Pricing American Options

• Pricing American option = solving optimal stopping problem
• Deterministic numerical methods (binomial lattices, PDE methods) in dimensions 1-3
• Lots of problems in dimension 5-80
• Monte Carlo potentially attractive --- how to combine simulation and dynamic programming?
• We give
  – overview of a class of related methods and ideas
  – a new result on a regression-based method
Monte Carlo for American Options: A Circle of Ideas

- Weighted backward induction on independent paths
  - Weights from likelihood ratios
  - Weights from calibration
  - Weights from regression
- High bias from peeking ahead
- Low bias from suboptimality
- Duality
Formulation

Exercise opportunities at $0 = t_0 < t_1 < \ldots < t_m$

Underlying Markov process $X(t)$, $X_i = X(t_i)$

Exercise at $t_i$ in state $x$ pays $h_i(x) \geq 0$

Value at date $i$ in state $x$

$$V_i(x) = \sup_{\tau \in T_i} E[h_\tau(X_\tau) | X_i = x]$$

$T_i = \{i, i+1, \ldots, m\} – \text{valued stopping times}$

Find $V_0(X_0)$
Dynamic Programming

\[ V_m = h_m \]
\[ V_i(x) = \max\{h_i(x), E[V_{i+1}(X_{i+1}) | X_i = x]\}, \quad i = 0, 1, \ldots, m - 1 \]

Continuation values

\[ C_i(x) = E[V_{i+1}(X_{i+1}) | X_i = x] \]
\[ V_i = \max(h_i, C_i) \]

\[ C_m = 0 \]
\[ C_i(x) = E[\max\{h_{i+1}(X_{i+1}), C_{i+1}(X_{i+1})\} | X_i = x] \]
Stochastic Mesh
Weighted Backward Induction

$$X_{ij} \leftarrow \cdots \leftarrow W_{jk} \cdots \leftarrow X_{i+1,k}, \hat{V}_{i+1,k}$$

Estimate
$$C_i(X_{ij}) = E[V_{i+1}(X_{i+1}) \mid X_i = X_{ij}]$$

using
$$\hat{C}_{ij} = \frac{1}{b} \sum_{k=1}^{b} W_{jk} \hat{V}_{i+1,k}$$

Weights correct for the fact that downstream nodes have the "wrong" distribution: $X_{i+1,k}$ does not have the law of $(X(t_{i+1}) \mid X(t_i) = X_{ij})$
High Bias

Basic algorithm

\[ \hat{V}_M = h_M \]

\[ \hat{V}_{ij} = \max \left\{ h_i(X_{ij}), \frac{1}{b} \sum_{k=1}^{b} W_{jk} \hat{V}_{i+1,k} \right\} \]

Suppose weights "work" for true value

\[ E[W^i_{jk} V_{i+1}(X_{i+1,k}) \mid X_i = X_{ij}] = C_i(X_i) \]

Then induction shows

\[ E[\hat{V}_{ij} \mid X_{ij}] \geq V_i(X_{ij}) \]

by Jensen's inequality and convexity of max
Low Bias: Mesh-Defined Stopping Rule

Simulate independent path
At each exercise date $t_i$ estimate continuation value

$$
\hat{C}_i(X(t_i)) = \frac{1}{b} \sum_{k=1}^{b} W_k^i(X(t_i)) \hat{V}_{i+1,k}
$$

Stopping rule \( \hat{\tau} = \min \left\{ i : h_i(X(t_i)) \geq \hat{C}_i(X(t_i)) \right\} \)

Estimate \( h_{\hat{\tau}}(X(t_{\hat{\tau}})) \) (separates decision from payoff)
Stopping rule is suboptimal so

$$
E[h_{\hat{\tau}}(X(t_{\hat{\tau}}))] \leq V_0(X_0)
$$
Interval Estimate
Choice of Weights

Unweighted average

\[
\frac{1}{b} \sum_{k=1}^{b} V_{i+1}(X_{i+1,k}) \rightarrow E[V_{i+1}(X_{i+1})] \neq E[V_{i+1}(X_{i+1}) \mid X_i = X_{ij}]
\]

