Merton’s Jump-Diffusion Model

• Empirically, stock returns tend to have fat tails, inconsistent with the Black-Scholes model’s assumptions.

• Stochastic volatility and jump processes have been proposed to address this problem.

• Merton’s (1976) jump-diffusion model is our focus.
Merton’s Jump-Diffusion Model (continued)

- This model superimposes a jump component on a diffusion component.

- The diffusion component is the familiar geometric Brownian motion.

- The jump component is composed of lognormal jumps driven by a Poisson process.
  - It models the rare but large changes in the stock price because of the arrival of important new information.
Merton’s Jump-Diffusion Model (continued)

• Let $S_t$ be the stock price at time $t$.

• The risk-neutral jump-diffusion process for the stock price follows

$$\frac{dS_t}{S_t} = (r - \lambda \bar{k}) \, dt + \sigma \, dW_t + k \, dq_t. \quad (101)$$

• Above, $\sigma$ denotes the volatility of the diffusion component.
Merton’s Jump-Diffusion Model (continued)

- The jump event is governed by a compound Poisson process $q_t$ with intensity $\lambda$, where $k$ denotes the magnitude of the random jump.

  - The distribution of $k$ obeys

    $$\ln(1 + k) \sim N(\gamma, \delta^2)$$

    with mean $\bar{k} \triangleq E(k) = e^{\gamma + \delta^2/2} - 1$.

- The model with $\lambda = 0$ reduces to the Black-Scholes model.
Merton’s Jump-Diffusion Model (continued)

• The solution to Eq. (101) on p. 761 is

\[ S_t = S_0 e^{(r - \lambda \bar{k} - \sigma^2/2) t + \sigma W_t} U(n(t)), \quad (102) \]

where

\[ U(n(t)) = \prod_{i=0}^{n(t)} (1 + k_i). \]

- \( k_i \) is the magnitude of the \( i \)th jump with \( \ln(1 + k_i) \sim N(\gamma, \delta^2) \).
- \( k_0 = 0 \).
- \( n(t) \) is a Poisson process with intensity \( \lambda \).
Merton’s Jump-Diffusion Model (concluded)

• Recall that \( n(t) \) denotes the number of jumps that occur up to time \( t \).

• As \( k > -1 \), stock prices will stay positive.

• The geometric Brownian motion, the lognormal jumps, and the Poisson process are assumed to be independent.
Tree for Merton’s Jump-Diffusion Model$^a$

- Define the $S$-logarithmic return of the stock price $S'$ as
  \[ \ln(S'/S). \]

- Define the logarithmic distance between stock prices $S'$ and $S$ as
  \[ | \ln(S') - \ln(S) | = | \ln(S'/S) |. \]

$^a$Dai (B82506025, R86526008, D8852600), C. Wang (F95922018), Lyuu, & Y. Liu (2010).
Tree for Merton’s Jump-Diffusion Model (continued)

- Take the logarithm of Eq. (102) on p. 763:

\[
M_t \triangleq \ln \left( \frac{S_t}{S_0} \right) = X_t + Y_t, \tag{103}
\]

where

\[
X_t \triangleq \left( r - \lambda \bar{k} - \frac{\sigma^2}{2} \right) t + \sigma W_t, \tag{104}
\]

\[
Y_t \triangleq \sum_{i=0}^{n(t)} \ln (1 + k_i). \tag{105}
\]

- It decomposes the \( S_0 \)-logarithmic return of \( S_t \) into the diffusion component \( X_t \) and the jump component \( Y_t \).
Tree for Merton’s Jump-Diffusion Model (continued)

- Motivated by decomposition (103) on p. 766, the tree construction divides each period into a diffusion phase followed by a jump phase.

- In the diffusion phase, $X_t$ is approximated by the BOPM.

- So $X_t$ makes an up move to $X_t + \sigma \sqrt{\Delta t}$ with probability $p_u$ or a down move to $X_t - \sigma \sqrt{\Delta t}$ with probability $p_d$. 

Tree for Merton’s Jump-Diffusion Model (continued)

- According to BOPM,
  
  \[ p_u = \frac{e^{\mu \Delta t} - d}{u - d}, \]
  
  \[ p_d = 1 - p_u, \]

  except that \( \mu = r - \lambda \bar{k} \) here.

- The diffusion component gives rise to diffusion nodes.

- They are spaced at \( 2\sigma \sqrt{\Delta t} \) apart such as the white nodes A, B, C, D, E, F, and G on p. 769.
White nodes are *diffusion nodes*. Gray nodes are *jump nodes*. In the diffusion phase, the solid black lines denote the binomial structure of BOPM; the dashed lines denote the trinomial structure. Here $m$ is set to one for simplicity. Only the double-circled nodes will remain after the construction. Note that a and b are diffusion nodes because no jump occurs in the jump phase.
Tree for Merton’s Jump-Diffusion Model (concluded)

- In the jump phase, $Y_{t+\Delta t}$ is approximated by moves from each diffusion node to $2m$ jump nodes that match the first $2m$ moments of the lognormal jump.

- The $m$ jump nodes above the diffusion node are spaced at $h \triangleq \sqrt{\gamma^2 + \delta^2}$ apart.

- The same holds for the $m$ jump nodes below the diffusion node.

- The gray nodes at time $\ell\Delta t$ on p. 769 are jump nodes.

- The size of the tree is $O(n^{2.5})$. 
Multivariate Contingent Claims

- They depend on two or more underlying assets.
- The basket call on \( m \) assets has the terminal payoff

\[
\max \left( \sum_{i=1}^{m} \alpha_i S_i(\tau) - X, 0 \right),
\]

where \( \alpha_i \) is the percentage of asset \( i \).

