Working Paper 9201

BINOMIAL APPROXIMATION IN FINANCIAL MODELS: COMPUTATIONAL SIMPLICITY AND CONVERGENCE

by Anlong Li

Anlong Li is a Ph.D. candidate in operations research at the Weatherhead School of Management at Case Western Reserve University, Cleveland, Ohio, and a research associate at the Federal Reserve Bank of Cleveland. The author would like to thank Peter Ritchken and James Thomson for helpful comments and suggestions. He retains sole responsibility for any remaining errors.

January 1992

Abstract

This paper explores the potential of transformation and other schemes in constructing a sequence of simple binomial processes that weakly converges to the desired diffusion limit. Convergence results are established for valuing both European and American contingent claims when the underlying asset prices are approximated by simple binomial processes. We also demonstrate how to construct reflecting and absorbing binomial processes to approximate diffusions with boundaries. Numerical examples show that the proposed simple approximations not only converge, but also give more accurate results than existing methods, such as that of Nelson and Ramaswamy (1990), especially for longer maturities.

Introduction

Binomial models were first introduced by Sharpe (1978) and Cox, Ross, and Rubinstein (1979) to price options on assets with lognormal prices. This approach is attractive for valuing both American contingent claims and options with alternative asset price processes for which a closed-form option pricing formula, such as that of Black and Schole (1973), is not available. Cox and Rubinstein (1985) conceptually extend their model to approximate general diffusion processes. However, the resulting lattice is complicated by the fact that the number of states grows exponentially from one period to the next. A simple way to avoid such complexity is to transform the process into one that can be easily approximated by computationally simple binomial lattices whose nodes grow linearly in number from period to period. This idea has been used by Nelson and Ramaswamy (1990) in binomial models, and by Hull and White (1990) in explicit finite-difference methods. Amin (1991) suggests transforming the time scale to overcome the computational complexity caused by time-dependent volatilities.

In this paper, we investigate simple binomial approximations from several perspectives. First, we identify the class of diffusions that can be simply approximated using the popular binomial models of Cox and Rubinstein (1985, chapter 7), with no transformation. This results in a much larger set of diffusions that can be used as the transformed processes; thus, Nelson and Ramaswamy's (1990) method is generalized. We then explore the possibility of achieving computational simplicity by directly adjusting the Cox and Rubinstein (1985) binomial model. It turns out that the adjusted binomial lattice is a second truncation of the transformation method, further confirming our belief that transformation is, in principle, essential for achieving computational simplicity. However, when the transformation is analytically intractable, the adjusted binomial model can serve as an approximation.

We also propose a different approach to resolve the singularity problem associated with the boundary of a diffusion. Such diffusions are approximated here by reflecting or absorbing binomial processes. Although Nelson and Ramaswamy (1990) have developed a multiple-jump scheme for such cases, unfortunately, numerical examples show that their approximations become coarse as the maturity lengthens. Theoretically, both approaches guarantee convergence; however, the method developed here does not become coarse for longer maturities.

Actually, the time increment can be chosen to make the binomial chain purely reflecting or absorbing. Thus, the binomial process will reach an approximating boundary in a given number of steps. The process is either reflected or stays at the approximating boundary, depending on the nature of the boundary. This is particularly attractive when applied to the implicit finite-difference method, because it prevents the process from getting too close to the ultimate boundary, and the calculated transition probability will stay within the interval [0,1].

If asset prices can be approximated by binomial processes, then the corresponding options on such assets can be approximated using the same lattice. For European options with a continuous payoff function, the continuous mapping theorem of weak convergence guarantees that the option price sequence obtained from the binomial lattice will converge to its continuous-time counterpart, as long as the binomial processes weakly converge to the diffusion limit. For American options, one has to show that the sequence of optimal exercise strategies obtained from the binomial approximation converges to the optimal exercise strategy in the diffusion limit. This is an issue that has not been thoroughly studied. Assuming the optimal strategies are the same for both the approximating binomial processes and the diffusion limit, one can use the intuitive argument that, before the early exercise, the limit of the option price sequence satisfies the partial differential equation for the option price in continuous time. However, the optimal strategies are not known beforehand, and it remains to be shown whether the optimal strategies on the approximation lattice converge to the optimal strategy in continuous time.

The paper is organized as follows: Section 1 reviews the basics of diffusion approximation. Section 2 discusses how transformation methods can be used to achieve computationally simple binomial approximations. Section 3 focuses on approximating diffusions with boundaries. Section 4 deals with convergence in approximating both European and American contingent claims. Section 5 provides numerical examples, and section 6 concludes the paper. Proofs can be found in the appendices.

1. Weak Convergence and Diffusion Approximation

Let $(\Omega,\mathcal{F},P)$ be a probability space and $\mathbb{R}^d$ be the d-dimensional Euclidean vector space. Let $W(t)$ be an $\mathbb{R}^d$-valued Wiener process defined on $(\Omega,\mathcal{F},P)$. Fix a finite time interval $[0,T]$. Then $\mathcal{F}_t = \mathcal{F}(W(s),\ 0 \le s \le t) \subset \mathcal{F}$ for all $0 \le t \le T$. An $\mathbb{R}^d$-valued diffusion process $Y(t)$ can be defined by the following differential equation:

$$dY(t) = \mu(t,Y(t))dt + \sigma(t,Y(t))dW(t), \qquad Y(0) = Y_0, \tag{1.1}$$

where $\mu(t,Y(t))$ and $\sigma(t,Y(t))$ are the instantaneous mean and standard deviation of $Y(t)$, respectively. Let $C^d[0,T]$ be the space of $\mathbb{R}^d$-valued continuous functions on $[0,T]$. Then $W(t)$ and $Y(t)$ have sample paths in $C^d[0,T]$.

The processes used to approximate $Y(t)$ do not necessarily have continuous paths. Let $D^d[0,T]$ be the space of $\mathbb{R}^d$-valued functions on $[0,T]$, which are right continuous and have left-hand limits, and let $Y^{(n)}(t)$ be a sequence of processes with sample paths in $D^d[0,T]$. We use "$\Longrightarrow$" to denote weak convergence. Then $Y^{(n)} \Longrightarrow Y$ if $PY^{(n)} \Longrightarrow PY$, where $PY^{(n)}$ and PY are the measures induced by $Y^{(n)}$ and Y, respectively. On the other hand, if $Y^{(n)}(t) \Longrightarrow Y(t)$, every finite distribution of $Y^{(n)}(t)$ will converge to that of $Y(t)$. For a more detailed discussion of weak convergence, see Billingsley (1968).

In financial models, diffusions are usually approximated by binomial or multinomial processes. Such processes are characterized by the following definitions.

Definition 1. (Multinomial tree) Let $J_1,\ldots,J_m$ be functions from $\mathbb{R}^d$ to $\mathbb{R}^d$, and let $0 = t_0 < t_1 < \cdots < t_n = T$ and $Y_0 \in \mathbb{R}^d$. An m-ary tree is constructed as follows. At time t_0, the starting node (the "root") is labeled Y_0. At time $t_k < T$, each node Y at the beginning of period k branches into m nodes (the "sons"), labeled $J_1(Y,t_k),\ldots,J_m(Y,t_k)$. We call such an arrangement an n-period, m-nomial tree (lattice), and refer to $(J_1,\ldots,J_m)$ as the generator of this tree.

Definition 2. (Multinomial process) Let $\{Y_k,\ 0 \le k \le n\}$ be a discrete Markov chain in $\mathbb{R}^d$ with transition function $\nu(y_k,t_k,\Gamma) = P\{Y_{k+1} = \Gamma \mid Y_k = y_k\}$. If Γ takes values only from $J_1(Y_k,t_k),\ \ldots,\ J_m(Y_k,t_k)$, where J_1,

..., J_m are given functions from $\mathbb{R}^d$ to $\mathbb{R}^d$, then Y_k is called an m-nomial Markov chain on the m-nomial tree generated by $(J_1,\dots,J_m)$. The process $Y(t) = Y_{[nt]}$ is called an m-nomial Markov process, where $[nt]$ is the largest integer less than or equal to nt.

In graph theory terminology, levels are used to measure the distance from a node to the root in a (rooted) tree. Here we use periods instead. We also draw the tree on two-dimensional Cartesian coordinates, with time on the x-axis and state on the y-axis. In this case, we picture the tree growing from left to right.

In each period k (or time t_k), there are m^k nodes. The number of nodes grows exponentially from one period to the next, but this number can be dramatically reduced if we combine those that have the same labels (values). Graphically, we no longer have a tree once this combination is performed. Thus, we use the term "lattice." Such a lattice is considered computationally simple if, after combination, the number of nodes grows linearly from period to period.

In these definitions, if $J_i(Y_k, t_k)$ and $\nu(Y_k, t_k, \Gamma)$ are not dependent on the time index t_k, then the m-nomial tree, the Markov chain $\{Y_k\}$, and the Markov process $Y(t)$ are nonhomogeneous.

For diffusion approximation, we present in lemma 1 a modified version of corollary 7.4.2 of Ethier and Kurtz (1986, pp. 355-356). This is the theoretical basis for the discrete approximations used in recent finance literature (see Nelson [1990] and He [1991]).

Lemma 1. Suppose the stochastic differential equation (1.1) has an *a.e.* unique solution for any given Y_0. Let $Y_k^{(n)}$, $0 \le k \le n$, be an m-nomial Markov chain with lattice generator $(J_1,\dots,J_m)$ and transition function $\nu(x,t,\Gamma)$. Set

$$\mu_n(x,t) = \frac{1}{n} \sum_{i=1}^{m} [J_i(x,t)-x]\nu(x,t,J_i(x,t)) \text{ and } \tag{1.2}$$

$$\sigma_n^2(x,t) = \frac{1}{n}\sum_{i=1}^{m} [J_i(x,t)-x][J_i(x,t)-x]^T\nu(x,t,J_i(x,t)). \tag{1.3}$$

Suppose for every $r > 0$,

$$\sup_{\|x\|\le r} \|J_i(x,t) - x\| \longrightarrow 0 \qquad \forall\ i \tag{1.4}$$

$$\sup_{\|x\|\le r} \|\mu_n(x,t) - \mu(x,t)\| \longrightarrow 0 \tag{1.5}$$

$$\sup_{\|x\|\le r} \|\sigma_n^2(x,t) - \sigma^2(x,t)\| \longrightarrow 0. \tag{1.6}$$

Define $Y^{(n)}(t) = Y^{(n)}([nt])$. Then $Y^{(n)}(t)$ converges in distribution to the solution of (1.1).

In order to apply lemma 1, one first has to check whether the underlying diffusion equation (1.1) has an *a.e.* unique solution (or whether the corresponding martingale problem is well posed). Because most diffusions in financial models have this property, condition (1.4) is trivially satisfied in most cases. Conditions (1.5) and (1.6), which are often referred to as the consistency conditions, state that the first two calculated local moments converge to that of the diffusion.

The next lemma is a direct consequence of weak convergence and is useful in proving convergence in option approximation.

Lemma 2. Let g be a real-valued, bounded, and continuous function on $D^d[0,T]$. Then $Y^{(n)} \Longrightarrow Y$ implies $g(Y^{(n)}) \longrightarrow g(Y)$.

