A tree-based method to price American options in the Heston model

Michel Vellekoop

Financial Engineering Laboratory, University of Twente, PO Box 217,
7500 AE Enschede, The Netherlands; email: m.h.vellekoop@math.utwente.nl

Hans Nieuwenhuis

University of Groningen, Faculty of Economics, PO Box 800,
9700 AV Groningen, The Netherlands; email: j.w.nieuwenhuis@eco.rug.nl

We develop an algorithm to price American options on assets that follow the stochastic volatility model defined by Heston. We use an approach which is based on a modification of a combined tree for stock prices and volatilities, where the number of nodes grows quadratically in the number of time steps. We show in a number of numerical tests that we get accurate results in a fast manner, and that features which are essential for the practical use of stock option pricing algorithms, such as the incorporation of cash dividends and a term structure of interest rates, can easily be incorporated.

1 INTRODUCTION

One of the most popular models for equity option pricing under stochastic volatility is the one defined by Heston (1993):

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_t^1 \quad (1.1)$$

$$dV_t = \kappa(\theta - V_t) dt + \omega \sqrt{V_t} dW_t^2 \quad (1.2)$$

In this model for the stock price process S and squared volatility process V the processes W^1 and W^2 are standard Brownian motions that may have a non-zero correlation coefficient ρ , while μ , κ , θ and ω are known strictly positive parameters.

The popularity of this model can be explained to a large extent by the possibility of deriving option price formulas for European options in closed form using Fourier transforms (Carr and Madan (1999)). This closed form actually requires the numerical approximation of a certain integral, and some care has to be taken when dealing with the complex logarithm in this integral (see the recent papers by Kahl and Jackel (2005), Lord and Kahl (2006) and Albrecher *et al* (2006) on this issue). This

The authors would like to thank two anonymous referees who provided some suggestions that have significantly improved the paper.

method is a lot faster than methods in which the corresponding partial differential equation is solved numerically, so calibration of the Heston model to European options is a lot easier than calibration of other stochastic volatility models which do not admit closed form pricing functions.

For American options, or asset dynamics which involve the payment of cash dividends at fixed dividend dates instead of continuous dividend payments, such closed form pricing functions do not exist. Methods to determine the prices of such options in the Heston model therefore typically focus on numerical schemes to solve the Heston partial differential equation under early exercise constraints.

Clarke and Parrott (1999) formulate the American put pricing problem as a linear complementarity problem and use an implicit finite difference scheme combined with a multigrid procedure based on earlier work by Brandt and Cryer (1983) to find price approximations. Their method was further improved by Oosterlee (2003) who used Fourier analysis methods to optimize the smoothing procedure in the multigrid procedure. Forsyth, Vetzal and Zvan have used a penalty method to deal with the early exercise constraint, and showed that in the limit this is in fact equivalent to a linear complementarity formulation (Zvan *et al* (1998)). Ikonen and Toivanen (2004, 2007) used operator splitting methods to price American options in the Heston model. They find that the error induced by splitting does not reduce the accuracy when compared to Crank–Nicholson methods without splitting.

Monte Carlo simulation methods have also been used to price options in the Heston model. Broadie and Kaya (2006) formulated an unbiased simulation scheme which exploits the fact that the (conditional) distributions of the stochastic volatility process is known. However, their method may be rather time-consuming, as shown in a paper by Lord *et al* (2009). In that paper, a simpler method is proposed, in which a direct Euler method is modified in such a way that the variance is guaranteed to remain positive. Andersen (2008) has recently developed two new discretization schemes. These are based on the exact distributions studied by Broadie and Kaya, but make use of approximations to speed up computations. None of the papers above give speed and accuracy results for options with early exercise possibilities, but the Monte Carlo methods could in principle be adjusted to treat American options as well by the use of least squares regressions (Longstaff and Schwartz (2001)).

In this paper, we will develop an alternative method which is based on building a discrete time process that approximates the dynamics of the Heston model. Not only will we show that American options and options including cash dividends can be priced very fast on such trees, but we will also argue that the structure is very transparent, and therefore particularly easy to implement.

Approximating trees for the Heston model have been considered before, for example in the paper by Leisen (2000). In that paper a multinomial tree is constructed with four successor nodes per node at every timestep for the asset price process, and two successor nodes for the volatility process. The goal behind the construction of this tree is weak convergence to the correct joint distribution of asset price and volatility when the number of timesteps grows to infinity. However, Florescu and

Viens (2005) mention that since a transformation to eliminate the volatility is used, weak convergence cannot be proven along the lines suggested in the paper. They therefore propose a different algorithm (Florescu and Viens (2006)), which combines a tree with Monte Carlo simulations based on a finite number of particles between time steps, but this is computationally expensive.

Hilliard and Schwartz (1996) use a transformation of the S and V variables to end up with stochastic differential equations for the transformed processes which lead to recombining discretizations in a two-dimensional tree. The transformation itself is based on the one proposed earlier by Nelson and Ramaswamy (1990). However, their method has some numerical problems since negative probabilities may arise in the discretized model when certain constraints on the correlation coefficient are not satisfied. Moreover, since a lot of computational time is used for calculations in parts of the state space which have a low probability, the method is often very inefficient.

All these tree methods use a grid in which both the distance between (transformed) asset price grid points and (transformed) volatility grid points are of the order $\sqrt{\Delta t}$, where Δt denotes the discrete time step. To make sure that the stochastic process defined on the tree converges (weakly) to the correct continuous time process, second order moment matching conditions are used and this generates equations that need to be satisfied by the single time step transition probabilities under the risk-neutral measure.

An exception is the paper by Guan and Xiaoqiang (2000), which is based on a scheme introduced in earlier work by Finuance and Tomas (1996). In the Guan and Xiaoqiang paper, a tree is constructed in which the problem that nodes do not recombine is solved by an interpolation/extrapolation scheme. A fixed grid is defined, and whenever option prices are needed for values that are not on the grid, interpolation or extrapolation based on neighboring points is applied. Negative option prices may occur in the tree due to the interpolation, but these are then set to zero. The method is shown to work very well for short maturity options, where the extrapolation errors are not yet too large.

In this paper, we propose an extension to this method in which we use different interpolation schemes in order to exploit the regularity of the option payoff function. We define a flexible grid for the logarithm of the stock price $Z_t = \ln S_t$ and the squared volatility process V_t . In contrast to most earlier methods, we take a mesh size which is of order $o(\Delta t)$ at every timestep. The discrete time stochastic process that we define takes its values on this grid, and has 16 or 64 successor nodes in every point. This obviously means that we require quite some computation time per timestep. But the extra degrees of freedom that we create by our setup allow us to exploit the fact that the American option price function is once continuously differentiable in both S and V . This smoothness is used to improve the speed of convergence, which means that we require more computations per timestep but far less timesteps than are needed for alternative methods. We therefore retain the flexibility of the Guan and Xiaoqiang method when introducing for example a term

structure of interest rates and cash dividends, but at the same time we will avoid negative transition probabilities and get quicker convergence.