- Basket options are essentially options on a portfolio of stocks; they are index options.
- Option on the best of two risky assets and cash has a terminal payoff of \( \max(S_1(\tau), S_2(\tau), X) \).
## Multivariate Contingent Claims (concluded)\(^a\)

<table>
<thead>
<tr>
<th>Name</th>
<th>Payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange option</td>
<td>(\max(S_1(\tau) - S_2(\tau), 0))</td>
</tr>
<tr>
<td>Better-off option</td>
<td>(\max(S_1(\tau), \ldots, S_k(\tau), 0))</td>
</tr>
<tr>
<td>Worst-off option</td>
<td>(\min(S_1(\tau), \ldots, S_k(\tau), 0))</td>
</tr>
<tr>
<td>Binary maximum option</td>
<td>(I{ \max(S_1(\tau), \ldots, S_k(\tau)) &gt; X })</td>
</tr>
<tr>
<td>Maximum option</td>
<td>(\max(\max(S_1(\tau), \ldots, S_k(\tau)) - X, 0))</td>
</tr>
<tr>
<td>Minimum option</td>
<td>(\max(\min(S_1(\tau), \ldots, S_k(\tau)) - X, 0))</td>
</tr>
<tr>
<td>Spread option</td>
<td>(\max(S_1(\tau) - S_2(\tau) - X, 0))</td>
</tr>
<tr>
<td>Basket average option</td>
<td>(\max((S_1(\tau) + \cdots + S_k(\tau))/k - X, 0))</td>
</tr>
<tr>
<td>Multi-strike option</td>
<td>(\max(S_1(\tau) - X_1, \ldots, S_k(\tau) - X_k, 0))</td>
</tr>
<tr>
<td>Pyramid rainbow option</td>
<td>(\max(</td>
</tr>
<tr>
<td>Madonna option</td>
<td>(\max(\sqrt{(S_1(\tau) - X_1)^2 + \cdots + (S_k(\tau) - X_k)^2} - X, 0))</td>
</tr>
</tbody>
</table>

\(^a\)Lyuu & Teng (R91723054) (2011).

©2018 Prof. Yuh-Dauh Lyuu, National Taiwan University
Correlated Trinomial Model\textsuperscript{a}

- Two risky assets $S_1$ and $S_2$ follow

\[ \frac{dS_i}{S_i} = r \, dt + \sigma_i \, dW_i \]

in a risk-neutral economy, $i = 1, 2$.

- Let

\[ M_i \triangleq e^{r \Delta t}, \]

\[ V_i \triangleq M_i^2 (e^{\sigma_i^2 \Delta t} - 1). \]

- $S_i M_i$ is the mean of $S_i$ at time $\Delta t$.
- $S_i^2 V_i$ the variance of $S_i$ at time $\Delta t$.

\textsuperscript{a}Boyle, Evnine, & Gibbs (1989).
Correlated Trinomial Model (continued)

1. The value of $S_1 S_2$ at time $\Delta t$ has a joint lognormal distribution with mean $S_1 S_2 M_1 M_2 e^{\rho \sigma_1 \sigma_2 \Delta t}$, where $\rho$ is the correlation between $dW_1$ and $dW_2$.

2. Next match the 1st and 2nd moments of the approximating discrete distribution to those of the continuous counterpart.

3. At time $\Delta t$ from now, there are 5 distinct outcomes.
Correlated Trinomial Model (continued)

- The five-point probability distribution of the asset prices is

<table>
<thead>
<tr>
<th>Probability</th>
<th>Asset 1</th>
<th>Asset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>$S_1u_1$</td>
<td>$S_2u_2$</td>
</tr>
<tr>
<td>$p_2$</td>
<td>$S_1u_1$</td>
<td>$S_2d_2$</td>
</tr>
<tr>
<td>$p_3$</td>
<td>$S_1d_1$</td>
<td>$S_2d_2$</td>
</tr>
<tr>
<td>$p_4$</td>
<td>$S_1d_1$</td>
<td>$S_2u_2$</td>
</tr>
<tr>
<td>$p_5$</td>
<td>$S_1$</td>
<td>$S_2$</td>
</tr>
</tbody>
</table>

- As usual, impose $u_i d_i = 1$. 
Correlated Trinomial Model (continued)

- The probabilities must sum to one, and the means must be matched:

\[
\begin{align*}
1 &= p_1 + p_2 + p_3 + p_4 + p_5, \\
S_1 M_1 &= (p_1 + p_2) S_1 u_1 + p_5 S_1 + (p_3 + p_4) S_1 d_1, \\
S_2 M_2 &= (p_1 + p_4) S_2 u_2 + p_5 S_2 + (p_2 + p_3) S_2 d_2.
\end{align*}
\]
Correlated Trinomial Model (concluded)

- Let $R \overset{\Delta}{=} M_1 M_2 e^{\rho \sigma_1 \sigma_2 \Delta t}$.
- Match the variances and covariance:
  \[
  S_1^2 V_1 = (p_1 + p_2)((S_1 u_1)^2 - (S_1 M_1)^2) + p_5(S_1^2 - (S_1 M_1)^2) \\
  + (p_3 + p_4)((S_1 d_1)^2 - (S_1 M_1)^2),
  \]
  \[
  S_2^2 V_2 = (p_1 + p_4)((S_2 u_2)^2 - (S_2 M_2)^2) + p_5(S_2^2 - (S_2 M_2)^2) \\
  + (p_2 + p_3)((S_2 d_2)^2 - (S_2 M_2)^2),
  \]
  \[
  S_1 S_2 R = (p_1 u_1 u_2 + p_2 u_1 d_2 + p_3 d_1 d_2 + p_4 d_1 u_2 + p_5) S_1 S_2.
  \]
- The solutions appear on p. 246 of the textbook.
Correlated Trinomial Model Simplified

- Let $\mu'_i \equiv r - \sigma_i^2 / 2$ and $u_i \equiv e^{\lambda \sigma_i \sqrt{\Delta t}}$ for $i = 1, 2$.

