Matrix Computation

To set up a philosophy against physics is rash;
philosophers who have done so
have always ended in disaster.

— Bertrand Russell

Definitions and Basic Results

- Let $A \equiv [\, a_{ij} \,]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \boldsymbol{R}^{m \times n}$, denote an $m \times n$ matrix.

- It can also be represented as $[\, a_1, a_2, \ldots, a_n \,]$ where $a_i \in \boldsymbol{R}^m$ are vectors.

  – Vectors are column vectors unless stated otherwise.

- A is a square matrix when $m = n$.

- The rank of a matrix is the largest number of linearly independent columns.

Definitions and Basic Results (continued)

- A square matrix A is said to be symmetric if $A^{\mathrm{T}} = A$.

- A real $n \times n$ matrix

$$A \equiv [\, a_{ij} \,]_{i,j}$$

  is diagonally dominant if $|\, a_{ii} \,| > \sum_{j \neq i} |\, a_{ij} \,|$ for $1 \leq i \leq n$.

  – Such matrices are nonsingular.

- The identity matrix is the square matrix

$$I \equiv \mathrm{diag}[\, 1, 1, \ldots, 1 \,].$$

Definitions and Basic Results (concluded)

- A matrix has full column rank if its columns are linearly independent.

- A real symmetric matrix A is positive definite if

$$x^{\mathrm{T}} A x = \sum_{i,j} a_{ij} x_i x_j > 0$$

  for any nonzero vector x.

- A matrix A is positive definite if and only if there exists a matrix W such that $A = W^{\mathrm{T}} W$ and W has full column rank.

Cholesky Decomposition

- Positive definite matrices can be factored as

$$A = LL^{\mathrm{T}},$$

called the Cholesky decomposition.

  − Above, L is a lower triangular matrix.

Generation of Multivariate Normal Distribution

- Let $\boldsymbol{x} \equiv [\, x_1, x_2, \ldots, x_n \,]^{\mathrm{T}}$ be a vector random variable with a positive definite covariance matrix C.

- As usual, assume $E[\, \boldsymbol{x} \,] = \boldsymbol{0}$.

- This distribution can be generated by $P\boldsymbol{y}$.

  - $C = PP^{\mathrm{T}}$ is the Cholesky decomposition of C.[a]

  - $\boldsymbol{y} \equiv [\, y_1, y_2, \ldots, y_n \,]^{\mathrm{T}}$ is a vector random variable with a covariance matrix equal to the identity matrix.

_____

[a]What if C is not positive definite? See Lai and Lyuu (2007).

Generation of Multivariate Normal Distribution (concluded)

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^{\mathrm{T}}$.

- We start with independent standard normal distributions $y_1, y_2, \ldots, y_n$.

- Then $P[\, y_1, y_2, \ldots, y_n \,]^{\mathrm{T}}$ has the desired distribution.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives (p. 556).

- For example, the rainbow option on k assets has payoff

$$\max(\max(S_1, S_2, \ldots, S_k) - X, 0)$$

  at maturity.

- The closed-form formula is a multi-dimensional integral.[a]

---

[a]Johnson (1987).

Multivariate Derivatives Pricing (concluded)

- Suppose $dS_j/S_j = r\,dt + \sigma_j\,dW_j$, $1 \le j \le n$, where C is the correlation matrix for $dW_1, dW_2, \dots, dW_k$.

- Let $C = PP^{\mathrm{T}}$.

- Let ξ consist of k independent random variables from $N(0,1)$.

- Let $\xi' = P\xi$.

- Similar to Eq. (63) on p. 595,

$$S_{i+1} = S_i e^{(r - \sigma_j^2/2)\,\Delta t + \sigma_j\sqrt{\Delta t}\,\xi'_j}, \quad 1 \le j \le n.$$

Least-Squares Problems

- The least-squares (LS) problem is concerned with $\min_{x \in R^n} \| Ax - b \|$, where $A \in \boldsymbol{R}^{m \times n}$, $b \in \boldsymbol{R}^m$, $m \geq n$.

- The LS problem is called regression analysis in statistics and is equivalent to minimizing the mean-square error.

- Often written as

$$Ax = b.$$

Polynomial Regression

- In polynomial regression, $x_0 + x_1 x + \cdots + x_n x^n$ is used to fit the data $\{ (a_1, b_1), (a_2, b_2), \ldots, (a_m, b_m) \}$.

- This leads to the LS problem,

$$
\begin{bmatrix}
1 & a_1 & a_1^2 & \cdots & a_1^n \\
1 & a_2 & a_2^2 & \cdots & a_2^n \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & a_m & a_m^2 & \cdots & a_m^n
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
\vdots \\
x_n
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m
\end{bmatrix}.
$$

- Consult the text for solutions.

American Option Pricing by Simulation

- The continuation value of an American option is the conditional expectation of the payoff from keeping the option alive now.

- The option holder must compare the immediate exercise value and the continuation value.

- In standard Monte Carlo simulation, each path is treated independently of other paths.

- But the decision to exercise the option cannot be reached by looking at only one path alone.

The Least-Squares Monte Carlo Approach

- The continuation value can be estimated from the cross-sectional information in the simulation by using least squares.[a]

- The result is a function of the state for estimating the continuation values.

- Use the function to estimate the continuation value for each path to determine its cash flow.

- This is called the least-squares Monte Carlo (LSM) approach and is provably convergent.[b]

---

[a]Longstaff and Schwartz (2001).
[b]Clément, Lamberton, and Protter (2002).

A Numerical Example

- Consider a 3-year American put on a non-dividend-paying stock.

- The put is exercisable at years 0, 1, 2, and 3.

- The strike price $X = 105$.

- The annualized riskless rate is $r = 5\%$.

- The spot stock price is 101.

  - The annual discount factor hence equals 0.951229.

- We use only 8 price paths to illustrate the algorithm.
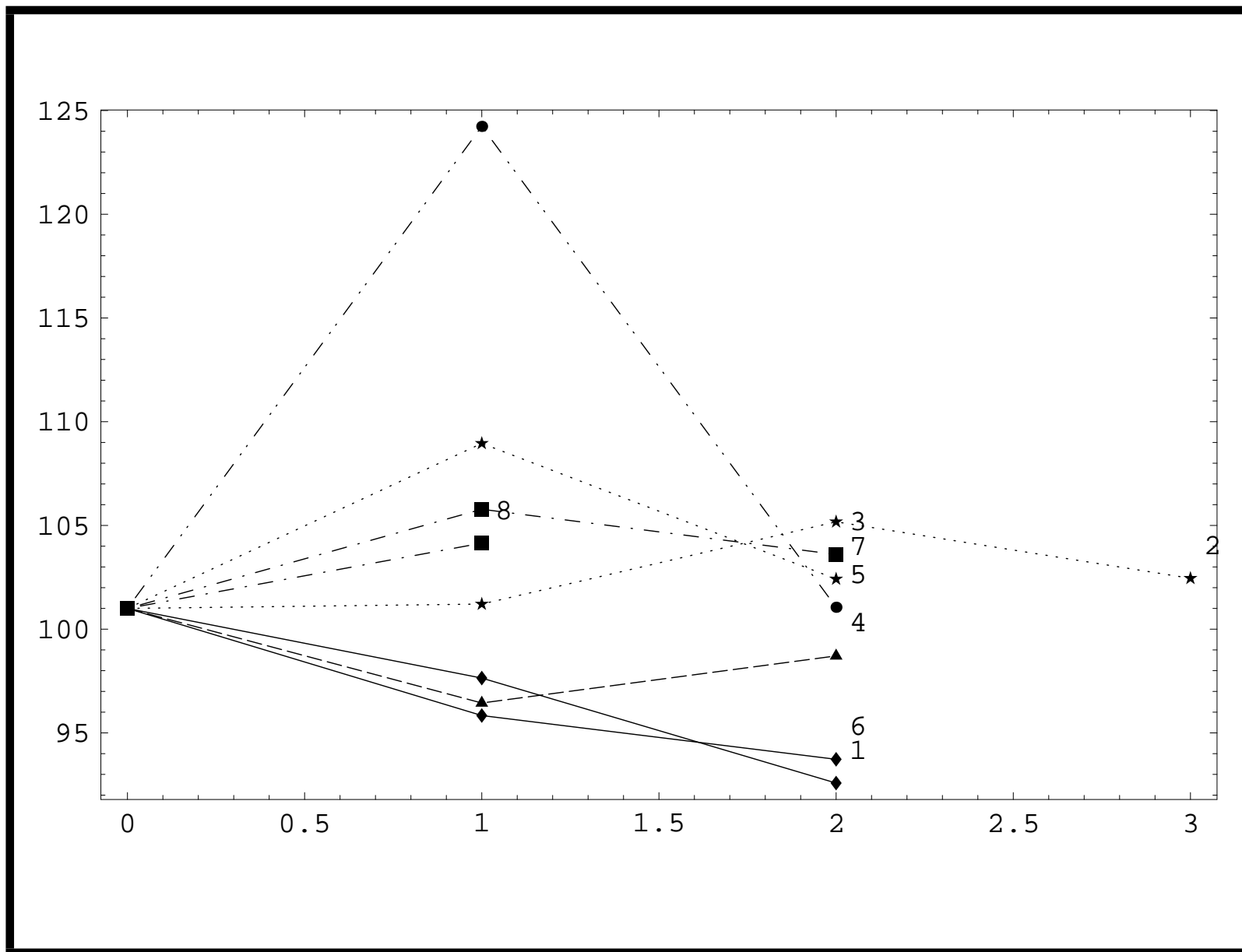
A Numerical Example (continued)

