SIMPLE BINOMIAL PROCESSES AS DIFFUSION APPROXIMATIONS IN FINANCIAL MODELS

by

Daniel B. Nelson
Krishna Ramaswamy

30-89

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH

Simple Binomial Processes as Diffusion Approximations in Financial Models

Daniel B. Nelson†
Krishna Ramaswamy‡

Abstract

A binomial approximation to a diffusion is defined as *computationally simple* if the number of nodes grows at most linearly in the number of time intervals. This paper shows how to construct computationally simple binomial processes which converge weakly to commonly employed diffusions in financial models. It also demonstrates the convergence of the sequence of bond and European option prices from these processes to the corresponding values in the diffusion limit. Numerical examples from the Constant Elasticity of Variance stock price and the Cox, Ingersoll, and Ross discount bond price are provided.

June, 1989
Revised March, 1990
Forthcoming, *The Review of Financial Studies*

† Graduate School of Business, The University of Chicago, 1101 E. 58th Street, Chicago, IL 60637
‡ The Wharton School of The University of Pennsylvania, Philadelphia, PA 19104
We thank George Constantinides, John Cox, Darrell Duffie, Dean Foster, Andrew Lo, Patrick Waldron, the referees, and seminar participants at Carnegie-Mellon, N.Y.U., The University of Chicago, The University of Waterloo and The Wharton School for helpful comments.

1 Introduction

The seminal work of Merton (1969) and Black and Scholes (1973) paved the way for the use of continuous-time models in finance. The usefulness of the underlying mathematical techniques has never been in doubt: the pricing of options and other contingent claims has relied heavily on these techniques. When Sharpe (1978) developed the binomial approach, the option pricing model became accessible to a much wider audience. Cox, Ross and Rubinstein (1979) showed that a suitably defined binomial model for the evolution of the stock price converges weakly[1] to a lognormal diffusion as the time between binomial jumps shrinks toward zero; and they also showed in this case that the European option's value in the binomial model converges to the value given by the Black-Scholes formula. Cox and Rubinstein (1985) exploit this approach to value American options on dividend paying stocks, and they also show how to employ the binomial approach when some of the other assumptions made by Black and Scholes are relaxed. In fact, Cox and Rubinstein demonstrate the connexion between the continuous-time valuation equation (which is the fundamental partial differential equation for the contingent claim) and the discrete time, one-period valuation formula developed under the assumption of a binomial model for stock prices — both are descriptions of the local behavior of the contingent claim's value in relation to the underlying asset.

In a normative sense, the binomial model has enabled users to value contingent claims in some restrictive settings where a closed form solution is unavailable. By the reasoning provided in Cox and Rubinstein (and made explicit in Brennan and Schwartz (1978)) this is formally equivalent to a numerical solution to the partial differential equation — the binomial model provides one such solution, requires elementary methods in implementation and has the splendid virtue of being pedagogically useful.

This is one reason why the stochastic differential equation defining the lognormal diffusion has become the workhorse in option pricing models. Binomial approximation[2] and valuation methods have been applied to other diffusions besides the lognormal (for example, the constant-elasticity of variance or CEV diffusion in Cox and Rubinstein (1985, p. 362)) — however, it turns out that the binomial tree structures available in these cases are com-

[1]By weak convergence we mean convergence in distribution; see footnote 7 for a discussion employing the notation used in this paper. The first part of the appendix provides the technical background for this definition.

[2]Tom Nagylaki pointed out that our use of the term "binomial process" is an abuse of terminology: in Cox, Ross, and Rubinstein (1979), the term was strictly correct, since the log of the stock price at any time period had a binomial distribution. This will not, in general, be the case for the diffusion approximations proposed in this paper. In the finance literature, however, the term "binomial process" has come to refer more generally to two-state models of the sort discussed in this paper (see, for example, Cox and Rubinstein (1985) p. 361).

putationally complex in that the number of nodes doubles at each time step. And there are still other diffusions employed in financial models (for example, for interest rates) for which the availability of a computationally simple sequence of binomial processes would be useful. We define a computationally simple tree (for an example, see Figure 1) as one where the number of nodes in the tree structure grows at most linearly with the number of time intervals.

This paper develops conditions under which a sequence of binomial processes converges weakly to a diffusion, and demonstrates a procedure that can be used to find a computationally simple binomial tree, given the diffusion limit to which we wish to take the sequence of such trees. In words, the conditions require that the instantaneous drift and the instantaneous variance of the diffusion process are well-behaved, and that the local drift and local variance in the binomial representation converge to the instantaneous drift and variance respectively; and because the sample paths of the limiting diffusion are continuous, we also require that the jump sizes converge to zero in a sensible way. Thus, the upward and downward jump, as well as the probability of an up-move in the binomial representation are chosen to match the local drift and variance. We meet the requirement that the tree be computationally simple by ensuring that, within the binomial representation, an up-move followed by a down-move causes a displacement in the value of the process that is the same when the moves take place in the reverse order. This is achieved by employing a transform of the process which takes the diffusion and removes its heteroskedasticity. Computational simplicity is achieved for the transformed (homoskedastic) process, and the inverse transform enables us to recover the original process. The sizes of the up and down moves, as well as the probability of an up-move, can depend on the level of the process, the behavior of the diffusion at certain boundaries, and on calendar time. The implementation of the binomial method is straightforward. We compare known solutions for options and bonds to those obtained numerically from the binomial model.

The paper is organized as follows. Section 2 develops the assumptions and presents the basic theorem which enables the construction of a sequence of binomial processes; it also gives the general conditions under which one can apply the transformation and construct computationally simple binomial processes. Two examples, which demonstrate how one can modify the binomial model to capture boundary behavior and retain computational simplicity, are also given. Section 3 provides the justification for employing the binomial model for valuation and also gives numerical solutions for three cases. Section 4 has some concluding comments. The proofs and technical details are collected in the Appendix.

2 Stochastic Differential Equations and Simple Binomial Approximations

In this Section, we state conditions for a sequence of binomial processes to converge weakly to a diffusion, and develop a technique for constructing computationally simple binomial diffusion approximations. We provide three examples: the Ornstein-Uhlenbeck process (for which the binomial representation is well known (see Cox and Miller (1984)), the CEV process introduced by Cox and Ross (1976), and the one-factor interest rate process of Cox, Ingersoll and Ross (1979).

2.1 Binomial Diffusion Approximations

Suppose we are given the stochastic differential equation

$$dy_t = \mu(y,t)\,dt + \sigma(y,t)\,dW_t, \tag{1}$$

where $\{W_t, t \geq 0\}$ is a standard Brownian motion, $\mu(y,t)$ and $\sigma(y,t) \geq 0$ are the instantaneous drift and standard deviation of y_t, and y_0 is a constant. We wish to find a sequence of binomial processes that converges in distribution to the process (1) over the time interval $[0,T]$. We first take the sequence of binomial processes as given, and give conditions to check whether the sequence converges to the diffusion (1). We then tackle the problem of constructing a sequence of binomial approximations, given a limit diffusion.

To fix matters, take the interval $[0,T]$, and chop it into n equal pieces of length $h \equiv T/n$. For each h consider a stochastic process $\{{}_hy_t\}$ on the time interval $[0,T]$, which is constant between nodes and, at any given node, either jumps up (down) some specified distance with probability q (respectively, $1-q$). For example, if we set $q = \frac{1}{2}$ and the up or down jump size equal to $\sqrt{h}$, it is well known that as $n \to \infty$, $\{{}_hy_t\}$ converges in distribution to a Brownian motion.

The sizes and probabilities of up or down jumps are specified as follows: define $q_h(y,hk)$, $Y_h^+(y,hk)$, and $Y_h^-(y,hk)$ to be scalar-valued functions defined on $\Re^1 \times [0,\infty)$ satisfying

$$0 \leq q_h(y,hk) \leq 1 \qquad \text{and} \qquad -\infty < Y_h^-(y,hk) \leq Y_h^+(y,hk) < \infty \tag{2}$$

for all $y \in \Re^1$ and all $k = 0, 1, 2, \ldots n$. The stochastic process followed by ${}_hy_t$ is given by

$${}_hy_0 = y_0, \quad \text{for all} \quad h, \tag{3}$$

$$ {}_hy_t = {}_hy_{kh}, \quad kh \le t < (k+1)h, \tag{4} $$

$$ P[\,{}_hy_{(k+1)h} = Y_h^+({}_hy_{hk}, hk) | hk,\, {}_hy_{hk}] = q_h({}_hy_{hk}, hk), \tag{5} $$

$$ P[\,{}_hy_{(k+1)h} = Y_h^-({}_hy_{hk}, hk) | hk,\, {}_hy_{hk}] = 1 - q_h({}_hy_{hk}, hk), \tag{6} $$

$$ P[\,{}_hy_{(k+1)h} = c | hk,\, {}_hy_{hk}] = 0, \quad \text{for} \quad c \neq Y_h^+({}_hy_{hk}, hk), \quad c \neq Y_h^-({}_hy_{hk}, hk). \tag{7} $$

The stochastic process ${}_hy_t$ is a step function with initial value y_0 which jumps only at times $h, 2h, 3h, \ldots$ At each jump the process can make one of two possible moves: up to a value Y_h^+ or down to a value Y_h^-. q_h is the probability of an upward move. Y_h^+, Y_h^-, and q_h are all allowed to depend on h, on the value of the process immediately before the jump (${}_hy_{hk}$), and on the time index hk. By the statement in equations (3) to (7) the process described is a Markov chain.

We apply a result[3] from Stroock and Srinivasa Varadhan (1979, Section 11.2) which states conditions under which $\{{}_hy_t\}_{h\downarrow 0}$ converges weakly to the y_t process in (1). To use this result we need assumptions about both the limiting stochastic differential equation and the sequence of Markov chains defined above. The first two assumptions ensure that the limiting stochastic differential equation (1) is well-behaved.[4]

Assumption 1 *The functions $\mu(y,t)$ and $\sigma(y,t)$ are continuous, and $\sigma(y,t)$ is always non-negative.*

Assumption 2 *With probability one, a solution $\{y_t\}$ of the stochastic integral equation*

$$ y_t = y_0 + \int_0^t \mu(y_s, s)\, ds + \int_0^t \sigma(y_s, s)\, dW_s, \tag{8} $$

exists for $0 < t < \infty$, and is distributionally unique.[5]

---

[3]Variations and extensions of these results are found in Kushner (1984), and Ethier and Kurtz (1986), Section 7.4.

[4]Most, but not all, of the stochastic differential equations commonly used in financial economics satisfy Assumptions 1 and 2. Consider, for example, the Brownian Bridge bond price process used in Ball and Torous (1983):

$$ dB_t = -\frac{dt}{T-t} + s\, dW_t. $$

This process is defined on the time interval $[0, T]$, where T is the maturity date of the bond. As $t \to T$ the drift rate explodes, violating Assumption 1. As Cheng (1989) has shown, however, this bond pricing process admits arbitrage.

[5]Assumption 2 is much weaker than the more familiar requirement of pathwise existence and uniqueness of solutions to (1), in that a realization of the Brownian motion $\{W_t\}$ need not map uniquely into a realization of the sample path of $\{Y_t\}$; many realizations may be possible, sharing a common distribution on the space of all continuous mappings from $[0, \infty)$ into $\Re^1$ — see Ethier and Kurtz (1986) Section 5.3, and Liptser and Shiryayev (1977) Section 4.4. Stroock and Srinivasa Varadhan (1979), Chapters 6, 7, 8, and 10 give conditions which imply that Assumption 2 holds. A number of these conditions are summarized in Nelson (1989, Appendix A).

Under Assumption 2, the distribution of the random process $\{y_t\}_{0 \leq t < T}$ is characterized by four things:

a. the starting point y_0,

b. the continuity (with probability one) of y_t as a random function of t,

c. the drift function $\mu(y, t)$, and

d. the diffusion function $\sigma^2(y, t)$.

If $\{{}_h y_t\}_{h \downarrow 0}$ is to converge in distribution to $\{y_t\}$, properties (a)-(d) must be matched in the limit. Specifically, we require:

a′. that ${}_h y_0 = y_0$ for all h,

b′. that the jump sizes of ${}_h y_t$ become small at a sufficiently rapid rate as $h \downarrow 0$,

c′. that the drift of ${}_h y_t$ converges (in a sense to be made precise below) to $\mu(y, t)$, and

d′. that the local variance of ${}_h y_t$ converge to $\sigma^2(y, t)$.

Note that a′ is assured by (3). To ensure b′, we make the following

Assumption 3 *For all $\delta > 0$, and all $T > 0$*

$$\lim_{h \downarrow 0} \sup_{|y| \leq \delta, 0 \leq t < T} \left| Y_h^+(y, t) - y \right| = 0, \tag{9}$$

$$\lim_{h \downarrow 0} \sup_{|y| \leq \delta, 0 \leq t < T} \left| Y_h^-(y, t) - y \right| = 0. \tag{10}$$

For c′ and d′, define for any $h > 0$ the local drift $\mu_h(y, t)$ and the local second moment[6] $\sigma_h^2(y, t)$ of the binomial process (3) – (7) by

$$\mu_h(y, t) \equiv \frac{q_h(y, t^*)[Y_h^+(y, t^*) - y] + (1 - q_h(y, t^*))[Y_h^-(y, t^*) - y]}{h} \quad \text{and} \tag{11}$$

$$\sigma_h^2(y, t) \equiv \frac{q_h(y, t^*)[Y_h^+(y, t^*) - y]^2 + (1 - q_h(y, t^*))[Y_h^-(y, t^*) - y]^2}{h}, \tag{12}$$

with $t^* \equiv h \cdot [t/h]$, where $[t/h]$ is the integer part of t/h. The next assumption requires that μ_h and σ_h^2 converge uniformly to μ and σ^2 on sets of the form $|y| \leq \delta, 0 \leq t < T$.