2. Binomial Approximation

2.1. Complexity of the Binomial Lattice

In the rest of this paper, we will consider only one-dimensional diffusions. Divide the time interval $[0,T]$ into n subintervals $[t_i, t_{i+1}]$ of equal length $h = T/n$, where $t_i = ih$, $i = 0,1,\ldots,n$. From definition 1, a binomial tree is generated by two functions, J^+ and J^-. Graphically, the building block at any node y looks like

$$y \;\text{———}\!\!\begin{cases} y_h^+ = J^+(y,t,\sqrt{h}) \\[2ex] y_h^- = J^-(y,t,\sqrt{h}). \end{cases} \tag{2.1}$$

For convenience, we call J^+ and J^- the up and down jumps, respectively. If the up-jump probability is $q(y,t)$, then the down-jump probability is $1 - q(y,t)$. After J^+ and J^- are constructed, $q(y,t)$ is calculated to satisfy consistency conditions (1.5) and (1.6).

The building block (2.1), together with the calculated transition probability $q(y,t)$, generates a binomial model. If the number of nodes grows linearly from period to period, then the model is computationally simple. In particular, the model is path-independent if, starting at any node, the binomial chain reaches the same state by following different paths, as long as these paths have the same number of up and down jumps. The model is considered stable if $q(y,t)$ falls between 0 and 1.

Assume J^+ and J^- are twice differentiable with respect to $\sqrt{h}$ for any given y and t. Then conditions (1.5) and (1.6) require that

$$J^{\pm}(y,t,\sqrt{h}) = y \pm \sigma(y,t)\sqrt{h} + O^{\pm}(h). \tag{2.2}$$

Omitting the term $O^{\pm}(h)$, we have the state-symmetric binomial model of Cox and Rubinstein (1985, chapter 7),

$$y_h^{\pm} = y \pm \sigma(y,t)\sqrt{h}, \tag{2.3}$$

with up-jump probability

$$q(y,t) = [1+ \mu(y,t)/\sigma(y,t)]\sqrt{h}/2. \tag{2.4}$$

An alternative is the probability-symmetric model, where both jumps have probability of 1/2:

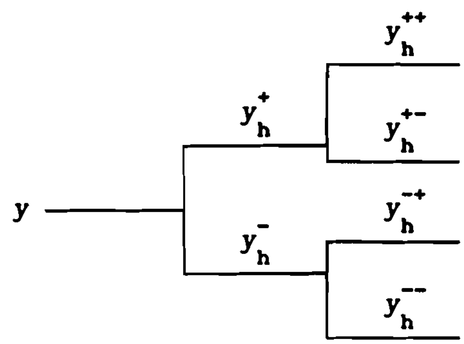
$$y_h^{\pm} = y \pm \sigma(y,t)\sqrt{h} + \mu(y,t)h. \tag{2.5}$$

Both models have certain advantages. The state-symmetric model does not incorporate the drift term $\mu(y,t)$ in the jumps. This coincides with the notion that option price does not depend on the expected stock return. However, state-symmetric models may be unstable, whereas probability-symmetric models are always stable.[1]

For Brownian motion and the geometric Wiener process, both models are computationally simple, with only $n + 1$ nodes in period n. In fact,

[1] Trigeorgis (1991) has developed a binomial model for Brownian motion that is both state-symmetric and stable.

the state-symmetric model is path-independent for homogeneous diffusions if and only if $\sigma(y)$ is linear in y.[2] However, this is not the case in general. For the state-symmetric model presented here, a three-period binomial lattice looks like



(2.6)

where

$$\begin{cases} y_h^{+-} = y^+ - \sigma(y^+, t+h)\sqrt{h} \\ y_h^{-+} = y^- + \sigma(y^-, t+h)\sqrt{h}. \end{cases} \tag{2.7}$$

Generally, y_h^{+-} and y_h^{-+} are not equal, so we have to use different nodes to represent them. The number of nodes in the lattice grows geometrically from period to period, exceeding one million in as few as 20 periods. As a result, we have a computationally complex lattice. The next two subsections discuss ways to resolve this problem.

---

[2] To see this, note that

$$\begin{aligned} y_h^{+-} - y_h^{-+} &= [2\sigma(y,t) - \sigma(y^+, t+h) - \sigma(y^-, t+h)]\sqrt{h} \\ &= -[\sigma''_{YY}(y,t)\sigma^2(y,t) + \sigma'_t(y,t)]\sqrt{h}^3 + O(h^2). \end{aligned}$$

For homogeneous diffusions, $\sigma(y,t) = \sigma(y)$. Thus, $y_h^{+-} - y_h^{-+} = 0$ implies $\sigma''_{YY}(y) = 0$, i.e., $\sigma(y)$ is linear. On the other hand, if $\sigma(y) = a + by$,

$$\begin{aligned} y^{+-} - y^{-+} &= y^+ - \sigma(y^+)\sqrt{h} - [y^- + \sigma(y^-)\sqrt{h}] \\ &= (y^+ - y^-) - [\sigma(y^+) + \sigma(y^-)]\sqrt{h} \\ &= (y^+ - y^-) - [2a + b(y^+ + y^-)]\sqrt{h} \\ &= 2(a+by)\sqrt{h} - 2(a+by)\sqrt{h} \\ &= 0. \end{aligned}$$

2.2. Adjusted Binomial Lattices

To reduce the complexity of the noncombining lattice for general diffusions, we make the following adjustment:

$$y_h^{\pm}(y,t) = y \pm \sigma(y,t)\sqrt{h} + \lambda(y,t)h, \qquad (2.8a)$$

where

$$\lambda(y,t) = \sigma(y,t)[\tfrac{1}{2}\sigma'_y(y,t) + \int \frac{\sigma'_t(y,t)}{\sigma^2(y,t)}\,dy + C] \qquad (2.8b)$$

and C is any constant. The binomial lattice (2.8a) is computationally simple not only for linear volatility functions, but also for the square root volatility function $\sigma(y,t) = \sigma\sqrt{y}$.

Equation (2.8b) is a necessary condition for path-independence. (See appendix B for details.) Generally, we need more adjustment terms in the lattice to close the gap between y_h^{+-} and y_h^{-+}. In the Constant Elasticity of Variance (CEV) model with $\gamma = 1 - k^{-1}$ (k is any positive integer), we need to include terms up to the order of $(\sqrt{h})^k$ in the adjustment to accomplish this. However, the difference $y_h^{+-} - y_h^{-+}$ in the adjusted lattice (2.8a) is usually as small as $o(h^2)$, and we can force the nodes to reconnect to obtain computational simplicity.[3] (See appendix C for further discussion.)

2.3. Binomial Lattices Generated by Transformations

In section 2.1, we showed that the Cox and Rubinstein binomial model (2.1) is computationally simple if and only if the volatility function of diffusion (1.1) is linear. For diffusions with general volatility functions, computational simplicity can be achieved through transformation. To do this, first identify a function f such that the transformed process $X(t) = f^{-1}(X(t),t)$ has a linear volatility. Then construct a sequence of simple binomial processes $X^{(n)}(t)$ that weakly converges to $X(t)$. If f is continuous, $Y^{(n)}(t) = f(X^{(n)}(t),t) \Longrightarrow Y(t)$.

---

[3] For example, we can take the average of Y_h^{+-} and Y_h^{-+} or simply pick either one of them.

We first consider the case in which $X(t)$ has unity volatility.[4] To identify the transformation f: $X(t) \mapsto Y(t)$, let $g = f^{-1}$. Applying Itô's formula, we have

$$dX(t) = [\mu(Y,t)\frac{\partial g}{\partial Y} + \frac{1}{2}\sigma^2(Y,t))\frac{\partial^2 g}{\partial Y^2} + \frac{\partial g}{\partial t}]dt + \sigma(Y,t)\frac{\partial g}{\partial Y}\, dW(t). \qquad (2.9)$$

Choose g such that $\sigma(Y,t)\frac{\partial g}{\partial Y} = 1$, which gives

$$x = g(y,t) = \int_0^Y \frac{dz}{\sigma(z,t)}. \qquad (2.10)$$

For convenience, we set the lower limit of the integral to zero. This gives $g(0,t) = 0$. As long as $\sigma(y,t) > 0$, g will be strictly increasing in y. Thus, the transformation f, which is the inverse of g, exists and is strictly increasing in x. Since $X(t)$ can be approximated by the simple model $x_h^{\pm} = x \pm \sqrt{h}$, the corresponding binomial model for $Y(t)$ is

$$y_h^{\pm} = f(x \pm \sqrt{h},\ t + h), \qquad (2.11)$$

with up-jump probability

$$q_h(y,t) = \frac{\mu(y,t)h + y - y_h^-}{y_h^+ - y_h^-}. \qquad (2.12)$$

Using Taylor's expansion for equation (2.11), we have

$$y_h^{\pm} = y \pm \sigma(y,t)\sqrt{h} + [\frac{1}{2}\sigma'_Y(y,t)\sigma(y,t) + \int_0^Y \frac{\sigma'_t(y,t)}{\sigma^2(y,t)}dy]h + o(h). \qquad (2.13)$$

Clearly, the adjusted lattice (2.8) is a truncation of equation (2.13).

The number of possible states of the binomial lattice (2.11) is at most $2n + 1$ for n partitions of the time interval $[0,T]$. Let $\mathcal{Y}^{(n)}$ be the state space for the binomial Markov chain $Y_k^{(n)}$; then

$$\mathcal{Y}^{(n)} = \{f(f^{-1}(y_0) + k\sqrt{h}) \mid -n \le k \le n\}. \qquad (2.14)$$

---

[4] This is the case examined in Nelson and Ramaswamy (1990) and Hull and White (1990).

Assumption 1. The diffusion equation (1.1) has an *a.e.* unique solution $Y(t)$ on $[0,T]$ for any given y_0.

Assumption 2. $\mu(y,t)$ is continuous, $\sigma(y,t)$ is nonnegative and twice differentiable, and the integral in equation (2.10) exists.

Assumption 3. For every $r > 0$, there exists an $h^* > 0$ such that for all $0 < h < h^*$, $0 \le q_h(y,t) \le 1$ for all $y \in \mathcal{Y}^{(n)}$.

The purpose of assumption 1 is obvious. Assumption 2 validates the use of the transformation. Assumption 3 guarantees that equation (2.12) defines a valid transition probability. To apply lemma 1, we need to check conditions (1.4) - (1.6). Condition (1.4) always holds, since f is continuous. From equations (2.12) and (2.13), we have

$$[(y_h^+ - y)q_h + (y_h^- - y)(1-q_h)]/h - \mu(y,t) = 0$$

$$[(y_h^+-y)^2 q_h + (y_h^- -y)^2(1-q_h)]/h - \sigma^2(y,t) = (q - \tfrac{1}{2})\sigma_y'\sigma^2\sqrt{h} + o(\sqrt{h}).$$

This implies that conditions (1.5) and (1.6) are satisfied. Thus, in applying lemma 1, we establish the following result:

Theorem 1. Let assumptions 1 - 3 hold. Let $Y_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial Markov chain with lattice generator (2.11) and transition probability (2.12). Define $Y^{(n)}(t) = Y_{[nt]}^{(n)}$. Then $Y^{(n)}(t) \Longrightarrow Y(t)$.