Stock price paths

Path	Year 0	Year 1	Year 2	Year 3
1	**101**	**97.6424**	**92.5815**	107.5178
2	**101**	**101.2103**	105.1763	**102.4524**
3	**101**	105.7802	**103.6010**	124.5115
4	**101**	**96.4411**	**98.7120**	108.3600
5	**101**	124.2345	**101.0564**	**104.5315**
6	**101**	**95.8375**	**93.7270**	**99.3788**
7	**101**	108.9554	**102.4177**	**100.9225**
8	**101**	**104.1475**	113.2516	115.0994
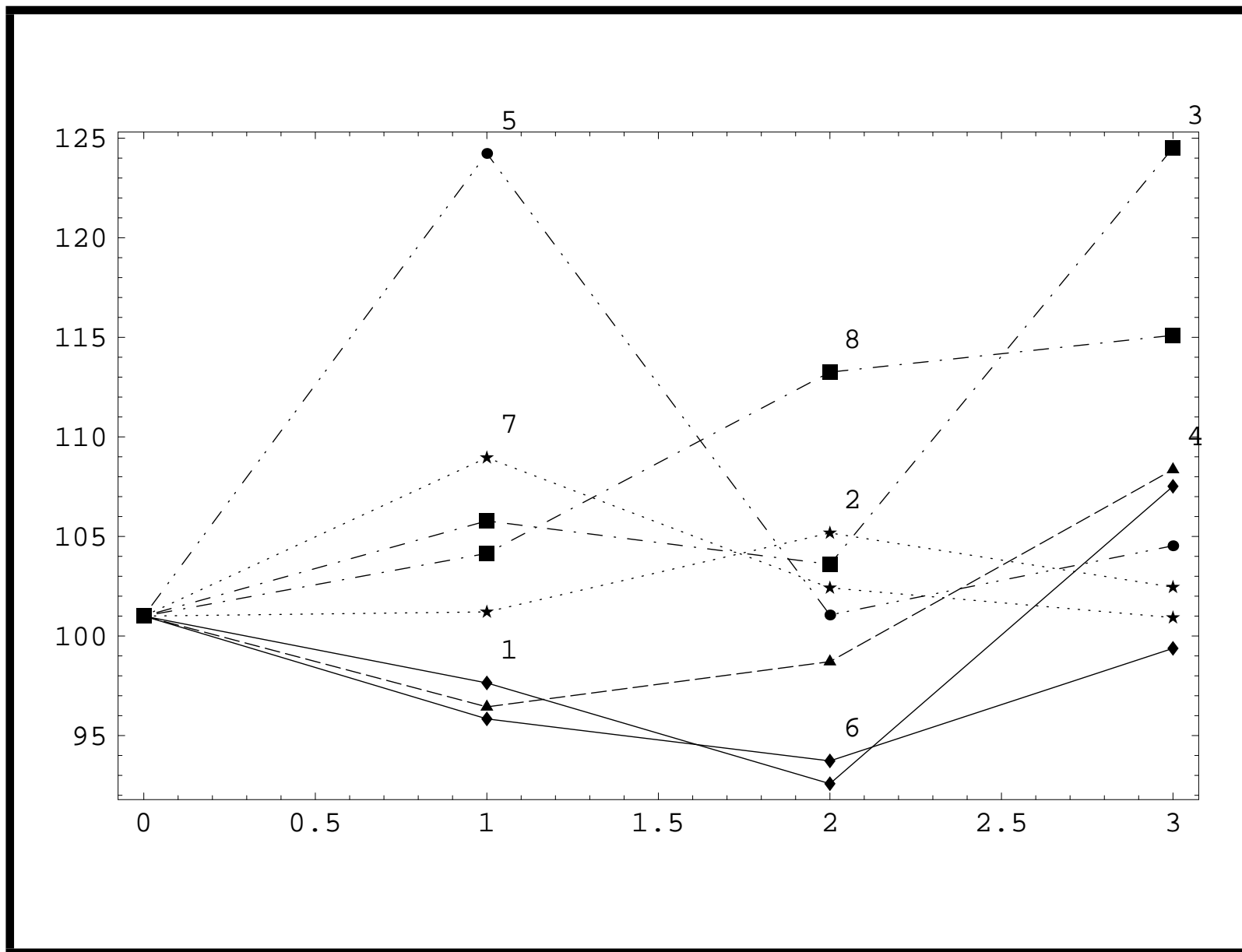
A Numerical Example (continued)

- We use the basis functions $1, x, x^2$.

  - Other basis functions are possible.[a]

- The plot next page shows the final estimated optimal exercise strategy given by LSM.

- We now proceed to tackle our problem.

- Our concrete problem is to calculate the cash flow along each path, using information from all paths.

---

[a]Laguerre polynomials, Hermite polynomials, Legendre polynomials, Chebyshev polynomials, Gedenbauer polynomials, and Jacobi polynomials.

A Numerical Example (continued)

Cash flows at year 3

Path	Year 0	Year 1	Year 2	Year 3
1	—	—	—	0
2	—	—	—	2.5476
3	—	—	—	0
4	—	—	—	0
5	—	—	—	0.4685
6	—	—	—	5.6212
7	—	—	—	4.0775
8	—	—	—	0

A Numerical Example (continued)

- The cash flows at year 3 are the exercise value if the put is in the money.

- Only 4 paths are in the money: 2, 5, 6, 7.

- Some of the cash flows may not occur if the put is exercised earlier, which we will find out step by step.

- Incidentally, the European counterpart has a value of

$$0.951229^3 \times \frac{2.5476 + 0.4685 + 5.6212 + 4.0775}{8} = 1.3680.$$

A Numerical Example (continued)

- We move on to year 2.

- For each state that is in the money at year 2, we must decide whether to exercise it.

- There are 6 paths for which the put is in the money: 1, 3, 4, 5, 6, 7.

- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.

  - If there were none, we would move on to year 1.

A Numerical Example (continued)

- Let x denote the stock prices at year 2 for those 6 paths.

- Let y denote the corresponding discounted future cash flows (at year 3) if the put is not exercised at year 2.

A Numerical Example (continued)

Regression at year 2

Path	x	y
1	92.5815	0×0.951229
2	—	—
3	103.6010	0×0.951229
4	98.7120	0×0.951229
5	101.0564	0.4685×0.951229
6	93.7270	5.6212×0.951229
7	102.4177	4.0775×0.951229
8	—	—

A Numerical Example (continued)

- We regress y on 1, x, and x^2.

- The result is

$$f(x) = 22.08 - 0.313114 \times x + 0.00106918 \times x^2.$$

- f estimates the continuation value conditional on the stock price at year 2.

- We next compare the immediate exercise value and the continuation value.

A Numerical Example (continued)

Optimal early exercise decision at year 2

Path	Exercise	Continuation
1	12.4185	$f(92.5815) = 2.2558$
2	—	—
3	1.3990	$f(103.6010) = 1.1168$
4	6.2880	$f(98.7120) = 1.5901$
5	3.9436	$f(101.0564) = 1.3568$
6	11.2730	$f(93.7270) = 2.1253$
7	2.5823	$f(102.4177) = 0.3326$
8	—	—

A Numerical Example (continued)

- Amazingly, the put should be exercised in all 6 paths: 1, 3, 4, 5, 6, 7.

- Now, any positive cash flow at year 3 should be set to zero for these paths as the put is exercised before year 3.

  - They are paths 5, 6, 7.

- Hence the cash flows on p. 646 become the next ones.

A Numerical Example (continued)

Cash flows at years 2 & 3

Path	Year 0	Year 1	Year 2	Year 3
1	—	—	12.4185	0
2	—	—	0	2.5476
3	—	—	1.3990	0
4	—	—	6.2880	0
5	—	—	3.9436	0
6	—	—	11.2730	0
7	—	—	2.5823	0
8	—	—	0	0

A Numerical Example (continued)

- We move on to year 1.

- For each state that is in the money at year 1, we must decide whether to exercise it.

- There are 5 paths for which the put is in the money: 1, 2, 4, 6, 8.

- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.

  – If there were none, we would move on to year 0.

A Numerical Example (continued)

- Let x denote the stock prices at year 1 for those 5 paths.

- Let y denote the corresponding discounted future cash flows if the put is not exercised at year 1.

- From p. 654, we have the following table.

A Numerical Example (continued)

Regression at year 1

Path	x	y
1	97.6424	12.4185×0.951229
2	101.2103	2.5476×0.951229^2
3	—	—
4	96.4411	6.2880×0.951229
5	—	—
6	95.8375	11.2730×0.951229
7	—	—
8	104.1475	0

A Numerical Example (continued)

- We regress y on 1, x, and x^2.

- The result is

$$f(x) = -420.964 + 9.78113 \times x - 0.0551567 \times x^2.$$

- f estimates the continuation value conditional on the stock price at year 1.

- We next compare the immediate exercise value and the continuation value.

A Numerical Example (continued)

Optimal early exercise decision at year 1

Path	Exercise	Continuation
1	7.3576	$f(97.6424) = 8.2230$
2	3.7897	$f(101.2103) = 3.9882$
3	—	—
4	8.5589	$f(96.4411) = 9.3329$
5	—	—
6	9.1625	$f(95.8375) = 9.83042$
7	—	—
8	0.8525	$f(104.1475) = -0.551885$

A Numerical Example (continued)

- The put should be exercised for 1 path only: 8.

- Now, any positive future cash flow should be set to zero for this path as the put is exercised before years 2 and 3.
  - But there is none.

- Hence the cash flows on p. 654 become the next ones.

- They also confirm the plot on p. 645.

A Numerical Example (continued)

Cash flows at years 1, 2, & 3

Path	Year 0	Year 1	Year 2	Year 3
1	—	0	12.4185	0
2	—	0	0	2.5476
3	—	0	1.3990	0
4	—	0	6.2880	0
5	—	0	3.9436	0
6	—	0	11.2730	0
7	—	0	2.5823	0
8	—	0.8525	0	0

A Numerical Example (continued)

- We move on to year 0.

- The continuation value is, from p 661,

$$(12.4185 \times 0.951229^2 + 2.5476 \times 0.951229^3$$
$$+1.3990 \times 0.951229^2 + 6.2880 \times 0.951229^2$$
$$+3.9436 \times 0.951229^2 + 11.2730 \times 0.951229^2$$
$$+2.5823 \times 0.951229^2 + 0.8525 \times 0.951229)/8$$
$$= \quad 4.66263.$$

A Numerical Example (concluded)

- As this is larger than the immediate exercise value of $105 - 101 = 4$, the put should not be exercised at year 0.

- Hence the put's value is estimated to be 4.66263.

- Compare this to the European put's value of 1.3680 (p. 647).

Time Series Analysis

The historian is a prophet in reverse.

— Friedrich von Schlegel (1772–1829)

Conditional Variance Models for Price Volatility

- Although a stationary model (see text for definition) has constant variance, its *conditional* variance may vary.

