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.

Monte Carlo Option Pricing

- For the pricing of European options on a dividend-paying stock, we may proceed as follows.
- Stock prices $S_1, S_2, S_3, \ldots$ at times $\Delta t, 2\Delta t, 3\Delta t, \ldots$ can be generated via

$$S_{i+1} = S_i e^{(\mu - \sigma^2/2)\,\Delta t + \sigma\sqrt{\Delta t}\,\xi}, \quad \xi \sim N(0,1) \tag{71}$$

  when $dS/S = \mu\,dt + \sigma\,dW$.
- Non-dividend-paying stock prices in a risk-neutral economy can be generated by setting $\mu = r$.
- Pricing Asian options is easy (see text).

Pricing American Options

- Standard Monte Carlo simulation is inappropriate for American options because of early exercise.
- It is difficult to determine the early-exercise point based on one single path.
- Monte Carlo simulation can be modified to price American options with small biases (see p. 683).[a]

[a] Longstaff and 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.

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).

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

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

  - $\text{Var}[\,Y_1\,]/2$ 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 ξ.
- This generates samples $x_1, x_2, \ldots, x_n$.
- The estimate is then $g(\boldsymbol{x})$, where $\boldsymbol{x} \equiv (x_1, x_2 \ldots, x_n)$.

Variance Reduction: Antithetic Variates (concluded)

- We do not sample n more numbers from ξ for the second estimate.
- The antithetic-variates method computes $g(\boldsymbol{x}')$ from the sample path $\boldsymbol{x}' \equiv (x'_1, x'_2 \ldots, x'_n)$ generated by $-\xi_1, -\xi_2, \ldots, -\xi_n$.
- We then output $(g(\boldsymbol{x}) + g(\boldsymbol{x}'))/2$.
- 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 $E[\,X \mid Z = z\,]$ can be efficiently and precisely computed.
- $E[\,X\,] = E[\,E[\,X \mid Z\,]\,]$ by the law of iterated conditional expectations.
- Hence the random variable $E[\,X \mid Z\,]$ 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 β.
  - β can scale the deviation $Y - \mu$ to arrive at an adjustment for X.
  - 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
$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] + \beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,], \tag{72}$$
- Hence W is less variable than X if and only if
$$\beta^2\,\mathrm{Var}[\,Y\,] + 2\beta\,\mathrm{Cov}[\,X, Y\,] < 0. \tag{73}$$
- The success of the scheme clearly depends on both β and the choice of Y.

Control Variates (concluded)

- 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 μ.
  - Discretely monitored geometric average-rate option vs. the continuously monitored geometric average-rate option given by formulas (31) on p. 314.
- 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 β

- Equation (72) on p. 642 is minimized when
$$\beta = -\mathrm{Cov}[\,X, Y\,]/\mathrm{Var}[\,Y\,],$$
which was called beta earlier in the book.
- For this specific β,
$$\mathrm{Var}[\,W\,] = \mathrm{Var}[\,X\,] - \frac{\mathrm{Cov}[\,X, Y\,]^2}{\mathrm{Var}[\,Y\,]} = \left(1 - \rho_{X,Y}^2\right)\mathrm{Var}[\,X\,],$$
where $\rho_{X,Y}$ is the correlation between X and Y.
- The stronger X and Y are correlated, the greater the reduction in variance.

Optimal Choice of β (continued)

- For example, if this correlation is nearly perfect (± 1), we could control X almost exactly, eliminating practically all of its variance.
- 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 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 $\beta > 0$.

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.

Quasi-Monte Carlo Methods

- The low-discrepancy sequences (or quasi-random sequences) address the above-mentioned problems.
- It is a deterministic version of the Monte Carlo method in that random samples are replaced by deterministic quasi-random points.
- If a smaller number of samples suffices as a result, efficiency has been gained.
- Aim is to select deterministic points for which the deterministic error bound is smaller than Monte Carlo's probabilistic error bound.