In the above theorem, assumption 3 can be replaced by conditions that are easier to verify. Either of the following is sufficient:

(*i*) The transformed process $X(t)$ has a locally bounded drift.
(*ii*) There exists an $\varepsilon > 0$ such that $\sigma(y,t) \ge \varepsilon$ for all y and $0<t<T$.

Condition (*i*) is somewhat weaker than condition (*ii*). For example, the geometric Wiener process does not satisfy condition (*ii*). However, the transformed process $X(t)$, a Brownian motion, satisfies condition (*i*).

Generally, one can transform the underlying diffusion Y into a new one, X, whose volatility function is either 1 or $a + bX$. The resulting binomial process will be slightly different, however. If f is the transformation to a diffusion with unity volatility, the corresponding binomial model will be given by equations (2.11) and (2.12). If the transformed process has linear volatility $a + bX$, $(b \neq 0)$, then the resulting binomial model is

$$y_h^{\pm} = f(f^{-1}(y) \pm \frac{\ln(1+b\sqrt{h})}{b}), \tag{2.15}$$

with the same transition probability as in equation (2.12). Using Taylor's expansion for equation (2.15), we have

$$y_h^{\pm} = y \pm \sigma(y)\sqrt{h} + \frac{1}{2}[\sigma_Y'(y) + b]\sigma(y)h + o(h). \tag{2.16}$$

Generally, a path-independent binomial model for Y would be

$$y_h^{\pm} = f(f^{-1}(y) \pm \sqrt{h} + o^{\pm}(\sqrt{h})),$$

where $o^{\pm}(\sqrt{h})/\sqrt{h} \longrightarrow 0$ as $h \longrightarrow 0$. (We may even construct examples in which the term $o^{\pm}(\sqrt{h})$ depends on y.)

Even though the Cox-Rubinstein model (2.1) is computationally simple for diffusions with linear volatility, there are certain reasons why one may prefer to transform these diffusions into those with unity volatility. On the other hand, $x_h^{\pm} = x \pm \sqrt{h}$ may not be the best choice for the transformed process. For example, Trigeorgis (1991) uses $x_h^{\pm} = x \pm \sqrt{h}(\sqrt{1 + \mu h})$ to achieve stability, where $\mu = r - \sigma^2/2$, r is the risk-free rate, and σ is the volatility of stock returns. Stability can also be achieved through time changes.

3. Singular Diffusions

Many diffusions in financial models have a lower boundary of 0. For example, stock prices and nominal interest rates are always assumed to be nonnegative. This is often modeled by allowing $\sigma(0,t) = 0$. If the drift term $\mu(0,t)$ equals zero as well, state 0 will serve as an absorbing boundary in many cases.[5] If the drift term is positive at 0,

---

[5] The geometric Wiener process is an exception because it has a natural boundary at 0. If the process starts from a positive state, it will never reach this boundary.

it will pull the process back from zero and is thus considered a reflecting boundary. There are also cases in between these two.

When $\sigma(y,t)$ is very close to zero for a small state y, the up-jump probability $q_h(y)$ in equation (2.12) may be pushed out of its meaningful range [0,1], and assumption 3 will be violated.[6] To avoid this problem, we use absorbing or reflecting binomial processes in the approximation. Specifically, we impose an approximating boundary y^*_t for the binomial process $Y^{(n)}$. Let x^*_t be the corresponding approximating boundary for the transformed process $X^{(n)}$. Then $y^*_t = f(x^*_t, t)$. For technical reasons, we may allow $X^{(n)}$ to be slightly below x^*_t on the lattice; thus, $Y^{(n)}$ may move slightly below y^*_t but remain above zero.[7]

Generally, y^*_t depends on the number of partitions n, the time t, and the nature of the true boundary 0. However, in the limit, we require y^*_t to approach zero for large n. The following two subsections examine reflecting and absorbing boundaries separately.

3.1. Reflecting Boundary

Assumption 3a. For any $r > 0$, there exists an $N > 0$ such that for any $h = T/n$ with $n > N$, and for any $t \in (0,T)$, there exists an x^*_t such that

$$0 \le q_h(f(x,t),t) \le 1, \quad x^*_t \le x < r. \tag{3.1}$$

This assumption allows the calculated transition probability in equation (2.12) to exceed 1 for very small states. At any state smaller than x^*_t, the binomial chain cannot jump down any farther. As a result, the first state below x^*_t serves as the reflecting boundary for the approximating binomial chain $X^{(n)}_k$. The resulting binomial lattice is

$$x^+_h = x + \sqrt{h} \tag{3.2a}$$

---

[6] Nelson and Ramaswamy (1990) suggest that the up jumps at lower states be moved higher (multi-jump) in the lattice to keep the transition probability between 0 and 1. The magnitude of the multiple jump reflects how "strongly" the drift pulls a small state away from zero.

[7] Actually, the step size h can be controlled so that the binomial process $X^{(n)}$ reaches the boundary x^* in exactly an integer number of steps. For details, see the examples in appendix E.

$$x_h^- = \begin{cases} x - \sqrt{h} & \text{if } x > x_t^* \\ x + j\sqrt{h} & \text{if } x \le x_t^*, \end{cases} \tag{3.2b}$$

where $j \ge 1$ is the smallest odd integer such that

$$\mu(f(x,t),t)h \le f(x+j\sqrt{h},t+h) - f(x,t). \tag{3.3}$$

The adjustment for x_h^- in equation (3.2b) ensures that the calculated transition probability $q_h(y,t)$ is between 0 and 1, with odd integer j showing the strength of the reflection. In most cases, $j \le 3$. When $j \ge 1$, the state $y_t^* = f(x_t^*,t)$ serves as a reflecting barrier for the binomial process $Y^{(n)}$. When $j = 1$, the binomial process jumps to a higher node in the lattice with probability 1. It is unlikely that $j = 1$ on a binomial lattice. However, we can choose h to make this happen. (See appendix E for details.) When $j = -1$, once the process reaches the boundary y_t^*, it will stay there with positive probability.

Theorem 2. Let assumptions 1, 2, and 3a hold. Suppose the j value in assumption 3a is bounded for all n, and let $X_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial Markov chain with lattice generator (3.2) and transition probability (2.12). Then $Y^{(n)}(t) = f(X_{[nt]}^{(n)},[nt]) \Longrightarrow Y(t)$.

Proof. We need to show that conditions (1.4) - (1.6) hold for all possible y on the lattice. Condition (1.4) holds because f is continuous. Since $q_h(y)$ is defined so that $0 \le q_h(y,t) \le 1$, condition (1.5) also holds. From the proof of theorem 1, condition (1.6) holds for all y corresponding to $x > x_t^*$, so we need only verify this condition for $y \le y_t^*$. From equation (3.3) and using Taylor's expansion, we have

$$\begin{aligned}\mu(y,t)h &\ge f(x+j\sqrt{h},t+h) - y \\ &= \sigma(y,t)j\sqrt{h} + \frac{1}{2}\sigma_Y'(y,t)\sigma(y,t)j^2h + f_t'(x,t)h + o(h).\end{aligned}$$

Thus, $\sigma(y,t) = O(\sqrt{h})$ and

$$\left|\sigma_n^2(y,t) - \sigma^2(y,t)\right| \le \{[f(x+\sqrt{h},t+h)-y)]^2 + [f(x+j\sqrt{h},t+h)-y]^2\}/h + \sigma^2(y,t)$$

$$\le (2+j^2)\sigma^2(y,t) + o(1) \longrightarrow 0.$$

This implies that condition (1.6) holds for $y \le y_t^*$. Q.E.D.

3.2. Absorbing Boundary

The absorbing case is relatively simple. Both the drift and volatility terms vanish at state 0. Thus, we need to prescribe state 0 as an absorbing barrier for the approximating binomial process.

Assumption 3b. For any $r > 0$, there exists an $N > 0$ such that for any $h = T/n$ with $n > N$, and for any $t \in (0,T)$,

$$0 \le q_h(f(x,t),t) \le 1, \quad 0 \le x < r \tag{3.4}$$

$$\mu(0,t) = \sigma(0,t) = 0. \tag{3.5}$$

For the transformed process, the binomial lattice is defined as

$$x_h^{\pm} = \begin{cases} x \pm \sqrt{h} & \text{if } x > 0 \\ x & \text{if } x \le 0. \end{cases} \tag{3.6}$$

Accordingly, for the original process Y,

$$y_h^{\pm} = f(x_h^{\pm}, t). \tag{3.7}$$

The up-jump probability is

$$q_h(y,t) = \begin{cases} \dfrac{\mu(y,t)h + y - y_h^-}{y_h^+ - y_h^-} & \text{if } y^+ \ne y^- \\ 1 & \text{if } y^+ = y^- \end{cases} \tag{3.8}$$

Note that when $Y^{(n)}$ reaches the absorbing boundary, it stays there with probability 1.

Theorem 3. Let assumptions 1, 2, and 3b hold. Let $X_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial chain defined by equations (3.6) and (3.8). Then $Y^{(n)}(t) = f(X_{[nt]}^{(n)}, [nt]) \Rightarrow Y(t)$.

Proof. As for theorem 2, we need only check equation (1.6). In fact,

$$\begin{aligned} \sigma_n^2(0,t) - \sigma^2(0,t) &= [f(0,t+h) - f(0,t)]^2/h \\ &= [f_t'(0,t)h + o(h)]^2/h \longrightarrow 0. \end{aligned}$$

Q.E.D.

4. Contingent Claim Approximation

4.1. European Options

Suppose the stock price follows the diffusion process

$$dS(t) = \mu(t,S(t))dt + \sigma(t,S(t))dW(t), \qquad S(0) = S_0 \tag{4.1}$$

and the discount bond price $B(t)$ evolves according to

$$dB(t) = r(t,S(t))dt, \qquad B(0) = 1. \tag{4.2}$$

Further assume no dividends on the stocks. If the terminal payoff of a European contingent claim at maturity T is $g(S(T))$, then at any time $t \le T$, the discounted terminal payoff is

$$G_{t,T}(S) = \exp[-\int_t^T r(u,S(u))du]g(S(T)). \tag{4.3}$$

Following Harrison and Kreps (1979) and Harrison and Pliska (1981), there is an equivalent martingale measure Q on $(\Omega,\mathcal{B})$ under which the price of this contingent claim is the expectation of $G_{t,T}(t,S)$. That is,

$$C^E_{t,T}(S) = \mathrm{E}_Q[G_{t,T}(S)]. \tag{4.4}$$

Under measure Q, the stochastic evolution of the stock prices follows the so-called pseudo process, which differs from process (4.1) only in the drift term. Specifically, under Q, the stock price process solves

$$dS(t) = r(t,S(t))S(t)dt + \sigma(t,S(t))d\hat{W}(t), \qquad S(0) = S_0. \tag{4.5}$$

Let $\{S^{(n)}\}$ be a sequence of binomial processes that weakly converges to S under Q. Assume, as before, that the time period is evenly divided into n periods of equal length h. For any n, consider a European contingent claim on a stock whose prices follow process $S^{(n)}$. Let $g(S^{(n)}(T))$ be the payoff of such a claim at maturity T, and $r(t,S^{(n)}(t))$ be the instantaneous return on the associated discount bond. Then, this claim can be priced by arbitrage using standard backward recursion on the approximating binomial lattice.

