Computational Finance

July 31, 2020

Syllabus

1. Numerical methods relevant to integration, differentiation and solving the partial differential equations of mathematical finance: examples of exact solutions including Black Scholes and its relatives, finite difference methods including algorithms and question of stability and convergence, treatment of near and far boundary conditions, the connection with binomial models, interest rate models, early exercise, and the corresponding free boundary problems, and a brief introduction to numerical methods for solving multi-factor models.

2. Simulation including random variable generation, variance reduction methods and statistical analysis of simulation output. Pseudo random numbers, Linear congruential generator, Mersenne twister RNG. The use of Monte Carlo simulation in solving applied problems on derivative pricing discussed in the current finance literature. The technical topics addressed include importance sampling, Monte Carlo integration, Simulation of Random walk and approximations to diffusion processes, martingale control variables, stratification, and the estimation of the "Greeks." Application areas include the pricing of American options, pricing interest rate dependent claims, and credit risk. The use of importance sampling for Monte Carlo simulation of VaR for portfolios of options.

3. Statistical Analysis of Financial Returns: Fat-tailed and skewed distributions, outliers, stylized facts of volatility, implied volatility surface, and volatility estimation using high frequency data.

4. Advanced topics: Estimating VaR and Expected Shortfall, Exact and bootstrap confidence intervals, Extreme value theory methods for risk management, fitting copulas to data.

Reference Texts:

1. P. Glasserman, Monte Carlo Methods in Financial Engineering

2. D. Ruppert, Statistics and Data Analysis for Financial Engineering

3. R. Carmona: Statistical Analysis of Financial Data in S-Plus

4. N. H. Chan, Time Series: Applications to Finance

5. R. S. Tsay, Analysis of Financial Time Series

6. J. Franke, W. K. Hardle and C. M. Hafner, Statistics of Financial Markets: An Introduction

7. Kerry Back: A course in derivative securities

General instructions

- Bring laptops to class.

- HW0: revise your R skills.

- Final will be cumulative.

- 5% of score is on class involvement including attendance.

- My email is rsen@isichennai.res.in and phone no is 9176620249. Please email me unless it is a very urgent situation.

Numerical Methods

Preliminaries

- Algorithm: convergence

- Error: Discretization, truncation, rounding

- Well/Ill conditioned problems

Finding Roots of a function

- Application: Find Implied Volatility

- Definition of Implied Volatility

- Program in R to calculate Black-Scholes price of European option (*blackscholes*)

- Bisection Algorithm

- Program in R to calculate Implied Volatility using bisection (*implied_bisect*)

- Check the "no-arbitrage" condition and display an error message when this is violated.

- Inputs: all parameters of *blackscholes* other than σ, C and error tolerance.

- Newton-Raphson

- Update with vega

- Secant Method

1 Preliminaries

1.1 Algorithms

An algorithm is a set of instructions to construct an approximate solution to a mathematical problem.

A basic requirement for an algorithm is that the error can be made as small as we like.

Usually, the higher the accuracy we demand, the greater is the amount of computation required.

An algorithm is convergent if it produces a sequence of values which converge to the desired solution of the problem.

Example Given $c > 1$ and $\epsilon > 0$, use the bisection method to seek an approximation to $\sqrt{c}$ with error not greater than ϵ.

Find $x = \sqrt{c}, c > 1$ constant.

Answer: $x = \sqrt{c} \iff x^2 = c \iff f(x) := x^2 - c = 0$

$\Rightarrow f(1) = 1 - c < 0$ and $f(c) = c^2 - c > 0$

$\Rightarrow \exists x_0 \in (1, c) s.t. f(x_0) = 0$

$f'(x) = 2x > 0, f$ monotonically increasing. So x_0 is unique.

Denote $In := [a_n, b_n]$ with $I_0 = [a_0, b_0] = [1, c]$

Let $x_n := (a_n + b_n)/2$.

- If $f(x_n) = 0$ then $x_0 = x_n$.

- If $f(a_n) \times f(x_n) > 0$ then $x_0 \in (x_n, b_n)$, let $a_{n+1} := x_n, b_{n+1} := b_n$.

- If $f(a_n) \times f(x_n) < 0$ then $x_0 \in (a_n, x_n)$, let $a_{n+1} := a_n, b_{n+1} := x_n$.

Length of $I_n : m(I_n) = 1/2m(I_{n-1}) = \cdots = 1/2^n m(I_0) = (c - 1)/2^n$

Algorithm stops if $m(I_n) < \epsilon$ and let $x_0 := x_n$.

Error as small as we like?

$x_0, x* \in I_n$ error $|x_0 - x * | \leq m(I_n) \to 0$ as $n \to \infty$.

1.2 Errors

There are various errors in computed solutions, such as

- discretization error (discrete approximation to continuous systems),

- truncation error (termination of an infinite process), and

- rounding error (finite digit limitation in computer arithmetic).

If a is a number and $\tilde{a}$ is an approximation to a, then the absolute error is $|a - \tilde{a}|$ and the relative error is $|a - \tilde{a}|/|a|$ provided $a \neq 0$.

Example: Discuss sources of errors in deriving the numerical solution of the non-linear differential equation $x' = f(x)$ on the interval $[a, b]$ with initial condition $x(a) = x_0$.

discretization error

$x' = f(x)$ [differential equation]

$(x(t + h) - x(t))/h = f(x(t))$ [difference equation]

$\text{DE} = |(x(t + h) - x(t))/h - x'(t)|$

truncation error $lim_{n \to \infty} x_n = x$, approximate x with x_N, N a large number.

$\text{TE} = |x - x_N|$

rounding error We cannot express x exactly, due to finite digit limitation. We get $\hat{x}$ instead.

$\text{RE} = |x - \hat{x}|$

Total error = DE + TE + RE.

1.3 Well/Ill conditioned problems

A problem is well-conditioned (or ill-conditioned) if every small perturbation of the data results in a small (or large) change in the solution.

Example: Show that the solution to equations $x + y = 2$ and $x + 1.01y = 2.01$ is ill-conditioned.

