# An Introduction to Computational Finance Without Agonizing Pain



P.A. Forsyth*

May 30, 2022

## Contents



*Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, paforsyt@uwaterloo.ca, www.scicom.uwaterloo.ca/ paforsyt, tel: (519) 888-4567x34415, fax: (519) 885-1208

*"Men wanted for hazardous journey, small wages, bitter cold, long months of complete darkness, constant dangers, safe return doubtful. Honour and recognition in case of success."* Advertisement placed by Earnest Shackleton in 1914. He received 5000 replies. An example of extreme risk-seeking behaviour. Hedging with options is used to mitigate risk, and would not appeal to members of Shackleton's expedition.

# 1 The First Option Trade

Many people think that options and futures are recent inventions. However, options have a long history, going back to ancient Greece.

As recorded by Aristotle in *Politics*, the fifth century BC philosopher Thales of Miletus took part in a sophisticated trading strategy. The main point of this trade was to confirm that philosophers could become rich if they so chose. This is perhaps the first rejoinder to the famous question *"If you are so smart, why aren't you rich?"* which has dogged academics throughout the ages.

Thales observed that the weather was very favourable to a good olive crop, which would result in a bumper harvest of olives. If there was an established Athens Board of Olives Exchange, Thales could have simply sold olive futures short (a surplus of olives would cause the price of olives to go down). Since the exchange did not exist, Thales put a deposit on all the olive presses surrounding Miletus. When the olive crop was harvested, demand for olive presses reached enormous proportions (olives were not a storable commodity). Thales then sublet the presses for a profit. Note that by placing a deposit on the presses, Thales was actually manufacturing an option on the olive crop, i.e. the most he could lose was his deposit. If had sold short olive futures, he would have been liable to an unlimited loss, in the event that the olive crop turned out bad, and the price of olives went up. In other words, he had an option on a future of a non-storable commodity.

# 2 The Black-Scholes Equation

This is the basic PDE used in option pricing. We will derive this PDE for a simple case below. Things get much more complicated for real contracts.

## 2.1 Background

Over the past few years derivative securities (options, futures, and forward contracts) have become essential tools for corporations and investors alike. Derivatives facilitate the transfer of financial risks. As such, they may be used to hedge risk exposures or to assume risks in the anticipation of profits. To take a simple yet instructive example, a gold mining firm is exposed to fluctuations in the price of gold. The firm could use a forward contract to fix the price of its future sales. This would protect the firm against a fall in the price of gold, but it would also sacrifice the upside potential from a gold price increase. This could be preserved by using options instead of a forward contract.

Individual investors can also use derivatives as part of their investment strategies. This can be done through direct trading on financial exchanges. In addition, it is quite common for financial products to include some form of embedded derivative. Any insurance contract can be viewed as a *put* option. Consequently, any investment which provides some kind of protection actually includes an option feature. Standard examples include deposit insurance guarantees on savings accounts as well as the provision of being able to redeem a savings bond at par at any time. These types of embedded options are becoming increasingly common and increasingly complex. A prominent current example are investment guarantees being offered by insurance companies ("segregated funds") and mutual funds. In such contracts, the initial investment is guaranteed, and gains can be locked-in (reset) a fixed number of times per year at the option of the contract holder. This is actually a very complex put option, known as a shout option. How much should an investor be willing to pay for this insurance? Determining the fair market value of these sorts of contracts is a problem in *option pricing*.

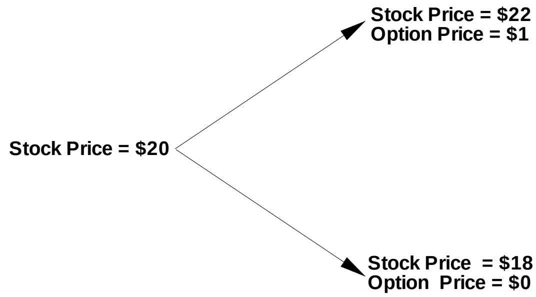



Figure 2.1: A simple case where the stock value can either be \$22 or \$18, with a European call option, $K =$ \$21.

## 2.2 Definitions

Let's consider some simple European put/call options. At some time $T$ in the future (the expiry or exercise date) the holder has the right, but not the obligation, to

- Buy an asset at a prescribed price $K$ (the exercise or strike price). This is a call option.
- Sell the asset at a prescribed price $K$ (the exercise or strike price). This is a put option.

At expiry time $T$, we know with certainty what the value of the option is, in terms of the price of the underlying asset $S$,

$$\begin{aligned} \text{Payoff} &= \max(S-K,0) \text{ for a call} \\ \text{Payoff} &= \max(K-S,0) \text{ for a put} \end{aligned} \tag{2.1}$$

Note that the payoff from an option is always non-negative, since the holder has a right but not an obligation. This contrasts with a forward contract, where the holder *must* buy or sell at a prescribed price.

## 2.3 A Simple Example: The Two State Tree

This example is taken from *Options, futures, and other derivatives*, by John Hull. Suppose the value of a stock is currently \$20. It is known that at the end of three months, the stock price will be either \$22 or \$18. We assume that the stock pays no dividends, and we would like to value a European call option to buy the stock in three months for \$21. This option can have only two possible values in three months: if the stock price is \$22, the option is worth \$1, if the stock price is \$18, the option is worth zero. This is illustrated in Figure 2.1.

In order to price this option, we can set up an imaginary portfolio consisting of the option and the stock, in such a way that there is no uncertainty about the value of the portfolio at the end of three months. Since the portfolio has no risk, the return earned by this portfolio must be the risk-free rate.

Consider a portfolio consisting of a long (positive) position of $\delta$ shares of stock, and short (negative) one call option. We will compute $\delta$ so that the portfolio is riskless. If the stock moves up to \$22 or goes down to \$18, then the value of the portfolio is

$$\begin{aligned} \text{Value if stock goes up} &= \$22\delta - 1 \\ \text{Value if stock goes down} &= \$18\delta - 0 \end{aligned} \tag{2.2}$$

So, if we choose $\delta = .25$, then the value of the portfolio is

$$\begin{aligned} \text{Value if stock goes up} &= \$22\delta - 1 = \$4.50 \\ \text{Value if stock goes down} &= \$18\delta - 0 = \$4.50 \end{aligned} \tag{2.3}$$

So, regardless of whether the stock moves up or down, the value of the portfolio is \$4.50. A risk-free portfolio must earn the risk free rate. Suppose the current risk-free rate is 12%, then the value of the portfolio today must be the present value of \$4.50, or

$$4.50 \times e^{-.12 \times .25} = 4.367$$

The value of the stock today is \$20. Let the value of the option be $V$. The value of the portfolio is

$$\begin{aligned} 20 \times .25 - V &= 4.367 \\ \rightarrow & \quad V = .633 \end{aligned}$$

## 2.4 A hedging strategy

So, if we sell the above option (we hold a short position in the option), then we can hedge this position in the following way. Today, we sell the option for \$.633, borrow \$4.367 from the bank at the risk free rate (this means that we have to pay the bank back \$4.50 in three months), which gives us \$5.00 in cash. Then, we buy .25 shares at \$20.00 (the current price of the stock). In three months time, one of two things happens

- The stock goes up to \$22, our stock holding is now worth \$5.50, we pay the option holder \$1.00, which leaves us with \$4.50, just enough to pay off the bank loan.
- The stock goes down to \$18.00. The call option is worthless. The value of the stock holding is now \$4.50, which is just enough to pay off the bank loan.

Consequently, in this simple situation, we see that the *theoretical* price of the option is the cost for the seller to set up portfolio, which will precisely pay off the option holder and any bank loans required to set up the hedge, at the expiry of the option. In other words, this is price which a hedger requires to ensure that there is always just enough money at the end to net out at zero gain or loss. If the market price of the option was higher than this value, the seller could sell at the higher price and lock in an instantaneous risk-free gain. Alternatively, if the market price of the option was lower than the theoretical, or *fair market* value, it would be possible to lock in a risk-free gain by selling the portfolio short. Any such arbitrage opportunities are rapidly exploited in the market, so that for most investors, we can assume that such opportunities are not possible (the *no arbitrage condition*), and therefore that the market price of the option should be the theoretical price.

Note that this hedge works regardless of whether or not the stock goes up or down. Once we set up this hedge, we don't have a care in the world. The value of the option is also independent of the probability that the stock goes up to \$22 or down to \$18. This is somewhat counterintuitive.

## 2.5 Brownian Motion

Before we consider a model for stock price movements, let's consider the idea of Brownian motion with drift. Suppose $X$ is a random variable, and in time $t \rightarrow t + dt$, $X \rightarrow X + dX$, where

$$dX = \alpha dt + \sigma dZ \tag{2.4}$$

where $\alpha dt$ is the drift term, $\sigma$ is the volatility, and $dZ$ is a random term. The $dZ$ term has the form

$$dZ = \phi\sqrt{dt} \tag{2.5}$$

where $\phi$ is a random variable drawn from a normal distribution with mean zero and variance one ($\phi \sim N(0,1)$, i.e. $\phi$ is normally distributed).

If $E$ is the expectation operator, then

$$E(\phi) = 0 \qquad E(\phi^2) = 1 \ . \tag{2.6}$$

Now in a time interval $dt$, we have

$$\begin{aligned} E(dX) &= E(\alpha dt) + E(\sigma dZ) \\ &= \alpha dt \ , \end{aligned} \tag{2.7}$$

and the variance of $dX$, denoted by $Var(dX)$ is

$$\begin{aligned} Var(dX) &= E([dX - E(dX)]^2) \\ &= E([\sigma dZ]^2) \\ &= \sigma^2 dt \ . \end{aligned} \tag{2.8}$$

Let's look at a discrete model to understand this process more completely. Suppose that we have a discrete lattice of points. Let $X = X_0$ at $t = 0$. Suppose that at $t = \Delta t$,

$$\begin{aligned} X_0 &\rightarrow X_0 + \Delta h \ ; \quad \text{with probability } p \\ X_0 &\rightarrow X_0 - \Delta h \ ; \quad \text{with probability } q \end{aligned} \tag{2.9}$$

where $p + q = 1$. Assume that

- $X$ follows a Markov process, i.e. the probability distribution in the future depends only on where it is now.
- The probability of an up or down move is independent of what happened in the past.
- $X$ can move only up or down $\Delta h$.

At any lattice point $X_0 + i\Delta h$, the probability of an up move is $p$, and the probability of a down move is $q$. The probabilities of reaching any particular lattice point for the first three moves are shown in Figure 2.2. Each move takes place in the time interval $t \rightarrow t + \Delta t$.

Let $\Delta X$ be the change in $X$ over the interval $t \rightarrow t + \Delta t$. Then

$$\begin{aligned} E(\Delta X) &= (p - q)\Delta h \\ E([\Delta X]^2) &= p(\Delta h)^2 + q(-\Delta h)^2 \\ &= (\Delta h)^2, \end{aligned} \tag{2.10}$$

so that the variance of $\Delta X$ is (over $t \rightarrow t + \Delta t$)

$$\begin{aligned} Var(\Delta X) &= E([\Delta X]^2) - [E(\Delta X)]^2 \\ &= (\Delta h)^2 - (p - q)^2(\Delta h)^2 \\ &= 4pq(\Delta h)^2 \ . \end{aligned} \tag{2.11}$$

Now, suppose we consider the distribution of $X$ after $n$ moves, so that $t = n\Delta t$. The probability of $j$ up moves, and $(n - j)$ down moves ($P(n,j)$) is

$$P(n,j) = \frac{n!}{j!(n-j)!} p^j q^{n-j} \tag{2.12}$$

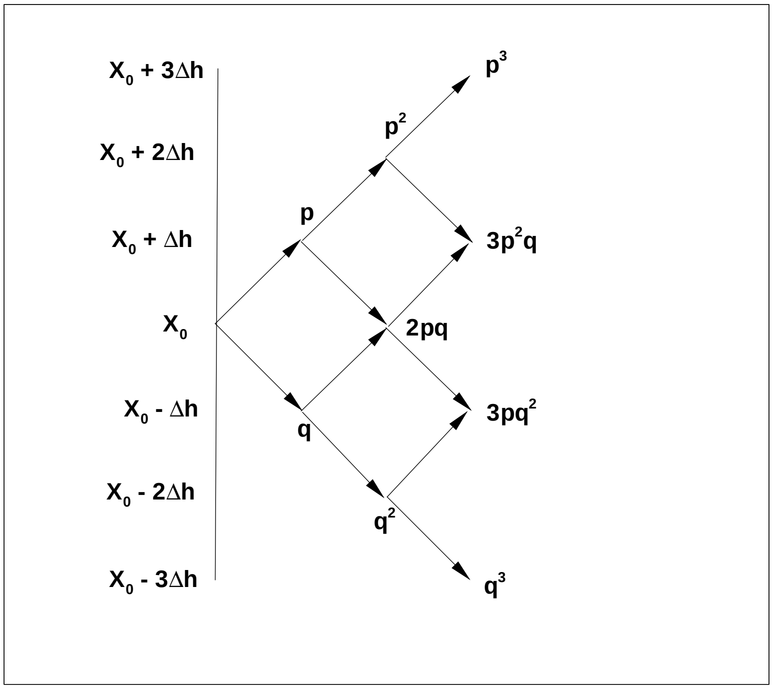



Figure 2.2: Probabilities of reaching the discrete lattice points for the first three moves.

which is just a binomial distribution. Now, if $X_n$ is the value of $X$ after $n$ steps on the lattice, then

$$\begin{aligned} E(X_n - X_0) &= nE(\Delta X) \\ Var(X_n - X_0) &= nVar(\Delta X) \ , \end{aligned} \tag{2.13}$$

which follows from the properties of a binomial distribution, (each up or down move is independent of previous moves). Consequently, from equations (2.10, 2.11, 2.13) we obtain

$$\begin{aligned} E(X_n - X_0) &= n(p-q)\Delta h \\ &= \frac{t}{\Delta t}(p-q)\Delta h \\ Var(X_n - X_0) &= n4pq(\Delta h)^2 \\ &= \frac{t}{\Delta t}4pq(\Delta h)^2 \end{aligned} \tag{2.14}$$

Now, we would like to take the limit at $\Delta t \to 0$ in such a way that the mean and variance of $X$, after a finite time $t$ is independent of $\Delta t$, and we would like to recover

$$\begin{aligned} dX &= \alpha dt + \sigma dZ \\ E(dX) &= \alpha dt \\ Var(dX) &= \sigma^2 dt \end{aligned} \tag{2.15}$$

as $\Delta t \to 0$. Now, since $0 \le p, q \le 1$, we need to choose $\Delta h = Const\ \sqrt{\Delta t}$. Otherwise, from equation (2.14) we get that $Var(X_n - X_0)$ is either 0 or infinite after a finite time. (Stock variances do not have either of these properties, so this is obviously not a very interesting case).

Let's choose $\Delta h = \sigma\sqrt{\Delta t}$, which gives (from equation (2.14))

$$\begin{aligned} E(X_n - X_0) &= (p-q)\frac{\sigma t}{\sqrt{\Delta t}} \\ Var(X_n - X_0) &= t4pq\sigma^2 \end{aligned} \tag{2.16}$$

Now, for $E(X_n - X_0)$ to be independent of $\Delta t$ as $\Delta t \to 0$, we must have

$$(p-q) = \ Const.\ \sqrt{\Delta t} \tag{2.17}$$

If we choose

$$p - q = \frac{\alpha}{\sigma}\sqrt{\Delta t} \tag{2.18}$$

we get

$$\begin{aligned} p &= \frac{1}{2}[1 + \frac{\alpha}{\sigma}\sqrt{\Delta t}] \\ q &= \frac{1}{2}[1 - \frac{\alpha}{\sigma}\sqrt{\Delta t}] \end{aligned} \tag{2.19}$$

Now, putting together equations (2.16-2.19) gives

$$\begin{aligned} E(X_n - X_0) &= \alpha t \\ Var(X_n - X_0) &= t\sigma^2(1 - \frac{\alpha^2}{\sigma^2}\Delta t) \\ &= t\sigma^2 \ \ ; \ \ \Delta t \to 0 \ . \end{aligned} \tag{2.20}$$

Now, let's imagine that $X(t_n) - X(t_0) = X_n - X_0$ is very small, so that $X_n - X_0 \simeq dX$ and $t_n - t_0 \simeq dt$, so that equation (2.20) becomes

$$\begin{aligned} E(dX) &= \alpha \ dt \\ Var(dX) &= \sigma^2 \ dt \ . \end{aligned} \tag{2.21}$$

which agrees with equations (2.7-2.8). Hence, in the limit as $\Delta t \to 0$, we can interpret the random walk for $X$ on the lattice (with these parameters) as the solution to the stochastic differential equation (SDE)

$$\begin{aligned} dX &= \alpha \ dt + \sigma \ dZ \\ & dZ = \phi\sqrt{dt}. \end{aligned} \tag{2.22}$$

Consider the case where $\alpha = 0, \sigma = 1$, so that $dX = dZ = \simeq Z(t_i) - Z(t_{i-1}) = Z_i - Z_{i-1} = X_i - X_{i-1}$. Now we can write

$$\int_0^t dZ = \lim_{\Delta t \to 0} \sum_i (Z_{i+1} - Z_i) = (Z_n - Z_0) \ . \tag{2.23}$$

From equation (2.20) ($\alpha = 0, \sigma = 1$) we have

$$\begin{aligned} E(Z_n - Z_0) &= 0 \\ Var(Z_n - Z_0) &= t \ . \end{aligned} \tag{2.24}$$

Now, if $n$ is large ($\Delta t \to 0$), recall that the binomial distribution (2.12) tends to a normal distribution. From equation (2.24), we have that the mean of this distribution is zero, with variance $t$, so that

$$\begin{aligned} (Z_n - Z_0) &\sim N(0,t) \\ &= \int_0^t dZ \ . \end{aligned} \tag{2.25}$$

In other words, after a finite time $t$, $\int_0^t dZ$ is normally distributed with mean zero and variance $t$ (the limit of a binomial distribution is a normal distribution).

Recall that have that $Z_i - Z_{i-1} = \sqrt{\Delta t}$ with probability $p$ and $Z_i - Z_{i-1} = -\sqrt{\Delta t}$ with probability $q$. Note that $(Z_i - Z_{i-1})^2 = \Delta t$, with certainty, so that we can write

$$(Z_i - Z_{i-1})^2 \simeq (dZ)^2 = \Delta t \ . \tag{2.26}$$

To summarize

- We can interpret the SDE

$$\begin{aligned} dX &= \alpha\ dt + \sigma\ dZ \\ & dZ = \phi\sqrt{dt}. \end{aligned} \tag{2.27}$$

  as the limit of a discrete random walk on a lattice as the timestep tends to zero.

- $Var(dZ) = dt$, otherwise, after any finite time, the $Var(X_n - X_0)$ is either zero or infinite.

- We can *integrate* the term $dZ$ to obtain

$$\begin{aligned} \int_0^t dZ &= Z(t) - Z(0) \\ &\sim N(0,t) \ . \end{aligned} \tag{2.28}$$

Going back to our lattice example, note that the total distance traveled over any finite interval of time becomes infinite,

$$E(|\Delta X|) \quad = \quad \Delta h \tag{2.29}$$

so that the the total distance traveled in $n$ steps is

$$\begin{aligned} n\Delta h &= \frac{t}{\Delta t}\Delta h \\ &= \frac{t\sigma}{\sqrt{\Delta t}} \end{aligned} \tag{2.30}$$

which goes to infinity as $\Delta t \to 0$. Similarly,

$$\frac{\Delta x}{\Delta t} \quad = \quad \pm\infty \ . \tag{2.31}$$

Consequently, Brownian motion is very jagged at every timescale. These paths are not differentiable, i.e. $\frac{dx}{dt}$ does not exist, so we cannot speak of

$$E(\frac{dx}{dt}) \tag{2.32}$$

but we can possibly define

$$\frac{E(dx)}{dt} \ . \tag{2.33}$$

We can verify that taking the limit as $\Delta t \to 0$ on the discrete lattice converges to the normal density. Consider the data in Table 2.1. The random walk on the lattice was simulated using a Monte Carlo approach. Starting at $X_0$, the particle was moved up with probability $p$ (2.19), and down with probability $(1-p)$. A random number was used to determine the actual move. At the next node, this was repeated, until we obtain the position of $X$ after $n$ steps, $X_n$. This is repeated many times. We can then determine the mean and variance of these outcomes (see Table 2.2). The mean and variance of $e^X$ have also been included, since this is relevant for the case of Geometric Brownian Motion, which will be studied in the next Section. A histogram of the outcomes is shown in Figure 2.5.

The Matlab M file used to generate the walk on the lattice is given in Algorithm 2.34.

| | |
|---|---|
| $T$ | 1.0 |
| $\sigma$ | .2 |
| $\alpha$ | .10 |
| $X_{init}$ | 0 |
| Number of simulations | 50000 |
| Number of timesteps | 400 |

Table 2.1: *Data used in simulation of discrete walk on a lattice.*

| Variable | Mean | Standard Deviation |
|---|---|---|
| X(T) | 0.10093 | 0.20035 |
| $e^{X(T)}$ | .22813 | 1.1286 |

Table 2.2: *Test results: discrete lattice walk, data in Table 2.1.*

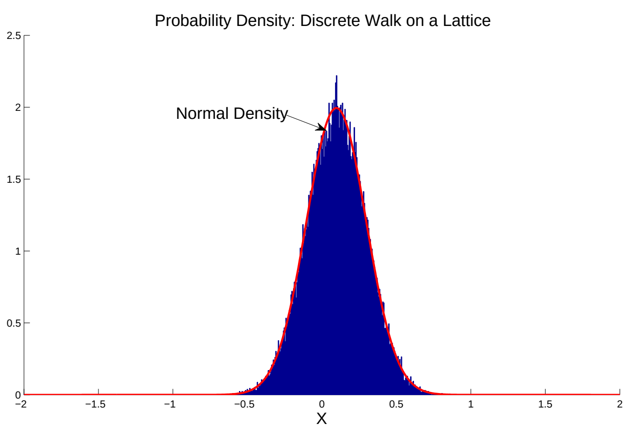



Figure 2.3: Normalized histogram of discrete lattice walk simulations. Normal density with mean .1, standard deviation .2 also shown.

**Vectorized M file For Lattice Walk**

```
 function [X_new] = walk_sim( N_sim,N,...
                mu,  T,  sigma, X_init)
%
%  N_sim   number of simulations
%  N       number of timesteps
%  X_init  initial value
%   T                expiry time
%   sigma            volatility
%   mu               drift
%
%   lattice factors
%
%
  delt = T/N;% timestep size
  up = sigma*sqrt(delt);
  down = - sigma*sqrt(delt);

  p = 1./2.*( 1. + mu/sigma*sqrt( delt ) );

  X_new = zeros(N_sim,1);
  X_new(1:N_sim,1)  = X_init;

  ptest = zeros(N_sim, 1);

 for i=1:N % timestep loop
    % now, for each timestep, generate info for
    % all simulations

     ptest(:,1) = rand(N_sim,1);
     ptest(:,1) = (ptest(:,1) <= p);  % = 1 if up move
                                      % = 0 if downmove
     X_new(:,1) = X_new(:,1) + ptest(:,1)*up + (1.-ptest(:,1))*down;

     %  end of generation of all data for all simulations
     %  for this timestep
 end % timestep loop
```

(2.34)

## 2.6 Geometric Brownian motion with drift

Of course, the actual path followed by stock is more complex than the simple situation described above. More realistically, we assume that the relative changes in stock prices (the returns) follow Brownian motion with drift. We suppose that in an infinitesimal time $dt$, the stock price $S$ changes to $S + dS$, where

$$\frac{dS}{S} = \mu dt + \sigma dZ \tag{2.35}$$

where $\mu$ is the drift rate, $\sigma$ is the volatility, and $dZ$ is the increment of a Wiener process,

$$dZ \quad = \quad \phi\sqrt{dt} \tag{2.36}$$

where $\phi \sim N(0, 1)$. Equations (2.35) and (2.36) are called geometric Brownian motion with drift. So, superimposed on the upward (relative) drift is a (relative) random walk. The degree of randomness is given

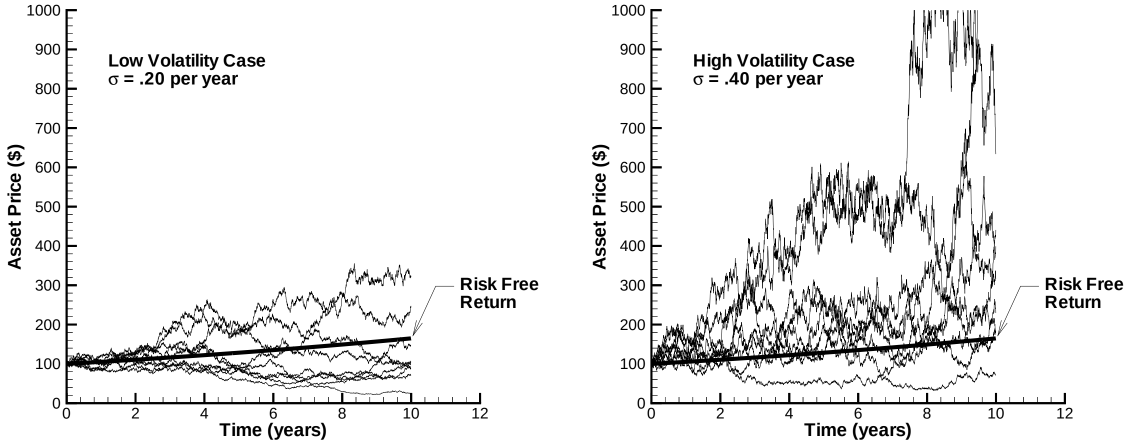



Figure 2.4: Realizations of asset price following geometric Brownian motion. Left: low volatility case; right: high volatility case. Risk-free rate of return $r = .05$.

by the volatility $\sigma$. Figure 2.4 gives an illustration of ten realizations of this random process for two different values of the volatility. In this case, we assume that the drift rate $\mu$ equals the risk free rate.

Note that

$$\begin{aligned} E(dS) &= E(\sigma S dZ + \mu S dt) \\ &= \mu S dt \\ &\quad \text{since } E(dZ) = 0 \end{aligned} \tag{2.37}$$

and that the variance of $dS$ is

$$\begin{aligned} Var[dS] &= E(dS^2) - [E(dS)]^2 \\ &= E(\sigma^2 S^2 dZ^2) \\ &= \sigma^2 S^2 dt \end{aligned} \tag{2.38}$$

so that $\sigma$ is a measure of the degree of randomness of the stock price movement.

Equation (2.35) is a *stochastic differential equation*. The normal rules of calculus don't apply, since for example

$$\begin{aligned} \frac{dZ}{dt} &= \phi \frac{1}{\sqrt{dt}} \\ &\rightarrow \infty \text{ as } dt \rightarrow 0 \quad . \end{aligned}$$

The study of these sorts of equations uses results from stochastic calculus. However, for our purposes, we need only one result, which is Ito's Lemma (see *Derivatives: the theory and practice of financial engineering*, by P. Wilmott). Suppose we have some function $G = G(S, t)$, where $S$ follows the stochastic process equation (2.35), then, in small time increment $dt$, $G \rightarrow G + dG$, where

$$dG = \left( \mu S \frac{\partial G}{\partial S} + \frac{\sigma^2 S^2}{2} \frac{\partial^2 G}{\partial S^2} + \frac{\partial G}{\partial t} \right) dt + \sigma S \frac{\partial G}{\partial S} dZ \tag{2.39}$$

An informal derivation of this result is given in the following section.

### 2.6.1 Ito's Lemma

We give an informal derivation of Ito's lemma (2.39). Suppose we have a variable $S$ which follows

$$dS = a(S,t)dt + b(S,t)dZ \tag{2.40}$$

where $dZ$ is the increment of a Weiner process.

Now since

$$dZ^2 \quad = \quad \phi^2 dt \tag{2.41}$$

where $\phi$ is a random variable drawn from a normal distribution with mean zero and unit variance, we have that, if $E$ is the expectation operator, then

$$E(\phi) = 0 \qquad E(\phi^2) = 1 \tag{2.42}$$

so that the expected value of $dZ^2$ is

$$E(dZ^2) = dt \tag{2.43}$$

Now, it can be shown (see Section 6) that in the limit as $dt \to 0$, we have that $\phi^2 dt$ becomes non-stochastic, so that with probability one

$$dZ^2 \to dt \quad \text{as } dt \to 0 \tag{2.44}$$

Now, suppose we have some function $G = G(S,t)$, then

$$dG = G_S dS + G_t dt + G_{SS}\frac{dS^2}{2} + ... \tag{2.45}$$

Now (from (2.40) )

$$\begin{aligned} (dS)^2 &= (adt + b\ dZ)^2 \\ &= a^2dt^2 + ab\ dZdt + b^2dZ^2 \end{aligned} \tag{2.46}$$

Since $dZ = O(\sqrt{dt})$ and $dZ^2 \to dt$, equation (2.46) becomes

$$(dS)^2 = b^2dZ^2 + O((dt)^{3/2}) \tag{2.47}$$

or

$$(dS)^2 \to b^2dt \ \ \text{as } dt \to 0 \tag{2.48}$$

Now, equations(2.40,2.45,2.48) give

$$\begin{aligned} dG &= G_S dS + G_t dt + G_{SS}\frac{dS^2}{2} + ... \\ &= G_S(a\ dt + b\ dZ) + dt(G_t + G_{SS}\frac{b^2}{2}) \\ &= G_S b\ dZ + (aG_S + G_{SS}\frac{b^2}{2} + G_t)dt \end{aligned} \tag{2.49}$$

So, we have the result that if

$$dS = a(S,t)dt + b(S,t)dZ \tag{2.50}$$

and if $G = G(S,t)$, then

$$dG = G_S b\ dZ + (a\ G_S + G_{SS}\frac{b^2}{2} + G_t)dt \tag{2.51}$$

Equation (2.39) can be deduced by setting $a = \mu S$ and $b = \sigma S$ in equation (2.51).

### 2.6.2 Some uses of Ito's Lemma

Suppose we have

$$dS = \mu dt + \sigma dZ \ . \tag{2.52}$$

If $\mu, \sigma = Const.$, then this can be integrated (from $t = 0$ to $t = t$) exactly to give

$$S(t) = S(0) + \mu t + \sigma(Z(t) - Z(0)) \tag{2.53}$$

and from equation (2.28)

$$Z(t) - Z(0) \quad \sim \quad N(0, t) \tag{2.54}$$

Note that when we say that we solve a stochastic differential equation exactly, this means that we have an expression for the distribution of $S(T)$.

Suppose instead we use the more usual geometric Brownian motion

$$dS = \mu S dt + \sigma S dZ \tag{2.55}$$

Let $F(S) = \log S$, and use Ito's Lemma

$$\begin{aligned} dF &= F_S S \sigma dZ + (F_S \mu S + F_{SS} \frac{\sigma^2 S^2}{2} + F_t) dt \\ &= (\mu - \frac{\sigma^2}{2}) dt + \sigma dZ \ , \end{aligned} \tag{2.56}$$

so that we can integrate this to get

$$F(t) = F(0) + (\mu - \frac{\sigma^2}{2}) t + \sigma(Z(t) - Z(0)) \tag{2.57}$$

or, since $S = e^F$,

$$S(t) = S(0) \exp[(\mu - \frac{\sigma^2}{2}) t + \sigma(Z(t) - Z(0))] \ . \tag{2.58}$$

Unfortunately, these cases are about the only situations where we can exactly integrate the SDE (constant $\sigma, \mu$).