Let $V^E_k(S^{(n)}_k)$ be the value of this claim at node $(kh, S^{(n)}_k)$ on the binomial lattice, where $S^{(n)}_k$ is the stock price at time kh. After one period, suppose the binomial chain jumps up to $(kh+h, S^{(n)+}_k)$ with probability p_k and then jumps down to $(kh+h, S^{(n)-}_k)$ with probability $1 - p_k$. To eliminate arbitrage, we have

$$V^E_k(S^{(n)}_k) = \exp[-r(kh, S^{(n)}_k)]E_k[V^E_{k+1}(S^{(n)}_{k+1})|S^{(n)}_k], \tag{4.6}$$

where

$$E_k[V^E_{k+1}(S^{(n)}_{k+1})|S^{(n)}_k] = p_k V^E_{k+1}(S^{(n)+}_k) + (1-p_k)V^E_{k+1}(S^{(n)-}_{k+1}).$$

The boundary condition is

$$V^E_k(S^{(n)}_n) = g(S^{(n)}_n).$$

Since $S^{(n)}(t) = S^{(n)}_{[t/h]}$ is Markovian and its sample paths are step functions, an induction argument yields

$$\begin{aligned} V^E_{t,T}(S^{(n)}) &= \exp[-\int_t^T r(u, S^{(n)}(u))du]g(S(T)) \\ &= V^E_{[t/h]}(S^{(n)}_{[t/h]}). \end{aligned}$$

Therefore, $V^E_{t,T}(S^{(n)}) = E_Q[G_{t,T}(S^{(n)})]$ is the value of the claim on $S^{(n)}$. If r and g are continuous functions of t and $S(t)$ on $[0,T]$, then $G_{t,T}(S)$ is continuous on $\mathbb{D}^d[t,T]$. Applying lemma 2, we have

Theorem 4. Suppose S is the *a.e.* unique solution of equation (4.1), r and g are continuous in t and $S(t)$ on $[0,T]$, and $\{S^{(n)}\}$ is a sequence of binomial processes that weakly converges to S under measure Q. Then

$$C^E_{t,T}(S^{(n)}) = E_Q[G_{t,T}(S^{(n)})] \rightarrow C^E_{t,T}(S). \tag{4.7}$$

4.2. Discount Bonds

A discount bond with maturity T can also be viewed as a European option. We consider it to be a contingent claim with a payoff of $1 for every state at time T. Assume that the instantaneous interest rate $r(t)$ follows

$$dr(t) = \mu(t,r(t))dt + \sigma(t,r(t))dW(t), \qquad r(0) = r_0. \tag{4.8}$$

Under the local expectation hypothesis,[8] at any time $t \le T$, the price of such a discount bond is

$$B_{t,T}(r) = \mathbf{E}_t \exp[-\int_t^T r(u)du]. \tag{4.9}$$

When the interest-rate process (4.8) is approximated by binomial processes, we can calculate the approximated price of the discount bond on the binomial lattice. Similar to the European options, we have the following convergent theorem for discount bond approximation:

Theorem 5. Suppose r is the *a.e.* unique solution of equation (4.8) and $\{r^{(n)}\}$ is a sequence of binomial processes that weakly converges to r under measure Q. Then

$$B_{t,T}(r^{(n)}) = \mathbf{E}_t \exp(-\int_t^T r^{(n)}(u)du) \longrightarrow B_{t,T}(r). \tag{4.10}$$

Similarly, $B_{t,T}(r^{(n)})$ can be calculated on the binomial lattice using backward recursion.

4.3. American Options

For American options, not all contracts will be held to maturity; early exercise may be optimal. An exercise strategy is best described by a stopping time, since the decision to exercise an option is based only on the information available up to that time. Let $\mathcal{T}_{0,T}$ be the class of $\{\mathcal{F}_t\}$-stopping times with values in $[0,T]$. Following the arbitrage argument of Karatzas (1988), there exists an optimal $\{\mathcal{F}\ \}$-stopping time ρ such that the time t price of an American option is

$$C^A_{t,T}(S,\rho) = \sup_{\tau\in\mathcal{T}_{0,T}} \{C^A_{t,T}(S,\tau)\}, \tag{4.11}$$

---

[8] See Ingersoll (1987) for a discussion of expectation hypotheses.

where

$$C^{A}_{t,T}(S,\tau) = E_Q[\exp(-\int_t^{\tau} r(u,S(u))du)g(S(\tau))] \qquad (4.12)$$

and $g(S(u))$ is the immediate payoff if the option is exercised at time u. Actually, we can restrict the optimization over a smaller class of stopping times than the class $\mathcal{T}_{0,T}$. For example, an American put option can be exercised immediately when the stock price falls below a critical boundary. Shiryayev (1978) has shown that the value $C^{A}_{t,T}(S,\rho)$ in equation (4.11) will not change if we consider only the class of stopping times $\mathcal{D}_{0,T}$ that take the form

$$\tau_D = inf\{t \le T: (t,S(t)) \in D\}, \qquad (4.13)$$

where D is a closed subset of $\mathbb{R}^{+}x[0,T]$. From equation (4.13), τ_D is the first time the process $S(t)$ reaches the stopping region D. Define the continuation region G as $\mathbb{R}^{+}x[0,T] - D$, and suppose the process $S(t)$ starts within G; that is, $(0,S(0)) \in G$. Then the option is exercised as soon as the stock price reaches the boundary $\partial G = \partial D$. Let ρ be the optimal stopping time in $\mathcal{D}_{0,T}$. If D^{*} is the corresponding optimal stopping region, then

$$\tau = inf\{t: \le T: (t,S(t)) \in D^{*}\}. \qquad (4.14)$$

In binomial approximation, the option can only be exercised at discrete times $t_k = kh$, $k = 0,1,\dots,n$. Let $\mathcal{T}^{(n)}_{0,T}$ be the subset of $\{\mathcal{F}_t\}$-stopping times with discrete values ih, $i \le n$. Then we have the following convergence theorem for American contingent claims:

Theorem 6. Suppose S is the *a.e.* unique solution of equation (4.5) and $S^{(n)}$ is a sequence of processes that weakly converges to S under measure Q. Suppose further that τ is continuous *a.e.* relative to the measure induced by the limit process $S(t)$, the boundary of the optimal stopping region. Define

$$C^{A}_{t,T}(S^{(n)},\rho^{(n)}) = \sup_{\tau\in\mathcal{T}^{(n)}_{0,T}} E_Q\{\exp[-\int_t^{\tau^{(n)}} r(u,S^{(n)}(u))du]g(S^{(n)}(\rho^{(n)}))\}. \qquad (4.15)$$

Then

$$C^{A}_{t,T}(S^{(n)},\rho^{(n)}) \longrightarrow C^{A}_{t,T}(S,\tau). \tag{4.16}$$

Proof. First, we need to discretize the optimal stopping time τ in order to compare it with $\rho^{(n)}$. Define

$$\tau^{(n)} = inf\{t = kh\colon\ k \le n,\ S^{(n)}(t) \le D^{*}\}. \tag{4.17}$$

Since $\rho^{(n)}$ is optimal for the price process $S^{(n)}$, we have

$$C^{A}_{t,T}(S^{(n)},\rho^{(n)}) \ge C^{A}_{t,T}(S^{(n)},\tau^{(n)}) \tag{4.18}$$

for all n. Using Skorokhod embedding (see Kushner [1990]), we can assume that $S^{(n)}$ and S are defined on the same probability space. Since τ is continuous *a.e.* relative to the measure induced by the limit process $S(t)$, by weak convergence, $\tau^{(n)} \longrightarrow \tau$ *a.e.* Further, since $C^{A}_{t,T}(S,\tau)$ is continuous in both S and τ, we have

$$C^{A}_{t,T}(S^{(n)},\tau^{(n)}) \longrightarrow C^{A}_{t,T}(S,\tau). \tag{4.19}$$

On the other hand, the sequence $\rho^{(n)}$ is tight because $0 < \rho^{(n)} \le T$. Let ρ be the limit of some convergent subsequence of $\{\rho^{(n)}\}$. Then

$$C^{A}_{t,T}(S^{(n)},\rho^{(n)}) \longrightarrow C^{A}_{t,T}(S,\rho). \tag{4.20}$$

Taking the limit in equation (4.18) yields

$$C^{A}_{t,T}(S,\rho) \ge C^{A}_{t,T}(S,\tau).$$

However, since τ is optimal under the price process S, we have

$$C^{A}_{t,T}(S,\rho) = C^{A}_{t,T}(S,\tau). \tag{4.21}$$

Notice that equation (4.21) does not depend on the subsequence. This completes the proof. Q.E.D.

For each n, the discrete optimal stopping problem (4.15) can be solved using dynamic programming on the binomial lattice for the approximating process $S^{(n)}$. Let $V^{A}_{k}(S^{(n)}_{k})$ be the value of the American claim on $S^{(n)}$ at node $(kh,\ S^{(n)}_{k})$ on the binomial lattice, where $S^{(n)}_{k}$ is the stock price at time kh. After one period, suppose the binomial chain jumps up to $(kh + h,\ S_k^{(n)+})$ with probability p_k and then jumps down to $(kh + h,\ S_k^{(n)-})$ with probability $1 - p_k$. The Bellman equation for the optimization problem in equation (4.15) is then

$$V_k^A(S_k^{(n)}) = \max\{g(S_k^{(n)}),\ \exp[-r(kh,S_k^{(n)})]\mathbf{E}_k[V_{k+1}^A(S_{k+1}^{(n)})|S_k^{(n)}]\}, \tag{4.22}$$

where

$$\mathbf{E}_k[V_{k+1}^A(S_{k+1}^{(n)})|S_k^{(n)}] = p_k V_{k+1}^A(S_k^{(n)+}) + (1-p_k)V_{k+1}^A(S_{k+1}^{(n)-}).$$

The boundary condition is

$$V_k^A(S_n^{(n)}) = g(S_n^{(n)}). \tag{4.23}$$

Since $S^{(n)}(t) = S_{[t/h]}^{(n)}$ is Markovian and its sample paths are step functions, an induction argument yields

$$\begin{aligned} V_{t,T}^A(S^{(n)}) &= \exp[-\int_t^T r(u,S^{(n)}(u))du]g(S(T)) \\ &= V_{[t/h]}^A(S_{[t/h]}^{(n)}). \end{aligned}$$

A crucial condition in theorem 6 is the continuity of the optimal stopping time τ on the optimal exercising boundary. A sufficient condition for τ to be continuous is that the paths of the diffusion $S(t)$ are tangent to the boundary ∂D with probability 0.[9] It is also sufficient if all the points on the boundary ∂D are regular for the diffusion $S(t)$. When $\sigma(S(t)) > 0$ on ∂D, a point on ∂D is regular if it can be reached by an open cone (Dynkin [1965]). For American put options on lognormal prices, Van Moerbeke (1976) shows that the optimal boundary is increasing and continuously differentiable in time. Thus, the open cone condition can be easily verified. Most diffusions and their optimal boundaries in financial models fall into this category. (See Kushner [1984 and 1990] for further discussion.)

---

[9] To see this, we need to show that $\forall\ \varepsilon,\delta>0$, $\exists\ N>0$ such that when $n>N$,

$$P(|\tau-\tau^{(n)}|>\delta) < \varepsilon.$$

Actually, for any given $\varepsilon,\delta>0$, if one of the processes S and $S^{(n)}$ hits the boundary K^* first, say at time t, then the other will hit the boundary within time interval $(t,t+\delta)$ with probability $1-\varepsilon$.