Suggests weights should be likelihood ratios relating conditional and unconditional distributions

By Radon-Nikodym Theorem, these are the only weights that work, if we consider a sufficiently rich class of \( V_{i+1} \)

Simplest case

\[
W_{jk}^i = \frac{f_i(X_{ij}, X_{i+1,k})}{g_{i+1}(X_{i+1,k})} = \frac{\text{transition density}}{\text{marginal density}}
\]

Convergence results (Broadie-Glasserman 1997, Avramidis-Matzinger 2002) for alternative constructions and LR weights

Central limit theorem? Choice of conditioning?
Weights Through Calibration

Transition density

- may be unknown
- may fail to exist (dim(state)>dim(Brownian motion))

Weights through calibration (B-G-Ha 2001)

Given known functions

\[ \psi(x) = E[\Psi(X_{i+1}) \mid X_i = x] \]

choose weights that “price” \( \Psi \) correctly
Weights Through Calibration

\[ \psi(x) = E[\Psi(X_{i+1}) \mid X_i = x] \]

Choose weights so that

\[ \frac{1}{b} \sum_{k=1}^{b} W_{jk}^i \Psi(X_{i+1,k}) = \psi(X_{ij}) \]

From all feasible solutions, choose one minimizing

\[ \sum_{k=1}^{b} H(W_{jk}^i), \text{ some convex } H \text{ (e.g., quadratic, entropy)} \]

These are maximally uniform weights (in majorization ordering)
Weights Through Calibration

- Multiperiod version of Weighted Monte Carlo (in the sense of Avellaneda et al.)
- Analysis of WMC in Glasserman-Yu (2003) shows that over a single period
  - In limit as $b$ increases, equivalent to a (signed) change of measure determined by $G = (H')^{-1}$
  \[
  \frac{1}{b} \sum_{k=1}^{b} W_{jk} V(X_{i+1,k}) \Rightarrow E[V(X_{i+1})G(\mu + \lambda' \Psi(X_{i+1})) | X_i]
  \]
  - Correctly prices $V$ in span of $\Psi$
  - For quadratic $H$, equivalent to regression on $\Psi$ rather than $\psi$
  - Regression on $\Psi$ rather than $\psi$ produces less-dispersed estimates

Multiple periods? Choice of objectives? CLT?
Regression-Based Methods


Posit approximation of the form

\[ C_i(x) \approx \beta' \psi(x) \]
\[ V_{i+1}(X_{i+1,j}) = \beta' \psi(X_{ij}) + \epsilon \]

Estimate coefficients through regression

\[ \psi(X_{i1}) \quad \ldots \quad \psi(X_{ib}) \]
\[ \hat{V}_{i+1,1} \quad \hat{V}_{i+1,2} \quad \hat{V}_{i+1,b} \]
Regression-Implied Mesh Weights

\[ \hat{\beta} = B_{\psi}^{-1} B_{\psi V}, \quad B_{\psi V} = \frac{1}{b} \sum_{k=1}^{b} \psi(X_{ik}) \hat{V}_{i+1,k} \]

SO

\[ \hat{C}_{ij} = \psi(X_{ij})' \hat{\beta} \]

\[ = \psi(X_{ij})' B_{\psi}^{-1} B_{\psi V} \]

\[ = \frac{1}{b} \sum_{k=1}^{b} \psi(X_{ij})' B_{\psi}^{-1} \psi(X_{ik}) \hat{V}_{i+1,k} \]

\[ W_{jk}^{i} \]
Backward Induction

Tsitsiklis-Van Roy (1999, 2001):

\[ \hat{V}_{ij} = \max(h_i(X_{ij}), \hat{C}_i(X_{ij})) \]

"High" mesh estimator with regression weights

Longstaff-Schwartz (2001):

\[ \hat{V}_{ij} = \begin{cases} h_i(X_{ij}), & \text{if } h_i(X_{ij}) \geq \hat{C}_i(X_{ij}) \\ \hat{V}_{i+1,j}, & \text{if } h_i(X_{ij}) < \hat{C}_i(X_{ij}) \end{cases} \]