[6]This is not the local variance, because the moment is centered around y and not around the conditional mean. As $h \downarrow 0$, however, the local variance and second moment approach the same limit.

Assumption 4 *For every $T > 0$ and every $\delta > 0$,*

$$\lim_{h\downarrow 0} \sup_{|y|\leq\delta, 0\leq t<T} |\mu_h(y,t) - \mu(y,t)| = 0, \tag{13}$$

and

$$\lim_{h\downarrow 0} \sup_{|y|\leq\delta, 0\leq t<T} \left|\sigma_h^2(y,t) - \sigma^2(y,t)\right| = 0. \tag{14}$$

Theorem 1 *Under Assumptions 1 through 4, $\{{}_h y_t\} \Rightarrow \{y_t\}$, where "$\Rightarrow$" denotes weak convergence (i.e., convergence in distribution[7]) and $\{y_t\}$ is the solution of (1).* □

As an example, consider the well-known Ornstein-Uhlenbeck process (the continuous time version of the first-order autoregressive process), employed in the bond pricing model of Vasicek (1977):

$$dy_t = \beta(\alpha - y_t)\,dt + \sigma\,dW_t, \tag{15}$$

where β is non-negative, and y_0 is fixed. Define a sequence $\{{}_h y_t\}_{h\downarrow 0}$ of binomial approximations to (15) with common initial value y_0 and

$$Y_h^+(y,t) \equiv y + \sigma\sqrt{h}, \tag{16}$$

$$Y_h^-(y,t) \equiv y - \sigma\sqrt{h}, \tag{17}$$

and let

$$q_h \equiv \begin{cases} (1/2) + \sqrt{h}\beta(\alpha - y)/2\sigma & \text{if} \quad 0 \leq (1/2) + \sqrt{h}\beta(\alpha - y)/2\sigma \leq 1, \\ 0 & \text{if} \quad (1/2) + \sqrt{h}\beta(\alpha - y)/2\sigma < 0, \\ 1 & \text{otherwise.} \end{cases} \tag{18}$$

The probability q_h is chosen to match the drift; it is censored if it falls outside $[0,1]$. It is straightforward to verify that Assumption 2 is satisfied (Arnold (1974), Section 8.3), and to show that Assumptions 1 and 3 hold. The local drift and second moment are

$$\mu_h(y) = \begin{cases} \beta(\alpha - y) & \text{if} \quad 0 < q_h < 1, \\ \sigma/\sqrt{h} & \text{if} \quad q_h = 1, \\ -\sigma/\sqrt{h} & \text{if} \quad q_h = 0, \end{cases} \tag{19}$$

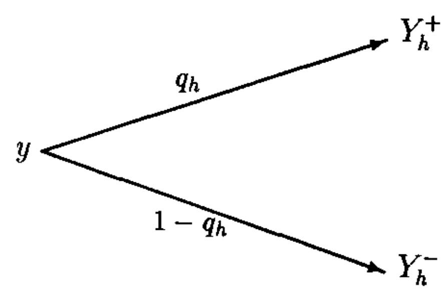
[7] "Convergence in distribution" means that the probability measures corresponding to the sequence of $\{{}_h y_t\}$ processes converge weakly to the probability measure of the $\{y_t\}$ process in (1); this is in a space of functions that are continuous from the right with finite left limits, endowed with the Skorohod metric (the Appendix provides further definitions). Weak convergence implies, for example, that given times $t_1, t_2, \ldots t_n > 0$, the joint distributions of $\{{}_h y_{t_1}, {}_h y_{t_2}, \ldots {}_h y_{t_n}\}$ converge to the joint distribution of $(y_{t_1}, y_{t_2}, \ldots y_{t_n})$ as $h \downarrow 0$. More generally, weak convergence implies that if $f(\cdot)$ is a continuous functional, then $f(\{{}_h y_t\})$ converges in distribution to $f(\{y_t\})$ as $h \downarrow 0$. For a discussion of the implications of weak convergence, see Billingsley (1968).

and

$$\sigma_h^2(y) = \sigma^2. \tag{20}$$

By definition, for any $\delta > 0$, q_h converges uniformly to $\frac{1}{2}$ on the set $|y| \leq \delta$ as $h \downarrow 0$. Therefore the local drift of the binomial process $\{_h y_t\}$ converges uniformly on compact sets to the instantaneous drift of the stochastic differential equation; and the local second moment identically equals the instantaneous variance, so Assumption 4 holds. We then apply Theorem 1 to conclude that $\{_h y_t\} \Rightarrow \{y_t\}$ as $h \downarrow 0$.

The intuition underlying the construction of a *simple* binomial sequence is uncomplicated. Suppose, following the suggestion in Cox and Rubinstein (1985), Section 7.1, we use the binomial jumps described by



as the basic building block for a binomial tree, where

$$Y_h^+ \equiv y + \sqrt{h}\sigma(y,t), \tag{21}$$

$$Y_h^- \equiv y - \sqrt{h}\sigma(y,t), \quad \text{and} \tag{22}$$

$$q_h \equiv \frac{1}{2} + \sqrt{h}\mu(y,t)/2\sigma(y,t). \tag{23}$$

In (21)–(23), h is the time interval between jumps, and q_h is the probability of a jump to Y_h^+. The total displacement is $\sqrt{h}[-\sigma(y,t) + \sigma(Y_h^-, t+h)]$ if an up move follows a down move, and it is $\sqrt{h}[\sigma(y,t) - \sigma(Y_h^+, t+h)]$ if a down move follows an up move. In general, these are not equal, so the branches of the binomial tree do not re-connect and the number of nodes doubles at each time step. However, whenever Assumptions 1–4 are satisfied by this binomial sequence (which is often the case), weak convergence will follow. But such a computationally complex tree is useless for purposes such as option pricing: after only 20 periods, the process could take more that a million different values, and after 40 periods more than a trillion values. A computationally simple binomial representation would allow the process to take at most 21 and 41 values after 20 and 40 periods respectively.

The definitions in (21)–(23) lead to a computationally complex tree because the step-sizes are proportional to the state-dependent conditional standard deviation $\sigma(y,t)$. Note, however, that if $\sigma(y,t)$ is constant, as it was in the approximation developed above for the Ornstein-Uhlenbeck process, then the displacements are equal — so computational simplicity is retained. This suggests that a transformation that purges the original stochastic differential equation (1) of conditional heteroskedasticity will permit us to construct a computationally simple tree.

2.2 Retaining Computational Simplicity: The basic intuition

To this end, consider a transform $X(y,t)$ which differentiable twice in y and once in t. We have by Itô's Lemma

$$\begin{aligned} dX(y_t,t) &= \left(\mu(y_t,t)\frac{\partial X(y_t,t)}{\partial y} + \frac{1}{2}\sigma^2(y_t,t)\frac{\partial^2 X(y_t,t)}{\partial y^2} + \frac{\partial X(y_t,t)}{\partial t}\right) dt \\ &\quad + \left(\sigma(y_t,t)\frac{\partial X(y_t,t)}{\partial y}\right) dW_t. \end{aligned} \tag{24}$$

Now choose $X(y,t)$ to satisfy

$$X(y,t) = \int^{y} \frac{dZ}{\sigma(Z,t)} \tag{25}$$

on the support of y. Then the term $\frac{\partial X(y,t)}{\partial y}\sigma(y,t)\,dW_t$ in (24) becomes dW_t, and the instantaneous volatility of the transformed process $x_t \equiv X(y_t,t)$ is constant. In this case, we can develop a computationally simple binomial tree for x where the second moment of the local change in x is constant at every node. To arrive at the sequence of binomial processes on y, we transform from x back to y by defining

$$Y(x,t) \equiv \{y : X(y,t) = x\}. \tag{26}$$

It is easy to see that $\frac{\partial Y}{\partial x} = \sigma(y,t)$, and by Assumption 1, this means that $Y(x,t)$ is weakly monotone in x for a fixed t. Then we can use the transform in (26) to define a tree for y which takes the form shown in Figure 2, so that

$$Y_h^+(x,t) = Y(x+\sqrt{h}, t+h) \qquad \text{and} \tag{27}$$

$$Y_h^-(x,t) = Y(x-\sqrt{h}, t+h). \tag{28}$$

Note that the tree for y has inherited the computational simplicity that the tree for x displays. Using the fact that $\frac{\partial Y(x,t)}{\partial x} = \sigma(Y(x,t),t)$, a Taylor's series expansion of Y_h^+ and Y_h^- around $h=0$ yields

$$Y_h^{\pm}(x,t) = Y(x,t) \pm \sigma(Y(x,t),t)\sqrt{h} + O(h), \quad \text{and} \tag{29}$$

$$\sigma_h^2(Y(x,t),t) = \sigma^2(Y(x,t),t) + O(\sqrt{h}). \tag{30}$$

This shows that the local second moment of ${}_h y_t$ converges to the instantaneous variance $\sigma^2(y,t)$ as $h \downarrow 0$. Finally, to get the local drift to match the drift of the limit diffusion, we need

$$\mu_h(y,t) \to \mu(y,t) \tag{31}$$

uniformly on $\{(y,t) : |y|, t < \delta\}$ for every $\delta > 0$. We *tentatively* choose

$$q_h \equiv \frac{h\mu(Y(x,t),t) + Y(x,t) - Y_h^-(x,t)}{Y_h^+(x,t) - Y_h^-(x,t)}, \tag{32}$$

which, if it is a legitimate probability (i.e., between 0 and 1) sets the local drift *exactly* equal to the drift of the limiting diffusion (1). This device — the use of a transform, its inverse and the choice of the probability q_h — enables one to construct a computationally simple binomial approximation. It turns out to be a useful device in many commonly employed diffusions in finance, where a transformation like (25) is readily available. A straightforward example of this transformation is for the lognormal diffusion, where $\mu(y,t) = \mu\, y$ and $\sigma(y,t) = \sigma\, y$. The transformation is simply $X(y) = \sigma^{-1}\log y$, and the inverse transformation is $Y(x) = e^{\sigma x}$. This was the transformation employed by Cox, Ross, and Rubinstein (1979) to obtain a computationally simple tree. Such transformations can be made for other diffusions, even if their drift and diffusion functions depend on t.

Our specification of Y, Y_h^+, Y_h^-, and q_h has been tentative, since these functions often have to be modified in individual cases. For example, since q_h is a probability, it must lie between zero and one, whereas the value implied by (32) may not. We must sometimes also allow x to jump up or down by a quantity greater than $\sqrt{h}$ in order to maintain the drift rate. Furthermore, the diffusion may have a boundary at 0 (or some other value). At such a boundary $\sigma(\cdot,t) = 0$ and the transformation (25) may need to be modified. The next task is to formally state sufficient conditions for a sequence of *computationally simple* binomial processes to satisfy the conditions of Theorem 1. This is the focus of the next section: to implement the transformation just outlined in a general way.[8]

2.3 Retaining Computational Simplicity: A general treatment

The principal complications that arise in implementing our strategy come from singularities in $\sigma(\cdot,\cdot)$; for example, $\sigma(y^*,t) = 0$ for some y^*. Such singularities are usually associated with boundaries on the support of the process, and often arise in financial economics; for example, with limited liability and in the absence of arbitrage zero must be a lower boundary for stock prices and nominal interest rates.

[8]Readers less interested in the technical development of the approximations may wish to skip to Section 2.4, which presents simple examples of the technique.

There is a large variety of possible boundary behaviors (see Karlin and Taylor (1981)), so it is necessary to confine our attention to the cases likely to be most useful in finance. First, we consider the case in which $\sigma(\cdot,\cdot)$ has no singularities on $\Re^1 \times [0,\infty)$. (This is the case, for example, for the Ornstein-Uhlenbeck diffusion considered earlier.) Then, we consider the case in which $\sigma(0,t)=0$ and $\mu(0,t)\geq 0$ for all t, implying a lower boundary at zero on the support of the limiting diffusion.

CASE 1. No Singularities in $\sigma(y,t)$

As in Section 2.2, we define the $X(y,t)$ function, along with x values corresponding to extreme values of y

$$X(y,t) \equiv \int^y dZ/\sigma(Z,t), \tag{33}$$

$$x^U(t) \equiv \lim_{y\to\infty} X(y,t), \tag{34}$$

$$x^L(t) \equiv \lim_{y\to-\infty} X(y,t). \tag{35}$$

The following assumption is convenient, and can be relaxed at the expense of simplicity.

Assumption 5 *$x^U(t)$ and $x^L(t)$ are constants.*

The definition of the inverse transform in (26) is now modified to read

$$Y(x,t) \equiv \begin{cases} y : X(y,t)=x & \text{if} \quad x^L < x < x^U \\ \infty & \text{if} \quad x^U \leq x \\ -\infty & \text{if} \quad x \leq x^L \end{cases} \tag{36}$$

We retain the definitions of $Y_h^{\pm}(x,t)$ and $q_h(x,t)$ given in (27), (28) and (32) respectively, except that we censor the $q_h(x,t)$:

$$q_h^*(x,t) = \max\{0, \min[1, q_h(x,t)]\} \tag{37}$$

This specifies the sequence of binomial approximations for this case.[9]

Our strategy is as follows: we will apply Theorem 1, so we must verify its four assumptions. Recall that the first two conditions relate to the stochastic differential equation which serves as the limit, and the last two relate to the sequence of binomial approximations (which now must involve the transformation introduced to buy computational simplicity). To verify Assumptions 1 and 2 for the current Case, we employ Assumptions 6 and 7.

