b,*]

[a] *Rheinische Friedrich-W. Univ. Bonn, Inst. für Gesellschafts- und Wirtschaftswissenschaften Adenauerallee 24-42, D-53113 Bonn, Germany*
[b] *CREST Paris, France*

Abstract

Leisen and Reimer (1996) suggested to consider the order of convergence as a measure of convergence speed for European call options. In this paper we study in a first step the problem of determining the order of convergence in pricing American put options for several approaches in the literature. We will then examine in detail extrapolation and the Control Variate technique for improving convergence and will explain their pitfalls. Since the investigation reveals the need for smooth converging models in order to get smaller initial errors, such a model is constructed. The different approaches are then tested: simulations exhibit up to 100 times smaller initial errors.

JEL classification: G13

Keywords: Binomial model; Order of convergence; Smoothing; Extrapolation; Control Variate technique

1. Introduction

In their celebrated work Black and Scholes (1973) introduced a new framework into the theory of option valuation using the notions of hedging and arbitrage-free pricing. Later Harrison and Kreps (1979) and Harrison and

* E-mail: leisen@addi.finasto.uni-bonn.de

0165-1889/98/$19.00
PII S0165-1889(98)00019-0

Pliska (1981) developed the concept of the equivalent martingale measure. This concept gave an elegant technique to express and solve pricing problems. Bensoussan (1984) and Karatzas (1988) generalized this technique to the case of the American put option. In this context the price is determined by an optimal stopping problem; the price-process can be described as the smallest supermartingale majorant to the discounted payoff ('Snell enveloppe'). This problem was already studied by McKean (1965) and transformed into a free boundary problem. Moreover he represented the stopping time in terms of the so called early-exercise boundary and the option price as a function of this boundary. Van Moerbeke (1976) derived properties of the boundary. After McKean (1965) many authors were dealing with representations of the price in terms of the boundary; a very intuitive one was given recently by Carr et al. (1992). For an overview of the state-of-the-art in continuous time we refer the reader to Myneni (1992).

Though the American put option is of great interest in practice, up to now no closed-form or analytical solution to the price nor to the boundary is known. Therefore there is an abundance of numerical work on this subject.

A straightforward approach is dealing with analytic approximations. The best known of these are quadratic approximations which were developed by MacMillan (1986) and extended by Barone–Adesi and Whaley (1987). However such approximations cannot be made arbitrarily accurate.

Another approach starts from a discretization of the partial differential equation describing the free boundary problem. This method of finite differences was originally proposed by Brennan and Schwartz (1977). Using variational inequalities the algorithm was justified completely only recently by Jaillet et al. (1990).

This paper sticks to the broad field of binomial models, the first of which was proposed by Cox et al. (1979) (CRR). They are constructed in such a way that if the time between two trading dates shrinks to zero, convergence (weakly in distribution) to their continuous counterpart is achieved. In these models American put options can be priced very easily by the Bellman principle of dynamic optimization, which is justified intuitively from arbitrage arguments.

Though in the case of European call and put options, convergence of prices is ensured very easily from weak convergence of the processes, things are much more complicated in the case of the American put option, since in general convergence of maxima over expectations on functionals on the processes – which are the prices – cannot be derived from weak convergence only (Aldous, 1981). However, a proof can be deducted from Kushner (1977) in a slightly different context and more recently from Lamberton and Pagès (1990).

There are numerous binomial approaches and extensions. One mainstream is dealing with 'better' price approximations as in CRR. Jarrow and Rudd (1983), pp. 183–188 (JR) adjusted this model to account for the local drift term. Tian (1993) argued that matching discrete and continuous local moments should yield 'better' convergence. Actually, though these works worry about better convergence, none of them resolved it fully for the lack of a proper definition.

These problems were addressed by Leisen and Reimer (1996). They measured the speed of convergence by the concept of order of convergence. It was shown a general theorem for determining it in the case of the European call option. Using this they concluded that in this sense the presented models of CRR, JR and Tian are equal; they all converge with order one. In a second step, a model with the higher order of convergence two was constructed.

In this paper, we first give a short introduction to the (discrete and continuous) models and the basic notation (Section 2). In the next step, we will then extend the theorem derived by Leisen and Reimer (1996) to the case of the American put option. This leads to determine order of convergence one for the models of CRR and one from above resp. 1/2 from below for the models of JR and Tian (Section 3). The information about the type of convergence is then used for an error representation. This allows to analyze in detail two ad hoc improvements common in practice: The Richardson extrapolation and the Control Variate technique introduced by Hull and White (1988). Since the analysis reveals the need for smoothing the convergence behavior of price calculations, we construct a new model for calculating European put option prices (Section 4). Though this model is very simple, it yields order of convergence two by extrapolation. In Section 5 we present a numerical analysis of different binomial models for the American put option. It turns out that extrapolation yields initial errors that are up to 100 times smaller than those using previous binomial models.

2. The framework

Throughout the following paper we suppose a constant interest rate $r \geq 0$ and a constant volatility $\sigma > 0$, to be given. Continuous capital markets are modelled by a stock price process $(S_t)_{t \geq 0}$ following geometric Brownian motion, i.e.:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t$$

where $(W_t)_{t \geq 0}$ is a standard Wiener process on some probability space $(\Omega, \mathscr{F}, Q)$. Please note that we immediately introduced the risk neutral probability measure Q according to Harrison and Pliska (1981).

In this model the price $P^e(t, S)$ of a European put with strike K when time-to-maturity equals $T - t$ and the stock-value equals S is the well known Black–Scholes formula:

$$P^e(t, S) = Ke^{-r(T-t)}\mathscr{N}(-d_2) - S\mathscr{N}(-d_1),$$

$$d_{1,2} = \frac{\ln(S/K) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}},$$

where $\mathcal{N}(\cdot)$ is the cumulative standard normal distribution function. Things get complicated when dealing with American put options. Suppose we are given a fixed American put with strike K and maturity date T. Denote the price function by $P^a(t,S)$. From Van Moerbeke (1976) it follows that there exists a critical stock price B_t, below which the option should always be exercised ($P^a(t,S)=(K-S)^+$ for $S \le B_t$) and above which it should never be exercised ($P^a(t,S)>(K-S)^+$ for $S > B_t$). The function $t \mapsto B_t$ is a smooth, nondecreasing function of time t which terminates in the strike price ($B_T = K$). It is called the (early-exercise) *boundary*.

The boundary separates the domain $\mathscr{D}=[0,T]\times\mathbb{R}^+$ into the continuation region $\mathscr{C}:=\{(t,S)\in\mathscr{D}\,|\,S>B_t\}$ and the stopping region $\mathscr{S}:=\{(t,S)\in\mathscr{D}\,|\,S\le B_t\}$. Binomial models are a description of discrete asset price dynamics. They specify a number n of trading dates. Trading occurs only at the equidistant spots of time $t_i^n \in \mathscr{T}^n := \{0=t_0^n,\dots,t_n^n=T\}$ with $t_{i+1}^n - t_i^n := \Delta t_n := T/n$ $(i=0,\dots,n-1)$. In order to achieve a complete market model, the one-period returns $\overline{R}_{n,i}$ $(i=1,\dots,n)$ are modelled by two point iid binomial random variables

$$\overline{R}_{n,i} = \begin{cases} u_n & \text{with probability} \quad \overline{q}_n, \\ d_n & \text{with complementary probability} \quad 1-\overline{q}_n \end{cases}$$

on a suitable probability space $(\overline{\Omega},\overline{\mathscr{F}},\overline{Q})$. Therefore the discrete asset price dynamics is $(\overline{S}_{n,k})_k$ where the price $\overline{S}_{n,k}$ at time t_k^n is described by

$$\overline{S}_{n,k} = S_0 \prod_{i=1}^{k} \overline{R}_{n,i}.$$

The specification of the one-period returns is a complete description of the discrete dynamics. We call a finite sequence $\overline{R}_n = (\overline{R}_{n,i})_{i=1,\dots,n}$ a *lattice* (*tree*). Observing that

$$P_n^a(t,S) = \max\{S, \mathrm{E}[P_n^a(t+\Delta t_n, \overline{R}_{n,1}S)]\}$$

and $P_n^a(T,\cdot)=f$, the American put option price can easily be calculated backward in time.

In the sequel we will suppose always that a whole sequence of lattices is given. One should think of it as a triangular array

$$\begin{array}{cccc} \overline{R}_{1,1} & & & \\ \overline{R}_{2,1} & \overline{R}_{2,2} & & \\ \overline{R}_{3,1} & \overline{R}_{3,2} & \overline{R}_{3,3} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

where each row represents a lattice.

