Control Variates

- Use the analytic solution of a "similar" yet "simpler" problem to improve the solution.
- Suppose we want to estimate $E[X]$ and there exists a random variable Y with a known mean $\mu \stackrel{\Delta}{=} E[Y]$.
- Then $W \stackrel{\Delta}{=} X + \beta(Y - \mu)$ can serve as a "controlled" estimator of $E[X]$ for any constant β.
  - However β is chosen, W remains an unbiased estimator of $E[X]$ as

$$E[W] = E[X] + \beta E[Y - \mu] = E[X].$$

Control Variates (continued)

- Note that

$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] + \beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,], \tag{115}$$

- Hence W is less variable than X if and only if

$$\beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,] < 0. \tag{116}$$

Control Variates (concluded)

- The success of the scheme clearly depends on both β and the choice of Y.
  - American options can be priced by choosing Y to be the otherwise identical European option and μ the Black-Scholes formula.[a]
  - Arithmetic Asian options can be priced by choosing Y to be the otherwise identical geometric Asian option's price and $\beta = -1$.
- This approach is much more effective than the antithetic-variates method.[b]

[a] Hull & White (1988).
[b] Boyle, Broadie, & Glasserman (1997).

Choice of Y

- In general, the choice of Y is ad hoc,[a] and experiments must be performed to confirm the wisdom of the choice.
- Try to match calls with calls and puts with puts.[b]
- On many occasions, Y is a discretized version of the derivative that gives μ.
  - Discretely monitored geometric Asian option vs. the continuously monitored version.[c]
- The discrepancy can be large (e.g., lookback options).[d]

---

[a] But see Dai (B82506025, R86526008, D8852600), Chiu (R94922072), & Lyuu (2015, 2018).

[b] Contributed by Ms. Teng, Huei-Wen (R91723054) on May 25, 2004.

[c] Priced by formulas (53) on p. 424.

[d] Contributed by Mr. Tsai, Hwai (R92723049) on May 12, 2004.

Optimal Choice of β

- Equation (115) on p. 856 is minimized when

$$\beta = -\mathrm{Cov}[\,X, Y\,]/\mathrm{Var}[\,Y\,].$$

  - It is called beta in the book.

- For this specific β,

$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] - \frac{\mathrm{Cov}[\,X, Y\,]^2}{\mathrm{Var}[\,Y\,]} = \left(1 - \rho_{X,Y}^2\right) \mathrm{Var}[\,X\,],$$

  where $\rho_{X,Y}$ is the correlation between X and Y.

Optimal Choice of β (continued)

- Note that the variance can never be increased with the optimal choice.
- Furthermore, the stronger X and Y are correlated, the greater the reduction in variance.
- For example, if this correlation is nearly perfect (± 1), we could control X almost exactly.

Optimal Choice of β (continued)

- Typically, neither $\mathrm{Var}[\,Y\,]$ nor $\mathrm{Cov}[\,X, Y\,]$ is known.
- Therefore, we cannot obtain the maximum reduction in variance.
- We can guess these values and hope that the resulting W does indeed have a smaller variance than X.
- A second possibility is to use the simulated data to estimate these quantities.
  – How to do it efficiently in terms of time and space?

Optimal Choice of β (concluded)

- Observe that $-\beta$ has the same sign as the correlation between X and Y.
- Hence, if X and Y are positively correlated, $\beta < 0$, then X is adjusted downward whenever $Y > \mu$ and upward otherwise.
- The opposite is true when X and Y are negatively correlated, in which case $\beta > 0$.
- Suppose a suboptimal $\beta + \epsilon$ is used instead.
- The variance increases by only $\epsilon^2 \text{Var}[Y]$.[a]

[a] Han & Y. Lai (2010).

A Pitfall

- A potential pitfall is to sample X and Y *independently*.
- In this case, $\mathrm{Cov}[\,X, Y\,] = 0$.
- Equation (115) on p. 856 becomes

$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] + \beta^2\, \mathrm{Var}[\,Y\,].$$

- So whatever Y is, the variance is *increased*!
- Lesson: X and Y must be correlated.

Problems with the Monte Carlo Method

- The error bound is only probabilistic.
- The probabilistic error bound of $O(1/\sqrt{N}\,)$ does not benefit from regularity of the integrand function.
- The requirement that the points be independent random samples are wasteful because of clustering.
- In reality, pseudorandom numbers generated by completely deterministic means are used.
- Monte Carlo simulation exhibits a great sensitivity on the seed of the pseudorandom-number generator.

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 $A \stackrel{\Delta}{=} [\, a_{ij} \,]_{1 \leq i \leq m, 1 \leq j \leq n}$, 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

$$A \stackrel{\Delta}{=} [\, a_{ij} \,]_{i,j}$$

  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

$$I \stackrel{\Delta}{=} \mathrm{diag}[\, 1, 1, \ldots, 1 \,].$$

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 Distribution

- Let $\boldsymbol{x} \stackrel{\Delta}{=} [\, x_1, x_2, \ldots, x_n \,]^{\mathrm{T}}$ be a vector random variable with a positive definite covariance matrix C.
- As usual, assume $E[\,\boldsymbol{x}\,] = \mathbf{0}$.
- This covariance structure can be matched by $P\boldsymbol{y}$.
  - $\boldsymbol{y} \stackrel{\Delta}{=} [\, y_1, y_2, \ldots, y_n \,]^{\mathrm{T}}$ is a vector random variable with a covariance matrix equal to the identity matrix.
  - $C = PP^{\mathrm{T}}$ is the Cholesky decomposition of C.[a]

---

[a]What if C is not positive definite? See Y. Y. Lai (`R93942114`) & Lyuu (2007).