[9]Note that if h is very large, it is possible that the steps are such that *both* Y_h^+ and Y_h^- are infinite. We assume that we can choose h small enough to avoid this, so that q_h is well-defined.

Assumption 6 *$\mu(y,t)$ and $\sigma^2(y,t)$ are continuous everywhere. For every $R > 0$ and every $T > 0$, there is a number $\Lambda_{T,R} > 0$ such that*

$$0 \leq \inf_{0 \leq t \leq T, |y| \leq R} \sigma(y,t) - \Lambda_{T,R}. \tag{38}$$

(38) is a non-singularity assumption — it ensures that $\sigma(y,t)$ is bounded away from zero except at $t = \infty$ and/or $|y| = \infty$. Note that in this case $Y(x,t)$ is a strictly monotone increasing function of x for fixed t.

We must also ensure that the process for y does not explode to infinity in finite time. Stroock and Srinivasa Varadhan (1979, Theorems 10.2.1 and 10.2.3) provide two sufficient conditions for non-explosion. One of these, a Lyapunov condition, is given in the Appendix. For now we explicitly rule out this behavior:

Assumption 7 *All solutions of (1) share the property that for all T, $0 < T < \infty$,*

$$\lim_{B \to \infty} P\left(\sup_{0 \leq t \leq T} |y_t| > B \right) = 0. \tag{39}$$

To verify that Assumptions 3 and 4 hold, expand $Y_h^+(x,t) - Y(x,t)$ and $Y_h^-(x,t) - Y(x,t)$ as functions of $\sqrt{h}$ in a Taylor's series around $\sqrt{h} = 0$:

$$Y_h^{\pm}(X(y,t),t) - y = \pm\sqrt{h}\sigma(y,t) + O(h).$$

As in Section 2.2, this gets the local variance right and the stepsizes small as $h \downarrow 0$. Since $\sigma(y,t)$ is bounded away from zero on bounded sets (by Assumption 5), it is unnecessary to truncate q_h on bounded sets (y,t) when h is small — so the drift matches as well. This is the line of the argument in the proof. In order for the Taylor's series argument to go through, however, we need regularity conditions on the diffusion function $\sigma^2(y,t)$ and the transformation $Y(x,t)$. This is the basis for the next assumption.

Assumption 8 *The first and second order partial derivatives*[10]

$$\sigma_y, \sigma_t, \sigma_{yt}, \sigma_{tt}, Y_x, Y_{xx}, Y_t, Y_{tt}, Y_{xt}$$

are well-defined and locally bounded[11] *for all $(y,t) \in \Re^1 \times [0,\infty)$.*

The theorem for Case 1 can now be stated.

[10]The definitions of these partial derivatives are collected in the Appendix.
[11]By "locally bounded" we mean bounded on bounded (y,t) sets.

Theorem 2 *Let Assumptions 5–8 hold. For $h > 0$, define the x–tree as in Figure 1, with ${}_h x_0 \equiv X(y_0, 0)$, and the transitions for the x process given by*

$$ {}_h x_{h(k+1)} = \begin{cases} {}_h x_{hk} + \sqrt{h} & \text{with probability} \quad q_h^*({}_h x_{hk}, hk) \\ {}_h x_{hk} - \sqrt{h} & \text{with probability} \quad 1 - q_h^*({}_h x_h k, hk) \end{cases} \tag{40} $$

Define the y–tree as in Figure 2. That is, for $h > 0$, define ${}_h y_t \equiv Y({}_h x_{hk}, hk)$ for $hk \leq t < h(k+1)$. By construction, $\{{}_h y_t\}$ is computationally simple. Then $\{{}_h y_t\} \Rightarrow \{y_t\}$ as $h \downarrow 0$, where $\{y_t\}$ is the solution to (1). □

CASE 2. A Singularity at y=0: $\sigma(0,t) = 0,\ \mu(0,t) \geq 0$

In this case the diffusion coefficient vanishes at a lower boundary (zero), but the drift rate might serve to "return" the process above it. This would be a reasonable specification for a process on the price of an asset or on the nominal interest rate. To handle this case, we must modify some of the definitions and assumptions given earlier. The lower limit for x is redefined as

$$ x^L(t) \equiv \lim_{y \downarrow 0} X(y,t), \tag{41} $$

and the inverse transform (which is now a weakly monotone function of x) defined in (26) as

$$ Y(x,t) \equiv \begin{cases} y : X(y,t) = x & \text{if} \quad x^L < x < x^U \\ \infty & \text{if} \quad x^U \leq x \\ 0 & \text{if} \quad x \leq x^L \end{cases} \tag{42} $$

As before we assume that x^L and x^U don't depend on t.

An important aspect of Case 2 relates to the stepsizes: thus far they are (approximately) proportional to $\sigma(y,t)$. But if $\sigma(y,t)$ is very small near $y = 0$ and $\mu(y,t)$ is not small, we may need to take *multiple* jumps in this region in order to match the drift of the limit diffusion. Choose $x^B > x^L$, and define the function $J_h^+(x,t)$ as

$$ J_h^+(x,t) \equiv \begin{cases} \text{the smallest, odd, positive, integer } j \text{ such that} & \\ \quad Y(x + j\sqrt{h}, t+h) - Y(x,t) \geq \mu(Y(x,t),t) \cdot h, & \text{if} \quad x < x^B \\ \quad 1 & \text{if} \quad x \geq x^B. \end{cases} \tag{43} $$

$J_h^+(x,t)$ is the minimum number of *upward* jumps that keeps the jump probability q_h less than one without censoring; and it is *odd* so that the jump moves the process to an existing node on the tree. By permitting these multiple jumps in a restricted region near 0, we retain computational simplicity; at large values of y (corresponding to $x > x^B$) we disallow multiple upward jumps, because if J_h^+ is unbounded it might increase the number of nodes at a rate rapid enough to affect computational simplicity. Similarly, define $J_h^-(x,t)$ by

$$ J_h^-(x,t) \equiv \begin{cases} \text{the smallest, odd, positive, integer } j \text{ such that} \\ \textit{either} \quad (a) \quad Y(x,t) - Y(x - j\sqrt{h}, t+h) \leq \mu(Y(x,t),t) \cdot h \\ \textit{or} \quad (b) \quad Y(x - j\sqrt{h}, t+h) = 0. \end{cases} \tag{44} $$

$J_h^-(x,t)$ is the minimum number[12] of *downward* jumps that either keeps the probability q_h positive (without censoring) or forces the down-state value for Y_h^- to zero. The transitions in the value for y are then restated as

$$Y_h^{\pm}(x,t) \equiv Y(x \pm J_h^{\pm}(x,t) \cdot \sqrt{h},\, t+h), \tag{45}$$

and we retain the definition of q_h^* given in (32) and (37).

Note that Assumption 6 is incompatible with $\sigma(0,t) = 0$ and a replacement must be found to guarantee that Assumptions 1 and 2 are satisfied. The following Lipschitz condition, combined with Assumption 7, guarantees that Assumptions 1 and 2 are satisfied:

Assumption 9 *Let $\sigma(y,t)$ and $\mu(y,t)$ be continuous on $\Re^1 \times [0,\infty)$. There exists an increasing, non-negative function $\rho(u)$ from $[0,\infty)$ into $[0,\infty)$ such that*

$$\rho(u) > 0 \quad \textit{for} \quad u > 0, \quad \textit{and} \tag{46}$$

$$\lim_{\epsilon \downarrow 0} \int_{\epsilon}^{1} [\rho(u)]^{-2}\, du = \infty. \tag{47}$$

Further, for every $R > 0$ and $T > 0$, there exists a number $\Lambda_{T,R} > 0$ such that

$$\sup_{|y^*| \le R, |y| \le R, 0 \le t \le T} |\sigma(y^*,t) - \sigma(y,t)| - \Lambda_{T,R}\, \rho(|y - y^*|) \le 0, \qquad \textit{and} \tag{48}$$

$$\sup_{|y^*| \le R, |y| \le R, 0 \le t \le T} |\mu(y^*,t) - \mu(y,t)| - \Lambda_{T,R}\, |y - y^*| \le 0.^{13} \tag{49}$$

To carry out the Taylor's series argument and to handle the singularity at $y = 0$, we alter Assumption 8 as follows:

Assumption 10 *On every compact subset of $\{(y,t) : 0 < y < \infty, 0 \le t < \infty\}$, $\sigma_y, \sigma_t, \sigma_{yt}, \sigma_{tt}$ exist and are bounded, and $\sigma(y,t)$ is bounded and bounded away from zero. There exists a $\Delta > 0$ such that for every $T > 0$,*

$$\inf_{0 \le t \le T, 0 \le y \le \Delta} \sigma_y(y,t) > 0.$$

Furthermore, $Y_{xx}, Y_t, Y_{tt}, Y_{xt}$ exist for all $(y,t) \in [0,\infty) \times [0,\infty)$ and are bounded on bounded sets. For all $t \ge 0, \sigma(0,t) = 0$ and $\mu(0,t) \ge 0$.

---

[12] Using Assumption 9 it is easy to show, given x, t, h, that J_h^+ and J_h^- exist and are finite.

[13] It is easy to show that the square root diffusion

$$dr = \kappa(\mu - r)\, dt + \sigma\sqrt{r}\, dW$$

discussed in Section 2.4 satisfies Assumption 9, using $\rho(z) = \sqrt{z}$. On the other hand, the "double square root" process in Longstaff (1989)

$$dr = \kappa(\mu - \sqrt{r})\, dt + \sigma\sqrt{r}\, dW$$

satisfies (48) (again using $\rho(z) = \sqrt{z}$) but does not satisfy (49).

Assumption 10 weakens Assumption 8 by allowing $\sigma_y(0,t), \sigma_t(0,t), \sigma_{yt}(0,t)$, and $\sigma_{tt}(0,t)$ to be infinite. We also impose the restriction that σ_y be positive in some neighborhood of $y = 0$. Note, however, that $Y_{xx}(0,t), Y_t(0,t), Y_{tt}(0,t), Y_{xt}(0,t)$ must still be finite.

The theorem for Case 2 can now be stated.

Theorem 3 *Let Assumptions 5, 7, 9, and 10 hold, and assume $y_0 > 0$. Define ${}_h x_{hk}$ and ${}_h y_t$ as in Theorem 2, replacing relations (35) and (36) with (41) and (42) respectively, and using (45) to define $Y_h^{\pm}$. Then $\{{}_h y_t\} \Rightarrow \{y_t\}$ as $h \downarrow 0$; and if $x^B < \infty$, $\{{}_h y_t\}$ is computationally simple by construction. Further, 0 bounds the support of ${}_h y_t$ and y_t from below:*

$$P\left(\inf_{0 \le t < \infty} {}_h y_t < 0\right) = P\left(\inf_{0 \le t < \infty} y_t < 0\right) = 0. \qquad \square \tag{50}$$

Between them, Theorems 2 and 3 show how to construct computationally simple approximations for diffusions encountered in many applications in finance. The obvious extension of the results of Theorems 2 and 3 is to cases where an *upper* boundary also applies: for example, in modelling the price of a discount bond. An upper boundary where the drift μ is non-positive and the standard deviation σ is zero can be handled by modifying the arguments in Case 2 of Section 2.3. These modifications are straightforward, and they will generally require the use of multiple *downward* jumps near the upper boundary.

2.4 Examples

The CEV Stock Price Process

In this model the stock price is assumed to follow

$$ds_t = \mu s_t \, dt + \sigma s_t^{\gamma} \, dW_t, \qquad 0 < \gamma \le 1, \tag{51}$$

where s_0 is positive. Here $\sigma(0) = 0$; and as long as $\gamma \ge \frac{1}{2}$ (which we assume hereafter) the process is trapped at zero once it gets there, and the regularity conditions of Theorem 3 can be shown to hold. Our x–transform is given by

$$X(s) \equiv \sigma^{-1} \int^{s} Z^{-\gamma} \, dZ = \frac{s^{1-\gamma}}{\sigma(1-\gamma)}. \tag{52}$$

We define $x_0 = X(s_0)$, and draw out our x-tree as in Figure 1. The inverse transform is given by[14]

$$S(x) \equiv \begin{cases} [\sigma(1-\gamma)x]^{\frac{1}{1-\gamma}} & \text{if } \; x > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{53}$$

[14]Bates (1988) independently developed a similar transformation in the context of pricing American options on futures contracts.

which corresponds to Figure 2, with S replacing Y. We employ the definitions for the multiple jumps $J_h^{\pm}$ given in (43) and (44), replacing Y with S; and we define the functions

$$S_h^+(x) \equiv S(x + J_h^+\sqrt{h}), \quad \text{and} \tag{54}$$

$$S_h^-(x) \equiv S(x - J_h^-\sqrt{h}). \tag{55}$$

It remains to specify the probability q_h. For $x > 0$ set

$$q_h^* \equiv \frac{h\mu S(x) + S(x) - S_h^-(x)}{S_h^+(x) - S_h^-(x)}. \tag{56}$$

Then define q_h by

$$q_h \equiv \begin{cases} q_h^* & \text{if} \quad x > 0 \quad \textit{and} \quad 0 < q_h^* < 1 \\ 0 & \text{if} \quad \textit{either} \quad x \leq 0 \quad \textit{or} \quad q_h^* \leq 0 \\ 1 & \text{if} \quad q_h^* > 1. \end{cases} \tag{57}$$

These definitions ensure that q_h is a legitimate probability and that if ${}_h s_t$ reaches 0 it stays trapped there. We now apply Theorem 3 to the sequence of binomial processes for s.