$x + y = 2, x + 1.01y = 2.01 \Rightarrow x = 1, y = 1$

Change 2.01 to 2.02

$x + y = 2, x + 1.01y = 2.02 \Rightarrow x = 0, y = 2$

I.e. 0.5% change in data produces 100% change in solution: ill-conditioned !

reason: $\det(1\ 1\ 1\ 1.01) = 0.01$, nearly singular.

2 Finding Roots of a function

Application: Find Implied Volatility
Let C denote the market price of a European Call option. The implied volatility σ satisfies

$$\text{Black\_Scholes\_Call}(S, K, r, T, \sigma, 0) - C = 0 \tag{1}$$

```
blackscholes <- function(S, X, rf, T, sigma, CallPut)
```

Bisection Algorithm
Start with upper and lower bounds for σ and repeatedly bisect the interval containing σ, each time finding a new upper and lower bound.
First we need to check the "no-arbitrage" condition $C + e^{-rT}K \geq S$ and display an error message when this is violated.
Inputs: all parameters of Black_Scholes_Call other than σ, C and error tolerance.

```
implied_bisect <- function(S, X, rf, T, CallPut, price, tol, cond)
```

Newton-Raphson
Faster root-finding algorithm than bisection.
General Newton-Raphson update

$$x_1 = x_0 - f(x_0)/f'(x_0) \tag{2}$$

. This amounts to approximating the function as being linear and using the root of the approximation as the updated guess. In case of Option Prices, the derivative of the option price with respect to volatility is called vega. For the European option price under the Black-Scholes model this equals

$$\nu = S\sqrt{(T)}\phi(d_1) \tag{3}$$

where ϕ is the normal pdf.
So updated statement is

```
guess=guess-call/vega
```

Secant Method
A similar idea that does not require the computation of derivative is to keep track of two most recent (x,f(x)) and use the difference quotient.
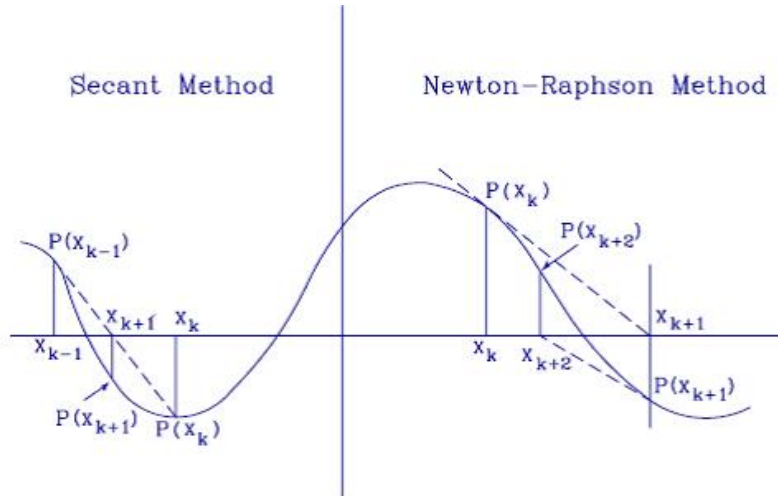
```
vega=(call-prior_call)/(guess-prior_guess)
```

Figure 3.1 depicts a typical polynomial with real roots. Construct the tangent to the curve at the point x_k and extend this tangent to the x-axis. The crossing point x_k+1 represents an improved value for the root in the Newton-Raphson algorithm. The point x_k-1 can be used to construct a secant providing a second method for finding an improved value of x.

Program for computing Black-Scholes option price

```
blackscholes <- function(S, X, rf, T, sigma, CallPut) {
#CallPut is 0 for Call and 1 for Put
d1 <- (log(S/X)+(rf-sigma^2/2)*T)/(sigma*sqrt(T))
d2 <- d1 - sigma * sqrt(T)
if(CallPut == 0)
value<- max(S*pnorm(d1) - X*exp(-rf*T)*pnorm(d2), 0)
else
value<- max(X*exp(-rf*T) * pnorm(-d2) - S*pnorm(-d1), 0)
return(value)
}
```

Program for computing implied volatility by bisection method

```
implied_bisect <- function(S, X, rf, T, CallPut, price, tol, cond) {
#cond=1 if upper-lower is used for tolerence, otherwise abs(fguess) is used for to
if(CallPut==0 && price<S-X*exp(-rf*T)|| CallPut!=0 && price<X*exp(-rf*T)-S)
guess=0
else
#finding the upper bound
 lower=0
 flower=blackscholes(S, X, rf, T,lower,CallPut)- price
 upper=1
 fupper=blackscholes(S, X, rf, T,upper,CallPut)- price
```

```
  while(flower*fupper>0){
  upper=upper*2
  fupper=blackscholes(S, X, rf, T,upper,CallPut)- price}
#updating to bisected interval
 guess=0.5*(upper+lower)
 fguess=blackscholes(S, X, rf, T,guess,CallPut)- price
  if (cond==1)
   while(upper-lower>tol){
    if(fguess*flower<0)
    upper=guess
    else
    lower=guess
    end
   flower=blackscholes(S, X, rf, T,lower,CallPut)- price
   fupper=blackscholes(S, X, rf, T,upper,CallPut)- price
   guess=0.5*(upper+lower)
   fguess=blackscholes(S, X, rf, T,guess,CallPut)- price}
  else
   while(abs(fguess)>tol){
    if(fguess*flower<0)
    upper=guess
    else
    lower=guess
    end
   flower=blackscholes(S, X, rf, T,lower,CallPut)- price
   fupper=blackscholes(S, X, rf, T,upper,CallPut)- price
   guess=0.5*(upper+lower)
   fguess=blackscholes(S, X, rf, T,guess,CallPut)- price}
  end
 end
return(guess)
}
```

Homework 1 Write the program for computing implied volatility using Newton Raphson and Secant methods.

3 Numerical Integration

3.1 Motivation

Numerical integration is a standard topic in numerical analysis. Classical approaches to numerical integration based on quadrature formulas are deterministic.