- Take for example an AR(1) process $X_t = aX_{t-1} + \epsilon_t$ with $|a| < 1$.

  - Here, ϵ_t is a stationary, uncorrelated process with zero mean and constant variance σ^2.

- The conditional variance,

$$\text{Var}[\, X_t \mid X_{t-1}, X_{t-2}, \ldots \,],$$

equals σ^2, which is smaller than the *unconditional* variance $\text{Var}[\, X_t \,] = \sigma^2/(1 - a^2)$.

Conditional Variance Models for Price Volatility (concluded)

- In the lognormal model, the conditional variance evolves independently of past returns.

- Suppose we assume that conditional variances are deterministic functions of past returns:

$$V_t = f(X_{t-1}, X_{t-2}, \dots)$$

for some function f.

- Then V_t can be computed given the information set of past returns:

$$I_{t-1} \equiv \{X_{t-1}, X_{t-2}, \dots\}.$$

ARCH Models[a]

- An influential model in this direction is the autoregressive conditional heteroskedastic (ARCH) model.

- Assume that $\{U_t\}$ is a Gaussian stationary, uncorrelated process.

---

[a]Engle (1982), co-winner of the 2003 Nobel Prize in Economic Sciences.

ARCH Models (continued)

- The ARCH(p) process is defined by

$$X_t - \mu = \left(a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2 \right)^{1/2} U_t,$$

  where $a_1, \ldots, a_p \geq 0$ and $a_0 > 0$.
  - Thus $X_t \,|\, I_{t-1} \sim N(\mu, V_t^2)$.

- The variance V_t^2 satisfies

$$V_t^2 = a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2.$$

- The volatility at time t as estimated at time $t-1$ depends on the p most recent observations on squared returns.

ARCH Models (concluded)

- The ARCH(1) process

$$X_t - \mu = (a_0 + a_1(X_{t-1} - \mu)^2)^{1/2} U_t$$

  is the simplest.

- For it,

$$\mathrm{Var}[\, X_t \mid X_{t-1} = x_{t-1} \,] = a_0 + a_1(x_{t-1} - \mu)^2.$$

- The process $\{\, X_t \,\}$ is stationary with finite variance if and only if $a_1 < 1$, in which case $\mathrm{Var}[\, X_t \,] = a_0/(1 - a_1)$.

GARCH Models[a]

- A very popular extension of the ARCH model is the generalized autoregressive conditional heteroskedastic (GARCH) process.

- The simplest GARCH$(1, 1)$ process adds $a_2 V_{t-1}^2$ to the ARCH(1) process, resulting in

$$V_t^2 = a_0 + a_1 (X_{t-1} - \mu)^2 + a_2 V_{t-1}^2.$$

- The volatility at time t as estimated at time $t - 1$ depends on the squared return and the estimated volatility at time $t - 1$.

---

[a]Bollerslev (1986); Taylor (1986).

GARCH Models (concluded)

- The estimate of volatility averages past squared returns by giving heavier weights to recent squared returns (see text).

- It is usually assumed that $a_1 + a_2 < 1$ and $a_0 > 0$, in which case the unconditional, long-run variance is given by $a_0/(1 - a_1 - a_2)$.

- A popular special case of GARCH$(1, 1)$ is the exponentially weighted moving average process, which sets a_0 to zero and a_2 to $1 - a_1$.

- This model is used in J.P. Morgan's RiskMetrics™.

GARCH Option Pricing

- Options can be priced when the underlying asset's return follows a GARCH process.

- Let S_t denote the asset price at date t.

- Let h_t^2 be the conditional variance of the return over the period $[\, t, t+1\,]$ given the information at date t.

  - "One day" is merely a convenient term for any elapsed time Δt.

GARCH Option Pricing (continued)

- Adopt the following risk-neutral process for the price dynamics:[a]

$$\ln \frac{S_{t+1}}{S_t} = r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}, \tag{66}$$

where

$$
\begin{aligned}
h_{t+1}^2 &= \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2, & (67) \\
\epsilon_{t+1} &\sim N(0,1) \text{ given information at date } t, \\
r &= \text{daily riskless return,} \\
c &\geq 0.
\end{aligned}
$$

---

[a]Duan (1995).

GARCH Option Pricing (continued)

- The five unknown parameters of the model are c, h_0, β_0, β_1, and β_2.

- It is postulated that $\beta_0, \beta_1, \beta_2 \geq 0$ to make the conditional variance positive.

- The above process, called the nonlinear asymmetric GARCH model, generalizes the GARCH$(1, 1)$ model (see text).

GARCH Option Pricing (continued)

- It captures the volatility clustering in asset returns first noted by Mandelbrot (1963).[a]

  – When $c = 0$, a large ϵ_{t+1} results in a large h_{t+1}, which in turns tends to yield a large h_{t+2}, and so on.

- It also captures the negative correlation between the asset return and changes in its (conditional) volatility.[b]

  – For $c > 0$, a positive ϵ_{t+1} (good news) tends to decrease h_{t+1}, whereas a negative ϵ_{t+1} (bad news) tends to do the opposite.

---

[a] "... large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes ..."

[b] Noted by Black (1976): Volatility tends to rise in response to "bad news" and fall in response to "good news."

GARCH Option Pricing (concluded)

- With $y_t \equiv \ln S_t$ denoting the logarithmic price, the model becomes

$$y_{t+1} = y_t + r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}. \tag{68}$$

- The pair (y_t, h_t^2) completely describes the current state.

- The conditional mean and variance of y_{t+1} are clearly

$$E[\, y_{t+1} \,|\, y_t, h_t^2 \,] \;=\; y_t + r - \frac{h_t^2}{2}, \tag{69}$$

$$\mathrm{Var}[\, y_{t+1} \,|\, y_t, h_t^2 \,] \;=\; h_t^2. \tag{70}$$

The Ritchken-Trevor (RT) Algorithm[a]

- The GARCH model is a continuous-state model.

- To approximate it, we turn to trees with *discrete* states.

- Path dependence in GARCH makes the tree for asset prices explode exponentially (why?).

- We need to mitigate this combinatorial explosion.

---

[a]Ritchken and Trevor (1999).
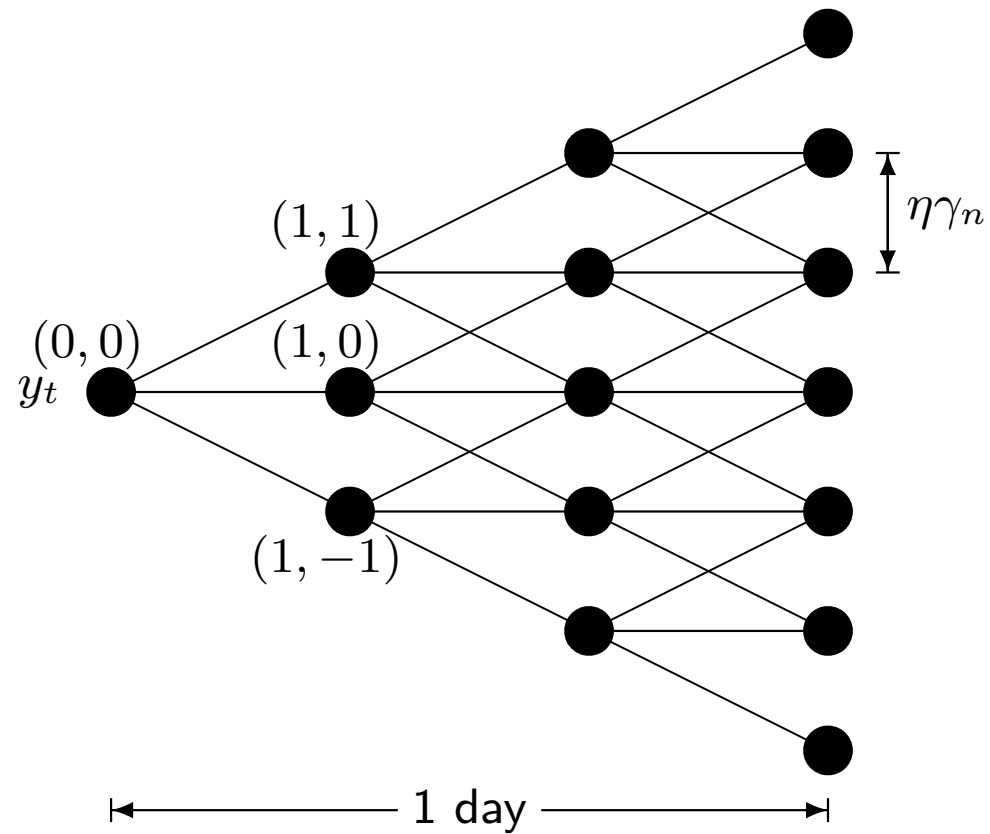
The Ritchken-Trevor Algorithm (continued)

- Partition a day into n periods.

- Three states follow each state (y_t, h_t^2) after a period.

- As the trinomial model combines, $2n + 1$ states at date $t + 1$ follow each state at date t (recall p. 539).

- These $2n + 1$ values must approximate the distribution of (y_{t+1}, h_{t+1}^2).

- So the conditional moments (69)–(70) at date $t + 1$ on p. 677 must be matched by the trinomial model to guarantee convergence to the continuous-state model.

The Ritchken-Trevor Algorithm (continued)

- It remains to pick the jump size and the three branching probabilities.

- The role of σ in the Black-Scholes option pricing model is played by h_t in the GARCH model.

- As a jump size proportional to $\sigma/\sqrt{n}$ is picked in the BOPM, a comparable magnitude will be chosen here.

- Define $\gamma \equiv h_0$, though other multiples of h_0 are possible, and

$$\gamma_n \equiv \frac{\gamma}{\sqrt{n}}.$$

- The jump size will be some integer multiple η of γ_n.

- We call η the jump parameter (p. 681).

The seven values on the right approximate the distribution of logarithmic price y_{t+1}.

The Ritchken-Trevor Algorithm (continued)

- The middle branch does not change the underlying asset's price.

- The probabilities for the up, middle, and down branches are

$$
p_u \quad = \quad \frac{h_t^2}{2\eta^2\gamma^2} + \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}, \tag{71}
$$

$$
p_m \quad = \quad 1 - \frac{h_t^2}{\eta^2\gamma^2}, \tag{72}
$$

$$
p_d \quad = \quad \frac{h_t^2}{2\eta^2\gamma^2} - \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}. \tag{73}
$$

The Ritchken-Trevor Algorithm (continued)

- It can be shown that:

  - The trinomial model takes on $2n + 1$ values at date $t + 1$ for y_{t+1}.