Corollary 3.1 Define the sequence of $\{{}_h s_t\}$ processes by (52) – (57) and (3) to (7), replacing ${}_h y_t$ with ${}_h s_t$. As $h \downarrow 0$, $\{{}_h s_t\} \Rightarrow \{s_t\}$, the solution of (51). □

The CIR diffusion on the short rate

Consider now the autoregressive "square root" interest rate process used by Cox, Ingersoll, and Ross (1979)[15]:

$$dr = \kappa(\mu - r)\,dt + \sigma\sqrt{r}\,dW, \tag{58}$$

with $\kappa \geq 0, \mu \geq 0$, and the initial value of $r = r_0$, a non-negative constant. The necessary transformation is

$$X(r) \equiv \int^r \frac{dZ}{\sigma\sqrt{Z}} = \frac{2\sqrt{r}}{\sigma} \tag{59}$$

with $x_0 = X(r_0)$. Zero is a lower boundary for r. As outlined in Section 2.3, we define the inverse transform

$$R(x) \equiv \begin{cases} \frac{\sigma^2 x^2}{4} & \text{if} \quad x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{60}$$

Because the drift in (58) does not vanish as $r \downarrow 0$, 0 is not an absorbing state for r unless either κ or μ equals zero. This illustrates why it was necessary to introduce multiple jumps in Section 2.3. Suppose that we are at node c in Figure 3. At this node, $x < 0$, so $R(x) = 0$. The usual upward jump of $\sqrt{h}$ would take us to node k, at which $R(x)$ still equals zero. Clearly, if there is a positive drift in the process at $r = 0$ (which is true if κ and μ are

[15] Ball (1989) develops a different binomial approximation for this diffusion; he exploits knowledge of properties of the conditional distribution of the interest rate.

$$dr = \kappa(\mu - r)\,dt + \sigma r\,dW. \tag{65}$$

There is no known closed form for the conditional distribution of r_t for this process, but the binomial approximation would allow us to price contingent claims for which there is no known pricing formula.

[16]Note that at large values of r (and hence of x) the drift is negative, and we avoid multiple upward jumps in that region.

3 Applications of the Binomial Method to Valuation Models

In this Section, we examine models for option values and for default-free bonds, employing the binomial model described in Section 2 for the relevant diffusions. Unfortunately, the theorems in Section 2 speak strictly to the weak convergence of the sequence of the binomial models to the underlying diffusion, and do not directly apply to the convergence of values of options and bonds[17]. In this Section, however, we adapt Theorems 1 and 3 to prove convergence of binomial bond and European option prices to their diffusion limits. In both applications below (which follow the diffusions studied in Section 2.4), we provide numerical evaluations of the binomial method.

3.1 Option Pricing

The CEV diffusion defined in (51) is an example where a computationally simple binomial tree can be constructed and employed in option valuation. Furthermore, since a formula for the value of European call options on a non-dividend paying stock is available in this case, (Cox and Rubinstein (1985) p. 363) the results can be readily verified.

Let the stock price be s, and let $r \geq 0$ be the constant, continuously compounded riskless interest rate. The valuation procedure for the European call option, following the arguments in Black and Scholes (1973) and Merton (1973) requires that the call value $C(s, T-t)$ satisfies the partial differential equation (PDE)

$$\frac{1}{2}\sigma^2 s^{2\gamma} C_{ss} + rsC_s - rC = -C_t \tag{66}$$

subject to standard terminal and boundary conditions. The binomial method leads to the requirement that at every node, the call values satisfy the one period valuation formula leading to relation (67) is well known – it requires the construction of a non-anticipating, self-financing portfolio of the risky asset and a riskless asset which delivers the option's payoff at maturity (see Cox and Rubinstein (1985)).

Before passing to the numerical solutions, we present the argument justifying the binomial method for European option pricing: Theorem 4 below shows that the sequence of solutions to the difference equation (67) converges to the solution of the PDE, subject to the appropriate boundary conditions. Unfortunately, we have not been able to extend this theorem to cover American options rigorously. When we permit premature exercise and hence a free boundary, the binomial procedure has performed well in experiments, but there is no guarantee that it always will.

It is well known that the value of the drift rate $\mu(s,t)$ does not affect the option value: within the binomial representation of the stock price process, $\mu(s,t)$ affects the probability of an up-move, but this probability does not enter the valuation procedure for the call option. The valuation procedure depends on the pseudo-probabilities (see Cox and Rubinstein (1985), and Harrison and Kreps (1979) for the connexion to the equivalent martingale measure). As this argument shows, the local mean and second moment of the binomial representation of S must pass to the drift and variance rate for the risk-neutralized diffusion.

If the payoff on the contingent claim depends only on the final stock price, which is true for European options, Theorems 1 or 3 can be used to price the claim. Here we start with a stock price process of the form

$$ds_t = \mu(s_t,t)\,dt + \sigma(s_t,t)\,dW_t \quad \text{fixed} \quad s_0, \tag{68}$$

and its risk-neutral counterpart:

$$d\mathcal{S}_t = r \cdot \mathcal{S}_t\,dt + \sigma(\mathcal{S}_t,t)\,dW_t,\ \mathcal{S}_0 = s_0. \tag{69}$$

Suppose (69) satisfies[18] the assumptions of Theorem 3. We then create an approximation ${}_h\mathcal{S}_t$ to the process for $\mathcal{S}_t$. To accomplish this, define the tree for $\{{}_h\mathcal{S}_t\}$ as in Theorem 3 (replacing y with $\mathcal{S}$ where necessary). In order to preclude arbitrage between the stock and the riskless asset in each economy (indexed by h), we now permit upward jumps everywhere so that $x^B = \infty$ in (43): by doing so we avoid the undesirable feature of having to truncate

[18]The process (68) itself should not permit arbitrage, and the use of the equivalent martingale measure relies on this; see Harrison and Kreps (1979). One implication of the no-arbitrage requirement is that $\mu(0,t) = 0$.

the probabilities in each economy. Define[19]

$$J_h^+(x,t) \equiv \begin{cases} \text{the smallest, odd, positive, integer } j \text{ such that} \\ \quad \mathcal{S}(x + j\sqrt{h}, t+h) - exp(rh)\cdot \mathcal{S}(x,t) \geq 0, \end{cases} \tag{70}$$

$$J_h^-(x,t) \equiv \begin{cases} \text{the smallest, odd, positive, integer } j \text{ such that} \\ \quad \mathcal{S}(x - j\sqrt{h}, t+h) - exp(rh)\cdot \mathcal{S}(x,t) \leq 0 \end{cases} \tag{71}$$

Define also

$$p_h \equiv \begin{cases} \frac{[\mathcal{S}(x,t)e^{rh} - \mathcal{S}_h^-(x,t)]}{[\mathcal{S}_h^+(x,t) - \mathcal{S}_h^-(x,t)]} & if \quad \mathcal{S}_h^+ > 0 \\ 0 & otherwise, \end{cases} \tag{72}$$

where p_h is the risk-neutral probability implied by the arbitrage argument of Cox and Rubinstein (1985). To rule out arbitrage $0 \leq p_h \leq 1$ globally for the process for ${}_h\mathcal{S}_t$, and this is guaranteed by (70)-(71). We define p_h to be the probability of an upward jump for the $\{{}_h\mathcal{S}_t\}$ process.

Using the arguments in Cox, Ross, and Rubinstein (1979) that the absence of arbitrage implies that the prices at time 0 of European Put and Call options on ${}_h\mathcal{S}_t$ with expiration at date T (which is an integer multiple of h) and striking price $K \geq 0$ are given by

$$e^{-rT}\cdot \mathcal{E}_{0,h}[(K - {}_h\mathcal{S}_T)^+] \quad \text{and} \quad e^{-rT}\cdot \mathcal{E}_{0,h}[({}_h\mathcal{S}_T - K)^+]$$

respectively; here $\mathcal{E}_{0,h}$ is the risk-neutralised, time 0 expectations operator. Theorem 3 allows us to conclude that $\{{}_h\mathcal{S}_T\} \Rightarrow \{\mathcal{S}_T\}$. Finally, since the terminal payoff for the put is uniformly bounded in h (i.e., the put price is always less than the exercise price K), Theorem 25.12 in Billingsley (1986) allows us to conclude that $e^{-rT}\cdot \mathcal{E}_{0,h}[(K - {}_h\mathcal{S}_T)^+] \to e^{-rT}\cdot \mathcal{E}_0[(K - \mathcal{S}_T)^+]$ as $h \downarrow 0$.[20] This is the basis for the following

Theorem 4 [21] *Let the process (69) satisfy the conditions of Theorem 3. Define the ${}_h\mathcal{S}_t$ as indicated above. Then $\{{}_h\mathcal{S}_t\} \Rightarrow \{\mathcal{S}_t\}$, the put value*

$$e^{-rT}\cdot \mathcal{E}_{0,h}[(K - {}_h\mathcal{S}_T)^+] \to e^{-rT}\cdot \mathcal{E}_0[(K - \mathcal{S}_T)^+]$$

and the call value

$$e^{-rT}\cdot \mathcal{E}_{0,h}[({}_h\mathcal{S}_T - K)^+] \to e^{-rT}\cdot \mathcal{E}_0[(\mathcal{S}_T - K)^+]. \quad \square$$

[19] Note that the jumps defined in $J_h^{\pm}$ are consistent with a no-arbitrage condition in each of the sequence of economies indexed by h. Note also that a binomial approximation for s_t can be designed using the arguments in Sections 2.2 and 2.3, but this is unnecessary for our purposes here.

[20] This result extends to European calls as well, since European put-call parity allows us to conclude that $e^{-rT}\cdot \mathcal{E}_{0,h}[({}_h\mathcal{S}_T - K)^+] \to e^{-rT}\cdot \mathcal{E}_0[(\mathcal{S}_T - K)^+]$.

[21] This theorem is related to recent results of He (1989), who considers convergence of prices of a contingent claim with a terminal payoff function $g(s_T)$ satisfying Lipschitz conditions. His results also apply to the multivariate case. On the other hand, he imposes severe restrictions on the stock price process, excluding, for example, the CEV stock price process. Boyle, Evnine and Gibbs (1989) develop a discrete distribution to approximate the multivariate lognormal diffusion and apply it to contingent claims valuation.

Duffie and Protter (1988) pose a related question: suppose the stock price process (for any given h) is $\{{}_hs_t\}$, and it converges weakly to $\{s_t\}$ as $h \downarrow 0$ for some limit process $\{s_t\}$. To price an option on $\{{}_hs_t\}$, suppose we set up the hedge portfolio incorrectly — we use the hedging rule that would be correct if the underlying stock price process were $\{s_t\}$. Duffie and Protter show that under certain conditions, the risk introduced by using the incorrect hedging rule vanishes as $h \downarrow 0$. This is a reassuring result, since any model's description of stock price movements is at best approximately correct. Our Theorem 4 is concerned with *exact* arbitrage pricing[22] for a sequence of stock price processes $\{{}_hs_t\}$ and its limit process $\{s_t\}$.

With American options, we cannot rely on Theorem 4, and are forced to indicate how the discrete valuation equation (67) converges to the PDE in which now premature exercise might be optimal. In order to show how the discrete valuation is related to the PDE, expand the call's value[23] $C(S_h^+, T-t-h)$ in a Taylor's series around $(s, T-t)$, retaining terms of order h or greater:

$$C(S_h^+, T-t-h) = C(s, T-t) + (S_h^+ - s)C_s\Big|_{s,T-t} + \frac{1}{2}(S_h^+ - s)^2 C_{ss}\Big|_{s,T-t} + hC_t|_{s,T-t} + o(h), \tag{73}$$

and similarly for $C(S_h^-, T-t-h)$. Substitute these expressions into (67), divide through by h, and take h to zero; we pass to the PDE. This connexion between the two valuation equations follows the argument given by Cox and Rubinstein (1985, pp. 208-9). It allows us to interpret the binomial model as a numerical method for the solution of the PDE. This argument is not rigorous, but it suggests the usefulness of the binomial method in the valuation of American options. We can calibrate the approximation by comparing the American option values to the values obtained from an alternate numerical procedure, such as the method of finite differences.

To check the numerical accuracy of the binomial method for the CEV process, we chose the following parameters: (i) an annual rate of interest of 5%, (ii) values for γ of 0.5 (the square root diffusion) and 0.875 (the average of the values reported by Gibbons and Jacklin (1989)), (iii) three values of σ, chosen such that the initial, annualized instantaneous standard deviations correspond to 0.2, 0.3, and 0.4. We fix the initial stock price at \$40, and for each

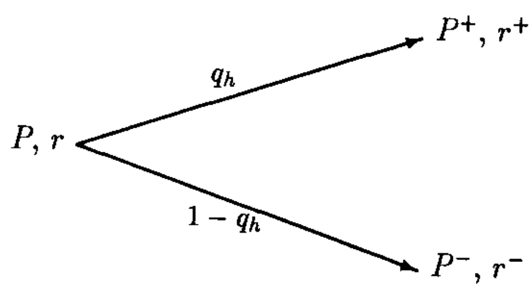
[22] We have required exact arbitrage pricing in each economy (indexed by h) in the definitions in (70)-(72), in the spirit of the development in Cox, Ross and Rubinstein (1979). These definitions do not ensure computational simplicity in every case; however, in the CEV application given below simplicity is achieved for conventional parameter values. Of course, a simple binomial approximation to (68) can be readily found from the methods in Section 2.