The paper is organized as follows. In the next section we define the simplest version of our algorithm and in Section 3 we prove that the method is guaranteed to converge to the correct value. In Section 4 we propose a modification which makes the convergence a lot quicker. Section 5 then discusses some numerical experiments that have been performed with the schemes and we end with conclusions and recommendations for further research.

2 THE ALGORITHM

The Heston model for squared volatility process V and log stock price process Z is:

$$\begin{aligned}dV_t &= \kappa(\theta - V_t) dt + \omega\sqrt{V_t} dW_t^1 \\dZ_t &= (r - \frac{1}{2}V_t) dt + \sqrt{V_t} dW_t^2\end{aligned}$$

for given $(V_0, Z_0) = (\sigma^2, \ln S_0)$, where $\{(W_t^1, W_t^2), t \in [0, T]\}$ are correlated standard Brownian motion processes with correlation coefficient ρ under the risk-neutral measure $\mathbb{Q}$, and κ, ω, θ and r are strictly positive constants. The Heston model with stochastic interest rates will not be considered here, see Boyarchenko and Levendorskii (2007) for results on that more complicated problem.

The value of the European option with maturity T and payoff, in terms of stock and volatility values, given by a function $\Phi: \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}$ equals (see for example Karatzas and Shreve (1998)):

$$E_t(V_t, Z_t) = p(t, T)\mathbb{E}^{\mathbb{Q}}[\Phi(e^{Z_T}, \sqrt{V_T}) | \mathcal{F}_t] \quad (2.1)$$

with $p(t, T) = e^{-r(T-t)}$ the discount rate, and $(\mathcal{F}_t)_{t \in [0, T]}$ the filtration generated by the process $\{(V_t, Z_t), t \in [0, T]\}$. The righthand side can be shown to be indeed a function of t, V_t and Z_t only, because of the Markov property of (V_t, Z_t) . The value of the American option is:

$$A_t(V_t, Z_t) = \text{ess. sup}_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}[p(t, \tau) \Phi(e^{Z_\tau}, \sqrt{V_\tau}) | \mathcal{F}_t] \quad (2.2)$$

where $\mathcal{T}_{t,T}$ is set of all $(\mathcal{F}_t)_{t \in [0, T]}$ -stopping times that take values in $[t, T]$.

We now define the following discrete time stochastic processes ($k = 0, \dots, n-1$):

$$\begin{aligned}V_{k+1}^n &= V_k^n + \kappa(\theta - (V_k^n)^+) \Delta t^n + Y_{k+1}^{n,1} \omega \sqrt{(V_k^n)^+ \Delta t^n} \\Z_{k+1}^n &= Z_k^n + (r - \frac{1}{2}V_k^n) \Delta t^n + Y_{k+1}^{n,2} \sqrt{(V_k^n)^+ \Delta t^n}\end{aligned} \quad (2.3)$$

with $(V_0^n, Z_0^n) = (V_0, Z_0)$ and where the variables $(Y_k^{n,1}, Y_k^{n,2})$ are independent and identically distributed in k , with:

$$\mathbb{Q}^n(Y_k^{n,1} = +1, Y_k^{n,2} = +1) = \frac{1}{4}(1 + \rho) \quad (2.4)$$

$$\mathbb{Q}^n(Y_k^{n,1} = -1, Y_k^{n,2} = +1) = \frac{1}{4}(1 - \rho) \quad (2.5)$$

$$\mathbb{Q}^n(Y_k^{n,1} = +1, Y_k^{n,2} = -1) = \frac{1}{4}(1 - \rho) \quad (2.6)$$

$$\mathbb{Q}^n(Y_k^{n,1} = -1, Y_k^{n,2} = -1) = \frac{1}{4}(1 + \rho) \quad (2.7)$$

under a new pricing measure $\mathbb{Q}^n$. Notice that we have taken the positive part of the variance process $(V_k^n)^+ = \max\{0, V_k^n\}$ to ensure positivity of this process. This choice corresponds to the full truncation scheme of Lord *et al* (2009).

This seems to be a very natural discrete time stochastic process to approximate the Heston model in continuous time, but it is of course non-recombining. The number of possible values for the process (V^n, Z^n) grows exponentially in the number of timesteps and it is therefore not very well suited to do actual computations. We therefore modify it as follows.

Let:

$$z_k^{n,\max} = \max\{z: \mathbb{Q}^n(Z_k^n = z) > 0\}$$

and define $z_k^{n,\min}$, $v_k^{n,\max}$ and $v_k^{n,\min}$ analogously. We take:

$$\Delta z_k^n = (z_k^{n,\max} - z_k^{n,\min})/m_z$$

$$\Delta v_k^n = (v_k^{n,\max} - v_k^{n,\min})/m_v$$

for certain $m_v, m_z \in \mathbb{N}^+$ which describe how fine the mesh is that we will take, and define the set of gridpoints in $\mathbb{R}^2$:

$$\hat{\mathcal{S}}_k^n = \{(v_k^{n,\min} + i\Delta v_k^n, z_k^{n,\min} + j\Delta z_k^n) \mid i = 0, \dots, m_v, j = 0, \dots, m_z\} \quad (2.8)$$

For any set $\mathcal{S} = \mathcal{S}_x \times \mathcal{S}_y \subset \mathbb{R}^2$ of gridpoints and functions $f: \mathcal{S} \rightarrow \mathbb{R}$ we denote by $\mathcal{L}^{\mathcal{S}}[f]: \mathcal{C}(\mathcal{S}) \rightarrow \mathbb{R}$ the piecewise bilinear interpolating function corresponding to this function on the grid, where $\mathcal{C}(\mathcal{S})$ denotes the convex hull of $\mathcal{S}$. This means that:

$$\begin{aligned} \mathcal{L}^{\mathcal{S}}[f](x, y) = & c_{00}^{\mathcal{S}}(x, y) f(x_0^{\mathcal{S}}(x), y_0^{\mathcal{S}}(y)) + c_{10}^{\mathcal{S}}(x, y) f(x_1^{\mathcal{S}}(x), y_0^{\mathcal{S}}(y)) \\ & + c_{01}^{\mathcal{S}}(x, y) f(x_0^{\mathcal{S}}(x), y_1^{\mathcal{S}}(y)) + c_{11}^{\mathcal{S}}(x, y) f(x_1^{\mathcal{S}}(x), y_1^{\mathcal{S}}(y)) \end{aligned}$$

where the functions $x_0^{\mathcal{S}}: \mathcal{C}(\mathcal{S}) \rightarrow \mathbb{R}$ and $y_0^{\mathcal{S}}: \mathcal{C}(\mathcal{S}) \rightarrow \mathbb{R}$ are defined in such a way that:

$$\begin{aligned} (x_0^{\mathcal{S}}(x), y_0^{\mathcal{S}}(y)) & \in \mathcal{S}, & (x_1^{\mathcal{S}}(x), y_1^{\mathcal{S}}(y)) & \in \mathcal{S}, \\ x_0^{\mathcal{S}}(x) & \leq x \leq x_1^{\mathcal{S}}(x), & y_0^{\mathcal{S}}(y) & \leq y \leq y_1^{\mathcal{S}}(y) \end{aligned}$$

ie, x and y are always in between the adjacent grid points $x_0^S(x)$, $x_1^S(x)$ and $y_0^S(y)$, $y_1^S(y)$ respectively. The functions c_{ij}^S are defined as:

$$\begin{aligned} c_{00}^S(x, y) &= (1 - \tilde{x})(1 - \tilde{y}) \\ c_{10}^S(x, y) &= \tilde{x}(1 - \tilde{y}) \\ c_{01}^S(x, y) &= (1 - \tilde{x})\tilde{y} \\ c_{11}^S(x, y) &= \tilde{x}\tilde{y} \end{aligned}$$

with:

$$\begin{aligned} \tilde{x} &= \frac{x - x_0^S(x)}{x_1^S(x) - x_0^S(x)} \\ \tilde{y} &= \frac{y - y_0^S(y)}{y_1^S(y) - y_0^S(y)} \end{aligned}$$

We define for $u \in \{0, 1\}$:

$$\begin{aligned} h_u^{k,n}(x) &= x - x_u^{\hat{S}_k^n}(x) \\ g_u^{k,n}(y) &= y - y_u^{\hat{S}_k^n}(y) \end{aligned}$$

Obviously, $|h_u^{k,n}(x)| < \Delta v_k^n$ and $|g_u^{k,n}(x)| < \Delta z_k^n$, and we will need these expressions in the proof of Theorem 2 in the next section.

We take a new process $(\tilde{V}^n, \tilde{Z}^n)$ on the space $\hat{S}_k^n$, as follows. First we define:

$$\begin{aligned} v^n(y, v, z) &= v + \kappa(\theta - v^+) \Delta t^n + y\omega\sqrt{v^+ \Delta t^n} \\ z^n(y, v, z) &= z + (r - \frac{1}{2}v) \Delta t^n + y\sqrt{v^+ \Delta t^n} \end{aligned}$$

where $\Delta t^n = T/n$. Let:

$$(\tilde{V}_{k+1}^n, \tilde{Z}_{k+1}^n) = (x_{Y_{k+1}^{n,3}}^{\hat{S}_{k+1}^n}(v^n(Y_{k+1}^{n,1}, \tilde{V}_k^n, \tilde{Z}_k^n)), y_{Y_{k+1}^{n,4}}^{\hat{S}_{k+1}^n}(z^n(Y_{k+1}^{n,2}, \tilde{V}_k^n, \tilde{Z}_k^n))) \quad (2.9)$$

where under a new measure $\tilde{\mathbb{Q}}^n$, the processes $(Y^{n,1}, Y^{n,2})$ have the same distribution as under $\mathbb{Q}^n$:

$$\tilde{\mathbb{Q}}^n(Y_{k+1}^{n,1} = i, Y_{k+1}^{n,2} = j) = \frac{1}{4}(1 + ij\rho) \quad i, j \in \{-1, 1\} \quad (2.10)$$

and are independent for different values of k , while the $(Y^{n,3}, Y^{n,4})$ have the following conditional distribution:

$$\begin{aligned} \tilde{\mathbb{Q}}^n(Y_{k+1}^{n,3} = i, Y_{k+1}^{n,4} = j \mid (Y_{k+1}^{n,1}, Y_{k+1}^{n,2}, \tilde{V}_k^n, \tilde{Z}_k^n)) \\ = c_{ij}^{\hat{S}_{k+1}^n}(v^n(Y_{k+1}^{n,1}, \tilde{V}_k^n, \tilde{Z}_k^n), z^n(Y_{k+1}^{n,2}, \tilde{V}_k^n, \tilde{Z}_k^n)) \quad i, j \in \{0, 1\} \end{aligned} \quad (2.11)$$

FIGURE 1 Construction of the risk-neutral probability measure on the grid.

and are also independent for different values of k . Note that this means that in every time step there are $4 \times 4 = 16$ new possible values for $(\tilde{V}_{k+1}^n, \tilde{Z}_{k+1}^n)$ based on the current value of $(\tilde{V}_k^n, \tilde{Z}_k^n)$, since every one of the four next values is split over four new points. The joint probabilities follow from (2.10) and (2.11):

$$\begin{aligned} \tilde{\mathbb{Q}}^n(Y_{k+1}^{n,1} = i_1, Y_{k+1}^{n,2} = i_2, Y_{k+1}^{n,3} = i_3, Y_{k+1}^{n,4} = i_4, | (\tilde{V}_k^n, \tilde{Z}_k^n)) \\ = \frac{1}{4}(1 + i_1 i_2 \rho) c_{i_3 i_4}^{\hat{S}_{k+1}^n}(v^n(i_1, \tilde{V}_k^n, \tilde{Z}_k^n), z^n(i_2, \tilde{V}_k^n, \tilde{Z}_k^n)) \end{aligned} \quad (2.12)$$

for $i_1, i_2 \in \{-1, 1\}$ and $i_3, i_4 \in \{0, 1\}$.

Figure 1 illustrates how the construction is carried out. Starting from a point $(\tilde{V}_k^n, \tilde{Z}_k^n)$, the values of the stochastic variables $Y_{k+1}^{n,1} \in \{-1, 1\}$ and $Y_{k+1}^{n,2} \in \{-1, 1\}$ determine four successor points with coordinates:

$$v_{\text{new}} = v^n(Y_{k+1}^{n,1}, \tilde{V}_k^n, \tilde{Z}_k^n) \quad (2.13)$$

$$z_{\text{new}} = z^n(Y_{k+1}^{n,2}, \tilde{V}_k^n, \tilde{Z}_k^n) \quad (2.14)$$

that are indicated here with black dots. We recombine by assigning probabilities to the four nearest neighbors, which are indicated using black squares. These four probabilities are given by $c_{ij}^S(v_{\text{new}}, z_{\text{new}})$ where $i, j \in \{0, 1\}$.

In summary, the algorithm would thus consist of the following steps:

- 1) Choose the number of timesteps n and mesh sizes (m_z, m_v) . For given starting values $(V_0^n, Z_0^n) = (\sigma^2, \ln S_0)$ iterate Equation (2.9) over all timesteps $k = 0, \dots, n-1$ and determine for every k the maximal and minimal values of the V^n and Z^n process, ie, $z_k^{n,\max}, z_k^{n,\min}, v_k^{n,\max}$ and $v_k^{n,\min}$.
- 2) Use these values to define the grid distances Δz_k^n and Δv_k^n for a given grid size (m_z, m_v) ; this now defines the grid $\mathcal{S}_k^n$ as in (2.8) for every timestep $k = 0, \dots, n$.
- 3) Fill in the option values for the grid on the final step, ie, on $\mathcal{S}_n^n$.
- 4) Work backwards on the tree: from every node (v, z) on the grid at timestep k we can reach 16 possible successor nodes at time $k+1$, as defined in (2.9), which have probabilities given by (2.12). The option price in node (v, z) is the discounted expected value over these successor nodes, or (if the option can be exercised before maturity) the payoff upon exercise in that node whenever that gives a larger value.