A method which assigns to each refinement n a lattice is called a *lattice approach*. In order to compare it with the continuous model, denote for any $n \in \mathbb{N}$ for $i = 1, \dots, n$ by $R_{n,i}$ the continuous return between times t_i^n and t_{i+1}^n. For n fixed they are iid random variables on $(\Omega, \mathcal{F}, Q)$ such that $S_{t_k^n} = S_0 \prod_{i=1}^{k} R_{n,i}$, $\forall k = 0, \dots, n$. Several different lattice-approaches have been proposed. The model of CRR uses

$$u_n = \exp\{\sigma\sqrt{\Delta t_n}\},$$

$$d_n = \exp\{-\sigma\sqrt{\Delta t_n}\}.$$

To take into account the risk-neutrality argument of Harrison and Pliska (1981), the expected one-period return $\overline{\mathrm{E}}[\overline{R}_{n,1}]$ must be equal to the one period return of the riskless bond $r_n = \exp\{r\,\Delta t_n\}$. This amounts to setting $\overline{q}_n = (u_n - r_n)/(u_n - d_n)$. The risk-neutrality argument amounts to matching discrete and continuous first moments. Tian's parameter selection requires the second and third moments to be equal, too:

$$u_n = \frac{r_n v_n}{2}\left(v_n + 1 + \sqrt{v_n^2 + 2v_n - 3}\right),$$

$$d_n = \frac{r_n v_n}{2}\left(v_n + 1 - \sqrt{v_n^2 + 2v_n - 3}\right),$$

where

$$v_n = \exp\{\sigma^2 \Delta t_n\}.$$

JR argue in terms of gross return. Adding the local drift term $\mu'\,\Delta t_n$ yields

$$u_n = \exp\{\sigma\sqrt{\Delta t_n} + \mu'\,\Delta t_n\},$$

$$d_n = \exp\{-\sigma\sqrt{\Delta t_n} + \mu'\,\Delta t_n\},$$

where

$$\mu' = r - \sigma^2/2.$$

Moreover they have $\overline{q}_n = \frac{1}{2}$.

3. Characterization of errors

Now suppose we are given some fixed stock S_0 and a contingent claim. Denote its continuous time price by p_∞. Moreover suppose we study a lattice-approach yielding a sequence $(\overline{R}_n)_n$ of lattices. From this sequence we can calculate a sequence $(p_n)_n$ of discrete prices. We know from Kushner (1977) and

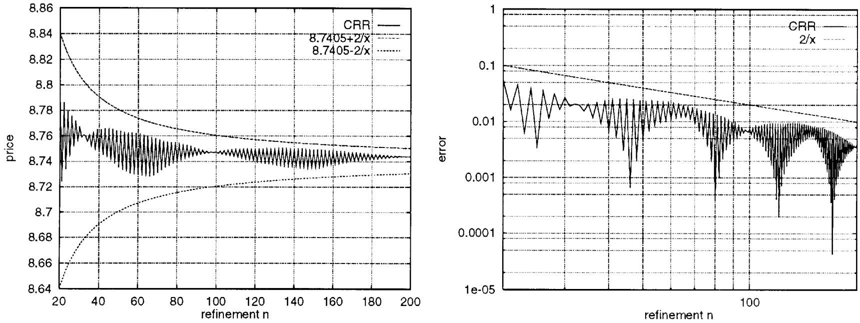

Fig. 1. Price, error and bounding error functions depending on refinement.

Lamberton and Pagès (1990) that discrete American put prices converge to the continuous price p_∞. That means, if we denote by $e_n := |p_\infty - p_n|$ the error each lattice produces, we have $\lim_{n\to\infty} e_n = 0$. (Fig. 1). A straightforward way to measure convergence speed is to compare it with those of the sequences $(1/n)_n, (1/n^2)_n, \ldots$. That is, we use the mathematical concept of 'order of convergence'. Restated in our specific case, we adopt the following:

Definition 3.1. Let $(\overline{R}_n)_n$ be a sequence of lattices. A sequence of prices $(p_n)_n$ calculated from the lattices *converges with order* $\rho > 0$ if there exists a constant $\kappa > 0$ such that

$$\forall n \in \mathbb{N}: \quad e_n \leq \kappa / n^\rho.$$

In the sequel we will often write shortly $e_n = \mathcal{O}(1/n^\rho)$ for this.

Please note that convergence of prices is implied by any order of convergence greater than 0. Moreover we remark that higher order means 'quicker' convergence. Thus the theoretical concept of order of convergence is not unique: a lattice approach with order ρ has also order $\tilde{\rho} \leq \rho$.

Though the concept of order of convergence may seem very theoretical, indeed it is easy to observe in actual simulations. Because of $\log \kappa/n^\rho = \log \kappa - \rho \log n$ the bounding function κ/n^ρ becomes a straight line with slope equal to $(-\rho)$ and shift κ on a log–log scale. So when plotting e_n on a log–log scale, determining the order of convergence consists in looking for the slope of a suitable bounding straight line. In Fig. 1 we calculated American put option prices (and their errors) of the continuous-time solution with the CRR-model with the following parameters: $S = 100$, $K = 105$, $T = 1$, $r = 0.05$, $\sigma = 0.2$. The refinement is iterated from $n = 10, \ldots, 200$. Moreover, we plotted the function $2/x$ as an upper bounding error line. This suggests that the order of convergence is equal to one.

Leisen and Reimer (1996) were looking for factors derived from the lattice approach under consideration that determine the order of convergence for European call options. The following (pseudo-)moments turned out to fulfill this.

Definition 3.2. For a sequence of lattices $(\overline{R}_n)_{n\in\mathbb{N}}$ we call for all $n\in\mathbb{N}$:

Moments:

$$\mathfrak{m}_n^1 := \overline{\mathrm{E}}[\overline{R}_{n,1} - 1] - \mathrm{E}[R_{n,1} - 1],$$

$$\mathfrak{m}_n^2 := \overline{\mathrm{E}}[(\overline{R}_{n,1} - 1)^2] - \mathrm{E}[(R_{n,1} - 1)^2],$$

$$\mathfrak{m}_n^3 := \overline{\mathrm{E}}[(\overline{R}_{n,1} - 1)^3] - \mathrm{E}[(R_{n,1} - 1)^3],$$

Pseudo-moment:

$$\mathfrak{p}_n := \overline{\mathrm{E}}[(\ln \overline{R}_{n,1})(\overline{R}_{n,1} - 1)^3].$$

These moments are mainly the differences between the ordinary moments of the discrete and continuous approaches. Therefore, they represent a generalization of the ordinary moments. The form of the pseudo-moment is of technical nature as it resulted from the proof of Theorem 3.1. Please note that $\mathfrak{m}_n^1 = 0$ from the risk neutrality argument of Harrison and Pliska (1981).

In the case here, where we have a discrete approximation of a continuous framework, it turns out that the order of convergence is determined by the difference of the ordinary moments, i.e. by that of our moments. This is exactly what Theorem 3.1 which was stated and proven in Leisen and Reimer (1996) says.

Theorem 3.1. *Let* $(\overline{R}_n)_{n\in\mathbb{N}}$ *be a sequence of lattices and* $\mathfrak{m}_n^2, \mathfrak{m}_n^3, \mathfrak{p}_n$ *its respective (pseudo-) moments. Then the order of convergence in calculating European call option prices is the smallest order contained in* $\mathfrak{m}_n^2, \mathfrak{m}_n^3$ *and* $\mathfrak{p}_n$, *reduced by* 1, *but not smaller than* 1, *i.e.*:

$$\exists \kappa(S_0, K, r, \sigma, T): e_n \leq \kappa\{n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n\}.$$

Theorem 3.2. *Under the assumptions of Theorem* 3.1 *the same results hold for European put options.*

Proof. This is an immediate consequence of put-call parity. □

Proposition 3.1. *The lattice-approaches of CRR, JR and Tian satisfy*

$$\mathfrak{m}_n^2 = \mathcal{O}(1/n^2), \qquad \mathfrak{m}_n^3 = \mathcal{O}(1/n^2), \qquad \mathfrak{p}_n = \mathcal{O}(1/n^2).$$

Proof. See the Appendix in Leisen and Reimer (1996). □

Theorem 3.2 and Proposition 3.1 immediately yield

Corollary 3.1. *European put option prices calculated using the lattice-approaches of CRR, JR and Tian converge with order one.*

The question whether it would be possible to strengthen the result of Theorem 3.2 in order to prove higher order of convergence may now arise. We now state a Theorem which says that this bound is actually the best achievable. The idea and the proof are from David Heath.

Theorem 3.3. *Suppose a fixed initial stock-price S, interest rate r and volatility σ, as well as a sequence $(\overline{R}_n)_{n\in\mathbb{N}}$ of lattices with $u_n/d_n = 1 + \mathcal{O}(\sqrt{\Delta t_n})$. Then there exists a strike price K such that prices calculated for this European put option have error $e_n \geq c^*/n$ for a suitable constant $c^* \in \mathbb{R}$.*

Proof. Holding all other parameters fixed, this proof will study the dependence of the price $P^{\mathrm{e}}(K)$ of a European put option on its strike. It is a strictly convex function. We deduce from this that there are some $C > 0$, $K_1 < K_2$ such that

$$\frac{\partial^2 P^{\mathrm{e}}(K)}{\partial K^2} \geq C \quad \text{for all } K \in [K_1, K_2].$$

Let $K^* = (K_1 + K_2)/2$ and I_n denote the interval between successive terminal stock prices which contain K^*, where we suppose a sufficiently high refinement $n \geq n_0$.

Then

$$|I_n| \geq \left(\frac{u_n}{d_n} - 1\right)K^*.$$

P_n^{e} is a linear function in the strike price K on the interval I_n. Therefore

$$\frac{\partial^2 P_n^{\mathrm{e}}(K)}{\partial K^2} = 0 \quad \text{on } I_n.$$

Let $e_n(K) := P^{\mathrm{e}}(K) - P_n^{\mathrm{e}}(K)$ denote the error depending on the strike price K. Then

$$\frac{\partial^2 e_n}{\partial K^2} = \frac{\partial^2 P^{\mathrm{e}}(K)}{\partial K^2} \geq C \quad \text{on } I_n \subset [K_1, K_2].$$

Integrating twice yields

$$\sup_{K\in I_n} |e_n| \geq \frac{|I_n|^2}{16} \inf_{K\in I_n} \left|\frac{\partial^2 e_n}{\partial K^2}\right| \geq \frac{C\ |I_n|^2}{16} \geq \frac{C\ |K^*|^2}{16}\left(\frac{u_n}{d_n}-1\right)^2 \geq \frac{C^*}{n}$$

for a suitable constant C^*. □

Obviously the lattice approaches of CRR, JR and Tian fulfill the condition $u_n/d_n = 1 + \mathcal{O}(\sqrt{\Delta t_n})$ in Theorem 3.3. The following two theorems will state a result similar to that of Theorem 3.2 for the American put option.

