Monte Carlo Simulation[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.

---

[a] A top 10 algorithm (Dongarra & Sullivan, 2000).

The Big Idea

- Assume $X_1, X_2, \ldots, X_n$ have a joint distribution.
- $\theta \stackrel{\Delta}{=} E[\, g(X_1, X_2, \ldots, X_n)\,]$ for some function g is desired.
- We generate

$$\left(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\right), \quad 1 \leq i \leq N$$

independently with the same joint distribution as $(X_1, X_2, \ldots, X_n)$.

- Output $\overline{Y} \stackrel{\Delta}{=} (1/N) \sum_{i=1}^{N} Y_i$, where

$$Y_i \stackrel{\Delta}{=} g\left(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}\right).$$

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, $\overline{Y}$, satisfies $E[\,\overline{Y}\,] = \theta$, it can be used to estimate θ.
- 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.

---

[a]This may not be an issue if the financial derivative only requires discrete sampling along time, such as the *discrete* barrier option.

Accuracy and Number of Replications

- The statistical error of the sample mean $\overline{Y}$ of the random variable Y grows as $1/\sqrt{N}$.
  - Because $\text{Var}[\,\overline{Y}\,] = \text{Var}[\,Y\,]/N$.
- In fact, this convergence rate is asymptotically optimal.[a]
- So the variance of the estimator $\overline{Y}$ 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 $c > 0$.
- 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 for n large.

Monte Carlo Option Pricing

- 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

$$\begin{aligned} & S_{i+1} \\ = \quad & S_i e^{(\mu - \sigma^2/2)\, \Delta t + \sigma \sqrt{\Delta t}\, \xi}, \quad \xi \sim N(0,1), \end{aligned} \tag{124}$$

by Eq. (87) on p. 621.

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]
- Negative stock prices are also possible.[b]
- In practice, this is not expected to be a major problem as long as Δt is sufficiently small.

---

[a] Contributed by Mr. Tai, Hui-Chin (`R97723028`) on April 22, 2009.

[b] Contributed by Mr. Chen, Yu-Hsing (`B06901048`, `R11922045`) on May 5, 2023.

Monte Carlo Option Pricing (continued)

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

$C := 0$; {Accumulated terminal option value.}
for $i = 1, 2, 3, \ldots, N$ **do**
  $P := S \times e^{(r - \sigma^2/2)\, T + \sigma \sqrt{T}\, \xi}$, $\xi \sim N(0, 1)$;
  $C := C + \max(P - X, 0)$;
end for
`return` Ce^{-rT}/N;

[a]It is sometimes called a one-shot simulation (Brigo & Mercurio, 2006).

Monte Carlo Option Pricing (concluded)

Pricing Asian options is also easy.

$C := 0;$
for $i = 1, 2, 3, \ldots, N$ **do**
  $P := S;\ M := S;$
  for $j = 1, 2, 3, \ldots, n$ **do**
    $P := P \times e^{(r-\sigma^2/2)(T/n)+\sigma\sqrt{T/n}\,\xi};$
    $M := M + P;$
  end for
  $C := C + \max(M/(n+1) - X, 0);$
end for
`return` $Ce^{-rT}/N;$

How about American Options?

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise.
  - 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.[a]
- But Monte Carlo simulation can be modified to price American options with small biases.[b]

---

[a]Unless, of course, the exercise boundary is given (recall pp. 403ff). Contributed by Mr. Chen, Tung-Li (D09922014) on May 5, 2023.

[b]Longstaff & Schwartz (2001). See pp. 927ff.

Obtaining Profit and Loss of Delta Hedge[a]

- Profit and loss of delta hedge should be calculated under the real-world probability measure.[b]
- So stock prices should be sampled from

$$\frac{dS}{S} = \mu\, dt + \sigma\, dW.$$

- Suppose backward induction on a tree under the risk-neutral measure is performed for the delta.[c]

---

[a]Contributed by Mr. Lu, Zheng-Liang (`D00922011`) on August 12, 2021.

[b]Recall p. 713.

[c]Because, say, no closed-form formulas are available for the delta.

Obtaining Profit and Loss of Delta Hedge (concluded)

- Note that one needs a delta per stock price.
- So Nn trees are needed for the distribution of the profit and loss from N paths with $n + 1$ stock prices per path.
- These are a lot of trees!
- How to do it efficiently to generate plots like that on p. 656?