5. Numerical Examples

In this section, we apply the method developed in sections 3 and 4 in order to approximate discount bond and stock option prices. Since the diffusion processes in this section are homogeneous, the time argument t will be dropped wherever appropriate.

5.1. Bond Pricing

Suppose the instantaneous interest rate follows the mean reverting square root (MRSR) process

$$dy = \kappa(\mu-y)\ dt + \sigma\sqrt{y}\ dW, \tag{5.1}$$

where κ, μ, $\sigma > 0$. From Feller's boundary classification, state 0 is an inaccessible reflecting boundary when $2\kappa\mu/\sigma^2 \geq 1$. Otherwise, state 0 is accessible and either reflecting or absorbing.

From equation (2.10), the transformation for this process is

$$f(x) = \begin{cases} \sigma^2x^2/4 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{5.2}$$

Let $\phi = 4\kappa\mu/\sigma^2 - 1$. Then the transformed process $X(t)$ follows

$$dX = \frac{1}{2}(\phi/X - \kappa X)dt + dW. \tag{5.3}$$

Except for some small states, the binomial generator is

$$y^{\pm} = (\sqrt{y} \pm \sigma\sqrt{h}/2)^2 \tag{5.4}$$

and the up-jump probability is

$$q_h(y) = \frac{1}{2} + \frac{\kappa(\mu-y) - \sigma^2/4}{2\sigma\sqrt{y}}\sqrt{h} \tag{5.5}$$

$$= \frac{1}{2} + \frac{\phi - \kappa x^2}{4x}\sqrt{h}.$$

Case 1. $\phi > 0$. From equation (5.5), $q_h(y)$ is strictly decreasing for all $y > 0$. For any given large number $r > 0$, we can always choose a sufficiently small h such that $q_h(y) > 0$ for all $0 < y < r$. This is true because $q_h(y)$ is decreasing and $\lim_{h\to 0} q_h(r) = 1/2$. Actually, when

$\kappa\sigma T < 16$, $q_h(y) > 0$ for all possible y values on the binomial lattice.[10]

On the other hand, for any small h, a reflecting boundary x^* can be calculated by setting $q(y) = 1$, or $\kappa x^2 + (2/\sqrt{h})x - \phi = 0$. This gives

$$x^* = (-1/\sqrt{h} + \sqrt{1/h + \kappa\phi})/\kappa = \phi\sqrt{h}/2 + o(h). \tag{5.6}$$

Let $y^* = f(x^*)$; then $0 \le q_h(y) < 1$ as long as $y^* < y < r$. Therefore, assumption 3a is satisfied, and from theorem 2 we have the following result:

Corollary 1. For the MRSR process (5.1) with $\phi > 0$, let x^* be defined by equation (5.6) with $x^* < x_0$. Let $X_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial Markov chain with lattice generator (3.2) and transition probability (2.12). Then $Y^{(n)}(t)\ f(X_{[nt]}^{(n)}) \Longrightarrow Y(t)$.

Case 2. $-1 < \phi \le 0$ and $\mu(0)h = \kappa\mu h \le \sigma^2 h/4 = f(\sqrt{h})$. In this case, the reflecting barrier is $x^* = 0$. The binomial Markov chain with the lattice generator for the transformed process is defined by equation (3.2). By definition, $0 \le q_n(y) \le 1$ when $x < 0$. For $x > 0$,

$$q_h(y) = \begin{cases} \dfrac{\kappa(\mu-y)h + y}{(\sqrt{y} + \sigma\sqrt{h}/2)^2} & \text{if } 0 \le y < \sigma\sqrt{h}/2 \\[2ex] \dfrac{1}{2} + \dfrac{\kappa(\mu-y) - \sigma^2/4}{2\sigma\sqrt{y}}\sqrt{h} & \text{if } y \ge \sigma\sqrt{h}/2. \end{cases} \tag{5.7}$$

For $0 \le y < \sigma\sqrt{h}/2$, we have $0 \le q_n(y) < (\kappa\mu h + y)/(y + \sigma^2 h/4) \le 1$, since $\sigma\sqrt{h}/2 < \mu$ for small h. For $y \ge \sigma\sqrt{h}/2$, $q_n(y) = \frac{1}{2} + \frac{\phi - \kappa X^2}{4X}\sqrt{h}$ is concave

---

[10] The largest state in $\mathcal{Y}^{(n)}$ is $y_{max} = f(n\sqrt{h}) = f(\sqrt{T/h})$. Then

$$\lim_{h\to 0} hy_{max} = \lim_{h\to 0} h(\sqrt{y_0} + \sigma n\sqrt{h}/2)^2 = \sigma^2 T/4$$

$$\lim_{h\to 0} q_n(y_{max}) = \lim_{h\to 0} [\frac{1}{2} + \frac{\kappa(\mu-y_{max}) - \sigma^2/4}{2\sigma\sqrt{y_{max}}}\sqrt{h}] = \frac{1}{2} - \kappa\sigma T/8.$$

Thus, as long as $\kappa\sigma T < 16$, there exists an $h^* > 0$ such that $0 \le q\ (y) \le 1$ for all $0 < y < h^*$ and $y \in \mathcal{Y}^{(n)}$

and bounded from above by 1/2. Actually, for large y, the condition $\kappa\sigma T < 16$ will guarantee $q_n(y) \geq 0$. Therefore, assumption 3b is satisfied, and we have the following result:

Corollary 2. For the MRSR process (5.1) with $-1 < \phi \leq 0$, let $x^* = 0$. Let $X_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial Markov chain with lattice generator (3.2) and transition probability (5.7). Then $Y^{(n)}(t)$ $f(X_{[nt]}^{(n)})$ $\Longrightarrow Y(t)$.

We now turn to approximating the discount bond price. Suppose the local expectation hypothesis holds. Then the time t price of a discount bond that matures at time T is

$$B(t,Y) = \mathbf{E}_t[\exp(-\int_t^T Y(u)du)]. \tag{5.8}$$

Let $\{Y^{(n)}\}$ be the sequence of binomial processes in either corollary 1 or corollary 2. Then the approximated bond price is

$$B(t,Y^{(n)}) = \mathbf{E}_t[\exp(-\int_t^T Y^{(n)}(u)du)]. \tag{5.9}$$

Like the European option, the bond price $B(t,Y^{(n)})$ is calculated using backward recursion on the binomial lattice for $Y^{(n)}$. At node (t_k,y),

$$B(t_k,y) = [q_h(y)B(t_{k+1},y_h^+) + (1-q_h(y))B(t_{k+1},y_h^-)]\exp(-yh). \tag{5.10}$$

The boundary condition is $B(T,y) = 1$.

Table 1 shows the approximated prices of a discount bond when the instantaneous interest rate follows the MRSR process (5.1). The first four columns specify the same parameters as in Nelson and Ramaswamy (1990). The volatility σ and the initial interest rate y_0 are annualized, while the maturity T is measured in months. The next three columns display the bond prices obtained using several different numbers of partitions in the approximation. The last column contains the theoretical values calculated using the formula of Cox, Ingersoll, and Ross (1985).

This table clearly illustrates the convergence of the approximated discount bond prices to the corresponding theoretical values for a wide range of case parameters. Compared with Nelson and Ramaswamy's results using the same parameters, our approximations are much more accurate, especially for higher σ, κ, and T values.

Table 1. Discount Bond Prices

κ	σ	T	y_0	n: 5	n: 50	n: 100	CIR
0.01	0.10	1	0.05	99.5841	99.5841	99.5840	99.5841
0.01	0.10	1	0.11	99.0876	99.0876	99.0876	99.0876
0.01	0.10	6	0.05	97.5288	97.5284	97.5283	97.5284
0.01	0.10	6	0.11	94.6529	94.6541	94.6542	94.6541
0.01	0.10	12	0.05	95.1172	95.1166	95.1167	95.1166
0.01	0.10	12	0.11	89.6059	89.6123	89.6127	89.6129
0.01	0.10	60	0.05	78.2003	78.3296	78.3373	78.3412
0.01	0.10	60	0.11	58.7689	59.0976	59.1167	59.1262
0.10	0.10	1	0.05	99.5834	99.5832	99.5832	99.5831
0.10	0.10	1	0.11	99.0883	99.0886	99.0885	99.0887
0.10	0.10	6	0.05	97.5028	97.4967	97.4964	97.4961
0.10	0.10	6	0.11	94.6782	94.6848	94.6851	94.6855
0.10	0.10	12	0.05	95.0166	94.9948	94.9937	94.9930
0.10	0.10	12	0.11	89.7000	89.7258	89.7272	89.7278
0.10	0.10	60	0.05	76.2518	76.0230	76.0114	76.0057
0.10	0.10	60	0.11	59.9993	60.4174	60.4401	60.4514
0.10	0.50	1	0.05	99.5835	99.5833	99.5834	99.5832
0.10	0.50	1	0.11	99.0885	99.0887	99.0888	99.0887
0.10	0.50	6	0.05	97.5193	97.5194	97.5193	97.5194
0.10	0.50	6	0.11	94.7138	94.7327	94.7338	94.7344
0.10	0.50	12	0.05	95.1375	95.1603	95.1619	95.1624
0.10	0.50	12	0.11	89.9539	90.0635	90.0697	90.0729
0.10	0.50	60	0.05	82.0624	83.3398	83.4165	83.4422
0.10	0.50	60	0.11	69.8774	72.2952	72.4238	72.4837
0.50	0.50	1	0.05	99.5802	99.5793	99.5793	99.5792
0.50	0.50	1	0.11	99.0918	99.0928	99.0928	99.0929
0.50	0.50	6	0.05	97.4081	97.3865	97.3854	97.3848
0.50	0.50	6	0.11	94.8171	94.8523	94.8542	94.8552
0.50	0.50	12	0.05	94.7177	94.6644	94.6618	94.6605
0.50	0.50	12	0.11	90.2849	90.4130	90.4199	90.4234
0.50	0.50	60	0.05	74.2063	74.7386	74.7885	74.8086
0.50	0.50	60	0.11	67.6248	68.4996	68.5585	68.6002

n = number of partitions

Interest rate follows equation (5.1)

Face value of the bond = $100

CIR = accurate value derived from Cox, Ingersoll, and Ross (1985)

5.2. Stock Options

Consider the CEV stock price process

$$dY = \mu Y dt + \sigma Y^{\gamma} dW, \qquad 0 < \gamma \le 1. \tag{5.11}$$

When $\frac{1}{2} \le \gamma < 1$ (which we assume hereafter), state 0 is an absorbing boundary. The transformation function is

$$y = f(x) = \begin{cases} [\sigma(1-\gamma)x]^{1/(1-\gamma)} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0, \end{cases}$$

with inverse $x = g(y) = \frac{y^{1-\gamma}}{\sigma(1-\gamma)}$, $(y \ge 0)$. The transformed process $X(t)$ follows the diffusion process

$$dX = [\mu(1-\gamma)X - \frac{1}{2(1-\gamma)X}]dt + dW. \tag{5.12}$$

The approximating binomial processes $X^{(n)}$ and $Y^{(n)}$ are defined by equations (3.6) and (3.7). Since $-0.5 \le \gamma < 1$, it can be shown that $q_h(y)$ is increasing for $y \ge 0$.[11] For any given number $r > 0$, we can always choose a sufficiently small h such that $q_h(y) > 0$ for all $0 < y < r$. This is true because $q_h(y)$ is decreasing and $\lim_{h\to 0} q_h(r) = 1/2$. Moreover, let y_{max} be the largest state on the binomial lattice. Then

$$\lim_{h\to 0} y_{max}^{1-\gamma}\sqrt{h} = \lim_{h\to 0} [y_0^{1-\gamma} + n(1-\gamma)\sigma\sqrt{h}]\sqrt{h} = (1-\gamma)\sigma T$$