We have seen that option pricing requires computing an expected value under a risk-neutral measure, but an expected value is actually an integral.

The expected value of a function $g(.)$ of a random variable X with probability density $f_X(x)$ is $\int_{-\infty}^{\infty} g(x) f_X(x) dx$. In really simple cases, we may find an analytical solution, like in the Black- Scholes case.

If the random variable X is a scalar, classical deterministic methods work quite well. When expectation is taken with respect to a random vector and we must integrate over a high-dimensional space, random sampling may be necessary. We shall see these in the simulation chapter as Monte-Carlo methods.

Numerical integration may be implicitly used to estimate probabilities. If A is an event which may occur or not depending on a random variable X, then $P(A) = \int_{-\infty}^{\infty} I_A(x) f_X(x) dx$, where $I_A(x)$ is the indicator function for event A (taking the value 1 if A occurs when $X = x$, 0 otherwise). When A is a rare event, clever strategies are needed to get an accurate estimate with a reasonable computational effort.

Finally, there are situations in which we define a function by an integral. A typical case is the expected value of a function depending on a control variable (modeling our decisions) and a random variable (modeling what we cannot control)

$$H(z) = E(g(X, z)) = \int_{-\infty}^{\infty} g(x, z) f_X(x) dx$$

This is quite common in stochastic optimization and dynamic programming, whereby we want to find a maximizer (or minimizer) of $H(z)$, and this calls for a suitable approximation of H by discretization of the continuous distribution. In other words, we want to generate a discrete set of scenarios yielding a reasonable approximation of the underlying uncertainty. Numerical methods such as Gaussian quadrature are helpful here.

3.2 Introduction

Consider the problem of approximating the value of a definite integral like

$$I[f] = \int_a^b f(x)dx$$

over a bounded interval $[a, b]$ for a function f of a single variable. Since the integration is a linear operator, it is natural to look for an approximation preserving this property. Using a finite number of values of f over a set of nodes x_j such that $a = x_0 < x_1 < \cdots < x_n = b$, we may define a quadrature formula such as

$$Q[f] = \sum_{j=0}^{n} w_j f(x_j)$$

A quadrature formula is characterized by the weights w_j and by the nodes x_j. To be precise, a quadrature formula like the one we are describing is called a closed formula, since evaluation of the function in the extreme points of the interval is used. Sometimes, open formulas are used when the function is not well-behaved near a or b, or when we are integrating on an infinite interval. Any quadrature formula is characterized by a truncation error:

$$E = I[f] - Q[f]$$

A reasonable requirement is that the error should be zero for sufficiently simple functions such as polynomials. We may define the order of a certain quadrature formula as the maximum degree m such that the truncation error is zero for all the polynomials of degree n or less. In other words, if the original function is substituted by an interpolating polynomial, we should not commit any error in integrating the polynomial.

3.3 Classical interpolatory formulas

One way to derive quadrature formula is to consider equally spaced nodes:

$$x_j = a + jh, j = 0, 1, 2, \cdots, n$$

where $h = (b - a)/n$; also let $f_j = f(x_j)$. Selecting equally spaced nodes yields the set of Newton-Cotes quadrature formulas.

Given those $n + 1$ nodes, we may consider the interpolating polynomial $P_n(x)$ using Lagrange polynomials of degree n.

For a given set of distinct points x_j and numbers f_j, the Lagrange polynomial is the polynomial of lowest degree that assumes at each point x_j the corresponding

value f_j. Given a set of $n+1$ data points $((x_0, f_0), \cdots, (x_n, f_n))$ where no two x_j are the same, the interpolation polynomial in the Lagrange form is a linear combination

$$P_n(x) := \sum_{j=0}^{n} f_j \ell_j(x)$$

of Lagrange basis polynomials

$$\ell_j(x) := \prod_{\substack{0 \le m \le n \\ m \ne j}} \frac{x - x_m}{x_j - x_m} = \frac{(x - x_0)}{(x_j - x_0)} \cdots \frac{(x - x_{j-1})}{(x_j - x_{j-1})} \frac{(x - x_{j+1})}{(x_j - x_{j+1})} \cdots \frac{(x - x_n)}{(x_j - x_n)},$$

where $0 \le j \le n$.

Then we may compute the correct weights as follows:

$$\int_a^b f(x)dx \approx \int_a^b P_n(x)dx = \int_a^b \sum_{j=0}^{n} f_j \ell_j(x)dx = \sum_{j=0}^{n} f_j [\int_a^b \ell_j(x)dx] = \sum_{j=0}^{n} w_j f_j$$

3.3.1 Trapezoidal rule

Consider the case of two nodes only, $x_0 = a$ and $x_1 = b$. Here we are just interpolating f by a straight line:

$$P_1(x) = f_0 \frac{(x - x_1)}{(x_0 - x_1)} + f_1 \frac{(x - x_0)}{(x_1 - x_0)}. \tag{4}$$

A straightforward calculation yields $\int_{x_0}^{x_1} P_1(x) = h(\frac{f1+f0}{2})$. Actually, what we are saying is that we may approximate the area below the function using trapezoidal elements and the formula above gives the area of one element. Applying the idea to more subintervals, we get the trapezoidal rule:

$$Q[f] = h(\frac{1}{2}f_0 + \sum_{j=1}^{n-1} f_j + \frac{1}{2}f_n). \tag{5}$$

Given any quadrature formula for an interval: we may get a composite formula by applying the same pattern to small subintervals of a large one.