Theorem 3.4. *Let* $(\overline{R}_n)_{n\in\mathbb{N}}$ *be a sequence of lattices and* $\mathfrak{m}_n^2, \mathfrak{m}_n^3, \mathfrak{p}_n$ *its respective (pseudo-) moments. Then there exists a constant* $\kappa_u(S_0, K, r, \sigma, T)$ *such that*

$$P^{\mathrm{a}}(0, S_0) - P_n^{\mathrm{a}}(0, S_0) \leq \kappa_u\{n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n\}.$$

Proof. Denote by $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{Q})$ the product of $(\Omega, \mathscr{F}, Q)$ and $(\overline{\Omega}, \overline{\mathscr{F}}, \overline{Q})$. For all $n \in \mathbb{N}$ and $k = 0, \ldots, n$ let $\mathscr{A}_{n,k} = \sigma(\overline{S}_{n,i} \,|\, i \leq k)$ denote the information structure. From Carr et al. (1992) we know that the price of the American put can be decomposed into the price of a European put and the early-exercise premium π, which takes the form

$$\pi = rK \int_0^T \mathrm{e}^{-rt'} \mathcal{N}(b_{2,0}(S_0, t'))\, \mathrm{d}t'$$

where $b_{2,0}(x,t') = [\ln Bt/x - (r - \sigma^2/2)t']/\sigma\sqrt{t'}$. Lemma A.2 in the Appendix tells us that stopping the discrete process $(\overline{S}_{n,k})_{k=0,\ldots,n}$ according to the rule $(B_{t_k^n})_{k=0,\ldots,n}$ yields the premium

$$\pi_n^B = \sum_{k=0}^{n-1} \mathrm{e}^{-rt_k^n} K(1 - \mathrm{e}^{-r\Delta t}) \hat{Q}[\overline{S}_{n,k} < B_{t_k^n}] + \mathcal{O}(\Delta t_n).$$

The optimal stopping policy, however, yields the higher premium π_n. Therefore, we have according to Lemma A.5 in the Appendix:

$$\exists \kappa_u(S_0, K, r, \sigma, T)\colon \pi - \pi_n \leq \kappa_u \Delta t_n.$$

Since

$$|P^{\mathrm{a}}(0, S_0) - P_n^{\mathrm{a}}(0, S_0)| \leq |P^{\mathrm{e}}(0, S_0) - P_n^{\mathrm{e}}(0, S_0)| + |\pi - \pi_n|.$$

Theorem 3.4 now follows immediately from Theorem 3.1. □

Theorem 3.5. Let $(\overline{R}_n)_{n\in\mathbb{N}}$ be a sequence of lattices and $\mathfrak{m}_n^2, \mathfrak{m}_n^3, \mathfrak{p}_n$ its respective (pseudo-)moments. Then there exists a constant $\kappa_1(S_0, K, r, \sigma, T)$ such that

$$P^{a}(0,S_0) - P_n^{a}(0,S_0) \geq \kappa_1\left\{n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + \frac{1}{\sqrt{n}}\right\}.$$

If $(\overline{R}_n)_n$ is constructed according to the CRR lattice approach, then we have the stronger result that there exists a constant $\kappa_1(S_0, K, r, \sigma, T)$ such that

$$P^{a}(0,S_0) - P_n^{a}(0,S_0) \geq \kappa_1\{n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n\}.$$

Proof. According to Lemma A.6 in the Appendix:

$$\mathcal{O}(\Delta t_n) = \hat{\mathrm{E}}[P^{a}(t_k^n, \overline{S}_{n,k}) - P_n^{a}(t_k^n, \overline{S}_{n,k})].$$

According to Lemma A.1 in the Appendix we see that the right-hand side of this expression is equal to

$$\begin{aligned}
&\hat{\mathrm{E}}\Bigg[P^{e}(t_k^n, \overline{S}_{n,k}) + rK\int_{t_k^n}^{T} \mathrm{e}^{-r(t-t_k^n)}\mathcal{N}(b_{2,t_k^n}(\overline{S}_{n,k}, t))\,\mathrm{d}t - P_n^{e}(t_k^n, \overline{S}_{n,k})\\
&\quad - K(1-\mathrm{e}^{-r\Delta t_n})\sum_{i=k}^{n} \mathrm{e}^{-r(t_i^n - t_k^n)}\hat{Q}[\overline{S}_{n,i} \leq \overline{B}_{n,i}]\Bigg] + \mathcal{O}(\Delta t_n)\\
&\quad = \hat{\mathrm{E}}[P^{e}(t_k^n, \overline{S}_{n,k}) - P^{e}(t_k^n, S_{t_k^n})]\\
&\qquad + \underbrace{\hat{\mathrm{E}}[P^{e}(t_k^n, S_{t_k^n})]}_{=P^{e}(0,S_0)} - \underbrace{\hat{\mathrm{E}}[P_n^{e}(t_k^n, \overline{S}_{n,k})]}_{=P_n^{e}(0,S_0)}\\
&\qquad + rK\int_{t_k^n}^{T} \hat{\mathrm{E}}[\mathrm{e}^{-r(t-t_k^n)}\mathcal{N}(b_{2,t_k^n}(\overline{S}_{n,k},t))]\,\mathrm{d}t\\
&\qquad - K(1-\mathrm{e}^{-r\Delta t_n})\sum_{i=k}^{n} \mathrm{e}^{-r(t_i^n - t_k^n)}\hat{Q}[\overline{S}_{n,i} \leq \overline{B}_{n,i}]\\
&\qquad + \mathcal{O}(\Delta t_n)\\
&\quad = \hat{\mathrm{E}}[P^{e}(t_k^n, \overline{S}_{n,k}) - P^{e}(t_k^n, S_{t^n_k})]\\
&\qquad + P^{a}(0,S_0) - P_n^{a}(0,S_0)
\end{aligned}$$

$$- rK\int_0^{t_k^n} e^{-rt}\mathcal{N}(b_{2,0}(S_0,t))\,dt + K(1-e^{-r\Delta t_n})\sum_{i=0}^{k} e^{-rt_i^n}\hat{Q}[\overline{S}_{n,i} \le \overline{B}_{n,i}]$$
$$+\ \mathcal{O}(\Delta t_n).$$

The proof of Theorem 1 in Leisen and Reimer (1996) contains as a special case the estimation of

$$\hat{\mathrm{E}}[P^e(t_k^n,\overline{S}_{n,k}) - P^e(t_k^n,S_{t_k^n})] = \mathcal{O}(n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n).$$

The assertion now follows immediately from an application of the trapezoidal formula of numerical integration (as in the proof of Lemma A.5 in the Appendix) and from Lemmata A.3 and A.7 in the Appendix. □

Theorems 3.4 and 3.5 together with Proposition 3.1 imply immediately the following two corollaries:

Corollary 3.2. *American put option prices calculated using the lattice approach of CRR converge with order one.*

Corollary 3.3. *American put option prices calculated using the lattice approaches of JR and Tian converge with order one from above and order* 1/2 *from below.*

These results improve on that of Lamberton (1995), who proved in the case of the CRR model for the lower bound an order of 2/3 and for the upper bound, 1/2. Moreover our results apply to general lattice approaches.

4. How to decrease errors properly

Actually error pictures like Fig. 1 and simulations performed by Broadie and Detemple (1996) suggest that the order of convergence is also one for the models of JR and Tian. We will subsequently assume that this holds. Then the results in the previous section tell us that for a certain class of models, calculating either American or European put option prices, the error e_n has the form $\kappa_1(n)/n$ + *higher-order terms* for a suitable bounded function κ_1.

To take advantage of this information, let us suppose in a first approximation that $p_n = \kappa_1/n + p_\infty$. For any given refinement n this equation contains two unknowns: the constant κ_1 and the correct value p_∞. In order to find a unique solution, we need a pair of refinements (n_1,n_2) with $n_2 > n_1$ and corresponding prices (p_{n_1},p_{n_2}). Denoting the approximation for p_∞ by $p_{(n_1,n_2)}$ we have the following system of equations:

$$\kappa_1/n_1 + p_{(n_1,n_2)} = p_{n_1},$$
$$\kappa_1/n_2 + p_{(n_1,n_2)} = p_{n_2}.$$

Resolving yields

$$p_{(n_1,n_2)} = p_{n_2} - \frac{(p_{n_1} - p_{n_2})n_1}{n_2 - n_1}$$

$$= \frac{n_2 p_{n_2} - n_1 \ p_{n_1}}{n_2 - n_1}.$$

We will refer to this as the extrapolation rule. Unless otherwise stated, we take the pair $(n, 2n)$. This is commonly referred to as the Richardson extrapolation (Kloeden and Platen, 1992).

The above analysis needs to be refined for two reasons. The first stems from the fact that in general the constant will depend on the refinement, whereas above, we replaced the function $\kappa_1(n)$ by a constant κ_1. The second stems from the *higher-order terms*. Since these may distort extrapolation, our rule may no longer be optimal. Therefore, a detailed analysis of the error $e(n_1, n_2) = p_{(n_1, n_2)} - p_\infty$ is needed.