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 gamma.[a]

[a]For multivariate derivatives.

Problems with the Bump-and-Revalue Method

- Consider the binary option with payoff

$$\begin{cases} 1, & \text{if } S(T) > X, \\ 0, & \text{otherwise.} \end{cases}$$

- Then, if common random numbers are used,

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

- 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 *minus* the derivative of the European call with respect to X.
  - It also equals $X\tau$ times the rho of a European call (p. 364).

- 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:

$$e^{-r\tau}\,E\Bigg[\,\frac{P(S_1+\epsilon_1, S_2+\epsilon_2)-P(S_1-\epsilon_1, S_2+\epsilon_2)}{4\epsilon_1\epsilon_2}$$

$$\frac{-P(S_1+\epsilon_1, S_2-\epsilon_2)+P(S_1-\epsilon_1, S_2-\epsilon_2)}{}\Bigg].$$

Gamma (continued)

- Choosing an ϵ of the right magnitude can be challenging.
  - If ϵ is too large, inaccurate Greeks result.
  - If ϵ 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 (in some sense)

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

  holds for all $i > 0$, where θ is a parameter of interest.[a]
  - A common requirement is Lipschitz continuity.[b]
- Then Greeks become integrals.
- As a result, we avoid ϵ, finite differences, and resimulation.

---

[a] The $\partial^i P(S)/\partial \theta^i$ within $E[\,\cdot\,]$ may not be partial differentiation in the classic sense.

[b] Broadie & Glasserman (1996).

Gamma (continued)

- This is indeed possible for a broad class of payoff functions.[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).

[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

$$
\begin{aligned}
& \frac{\partial}{\partial \theta} \int_{\Re} h(\theta, x) \prod_{j \in \mathcal{B}} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x)\, f(x)\, dx \\
= \quad & \int_{\Re} h_\theta(\theta, x) \prod_{j \in \mathcal{B}} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x)\, f(x)\, dx \\
& + \sum_{l \in \mathcal{B}} \left[h(\theta, x) J_l(\theta, x) \prod_{j \in \mathcal{B} \setminus l} \mathbf{1}_{\{g_j(\theta, x) > 0\}}(x)\, f(x) \right]_{x = \chi_l(\theta)},
\end{aligned}
$$

where

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

Gamma (concluded)

- Similar results have been derived for Levy processes.[a]
- Formulas are also available for credit derivatives.[b]
- In queueing networks, this is called infinitesimal perturbation analysis (IPA).[c]

---

[a] Lyuu, Teng (R91723054), & S. Wang (2013).
[b] Lyuu, Teng (R91723054), Tseng, & S. Wang (2014, 2019).
[c] Cao (1985); Y. C. Ho & Cao (1985).

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 $\max(S(t_n) - X, 0)$.
- Repeat these steps and average the payoffs for a Monte Carlo estimate.

$C := 0$;

for $i = 1, 2, 3, \ldots, N$ **do**

  $P := S$; hit $:= 0$;

  for $j = 1, 2, 3, \ldots, n$ **do**

    $P := P \times e^{(r-\sigma^2/2)\,(T/n)+\sigma\sqrt{(T/n)}\,\xi}$; {By Eq. (124) on p. 860.}

    if $P \geq H$ **then**

      hit $:= 1$;

      `break`;

    end if

  end for

  if hit $= 0$ **then**

    $C := C + \max(P - X, 0)$;

  end if

end for

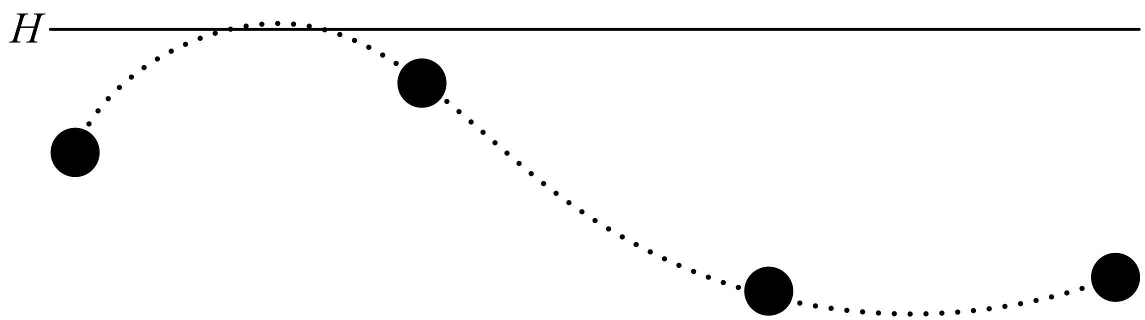
`return` Ce^{-rT}/N;

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).
  - Hence the knock-out probability is underestimated.