### 2.6.3 Some more uses of Ito's Lemma

We can often use Ito's Lemma and some algebraic tricks to determine some properties of distributions. Let

$$dX = a(X, t) \ dt + b(X, t) \ dZ \ , \tag{2.59}$$

then if $G = G(X)$, then

$$dG = \left[ a G_X + G_t + \frac{b^2}{2} G_{XX} \right] \ dt + G_X b \ dZ \ . \tag{2.60}$$

If $E[X] = \bar{X}$, then ($b(X, t)$ and $dZ$ are independent)

$$\begin{aligned} E[dX] &= d \ E[S] = d\bar{X} \\ &= E[a \ dt] + E[b] \ E[dZ] \\ &= E[a \ dt] \ , \end{aligned} \tag{2.61}$$

so that

$$\begin{aligned}\frac{d\,\bar{X}}{dt} &= E[a] = \bar{a} \\ \bar{X} = E\left[\int_0^t a\ dt\right] & . \end{aligned} \tag{2.62}$$

Let $\bar{G} = E[(X - \bar{X})^2] = var(X)$, then

$$\begin{aligned} d\bar{G} &= E\left[dG\right] \\ &= E[2(X - \bar{X})a - 2(X - \bar{X})\bar{a} + b^2]\ dt + E[2b(X - \bar{X})]E[dZ] \\ &= E[b^2\ dt] + E[2(X - \bar{X})(a - \bar{a})\ dt]\ , \end{aligned} \tag{2.63}$$

which means that

$$\bar{G} = var(X) = E\left[\int_0^t b^2\ dt\right] + E\left[\int_0^t 2(a - \bar{a})(X - \bar{X})\ dt\right]\ . \tag{2.64}$$

In a particular case, we can sometimes get more useful expressions. If

$$dS = \mu S\ dt + \sigma S\ dZ \tag{2.65}$$

with $\mu, \sigma$ constant, then

$$\begin{aligned} E[dS] &= d\bar{S} = E[\mu S]\ dt \\ &= \mu\bar{S}\ dt\ , \end{aligned} \tag{2.66}$$

so that

$$\begin{aligned} d\bar{S} &= \mu\bar{S}\ dt \\ \bar{S} &= S_0 e^{\mu t}\ . \end{aligned} \tag{2.67}$$

Now, let $G(S) = S^2$, so that $E[G] = \bar{G} = E[S^2]$, then (from Ito's Lemma)

$$\begin{aligned} d\,\bar{G} &= E[2\mu S^2 + \sigma^2 S^2]\ dt + E[2S^2\sigma]E[dZ] \\ &= E[2\mu S^2 + \sigma^2 S^2]\ dt \\ &= (2\mu + \sigma^2)\bar{G}\ dt\ , \end{aligned} \tag{2.68}$$

so that

$$\begin{aligned} \bar{G} &= \bar{G}_0 e^{(2\mu+\sigma^2)t} \\ & E[S^2] = S_0^2 e^{(2\mu+\sigma^2)t}\ . \end{aligned} \tag{2.69}$$

From equations (2.67) and (2.69) we then have

$$\begin{aligned} var(S) &= E[S^2] - (E[S])^2 \\ &= E[S^2] - \bar{S}^2 \\ &= S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1) \\ &= \bar{S}^2(e^{\sigma^2 t} - 1)\ . \end{aligned} \tag{2.70}$$

One can use the same ideas to compute the skewness, $E[(S - \bar{S})^3]$. If $G(S) = S^3$ and $\bar{G} = E[G(S)] = E[S^3]$, then

$$\begin{aligned} d\bar{G} &= E[\mu S \cdot 3S^2 + \sigma^2 S^2/2 \cdot 3 \cdot 2S]\ dt + E[3S^2\sigma S]E[dZ] \\ &= E[3\mu S^3 + 3\sigma^2 S^3] \\ &= 3(\mu + \sigma^2)\bar{G}\ , \end{aligned} \tag{2.71}$$

so that

$$
\begin{aligned}
\bar{G} &= E[S^3] \\
&= S_0^3 e^{3(\mu+\sigma^2)t} \ . \qquad (2.72)
\end{aligned}
$$

We can then obtain the skewness from

$$
\begin{aligned}
E[(S-\bar{S})^3] &= E[S^3 - 2S^2\bar{S} - 2S\bar{S}^2 + \bar{S}^3] \\
&= E[S^3] - 2\bar{S}E[S^2] - \bar{S}^3 \ . \qquad (2.73)
\end{aligned}
$$

Equations (2.67, 2.69, 2.72) can then be substituted into equation (2.73) to get the desired result.

### 2.6.4 More on GBM with constant coefficients

Suppose

$$
dS = \mu S \ dt + \sigma S \ dZ \qquad (2.74)
$$

then we know from equation (2.67) that

$$
E[S] = \bar{S} = S_0 e^{\mu t} \ . \qquad (2.75)
$$

Let $X = \log S$, then from Ito's Lemma we get

$$
dX = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma \ dZ \ , \qquad (2.76)
$$

which can be integrated (assuming $\mu, \sigma$ are constants)

$$
X = X_0 + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma Z(t) \ , \qquad (2.77)
$$

where we have assumed (without loss of generality) that $Z(0) = 0$. In terms of $S$, equation (2.77) becomes

$$
S = S_0 e^{(\mu-\sigma^2/2)t} e^{\sigma Z(t)} \ . \qquad (2.78)
$$

Comparing equation (2.75) and (2.78), this implies that

$$
E[e^{\sigma Z(t)}] = e^{(\sigma^2/2)t} \ . \qquad (2.79)
$$

We can verify this directly. Let

$$
Y = g(W,t) = e^{\sigma W} \ ; \ W = Z(t) \ ; \ dW = dZ \ . \qquad (2.80)
$$

Ito's Lemma then gives

$$
\begin{aligned}
dY &= g_t \ dt + g_W \ dW + \frac{g_{WW}}{2} \ dW^2 \\
&= 0 + \sigma e^{\sigma W} \ dZ + \frac{\sigma^2}{2} e^{\sigma W} \ dt \\
&= \sigma Y \ dZ + \frac{\sigma^2}{2} Y \ dt \ , \qquad (2.81)
\end{aligned}
$$

so that

$$
E[dY] = d\bar{Y} = \frac{\sigma^2}{2} \bar{Y} \ dt \qquad (2.82)
$$

which gives

$$\begin{aligned}\bar{Y} &= Y_0 e^{(\sigma^2/2)t} \\ &= e^{(\sigma^2/2)t}\end{aligned} \tag{2.83}$$

where once again, we assume that $W_0 = Z_0 = 0$. Finally, from the definition of $Y$ we have

$$E[e^{\sigma Z(t)}] = e^{(\sigma^2/2)t} \tag{2.84}$$

which agrees with equation (2.79).

### 2.6.5 Integration by Parts

Let $X(t)$, $Y(t)$ be two stochastic variables, with $X(t_i) = X_i$ and $Y(t_i) = Y_i$, then

$$(X_{i+1} - X_i)(Y_{i+1} - Y_i) = X_{i+1}Y_{i+1} - X_iY_i - X_i(Y_{i+1} - Y_i) - Y_i(X_{i+1} - X_i) \ . \tag{2.85}$$

Hence

$$\sum_{i=1}^{i=N}(X_{i+1} - X_i)(Y_{i+1} - Y_i) = X_{N+1}Y_{N+1} - X_1Y_1 - \textstyle\sum_{i=1}^{i=N} X_i(Y_{i+1} - Y_i) - \sum_{i=1}^{i=N} Y_i(X_{i+1} - X_i) \ . \tag{2.86}$$

Let $\Delta t \to 0$, then the sums in equation (2.86) become Ito stochastic integrals

$$\int_0^T dX(t')dY(t') = [XY]_0^T - \int_0^T X(t')dY(t') - \int_0^T Y(t')dX(t') \ , \tag{2.87}$$

which we can write as the Ito integration by parts rule

$$d(XY) = Y\ dX + X\ dY + dX\ dY \ . \tag{2.88}$$

Note the extra term $dX\ dY$ in equation (2.88) compared with the non-stochastic integration by parts rule.

## 2.7 The Black-Scholes Analysis

Assume

- The stock price follows geometric Brownian motion, equation (2.35).
- The risk-free rate of return is a constant $r$.
- There are no arbitrage opportunities, i.e. all risk-free portfolios must earn the risk-free rate of return.
- Short selling is permitted (i.e. we can own negative quantities of an asset).

Suppose that we have an option whose value is given by $V = V(S, t)$. Construct an imaginary portfolio, consisting of one option, and a number of $(-(\alpha^h))$ of the underlying asset. (If $(\alpha^h) > 0$, then we have sold the asset short, i.e. we have borrowed an asset, sold it, and are obligated to give it back at some future date).

The value of this portfolio $P$ is

$$P = V - (\alpha^h)S \tag{2.89}$$

In a small time $dt$, $P \to P + dP$,

$$dP = dV - (\alpha^h)dS \tag{2.90}$$

Note that in equation (2.90) we not included a term $(\alpha^h)_S S$. This is actually a rather subtle point, since we shall see (later on) that $(\alpha^h)$ actually depends on $S$. However, if we think of a real situation, at any instant in time, we must choose $(\alpha^h)$, and then we hold the portfolio while the asset moves randomly. So, equation (2.90) is actually the change in the value of the portfolio, not a differential. If we were taking a true differential then equation (2.90) would be

$$dP = dV - (\alpha^h)dS - Sd(\alpha^h)$$

but we have to remember that $(\alpha^h)$ does not change over a small time interval, since we pick $(\alpha^h)$, and then $S$ changes randomly. We are not allowed to *peek into the future*, (otherwise, we could get rich without risk, which is not permitted by the no-arbitrage condition) and hence $(\alpha^h)$ is not allowed to contain any information about future asset price movements. The principle of *no peeking into the future* is why Ito stochastic calculus is used. Other forms of stochastic calculus are used in Physics applications (i.e. turbulent flow).

Substituting equations (2.35) and (2.39) into equation (2.90) gives

$$dP = \sigma S \left(V_S - (\alpha^h)\right) dZ + \left(\mu S V_S + \frac{\sigma^2 S^2}{2} V_{SS} + V_t - \mu(\alpha^h) S\right) dt \tag{2.91}$$

We can make this portfolio riskless over the time interval $dt$, by choosing $(\alpha^h) = V_S$ in equation (2.91). This eliminates the $dZ$ term in equation (2.91). (This is the analogue of our choice of the amount of stock in the riskless portfolio for the two state tree model.) So, letting

$$(\alpha^h) = V_S \tag{2.92}$$

then substituting equation (2.92) into equation (2.91) gives

$$dP = \left(V_t + \frac{\sigma^2 S^2}{2} V_{SS}\right) dt \tag{2.93}$$

Since $P$ is now risk-free in the interval $t \to t + dt$, then no-arbitrage says that

$$dP = rPdt \tag{2.94}$$

Therefore, equations (2.93) and (2.94) give

$$rPdt = \left(V_t + \frac{\sigma^2 S^2}{2} V_{SS}\right) dt \tag{2.95}$$

Since

$$P = V - (\alpha^h)S = V - V_S S \tag{2.96}$$

then substituting equation (2.96) into equation (2.95) gives

$$V_t + \frac{\sigma^2 S^2}{2} V_{SS} + rSV_S - rV = 0 \tag{2.97}$$

which is the Black-Scholes equation. Note the rather remarkable fact that equation (2.97) is independent of the drift rate $\mu$.

Equation (2.97) is solved backwards in time from the option expiry time $t = T$ to the present $t = 0$.

## 2.8 Hedging in Continuous Time

We can construct a hedging strategy based on the solution to the above equation. Suppose we sell an option at price $V$ at $t = 0$. Then we carry out the following

- We sell one option worth $V$. (This gives us $V$ in cash initially).

- We borrow $(S\frac{\partial V}{\partial S} - V)$ from the bank.
- We buy $\frac{\partial V}{\partial S}$ shares at price $S$.

At every instant in time, we adjust the amount of stock we own so that we always have $\frac{\partial V}{\partial S}$ shares. Note that this is a *dynamic* hedge, since we have to continually rebalance the portfolio. Cash will flow into and out of the bank account, in response to changes in $S$. If the amount in the bank is positive, we receive the risk free rate of return. If negative, then we borrow at the risk free rate.

So, our hedging portfolio will be

- Short one option worth $V$.
- Long $\frac{\partial V}{\partial S}$ shares at price $S$.
- $V - S\frac{\partial V}{\partial S}$ cash in the bank account.

At any instant in time (including the terminal time), this portfolio can be liquidated and any obligations implied by the short position in the option can be covered, at *zero gain or loss*, regardless of the value of $S$. Note that given the receipt of the cash for the option, this strategy is *self-financing*.

## 2.9 The option price

So, we can see that the price of the option valued by the Black-Scholes equation is the market price of the option at any time. If the price was higher then the Black-Scholes price, we could construct the hedging portfolio, dynamically adjust the hedge, and end up with a positive amount at the end. Similarly, if the price was lower than the Black-Scholes price, we could short the hedging portfolio, and end up with a positive gain. By the *no-arbitrage* condition, this should not be possible.

Note that we are *not* trying to predict the price movements of the underlying asset, which is a random process. The value of the option is based on a hedging strategy which is *dynamic*, and must be continuously rebalanced. The price is the cost of setting up the hedging portfolio. The Black-Scholes price is *not* the expected payoff.

The price given by the Black-Scholes price is *not* the value of the option to a speculator, who buys and holds the option. A speculator is making bets about the underlying drift rate of the stock (note that the drift rate does not appear in the Black-Scholes equation). For a speculator, the value of the option is given by an equation similar to the Black-Scholes equation, except that the drift rate appears. In this case, the price can be interpreted as the *expected payoff* based on the guess for the drift rate. But this is art, not science!

## 2.10 American early exercise

Actually, most options traded are *American* options, which have the feature that they can be exercised at any time. Consequently, an investor acting optimally, will always exercise the option if the value falls below the payoff or exercise value. So, the value of an American option is given by the solution to equation (2.97) with the additional constraint

$$V(S,t) \geq \begin{cases} \max(S-K,0) & \text{for a call} \\ \max(K-S,0) & \text{for a put} \end{cases} \tag{2.98}$$

Note that since we are working backwards in time, we *know* what the option is worth in future, and therefore we can determine the optimal course of action.

In order to write equation (2.97) in more conventional form, define $\tau = T - t$, so that equation (2.97) becomes

$$\begin{aligned}
V_\tau &= \frac{\sigma^2 S^2}{2} V_{SS} + rSV_S - rV \\
V(S, \tau = 0) &= \begin{cases} \max(S-K, 0) & \text{for a call} \\ \max(K-S, 0) & \text{for a put} \end{cases} \\
V(0, \tau) &\to V_\tau = -rV \\
V(S = \infty, \tau) &\to \begin{cases} \simeq S & \text{for a call} \\ \simeq 0 & \text{for a put} \end{cases}
\end{aligned} \tag{2.99}$$

If the option is American, then we also have the additional constraints

$$V(S, \tau) \geq \begin{cases} \max(S-K, 0) & \text{for a call} \\ \max(K-S, 0) & \text{for a put} \end{cases} \tag{2.100}$$

Define the operator

$$LV \equiv V_\tau - (\frac{\sigma^2 S^2}{2} V_{SS} + rSV_S - rV) \tag{2.101}$$

and let $V(S, 0) = V^*$. More formally, the American option pricing problem can be stated as

$$\begin{aligned}
LV &\geq 0 \\
V - V^* &\geq 0 \\
(V - V^*)LV &= 0
\end{aligned} \tag{2.102}$$

# 3 The Risk Neutral World

Suppose instead of valuing an option using the above no-arbitrage argument, we wanted to know the expected value of the option. We can imagine that we are buying and holding the option, and not hedging. If we are considering the value of risky cash flows in the future, then these cash flows should be discounted at an appropriate discount rate, which we will call $\rho$ (i.e. the riskier the cash flows, the higher $\rho$).

Consequently the value of an option today can be considered to the be the discounted future value. This is simply the old idea of net present value. Regard $S$ today as known, and let $V(S + dS, t + dt)$ be the value of the option at some future time $t + dt$, which is uncertain, since $S$ evolves randomly. Thus

$$V(S, t) = \frac{1}{1 + \rho dt} E(V(S + dS, t + dt)) \tag{3.1}$$

where $E(...)$ is the expectation operator, i.e. the expected value of $V(S + dS, t + dt)$ given that $V = V(S, t)$ at $t = t$. We can rewrite equation (3.1) as (ignoring terms of $o(dt)$, where $o(dt)$ represents terms that go to zero faster than $dt$ )

$$\rho dt V(S, t) = E(V(S, t) + dV) - V(S, t) \ . \tag{3.2}$$

Since we regard $V$ as the expected value, so that $E[V(S, t)] = V(S, t)$, and then

$$E(V(S, t) + dV) - V(S, t) = E(dV) \ , \tag{3.3}$$

so that equation (3.2) becomes

$$\rho dt V(S, t) = E(dV) \ . \tag{3.4}$$

Assume that

$$\frac{dS}{S} = \mu dt + \sigma dZ \ . \tag{3.5}$$

From Ito's Lemma (2.39) we have that

$$dV = \left( V_t + \frac{\sigma^2 S^2}{2} V_{SS} + \mu S V_S \right) dt + \sigma S V_S \; dZ \; . \tag{3.6}$$

Noting that

$$E(dZ) = 0 \tag{3.7}$$

then

$$E(dV) = \left( V_t + \frac{\sigma^2 S^2}{2} V_{SS} + \mu S V_S \right) dt \; . \tag{3.8}$$

Combining equations (3.4-3.8) gives

$$V_t + \frac{\sigma^2 S^2}{2} V_{SS} + \mu S V_S - \rho V = 0 \; . \tag{3.9}$$

Equation (3.9) is the PDE for the *expected value* of an option. If we are not hedging, maybe this is the value that we are interested in, not the no-arbitrage value. However, if this is the case, we have to estimate the drift rate $\mu$, and the discount rate $\rho$. Estimating the appropriate discount rate is always a thorny issue.

Now, note the interesting fact, if we set $\rho = r$ and $\mu = r$ in equation (3.9) then we simply get the Black-Scholes equation (2.97).

This means that the no-arbitrage price of an option is identical to the expected value if $\rho = r$ and $\mu = r$. In other words, we can determine the no-arbitrage price by pretending we are living in a world where all assets drift at rate $r$, and all investments are discounted at rate $r$. This is the so-called *risk neutral* world.

This result is the source of endless confusion. It is best to think of this as simply a mathematical fluke. This does not have any reality. Investors would be very stupid to think that the drift rate of risky investments is $r$. I'd rather just buy risk-free bonds in this case. There is in reality no such thing as a risk-neutral world. Nevertheless, this result is useful for determining the no-arbitrage value of an option using a Monte Carlo approach. Using this numerical method, we simply assume that

$$dS = rSdt + \sigma S dZ \tag{3.10}$$

and simulate a large number of random paths. If we know the option payoff as a function of $S$ at $t = T$, then we compute

$$V(S, 0) = e^{-rT} E_Q(V(S, T)) \tag{3.11}$$

which should be the no-arbitrage value.

Note the $E_Q$ in the above equation. This makes it clear that we are taking the expectation in the *risk neutral* world (the expectation in the Q measure). This contrasts with the real-world expectation (the P measure).

Suppose we want to know the expected value (in the real world) of an asset which pays $V(S, t = T)$ at $t = T$ in the future. Then, the expected value (today) is given by solving

$$V_t + \frac{\sigma^2 S^2}{2} V_{SS} + \mu S V_S = 0 \; . \tag{3.12}$$

where we have dropped the discounting term. In particular, suppose we are going to receive $V = S(t = T)$, i.e. just the asset at $t = T$. Assume that the solution to equation (3.12) is $V = Const. \; A(t)S$, and we find that

$$V = Const. \; S e^{\mu(T-t)} \; . \tag{3.13}$$

Noting that we receive $V = S$ at $t = T$ means that

$$V = Se^{\mu(T-t)} . \tag{3.14}$$

Today, we can acquire the asset for price $S(t = 0)$. At $t = T$, the asset is worth $S(t = T)$. Equation (3.14) then says that

$$E[V(S(t=0), t=0)] = E[S(t=0)] = S(t=0)e^{\mu(T)} \tag{3.15}$$

In other words, if

$$dS = S\mu \; dt + S\sigma \; dZ \tag{3.16}$$

then (setting $t = T$)

$$E[S] = Se^{\mu t} . \tag{3.17}$$

Recall that the exact solution to equation (3.16) is (equation (2.58))

$$S(t) = S(0) \exp[(\mu - \frac{\sigma^2}{2})t + \sigma(Z(t) - Z(0))] . \tag{3.18}$$

So that we have just shown that $E[S] = Se^{\mu t}$ by using a simple PDE argument and Ito's Lemma. Isn't this easier than using brute force statistics? PDEs are much more elegant.

# 4 Monte Carlo Methods

This brings us to the simplest numerical method for computing the no-arbitrage value of an option. Suppose that we assume that the underlying process is

$$\frac{dS}{S} = rdt + \sigma dZ \tag{4.1}$$

then we can simulate a path forward in time, starting at some price today $S^0$, using a forward Euler timestepping method ($S^i = S(t_i)$)

$$S^{i+1} = S^i + S^i(r\Delta t + \sigma\phi^i\sqrt{\Delta t}) \tag{4.2}$$

where $\Delta t$ is the finite timestep, and $\phi^i$ is a random number which is $N(0, 1)$. Note that at each timestep, we generate a new random number. After $N$ steps, with $T = N\Delta t$, we have a single realized path. Given the payoff function of the option, the value for this path would be

$$Value = Payoff(S^N) . \tag{4.3}$$

For example, if the option was a European call, then

$$\begin{aligned} Value &= \max(S^N - K, 0) \\ K &= \text{ Strike Price} \end{aligned} \tag{4.4}$$

Suppose we run a series of trials, $m = 1, ..., M$, and denote the payoff after the $m'th$ trial as $payoff(m)$. Then, the no-arbitrage value of the option is

$$\begin{aligned} Option\ Value &= e^{-rT}E(payoff) \\ &\simeq e^{-rT}\frac{1}{M}\sum_{m=1}^{m=M} payoff(m) . \end{aligned} \tag{4.5}$$

Recall that these paths are not the real paths, but are the risk neutral paths.

Now, we should remember that we are

1. approximating the solution to the SDE by forward Euler, which has $O(\Delta t)$ truncation error.

2. approximating the expectation by the mean of many random paths. This Monte Carlo error is of size $O(1/\sqrt{M})$, which is slowly converging.

There are thus two sources of error in the Monte Carlo approach: timestepping error and sampling error.

The slow rate of convergence of Monte Carlo methods makes these techniques unattractive except when the option is written on several (i.e. more than three) underlying assets. As well, since we are simulating forward in time, we cannot know at a given point in the forward path if it is optimal to exercise or hold an American style option. This is easy if we use a PDE method, since we solve the PDE backwards in time, so we always know the *continuation value* and hence can act optimally. However, if we have more than three factors, PDE methods become very expensive computationally. As well, if we want to determine the effects of discrete hedging, for example, a Monte Carlo approach is very easy to implement.

The error in the Monte Carlo method is then

$$
\begin{aligned}
Error &= O\left(\max(\Delta t, \frac{1}{\sqrt{M}})\right) \\
&\quad \Delta t = \text{ timestep} \\
&\quad M = \text{ number of Monte Carlo paths}
\end{aligned}
\tag{4.6}
$$

Now, it doesn't make sense to drive the Monte Carlo error down to zero if there is $O(\Delta t)$ timestepping error. We should seek to balance the timestepping error and the sampling error. In order to make these two errors the same order, we should choose $M = O(\frac{1}{(\Delta t)^2})$. This makes the total error $O(\Delta t)$. We also have that

$$
\begin{aligned}
\text{Complexity} &= O\left(\frac{M}{\Delta t}\right) \\
&= O\left(\frac{1}{(\Delta t)^3}\right) \\
\Delta t &= O\left((\text{Complexity})^{-1/3}\right)
\end{aligned}
\tag{4.7}
$$

and hence

$$
\text{Error} = O\left(\frac{1}{(\text{ Complexity})^{1/3}}\right) . \tag{4.8}
$$

In practice, the convergence in terms of timestep error is often not done. People just pick a timestep, i.e. one day, and increase the number of Monte Carlo samples until they achieve convergence in terms of sampling error, and ignore the timestep error. Sometimes this gives bad results!

Note that the exact solution to Geometric Brownian motion (2.58) has the property that the asset value $S$ can never reach $S = 0$ if $S(0) > 0$, in any finite time. However, due to the approximate nature of our Forward Euler method for solving the SDE, it is possible that a negative or zero $S^i$ can show up. We can do one of three things here, in this case

- Cut back the timestep at this point in the simulation so that $S$ is positive.
- Set $S = 0$ and continue. In this case, $S$ remains zero for the rest of this particular simulation.
- Use Ito's Lemma, and determine the SDE for $\log S$, i.e. if $F = \log S$, then, from equation (2.56), we obtain (with $\mu = r$)

$$
dF = (r - \frac{\sigma^2}{2})dt + \sigma dZ , \tag{4.9}
$$

  so that now, if $F < 0$, there is no problem, since $S = e^F$, and if $F < 0$, this just means that $S$ is very small. We can use this idea for any stochastic process where the variable should not go negative.

Usually, most people set $S = 0$ and continue. As long as the timestep is not too large, this situation is probably due to an event of low probability, hence any errors incurred will not affect the expected value very much. If negative $S$ values show up many times, this is a signal that the timestep is too large.

In the case of simple Geometric Brownian motion, where $r, \sigma$ are constants, then the SDE can be solved exactly, and we can avoid timestepping errors (see Section 2.6.2). In this case

$$S(T) = S(0) \exp[(r - \frac{\sigma^2}{2})T + \sigma\phi\sqrt{T}] \tag{4.10}$$

where $\phi \sim N(0, 1)$. I'll remind you that equation (4.10) is exact. For these simple cases, we should always use equation (4.10). Unfortunately, this does not work in more realistic situations.

Monte Carlo is popular because

- It is simple to code. Easily handles complex path dependence.
- Easily handles multiple assets.

The disadvantages of Monte Carlo methods are

- It is difficult to apply this idea to problems involving optimal decision making (e.g. American options).
- It is hard to compute the Greeks ($V_S, V_{SS}$), which are the hedging parameters, very accurately.
- MC converges slowly.

## 4.1 Monte Carlo Error Estimators

The sampling error can be estimated via a statistical approach. If the estimated mean of the sample is

$$\hat{\mu} = \frac{e^{-rT}}{M} \sum_{m=1}^{m=M} payoff(m) \tag{4.11}$$

and the standard deviation of the estimate is

$$\omega = \left( \frac{1}{M-1} \sum_{m=1}^{m=M} (e^{-rT} payoff(m) - \hat{\mu})^2 \right)^{1/2} \tag{4.12}$$

then the 95% confidence interval for the actual value $V$ of the option is

$$\hat{\mu} - \frac{1.96\omega}{\sqrt{M}} < V < \hat{\mu} + \frac{1.96\omega}{\sqrt{M}} \tag{4.13}$$

Note that in order to reduce this error by a factor of 10, the number of simulations must be increased by 100.

The timestep error can be estimated by running the problem with different size timesteps, comparing the solutions.

## 4.2 Random Numbers and Monte Carlo

There are many good algorithms for generating random sequences which are uniformly distributed in $[0, 1]$. See for example, (*Numerical Recipes in C++.*, Press et al, Cambridge University Press, 2002). As pointed out in this book, often the system supplied random number generators, such as *rand* in the standard C library, and the infamous *RANDU* IBM function, are extremely bad. The Matlab functions appear to be quite good. For more details, please look at (*Park and Miller*, ACM Transactions on Mathematical Software, 31 (1988) 1192-1201). Another good generator is described in (Matsumoto and Nishimura, *"The Mersenne*

*Twister: a 623 dimensionally equidistributed uniform pseudorandom number generator,"* ACM Transactions on Modelling and Computer Simulation, 8 (1998) 3-30.) Code can be downloaded from the authors Web site.

However, we need random numbers which are normally distributed on $[-\infty, +\infty]$, with mean zero and variance one ($N(0,1)$).

Suppose we have uniformly distributed numbers on $[0,1]$, i.e. the probability of obtaining a number between $x$ and $x + dx$ is

$$\begin{aligned} p(x)dx &= dx \ ; \quad 0 \leq x \leq 1 \\ &= 0 \ ; \quad \text{otherwise} \end{aligned} \tag{4.14}$$

Let's take a function of this random variable $y(x)$. How is $y(x)$ distributed? Let $\hat{p}(y)$ be the probability distribution of obtaining $y$ in $[y, y + dy]$. Consequently, we must have (recall the law of transformation of probabilities)

$$p(x)|dx| = \hat{p}(y)|dy|$$

or

$$\hat{p}(y) = p(x)\left|\frac{dx}{dy}\right| \ . \tag{4.15}$$

Suppose we want $\hat{p}(y)$ to be normal,

$$\hat{p}(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \ . \tag{4.16}$$

If we start with a uniform distribution, $p(x) = 1$ on $[0,1]$, then from equations (4.15-4.16) we obtain

$$\frac{dx}{dy} = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \ . \tag{4.17}$$

Now, for $x \in [0,1]$, we have that the probability of obtaining a number in $[0,x]$ is

$$\int_0^x dx' = x \ , \tag{4.18}$$

but from equation (4.17) we have

$$dx' = \frac{e^{-(y')^2/2}}{\sqrt{2\pi}} dy' \ . \tag{4.19}$$

So, there exists a $y$ such that the probability of getting a $y'$ in $[-\infty, y]$ is equal to the probability of getting $x'$ in $[0,x]$,

$$\int_0^x dx' = \int_{-\infty}^{y} \frac{e^{-(y')^2/2}}{\sqrt{2\pi}} dy' \ , \tag{4.20}$$

or

$$x = \int_{-\infty}^{y} \frac{e^{-(y')^2/2}}{\sqrt{2\pi}} dy' \ . \tag{4.21}$$

So, if we generate uniformly distributed numbers $x$ on $[0,1]$, then to determine $y$ which are $N(0,1)$, we do the following

- Generate $x$
- Find $y$ such that

$$x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(y')^2/2} dy' \quad . \tag{4.22}$$

We can write this last step as

$$y = F(x) \tag{4.23}$$

where $F(x)$ is the inverse cumulative normal distribution.

## 4.3 The Box-Muller Algorithm

Starting from random numbers which are uniformly distributed on $[0, 1]$, there is actually a simpler method for obtaining random numbers which are normally distributed.