  - These values have a matching mean for y_{t+1}.

  - These values have an asymptotically matching variance for y_{t+1}.

- The central limit theorem thus guarantees the desired convergence as n increases.

The Ritchken-Trevor Algorithm (continued)

- We can dispense with the intermediate nodes *between* dates to create a $(2n+1)$-nomial tree (p. 685).

- The resulting model is multinomial with $2n+1$ branches from any state (y_t, h_t^2).

- There are two reasons behind this manipulation.

  - Interdate nodes are created merely to approximate the continuous-state model after one day.

  - Keeping the interdate nodes results in a tree that can be as much as n times larger.
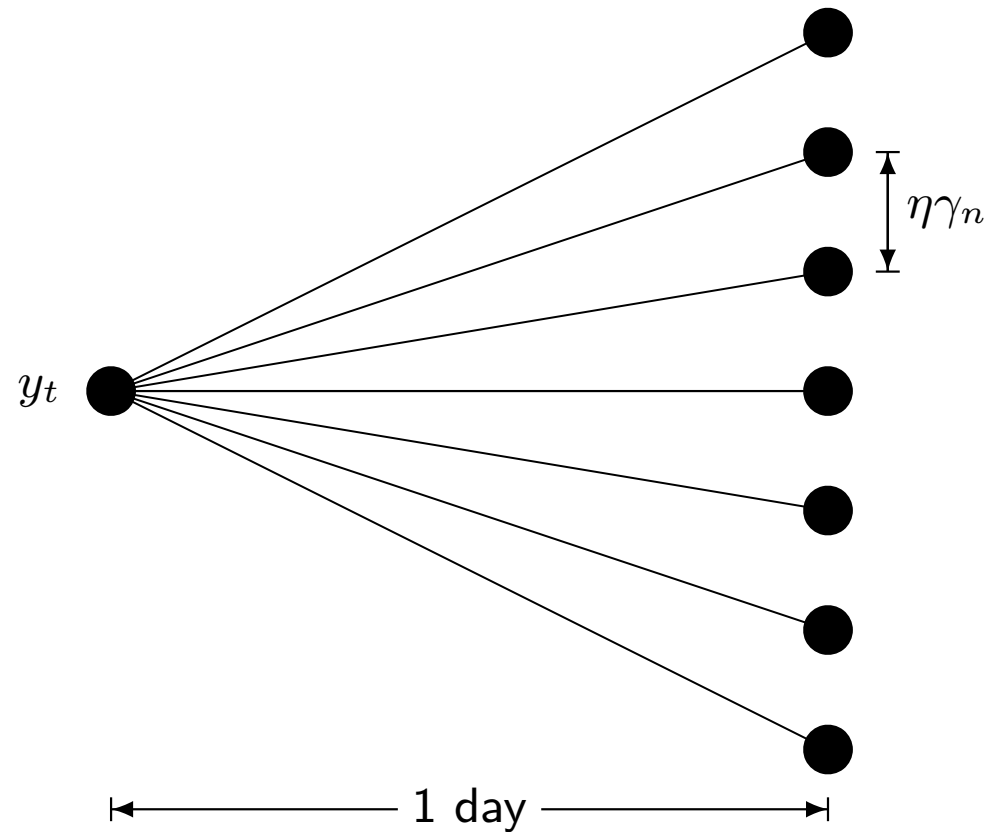
This heptanomial tree is the outcome of the trinomial tree on p. 681 after its intermediate nodes are removed.

The Ritchken-Trevor Algorithm (continued)

- A node with logarithmic price $y_t + \ell \eta \gamma_n$ at date $t+1$ follows the current node at date t with price y_t for some $-n \le \ell \le n$.

- To reach that price in n periods, the number of up moves must exceed that of down moves by exactly ℓ.

- The probability that this happens is

$$P(\ell) \equiv \sum_{j_u, j_m, j_d} \frac{n!}{j_u! \, j_m! \, j_d!} \, p_u^{j_u} p_m^{j_m} p_d^{j_d},$$

with $j_u, j_m, j_d \ge 0$, $n = j_u + j_m + j_d$, and $\ell = j_u - j_d$.

The Ritchken-Trevor Algorithm (continued)

- A particularly simple way to calculate the $P(\ell)$s starts by noting that

$$(p_u x + p_m + p_d x^{-1})^n = \sum_{\ell=-n}^{n} P(\ell)\, x^{\ell}. \qquad (74)$$

- So we expand $(p_u x + p_m + p_d x^{-1})^n$ and retrieve the probabilities by reading off the coefficients.

- It can be computed in $O(n^2)$ time.

The Ritchken-Trevor Algorithm (continued)

- The updating rule (67) on p. 674 must be modified to account for the adoption of the discrete-state model.

- The logarithmic price $y_t + \ell\eta\gamma_n$ at date $t+1$ following state (y_t, h_t^2) at date t has a variance equal to

$$h_{t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon'_{t+1} - c)^2, \qquad (75)$$

  − Above,

$$\epsilon'_{t+1} = \frac{\ell\eta\gamma_n - (r - h_t^2/2)}{h_t}, \quad \ell = 0, \pm 1, \pm 2, \ldots, \pm n,$$

  is a discrete random variable with $2n+1$ values.

The Ritchken-Trevor Algorithm (continued)

- Different conditional variances h_t^2 may require different η so that the probabilities calculated by Eqs. (71)–(73) on p. 682 lie between 0 and 1.

- This implies varying jump sizes.

- The necessary requirement $p_m \geq 0$ implies $\eta \geq h_t/\gamma$.

- Hence we try

$$\eta = \lceil h_t/\gamma \rceil, \lceil h_t/\gamma \rceil + 1, \lceil h_t/\gamma \rceil + 2, \ldots$$

until valid probabilities are obtained or until their nonexistence is confirmed.
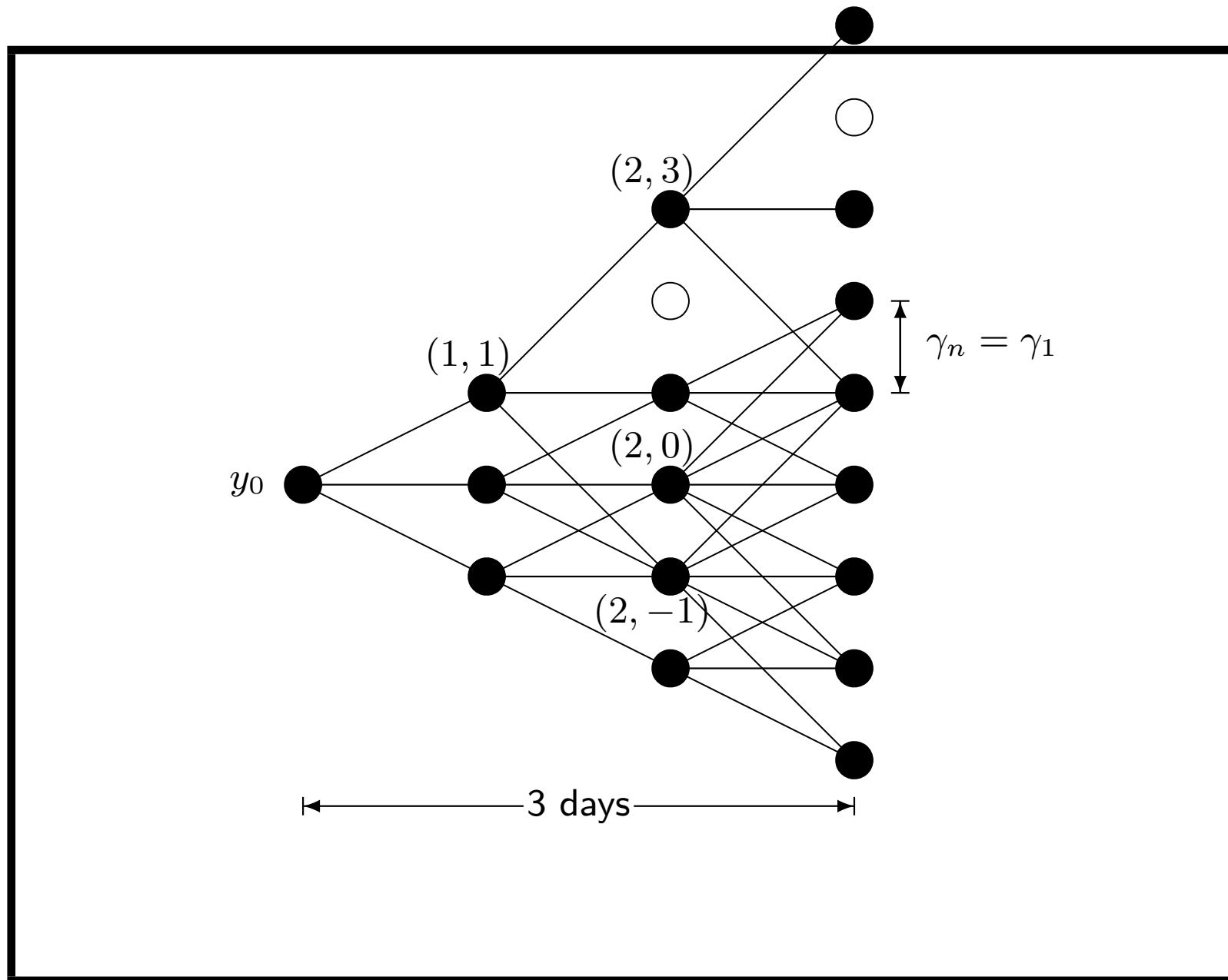
The Ritchken-Trevor Algorithm (continued)

- The sufficient and necessary condition for valid probabilities to exist is[a]

$$\frac{|\, r - (h_t^2/2)\,|}{2\eta\gamma\sqrt{n}} \leq \frac{h_t^2}{2\eta^2\gamma^2} \leq \min\left(1 - \frac{|\, r - (h_t^2/2)\,|}{2\eta\gamma\sqrt{n}}, \frac{1}{2}\right).$$

- Obviously, the magnitude of η tends to grow with h_t.