---

[a]Shevchenko (2003).


H

Biases in Pricing Continuously Monitored Options with Monte Carlo (concluded)

- The bias can be lowered by increasing the number of observations along the sample path.
  - For trees, the knock-out probability may *decrease* as the number of time steps is increased.
- 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.
  - Formally,

$$p \stackrel{\Delta}{=} \text{Prob}[\, S(t) < H, 0 \leq t \leq T \,|\, S(t_0), S(t_1), \ldots, S(t_n) \,].$$

- This methodology is called the Brownian bridge approach.

Brownian Bridge Approach to Pricing Barrier Options (continued)

- As a barrier is not hit over a time interval if and only if the maximum stock price over that period is at most 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 22 *Assume S follows $dS/S = \mu\, dt + \sigma\, dW$ and define*[a]

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

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

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

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

$$\text{Prob}\left[\min_{t \le u \le t+\Delta t} S(u) > h \,\middle|\, S(t), S(t+\Delta t) \right] = 1 - \zeta(h).$$

---

[a] Here, Δt is an arbitrary positive real number.

Brownian Bridge Approach to Pricing Barrier Options (continued)

- Lemma 22 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[a] call, 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 $S(0) < H$ by definition.

Brownian Bridge Approach to Pricing Barrier Options (continued)

The following algorithm works for up-and-out *and* down-and-out calls.

$C := 0;$
for $i = 1, 2, 3, \ldots, N$ **do**
  $P := S \times e^{(r-q-\sigma^2/2)\,T+\sigma\sqrt{T}\;\xi(\,)};$
  if $(S < H$ and $P < H)$ or $(S > H$ and $P > H)$ **then**
    $C := C + \max(P - X, 0) \times \left\{1 - \exp\left[-\frac{2\ln(H/S) \times \ln(H/P)}{\sigma^2 T}\right]\right\};$
  end if
end for
`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 $(t_i, t_{i+1}]$, $0 \leq i < m$.
- This option contains m barriers.
- Multiply the probabilities for the m time intervals to obtain the desired probability adjustment term.

Pricing Barrier Options without Brownian Bridge

- Let T_h denote the amount of time for a process X_t to hit h for the *first* time.
- It is called the first passage time or the first hitting time.
- Suppose X_t is a (μ, σ) Brownian motion:

$$dX_t = \mu\, dt + \sigma\, dW_t, \quad t \geq 0.$$

Pricing Barrier Options without Brownian Bridge (continued)

- The first passage time T_h follows the inverse Gaussian (IG) distribution with probability density function:[a]

$$\frac{|\, h - X(0) \,|}{\sigma t^{3/2} \sqrt{2\pi}}\, e^{-(h - X(0) - \mu x)^2 / (2\sigma^2 x)}.$$

- For pricing a barrier option with barrier H by simulation, the density function becomes

$$\frac{|\, \ln(H/S(0)) \,|}{\sigma t^{3/2} \sqrt{2\pi}}\, e^{-\left[\, \ln(H/S(0)) - (r - \sigma^2/2)\, x \,\right]^2 / (2\sigma^2 x)}.$$

---

[a] A. N. Borodin & Salminen (1996), with Laplace transform $E[\, e^{-\lambda T_h} \,] = e^{-|\, h - X(0) \,| \sqrt{2\lambda}}$, $\lambda > 0$.

Pricing Barrier Options without Brownian Bridge (concluded)

- Draw an x from this distribution.[a]
- If $x > T$, a knock-in option fails to knock in, whereas a knock-out option does not knock out.
- If $x \leq T$, the opposite is true.
- If the barrier option survives at maturity T, then draw an $S(T)$ to calculate its payoff.
- Repeat the above process and average the discounted payoff.

[a]The IG distribution can be very efficiently sampled (Michael, Schucany, & Haas, 1976).