Having thus defined a discrete-time homogeneous Markov process, we will now investigate the weak convergence of this stochastic process in the next section.

3 WEAK CONVERGENCE

From now on, we take $T = 1$ without loss of generality. We will use the following theorem by Stroock and Varadhan (1979, Section 11.2).

THEOREM 1 *Let $\mathbb{Q}^n$ be a probability measure and Π_n a transition function on $\mathbb{R}^d$ such that $\mathbb{Q}^n$ -a.s. $X_0^n = x_0$ and for all $k > 0$:*

$$X_t^n = (k+1 - tn)X_{k/n} + (tn - k)X_{(k+1)/n}, \quad \frac{k}{n} \leq t \leq \frac{k+1}{n} \quad (3.1)$$

Define for all $k > 0$ and all Borel sets A in $\mathbb{R}^d$:

$$\Pi_n(X_{k/n}, A) = \mathbb{Q}^n(X_{(k+1)/n} \in A \mid \mathcal{F}_{k/n}) \quad (3.2)$$

where $\mathcal{F}_{k/n} = \sigma(\{X_t, 0 \leq t \leq k/n\})$ and assume that for all $R > 0$, $\epsilon > 0$ and certain continuous functions $a: \mathbb{R}^d \rightarrow S_d$ and $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$:

$$\lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} \|a_n(x) - a(x)\| = 0, \quad a_n^{ij}(x) = n \int_{\|y-x\| \leq 1} (y_i - x_i)(y_j - x_j) \Pi_n(x, dy) \quad (3.3)$$

$$\lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} \|b_n(x) - b(x)\| = 0, \quad b_n^i(x) = n \int_{\|y-x\| \leq 1} (y_i - x_i) \Pi_n(x, dy) \quad (3.4)$$

$$\lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} \Delta_n^\epsilon(x) = 0, \quad \Delta_n^\epsilon(x) = n \Pi^n(x, \{y \in \mathbb{R}^d: \|y - x\| > \epsilon\}) \quad (3.5)$$

If the martingale problem for a and b has exactly one solution $\mathbb{Q}_x$ for every initial value $x_0 = x$ then $\mathbb{Q}^n$ converges weakly to $\mathbb{Q}_{x_0}$, uniformly on compact subsets of $\mathbb{R}^d$.

We take $\hat{V}_{k/n}^n = \tilde{V}_k^n$ and $\hat{Z}_{k/n}^n = \tilde{Z}_k^n$ and define:

$$\begin{aligned}\hat{V}_t^n &= (k+1 - tn)\hat{V}_{k/n}^n + (tn - k)\hat{V}_{(k+1)/n}^n, & \frac{k}{n} \leq t \leq \frac{k+1}{n} \\ \hat{Z}_t^n &= (k+1 - tn)\hat{Z}_{k/n}^n + (tn - k)\hat{Z}_{(k+1)/n}^n\end{aligned}$$

as in the previous theorem. This simply means that we use linear interpolation in time to transform the discrete time processes $\tilde{V}^n$ and $\tilde{Z}^n$ to continuous time processes $\hat{V}^n$ and $\hat{Z}^n$ and we can then study the convergence properties in the space of continuous processes on $[0, 1]$.

Our main result is the following.

THEOREM 2 *Assume that:*

$$\lim_{n \rightarrow \infty} \Delta t^n = 0, \quad \lim_{n \rightarrow \infty} \max_{k=1, \dots, n-1} \frac{\Delta v_k^n}{\Delta t^n} = 0, \quad \lim_{n \rightarrow \infty} \max_{k=1, \dots, n-1} \frac{\Delta z_k^n}{\Delta t^n} = 0 \quad (3.6)$$

Then the processes $(\hat{V}^n, \hat{Z}^n)$ converge weakly to the process (V, Z) , ie:

$$((\hat{V}^n, \hat{Z}^n) | \tilde{\mathbb{Q}}^n) \xrightarrow{\mathcal{W}} ((V, Z) | \mathbb{Q}) \quad (3.7)$$

where $\mathcal{W}$ stands for weak convergence in the space of continuous processes on $[0, 1]$.

PROOF Define $X_{k/n}^n = (\tilde{V}_k^n, \tilde{Z}_k^n)$ and:

$$\Delta X_k^n = X_{(k+1)/n}^n - X_{k/n}^n \quad (3.8)$$

Take $x = (v, z)$. Since the drift and diffusion functions of the Heston model:

$$b(x) = \begin{bmatrix} \kappa(\theta - v) \\ r - \frac{1}{2}v \end{bmatrix}, \quad a(x) = v \begin{bmatrix} \omega^2 & \rho\omega \\ \rho\omega & 1 \end{bmatrix} \quad (3.9)$$

define a unique solution for the associated martingale problem (see for example Kallsen (2004)), we only need to check the conditions stated in the previous theorem. This boils down to proving that for all $R > 0$ and all $\epsilon > 0$ we have:

$$\begin{aligned}0 &= \lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} \|b(x) - n\mathbb{E}^{\tilde{\mathbb{Q}}^n}[(\Delta X_k^n) \mathbf{1}_{\{\|\Delta X_k^n\| \leq 1\}} | X_{k/n}^n = x]\| \\ 0 &= \lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} \|a(x) - n\mathbb{E}^{\tilde{\mathbb{Q}}^n}[(\Delta X_k^n)(\Delta X_k^n)^T \mathbf{1}_{\{\|\Delta X_k^n\| \leq 1\}} | X_{k/n}^n = x]\| \\ 0 &= \lim_{n \rightarrow \infty} \sup_{\|x\| \leq R} n\tilde{\mathbb{Q}}^n(\|\Delta X_k^n\| > \epsilon | X_{k/n}^n = x)\end{aligned}$$

Let $X_{k/n}^n = x = (v, z)$. Then:

$$\begin{aligned}\Delta X_k^n &= \begin{bmatrix} \tilde{V}_{k+1}^n - \tilde{V}_k^n \\ \tilde{Z}_{k+1}^n - \tilde{Z}_k^n \end{bmatrix} = \begin{bmatrix} x_{Y_{k+1}^{n,3}}^{\hat{S}_{k+1}^n}(v^n(Y_{k+1}^{n,1}, v, z)) - v \\ y_{Y_{k+1}^{n,4}}^{\hat{S}_{k+1}^n}(z^n(Y_{k+1}^{n,2}, v, z)) - z \end{bmatrix} \\ &= \begin{bmatrix} v^n(v, z, Y_{k+1}^{n,1}) - v + h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ z^n(v, z, Y_{k+1}^{n,2}) - z + g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix}\end{aligned}$$

so:

$$\Delta X_k^n = \begin{bmatrix} \kappa(\theta - v^+) \Delta t^n + Y_{k+1}^{n,1} \omega \sqrt{v^+ \Delta t^n} \\ (r - \frac{1}{2}v) \Delta t^n + Y_{k+1}^{n,2} \sqrt{v^+ \Delta t^n} \end{bmatrix} + \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix} \quad (3.10)$$

with:

$$\tilde{\mathbb{Q}}^n(Y_{k+1}^{n,1} = i, Y_{k+1}^{n,2} = j) = \frac{1}{4}(1 + ijp) \quad i, j \in \{-1, 1\} \quad (3.11)$$