Problems with Quasi-Monte Carlo Methods

- Their theories are valid for integration problems, but may not be directly applicable to simulations because of the correlations between points in a quasi-random sequence.
- This problem may be overcome by writing the desired result as an integral.
- But the integral often has a very high dimension.

Problems with Quasi-Monte Carlo Methods (concluded)

- The improved accuracy is generally lost for problems of high dimension or problems in which the integrand is not smooth.
- No theoretical basis for empirical estimates of their accuracy, a role played by the central limit theorem in the Monte Carlo method.

Assessment

- The results are somewhat mixed.
- The application of such methods in finance seems promising.
- A speed-up as high as 1,000 over the Monte Carlo method, for example, is reported.
- The success of the quasi-Monte Carlo method when compared with traditional variance-reduction techniques is problem dependent.
- For example, the antithetic-variates method outperforms the quasi-Monte Carlo method in bond pricing.

Matrix Computation

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

Definitions and Basic Results

- Let $A \equiv [\, a_{ij} \,]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \boldsymbol{R}^{m \times n}$, denote an $m \times n$ matrix.
- It can also be represented as $[\, a_1, a_2, \ldots, a_n \,]$ where $a_i \in \boldsymbol{R}^m$ are vectors.
  - Vectors are column vectors unless stated otherwise.
- A is a square matrix when $m = n$.
- The rank of a matrix is the largest number of linearly independent columns.
- An $m \times n$ matrix is rank deficient if its rank is less than $\min(m, n)$; otherwise, it has full rank.

Definitions and Basic Results (continued)

- A square matrix A is said to be symmetric if $A^{\mathrm{T}} = A$.
- A real $n \times n$ matrix $A \equiv [\, a_{ij} \,]_{i,j}$ is diagonally dominant if $|\, a_{ii} \,| > \sum_{j \neq i} |\, a_{ij} \,|$ for $1 \leq i \leq n$.
  - Such matrices are nonsingular.
- A diagonal $m \times n$ matrix $D \equiv [\, d_{ij} \,]_{i,j}$ may be denoted by $\mathrm{diag}[\, D_1, D_2, \ldots, D_q \,]$, where $q \equiv \min(m, n)$ and $D_i = d_{ii}$ for $1 \leq i \leq q$.
- The identity matrix is the square matrix

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

Diagonal Matrices

$$\begin{bmatrix} \times & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} \times & 0 & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \end{bmatrix} \quad \begin{bmatrix} \times & 0 & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

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.
- It is known that 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.

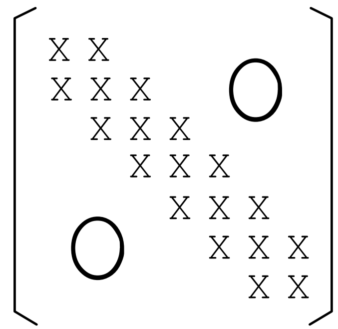
Gaussian Elimination[a]

- Gaussian elimination is a standard method for solving a linear system $Ax = b$, where $A \in \boldsymbol{R}^{n \times n}$.
- The total running time is $O(n^3)$.
- The space complexity is $O(n^2)$.

[a]Carl Friedrich Gauss (1777–1855) in 1809.

Banded Linear Systems

- Matrix A is banded if all the nonzero elements are placed near the diagonal of the matrix.
- We say $A = [\, a_{ij} \,]_{i,j}$ has upper bandwidth u if $a_{ij} = 0$ for $j - i > u$ and lower bandwidth l if $a_{ij} = 0$ for $i - j > l$.
  - A tridiagonal matrix, for instance, has upper bandwidth one and lower bandwidth one.
- For banded matrices, Gaussian elimination can be easily modified to run in $O(nul)$ time.