Brownian Bridge Approach to Pricing Lookback Options[a]

- By Lemma 22(1) (p. 885),

$$
\begin{aligned}
F_{\max}(y) &\stackrel{\Delta}{=} \text{Prob}\left[\max_{0\leq t\leq T} S(t) < y \,|\, S(0), S(T)\right] \\
&= 1 - \exp\left[-\frac{2\ln(y/S(0))\ln(y/S(T))}{\sigma^2 T}\right].
\end{aligned}
$$

- So $F_{\max}$ is the conditional distribution function of the maximum stock price.

[a] El Babsiri & Noel (1998).

Brownian Bridge Approach to Pricing Lookback Options (continued)

- A random variable with that distribution can be generated by $F_{\max}^{-1}(x)$, where x is uniformly distributed over $(0, 1)$.[a]
- Note that

$$x = 1 - \exp\left[-\frac{2\ln(y/S(0))\ln(y/S(T))}{\sigma^2 T} \right].$$

[a]This is called the inverse-transform technique (see p. 259 of the textbook).

Brownian Bridge Approach to Pricing Lookback Options (continued)

- Equivalently,

$$
\begin{aligned}
& \ln(1-x) \\
= \quad & -\frac{2\ln(y/S(0))\ln(y/S(T))}{\sigma^2 T} \\
= \quad & -\frac{2}{\sigma^2 T}\{\,[\,\ln(y)-\ln S(0)\,]\,[\,\ln(y)-\ln S(T)\,]\,\}.
\end{aligned}
$$

Brownian Bridge Approach to Pricing Lookback Options (continued)

- There are two solutions for $\ln y$.
- But only one is consistent with $y \geq \max(S(0), S(T))$:

$$\ln y = \frac{\ln(S(0)\, S(T)) + \sqrt{\left(\ln \frac{S(T)}{S(0)}\right)^2 - 2\sigma^2 T \ln(1-x)}}{2}.$$

Brownian Bridge Approach to Pricing Lookback Options (concluded)

The following algorithm works for the lookback put on the maximum.

$C := 0;$
for $i = 1, 2, 3, \ldots, N$ **do**
  $P := S \times e^{(r-q-\sigma^2/2)\,T+\sigma\sqrt{T}\,\xi(\,)}$; {By Eq. (124) on p. 860.}
  $Y := \exp\left[\frac{\ln(SP)+\sqrt{\left(\ln\frac{P}{S}\right)^2-2\sigma^2 T\ln[\,1-U(0,1)\,]}}{2}\right];$
  $C := C + (Y - P);$
end for
`return` $Ce^{-rT}/N;$

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 work 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 want to estimate $E[\, g(X_1, X_2, \ldots, X_n) \,]$.
- Let Y_1 and Y_2 be random variables with the same distribution as $g(X_1, X_2, \ldots, X_n)$.
- Then

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

  - $\mathrm{Var}[\, Y_1 \,]/2$ is the variance of the Monte Carlo method with two *independent* replications.
- The variance $\mathrm{Var}[\, (Y_1 + Y_2)/2 \,]$ is smaller than $\mathrm{Var}[\, Y_1 \,]/2$ 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 first path's random numbers.
- 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.
- Suppose one simulation run has realizations $\xi_1, \xi_2, \ldots, \xi_n$ for the normally distributed fluctuation term ξ.
- This generates samples $x_1, x_2, \ldots, x_n$.
- The first estimate is then $g(\boldsymbol{x})$, where $\boldsymbol{x} \stackrel{\Delta}{=} (x_1, x_2 \ldots, x_n)$.

Variance Reduction: Antithetic Variates (concluded)

- The antithetic-variates method does not sample n more numbers from ξ for the second estimate $g(\boldsymbol{x}')$.
- Instead, generate the sample path $\boldsymbol{x}' \stackrel{\Delta}{=} (x'_1, x'_2 \ldots, x'_n)$ from $-\xi_1, -\xi_2, \ldots, -\xi_n$.
- Compute $g(\boldsymbol{x}')$.
- Output $(g(\boldsymbol{x}) + g(\boldsymbol{x}'))/2$.
- Repeat the above steps.