- The following simpler scheme is good enough:

$$
\begin{align*}
\hat{p}^1 &= \frac{1}{4} \left[ \frac{1}{\lambda^2} + \frac{\sqrt{\Delta t}}{\lambda} \left( \frac{\mu'_1}{\sigma_1} + \frac{\mu'_2}{\sigma_2} \right) + \frac{\rho}{\lambda^2} \right], \\
\hat{p}^2 &= \frac{1}{4} \left[ \frac{1}{\lambda^2} + \frac{\sqrt{\Delta t}}{\lambda} \left( \frac{\mu'_1}{\sigma_1} - \frac{\mu'_2}{\sigma_2} \right) - \frac{\rho}{\lambda^2} \right], \\
\hat{p}^3 &= \frac{1}{4} \left[ \frac{1}{\lambda^2} + \frac{\sqrt{\Delta t}}{\lambda} \left( -\frac{\mu'_1}{\sigma_1} - \frac{\mu'_2}{\sigma_2} \right) + \frac{\rho}{\lambda^2} \right], \\
\hat{p}^4 &= \frac{1}{4} \left[ \frac{1}{\lambda^2} + \frac{\sqrt{\Delta t}}{\lambda} \left( -\frac{\mu'_1}{\sigma_1} + \frac{\mu'_2}{\sigma_2} \right) - \frac{\rho}{\lambda^2} \right], \\
\hat{p}^5 &= 1 - \frac{1}{\lambda^2}.
\end{align*}
$$

\footnote{Madan, Milne, & Shefrin (1989).}
Correlated Trinomial Model Simplified (continued)

- All of the probabilities lie between 0 and 1 if and only if
  \[ -1 + \lambda \sqrt{\Delta t} \left| \frac{\mu_1'}{\sigma_1} + \frac{\mu_2'}{\sigma_2} \right| \leq \rho \leq 1 - \lambda \sqrt{\Delta t} \left| \frac{\mu_1'}{\sigma_1} - \frac{\mu_2'}{\sigma_2} \right| \]
  \[ 1 \leq \lambda \]  

- We call a multivariate tree (correlation-) optimal if it guarantees valid probabilities as long as
  \[ -1 + O(\sqrt{\Delta t}) < \rho < 1 - O(\sqrt{\Delta t}) \]

such as the above one.\(^a\)

\(^a\)W. Kao (R98922093) (2011); W. Kao (R98922093), Lyuu, & Wen (D94922003) (2014).
Correlated Trinomial Model Simplified (continued)

- But this model cannot price 2-asset 2-barrier options accurately.\(^a\)

- Few multivariate trees are both optimal and able to handle multiple barriers.\(^b\)

- An alternative is to use orthogonalization.\(^c\)

\(^a\)See Y. Chang (B89704039, R93922034), Hsu (R7526001, D89922012), & Lyuu (2006); W. Kao (R98922093), Lyuu, & Wen (D94922003) (2014) for solutions.

\(^b\)See W. Kao (R98922093), Lyuu, & Wen (D94922003) (2014) for one.

\(^c\)Hull & White (1990); Dai (B82506025, R86526008, D8852600), C. Wang (F95922018), & Lyuu (2013).
Correlated Trinomial Model Simplified (concluded)

- Suppose we allow each asset’s volatility to be a function of time.a

- There are $k$ assets.

- Can you build an optimal multivariate tree that can handle a barrier on each asset in time $O(n^{k+1})$?b

---

a Recall p. 303.
b See Y. Zhang (R05922052) (2018) for a complete solution.
Extrapolation

- It is a method to speed up numerical convergence.
- Say $f(n)$ converges to an unknown limit $f$ at rate of $1/n$:

$$f(n) = f + \frac{c}{n} + o\left(\frac{1}{n}\right).$$  \hspace{1cm} (108)

- Assume $c$ is an unknown constant independent of $n$.
  - Convergence is basically monotonic and smooth.
Extrapolation (concluded)

- From two approximations $f(n_1)$ and $f(n_2)$ and ignoring the smaller terms,

$$f(n_1) = f + \frac{c}{n_1},$$

$$f(n_2) = f + \frac{c}{n_2}.$$  

- A better approximation to the desired $f$ is

$$f = \frac{n_1 f(n_1) - n_2 f(n_2)}{n_1 - n_2}. \quad (109)$$

- This estimate should converge faster than $1/n$.\(^a\)

- The Richardson extrapolation uses $n_2 = 2n_1$.

\(^a\)It is identical to the forward rate formula (22) on p. 147!
Improving BOPM with Extrapolation

- Consider standard European options.
- Denote the option value under BOPM using $n$ time periods by $f(n)$.
- It is known that BOPM convergences at the rate of $1/n$, consistent with Eq. (108) on p. 782.
- But the plots on p. 294 (redrawn on next page) demonstrate that convergence to the true option value oscillates with $n$.
- Extrapolation is inapplicable at this stage.
Improving BOPM with Extrapolation (concluded)

• Take the at-the-money option in the left plot on p. 785.

• The sequence with odd \( n \) turns out to be monotonic and smooth (see the left plot on p. 787).\(^a\)

• Apply extrapolation (109) on p. 783 with \( n_2 = n_1 + 2 \), where \( n_1 \) is odd.

• Result is shown in the right plot on p. 787.

• The convergence rate is amazing.

• See Exercise 9.3.8 of the text (p. 111) for ideas in the general case.

\(^a\)This can be proved (L. Chang & Palmer, 2007).
Numerical Methods
All science is dominated by the idea of approximation.

— Bertrand Russell
Finite-Difference Methods

• Place a grid of points on the space over which the desired function takes value.

• Then approximate the function value at each of these points (p. 791).

• Solve the equation numerically by introducing difference equations in place of derivatives.
Example: Poisson’s Equation

• It is $\partial^2 \theta / \partial x^2 + \partial^2 \theta / \partial y^2 = -\rho(x, y)$, which describes the electrostatic field.

• Replace second derivatives with finite differences through central difference.

• Introduce evenly spaced grid points with distance of $\Delta x$ along the $x$ axis and $\Delta y$ along the $y$ axis.