[23] Because we're assuming a nondividend paying stock, this value also applies to American call options. For American puts, however, the one period valuation formula in (67) must be replaced by the immediate exercise value if the latter dominates — and hence an optimal exercise policy found as part of the problem, thus constituting a free boundary.

combination of parameter values, we compute the option values at striking prices of \$35, \$40 and \$45. Table 1 displays the results for European Calls and American Puts. Formula values for European options under the square root diffusion are available in Cox and Rubinstein, p. 364. For comparison, we computed the values of European call options for $\gamma = 0.875$ and for all the American puts numerically, using the implicit finite difference method to solve the PDE. The binomial method gives answers accurate to within \$0.01–\$0.02 for the chosen maturities of 1 and 4 months, as long as 50 time intervals are used. The approximation deteriorates as the maturity is lengthened, and the binomial method gives coarse answers for 5 time steps at these parameter values. [24]

3.2 Bond Pricing

The diffusion in (58) proposed by Cox, Ingersoll and Ross (1979) is one of several models for the nominal short term interest rate which can be employed to value a stream of default-free cash flows. The binomial valuation method imposes that the value of this stream at any stage is equal to the expected future value (at the two subsequent nodes) discounted at the risk-adjusted rate. In general, the one-step binomial tree can be represented as



where P is the value of the claim, and r is nominal short term rate of interest, and the suffixes + and − apply to these quantities at the subsequent time node, after an up and down move respectively.[25] The valuation method states that

$$P = \frac{[q_h P^+] + [(1 - q_h)P^-]}{(1 + r^* h)}, \tag{74}$$

[24]The binomial routine, with the transformation defined as in (53) – (55) and the jumps defined in (70) – (71), was implemented in GAUSS on a personal computer. Valuations of at-the-money American Puts (with 4 months to expiration, $\sigma = 0.4$ and $\gamma = 0.875$) with the binomial model required 0.22, 4.01 and 66.57 seconds for values of n at 5, 50 and 250 respectively. Accuracy comparable (within 0.1 cents) to the valuation with n at 50 was obtained by a solution to the PDE for the American Put using the implicit method of finite differences, reported in Table 1, in 93.6 seconds. These figures for accuracy and execution time should not be taken as representative at all parameter values. The GAUSS code for the binomial method used in the tables is available from the authors on request.

[25]Note that since P moves inversely with r, $P^+ < P < P^-$.

where r^* is the risk-adjusted discount rate, $r^* \equiv r + r \cdot g(t, T-t)$, and $r \cdot g(t, T-t)$ represents the instantaneous risk premium. We assume that it is a bounded, continuous function of the time index t and the time to maturity $T-t$, satisfying $g(t, T-t) \geq -1$, thus ensuring that r^* is non-negative. If $g(t, T-t) \equiv 0$, then the Local Expectations Hypothesis applies.[26]

The value P at any node must be augmented by any cash distribution that might occur at that node; in the numerical example below we value a discount instrument, and therefore there are no cash distributions. Again, one can rearrange the one period valuation equation, expand P^+ and P^- in a Taylor's series around $P(r, T-t)$, and pass to the PDE, which is the valuation equation for this asset.

As was the case for the CEV European option pricing model, we can use a version[27] of Theorem 3 to show that for a discount bond, the sequence of bond prices produced by the binomial model converges to the bond price produced by the diffusion limit. This would then justify the use of the binomial method in this context.

Consider the value at time 0 of a pure discount bond that pays \$1 at time T. The binomial pricing procedure implies (given $h \equiv T/n$)

$$_hP_0 = 1 \cdot E_{0,h}\left[\exp\left\{-h\sum_{j=0}^{n-1}\left({}_hr_{jh} + {}_hr_{jh} \cdot g(jh, (n-j)h)\right)\right\}\right]. \tag{75}$$

In continuous time, we have (see Cox, Ingersoll and Ross (1981))

$$P_0 = 1 \cdot E_0\left[\exp\left\{-\int_0^T \left(r_t + r_t \cdot g(t, T-t)\right)\, dt\right\}\right]. \tag{76}$$

To show that $_hP_0 \to P_0$ as $h \downarrow 0$, we first define the stochastic process $_hy_t$, by

$$_hy_0 \equiv 1, \tag{77}$$

$$_hy_t = {}_hy_{kh}, \quad kh \leq t < (k+1)h, \quad \text{and} \tag{78}$$

$$_hy_{h(k+1)} = {}_hy_{hk} \cdot \exp\left[-h \cdot \left({}_hr_{hk} + {}_hr_{hk} \cdot g(kh, (n-k)h)\right)\right]. \tag{79}$$

It is easy to check that

$$_hy_T = \exp\left[-h\sum_{j=0}^{n-1}\left({}_hr_{jh} + {}_hr_{jh} \cdot g(jh, (n-j)h)\right)\right] \tag{80}$$

[26]For a discussion of the risk premiums which are consistent with a no-arbitrage condition, see Ingersoll (1987) Chapter 18.

[27]While Theorem 1 dealt only with univariate processes, more general theorems are readily available — see Stroock and Srinivasa Varadhan (1979, Section 11.2), Ethier and Kurtz (1986, Section 7.4) and Nelson (1989, Theorem 2.1). The significant change is that the local second moment is a matrix, and is required to converge to the instantaneous covariance matrix of the diffusion.

$$ {}_hP_0 = E_{0,h}[{}_hy_T]. \tag{81} $$

${}_hy_T$ is uniformly bounded from above by 1 and below by 0, so if the vector[28] Markov process $\{{}_hy_t, {}_hr_t\}$ converges weakly to a well-behaved diffusion, then ${}_hP_0 = E_{0,h}[{}_hy_T]$ converges to $P_0 = E_0[y_T]$ as $h \downarrow 0$. With this as background, the following Theorem justifies the recursive valuation procedure in the binomial model:

Theorem 5 *Suppose that the interest rate process takes the form*

$$ dr_t = \mu(r_t, t)\, dt + \sigma(r_t, t)\, dW_t \tag{82} $$

where r_0 is a non-negative constant and $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ satisfy the conditions of Theorem 3. Define the process $\{{}_hy_t\}$ as in (77)–(79), and construct the approximating binomial sequence $\{{}_hr_t\}$ as in Theorem 3. Then $\{{}_hy_t, {}_hr_t\} \Rightarrow \{y_t, r_t\}$, where $\{r_t\}$ is the solution to (82) and y_t satisfies

$$ dy_t = -y_t \cdot r_t \cdot \{1 + g(t, T-t)\}\, dt, \quad y_0 = 1. $$

Further, ${}_hP_0 \rightarrow P_0$ as $h \downarrow 0$. □

We examined the numerical accuracy of the binomial model in valuing a discount bond with a face value of \$100 using the CIR interest rate process. The parameter values were set as follows: (i) The value of κ was varied from 0.01 to 0.5. A value of zero produces a martingale; (ii) The value of σ was varied from 0.1 to 0.5; (iii) The long run mean μ was fixed at 8%; and (iv) $g(t, T-t) \equiv 0$, so the Local Expectations Hypothesis holds. These values cover (and go well outside) the range of parameter values reported for nominal Treasury securities by Pearson and Sun (1989). Two initial values of the interest rate, r_0, were chosen: 5% and 11%. The maturities of the instruments chosen were 1, 6, 12, and 60 months. The binomial method was implemented in GAUSS for values of n, the number of steps, ranging from 5 to 200. Cox, Ingersoll and Ross (1985) provide a formula for this bond's value.

Table 2 reports the computed values. The column under "CIR" reports the known solution value for the parameters in that row. The binomial procedure provides accurate solutions, especially for short maturity bills. For given values of n and the bill's maturity, the error increases as σ increases, as is to be expected: in the limiting diffusion process, the distribution of r_t for any t is continuous, and our approximation replaces this with a discrete distribution. For any given h, the higher σ is, the further apart are the values that we let ${}_hr_t$ take, making the approximation more coarse.

[28] *ibid.*

The binomial method can be quickly adapted to compute values for contingent claims on fixed income securities. Because the procedure is relatively flexible, it can be employed for alternative diffusion processes as indicated in Section 2.

4 Conclusion

Sharpe's insight, in the development of the binomial approach, has led to the use of the binomial model in many normative applications in finance, especially in option pricing. Its simplicity and its flexibility are considerable virtues. Unfortunately, the approach has been restricted in its use to those situations in which the underlying asset's price follows a lognormal process in continuous time. This paper presents conditions under which a sequence of binomial processes converges weakly to a diffusion, and shows constructively how one can employ a transformation to produce computationally simple binomial processes. The transformation is relatively straightforward: the construction of the binomial process requires the sizes of the up and down jumps (and the associated probability) to be such that its local drift and second moment converge to the drift and variance of the desired diffusion, and that the jump size goes to zero as the jumps become more frequent. The diffusion's behavior at the boundaries will, in general, require us to modify the transformation and allow multiple jumps in the binomial tree.

In the context of financial models (especially option pricing models), the binomial method numerically solves a partial differential equation for the value of some asset. The methods in this paper permit one to solve such PDEs for alternative underlying diffusions; and these methods might be useful in other contexts as well. For example, we might wish to put a diffusion process on aggregate consumption, and derive bond pricing formulæ by looking at the expected marginal rates of substitution of a representative consumer-investor. The methods in this paper are most useful in such cases, especially when an analytical solution to the problem remains elusive.

APPENDIX

The Formal Setup in Section 2

Let $\mathcal{D}$ be the space of mappings from $[0,\infty)$ into $\Re^1$ that are continuous from the right with finite left limits; $\mathcal{D}$ is a metric space when endowed with the Skorohod metric (Billingsley

(1968)). For each $h > 0$, let M_{kh} be the σ–algebra generated by hk, ${}_hy_0$, ${}_hy_h$, ${}_hy_{2h} \cdots {}_hy_{kh}$, and let $\mathcal{B}$ denote the Borel sets on $\Re^1$. Let $q_h(y, hk), Y_h^+(y, hk)$, and $Y_h^-(y, hk)$ be scalar-valued functions defined on $[0, \infty) \times \Re^1$ satisfying (2), (3) for all $y \in \Re^1$ and all $k = 0, 1, 2, \ldots n$.

Let P_h be the probability measure on $\mathcal{D}$ such that the following hold with probability 1 for $k = 0, 1, \ldots n$:

$$P_h[{}_hy_0 = y_0] = 1, \tag{83}$$
$$P_h[{}_hy_t = {}_hy_{kh}, kh \le t < (k+1)h] = 1, \tag{84}$$
$$P_h[{}_hy_{(k+1)h} = Y_h^+({}_hy_{hk}, hk)|M_{kh}] = q_h({}_hy_{hk}, hk), \tag{85}$$
$$P_h[{}_hy_{(k+1)h} = Y_h^-({}_hy_{hk}, hk)|M_{kh}] = 1 - q_h({}_hy_{hk}, hk), \tag{86}$$
$$P_h[{}_hy_{(k+1)h} = c|M_{kh}] = 0 \quad \text{for} \quad c \ne Y_h^+(hyhk, hk), c \ne Y_h^-(hyhk, hk). \tag{87}$$

Lyapunov Condition Sufficient for Assumption 7

The following is a Theorem in Stroock and Srinivasa Varadhan (1979, Theorem 10.2.1), justifying Assumption 7:

Assume there exists a non-negative function $\phi(y, t)$ that is differentiable with respect to t and twice differentiable with respect to y, such that for each $T > 0$,

$$\lim_{|y|\to\infty} \inf_{0 \le t \le T} \phi(y, t) = \infty, \tag{88}$$

and a positive, locally bounded function $\lambda(T)$ such that for each $T > 0$, for all $y \in \Re^1$ and all $t, 0 \le t \le T$,

$$\mu(y, t)\frac{\partial \phi(y, t)}{\partial y} + \frac{1}{2}\sigma^2(y, t)\frac{\partial^2 \phi(y, t)}{\partial y^2} + \frac{\partial \phi(y, t)}{\partial t} \le \lambda(T)\phi(y, t). \tag{89}$$

Then Assumption 7 holds. □

The Partial Derivatives in Assumptions 8 & 10

The following definitions apply to the partial derivatives (needed in the Taylor's series expansion) and employed in Assumptions 8 and 10: $\sigma_y \equiv \frac{\partial \sigma(y,t)}{\partial y}$, $\sigma_t \equiv \frac{\partial \sigma(y,t)}{\partial t}$, $\sigma_{yt} \equiv \frac{\partial^2 \sigma(y,t)}{\partial y \partial t}$, $\sigma_{yy} \equiv \frac{\partial^2 \sigma(y,t)}{\partial y^2}$, $\sigma_{tt} \equiv \frac{\partial^2 \sigma(y,t)}{\partial t^2}$. And the corresponding partial derivatives for $Y(x, t)$ are readily computed, using the implicit function theorem; we give two of these which are needed below.

$$Y_x \equiv \left.\frac{\partial Y(x, t)}{\partial x}\right|_{x = X(y,t)} = \sigma(y, t) \tag{90}$$

$$Y_{xx} \equiv \left.\frac{\partial^2 Y(x,t)}{\partial x^2}\right|_{x=X(y,t)} = \sigma(y,t)\cdot\sigma_y(y,t) \tag{91}$$

If the functions σ_y, σ_t *etc.* are not well-defined or are infinite at a point (y^*, t), we can often extend these definitions in the obvious way by taking limits. For example, if $\sigma(y^*, t) = 0$ but $\sigma_y(y^*, t) = \infty$, then define $Y_{xx}(y^*, t) \equiv lim_{y \to y^*} Y_{xx}(y, t)$, if this limit exists.