If $p(x)$ is the probability of finding $x \in [x, x + dx]$ and if $y = y(x)$, and $\hat{p}(y)$ is the probability of finding $y \in [y, y + dy]$, then, from equation (4.15) we have

$$|p(x)dx| = |\hat{p}(y)dy| \tag{4.24}$$

or

$$\hat{p}(y) = p(x) \left| \frac{dx}{dy} \right| \quad . \tag{4.25}$$

Now, suppose we have two original random variables $x_1, x_2$, and let $p(x_i, x_2)$ be the probability of obtaining $(x_1, x_2)$ in $[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2]$. Then, if

$$\begin{aligned} y_1 &= y_1(x_1, x_2) \\ y_2 &= y_2(x_1, x_2) \end{aligned} \tag{4.26}$$

and we have that

$$\hat{p}(y_1, y_2) = p(x_1, x_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \tag{4.27}$$

where the Jacobian of the transformation is defined as

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = det \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \tag{4.28}$$

Recall that the Jacobian of the transformation can be regarded as the scaling factor which transforms $dx_1 \ dx_2$ to $dy_1 \ dy_2$, i.e.

$$dx_1 \ dx_2 = \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy_1 \ dy_2 \ . \tag{4.29}$$

Now, suppose that we have $x_1, x_2$ uniformly distributed on $[0, 1] \times [0, 1]$, i.e.

$$p(x_1, x_2) = U(x_1)U(x_2) \tag{4.30}$$

where

$$\begin{aligned} U(x) &= 1 \quad ; \quad 0 \leq x \leq 1 \\ &= 0 \quad ; \quad \text{otherwise} \ . \end{aligned} \tag{4.31}$$

We denote this distribution as $x_1 \sim U[0,1]$ and $x_2 \sim U[0,1]$.

If $p(x_1, x_2)$ is given by equation (4.30), then we have from equation (4.27) that

$$\hat{p}(y_1, y_2) = \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \tag{4.32}$$

Now, we want to find a transformation $y_1 = y_1(x_1, x_2), y_2 = y_2(x_1, x_2)$ which results in normal distributions for $y_1, y_2$. Consider

$$\begin{aligned} y_1 &= \sqrt{-2 \log x_1} \cos 2\pi x_2 \\ y_2 &= \sqrt{-2 \log x_1} \sin 2\pi x_2 \end{aligned} \tag{4.33}$$

or solving for $(x_2, x_2)$

$$\begin{aligned} x_1 &= \exp\left( \frac{-1}{2}(y_1^2 + y_2^2) \right) \\ x_2 &= \frac{1}{2\pi} \tan^{-1} \left[ \frac{y_2}{y_1} \right] . \end{aligned} \tag{4.34}$$

After some tedious algebra, we can see that (using equation (4.34))

$$\left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-y_2^2/2} \tag{4.35}$$

Now, assuming that equation (4.30) holds, then from equations (4.32-4.35) we have

$$\hat{p}(y_1, y_2) = \frac{1}{\sqrt{2\pi}} e^{-y_1^2/2} \frac{1}{\sqrt{2\pi}} e^{-y_2^2/2} \tag{4.36}$$

so that $(y_1, y_2)$ are independent, normally distributed random variables, with mean zero and variance one, or

$$y_1 \sim N(0,1) \quad ; \quad y_2 \sim N(0,1) . \tag{4.37}$$

This gives the following algorithm for generating normally distributed random numbers (given uniformly distributed numbers):

**Box Muller Algorithm**

Repeat
  Generate $u_1 \sim U(0,1)$, $u_2 \sim U(0,1)$
  $\theta = 2\pi u_2$, $\rho = \sqrt{-2 \log u_1}$
  $z_1 = \rho \cos \theta$; $z_2 = \rho \sin \theta$
End Repeat

(4.38)

This has the effect that $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$.

Note that we generate two draws from a normal distribution on each pass through the loop.

### 4.3.1 An improved Box Muller

The algorithm (4.38) can be expensive due to the trigonometric function evaluations. We can use the following method to avoid these evaluations. Let

$$\begin{aligned} U_1 \sim U[0,1] \quad &; \quad U_2 \sim U[0,1] \\ V_1 = 2U_1 - 1 \quad &; \quad V_2 = 2U_2 - 1 \end{aligned} \tag{4.39}$$

which means that $(V_1, V_2)$ are uniformly distributed in $[-1, 1] \times [-1, 1]$. Now, we carry out the following procedure

**Rejection Method**

```
Repeat
    If ( V_1^2 + V_2^2 < 1 )
        Accept
    Else
        Reject
    Endif
End Repeat
```

(4.40)

which means that if we define $(V_1, V_2)$ as in equation (4.39), and then process the pairs $(V_1, V_2)$ using algorithm (4.40) we have that $(V_1, V_2)$ are uniformly distributed on the disk centered at the origin, with radius one, in the $(V_1, V_2)$ plane. This is denoted by

$$(V_1, V_2) \quad \sim \quad D(0, 1) \ . \tag{4.41}$$

If $(V_1, V_2) \sim D(0, 1)$ and $R^2 = V_1^2 + V_2^2$, then the probability of finding $R$ in $[R, R + dR]$ is

$$\begin{aligned} p(R)\ dR &= \frac{2\pi R\ dR}{\pi(1)^2} \\ &= 2R\ dR \ . \end{aligned} \tag{4.42}$$

From the fundamental law of transformation of probabilities, we have that

$$\begin{aligned} p(R^2)d(R^2) &= p(R)dR \\ &= 2R\ dR \end{aligned} \tag{4.43}$$

so that

$$\begin{aligned} p(R^2) &= \frac{2R}{\frac{d(R^2)}{dR}} \\ &= 1 \end{aligned} \tag{4.44}$$

so that $R^2$ is uniformly distributed on $[0, 1]$, $(R^2 \sim U[0, 1])$.

As well, if $\theta = \tan^{-1}(V_2/V_1)$, i.e. $\theta$ is the angle between a line from the origin to the point $(V_1, V_2)$ and the $V_1$ axis, then $\theta \sim U[0, 2\pi]$. Note that

$$\begin{aligned} \cos\theta &= \frac{V_1}{\sqrt{V_1^2 + V_2^2}} \\ \sin\theta &= \frac{V_2}{\sqrt{V_1^2 + V_2^2}} \ . \end{aligned} \tag{4.45}$$

Now in the original Box Muller algorithm (4.38),

$$\begin{aligned} \rho = \sqrt{-2\log U_1} \quad &; \quad U_1 \sim U[0, 1] \\ \theta = 2\Pi U_2 \quad &; \quad U_2 \sim U[0, 1] \ , \end{aligned} \tag{4.46}$$

but $\theta = \tan^{-1}(V_2/V_1) \sim U[0, 2\pi]$, and $R^2 = U[0,1]$. Therefore, if we let $W = R^2$, then we can replace $\theta, \rho$ in algorithm (4.38) by

$$\begin{aligned} \theta &= \tan^{-1}\left(\frac{V_2}{V_1}\right) \\ \rho &= \sqrt{-2\log W} \ . \end{aligned} \tag{4.47}$$

Now, the last step in the Box Muller algorithm (4.38) is

$$\begin{aligned} Z_1 &= \rho\cos\theta \\ Z_2 &= \rho\sin\theta \ , \end{aligned} \tag{4.48}$$

but since $W = R^2 = V_1^2 + V_2^2$, then $\cos\theta = V_1/R$, $\sin\theta = V_2/R$, so that

$$\begin{aligned} Z_1 &= \rho\frac{V_1}{\sqrt{W}} \\ Z_2 &= \rho\frac{V_2}{\sqrt{W}} \ . \end{aligned} \tag{4.49}$$

This leads to the following algorithm

**Polar form of Box Muller**

Repeat

Generate $U_1 \sim U[0,1]$, $U_2 \sim U[0,1]$.
Let

$$\begin{aligned} V_1 &= 2U_1 - 1 \\ V_2 &= 2U_2 - 1 \\ W &= V_1^2 + V_2^2 \end{aligned}$$

If( $W < 1$) then

$$\begin{aligned} Z_1 &= V_1\sqrt{-2\log W/W} \\ Z_2 &= V_2\sqrt{-2\log W/W} \end{aligned} \tag{4.50}$$

End If

End Repeat

Consequently, $(Z_1, Z_2)$ are independent (uncorrelated), and $Z_1 \sim N(0,1)$, and $Z_2 \sim N(0,1)$. Because of the rejection step (4.40), about $(1 - \pi/4)$ of the random draws in $[-1,+1] \times [-1,+1]$ are rejected (about 21%), but this method is still generally more efficient than brute force Box Muller.

## 4.4 Speeding up Monte Carlo

Monte Carlo methods are slow to converge, since the error is given by

$$\text{Error} = O(\frac{1}{\sqrt{M}})$$

where $M$ is the number of samples. There are many methods which can be used to try to speed up convergence. These are usually termed Variance Reduction techniques.

Perhaps the simplest idea is the Antithetic Variable method. Suppose we compute a random asset path

$$S^{i+1} = S^i \mu \Delta t + S^i \sigma \phi^i \sqrt{\Delta t}$$

where $\phi^i$ are $N(0,1)$. We store all the $\phi^i, i = 1, ...,$ for a given path. Call the estimate for the option price from this sample path $V^+$. Then compute a second sample path where $(\phi^i)' = -\phi^i, i = 1, ...,$. Call this estimate $V^-$. Then compute the average

$$\bar{V} = \frac{V^+ + V^-}{2} ,$$

and continue sampling in this way. Averaging over all the $\bar{V}$, slightly faster convergence is obtained. Intuitively, we can see that this *symmetrizes* the random paths.

Let $X^+$ be the option values obtained from all the $V^+$ simulations, and $X^-$ be the estimates obtained from all the $V^-$ simulations. Note that $Var(X^+) = Var(X^-)$ (they have the same distribution). Then

$$\begin{aligned} Var(\frac{X^+ + X^-}{2}) &= \frac{1}{4}Var(X^+) + \frac{1}{4}Var(X^-) + \frac{1}{2}Cov(X^+, X^-) \\ &= \frac{1}{2}Var(X^+) + \frac{1}{2}Cov(X^+, X^-) \end{aligned} \tag{4.51}$$

which will be smaller than $Var(X^+)$ if $Cov(X^+, X^-)$ is nonpositive. Warning: this is not always the case. For example, if the payoff is not a monotonic function of $S$, the results may actually be worse than crude Monte Carlo. For example, if the payoff is a capped call

$$\begin{aligned} \text{payoff} &= \min(K_2, \max(S - K_1, 0)) \\ & K_2 > K_1 \end{aligned}$$

then the antithetic method performs poorly.

Note that this method can be used to estimate the mean. In the MC error estimator (4.13), compute the standard deviation of the estimator as $\omega = \sqrt{Var(\frac{X^+ + X^-}{2})}$.

However, if we want to estimate the distribution of option prices (i.e. a probability distribution), then we should not average each $V^+$ and $V^-$, since this changes the variance of the actual distribution.

If we want to know the actual variance of the distribution (and not just the mean), then to compute the variance of the distribution, we should just use the estimates $V^+$, and compute the estimate of the variance in the usual way. This should also be used if we want to plot a histogram of the distribution, or compute the Value at Risk.

## 4.5 Estimating the mean and variance

An estimate of the mean $\bar{x}$ and variance $s_M^2$ of $M$ numbers $x_1, x_2, ..., x_M$ is

$$\begin{aligned} s_M^2 &= \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \bar{x})^2 \\ \bar{x} &= \frac{1}{M}\sum_{i=1}^{M} x_i \end{aligned} \tag{4.52}$$

Alternatively, one can use

$$s_M^2 = \frac{1}{M-1}\left(\sum_{i=1}^{M} x_i^2 - \frac{1}{M}\left(\sum_{i=1}^{M} x_i\right)^2\right) \tag{4.53}$$

which has the advantage that the estimate of the mean and standard deviation can be computed in one loop.

Something else to be aware of, is that doing a simple sum of many numbers is prone to accumulation of roundoff errors. A simple remedy is to do pairwise accumulation of the sum. See, for example, *Accurate floating point summation*, P. Linz, Comm. ACM, vol 13 (1970) 361-362, or *The accuracy of floating point summation,*, N. Higham, SIAM J. Sci. Computing, vol 14 (1993) 783-799.

In order to avoid roundoff, the following method is suggested by Seydel (R. Seydel, Tools for Computational Finance, Springer, 2002). Set

$$\alpha_1 = x_1 \quad ; \quad \beta_1 = 0 \tag{4.54}$$

then compute recursively

$$\begin{aligned} \alpha_i &= \alpha_{i-1} + \frac{x_i - \alpha_{i-1}}{i} \\ \beta_i &= \beta_{i-1} + \frac{(i-1)(x_i - \alpha_{i-1})^2}{i} \end{aligned} \tag{4.55}$$

so that

$$\begin{aligned} \bar{x} &= \alpha_M \\ s_M^2 &= \frac{\beta_M}{M-1} \end{aligned} \tag{4.56}$$

## 4.6 Low Discrepancy Sequences

In a effort to get around the $\frac{1}{\sqrt{M}}$, ($M =$ number of samples) behaviour of Monte Carlo methods, quasi-Monte Carlo methods have been devised.

These techniques use a deterministic sequence of numbers (low discrepancy sequences). The idea here is that a Monte Carlo method does not fill the sample space very evenly (after all, its random). A low discrepancy sequence tends to sample the space in a orderly fashion. If $d$ is the dimension of the space, then the worst case error bound for an LDS method is

$$\text{Error} \quad = O\left( \frac{(\log M)^d}{M} \right) \tag{4.57}$$

where $M$ is the number of samples used. Clearly, if $d$ is small, then this error bound is (at least asymptotically) better than Monte Carlo.

LDS methods generate numbers on $[0, 1]$. We cannot use the Box-Muller method in this case to produce normally distributed numbers, since these numbers are deterministic. We have to invert the cumulative normal distribution in order to get the numbers distributed with mean zero and standard deviation one on $[-\infty, +\infty]$. So, if $F(x)$ is the inverse cumulative normal distribution, then

$$\begin{aligned} x_{LDS} &= \text{uniformly distributed on } [0,1] \\ y_{LDS} &= F(x_{LDS}) \text{ is } N(0,1)\ . \end{aligned} \tag{4.58}$$

Another problem has to do with the fact that if we are stepping through time, i.e.

$$\begin{aligned} S^{n+1} &= S^n + S^n(r\Delta t + \phi\sigma\sqrt{\Delta t}) \\ & \phi = N(0,1) \end{aligned} \tag{4.59}$$

with, say, $N$ steps in total, then we need to think of this as a problem in $N$ dimensional space. In other words, the $k-th$ timestep is sampled from the $k-th$ coordinate in this $N$ dimensional space. We are trying to uniformly sample from this $N$ dimensional space.

Let $\hat{x}$ be a vector of LDS numbers on $[0,1]$, in $N$ dimensional space

$$\hat{x} = \begin{bmatrix} x_1 \\ x_2 \\ | \\ x_N \end{bmatrix} . \tag{4.60}$$

So, an LDS algorithm would proceed as follows, for the $j'th$ trial

- Generate $\hat{x}^j$ (the $j'th$ LDS number in an $N$ dimensional space).
- Generate the normally distributed vector $\hat{y}^j$ by inverting the cumulative normal distribution for each component

$$\hat{y}^j = \begin{bmatrix} F(x_1^j) \\ F(x_2^j) \\ | \\ F(x_N^j) \end{bmatrix} \tag{4.61}$$

- Generate a complete sample path $k = 0, ..., N-1$

$$S_j^{k+1} = S_j^k + S_j^k(r\Delta t + \hat{y}_{k+1}^j \sigma \sqrt{\Delta t}) \tag{4.62}$$

- Compute the payoff at $S = S_j^N$

The option value is the average of these trials.

There are a variety of LDS numbers: Halton, Sobol, Niederrieter, etc. Our tests seem to indicate that Sobol is the best.

Note that the worst case error bound for the error is given by equation (4.57). If we use a reasonable number of timesteps, say $50-100$, then, $d = 50-100$, which gives a very bad error bound. For $d$ large, the numerator in equation (4.57) dominates. The denominator only dominates when

$$M \simeq e^d \tag{4.63}$$

which is a very large number of trials for $d \simeq 100$. Fortunately, at least for path-dependent options, we have found that things are not quite this bad, and LDS seems to work if the number of timesteps is less than $100-200$. However, once the dimensionality gets above a few hundred, convergence seems to slow down.

## 4.7 Correlated Random Numbers

In many cases involving multiple assets, we would like to generate correlated, normally distributed random numbers. Suppose we have $i = 1, ..., d$ assets, and each asset follows the simulated path

$$S_i^{n+1} = S_i^n + S_i^n(r\Delta t + \phi_i^n \sigma_i \sqrt{\Delta t}) \tag{4.64}$$

where $\phi_i^n$ is $N(0,1)$ and

$$E(\phi_i^n \phi_j^n) = \rho_{ij} \tag{4.65}$$

where $\rho_{ij}$ is the correlation between asset $i$ and asset $j$.

Now, it is easy to generate a set of $d$ uncorrelated $N(0,1)$ variables. Call these $\epsilon_1, ..., \epsilon_d$. So, how do we produce correlated numbers? Let

$$[\Psi]_{ij} = \rho_{ij} \tag{4.66}$$

be the matrix of correlation coefficients. Assume that this matrix is SPD (if not, one of the random variables is a linear combination of the others, hence this is a degenerate case). Assuming $\Psi$ is SPD, we can Cholesky factor $\Psi = LL^t$, so that

$$\rho_{ij} = \sum_k L_{ik} L^t_{kj} \tag{4.67}$$

Let $\bar{\phi}$ be the vector of correlated normally distributed random numbers (i.e. what we want to get), and let $\bar{\epsilon}$ be the vector of uncorrelated $N(0,1)$ numbers (i.e. what we are given).

$$\bar{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ | \\ \phi_d \end{bmatrix} \; ; \; \bar{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ | \\ \epsilon_d \end{bmatrix} \tag{4.68}$$

So, given $\bar{\epsilon}$, we have

$$E(\epsilon_i \epsilon_j) = \delta_{ij}$$

where

$$\begin{aligned} \delta_{ij} &= 0 \; ; \; \text{ if } i \neq j \\ &= 1 \; ; \; \text{ if } i = j \; . \end{aligned}$$

since the $\epsilon_i$ are uncorrelated. Now, let

$$\phi_i = \sum_j L_{ij} \epsilon_j \tag{4.69}$$

which gives

$$\begin{aligned} \phi_i \phi_k &= \sum_j \sum_l L_{ij} L_{kl} \epsilon_l \epsilon_j \\ &= \sum_j \sum_l L_{ij} \epsilon_l \epsilon_j L^t_{lk} \; . \end{aligned} \tag{4.70}$$

Now,

$$\begin{aligned} E(\phi_i \phi_k) &= E\left[ \sum_j \sum_l L_{ij} \epsilon_l \epsilon_j L^t_{lk} \right] \\ &= \sum_j \sum_l L_{ij} E(\epsilon_l \epsilon_j) L^t_{lk} \\ &= \sum_j \sum_l L_{ij} \delta_{lj} L^t_{lk} \\ &= \sum_l L_{il} L^t_{lk} \\ &= \rho_{ij} \end{aligned} \tag{4.71}$$

It is easy to show that

$$\begin{aligned} E(\phi_i) &= 0 \\ E(\phi_i^2) &= \rho_{ii} = 1 \; . \end{aligned} \tag{4.72}$$

So, in order to generate correlated $N(0,1)$ numbers:

- Factor the correlation matrix $\Psi = LL^t$
- Generate uncorrelated $N(0,1)$ numbers $\epsilon_i$
- Correlated numbers $\phi_i$ are given from

$$\bar{\phi} \quad = \quad L\bar{\epsilon}$$

## 4.8 Integration of Stochastic Differential Equations

Up to now, we have been fairly slack about defining what we mean by convergence when we use forward Euler timestepping (4.2) to integrate

$$dS = \mu S \; dt + \sigma S \; dZ \; . \tag{4.73}$$

The forward Euler algorithm is simply

$$S^{i+1} \quad = \quad S^i + S^i(\mu h + \phi^i \sqrt{h}) \tag{4.74}$$

where $h = \Delta t$ is the finite timestep. For a good overview of these methods, check out (*"An algorithmic introduction to numerical simulation of stochastic differential equations,"* by D. Higham, SIAM Review vol. 43 (2001) 525-546). This article also has some good tips on solving SDEs using Matlab, in particular, taking full advantage of the vectorization of Matlab. Note that eliminating as many *for* loops as possible (i.e. computing all the MC realizations for each timestep in a vector) can reduce computation time by orders of magnitude.

Before we start defining what we mean by convergence, let's consider the following situation. Recall that

$$dZ = \phi\sqrt{dt} \tag{4.75}$$

where $\phi$ is a random draw from a normal distribution with mean zero and variance one. Let's imagine generating a set of $Z$ values at discrete times $t_i$, e.g. $Z(t_i) = Z_i$, by

$$Z_{i+1} \quad = \quad Z_i + \phi\sqrt{\Delta t} \; . \tag{4.76}$$

Now, these are all legitimate points along a Brownian motion path, since there is no timestepping error here, in view of equation (2.54). So, this set of values $\{Z_0, Z_1, ...,\}$ are valid points along a Brownian path. Now, recall that the exact solution (for a given Brownian path) of equation (4.73) is given by equation (2.58)

$$S(T) = S(0) \exp[(\mu - \frac{\sigma^2}{2})t + \sigma(Z(T) - Z(0))] \tag{4.77}$$

where $T$ is the stopping time of the simulation.

Now if we integrate equation (4.73) using forward Euler, with the discrete timesteps $\Delta t = t_{i+1} - t_i$, using the realization of the Brownian path $\{Z_0, Z_1, ...,\}$, we will not get the exact solution (4.77). This is because even though the Brownian path points are exact, time discretization errors are introduced in equation (4.74). So, how can we systematically study convergence of algorithm (4.74)? We can simply take smaller timesteps. However, we want to do this by *filling in* new $Z$ values in the Brownian path, while keeping the old values (since these are perfectly legitimate values). Let $S(T)^h$ represent the forward Euler solution (4.74) for a fixed timestep $h$. Let $S(T)$ be the exact solution (4.77). As $h \to 0$, we would expect $S(T)^h \to S(T)$, for a given path.

### 4.8.1 The Brownian Bridge

So, given a set of valid $Z_k$, how do we refine this path, keeping the existing points along this path? In particular, suppose we have two points $Z_i, Z_k$, at $(t_i, t_k)$, and we would like to generate a point $Z_j$ at $t_j$, with $t_i < t_j < t_k$. How should we pick $Z_j$? What density function should we use when generating $Z_j$, given that $Z_k$ is known?

Let $x, y$ be two draws from a normal distribution with mean zero and variance one. Suppose we have the point $Z(t_i) = Z_i$ and we generate $Z(t_j) = Z_j$, $Z(t_k) = Z_k$ along the Wiener path,

$$Z_j = Z_i + x\sqrt{t_j - t_i} \tag{4.78}$$

$$Z_k = Z_j + y\sqrt{t_k - t_j} \tag{4.79}$$

$$Z_k = Z_i + x\sqrt{t_j - t_i} + y\sqrt{t_k - t_j} \; . \tag{4.80}$$

So, given $(x, y)$, and $Z_i$, we can generate $Z_j, Z_k$. Suppose on the other hand, we have $Z_i$, and we generate $Z_k$ directly using

$$Z_k = Z_i + z\sqrt{t_k - t_i} \; , \tag{4.81}$$

where $z$ is $N(0, 1)$. Then how do we generate $Z_j$ using equation (4.78)? Since we are now specifying that we know $Z_k$, this means that our method for generating $Z_j$ is constrained. For example, given $z$, we must have that, from equations (4.80) and (4.81)

$$y = \frac{z\sqrt{t_k - t_i} - x\sqrt{t_j - t_i}}{\sqrt{t_k - t_j}} \; . \tag{4.82}$$

Now the probability density of drawing the pair $(x, y)$ given $z$, denoted by $p(x, y|z)$ is

$$p(x, y|z) = \frac{p(x)p(y)}{p(z)} \tag{4.83}$$

where $p(..)$ is a standard normal distribution, and we have used the fact that successive increments of a Brownian process are uncorrelated.

From equation (4.82), we can write $y = y(x, z)$, so that $p(x, y|z) = p(x, y(x, z)|z)$

$$\begin{aligned} p(x, y(x, z)|z) &= \frac{p(x)p(y(x, z))}{p(z)} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x^2 + y^2 - z^2)\right] \end{aligned} \tag{4.84}$$

or (after some algebra, using equation (4.82))

$$\begin{aligned} p(x|z) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - \alpha z)^2/\beta^2\right] \\ \alpha &= \sqrt{\frac{t_j - t_i}{t_k - t_i}} \\ \beta &= \sqrt{\frac{t_k - t_j}{t_k - t_i}} \end{aligned} \tag{4.85}$$

so that $x$ is normally distributed with mean $\alpha z$ and variance $\beta^2$. Since

$$z = \frac{Z_k - Z_i}{\sqrt{t_k - t_i}} \tag{4.86}$$

we have that $x$ has mean

$$E(x) = \frac{\sqrt{t_j - t_i}}{t_k - t_i}(Z_k - Z_i) \tag{4.87}$$

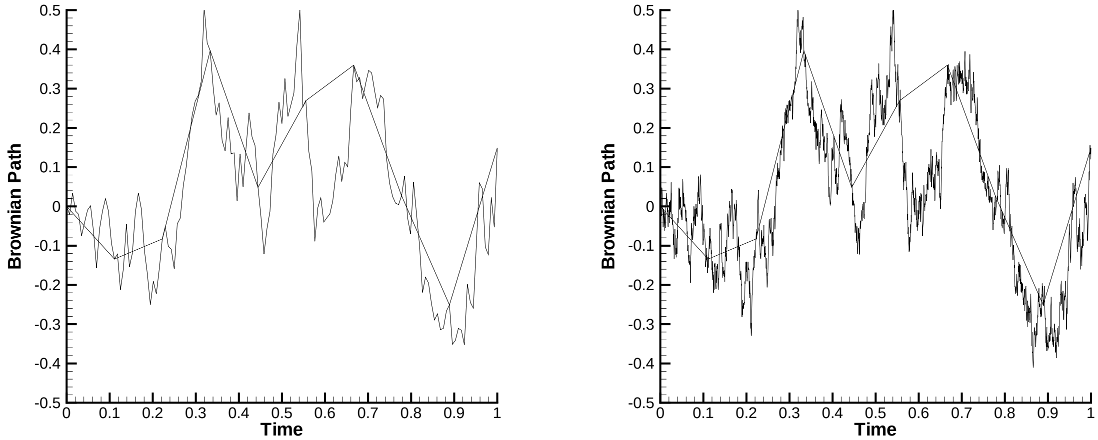



Figure 4.1: *Effect of adding more points to a Brownian path using a Brownian bridge. Note that the small timestep points match the coarse timestep points. Left: each coarse timestep is divided into 16 substeps. Right: each coarse timestep divided into 64 substeps.*

and variance

$$E[(x - E(x))^2] \quad = \quad \frac{t_k - t_j}{t_k - t_i} \tag{4.88}$$

Now, let

$$x \quad = \quad \frac{\sqrt{t_j - t_i}}{t_k - t_i}(Z_k - Z_i) + \phi\sqrt{\frac{t_k - t_j}{t_k - t_i}} \tag{4.89}$$

where $\phi$ is $N(0,1)$. Clearly, $x$ satisfies equations (4.87) and (4.89). Substituting equation (4.89) into (4.78) gives

$$Z_j \quad = \quad \left(\frac{t_k - t_j}{t_k - t_i}\right) Z_i + \left(\frac{t_j - t_i}{t_k - t_i}\right) Z_k + \phi\sqrt{\frac{(t_j - t_i)(t_k - t_j)}{(t_k - t_i)}} \tag{4.90}$$

where $\phi$ is $N(0,1)$. Equation (4.90) is known as the Brownian Bridge.

Figure 4.1 shows different Brownian paths constructed for different timestep sizes. An initial coarse path is constructed, then the fine timestep path is constructed from the coarse path using a Brownian Bridge. By construction, the final timestep path will pass through the coarse timestep nodes.

Figure 4.2 shows the asset paths integrated using the forward Euler algorithm (4.74) fed with the Brownian paths in Figure 4.1. In this case, note that the fine timestep path does not coincide with the coarse timestep nodes, due to the timestepping error.

### 4.8.2 Strong and Weak Convergence

Since we are dealing with a probabilistic situation here, it is not obvious how to define convergence. Given a number of points along a Brownian path, we could imagine refining this path (using a Brownian Bridge), and then seeing if the solution converged to exact solution. For the model SDE (4.73), we could ask that

$$E\left[|S(T) - S^h(T)|\right] \quad \leq \quad Const.\ h^{\gamma} \tag{4.91}$$

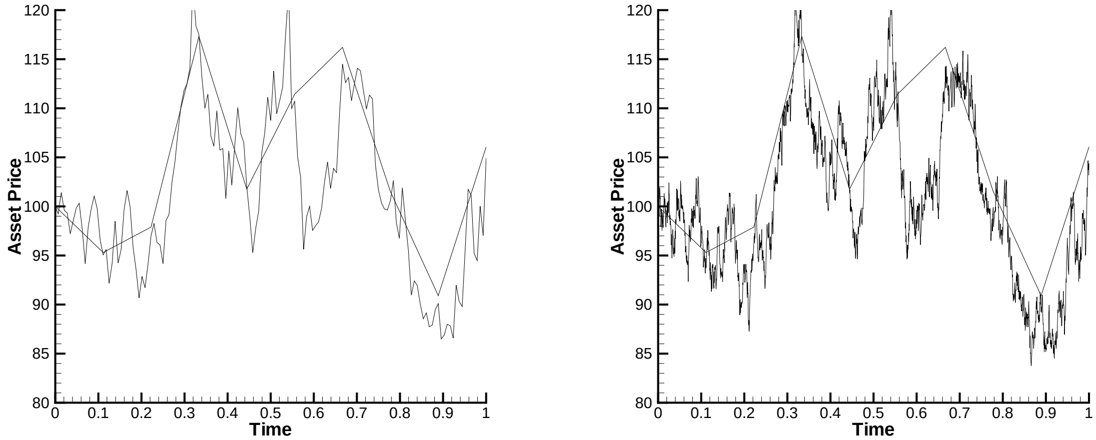



Figure 4.2: *Brownian paths shown in Figure 4.1 used to determine asset price paths using forward Euler timestepping (4.74). In this case, note that the asset paths for fine and coarse timestepping do not agree at the final time (due to the timestepping error). Eventually, for small enough timesteps, the final asset value will converge to the exact solution to the SDE. Left: each coarse timestep is divided into 16 substeps. Right: each coarse timestep divided into 64substeps.*

where the expectation in equation (4.91) is over many Brownian paths, and $h$ is the timestep size. Note that $S(T)$ is the exact solution along a particular Brownian path; the same path used to compute $S^h(T)$. Criterion (4.91) is called *strong convergence*. A less strict criterion is

$$|E\left[S(T)\right] - E\left[S^h(T)\right]| \quad \leq \quad Const. \ h^{\gamma} \tag{4.92}$$

It can be shown that using forward Euler results in weak convergence with $\gamma = 1$, and strong convergence with $\gamma = .5$.

