Lévy-Sheffer Systems and the Longstaff-Schwartz Algorithm for American Option Pricing

Stefan Gerhold
TU Vienna

October 18, 2008

Overview

- Motivation
- Dynamic programming
- The Longstaff-Schwartz algorithm
- Overview of convergence results
- New convergence results for certain Lévy processes

Motivation

- Many financial instruments have Bermudan features (Callable, flippable)
- Bermudan swaptions
- Cancellable structured notes
- Structured notes with flip options

A Typical Structured Note

- 25 semiannual coupons

$$xx \cdot (EUR10Y - EUR2Y), \text{floored at } xx, \text{ceiled at } xx.$$

- structure-wide floor of $xx\%$ on the sum of all coupons (**path-dependence!**)
- Flip option: Issuer has the right to permanently change the coupon to a floater on any coupon date (**optionality!**).
- Pricing by the LIBOR market model

Algorithms

- An optimal stopping problem has to be solved
- Lattices
- PDEs (free boundary problems)
- Many practical problems have dimension between 5 and 80
- For these the only practical approach is Monte Carlo
- Longstaff-Schwartz (2001): Monte Carlo + dynamic programming
- Main difficulty with Monte-Carlo: How to compute conditional expectations?

Dynamic Programming

- Underlying S_t (may be multi-dimensional)
- Exercise times $t_1, \ldots, t_m$
- discrete time steps
- Backward induction
- Successively price options with exercise times $t_n, \ldots, t_m$ $(n = m, \ldots, 1)$

Dynamic Programming

- $V_n(x)$ = value of the option with exercise times $t_n, \ldots, t_m$, at time t_n in state $S_{t_n} = x$.
- $h_n(x)$ = payoff function, if the option is exercised at time t_n.
- We have
$$V_n(x) = \max\{h_n(x), C_n(x)\},$$
where $C_n(x)$ is the value of keeping the option.
- (Side remark: h_n may not have a closed form expression in practice.)

Dynamic Programming

- Continuation value

$$C_n(x) = \mathbf{E}[V_{n+1}(S_{t_{n+1}}) \mid S_{t_n} = x]$$

- Backward induction

$$\begin{aligned} C_m(x) &= 0, \\ C_n(x) &= \mathbf{E}[\max\{h_{n+1}(S_{t_{n+1}}), C_{n+1}(S_{t_{n+1}})\} \mid S_{t_n} = x]. \end{aligned}$$

- Option value at time t_0 is

$$\max\{h_0(S_{t_0}), C_0(S_{t_0})\}.$$

The Longstaff-Schwartz Algorithm

- Approximate optimal exercise strategy by dynamic programming
- Generate and store Monte Carlo paths
- **Assumption: Continuation value is simple function of current value of state variables.**
- Set up a linear combination of basis function
- Estimate coefficients by regression **across all paths**
- Similar ideas appeared in Carriere (1996) and Tsitsiklis & van Roy (2001)

The Longstaff-Schwartz Algorithm

- Approximate continuation values

$$C_n(x) \approx \sum_{k=0}^{K} \beta_{nk} \psi_{nk}(x) = \beta_n^{\mathrm{T}} \psi_n(x),$$

- Regression coefficients

$$\beta_n = (\beta_{n0}, \ldots, \beta_{nK})^{\mathrm{T}}$$

- Basis functions

$$\psi_n(x) = (\psi_{n0}(x), \ldots, \psi_{nK}(x))^{\mathrm{T}}$$

The Longstaff-Schwartz Algorithm

- n exercise dates, K basis functions, N Monte Carlo paths
- Estimate coefficients of continuation values by cross-sectional regression
- High bias from peeking ahead
- Low bias from suboptimality
- Typical regressors in interest rate modeling: three forward (or swap) rates, of short, middle, and long tenor

The Longstaff-Schwartz Algorithm

- Approximation one: replace conditional expectations in the dynamic programming principle by projections on a finite set of functions taken from a suitable basis
- Approximation two: use Monte- Carlo simulations and least squares regression to compute the value function of the first approximation.

Convergence Results

- K basis functions, N Monte Carlo paths
- Partial results by Longstaff, Schwartz (2001) (2 exercise times)
- Clément, Lamberton, Protter (2002):
  - Almost sure convergence of first approximation, as K to infinity
  - K fixed: N to infinity, almost sure convergence of the MC procedure to the value function of approximation 1.

Choice of Basis Functions

- Choice of basis functions?
- Examples in the literature: Laguerre polynomials, decreasing exponential functions, etc.
- Numerical study by Stentoft (2004)
- Few rigorous results on good choices
- Glasserman, Yu (2004): **How many** basis functions should one use?