Generation of Multivariate Distribution (concluded)

- For example, suppose
$$C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$
- Then
$$P = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$
as $PP^{\mathrm{T}} = C$.[a]

[a] Recall Eq. (27) on p. 174.

Generation of Multivariate Normal Distribution

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^{\mathrm{T}}$.
  - First, generate 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.
  - These steps can then be repeated.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives (pp. 772ff).
- 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); Chen (D95723006) & Lyuu (2009).

Multivariate Derivatives Pricing (concluded)

- Suppose $dS_j/S_j = r\,dt + \sigma_j\,dW_j$, $1 \leq j \leq k$, where C is the correlation matrix for $dW_1, dW_2, \ldots, dW_k$.
- Let $C = PP^{\mathrm{T}}$.
- Let ξ consist of k independent random variables from $N(0,1)$.
- Let $\xi' = P\xi$.
- Similar to Eq. (114) on p. 816,

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

Least-Squares Problems

- The least-squares (LS) problem is concerned with

$$\min_{x \in R^n} \| Ax - b \|,$$

where $A \in \boldsymbol{R}^{m \times n}$, $b \in \boldsymbol{R}^m$, and $m \geq n$.

- 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 p. 273 of the textbook 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 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.

[a] Longstaff & Schwartz (2001).

The Least-Squares Monte Carlo Approach (concluded)

- The LSM is provably convergent.[a]
- The LSM can be easily parallelized.[b]
  - Partition the paths into subproblems and perform LSM on each of them independently.
  - The speedup is close to linear (i.e., proportional to the number of cores).
- Surprisingly, accuracy is not affected.

---

[a] Clément, Lamberton, & Protter (2002); Stentoft (2004).

[b] K. Huang (B96902079, R00922018) (2013); C. W. Chen (B97902046, R01922005) (2014); C. W. Chen (B97902046, R01922005), K. Huang (B96902079, R00922018) & Lyuu (2015).