Proposition 4.1. *Suppose* $e_n = \kappa_1(n)/n + \kappa_2(n)/n^2$ *where* $\kappa_1, \kappa_2 : \mathbb{N} \to \mathbb{R}$ *are suitable functions. Then*

$$e_{(n_1,n_2)} = \frac{\kappa_1(n_2) - \kappa_1(n_1)}{n_2 - n_1} + \frac{n_1\kappa_2(n_2) - n_2\kappa_2(n_1)}{n_1 n_2 (n_2 - n_1)}.$$

Proof. It is obvious that extrapolation yields the error:

$$e_{(n_1,n_2)} = \frac{n_2 \ e_2 \ - n_1 e_1}{n_2 \ - n_1}$$

$$= \frac{n_2\left(\frac{\kappa_1(n_2)}{n_2} + \frac{\kappa_2(n_2)}{n_2^2}\right) - n_1\left(\frac{\kappa_1(n_1)}{n_1} + \frac{\kappa_2(n_1)}{n_1^2}\right)}{n_2 \ - n_1}.$$

The statement of the proposition follows immediately from this. □

From Corollary 3.2 in the previous section it is clear that for the lattice approaches of CRR, $b_\mathrm{l}^\mathrm{a} := \liminf_{n\to\infty}\kappa_1(n)$ and $b_\mathrm{u}^\mathrm{a} := \limsup_{n\to\infty}\kappa_1(n)$ exist and are finite. For the approaches of JR and Tian it follows from Corollary 3.3 only that b_u^a is finite. However, according to the assumption at the beginning of this section, we have b_l^a finite, too. The proposition tells us that the absolute first-order error resulting from extrapolation is bounded by $|e_{(n_1,n_2)}| \le (b_\mathrm{u}^\mathrm{a} - b_\mathrm{l}^\mathrm{a})/(n_2 - n_1)$. For the Richardson extrapolation we get the error estimate $|e_{(n,2n)}| < (b_\mathrm{u}^\mathrm{a} - b_\mathrm{l}^\mathrm{a})/n$. This means that extrapolation replaces the constant $|b_\mathrm{u}^\mathrm{a}| \vee |b_\mathrm{l}^\mathrm{a}|$ by $|b_\mathrm{u}^\mathrm{a} - b_\mathrm{l}^\mathrm{a}|$. Therefore extrapolating makes sense only if

$|b_{\rm u}^{\rm a} - b_{\rm l}^{\rm a}| < \max\{|b_{\rm u}^{\rm a}|, |b_{\rm l}^{\rm a}|\}$ and our aim in constructing new models should be to get models with a very small $|b_{\rm u}^{\rm a} - b_{\rm l}^{\rm a}|$.

Please note that this observation explains the (obvious) fact that for the CRR model extrapolation does not make sense, since there we typically have $b_{\rm l}^{\rm a} < 0 < b_{\rm u}^{\rm a}$ yielding $|b_{\rm u}^{\rm a} - b_{\rm l}^{\rm a}| > b_{\rm u}^{\rm a}$. The same holds for the JR and Tian models.

Actually there is an optimal case, in which $b_{\rm u}^{\rm a} = b_{\rm l}^{\rm a}$. If κ_2 is bounded, an immediate consequence of Proposition 4.1 is that extrapolated prices converge with order two. Whereas in general we need to select n_2 such that $n_2 - n_1 = \mathcal{O}(n_1)$ to get a series of extrapolated prices converging to the true price p_∞, in this special case it is possible to select n_2 such that $n_2 - n_1 = const.$ and still get convergence of prices. Under the additional assumption that $\kappa_2 = const.$ we even get the scheme converging with order two. This is very interesting since the extra amount of computation time needed for extrapolation becomes comparable to that needed to calculate the price for n_1. We should therefore try to construct new models with $b_{\rm u}^{\rm a} = b_{\rm l}^{\rm a}$, for which the error picture looks 'smooth'. This is why we loosely speak of smoothing options when constructing better performing models.

Another major approach for improving results is the Control Variate technique (CV) proposed by Hull and White (1988). This technique uses the same lattice with refinement n to calculate the price approximations $P_n^{\rm a}$ of the American and $P_n^{\rm e}$ of the European put option. It is inspired by the observation that the order of convergence is the same for the European and American put. Then it is *assumed* that errors to the true prices are approximately equal, i.e.

$$P^{\rm a}(0,S_0) - P_n^{\rm a}(0,S_0) \approx P^{\rm e}(0,S_0) - P_n^{\rm e}(0,S_0)$$

$$\Rightarrow P_n^{\rm a}(0,S_0) \approx P_n^{\rm a}(0,S_0) + P^{\rm e}(0,S_0) - P_n^{\rm e}(0,S_0).$$

However, looking closely at the errors we immediately get the following.

Proposition 4.2. *Suppose* $e_n^{\rm a} = \kappa_1^{\rm a}(n)/n$, $e_n^{\rm e} = \kappa_1^{\rm e}(n)/n$ *where* $\kappa_1^{\rm a}, \kappa_1^{\rm e}: \mathbb{N} \to \mathbb{R}$ *are suitable functions. Then*

$$e_n^{\rm CV} = \frac{\kappa_1^{\rm a}(n) - \kappa_1^{\rm e}(n)}{n}.$$

The price calculated using the CV technique will be a fine estimate only if good and bad price approximations follow at the same rhythm for European and American puts. However, in general this will not hold. In order to perform a similar analysis as previously done for extrapolation we deduce from Theorem 1 of Leisen and Reimer (1996) (see Theorems 3.1 and 3.2) that $b_{\rm l}^{\rm e} := \liminf_{n\to\infty} \kappa_1(n)$ and $b_{\rm u}^{\rm e} := \limsup_{n\to\infty} \kappa_1(n)$ exist and are finite. Then

$$|e_n^{\rm CV}| \le \frac{(|b_{\rm u}^{\rm a}| \vee |b_{\rm u}^{\rm e}|) - (|b_{\rm l}^{\rm a}| \wedge |b_{\rm l}^{\rm e}|)}{n}.$$

Therefore the CV technique replaces the constant $|b^{a}_{u}| \vee |b^{a}_{l}|$ by $(|b^{a}_{u}| \vee |b^{e}_{u}|) - (|b^{a}_{l}| \wedge |b^{e}_{l}|)$ and all the conclusions drawn from Proposition 4.1 for extrapolation carry over to the CV technique. Specifically we also have to smooth the option, i.e. reduce price oscillations as much as possible, in order to get better performing models. In the sequel we will consider only extrapolation and show a way of how to smooth the option at least partially.

We have according to Carr et al. (1992) and Lemma 1 in the Appendix:

$$P^{a}(0, S_0) = P^{e}(0, S_0) + rK \int_0^T e^{-rt} Q[S_t \leq Bt]\, dt,$$

$$P^{a}_{n}(0, S_0) = P^{e}_{n}(0, S_0) + rK \sum_{j=0}^{n} e^{-rt^{n}_{k}} \overline{Q}[\overline{S}_{n,j} \leq \overline{B}_{n,j}] + \mathcal{O}(\Delta t_n).$$

This means that errors result both from approximating the European put component as well as from the early exercise premium, whereas the errors in the early exercise premium component result from approximating the value of the cash-or-nothing options $Q[S_t \leq B_{t^{n}_{j}}] - \overline{Q}[\overline{S}_{n,j} \leq \overline{B}_{n,j}]$.

With barrier option valuation, Derman et al. (1995) argue that price oscillations result from the fact that a specific lattice under consideration implicitly determines the class of possible option contracts which can be priced, since exercise is only possible at nodes in the tree grid. They call this the 'quantization error'. More specifically, in the case of the European call option, Leisen and Reimer (1996) determined as the origin of these errors the following: when taking a close look at terminal nodes, especially at the nodes around the strike price K, we see that with varying n, nodes shift upwards and downwards. Since they contain the whole probability mass, this causes the distortions.

Improving results for cash-or-nothing options is difficult, since we do not know the exercise boundaries B, resp. $\overline{B}$. However, we can profit from this observation in constructing a model which improves at least the European put component. This can be done by ensuring that the strike always lies fixed at a specific node, the center of the tree. In order to do this consistently we must assume that n is even. Therefore, suppose we are given a refinement n with n even and u_n, d_n according to CRR, i.e. $u_n = \exp\{\sigma\sqrt{\Delta t_n}\}$, $d_n = 1/u_n$. Remember JR who adjusted the local drift term to match the continuous drift term. We are interested in fixing the strike at the center of the tree at maturity. Thus the new parameter selection u'_n, d'_n should fulfill

$$u'_n = u_n e^{c_n},$$

$$d'_n = d_n e^{c_n},$$

$$S_0 (u'_n d'_n)^{n/2} = K.$$

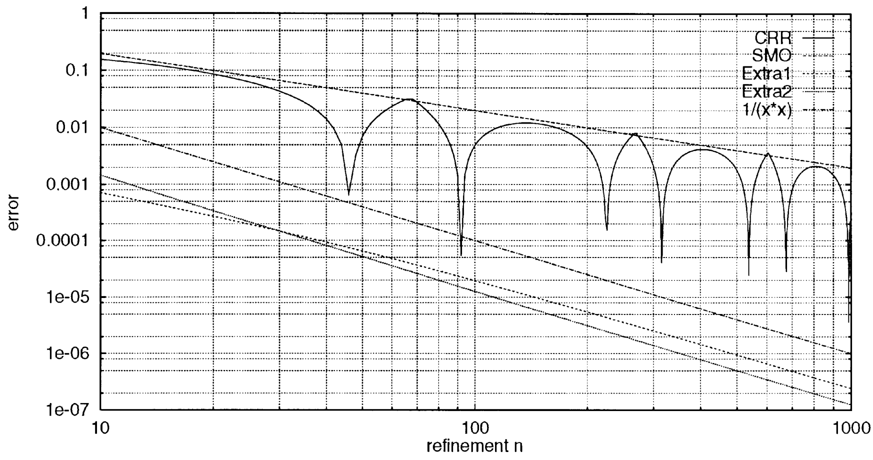

Fig. 2. Error study using CRR, SMO and two extrapolations of SMO.