• The finite difference form is

\[
-\rho(x_i, y_j) = \frac{\theta(x_{i+1}, y_j) - 2\theta(x_i, y_j) + \theta(x_{i-1}, y_j)}{(\Delta x)^2} \\
+ \frac{\theta(x_i, y_{j+1}) - 2\theta(x_i, y_j) + \theta(x_i, y_{j-1})}{(\Delta y)^2}.
\]
Example: Poisson’s Equation (concluded)

- In the above, \( \Delta x \triangleq x_i - x_{i-1} \) and \( \Delta y \triangleq y_j - y_{j-1} \) for \( i, j = 1, 2, \ldots \).

- When the grid points are evenly spaced in both axes so that \( \Delta x = \Delta y = h \), the difference equation becomes

\[
-h^2 \rho(x_i, y_j) = \theta(x_{i+1}, y_j) + \theta(x_{i-1}, y_j) \\
+ \theta(x_i, y_{j+1}) + \theta(x_i, y_{j-1}) - 4\theta(x_i, y_j).
\]

- Given boundary values, we can solve for the \( x_i \)'s and the \( y_j \)'s within the square \([ \pm L, \pm L ]\).

- From now on, \( \theta_{i,j} \) will denote the finite-difference approximation to the exact \( \theta(x_i, y_j) \).
Explicit Methods

- Consider the diffusion equation
  \[ D \left( \frac{\partial^2 \theta}{\partial x^2} \right) - \left( \frac{\partial \theta}{\partial t} \right) = 0, \quad D > 0. \]

- Use evenly spaced grid points \((x_i, t_j)\) with distances \(\Delta x\) and \(\Delta t\), where \(\Delta x = x_{i+1} - x_i\) and \(\Delta t = t_{j+1} - t_j\).

- Employ central difference for the second derivative and forward difference for the time derivative to obtain
  \[
  \left. \frac{\partial \theta(x, t)}{\partial t} \right|_{t=t_j} = \left( \frac{\theta(x, t_{j+1}) - \theta(x, t_j)}{\Delta t} \right) + \cdots, \tag{110}
  \]
  \[
  \left. \frac{\partial^2 \theta(x, t)}{\partial x^2} \right|_{x=x_i} = \left( \frac{\theta(x_{i+1}, t) - 2\theta(x_i, t) + \theta(x_{i-1}, t)}{(\Delta x)^2} \right) + \cdots \tag{111}
  \]
Explicit Methods (continued)

- Next, assemble Eqs. (110) and (111) into a single equation at \((x_i, t_j)\).

- But we need to decide how to evaluate \(x\) in the first equation and \(t\) in the second.

- Since central difference around \(x_i\) is used in Eq. (111), we might as well use \(x_i\) for \(x\) in Eq. (110).

- Two choices are possible for \(t\) in Eq. (111).

- The first choice uses \(t = t_j\) to yield the following finite-difference equation,

\[
\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \frac{\theta_{i+1,j} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2}.
\]

(112)
Explicit Methods (continued)

• The stencil of grid points involves four values, $\theta_{i,j+1}$, $\theta_{i,j}$, $\theta_{i+1,j}$, and $\theta_{i-1,j}$.

• Rearrange Eq. (112) on p. 795 as

$$\theta_{i,j+1} = \frac{D \Delta t}{(\Delta x)^2} \theta_{i+1,j} + \left( 1 - \frac{2D \Delta t}{(\Delta x)^2} \right) \theta_{i,j} + \frac{D \Delta t}{(\Delta x)^2} \theta_{i-1,j}.$$ 

• We can calculate $\theta_{i,j+1}$ from $\theta_{i,j}$, $\theta_{i+1,j}$, $\theta_{i-1,j}$, at the previous time $t_j$ (see exhibit (a) on next page).
Stencils

(a) $x_{i-1}$ $x_i$ $x_{i+1}$ $t_j$ $t_{j+1}$

(b) $x_{i-1}$ $x_i$ $x_{i+1}$ $t_j$ $t_{j+1}$
Explicit Methods (concluded)

• Starting from the initial conditions at $t_0$, that is, $\theta_{i,0} = \theta(x_i, t_0)$, $i = 1, 2, \ldots$, we calculate

  \[ \theta_{i,1}, \quad i = 1, 2, \ldots. \]

• And then

  \[ \theta_{i,2}, \quad i = 1, 2, \ldots. \]

• And so on.
Stability

- The explicit method is numerically unstable unless

\[ \Delta t \leq (\Delta x)^2 / (2D). \]

- A numerical method is unstable if the solution is highly sensitive to changes in initial conditions.

- The stability condition may lead to high running times and memory requirements.

- For instance, halving \( \Delta x \) would imply quadrupling \((\Delta t)^{-1}\), resulting in a running time 8 times as much.
Explicit Method and Trinomial Tree

• Recall that

\[ \theta_{i,j+1} = \frac{D \Delta t}{(\Delta x)^2} \theta_{i+1,j} + \left( 1 - \frac{2D \Delta t}{(\Delta x)^2} \right) \theta_{i,j} + \frac{D \Delta t}{(\Delta x)^2} \theta_{i-1,j}. \]

• When the stability condition is satisfied, the three coefficients for \( \theta_{i+1,j}, \theta_{i,j}, \text{ and } \theta_{i-1,j} \) all lie between zero and one and sum to one.

• They can be interpreted as probabilities.

• So the finite-difference equation becomes identical to backward induction on trinomial trees!
Explicit Method and Trinomial Tree (concluded)

- The freedom in choosing $\Delta x$ corresponds to similar freedom in the construction of trinomial trees.

- The explicit finite-difference equation is also identical to backward induction on a binomial tree.$^a$
  - Let the binomial tree take 2 steps each of length $\Delta t/2$.
  - It is now a trinomial tree.

$^a$Hilliard (2014).
Implicit Methods

• Suppose we use $t = t_{j+1}$ in Eq. (111) on p. 794 instead.

• The finite-difference equation becomes

$$\frac{\theta_{i,j+1} - \theta_{i,j}}{\Delta t} = D \frac{\theta_{i+1,j+1} - 2\theta_{i,j+1} + \theta_{i-1,j+1}}{(\Delta x)^2}.$$  \hspace{1cm} (113)

• The stencil involves $\theta_{i,j}$, $\theta_{i,j+1}$, $\theta_{i+1,j+1}$, and $\theta_{i-1,j+1}$.