Table 4.1 shows some test data used to integrate the SDE (4.73) using method (4.74). A series of Brownian paths was constructed, beginning with a coarse timestep path. These paths were systematically refined using the Brownian Bridge construction. Table 4.2 shows results where the strong and weak convergence errors are estimated as

$$\text{Strong Error} \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\left[|S(T)_i - S^h(T)_i|\right] \tag{4.93}$$

$$\text{Weak Error} \quad = \quad |\frac{1}{N}\sum_{i=1}^{N}\left[S(T)_i\right] - \frac{1}{N}\sum_{i=1}^{N}\left[S^h(T)_i\right]| \ , \tag{4.94}$$

where $S^h(T)_i$ is the solution obtained by forward Euler timestepping along the $i'th$ Brownian path, and $S(T)_i$ is the exact solution along this same path, and $N$ is the number of samples. Note that for equation (4.73), we have the exact solution

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[S(T)_i\right] \quad = \quad S_0 e^{\mu T} \tag{4.95}$$

but we do not replace the approximate sampled value of the limit in equation (4.94) by the theoretical limit

| | |
|---|---|
| $T$ | .25 |
| $\sigma$ | .4 |
| $\mu$ | .06 |
| $S_0$ | 100 |

Table 4.1: *Data used in the convergence tests.*

| Timesteps | Strong Error (4.91) | Weak Error (4.92) |
|---|---|---|
| 72 | .0269 | .00194 |
| 144 | .0190 | .00174 |
| 288 | .0135 | .00093 |
| 576 | .0095 | .00047 |

Table 4.2: *Convergence results, 100,000 samples used. Data in Table 4.1.*

(4.95). If we use enough Monte Carlo samples, we could replace the approximate expression

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} [S(T)_i]$$

by $S_0 e^{\mu T}$, but for normal parameters, the Monte Carlo sampling error is much larger than the timestepping error, so we would have to use an enormous number of Monte Carlo samples. Estimating the weak error using equation (4.94) will measure the timestepping error, as opposed to the Monte Carlo sampling error. However, for normal parameters, even using equation (4.94) requires a large number of Monte Carlo samples in order to ensure that the error is dominated by the timestepping error.

In Table 4.1, we can see that the ratio of the errors is about $\sqrt{2}$ for the strong error, and about two for the weak error. This is consistent with a convergence rate of $\gamma = .5$ for strong convergence, and $\gamma = 1.0$ for weak convergence.

## 4.9 Matlab and Monte Carlo Simulation

A straightforward implementation of Monte Carlo timestepping for solving the SDE

$$dS = \mu S\, dt + \sigma S\, dZ \tag{4.96}$$

in Matlab is shown in Algorithm (4.97). This code runs very slowly.

**Slow.m**

```
   randn('state',100);
   T = 1.00;              % expiry time
   sigma = 0.25;          % volatility
   mu = .10;              % P measure drift
   S_init = 100;          % initial value
   N_sim = 10000;         %   number of simulations
   N = 100;               %   number of timesteps
   delt  = T/N;           %   timestep

   drift = mu*delt;
   sigma_sqrt_delt = sigma*sqrt(delt);
   S_new = zeros(N_sim,1);

for m=1:N_sim
 S = S_init;
 for i=1:N % timestep loop

    S = S + S*( drift + sigma_sqrt_delt*randn(1,1) );

    S = max(0.0, S );
           % check to make sure that S_new cannot be < 0

 end % timestep loop
 S_new(m,1) = S;
end % simulation loop

 n_bin = 200;
 hist(S_new, n_bin);

 stndrd_dev = std(S_new);
 disp(sprintf('standard deviation:  %.5g\n',stndrd_dev));

 mean_S = mean(S_new);
 disp(sprintf('mean:  %.5g\n',stndrd_dev));
```

(4.97)

Alternatively, we can use Matlab's vectorization capabilities to interchange the timestep loop and the simulation loop. The innermost simulation loop can be replaced by vector statements, as shown in Algorithm (4.98). This runs much faster.

**Fast.m**

```
 randn('state',100);
  %
  T = 1.00;              % expiry time
  sigma = 0.25;          % volatility
  mu = .10;              % P measure drift
  S_init = 100;          % initial value
  N_sim = 10000;         %   number of simulations
  N = 100;               %   number of timesteps
  delt  = T/N;           %   timestep

  drift = mu*delt;
  sigma_sqrt_delt = sigma*sqrt(delt);

  S_old = zeros(N_sim,1);
  S_new = zeros(N_sim,1);

  S_old(1:N_sim,1) = S_init;

for i=1:N % timestep loop

    % now, for each timestep, generate info for
    % all simulations

    S_new(:,1) = S_old(:,1) +...
         S_old(:,1).*( drift + sigma_sqrt_delt*randn(N_sim,1) );

    S_new(:,1) = max(0.0, S_new(:,1) );
          % check to make sure that S_new cannot be < 0

   S_old(:,1) = S_new(:,1);
     %
     %  end of generation of all data for all simulations
     %  for this timestep

end % timestep loop

n_bin = 200;
hist(S_new, n_bin);

stndrd_dev = std(S_new);
disp(sprintf('standard deviation:  %.5g\n',stndrd_dev));

mean_S = mean(S_new);
disp(sprintf('mean:  %.5g\n',stndrd_dev));
```

(4.98)

# 5 The Binomial Model: Overview

We have seen that a problem with the Monte Carlo method is that it is difficult to use for valuing American style options. Recall that the holder of an American option can exercise the option at any time and receive the payoff. In order to determine whether or not it is worthwhile to hold the option, we have to compare the value of continuing to hold the option (the continuation value) with the payoff. If the continuation value is greater than the payoff, then we hold; otherwise, we exercise.

At any point in time, the continuation value depends on what happens in the future. Clearly, if we simulate forward in time, as in the Monte Carlo approach, we don't know what happens in the future, and hence we don't know how to act optimally. This is actually a dynamic programming problem. These sorts of problems are usually solved by proceeding from the end point backwards. We use the same idea here. We have to start from the terminal time and work backwards.

## 5.1 A Binomial Model Based on the Risk Neutral Walk

Recall that we can determine the no-arbitrage value of an option by pretending we live in a risk-neutral world, where risky assets drift at $r$ and are discounted at $r$. If we let $X = \log S$, then the risk neutral process for $X$ is (from equation (2.56) )

$$dX = (r - \frac{\sigma^2}{2})dt + \sigma dZ \ . \tag{5.1}$$

Now, we can construct a discrete approximation to this random walk using the lattice discussed in Section 2.5. In fact, all we have to do is let let $\alpha = r - \frac{\sigma^2}{2}$, so that equation (5.1) is formally identical to equation (2.4). In order to ensure that in the limit as $\Delta t \to 0$, we get the process (5.1), we require that the sizes of the random jumps are $\Delta X = \sigma\sqrt{\Delta t}$ and that the probabilities of up ($p$) and down ($q$) moves are

$$\begin{aligned} p^r &= \frac{1}{2}[1 + \frac{\alpha}{\sigma}\sqrt{\Delta t}] \\ &= \frac{1}{2}[1 + \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\Delta t}] \\ q^r &= \frac{1}{2}[1 - \frac{\alpha}{\sigma}\sqrt{\Delta t}] \\ &= \frac{1}{2}[1 - \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\Delta t}] \ , \end{aligned} \tag{5.2}$$

where we have denoted the risk neutral probabilities by $p^r$ and $q^r$ to distinguish them from the real probabilities $p, q$.

Now, we will switch to a more common notation. If we are at node $j$, timestep $n$, we will denote this node location by $X_j^n$. Recall that $X = \log S$, so that in terms of asset price, this is $S_j^n = e^{X_j^n}$.

Now, consider that at node $(j, n)$, the asset can move up with probability $p^r$ and down with probability $q^r$. In other words

$$\begin{aligned} S_j^n &\to S_{j+1}^{n+1} \ ; \quad \text{with probability } p^r \\ S_j^n &\to S_j^{n+1} \ ; \quad \text{with probability } q^r \end{aligned} \tag{5.3}$$

Now, since in Section 2.5 we showed that $\Delta X = \sigma\sqrt{\Delta t}$, so that ($S = e^X$)

$$\begin{aligned} S_{j+1}^{n+1} &= S_j^n e^{\sigma\sqrt{\Delta t}} \\ S_j^{n+1} &= S_j^n e^{-\sigma\sqrt{\Delta t}} \end{aligned} \tag{5.4}$$

or

$$S_j^n = S_0^0 e^{(2j-n)\sigma\sqrt{\Delta t}} \ ; \quad j = 0, .., n \tag{5.5}$$

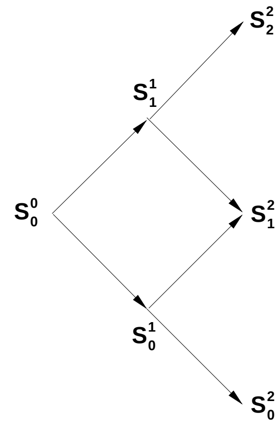



Figure 5.1: Lattice of stock price values

So, the first step in the process is to construct a tree of stock price values, as shown on Figure 5.1.

Associated with each stock price on the lattice is the option value $V_j^n$. We first set the value of the option at $T = N\Delta t$ to the payoff. For example, if we are valuing a put option, then

$$V_j^N = \max(K - S_j^N, 0) \; ; j = 0, ..., N \tag{5.6}$$

Then, we can use the risk neutral world idea to determine the no-arbitrage value of the option (it is the expected value in the risk neutral world). We can do this by working backward through the lattice. The value today is the discounted expected future value

**European Lattice Algorithm**

$$\begin{aligned} V_j^n &= e^{-r\Delta t}\left(p^r V_{j+1}^{n+1} + q^r V_j^{n+1}\right) \\ & n = N-1, ..., 0 \\ & j = 0, ..., n \end{aligned} \tag{5.7}$$

Rolling back through the tree, we obtain the value at $S_0^0$ today, which is $V_0^0$.

If the option is an American put, we can determine if it is optimal to hold or exercise, since we know the continuation value. In this case the rollback (5.7) becomes

**American Lattice Algorithm**

$$\begin{aligned} (V_j^n)^c &= e^{-r\Delta t}\left(p^r V_{j+1}^{n+1} + q^r V_j^{n+1}\right) \\ V_j^n &= \max\left((V_j^n)^c, \max(K - S_j^n, 0)\right) \\ & n = N-1, ..., 0 \\ & j = 0, ..., n \end{aligned} \tag{5.8}$$

which is illustrated in Figure 5.1.

The binomial lattice method has the following advantages

- It is very easy to code for simple cases.

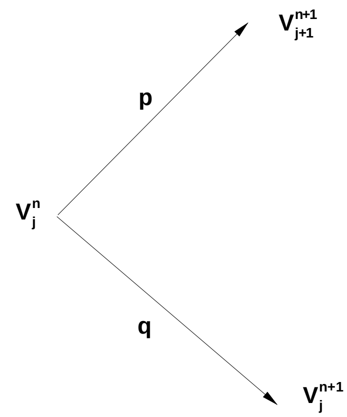



Figure 5.2: Backward recursion step.

- It is easy to explain to managers.
- American options are easy to handle.

However, the binomial lattice method has the following disadvantages

- Except for simple cases, coding becomes complex. For example, if we want to handle simple barrier options, things become nightmarish.
- This method is algebraically identical to an explicit finite difference solution of the Black-Scholes equation. Consequently, convergence is at an $O(\Delta t)$ rate.
- The *probabilities* $p^r, q^r$ are not real probabilities, they are simply the coefficients in a particular discretization of a PDE. Regarding them as probabilities leads to much fuzzy thinking, and complex wrong-headed arguments.

## 5.2 A No-arbitrage Lattice

We can also derive the lattice method directly from the discrete lattice model in Section 2.5. Suppose we assume that

$$dS = \mu S dt + \sigma S dZ \tag{5.9}$$

and letting $X = \log S$, we have that

$$dX = (\mu - \frac{\sigma^2}{2})dt + \sigma dZ \tag{5.10}$$

so that $\alpha = \mu - \frac{\sigma^2}{2}$ in equation (2.19). Now, let's consider the usual hedging portfolio at $t = n\Delta t, S = S_j^n$,

$$P_j^n = V_j^n - (\alpha^h) S_j^n \ , \tag{5.11}$$

where $V_j^n$ is the value of the option at $t = n\Delta t, S = S_j^n$. At $t = (n+1)\Delta t$,

$$
\begin{array}{rcll}
S_j^n & \rightarrow & S_{j+1}^{n+1} & ; \quad \text{with probability } p \\
S_j^n & \rightarrow & S_j^{n+1} & ; \quad \text{with probability } q \\
 & & S_{j+1}^{n+1} = S_j^n e^{\sigma\sqrt{\Delta t}} & \\
 & & S_j^{n+1} = S_j^n e^{-\sigma\sqrt{\Delta t}} &
\end{array}
$$

so that the value of the hedging portfolio at $t = n+1$ is

$$
\begin{aligned}
P_{j+1}^{n+1} &= V_{j+1}^{n+1} - (\alpha^h) S_{j+1}^{n+1} \quad ; \quad \text{with probability } p & (5.12)\\
P_j^{n+1} &= V_j^{n+1} - (\alpha^h) S_j^{n+1} \quad ; \quad \text{with probability } q\ . & (5.13)
\end{aligned}
$$

Now, as in Section 2.3, we can determine $(\alpha^h)$ so that the value of the hedging portfolio is independent of $p, q$. We do this by requiring that

$$
P_{j+1}^{n+1} = P_j^{n+1} \tag{5.14}
$$

so that

$$
V_{j+1}^{n+1} - (\alpha^h) S_{j+1}^{n+1} = V_j^{n+1} - (\alpha^h) S_j^{n+1}
$$

which gives

$$
(\alpha^h) = \frac{V_{j+1}^{n+1} - V_j^{n+1}}{S_{j+1}^{n+1} - S_j^{n+1}}\ . \tag{5.15}
$$

Since this portfolio is risk free, it must earn the risk free rate of return, so that

$$
\begin{aligned}
P_j^n &= e^{-r\Delta t} P_{j+1}^{n+1} \\
&= e^{-r\Delta t} P_j^{n+1}\ . 
\end{aligned} \tag{5.16}
$$

Now, substitute for $P_j^n$ from equation (5.11), with $P_{j+1}^{n+1}$ from equation (5.13), and $(\alpha^h)$ from equation (5.15) gives

$$
\begin{aligned}
V_j^n &= e^{-r\Delta t}\left(p^{r*} V_{j+1}^{n+1} + q^{r*} V_j^{n+1}\right) \\
& \quad p^{r*} = \frac{e^{r\Delta t} - e^{-\sigma\sqrt{\Delta t}}}{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}} \\
& \quad q^{r*} = 1 - p^{r*}\ .
\end{aligned} \tag{5.17}
$$

Note that $p^{r*}, q^{r*}$ do not depend on the real drift rate $\mu$, which is expected. If we expand $p^{r*}, q^{r*}$ in a Taylor Series, and compare with the $p^r, q^r$ in equations (5.2), we can show that

$$
\begin{aligned}
p^{r*} &= p^r + O((\Delta t)^{3/2}) \\
q^{r*} &= q^r + O((\Delta t)^{3/2})\ .
\end{aligned} \tag{5.18}
$$

After a bit more work, one can show that the value of the option at $t = 0$, $V_0^0$ using either $p^{r*}, q^{r*}$ or $p^r, q^r$ is the same to $O(\Delta t)$, which is not surprising, since these methods can both be regarded as an explicit finite difference approximation to the Black-Scholes equation, having truncation error $O(\Delta t)$. The definition $p^{r*}, q^{r*}$ is the common definition in finance books, since the tree has no-arbitrage.

What is the meaning of a *no-arbitrage* tree? If we are sitting at node $S_j^n$, and assuming that there are only two possible future states

$$\begin{aligned} S_j^n &\rightarrow S_{j+1}^{n+1} \;; \quad \text{with probability } p \\ S_j^n &\rightarrow S_j^{n+1} \;; \quad \text{with probability } q \end{aligned}$$

then using $(\alpha^h)$ from equation (5.15) guarantees that the hedging portfolio has the same value in both future states.

But let's be a bit more sensible here. Suppose we are hedging a portfolio of RIM stock. Let $\Delta t =$ one day. Suppose the price of RIM stocks is \$10 today. Do we actually believe that tomorrow there are only two possible prices for Rim stock

$$\begin{aligned} S_{up} &= 10e^{\sigma\sqrt{\Delta t}} \\ S_{down} &= 10e^{-\sigma\sqrt{\Delta t}} \;? \end{aligned} \tag{5.19}$$

Of course not. This is obviously a highly simplified model. The fact that there is no-arbitrage in the context of the simplified model does not really have a lot of relevance to the real-world situation. The best that can be said is that if the Black-Scholes model was perfect, then we have that the portfolio hedging ratios computed using either $p^r, q^r$ or $p^{r*}, q^{r*}$ are both correct to $O(\Delta t)$.

## 5.3 A Drifting Lattice

All of the lattice methods we have discussed so far can be described in the following general way.

$$\begin{aligned} S_j^n &\rightarrow S_{j+1}^{n+1} = S_j^n u \;; \quad \text{with probability} \;\; p \\ S_j^n &\rightarrow S_j^{n+1} = S_j^n d \;; \quad \text{with probability} \;\; (1-p) \\ & S_j^n = S_0^0 u^j d^{n-j} \;, \end{aligned} \tag{5.20}$$

where $u, d$ are the up and down multipliers. For example, the lattice of Section 5.1, can be described by

$$\begin{aligned} u &= e^{\sigma\sqrt{\Delta t}} \\ d &= \frac{1}{u} \\ p &= \frac{1}{2}\left[1 + \left(\frac{r}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\Delta t}\right] \;, \end{aligned} \tag{5.21}$$

while the no-arbitrage lattice of Section 5.2 can be described by

$$\begin{aligned} u &= e^{\sigma\sqrt{\Delta t}} \\ d &= \frac{1}{u} \\ a &= e^{r\Delta t} \\ p &= \frac{a-d}{u-d} \;. \end{aligned} \tag{5.22}$$

A problem with parameters (5.21) or (5.22) is that if $r\Delta t > \sigma\sqrt{\Delta t}$, then $p > 1$, which is obviously undesirable. In fact, this leads to an unstable recursion. This can be avoided by using a *drifting lattice.*

Recall our random walk on a lattice in Section 2.5. Let $X_j^n = \log(S_j^n)$, so we will first determine our lattice parameters for $X$ and then use Ito's Lemma to convert this to the GBM case. Recall that

$$
\begin{aligned}
X_{j+1}^{n+1} &\rightarrow X_j^n + \sigma\sqrt{\Delta t}\ ; \quad \text{with probability} \ \ p \\
X_j^{n+1} &\rightarrow X_j^n - \sigma\sqrt{\Delta t}\ ; \quad \text{with probability} \ \ (1-p) \\
& \qquad p = \frac{1}{2}\left[1 + \frac{\alpha}{\sigma}\sqrt{\Delta t}\right] . 
\end{aligned} \tag{5.23}
$$

Define

$$
\begin{aligned}
\Delta X_j^n &= (X^{n+1}|X_j^n) - X_j^n \\
&= \{\sigma\sqrt{\Delta t}, -\sigma\sqrt{\Delta t}\} .
\end{aligned} \tag{5.24}
$$

Recall that

$$
\begin{aligned}
E[\Delta X_j^n] &= \alpha \Delta t \\
Var[\Delta X_j^n] &= \sigma^2 \Delta t ,
\end{aligned} \tag{5.25}
$$

and that the walk on the discrete lattice converges to the solution of the SDE

$$
dX = \alpha \ dt + \sigma \ dZ \tag{5.26}
$$

as $\Delta t \rightarrow 0$. Now, suppose that the drift $\alpha = 0$, then from equation (5.23) $p = 1/2$, and the walk converges to the SDE

$$
dX = \sigma \ dZ \ ; \quad \Delta t \rightarrow 0 . \tag{5.27}
$$

We would like to leave $p = 1/2$, but change the lattice so that we recover the original drift rate. Consider the following

$$
\begin{aligned}
X_{j+1}^{n+1} &= X_j^n + \sigma\sqrt{\Delta t} + \alpha\Delta t \\
X_j^{n+1} &= X_j^n - \sigma\sqrt{\Delta t} + \alpha\Delta t \\
& \Delta X_j^n = \{\sigma\sqrt{\Delta t} + \alpha\Delta t, -\sigma\sqrt{\Delta t} + \alpha\Delta t\} .
\end{aligned} \tag{5.28}
$$

Consequently, (noting that we leave $p = 1/2$)

$$
\begin{aligned}
E[\Delta X_j^n] &= p\left[\sigma\sqrt{\Delta t} + \alpha\Delta t\right] + (1-p)\left[-\sigma\sqrt{\Delta t} + \alpha\Delta t\right] \\
&= \alpha\Delta t .
\end{aligned} \tag{5.29}
$$

In addition

$$
\begin{aligned}
Var[\Delta X_j^n] &= E\left[(\Delta X_j^n)^2\right] - \left(E[\Delta X_j^n]\right)^2 \\
&= p\left[(\sigma\sqrt{\Delta t} + \alpha\Delta t)^2\right] + (1-p)\left[(-\sigma\sqrt{\Delta t} + \alpha\Delta t)^2\right] - (\alpha\Delta t)^2 \\
&= \sigma^2\Delta t .
\end{aligned} \tag{5.30}
$$

Consequently, the walk with $p = 1/2$ and $X_j^{n+1}, X_{j+1}^{n+1}$ given by equation (5.28) converges (as $\Delta t \rightarrow 0$) to the solution of the SDE

$$
dX = \alpha \ dt + \sigma \ dZ . \tag{5.31}
$$

| | |
|---|---|
| $T$ | 1.0 |
| $\sigma$ | .3 |
| $S_0$ (initial price) | 100 |
| $r$ | $= .02$ |
| $K$ | 100 |
| Option Type | Put |

Table 5.1: Data used in lattice tests

Now, we can price options by taking the expectation in the risk neutral world. Assuming the underlying asset follows GBM, then in the risk neutral process for the price $S$ of the asset is given by

$$\frac{dS}{S} = r\ dt + \sigma\ dZ\ . \tag{5.32}$$

Let $X = \log S$, then from Ito's Lemma we have

$$dX = \left(r - \frac{\sigma^2}{2}\right)dt + \sigma\ dZ\ . \tag{5.33}$$

This corresponds the risk neutral walk on the lattice with parameters

$$\begin{aligned} \alpha &= r - \frac{\sigma^2}{2} \\ X_j^n &= \log S_j^n\ . \end{aligned} \tag{5.34}$$

Now, substitute equation (5.34) into equation (5.28) to obtain

$$\begin{aligned} S_j^n &\rightarrow S_{j+1}^{n+1} = S_j^n u\ ;\ \text{ with probability } p \\ S_j^n &\rightarrow S_j^{n+1} = S_j^n d\ ;\ \text{ with probability } (1-p) \\ & \quad p = \frac{1}{2} \\ & \quad u = \exp[\sigma\sqrt{\Delta t} + (r - \sigma^2/2)\Delta t] \\ & \quad d = \exp[-\sigma\sqrt{\Delta t} + (r - \sigma^2/2)\Delta t]\ . \end{aligned} \tag{5.35}$$

Consequently, we can use the usual lattice backwards recursion with parameters (5.35), and this will converge to the correct price.

Compared to the lattice with parameters (5.22), the lattice with parameters (5.35) will be stable for any values of $r, \sigma, \Delta t$. However, the lattice constructed using parameters (5.35) is unbalanced, in the sense that $S_j^n ud \neq S_j^n$.

### 5.3.1 Numerical Comparison: No-arbitrage Lattice and Drifting Lattice

We use the test data in Table 5.1. Table 5.2 shows a convergence study for both lattice methods. We expect that the error in the lattice method is $O(\Delta t)$. In other words

$$V^{tree}(\Delta t, S, t) = V^{exact}(S,t) + C_1\Delta t + C_2(\Delta t)^2 \tag{5.36}$$

where $C_1$ is independent of $\Delta t$. This suggests that

$$\lim_{\Delta t \rightarrow 0} \frac{V^{tree}((\Delta t)/2, S_0^0, t=0) - V^{tree}(\Delta t, S_0^0, t=0)}{V^{tree}((\Delta t)/4, S_0^0, t=0) - V^{tree}((\Delta t)/2, S_0^0, t=0)} = 2\ . \tag{5.37}$$

| Timesteps | Value | Change | Ratio |
|---|---|---|---|
| American Put: No-arbitrage lattice (5.22) | | | |
| 500 | 11.0093931 | | |
| 1000 | 11.0113180 | 0.0037 | |
| 2000 | 11.0122835 | 9.6550e-04 | 3.80 |
| 4000 | 11.0127632 | 4.7970e-04 | 2.01 |
| 8000 | 11.0130028 | 2.3960e-04 | 2.00 |
| 16000 | 11.0131227 | 1.1990e-04 | 2.00 |
| European Put: No arbitrage lattice (5.22) | | | |
| 500 | 10.8355351 | | |
| 1000 | 10.8384915 | 0.0029564 | |
| 2000 | 10.839970 | 0.001478 | 2.00 |
| 4000 | 10.8407094 | 7.394e-04 | 2.00 |
| 8000 | 10.8410790 | 3.6960e-4 | 2.00 |
| 16000 | 10.8412639 | 1.8489e-04 | 1.99 |
| American Put: Drifting lattice (5.35) | | | |
| 500 | 11.0114885 | | |
| 1000 | 11.0157093 | 0.001882 | |
| 2000 | 11.0133918 | -0.002317 | -0.81225 |
| 4000 | 11.0139094 | 5.17599e-04 | -4.476 |
| 8000 | 11.0134913 | -4.1809e-04 | -1.2380 |
| 16000 | 11.0133705 | -1.20799e-04 | 3.46 |
| European Put: Drifting Lattice (5.35) | | | |
| 500 | 10.8383929 | | |
| 1000 | 10.8435056 | 0.00511 | |
| 2000 | 10.8413199 | -0.002185 | -2.344 |
| 4000 | 10.8420495 | 7.2909e-04 | -2.996 |
| 8000 | 10.8416523 | -3.972e-04 | -1.835 |
| 16000 | 10.8415488 | -1.0349e-04 | 3.8 |

Table 5.2: Option value at $S = 100$, $t = 0$, data in Table 5.1. Change is the change from the previous solution computed using 1/2 the timesteps. Ratio is ratio of successive changes.

We can see from Table 5.2 that for sufficiently small $\Delta t$, this appears to be true for the no-arbitrage lattice parameters. However, convergence seems to be somewhat erratic for the drifting lattice parameters (5.35). A lattice method can be viewed as an explicit finite difference method. Due to the fact that the payoff has a discontinuous derivative at the strike, finite difference theory suggests that in order to achieve a good rate of convergence, there should always be a lattice node at the strike. If we use an even number of timesteps, then this is true for a no-arbitrage lattice. However, this is not true for a drifting lattice. We suspect that this explains the poor convergence for this case.

To verify the importance of having a node at the strike, let's use the no-arbitrage parameters (5.22), and price a European put option with the parameters in Table 5.1, except that we set $S_0 = 92$. In general, in this case, there will not be a node at the strike. The results are shown in Table 5.3. The convergence is very erratic.

## 5.4 Smoothing the Payoff

As we can see from the numerical results in Section 5.3.1, it is sometimes not possible to ensure that the lattice has a node at the strike, which causes erratic convergence behaviour.

From the analysis of finite difference methods for PDEs, it is known that good convergence can usually

| Timesteps | Value | Change | Ratio |
|---|---|---|---|
| European Put: No-arbitrage lattice (5.22) | | | |
| 500 | 14.5868102 | | |
| 1000 | 14.5907463 | 0.00393609 | |
| 2000 | 14.5887935 | -0.0019528 | -2.01 |
| 4000 | 14.5885792 | -2.14299e-04 | 9.11 |
| 8000 | 14.5886970 | 1.17799e-04 | - 1.81 |
| 16000 | 14.5885398 | -1.57200e-04 | -0.749 |

Table 5.3: Option value at $S = 92$, $t = 0$, data in Table 5.1. Change is the change from the previous solution computed using 1/2 the timesteps. Ratio is ratio of successive changes.

be restored if the payoff is smoothed appropriately.