The third equation tells us $c_n = \ln(K/S_0)/n$. The equivalent martingale measure is obtained by setting $\overline{q}'_n = (r_n - d'_n)/(u'_n - d'_n)$.

In the sequel this model will be referred to as SMO.

In Fig. 2 we plotted the error for the CRR and SMO models in calculating European put option prices with the following parameters: $S = 100$, $K = 105$, $T = 1$, $r = 0.05$, $\sigma = 0.2$, iterating refinement $n = 10, \dots, 1000$. Since SMO is defined only for n even, we restricted ourselves to even refinements. Moreover both approaches are compared to the following two extrapolations of SMO: EXTRA1 uses the pair $(n, 2n)$, while EXTRA2 uses the pair $(n, n + 40)$. Actually we see that the SMO model performs badly, i.e. it always yields higher errors than CRR. This stems from the fact that we did not care for implementing the local variance properly.

It would be possible to construct smooth models which behave much better. However we preferred this approach for its very simple and intuitive construction. Moreover, we are able to see very drastically the effects of extrapolation, since the 'slow' convergence speed of SMO disappears completely. When comparing it with the function $1/x^2$ we see that both extrapolations converge with order two in the long run. That is, our remarks made after Proposition 4.1, which told us that it is possible to improve by one the order of convergence, even with the simple EXTRA2, are completely justified.

For the American put option we may not expect to get smooth results, but at least 'smoother ones', replacing the constant by a much smaller one. This means that we are starting with a lower initial error, which means in turn, that a lower refinement already yields the same precision level. A comparison of different lattice approaches for the American put will be studied in more detail in the next section.

5. Numerical evaluation

The simulations in the previous sections all clung to fixed parameter constellations (σ, T, S, K, r). They give us a very good intuition about the convergence behavior. However it is quite difficult to evaluate suitably the additional time needed for extrapolation. Moreover financial institutions being faced typically by a large book of options to price, would like to get an idea of the overall error of this and the time necessary to price the options. Therefore, we will now perform a simulation study suggested by Broadie and Detemple (1996).

To start we choose a sample $\mathscr{S} = \{s_i \mid i \in I\}$ of parameter constellations $s_i = (\sigma_i, T_i, S_i, K_i = 100, r_i)$. Each of these samples represents an option contract. It has price p_i which we calculate using CRR with a refinement of 15,000 steps. In order to evaluate the accuracy of a specific lattice approach and a refinement n we proceed as follows: For each $s_i \in \mathscr{S}$ we calculate a price approximation $\hat{p}_i$ and a relative error $e_i = (\hat{p}_i - p_i)/p_i$. Relative errors do not change if S and K are scaled by the same factor, i.e. only the ratio S/K is of interest. Therefore, it is a suitable restriction to set $K = 100$ in the whole sample. Calculating the relative root-mean-squared (RMS) error

$$RMS = \sqrt{\left(\sum_{i \in I} e_i^2\right) \Big/ |I|}$$

over the sample $\mathscr{S}$ gives us for each lattice approach and refinement n a measure of its accuracy.

Computation speed is expressed by the number of option prices calculated per second. Since we stick to tree models with identical structure except for the tree parameters, we need not care for tuning the computer implementation of our methods.

Then we plot for each lattice approach and refinement accuracy against computation time. By connecting the points for one lattice approach we get a line. We choose the following distribution of parameters for the whole sample. Volatility is distributed uniformly between 0.1 and 0.6. Time-to-maturity is with probability 0.75 uniform between 0.1 and 1.0 yr and with probability 0.25 uniform between 1.0 and 5.0 yr. 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. 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. Each parameter is selected independently of the others. This selection of parameters matches the choice of Broadie and Detemple (1996) except for dividends which we do not consider here. To make relative errors meaningful, that is to avoid senseless distortions because of very small option prices, options with $c_i \leq 0.50$ did not enter the sample. In the total we have a sample of 1116 option contracts. We limited our study to refinements $n =$ 24, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000.

We compared the CRR model with the SMO model and its extrapolation. Moreover we measured them against the PP model suggested by Leisen and Reimer (1996) and its extrapolation. The latter model was constructed using the works of Pratt (1968) and Peizer and Pratt (1968) on inverted normal approximations, so as to yield order of convergence two for the European put option.

To account for different behavior with long/short maturities respectively in/out-of-the money options, we split the whole sample $\mathscr{S}$ into 4 samples $\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_3, \mathscr{S}_4$. The first two samples $\mathscr{S}_1, \mathscr{S}_2$ contain those parameter constellations with a time-to-maturity of $T \leq 0.2$. $\mathscr{S}_1$ contains those constellations representing out-of-the-money American put options ($S \geq 100$). This represents a sample of 89 parameter constellations. Simulation results for the sample $\mathscr{S}_1$ are presented in Fig. 3. We see that SMO yields results that are 3 times worse than CRR, whereas PP yields 10 times better results than CRR. Surprisingly however, extrapolating SMO and PP yields results that are again approximately 10 times better than PP, i.e. in total they have an initial error 100 times lower than CRR. Moreover, we see immediately that extrapolation has a

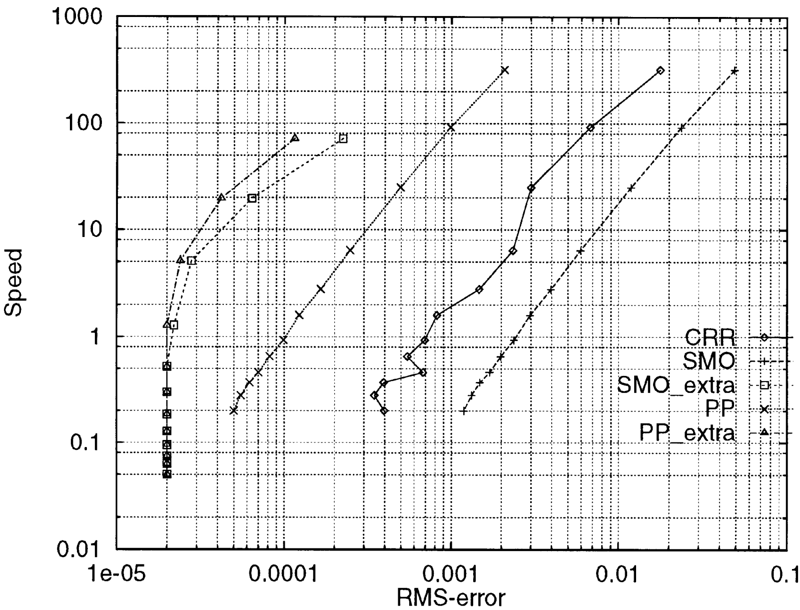

Fig. 3. Efficiency for out-of-the-money American put option contracts with short time-to-maturity.

tremendous effect on the error, since using it in a 200 step (together with a 400 step) tree exceeds already the precision level of a CRR tree with a refinement of 15,000, such that we could have dropped higher extrapolations. The second sample $\mathscr{S}_2$ contains in-the-money American put option contracts ($S \leq 100$) (150 contracts) and is presented in Fig. 4. Here the effects of extrapolation are still astonishing. Although extrapolating the PP and SMO models yields only 3 times better results than PP, this yields 10 times better results than CRR. Thus we are winning a factor 30 in comparison to CRR.

The last two samples $\mathscr{S}_3$ and $\mathscr{S}_4$ are similarly organized and plotted in Figs. 5 and 6. They deal with options with a long time-to-maturity $T \geq 0.2$. In the case of out-of-the-money options (Fig. 5) we see that PP performs 3 times better than CRR and that extrapolating PP and SMO improves this again by a factor of 3 in comparison to PP, yet. Therefore, the latter performs approximately 10 times better than CRR. In the case of in-the-money options (Fig. 6) extrapolation of PP and SMO improves the results by a factor of 2 in comparison to CRR, whereas PP shows only an improvement of 1.5.

Generally spoken, out-of-the-money options converge much smoother and therefore yield much better convergence results with extrapolation. Moreover, we want to remark that extrapolation with $n = 24$ actually ensures in all cases that the error is less than 0.01. This means that we already attain a sufficiently

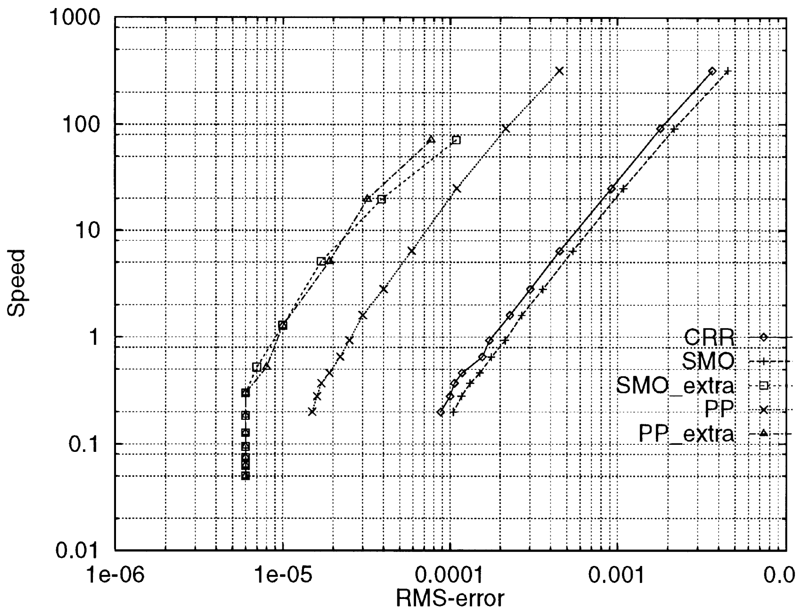

Fig. 4. Efficiency for in-the-money American put option contracts with short time-to-maturity.