• This method is implicit:
  
  – The value of any one of the three quantities at $t_{j+1}$ cannot be calculated unless the other two are known.
  
  – See exhibit (b) on p. 797.
Implicit Methods (continued)

• Equation (113) can be rearranged as

\[ \theta_{i-1,j+1} - (2 + \gamma) \theta_{i,j+1} + \theta_{i+1,j+1} = -\gamma \theta_{i,j}, \]

where \( \gamma \triangleq (\Delta x)^2/(D\Delta t). \)

• This equation is unconditionally stable.

• Suppose the boundary conditions are given at \( x = x_0 \) and \( x = x_{N+1}. \)

• After \( \theta_{i,j} \) has been calculated for \( i = 1, 2, \ldots, N \), the values of \( \theta_{i,j+1} \) at time \( t_{j+1} \) can be computed as the solution to the following tridiagonal linear system,
Implicit Methods (continued)

\[
\begin{bmatrix}
  a & 1 & 0 & \cdots & \cdots & \cdots & 0 \\
  1 & a & 1 & 0 & \cdots & \cdots & 0 \\
  0 & 1 & a & 1 & 0 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
  0 & \cdots & \cdots & 0 & 1 & a & 1 \\
  0 & \cdots & \cdots & 0 & 1 & a & a
\end{bmatrix}
\begin{bmatrix}
  \theta_{1,j+1} \\
  \theta_{2,j+1} \\
  \theta_{3,j+1} \\
  \vdots \\
  \vdots \\
  \theta_{N,j+1}
\end{bmatrix}
= 
\begin{bmatrix}
  -\gamma \theta_{1,j} & -\theta_{0,j+1} \\
  -\gamma \theta_{2,j} & -\gamma \theta_{3,j} \\
  \vdots & \vdots \\
  -\gamma \theta_{N,j} & -\theta_{N+1,j+1}
\end{bmatrix},
\]

where \( a \triangleq -2 - \gamma \).
Implicit Methods (concluded)

• Tridiagonal systems can be solved in $O(N)$ time and $O(N)$ space.
  – Never invert a matrix to solve a tridiagonal system.

• The matrix above is nonsingular when $\gamma \geq 0$.
  – A square matrix is nonsingular if its inverse exists.
Crank-Nicolson Method

- Take the average of explicit method (112) on p. 795 and implicit method (113) on p. 802:

\[
\frac{\theta_{i,j}^{+1} - \theta_{i,j}}{\Delta t} = \frac{1}{2} \left( D \frac{\theta_{i+1,j}^{+1} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2} + D \frac{\theta_{i+1,j}^{+1} - 2\theta_{i,j} + \theta_{i-1,j}}{(\Delta x)^2} \right).
\]

- After rearrangement,

\[
\gamma \theta_{i,j}^{+1} - \frac{\theta_{i+1,j}^{+1} - 2\theta_{i,j} + \theta_{i-1,j}}{2} = \gamma \theta_{i,j} + \frac{\theta_{i+1,j}^{+1} - 2\theta_{i,j} + \theta_{i-1,j}}{2}.
\]

- This is an unconditionally stable implicit method with excellent rates of convergence.
Stencil

\[ x_{i+1} \quad x_i \quad x_{i+1} \]

\[ t_j \quad t_{j+1} \]
Numerically Solving the Black-Scholes PDE (86) on p. 651

- See text.
- Brennan and Schwartz (1978) analyze the stability of the implicit method.
Monte Carlo Simulation\textsuperscript{a}

- Monte Carlo simulation is a sampling scheme.
- In many important applications within finance and without, Monte Carlo is one of the few feasible tools.
- When the time evolution of a stochastic process is not easy to describe analytically, Monte Carlo may very well be the only strategy that succeeds consistently.

\textsuperscript{a}A top 10 algorithm (Dongarra & Sullivan, 2000).
The Big Idea

- Assume $X_1, X_2, \ldots, X_n$ have a joint distribution.
- $\theta \triangleq E[g(X_1, X_2, \ldots, X_n)]$ for some function $g$ is desired.
- We generate
  
  \[
  \left( x^{(i)}_1, x^{(i)}_2, \ldots, x^{(i)}_n \right), \quad 1 \leq i \leq N
  \]
  
  independently with the same joint distribution as $(X_1, X_2, \ldots, X_n)$.
- Set
  
  \[
  Y_i \triangleq g \left( x^{(i)}_1, x^{(i)}_2, \ldots, x^{(i)}_n \right).
  \]
The Big Idea (concluded)

• $Y_1, Y_2, \ldots, Y_N$ are independent and identically distributed random variables.

• Each $Y_i$ has the same distribution as

$$Y \triangleq g(X_1, X_2, \ldots, X_n).$$

• Since the average of these $N$ random variables, $\bar{Y}$, satisfies $E[\bar{Y}] = \theta$, it can be used to estimate $\theta$.

• The strong law of large numbers says that this procedure converges almost surely.

• The number of replications (or independent trials), $N$, is called the sample size.
Accuracy

- The Monte Carlo estimate and true value may differ owing to two reasons:
  1. Sampling variation.
  2. The discreteness of the sample paths.a
- The first can be controlled by the number of replications.
- The second can be controlled by the number of observations along the sample path.

aThis may not be an issue if the financial derivative only requires discrete sampling along the time dimension, such as the discrete barrier option.
Accuracy and Number of Replications

- The statistical error of the sample mean $\bar{Y}$ of the random variable $Y$ grows as $1/\sqrt{N}$.
  - Because $\text{Var} [\bar{Y}] = \text{Var}[Y]/N$.
- In fact, this convergence rate is asymptotically optimal.\(^a\)
- So the variance of the estimator $\bar{Y}$ can be reduced by a factor of $1/N$ by doing $N$ times as much work.
- This is amazing because the same order of convergence holds independently of the dimension $n$.