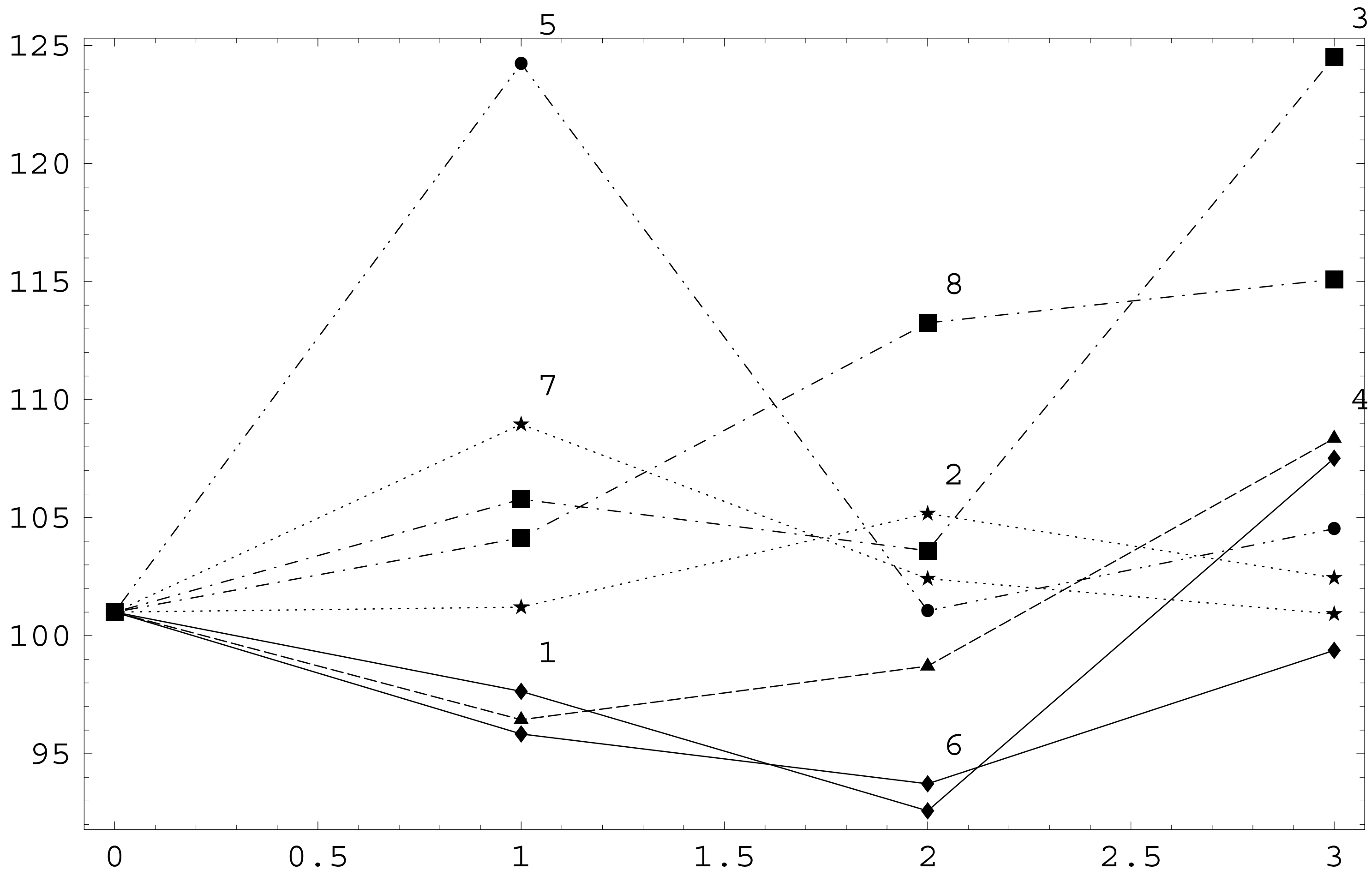
A Numerical Example

- 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 annual discount factor hence equals 0.951229.
- The current stock price is 101.
- 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



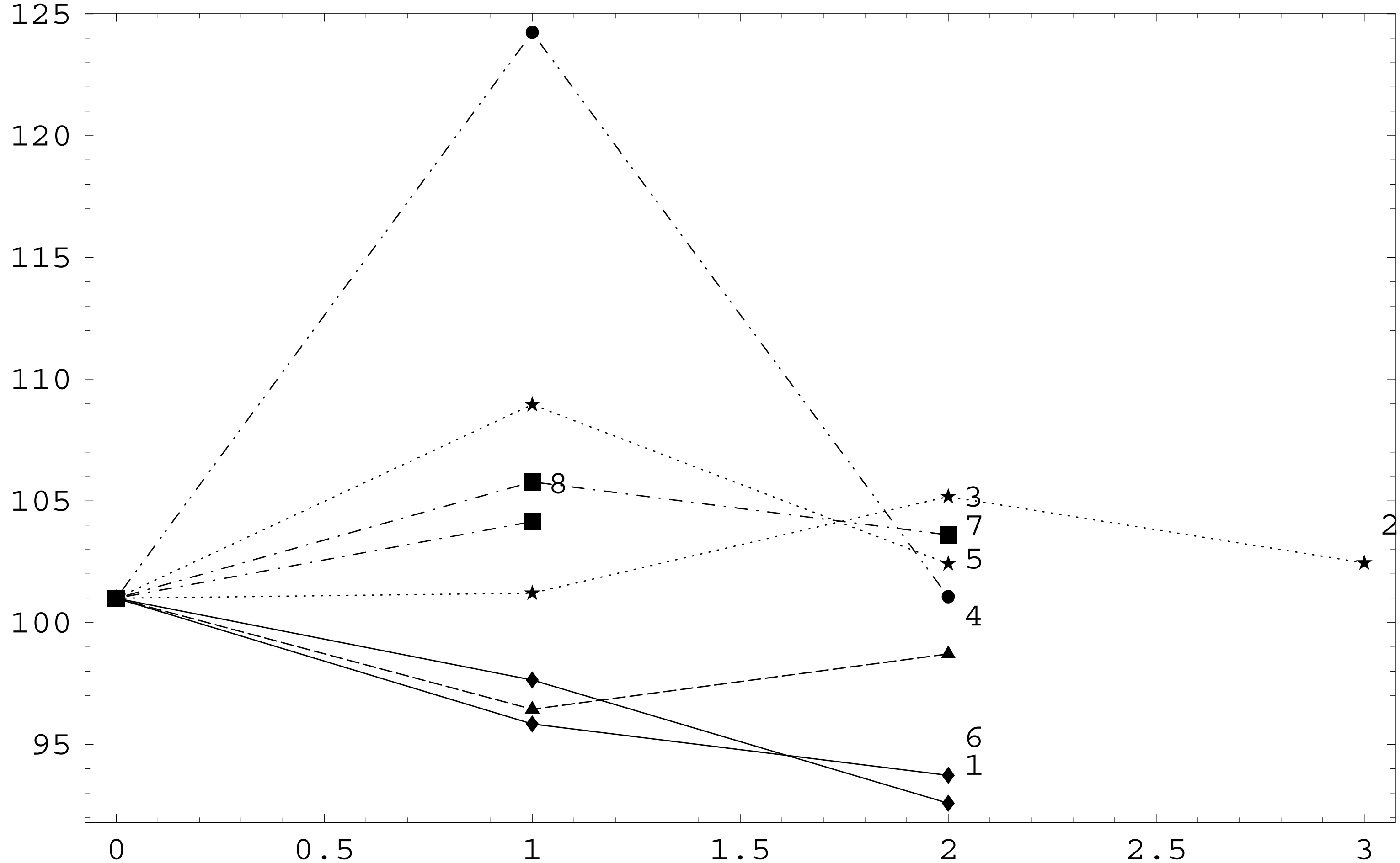
125
120
115
110
105
100
95
0
0.5
1
1.5
2
2.5
3
5
3
8
7
4
2
1
6

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.
- The idea 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.