- The plot on p. 691 uses $n = 1$ to illustrate our points for a 3-day model.

- For example, node $(1, 1)$ of date 1 and node $(2, 3)$ of date 2 pick $\eta = 2$.

---

[a]Lyuu and Wu (2003).

$(2,3)$

$(1,1)$

$\gamma_n = \gamma_1$

$(2,0)$

y_0

$(2,-1)$

3 days

The Ritchken-Trevor Algorithm (continued)

- The topology of the tree is not a standard combining multinomial tree.

- For example, a few nodes on p. 691 such as nodes $(2, 0)$ and $(2, -1)$ have multiple jump sizes.

- The reason is the path dependence of the model.
  - Two paths can reach node $(2, 0)$ from the root node, each with a different variance for the node.
  - One of the variances results in $\eta = 1$, whereas the other results in $\eta = 2$.

The Ritchken-Trevor Algorithm (concluded)

- The possible values of h_t^2 at a node are exponential nature.

- To address this problem, we record only the maximum and minimum h_t^2 at each node.[a]

- Therefore, each node on the tree contains only two states $(y_t, h_{\max}^2)$ and $(y_t, h_{\min}^2)$.

- Each of $(y_t, h_{\max}^2)$ and $(y_t, h_{\min}^2)$ carries its own η and set of $2n + 1$ branching probabilities.

---

[a]Cakici and Topyan (2000).

Negative Aspects of the Ritchken-Trevor Algorithm[a]

- A small n may yield inaccurate option prices.

- But the tree will grow exponentially if n is large enough.
  - Specifically, $n > (1 - \beta_1)/\beta_2$ when $r = c = 0$.

- A large n has another serious problem: The tree cannot grow beyond a certain date.

- Thus the choice of n may be limited in practice.

- The RT algorithm can be modified to be free of shortened maturity and (to some extent) exponential complexity.[b]
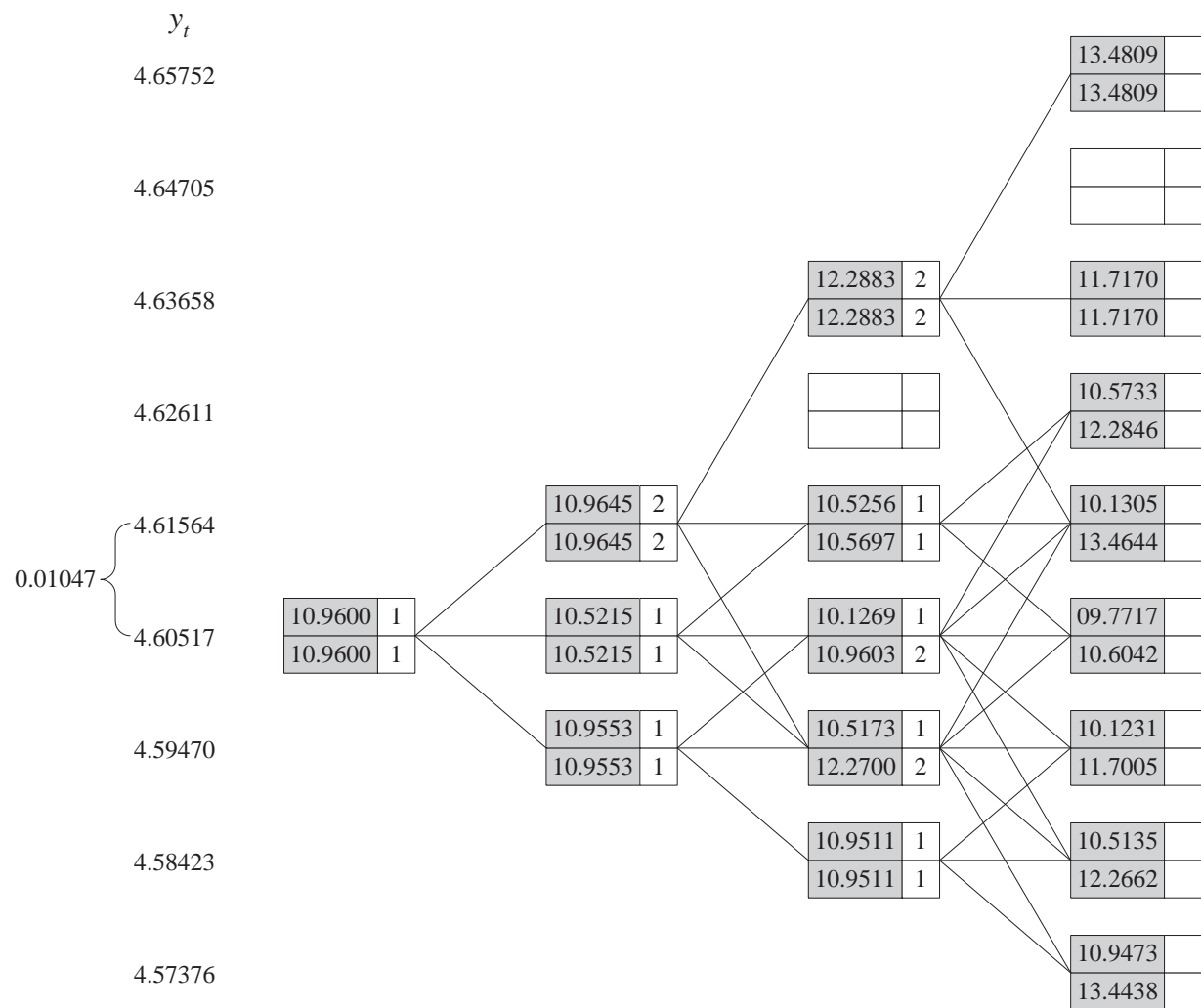
---

[a]Lyuu and Wu (2003); Lyuu and Wu (2005).
[b]It is only quadratic if n is not too large!

Numerical Examples

- Assume $S_0 = 100$, $y_0 = \ln S_0 = 4.60517$, $r = 0$, $h_0^2 = 0.0001096$, $\gamma = h_0 = 0.010469$, $n = 1$, $\gamma_n = \gamma/\sqrt{n} = 0.010469$, $\beta_0 = 0.000006575$, $\beta_1 = 0.9$, $\beta_2 = 0.04$, and $c = 0$.

- A daily variance of 0.0001096 corresponds to an annual volatility of $\sqrt{365 \times 0.0001096} \approx 20\%$.

- Let $h^2(i, j)$ denote the variance at node (i, j).

- Initially, $h^2(0, 0) = h_0^2 = 0.0001096$.

Numerical Examples (continued)

- Let $h_{\max}^2(i, j)$ denote the maximum variance at node (i, j).

- Let $h_{\min}^2(i, j)$ denote the minimum variance at node (i, j).

- Initially, $h_{\max}^2(0, 0) = h_{\min}^2(0, 0) = h_0^2$.

- The resulting three-day tree is depicted on p. 697.

y_t

4.65752

4.64705

4.63658

4.62611

4.61564

0.01047

4.60517

4.59470

4.58423

4.57376

13.4809	
13.4809	

12.2883	2
12.2883	2

11.7170	
11.7170	

10.5733	
12.2846	

10.9645	2
10.9645	2

10.5256	1
10.5697	1

10.1305	
13.4644	

10.9600	1
10.9600	1

10.5215	1
10.5215	1

10.1269	1
10.9603	2

09.7717	
10.6042	

10.9553	1
10.9553	1

10.5173	1
12.2700	2

10.1231	
11.7005	

10.9511	1
10.9511	1

10.5135	
12.2662	

10.9473	
13.4438	

A top (bottom) number inside a gray box refers to the minimum (maximum, respectively) variance $h_{\min}^2$ ($h_{\max}^2$, respectively) for the node. Variances are multiplied by 100,000 for readability. A top (bottom) number inside a white box refers to η corresponding to $h_{\min}^2$ ($h_{\max}^2$, respectively).

Numerical Examples (continued)

- Let us see how the numbers are calculated.

- Start with the root node, node $(0, 0)$.

- Try $\eta = 1$ in Eqs. (71)–(73) on p. 682 first to obtain

$$
\begin{aligned}
p_u &= 0.4974, \\
p_m &= 0, \\
p_d &= 0.5026.
\end{aligned}
$$

- As they are valid probabilities, the three branches from the root node use single jumps.

Numerical Examples (continued)

- Move on to node $(1, 1)$.

- It has one predecessor node—node $(0, 0)$—and it takes an up move to reach the current node.

- So apply updating rule (75) on p. 688 with $\ell = 1$ and $h_t^2 = h^2(0, 0)$.

- The result is $h^2(1, 1) = 0.000109645$.

Numerical Examples (continued)

- Because $\lceil h(1,1)/\gamma \rceil = 2$, we try $\eta = 2$ in Eqs. (71)–(73) on p. 682 first to obtain

$$
\begin{aligned}
p_u &= 0.1237, \\
p_m &= 0.7499, \\
p_d &= 0.1264.
\end{aligned}
$$

- As they are valid probabilities, the three branches from node $(1, 1)$ use double jumps.

Numerical Examples (continued)

- Carry out similar calculations for node $(1, 0)$ with $\ell = 0$ in updating rule (75) on p. 688.

- Carry out similar calculations for node $(1, -1)$ with $\ell = -1$ in updating rule (75).

- Single jump $\eta = 1$ works in both nodes.

- The resulting variances are

$$
\begin{aligned}
h^2(1, 0) &= 0.000105215, \\
h^2(1, -1) &= 0.000109553.
\end{aligned}
$$

Numerical Examples (continued)

- Node $(2, 0)$ has 2 predecessor nodes, $(1, 0)$ and $(1, -1)$.

- Both have to be considered in deriving the variances.

- Let us start with node $(1, 0)$.

- Because it takes a middle move to reach the current node, we apply updating rule (75) on p. 688 with $\ell = 0$ and $h_t^2 = h^2(1, 0)$.

- The result is $h_{t+1}^2 = 0.000101269$.

Numerical Examples (continued)

- Now move on to the other predecessor node $(1, -1)$.

- Because it takes an up move to reach the current node, apply updating rule (75) on p. 688 with $\ell = 1$ and $h_t^2 = h^2(1, -1)$.

- The result is $h_{t+1}^2 = 0.000109603$.