Since $\Delta t^n = 1/n$ we see from (3.10) that for $n > n_1$ with n_1 large enough we have $\|\Delta X_k^n\| \leq 1$ since $|Y_{k+1}^{n,u}| = 1$ and at the same time:

$$|h_u^{k+1,n}(v)| < \Delta v_{k+1}^n, \quad |g_u^{k+1,n}(z)| < \Delta z_{k+1}^n \quad (3.12)$$

which implies that $h_u^{k+1,n} = o(\Delta t^n)$ and $g_u^{k+1,n} = o(\Delta t^n)$ for $u \in \{0, 1\}$. But this means that for $n > n_1$ we find:

$$\begin{aligned}b(x) - n\mathbb{E}^{\tilde{\mathbb{Q}}^n}[(\Delta X_k^n) \mathbf{1}_{\{\|\Delta X_k^n\| \leq 1\}} \mid X_{k/n}^n = x] \\ = n\mathbb{E}^{\tilde{\mathbb{Q}}^n} \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix} + \begin{bmatrix} \kappa(v - v^+) \\ 0 \end{bmatrix} \end{aligned} \quad (3.13)$$

The norm of the first term clearly converges to zero. So does the second one, since the values of v are generated by the function $v^n(y, \cdot, \cdot)$ with $y \in \{-1, 1\}$, which has a minimum value:

$$\Delta t^n \left(\kappa\theta - \frac{\omega^2}{4(1 - \kappa \Delta t^n)} \right)$$

If this value is positive v itself can never become negative so $v - v^+ = 0$ and otherwise we have for $\Delta t^n < 1/(2\kappa)$ that $|v - v^+| \leq \frac{1}{2}\omega^2 \Delta t^n$ which converges to zero.

The second condition is satisfied as well, because:

$$a(x) - n\mathbb{E}^{\tilde{\mathbb{Q}}^n}[(\Delta X_k^n)(\Delta X_k^n)^T \mathbf{1}_{\{\|\Delta X_k^n\| \leq 1\}} | X_{k/n}^n = x] = \begin{bmatrix} \tilde{a}_{11}^n & \tilde{a}_{12}^n \\ \tilde{a}_{12}^n & \tilde{a}_{22}^n \end{bmatrix} \quad (3.14)$$

for $n > n_1$, with:

$$\begin{aligned} \tilde{a}_{11}^n &= v\omega^2 - \mathbb{E}^{\tilde{\mathbb{Q}}^n} \frac{1}{\Delta t^n} (\kappa(\theta - v^+) \Delta t^n + Y_{k+1}^{n,1} \omega \sqrt{v^+ \Delta t^n} + o(\Delta t^n))^2 \\ \tilde{a}_{12}^n &= v\rho\omega - \mathbb{E}^{\tilde{\mathbb{Q}}^n} \frac{1}{\Delta t^n} \left(\left(r - \frac{1}{2}v \right) \Delta t^n + Y_{k+1}^{n,2} \sqrt{v^+ \Delta t^n} + o(\Delta t^n) \right) \\ &\quad \cdot (\kappa(\theta - v^+) \Delta t^n + Y_{k+1}^{n,1} \omega \sqrt{v^+ \Delta t^n} + o(\Delta t^n)) \\ \tilde{a}_{22}^n &= v - \mathbb{E}^{\tilde{\mathbb{Q}}^n} \frac{1}{\Delta t^n} \left(\left(r - \frac{1}{2}v \right) \Delta t^n + Y_{k+1}^{n,2} \sqrt{v^+ \Delta t^n} + o(\Delta t^n) \right)^2 \end{aligned}$$

But expanding terms then gives:

$$\begin{aligned} \tilde{a}_{11}^n &= (v - v^+)\omega^2 + O(\sqrt{\Delta t^n}) \\ \tilde{a}_{12}^n &= (v - v^+)\rho\omega + O(\sqrt{\Delta t^n}) \\ \tilde{a}_{22}^n &= v - v^+ + O(\sqrt{\Delta t^n}) \end{aligned}$$

which shows convergence in norm, since we already established that $|v - v^+| \leq \frac{1}{2}\omega^2 \Delta t^n$. The third condition is trivial since for $\|x\| < R$:

$$\begin{aligned} &\tilde{\mathbb{Q}}^n(\|\Delta X_k^n\| > \epsilon | X_{k/n}^n = x) \\ &= \tilde{\mathbb{Q}}^n \left(\left\| \begin{bmatrix} \kappa(\theta - v^+) \Delta t^n + Y_{k+1}^{n,1} \omega \sqrt{v^+ \Delta t^n} \\ (r - \frac{1}{2}v) \Delta t^n + Y_{k+1}^{n,2} \sqrt{v^+ \Delta t^n} \end{bmatrix} + \begin{bmatrix} h_{Y_{k+1}^{n,3}}^{k+1,n}(v) \\ g_{Y_{k+1}^{n,4}}^{k+1,n}(z) \end{bmatrix} \right\| > \epsilon \right) \end{aligned}$$

and this is actually equal to zero if we take n large enough, since the vector of which the norm is taken can be made arbitrarily small since $|Y_{k+1}^{n,u}| = 1$ for $u \in \{1, 2\}$ while $h_u^{k+1,n} = o(\Delta t^n)$ and $g_u^{k+1,n} = o(\Delta t^n)$ for $u \in \{0, 1\}$. $\square$

This shows that for our tree method convergence to the correct value is guaranteed. However, the speed of convergence may be improved considerably by using a slightly more complicated setup, as we will explain in the next section.

4 SPEEDING UP THE RATE OF CONVERGENCE

The proof for weak convergence given in the preceding section relies on the fact that we use a bilinear interpolation for option price values which fall outside grid points, as can clearly be seen from the functions c_{ij}^S that we defined in Section 2. This interpolation guarantees that the interpolating function is continuous for all values of the stock price and volatility that are considered. It also guarantees that no negative

probabilities will occur and that the interpolated values are positive if function values at grid points are positive, so the check for negative option values mentioned in the Guan and Xiaoqiang paper is not necessary in our case.

Although the scheme guarantees that we use approximations that are continuous in both S and V , these approximations will obviously not be differentiable. The American option price function, however, is known to be once differentiable in both S and V . We could thus try to use more elaborate interpolation schemes which result in interpolating functions that have the same property in order to improve the rate of convergence.

