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: \( \textbf{for } i = 1, 2, 3, \ldots, N \textbf{ do} \)
3: \( P := S; \) hit := 0;
4: \( \textbf{for } j = 1, 2, 3, \ldots, n \textbf{ do} \)
5: \( P := P \times e^{(r-\sigma^2/2)(T/n)+\sigma\sqrt{T/n}} \xi; \) \{By Eq. (120) on p. 853.\}
6: \( \textbf{if } P \geq H \textbf{ then} \)
7: \( \text{hit := 1;} \)
8: \( \text{break;} \)
9: \( \textbf{end if} \)
10: \( \textbf{end for} \)
11: \( \textbf{if } \text{hit} = 0 \textbf{ then} \)
12: \( C := C + \max(P - X, 0); \)
13: \( \textbf{end if} \)
14: \( \textbf{end for} \)
15: \( \textbf{return } Ce^{-rT}/N; \)
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 knock-out probabilities are underestimated.

$^a$Shevchenko (2003).
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 probabilities 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.

• This methodology is called the Brownian bridge approach.

• Formally, we have

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

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

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

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

\[ ^{a}\text{Here, } \Delta t \text{ is an arbitrary positive real number.} \]
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\(^a\) 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.

1: \( C := 0; \)
2: \( \text{for } i = 1, 2, 3, \ldots, N \text{ do} \)
3: \( P := S \times e^{(r-q-\sigma^2/2)T+\sigma\sqrt{T} \xi()}; \)
4: \( \text{if } (S < H \text{ and } P < H) \text{ or } (S > H \text{ and } P > H) \text{ then} \)
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: \( \text{end if} \)
7: \( \text{end for} \)
8: \( \text{return } Ce^{-rT}/N; \)
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 contains $n$ barriers.

• Multiply the probabilities for the $n$ 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 $(\mu, \sigma)$ 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:

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

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

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

---

*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 many times to average the discounted payoff.

---

$^a$The IG distribution can be very efficiently sampled (Michael, Schucany, & Haas, 1976).
Brownian Bridge Approach to Pricing Lookback Options\textsuperscript{a}

- By Lemma 21(1) (p. 876),

\[
F_{\text{max}}(y) \triangleq \text{Prob}\left[\max_{0 \leq t \leq T} S(t) < y \mid S(0), S(T)\right] = 1 - \exp\left[-\frac{2 \ln(y/S(0)) \ln(y/S(T))}{\sigma^2 T}\right].
\]

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

\textsuperscript{a}El Babsiri & Noel (1998).
Brownian Bridge Approach to Pricing Lookback Options (continued)

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

- In other words,

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

\(^{\text{a}}\)

\(^{\text{a}}\)This is called the inverse-transform technique (see p. 259 of the textbook).
Brownian Bridge Approach to Pricing Lookback Options (continued)

• Equivalently,

\[
\ln(1 - x) = -\frac{2 \ln(y/S(0)) \ln(y/S(T))}{\sigma^2 T} = -\frac{2}{\sigma^2 T} \left\{ \left[ \ln(y) - \ln S(0) \right] \left[ \ln(y) - \ln S(T) \right] \right\}.
\]
Brownian Bridge Approach to Pricing Lookback Options (continued)

- There are two solutions for $\ln y$.

- But only one is consistent with $y \geq \max(S(0), S(T))$:

$$\ln y = \ln(S(0) S(T)) + \sqrt{\left(\ln \frac{S(T)}{S(0)}\right)^2 - 2\sigma^2 T \ln(1 - x)}.$$
Brownian Bridge Approach to Pricing Lookback Options (concluded)

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

1: $C := 0$;
2: for $i = 1, 2, 3, \ldots, N$ do
3: $P := S \times e^{(r-q-\sigma^2/2)T+\sigma\sqrt{T} \xi(\cdot)}; \{\text{By Eq. (120) on p. 853.}\}$
4: $Y := \exp \left[ \ln(SP)+\sqrt{(\ln \frac{P}{S})^2-2\sigma^2T\ln[1-U(0,1)]} \right];$
5: $C := C + (Y - P);$ 
6: end for
7: return $Ce^{-rT}/N;$
Pricing Lookback Options without Brownian Bridge

• Suppose we do not draw $S(T)$ in simulation.

