Variance Reduction: Antithetic Variates

• We are interested in estimating $E[g(X_1, X_2, \ldots, X_n)]$, where $X_1, X_2, \ldots, X_n$ are independent.

• Let $Y_1$ and $Y_2$ be random variables with the same distribution as $g(X_1, X_2, \ldots, X_n)$.

• Then
  \[
  \text{Var} \left[ \frac{Y_1 + Y_2}{2} \right] = \frac{\text{Var}[Y_1]}{2} + \frac{\text{Cov}[Y_1, Y_2]}{2}.
  \]

  – $\frac{\text{Var}[Y_1]}{2}$ is the variance of the Monte Carlo method with two (independent) replications.

• The variance $\text{Var}[\frac{(Y_1 + Y_2)}{2}]$ is smaller than $\frac{\text{Var}[Y_1]}{2}$ when $Y_1$ and $Y_2$ are negatively correlated.
Variance Reduction: Antithetic Variates (continued)

- For each simulated sample path $X$, a second one is obtained by reusing the random numbers on which the first path is based.
- This yields a second sample path $Y$.
- Two estimates are then obtained: One based on $X$ and the other on $Y$.
- If $N$ independent sample paths are generated, the antithetic-variates estimator averages over $2N$ estimates.
Variance Reduction: Antithetic Variates (continued)

• Consider process \( dX = a_t \, dt + b_t \sqrt{dt} \, \xi \).

• Let \( g \) be a function of \( n \) samples \( X_1, X_2, \ldots, X_n \) on the sample path.

• We are interested in \( E[g(X_1, X_2, \ldots, X_n)] \).

• Suppose one simulation run has realizations \( \xi_1, \xi_2, \ldots, \xi_n \) for the normally distributed fluctuation term \( \xi \).

• This generates samples \( x_1, x_2, \ldots, x_n \).

• The estimate is then \( g(\mathbf{x}) \), where \( \mathbf{x} \equiv (x_1, x_2 \ldots, x_n) \).
Variance Reduction: Antithetic Variates (concluded)

- The antithetic-variates method does not sample \( n \) more numbers from \( \xi \) for the second estimate \( g(x') \).
- Instead, generate the sample path \( x' \equiv (x'_1, x'_2, \ldots, x'_n) \) from \( -\xi_1, -\xi_2, \ldots, -\xi_n \).
- Compute \( g(x') \).
- Output \( (g(x) + g(x'))/2 \).
- Repeat the above steps for as many times as required by accuracy.
Variance Reduction: Conditioning

- We are interested in estimating $E[X]$.
- Suppose here is a random variable $Z$ such that $E[X | Z = z]$ can be efficiently and precisely computed.
- $E[X] = E[E[X | Z]]$ by the law of iterated conditional expectations.
- Hence the random variable $E[X | Z]$ is also an unbiased estimator of $E[X]$. 
Variance Reduction: Conditioning (concluded)

• As

\[ \text{Var}[E[X \mid Z]] \leq \text{Var}[X], \]

\[ E[X \mid Z] \] has a smaller variance than observing \( X \) directly.

• First obtain a random observation \( z \) on \( Z \).

• Then calculate \( E[X \mid Z = z] \) as our estimate.
  – There is no need to resort to simulation in computing \( E[X \mid Z = z] \).

• The procedure can be repeated a few times to reduce the variance.
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 $\mu \equiv E[Y]$.

- Then $W \equiv X + \beta(Y - \mu)$ can serve as a “controlled” estimator of $E[X]$ for any constant $\beta$.
  
  - However $\beta$ 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

\[ \text{Var}[W] = \text{Var}[X] + \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X,Y], \]

(64)

• Hence \( W \) is less variable than \( X \) if and only if

\[ \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X,Y] < 0. \]

(65)
Control Variates (concluded)

- The success of the scheme clearly depends on both $\beta$ and the choice of $Y$.
- For example, arithmetic average-rate options can be priced by choosing $Y$ to be the otherwise identical geometric average-rate option’s price and $\beta = -1$.
- This approach is much more effective than the antithetic-variates method.
Choice of $Y$

- In general, the choice of $Y$ is ad hoc, and experiments must be performed to confirm the wisdom of the choice.
- Try to match calls with calls and puts with puts.\(^a\)
- On many occasions, $Y$ is a discretized version of the derivative that gives $\mu$.
  - Discretely monitored geometric average-rate option vs. the continuously monitored geometric average-rate option given by formulas (29) on p. 332.
- For some choices, the discrepancy can be significant, such as the lookback option.\(^b\)

\(^a\)Contributed by Ms. Teng, Huei-Wen (R91723054) on May 25, 2004.
\(^b\)Contributed by Mr. Tsai, Hwai (R92723049) on May 12, 2004.
Optimal Choice of $\beta$

- Equation (64) on p. 632 is minimized when

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

which was called beta in the book.