A natural candidate for this interpolation would be the one based on bicubic splines instead of bilinear ones. For every option value we need to approximate, we need not four but 16 surrounding points. We will make sure that the interpolating polynomials we use fit the grid points exactly, but we also impose continuity of the first order derivatives in the S and V directions, and continuity of the cross-derivative with respect to S and V , ie, $\partial^2 f / \partial S \partial V$ at all grid points. This means that the interpolating polynomial $p: \mathbb{R}^2 \rightarrow \mathbb{R}$ used to interpolate a function f between the grid points (x, y) , $(x + \Delta x, y)$, $(x, y + \Delta y)$ and $(x + \Delta x, y + \Delta y)$ should satisfy for all $a \in \{x, x + \Delta x\}$, $b \in \{y, y + \Delta y\}$:

$$\begin{aligned} p(a, b) &= f(a, b) \\ \partial_1 p(a, b) &= \frac{f(a + \Delta x, b) - f(a - \Delta x, b)}{2\Delta x} \\ \partial_2 p(a, b) &= \frac{f(a, b + \Delta y) - f(a, b - \Delta y)}{2\Delta y} \\ \partial_1 \partial_2 p(a, b) &= \frac{f(a + \Delta x, b + \Delta y) + f(a - \Delta x, b - \Delta y) - f(a - \Delta x, b) - f(a, b - \Delta y)}{4\Delta x \Delta y} \end{aligned}$$

In terms of our earlier notation, this gives after some trivial calculations that $p(x, y) = \mathcal{L}^S[f](x, y)$ with:

$$\mathcal{L}^S[f](x, y) = \sum_{i=-1}^2 \sum_{j=-1}^2 c_{ij}^S(x, y) f(x_i^S(x), y_j^S(y)) \quad (4.1)$$

where the functions x_0^S and y_0^S are defined as before, $x_{-1}^S(x) = x_0^S(x) - \Delta x$, $x_2^S(x) = x_1^S(x) + \Delta x$, and similar definitions hold for y_{-1}^S and y_2^S . The interpolating functions are given by the basis functions:

$$\begin{aligned} c_{00}^S(x, y) &= \frac{1}{4}(\tilde{x} - 1)(\tilde{y} - 1)(3\tilde{x}^2 - 2\tilde{x} - 2)(3\tilde{y}^2 - 2\tilde{y} - 2) \\ c_{-1,-1}^S(x, y) &= \frac{1}{4}\tilde{x}(\tilde{x} - 1)^2\tilde{y}(\tilde{y} - 1)^2 \\ c_{0,-1}^S(x, y) &= -\frac{1}{4}(\tilde{x} - 1)(3\tilde{x}^2 - 2\tilde{x} - 2)\tilde{y}(\tilde{y} - 1)^2 \end{aligned}$$

FIGURE 2 Spline interpolation functions.

while the other ones follow from obvious symmetries:

$$\begin{aligned} c_{i,j}(\tilde{x}, \tilde{y}) &= c_{j,i}(\tilde{y}, \tilde{x}) \\ c_{i,j}(\tilde{x}, \tilde{y}) &= c_{1-i,j}(1 - \tilde{x}, \tilde{y}) \\ c_{i,j}(\tilde{x}, \tilde{y}) &= c_{i,1-j}(\tilde{x}, 1 - \tilde{y}) \end{aligned}$$

By using this interpolation scheme we can now define a tree in which every node has 64 successor nodes (although they may not all be distinct, of course) in the same way as outlined before. This raises the obvious question of whether we can again show weak convergence to the processes (V, Z) under the risk-neutral measure.

Figure 2 shows the shape of the basic spline interpolation functions and we can immediately see that we cannot give a similar weak convergence proof as for the earlier case. In our proof we made use of the fact that the earlier functions c_{ij} all map into $[0, 1]$ while the sum over all i and j of $c_{ij}(x, y)$ equals 1 for all x and y , ie, the functions $\{c_{ij}, i, j \in [0, 1]\}$ form a partition of unity. This allowed us to use the interpolation scheme to redistribute probability mass over neighboring points, to create a stochastic process and then use known weak convergence results for such processes. However, we see that when we use the new interpolation based on 64 successor nodes, some of the weights may be negative (see the plot of $c_{0,-1}$). This means that we can no longer guarantee that our procedure results in a stochastic process under a risk-neutral measure which is positive everywhere. The processes S and V remain positive by our construction, but we would end up with a signed risk-neutral measure $\tilde{\mathbb{Q}}^n$, and Theorem 1 can then no longer be applied.

Numerical error analysis helps to illustrate the point. If some weights c_{ij} are negative, numerical errors that we introduce in our scheme may increase over time. Indeed, we would like to have that $\mathbb{E}^{\tilde{\mathbb{Q}}^n} \mathcal{L}^S[\Delta f] \leq \max_S \Delta f$ for an error function Δf on the grid. This is satisfied for our earlier scheme with 16 successor nodes, but not for the new one. It would therefore be better to use a local scheme which gives positive values for all positive functions and is C^1 at the same time. Such schemes

exist (Schmidt (1995)) but they require a lot of computation time and are therefore not very efficient. That's why we prefer to use the interpolation method described above. Notice that if the errors Δf do not change signs in neighboring points all the time, the errors can be expected to decrease on average. We will see in the numerical results of the next section that this is indeed what is observed in practice. In fact, the error decreases so fast on average that we will often obtain quicker convergence than for the case with only 16 successor nodes.

5 NUMERICAL RESULTS

5.1 Standard benchmark

The first test case we considered is the same as in earlier papers that deal with American put options in the Heston model (Brandt and Cryer (1983); Clarke and Parrott (1999); Ikonen and Toivanen (2007); Oosterlee (2003); Zvan *et al* (1998)). The relevant parameter values taken in all those papers are:

$$\kappa = 5, \quad \theta = 0.16, \quad \omega = 0.9, \quad \rho = 0.1, \quad r = 0.10 \quad (5.1)$$

The strike is $K = 10$ and different values of the starting volatility V_0 and initial stock price S_0 were considered. The time to maturity is three months so $T = 0.25$.

Tables 1 and 2 on page 15 give the option prices, and relative errors for the method involving 16 successors for every state, while Tables 3 and 4 (see page 16) provide the prices generated by the method of the previous section with 64 successors per state. Results are provided for an initial volatility value of $\sqrt{V_0} = 25\%$. We give the results for different initial stock prices $S_0 \in \{8, 9, 10, 11, 12\}$, ie, the values that were used in the earlier papers, and for four different number of timesteps and grid sizes. Notice that we multiply the grid point distances by $\frac{1}{2}$ for the S and V variables, but the timesteps by $\sqrt{1/2}$ to satisfy the conditions of Theorem 2.

To get error estimates we used exact Heston option prices for the European case. For American options, the results by the other authors are indicated by ZFV (Zven, Forsyth, Vetzal), IT-PSOR (Ikonen and Toivanen PSOR method), OO (Oosterlee) and CP (Clarke and Parrott). Error values for the American options are based on the results by Zven, Forsyth, and Vetzal, the values reported by the other authors can be seen to be close.