For simplicity, we will derive the smoothing technique for the case of the no-arbitrage lattice (5.22). The same final expressions will also be valid for the drifting lattice parameters (5.35). For the no-arbitrage lattice, with $N = T/(\Delta t)$,

$$S_j^N = S_0^0 e^{(2j-N)\sigma\sqrt{\Delta t}} \tag{5.38}$$

so that, letting $X_j^N = \log(S_j^N)$, then

$$\begin{aligned} X_{j+1}^N &= X_j^N + 2\sigma\sqrt{\Delta t} \\ X_{j-1}^N &= X_j^N - 2\sigma\sqrt{\Delta t} \ . \end{aligned} \tag{5.39}$$

Let $g(X)$ represent the payoff function. For example,

$$g(X) = \begin{cases} \max(e^X - K, 0) & \text{Call} \\ \max(K - e^X, 0) & \text{Put} \end{cases} , \tag{5.40}$$

with strike $K$. Now, a simple way to smooth the payoff is take its average value around node $X_j^N$. Let $\hat{g}(x)$ denote the smoothed payoff. Then

$$\hat{g}(X_j^n) = \frac{1}{2\sigma\sqrt{\Delta t}} \int_{X_j^N - \sigma\sqrt{\Delta t}}^{X_j^N + \sigma\sqrt{\Delta t}} g(x)\ dx \ . \tag{5.41}$$

Let $u = X - X_j^N$, then equation (5.41) becomes

$$\hat{g}(X_j^n) = \hat{g}_j^N = \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} g(X_j^N + u)\ du \ . \tag{5.42}$$

For a put, the smoothed payoff $\hat{P}(X_j^N)$ would be

$$\begin{aligned} \hat{P}(X_j^N) &= \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} \max(K - e^{X_j^N} e^u, 0)\ du \\ &= \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} \max(K - S_j^N e^u, 0)\ du \ , \end{aligned} \tag{5.43}$$

and the smoothed call payoff $\hat{C}(X_j^N)$ is

$$\hat{C}(X_j^N) = \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} \max(S_j^N e^u - K, 0)\ du \ . \tag{5.44}$$

For a put, from equation (5.43),

$$\begin{aligned} \max(K - S_j^N e^u, 0) &= 0 \;\; ; \;\; u \in [-\sigma\sqrt{\Delta t}, +\sigma\sqrt{\Delta t}] \\ & \quad S_j^N e^{-\sigma\sqrt{\Delta t}} > K \; . \end{aligned} \tag{5.45}$$

On the other hand,

$$\begin{aligned} \max(K - S_j^N e^u, 0) &= K - S_j^N e^u \;\; ; \;\; u \in [-\sigma\sqrt{\Delta t}, +\sigma\sqrt{\Delta t}] \\ & \quad S_j^N e^{+\sigma\sqrt{\Delta t}} < K \; . \end{aligned} \tag{5.46}$$

In this case,

$$\begin{aligned} \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} \left( K - S_j^N e^u \right) du &= \frac{1}{2\sigma\sqrt{\Delta t}} \left( K 2\sigma\sqrt{\Delta t} - S_j^N \left( e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}} \right) \right) \\ &= K - S_j^N \left( \frac{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}{2\sigma\sqrt{\Delta t}} \right) \; . \end{aligned} \tag{5.47}$$

If $S_j^N e^{-\sigma\sqrt{\Delta t}} \leq K \leq S_j^N e^{\sigma\sqrt{\Delta t}}$ then

$$\begin{aligned} & \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{+\sigma\sqrt{\Delta t}} \max(K - S_j^N e^u, 0) \; du \\ &= \frac{1}{2\sigma\sqrt{\Delta t}} \int_{-\sigma\sqrt{\Delta t}}^{\log(K/S_j^N)} \left( K - S_j^N e^u \right) du \\ &= \frac{1}{2\sigma\sqrt{\Delta t}} \left( K \left[ \log(K/S_j^N) + \sigma\sqrt{\Delta t} \right] - S_j^N \left[ K/S_j^N - e^{-\sigma\sqrt{\Delta t}} \right] \right) \; . \end{aligned} \tag{5.48}$$

Putting this all together

$$\hat{P}_j^N = \begin{cases} 0 & S_j^N e^{-\sigma\sqrt{\Delta t}} > K \\ K - S_j^N \left( \frac{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}{2\sigma\sqrt{\Delta t}} \right) & S_j^N e^{+\sigma\sqrt{\Delta t}} < K \\ \frac{1}{2\sigma\sqrt{\Delta t}} \left( K \left[ \log(K/S_j^N) + \sigma\sqrt{\Delta t} \right] - S_j^N \left[ (K/S_j^N) - e^{-\sigma\sqrt{\Delta t}} \right] \right) & S_j^N e^{-\sigma\sqrt{\Delta t}} \leq K \leq S_j^N e^{\sigma\sqrt{\Delta t}} \end{cases} \tag{5.49}$$

Similarly, the smoothed call payoff $\hat{C}_j^N$ is

$$\hat{C}_j^N = \begin{cases} 0 & S_j^N e^{+\sigma\sqrt{\Delta t}} < K \\ S_j^N \left( \frac{e^{\sigma\sqrt{\Delta t}} - e^{-\sigma\sqrt{\Delta t}}}{2\sigma\sqrt{\Delta t}} \right) - K & S_j^N e^{-\sigma\sqrt{\Delta t}} > K \\ \frac{1}{2\sigma\sqrt{\Delta t}} \left( S_j^N \left[ e^{\sigma\sqrt{\Delta t}} - (K/S_j^N) \right] - K \left[ \sigma\sqrt{\Delta t} - \log(K/S_j^N) \right] \right) & S_j^N e^{-\sigma\sqrt{\Delta t}} \leq K \leq S_j^N e^{\sigma\sqrt{\Delta t}} \end{cases} \tag{5.50}$$

Now, let's repeat the test shown in Table 5.3, this time using the smoothed payoffs. The results are shown in Table 5.4. The convergence is much improved, and is now close to first order.

As another check, we repeat the test for the American option in Table 5.2, using the Drifting lattice (5.35) with the smoothed payoff. These results are given in Table 5.5, for the case $S_0 = 100$. Again, the convergence is close to first order, and is much smoother than without smoothing.

| Timesteps | Value | Change | Ratio |
|---|---|---|---|
| European Put: No-arbitrage lattice (5.22) | | | |
| 500 | 14.5921428 | | |
| 1000 | 14.5902654 | -0.0018774 | |
| 2000 | 14.5893256 | -0.00093979 | 1.997 |
| 4000 | 14.5888575 | -0.00046810 | 2.007 |
| 8000 | 14.5886207 | -0.00023749 | 1.971 |
| 16000 | 14.5885031 | -0.0001169 | 2.031 |
| American Put: No-arbitrage lattice (5.22) | | | |
| 500 | 14.8598750 | | |
| 1000 | 14.8573440 | -0.00253 | |
| 2000 | 14.8559674 | -0.001376 | 1.839 |
| 4000 | 14.8552981 | -0.0006693 | 2.057 |
| 8000 | 14.8549594 | -0.0003387 | 1.976 |
| 16000 | 14.8547869 | -0.0001725 | 1.963 |

Table 5.4: Option value at $S = 92$, $t = 0$, data in Table 5.1. Change is the change from the previous solution computed using 1/2 the timesteps. Ratio is ratio of successive changes. Smoothed payoff.

| Timesteps | Value | Change | Ratio |
|---|---|---|---|
| American Put: Drifting lattice (5.35) | | | |
| 500 | 11.0190437 | | |
| 1000 | 11.0162151 | -0.002828599 | |
| 2000 | 11.0147342 | -0.001480900 | 1.910 |
| 4000 | 11.0140015 | -0.0007327 | 2.021 |
| 8000 | 11.0136218 | -0.0003797 | 1.929 |
| 16000 | 11.0134323 | -0.000189499 | 2.003 |

Table 5.5: Option value at $S = 100$, $t = 0$, data in Table 5.1. Change is the change from the previous solution computed using 1/2 the timesteps. Ratio is ratio of successive changes. Smoothed payoff. Drifting lattice.

| Timesteps | Value | Change | Extrapolated |
|---|---|---|---|
| American Put: Drifting lattice (5.35) | | | |
| 500 | 11.0190437 | | |
| 1000 | 11.0162151 | -0.002828599 | 11.0133865 |
| 2000 | 11.0147342 | -0.001480900 | 11.0132532 |
| 4000 | 11.0140015 | -0.0007327 | 11.0132687 |
| 8000 | 11.0136218 | - 0.0003797 | 11.0132421 |
| 16000 | 11.0134323 | -0.000189499 | 11.0132428 |
| 32000 | 11.0133381 | -9.419999e-5 | 11.0132439 |
| 64000 | 11.0132903 | -4.78000e-5 | 11.0132425 |

Table 5.6: Option value at $S = 100$, $t = 0$, data in Table 5.1. Change is the change from the previous solution computed using 1/2 the timesteps. Extrapolation refers to Richardson extrapolation (5.54). Smoothed payoff. Drifting lattice.

### 5.4.1 Richardson extrapolation

From Table (5.5) it seems we have a smoothly converging set of values for different timesteps $\Delta t$. Let $P_{\Delta t}$ be the lattice result with a timestep of size $\Delta t$. Then

$$P_{\Delta t} = P_{exact} + O(\Delta t) \ , \tag{5.51}$$

where $P_{exact}$ is the exact solution. For $\Delta t$ small, we can approximate equation (5.51) by

$$P_{\Delta t} = P_{exact} + C\Delta t + O((\Delta t)^2) \tag{5.52}$$

where $C$ is assumed to be a constant independent of $\Delta t$. Ignoring the $O((\Delta t)^2)$ term in equation (5.52), we can write equation (5.52) for two different timestep sizes

$$\begin{aligned} P_{\Delta t} &= P_{exact} + C\Delta t \\ P_{\Delta t/2} &= P_{exact} + C\frac{\Delta t}{2} \ , \end{aligned} \tag{5.53}$$

or, eliminating $C$ from equations (5.53) gives

$$P_{exact} = P_{\Delta t/2} + (P_{\Delta t/2} - P_{\Delta t}) \ . \tag{5.54}$$

Of course, we will not get the real exact solution, but we should end up with a better estimate than $P_{\Delta t/2}$. Table 5.6 shows the results of applying Richardson extrapolation for a sequence of runs with decreasing timesteps. Note that the extrapolated solution at $N = 2000$ appears to have about six digits correct, compared to the non-extrapolated result ($N = 2000$) which has about four digits correct.

## 5.5 Matlab Implementation

Here is the Matlab implementation of a European option, using the no-arbitrage lattice parameters, using loops.

**TreeSlow.m**

```
% Compute Black-Scholes option value using a binomial tree
%  European case
%  uses loops

S0 = 100;        % S0 - current stock price
K=100;           % K - strike
T = 1.0;         % T - expiry time
r = .02;         % r - interest rat
sigma = .3;      % sigma - volatility
opttype = 1;     % opttype - 0 for a call, otherwise a put
Nsteps = 10000;  % Nsteps - number of timesteps

delt = T/Nsteps;

%
%  tree parameters
%
     u = exp(sigma * sqrt(delt) );
     d = 1./u;
     a = exp( r*delt );
     p = (a - d)/(u - d);
%
% payoff at t=T
%
  W= zeros(Nsteps+1,1);

     %  W(j+1,1) corresponds to node S_{j}^N
     %               j=0,...,Nsteps

     for j=0: Nsteps
         W(j+1,1) = S0*u^(j)*d^(Nsteps -j);
             % offest by one since
             % matlab arrays start at 1
     end

    %vector operations

     if( opttype == 0)
         W = max( W - K, 0);
     else
         W = max( K - W, 0);
     end

% backward recursion

 for n=Nsteps-1:-1:0
      for j=0:n  % matlab arraya start at 1
        W(j+1,1) = exp(-r*delt)*( p*W(j+2,1) + (1-p)*W(j+1,1) );
      end
  end

  value = W(1);

 disp(sprintf('Tree Value: %.9g \n',value));
```

It is, of course, more efficient to vectorize Matlab code, i.e. avoid using explicit loops, as in *TreeFast.m*.

**TreeFast.m**

```
%
% Compute Black-Scholes option value using a binomial tree
%  European case
% vectorized code

S0 = 100;          % S0 - current stock price
K=100;             % K - strike
T = 1.0;           % T - expiry time
r = .02;           % r - interest rate
sigma = .3;        % sigma - volatility
opttype = 1;       % opttype - 0 for a call, otherwise a put
Nsteps = 10000;    % Nsteps - number of timesteps

delt = T/Nsteps;

%  tree parameters

     u = exp(sigma * sqrt(delt) );
     d = 1./u;
     a = exp( r*delt );
     p = (a - d)/(u - d);

%
% payoff at t=T
%
     W =  S0*d.^([Nsteps:-1:0]').*u.^([0:Nsteps]');
               % W is column vector of size Nsteps+1 X 1
     if( opttype == 0)
         W = max( W - K, 0);
     else
         W = max( K - W, 0);
     end


% backward recursion

  for i=Nsteps:-1:1
      W = exp(-r*delt)*( p*W(2:i+1) + (1-p)*W(1:i) );
  end

 value = W(1);

 disp(sprintf('Tree Value: %.9g \n',value));
```

### 5.5.1 American Case

Implementing the American case requires the payoff at each node in the tree. This can be done efficiently using vector statements. For example, since $S_{j+1}^{n+1} = S_j^n u$, then we generate the asset prices at timestep $n$ in terms of the asset prices at timestep $n+1$, using a vector statement.

**Code Fragment: American Option**

```
S = S0*d.^([Nsteps:-1:0]').*u.^([0:Nsteps]');  % column vector
                                               % S at t=N
  ....

 for i=Nsteps:-1:1
    W = exp(-r*delt)*( p*W(2:i+1) + (1-p)*W(1:i) );

    S = S(2:i+1,1)/u;  % S at timestep i

    % add American Check
    ...
 end
```

### 5.5.2 Discrete Fixed Amount Dividends

In real life, fixed (not proportional) dividends are usually announced in advance of the dividend date. Implementing this in a binomial tree is a source of much confusion. Many textbooks get this wrong. For a discussion of various approaches, and their problems, see (*Back to basics: a new approach to discrete dividend problem,* Wilmott Magazine (2003), Volume 9, Haug, Haug and Lewis; *Efficient Pricing of Derivatives on Assets with Discrete Dividends*, Applied Mathematical Finance (2006) Vol 13, pages 265-284, M. H. Vellekoop, J. W. Nieuwenhuis).

Some common suggestions are:

- Use a non-recombining tree (very slow)
- Use an *escrowed dividend* model (does not converge to the correct solution in some cases).

We go back to basics here. Suppose the announced dividend at $t_D$ is $D^*$. Let $S(t_D^+)$ be the value of the stock price the instant after the dividend is paid, and $S(t_D^-)$ be the value of the stock just before $t_D$. Then, by no-arbitrage

$$S(t_D^+) \quad = S(t_D^-) - D^* \ . \tag{5.55}$$

But equation (5.55) is not quite right. The actual dividend paid cannot be more than the stock price (otherwise there is an arbitrage opportunity). So, we modify equation (5.55)

$$S(t_D^+) \quad = S(t_D^-) - \min(D^* \ , S(t_D^-)). \tag{5.56}$$

Let $V(S,t)$ be the value of an option. By no-arbitrage, we must have

$$V(S(t_D^-), t_D^-) = V(S(t_D^+), t_D^+) \ . \tag{5.57}$$

We simply apply equation (5.57) at any dividend date. Of course, $S(t_D^+)$ may not correspond to a node in the tree, so we have to interpolate to get the value we need. If we use linear interpolation, this introduces an error of size $O(\Delta t)$, which is the same order as the usual $O(\Delta t)$ lattice error. A function which does this operation is shown in **dividend.m**.

| | |
|---|---|
| $T$ | 1.0 |
| $\sigma$ | .2 |
| $S_0$ (initial price) | 100 |
| $r$ | .05 |
| $K$ | 100 |
| Option Type | Put |
| American | Yes |
| Dividend date $t_D$ | 0.5 |
| Dividend value $D^*$ | 5.0 |

Table 5.7: Data used in discrete dividend example.

| Timesteps | Value | Change | Extrapolated |
|---|---|---|---|
| American Put: No-arbitrage lattice (5.22) | | | |
| 2000 | 8.44264346 | | |
| 4000 | 8.44181828 | -0.00082517 | 8.44099311 |
| 8000 | 8.44142316 | -0.00039511 | 8.44102805 |
| 16000 | 8.44120877 | -0.00021439 | 8.44099438 |

Table 5.8: Discrete dividend example. Option value at $S = 100$, $t = 0$, data in Table 5.7. Change is the change from the previous solution computed using 1/2 the timesteps. Extrapolation refers to Richardson extrapolation (5.54). Smoothed payoff.

**dividend.m**

```
 function W_out = dividend( W_in, S, div_value)

% W_in:   value of option at t^+
% S   :   asset prices
% div_value:    discrete dollar dividend

%
%   W_out: option value at t^-
%
%   assume American constraint applied in caller


 S_min = min(S);

 S_ex = S - div_value; % ex dividend stock value

 S_ex = max( S_ex, S_min);  % make sure that
                     % dividend payment does
                     % not cause S < S_min

 W_out = interp1( S,  W_in, S_ex);
```

### 5.5.3 Discrete Dividend Example

The parameters for this example are given in Table 5.7. Computational results are shown in Table 5.8.

## 5.6 Dynamic Programming

Note that in equation (5.8) we were able to determine the optimal strategy for exercising an American option by working backwards through the lattice, and comparing the value of immediate exercise with the value of continuing to hold. This is an example of the Dynamic Programming Principle.

This idea was originally developed by Richard Bellman. Here is a quote from his address to the American Mathematical society in 1953:

> *"An optimal policy has the property that, whatever the initial state and initial decision, the remaining decisions must constitute an optimal policy given the state resulting from that first decision."*

This can be stated more succinctly as

> **Principle of Optimality** *From any point on an optimal trajectory, the remaining trajectory is optimal for the corresponding problem initiated at that point.*

Here is a classic example of the application of dynamic programming. Suppose you desire to marry the richest person you meet. We make the assumptions that

- If you ask anyone to marry you, they will immediately accept.
- You are allowed to meet only $N$ persons.
- The wealth of each person is randomly uniformly distributed in $[0, 1]$.

What is the optimal strategy to maximize the expected wealth of your marriage partner?

One way to solve this problem would be think about what happens when you meet the first person. You could then compare that person's wealth with what you might expect to get later on if you rejected the first person. But, what you expect to get later on is a complicated mixture of all the possible decisions you could make later. This becomes very difficult to work out.

This is where the Principle of Optimality comes to the rescue. Let's consider the simple problem of what to do if you encounter potential partner $N - 1$. At this stage, you only have two choices

- Ask $N - 1$ to marry you.
- Marry partner $N$.

What should you do? If you marry partner $N$, then the expected wealth of $N$ is .5. Thus, you should marry $N - 1$ if his/her wealth is more than .5, otherwise marry $N$. The probability that partner $N - 1$ has wealth greater than .5 is .5. You thus marry partner $N - 1$ about $1/2$ the time, and $1/2$ the time you will marry partner $N$. The expected wealth of partner $N - 1$, given that you marry him/her is $(.5 + 1)/2 = .75$.

Let $V_N$ be the expected value of a partner at step $N$. In this case $V_N = .5$. Let $W_{N-1}$ be the wealth of a partner you marry at step $N - 1$ (assuming you stop at step $N - 1$). Let $\overline{W}_{N-1}$ be the expected wealth of your partner at step $N - 1$ assuming you act optimally. More precisely

$$\overline{W}_{N-1} = E[W_{N-1} | W_{N-1} > V_N] \tag{5.58}$$

where $E[\cdot]$ is the expectation operator. Let $p$ be the probability that $W_{N-1} > V_N$. Your expected partner wealth at step $N - 1$ is thus

$$\begin{aligned} V_{N-1} &= p \times \overline{W}_{N-1} + (1 - p) \times V_N && (5.59) \\ &= .5 \times .75 + .5 \times .5 \\ &= .625 && (5.60) \end{aligned}$$

Note that $V_{N-1}$ is the expected partner wealth assuming you act optimally at step $N - 1$.

Now, writing down equation (5.59) at step $N-2$ gives

$$\begin{aligned} V_{N-2} &= p \times \overline{W}_{N-2} + (1-p) \times V_{N-1} \\ &\quad p = Prob[W_{N-2} > V_{N-1}] \\ &\quad \overline{W}_{N-2} = E[W_{N-2} | W_{N-2} > V_{N-1}] \end{aligned} \tag{5.61}$$

Since $V_{N-1} = .625$, we have $p = (1-.625) = .375$, and $\overline{W}_{N-2} = (.625+1)/2 = .8125$. Your expected partner wealth at step $N-2$ is then

$$\begin{aligned} V_{N-2} &= .375 \times \overline{W}_{N-2} + .625 \times V_{N-1} \\ &= .375 \times .8125 + .625 \times .625 \\ &= .6953125 \end{aligned} \tag{5.62}$$

Again, note that $V_{N-2}$ is the expected partner wealth assuming you act optimally at steps $N-2, N-1$. We can continue working backwards to step $N-3, N-4$ and so on. In general

$$\begin{aligned} V_k &= p \times \overline{W}_k + (1-p) \times V_{k+1} \\ &\quad p = Prob[W_k > V_{k+1}] \\ &\quad \overline{W}_k = E[W_k | W_k > V_{k+1}] \ , \end{aligned} \tag{5.63}$$

where

$$\begin{aligned} V_k &= E\,[\text{Wealth of partner, at step } k\text{, assuming optimal strategy} \\ &\qquad \text{used at steps } k, k+1, ...] \\ W_k &= \text{Wealth of partner } k \text{ assuming it is optimal to stop at step } k \ . \end{aligned} \tag{5.64}$$

This algorithm is shown in Figure 5.3. If $N = 10$, then the expected partner wealth is about .861.

# 6 More on Ito's Lemma

In Section 2.6.1, we mysteriously made the infamous comment

> ...it can be shown that $dZ^2 \to dt$ as $dt \to 0$

In this Section, we will give some justification this remark. For a lot more details here, we refer the reader to *Stochastic Differential Equations*, by Bernt Oksendal, Springer, 1998.

We have to go back here, and decide what the statement

$$dX = \alpha dt + c dZ \tag{6.1}$$

really means. The only sensible interpretation of this is

$$X(t) - X(0) = \int_0^t \alpha(X(s), s) ds + \int_0^t c(X(s), s) dZ(s) \ . \tag{6.2}$$

where we can interpret the integrals as the limit, as $\Delta t \to 0$ of a discrete sum. For example,

$$\begin{aligned} \int_0^t c(X(s), s) dZ(s) &= \lim_{\Delta t \to 0} \sum_{j=0}^{j=N-1} c_j \Delta Z_j \\ &\quad c_j = c(X(t_j), t_j) \\ &\quad Z_j = Z(t_j) \\ &\quad \Delta Z_j = Z(t_{j+1}) - Z(t_j) \\ &\quad \Delta t = t_{j+1} - t_j \\ &\quad N = t/(\Delta t) \end{aligned} \tag{6.3}$$

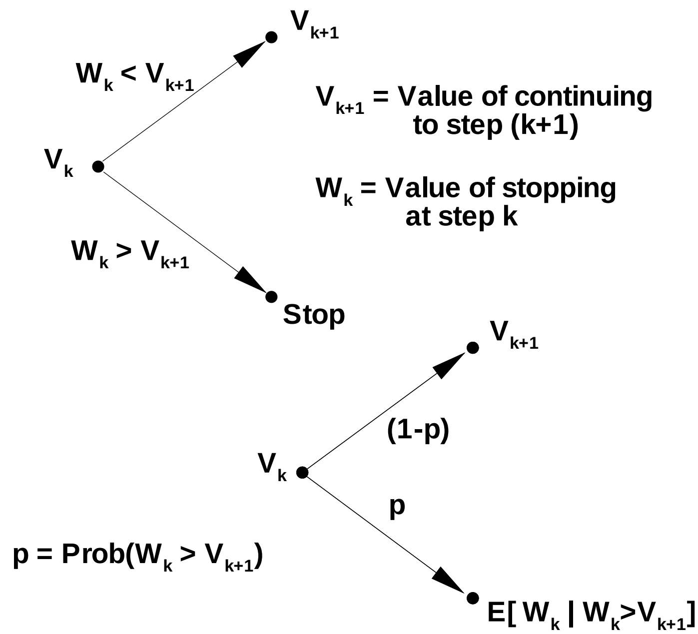



Figure 5.3: The dynamic programming algorithm. $V_{k+1}$ is the value of continuing, assuming the optimal strategy is followed at steps $k+1, k+2, ...$

In particular, in order to derive Ito's Lemma, we have to decide what

$$\int_0^t c(X(s), s)\ dZ(s)^2 \tag{6.4}$$

means. Replace the integral by a sum,

$$\int_0^t c(X(s), s)\ dZ(s)^2 \quad = \quad \lim_{\Delta t \to 0} \sum_{j=0}^{j=N-1} c(X_j, t_j) \Delta Z_j^2 \ . \tag{6.5}$$

Note that we have evaluated the integral using the left hand end point of each subinterval (the *no peeking into the future* principle).

From now on, we will use the notation

$$\sum_j \quad \equiv \quad \sum_{j=0}^{j=N-1} \ . \tag{6.6}$$

Now, we claim that

$$\int_0^t c(X(s), s) dZ^2(s) \quad = \quad \int_0^t c(X(s), s) ds \tag{6.7}$$

or

$$\lim_{\Delta t \to 0} \left[ \sum_j c_j \Delta Z_j^2 \right] \quad = \quad \lim_{\Delta t \to 0} \sum_j c_j \Delta t \tag{6.8}$$

which is what we mean by equation (6.7). i.e. we can say that $dZ^2 \to dt$ as $dt \to 0$.

Now, let's consider a finite $\Delta t$, and consider the expression

$$E\left[\left(\sum_j c_j \Delta Z_j^2 - \sum_j c_j \Delta t\right)^2\right] \tag{6.9}$$

If equation (6.9) tends to zero as $\Delta t \to 0$, then we can say that (in the mean square limit)

$$\begin{aligned}\lim_{\Delta t \to 0}\left[\sum_j c_j \Delta Z_j^2\right] &= \lim_{\Delta t \to 0}\sum_j c_j \Delta t \\ &= \int_0^t c(X(s), s)\; ds \end{aligned} \tag{6.10}$$

so that in this sense

$$\int_0^t c(X, s)\; dZ^2 = \int_0^t c(X, s)\; ds \tag{6.11}$$

and hence we can say that

$$dZ^2 \to dt \tag{6.12}$$

with probability one as $\Delta t \to 0$.

Now, expanding equation (6.9) gives

$$E\left[\left(\sum_j c_j \Delta Z_j^2 - \sum_j c_j \Delta t\right)^2\right] = \sum_{ij} E\left[c_j(\Delta Z_j^2 - \Delta t)c_i(\Delta Z_i^2 - \Delta t)\right] \ . \tag{6.13}$$

Now, note the following

- The increments of Brownian motion are uncorrelated, i.e. $Cov\left[\Delta Z_i\ \Delta Z_j\right] = 0,\ i \neq j$, which means that $Cov\left[\Delta Z_i^2\ \Delta Z_j^2\right] = 0$, or $E\left[(\Delta Z_j^2 - \Delta t)(\Delta Z_i^2 - \Delta t)\right] = 0,\ i \neq j$.
- $c_i = c(t_i, X(Z_i))$, and $\Delta Z_i$ are independent.

It then follows that for $i < j$

$$\begin{aligned} E\left[c_j(\Delta Z_j^2 - \Delta t)c_i(\Delta Z_i^2 - \Delta t)\right] &= E[c_i c_j(\Delta Z_i^2 - \Delta t)]E[(\Delta Z_j^2 - \Delta t)] \\ &= 0 \ . \end{aligned} \tag{6.14}$$

Similarly, if $i > j$

$$\begin{aligned} E\left[c_j(\Delta Z_j^2 - \Delta t)c_i(\Delta Z_i^2 - \Delta t)\right] &= E[c_i c_j(\Delta Z_j^2 - \Delta t)]E[(\Delta Z_i^2 - \Delta t)] \\ &= 0 \ . \end{aligned} \tag{6.15}$$

So that in all cases

$$E\left[c_j(\Delta Z_j^2 - \Delta t)c_i(\Delta Z_i^2 - \Delta t)\right] = \delta_{ij} E\left[c_i^2(\Delta Z_i^2 - \Delta t)^2\right] \ . \tag{6.16}$$

It also follows from the above properties that

$$E[c_j^2(\Delta Z_j^2 - \Delta t)^2] = E[c_j^2]\ E[(\Delta Z_j^2 - \Delta t)^2] \tag{6.17}$$

since $c_j$ and $(\Delta Z_j^2 - \Delta t)$ are independent.

Using equations (6.16-6.17), then equation (6.13) becomes

$$\sum_{ij} E\left[c_j(\Delta Z_j^2 - \Delta t)\ c_i(\Delta Z_i^2 - \Delta t)\right] \quad = \quad \sum_i E[c_i^2]\ E\left[(\Delta Z_i^2 - \Delta t)^2\right] \ . \tag{6.18}$$

Now,

$$\sum_i E[c_i^2]\ E\left[(\Delta Z_i^2 - \Delta t)^2\right] \quad = \quad \sum_i E[c_i^2]\ \left(E\left[\Delta Z_i^4\right] - 2\Delta t E\left[\Delta Z_i^2\right] + (\Delta t)^2\right) \ . \tag{6.19}$$

Recall that $\Delta Z$ is $N(0, \Delta t)$ ( normally distributed with mean zero and variance $\Delta t$) so that

$$\begin{aligned} E\left[(\Delta Z_i)^2\right] &= \Delta t \\ E\left[(\Delta Z_i)^4\right] &= 3(\Delta t)^2 \end{aligned} \tag{6.20}$$

so that equation (6.19) becomes

$$E\left[\Delta Z_i^4\right] - 2\Delta t E\left[\Delta Z_i^2\right] + (\Delta t)^2 \quad = \quad 2(\Delta t)^2 \tag{6.21}$$

and

$$\begin{aligned} \sum_i E[c_i^2]\ E\left[(\Delta Z_i^2 - \Delta t)^2\right] &= 2\sum_i E[c_i^2](\Delta t)^2 \\ &= 2\Delta t \left(\sum_i E[c_i^2]\Delta t\right) \\ &= O(\Delta t) \end{aligned} \tag{6.22}$$

so that we have

$$E\left[\left(\sum c_j \Delta Z_j^2 - \sum_j c_j \Delta t\right)^2\right] \quad = \quad O(\Delta t) \tag{6.23}$$

or

$$\lim_{\Delta t \to 0} E\left[\left(\sum c_j \Delta Z_j^2 - \int_0^t c(s, X(s)) ds\right)^2\right] \quad = \quad 0 \tag{6.24}$$

so that in this sense we can write

$$dZ^2 \quad \to \quad dt \quad ; \quad dt \to 0 \ . \tag{6.25}$$

# 7 Derivative Contracts on non-traded Assets and Real Options

The hedging arguments used in previous sections use the underlying asset to construct a hedging portfolio. What if the underlying asset cannot be bought and sold, or is non-storable? If the underlying variable is an interest rate, we can't store this. Or if the underlying asset is bandwidth, we can't store this either. However, we can get around this using the following approach.

## 7.1 Derivative Contracts

Let the underlying variable follow

$$dS = a(S,t)dt + b(S,t)dZ, \tag{7.1}$$

and let $F = F(S,t)$, so that from Ito's Lemma

$$dF = \left[ aF_S + \frac{b^2}{2} F_{SS} + F_t \right] dt + bF_S dZ, \tag{7.2}$$

or in shorter form

$$\begin{aligned} dF &= \mu dt + \sigma^* dZ \\ &\quad \mu = aF_S + \frac{b^2}{2} F_{SS} + F_t \\ &\quad \sigma^* = bF_S \ . \end{aligned} \tag{7.3}$$

Now, instead of hedging with the underlying asset, we will hedge one contract with another. Suppose we have two contracts $F_1, F_2$ (they could have different maturities for example). Then

$$\begin{aligned} dF_1 &= \mu_1 dt + \sigma_1^* dZ \\ dF_2 &= \mu_2 dt + \sigma_2^* dZ \\ &\quad \mu_i = a(F_i)_S + \frac{b^2}{2}(F_i)_{SS} + (F_i)_t \\ &\quad \sigma_i^* = b(F_i)_S \quad ; \quad i = 1,2 \quad . \end{aligned} \tag{7.4}$$

Consider the portfolio $\Pi$

$$\Pi = n_1 F_1 + n_2 F_2 \tag{7.5}$$

so that

$$\begin{aligned} d\Pi &= n_1 \ dF_1 + n_2 \ dF_2 \\ &= n_1(\mu_1 \ dt + \sigma_1^* \ dZ) + n_2(\mu_2 \ dt + \sigma_2^* \ dZ) \\ &= (n_1\mu_1 + n_2\mu_2) \ dt + (n_1\sigma_1^* + n_2\sigma_2^*) \ dZ \ . \end{aligned} \tag{7.6}$$

Now, to eliminate risk, choose

$$(n_1\sigma_1^* + n_2\sigma_2^*) = 0 \tag{7.7}$$

which means that $\Pi$ is riskless, hence

$$d\Pi = r\Pi \ dt \ , \tag{7.8}$$

so that, using equations (7.6-7.8), we obtain

$$(n_1\mu_1 + n_2\mu_2) = r(n_1 F_1 + n_2 F_2). \tag{7.9}$$

Putting together equations (7.7) and (7.9) gives

$$\begin{bmatrix} \sigma_1^* & \sigma_2^* \\ \mu_1 - rF_1 & \mu_2 - rF_2 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} . \tag{7.10}$$

Now, equation (7.10) only has a nonzero solution if the two rows of equation (7.10) are linearly dependent. In other words, there must be a $\lambda_S = \lambda_S(S,t)$ (independent of the type of contract) such that

$$\begin{aligned} \mu_1 - rF_1 &= \lambda_S \sigma_1^* \\ \mu_2 - rF_2 &= \lambda_S \sigma_2^* \ . \end{aligned} \tag{7.11}$$

Dropping the subscripts, we obtain

$$\frac{\mu - rF}{\sigma^*} = \lambda_S \tag{7.12}$$

Substituting $\mu, \sigma^*$ from equations (7.3) into equation (7.12) gives

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_S b) F_S - rF = 0 \ . \tag{7.13}$$

Equation (7.13) is the PDE satisfied by a derivative contract on any asset $S$. Note that it does not matter if we cannot trade $S$.