- We hence record

$$
\begin{aligned}
h_{\min}^2(2, 0) &= 0.000101269, \\
h_{\max}^2(2, 0) &= 0.000109603.
\end{aligned}
$$

Numerical Examples (continued)

- Consider state $h_{\max}^2(2,0)$ first.

- Because $\lceil h_{\max}(2,0)/\gamma \rceil = 2$, we first try $\eta = 2$ in Eqs. (71)–(73) on p. 682 to obtain

$$
\begin{aligned}
p_u &= 0.1237, \\
p_m &= 0.7500, \\
p_d &= 0.1263.
\end{aligned}
$$

- As they are valid probabilities, the three branches from node $(2,0)$ with the maximum variance use double jumps.

Numerical Examples (continued)

- Now consider state $h_{\min}^2(2,0)$.

- Because $\lceil h_{\min}(2,0)/\gamma \rceil = 1$, we first try $\eta = 1$ in Eqs. (71)–(73) on p. 682 to obtain

$$
\begin{aligned}
p_u &= 0.4596, \\
p_m &= 0.0760, \\
p_d &= 0.4644.
\end{aligned}
$$

- As they are valid probabilities, the three branches from node $(2,0)$ with the minimum variance use single jumps.

Numerical Examples (continued)

- Node $(2, -1)$ has 3 predecessor nodes.

- Start with node $(1, 1)$.

- Because it takes a down move to reach the current node, we apply updating rule (75) on p. 688 with $\ell = -1$ and $h_t^2 = h^2(1, 1)$.

- The result is $h_{t+1}^2 = 0.0001227$.

Numerical Examples (continued)

- Now move on to predecessor node $(1, 0)$.

- Because it also takes a down move to reach the current node, we apply updating rule (75) on p. 688 with $\ell = -1$ and $h_t^2 = h^2(1, 0)$.

- The result is $h_{t+1}^2 = 0.000105609$.

Numerical Examples (continued)

- Finally, consider predecessor node $(1, -1)$.

- Because it takes a middle move to reach the current node, we apply updating rule (75) on p. 688 with $\ell = 0$ and $h_t^2 = h^2(1, -1)$.

- The result is $h_{t+1}^2 = 0.000105173$.

- We hence record

$$
\begin{aligned}
h_{\min}^2(2, -1) &= 0.000105173, \\
h_{\max}^2(2, -1) &= 0.0001227.
\end{aligned}
$$

Numerical Examples (continued)

- Consider state $h_{\max}^2(2, -1)$.

- Because $\lceil h_{\max}(2, -1)/\gamma \rceil = 2$, we first try $\eta = 2$ in Eqs. (71)–(73) on p. 682 to obtain

$$
\begin{aligned}
p_u &= 0.1385, \\
p_m &= 0.7201, \\
p_d &= 0.1414.
\end{aligned}
$$

- As they are valid probabilities, the three branches from node $(2, -1)$ with the maximum variance use double jumps.

Numerical Examples (continued)

- Next, consider state $h_{\min}^2(2, -1)$.

- Because $\lceil h_{\min}(2, -1)/\gamma \rceil = 1$, we first try $\eta = 1$ in Eqs. (71)–(73) on p. 682 to obtain

$$
\begin{aligned}
p_u &= 0.4773, \\
p_m &= 0.0404, \\
p_d &= 0.4823.
\end{aligned}
$$

- As they are valid probabilities, the three branches from node $(2, -1)$ with the minimum variance use single jumps.

Numerical Examples (concluded)

- Other nodes at dates 2 and 3 can be handled similarly.

- In general, if a node has k predecessor nodes, then $2k$ variances will be calculated using the updating rule.

  - This is because each predecessor node keeps two variance numbers.

- But only the maximum and minimum variances will be kept.

Negative Aspects of the RT Algorithm Revisited[a]

- Recall the problems mentioned on p. 694.

- In our case, combinatorial explosion occurs when

$$n > \frac{1 - \beta_1}{\beta_2} = \frac{1 - 0.9}{0.04} = 2.5.$$

- Suppose we are willing to accept the exponential running time and pick $n = 100$ to seek accuracy.

- But the problem of shortened maturity forces the tree to stop at date 9!

---

[a]Lyuu and Wu (2003).
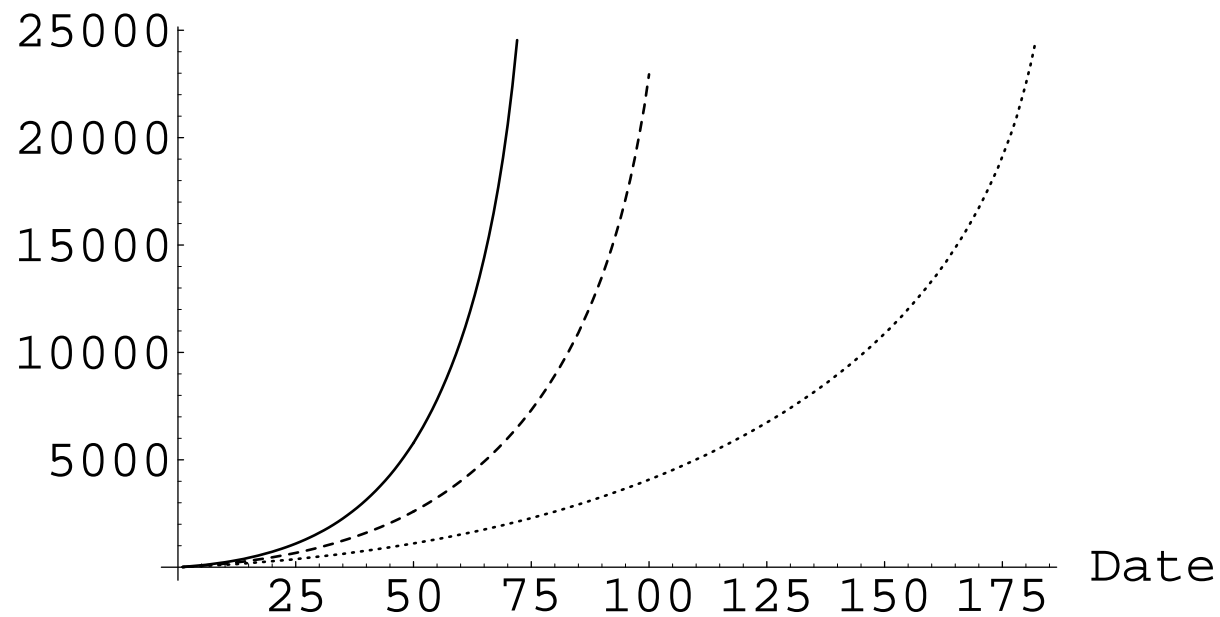
Dotted line: $n = 3$; dashed line: $n = 4$; solid line: $n = 5$.

Backward Induction on the RT Tree

- After the RT tree is constructed, it can be used to price options by backward induction.

- Recall that each node keeps two variances $h_{\max}^2$ and $h_{\min}^2$.

- We now increase that number to K equally spaced variances between $h_{\max}^2$ and $h_{\min}^2$ at each node.

- Besides the minimum and maximum variances, the other $K - 2$ variances in between are linearly interpolated.[a]

---

[a]In practice, log-linear interpolation works better (Lyuu and Wu (2005)). Log-cubic interpolation works even better (Liu (2005)).

Backward Induction on the RT Tree (continued)

- For example, if $K = 3$, then a variance of 10.5436×10^{-6} will be added between the maximum and minimum variances at node $(2, 0)$ on p. 697.

- In general, the kth variance at node (i, j) is

$$h_{\min}^2(i, j) + k \, \frac{h_{\max}^2(i, j) - h_{\min}^2(i, j)}{K - 1},$$

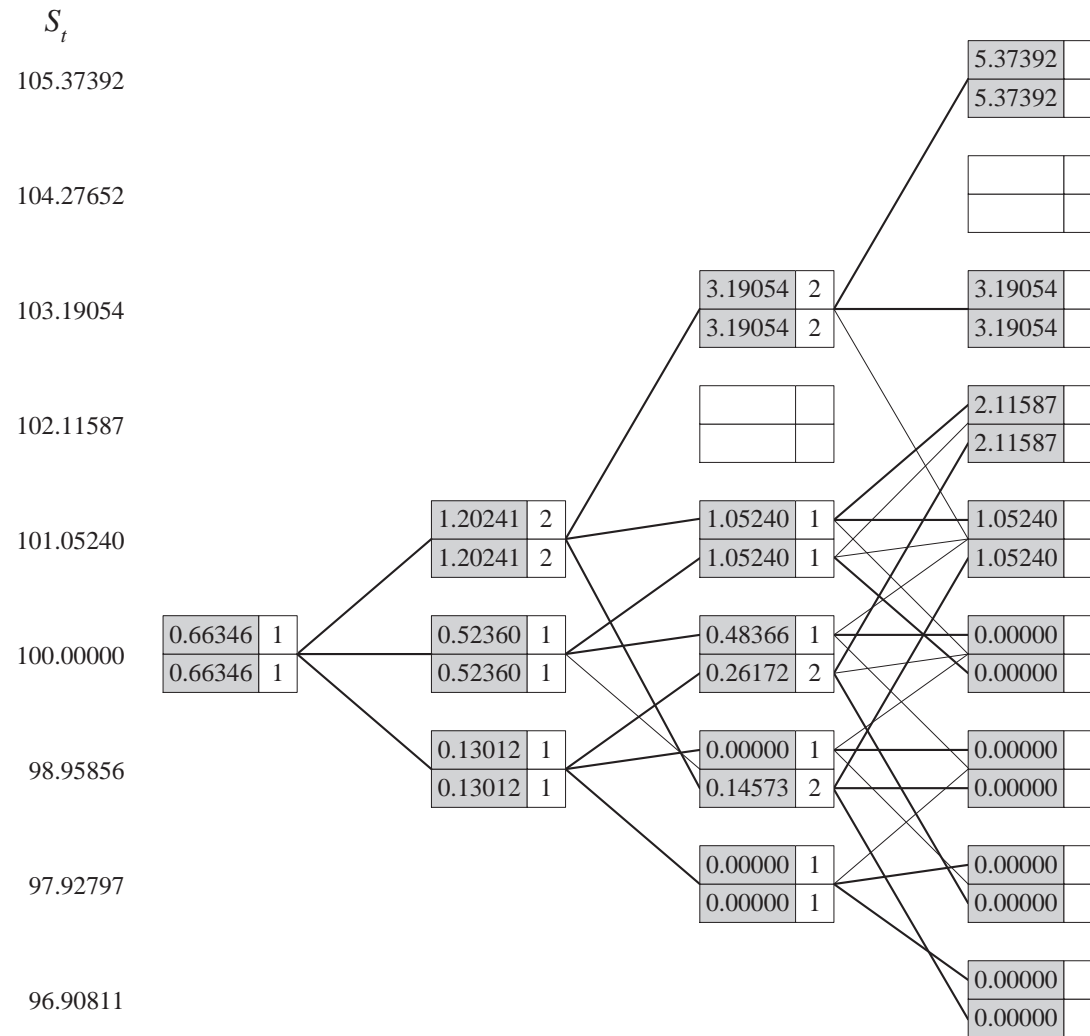$k = 0, 1, \ldots, K - 1.$

- Each interpolated variance's jump parameter and branching probabilities can be computed as before.