• Now, the distribution function of the maximum logarithmic stock price is

$$\text{Prob} \left[ \max_{0 \leq t \leq T} \frac{\ln S(t)}{S(0)} < y \right]$$

$$= 1 - N \left( \frac{-y + \left( r - q - \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}} \right)$$

$$- e^{-2y \left( r - q - \frac{\sigma^2}{2} \right)} \frac{2y \left( r - q - \frac{\sigma^2}{2} \right)}{\sigma^2} N \left( \frac{-y - \left( r - q - \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}} \right), \quad y \geq 0.$$ 

• The inverse of that is much harder to calculate.

\[\text{aA. N. Borodin & Salminen (1996).}\]
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 $\text{Var}[\frac{Y_1 + Y_2}{2}]$ 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 \( g(\boldsymbol{x}) \), 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 $\xi$ for the second estimate $g(x').$
- Instead, generate the sample path $x' \stackrel{\Delta}{=} (x'_1, x'_2, \ldots, x'_n)$ from $-\xi_1, -\xi_2, \ldots, -\xi_n.$
- Compute $g(x').$
- Output $(g(x) + g(x'))/2.$
- Repeat the above steps for as many times as required by accuracy.
Variance Reduction: Conditioning

• We are interested in estimating $E[X]$.

• Suppose here is a random variable $Z$ such that $E[X | Z = z]$ can be efficiently and precisely computed.

• $E[X] = E[E[X | Z]]$ by the law of iterated conditional expectations.

• Hence the random variable $E[X | Z]$ is also an unbiased estimator of $E[X]$. 
Variance Reduction: Conditioning (concluded)

• As

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

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

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

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

• The procedure can be repeated a few times to reduce the variance.
Control Variates

- Use the analytic solution of a “similar” yet “simpler” problem to improve the solution.

- Suppose we want to estimate $E[X]$ and there exists a random variable $Y$ with a known mean $\mu \triangleq E[Y]$.

- 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

\[
\text{Var}[W] = \text{Var}[X] + \beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X,Y],
\]  
(121)

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

\[
\beta^2 \text{Var}[Y] + 2\beta \text{Cov}[X,Y] < 0.
\]  
(122)
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.\(^a\)
  - Arithmetic Asian options can be priced by choosing $Y$ to be the otherwise identical geometric Asian option’s price and $\beta = -1$.

- This approach is much more effective than the antithetic-variates method.\(^b\)

\(^a\)Hull & White (1988).
\(^b\)Boyle, Broadie, & Glasserman (1997).
Choice of \( Y \)

- In general, the choice of \( Y \) is ad hoc,\(^a\) and experiments must be performed to confirm the wisdom of the choice.

- Try to match calls with calls and puts with puts.\(^b\)

- On many occasions, \( Y \) is a discretized version of the derivative that gives \( \mu \).
  - Discretely monitored geometric Asian option vs. the continuously monitored version.\(^c\)

- The discrepancy can be large (e.g., lookback options).\(^d\)

\(^a\)But see Dai (B82506025, R86526008, D8852600), C. Chiu (B90201037, R94922072), \& Lyuu (2015, 2018).

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

\(^c\)Priced by formulas (55) on p. 442.

\(^d\)Contributed by Mr. Tsai, Hwai (R92723049) on May 12, 2004.
Optimal Choice of $\beta$

- Equation (121) on p. 897 is minimized when

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

- It is called beta in the book.

- For this specific $\beta$,

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

where $\rho_{X,Y}$ is the correlation between $X$ and $Y$. 
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 $\epsilon^2 \text{Var}[Y]$.$^a$

$^a$Han & Y. Lai (2010).
A Pitfall

• A potential pitfall is to sample $X$ and $Y$ independently.
• In this case, $\text{Cov}[X, Y] = 0$.
• Equation (121) on p. 897 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 $O(1/\sqrt{N})$ 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 $A \triangleq [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \mathbb{R}^{m \times n}$, denote an $m \times n$ matrix.