Suppose that $F_2 = S$ is a traded asset. Then we can hedge with $S$, and from equation (7.11) we have

$$\mu_2 - rS = \lambda_S \sigma_2^* \tag{7.14}$$

and from equations (7.1) and (7.3) we have

$$\begin{aligned} \sigma_2^* &= b \\ \mu_2 &= a \end{aligned} \tag{7.15}$$

and so, using equations (7.11) and (7.15) , we have that

$$\lambda_S = \frac{a - rS}{b} \tag{7.16}$$

and equation (7.13) reduces to

$$F_t + \frac{b^2}{2} F_{SS} + rSF_S - rF = 0 \ . \tag{7.17}$$

Suppose

$$\begin{aligned} \mu &= F\mu' \\ \sigma^* &= F\sigma' \end{aligned} \tag{7.18}$$

so that we can write

$$dF = F\mu' dt + F\sigma' dZ \tag{7.19}$$

then using equation (7.18) in equation (7.12) gives

$$\mu' = r + \lambda_S \sigma' \tag{7.20}$$

which has the convenient interpretation that the expected return on holding (not hedging) the derivative contract $F$ is the risk-free rate plus extra compensation due to the riskiness of holding $F$. The extra return is $\lambda_S \sigma'$, where $\lambda_S$ is the market price of risk of $S$ (which should be the same for all contracts depending on $S$) and $\sigma'$ is the volatility of $F$. Note that the volatility and drift of $F$ are not the volatility and drift of the underlying asset $S$.

If we believe that the Capital Asset Pricing Model holds, then a simple minded idea is to estimate

$$\lambda_S = \rho_{SM} \lambda_M \tag{7.21}$$

where $\lambda_M$ is the price of risk of the market portfolio, and $\rho_{SM}$ is the correlation of returns between $S$ and the returns of the market portfolio.

Another idea is the following. Suppose we can find some companies whose main source of business is based on $S$. Let $q_i$ be the price of this companies stock at $t = t_i$. The return of the stock over $t_i - t_{i-1}$ is

$$R_i = \frac{q_i - q_{i-1}}{q_{i-1}} .$$

Let $R_i^M$ be the return of the market portfolio (i.e. a broad index) over the same period. We compute $\beta$ as the best fit linear regression to

$$R_i = \alpha + \beta R_i^M$$

which means that

$$\beta = \frac{Cov(R, R^M)}{Var(R^M)}. \tag{7.22}$$

Now, from CAPM we have that

$$E(R) = r + \beta \left[ E(R^M) - r \right] \tag{7.23}$$

where $E(...)$ is the expectation operator. We would like to determine the *unlevered* $\beta$, denoted by $\beta^u$, which is the $\beta$ for an investment made using equity only. In other words, if the firm we used to compute the $\beta$ above has significant debt, its riskiness with respect to $S$ is amplified. The unlevered $\beta$ can be computed by

$$\beta^u = \frac{E}{E + (1 - T_c) D} \beta \tag{7.24}$$

where

$$\begin{aligned} D &= \text{long term debt} \\ E &= \text{Total market capitalization} \\ T_c &= \text{Corporate Tax rate} \ . \end{aligned} \tag{7.25}$$

So, now the expected return from a pure equity investment based on $S$ is

$$E(R^u) = r + \beta^u \left[ E(R^M) - r \right] . \tag{7.26}$$

If we assume that $F$ in equation (7.19) is the company stock, then

$$\begin{aligned} \mu' &= E(R^u) \\ &= r + \beta^u \left[ E(R^M) - r \right] \end{aligned} \tag{7.27}$$

But equation (7.20) says that

$$\mu' = r + \lambda_S \sigma' . \tag{7.28}$$

Combining equations (7.27-7.27) gives

$$\lambda_S \sigma' = \beta^u \left[ E(R^M) - r \right] . \tag{7.29}$$

Recall from equations (7.3) and (7.18) that

$$\begin{aligned}\sigma^* &= F\sigma' \\ \sigma^* &= bF_S \ ,\end{aligned}$$

or

$$\sigma' = \frac{bF_S}{F} \ . \tag{7.30}$$

Combining equations (7.29-7.30) gives

$$\lambda_S = \frac{\beta^u \left[E(R^M) - r\right]}{\frac{bF_S}{F}} \ . \tag{7.31}$$

In principle, we can now compute $\lambda_S$, since

- The unleveraged $\beta^u$ is computed as described above. This can be done using market data for a specific firm, whose main business is based on $S$, and the firms balance sheet.
- $b(S,t)/S$ is the volatility rate of $S$ (equation (7.1)).
- $[E(R^M) - r]$ can be determined from historical data. For example, the expected return of the market index above the risk free rate is about 6% for the past 50 years of Canadian data.
- The risk free rate $r$ is simply the current T-bill rate.
- $F_S$ can be estimated by computing a linear regression of the stock price of a firm which invests in $S$, and $S$. Now, this may have to be unlevered, to reduce the effect of debt. If we are going to now value the real option for a specific firm, we will have to make some assumption about how the firm will finance a new investment. If it is going to use pure equity, then we are done. If it is a mixture of debt and equity, we should relever the value of $F_S$. At this point, we need to talk to a Finance Professor to get this right.

## 7.2 A Forward Contract

A forward contract is a special type of derivative contract. The holder of a forward contract agrees to buy or sell the underlying asset at some delivery price $K$ in the future. $K$ is determined so that the cost of entering into the forward contract is zero at its inception.

The payoff of a (long) forward contract expiring at $t = T$ is then

$$V(S, \tau = 0) = S(T) - K \ . \tag{7.32}$$

Note that there is no optionality in a forward contract.

The value of a forward contract is a contingent claim. and its value is given by equation (7.13)

$$V_t + \frac{b^2}{2} V_{SS} + (a - \lambda_S b) V_S - rV = 0 \ . \tag{7.33}$$

Now we can also use a simple no-arbitrage argument to express the value of a forward contract in terms of the original delivery price $K$, (which is set at the inception of the contract) and the current forward price $f(S, \tau)$. Suppose we are long a forward contract with delivery price $K$. At some time $t > 0$, ($\tau < T$), the forward price is no longer $K$. Suppose the forward price is $f(S, \tau)$, then the payoff of a long forward contract, entered into at ($\tau$) is

$$\text{Payoff} = S(T) - f(S(\tau), \tau) \ .$$

Suppose we are long the forward contract struck at $t = 0$ with delivery price $K$. At some time $t > 0$, we hedge this contract by going short a forward with the current delivery price $f(S, \tau)$ (which costs us nothing to enter into). The payoff of this portfolio is

$$S - K - (S - f) = f - K \tag{7.34}$$

Since $f, K$ are known with certainty at $(S, \tau)$, then the value of this portfolio today is

$$(f - K)e^{-r\tau} \ . \tag{7.35}$$

But if we hold a forward contract today, we can always construct the above hedge at no cost. Therefore,

$$V(S, \tau) \quad = \quad (f - K)e^{-r\tau} \ . \tag{7.36}$$

Substituting equation (7.36) into equation (7.33), and noting that $K$ is a constant, gives us the following PDE for the forward price (the delivery price which makes the forward contract worth zero at inception)

$$f_\tau = \frac{b^2}{2} f_{SS} + (a - \lambda_S b) f_S \tag{7.37}$$

with terminal condition

$$f(S, \tau = 0) = S \tag{7.38}$$

which can be interpreted as the fact that the forward price must equal the spot price at $t = T$.

Suppose we can estimate $a, b$ in equation (7.37), and there are a set of forward prices available. We can then estimate $\lambda_S$ by solving equation (7.37) and adjusting $\lambda_S$ until we obtain a good fit for the observed forward prices.

### 7.2.1 Convenience Yield

We can also write equation (7.37) as

$$f_t + \frac{b^2}{2} f_{SS} + (r - \delta) S f_S = 0 \tag{7.39}$$

where $\delta$ is defined as

$$\delta = r - \frac{a - \lambda_S b}{S} \ . \tag{7.40}$$

In this case, we can interpret $\delta$ as the *convenience yield* for holding the asset. For example, there is a convenience to holding supplies of natural gas in reserve.

### 7.2.2 Volatility of Forward Prices

From equation (7.37) we have that the forward price for a contract expiring at time $T$, at current time $t$, spot price $S(t)$ is given by

$$f(S, t) \quad = \quad E^Q[S(T)] \tag{7.41}$$

where $S$ follows the risk neutral process

$$dS \quad = \quad (a - \lambda_S b)\ dt + b\ dZ \ . \tag{7.42}$$

In other words. the *forward price is the risk neutral expected spot price at expiry.*

Now, using Ito's Lemma and assuming the risk neutral spot process (7.42) gives

$$df = \left(f_t + \frac{b^2}{2} f_{SS} + (a - \lambda_S b) f_S\right) dt + f_S b \; dZ \; . \tag{7.43}$$

But since $f$ satisfies equation (7.37), equation (7.43) becomes

$$\begin{aligned} df &= f_S b \; dZ \\ &= \hat{\sigma} f \; dZ \; , \end{aligned} \tag{7.44}$$

where the *effective volatility* of the forward price is

$$\hat{\sigma} = \frac{f_S b}{f} \; . \tag{7.45}$$

Note that from equation (7.44), the forward price has zero drift.

# 8 Discrete Hedging

In practice, we cannot hedge at infinitesimal time intervals. In fact, we would like to hedge as infrequently as possible, since in real life, there are transaction costs (something which is ignored in the basic Black-Scholes equation, but which can be taken into account and results in a nonlinear PDE).

## 8.1 Delta Hedging

Recall that the basic derivation of the Black-Scholes equation used a hedging portfolio where we hold $V_S$ shares. In finance, $V_S$ is called the option delta, hence this strategy is called delta hedging.

As an example, consider the hedging portfolio $P(t)$ which is composed of

- A short position in an option $-V(t)$.
- Long $\alpha(t)^h S(t)$ shares
- An amount in a risk-free bank account $B(t)$.

Initially, we have

$$\begin{aligned} P(0) = 0 &= -V(0) + \alpha(0)^h S(0) + B(0) \\ & \quad \alpha = V_S \\ & \quad B(0) = V(0) - \alpha(0)^h S(0) \end{aligned}$$

The hedge is rebalanced at discrete times $t_i$. Defining

$$\begin{aligned} \alpha_i^h &= V_S(S_i, t_i) \\ V_i &= V(S_i, t_i) \end{aligned}$$

then, we have to update the hedge by purchasing $\alpha_i - \alpha_{i-1}$ shares at $t = t_i$, so that updating our share position requires

$$S(t_i)(\alpha_i^h - \alpha_{i-1}^h)$$

in cash, which we borrow from the bank if $(\alpha_i^h - \alpha_{i-1}^h) > 0$. If $(\alpha_i^h - \alpha_{i-1}^h) < 0$, then we sell some shares and deposit the proceeds in the bank account. If $\Delta t = t_i - t_{i-1}$, then the bank account balance is updated by

$$B_i = e^{r\Delta t} B_{i-1} - S_i(\alpha_i^h - \alpha_{i-1}^h)$$

At the instant after the rebalancing time $t_i$, the value of the portfolio is

$$P(t_i) = -V(t_i) + \alpha(t_i)^h S(t_i) + B(t_i)$$

Since we are hedging at discrete time intervals, the hedge is no longer risk free (it is risk free only in the limit as the hedging interval goes to zero). We can determine the distribution of profit and loss ( P& L) by carrying out a Monte Carlo simulation. Suppose we have precomputed the values of $V_S$ for all the likely $(S, t)$ values. Then, we simulate a series of random paths. For each random path, we determine the discounted relative hedging error

$$error = \frac{e^{-rT^*} P(T^*)}{V(S_0, t = 0)} \tag{8.1}$$

After computing many sample paths, we can plot a histogram of relative hedging error, i.e. fraction of Monte Carlo trials giving a hedging error between $E$ and $E + \Delta E$. We can compute the variance of this distribution, and also the value at risk (VAR). VAR is the worst case loss with a given probability. For example, a typical VAR number reported is the maximum loss that would occur 95% of the time. In other words, find the value of $E$ along the x-axis such that the area under the histogram plot to the right of this point is .95× the total area.

As an example, consider the case of an American put option, $T = .25$, $\sigma = .3$, $r = .06, K = S_0 = 100$. At $t = 0$, $S_0 = 100$. Since there are discrete hedging errors, the results in this case will depend on the stock drift rate, which we set at $\mu = .08$. The initial value of the American put, obtained by solving the Black-Scholes linear complementarity problem, is \$5.34. Figure 8.1 shows the results for no hedging, and hedging once a month. The x-axis in these plots shows the relative P & L of this portfolio (i.e. P & L divided by the Black-Scholes price), and the y-axis shows the relative frequency.

$$\text{Relative P\& L} = \frac{\text{Actual P\& L}}{\text{Black-Scholes price}} \tag{8.2}$$

Note that the no-hedging strategy actually has a high probability of ending up with a profit (from the option writer's point of view) since the drift rate of the stock is positive. In this case, the hedger does nothing, but simply pockets the option premium. Note the sudden jump in the relative frequency at relative $P\&L = 1$. This is because the maximum the option writer stands to gain is the option premium, which we assume is the Black-Scholes value. The writer makes this premium for any path which ends up $S > K$, which is many paths, hence the sudden jump in probability. However, there is significant probability of a loss as well. Figure 8.1 also shows the relative frequency of the $P\&L$ of hedging once a month (only three times during the life of the option).

In fact, there is a history of Ponzi-like hedge funds which simply write put options with essentially no hedging. In this case, these funds will perform very well for several years, since markets tend to drift up on average. However, then a sudden market drop occurs, and they will *blow up*. Blowing up is a technical term for losing all your capital and being forced to get a real job. However, usually the owners of these hedge funds walk away with large bonuses, and the shareholders take all the losses.

Figure 8.2 shows the results for rebalancing the hedge once a week, and daily. We can see clearly here that the mean is zero, and variance is getting smaller as the hedging interval is reduced. In fact, one can show that the standard deviation of the hedge error should be proportional to $\sqrt{\Delta t}$ where $\Delta t$ is the hedge rebalance frequency.

As another example, it is interesting to examine the stock, bond, and portfolio values along a single stochastic path. Figure 8.3 shows these values for hedging one year put, at a very high rebalancing frequency.

## 8.2 Gamma Hedging

In an attempt to account for some the errors in delta hedging at finite hedging intervals, we can try to use second derivative information. The second derivative of an option value $V_{SS}$ is called the option gamma, hence this strategy is termed delta-gamma hedging.

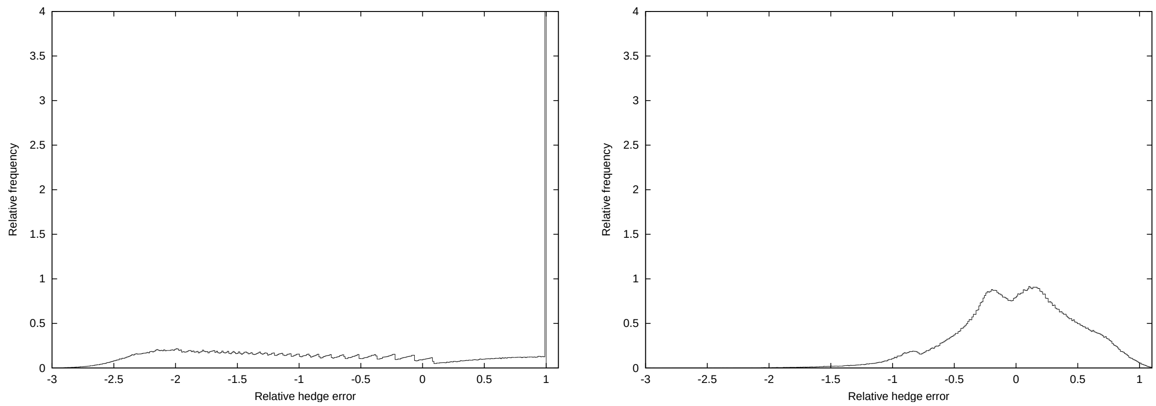



Figure 8.1: Relative frequency (y-axis) versus relative P&L of delta hedging strategies. Left: no hedging, right: rebalance hedge once a month. American put, $T = .25$, $\sigma = .3$, $r = .06$, $\mu = .08, K = S_0 = 100$. The relative P&L is computed by dividing the actual P&L by the Black-Scholes price.

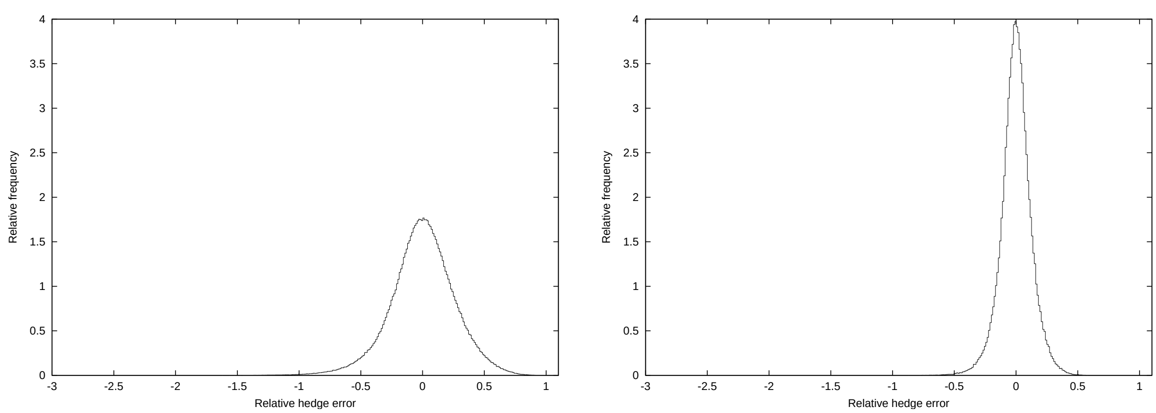



Figure 8.2: Relative frequency (y-axis) versus relative P&L of delta hedging strategies. Left: rebalance hedge once a week, right: rebalance hedge daily. American put, $T = .25$, $\sigma = .3$, $r = .06$, $\mu = .08, K = S_0 = 100$. The relative P&L is computed by dividing the actual P&L by the Black-Scholes price.

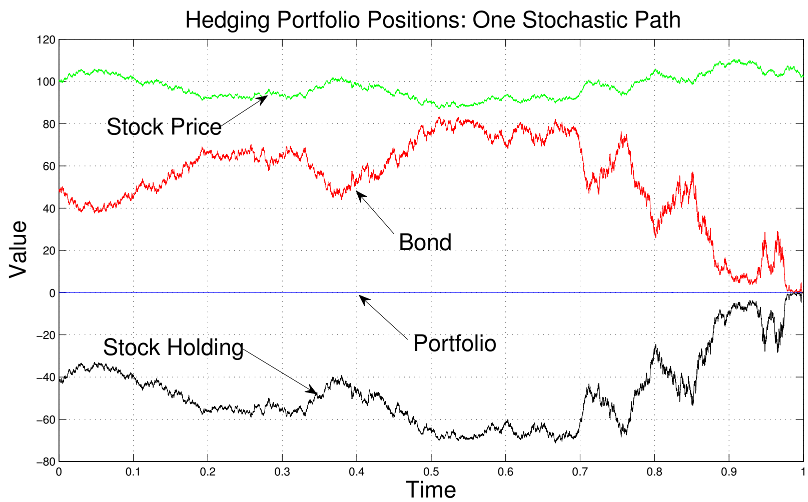



Figure 8.3: Values of components of the hedging portfolio along a single stochastic path. One year put. $K = 100$, $r = .02$, $\sigma = .20$, real world drift .10. Rebalanced 10000 times.

A gamma hedge consists of

- A short option position $-V(t)$.
- Long $\alpha^h S(t)$ shares
- Long $\beta$ another derivative security $I$.
- An amount in a risk-free bank account $B(t)$.

Now, recall that we consider $\alpha^h, \beta$ to be constant over the hedging interval (no peeking into the future), so we can regard these as constants (for the duration of the hedging interval).

The hedge portfolio $P(t)$ is then

$$P(t) = -V + \alpha^h S + \beta I + B(t)$$

Assuming that we buy and hold $\alpha^h$ shares and $\beta$ of the secondary instrument at the beginning of each hedging interval, then we require that

$$\begin{aligned} \frac{\partial P}{\partial S} &= -\frac{\partial V}{\partial S} + \alpha^h + \beta \frac{\partial I}{\partial S} = 0 \\ \frac{\partial^2 P}{\partial S^2} &= -\frac{\partial^2 V}{\partial S^2} + \beta \frac{\partial^2 I}{\partial S^2} = 0 \end{aligned} \tag{8.3}$$

Note that

- If $\beta = 0$, then we get back the usual delta hedge.
- In order for the gamma hedge to work, we need an instrument which has some gamma (the asset $S$ has second derivative zero). Hence, traders often speak of being long (positive) or short (negative) gamma, and try to buy/sell things to get gamma neutral.

So, at $t = 0$ we have

$$P(0) = 0 \Rightarrow B(0) = V(0) - \alpha_0^h S_0 - \beta_0 I_0$$

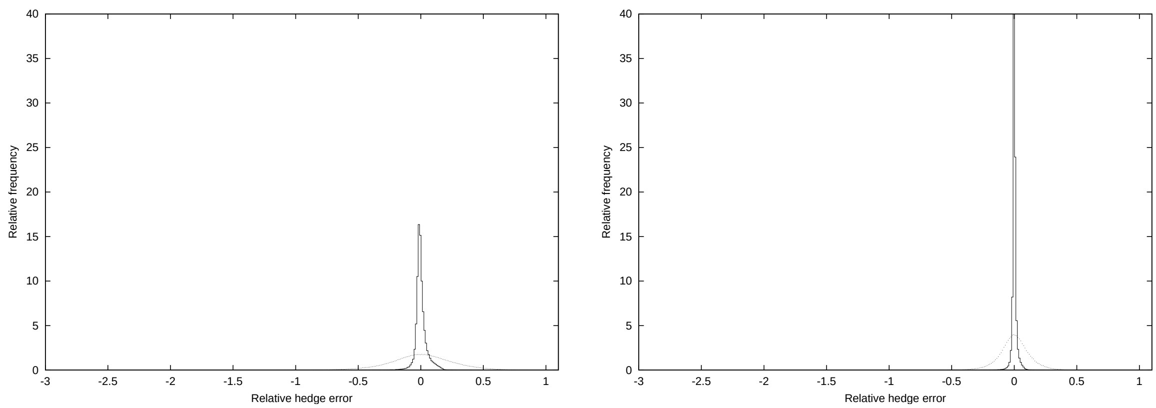



Figure 8.4: Relative frequency (y-axis) versus relative P&L of gamma hedging strategies. Left: rebalance hedge once a week, right: rebalance hedge daily. Dotted lines show the delta hedge for comparison. American put, $T = .25$, $\sigma = .3$, $r = .06$, $\mu = .08, K = 100, S_0 = 100$. Secondary instrument: European put option, same strike, $T = .5$ years. The relative P&L is computed by dividing the actual P&L by the Black-Scholes price.

The amounts $\alpha_i^h, \beta_i$ are determined by requiring that equation (8.3) hold

$$
\begin{aligned}
-(V_S)_i + \alpha_i^h + \beta_i (I_S)_i &= 0 \\
-(V_{SS})_i + \beta_i (I_{SS})_i &= 0
\end{aligned}
\tag{8.4}
$$

The bank account balance is then updated at each hedging time $t_i$ by

$$
B_i = e^{r\Delta t} B_{i-1} - S_i(\alpha_i^h - \alpha_{i-1}^h) - I_i(\beta_i - \beta_{i-1})
$$

We will consider the same example as we used in the delta hedge example. For an additional instrument, we will use a European put option written on the same underlying with the same strike price and a maturity of T=.5 years.

Figure 8.4 shows the results of gamma hedging, along with a comparison on delta hedging. In principle, gamma hedging produces a smaller variance with less frequent hedging. However, we are exposed to more *model error* in this case, since we need to be able to compute the second derivative of the theoretical price.

## 8.3 Vega Hedging

The most important parameter in the option pricing is the volatility. What if we are not sure about the value of the volatility? It is possible to assume that the volatility itself is stochastic, i.e.

$$
\begin{aligned}
dS &= \mu S dt + \sqrt{v} S dZ_1 \\
dv &= \kappa(\theta - v)dt + \sigma_v \sqrt{v} dZ_2
\end{aligned}
\tag{8.5}
$$

where $\mu$ is the expected growth rate of the stock price, $\sqrt{v}$ is its instantaneous volatility, $\kappa$ is a parameter controlling how fast $v$ reverts to its mean level of $\theta$, $\sigma_v$ is the "volatility of volatility" parameter, and $Z_1, Z_2$ are Wiener processes with correlation parameter $\rho$.

If we use the model in equation (8.5), the this will result in a two factor PDE to solve for the option price and the hedging parameters. Since there are two sources of risk $(dZ_1, dZ_2)$, we will need to hedge with the underlying asset and another option (Heston, *A closed form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Fin. Studies 6 (1993) 327-343).

Another possibility is to assume that the volatility is uncertain, and to assume that

$$
\sigma_{min} \le \sigma \le \sigma_{max},
$$

and to hedge based on a worst case (from the hedger's point of view). This results in an uncertain volatility model (Avellaneda, Levy, Paris, *Pricing and Hedging Derivative Securities in Markets with Uncertain Volatilities*, Appl. Math. Fin. 2 (1995) 77-88). This is great if you can get someone to buy this option at this price, because the hedger is always guaranteed to end up with a non-negative balance in the hedging portfolio. But you may not be able to sell at this price, since the option price is expensive (after all, the price you get has to cover the worst case scenario).

An alternative, much simpler, approach (and therefore popular in industry), is to construct a vega hedge. We assume that we know the volatility, and price the option in the usual way. Then, as with a gamma hedge, we construct a portfolio

- A short option position $-V(t)$.
- Long $\alpha^h S(t)$ shares
- Long $\beta$ another derivative security $I$.
- An amount in a risk-free bank account $B(t)$.

The hedge portfolio $P(t)$ is then

$$P(t) = -V + \alpha^h S + \beta I + B(t)$$

Assuming that we buy and hold $\alpha^h$ shares and $\beta$ of the secondary instrument at the beginning of each hedging interval, then we require that

$$\begin{aligned} \frac{\partial P}{\partial S} &= -\frac{\partial V}{\partial S} + \alpha^h + \beta \frac{\partial I}{\partial S} = 0 \\ \frac{\partial P}{\partial \sigma} &= -\frac{\partial V}{\partial \sigma} + \beta \frac{\partial I}{\partial \sigma} = 0 \end{aligned} \tag{8.6}$$

Note that if we assume that $\sigma$ is constant when pricing the option, yet do not assume $\sigma$ is constant when we hedge, this is somewhat inconsistent. Nevertheless, we can determine the derivatives in equation (8.6) numerically (solve the pricing equation for several different values of $\sigma$, and then finite difference the solutions).

In practice, we would sell the option priced using our best estimate of $\sigma$ (today). This is usually based on looking at the prices of traded options, and then backing out the volatility which gives back today's traded option price (this is the *implied volatility*). Then as time goes on, the implied volatility will likely change. We use the current implied volatility to determine the current hedge parameters in equation (8.6). Since this implied volatility has likely changed since we last rebalanced the hedge, there is some error in the hedge. However, taking into account the change in the hedge portfolio through equations (8.6) should make up for this error. This procedure is called delta-vega hedging.

In fact, even if the underlying process is a stochastic volatility, the vega hedge computed using a constant volatility model works surprisingly well (Hull and White, *The pricing of options on assets with stochastic volatilities,* J. of Finance, 42 (1987) 281-300).

## 8.4 A Stop-Loss Strategy

A simple minded idea for hedging options is based on the following idea. Suppose the writer has sold one call option with strike $K$. Assume that $S_0 < K$, i.e. the initial asset price is less than $K$. As soon as $S = K + \epsilon$, $\epsilon \ll 1$, the writer borrows $K + \epsilon$ and buys the stock. As soon as $S = K - \epsilon$, the writer sells the stock. If $S < K$ at expiry, the writer holds no stock, and does not owe the holder of the option anything. If $S > K$ at expiry, the writer sells the stock at price $S$, gives the holder $S - K$, and pays back the loan of $K$ used to buy the stock.

This looks like a good strategy. As far as a loan to buy the stock, the worst case for the financing cost of this would be $K(e^{rT} - 1)$, which will be small if $rT \ll 1$. In addition, note that buying occurs at price $K + \epsilon$, and selling occurs at price $K - \epsilon$, resulting in a transaction cost of $2\epsilon$ for each buy-sell trade. But if

| | |
|---|---|
| $T$ | .3846 |
| $\sigma$ | .2 |
| $\mu$ (real world drift) | .13 |
| $S_0$ (initial price) | 49 |
| $r$ | $= .05$ |
| $K$ | 50 |
| Option Type | Call |
| Early Exercise | No |

Table 8.1: Data used in the stop loss hedging strategy test.

| Hedging Frequency | 100 | 200 | 400 | 800 |
|---|---|---|---|---|
| Mean | -0.047345 | -0.04373 | -0.040295 | -0.040762 |
| Standard Deviation | 0.78175 | 0.77473 | 0.76954 | 0.77247 |
| VAR (95%) | -1.523 | -1.512 | -1.5059 | -1.5043 |
| CVAR (95%) | -2.0122 | -1.9983 | -1.9847 | -1.9974 |

Table 8.2: Statistics for the profit and loss, relative to the Black-Scholes price of 2.4005. Assumes 2.4005 received in cash initially.

we monitor the strategy closely, we can make $\epsilon$ as small as we like. Consequently, this strategy should cost less than the Black-Scholes price. We should be able to make millions.

Consider the following example, from the Hull book, with the data in Table 8.2. A Monte Carlo strategy was used to simulate this strategy (80,000 simulations). Table 8.1 shows the results in terms of the relative hedging error (8.1) for various numbers of rebalancing times. It is assumed that the hedger receives the Black-Scholes price in cash initially. Hence, a positive mean would indicate a relative profit compared to the Black-Scholes strategy.