125
120
115
110
105
100
95
0
0.5
1
1.5
2
2.5
3
8
3
7
5
2
4
6
1

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 (p. 882).
- 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

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

- $f(x)$ estimates the continuation value conditional on the stock price at year 2.
- We next compare the immediate exercise value and the continuation value.[a]

[a]The $f(102.4177)$ entry on the next page was corrected by Mr. Tu, Yung-Szu (B79503054, R83503086) on May 25, 2017.

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	$f(102.4177) = 1.2266$
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 or overridden for these paths as the put is exercised before year 3 (p. 882).
  - They are paths 5, 6, 7.
- The cash flows on p. 886 become the ones on next slide.

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 (p. 882).
- 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. 894, 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

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

- $f(x)$ 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.
  - Note that $f(104.1475) < 0$.
- Now, any positive future cash flow should be set to zero or overridden for this path.
  - But there is none.
- The cash flows on p. 894 become the ones on next slide.
- They also confirm the plot on p. 885.

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 901,

$$
\begin{aligned}
& (12.4185 \times 0.951229^2 + 2.5476 \times 0.951229^3 \\
& +1.3990 \times 0.951229^2 + 6.2880 \times 0.951229^2 \\
& +3.9436 \times 0.951229^2 + 11.2730 \times 0.951229^2 \\
& +2.5823 \times 0.951229^2 + 0.8525 \times 0.951229)/8 \\
= \; & 4.66263.
\end{aligned}
$$

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 with the European put's value of 1.3680 (p. 887).

Time Series Analysis

The historian is a prophet in reverse.
— Friedrich von Schlegel (1772–1829)

GARCH Option Pricing[a]

- 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.

---

[a]ARCH (autoregressive conditional heteroskedastic) is due to Engle (1982), co-winner of the 2003 Nobel Prize in Economic Sciences. GARCH (generalized ARCH) is due to Bollerslev (1986) and Taylor (1986). A Bloomberg quant said to me on Feb 29, 2008, that GARCH is seldom used in trading.

GARCH Option Pricing (continued)

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

$$\ln \frac{S_{t+1}}{S_t} = r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}, \tag{117}$$

where

$$\begin{aligned} h_{t+1}^2 &= \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2, && (118) \\ \epsilon_{t+1} &\sim N(0,1) \text{ given information at date } t, \\ r &= \text{daily riskless return}, \\ c &\geq 0. \end{aligned}$$

[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.
- There are other inequalities to satisfy (see text).
- The above process is called the nonlinear asymmetric GARCH (or NGARCH) model.

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 $y_t \stackrel{\Delta}{=} \ln S_t$ denoting the logarithmic price, the model becomes
$$y_{t+1} = y_t + r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}. \tag{119}$$
- The pair (y_t, h_t^2) completely describes the current state.
- The conditional mean and variance of y_{t+1} are clearly
$$E[\, y_{t+1} \,|\, y_t, h_t^2 \,] = y_t + r - \frac{h_t^2}{2}, \tag{120}$$
$$\mathrm{Var}[\, y_{t+1} \,|\, y_t, h_t^2 \,] = h_t^2. \tag{121}$$

GARCH Model: Inferences

- Suppose the parameters c, h_0, β_0, β_1, and β_2 are given.
- Then we can recover $h_1, h_2, \ldots, h_n$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ from the prices
$$S_0, S_1, \ldots, S_n$$
under the GARCH model (117) on p. 907.
- This property is useful in statistical inferences.

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 & Trevor (1999).

The RT 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, each state at date t is followed by $2n+1$ states at date $t+1$ (recall p. 703).
- These $2n+1$ values must approximate the distribution of (y_{t+1}, h_{t+1}^2).
- So the conditional moments (120)–(121) at date $t+1$ on p. 910 must be matched by the trinomial model to guarantee convergence to the continuous-state model.

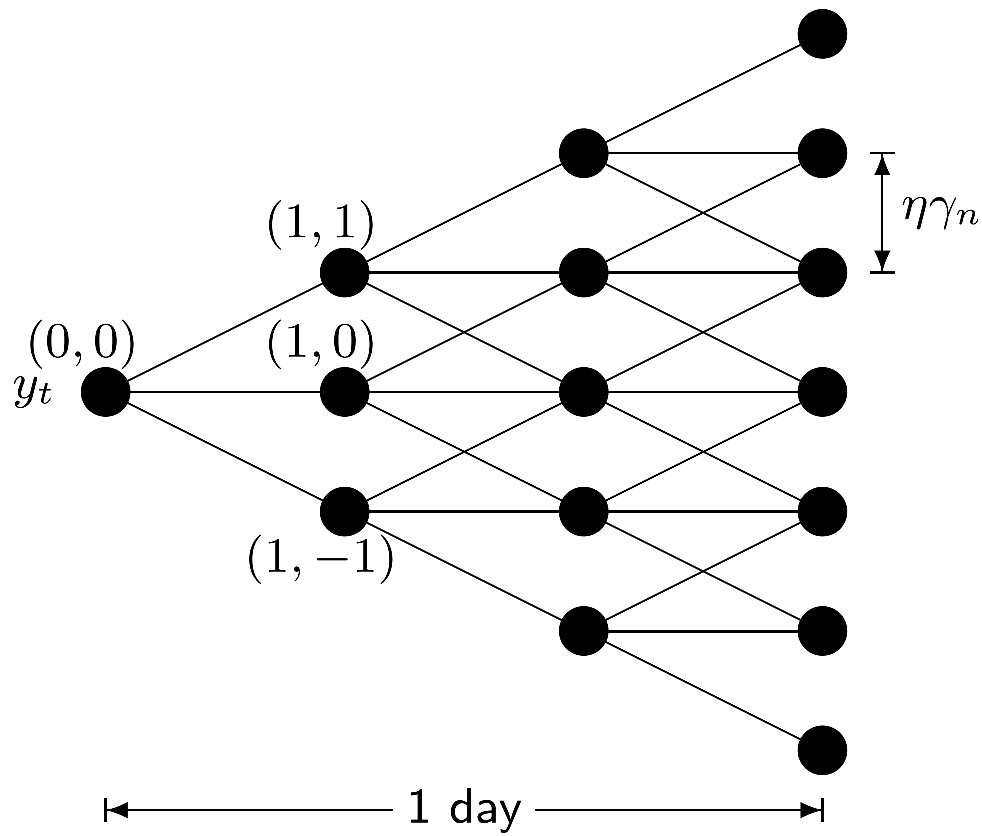
The RT Algorithm (continued)