• It can also be represented as $[a_1, a_2, \ldots, a_n]$ where $a_i \in \mathbb{R}^m$ are vectors.
  – Vectors are column vectors unless stated otherwise.

• $A$ is a square matrix when $m = n$.

• The rank of a matrix is the largest number of linearly independent columns.
Definitions and Basic Results (continued)

• A square matrix $A$ is said to be symmetric if $A^T = A$.

• A real $n \times n$ matrix

\[ A \triangleq [a_{ij}]_{i,j} \]

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

\[ I \triangleq \text{diag}[1, 1, \ldots, 1]. \]
Definitions and Basic Results (concluded)

- A matrix has full column rank if its columns are linearly independent.

- A real symmetric matrix $A$ is positive definite if

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

  for any nonzero vector $x$.

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

- Positive definite matrices can be factored as
  \[ A = LL^T, \]
  called the Cholesky decomposition.
  - Above, \( L \) is a lower triangular matrix.
Generation of Multivariate Distribution

- Let $\mathbf{x} \triangleq [x_1, x_2, \ldots, x_n]^T$ be a vector random variable with a positive definite covariance matrix $C$.

- As usual, assume $E[\mathbf{x}] = 0$.

- This covariance structure can be matched by $P\mathbf{y}$.
  - $\mathbf{y} \triangleq [y_1, y_2, \ldots, y_n]^T$ is a vector random variable with a covariance matrix equal to the identity matrix.
  - $C = PP^T$ is the Cholesky decomposition of $C$.

---

*aWhat if $C$ is not positive definite? See Y. Y. Lai (R93942114) & Lyuu (2007).*
Generation of Multivariate Distribution (concluded)

- For example, suppose
  \[
  C = \begin{bmatrix}
  1 & \rho \\
  \rho & 1 
  \end{bmatrix}.
  \]

- Then
  \[
  P = \begin{bmatrix}
  1 & 0 \\
  \rho & \sqrt{1 - \rho^2} 
  \end{bmatrix}
  \]

as \(PP^T = C\).

\(^a\)Recall Eq. (28) on p. 179.
Generation of Multivariate Normal Distribution

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = P P^T$.
  - First, generate independent standard normal distributions $y_1, y_2, \ldots, y_n$.
  - Then
    
    $P[y_1, y_2, \ldots, y_n]^T$

    has the desired distribution.
  - These steps can then be repeated.
Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives (pp. 809ff).

- 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); 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 \( C = PP^T. \)

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

• Let \( \xi' = P\xi. \)

• Similar to Eq. (120) on p. 853, 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. (120) on p. 853.
Least-Squares Problems

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

\[
\min_{x \in \mathbb{R}^n} \| Ax - b \|,
\]

where \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \), and \( m \geq n \).

• The LS problem is called regression analysis in statistics and is equivalent to minimizing the mean-square error.

• Often written as

\[
Ax = b.
\]
Polynomial Regression

- In polynomial regression, $x_0 + x_1 x + \cdots + x_n x^n$ is used to fit the data \{(a_1, b_1), (a_2, b_2), \ldots, (a_m, b_m)\}.

- This leads to the LS problem,

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

- Consult p. 273 of the textbook for solutions.
American Option Pricing by Simulation

- The continuation value of an American option is the conditional expectation of the payoff from keeping the option alive now.
- The option holder must compare the immediate exercise value and the continuation value.
- In standard Monte Carlo simulation, each path is treated independently of other paths.
- But the decision to exercise the option cannot be reached by looking at one path alone.
The Least-Squares Monte Carlo Approach

- The continuation value can be estimated from the cross-sectional information in the simulation by using least squares.\(^a\)

- The result is a function (of the state) for estimating the continuation values.

- Use the function to estimate the continuation value for each path to determine its cash flow.