3.3.2 Simpson's Rule

A quadrature formula based on $n+1$ nodes is by construction exact for polynomials of degree n. We may go the other way around, and build a formula by requiring a certain order. Consider the case

$$\int_0^1 f(x)dx = w_0 f(0) + w_1 f(0.5) + w_2 f(l),$$

11

and say we want a formula that is exact for polynomials of degree ≤ 2. Having fixed the nodes, we may find the weights by solving the system of linear equations:

$$1 = \int_0^1 1dx = w_0 + w_1 + w_2$$

$$1/2 = \int_0^1 xdx = 1/2w_1 + w_2$$

$$1/3 = \int_0^1 x^2dx = 1/4w_1 + w_2$$

which yields $w_o = 1/6, w_1 = 2/3, w2 = 1/6$. Applying the same idea on the interval $[a, b]$, we get Simpson's rule:

$$\int_a^b f(x)dx = \frac{b-a}{6}(f(a) + 4f(\frac{a+b}{2}) + f(b))$$

Simpson's rule may be applied to subintervals of (a, b) in order to get a composite formula.

3.4   Gaussian Quadrature

In Newton-Cotes formulas we fix nodes and try to find suitable weights so that the order of the formula is as large as possible. The rationale behind Gaussian quadrature is that if we do not fix nodes a priori, we essentially double the degrees of freedom, in such a way that the order can be more or less doubled. Furthermore, Gaussian quadrature formulas are developed with respect to a non-negative weight function $w(x)$. We look for a quadrature formula like $\int_a^b w(x)f(x) = \sum_{i=1}^n w_if(x_i)$ which is exact when f is a polynomial. The weight function $w(x)$ can be used to encapsulate undesired singularities of the integrand function. We will only outline the development of Gauss-Hermite quadrature, where $w(x) = e^{-x^2}$.

Now, how should we select the nodes and weights in in order to get a quadrature formula with maximum order. We should choose as nodes the n roots of a polynomial of order n, selected within a family of orthogonal polynomials with respect to the inner product:

$$< f, g >= \int_a^b w(x)f(x)g(x)dx.$$

It can be shown that a polynomial of degree k within that family has k distinct real roots. Furthermore, these roots are interleaved, in the sense that each of the $k - 1$ roots of the polynomial of degree $k - 1$ lies in an interval defined by a pair of consecutive roots of the polynomial of degree k . Using this choice of nodes, along with a proper choice of weights, yields a quadrature formula with order $2n - 1$.

To see this, consider a polynomial $q \in \Pi_n$, i.e., a polynomial of degree n, which is orthogonal to all polynomials in Π_{n-l}.

Any polynomial $f \in \Pi_{2n-1}$ can be divided by q, obtaining a quotient p and a remainder $r : f = qp+r$, where $p, r \in \Pi_{n-l}$. Now let us integrate wf by a quadrature formula on n nodes $x_i, i = 1, \cdots, n$, which are the zeros of q:

$$\int_a^b w(x)f(x)dx$$

$$= \int_a^b w(x)p(x)q(x)dx + \int_a^b w(x)r(x)dx$$

$$= 0 + \int_a^b w(x)r(x)dx \quad (q \text{ is orthogonal to } p)$$

$$= \sum_{i=1}^n w_i r(x_i) \quad (\text{quadrature is exact for } r \in \Pi_{n-1})$$

$$= \sum_{i=1}^n w_i f(x_i) \quad (x_i \text{ is zero of } q)$$

A family of orthogonal polynomials $p_j(x)$ may be built by the following procedure:

$$
\begin{aligned}
p_0(x) &= 1 \\
p_{j+1}(x) &= (x - a_j)p_j(x) - b_j(p_{j-1}(x), \quad j = 0, 1, 2, \cdots \\
\text{where } a_j &= \frac{< xp_j, p_j >}{< p_j, p_j >} \\
b_j &= \frac{< p_j, p_j >}{< p_{j-1}, p_{j-1} >}
\end{aligned}
$$

Here coefficient b_0 is arbitrary and it can be set to 0. At each step, the procedure generates a new polynomial of degree one plus the degree of the previous polynomial. In the end, we have a family of orthogonal polynomials, one for each degree. Actually there are different choices of normalizations yielding different families of polynomials.

In the Gauss-Hermite case, whereby $w(x) = e^{-x^2}$, applying the procedure above results in the following recursive procedure yielding a sequence of Hermite polynomials: $H_{j+l} = 2xH_j - 2jH_{j-1}$.

3.5   Applications

Tauchen, George, and Robert Hussey. "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models." Econometrica: Journal of the Econometric Society (1991): 371-396.

The paper develops a discrete state space solution method for a class of nonlinear rational expectations models. These models rarely admit explicit solutions.

As part of a larger study of monetary velocity in cash-in-advance models, Hodrick, Kockerlakota, and Lucas (1989) calibrate a Markov chain model for bivariate money growth and consumption growth. They find that with sixteen states of nature the Markov chain can adequately approximate a VAR(1) model fitted to annual data. Boudoukh and Whitelaw (1988), who use the quadrature technique and related ideas to study the pricing of mortgage-backed securities and American options. Though their securities have a particularly complicated path-dependent cash flow, in test cases they get very close approximations to exact solutions with state spaces as small as three points.

4 Basics of Mathematical Finance

4.1 Risk-Neutral Measure

Single Period Market: Consider a market in which K assets, labeled $A_1, A_2, \cdots, A_K$, are freely traded. Assume that one of these, say A_1, is riskless, that is, its value at time $t = 1$ does not depend on the market scenario. The share price of asset A_j at time $t = 0$ is S_0^j; without loss of generality, we may assume that $S_0^1 = 1$. Uncertainty about the behavior of the market is encapsulated in a finite set of N possible market scenarios, labeled $\omega_1, \omega_2, \cdots, \omega_N$. The share prices $S_1^2, S_1^3, \cdots, S_1^K$ of the $K-1$ assets at time $t = 1$ are functions of the market scenario.