- 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 $\gamma \stackrel{\Delta}{=} h_0$, though other multiples of h_0 are possible, and

$$\gamma_n \stackrel{\Delta}{=} \frac{\gamma}{\sqrt{n}}.$$

The RT Algorithm (continued)

- The jump size will be some integer multiple η of γ_n.
- We call η the jump parameter (see next page).
- Obviously, the magnitude of η grows with h_t.
- The middle branch does not change the underlying asset's price.



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

The RT Algorithm (continued)

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

$$p_u = \frac{h_t^2}{2\eta^2\gamma^2} + \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}, \tag{122}$$

$$p_m = 1 - \frac{h_t^2}{\eta^2\gamma^2}, \tag{123}$$

$$p_d = \frac{h_t^2}{2\eta^2\gamma^2} - \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}. \tag{124}$$

The RT 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 guarantees convergence as n increases.[a]

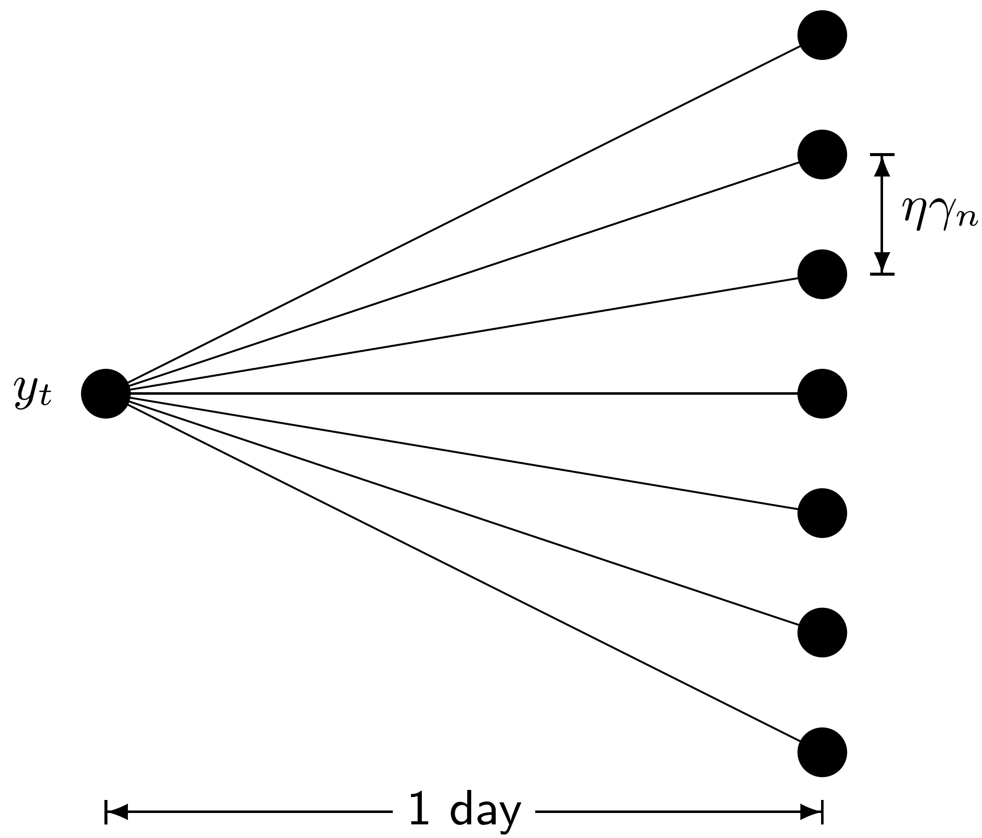
[a]Assume the probabilities are valid.

The RT Algorithm (continued)

- We can dispense with the intermediate nodes *between* dates to create a $(2n+1)$-nomial tree (p. 920).
- 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 n times larger.[a]

---

[a]Contrast it with the case on p. 391.