\(^a\)The Berry-Esseen theorem.
Accuracy and Number of Replications (concluded)

- In contrast, classic numerical integration schemes have an error bound of $O(N^{-c/n})$ for some constant $c > 0$.
  - $n$ is the dimension.

- The required number of evaluations thus grows exponentially in $n$ to achieve a given level of accuracy.
  - The curse of dimensionality.

- The Monte Carlo method is more efficient than alternative procedures for multivariate derivatives when $n$ is large.
For the pricing of European options on a dividend-paying stock, we may proceed as follows.

Assume
\[
\frac{dS}{S} = \mu \, dt + \sigma \, dW.
\]

Stock prices \( S_1, S_2, S_3, \ldots \) at times \( \Delta t, 2\Delta t, 3\Delta t, \ldots \) can be generated via

\[
S_{i+1} = S_i e^{(\mu - \sigma^2/2) \Delta t + \sigma \sqrt{\Delta t} \, \xi}, \quad \xi \sim N(0, 1).
\]  

(114)
Monte Carlo Option Pricing (continued)

- If we discretize $dS/S = \mu dt + \sigma dW$ directly, we will obtain
  
  $$S_{i+1} = S_i + S_i \mu \Delta t + S_i \sigma \sqrt{\Delta t} \, \xi.$$  

- But this is locally normally distributed, not lognormally, hence biased.\(^a\)

- In practice, this is not expected to be a major problem as long as $\Delta t$ is sufficiently small.

\(^a\)Contributed by Mr. Tai, Hui-Chin (R97723028) on April 22, 2009.
Monte Carlo Option Pricing (continued)

• Non-dividend-paying stock prices in a risk-neutral economy can be generated by setting $\mu = r$ and $\Delta t = T$.

1: $C := 0$; \{Accumulated terminal option value.\}
2: \textbf{for} $i = 1, 2, 3, \ldots, N$ \textbf{do}
3: \hspace{1em} $P := S \times e^{(r-\sigma^2/2)T+\sigma\sqrt{T}\xi}, \xi \sim N(0, 1)$;
4: \hspace{1em} $C := C + \max(P - X, 0)$;
5: \hspace{1em} \textbf{end for}
6: \textbf{return} $Ce^{-rT}/N$;
Monte Carlo Option Pricing (concluded)

- Pricing Asian options is also easy.

1: \( C := 0; \)
2: for \( i = 1, 2, 3, \ldots, N \) do
3: \( P := S; \quad M := S; \)
4: for \( j = 1, 2, 3, \ldots, n \) do
5: \( P := P \times e^{(r - \sigma^2/2)(T/n) + \sigma \sqrt{T/n} \xi}; \)
6: \( M := M + P; \)
7: end for
8: \( C := C + \max(M/(n + 1) - X, 0); \)
9: end for
10: return \( Ce^{-rT/N}; \)
How about American Options?

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise (why?).
  - Given a sample path \( S_0, S_1, \ldots, S_n \), how to decide which \( S_i \) is an early-exercise point?
  - What is the option price at each \( S_i \) if the option is not exercised?

- It is difficult to determine the early-exercise point based on one single path.

- But Monte Carlo simulation can be modified to price American options with small biases (pp. 872ff).\(^a\)

\(^a\)Longstaff & Schwartz (2001).
Delta and Common Random Numbers

• In estimating delta, it is natural to start with the finite-difference estimate

\[ e^{-r\tau} \frac{E[P(S + \epsilon)] - E[P(S - \epsilon)]}{2\epsilon}. \]

  – \( P(x) \) is the terminal payoff of the derivative security when the underlying asset’s initial price equals \( x \).

• Use simulation to estimate \( E[P(S + \epsilon)] \) first.

• Use another simulation to estimate \( E[P(S - \epsilon)] \).

• Finally, apply the formula to approximate the delta.

• This is also called the bump-and-revalue method.
Delta and Common Random Numbers (concluded)

- This method is not recommended because of its high variance.

- A much better approach is to use common random numbers to lower the variance:

\[ e^{-r\tau} E \left[ \frac{P(S + \epsilon) - P(S - \epsilon)}{2\epsilon} \right]. \]

- Here, the same random numbers are used for \( P(S + \epsilon) \) and \( P(S - \epsilon) \).

- This holds for gamma and cross gammas (for multivariate derivatives).
Problems with the Bump-and-Revalue Method

• Consider the binary option with payoff

\[
\begin{align*}
1, & \quad \text{if } S(T) > X, \\
0, & \quad \text{otherwise}.
\end{align*}
\]

• Then

\[
P(S+\epsilon)-P(S-\epsilon) = \begin{cases} 
1, & \quad \text{if } S + \epsilon > X \text{ and } S - \epsilon < X, \\
0, & \quad \text{otherwise}.
\end{cases}
\]

• So the finite-difference estimate per run for the (undiscounted) delta is 0 or \(O(1/\epsilon)\).

• This means high variance.
Problems with the Bump-and-Revalue Method (concluded)

- The price of the binary option equals

\[ e^{-r\tau} N(x - \sigma \sqrt{\tau}). \]

- It equals \textit{minus} the derivative of the European call with respect to \( X \).
- It also equals \( X\tau \) times the rho of a European call (p. 348).

- Its delta is

\[ \frac{N'(x - \sigma \sqrt{\tau})}{S\sigma \sqrt{\tau}}. \]
Gamma

- The finite-difference formula for gamma is
  \[ e^{-r\tau} E \left[ \frac{P(S + \epsilon) - 2 \times P(S) + P(S - \epsilon)}{\epsilon^2} \right]. \]

- For a correlation option with multiple underlying assets, the finite-difference formula for the cross gamma \( \partial^2 P(S_1, S_2, \ldots) / (\partial S_1 \partial S_2) \) is:
  \[ e^{-r\tau} E \left[ \frac{P(S_1 + \epsilon_1, S_2 + \epsilon_2) - P(S_1 - \epsilon_1, S_2 + \epsilon_2)}{4\epsilon_1\epsilon_2} \right. \]
  \[ -P(S_1 + \epsilon_1, S_2 - \epsilon_2) + \left. P(S_1 - \epsilon_1, S_2 - \epsilon_2) \right]. \]
Gamma (continued)

• Choosing an $\epsilon$ of the right magnitude can be challenging.
  – If $\epsilon$ is too large, inaccurate Greeks result.
  – If $\epsilon$ is too small, unstable Greeks result.