Backward Induction on the RT Tree (concluded)

- During backward induction, if a variance falls between two of the K variances, linear interpolation of the option prices corresponding to the two bracketing variances will be used as the approximate option price.

- The above ideas are reminiscent of the ones on p. 332, where we dealt with arithmetic average-rate options.

Numerical Examples

- We next use the numerical example on p. 697 to price a European call option with a strike price of 100 and expiring at date 3.

- Recall that the riskless interest rate is zero.

- Assume $K = 2$; hence there are no interpolated variances.

- The pricing tree is shown on p. 719 with a call price of 0.66346.

  - The branching probabilities needed in backward induction can be found on p. 720.

S_t

105.37392

104.27652

103.19054

102.11587

101.05240

100.00000

98.95856

97.92797

96.90811

0.66346	1
0.66346	1

1.20241	2
1.20241	2

0.52360	1
0.52360	1

0.13012	1
0.13012	1

3.19054	2
3.19054	2

1.05240	1
1.05240	1

0.48366	1
0.26172	2

0.00000	1
0.14573	2

0.00000	1
0.00000	1

5.37392	
5.37392	

3.19054	
3.19054	

2.11587	
2.11587	

1.05240	
1.05240	

0.00000	
0.00000	

0.00000	
0.00000	

0.00000	
0.00000	

0.00000	
0.00000	

rb[i][0]
rb[i][1]

rb[0][ ]	rb[1][ ]	rb[2][ ]	rb[3][ ]
0	−1	−2	−3
0	1	3	5

η[i][j][0]
η[i][j][1]

η[2][ ][ ] j

η[1][ ][ ]

η[0][ ][ ]

			j
		2	3
		2	
	2	1	2
	2	1	1
1	1	1	0
1	1	2	
	1	1	−1
	1	2	
		1	−2
		1	

$h^2[i][j][0]$
$h^2[i][j][1]$

$h^2[3][ ][ ]$ j

$h^2[2][ ][ ]$

$h^2[1][ ][ ]$

$h^2[0][ ][ ]$

			j
		13.4809	5
		13.4809	
			4
	12.2883	11.7170	3
	12.2883	11.7170	
		10.5733	2
		12.2846	
10.9645	10.5256	10.1305	1
10.9645	10.5697	13.4644	
10.9600	10.5215	10.1269	09.7717	0
10.9600	10.5215	10.9603	10.6042
	10.9553	10.5173	10.1231	−1
	10.9553	12.2700	11.7005
	10.9511	10.5135	−2
	10.9511	12.2662	
	10.9473		−3
	13.4438		

$p[i][j][0][1]$ $p[i][j][1][1]$
$p[i][j][0][0]$ $p[i][j][1][0]$
$p[i][j][0][-1]$ $p[i][j][1][-1]$

$p[2][ ][ ][ ]$ j

$p[1][ ][ ][ ]$

$p[0][ ][ ][ ]$

			j
		0.1387 0.1387	3
		0.7197 0.7197	
		0.1416 0.1416	
			2
	0.1237 0.1237	0.4777 0.4797	1
	0.7499 0.7499	0.0396 0.0356	
	0.1264 0.1264	0.4827 0.4847	
0.4974 0.4974	0.4775 0.4775	0.4596 0.1237	0
0.0000 0.0000	0.0400 0.0400	0.0760 0.7500	
0.5026 0.5026	0.4825 0.4825	0.4644 0.1263	
	0.4972 0.4972	0.4773 0.1385	−1
	0.0004 0.0004	0.0404 0.7201	
	0.5024 0.5024	0.4823 0.1414	
	0.4970 0.4970		−2
	0.0008 0.0008		
	0.5022 0.5022		

Numerical Examples (continued)

- Let us derive some of the numbers on p. 719.

- The option price for a terminal node at date 3 equals $\max(S_3 - 100, 0)$, independent of the variance level.

- Now move on to nodes at date 2.

- The option price at node $(2, 3)$ depends on those at nodes $(3, 5)$, $(3, 3)$, and $(3, 1)$.

- It therefore equals

$$0.1387 \times 5.37392 + 0.7197 \times 3.19054 + 0.1416 \times 1.05240 = 3.19054.$$

- Option prices for other nodes at date 2 can be computed similarly.

Numerical Examples (continued)

- For node $(1,1)$, the option price for both variances is

  $$0.1237 \times 3.19054 + 0.7499 \times 1.05240 + 0.1264 \times 0.14573 = 1.20241.$$

- Node $(1,0)$ is most interesting.

- We knew that a down move from it gives a variance of 0.000105609.

- This number falls between the minimum variance 0.000105173 and the maximum variance 0.0001227 at node $(2,-1)$ on p. 697.

Numerical Examples (continued)

- The option price corresponding to the minimum variance is 0.

- The option price corresponding to the maximum variance is 0.14573.

- The equation

$$x \times 0.000105173 + (1 - x) \times 0.0001227 = 0.000105609$$

  is satisfied by $x = 0.9751$.

- So the option for the down state is approximated by

$$x \times 0 + (1 - x) \times 0.14573 = 0.00362.$$

Numerical Examples (continued)

- The up move leads to the state with option price 1.05240.

- The middle move leads to the state with option price 0.48366.

- The option price at node $(1, 0)$ is finally calculated as

$$0.4775 \times 1.05240 + 0.0400 \times 0.48366 + 0.4825 \times 0.00362 = 0.52360.$$

Numerical Examples (concluded)

- It is possible for some of the three variances following an interpolated variance to exceed the maximum variance or be exceeded by the minimum variance.

- When this happens, the option price corresponding to the maximum or minimum variance will be used during backward induction.

- An interpolated variance may choose a branch that goes into a node that is *not* reached in the forward-induction tree-building phase.[a]

- In this case, the algorithm fails.

---

[a]Lyuu and Wu (2005).

Interest Rate Derivative Securities

What you are, you are only by contracts.
— Richard Wagner (1813–1883),
Der Ring des Nibelungen

Which shows that gambling's not a sin
provided that you always win.
— Roald Dahl (1916–1990),
"Snow-White and the Seven Dwarfs"

Term Structure Fitting

That's an old besetting sin;

they think calculating is inventing.

— Johann Wolfgang Goethe (1749–1832)

Introduction to Term Structure Modeling

The fox often ran to the hole
by which they had come in,
to find out if his body was still thin enough
to slip through it.
— *Grimm's Fairy Tales*

Outline

- Use the binomial interest rate tree to model stochastic term structure.

  - Illustrates the basic ideas underlying future models.

  - Applications are generic in that pricing and hedging methodologies can be easily adapted to other models.

- Although the idea is similar to the earlier one used in option pricing, the current task is more complicated.

  - The evolution of an entire term structure, not just a single stock price, is to be modeled.

  - Interest rates of various maturities cannot evolve arbitrarily or arbitrage profits may occur.

Issues

- A stochastic interest rate model performs two tasks.

  - Provides a stochastic process that defines future term structures without arbitrage profits.

  - "Consistent" with the observed term structures.

- The unbiased expectations theory, the liquidity preference theory, and the market segmentation theory can all be made consistent with the model.

History

- Methodology founded by Merton (1970).

- Modern interest rate modeling is often traced to 1977 when Vasicek and Cox, Ingersoll, and Ross developed simultaneously their influential models.

- Early models have fitting problems because they may not price today's benchmark bonds correctly.

- An alternative approach pioneered by Ho and Lee (1986) makes fitting the market yield curve mandatory.

- Models based on such a paradigm are called (somewhat misleadingly) arbitrage-free or no-arbitrage models.

Binomial Interest Rate Tree

- Goal is to construct a no-arbitrage interest rate tree consistent with the yields and/or yield volatilities of zero-coupon bonds of all maturities.

  - This procedure is called calibration.[a]

- Pick a binomial tree model in which the logarithm of the future short rate obeys the binomial distribution.

  - Exactly like the CRR tree.

- The limiting distribution of the short rate at any future time is hence lognormal.

---

[a]Derman (2004), "complexity without calibration is pointless."

Binomial Interest Rate Tree (continued)

- A binomial tree of future short rates is constructed.

- Every short rate is followed by two short rates in the following period (see next page).

- In the figure on p. 737 node A coincides with the start of period j during which the short rate r is in effect.

0.5　B　r_ℓ

r　A

0.5　C　r_h

| period $j-1$ | period j | period $j+1$ |
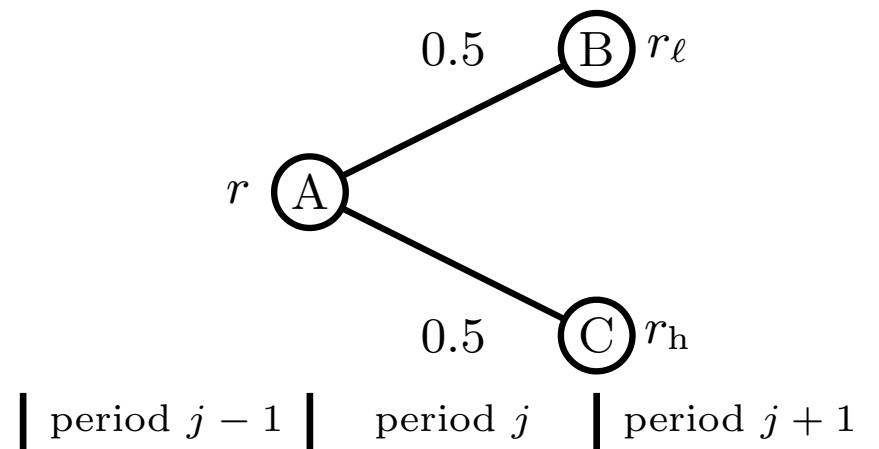
Binomial Interest Rate Tree (continued)

- At the conclusion of period j, a new short rate goes into effect for period $j + 1$.