Observe that, since asset A_1 is riskless, there is a constant r, called the riskless rate of return, such that the share price S_1^1 of A_1 in any scenario ω_i is

$$S_1^1(\omega_i) = e^r \forall \qquad i = 1, 2, \cdots, N. \tag{6}$$

Definition 1. *A* **portfolio** *is a vector* $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_K) \in \mathbb{R}^K$ *of K real numbers. The entry θ_j represents the number of shares of asset A_j that are owned; if $\theta_j < 0$ then the portfolio is said to be short $|\theta_j|$ shares of asset A_j .*

The value of the portfolio $\underline{\theta}$ at time $t = 0$ is

$$V_0(\underline{\theta}) = \sum_{j=1}^{K} \theta_j S_0^j \tag{7}$$

and the value of the portfolio θ at time $t = 1$ in market scenario ω_i is

$$V_1(\underline{\theta}; \omega_i) = \sum_{j=1}^{K} \theta_j S_1^j(\omega_i). \tag{8}$$

Definition 2. *An* **arbitrage** *is a portfolio $\underline{\theta}$ that "makes money from nothing", formally, a portfolio $\underline{\theta}$ such that either*

$$V_0(\underline{\theta}) \leq 0 \quad \text{and} \quad V_1(\underline{\theta}; \omega) > 0 \forall \qquad i = 1, 2, \cdots, N \tag{9}$$
$$\text{OR} \quad V_0(\underline{\theta}) < 0 \quad \text{and} \quad V_1(\underline{\theta}; \omega_j) \geq 0 \forall \qquad i = 1, 2, \cdots, N \tag{10}$$

Definition 3. *A probability distribution $\pi_i = \pi(\omega_i)$ on the set of possible market scenarios is said to be an* **equilibrium measure** *(or* **risk-neutral measure***) if, for every asset A, the share price of A at time $t = 0$ is the discounted expectation, under π, of the share price at time $t = 1$, that is, if*

$$S_0^j = e^{-r} \sum_{i=1}^{N} \pi(\omega_i) S_1^j(\omega_i) \forall \qquad j = 1, 2, \cdots, K. \tag{11}$$

4.2 Fundamental Theorems of Asset Pricing

Theorem 1. *(First Fundamental Theorem of Arbitrage Pricing) There exists an equilibrium measure if and only if arbitrages do not exist.*

The first implication is easy to prove. Suppose that there is an equilibrium measure π. Then for any portfolio $\underline{\theta}$, the portfolio values at time $t = 0$ and $t = 1$ are related by discounted expectation:

$$V_0(\underline{\theta}) = \sum_{i=1}^{N} \pi(\omega_i)e^{-r}V(\underline{\theta};\omega_i). \tag{12}$$

(To see this, just multiply equation (11) by θ_j , sum on j, and use the definitions of portfolio value in (7)-(8) above.) If $V(\underline{\theta};\omega_i) > 0$ for every market scenario ω_i (as must be the case for an arbitrage portfolio), then equation (12) implies that $V_0(\underline{\theta}) > 0$, and so $\underline{\theta}$ cannot be an arbitrage. Thus, arbitrages do not exist.

The second implication, that absence of arbitrages implies the existence of an equilibrium measure, is harder to prove and we shall skip it. The main ingredient is the Separating Hyperplane Theorem.

Example: The Call Option Let us consider the pricing of the European call option on an asset Stock. The strike price is K, and so the terminal value of the option is given by

$$V_1 = (S_1 - K)_+ \quad = \quad S_1 - K \quad \text{if} \quad S_1 \geq K \tag{13}$$
$$= \quad 0 \quad \text{if} \quad S_1 < K. \tag{14}$$

Two-Scenario Market: There are two possible market scenarios, ω_1, ω_2. The value of one share of Stock at time 1 is $S_1(\omega_i) = d_i$ in scenario ω_i, with $d_1 < d_2$. The riskless rate of return is r. By the fundamental theorem, in an arbitrage-free market, there is a probability distribution π on the two scenarios that determines prices by discounted expectation, and so, in particular,

$$S_0 = \pi(\omega_1)e^{-r}d_1 + \pi(\omega_2)e^{-r}d_2. \tag{15}$$

Because there are only two market scenarios, equation (15) uniquely determines the equilibrium measure π (note that π is a probability measure, so adds up to one):

$$\pi(\omega_1) \quad = \quad (d_2 - S_0e^r)/(d_2 - d_1), \tag{16}$$
$$\pi(\omega_2) \quad = \quad (S_0e^r - d_1)/(d_2 - d_1). \tag{17}$$

Finally, if the call option is to be freely traded, and if the market is to remain arbitrage-free, then its value at time $t = 0$ is also determined by discounted expectation. Since there is only one possible equilibrium measure, as in the last displayed equations, the value of the call at time $t = 0$ is

$$V_0 = \pi(\omega_2)(d_2 - K). \tag{18}$$

Three-Scenario Market: Consider now the pricing of the call option with strike K in the three-scenario market discussed earlier. If the only freely traded assets in the market were Stock and MoneyMarket, then the pricing formulas (11) would not uniquely determine the equilibrium distribution π, because formulas (11) provide only two equations in three unknowns. Thus, any probability distribution (π_1, π_2, π_3) on the three scenarios such that $S_0 e^r = d_1 \pi_1 + d_2 \pi_2 + d_3 \pi_3$ would be allowable as an equilibrium measure. Call the set of all such probability distributions Π. Then any element $\pi \in \Pi$ such that

$$v = e^{-r}(p_2(d_2 - K)_+ + p_3(d_3 - K)) \tag{19}$$

holds would be an equilibrium measure for the enlarged market in which the freely traded assets are Stock, MoneyMarket, and Call, where Call is the call option on Stock with strike K, provided the $t = 0$ price of Call is given by (19). By the Fundamental Theorem, any such market is arbitrage-free.

Definition 4. *Suppose the freely traded assets in a market are $A_1, \cdots, A_k$. A* **derivative security** *is a tradeable asset whose value V_1 at time $t = 1$ is a function $V(\omega_i)$ of the market scenario. In the language of probability, it is a random variable measurable on the sigma field generated by the asset prices.*

Definition 5. *Consider a market with freely traded assets $A_1, \cdots, A_k$ and B. A portfolio $\underline{\theta}$ in $A_1, \cdots, A_k$ is* **replicating portfolio** *for B if*

$$S_1^B(\omega_i) = \sum_{i=1}^{K} \theta_j S_1^j(\omega_i) \quad \forall i = 1, \cdots, N \tag{20}$$

Replicating portfolios enable financial institutions that sell asset B to **hedge** by using the following strategy. For each share of B sold, buy θ_j shares of asset A_j ad hold till time $t = 1$. Then at time $t = 1$, net gain = net loss =0. The institution selling B makes money by charging a fee or premium.