This heptanomial tree is the outcome of the trinomial tree on p. 916 after its intermediate nodes are removed.

The RT 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, where

$$-n \leq \ell \leq n.$$

- 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

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

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

The RT Algorithm (continued)

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

$$(p_u x + p_m + p_d x^{-1})^n = \sum_{\ell=-n}^{n} P(\ell)\, x^{\ell}. \tag{125}$$

  - Convince yourself that this trick does the "accounting" correctly.

- 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, if not less.

The RT Algorithm (continued)

- The updating rule (118) on p. 907 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) is associated with this variance:

$$h_{t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon'_{t+1} - c)^2, \qquad (126)$$

  – 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 RT Algorithm (continued)

- Different conditional variances h_t^2 may require different η so that the probabilities calculated by Eqs. (122)–(124) on p. 917 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.

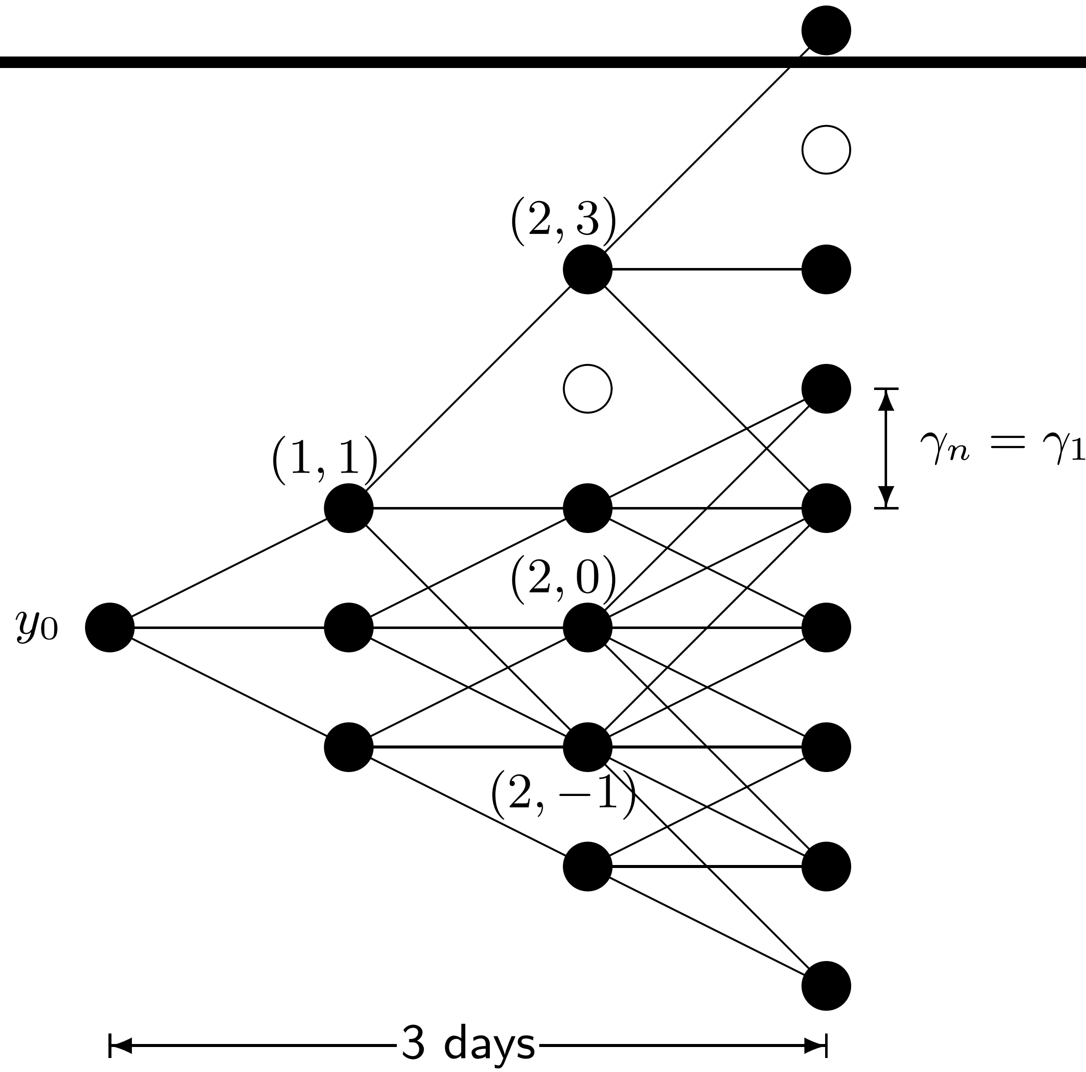
The RT Algorithm (continued)

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

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

- The plot on p. 926 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] C. Wu (`R90723065`) (2003); Lyuu & C. Wu (`R90723065`) (2003, 2005).



$(2,3)$
$(1,1)$
$\gamma_n = \gamma_1$
$(2,0)$
y_0
$(2,-1)$
3 days

The RT Algorithm (continued)

- The topology of the tree is not a standard combining multinomial tree.
- For example, a few nodes on p. 926 such as nodes $(2, 0)$ and $(2, -1)$ have *multiple* jump sizes.
- The reason is path dependency of the model.
  - Two paths can reach node $(2, 0)$ from the root node, each with a different variance for the node.
  - One variance results in $\eta = 1$.
  - The other results in $\eta = 2$.