From this table, we can see that the expected cost of the strategy is actually larger than the Black-Scholes strategy (the expected relative $P\&L$ is negative), and no matter how many times the hedge is rebalanced, the standard deviation seems to be about .77. This strategy is very bad.

Why does this not work? No matter where we start, there is a finite probability of crossing the strike price. The problem can be traced to the fact that we buy at $K+\epsilon$ and sell at $K-\epsilon$. As we make the hedging interval smaller, we can make $\epsilon$ smaller, but Brownian motion is infinitely jagged on any scale. Hence the number of times we buy and sell as the asset crosses $S = K$ becomes larger as the hedging interval becomes smaller. Recall from equation (2.30) that the total distance traveled by a particle undergoing Brownian motion, for a finite time, is infinite.

The probability density for the relative hedging error (8.1) is shown in Figure 8.5.

### 8.4.1 Profit and Loss: probability density, VAR and CVAR

Note that Figure 8.5 shows the probability density instead of the relative frequency. This is a more desirable way to show the results.

Suppose $P^i$ represents the $P\&L$ from the $i^{th}$ simulation. Then, we can construct a histogram of the results by first specifying a series of bins, and counting the number of occurrences in each bin. Suppose the $k^{th}$ bin lies in the interval $[a, b]$ of the x-axis. Then a histogram will show a bar indicating the number of occurrences of the $P$ in the interval $[a, b]$. Now

$$Prob[a \leq x \leq b] \simeq \frac{\text{Number of occurrences in bin } [a, b]}{\text{Total number of simulations}} . \tag{8.7}$$

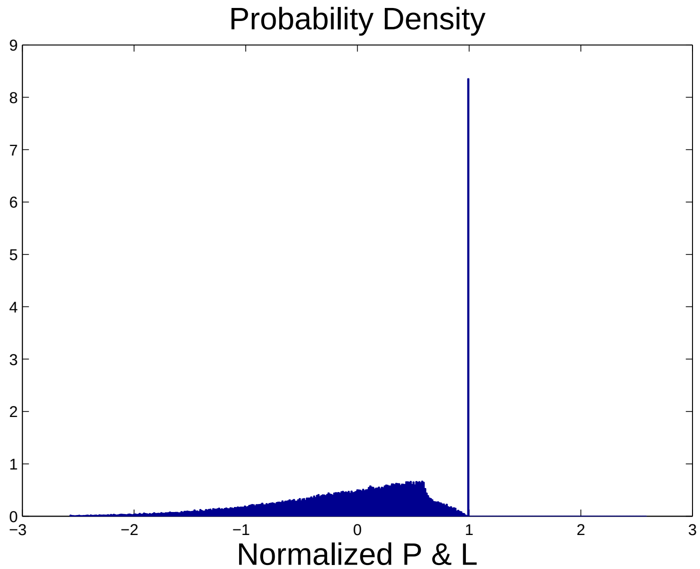



Figure 8.5: Probability density (y-axis) versus relative P&L of the stop loss strategy. The relative P&L is computed by dividing the actual P&L by the Black-Scholes price.

Let $p(x)\ dx$ be the probability of $P \in [x, x + dx]$, so

$$p(a)(b-a) \simeq \frac{\text{Number of occurrences in bin } [a,b]}{\text{Total number of simulations}} \ , \tag{8.8}$$

so that

$$p(a) \simeq \frac{1}{(b-a)} \frac{\text{Number of occurrences in bin } [a,b]}{\text{Total number of simulations}} \ . \tag{8.9}$$

Note that the Matlab functions `histc(X,EDGES)` , `bar(X,Y, 'hist')` are useful for producing probability density plots.

A useful way to measure tail risk, is to compute the Value at Risk (VAR) and the Conditional VAR (CVAR). Given an approximate probability density $p(x)$, the y% VAR is the point on the x-axis where

$$\int_{VAR}^{\infty} p(x)\ dx = \frac{y}{100} \ . \tag{8.10}$$

Typically, $y = 95\%, 99\%$. VAR can be interpreted as *" y% of the time, we end up with at least VAR"*.

Another measure of tail risk is CVAR. This is defined as

$$CVAR \quad = \quad \frac{\int_{-\infty}^{VAR} x\ p(x)\ dx}{\int_{-\infty}^{VAR} p(x)\ dx} \ . \tag{8.11}$$

The $y$% CVAR can be interpreted as the mean of the worst $(100 - y)$% scenarios. In Matlab, VAR can be easily computed by sorting the array of outcomes $P^i$. CVAR is then computed by averaging. It is interesting to note that CVAR can be computed in another way without first computing VAR.

### 8.4.2 Another way of computing CVAR

It is possible to compute CVAR without first computing VAR. We can compute CVAR in terms of a nice optimization problem, which makes the CVAR computation robust.

Let $VAR_\alpha$ be the value of VAR at the confidence level $\alpha$. For example, typically $\alpha = .95$ or $\alpha = .99$. If $p(x)$ is the density function of the $P\&L$ at time $T$, then

$$\int_{-\infty}^{VAR_\alpha} p(x)\ dx \quad = \quad 1 - \alpha \ , \tag{8.12}$$

and

$$\begin{aligned} CVAR_\alpha &= \frac{\int_{-\infty}^{VAR_\alpha} x\ p(x)\ dx}{1-\alpha} \\ &= \frac{\int_{-\infty}^{VAR_\alpha} x\ p(x)\ dx}{\int_{-\infty}^{VAR_\alpha} p(x)\ dx} \ . \end{aligned} \tag{8.13}$$

Now, consider the function

$$\begin{aligned} f(\alpha) &= \sup_y \left\{ y + \frac{1}{1-\alpha} E[\min(x-y, 0)] \right\} \\ &= \sup_y \left\{ y + \frac{1}{1-\alpha} \int_{-\infty}^{y} (x-y) p(x)\ dx \right\} \\ &= \sup_y g(y, \alpha) \ , \end{aligned} \tag{8.14}$$

where

$$g(y, \alpha) \quad = \quad y + \frac{1}{1-\alpha} \int_{-\infty}^{y} (x-y) p(x)\ dx \ , \tag{8.15}$$

and where $x = X(T)$, i.e. the value of the $P\&L$ at $t = T$, and $E[\cdot]$ is the expectation. Consequently,

$$\frac{\partial g}{\partial y} \quad = \quad 1 - \frac{1}{1-\alpha} \int_{-\infty}^{y} p(x)\ dx \ . \tag{8.16}$$

Now, note that if $y = VAR_\alpha$

$$1 - \frac{1}{1-\alpha} \int_{-\infty}^{VAR_\alpha} p(x)\ dx = 0 \ , \tag{8.17}$$

from equation (8.12), hence the point $y = VAR_\alpha$ is a local maximum or minimum of $g(y, \alpha)$. If $y < VAR_\alpha$, then

$$\int_{-\infty}^{y} p(x)\ dx \leq (1-\alpha) \ , \tag{8.18}$$

hence (from equations (8.16 and 8.18))

$$\frac{\partial g}{\partial y} \quad \geq \quad 0 \quad ; y < VAR_\alpha \ , \tag{8.19}$$

and similarly, if $y > VAR_\alpha$, then

$$\int_{-\infty}^{y} p(x)\ dx \geq (1-\alpha) \ , \tag{8.20}$$

hence from equations (8.16) and (8.20),

$$\frac{\partial g}{\partial y} \quad \leq \quad 0 \quad ; y > VAR_\alpha \ . \tag{8.21}$$

As a result, the point $y = VAR_\alpha$ is a global maximum of $g(y, \alpha)$, so that (noting equation (8.13))

$$
\begin{aligned}
\sup_y g(y,\alpha) &= VAR_\alpha + \frac{1}{1-\alpha}\int_{-\infty}^{VAR_\alpha} (x - VAR_\alpha) p(x)\ dx \\
&= \frac{1}{1-\alpha}\int_{-\infty}^{VAR_\alpha} x\ p(x)\ dx \\
&= \frac{\int_{-\infty}^{VAR_\alpha} x\ p(x)\ dx}{\int_{-\infty}^{VAR_\alpha} p(x)\ dx} \\
&= CVAR_\alpha\ . \qquad (8.22)
\end{aligned}
$$

## 8.5 Collateralized deals

To avoid counterparty risk, it is common to enter into a Collateralized deal.

Consider a seller of an option who hedges (H), and a buyer (B) who does not hedge. H sells an option worth $V(t)$ to B.

We assume that everyone borrows/lends at the rate $r$, but the arguments below can be generalized to different rates.

We assume the existence of a collateral account $\Pi_C(t)$. Mathematically, the collateral account is a virtual bookkeeping entry, since we assume that *rehypothecation* is allowed, so that any cash flows which flow due to collateral posting can be re-used by the receiver of these cash flows, to, for example, reduce borrowing.

If the collateral account is negative, then this means that we owe this amount to the poster of the collateral.

We assume full collateralization. This can also be generalized to the partial collateralization case.

### 8.5.1 Hedger

The hedger's portfolio is

$$
\begin{aligned}
P_H &= -V + \alpha S + \mathcal{B}_H + \Pi_C \\
& V = \text{ Price of option} \\
& S = \text{ price of underlying} \\
& \alpha = \text{ number of units of underlying} \\
& \mathcal{B}_H = \text{ hedger's bank account} \\
& \Pi_C = \text{ collateral account} \qquad (8.23)
\end{aligned}
$$

For definiteness, we will consider an example where $V(t) > 0$ and $\Pi_C > 0$, i.e. the case where H sells an option worth $V$ to B. However, this need not be the case in general.

At $t = 0$, B pays H for the option:

$$
\begin{aligned}
P_H &= -V(0) + 0 \cdot S(0) + \mathcal{B}_H(0) + \Pi_C(0) \\
& \mathcal{B}_H(0) = +V(0) \\
& \Pi_C(0) = 0 \\
& \alpha(0) = 0 \qquad (8.24)
\end{aligned}
$$

For all times $t + \Delta t$ (this is also valid for $t = 0, \Delta t = 0^+$, so that the equations below are valid for $t + \Delta t = 0^+$.

- H receives interest on the collateral and gets interest on $\mathcal{B}_H$

$$
\mathcal{B}'_H = \mathcal{B}_H(t) e^{r\Delta t} + \Pi_C(t) r \Delta t \qquad (8.25)
$$

- H rebalances the delta hedge by buying $\alpha(t+\Delta t)$ shares at price $S(t+\Delta t)$,

$$\mathcal{B}''_H = \mathcal{B}'_H - (\alpha(t+\Delta t) - \alpha(t))S(t+\Delta t) \tag{8.26}$$

- H transfers the change in collateral to B. H first transfers cash to B, and then the virtual bookkeeping account $\Pi_C$ is updated.

$$\begin{aligned} \mathcal{B}_H(t+\Delta t) &= \mathcal{B}''_H - (V(t+\Delta t) - \Pi_C(t)) \\ &= \mathcal{B}''_H - (V(t+\Delta t) - V(t)) \\ \Pi_C(t+\Delta t) &= \Pi_C(t) + (V(t+\Delta t) - \Pi_C(t)) \\ &= V(t+\Delta t) \end{aligned} \tag{8.27}$$

Note that at $t = 0^+$, H transfers the value of the option back to B, then borrows $-\alpha(0)S(0)$ to start the delta hedge.

$$\begin{aligned} \alpha(0^+) &= V_S(0) \\ \mathcal{B_H} &= -\alpha(0^+)S(0) \\ \Pi_C(0^+) &= +V(0) \end{aligned} \tag{8.28}$$

so that

$$P_H(0^+) = -V(0) + \underbrace{\alpha(0^+)S(0)}_{share\ position} - \underbrace{\alpha(0^+)S(0)}_{\mathcal{B_H}} + \underbrace{V(0)}_{\Pi_C} \tag{8.29}$$

so that $P_H = 0$. This will hold at all future times, after cash flows have been exchanged. From H's point of view, if $B$ defaults, then the long collateral position $\Pi_C$ cancels the short option position,

$$P_H(t) = \overbrace{-V(t) + \underbrace{V(t)}_{\Pi_C}}^{net\ position\ owed\ to\ H} + \underbrace{\alpha(t)S(t)}_{shares} + \underbrace{(-\alpha(t)S(t))}_{\mathcal{B}_H} \tag{8.30}$$

and the stock position can be liquidated to cancel the bank loan. Note that all the above arguments are valid if $V < 0$, i.e. we can either be long or short the option.

### 8.5.2 Buyer B

We assume that the buyer does not hedge. The buyer's portfolio is

$$\begin{aligned} P_H &= +V + \mathcal{B}_B - \Pi_C \\ & V = \text{ Price of option} \\ & \mathcal{B}_B = \text{ buyer's bank account} \\ & \Pi_C = \text{ collateral account} \end{aligned} \tag{8.31}$$

For all times $t + \Delta t$ (this is also valid for $t = 0, \Delta t = 0^+$, so that this also holds for $t + \Delta t = 0^+$.

- B gets interest on $\mathcal{B}_B$ and pays interest to H on the collateral

$$\mathcal{B}'_B = \mathcal{B}_B(t)e^{r\Delta t} - \Pi_C(t)r\Delta t \tag{8.32}$$

- B receives the change in collateral from H, which is immediately deposited in $\mathcal{B}_B$. The collateral bookkeeping entry $\Pi_C$ is updated.

$$\begin{aligned} \mathcal{B}_B(t+\Delta t) &= \mathcal{B}'_B + (V(t+\Delta t) - \Pi_C(t)) \\ &= \mathcal{B}'_B + (V(t+\Delta t) - V(t))] \\ \Pi_C(t+\Delta t) &= \Pi_C(t) + (V(t+\Delta t) - \Pi_C(t)) \\ &= V(t+\Delta t) \end{aligned} \tag{8.33}$$

We also write equations (8.32 - 8.33) as

$$\mathcal{B}_B(t+\Delta t) = \mathcal{B}_B(t) + \mathcal{B}_B(t) r \Delta t - V(t) r \Delta t + (V(t+\Delta t) - V(t)) \ . \tag{8.34}$$

Note that at $t = 0$, we have

$$\begin{aligned} \mathcal{B}_B &= -V(0) \\ \Pi_C(0^+) &= 0 \end{aligned} \tag{8.35}$$

i.e. the B borrows from the bank account to buy the option from H, so that

$$P_B(0) = +V(0) + \underbrace{(-V(0))}_{\mathcal{B}_B} + \underbrace{0}_{\Pi_C} \tag{8.36}$$

At $t = 0^+$, H posts collateral to B. B then immediately uses the cash to payoff the loan, and posts a virtual amount to the collateral account, i.e. this amount is owed to H.

$$P_B(0^+) = +V(0) + \underbrace{0}_{\mathcal{B}_B} + \underbrace{(-V(0))}_{\Pi_C} \tag{8.37}$$

so that $P_B = 0$.

Note that $\mathcal{B}_B$ satisfies (from equation (8.34))

$$\frac{d\mathcal{B}_B}{dt} = r\mathcal{B}_B - \underbrace{rV}_{collateral\ interest} + \underbrace{\frac{dV}{dt}}_{change\ in\ posted\ collateral} \tag{8.38}$$

It can be easily verified that the solution to this equation is ($\mathcal{B}_B(0^+) = 0$)

$$\mathcal{B}_B(t) = V(t) - V(0)e^{rt} \tag{8.39}$$

At any time, after the cash flows have been exchanged, if H defaults, then the collateral account cancels out the long option position,

$$P_B(t) = \overbrace{+V(t) + \underbrace{(-V(t))}_{\Pi_C}}^{net\ owed\ to\ B} + \overbrace{\underbrace{(V(t) - V(0)e^{rt})}_{\mathcal{B}_B}}^{net\ value\ to\ B} \tag{8.40}$$

and B is left with $V(t) - V(0)e^{rt}$, i.e. the value of the option less the cost of borrowing to fund the initial option price. This is effectively the transaction which occurred at $t = 0^+$, so B does not care if H defaults.

# 9 Jump Diffusion

Recall that if

$$dS = \mu S dt + \sigma S\ dZ \tag{9.1}$$

then from Ito's Lemma we have

$$d[\log S] = [\mu - \frac{\sigma^2}{2}]\ dt + \sigma\ dZ. \tag{9.2}$$

Now, suppose that we observe asset prices at discrete times $t_i$, i.e. $S(t_i) = S_i$, with $\Delta t = t_{i+1} - t_i$. Then from equation (9.2) we have

$$\begin{aligned} \log S_{i+1} - \log S_i &= \log(\frac{S_{i+1}}{S_i}) \\ &\simeq [\mu - \frac{\sigma^2}{2}] \, \Delta t + \sigma\phi\sqrt{\Delta t} \end{aligned} \tag{9.3}$$

where $\phi$ is $N(0,1)$. Now, if $\Delta t$ is sufficiently small, then $\Delta t$ is much smaller than $\sqrt{\Delta t}$, so that equation (9.3) can be approximated by

$$\begin{aligned} \log(\frac{S_{i+1} - S_i + S_i}{S_i}) &= \log(1 + \frac{S_{i+1} - S_i}{S_i}) \\ &\simeq \sigma\phi\sqrt{\Delta t}. \end{aligned} \tag{9.4}$$

Define the return $R_i$ in the period $t_{i+1} - t_i$ as

$$R_i = \frac{S_{i+1} - S_i}{S_i} \tag{9.5}$$

so that equation (9.4) becomes

$$\log(1 + R_i) \simeq R_i = \sigma\phi\sqrt{\Delta t}.$$

Consequently, a plot of the discretely observed returns of $S$ should be normally distributed, if the assumption (9.1) is true. In Figure 9.1 we can see a histogram of log monthly returns from the $TSX$ composite index for the period $1979 - 20124$. The histogram has been scaled to zero mean and unit standard deviation. A standard normal distribution is also shown. Note that for real data, there is a higher peak, and fatter tails than the normal distribution. This means that there is higher probability of zero return, or a large gain or loss compared to a normal distribution.

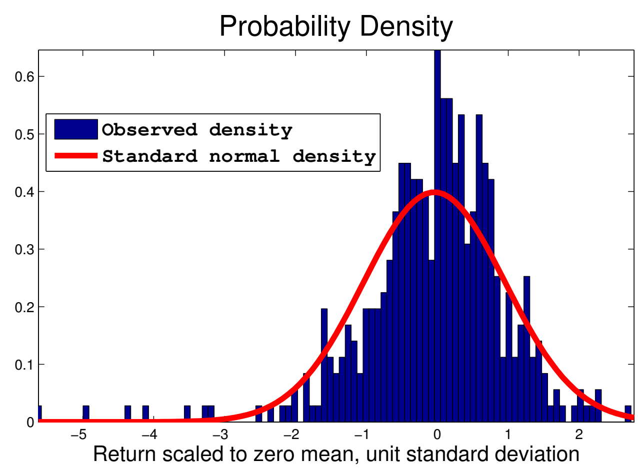



Figure 9.1: Probability density functions for the TSX composite log monthly returns $1979 - 2014$, scaled to zero mean and unit standard deviation and the standardized Normal distribution.

As $\Delta t \to 0$, Geometric Brownian Motion (equation (9.1)) assumes that the probability of a large return also tends to zero. The amplitude of the return is proportional to $\sqrt{\Delta t}$, so that the tails of the distribution become unimportant.

But, in real life, we can sometimes see very large returns (positive or negative) in small time increments. It therefore appears that Geometric Brownian Motion (GBM) is missing something important.

## 9.1 The Poisson Process

Consider a process where most of the time nothing happens (contrast this with Brownian motion, where some changes occur at any time scale we look at), but on rare occasions, a jump occurs. The jump size does not depend on the time interval, but the probability of the jump occurring does depend on the interval.

More formally, consider the process $dq$ where, in the interval $[t, t + dt]$,

$$\begin{aligned} dq &= 1 \ ; \quad \text{with probability } \lambda dt \\ &= 0 \ ; \quad \text{with probability } 1 - \lambda dt. \end{aligned} \tag{9.6}$$

Note, once again, that size of the Poisson outcome does not depend on $dt$. Also, the probability of a jump occurring in $[t, t + dt]$ goes to zero as $dt \to 0$, in contrast to Brownian motion, where some movement always takes place (the probability of movement is constant as $dt \to 0$), but the size of the movement tends to zero as $dt \to 0$. For future reference, note that

$$\begin{aligned} E[dq] &= \lambda\ dt \cdot 1 + (1 - \lambda\ dt) \cdot 0 \\ &= \lambda\ dt \end{aligned} \tag{9.7}$$

and

$$\begin{aligned} Var(dq) &= E[(dq - E[dq])^2] \\ &= E[(dq - \lambda\ dt)^2] \\ &= (1 - \lambda\ dt)^2 \cdot \lambda\ dt + (0 - \lambda\ dt)^2 \cdot (1 - \lambda\ dt) \\ &= \lambda\ dt + O((dt)^2)\ . \end{aligned} \tag{9.8}$$

Now, suppose we assume that, along with the usual GBM, occasionally the asset jumps, i.e. $S \to JS$, where $J$ is the size of a (proportional) jump. We will restrict $J$ to be non-negative.

Suppose a jump occurs in $[t, t + dt]$, with probability $\lambda dt$. Let's write this jump process as an SDE, i.e.

$$[dS]_{jump} = (J - 1)S\ dq$$

since, if a jump occurs

$$\begin{aligned} S_{after\ jump} &= S_{before\ jump} + [dS]_{jump} \\ &= S_{before\ jump} + (J - 1)S_{before\ jump} \\ &= JS_{before\ jump} \end{aligned} \tag{9.9}$$

which is what we want to model. So, if we have a combination of GBM and a rare jump event, then

$$dS = \mu S\ dt + \sigma S\ dZ + (J - 1)S\ dq \tag{9.10}$$

Assume that the jump size has some known probability density $g(J)$, i.e. given that a jump occurs, then the probability of a jump in $[J, J + dJ]$ is $g(J)\ dJ$, and

$$\int_{-\infty}^{+\infty} g(J)\ dJ = \int_0^{\infty} g(J)\ dJ = 1 \tag{9.11}$$

since we assume that $g(J) = 0$ if $J < 0$. For future reference, if $f = f(J)$, then the expected value of $f$ is

$$E[f] = \int_0^{\infty} f(J) g(J)\ dJ\ . \tag{9.12}$$

The process (9.10) is basically geometric Brownian motion (a continuous process) with rare discontinuous jumps. Some example realizations of jump diffusion paths are shown in Figure 9.2.

Figure 9.3 shows the price followed by a listed drug company. Note the extreme price changes over very small periods of time.

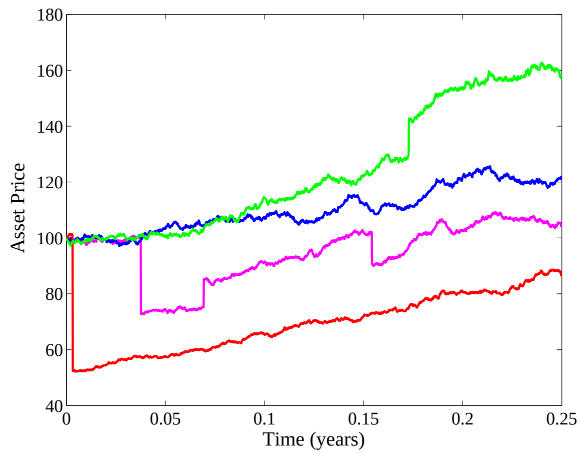



Figure 9.2: Some realizations of a jump diffusion process which follows equation (9.10).

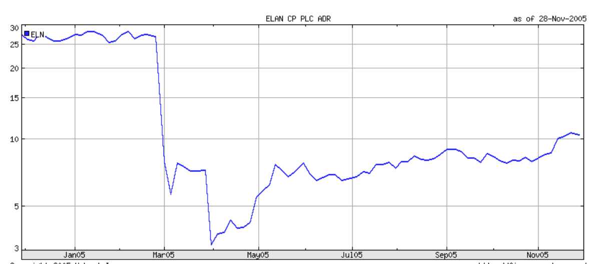



Figure 9.3: Actual price of a drug company stock. Compare with simulation of a jump diffusion in Figure 9.2.

## 9.2 The Jump Diffusion Pricing Equation

Now, form the usual hedging portfolio

$$P = V - \alpha S . \tag{9.13}$$

Now, consider

$$[dP]_{total} = [dP]_{Brownian} + [dP]_{jump} \tag{9.14}$$

where, from Ito's Lemma

$$[dP]_{Brownian} = [V_t + \frac{\sigma^2 S^2}{2} V_{SS}]dt + [V_S - \alpha](\mu S\ dt + \sigma S\ dZ) \tag{9.15}$$

and, noting that the jump is of finite size,

$$[dP]_{jump} = [V(JS, t) - V(S, t)]\ dq - \alpha(J - 1)S\ dq . \tag{9.16}$$

If we hedge the Brownian motion risk, by setting $\alpha = V_S$, then equations (9.14-9.16) give us

$$dP = [V_t + \frac{\sigma^2 S^2}{2} V_{SS}]dt + [V(JS, t) - V(S, t)]dq - V_S(J - 1)S\ dq . \tag{9.17}$$

So, we still have a random component ($dq$) which we have not hedged away. Let's take the expected value of this change in the portfolio, e.g.

$$E(dP) = [V_t + \frac{\sigma^2 S^2}{2} V_{SS}]dt + E[V(JS, t) - V(S, t)]E[dq] - V_S S E[J - 1]E[dq] \tag{9.18}$$

where we have assumed that probability of the jump and the probability of the size of the jump are independent. Defining $E(J - 1) = \kappa$, then we have that equation (9.18) becomes

$$E(dP) = [V_t + \frac{\sigma^2 S^2}{2} V_{SS}]dt + E[V(JS, t) - V(S, t)]\lambda\ dt - V_S S\kappa\lambda\ dt . \tag{9.19}$$

Now, we make a rather interesting assumption. Assume that an investor holds a diversified portfolio of these hedging portfolios, for many different stocks. If we make the rather dubious assumption that these jumps for different stocks are uncorrelated, then the variance of this portfolio of portfolios is small, hence there is little risk in this portfolio. Hence, the expected return should be

$$E[dP] = rP\ dt . \tag{9.20}$$

Now, equating equations (9.19 and (9.20) gives

$$V_t + \frac{\sigma^2 S^2}{2} V_{SS} + V_S[rS - S\kappa\lambda] - (r + \lambda)V + E[V(JS, t)]\lambda = 0 . \tag{9.21}$$

Using equation (9.12) in equation (9.21) gives

$$V_t + \frac{\sigma^2 S^2}{2} V_{SS} + V_S[rS - S\kappa\lambda] - (r + \lambda)V + \lambda \int_0^\infty g(J)V(JS, t)\ dJ\ = 0 . \tag{9.22}$$

Equation (9.22) is a Partial Integral Differential Equation (PIDE).

A common assumption is to assume that $g(J)$ is log normal,

$$g(J) = \frac{\exp\left(-\frac{(\log(J) - \hat{\mu})^2}{2\gamma^2}\right)}{\sqrt{2\pi}\gamma J}. \tag{9.23}$$

where, some algebra shows that

$$E(J-1) = \kappa = \exp(\hat{\mu} + \gamma^2/2) - 1 \ . \tag{9.24}$$

Now, what about our dubious assumption that jump risk was diversifiable? In practice, we can regard $\sigma, \hat{\mu}, \gamma, \lambda$ as parameters, and fit them to observed option prices. If we do this, (see L. Andersen and J. Andreasen, Jump-Diffusion processes: Volatility smile fitting and numerical methods, Review of Derivatives Research (2002), vol 4, pages 231-262), then we find that $\sigma$ is close to historical volatility, but that the fitted values of $\lambda, \hat{\mu}$ are at odds with the historical values. The fitted values seem to indicate that investors are pricing in larger more frequent jumps than has been historically observed. In other words, actual prices seem to indicate that investors do require some compensation for jump risk, which makes sense. In other words, these parameters contain a market price of risk.

Consequently, our assumption about jump risk being diversifiable is not really a problem if we fit the jump parameters from market (as opposed to historical) data, since the market-fit parameters will contain some effect due to risk preferences of investors.

One can be more rigorous about this if you assume some utility function for investors. See (Alan Lewis, *Fear of jumps,* Wilmott Magazine, December, 2002, pages 60-67) or (V. Naik, M. Lee, *General equilibrium pricing of options on the market portfolio with discontinuous returns,* The Review of Financial Studies, vol 3 (1990) pages 493-521.)

## 9.3 An Alternate Derivation of the Pricing Equation for Jump Diffusion

We will give a pure hedging argument in this section, in order to derive the PIDE for jump diffusion. Initially, we suppose that there is only one possible jump size $J$, i.e. after a jump, $S \rightarrow JS$, where $J$ is a known constant. Suppose

$$dS = a(S,t)dt + b(S,t)dZ + (J-1)S\ dq, \tag{9.25}$$

where $dq$ is the Poisson process. Consider a contract on $S$, $F(S,t)$, then

$$dF = \left[ aF_S + \frac{b^2}{2} F_{SS} + F_t \right] dt + bF_S dZ + [F(JS,t) - F(S,t)]\ dq, \tag{9.26}$$

or, in more compact notation

$$\begin{aligned} dF &= \mu\ dt + \sigma^*\ dZ + \Delta F\ dq \\ & \mu = aF_S + \frac{b^2}{2} F_{SS} + F_t \\ & \sigma^* = bF_S \\ & \Delta F = [F(JS,t) - F(S,t)] \ . \end{aligned} \tag{9.27}$$

Now, instead of hedging with the underlying asset, we will hedge one contract with another. Suppose we have three contracts $F_1, F_2, F_3$ (they could have different maturities for example).