- For this specific $\beta$,

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

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

- The stronger $X$ and $Y$ are correlated, the greater the reduction in variance.
Optimal Choice of $\beta$ (continued)

- For example, if this correlation is nearly perfect ($\pm 1$), we could control $X$ almost exactly.
- Typically, neither $\text{Var}[Y]$ nor $\text{Cov}[X,Y]$ 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.
Optimal Choice of $\beta$ (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 $\beta > 0$. 
Problems with the Monte Carlo Method

- The error bound is only probabilistic.
- The probabilistic error bound of $\sqrt{N}$ 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 $A \equiv [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \mathbb{R}^{m \times n}$, denote an $m \times n$ matrix.

- It can also be represented as $[a_1, a_2, \ldots, a_n]$ where $a_i \in \mathbb{R}^m$ 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 $A^T = A$.

- 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

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

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

for any nonzero vector $x$.

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

- Positive definite matrices can be factored as

\[ A = LL^T, \]

called the Cholesky decomposition.

- Above, \( L \) is a lower triangular matrix.
Generation of Multivariate Distribution

• Let \( \mathbf{x} \equiv [x_1, x_2, \ldots, x_n]^T \) be a vector random variable with a positive definite covariance matrix \( C \).

• As usual, assume \( E[\mathbf{x}] = 0 \).

• This distribution can be generated by \( P \mathbf{y} \).
  
  - \( C = P P^T \) is the Cholesky decomposition of \( C \).\(^a\)
  
  - \( \mathbf{y} \equiv [y_1, y_2, \ldots, y_n]^T \) 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

• Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^T$.

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

• Then $P[y_1, y_2, \ldots, y_n]^T$ has the desired distribution.
Multivariate Derivatives Pricing

• Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives (p. 566).

• 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 \( \frac{dS_j}{S_j} = r \, dt + \sigma_j \, dW_j \), \( 1 \leq j \leq n \), where \( C \) is the correlation matrix for \( dW_1, dW_2, \ldots , dW_k \).

• Let \( C = PP^T \).

• Let \( \xi \) consist of \( k \) independent random variables from \( N(0, 1) \).

• Let \( \xi' = P\xi \).

• Similar to Eq. (63) on p. 606,

\[
S_{i+1} = S_i e^{(r - \sigma_j^2/2) \Delta t + \sigma_j \sqrt{\Delta t} \, \xi_j'}, \quad 1 \leq j \leq n.
\]
Least-Squares Problems

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

- 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)

<table>
<thead>
<tr>
<th>Path</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101</td>
<td>97.6424</td>
<td>92.5815</td>
<td>107.5178</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>101.2103</td>
<td>105.1763</td>
<td>102.4524</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
<td>105.7802</td>
<td>103.6010</td>
<td>124.5115</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>96.4411</td>
<td>98.7120</td>
<td>108.3600</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>124.2345</td>
<td>101.0564</td>
<td>104.5315</td>
</tr>
<tr>
<td>6</td>
<td>101</td>
<td>95.8375</td>
<td>93.7270</td>
<td>99.3788</td>
</tr>
<tr>
<td>7</td>
<td>101</td>
<td>108.9554</td>
<td>102.4177</td>
<td>100.9225</td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>104.1475</td>
<td>113.2516</td>
<td>115.0994</td>
</tr>
</tbody>
</table>
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, Chebyshhev polynomials, Gedenbauer polynomials, and Jacobi polynomials.
A Numerical Example (continued)