We see from these tables that the method with 64 successor nodes usually performs considerably better than the method with 16 successor nodes. The second method gives option prices with errors smaller than one cent (or maximally 1%) even for the smallest number of timesteps. The computational time for the second case is of the same order of magnitude as for the first method. The CPU times for the two methods are approximately 0.19 seconds for the method with 16 successors and 0.35 seconds for the method with 64 successors for the smallest grid size, while on the largest grids they are 17 and 27 seconds respectively.¹

¹The algorithms were implemented in Matlab on a Pentium 4, 2 GHz machine.

TABLE 1 European option, 16 Successors, $\sqrt{V_0} = 0.25$.

Steps s	Steps v	Steps t	S_0				
			8	9	10	11	12
125	6	25	1.8427	1.0572	0.5153	0.2267	0.0975
250	12	35	1.8417	1.0543	0.5103	0.2206	0.0920
500	24	50	1.8408	1.0527	0.5077	0.2169	0.0884
1,000	48	71	1.8402	1.0515	0.5060	0.2142	0.0857
		(Exact)	1.8389	1.0483	0.5015	0.2082	0.0804
		(Error)	0.0038	0.0089	0.0138	0.0186	0.0171
			0.0028	0.0060	0.0088	0.0124	0.0115
			0.0019	0.0044	0.0062	0.0087	0.0080
			0.0013	0.0032	0.0045	0.0060	0.0053

TABLE 2 American option, 16 successors, $\sqrt{V_0} = 0.25$.

Steps s	Steps v	Steps t	S_0				
			8	9	10	11	12
125	6	25	1.9933	1.1189	0.5366	0.2333	0.0995
250	12	35	1.9948	1.1149	0.5307	0.2268	0.0938
500	24	50	1.9960	1.1128	0.5276	0.2230	0.0902
1,000	48	71	1.9970	1.1112	0.5254	0.2201	0.0875
			2.0000	1.1076	0.5202	0.2138	0.0821 ZFV
			2.0000	1.1074	0.5190	0.2130	0.0818 IT-PSOR
			2.0000	1.1070	0.5170	0.2120	0.0815 OO
			2.0000	1.1080	0.5316	0.2261	0.0907 CP
		(Error)	−0.0067	0.0113	0.0164	0.0195	0.0174 (ZFV)
			−0.0052	0.0073	0.0105	0.0130	0.0117
			−0.0040	0.0052	0.0074	0.0092	0.0081
			−0.0030	0.0036	0.0052	0.0063	0.0054

5.2 Optimal exercise surface

In a second experiment, we have approximated the optimal exercise surface for the Heston model. To make the option pricing problem a bit more realistic, we introduced a deterministic term structure of interest² and cash dividend payments for the underlying stock. If we assume that dividends will be paid at times t_i ($i = 1, \dots, m$) with $t < t_1 < t_2 < \dots < t_m < T$ and that the dividend paid at time t_i equals D_i then the ex-dividend stock price dynamics under the risk-neutral measure

²We have not explicitly included time-varying rates in our setup since this involves cumbersome notation in the proofs but one can easily check that whenever the deterministic function that describes the short rate is Lipschitz continuous, all our results go through.

TABLE 3 European option, 64 successors, $\sqrt{V_0} = 0.25$.

Steps s	Steps v	Steps t	S_0				
			8	9	10	11	12
125	6	25	1.8378	1.0474	0.5001	0.2080	0.0802
250	12	35	1.8378	1.0479	0.5012	0.2079	0.0797
500	24	50	1.8380	1.0481	0.5015	0.2079	0.0798
1,000	48	71	1.8382	1.0481	0.5016	0.2080	0.0799
		(Exact)	1.8389	1.0483	0.5015	0.2082	0.0804
		(Error)	-0.0011	-0.0009	-0.0013	-0.0002	-0.0003
			-0.0011	-0.0004	-0.0002	-0.0003	-0.0007
			-0.0009	-0.0002	0.0000	-0.0003	-0.0006
			-0.0007	-0.0002	0.0001	-0.0002	-0.0005

TABLE 4 American option, 64 successors, $\sqrt{V_0} = 0.25$.

Steps s	Steps v	Steps t	S_0				
			8	9	10	11	12
125	6	25	1.9922	1.1073	0.5192	0.2133	0.0815
250	12	35	1.9942	1.1075	0.5199	0.2133	0.0812
500	24	50	1.9957	1.1076	0.5201	0.2133	0.0813
1,000	48	71	1.9968	1.1076	0.5202	0.2134	0.0815
			2.0000	1.1076	0.5202	0.2138	0.0821
			2.0000	1.1074	0.5190	0.2130	0.0818
			2.0000	1.1070	0.5170	0.2120	0.0815
			2.0000	1.1080	0.5316	0.2261	0.0907
		(Error)	-0.0078	-0.0003	-0.0010	-0.0005	-0.0006
			-0.0058	-0.0001	-0.0003	-0.0005	-0.0009
			-0.0043	0.0000	-0.0001	-0.0005	-0.0008
			-0.0032	0.0000	0.0000	-0.0004	-0.0006

change to:

$$dS_t = r_t S_t \, dt + \sqrt{V_t} S_t \, dW_t^1 - \sum_{i=1}^m \delta_i(S_{t_i-}) \, d\mathbf{1}_{\{t \geq t_i\}}$$

(5.2)

where the dividend payoff function δ_i is:

$$\delta_i(x) = \min\{D_i, x\}$$

to make sure that the ex-dividend stock price S_{t_i} does not become negative.

We took a maturity of one year so $T = 1$ and the dividend dates were $t_1 = 0.25$, $t_2 = 0.50$ and $t_3 = 0.75$ respectively. The dividend payment was always equal to $D_i = 0.20$, ($i = 1, 2, 3$). The adjustments made to cover the dividend dates

FIGURE 3 Optimal exercise boundary for the American put.

were based on standard interpolation techniques, see Vellekoop and Nieuwenhuis (2006) for a detailed description. We left the other parameters as in (5.1), and we took $\sqrt{V_0} = 50\%$ and $S_0 = K = 10$. The term structure was defined in terms of a deterministic short rate equal to:

$$r_t = 0.10 - 0.08 e^{-2t/T}$$

so the discount rates at time t for times $u > t$ were given by:

$$p(t, u) = e^{-\int_t^u r_s ds}$$

The result is shown in Figure 3. We can clearly see that the optimal exercise boundary for the American put is discontinuous at the dividend dates. Moreover, the figure shows that for higher volatility values we tend to exercise a bit later when close to maturity, since the lines describing the surface are not parallel around $t = T$.

5.3 Speed versus accuracy for fitted parameter values

In a last numerical study, we compared the speed versus accuracy for our tree method to a method based on partial differential equations. The Heston partial differential equation with a free boundary for early exercise was solved numerically using the operator splitting method of Ikonen and Toivanen (2007). The performance of the other methods in the previous section have comparable speed versus accuracy tradeoffs, see Ikonen and Toivanen (2007) for details. We used sparse matrix

implementations and experimented with the grid sizes to improve the speed of this method.