- This is called the least-squares Monte Carlo (LSM) approach.

\(^a\)Longstaff & Schwartz (2001).
The Least-Squares Monte Carlo Approach (concluded)

- The LSM is provably convergent.\textsuperscript{a}

- The LSM can be easily parallelized.\textsuperscript{b}
  - Partition the paths into subproblems and perform LSM on each of them independently.
  - The speedup is close to linear (i.e., proportional to the number of cores).

- Surprisingly, accuracy is not affected.

\textsuperscript{a}Clément, Lamberton, & Protter (2002); Stentoft (2004).
\textsuperscript{b}K. Huang (B96902079, R00922018) (2013); C. W. Chen (B97902046, R01922005) (2014); C. W. Chen (B97902046, R01922005), K. Huang (B96902079, R00922018) & Lyuu (2015).
A Numerical Example

- Consider a 3-year American put on a non-dividend-paying stock.
- The put is exercisable at years 0, 1, 2, and 3.
- The strike price $X = 105$.
- The annualized riskless rate is $r = 5\%$.
  - The annual discount factor hence equals $0.951229$.
- The current stock price is 101.
- We use 8 price paths to illustrate the algorithm.
## A Numerical Example (continued)

Stock price paths

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

- We use the basis functions $1, x, x^2$.
  - Other basis functions are possible.\(^a\)

- The plot next page shows the final estimated optimal exercise strategy given by LSM.

- We now proceed to tackle our problem.

- The idea is to calculate the cash flow along each path, using information from all paths.

\(^a\)Laguerre polynomials, Hermite polynomials, Legendre polynomials, Chebyshev polynomials, Gedenbauer polynomials, or Jacobi polynomials.
A Numerical Example (continued)

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

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

\[
0.951229^3 \times \frac{2.5476 + 0.4685 + 5.6212 + 4.0775}{8} = 1.3680.
\]
A Numerical Example (continued)

- We move on to year 2.
- For each state that is in the money at year 2, we must decide whether to exercise it.
- There are 6 paths for which the put is in the money: 1, 3, 4, 5, 6, 7 (p. 923).
- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.
  - If there were none, move on to year 1.
A Numerical Example (continued)

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

Regression at year 2

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

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

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

- $f(x)$ 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

<table>
<thead>
<tr>
<th>Path</th>
<th>Exercise</th>
<th>Continuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.4185</td>
<td>$f(92.5815) = 2.2558$</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>1.3990</td>
<td>$f(103.6010) = 1.1168$</td>
</tr>
<tr>
<td>4</td>
<td>6.2880</td>
<td>$f(98.7120) = 1.5901$</td>
</tr>
<tr>
<td>5</td>
<td>3.9436</td>
<td>$f(101.0564) = 1.3568$</td>
</tr>
<tr>
<td>6</td>
<td>11.2730</td>
<td>$f(93.7270) = 2.1253$</td>
</tr>
<tr>
<td>7</td>
<td>2.5823</td>
<td>$f(102.4177) = 1.2266$</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
A Numerical Example (continued)

• Amazingly, the put should be exercised in all 6 paths: 1, 3, 4, 5, 6, 7.

• Now, any positive cash flow at year 3 vanishes for these paths as the put is exercised before year 3 (p. 923).
  – They are paths 5, 6, 7.

• The cash flows on p. 927 become the ones on next slide.
A Numerical Example (continued)

Cash flows at years 2 & 3

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

- We move on to year 1.
- For each state that is in the money at year 1, we must decide whether to exercise it.
- There are 5 paths for which the put is in the money: 1, 2, 4, 6, 8 (p. 923).
- Only in-the-money paths will be used in the regression because they are where early exercise is relevant.
  - If there were none, 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. 935, we have the following table.
A Numerical Example (continued)

Regression at year 1

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

• We regress $y$ on 1, $x$, and $x^2$.

• The result is

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

• $f(x)$ 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

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

- The put should be exercised for 1 path only: 8.
  - Note that $f(104.1475) < 0$.

- Now, any positive future cash flow vanishes for this path.
  - But there is none.