---

[11] To see this, first consider that $0 < x \le \sqrt{h}$. Then

$$q_h(y) = (\mu h+1)y/y^{+} = (\mu h+1)[x/(x+\sqrt{h})]^{1/(1-\gamma)},$$

which is increasing in y. When $x \le \sqrt{h}$, let $z = (1-\gamma)\sigma\sqrt{h}y^{\gamma-1}$. Then

$$q_h(y) = \frac{\mu h + 1 - (1-z)^m}{(1+z)^m - (1-z)^m}$$

and

$$\frac{dq(y)}{dy} = \frac{-(\mu h+1)[(1+z)^{m-1}-(1-z)^{m-1}] + 2(1+z)^m(1-z)^m}{[(1+z)^m - (1-z)^m]^2/m}\frac{dz}{dy}.$$

For any $m > 1$ and $z > 0$, we have $(1+z)^{m-1} - (1-z)^{m-1} > 2$, $2(1+z)^m(1-z)^m < 2$, and $dz/dy < 0$. This implies that $dq_h(y)/dy > 0$.

and

$$\lim_{h \to 0} q(y_{max}) = \frac{1}{2} + (1-\gamma)\mu T/2.$$

Consequently, $q(y) < 1$ if $(1-\gamma)\mu T < 1$. Assumption 3b is now satisfied. This leads to the following result:

Corollary 3. For the CEV process (5.11), suppose $-0.5 \le \gamma < 1$. Let $Y_k^{(n)}$, $k = 0,1,\ldots,n$ be the binomial Markov chain with lattice generator (3.7) and transition probability (3.8). Then $Y^{(n)}(t)\ f(X_{[nt]}^{(n)}) \Rightarrow Y(t)$.

Having set up a converging binomial lattice for the stock price, options on the stock can be approximated using backward recursion on the lattice, as described in section 4. Table 2 shows the approximated values of the call and put options. We fix $\gamma = 0.5$ and set the annual risk-free rate at 5 percent. The parameter σ is standardized such that the initial annual volatility of the stock return is 20 percent. The initial stock price is \$40. The strike prices (X) range from \$35 to \$45, and the maturities are one, four, and seven months. The first three columns specify case parameters, the next three display call prices for different numbers of partitions in the approximation, and the seventh column reports the theoretical call prices from Cox and Rubinstein (1985, p. 364). Nelson and Ramaswamy (1990) tabulated approximations only for maturities of one and four months, and their results show the same degree of accuracy as ours. For longer maturities, they reported coarse approximations without tabulating the results. The results displayed here for a seven-month maturity clearly illustrate that our approximations not only converge but are also very accurate.

Columns eight through 11 are the prices of American put options. The last column contains the approximated values obtained by Nelson and Ramaswamy (1990) using finite-difference methods. (Again, they did not tabulate the results for a seven-month maturity.) The trends of our findings for all maturities clearly illustrate that our binomial approximations converge as n increases.

Table 2. Call and Put Option Prices

			Call Prices				Put Prices			
σ	X	T	n=5	n=50	n=100	CR	n=5	n=50	n=100	NR
0.2	35	1	5.142	5.150	5.150	5.15	0.000	0.008	0.008	0.008
0.2	40	1	1.049	0.998	1.000	1.00	0.900	0.849	0.851	0.842
0.2	45	1	0.012	0.018	0.019	0.02	5.000	5.000	5.000	5.000
0.3	35	1	5.212	5.232	5.232	5.23	0.070	0.090	0.091	0.088
0.3	40	1	1.531	1.455	1.458	1.45	1.383	1.305	1.307	1.296
0.3	45	1	0.103	0.141	0.142	0.14	5.032	5.044	5.045	5.042
0.4	35	1	5.434	5.412	5.418	5.42	0.294	0.272	0.277	0.270
0.4	40	1	2.013	1.911	1.916	1.92	1.865	1.761	1.765	1.751
0.4	45	1	0.416	0.386	0.385	0.38	5.286	5.256	5.255	5.245
0.2	35	4	5.803	5.782	5.787	5.79	0.245	0.223	0.228	0.223
0.2	40	4	2.256	2.159	2.163	2.17	1.661	1.573	1.576	1.570
0.2	45	4	0.496	0.450	0.469	0.47	5.060	5.064	5.064	5.060
0.3	35	4	6.358	6.319	6.314	6.31	0.801	0.765	0.761	0.751
0.3	40	4	3.208	3.060	3.067	3.07	2.617	2.472	2.476	2.470
0.3	45	4	1.220	1.194	1.189	1.18	5.696	5.645	5.641	5.628
0.4	35	4	6.943	6.976	6.991	6.99	1.386	1.430	1.443	1.430
0.4	40	4	4.161	3.963	3.972	3.98	3.572	3.374	3.380	3.374
0.4	45	4	1.909	2.001	2.009	2.00	6.390	6.407	6.414	6.345
0.2	35	7	6.492	6.436	6.444	6.43	0.522	0.472	0.477	
0.2	40	7	3.114	2.992	2.998	3.00	2.077	1.981	1.983	
0.2	45	7	1.091	1.041	1.036	1.04	5.256	5.215	5.214	
0.3	35	7	7.218	7.247	7.262	7.26	1.255	1.298	1.310	
0.3	40	7	4.359	4.170	4.179	4.19	3.330	3.155	3.159	
0.3	45	7	2.039	2.130	2.140	2.13	6.124	6.137	6.144	
0.4	35	7	8.068	8.248	8.227	8.23	2.172	2.305	2.284	
0.4	40	7	5.604	5.350	5.362	5.37	4.581	4.333	4.341	
0.4	45	7	3.140	3.311	3.281	3.29	7.131	7.257	7.231	

n = number of partitions
Stock prices follow equation (5.11) with γ = 0.5
σ = initial annual volatility of stock return
X = strike price
T = maturity in months
Initial stock price = $40
Annual interest rate = 5 percent
CR = accurate value adopted from Cox and Rubinstein (1985, p. 364)
NR = Nelson and Ramaswamy's (1990) approximations

6. Concluding Remarks

We have demonstrated that transformation is a useful tool for simplifying binomial models in diffusion approximations. We have also shown that singular diffusions are better approximated by reflecting or absorbing binomial processes. This is a promising result, and the idea can also be easily applied to finite-difference methods.

An alternative way to achieve computational simplicity within this framework is through lattice adjustment. For one-dimensional diffusions, this may be less efficient than the transformation method. Nonetheless, it may be worthwhile to develop an adjustment scheme for general multidimensional diffusions for which the transformation method fails.

Another contribution of this paper is the convergence result established in approximating American contingent claims. In almost all cases, the optimal early exercise boundaries cannot be analytically solved. However, the approach taken here does require an analytical formula for the boundary. All that is needed is the continuity of the first hitting time with respect to the sample path of the diffusion.

Appendix A. An Invariant Result

This appendix shows that the probability-symmetric binomial model (2.5) is invariant under time-homogeneous transformation. That is, path-independence cannot be achieved by a time-transformation while probability-symmetry is preserved. To see this, rewrite equation (2.5) as

$$y_h^{\pm} = y \pm \sigma(y,t)\sqrt{h} + \mu(y,t)h. \tag{A.1}$$

Consider the difference between the two states when one follows the up-then-down path and the other follows the down-then-up path:

$$y^{+-} - y^{-+} = [2\mu'_Y\sigma - 2\sigma'_Y\mu - \sigma''_{YY}\sigma^2 - 2\sigma'_t]\sqrt{h}^3 + o(h^2). \tag{A.2}$$

A necessary condition for path-independence is

$$2\mu'_Y\sigma - 2\sigma'_Y\mu - \sigma''_{YY}\sigma^2 - 2\sigma'_t = 0, \tag{A.3}$$

which is equivalent to

$$\mu - [\tfrac{1}{2}\sigma'_Y + \int(\sigma'_t/\sigma^2)dy]\sigma = 0. \tag{A.4}$$

Suppose a transformation $x = g(y,t)$ is employed. The transformed $X = g(Y,t)$ diffusion then follows

$$dX = M(X)dt + S(X)dW, \tag{A.5}$$

where

$$M(X) = \mu g'_Y + \tfrac{1}{2}\sigma^2 g''_{YY} + g'_t \text{ and}$$
$$S(X) = \sigma g'_Y.$$

Note that

$$\int(S'_t/S^2)dx = \int[(\sigma_t g'_Y + \sigma g''_{tY})/(\sigma g'_Y)^2]dx$$
$$= \int[\sigma_t/\sigma^2 + g''_{tY}/(\sigma g'_Y)]dy \qquad \text{(since } dx = g'_Y dy)$$
$$= \int(\sigma_t/\sigma^2)dy + g'_t/S - \int g'_t\frac{\partial}{\partial y}[\frac{1}{\sigma g'_Y}]dy.$$

We have

$$M - [\tfrac{1}{2}S'_X + \int (S'_t/S^2)dx]S = \mu g'_Y + \tfrac{1}{2}\sigma^2 g''_{YY} + g'_t - [\tfrac{1}{2}\sigma'_Y g'_Y + \tfrac{1}{2}\sigma g''_{YY}]\sigma$$

$$- \{\int (\sigma_t/\sigma^2)dy + g'_t/S - \int g'_t \frac{\partial}{\partial y}[\frac{1}{\sigma g'_Y}]dy\}\sigma g'_Y$$

$$= \{\mu/\sigma - \tfrac{1}{2}\sigma'_Y - \int (\sigma_t/\sigma^2)dy - \int g'_t \frac{\partial}{\partial y}[\frac{1}{\sigma g'_Y}]dy\}\sigma g'_Y.$$

Therefore, similar to equation (A.4), a necessary condition for the transformed binomial lattice to be path-independent is

$$\mu/\sigma - \tfrac{1}{2}\sigma'_Y - \int (\sigma_t/\sigma^2)dy - \int g'_t \frac{\partial}{\partial y}[\frac{1}{\sigma g'_Y}]dy = 0. \tag{A.6}$$

Obviously, under any homogeneous transformation ($g'_t = 0$), conditions (A.4) and (A.6) are the same. That is, the original and the transformed processes become path-independent at the same time. However, a nonhomogeneous transformation may make a difference.