To make sure that the Heston parameter values were realistic, we took them from an earlier study by Bakshi *et al* (1997), in which the Heston model had been fitted to real option data on the S&P 500. The values reported in that study were:

$$\kappa = 1.15, \quad \theta = 0.0348, \quad \omega = 0.39, \quad \rho = -0.64, \quad r = 0.04, \quad \sqrt{V_0} = 0.1866$$

and these were used to price the options. We took $T = 0.25$ years to maturity and at-the-money options with $S = 100$. The strikes chosen were $K \in \{90, 100, 110\}$. Figure 4 (see page 19) shows the relative errors of the option prices against the computational time per option, both in logarithmic scaling. We did not include dividends here so the European call price could be determined exactly. For the American put prices we applied the different methods with increasing accuracy until they agreed on all relevant digits. We also implemented Monte Carlo methods based on Longstaff and Schwartz (2001), but these were never found to be competitive.

We see that the order of convergence of the partial differential equation method is better than the tree methods, but that the tree methods perform better when prices need to be generated fast. We also note that partial differential equation methods would become slower if a term structure of interest would be included, since the large matrices needed to implement the differential operator have to be reconstructed at every timestep, due to the changing of the short rate. We therefore believe that our approach would be competitive in practical applications. At the same time, the numerical analysis shown here suggests that if a lot of computational time is available per option, for example when option contracts are not traded online but over-the-counter, the partial differential equation-based method may be a better choice.

6 CONCLUSIONS

In this paper we have shown that tree-based methods can be used to find the price of American options in the Heston model. These methods are fast and simple to implement, and they can easily deal with problems which are not time-homogeneous, such as option pricing problems which involve a term structure of interest rates and an underlying asset that pays cash dividends.

The basic idea behind our method, to approximate a diffusion process on a grid which has a finer mesh for ΔS and ΔV for the stock and volatility than the usual square root of the time spacing Δt , may be applied to other methods as well. However, as we have seen, a proof of weak convergence based on Stroock and Varadhan's martingale problem can only be applied if we use approximation schemes which do not generate signed measures and these are not always the most efficient ones in terms of computation time. Finding other approximations with the desired properties and testing their efficiency when used to price American options in stochastic volatility models may therefore be an interesting direction for further research.

FIGURE 4 Time versus accuracy for European calls and American puts for different strikes.

Dashed = operator splitting, solid = 64 successors tree, dashed dots = 16 successors tree.

REFERENCES

- Albrecher, H., Mayer, P., Schoutens, W., and Tistaert, J. (2006). The little Heston trap. Technical Report, K. U. Leuven.
- Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *The Journal of Computational Finance* **11**(3), 1–42.
- Bakshi, Gu., Cao, C., and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance* **52**(5), 2003–2049.
- Boyarchenko, S., and Levendorskii, S. (2007). American options in the Heston model with stochastic interest rate. Working Paper.
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1031282.
- Brandt, C., and Cryer, C. W. (1983). Multigrid algorithms for the solution of linear complementarity problems arising from free boundary problems. *SIAM Journal of Scientific and Statistical Computing* **4**, 655–684.
- Broadie, M., and Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research* **54**, 217–231.
- Carr, P., and Madan, D. (1999). Option valuation using the fast Fourier transform. *The Journal of Computational Finance* **2**(4), 61–73.
- Clarke, N., and Parrott, K. (1999). Multigrid for American option pricing with stochastic volatility. *Applied Mathematical Finance* **6**(3), 177–195.
- Finuance, T. J., and Tomas, M. J. (1996). American stochastic volatility call option pricing: a lattice based approach. *Review of Derivatives Research* **1**, 183–201.
- Florescu, I., and Viens, F. (2005). A binomial tree approach to stochastic volatility driven model of the stock price. *Analele Universitatii din Craiova Seria Matematica-Informatica* **32**, 126–142.
- Florescu, I., and Viens, F. (2006). Stochastic volatility: option pricing using a multinomial recombining tree. Working Paper, Purdue University.
- Guan, L. K., and Xiaoqiang, G. (2000). Pricing American options with stochastic volatility: evidence from S&P futures options. *Journal of Futures Markets* **20**, 625–659.
- Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327–343.
- Hilliard, J. E., and Schwartz, A. (1996). Binomial option pricing under stochastic volatility and correlated state variables. *Journal of Derivatives* Fall, 23–39.
- Ikonen, S., and Toivanen, J. (2004). Operator splitting methods for pricing American options with stochastic volatility. Technical Report B11/2004, University of Jyväskylä.
- Ikonen, S., and Toivanen, J. (2007). Componentwise splitting methods for pricing American options under stochastic volatility. *International Journal of Theoretical and Applied Finance* **10**, 331–361.
- Ikonen, S., and Toivanen, J. (2007). Efficient numerical methods for pricing American options under stochastic volatility. *Numerical Methods for Partial Differential Equations* **24**, 104–126.

- Kahl, C., and Jackel, P. (2005). Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, 94–103.
- Kallsen, J. (2004). A didactic note on affine stochastic volatility models. Working Paper, TU München.
- Karatzas, I., and Shreve, S. (1998). *Methods of Mathematical Finance*. Springer-Verlag.
- Leisen, D. P. J. (2000). Stock evolution under stochastic volatility: a discrete approach. *Journal of Derivatives* **8**, 8–27.
- Longstaff, F. A., and Schwartz, E. S. (2001). Valuing American options by simulation: a simple least-squares method. *Review of Financial Studies* **14**, 113–147.
- Lord, R., and Kahl, C. (2006). Why the rotation count algorithm works. Working Paper, University of Wuppertal.
- Lord, R., Koekkoek, R., and Van Dijk, D. (2009). A comparison of biased simulation schemes for stochastic volatility models. *Journal of Quantitative Finance* iFirst, DOI: 10.1080/14697680802392496.
- Nelson, D., and Ramaswamy, K. (1990). Simple binomial processes as diffusion approximations in financial models. *Review of Financial Studies* **3**, 393–430.
- Oosterlee, C. W. (2003). On multigrid for linear complementarity problems with application to American-style options. *Electronic Transactions on Numerical Analysis* **15**, 165–185.
- Schmidt, J. W. (1995). Positive and S-convex C-1-interpolation of gridded 3-dimensional data. *Numerical Functional Analysis and Optimization* **16**, 233–246.
- Stroock, D. W., and Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer, Berlin.
- Vellekoop, M. H., and Nieuwenhuis, J. W. (2006). Efficient pricing of derivatives on assets with discrete dividends. *Applied Mathematical Finance* **13**, 265–284.
- Zvan, R., Forsyth, P., and Vetzal, K. (1998). A penalty method for American options with stochastic volatility. *Journal of Computational and Applied Mathematics* **92**, 199–218.