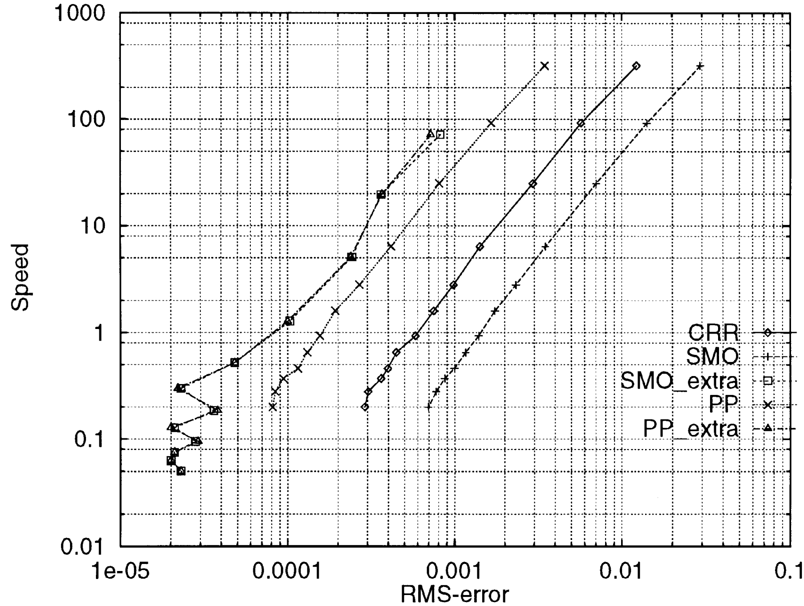

Fig. 5. Efficiency for out-of-the-money American put option contracts with long time-to-maturity.

high precision level, since in practice the results from discrete and continuous models can no longer be distinguished.

6. Conclusion

In this paper we examined the order of convergence for price calculations of the American put option. The results of Leisen and Reimer (1996) were extended to the American put option. It was thus shown that the models of CRR, JR and Tian are similar. In a next step we used this for an extrapolation rule and its error analysis. Here we saw the astonishing effects that a proper extrapolation may have. Actually, although the approach we have taken here is rather simple it already yields up to 100 times better results than the existing approaches. Better smoothing should improve this further and achieve order of convergence two as in the case of the European put option.

Acknowledgements

I am grateful to David Heath, Damien Lamberton, Stylianos Perrakis, Matthias Reimer and Dieter Sondermann for comments and assistance. Financial

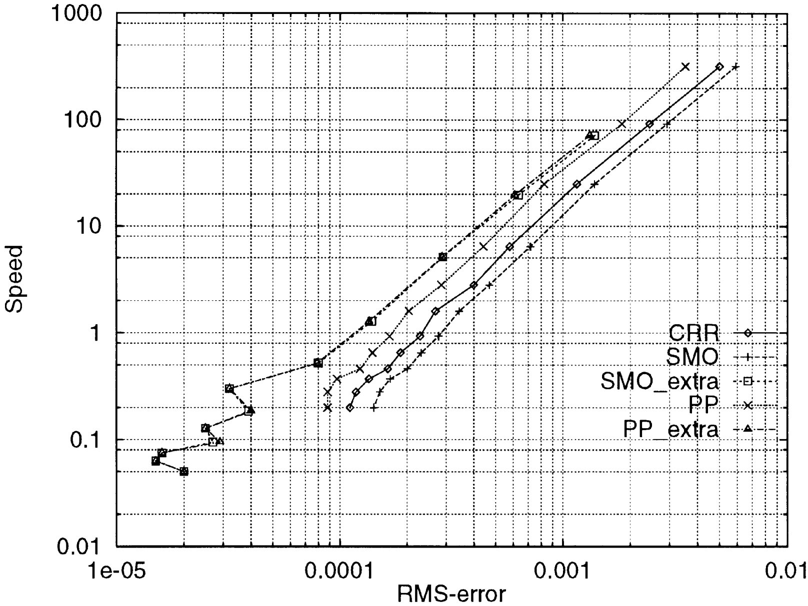

Fig. 6. Efficiency for in-the-money American put option contracts with long time-to-maturity.

support from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at Rheinische Friedrich-Wilhelms-Universität Bonn, is gratefully acknowledged.

Appendix A.

Let $\mathscr{C} := \{(n,k) | k = \min\{i | B_{t^n_i} < \overline{B}_{n,i}\}$ if this set is not empty$\}$, $\mathscr{D}_{n,k} := \{S_0 u_n^j d_n^{n-j} | 0 \le j \le k\}$, and denote by $\mathfrak{n}$ the density of the normal distribution function. For $i = 0, \ldots, n$ we will call $I_{n,i} := \{t^n_k \in \mathscr{T}^n \,|\, S_0 u_n^i d_n^{n-i} \le B_{t^n_k} < S_0 u_n^{i+1} d_n^{n-(i+1)}\}$ a domain. Domains are disjoint intervals, i.e. $I_{n,i} = [l_{n,i}, r_{n,i}[$ for suitable $l_{n,i} < r_{n,i}$. Obviously, we have $l_{n,0} < r_{n,0} = l_{n,1} < \cdots < r_{n,i}$.

Lemma A.1.

$$\pi_n(t^n_i, \overline{S}_{n,i}) = \hat{\mathrm{E}}\left[\sum_{k=0}^{n-1} \mathrm{e}^{-rt^n_k} K(1 - \mathrm{e}^{-r\Delta t_n}) \hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}] | \overline{S}_{n,i}\right] + \mathscr{O}(\Delta t_n).$$

Proof. For the cases $\overline{S}_{n,k} \le \overline{B}_{n,k}$, $u_n \overline{S}_{n,k} > \overline{B}_{n,k+1}$ we have

$$P_n^{\mathrm{a},B}(t_{k+1}^n, \overline{R}_{n,k+1}\overline{S}_{n,k}) = f(\overline{R}_{n,k+1}\overline{S}_{n,k}) + \mathcal{O}(\sqrt{\Delta t_n}) \quad \text{since } u_n - 1 = \mathcal{O}(\sqrt{\Delta t_n})$$

$\Rightarrow \hat{\mathrm{E}}[(K - \overline{S}_{n,k})^+ - \mathrm{e}^{-r\,\Delta t_n} P_n^{\mathrm{a}}(t_{k+1}^n, \overline{R}_{n,k+1}\overline{S}_{n,k}) | \mathscr{A}_{n,k}] = \mathcal{O}(\sqrt{\Delta t_n}) K(1 - \mathrm{e}^{-r\Delta t_n}) = \mathcal{O}(\sqrt{\Delta t_n}^3)$ since $1 - \mathrm{e}^{-r\Delta t_n} = \mathcal{O}(\Delta t_n)$.

If $\overline{S}_{n,k} \le \overline{B}_{n,k}$, $u_n \overline{S}_{n,k} \le \overline{B}_{n,k+1}$

$$\begin{aligned} &\hat{\mathrm{E}}[(K - \overline{S}_{n,k})^+ - \mathrm{e}^{-r\,\Delta t_n} P_n^{\mathrm{a}}(t_{k+1}^n, \overline{R}_{n,k+1}\overline{S}_{n,k}) | \mathscr{A}_{n,k}] \\ &\quad = \hat{\mathrm{E}}[(K - \overline{S}_{n,k}) - \mathrm{e}^{-r\,\Delta t_n}(K - \overline{R}_{n,k+1}\overline{S}_{n,k}) | \mathscr{A}_{n,k}] \\ &\quad = (K - \overline{S}_{n,k}) - \mathrm{e}^{-r\,\Delta t_n}(K - \mathrm{e}^{r\,\Delta t_n}\overline{S}_{n,k}) = K(1 - \mathrm{e}^{-r\,\Delta t_n}). \end{aligned}$$

From the arguments of Harrison and Pliska (1981) we have

$$\begin{aligned} \pi_n &= P_n^{\mathrm{a}}(0, S_0) - \mathrm{e}^{-rT}\hat{\mathrm{E}}[(K - \overline{S}_{n,n})^+] \\ &= P_n^{\mathrm{a}}(0, S_0) - \mathrm{e}^{-rT}\hat{\mathrm{E}}[\underbrace{P_n^{\mathrm{e}}(T, \overline{S}_{n,n})}_{=P_n^{\mathrm{a}}(T, \overline{S}_{n,n})}] \\ &= \hat{\mathrm{E}}\left[\sum_{k=0}^{n-1} \mathrm{e}^{-rt_k^n}\hat{\mathrm{E}}[P_n^{\mathrm{a}}(t_k^n, \overline{S}_{n,k}) - \mathrm{e}^{-r\,\Delta t_n} P_n^{\mathrm{a}}(t_{k+1}^n, \overline{S}_{n,k+1}) | \mathscr{A}_{n,k}]\right]. \end{aligned}$$