• This phenomenon is sometimes called the curse of differentiation.\(^a\)

\(^a\)Aït-Sahalia & Lo (1998); Bondarenko (2003).
Gamma (continued)

• In general, suppose

\[
\frac{\partial^i}{\partial \theta^i} e^{-r\tau} E[P(S)] = e^{-r\tau} E \left[ \frac{\partial^i P(S)}{\partial \theta^i} \right]
\]

holds for all \( i > 0 \), where \( \theta \) is a parameter of interest.

– A common requirement is Lipschitz continuity.\(^a\)

• Then formulas for the Greeks become integrals.

• As a result, we avoid \( \epsilon \), finite differences, and resimulation.

\(^a\) Broadie & Glasserman (1996).
Gamma (continued)

- This is indeed possible for a broad class of payoff functions.\(^\text{a}\)
  - Roughly speaking, any payoff function that is equal to a sum of products of differentiable functions and indicator functions with the right kind of support.
  - For example, the payoff of a call is
    \[
    \max(S(T) - X, 0) = (S(T) - X)I\{S(T) - X \geq 0\}.
    \]
  - The results are too technical to cover here (see next page).

\(^\text{a}\)Teng (R91723054) (2004); Lyuu & Teng (R91723054) (2011).
Gamma (continued)

- Suppose \( h(\theta, x) \in \mathcal{H} \) with pdf \( f(x) \) for \( x \) and \( g_j(\theta, x) \in \mathcal{G} \) for \( j \in \mathcal{B} \), a finite set of natural numbers.

- Then

\[
\frac{\partial}{\partial \theta} \int_{\mathbb{R}} h(\theta, x) \prod_{j \in \mathcal{B}} 1\{g_j(\theta, x) > 0\}(x) f(x) \, dx = \\
\int_{\mathbb{R}} h_\theta(\theta, x) \prod_{j \in \mathcal{B}} 1\{g_j(\theta, x) > 0\}(x) f(x) \, dx + \\
\sum_{l \in \mathcal{B}} \left[ h(\theta, x) J_l(\theta, x) \prod_{j \in \mathcal{B} \setminus l} 1\{g_j(\theta, x) > 0\}(x) f(x) \right]_{x = \chi_l(\theta)},
\]

where

\[
J_l(\theta, x) = \text{sign} \left( \frac{\partial g_l(\theta, x)}{\partial x_k} \right) \frac{\partial g_l(\theta, x)/\partial \theta}{\partial g_l(\theta, x)/\partial x} \text{ for } l \in \mathcal{B}.
\]
Gamma (concluded)

• Similar results have been derived for Levy processes.$^a$

• Formulas are also recently obtained for credit derivatives.$^b$

• In queueing networks, this is called infinitesimal perturbation analysis (IPA).$^c$

---

$^b$Lyuu, Teng (R91723054), & Tzeng (2014).
Biases in Pricing Continuously Monitored Options with Monte Carlo

- We are asked to price a continuously monitored up-and-out call with barrier $H$.
- The Monte Carlo method samples the stock price at $n$ discrete time points $t_1, t_2, \ldots, t_n$.
- A sample path

$$S(t_0), S(t_1), \ldots, S(t_n)$$

is produced.

- Here, $t_0 = 0$ is the current time, and $t_n = T$ is the expiration time of the option.
Biases in Pricing Continuously Monitored Options with Monte Carlo (continued)

• If all of the sampled prices are below the barrier, this sample path pays \( \max(S(t_n) - X, 0) \).

• Repeating these steps and averaging the payoffs yield a Monte Carlo estimate.
1: $C := 0$;  
2: for $i = 1, 2, 3, \ldots, N$ do  
3: \hspace{1em} $P := S; \ \text{hit} := 0;$  
4: \hspace{1em} for $j = 1, 2, 3, \ldots, n$ do  
5: \hspace{2em} $P := P \times e^{(r-\sigma^2/2)(T/n)+\sigma\sqrt(T/n)} \xi;$  
6: \hspace{2em} if $P \geq H$ then  
7: \hspace{3em} \text{hit} := 1;  
8: \hspace{3em} break;  
9: \hspace{2em} end if  
10: \hspace{1em} end for  
11: \hspace{1em} if hit = 0 then  
12: \hspace{2em} $C := C + \max(P - X, 0);$  
13: \hspace{2em} end if  
14: end for  
15: return $Ce^{-rT}/N;$
Biases in Pricing Continuously Monitored Options with Monte Carlo (continued)

- This estimate is biased.\(^a\)
  - Suppose none of the sampled prices on a sample path equals or exceeds the barrier \(H\).
  - It remains possible for the continuous sample path that passes through them to hit the barrier between sampled time points (see plot on next page).

\(^a\)Shevchenko (2003).
Biases in Pricing Continuously Monitored Options with Monte Carlo (concluded)

• The bias can certainly be lowered by increasing the number of observations along the sample path.

• However, even daily sampling may not suffice.