The RT Algorithm (concluded)

- The number of possible values of h_t^2 at a node can be exponential.
  - Because each path brings a different variance h_t^2.
- 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 & Topyan (2000). But see p. 963 for a potential problem.

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 quite limited in practice.
- The RT algorithm can be modified to be free of shortened maturity and exponential complexity.[b]

---

[a] Lyuu & C. Wu (R90723065) (2003, 2005).

[b] Its size is only $O(n^2)$ if $n \leq (\sqrt{(1 - \beta_1)/\beta_2} - c)^2$!

Numerical Examples

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

Numerical Examples (continued)

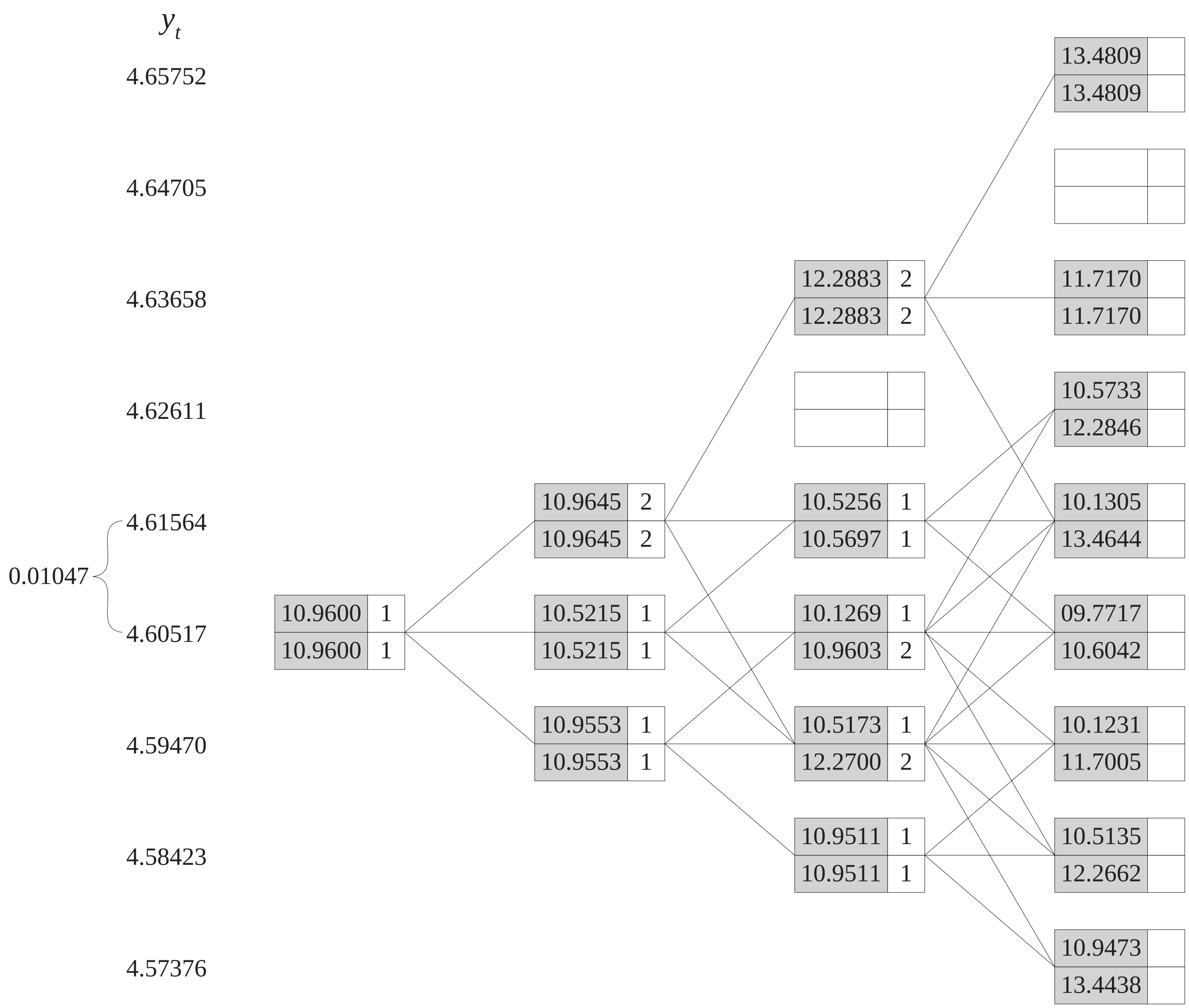
- A daily variance of 0.0001096 corresponds to an annual volatility of

$$\sqrt{365 \times 0.0001096} \approx 20\%.$$

- 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 $h^2_{\max}(i, j)$ denote the maximum variance at node (i, j).
- Let $h^2_{\min}(i, j)$ denote the minimum variance at node (i, j).
- Initially, $h^2_{\max}(0, 0) = h^2_{\min}(0, 0) = h^2_0$.
- The resulting 3-day tree is depicted on p. 933.