This implies according to the case study above:

$$\begin{aligned} &\hat{\mathrm{E}}[P_n^{\mathrm{a}}(t_k^n, \overline{S}_{n,k}) - \mathrm{e}^{-r\Delta t_n} P_n^{\mathrm{a}}(t_{k+1}^n, \overline{S}_{n,k+1}) | \mathscr{A}_{n,k}] \\ &\quad = \hat{\mathrm{E}}[1_{\overline{S}_{n,k} \le \overline{B}_{n,k}}((K - \overline{S}_{n,k})^+ - \mathrm{e}^{-r\,\Delta t_n} P_n^{\mathrm{a}}(t_{k+1}^n, \overline{S}_{n,k+1})) | \mathscr{A}_{n,k}] \\ &\quad = \hat{\mathrm{E}}[1_{\overline{S}_{n,k} \le \overline{B}_{n,k}} K(1 - \mathrm{e}^{-r\,\Delta t_n}) | \mathscr{A}_{n,k}] + \mathcal{O}(\sqrt{\Delta t_n}^3) \\ &\quad = \hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}] K(1 - \mathrm{e}^{-r\,\Delta t_n}) + \mathcal{O}(\sqrt{\Delta t_n}^3). \end{aligned}$$

Since

$$\begin{aligned} &\hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}, u_n \overline{S}_{n,k} > \overline{B}_{n,k+1}] = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) \\ &\quad \Rightarrow \sum_{k=1}^{n} \hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}, u_n \overline{S}_{n,k} > \overline{B}_{n,k+1}] = \mathcal{O}\left(\frac{1}{\sqrt{\Delta t_n}}\right), \end{aligned}$$

the assertion follows. □

Lemma A.2. Stopping the discrete process $(\overline{S}_{n,k})_{k=0,\dots,n}$ *according to the rule* $(B_{t_k^n})_{k=0,\dots,n}$ *yields the premium*

$$\pi_n^B = \sum_{k=0}^{n-1} \mathrm{e}^{-rt_k^n} K(1 - \mathrm{e}^{-r\,\Delta t_n}) \hat{Q}[\overline{S}_{n,k} < B_{t_k^n}] + \mathcal{O}(\Delta t_n).$$

Proof. Follows exactly as that of Lemma A.1. □

Lemma A.3.

$$\mathcal{N}(b_{2,0}(S_0,t_k^n)) - \hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}] = \mathcal{O}(\Delta t_n) + \mathcal{N}(b_{2,0}(S_0,t_k^n)) - \mathcal{N}(z_{2,0}(S_0,t_k^n))$$

where

$$z_{2,0}(S_0,t_k^n) = \frac{\ln \overline{B}_{n,k}/S_0 - n\ln d_n}{(\ln u_n - \ln d_n)\sigma_k} - \frac{n\overline{q}_n}{\sigma_k} + \frac{1}{\sigma_k}$$

and

$$\sigma_k = \sqrt{k\overline{q}_n(1-\overline{q}_n)}.$$

Proof.

$$\hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}] = \hat{Q}\left[\prod_{j=1}^{k} \overline{R}_{n,j} \le \frac{\overline{B}_{n,k}}{S_0}\right] = \hat{Q}\left[\sum_{j=1}^{k} \ln \overline{R}_{n,j} \le \ln \frac{\overline{B}_{n,k}}{S_0}\right].$$

Moreover,

$$\overline{B}_{n,k} = S_0 u_n^{j_k} d_n^{n-j_k}$$

$$\Rightarrow \frac{\overline{B}_{n,k}}{S_0} = u_n^{j_k} d_n^{n-j_k}$$

$$\Rightarrow \ln(\overline{B}_{n,k}/S_0) = j_k \ln u_n + (n-j_k)\ln d_n$$

$$= j_k(\ln u_n - \ln d_n) + n\ln d_n$$

$$\Rightarrow j_k = \frac{\ln(\overline{B}_{n,k}/S_0) - n\ln d_n}{\ln u_n - \ln d_n}.$$

Obviously σ_k is the variance of

$$\sum_{j=1}^{k} \ln \overline{R}_{n,j} \quad \text{and} \quad z_{2,0}(S_0,t_k^n) = \frac{j_k - n\overline{q}_n}{\sigma_k} + \frac{1}{2\sigma_k}.$$

Let $z_{1,0}(S_0,t_k^n) = -n\overline{q}_n/\sigma_k - 1/2\sigma_k$. Then, according to Prohorov and Rozanov (1969), (pp. 183,184) we have

$$\mathcal{N}(b_{2,0}(S_0,t_k^n)) - \hat{Q}[\overline{S}_{n,k} \le \overline{B}_{n,k}]$$

$$= \underbrace{\mathcal{N}(z_{2,0}(S_0,t_k^n)) - \mathcal{N}(z_{1,0}(S_0,t_k^n)) - \hat{Q}\left[\overline{S}_{n,k} \le \overline{B}_{n,k}\right]}_{=\mathcal{O}(\Delta t_n)}$$

$$- \mathcal{N}(z_{2,0}(S_0,t_k^n)) + \mathcal{N}(b_{2,0}(S_0,t_k^n)) + \underbrace{\mathcal{N}(z_{1,0}(S_0,t_k^n))}_{=\mathcal{O}(\Delta t_n^2)}$$

□

Lemma A.4.

$$\exists \kappa \in \mathbb{R}: \quad \mathcal{N}(b_{2,0}(S_0, t_k^n)) - \hat{Q}[\bar{S}_{n,k} \leq B_{t^n_k}] \leq \kappa \Delta t_n.$$

Proof. Denote $z'_{2,0}(S_0, t_k^n) = b_{2,0}(S_0, t_k^n) - n\bar{q}_n/\sigma_k + 1/\sigma_k$. Similar to Lemma A.3 we have

$$\mathcal{N}(b_{2,0}(S_0, t_k^n)) - \hat{Q}[\bar{S}_{n,k} \leq B_{t^n_k}] = \mathcal{O}(\Delta t_n) + \mathcal{N}(b_{2,0}(S_0, t_k^n)) - \mathcal{N}(z'_{2,0}(S_0, t_k^n)).$$

The assertion follows immediately from the observation $b_{2,0}(S_0, t_k^n) \leq z'_{2,0}(S_0, t_k^n)$. □

Lemma A.5. There exists a constant $\kappa(S_0, K, r, \sigma, T) \in \mathbb{R}$ *such that*

$$\pi - \hat{\mathrm{E}}\left[\sum_{k=0}^{n-1} \mathrm{e}^{-rt^n_k} K(1 - \mathrm{e}^{-r\Delta t_n}) \hat{Q}[\bar{S}_{n,k} < B_{t^n_k} | \mathscr{A}_{n,k}]\right] \leq \kappa\, \Delta t_n.$$

Proof. From a series expansion of the exponential function we get $K(1 - \mathrm{e}^{-r\,\Delta t_n}) = Kr\,\Delta t_n + \mathcal{O}(\Delta t_n^2)$. From Lemma A.4 it immediately follows that the term we want to estimate is less than

$$\mathcal{O}(\Delta t_n) + rK \sum_{k=0}^{n-1} \mathrm{e}^{-rt^n_k} \mathcal{N}(b_{2,0}(S_0, t_{k+1}^n))\, \Delta t_n$$

where

$$b_{2,0}(x, t) = (\ln(B_t/x) - (r - \tfrac{\sigma^2}{2})t)/\sigma\sqrt{t}.$$

The summation term can be viewed as an approximation to the respective integral. From the trapezoidal formula of numerical integration we get immediately that it equals for a suitable $\xi \in [0, T]$

$$rK \int_0^T \mathrm{e}^{-rt} \mathcal{N}(b_{2,0}(S_0, t))\, \mathrm{d}t + \Delta t_n^2 \mathcal{N}(b_{2,0}(S_0, \xi)) + \mathcal{O}(\Delta t_n).$$

Since the normal-function is bounded, we have proven the lemma. □

Lemma A.6.

$$\exists \kappa \in \mathbb{R}\ \ \forall (n,k) \in \mathscr{C}: \quad E[P^{\mathrm{a}}(t_k^n, \bar{S}_{n,k}) - P_n^{\mathrm{a}}(t_k^n, \bar{S}_{n,k})] \geq \kappa\, \Delta t_n.$$

Proof. Let $\kappa_0 := \mathrm{n}(B_0)/S_0$. For $\bar{S} \in \mathscr{D}_{n,k}$ ($n \in \mathbb{N}, 0 \leq k \leq n$) define

$$\Delta^{\mathrm{e}}_{n,k}(\bar{S}) := P^{\mathrm{e}}(t_k^n, \bar{S}) - P_n^{\mathrm{e}}(t_k^n, \bar{S}),$$

$$\Delta^{\mathrm{p}}_{n,k}(\bar{S}) := \pi(t_k^n, \bar{S}) - \pi_n(t_k^n, \bar{S}).$$

According to Leisen and Reimer (1996)

$$\exists \kappa_1 \in \mathbb{R} \;\; \forall n \in \mathbb{N} \;\; \forall 0 \le k \le n \;\; \forall \bar{S} \in \mathcal{D}_{n,k}\colon \; \kappa_1(n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n) \ge |\Delta^{\mathrm{e}}_{n,k}(\bar{S})|.$$

Now take $(n,k) \in \mathscr{C}$. Since $B_{t^n_k} < \bar{B}_{n,k}$ we have

$$\Delta^{\mathrm{e}}_{n,k}(\bar{B}_{n,k}) + \Delta^{\mathrm{p}}_{n,k}(\bar{B}_{n,k}) \ge 0.$$

This implies

$$\kappa_1(n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n) \ge \Delta^{\mathrm{e}}_{n,k}(\bar{B}_{n,k})$$

$$\ge -\kappa_0 \Delta^{\mathrm{p}}_{n,k}(\bar{B}_{n,k})$$

$$\Rightarrow -\tfrac{\kappa_1}{\kappa_0}(n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n) \le \Delta^{\mathrm{p}}_{n,k}(\bar{B}_{n,k})$$

$$\Rightarrow \Delta^{\mathrm{e}}_{n,k}(\bar{B}_{n,k}) + \Delta^{\mathrm{e}}_{n,k}(\bar{B}_{n,k}) \ge (\kappa_1 - \kappa_1/\kappa_0)(n(\mathfrak{m}_n^2 + \mathfrak{m}_n^3 + \mathfrak{p}_n) + 1/n). \quad \square$$

Lemma A.7. For the CRR lattice approach, there exists $\kappa \in \mathbb{R}$ such that for $(n,k) \in \mathscr{C}$ and $i \in \mathbb{N}$ with $I_{n,i} \subset [0, t^n_k]$ we have

$$\sum_{i=0}^{k} \mathcal{N}(b_{2,0}(S_0, t^n_k)) - \mathcal{N}(z_{2,0}(S_0, t^n_k)) \ge \kappa.$$

For any lattice approach the same result holds with $\kappa/\sqrt{\Delta t_n}$ instead of κ.