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
$$\mathrm{Var}[\,E[\,X \mid Z\,]\,] \leq \mathrm{Var}[\,X\,],$$
$E[\,X \mid Z\,]$ 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 is repeated 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 $\mu \stackrel{\Delta}{=} E[\,Y\,]$.
- 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

$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] + \beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,], \tag{125}$$

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

$$\beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,] < 0. \tag{126}$$

Control Variates (concluded)

- The success of the scheme clearly depends on both β and the choice of Y.
  - For pricing American options, choose Y to be the otherwise identical European option and μ the Black-Scholes formula.[a]
  - For pricing Arithmetic Asian options, choose Y to be the otherwise identical geometric Asian option, μ the formula (58) on p. 447, and $\beta = -1$.
- This approach is often 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 assess 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 T. Dai (`B82506025`, `R86526008`, `D8852600`), C. Chiu (`B90201037`, `R94922072`), & Lyuu (2015, 2018).

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

[c] Priced by formulas (58) on p. 447.

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

Optimal Choice of β

- Equation (125) on p. 905 is minimized when

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

  – It is called beta.

- For this specific β,

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

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

Optimal Choice of β (continued)

- The variance can never increase with the optimal choice.
- 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 $\text{Var}[\,Y\,]$ nor $\text{Cov}[\,X, Y\,]$ is known.
- So we cannot hope to obtain the maximum reduction in variance.
- We can guess a β 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 $\text{Var}[\,Y\,]$ and $\text{Cov}[\,X, Y\,]$.
  - 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 (125) on p. 905 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 or Cholesky factorization.

  - Above, L is a lower triangular matrix.
  - It can be computed by Crout's algorithm in quadratic time.[a]

[a]Golub & Van Loan (1989).

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 $PP^{\mathrm{T}} = C$, where[a]

$$P = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}.$$

---

[a] Recall Eq. (28) on p. 180.

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.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives.[a]
- 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.[b]

---

[a] Recall pp. 822ff.

[b] Johnson (1987); C. Y. 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. (124) on p. 860, for each asset $1 \leq j \leq k$,

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

by Eq. (124) on p. 860.

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 abbreviated 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 exercise decision cannot be reached by looking at just one path.

The Least-Squares Monte Carlo Approach

- Estimate the continuation value from the cross-sectional information in the simulation with least squares.[a]
- The result is a function of the state for estimating it.
- Use the estimated continuation value for each path to determine its cash flow.
- This is called least-squares Monte Carlo (LSM).

[a] Longstaff & Schwartz (2001).

The Least-Squares Monte Carlo Approach (concluded)

- LSM is provably convergent.[a]
- LSM can be easily parallelized.[b]
  - Partition the paths into subproblems and perform LSM on each 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).