Proofs

PROOF OF THEOREM 1: The proof is a direct application of Nelson (1989), Theorem 2.1, which is itself an adaptation of results in Stroock and Srinivasa Varadhan (1979). Assumptions 1, 3, and 4 are equivalent in the current context to Nelson (1989) Assumption 1. Our Assumption 2 is the equivalent of Nelson (1989) Assumption 4. Nelson (1989) Assumptions 2 and 3 are trivially satisfied. The Theorem then follows directly. *Q.E.D.*

PROOF OF THEOREM 2: Since Assumption 1 is obvious, our first task is to verify Assumption 2. By Stroock and Srinivasa Varadhan Theorems 7.2.1 and 10.2.1, and Ethier and Kurtz (1986) Corollary 5.3.4, Assumptions 6 and 7 together are sufficient to ensure that Assumption 2 is satisfied. Next, expand $Y_h^{\pm}(x,t) - Y(x,h)$ as a function of $z \equiv \sqrt{h}$ in a Taylor's series around $\sqrt{h} = 0$. For $x^L + \sqrt{h} < x < x^U - \sqrt{h}$ we have, for some η, $0 \leq \eta \leq 1$

$$Y_h^{\pm}(x,t) - Y(x,t) = \pm\sqrt{h}\left.\frac{\partial Y_h^{\pm}(x,t)}{\partial z}\right|_{z=0} + \frac{h}{2}\left.\frac{\partial^2 Y_h^{\pm}(x,t)}{\partial z^2}\right|_{z=\eta\sqrt{h}}$$

$$= \left.\left[\pm\sqrt{h}Y_x + 2hY_t\right]\right|_{Y(x,t),t} + \frac{h}{2}\left.\left[Y_{xx} \pm 4\sqrt{h}\left(Y_t + Y_{xt} + \sqrt{h}Y_{tt}\right)\right]\right|_{Y(x\pm\eta\sqrt{h},t+\eta^2 h),t+\eta^2 h} \tag{92}$$

Note that $Y_x = \sigma(y,t)$. Assumption 8 guarantees that $Y_t, Y_{xx}, Y_{xt}, Y_{tt}$ exist and are locally bounded, so it follows that

$$Y_h^{\pm}(x,t) - Y(x,h) = \pm\sqrt{h}\sigma\left(Y(x,t),t\right) + O(h) \tag{93}$$

where the $O(h)$ term vanishes uniformly on sets of the form $\{(x,t) : (Y(x,t),t) \in F\}$, with F any bounded subset of $\Re^1 \times [0, \infty]$. This establishes convergence of the local second moment $\sigma_h^2(y,t)$ to $\sigma^2(y,t)$ uniformly on compact sets, and also verifies Assumption 3. As in Section 2, the convergence of the local drift μ_h to μ uniformly on compacts is ensured if $0 \leq q_h \leq 1$. Rearranging (32), we have

$$0 \leq q_h(x,t) \leq 1 \Leftrightarrow \frac{\left(Y_h^-(x,t) - Y(x,t)\right)}{h} \leq \mu(Y(x,t),t) \leq \frac{\left(Y_h^+(x,t) - Y(x,t)\right)}{h}, \tag{94}$$

and using (93), this is equivalent to

$$-\frac{\sigma(Y(x,t),t)}{\sqrt{h}} + O(1) \leq \mu(Y(x,t),t) \leq \frac{\sigma(Y(x,t),t)}{\sqrt{h}} + O(1),$$

where the $O(1)$ terms are bounded on bounded (y,t) sets. Since μ is locally bounded and σ is locally bounded away from 0 (by Assumptions 5 and 6), we have $\mu_h = \mu$ for sufficiently small h on any bounded (y,t) set, concluding the proof. *Q.E.D.*

PROOF OF THEOREM 3: Again, Assumption 1 is obvious. Assumption 2 follows the argument in the proof to Theorem 2, except that we use Stroock and Srinivasa Varadhan's Theorem 8.2.1 in place of their Theorem 7.2.1. It remains to verify Assumptions 3 and 4 and (50). Consider (50) first: ${}_hy_t$ is non-negative for all h and t by construction. Assume for the time being that for all $y < 0, \mu(y,t) = \mu(0,t) \geq 0$ and $\sigma(y,t) = \sigma(0,t) = 0$. Define

$$f(y) \equiv -y^3 \cdot I(y \leq 0) \tag{95}$$

where $I(y \leq 0)$ is an indicator function that equals one when $y \leq 0$ and zero otherwise. Since $f(y)$ is twice differentiable and $y_0 \geq 0$, we have by Itô's Lemma

$$f(y_t) = -\int_0^t 3I(y_s \leq 0)\left(y_s^2\mu(y_s,s) + y_s\sigma^2(y_s,s)\right) ds - \int_0^t 3I(y_s \leq 0)y_s^2\sigma(y_s,s)\, dW_s. \tag{96}$$

Since $\mu(y,s) \geq 0$ and $\sigma(y,s) = 0$ for $y \leq 0$, and since $f(y_t)$ is continuous a.s., we have

$$f(y_t) = -\int_0^t 3I(y_s \leq 0)y_s^2\mu(y_s,s)\, ds \leq 0 \quad a.s.$$

But by its definition in (95), $f(y_t) \geq 0$. We conclude therefore that $f(y_t) = 0 \quad a.s.$ for all t. Since, by Itô's Lemma, the path of $f(y_t)$ is continuous a.s., we have

$$\inf_{t\geq 0} f(y_t) = \sup_{t\geq 0} f(y_t) = 0 \quad a.s., \tag{97}$$

and therefore

$$\inf_{t\geq 0} y_t \geq 0 \quad a.s. \tag{98}$$

Note that our assumption that $\mu(y,s) \geq 0$ and $\sigma(y,s) = 0$ for $y < 0$ was not essential, since (97) and (98) imply that if y_t starts non-negative, it stays non-negative with probability one. All we need is for μ and σ to be continuous at $y = 0$, and $\mu(0,s) \geq 0$ and $\sigma(0,s) = 0$. This concludes the proof of (50).

As before, Assumption 3 follows when we show that $Y_h^{\pm} - y = O(\sqrt{h})$ uniformly on compacts, which will follow from the verification of Assumption 4. First we verify that μ_h converges to μ uniformly on compacts. As before, this holds as long as

$$\frac{\left(Y_h^-(x,t) - Y(x,t)\right)}{h} \leq \mu(Y(x,t),t) + o(1) \leq \frac{\left(Y_h^+(x,t) - Y(x,t)\right)}{h} \tag{99}$$

for small h uniformly on compacts. In regions bounded away from $y = 0$, (99) holds by virtue of the argument in the proof to Theorem 2. In the neighborhood of $y = 0$ the upper inequality is satisfied because J_h^+ was chosen to do so. The same holds for the lower inequality, except when $Y_h^- = 0$. In this case we require

$$-\frac{y}{h} \leq \mu + o(1). \tag{100}$$

Assumption 9 implies there exists a $\Lambda_{T,R}$ such that $|\mu| \leq y\Lambda_{T,R}$. Rearranging (100) we require

$$-\left(\Lambda_{T,R} + \frac{1}{h}\right) y \leq o(1),$$

which clearly holds for small h.

Finally, we must verify that $\sigma_h^2 \to \sigma^2$ uniformly on compacts. First we consider values of y bounded away from 0. Choose a function $c(h, \lambda)$ such that for every λ, $0 < \lambda < \infty$,

$$\lim_{h \downarrow 0} \inf_{0 \leq t \leq \lambda, c(h,\lambda) \leq y \leq \lambda} \frac{\sigma(y,t)}{\sqrt{h}} = \infty$$

and

$$\lim_{h \downarrow 0} c(h, \lambda) = 0.$$

Assumption 10 guarantees that such a function exists. Define the set

$$Q(h, \lambda) = \{(y,t) : 0 \leq t \leq \lambda,\ c(h,\lambda) \leq y \leq \lambda\}.$$

By the same Taylor's series argument as in the proof of Theorem 1, we can show that on $Q(h, \lambda)$

$$O(1) - \frac{J_h^- \sigma}{\sqrt{h}} \leq \mu \leq O(1) + \frac{J_h^+ \sigma}{\sqrt{h}}.$$

By the construction of $Q(h, \lambda)$, $\frac{\sigma}{\sqrt{h}} \to \infty$ as $h \downarrow 0$, so that $J_h^{\pm} \to 1$ uniformly on $Q(h, \lambda)$. We then have

$$Y_h^{\pm}(X(y,t),t) - y = \pm\sqrt{h}\sigma(y,t) + O(h), \tag{101}$$

implying that $\sigma_h^2 - \sigma^2 \to 0$ uniformly on compacts outside a shrinking neighborhood of $y = 0$. We are left to check convergence of σ_h^2 when $y = 0$ or when y lies in some shrinking neighborhood of 0.

Now consider the case $y = 0$. Then $y = Y_h^- = \sigma(0,t) = 0 \leq \mu(0,t)$. Here J_h^+ is selected to the smallest odd positive integer that satisfies $\mu \leq Y_h^+/h$, implying that $Y_h^+ = o(\sqrt{h})$ and that $\frac{(Y_h^+)^2}{h} = o(1)$.

To show that $\sigma_h^2 \to \sigma^2$ on every shrinking neighborhood of $y = 0$, we show that $\frac{(Y_h^+ - y)^2}{h}$ and $\frac{(Y_h^- - y)^2}{h}$ converge to σ^2 in this region. As before, expand Y_h^+ and Y_h^- as functions of $\sqrt{h}$ in a Taylor's series around $\sqrt{h} = 0$. For some $\eta, \lambda, 0 \leq \lambda \leq 1, 0 \leq \eta \leq 1$ we have

$$Y_h^+(x,t) - Y(x,t) = \left[J_h^+ \sqrt{h} Y_x + 2hY_t\right]\Big|_{Y(x,t),t} + \frac{h}{2} J^2 Y_{xx}\Big|_{Y(x+\eta\sqrt{h},t+\eta^2 h),t+\eta^2 h} + O(h^{3/2}) \tag{102}$$

and

$$Y_h^-(x,t) - Y(x,t) = \left[-J_h^-\sqrt{h}Y_x + 2hY_t\right]\Big|_{Y(x,t),t} + \frac{h}{2}Y_{xx}\Big|_{Y(x-\lambda\sqrt{h},t+\lambda^2 h),t+\lambda^2 h} + O(h^{3/2}), \quad (103)$$

where the $h^{-3/2}O(h^{3/2})$ is bounded uniformly on compact sets.

Now consider the convergence of $(Y_h^- - y)^2/h$. By (103) we have

$$Y_h^- - y \equiv Y(x - J_h^-\sqrt{h}, t+h) - y = -J_h^-\sigma(y,t)\sqrt{h} + O(h),$$

where J_h^- is constrained to satisfy

$$-J_h^-\sigma(y,t)\sqrt{h} + O(h) \leq \mu h. \quad (104)$$

If $J_h^- \neq 1$ it is because of (104), in which case

$$Y_h^- - y = \mu h + \epsilon,$$

where ϵ is an overshooting error and arises because J_h^- is constrained to be an odd integer. From (103) however, it is easy to check that ϵ is bounded by $-2\sigma(y,t)\sqrt{h} + O(h)$. Therefore

$$\left|Y_h^- - y\right| \leq \left|\mu h + 2\sigma(y,t)\sqrt{h} + O(h)\right|$$

so that

$$\frac{\left(Y_h^- - y\right)^2}{h} - \sigma^2(y,t) = 3\sigma^2(y,t) + O(\sqrt{h}). \quad (105)$$

But we're considering values of y in a shrinking neighborhood around 0, and $\sigma^2(y,t) \to 0$ as $y \to 0$, so the LHS of (105) converges to 0 as required. A similar argument (using (102)) shows the convergence of $(Y_h^+ - y)^2/h$. *Q.E.D.*

PROOF OF COROLLARY 3.1: To satisfy the conditions of Theorem 3, we must extend the process to the whole real line by defining the new stochastic differential equation

$$ds_t = \mu s_t\, dt + \sigma|s_t|^\gamma\, dW_t, \quad (106)$$

which coincides with (51) when $s \geq 0$. Assumption 5 is trivially satisfied. Use $\lambda \equiv \max[|\sigma^2 + 2\mu|, 1]$ and $\phi(S) \equiv S^2 + \max[\sigma^2, 1]$. in the Lyapunov condition given earlier, which satisfies Assumption 7. When $\gamma \geq \frac{1}{2}$, Assumptions 9 and 10 are readily verified, with $\rho(z) \equiv z^\gamma$, and Theorem 3 applies. *Q.E.D.*