Decompositions

- Gaussian elimination can be used to factor any square matrix all of whose leading principal submatrices are nonsingular into a product of a lower triangular matrix L and an upper triangular matrix U:

$$A = LU.$$

- This is called the LU decomposition.
- The conditions are satisfied by positive definite matrices and diagonally dominant matrices.
- Positive definite matrices can in fact be factored as

$$A = LL^{\mathrm{T}}, \tag{74}$$

called the Cholesky decomposition.

Orthogonal and Orthonormal Matrices

- A vector set $\{ x_1, x_2, \ldots, x_p \}$ is orthogonal if all its vectors are nonzero and the inner products $x_i^{\mathrm{T}} x_j$ equal zero for $i \neq j$.
- It is orthonormal if, furthermore,

$$x_i^{\mathrm{T}} x_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- A real square matrix Q is orthogonal if $Q^{\mathrm{T}} Q = I$.
- For such matrices, $Q^{-1} = Q^{\mathrm{T}}$ and $QQ^{\mathrm{T}} = I$.

Generation of Multivariate Normal Distribution

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

$$\mathrm{Cov}[P\boldsymbol{y}] = P\,\mathrm{Cov}[\boldsymbol{y}]\,P^{\mathrm{T}} = PP^{\mathrm{T}} = C.$$

Generation of Multivariate Normal Distribution (concluded)

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^{\mathrm{T}}$.
- We start with independent standard normal distributions $y_1, y_2, \ldots, y_n$.
- Then $P[y_1, y_2, \ldots, y_n]^{\mathrm{T}}$ has the desired distribution.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives (p. 567).
- For example, the rainbow option on k assets has payoff

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

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

[a] Johnson (1987).

Multivariate Derivatives Pricing (concluded)

- Suppose $dS_j/S_j = r\,dt + \sigma_j\,dW_j$, $1 \leq j \leq n$, 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. (71) on p. 631,

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

Least-Squares Problems

- The least-squares (LS) problem is concerned with $\min_{x \in R^n} \parallel Ax - b \parallel$, where $A \in \boldsymbol{R}^{m \times n}$, $b \in \boldsymbol{R}^m$, $m \geq n$.
- The LS problem is called regression analysis in statistics and is equivalent to minimizing the mean-square error.
- Often stated as $Ax = b$, the LS problem is overdetermined when there are more equations than unknowns ($m > n$).

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,

$$\begin{bmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^n \\ 1 & a_2 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_m & a_m^2 & \cdots & a_m^n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \tag{75}$$

Normal Equations

- Since Ax is a linear combination of A's columns with coefficients $x_1, x_2, \ldots, x_n$, the LS problem finds the minimum distance between b and A's column space.
- A solution x_{LS} must identify a point Ax_{LS} which is at least as close to b as any other point in the column space.
- Therefore, the error vector $Ax_{\text{LS}} - b$ must be perpendicular to that space.

Normal Equations (concluded)

- This means

$$(Ay)^{\mathrm{T}}(Ax_{\text{LS}} - b) = y^{\mathrm{T}}\,(A^{\mathrm{T}}Ax_{\text{LS}} - A^{\mathrm{T}}b) = \mathbf{0}$$

for all y.

- We conclude that any solution x must satisfy the normal equations,

$$A^{\mathrm{T}}Ax = A^{\mathrm{T}}b. \tag{76}$$

Numerical Solutions to LS

- The LS problem is called the full-rank least-squares problem when A has full column rank.
  - Consider the polynomial regression (75) on p. 669.
  - The $m \times n$ matrix has full column rank as long as $a_1, a_2, \ldots, a_m$ contain at least n distinct numbers.
- Since $A^{\mathrm{T}}A$ is then nonsingular, the normal equations (76),

$$A^{\mathrm{T}}Ax = A^{\mathrm{T}}b,$$

can be solved, say, by Gaussian elimination.

Numerical Solutions to LS (concluded)

- The unique solution for normal equations is

$$x_{\text{LS}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

- This is called the ordinary least-squares (OLS) estimator.
- As $A^{\mathrm{T}}A$ is positive definite, the normal equations can be solved by the Cholesky decomposition (p. 662).
- This approach is usually not recommended because its numerical stability is lower than the alternative SVD approach (see text).