- The cash flows on p. 935 become the ones on next slide.

- They also confirm the plot on p. 926.
### A Numerical Example (continued)

#### Cash flows at years 1, 2, & 3

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

- We move on to year 0.
- The continuation value is, from p 942,

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

\[
= 4.66263.
\]
A Numerical Example (concluded)

- As this is larger than the immediate exercise value of
  \[105 - 101 = 4,\]
  the put should not be exercised at year 0.
- Hence the put’s value is estimated to be 4.66263.
- Compare this with the European put’s value of 1.3680 (p. 928).
Time Series Analysis
The historian is a prophet in reverse.
— Friedrich von Schlegel (1772–1829)
GARCH Option Pricing

• Options can be priced when the underlying asset’s return follows a GARCH process.

• Let $S_t$ denote the asset price at date $t$.

• Let $h_t^2$ be the conditional variance of the return over the period $[t, t + 1)$ given the information at date $t$.
  
  - “One day” is merely a convenient term for any elapsed time $\Delta t$.

---

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

- Adopt the following risk-neutral process for the price dynamics:\(^a\)

\[
\ln \frac{S_{t+1}}{S_t} = r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}, \tag{123}
\]

where

\[
h_{t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2, \tag{124}
\]

\[
\epsilon_{t+1} \sim N(0, 1) \text{ given information at date } t,
\]

\[
r = \text{daily riskless return},
\]

\[
c \geq 0.
\]

\(^a\)Duan (1995).
GARCH Option Pricing (continued)

- The five unknown parameters of the model are $c$, $h_0$, $\beta_0$, $\beta_1$, and $\beta_2$.

- It is postulated that $\beta_0, \beta_1, \beta_2 \geq 0$ to make the conditional variance positive.

- There are other inequalities to satisfy (see text).

- The above process is called the nonlinear asymmetric GARCH (or NGARCH) model.
GARCH Option Pricing (continued)

- It captures the volatility clustering in asset returns first noted by Mandelbrot (1963).\(^a\)
  - When \(c = 0\), a large \(\epsilon_{t+1}\) results in a large \(h_{t+1}\), which in turns tends to yield a large \(h_{t+2}\), and so on.

- It also captures the negative correlation between the asset return and changes in its (conditional) volatility.\(^b\)
  - For \(c > 0\), a positive \(\epsilon_{t+1}\) (good news) tends to decrease \(h_{t+1}\), whereas a negative \(\epsilon_{t+1}\) (bad news) tends to do the opposite.

\(^a\)“... large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes ...”

\(^b\)Noted by Black (1976): Volatility tends to rise in response to “bad news” and fall in response to “good news.”
GARCH Option Pricing (continued)

• This is called the leverage effect.
  – A falling stock price raises the fixed costs, relatively speaking.a

• With $y_t \triangleq \ln S_t$ denoting the logarithmic price, the model becomes

$$y_{t+1} = y_t + r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}.$$  \hspace{1cm} (125)

• The pair $(y_t, h_t^2)$ completely describes the current state.

\(^a\)Black (1992).
GARCH Option Pricing (concluded)

- The conditional mean and variance of $y_{t+1}$ are clearly

$$
E[y_{t+1} \mid y_t, h_t] = y_t + r - \frac{h_t^2}{2}, \quad (126)
$$

$$
\text{Var}[y_{t+1} \mid y_t, h_t] = h_t^2. \quad (127)
$$

- Finally, given $(y_t, h_t^2)$, the correlation between $y_{t+1}$ and $h_{t+1}$ equals

$$
- \frac{2c}{\sqrt{2 + 4c^2}},
$$

which is negative for $c > 0$. 
GARCH Model: Inferences

- Suppose the parameters $c, h_0, \beta_0, \beta_1,$ and $\beta_2$ are given.
- Then we can recover $h_1, h_2, \ldots, h_n$ and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ from the prices $S_0, S_1, \ldots, S_n$ under the GARCH model (123) on p. 948.
- This property is useful in statistical inferences.