Appendix B: Adjusted Binomial Lattice

Consider the binomial approximation of diffusion process (1.1). The adjusted binomial lattice (2.8) is obtained by adding an extra term $\lambda(y,t)h$ to the state-symmetric building block (2.3). To determine the local adjustment term $\lambda(y,t)$, calculate the up-down state (y^{+-}) and down-up state (y^{-+}) for the adjusted binomial model (2.8):

$$\begin{cases} y_h^{+-} = y^+ - \sigma(y^+,t)\sqrt{h} + \lambda(y^+,t)h \\ y_h^{-+} = y^- + \sigma(y^-,t)\sqrt{h} + \lambda(y^-,t)h. \end{cases} \tag{B.1}$$

The difference (gap) between these two states is

$$\begin{aligned} y^{+-} - y^{-+} &= [\lambda(y^+,t+h)-\lambda(y^-,t+h)]h + [2\sigma(y,t)-\sigma(y^+,t+h)-\sigma(y^-,t+h)]\sqrt{h} \\ &= [2\lambda_Y'(y,t)\sigma(y,t)-2\sigma_Y'(y,t)\lambda(y,t)-\sigma_{YY}''(y,t)\sigma^2(y,t)-2\sigma_t'(y,t)]\sqrt{h}^3 \\ &\quad + o(h^2). \end{aligned} \tag{B.2}$$

Therefore, the difference between the two expected merging states is of order $\sqrt{h}^3$, while that between nonmerging states is of order $\sqrt{h}$. To close this gap, we choose $\lambda(y,t)$ such that the coefficient of $\sqrt{h}^3$ in (B.2) becomes zero. Or, equivalently,

$$\frac{2\lambda_Y'(Y,t)\sigma(Y,t)-2\sigma_Y'(Y,t)\lambda(Y,t)}{\sigma^2(Y,t)} = \sigma_{YY}''(Y,t) + 2\sigma_t'(Y,t)/\sigma^2(Y,t). \tag{B.3}$$

Integrate both sides with respect to y and rearrange. Then

$$\lambda(y,t) = \sigma(y,t)[\tfrac{1}{2}\sigma_Y'(y,t) + \int \frac{\sigma_t'(y,t)}{\sigma^2(y,t)}\,dy]. \tag{B.4}$$

When $\sigma(y,t) = \sigma(y)$ does not depend on t, we simply choose

$$\lambda(y) = \tfrac{1}{2}\sigma(y)[\sigma_Y'(y) + C], \tag{B.5}$$

where C is a constant. For simplicity, C is set to 0. If we set $C = b$, then the adjusted binomial lattice (2.8) will capture the first three terms in the Taylor expansion of the alternative binomial model (2.16).

Appendix C. Pseudo Path-Independent Model

As noted in section 2, the adjusted binomial lattice may not be path-independent. However, under mildly smooth conditions, the gap between the up-then-down state (y^{+-}) and the down-then-up state (y^{-+}) is sufficiently small after the adjustment. By ignoring such minor differences, we obtain a pseudo path-independent lattice. To illustrate this idea, we specify a procedure for reconnecting the nodes as follows.

Pseudo Path-Independent Algorithm:

Step 1. Starting from the initial node y_0 at time 0, branch into two nodes using the adjusted jumping scheme in (2.8). Denote the two nodes in period 1 by $y_{1,0}(h)$ and $y_{1,1}(h)$:

$$\begin{aligned} y_{1,0}(h) &= y_0(h) + [\sigma(y_0)\sigma_Y(y_0)/2\ ]h - \sigma(y_0)\sqrt{h} \\ y_{1,1}(h) &= y_0(h) + [\sigma(y_0)\sigma_Y(y_0)/2\ ]h + \sigma(y_0)\sqrt{h}. \end{aligned} \qquad (C.1)$$

Step 2. At the end of period k (or time $t = kh$, $k \geq 1$), there are $k + 1$ nodes $y_{k,j}(h)$, $(0 \leq j \leq k)$. Construct the nodes for period $k + 1$ as follows:

$$\begin{aligned} y_{k+1,0}(h) &= y_{k,0}(h) + \tfrac{1}{2}\sigma(y_{k,0})\sigma_Y(y_{k,0})h - \sigma(y_{k,0})\sqrt{h} \\ y_{k+1,j+1}(h) &= y_{k,j}(h) + \tfrac{1}{2}\sigma(y_{k,j})\sigma_Y(y_{k,j})h + \sigma(y_{k,j})\sqrt{h}, \end{aligned} \qquad (C.2)$$

$$(0 \leq j \leq k)$$

(That is, except for the "bottom" node in period k+1, all other nodes are computed from upward moves from the previous period.)

Step 3. In period k, if the process is at state $y_{k,j}$, $j = 0,\ldots,k$, the lattice generator is

$$y_{k,j} \begin{cases} \xrightarrow{\ q_{k,j}\ } y_{k+1,j+1} \\ \xrightarrow{\ 1-q_{k,j}\ } y_{k+1,j} \end{cases} \qquad (C.3)$$

where

$$q_{k,j} = 0.5 + \frac{\mu(Y_{k,j},t) - \lambda(Y_{k,j},t)}{2\sigma(Y_{k,j},t)}\sqrt{h}. \tag{C.4}$$

Step 4. Repeat steps 2 and 3 until $k = n$.

In the above construction, we ignore the actual down jumps in the adjusted lattice except for the one that is always down. Note that an actual down jump in period k would create additional k nodes. Let $\{y^*_{k+1,j}(h),\ j = 0,1,\ldots,k\}$ represent these nodes. Then

$$y^*_{k+1,j}(h) = y_{k,j}(h) + \tfrac{1}{2}\sigma(y_{k,j})\sigma_Y(y_{k,j})h - \sigma(y_{k,j})\sqrt{h}, \quad 0 \le j \le k. \tag{C.5}$$

In general, $y^*_{k+1,j}(h) \neq y_{k+1,j}(h)$. The difference, however, is negligible. Therefore, in the above pseudo path-independent lattice, we ignore the $y^*(h)$ values completely and force the down jumps to reconnect with the up jumps from the immediate state below. Specifically, we bend the downward branches by

$$\Delta y_{k,j}(h) = y^*_{k+1,j}(h) - y_{k+1,j-1}(h), \qquad 1 \le j \le k \tag{C.6}$$

such that $k + 1$ pairs of nodes reconnect at time $(k + 1)h$.

In summary, for the pseudo path-independent algorithm, the up and down jumps at node j in period k are

$$y^+_{k,j}(h) = y_{k,j}(h) + \tfrac{1}{2}\sigma(y_{k,j})\sigma_Y(y_{k,j})h + \sigma(y_{k,j})\sqrt{h} \tag{C.7}$$

$$y^-_{k,j}(h) = y_{k,j-1}(h) + \tfrac{1}{2}\sigma(y_{k,j-1})\sigma_Y(y_{k,j-1})h + \sigma(y_{k,j-1})\sqrt{h} + \Delta y_{k,j}(h). \tag{C.8}$$

Lemma 3. If $\sigma(y,t)$ has locally bounded partial derivatives up to the fourth order in y and up to the second order in t, then

(1) in the adjusted lattice (2.8),

$$\Delta_2(y,h) = y^{+-} - y^{-+} = o(h^2) \text{ and} \tag{C.9}$$

(2) in the pseudo path-independent algorithm,

$$\Delta_{k,j}(h) = o(h), \quad (k > 1, \ 1 \le j \le k). \tag{C.10}$$

Proof: Equation (C.9) follows directly from the discussion in section 2.2. To prove equation (C.10), note that

$$\Delta_{k,1}(h) = \Delta_2(y_{k-1,0}, h) = o(h^2). \qquad \forall\, k > 0.$$

Repeatedly using equation (C.9) yields

$$\begin{aligned}
\Delta_{k,2}(h) &= \Delta_2(y_{k-2,1}, h) + \Delta_{k-1,1}(h)[1+O(h)] \\
&= o(h^2) + o(h^2) \\
&= 2\ o(h^2) \\
\cdots \\
\Delta_{k,j}(h) &= \Delta_2(y_{k-2,j-1}, h) + \Delta_{k-1,j-1}(h)[1+O(h)] \\
&= o(h^2) + (j-1)o(h^2) \\
&= j\ o(h^2).
\end{aligned}$$

Since $j \le n = T/h$, in the worst case, we have

$$\begin{aligned}
\Delta_{k,j}(h) &= n\ o(h^2) \\
&= o(h) \qquad \text{for all } k > 1, \text{ and } j = 1,\ldots,k.
\end{aligned}$$

Proposition C. Suppose the diffusion equation (1.1) has an *a.e.* unique solution $Y(t)$ for any given $Y(0)$. Let $Y^{(n)}(k)$ be the binomial Markov chain corresponding to the pseudo path-independent lattice in (C.7) - (C.8) and the transition probability (2.12). Define $Y^{(n)}(t) = Y^{(n)}_{[nt]}$. Then $Y^{(n)}$ weakly converges to $Y(t)$.

Proof. From lemma 2, we can rewrite equation (C.8) as

$$y^-_{k,j}(h) = y_{k,j}(h) + \tfrac{1}{2}\sigma(y_{k,j})\sigma_Y(y_{k,j})h + \sigma(y_{k,j})\sqrt{h} + O(\sqrt{h}^3).$$

To simplify the notation, we use y^+, y^-, and y for $y^+_{k,j}(h)$, $y^-_{k,j}(h)$, and $y_{k,j}(h)$, respectively. Recall the transition probability (C.4). We can calculate the local drift $\mu_h(y,t)$ and second moment $\sigma_h(y,t)$ as follows:

$$\mu_h(y,t) = \{q(y,t)[y^+ - y] + (1 - q(y,t))[y^- - y]\}/h$$

$$= \{q(y,t)[\lambda(y,t)h + \sigma(y,t)\sqrt{h}]$$

$$+ [1 - q(y,t)][\lambda(y,t)h - \sigma(y,t)\sqrt{h} + O(\sqrt{h}^3)]\}/h$$

$$= \{\lambda(y,t)h + [2q(y,t)-1]\sigma(y,t)\sqrt{h} + O(\sqrt{h}^3)\}/h$$

$$= \mu(y,t) + O(\sqrt{h})$$

$$\sigma_h^2(y,t) = \{q(y,t)[y^+ - y]^2 + (1 - q(y,t))[y^- - y]^2\}/h$$

$$= \{q(y,t)[\lambda(y,t)h + \sigma(y,t)\sqrt{h}]^2$$

$$+ [1 - q(y,t)][\lambda(y,t)h - \sigma(y,t)\sqrt{h} + O(\sqrt{h}^3)]^2\}/h$$

$$= \{\sigma^2(y,t)h + O(h^2)\}/h$$

$$= \sigma^2(y,t) + O(h).$$

We use the fact that $[q(y,t) - 1]\sigma(y,t) = [\mu(y,t) - \lambda(y,t)]\sqrt{h}$. Thus, the local drift $\mu_h(y,t)$ and second moment $\sigma_h(y,t)$ converge to the true drift $\mu(y,t)$ and moment $\sigma(y,t)$. From lemma 1, the pseudo path-independent binomial process weakly converges to the corresponding diffusion.

Appendix D: Proof of Equations (2.13) and (2.16)

To prove equation (2.13), use the Taylor expansion for (2.11); then

$$y_h^{\pm} = f(x,t) + \frac{\partial f(x,t)}{\partial x}\sqrt{h} + \frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2}h + \frac{\partial f(x,t)}{\partial t}h + o(h). \tag{D.1}$$

Note that $f(x,t) = y$ and that the two partial derivatives of f with respect to y in the above equation are

$$\frac{\partial f(x,t)}{\partial x} = \sigma(y,t) \quad \text{and} \quad \frac{\partial^2 f(x,t)}{\partial x^2} = \sigma_Y'(y,t)\sigma(y,t).$$

To find $\frac{\partial f(x,t)}{\partial t}$, note that $x = g(y,t)$. Thus, $0 = \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial g}{\partial t}$, and

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial y}{\partial t} = -\frac{\partial g/\partial t}{\partial g/\partial y} = \sigma(y,t)\int_0^Y \frac{\sigma_t'(y,t)}{\sigma^2(y,t)}\,dy.$$

Substituting these derivatives into equation (D.1) yields equation (2.13).