Interleaves "high" and "low"

Numerically, T-V with second pass similar to L-S
More on Weights

Special form of regression weights allows fast calc.

\[
\hat{C}_{ij} = \frac{1}{b} \sum_{k=1}^{b} \psi(X_{ij})' B_{\psi}^{-1} \psi(X_{ik}) \hat{V}_{i+1,k}, \quad O(b^2)
\]

\[
\hat{C}_{ij} = \beta' \psi(X_{ij}), \quad O(bK), \quad K = \text{dim } \psi \ll b
\]

big advantage

Requires little to get some price
More on Weights

- Regression weights ignore spacing $t_{i+1} - t_i$
- Weights can be surprising

![Graph showing comparison between Regression Weights (5 polynomials) and LR Weights](image)
Convergence

- Convergence of regression-based methods requires increase in $K$ as well as $b$
- Clément, Lamberton, Protter (2002) prove convergence of L-S method as first $b$ then $K$ increase
- What about rate as both increase together?
- We consider (joint work with Bin Yu)
  - Brownian motion, geometric Brownian motion
  - polynomial $\psi$
  - worst-case convergence
  - single-period and multiple periods
Formulation: Single-Period Brownian Case

- Dates \( t_1 < t_2 \)
- Hermite polynomials \( \psi_k(X(t_1)), k = 0,1,...,K \)
  normalized so \( E[\psi_k^2(X(t_1))] = 1 \), makes \( B_\psi = I \)
- Downstream value spanned:
  \[
  Y = \sum_{k=0}^{K} a_k \psi_k(X(t_2))
  \]
  \[
  E[Y \mid X(t_1)] = \sum_{k=0}^{K} \beta_k \psi_k(X(t_1))
  \]
- Criterion:
  \[
  \text{MSE}(\hat{\beta}) = \sum_k \text{Var}[\hat{\beta}_k] = \int [\beta' \psi(x) - \beta' \psi(x)]^2 \phi(x) \, dx
  \]
Theorem

Let \( c = 2 \log \left(2 + \sqrt{t_2 / t_1}\right)\), \( N = \# \text{paths} \)

If \( K_N = (1 - \delta) \log N / c \), for some \( \delta > 0 \)

\[
\lim_{N \to \infty} \sup_{\|\beta\| = 1} \text{MSE}(\hat{\beta}) = 0
\]

If \( K_N = (1 + \delta) \log N / c \), for some \( \delta > 0 \)

\[
\lim_{N \to \infty} \sup_{\|\beta\| = 1} \text{MSE}(\hat{\beta}) = \infty
\]

Worst case attained at \( \beta = (0,...,0,1)' \)
# Estimated MSE

<table>
<thead>
<tr>
<th>$N$</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
<th>16000</th>
<th>32000</th>
<th>64000</th>
<th>128000</th>
</tr>
</thead>
<tbody>
<tr>
<td>K=1</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>1.67</td>
<td>0.31</td>
<td>0.17</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>5.6</td>
<td>3.0</td>
<td>1.6</td>
<td>1.73</td>
<td>0.36</td>
<td>0.18</td>
<td>0.09</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>52.7</td>
<td>23.4</td>
<td>13.5</td>
<td>6.0</td>
<td>3.1</td>
<td>1.5</td>
<td>0.8</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>6</td>
<td>427.2</td>
<td>155.7</td>
<td>93.3</td>
<td>38.4</td>
<td>24.0</td>
<td>10.8</td>
<td>6.2</td>
<td>3.1</td>
<td>1.5</td>
</tr>
<tr>
<td>7</td>
<td>2403</td>
<td>1202</td>
<td>600.8</td>
<td>300.4</td>
<td>150.2</td>
<td>75.1</td>
<td>37.5</td>
<td>18.8</td>
<td>9.4</td>
</tr>
<tr>
<td>8</td>
<td>11447</td>
<td>5723</td>
<td>2862</td>
<td>1431</td>
<td>715.4</td>
<td>357.7</td>
<td>178.9</td>
<td>89.4</td>
<td>44.7</td>
</tr>
<tr>
<td>9</td>
<td>9856</td>
<td>4928</td>
<td>2464</td>
<td>1232</td>
<td>616</td>
<td>308</td>
<td>154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6109</td>
<td>3054</td>
<td>1527</td>
<td>764</td>
<td>381</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>2810</td>
<td>1405</td>
<td>702</td>
<td>447</td>
<td>236</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bound 2.5 2.8 3.1 3.4 3.7 3.9 4.2 4.5 4.8
Theorem: Lognormal Case