PROOF OF COROLLARY 3.2: Again, we must extend the diffusion to all of $\Re^1$ in order to apply Theorem 3. Therefore define the diffusion

$$dr = \kappa(\mu - r)\, dt + \sigma\sqrt{|r|}\, dW, \quad (107)$$

with the same positive initial value r_0. For verifying Assumption 7, we use the Lyapunov condition and set $\lambda \equiv \max[1, |\sigma^2 + 2\kappa(\mu+1)|]$, and $\phi(r) \equiv \max[1, |\sigma^2 + 2\kappa\mu|] + r^2$. Assumption 9 is easily verified with $\rho(z) \equiv \sqrt{z}$, and Assumptions 5 and 10 are easily verified. The theorem now follows as a special case of Theorem 3. *Q.E.D.*

PROOF OF THEOREM 4: $\{{}_hS_t\} \Rightarrow \{S_t\}$ follow by direct application of Theorem 3. Convergence of the put price follows from Billingsley (1986), Theorem 25.12, because of the uniform boundedness of the put's terminal value at T. Convergence of the call price then follows from put-call parity. *Q.E.D*

PROOF OF THEOREM 5: We need to employ a multivariate version of Theorem 1, which is available in Nelson (1989, Appendix A). Assumption 1 of Theorem 1 (now requiring continuity of the drift function as a vector and the diffusion function as a matrix) is obviously met. The local drift of ${}_hy_t$ is

$$\frac{y \cdot (\exp(-r \cdot [1 - g(t, T-t)] \cdot h) - 1)}{h} = -r \cdot (1 - g(t, T-t)) \cdot y + O(h), \tag{108}$$

where $O(h)$ vanishes uniformly on compacts. The local second moment of ${}_hy_t$ is

$$\frac{y^2 \cdot (exp(-r \cdot [1 - g(t, T-t)] \cdot h) - 1)^2}{h} = O(h). \tag{109}$$

which also implies that the jump size vanishes uniformly on compacts. And the local cross-moment of y and r vanishes, satisfying Assumptions 3 and 4. All that remains to invoke the multivariate version of Theorem 1 is to prove that the multivariate analog of Assumption 2 holds. We have already seen in Theorem 3 that $\{{}_hr_t \Rightarrow r_t\}$, and that Assumption 2 is satisfied by the $\{r_t\}$ process. Note that there is no feedback from the y back to r. For any realization of the sample path of $\{r_t\}$,

$$y_t = \exp\left\{-\int_0^t [r_s + r_s \cdot g(s, T-t)]\, ds\right\}, \tag{110}$$

exists and is unique as a Riemann-Stieltjes integral. From Theorem 25.12 in Billingsley (1986), ${}_hP_0 \rightarrow P_0$, as required. *Q.E.D.*

References

[1] Arnold, L. (1974), *Stochastic Differential Equations: Theory and Applications*, Wiley, New York.

[2] Ball, C.A. (1989), "A Branching Model for Bond Price Dynamics and Contingent Claim Pricing," mimeo, Graduate School of Business, University of Michigan, Ann Arbor.

[3] Ball, C.A., and W.N. Torous (1983), "Bond Price Dynamics and Options," *Journal of Financial and Quantitative Analysis,* 18, 4, 517-531.

[4] Bates, D.S. (1988), "The Crash Premium: Option Pricing Under Asymmetric Processes, with Applications to Options on Deutschmark Futures," Rodney L. White Center Working Paper 36-88, The Wharton School, University of Pennsylvania.

[5] Billingsley, P. (1968), *Convergence of Probability Measures,* Wiley, New York.

[6] Billingsley, P. (1986), *Probability and Measure,* (2d Ed.), Wiley, New York.

[7] Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy,* 81, 637-659.

[8] Boyle, P.P., J. Evnine and S. Gibbs (1979), " Numerical Evaluation of Multivariate Contingent Claims," *The Review of Financial Studies,* 2, 2, 241-50.

[9] Brennan, M.J. and E.S. Schwartz (1978), "Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis," *Journal of Financial and Quantitative Analysis,* 13, 3, 461-474.

[10] Cheng, S. (1989), "On the Feasibility of Arbitrage-Based Option Pricing When Stochastic Bond Price Processes are Involved," mimeo, Columbia University, Graduate School of Business.

[11] Cox, D. and H.D. Miller (1984), *The Theory of Stochastic Processes,* Science Paperbacks, New York.

[12] Cox, J.C., J.E. Ingersoll, Jr. and S.A. Ross (1979), "Duration and the Measurement of Basis Risk," *Journal of Business,* 52, 1, 51-61.

[13] Cox, J.C., J.E. Ingersoll, Jr. and S.A. Ross (1981), "A Re-examination of Traditional Hypotheses about the Term Structure of Interest Rates," *Journal of Finance,* 35, 389-403.

[14] Cox, J.C., J.E. Ingersoll, Jr. and S.A. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica,* 53, 363-384.

[15] Cox, J.C., S.A. Ross, and M. Rubinstein (1979), "Options Pricing: A Simplified Approach," *Journal of Financial Economics,* 7, 229-263.

[16] Cox, J.C. and S.A. Ross (1976), "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics,* 3, 145-166.

[17] Cox, J.C. and M. Rubinstein (1985), *Options Markets,* Prentice Hall, New Jersey.

[18] Duffie, D. and P. Protter (1988), "From Discrete to Continuous Time Finance: Weak Convergence of the Financial Gain Process," mimeo, Graduate School of Business, Stanford University.

[19] Ethier, S.N. and Kurtz, T.G. (1986), *Markov Processes: Characterization and Convergence,* Wiley, New York.

[20] Gibbons, M. and C. Jacklin, (1989), "CEV Diffusion Estimation," mimeo, Graduate School of Business, Stanford University.

[21] Harrison, J.M. and D.M. Kreps (1979), "Martingales and Arbitrage in Multiperiod Security Markets," *Journal of Economic Theory,* 20, 381-408.

[22] He, H. (1989), "Convergence from Discrete to Continuous Time Financial Models," Finance Working Paper 190, Institute of Business and Economic Research, University of California, Berkeley.

[23] Ingersoll, Jr., J.E. (1987), *Theory of Financial Decision Making,* Rowman and Littlefield, Totowa, N.J.

[24] Karlin, S. and H.M. Taylor (1981), *A Second Course in Stochastic Processes,* Academic Press, New York.

[25] Kushner, H.J. (1984), *Approximation and Weak Convergence Methods for Random Processes,* MIT Press, Cambridge, Mass.

[26] Liptser, R.S. and A.N. Shiryayev (1977), *Statistics of Random Processes,* Volume I, Springer Verlag, New York.

[27] Longstaff, F.A. (1989), "A Nonlinear General Equilibrium Model of the Term Structure of Interest Rates," *Journal of Financial Economics,* 23, 2, pp. 195-224.

[28] Merton, R.C. (1969), "Lifetime Portfolio Selection Under Uncertainty: the Continuous Time Case," *Review of Economic Statistics,* 51, 247-257.

[29] Merton, R.C. (1973), "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science,* 4, Spring, 141-183.

[30] Nelson, D.B. (1989), "ARCH Models as Diffusion Approximations," Graduate School of Business, University of Chicago, Working Paper Series in Economics and Econometrics, 89-77, forthcoming, *Journal of Econometrics.*

[31] Pearson, N., and T.S. Sun (1989), "A Test of the Cox, Ingersoll and Ross Model of the Term Structure of Interest Rates Using the Method of Maximum Likelihood," mimeo, Sloan School of Management, M.I.T.

[32] Sharpe, W.F. (1978), *Investments,* (2d Ed.), Prentice Hall, Englewood Cliffs, N.J.

[33] Stroock, D.W. and S.R. Srinivasa Varadhan, (1979), *Multidimensional Diffusion Processes,* Springer-Verlag, Berlin.

[34] Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics,* 5, 2, 177-188.

Table 1. Values of American Call and Put Options on Stock for the CEV Process
Non Dividend Paying Stock, Binomial Method
Stock Price = 40; Interest Rate = 5%
Strike Price K = 35, 40, 45

Panel A: Call Option Values

$(T-t)$		Gamma(γ)=0.5 1 Month			Gamma(γ)=0.5 4 Months			Gamma(γ)=0.875 1 Month			Gamma(γ)=0.875 4 Months		
σ'	n / K	5	50	∞	5	50	∞	5	50	PDE	5	50	PDE
0.2	35	5.142	5.149	5.15	5.798	5.777	5.79	5.142	5.148	5.147	5.773	5.759	5.759
	40	1.049	0.999	1.00	2.249	2.152	2.17	1.049	0.998	0.990	2.250	2.151	2.153
	45	0.012	0.018	0.02	0.491	0.466	0.47	0.016	0.021	0.021	0.526	0.495	0.489
0.3	35	5.212	5.231	5.23	6.352	6.311	6.31	5.207	5.222	5.218	6.310	6.264	6.255
	40	1.531	1.454	1.46	3.199	3.051	3.00	1.531	1.454	1.446	3.200	3.050	3.057
	45	0.103	0.141	0.14	1.213	1.186	1.18	0.109	0.156	0.154	1.260	1.238	1.225
0.4	35	5.434	5.412	5.42	6.935	6.966	6.99	5.404	5.390	5.387	6.876	6.896	6.903
	40	2.012	1.911	1.92	4.149	3.951	3.98	2.012	1.911	1.901	4.149	3.949	3.961
	45	0.416	0.389	0.38	1.901	1.990	2.00	0.447	0.413	0.401	1.961	2.068	2.061

Panel B: Put Option Values

$(T-t)$		Gamma(γ)=0.5 1 Month			Gamma(γ)=0.5 4 Months			Gamma(γ)=0.875 1 Month			Gamma(γ)=0.875 4 Months		
σ'	n / K	5	50	PDE	5	50	PDE	5	50	PDE	5	50	PDE
0.2	35	0.000	0.007	0.008	0.243	0.221	0.223	0.000	0.006	0.007	0.218	0.202	0.203
	40	0.900	0.849	0.842	1.656	1.570	1.570	0.900	0.849	0.842	1.656	1.571	1.571
	45	5.000	5.000	5.000	5.058	5.063	5.060	5.000	5.000	5.000	5.085	5.082	5.078
0.3	35	0.070	0.090	0.088	0.797	0.761	0.751	0.065	0.081	0.079	0.755	0.714	0.706
	40	1.382	1.305	1.296	2.610	2.466	2.470	1.383	1.305	1.297	2.613	2.468	2.471
	45	5.032	5.044	5.042	5.691	5.641	5.628	5.038	5.056	5.052	5.743	5.693	5.680
0.4	35	0.293	0.272	0.270	1.380	1.423	1.430	0.264	0.249	0.248	1.322	1.352	1.359
	40	1.865	1.761	1.751	3.565	3.367	3.374	1.865	1.761	1.751	3.565	3.367	3.374
	45	5.286	5.256	5.245	6.384	6.400	6.395	5.317	5.282	5.269	6.449	6.479	6.472

Notes:

1. The diffusion corresponding to the CEV is defined as $dS = \mu S\,dt + \sigma S^{\gamma}\,dz$.
2. The value of σ is set so that the annual standard deviation (σ') is 0.2, 0.3 and 0.4 at the current stock price of 40. That is, $\sigma S^{\gamma} = 40\sigma'$. There are n steps in the binomial method.
3. For $\gamma = 0.5$, the column under $n = \infty$ for calls corresponds to the formula value of a European call option for the square root process; the values are taken from Cox & Rubinstein [1985, p364].
4. The column under n=PDE corresponds to the numerical solution of the partial differential equation for the option, using the implicit, finite difference method. The mesh interval along the time dimension was 0.5 days, and the mesh interval along the stock price dimension was 20 cents.

Table 2. Values of Discount Bonds for the Cox, Ingersoll, & Ross Term Structure Model, using the Binomial Method

Kappa κ	Sigma σ	Maturity Months	r_0	n = 5	n = 50	n = 100	n = 200	CIR
0.01	0.1	1	0.05	99.5843	99.5841	99.5841	99.5841	99.5841
0.01	0.1	1	0.11	99.0884	99.0877	99.0877	99.0877	99.0876
0.01	0.1	6	0.05	97.5351	97.5290	97.5287	97.5285	97.5284
0.01	0.1	6	0.11	94.6818	94.6570	94.6556	94.6549	94.6542
0.01	0.1	12	0.05	95.1427	95.1192	95.1179	95.1173	95.1166
0.01	0.1	12	0.11	89.7158	89.6235	89.6183	89.6157	89.6131
0.01	0.1	60	0.05	78.7935	78.3999	78.3771	78.3665	78.3451
0.01	0.1	60	0.11	60.5034	59.2818	59.2092	59.1727	59.1358
0.1	0.1	1	0.05	99.5835	99.5832	99.5832	99.5832	99.5832
0.1	0.1	1	0.11	99.0892	99.0886	99.0886	99.0886	99.0886
0.1	0.1	6	0.05	97.5092	97.4973	97.4967	97.4963	97.4960
0.1	0.1	6	0.11	94.7067	94.6877	94.6866	94.6860	94.6855
0.1	0.1	12	0.05	95.0430	94.9975	94.9950	94.9937	94.9924
0.1	0.1	12	0.11	89.8079	89.7367	89.7327	89.7307	89.7287
0.1	0.1	60	0.05	76.9130	76.0941	76.0486	76.0236	76.0000
0.1	0.1	60	0.11	61.6259	60.5878	60.5255	60.4942	60.4628
0.1	0.5	1	0.05	99.5837	99.5833	99.5833	99.5833	99.5833
0.1	0.5	1	0.11	99.0894	99.0889	99.0888	99.0888	99.0888
0.1	0.5	6	0.05	97.5310	97.5287	97.5287	97.5293	97.5194
0.1	0.5	6	0.11	94.7531	94.7384	94.7375	94.7370	94.7349
0.1	0.5	12	0.05	95.2109	95.2503	95.2522	95.2535	95.1632
0.1	0.5	12	0.11	90.1275	90.0911	90.1109	90.1101	90.0762
0.1	0.5	60	0.05	84.2543	87.1194	87.3946	87.4445	83.4832
0.1	0.5	60	0.11	72.9138	75.1890	75.3230	74.7616	72.5572
0.5	0.5	1	0.05	99.5804	99.5793	99.5793	99.5792	99.5792
0.5	0.5	1	0.11	99.0927	99.0928	99.0929	99.0929	99.0929
0.5	0.5	6	0.05	97.4201	97.3993	97.4159	97.3959	97.3843
0.5	0.5	6	0.11	94.8543	94.8598	94.8616	94.8618	94.8562
0.5	0.5	12	0.05	94.8006	94.9097	94.8167	94.9745	94.6592
0.5	0.5	12	0.11	90.4403	90.5151	90.5191	90.5427	90.4269
0.5	0.5	60	0.05	76.6883	76.0322	83.2037	86.9128	74.8289
0.5	0.5	60	0.11	72.5554	75.1403	73.2344	77.4502	68.6361

Notes:

1. The diffusion employed in the Cox, Ingersoll & Ross model is $dr = \kappa(\mu - r)\,dt + \sigma\sqrt{r}\,dz$.
2. The value of μ, the long run mean rate, is set at 8%. r_0 is the initial value of the interest rate.
3. The Local Expectations Hypothesis is applied to the valuation of a pure discount bond with a face value of $100, using the number of time steps indicated by n in the binomial model.
4. The column under 'CIR' indicates the value given by the formula in Cox, Ingersoll and Ross(1985) for the corresponding parameter values.