Definition 6. *A market is* **complete** *if every derivative security can be hedged.*

Theorem 2. *(Second Fundamental Theorem of Asset Pricing): Consider a market model that has a risk-neutral measure. The market is complete iff the risk-neutral measure is unique.*

Proof: See Shreve page 232.

Remark 1: Risk-neutral measure may be, and generally is, different from the physical measure. Under the physical measure, in general, expected returns of stocks are more than that on bonds. This is measured by the market price of risk. This is related to risk/return relations as in CAPM. But for pricing derivatives, one has to use the risk-neutral measure.

Remark 2: Market is complete iff $A_1, \cdots, A_K$ **linerly** span the space of derivative securities. In general, the latter are random variables, that is functions, but not necessarily linear.

Remark 3: We have done the whole treatment in a two-period set-up. All the results can be generalized to a continuous time setup. The mathematics get very involved, but the concepts are the same.

4.3 Brownian Motion

Why Brownian motion?

- In real markets, trading takes place in continuous time. Problems of pricing and hedging derivative securities in continuous-time markets require continuous-time models.

- In equilibrium, the discounted price process of any tradeable asset, is a martingale.

- The prices of traded assets vary continuously with time and have finite quadratic variation.

Brownian motion now rears its head for the following basic reason, a fundamental theorem of Paul Levy:

Theorem Every continuous-time martingale with continuous paths and finite quadratic variation is a time-changed Brownian motion.

Definition A standard Brownian (or a standard Wiener process) is a stochastic process $\{W_t\}_{t \geq 0}$ (that is, a family of random variables W_t, indexed by nonnegative real numbers t, defined on a common probability space (Ω, F, P)) with the following properties:

1. $W_0 = 0$

2. With probability 1, the function $t \to W_t$ is continuous in t.

3. The process has stationary, independent increments.

4. The increment $W_{t+s} - W_s$ has the Normal$(0, t)$ distribution.

The term independent increments means that for every choice of nonnegative real numbers $0 \leq s_1 < t_1 \leq s_2 < t_2 = \cdots = s_n < t_n < \infty$, the increment random variables $W_{t_1} - W_{s_1}, W_{t_2} - W_{s_2}, \cdots, W_{t_n} - W_{s_n}$, are jointly independent.

The term stationary increments means that for any $0 < s, t < \infty$ the distribution of the increment $W_{t+s} - W_s$ has the same distribution as $W_t - W_0 = W_t$.

It should not be obvious that properties (1)–(4) in the definition of a standard Brownian motion are mutually consistent, so it is not a priori clear that a standard Brownian motion exists. (The main issue is to show that properties (3)–(4) do not preclude the possibility of continuous paths.) That it does exist was first proved by N. Wiener in about 1920. But notice that properties (3) and (4) are compatible. This follows from the following elementary property of the normal distributions: If X, Y are independent, normally distributed random variables with means μ_X, μ_Y and variances σ_X^2, σ_Y^2 then the random variable $X + Y$ is normally distributed with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$.

4.3.1 Quadratic Variation of Brownian Paths

Fix $t > 0$, and let $\Pi = \{t_0, t_1, t_2, \cdots, t_n\}$ be a partition of the interval $[0, t]$. For each natural number n, define the $n-$th dyadic partition $D_n[0, t]$ to be the partition consisting of the dyadic rationals $k/2^n$ of depth n (here k is an integer) that are between 0 and t (with t added if it is not a dyadic rational of depth n). Let $X(s)$ be any process indexed by s. For any partition Π, define the quadratic variation of X relative to Π by

$$QV(X; \Pi) = \sum_{j=1}^{n} (X(t_j) - X(t_{j-1}))^2$$

Proposition Let $\{W(t)\}_{t \geq 0}$ be a standard Brownian motion. For each $t > 0$, with probability one,

$$\lim_{n \to \infty} QV(W; D_n[0, t]) = t.$$

The primary significance of this result is that it is the key to the Itˆo formula of stochastic calculus. It also implies that the Brownian path cannot be of bounded variation in any interval, and so in particular is not differentiable in any interval. (Even more is known: the Brownian path is nowhere differentiable.)

Partial Proof. We will first prove the weaker statement that the convergence holds in probability. To simplify things, assume that $t = 1$. Then for each $n \geq 1$, the random variables $\xi_{n,k} := 2^n (W(k/2^n) - W((k-1)/2^n))^2$, $k = 1, 2, ..., 2^n$ are independent, identically distributed χ^2 with one degree of freedom (that is, they are distributed as the square of a standard normal random variable).

Observe that $E\xi_{n,k} = 1$. Now $QV(W; D_n[0, 1]) = 2^{-n} \sum_{k=1}^{2^n} \xi_{n,k}$. The right side of this equation is the average of $2n$ independent, identically distributed random variables, and so the Weak Law of Large Numbers implies convergence in probability to the mean of the χ^2 distribution with one degree of freedom, which equals 1. The stronger statement that the convergence holds with probability one can easily be deduced from the Chebyshev inequality and the Borel–Cantelli lemma.

Homework 2

1. In a single period binary market with STOCK and BOND, obtain the risk-neutral distribution. What are the conditions on the possible prices at time 1? Price a European CALL with strike K. Prove this with an arbitrage argument.