An Intuitive Methodology

- Let $\Phi(x) \equiv (1/2) \parallel Ax - b \parallel^2$.
- Define its gradient vector as

$$\bigtriangledown\Phi(x) \equiv \left[\frac{\partial\Phi(x)}{\partial x_1}, \frac{\partial\Phi(x)}{\partial x_2}, \ldots, \frac{\partial\Phi(x)}{\partial x_n} \right]^{\mathrm{T}}.$$

- Then normal equations are exactly $\bigtriangledown\Phi(x) = \mathbf{0}$.
- This method based on calculus can often be derived without appealing to normal equations.

An Intuitive Methodology (continued)

- Take the polynomial regression on p. 669.
- The mean-square error is

$$\Phi(x_0, \ldots, x_n) = \sum_{i=1}^{m} \left[(x_0 + x_1 a_i + \cdots + x_n a_i^n) - b_i \right]^2.$$

- To minimize it, we set

$$\frac{\partial\Phi}{\partial x_j} = 0$$

for $0 \leq j \leq n$.

An Intuitive Methodology (continued)

- These equalities result in

$$\begin{aligned}
\sum_{i=1}^{m} \left[(x_0 + x_1 a_i + \cdots + x_n a_i^n) - b_i \right] &= 0,\\
\sum_{i=1}^{m} a_i \left[(x_0 + x_1 a_i + \cdots + x_n a_i^n) - b_i \right] &= 0,\\
&\vdots\\
\sum_{i=1}^{m} a_i^n \left[(x_0 + x_1 a_i + \cdots + x_n a_i^n) - b_i \right] &= 0.
\end{aligned}$$

An Intuitive Methodology (continued)

- They lead to the linear system,

$$\begin{bmatrix}
\sum_{i=1}^m 1 & \sum_{i=1}^m a_i & \sum_{i=1}^m a_i^2 & \cdots & \sum_{i=1}^m a_i^n\\
\sum_{i=1}^m a_i & \sum_{i=1}^m a_i^2 & \sum_{i=1}^m a_i^3 & \cdots & \sum_{i=1}^m a_i^{n+1}\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
\sum_{i=1}^m a_n & \sum_{i=1}^m a_i^{n+1} & \sum_{i=1}^m a_i^{n+2} & \cdots & \sum_{i=1}^m a_i^{2n}
\end{bmatrix}
\begin{bmatrix} x_0\\ x_1\\ \vdots\\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^m b_i\\ \sum_{i=1}^m a_i b_i\\ \vdots\\ \sum_{i=1}^m a_i^n b_i \end{bmatrix}.$$

- It can be solved by Gaussian elimination.

An Intuitive Methodology (continued)

- Polynomial regression uses $1, x, \ldots, x^n$ as the basis functions.
- In general, we can use $f_0(x), f_1(x), \ldots, f_n(x)$ as the basis functions.
- The mean-square error is

$$\begin{aligned}&\Phi(x_0, \ldots, x_n)\\ = &\sum_{i=1}^{m} \left[\, (x_0 f_0(a_i) + x_1 f_1(a_i) + \cdots + x_n f_n(a_i)) - b_i \,\right]^2 .\end{aligned}$$

- To minimize it, we again set

$$\frac{\partial \Phi}{\partial x_j} = 0, \quad 0 \leq j \leq n.$$

An Intuitive Methodology (continued)

- These equalities result in

$$\begin{aligned}
\sum_{i=1}^{m} f_0(a_i) \left[\, (x_0 f_0(a_i) + x_1 f_1(a_i) + \cdots + x_n f_n(a_i)) - b_i \,\right] &= 0,\\
\sum_{i=1}^{m} f_1(a_i) \left[\, (x_0 f_0(a_i) + x_1 f_1(a_i) + \cdots + x_n f_n(a_i)) - b_i \,\right] &= 0,\\
&\vdots\\
\sum_{i=1}^{m} f_n(a_i) \left[\, (x_0 f_0(a_i) + x_1 f_1(a_i) + \cdots + x_n f_n(a_i)) - b_i \,\right] &= 0.
\end{aligned}$$