y_t
4.65752
4.64705
4.63658
4.62611
4.61564
0.01047
4.60517
4.59470
4.58423
4.57376
10.9600 1
10.9600 1
10.9645 2
10.9645 2
10.5215 1
10.5215 1
10.9553 1
10.9553 1
12.2883 2
12.2883 2
10.5256 1
10.5697 1
10.1269 1
10.9603 2
10.5173 1
12.2700 2
10.9511 1
10.9511 1
13.4809
13.4809
11.7170
11.7170
10.5733
12.2846
10.1305
13.4644
09.7717
10.6042
10.1231
11.7005
10.5135
12.2662
10.9473
13.4438

- A top number inside a gray box refers to the minimum variance $h^2_{\min}$ for the node.
- A bottom number inside a gray box refers to the maximum variance $h^2_{\max}$ for the node.
- Variances are multiplied by 100,000 for readability.
- The top number inside a white box refers to the η for $h^2_{\min}$.
- The bottom number inside a white box refers to the η for $h^2_{\max}$.

Numerical Examples (continued)

- Let us see how the numbers are calculated.
- Start with the root node, node $(0, 0)$.
- Try $\eta = 1$ in Eqs. (122)–(124) on p. 917 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 (126) on p. 923 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. (122)–(124) on p. 917 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 (126) on p. 923.
- Carry out similar calculations for node $(1, -1)$ with $\ell = -1$ in updating rule (126).
- Single jump $\eta = 1$ works for 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 (126) on p. 923 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 (126) on p. 923 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. (122)–(124) on p. 917 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 $h^2_{\min}(2,0)$.
- Because $\lceil h_{\min}(2,0)/\gamma \rceil = 1$, we first try $\eta = 1$ in Eqs. (122)–(124) on p. 917 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 *one* down move to reach the current node, we apply updating rule (126) on p. 923 with $\ell = -1$ and $h_t^2 = h^2(1, 1)$.[a]
- The result is $h_{t+1}^2 = 0.0001227$.

---

[a]Note that it is *not* $\ell = -2$. The reason is that $h(1, 1)$ has $\eta = 2$ (p. 937).

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 (126) on p. 923 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 (126) on p. 923 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 $h^2_{\max}(2,-1)$.
- Because $\lceil h_{\max}(2,-1)/\gamma \rceil = 2$, we first try $\eta = 2$ in Eqs. (122)–(124) on p. 917 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 $h^2_{\min}(2, -1)$.
- Because $\lceil h_{\min}(2, -1)/\gamma \rceil = 1$, we first try $\eta = 1$ in Eqs. (122)–(124) on p. 917 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 up to $2k$ 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]

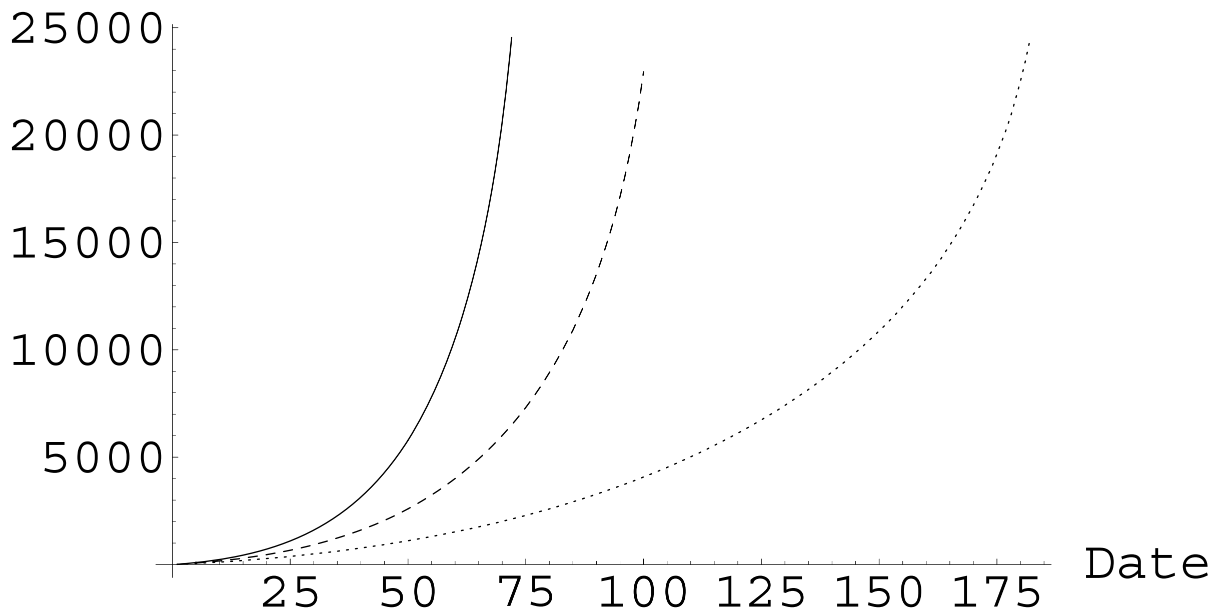
- Recall the problems mentioned on p. 929.
- In our case, combinatorial explosion occurs when

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

(see the next plot).

- 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 & C. Wu (R90723065) (2003, 2005).



Dotted line: $n = 3$; dashed line: $n = 4$; solid line: $n = 5$.