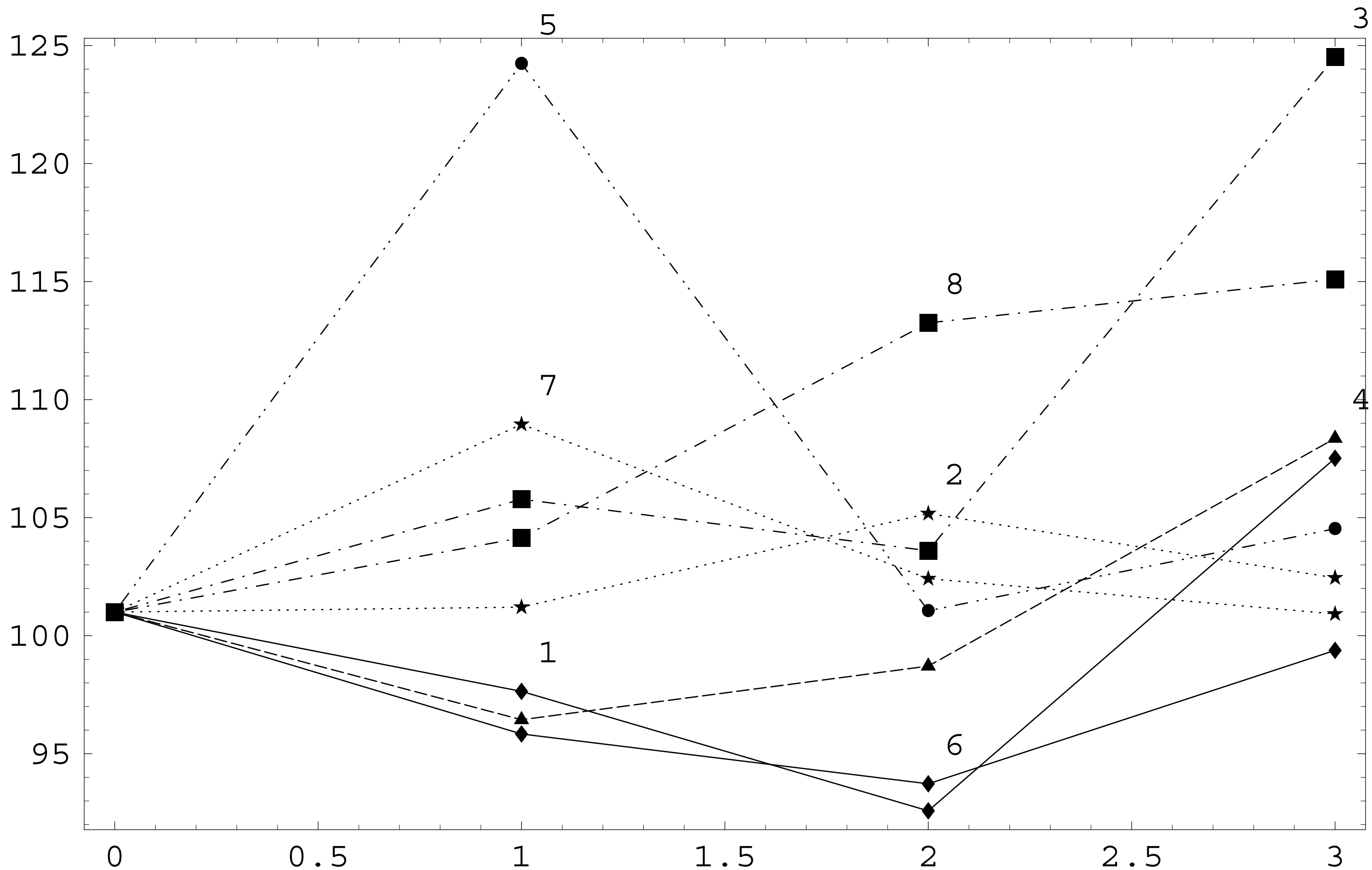
A Numerical Example

- 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 equals 0.951229.
- The current stock price is 101.
- We use 8 price paths to illustrate the algorithm.

A Numerical Example (continued)