An Intuitive Methodology (continued)

- They lead to the linear system,

$$\begin{bmatrix}
\sum_{i=1}^{m} f_0(a_i) f_0(a_i) & \cdots & \sum_{i=1}^{m} f_0(a_i) f_n(a_i)\\
\sum_{i=1}^{m} f_1(a_i) f_0(a_i) & \cdots & \sum_{i=1}^{m} f_1(a_i) f_n(a_i)\\
\vdots & \ddots & \vdots\\
\sum_{i=1}^{m} f_n(a_i) f_0(a_i) & \cdots & \sum_{i=1}^{m} f_n(a_i) f_n(a_i)
\end{bmatrix}
\begin{bmatrix} x_0\\ x_1\\ \vdots\\ x_n \end{bmatrix}
= \begin{bmatrix}
\sum_{i=1}^{m} f_0(a_i)\, b_i\\
\sum_{i=1}^{m} f_1(a_i)\, b_i\\
\vdots\\
\sum_{i=1}^{m} f_n(a_i)\, b_i
\end{bmatrix}.$$

- It can be solved by Gaussian elimination.

An Intuitive Methodology (continued)

- Popular types of basis functions include: Laguerre polynomials, Hermite polynomials, Legendre polynomials, Chebyshev polynomials, Gedenbauer polynomials, and Jacobi polynomials.
- Again, in general, the SVD approach is more stable.

An Intuitive Methodology (concluded)

- And the LS formulation is

$$\begin{bmatrix} f_0(a_1) & f_1(a_1) & f_2(a_1) & \cdots & f_n(a_1) \\ f_0(a_2) & f_1(a_2) & f_2(a_2) & \cdots & f_n(a_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_0(a_m) & f_1(a_m) & f_2(a_m) & \cdots & f_n(a_m) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

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).

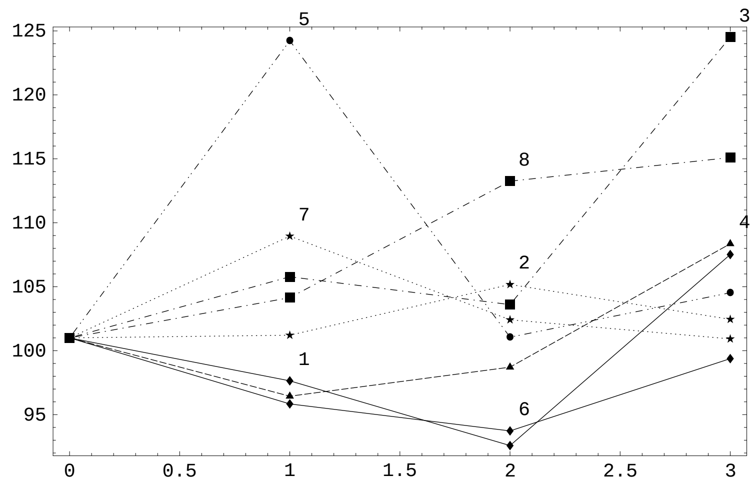
A Numerical Example

- Consider a 3-year American put on a non-dividend-paying stock.
- The put is exercisable at years 1, 2, and 3.
- The strike price $X = 105$.
- The annualized riskless rate is $r = 5\%$.
- The spot stock price is 101.
  - The annual discount factor hence equals 0.951229.
- We use only 8 price paths to illustrate the algorithm.