- This may take one of two possible values:
  - r_ℓ: the "low" short-rate outcome at node B.
  - r_h: the "high" short-rate outcome at node C.

- Each branch has a fifty percent chance of occurring in a risk-neutral economy.

Binomial Interest Rate Tree (continued)

- We shall require that the paths combine as the binomial process unfolds.

- The short rate r can go to r_h and r_ℓ with equal risk-neutral probability $1/2$ in a period of length Δt.

- Hence the volatility of $\ln r$ after Δt time is

$$\sigma = \frac{1}{2} \frac{1}{\sqrt{\Delta t}} \ln\left(\frac{r_\mathrm{h}}{r_\ell}\right)$$

  (see Exercise 23.2.3 in text).

- Above, σ is annualized, whereas r_ℓ and r_h are period based.

Binomial Interest Rate Tree (continued)

- Note that

$$\frac{r_{\mathrm{h}}}{r_\ell} = e^{2\sigma\sqrt{\Delta t}}.$$

- Thus greater volatility, hence uncertainty, leads to larger r_{h}/r_ℓ and wider ranges of possible short rates.

- The ratio r_{h}/r_ℓ may depend on time if the volatility is a function of time.

- Note that r_{h}/r_ℓ has nothing to do with the current short rate r if σ is independent of r.

Binomial Interest Rate Tree (continued)

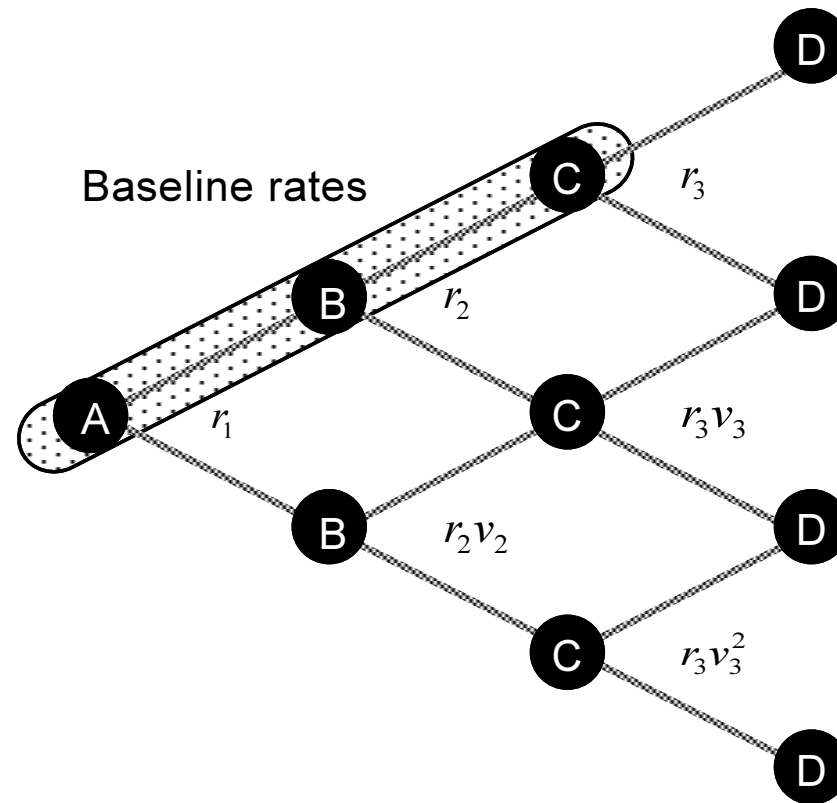- In general there are j possible rates in period j,

$$r_j, \, r_j v_j, \, r_j v_j^2, \, \ldots, \, r_j v_j^{j-1},$$

where

$$v_j \equiv e^{2\sigma_j \sqrt{\Delta t}} \tag{76}$$

is the multiplicative ratio for the rates in period j (see figure on next page).

- We shall call r_j the baseline rates.

- The subscript j in σ_j is meant to emphasize that the short rate volatility may be time dependent.

Binomial Interest Rate Tree (concluded)

- In the limit, the short rate follows the following process,

$$r(t) = \mu(t)\, e^{\sigma(t)\, W(t)}, \tag{77}$$

  in which the (percent) short rate volatility $\sigma(t)$ is a deterministic function of time.

- As the expected value of $r(t)$ equals $\mu(t)\, e^{\sigma(t)^2 (t/2)}$, a declining short rate volatility is usually imposed to preclude the short rate from assuming implausibly high values.

- Incidentally, this is how the binomial interest rate tree achieves mean reversion.

Memory Issues

- Path independency: The term structure at any node is independent of the path taken to reach it.

- So only the baseline rates r_i and the multiplicative ratios v_i need to be stored in computer memory.

- This takes up only $O(n)$ space.[a]

- Storing the whole tree would have taken up $O(n^2)$ space.

  - Daily interest rate movements for 30 years require roughly $(30 \times 365)^2/2 \approx 6 \times 10^7$ double-precision floating-point numbers (half a gigabyte!).

---

[a]Throughout this chapter, n denotes the depth of the tree.

Set Things in Motion

- The abstract process is now in place.

- Now need the annualized rates of return associated with the various riskless bonds that make up the benchmark yield curve and their volatilities.

- In the U.S., for example, the on-the-run yield curve obtained by the most recently issued Treasury securities may be used as the benchmark curve.

Set Things in Motion (concluded)

- The term structure of (yield) volatilities[a] can be estimated from either the historical data (historical volatility) or interest rate option prices such as cap prices (implied volatility).

- The binomial tree should be consistent with both term structures.

- Here we focus on the term structure of interest rates.

---

[a]Or simply the volatility (term) structure.

Model Term Structures

- The model price is computed by backward induction.

- Refer back to the figure on p. 737.

- Given that the values at nodes B and C are P_{B} and P_{C}, respectively, the value at node A is then

$$\frac{P_{\mathrm{B}} + P_{\mathrm{C}}}{2(1+r)} + \text{cash flow at node A}.$$

- We compute the values column by column without explicitly expanding the binomial interest rate tree (see figure next page).

- This takes quadratic time and linear space.

Cash flows:

$$C$$
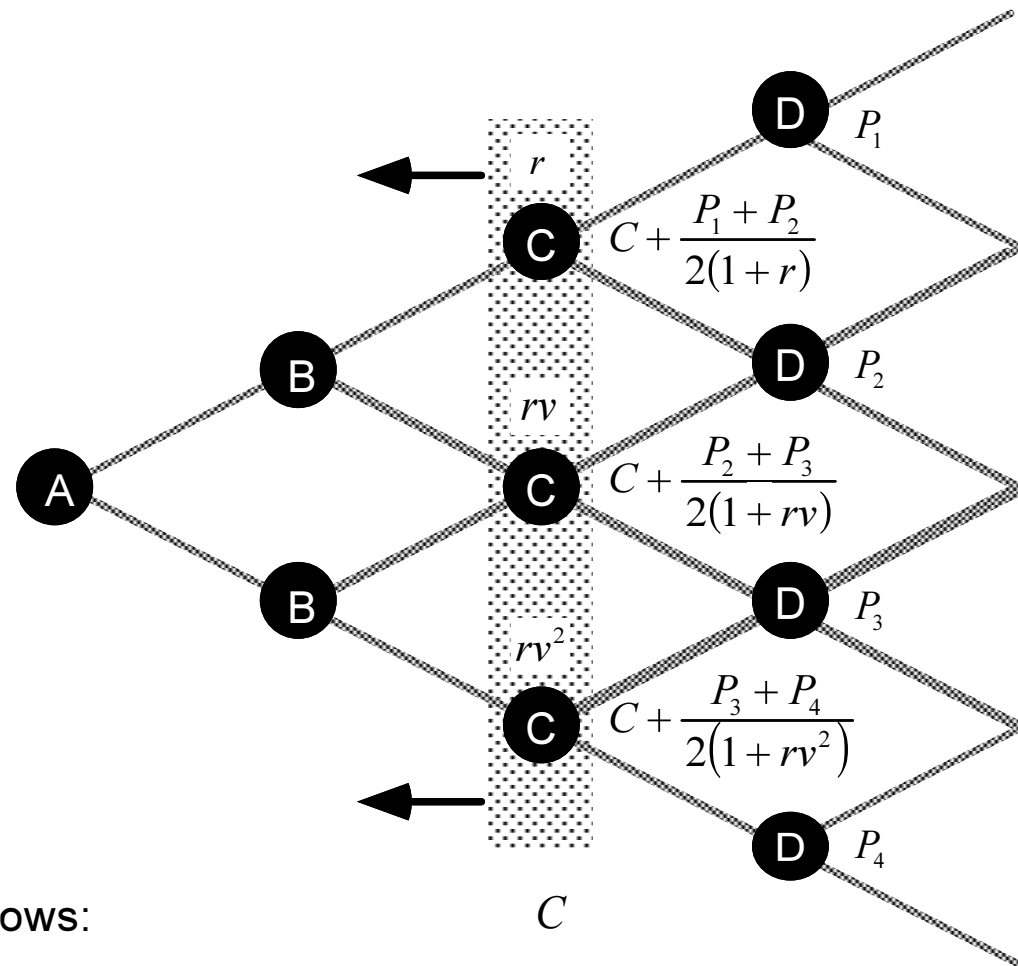
Term Structure Dynamics

- An n-period zero-coupon bond's price can be computed by assigning \$1 to every node at period n and then applying backward induction.

- Repeating this step for $n = 1, 2, \ldots$, one obtains the market discount function implied by the tree.

- The tree therefore determines a term structure.

- It also contains a term structure dynamics.

  - Taking any node in the tree as the current state induces a binomial interest rate tree and, again, a term structure.