2. Let $f(x) = P_1(x) + E(x)$, where P_1 is given in equation 4. Show that $\mid E(x) \mid \leq f''(\xi) * (x-a)(x-b)/2$ with $\xi \in (a, b)$. Assume that $|f''| \leq M$ is bounded. Then show that $\mid \int_a^b E(x)dx \mid \leq= M/12(b-a)^3$. The composite formula of Trapezoid rule in equation 5, has corresponding error $\mid I[f] - Q[f] \mid \leq M(b-a)/12h^2$

3. Derive the conditional distribution of W_{t+s} given W_s.

5 Numerical Differentiation

- bond duration

- delta, gamma

- sensitivity analysis

- optimization

This is taught from chapter 5 of the lecture notes of Prof Levy Introduction to Numerical Analysis available here http://www.math.umd.edu/ dlevy/books/na.pdf

Homework 3

1. Use the differentiation by integration formula (5.10) with n = 2 while approximating the derivative at x_1, to obtain the second-order centered approximation of the first-derivative.

2. Evaluate the quadratic variation of Brownian motion and exponential(x) on [0,1] for different grid sizes and plot.

3. Compute the histogram of profits from N simulations of a discretely balanced delta hedge for the Black Scholes call.

6 Numerical solution of PDE

Some common sensitivity measures of an option price C (the rate of change of the option price with respect to the underlying factors) are defined by $Delta = \delta = \frac{\partial C}{\partial S}$, $Gamma = \gamma = \frac{\partial^2 C}{\partial S^2}$, $vega = \nu = \frac{\partial C}{\partial \sigma}$, $rho = \rho = \frac{\partial C}{\partial r}$, $theta = \theta = -\frac{\partial C}{\partial T}$

6.1 examples of exact solutions

HW: Find explicit formulae for the Greeks in case of the Black-Scholes European call option.

- Deriving the Black Scholes pde

- Discretizing the PDE

- Implicit and Explicit Methods

- Crank-Nicolson method

- Random walk model and assumptions

- CRR model

- Binomial approximation to BSE

- N-period option pricing and approximation to BS

- Local truncation error

- Stability

- methods for solving multi-factor models

- interest rate models

- early exercise

- the corresponding free boundary problems

A good reference is Chapter 5 of Numerical Methods in Finance and Economics: A MATLAB-Based Introduction by Paolo Brandimarte.

For pricing of American Put Option and the free boundary problem, refer to The Pricing of the American Option by Ravi Myneni in The Annals of Applied Probability, Vol. 2, No. 1 (Feb., 1992), pp. 1-23. Particularly Lemma 5.1 explains the super-martingale property.

Homework 4

1. Derive the Black Scholes PDE and Transform BS pde to heat equation. Abhishek, Kritika, Subhendu

2. Derive the CRR model from conditions on first two momemts of Binomial and extra condition when $ud = 1$. Simplify this when $o(\Delta t)$ terms are ignored. Damitri, Tathagata

3. Show that Binomial Option pricing formula approximately solves the BS equation. Saikat

4. Show that Binomial Option price approximates the BS option price.

5. Crank Nicholson Algorithm for BSPDE by direct discretization. Abhishek, Tathagata, Saikat, Kritika

6. Crank Nicholson Algorithm for BSPDE after transforming to Heat Equation

7. Crank Nicholson Algorithm for American option. Refer to Matlab code in section 9.5 of Matlab book.

8. Evaluate the American Put on Spread option using Binomial model. Damitri

Simulation methods

Materials from the following pages of Glasserman: 11-12, 49-52, 58-69, 75-87, 90-98, 102-105,108-109,194-238.

7 Topics in Univariate iid returns

Read Chapter 6.5 of Embrechts Klueppelberg Mikosch.

7.1 VaR and Expected shortfall

The most common reported measure of risk is Value-at-Risk (VaR). The VaR of a portfolio is the amount risked over some period of time with a fixed probability. VaR provides a more sensible measure of the risk of the portfolio than variance since it focuses on losses.

The VaR of a portfolio measures the value which an investor would lose with some small probability, usually between 1 and 10%, over a selected period of time.

Because the VaR represents a hypothetical loss, it is usually a positive number.

The α Value-at-Risk (VaR) of a portfolio is defined as the largest number such that the probability that the loss in portfolio value over some period of time is greater than the VaR is α, that is $P(-R_t > VaR) = \alpha$ where $R_t = W_t - W_{t-1}$ is the change in the value of the portfolio, W_t and the time span depends on the application (e.g. one day or two weeks).

Let r_t represents the return on a portfolio, that is, $r_t = (P_t - P_0)/P_0$ with P_t denoting the price at time t. The α-VaR is $-q_\alpha(r_t)$ where $q_\alpha(r_t)$ is the α-quantile of the portfolio's return. In most cases α is chosen to be some small quantile eg, 1, 5 or 10%, and so VaR should generally be positive.

Expected Shortfall (ES) is defined as the expected value of the portfolio loss given a Value-at-Risk exceedance has occurred. That is

$$ES = E(r_t | r_t < -VaR).$$

7.2 Estimation

Parametric Estimation The simplest form of VaR specifies a parametric model for the distribution of returns and derives the VaR from the α-quantile of this distribution. For example, if $r_t \sim \mathcal{N}(\mu, \sigma^2)$, the α-VaR is $-\mu - \sigma\Phi^{-1}(\alpha)$ and the parameters can be directly estimated using Maximum likelihood with the usual estimators,

In a general parametric VaR model, some distribution for returns which depends on a set of unknown parameters θ is assumed, that is $r_t \sim F_{(\theta)}$ and parameters are estimated by maximum likelihood. The VaR is then $-F_\alpha^{-1}$ where F_α^{-1} is the α-quantile of the estimated distribution.

The models are parsimonious and the parameters estimates are precise yet finding a specification which necessarily includes the true distribution is difficult (or impossible).

7.3 Nonparametric Estimation

At the other end of the spectrum is a pure nonparametric estimate of the VaR using the α-quantile of the empirical distribution of r_t

These estimates are rough and a single new data point may produce very different VaR estimates. Smoothing the estimated quantile using a kernel density generally improves the precision of the estimate when compared to one calculated directly on the sorted returns. This is particularly true if the sample is small.

The advantage of nonparametric estimates of VaR is that they are generally consistent under very weak conditions and that they are trivial to compute. The disadvantage is that the VaR estimates can be poorly estimated – or equivalently that very large samples are needed for estimated quantiles to be accurate – particularly for 1% VaRs (or smaller).