• The computational cost also rises as a result.
Brownian Bridge Approach to Pricing Barrier Options

- We desire an unbiased estimate which can be calculated efficiently.
- The above-mentioned payoff should be multiplied by the probability $p$ that a continuous sample path does not hit the barrier conditional on the sampled prices.
- This methodology is called the Brownian bridge approach.
- Formally, we have

$$p \triangleq \operatorname{Prob}[S(t) < H, 0 \leq t \leq T \mid S(t_0), S(t_1), \ldots, S(t_n)].$$
Brownian Bridge Approach to Pricing Barrier Options (continued)

- As a barrier is hit over a time interval if and only if the maximum stock price over that period is at least $H$,

$$p = \text{Prob} \left[ \max_{0 \leq t \leq T} S(t) < H \mid S(t_0), S(t_1), \ldots, S(t_n) \right].$$

- Luckily, the conditional distribution of the maximum over a time interval given the beginning and ending stock prices is known.
Brownian Bridge Approach to Pricing Barrier Options (continued)

**Lemma 21** Assume $S$ follows $dS/S = \mu dt + \sigma dW$ and define

$$
\zeta(x) \triangleq \exp \left[ - \frac{2 \ln(x/S(t)) \ln(x/S(t+\Delta t))}{\sigma^2 \Delta t} \right].
$$

(1) If $H > \max(S(t), S(t + \Delta t))$, then

$$
\text{Prob} \left[ \max_{t \leq u \leq t + \Delta t} S(u) < H \ \bigg| \ S(t), S(t + \Delta t) \right] = 1 - \zeta(H).
$$

(2) If $h < \min(S(t), S(t + \Delta t))$, then

$$
\text{Prob} \left[ \min_{t \leq u \leq t + \Delta t} S(u) > h \ \bigg| \ S(t), S(t + \Delta t) \right] = 1 - \zeta(h).
$$
Brownian Bridge Approach to Pricing Barrier Options (continued)

- Lemma 21 gives the probability that the barrier is not hit in a time interval, given the starting and ending stock prices.
- For our up-and-out call,\(^a\) choose \(n = 1\).
- As a result,

\[
p = \begin{cases} 
1 - \exp \left[ -\frac{2 \ln(H/S(0)) \ln(H/S(T))}{\sigma^2 T} \right], & \text{if } H > \max(S(0), S(T)), \\
0, & \text{otherwise.}
\end{cases}
\]

\(^a\)So \(S(0) < H\).
Brownian Bridge Approach to Pricing Barrier Options
(continued)

The following algorithms works for up-and-out and
down-and-out calls.

1: \( C := 0; \)
2: for \( i = 1, 2, 3, \ldots, N \) do
3: \( P := S \times e^{(r-q-\sigma^2/2)T+\sigma\sqrt{T}\xi()} ; \)
4: if \( (S < H \text{ and } P < H) \) or \( (S > H \text{ and } P > H) \) then
5: \( C := C + \max(P - X, 0) \times \left\{ 1 - \exp \left[ -\frac{2\ln(H/S) \times \ln(H/P)}{\sigma^2 T} \right] \right\} ; \)
6: end if
7: end for
8: return \( Ce^{-rT}/N; \)

Brownian Bridge Approach to Pricing Barrier Options (concluded)

- The idea can be generalized.
- For example, we can handle more complex barrier options.
- Consider an up-and-out call with barrier $H_i$ for the time interval $(t_i, t_{i+1}]$, $0 \leq i < n$.
- This option thus contains $n$ barriers.
- Multiply the probabilities for the $n$ time intervals to obtain the desired probability adjustment term.
Variance Reduction

• The statistical efficiency of Monte Carlo simulation can be measured by the variance of its output.

• If this variance can be lowered without changing the expected value, fewer replications are needed.

• Methods that improve efficiency in this manner are called variance-reduction techniques.

• Such techniques become practical when the added costs are outweighed by the reduction in sampling.
Variance Reduction: Antithetic Variates

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

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

- Then

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

  - $\text{Var}[Y_1]/2$ is the variance of the Monte Carlo method with two independent replications.

- The variance $\text{Var}[ (Y_1 + Y_2)/2 ]$ is smaller than $\text{Var}[Y_1]/2$ 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 \( \mathbf{x} \triangleq (x_1, x_2 \ldots, x_n) \).
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 $x' \triangleq (x'_1, x'_2, \ldots, x'_n)$ 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

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

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

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

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

• 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 $\mu \triangleq E[Y]$.

- Then $W \triangleq X + \beta(Y - \mu)$ 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

\[ \var[W] = \var[X] + \beta^2 \var[Y] + 2\beta \cov[X,Y], \]  

(115)

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

\[ \beta^2 \var[Y] + 2\beta \cov[X,Y] < 0. \]  

(116)
Control Variates (concluded)

- The success of the scheme clearly depends on both $\beta$ and the choice of $Y$.
  - American options can be priced by choosing $Y$ to be the otherwise identical European option and $\mu$ the Black-Scholes formula.\textsuperscript{a}
  - 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.\textsuperscript{b}

\textsuperscript{a}Hull & White (1988).
\textsuperscript{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 $\mu$.
  - Discretely monitored geometric average-rate option vs. the continuously monitored geometric average-rate option given by formulas (53) on p. 424.

\(^a\)But see Dai (B82506025, R86526008, D8852600), Chiu (R94922072), & Lyuu (2015).
\(^b\)Contributed by Ms. Teng, Huei-Wen (R91723054) on May 25, 2004.
Optimal Choice of $\beta$

- For some choices, the discrepancy can be significant, such as the lookback option.\(^a\)

- Equation (115) on p. 850 is minimized when

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

- It is 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 $Y$.

\(^a\)Contributed by Mr. Tsai, Hwai (R92723049) on May 12, 2004.
Optimal Choice of $\beta$ (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 ($\pm 1$), we could control $X$ almost exactly.
Optimal Choice of $\beta$ (continued)

• 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.
  – How to do it efficiently in terms of time and space?
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 $\beta > 0$.

- Suppose a suboptimal $\beta + \epsilon$ is used instead.

- The variance increases by only $\epsilon^2 \text{Var}[Y]$.

---

*aHan & Y. Lai (2010).*
A Pitfall

• A potential pitfall is to sample $X$ and $Y$ independently.
• In this case, $\text{Cov}(X, Y) = 0$.
• Equation (115) on p. 850 becomes

$$\text{Var}(W) = \text{Var}(X) + \beta^2 \text{Var}(Y).$$

• 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 $\sqrt{N}$ 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.