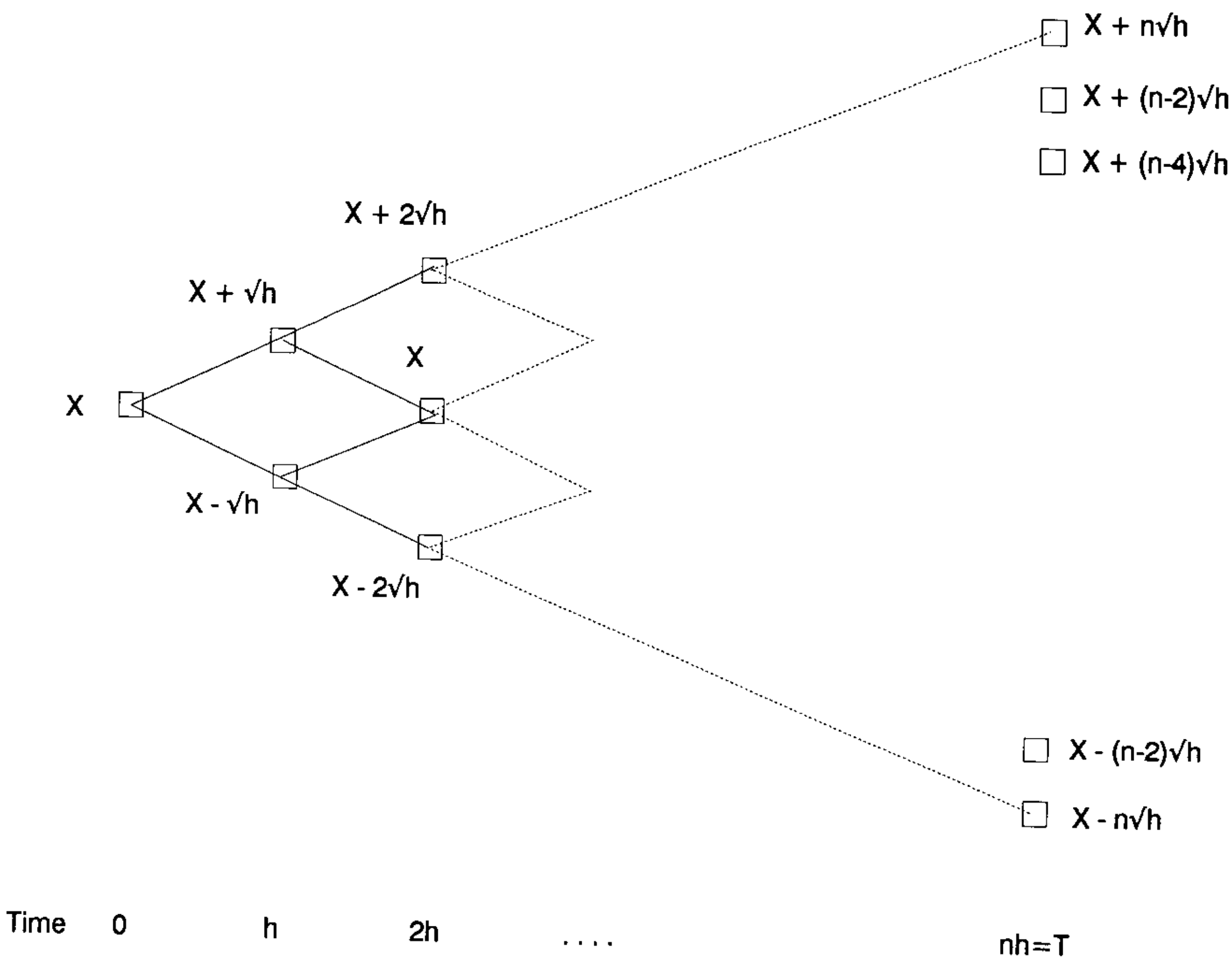


Figure 1. A Simple Binomial Tree Structure for X

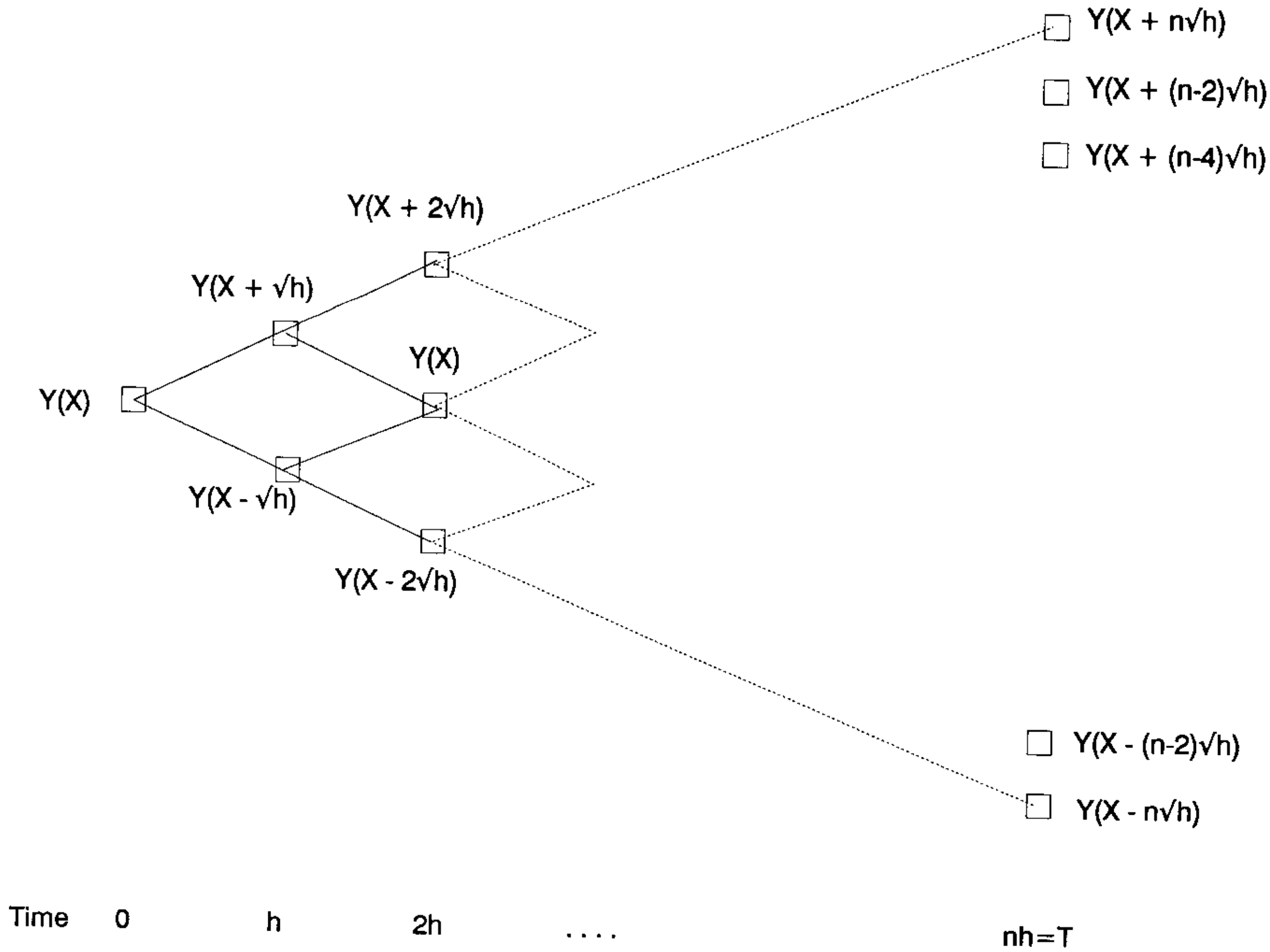


Figure 2. A Simple Binomial Tree Structure for $y = Y(X)$

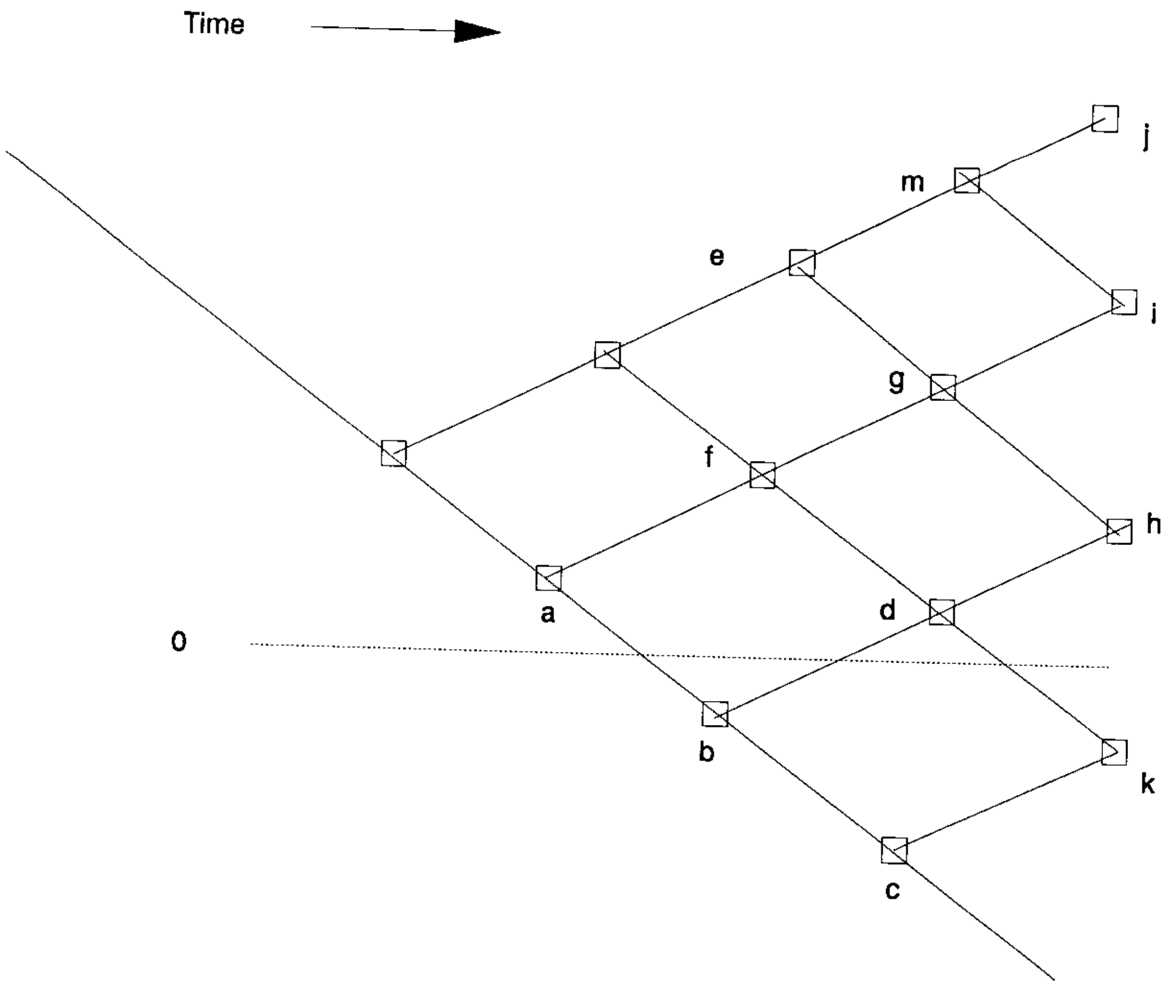


Figure 3. A Binomial Tree with a Lower Boundary at 0