Stock price paths

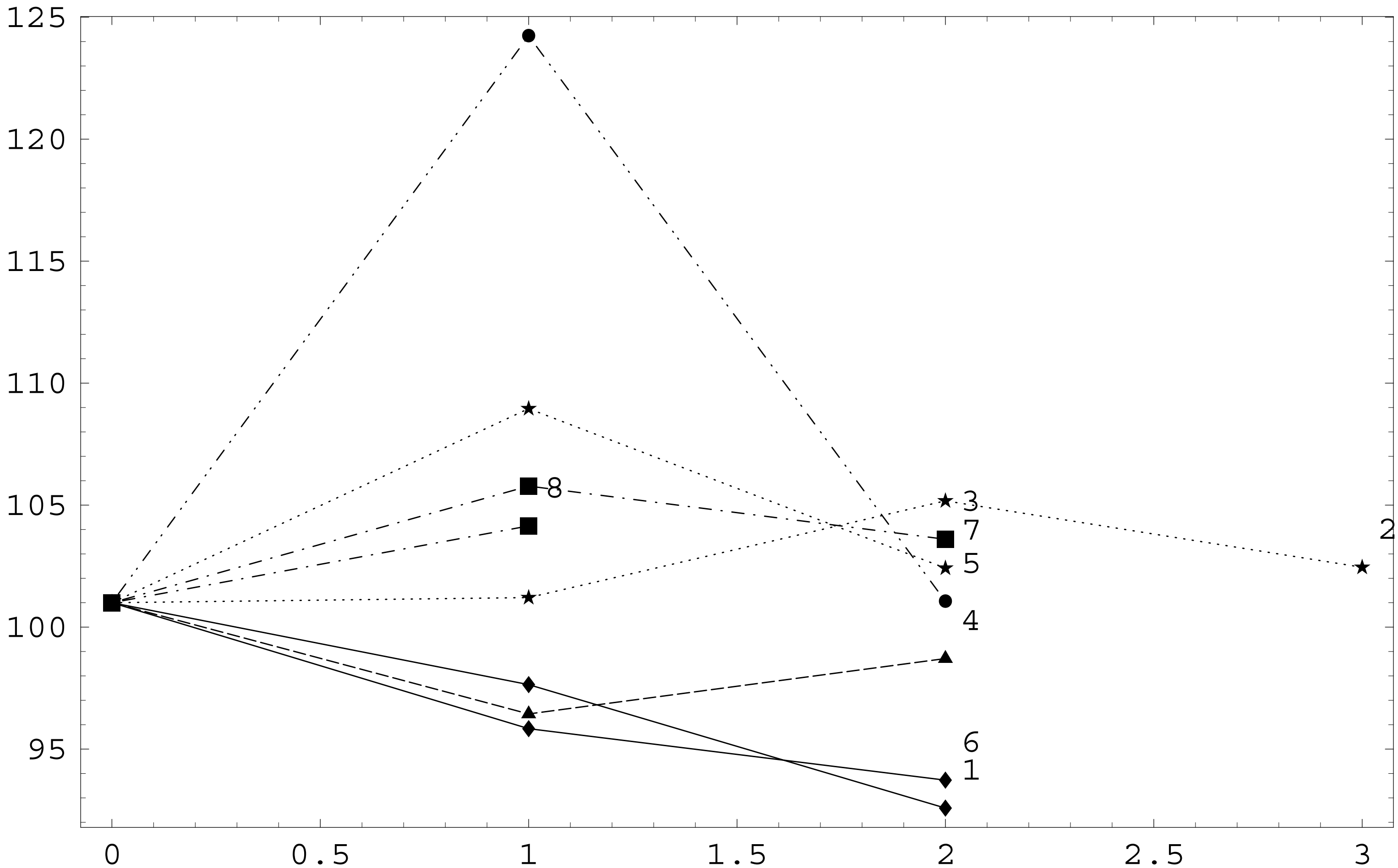
Path	Year 0	Year 1	Year 2	Year 3
1	**101**	**97.6424**	**92.5815**	107.5178
2	**101**	**101.2103**	105.1763	**102.4524**
3	**101**	105.7802	**103.6010**	124.5115
4	**101**	**96.4411**	**98.7120**	108.3600
5	**101**	124.2345	**101.0564**	**104.5315**
6	**101**	**95.8375**	**93.7270**	**99.3788**
7	**101**	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* in-the-money paths.

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


125
120
115
110
105
100
95
0
0.5
1
1.5
2
2.5
3
8
3
7
2
5
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 put's payoffs.
- Only 4 paths are in the money: 2, 5, 6, 7.
- Some of the cash flows may not materialize if the put is exercised earlier, which we will find out later.
- Incidentally, the *European* counterpart has a value of

$$\begin{aligned} & 0.951229^3 \times \frac{2.5476 + 0.4685 + 5.6212 + 4.0775}{8} \\ = \; & 1.3680. \end{aligned}$$

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.[a]
- Only in-the-money paths will be used in the regression because they are where early exercise is possible.
  - If there were none, move on to year 1.

---

[a] Recall p. 931.

A Numerical Example (continued)

- Let x denote the stock price at year 2 for each of those 6 paths.
- Let y denote the corresponding discounted future cash flow (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 estimated 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)

- The put should be exercised in all 6 paths: 1, 3, 4, 5, 6, 7.
- Now, any positive cash flow at year 3 vanishes for these paths as the put has been exercised before it.[a]
  - They are paths 5, 6, 7.
- The cash flows on p. 935 become the ones on next slide.

---

[a]Recall p. 931.

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	$\boxed{0}$
6	—	—	11.2730	$\boxed{0}$
7	—	—	2.5823	$\boxed{0}$
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.[a]
- Only in-the-money paths will be used in the regression because they are where early exercise is possible.
  - If there were none, move on to year 0.

[a] Recall p. 931.

A Numerical Example (continued)

- Let x denote the stock price at year 1 for each of those 5 paths.
- Let y denote the corresponding discounted future cash flow if the put is not exercised at year 1.
- From p. 943, 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	0×0.951229

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 estimated 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 its $f(104.1475) < 0$.
- Now, any positive future cash flow vanishes for this path.
  - But there is none.
- The cash flows on p. 943 become the ones on next slide.
- They also confirm the plot on p. 934.

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 950,

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

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.[a]

[a] Recall p. 936.