Extreme Value Theory (EVT) in finance

EVT has two significant results. First, the asymptotic distribution of a series of maxima (minima) is modelled and under certain conditions, the distribution of the standardized maximum of the series is shown to converge to the Gumbel, Frechet, or Weibull distributions. A standard form of these three distributions is called the generalized extreme value (GEV) distribution. The second significant result concerns the distribution of excess over a given threshold, where one is interested in modelling the behaviour of the excess loss once a high threshold (loss) is reached. This result is used to estimate the very high quantiles (0.999 and higher). EVT shows that the limiting distribution is a generalized Pareto distribution (GPD).

Fisher Tippet Theorem: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (iid) random variables. and $M_n = Max(X_1, \cdots, X_n)$. If there exist constants $c_n > 0, d_n \in R$, and some non-degenerate distribution H such that $c_n^{-1}(M_n - d_n)$ converges in law (weakly) to H, then H belongs to the type of one of the following dfs:

$$
\begin{aligned}
\Phi_\alpha(x) &= exp[-x^{-\alpha}]I(x > 0) \quad \text{for} \quad \alpha > 0 \quad \text{Frechet} \\
\Xi_\alpha(x) &= exp[-(-x)^\alpha]I(x \le 0) \quad \text{for} \quad \alpha > 0 \quad \text{Weibull} \\
\Lambda(x) &= exp[-exp(-x)] \quad \text{Gumbel}
\end{aligned}
$$

In this case we say that F belongs to the maximum domain of attraction of H and write $F \in MDA(H)$. This result is very significant, since the asymptotic distribution of the maximum always belongs to one of these three distributions, whatever the original distribution. The asymptotic distribution of the maximum can be estimated without making any assumptions about the nature of the original distribution of the observations (unlike with parametric VaR methods), that distribution being generally unknown.

The Frechet distribution corresponds to fat-tailed distributions and has been found to be the most appropriate for fat-tailed financial data.

For a cumulative distribution function (cdf) F, let us introduce the following notations:

1. $\overline{F}(x) = 1 - F(x)$

2. $F_u(x) = Pr[X - u \le x | X > u] = (F(x + u) - F(u))/(1 - F(u))$

A positive function L is slowly varying if, for any $t > 0$, $lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1$. Next we define regularly varying (RV) cdfs. For $\xi > 0$, $F \in RV(\xi)$ if for some slowly varying function L,

$$\overline{F}(x) = x^{-1/\xi} L(x)$$

Theorem (3.3.7 Embrechts et al) The df F belongs to the maximum domain of attraction of Φ_α if and only if $F \in RV(\alpha)$

For $\xi > 0$, the cdf of a generalized Pareto distribution (GPD) is given by

$$G_{\xi,\beta}(x) = 1 - (1 + \frac{\xi x}{\beta})^{-1/\xi}$$

THEOREM (Pickands - Balkema - de Haan) $F \in RV(\xi)$ if and only if there exists a function $\beta(u)$ such that

$$lim_{u \to \infty} sup_{0 \le x < \infty} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$$

The EVT uses PBD Theorem in order to find an approximation to the tail of the loss data. Given a loss data set, one selects a threshold u for which a GPD is a good approximation of the tail. The losses below the threshold are modeled using an empirical distribution; the losses above the threshold are modeled using the GPD.

Now we proceed to discuss the convergence required for quantiles and shortfall estimation.

For a cdf F, denote by $EVT[F, u]$ the approximation to F obtained by using the EVT tail:

$$
\begin{aligned}
EVT[F, u](x) \quad = \quad & F(x) \qquad x \le u \\
& F(u) + (1 - F(u))G_{\xi \beta(u)}(x - u) \quad x > u
\end{aligned}
$$

EVT approximation leads to uniform relative quantile (URQ) convergence if, for any $\epsilon > 0$, there exists a threshold u_0 such that for any $u \ge u_0$ and any quantile q,

$$(1 - \epsilon)Quantile(EVT[F, u], q) \le Quantile(F, q) \le (1 + \epsilon)Quantile(EVT[F, u], q)$$

For $F \in RV(\xi)$, EVT approximation leads to the URQ convergence only if $0 < lim_{x \to \infty} \overline{F}(x) x^{1/\xi} < \infty$.

URQ convergence implies convergence in mean. That is, URQ convergence of EVT approximation implies

$$(1 - \epsilon)Mean(EVT[F, u]) \le Mean(F) \le (1 + \epsilon)Mean(EVT[F, u])$$

8 Multivariate

1. Chapter 17 of Hardle for Copula

2. chapter 11.4 of Hardle for CAPM

3. Chapter 4.8 of Carmona for Term Structure

9 Univariate Time Series

1. chapter 11.6 for Unit root

2. Chapter 13 of Hardle for ARCH/GARCH models

3. chapter 8.3 of Carmona for Stochastic Volatility models

10 Multivariate Time Series

1. chapter 7.1.3 of Carmona for Vector AR

2. chapter 7.1.6 of Carmona for Cointegration

3. chapter 13.4 of Hardle for multivariate GARCH

11 High Frequency data

pages 4,5,7,10,16-18,24,44,45,46,47, 53 of The Econometrics of High Frequency Data by Myklnd and Zhang

Homework 5

1. Obtain the optimal allocation for stratified sampling (pg 217 Glasserman)

2. Prove Thm 17.3 of Hardle

3. 1.3 of Carmona

4. 1.6 of Carmona

5. Refer to Fig 1.13 of Carmona. Using S&P500 weekly log returns report Q-Q plot, kernel density and histogram.

6. Fit bivariate normal with Gumbel copula to utilities.asc Then simulate 1000 observations from tis distribution. Report the q-percentile with $q = 2$ for $X+Y$ under (1) bivariate normal MLE, (2) Simulation.

7. Repeat previous problem with GPD marginals.

Homework 6

From Carmona R book 4.12, 6.11, 7.7, 8.4