Proof. We define

- $B^l_{n,i} := B_{l_{n,i}}$, $B^r_{n,i} := B_{r_{n,i}}$.
- $g_{n,i}\colon I_{n,i} \to [0, \infty]$ the function

$$g_{n,i}(t) := \ln B^l_{n,i} + \mu_{n,i}(t - l_{n,i})$$

where

$$\mu_{n,i} := \frac{\ln B^r_{n,i} - \ln B^l_{n,i}}{r_{n,i} - l_{n,i}}.$$

- $\nu_{n,i} := (r_{n,i} - l_{n,i})/\Delta t_n$ the number of discrete time points in $I_{n,i}$.
- $\bar{\bar{B}}_{n,k} := \max\{\bar{S} \in \mathcal{D}_{n,k} \mid \bar{S} \le B_{t^n_k}\}$ the highest node below $B_{t^n_k}$.
- $\rho_n(k)$ the number of up-steps necessary to reach $\bar{\bar{B}}_{n,k}$, i.e. $\bar{\bar{B}}_{n,k} = S_0 u_n^{\rho_n(k)} d_n^{k-\rho_n(k)}$.

We will present our argumentation only for the CRR model. The general case follows similarly.

Take $(n,k)\in\mathscr{C}$ and $i\in\mathbb{N}$ and $I_{n,i}\subset[0,t^n_k]$. Let us assume in the sequel that $B^r_{n,i}=S_0u_n^{i+1}d_n^{n-(i+1)}$ and $B^l_{n,i}=S_0u_n^id_n^{n-i}$. This yields an error of order $\mathcal{O}(\Delta t_n)$.

On $I_{n,i}$ $\overline{\overline{B}}$ is alternately equal to $S_0u_n^id_n^{n-i}$ and $S_0u_n^{i-1}d_n^{n-(i-1)}$. Therefore for $l_{n,i}+j\Delta t_n\in I_{n,i}$ ($\Leftrightarrow 0\le j\le v_{n,i}$) we have

$$\frac{\ln\ g_{n,i}(l_{n,i}+j\ \Delta t_n\ )}{\overline{\overline{B}}_{n,j}}=\begin{cases}(\ln u_n)\frac{j}{v_{n,i}} & j \text{ even},\\ (\ln u_n)\frac{j}{v_{n,i}}+\ln\frac{1}{d_n} & j \text{ odd}.\end{cases}$$

Since for CRR we have $\ln u_n=\ln d_n=\sigma\sqrt{\Delta t_n}$, this implies

$$\sum_{t^n{}_j\in I_{n,i}}\frac{\ln g_{n,i}(l_{n,i}+j\,\Delta t_n)/\overline{\overline{B}}_{n,j}}{\sigma\sqrt{\Delta t_n}}-1=\sum_{j=0}^{v_{n,i}}\frac{j}{v_{n,i}}+\frac{v_{n,i}}{2}-v_{n,i}$$

$$=\frac{v_{n,i}(v_{n,i}+1)}{2\ v_{n,i}}-\frac{v_{n,i}}{2}=\frac{1}{2}.$$

Since $u_n=\exp\{+\sigma\sqrt{\Delta t_n}\}$ we have $|I_{n,i}|=\mathcal{O}(\sqrt{\Delta t_n})$.

Thus

$$\sum_{i=0}^{k}\mathscr{N}(b_{2,0}(S_0,t^n_k))-\mathscr{N}(z_{2,0}(S_0,t^n_k))=\sum_{t^n{}_j\in I_{n,i}}\frac{\ln B_{t^n{}_j}/\overline{B}_{n,j}}{\sigma\sqrt{\Delta t_n}}-1$$

$$\ge\sum_{t^n{}_j\in I_{n,i}}\frac{\ln B_{t^n{}_j}/\overline{\overline{B}}_{n,j}}{\sigma\sqrt{\Delta t_n}}-1\ge\sum_{t^n{}_j\in I_{n,i}}\frac{\ln\ g_{n,i}(l_{n,i}+j\Delta t_n\ )/\overline{\overline{B}}_{n,j}}{\sigma\sqrt{\Delta t_n}}-1+2\kappa_1$$

$$=\frac{1}{2}+2\kappa_1$$

where the last inequality and $\kappa_1\in\mathbb{R}$ is induced by the observation that $|g_{n,i}-B_{t^n_i}|\le|B^r_{n,i}-B^e_{n,i}|=\mathcal{O}(\sqrt{\Delta t_n})$.

Since $z_{2,0}(S_0,t^n_k)-b_{2,0}(S_0,t^n_k)=\mathcal{O}(\sqrt{\Delta t_n})$ we have

$$\mathscr{N}(z_{2,0}(S_0,t^n_k))-\mathscr{N}(b_{2,0}(S_0,t^n_k))$$

$$=(\mathfrak{n}(b_{2,0}(S_0,r_{n,i}))+\mathcal{O}(\sqrt{\Delta t_n}))(z_{2,0}(S_0,t^n_k)-b_{2,0}(S_0,t^n_k))$$

$$=\mathfrak{n}(b_{2,0}(S_0,r_{n,i}))(z_{2,0}(S_0,t^n_k)-b_{2,0}(S_0,t^n_k))+\mathcal{O}(\sqrt{\Delta t_n}).$$

The assertion follows now from the fact that $\sqrt{\Delta t_n}\sum_{i=0}^n\sqrt{r_{n,i}}^{-1}$ is uniformly bounded. □

References

Aldous, D.J, 1981. Weak Convergence and the Theory of Processes. University of California, Berkeley.

Barone-Adesi, G., Whaley, R., 1987. Efficient analytic approximation of American option values. Journal of Finance 42, 301–320.

Bensoussan, A., 1984. On the theory of option pricing. Acta Applicandae Mathematicae 2, 139–158.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Brennan, M., Schwartz, E., 1977. The valuation of American put options. Journal of Finance 32, 449–462.

Broadie, M., Detemple, J., 1996. American option evaluation new bounds, approximations, and a comparison of existing methods. Review of Financial Studies 9, 1211–1250.

Carr, P., Jarrow, R., Myneni, R., 1992. Alternative characterization of American put options. Mathematical Finance 2, 87–106.

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

Derman, E., Kani, I., Ergener, D., Bardhan, I., 1995. Enhanced numerical methods for options with barriers. Financial Analysts Journal 51, 65–74.

Harrison, J.M., Kreps, D., 1979. Martingales and Arbitrage in multiperiod security markets. Journal of Economic Theory 20, 381–408.

Harrison, J.M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. Stochastic Processes and their Applications 11, 215–260.

Hull, J.C., White, A., 1988. The use of the control variate technique in option pricing. Journal of Financial and Quantitative Analysis 23, 237–251.

Jaillet, P., Lamberton, D., Lapeyre, B., 1990. Variational inequalities and the pricing of the American options. Acta Applicandae Mathematicae 21, 163–289.

Jarrow, R., Rudd, A., 1983. Option Pricing. Irwin, Homewood.

Kloeden, P.E., Platen, E., 1992. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.

Kushner, H.J., 1977. Probability Methods for Approximations in Stochastic Control and for Elliptic Equations. Academic Press, New York.

Lamberton D., 1995. Error Estimates for the Binomial Approximation of American Put Options. Working paper, Université de Marne-la-Vallée.

Lamberton, D., Pagès, G., 1990. Sur l'approximation des réduites, annales de l' institut Henri Poincaré – Probabilité Statistique 26, 331–355.

Leisen, D.P.J., Reimer, M., 1996. Binomial models for option valuation-examining and improving convergence. Applied Mathematical Finance 3, 319–346.

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

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.

Myneni, R., 1992. The pricing of the American option. Annals of Applied Probability 2, 1–23.

Peizer, D.B., Pratt, J.W, 1968. A normal approximation for binomial, F, Beta, and other common related tail probabilities I. Journal of the American Statistical Association 63, 1416–1456.

Pratt, J.W., 1968. A normal approximation for binomial, F, Beta, and other common, related tail probabilities II. Journal of the American Statistical Association 63, 1457–1483.

Prohorov, Y.V., Rozanov, Y.A., 1969. Probability Theory. Springer, Berlin.

Tian, Y., 1993. A modified lattice approach to option pricing. Journal of Futures Markets 13, 563–577.

VanMoerbeke, P.L.J., 1976. On optimal stopping and free boundary problems. Archive for Rational Mechanics and Analysis 60, 101–148.