Results of Glasserman and Yu (2004)

- N Monte Carlo paths, K basis functions
- Underlying process S_t is (geometric) Brownian motion
- Basis functions are Hermite polynomials (Brownian motion)

$$H_1(x), \dots, H_K(x)$$

  or monomials (geometric Brownian motion)

$$x^1, \dots, x^K.$$

- Investigate convergence of mean square error as $N, K \to \infty$.

Results of Glasserman and Yu (2004)

- Simple models, but precise results
- Assume that there is an exact representation

$$C_n(x) = \sum_{k=0}^{K} \beta_{nk}\psi_{nk}(x) = \beta_n^{\mathrm{T}}\psi_n(x).$$

- Estimate mean square error $MSE = \mathbf{E}[|\beta - \hat{\beta}|^2]$
- Some simplifying assumptions
- In practice, the convergence behavior will be even worse

Results of Glasserman and Yu (2004) and SG (2008)

- What is the highest K for which $MSE \to 0$ as $N, K \to \infty$?

- Geometric Brownian motion: $\sqrt{\log N}$ (Gl., Yu 2004)
- Brownian motion: $\log N$ (Gl., Yu 2004)

- Geometric Poisson process: $\log N / \log \log N$ (SG 2008)
- Geometric Gamma process: $\log N / \log \log N$ (SG 2008)
- Geometric Pascal process: $\log N / \log \log N$ (SG 2008)
- Geometric Meixner process: $\log N / \log \log N$ (SG 2008)

Results of Glasserman and Yu (2004) and SG (2008)

- Exponential increase in number of paths as number of basis functions increases
- Practical consequence: Do not use too many basis functions!
- Proofs depend on estimations of moments

$$E[\psi_{nj}(S_{t_n})\psi_{mk}(S_{t_m})], \qquad E[\psi_{nj}(S_{t_n})^2\psi_{mk}(S_{t_m})^2]$$

- ψ_{nk} is the k-th basis function at the n-th exercise opportunity.
- Simplifies greatly if
  - $(\psi_{nk}(S_{t_n}))_{0\le n\le m}$ is a martingale for each k
  - $(\psi_{nk})_{k\in\mathbb{N}}$ is orthogonal w.r.t. the distribution of S_{t_n}

The Models

- Geometric Poisson process

$$S_t = S_0 \exp(N_t), \qquad N_t \quad \text{standard Poisson process.}$$

  Increments of N_t are Poisson distributed.

- Geometric Gamma process

$$S_t = S_0 \exp(G_t), \qquad G_t \quad \text{Gamma process.}$$

  Increments of G_t are Gamma distributed.

- Multiplication of S_t with a deterministic function of t is permitted in all cases.

The Models

- Geometric Pascal process

$$S_t = S_0 \exp(P_t), \qquad P_t \quad \text{Pascal (NegBin) process.}$$

  Increments of P_t are negative binomially distributed.

- Geometric Meixner process

$$S_t = S_0 \exp(H_t), \qquad H_t \quad \text{Meixner process.}$$

  Increments of H_t are Meixner distributed.

- Multiplication of S_t with a deterministic function of t is permitted in all cases.

Properties of the Meixner Process

- Meixner distribution is an orthogonality measure of the Meixner-Pollaczek polynomials
- Density of the Meixner distribution:

$$f(x) = \text{const} \cdot \exp\left(\frac{b(x-m)}{a}\right) \cdot \left|\Gamma\left(d + \frac{\mathrm{i}(x-m)}{a}\right)\right|^2, \qquad x \in \mathbb{R}.$$

- Semiheavy tails
- First application to finance by Schoutens (2002)
- Pure jump process
- Good fit to log-returns of stocks

The Models

- Poisson, Gamma, Pascal, Meixner processes: Distributions have semi-heavy tails
- The associated geometric processes do not have moments of all orders
- Hence no convergence analysis with polynomial basis functions
- We will use basis functions of logarithmic growth
- Recall:
  - $(\psi_{nk}(S_{t_n}))_{0\leq n\leq m}$ should be a martingale for each k.
  - $(\psi_{nk})_{k\in\mathbb{N}}$ should be orthogonal w.r.t. the distribution of S_{t_n}.