Cash flows at year 3

<table>
<thead>
<tr>
<th>Path</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2.5476</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.4685</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5.6212</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>4.0775</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Path</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92.5815</td>
<td>$0 \times 0.951229$</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>103.6010</td>
<td>$0 \times 0.951229$</td>
</tr>
<tr>
<td>4</td>
<td>98.7120</td>
<td>$0 \times 0.951229$</td>
</tr>
<tr>
<td>5</td>
<td>101.0564</td>
<td>$0.4685 \times 0.951229$</td>
</tr>
<tr>
<td>6</td>
<td>93.7270</td>
<td>$5.6212 \times 0.951229$</td>
</tr>
<tr>
<td>7</td>
<td>102.4177</td>
<td>$4.0775 \times 0.951229$</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Path</th>
<th>Exercise</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.4185</td>
<td>$f(92.5815) = 2.2558$</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>1.3990</td>
<td>$f(103.6010) = 1.1168$</td>
</tr>
<tr>
<td>4</td>
<td>6.2880</td>
<td>$f(98.7120) = 1.5901$</td>
</tr>
<tr>
<td>5</td>
<td>3.9436</td>
<td>$f(101.0564) = 1.3568$</td>
</tr>
<tr>
<td>6</td>
<td>11.2730</td>
<td>$f(93.7270) = 2.1253$</td>
</tr>
<tr>
<td>7</td>
<td>2.5823</td>
<td>$f(102.4177) = 0.3326$</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
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. 658 become the next ones.
A Numerical Example (continued)

Cash flows at years 2 & 3

<table>
<thead>
<tr>
<th>Path</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>—</td>
<td>12.4185</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>2.5476</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
<td>1.3990</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>—</td>
<td>6.2880</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
<td>3.9436</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>—</td>
<td>11.2730</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
<td>2.5823</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
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. 666, we have the following table.
A Numerical Example (continued)

Regression at year 1

<table>
<thead>
<tr>
<th>Path</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97.6424</td>
<td>$12.4185 \times 0.951229$</td>
</tr>
<tr>
<td>2</td>
<td>101.2103</td>
<td>$2.5476 \times 0.951229^2$</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>96.4411</td>
<td>$6.2880 \times 0.951229$</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>95.8375</td>
<td>$11.2730 \times 0.951229$</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>104.1475</td>
<td>0</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Path</th>
<th>Exercise</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.3576</td>
<td>$f(97.6424) = 8.2230$</td>
</tr>
<tr>
<td>2</td>
<td>3.7897</td>
<td>$f(101.2103) = 3.9882$</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>8.5589</td>
<td>$f(96.4411) = 9.3329$</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>9.1625</td>
<td>$f(95.8375) = 9.83042$</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>0.8525</td>
<td>$f(104.1475) = -0.551885$</td>
</tr>
</tbody>
</table>
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. 666 become the next ones.

• They also confirm the plot on p. 657.
A Numerical Example (continued)

Cash flows at years 1, 2, & 3

<table>
<thead>
<tr>
<th>Path</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>0</td>
<td>12.4185</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>2.5476</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>0</td>
<td>1.3990</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>0</td>
<td>6.2880</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>—</td>
<td>0</td>
<td>3.9436</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>0</td>
<td>11.2730</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>—</td>
<td>0</td>
<td>2.5823</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>0.8525</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
A Numerical Example (continued)

• We move on to year 0.

• The continuation value is, from p 673,

\[
(12.4185 \times 0.951229^2 + 2.5476 \times 0.951229^3 \\
+ 1.3990 \times 0.951229^2 + 6.2880 \times 0.951229^2 \\
+ 3.9436 \times 0.951229^2 + 11.2730 \times 0.951229^2 \\
+ 2.5823 \times 0.951229^2 + 0.8525 \times 0.951229) / 8
\]

\[= 4.66263.\]
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. 659).
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, $\epsilon_t$ is a stationary, uncorrelated process with zero mean and constant variance $\sigma^2$.

- The conditional variance,
  \[
  \text{Var}[X_t \mid X_{t-1}, X_{t-2}, \ldots],
  \]
  equals $\sigma^2$, 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:
  \[ V_t = f(X_{t-1}, X_{t-2}, \ldots) \]
  for some function \( f \).
- Then \( V_t \) can be computed given the information set of past returns:
  \[ I_{t-1} \equiv \{ X_{t-1}, X_{t-2}, \ldots \}. \]
ARCH Models\textsuperscript{a}

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

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

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

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

\textsuperscript{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™.