A Numerical Example (continued)

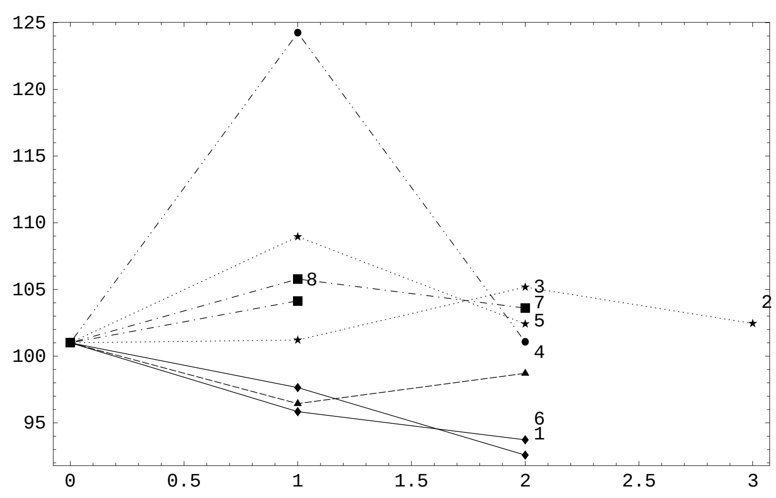
Stock price paths

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



A Numerical Example (continued)

- We use the basis functions $1, x, x^2$.
  - Other basis functions are possible (p. 681).
- 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 Numerical Example (continued)

Cash flows at year 3

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

A Numerical Example (continued)

- The cash flows at year 3 are the exercise value if the put is in the money.
- Only 4 paths are in the money: 2, 5, 6, 7.
- Some of the cash flows may not occur if the put is exercised earlier, which we will find out step by step.
- Incidentally, the European counterpart has a value of

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

A Numerical Example (continued)

- We move on to year 2.
- For each state that is in the money at year 2, we must decide whether to exercise it.
- There are 6 paths for which the put is in the money: 1, 3, 4, 5, 6, 7.
- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.
  - If there were none, we would move on to year 1.

A Numerical Example (continued)

- Let x denote the stock prices at year 2 for those 6 paths.
- Let y denote the corresponding discounted future cash flows (at year 3) if the put is not exercised at year 2.

A Numerical Example (continued)

Regression at year 2

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

A Numerical Example (continued)

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

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

- f estimates the continuation value conditional on the stock price at year 2.
- We next compare the immediate exercise value and the continuation value.

A Numerical Example (continued)

Optimal early exercise decision at year 2

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

A Numerical Example (continued)

- Amazingly, the put should be exercised in all 6 paths: 1, 3, 4, 5, 6, 7.
- Now, any positive cash flow at year 3 should be set to zero for these paths as the put is exercised before year 3.
  - They are paths 5, 6, 7.
- Hence the cash flows on p. 690 become the next ones.

A Numerical Example (continued)

Cash flows at years 2 & 3

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

A Numerical Example (continued)

- We move on to year 1.
- For each state that is in the money at year 1, we must decide whether to exercise it.
- There are 5 paths for which the put is in the money: 1, 2, 4, 6, 8.
- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.
  - If there were none, we would move on to year 0.

A Numerical Example (continued)

- Let x denote the stock prices at year 1 for those 5 paths.
- Let y denote the corresponding discounted future cash flows if the put is not exercised at year 1.
- From p. 698, we have the following table.

A Numerical Example (continued)

Regression at year 1

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

A Numerical Example (continued)

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

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

- f estimates the continuation value conditional on the stock price at year 1.
- We next compare the immediate exercise value and the continuation value.

A Numerical Example (continued)

Optimal early exercise decision at year 1

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

A Numerical Example (continued)

- The put should be exercised for 1 path only: 8.
- Now, any positive future cash flow should be set to zero for this path as the put is exercised before years 2 and 3.
  - But there is none.
- Hence the cash flows on p. 698 become the next ones.
- They also confirm the plot on p. 689.

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 705,

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

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.
- This is much larger than the European put's value of 1.3680 (p. 691).