Lévy-Sheffer Systems

- Sheffer Systems (Sheffer 1937, Meixner 1934): Given analytic functions f and u, what are the polynomials Q_k defined by

$$\sum_{k=0}^{\infty} Q_k(x)\frac{z^k}{k!} = f(z)\exp(xu(z))?$$

- Lévy-Sheffer Systems (Schoutens 2000): Define $Q_m(x,t)$ by

$$\sum_{k=0}^{\infty} Q_k(x,t)\frac{z^k}{k!} = f(z)^t\exp(xu(z)).$$

Lévy-Sheffer Systems

- Lévy-Sheffer Systems (Schoutens 2000): Define $Q_m(x,t)$ by

$$\sum_{k=0}^{\infty} Q_k(x,t)\frac{z^k}{k!} = f(z)^t \exp(xu(z)).$$

- Assume: f, u analytic at zero
- Assume: $1/f(u^{-1}(i\theta))$ is the characteristic function of an infinitely divisible distribution, defines a Lévy process X_t
- Defines $Q_m(x,t)$, polynomial in x
- Martingale property

$$E[Q_k(X_t,t) \mid X_s] = Q_k(X_s,s)$$

Examples of Lévy-Sheffer Systems

- Hermite polynomials, Brownian Motion
- Charlier polynomials, Poisson process
- Laguerre polynomials, Gamma process
- Meixner polynomials, Pascal process
- Meixner-Pollaczek polynomials, Meixner process

Martingale Properties

- Charlier, Laguerre, Meixner, Meixner-Pollaczek polynomials

$$\mathbf{E}[C_k(N_t, t) \mid N_s] = \left(\frac{s}{t}\right)^k C_k(N_s, s),$$
$$\mathbf{E}[L_k^{(t-1)}(G_t) \mid G_s] = L_k^{(s-1)}(G_s),$$
$$\mathbf{E}[M_k(P_t; t, q) \mid P_s] = \frac{(s)_k}{(t)_k} M_k(P_s; s, q),$$
$$\mathbf{E}[P_k(H_t; t, \zeta) \mid H_s] = P_k(H_s; s, \zeta),$$

- Connect moments at different exercise times, useful in the analysis

Convergence Results (SG 2008)

- N paths, K basis functions
- S_t geometric Poisson process, $\psi_{nk}(x) = t_n^k C_k(\log x, t_n)$ (Charlier polynomials)
- Put $(u, v) = (10, 4)$.
- **If $N \geq K^{(u+\varepsilon)K}$, then the mean square error tends to zero.**
- **If $N \leq K^{(v-\varepsilon)K}$, then the mean square error tends to infinity.**
- For the geometric Gamma, Pascal, and Meixner process, replace (u, v) by $(8, 8)$, $(11, 7)$, and $(8, 8)$, respectively.

Results of Glasserman and Yu (2004) and SG (2008)

- What is the highest K for which $MSE \to 0$ as $N, K \to \infty$?

- Geometric Brownian motion: $\sqrt{\log N}$ (Gl., Yu 2004)
- Brownian motion: $\log N$ (Gl., Yu 2004)

- Geometric Poisson process: $\log N / \log \log N$ (SG 2008)
- Geometric Gamma process: $\log N / \log \log N$ (SG 2008)
- Geometric Pascal process: $\log N / \log \log N$ (SG 2008)
- Geometric Meixner process: $\log N / \log \log N$ (SG 2008)

Proof Ingredients of the Convergence Results

- A formula of Zeng (1992) for linearization coefficients of Meixner-Pollaczek polynomials
- Linearization coefficients are the coefficients a_j in

$$\psi_{nk}^2 = \sum_{j=0}^{k} a_j \psi_{nj}.$$

- A classical formula connecting Laguerre polynomials with different parameters
- Estimate norm of the inverse of a tridiagonal matrix
- Estimate some involved sums

Conclusion

- Large number of high degree basis functions is detrimental for convergence
- This holds for several models, for which the convergence analysis is feasible
- If the results extend to higher dimensions, LS might be questionable
- N, the number of paths, increases slowest if S_t is Brownian motion
- Reason: Linearization coefficients of the Hermite polynomials grow slowest

References

- J.F. Carriere: Valuation of the early-exercise price for options using simulations and nonparametric regression. Insurance Math. Econom. 19 (1996).
- F.A. Longstaff, E.S. Schwartz: Valuing American Options by Simulation: A Simple Least-Squares Approach. The Review of Financial Studies 12 (2001).
- E. Clément, D. Lamberton, P. Protter: An analysis of a least squares regression method for American option pricing. Finance Stochast. 6 (2002).
- P. Glasserman, B. Yu: Number of paths versus number of basis functions in American option pricing. Ann. Appl. Probab. 14 (2004).
- S. Gerhold, Lévy-Sheffer Systems and the Longstaff-Schwartz Algorithm for American Option Pricing, 2008. Submitted.