$$\psi_k(X(t), t) = X(t)^k e^{-k^2 t/2}$$

If \( K_N = \sqrt{\frac{(1 - \delta) \log N}{5t_1 + t_2}} \), for some \( \delta > 0 \)

$$\lim_{N \to \infty} \sup_{\|\beta\|=1} \text{MSE}(\hat{\beta}) = 0$$

If \( K_N = \sqrt{\frac{(1 + \delta) \log N}{3t_1 + t_2}} \), for some \( \delta > 0 \)

$$\lim_{N \to \infty} \sup_{\|\beta\|=1} \text{MSE}(\hat{\beta}) = \infty$$

i.e., need \( N = O(\exp(K^2)) \)
Multiperiod Results

- Independent paths at each step (for tractability)
- “Spanning” assumption not preserved by backward induction
- We assume

\[ E[h_n^4(X_n)] \leq (t_n / t_{n-1})^{2K} E[\psi_{nK}(X_n)^4] \]

- Get general bound on error

\[ E[\|C_n - \hat{C}_n\|^2] \]

- Single-period rates sufficient for convergence of multiperiod problem
Comments

- No free lunch
- Just polynomials? Negative rate determined by "co-kurtosis"

\[
\frac{E[\psi_K(X(t_2),t_2)^2 \psi_K(X(t_1),t_1)^2]}{(E[\psi_K(X(t_2),t_2)\psi_K(X(t_1),t_1)])^2}
\]

Numerically find exponential growth for some bounded families

- How relevant is worst case?
- Egloff (2003) gets more positive conclusion with bounds based on VC-dimension (for bounded payoffs)
Concluding Remarks

• Several techniques turn out to be closely related
• Differ in
  – choice of weights
  – combinations of high- and low-bias elements
• Other strategies for choosing weights?
• Dual formulation (Haugh-Kogan 2001, Rogers 2001)
  – minimizing over martingales instead of maximizing over stopping times
  – combine with any suboptimal policy to get bounds
  – when regression is valid, residuals give optimal martingale
  – "High" estimate has same form as dual
• Lots of questions about convergence remain open
**Dual Formulation**

Haugh & Kogan (2001), Rogers (2001)
For any martingale $0 = M_0, M_1, \ldots, M_m$

$$V_0(X_0) \leq E\left[\max_{0 \leq k \leq m} \{ h_k(X_k) - M_k \} \right]$$

Equality if $M$ is the martingale part of $V$ (in the sense of Doob-Meyer decomposition)

Andersen-Broadie (2001): Given any stopping rule, use martingale part of (suboptimal) value; this gives upper bound to accompany low estimate

Is this upper bound related to other methods?
More on Duality

• If the regression is valid in the sense that

\[ V_{i+1}(X_{i+1}) = \beta' \psi(X_i) + \varepsilon_{i+1}, \quad E[\varepsilon_{i+1} \mid X_i] = 0 \]

then residuals define the optimal martingale

\[ M_i = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_i \]

Dual is the high estimator

• Every high estimator can be written as

\[
\max_{0 \leq k \leq m} \left\{ h_k(X_k) - \sum_{i=1}^{k} \varepsilon_i \right\}, \quad \varepsilon_{i+1} = \hat{V}_{i+1}(X_{i+1}) - \hat{C}_i(X_i)
\]

these may not be martingale differences