Next, we prove equation (2.16). In equation (2.9), we choose g such that $X(t)$ has a linear volatility function:

$$\sigma(y)\frac{\partial g}{\partial Y} = a + bx \tag{D.2}$$

or

$$\frac{1}{b}\ln(a+bx) = g(y) = \int^Y \frac{dZ}{\sigma(Z)}. \tag{D.3}$$

To approximate $X(t)$, we use the Cox-Rubinstein binomial model (2.13):

$$x_h^{\pm} = x \pm (a + bx)\sqrt{h}. \tag{D.4}$$

The corresponding binomial model for $y(t)$ is

$$y_h^{\pm} = f(x \pm (a + bx)\sqrt{h}) \tag{D.5}$$

with up-jump probability (2.12). Using the Taylor expansion on the above equation, we obtain equation (2.16). Here, the derivatives used are

$$\frac{\partial f(X)}{\partial X} = \frac{\sigma(Y)}{a + bX} \quad \text{and} \quad \frac{\partial^2 f(X)}{\partial X^2} = \frac{\frac{\partial \sigma(Y)}{\partial Y}}{a + bX}\frac{\partial f(X)}{\partial X} - \frac{b\sigma(Y)}{(a+bX)^2} = \frac{[\sigma_Y'(Y)+b]\sigma(Y)}{(a + bX)^2}.$$

Appendix E: Controlling Step Size

E1. Reflecting Boundary

This subsection shows how to control the partition size so that the approximating binomial chain is a true random walk with a reflecting barrier. Even though the process to be approximated has a reflecting barrier at 0, the barrier for the Markov chain is a small positive state that approaches 0 as the partition size diminishes. Such a barrier can be constructed by solving $q(y) = 1$. Let y^* be the solution. We can choose a partition size h such that y^* will be a state for the binomial process. Thus, when the process reaches y^*, it can be reflected with probability 1.

This is best explained by way of example. Let $Y(t)$ be the MRSR process (5.1). The transformed process $X(t)$ is given by equation (5.3). The approximated boundary x^* can then be calculated using equation (5.6). Suppose the transformed process $X(t)$ starts from $X(0)$ and hits the boundary x^* in exactly m steps by following an always-down path. That is, $X(0) - m\sqrt{h} = x^*$, which gives

$$\sqrt{h} = \frac{X(0)-x^*}{m}. \tag{E.1}$$

The number of partitions of the time period $[0,T]$ would be the largest integer that is less than or equal to T/h; i.e., $n = [T/h]$. Since T/m may not be an integer, we simply assume that the binomial process stays at the same state on the residual interval $[nh,T]$.

If we choose $x^* = \phi\sqrt{h}/2$, then $\sqrt{h} = \dfrac{X(0)}{m+\phi/2}$. Starting with $X(0)$, after m steps, the always-down state will be

$$x(mh) = x^* = \phi\sqrt{h}/2 \tag{E.2}$$

or

$$y(mh) = y^* = (\phi\sigma)^2 h/16, \tag{E.3}$$

and

$$q(y^*,t) = \frac{1}{2} + \frac{\phi\sigma^2/4 - \kappa y^*}{2\sigma\sqrt{y^*}}\sqrt{h} = 1 - \frac{\kappa\phi h}{8}. \tag{E.4}$$

If we choose $q(y^*, mh) = 1$, the binomial process will never go down any farther once it reaches y^*. With probability 1, the process jumps up to

$$y^+(mh + h) = (x^* + \sqrt{h})^2\sigma^2/4. \qquad \text{(E.5)}$$

The corresponding local drift $\mu_h(y^*, t)$ and second moment $\sigma_h(y^*, t)$ are

$$\mu_h(y^*, mh) = [y^+(mh + h) + y^*]/h = \kappa\mu \text{ and} \qquad \text{(E.6a)}$$

$$\sigma_h^2(y^*, mh) = [y^+(mh + h) - y^*]^2/h = (\kappa\mu)^2 h. \qquad \text{(E.6b)}$$

The true drift and variance at state y^* are

$$\mu(y^*, mh) = \kappa(\mu - y^*) = \kappa\mu - \kappa(\phi\sigma)^2 h/16 = \kappa\mu + O(h) \text{ and} \qquad \text{(E.7a)}$$

$$\sigma^2(y^*, mh) = \sigma^2 y^* = \phi^2\sigma^4 h/16 = O(h). \qquad \text{(E.7b)}$$

Thus, $\mu_h(y, t)$ and $\sigma_h(y, t)$ converge to $\mu(y, t)$ and $\sigma(y, t)$, respectively.

Proposition E1. Suppose the binomial lattice for the MRSR process (5.1) is generated by (3.2a) and (3.2b). Let $\phi = 4\kappa\mu/\sigma^2 - 1 > 0$. Suppose the transformed process $X(t)$ starts at t with $X(0)$ such that $m = \sqrt{n/T}\, X(0) - \phi/2$ is an integer less than n. Define an approximated boundary y^* as in equation (E.3). Let the transition probability be defined by (2.12) when $y > y^*$. At the approximated boundary y^*, set $q(y^*, t) = 1$. Then the resulting binomial process weakly converges to $y(t)$.

E2. Absorbing/Reflecting Boundary

Again, we use the MRSR process (5.1) to illustrate our method. Assume $-1 \le \phi = 4\kappa\mu/\sigma^2 - 1 < 0$. Then $y = 0$ is a sticky boundary. Let y^* be the small state such that one up jump from 0 to y^* with probability 1 matches the local mean exactly with the drift. That is,

$$y^* = \kappa(\mu - 0)h = \frac{\sigma^2(1+\phi)}{4} h, \qquad \text{(E.8)}$$

or equivalently, $x^* = \sqrt{1+\phi}\sqrt{h}$ for the transformed process. We control the step size h such that if the process starts from $X(0)$ and follows an always-down path, it will hit the small state y^* in exactly m steps.

That is, $X(0) - m\sqrt{h} = x^*$. For any state y above y^*, $0 \le q(y,t) \le 1$. At x^*, X can either jump up to $x^* + \sqrt{h}$ with probability

$$q(y^*,t) = \frac{2(1+\phi)}{2+2\sqrt{1+\phi}+\phi} + o(h), \tag{E.9}$$

or it can jump down to 0 with probability $1 - q(y^*,t)$. The true drift and variance at state y^* are

$$\mu(y^*,mh) = \kappa(\mu - y^*) = \kappa\mu + \kappa\frac{\sigma^2(1+\phi)}{4}h = \kappa\mu + O(h) \text{ and} \tag{E.10a}$$

$$\sigma^2(y^*,mh) = \sigma^2 y^* = \frac{\sigma^4(1+\phi)}{4}h = O(h). \tag{E.10b}$$

The corresponding local drift $\mu_h(y^*,t)$ and second moment $\sigma_h(y^*,t)$ are

$$\begin{aligned}\mu_h(y^*,mh) &= \{(y^*,t)[\frac{\sigma^2(x^*+\sqrt{h})^2}{4} - \frac{\sigma^2(x^*)^2}{4}] + [1-q(y^*,t)][-\frac{\sigma^2(x^*)^2}{4}]\}/h \\ &= q(y^*,t)[\frac{\sigma^2(x^*+\sqrt{h})^2}{4}] - \frac{\sigma^2(x^*)^2}{4} = \kappa\mu \text{ and}\end{aligned} \tag{E.11a}$$

$$\begin{aligned}\sigma_h^2(y^*,mh) &= \{q(y^*,t)[\frac{\sigma^2(x^*+\sqrt{h})^2}{4} - \frac{\sigma^2(x^*)^2}{4}]^2 + [1-q(y^*,t)][\frac{-\sigma^2(x^*)^2}{4}]^2\}/h \\ &= O(h).\end{aligned} \tag{E.11b}$$

Proposition E2. Suppose the binomial lattice for the MRSR process (5.1) is generated by (5.4). Let $\phi = 4\kappa\mu/\sigma^2 - 1 > 0$. Suppose the transformed process $X(t)$ starts with $X(0)$ such that $m = \sqrt{n/T}\, X(0) - \sqrt{1+\phi}\,\sqrt{h}$ is an integer less than n. Define an approximated boundary y^* as in equation (E.8). Let the transition probability be defined by equation (5.5) when $y > y^*$, and let $q(y^*,t)$ be given by equation (E.9). Set $q(0,t) = 0$. Then the resulting binomial process weakly converges to $Y(t)$.

References

Amin, K., 1991, "On the Computation of Continuous Time Option Prices Using Discrete Approximations," *Journal of Financial and Quantitative Analysis*, 26, 4, 477-495.

Billingsley, P., 1968, *Convergence of Probability Measures*, John Wiley & Sons, New York.

Black F. and M. Schole, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economics*, 81, 637-659.

Cox, J. C., Ingersoll, J. E., Jr., and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

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

Cox, J. C. and M. Rubinstein, 1985, *Options Markets*, Prentice-Hall, Englewood Cliffs, N.J.

Dynkin, E. B., 1965, *Markov Processes*, Springer-Verlag, Berlin.

Ethier, S. N., and T. G. Kurtz, 1986, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York.

Ingersoll, J. E., Jr., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield, Totowa, N.J.

Harrison, J. M. and D. M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381-408.

Harrison, J. M. and S. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 215-260.

He, H., 1991, "Convergence from Discrete to Continuous Time Financial Models," *Review of Financial Studies*, 3, 523-546.

Hull, J. and A. White, 1990, "Valuing Derivative Securities Using the Explicit Finite Difference Method," *Journal of Financial and Quantitative Analysis*, 25, 1, 87-99.

Karatzas, I., 1988, "On the Pricing of American Options," *Applied Mathematics and Optimization*, 17, 37-60.

Kushner, H. J., 1984, *Approximation and Weak Convergence Methods for Random Processes*, MIT Press, Cambridge, Mass.

Kushner, H. J., 1990, "Numerical Methods for Stochastic Control Problems in Continuous Time," *SIAM Journal of Control and Optimization*, 28, 999-1048.

Nelson, D.B., 1990, "ARCH Models as Diffusion Approximation," *Journal of Econometrics*, 45, 7-38.

Nelson, D.B. and K. Ramaswamy, 1990, "Simple Binomial Processes as Diffusion Approximations in Financial Models," *Review of Financial Studies*, 3, 393-430.

Sharpe, W. F., 1978, *Investment*, Prentice-Hall, Englewood Cliffs, N.J.

Shiryayev, A. N., 1978, *Optimal Stopping Rules*, Springer-Verlag, New York.

Trigeorgis, L., 1991, "A Log-Transformed Binomial Numerical Analysis Method for Valuing Complex Multi-Option Investments," *Journal of Financial and Quantitative Analysis*, 26, 3, 309-326.

Van Moerbeke, P., 1976, "On Optimal Stopping and Free Boundary Problems," *Archive for Rational Mechanics and Analysis*, 60, 101-148.