Consider the portfolio $\Pi$

$$\Pi = n_1 F_1 + n_2 F_2 + n_3 F_3 \tag{9.28}$$

so that

$$\begin{aligned} d\Pi &= n_1\ dF_1 + n_2\ dF_2 + n_3\ dF_3 \\ &= n_1(\mu_1\ dt + \sigma_1^*\ dZ + \Delta F_1\ dq) \\ & \quad + n_2(\mu_2\ dt + \sigma_2^*\ dZ + \Delta F_2\ dq) \\ & \quad + n_3(\mu_3\ dt + \sigma_3^*\ dZ + \Delta F_3\ dq) \\ &= (n_1\mu_1 + n_2\mu_2 + n_3\mu_3)\ dt \\ & \quad + (n_1\sigma_1^* + n_2\sigma_2^* + n_3\sigma_3^*)\ dZ \\ & \quad + (n_1\Delta F_1 + n_2\Delta F_2 + n_3\Delta F_3)\ dq \ . \end{aligned} \tag{9.29}$$

Eliminate the random terms by setting

$$
\begin{aligned}
(n_1 \Delta F_1 + n_2 \Delta F_2 + n_3 \Delta F_3) &= 0 \\
(n_1 \sigma_1^* + n_2 \sigma_2^* + n_3 \sigma_3^*) &= 0 \ .
\end{aligned} \tag{9.30}
$$

This means that the portfolio is riskless, hence

$$
d\Pi = r\Pi \ dt \ , \tag{9.31}
$$

hence (using equations (9.29-9.31))

$$
(n_1 \mu_1 + n_2 \mu_2 + n_3 \mu_3) = (n_1 F_1 + n_2 F_2 + n_3 F_3) r \ . \tag{9.32}
$$

Putting together equations (9.30) and (9.32), we obtain

$$
\begin{bmatrix} \sigma_1^* & \sigma_2^* & \sigma_3^* \\ \Delta F_1 & \Delta F_2 & \Delta F_3 \\ \mu_1 - rF_1 & \mu_2 - rF_2 & \mu_3 - rF_3 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{9.33}
$$

Equation (9.33) has a nonzero solution only if the rows are linearly dependent. There must be $\lambda_B(S,t), \lambda_J(S,t)$ such that

$$
\begin{aligned}
(\mu_1 - rF_1) &= \lambda_B \sigma_1^* - \lambda_J \Delta F_1 \\
(\mu_2 - rF_2) &= \lambda_B \sigma_2^* - \lambda_J \Delta F_2 \\
(\mu_3 - rF_3) &= \lambda_B \sigma_3^* - \lambda_J \Delta F_3 \ .
\end{aligned} \tag{9.34}
$$

(We will show later on that $\lambda_J \geq 0$ to avoid arbitrage). Dropping subscripts, we have

$$
(\mu - rF) = \lambda_B \sigma^* - \lambda_J \Delta F \tag{9.35}
$$

and substituting the definitions of $\mu, \sigma^*, \Delta F$, from equations (9.27), we obtain

$$
F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF + \lambda_J [F(JS,t) - F(S,t)] = 0 \ . \tag{9.36}
$$

Note that $\lambda_J$ will not be the real world intensity of the Poisson process, but $J$ will be the real world jump size.

In the event that, say, $F_3 = S$ is a traded asset, we note that in this case

$$
\begin{aligned}
\sigma_3^* &= b \\
\mu_3 &= a \\
\Delta F_3 &= (J-1) S \ .
\end{aligned} \tag{9.37}
$$

From equation (9.34) we have that

$$
(\mu_3 - rF_3) = \lambda_B \sigma_3^* - \lambda_J \Delta F_3 \ , \tag{9.38}
$$

or, using equation (9.37),

$$
a - \lambda_B b = rS - \lambda_J (J-1) S \ . \tag{9.39}
$$

Substituting equation (9.39) into equation (9.36) gives

$$
F_t + \frac{b^2}{2} F_{SS} + [r - \lambda_J (J-1)] S F_S - rF + \lambda_J [F(JS,t) - F(S,t)] = 0 \ . \tag{9.40}
$$

Note that equation (9.36) is valid if the underlying asset cannot be used to hedge, while equation (9.40) is valid only if the underlying asset can be used as part of the hedging portfolio.

Let $\tau = T - t$, and set $a = 0, b = 0, r = 0$ in equation (9.36), giving

$$F_\tau = \lambda_J [F(JS, \tau) - F(S, \tau)] \ . \tag{9.41}$$

Now, suppose that

$$\begin{aligned} F(S, \tau = 0) &= 0 \ ; \quad \text{if } S \geq K \\ &= 1 \ ; \quad \text{if } S < K \ . \end{aligned} \tag{9.42}$$

Now, consider the asset value $S^* > K$, and let $J = K/(2 * S^*)$. Imagine solving equation (9.41) to an infinitesimal time $\tau = \epsilon \ll 1$. We will obtain the following value for $F$,

$$F(S^*, \epsilon) \simeq \epsilon \lambda_J \ . \tag{9.43}$$

Since the payoff is nonnegative, we must have $\lambda_J \geq 0$ to avoid arbitrage.

Now, suppose that there are a finite number of jump states, i.e. after a jump, the asset may jump to to any state $J_i S$

$$S \to J_i S \ ; \quad i = 1, ..., n \ . \tag{9.44}$$

Repeating the above arguments, now using $n + 2$ hedging instruments in the hedging portfolio

$$\Pi = \sum_{i=1}^{i=n+2} n_i F_i \tag{9.45}$$

so that the diffusion and jumps are hedged perfectly, we obtain the following PDE

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF + \sum_{i=1}^{i=n} \lambda_J^i [F(J_i S, t) - F(S, t)] = 0 \ . \tag{9.46}$$

If we can use the underlying to hedge, then we get the analogue of equation (9.40)

$$F_t + \frac{b^2}{2} F_{SS} + (rS - \sum_{i=1}^{i=n} \lambda_J^i S(J_i - 1)) F_S - rF + \sum_{i=1}^{i=n} \lambda_J^i [F(J_i S, t) - F(S, t)] = 0 \ . \tag{9.47}$$

Now, let

$$\begin{aligned} p(J_i) &= \frac{\lambda_J^i}{\sum_{i=1}^{i=n} \lambda_J^i} \\ \lambda^* &= \sum_{i=1}^{i=n} \lambda_J^i \end{aligned} \tag{9.48}$$

then we can write equation (9.46) as

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF + \lambda^* \sum_{i=1}^{i=n} p(J_i) [F(J_i S, t) - F(S, t)] = 0 \ . \tag{9.49}$$

Note that since $\lambda_J^i \geq 0$, $p(J_i) \geq 0$, and $\lambda^* \geq 0$.

Taking the limit as the number of jump states tends to infinity, then $p(J)$ tends to a continuous distribution, so that equation (9.49) becomes

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF + \lambda^* \int_0^\infty p(J)[F(JS,t) - F(S,t)] \; dJ = 0 \; . \tag{9.50}$$

It is convenient to rewrite equation (9.50) in a slightly different form. Suppose we redefine $\lambda_B$ as follows

$$\lambda_B \quad = \quad \lambda'_B + \lambda^* \frac{E[J-1]S}{b} \tag{9.51}$$

where

$$E[J-1] = \int_0^\infty p(J)(J-1) \; dJ \; . \tag{9.52}$$

Substituting equation (9.51) into equation (9.50) gives

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda'_B b - \lambda^* E[J-1]S) F_S - rF + \lambda^* \int_0^\infty p(J)[F(JS,t) - F(S,t)] \; dJ = 0 \; . \tag{9.53}$$

Note that in the case that $F_{SS} = 0$ (which would normally be the case for $S \to \infty$), then equation (9.53) reduces to

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF = 0 \; , \tag{9.54}$$

so that the term $\lambda^* E[J-1]S$ in the drift term cancels the integral term, leaving the equation independent of $\lambda^*$. This is very convenient for applying numerical boundary conditions. The PIDE (9.53) can also be written as

$$F_t + \frac{b^2}{2} F_{SS} + (a - \lambda'_B b) F_S - rF + \lambda^* \int_0^\infty p[J][F(JS,t) - F(S,t) - (J-1)SF_S] \; dJ = 0 \tag{9.55}$$

which is valid for infinite activity processes.

In the case where we can hedge with the underlying asset $S$, we obtain

$$F_t + \frac{b^2}{2} F_{SS} + (r - \lambda^* E[J-1]) S F_S - rF + \lambda^* \int_0^\infty p(J)[F(JS,t) - F(S,t)] \; dJ = 0 \; . \tag{9.56}$$

Note that $\lambda^*$ and $p(J)$ are not the real arrival rate and real jump size distributions, since they are based on hedging arguments which eliminate the risk. Consequently, $\lambda^*$, $p(J)$ must be obtained by calibration to market data.

## 9.4 Simulating Jump Diffusion

Let's rewrite equation (9.10) as

$$\frac{dS_t}{S_{t^-}} \quad = \quad \mu \; dt + \sigma \; dZ + (J_t - 1) \; dq \; . \tag{9.57}$$

where

$$\begin{aligned} dq &= 1 \; ; \quad \text{with probability } \lambda dt \\ &= 0 \; ; \quad \text{with probability } 1 - \lambda dt. \end{aligned} \tag{9.58}$$

and where $S_{t^-}$ denotes the value of $S$ immediately before a possible jump occurs, $S_t$ is the value immediately after a jump occurs, and $J_t$ represents a possible jump size at time $t$.

In other words, if a jump occurs at time $t$, then equation (9.57) can be interpreted as

$$S_t = S_{t^-} J_t \ . \tag{9.59}$$

It will be easier to consider $X_t = \log S_t$. Suppose no jump occurs in $[t, t+\ dt]$, then equation (9.57) can be written as (using Ito's Lemma)

$$dX_t = (\mu - \sigma^2/2)\ dt + \sigma\ dZ \ . \tag{9.60}$$

If a jump occurs in $[t, t+\ dt]$, then equation (9.59) implies

$$\begin{aligned} X_t &= X_{t^-} + \log J_t \\ & dX_t = X_t - X_{t^-} \\ dX_t &= \log J_t \ . \end{aligned} \tag{9.61}$$

Putting together equations (9.60) and (9.61) gives

$$dX_t = (\mu - \sigma^2/2)\ dt + \sigma\ dZ + \log J_t\ dq \tag{9.62}$$

or writing this in terms of $S_t$, we get

$$S_t = S_{t^-} e^{(\ (\mu - \sigma^2/2)\ dt + \sigma(Z(t+dt) - Z(t))\ )} J_t \tag{9.63}$$

where we assume that only one possible jump can occur in $[t, t+dt]$, and that $J_t = 1$ if no jump occurs in $[t, t+dt]$.

A common assumption is that

$$\log J_t \sim N(\mu_J, \sigma_J^2) \tag{9.64}$$

i.e. that $J_t$ is lognormal, with mean $\mu_J$ and variance $\sigma_J^2$.

### 9.4.1 Compensated Drift

Recall equation (9.57)

$$\frac{dS_t}{S_{t^-}} = \mu\ dt + \sigma\ dZ + (J_t - 1)\ dq \ . \tag{9.65}$$

Now

$$\begin{aligned} E\left[\frac{dS_t}{S_{t^-}}\right] &= \mu\ dt + 0 + E[J_t - 1]E[dq] \\ &= \mu\ dt + E[J_t - 1]\lambda\ dt \ , \end{aligned} \tag{9.66}$$

where we assume that $dq$ and $J_t$ are independent. It is usual to redefine the drift term in terms of the compensated drift $\mu_c$

$$\begin{aligned} \mu &= \mu_c - \lambda\kappa \\ & \kappa = E[J_t - 1] \ . \end{aligned} \tag{9.67}$$

Now, with definition (9.67), equation (9.65) becomes

$$\frac{dS_t}{S_{t^-}} = (\mu_c - \lambda\kappa)\ dt + \sigma\ dZ + (J_t - 1)\ dq \ , \tag{9.68}$$

and

$$
\begin{aligned}
E\left[\frac{dS_t}{S_{t^-}}\right] &= (\mu_c - \lambda\kappa)\ dt + 0 + \lambda\kappa\ dt \\
&= \mu_c\ dt\ . 
\end{aligned} \tag{9.69}
$$

In this way

$$
E\left[\frac{dS_t}{S_{t^-}}\right] = \mu_c\ dt\ , \tag{9.70}
$$

so that $\mu_c\ dt$ is the expected return in $[t, t+dt]$.

If $\log J_t \sim N(\mu_J, \sigma_J^2)$, then

$$
\kappa = \exp(\sigma_J^2/2 + \mu_j) - 1\ . \tag{9.71}
$$

In terms of $X_t = \log S_t$, process (9.68) is

$$
dX_t = (\mu_c - \sigma^2/2 - \lambda\kappa)\ dt + \sigma\ dZ + \log J_t\ dq \tag{9.72}
$$

### 9.4.2 Contingent Claims Pricing

If we are interested in pricing a contingent claim, then we should compute the expected value in the risk neutral world. The requirement is then that

$$
E\left[\frac{dS_t}{S_{t^-}}\right] = \mu_c\ dt\ = r\ dt \tag{9.73}
$$

where $r$ is the risk free rate.

The stochastic process in the risk neutral world is then

$$
\frac{dS_t}{S_{t^-}} = (r - \lambda^{\mathbb{Q}}\kappa^{\mathbb{Q}})\ dt + \sigma\ dZ + (J_t^{\mathbb{Q}} - 1)\ dq\ , \tag{9.74}
$$

where $\lambda^{\mathbb{Q}}, \kappa^{\mathbb{Q}}, J_t^{\mathbb{Q}}$ are all risk adjusted quantities.

## 9.5 Matlab Code: Jump Diffusion

We give Matlab code for simulation jump diffusions, assuming process (9.72), with jump size given by (9.64) in (9.76).

We are assuming that

$$
\lambda\Delta t \ll 1 \tag{9.75}
$$

so that the probability of having more than one jump in $[t, t+dt]$ is negligible.

Note that there are more efficient ways of simulating jump diffusions, for special cases (i.e. volatility constant), but algorithm (9.76) is very simple and can be easily generalized.

**Vectorized M file For Jump Diffusion**

```
   randn('state',100);
   rand('state', 10);

   T = 1.00;              % expiry time
   sigma = 0.25;          % volatility
   mu = .10;              % P measure drift
   S_init = 100;          % initial value

    % jump size: log normal distribution

   lambda = .1 ;        %jump size arrival rate
   jump_vol = .40; ;    %stdrd dev of jump size
   jump_mean = -.9;     %mean of jump size

   N_sim = 100000;    %   number of simulations
   N = 250;           %   number of timesteps
   delt  = T/N;       %   timestep

%
%    compensated drift E[J-1]
%
   kappa = exp(.5*jump_vol*jump_vol + jump_mean) - 1.;
%
%   compensated drift for X = log(S)

   drift = (mu - sigma*sigma/2.0 - lambda*kappa);

%
%     X = log(S)
%
   X_old(1:N_sim,1) = log(S_init);
   X_new(1:N_sim,1) = zeros(N_sim, 1);

    jump_chek = zeros(N_sim, 1);
    jump_size = zeros(N_sim, 1);
    jump_mask = zeros( N_sim, 1);

for i=1:N % timestep loop
     jump_chek(:,1) = rand(N_sim,1);
     jump_mask(:,1) = ( jump_chek(:,1) <= lambda*delt);

     jump_size(:,1) = jump_mean  +  jump_vol*randn(N_sim,1);
     jump_size = jump_size.*jump_mask;

     X_new(:,1) = X_old(:,1) + drift*delt +sigma*sqrt(delt)*randn(N_sim,1) +...
                        jump_size(:,1);
     X_old(:,1) = X_new(:,1);
 end % timestep loop

 S(:,1) = exp( X_new(:,1) );

 n_bin = 200;
 hist(S, n_bin);

 stndrd_dev = std(S);
 disp(sprintf('standard deviation:  %.5g\n',stndrd_dev));
```



```
 mean_S = mean(S);
 disp(sprintf('mean:  %.5g\n',mean_S));
```

(9.76)

Code (9.76) assumes that the timestep is small, which means that the probability of a jump occurring in a step is $\simeq \lambda\ \Delta t$. However, it is more accurate the determine the number of jumps which occur in $[t, t+\Delta t]$ where $\Delta t$ is not necessarily small, using a Poisson distribution of the number of jumps. Code fragment (9.77) can do this job, but note that this is not vectorized.

**Code Fragment M file For Jump Diffusion: Finite $\Delta t$**

```
for i=1:N % timestep loop

   Num_jumps = poissrnd(lambda*delt,N_sim,1);
          % Nx1 vector, number of jumps in this step

   jump_size(1:N_sim, 1) = 0.0;
   for j=1:N_sim
      jump_size(j,1) = sum(  normrnd(jump_mean, ...
                           jump_vol, 1, Num_jumps(j,1)) ,2 );
           % sum of jumps for this path
   end

   X_new(:,1) = X_old(:,1) + drift*delt + ...
                sigma*sqrt(delt)*randn(N_sim,1) +...
                        jump_size(:,1);

   X_old(:,1) = X_new(:,1);

end
```

(9.77)

## 9.6 Poisson Distribution

Suppose we want to write our own code for sampling the number of jumps which occur in $[t, t+\Delta t]$ from a Poisson distribution. We can use the fundamental law of transformation of probabilities, and use the discrete cumulative distribution function. Suppose that the intensity of the Poisson process is $\lambda$ and the timestep is $\Delta t$. Then, the probability that the number of jumps $N$ in $[t, t+\Delta t]$ is at most $k$ is

$$Pr(N \leq k) = \sum_{j=0}^{j=k} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^j}{j!} \tag{9.78}$$

Therefore, we can sample the number of jumps in $[t, t+\Delta t]$ from a Poisson distribution using the following pseudo code described in Algorithm 9.1.

$p = e^{-\lambda\Delta t}$; $cum = p$ ; $U \simeq U[0,1]$ ; $N = 0$
**while** $cum < U$ **do**
  $N := N + 1$
  $p := p(\lambda\Delta t)/N$
  $cum := cum + p$
**end while**
**return** $(N)$

**Algorithm 9.1:** An algorithm for generating the number of jumps in $[t, t+\Delta t]$, sampled from a Poisson distribution with intensity $\lambda$.

# 10 Regime Switching

Of course, volatility is not constant in the real world. It is possible to combine jumps in the asset price with jumps in volatility and stochastic volatility. This leads to a two factor pricing PIDE for the option price.

A simpler approach is to assume that the volatility jumps between a number of *regimes* or volatility states. Let the value of a contingent claim be given by $F(\sigma, S, t)$, where we have allowed the volatility $\sigma$ to vary. Suppose

$$\begin{aligned} dS &= a\ dt + b\ dZ + (J_S - 1)S\ dq \\ d\sigma &= (J_\sigma - 1)\sigma\ dq\ , \end{aligned} \tag{10.1}$$

where $dq$ is a Poisson process and $dZ$ is the increment of a Weiner process. Here $a = a(\sigma, S, t)$, $b = b(\sigma, S, t)$. Note that the same $dq$ drives the jump in $S$ and the jump in $\sigma$. Following the same steps as in deriving equation (9.27) we obtain

$$\begin{aligned} dF &= \mu\ dt + \sigma^*\ dZ + \Delta F\ dq \\ &\mu = aF_S + \frac{b^2}{2}F_{SS} + F_t \\ &\sigma^* = bF_S \\ &\Delta F = [F(J_\sigma\sigma, J_S S, t) - F(\sigma, S, t)]\ . \end{aligned} \tag{10.2}$$

We follow the same steps as in the derivation of the jump diffusion PIDE in equations (9.29-9.36), i.e. we construct a hedge portfolio with three contracts $F_1, F_2, F_3$, and we do not assume that we can trade in the underlying. Eliminating the random terms gives rise to the analogue of equation (9.33) and hence a solution exists only if one of the equations is a linear combination of the other equations, which results in

$$(\mu - rF) = \lambda_B\sigma^* - \lambda_J\Delta F \tag{10.3}$$

and substituting the definitions of $\sigma^*, \mu$ from equation (10.2) gives

$$F_t + \frac{b^2}{2}F_{SS} + (a - \lambda_B b)F_S - rF + \lambda_J[F(J_\sigma\sigma, J_S S, t) - F(\sigma, S, t)] = 0\ . \tag{10.4}$$

In the event that, say, $F_3 = S$ is a traded asset, we note that in this case

$$\begin{aligned} \sigma_3^* &= b \\ \mu_3 &= a \\ \Delta F_3 &= (J_S - 1)S\ . \end{aligned} \tag{10.5}$$

Substituting equation (10.5) into equation (10.3) gives

$$a - \lambda_B b = rS - \lambda_J(J_S - 1)S\ . \tag{10.6}$$

Substituting equation (10.6) into equation (10.4) gives

$$F_t + \frac{b^2}{2}F_{SS} + [r - \lambda_J(J_S - 1)]SF_S - rF + \lambda_J[F(J_\sigma\sigma, J_S S, t) - F(\sigma, S, t)] = 0\ . \tag{10.7}$$

Note that if $J_S = 1$ (no jump in $S$, but a regime switch) then the term $\lambda_J(J_S - 1)$ disappears in the drift term.

We can repeat the above arguments with jumps from a given regime with volatility $\sigma$ to several possible regimes $J_\sigma^i\sigma$, $i = 1, \ldots, p$. Each possible transition $\sigma \to J_\sigma^i\sigma$ is driven by a Poisson process $dq^i$. We assume that $dq^i$ and $dq^j$ are independent. In this case, we have

$$
\begin{aligned}
dS &= a\ dt + b\ dZ + \sum_{i=1}^{i=p} (J_S^i - 1) S\ dq^i \\
d\sigma &= \sum_{i=1}^{i=p} (J_\sigma - 1) \sigma\ dq^i\ ,
\end{aligned}
\tag{10.8}
$$

Following the by now familiar steps, we obtain

$$
F_t + \frac{b^2}{2} F_{SS} + (a - \lambda_B b) F_S - rF + \sum_i \lambda_J^i [F(J_\sigma^i \sigma, J_S^i S, t) - F(\sigma, S, t)] = 0\ . \tag{10.9}
$$

Note that in general $\lambda_J^i = \lambda_J^i(\sigma, S)$, $J_\sigma^i = J_\sigma^i(\sigma, S)$, $J_S^i = J_S^i(\sigma, S)$, and $\lambda_B = \lambda_B(\sigma, S)$. Now, suppose we have only a finite number of possible regimes $\sigma_k$, $k = 1, \dots, p$. Let

$$
\begin{aligned}
b(\sigma_k, S, t) &= b_k \\
\lambda_B^k(\sigma_k, S) &= \lambda_B^k(S)) \\
F(\sigma_k, S, t) &= F^k(S, t) \\
\lambda_J^i(\sigma_k, S, t) &= \lambda_J^{k \to i} \\
J_\sigma^i(\sigma_k, S, t) &= J_\sigma^{k \to i} \\
J_S^i(\sigma_k, S, t) &= J_S^{k \to i}\ .
\end{aligned}
\tag{10.10}
$$

Rewriting equation (10.9) using equation (10.10) gives

$$
F_t^k + \frac{b_k^2}{2} F_{SS}^k + (a_k - \lambda_B^k b_k) F_S^k - rF^k + \sum_i \lambda_J^{k \to i} [F^i(J_S^{k \to i} S, t) - F^k(S, t)] = 0\ . \tag{10.11}
$$

If we can hedge with the underlying, then the usual arguments give

$$
a_k - \lambda_B^k b_k = rS - \sum_i \lambda_J^{k \to i} (J_S^{k \to i} - 1) S\ . \tag{10.12}
$$

Substituting equation (10.12) into equation (10.11) gives

$$
F_t^k + \frac{b_k^2}{2} F_{SS}^k + (r - \sum_i \lambda_J^{k \to i} (J_S^{k \to i} - 1)) S F_S^k - rF^k + \sum_i \lambda_J^{k \to i} [F^i(J_S^{k \to i} S, t) - F^k(S, t)] = 0\ . \tag{10.13}
$$

If we have only a small number of regimes, we are effectively solving a small number of coupled 1-d PDEs. In principle, the $J_S^{k \to i}, \sigma_k$ are $P$ measure parameters, while the $\lambda_J^{k \to i}$ is a $Q$ measure parameter. Often, we choose

$$
\begin{aligned}
b(\sigma, S, t) &= \sigma S \\
&\quad b_k = \sigma_k S\ .
\end{aligned}
\tag{10.14}
$$

We can also determine the $\sigma_k, J_S^{k \to i}, \lambda_J^{k \to i}$ by calibration to market prices.

# 11 Mean Variance Portfolio Optimization

An introduction to Computational Finance would not be complete without some discussion of Portfolio Optimization. Consider a risky asset which follows Geometric Brownian Motion with drift

$$
\frac{dS}{S} = \mu\ dt + \sigma\ dZ\ , \tag{11.1}
$$

where as usual $dZ = \phi\sqrt{dt}$ and $\phi \sim N(0,1)$. Suppose we consider a fixed finite interval $\Delta t$, then we can write equation (11.1) as

$$
\begin{aligned}
R &= \mu' + \sigma'\phi \\
&\quad R = \frac{\Delta S}{S} \\
&\quad \mu' = \mu \Delta t \\
&\quad \sigma' = \sigma\sqrt{\Delta t} \ ,
\end{aligned}
\tag{11.2}
$$

where $R$ is the actual return on the asset in $[t, t+\Delta t]$, $\mu'$ is the expected return on the asset in $[t, t+\Delta t]$, and $\sigma'$ is the standard deviation of the return on the asset in $[t, t+\Delta t]$.

Now consider a portfolio of $N$ risky assets. Let $R^i$ be the return on asset $i$ in $[t, t+\Delta t]$, so that

$$
R^i = \mu_i' + \sigma_i'\phi_i \tag{11.3}
$$

Suppose that the correlation between asset $i$ and asset $j$ is given by $\rho_{ij} = E[\phi_i\phi_j]$. Suppose we buy $x_i$ of each asset at $t$, to form the portfolio $P$

$$
P = \sum_{i=1}^{i=N} x_i S_i \ . \tag{11.4}
$$

Then, over the interval $[t, t+\Delta t]$

$$
\begin{aligned}
P + \Delta P &= \sum_{i=1}^{i=N} x_i S_i (1 + R^i) \\
\Delta P &= \sum_{i=1}^{i=N} x_i S_i R^i \\
\frac{\Delta P}{P} &= \sum_{i=1}^{i=N} w_i R^i \\
&\quad w_i = \frac{x_i S_i}{\sum_{j=1}^{j=N} x_j S_j}
\end{aligned}
\tag{11.5}
$$

In other words, we divide up our total wealth $W = \sum_{i=1}^{i=N} x_i S_i$ into each asset with weight $w_i$. Note that $\sum_{i=1}^{i=N} w_i = 1$.

To summarize, given some initial wealth at $t$, we suppose that an investor allocates a fraction $w_i$ of this wealth to each asset $i$. We assume that the total wealth is allocated to this risky portfolio $P$, so that

$$
\begin{aligned}
\sum_{i=1}^{i=N} w_i &= 1 \\
P &= \sum_{i=1}^{i=N} x_i S_i \\
R_p = \frac{\Delta P}{P} &= \sum_{i=1}^{i=N} w_i R^i \ .
\end{aligned}
\tag{11.6}
$$

The expected return on this portfolio $\overline{R_p}$ in $[t, t+\Delta t]$ is

$$
\overline{R_p} = \sum_{i=1}^{i=N} w_i \mu_i' \ , \tag{11.7}
$$

while the variance of $R_p$ in $[t, t+\Delta t]$ is

$$Var(R_p) = \sum_{i=1}^{i=N}\sum_{j=1}^{j=N} w_i w_j \sigma'_i \sigma'_j \rho_{ij} \ . \tag{11.8}$$

## 11.1 Special Cases

Suppose the assets all have zero correlation with one another, i.e. $\rho_{ij} \equiv 0, \forall i \neq j$ (of course $\rho_{ii} = 1$). Then equation (11.8) becomes

$$Var(R_p) = \sum_{i=1}^{i=N} (\sigma'_i)^2 (w_i)^2 \ . \tag{11.9}$$

Now, suppose we equally weight all the assets in the portfolio, i.e. $w_i = 1/N, \forall i$. Let $max_i \sigma'_i = \sigma'_{max}$, then

$$\begin{aligned} Var(R_p) &= \frac{1}{N^2}\sum_{i=1}^{i=N}(\sigma'_i)^2 \\ &\leq \frac{N(\sigma'_{max})^2}{N^2} \\ &= O\left(\frac{1}{N}\right) \ , \end{aligned} \tag{11.10}$$

so that in this special case, if we diversify over a large number of assets, the standard deviation of the portfolio tends to zero as $N \to \infty$.

Consider another case: all assets are perfectly correlated, $\rho_{ij} = 1, \forall i, j$. In this case

$$\begin{aligned} Var(R_p) &= \sum_{i=1}^{i=N}\sum_{j=1}^{j=N} w_i w_j \sigma'_i \sigma'_j \\ &= \left[\sum_{j=1}^{j=N} w_j \sigma'_j\right]^2 \end{aligned} \tag{11.11}$$

so that if $sd(R) = \sqrt{Var(R)}$ is the standard deviation of $R$, then, in this case

$$sd(R_p) = \sum_{j=1}^{j=N} w_j \sigma'_j \ , \tag{11.12}$$

which means that in this case the standard deviation of the portfolio is simply the weighted average of the individual asset standard deviations.

In general, we can expect that $0 < |\rho_{ij}| < 1$, so that the standard deviation of a portfolio of assets will be smaller than the weighted average of the individual asset standard deviation, but larger than zero.

This means that diversification will be a good thing (as Martha Stewart would say) in terms of risk versus reward. In fact, a portfolio of as little as $10 - 20$ stocks tends to reap most of the benefits of diversification.

## 11.2 The Portfolio Allocation Problem

Different investors will choose different portfolios depending on how much risk they wish to take. However, all investors like to achieve the highest possible expected return for a given amount of risk. We are assuming that risk and standard deviation of portfolio return are synonymous.

Let the covariance matrix $C$ be defined as

$$[C]_{ij} = C_{ij} = \sigma'_i \sigma'_j \rho_{ij} \tag{11.13}$$

and define the vectors $\bar{\mu} = [\mu'_1, \mu'_2, ..., \mu'_N]^t$, $\bar{w} = [w_1, w_2, ..., w_N]^t$. In theory, the covariance matrix should be symmetric positive semi-definite. However, measurement errors may result in $C$ having a negative eigenvalue, which should be fixed up somehow.

The expected return on the portfolio is then

$$\overline{R_p} = \bar{w}^t \bar{\mu} \ , \tag{11.14}$$

and the variance is

$$Var(R_p) \quad = \quad \bar{w}^t C \bar{w} \ . \tag{11.15}$$

We can think of portfolio allocation problem as the following. Let $\alpha$ represent the degree with which investors want to maximize return at the expense of assuming more risk. If $\alpha \to 0$, then investors want to avoid as much risk as possible. On the other hand, if $\alpha \to \infty$, then investors seek only to maximize expected return, and don't care about risk. The portfolio allocation problem is then (for given $\alpha$) find $\bar{w}$ which satisfies

$$\min_{\bar{w}} \bar{w}^t C \bar{w} - \alpha \bar{w}^t \bar{\mu} \tag{11.16}$$

subject to the constraints

$$\sum_i w_i \quad = \quad 1 \tag{11.17}$$

$$L_i \leq \quad w_i \quad \leq U_i \ \ ; \ \ i = 1, ..., N \ . \tag{11.18}$$

Constraint (11.17) is simply equation (11.6), while constraints (11.18) may arise due to the nature of the portfolio. For example, most mutual funds can only hold long positions ($w_i \geq 0$), and they may also be prohibited from having a large position in any one asset (e.g. $w_i \leq .20$). Long-short hedge funds will not have these types of restrictions. For fixed $\alpha$, equations (11.16-11.18) constitute a *quadratic programming* problem.

Let

$$\begin{aligned} sd(R_p) \quad &= \quad \text{standard deviation of } R_p \\ &= \quad \sqrt{Var(R_p)} \end{aligned} \tag{11.19}$$

We can now trace out a curve on the $(sd(R_p), \overline{R_p})$ plane. We pick various values of $\alpha$, and then solve the quadratic programming problem (11.16-11.18). Figure 11.1 shows a typical curve, which is also known as the *efficient frontier*. The data used for this example is

$$\begin{aligned} \bar{\mu} = \begin{bmatrix} .15 \\ .20 \\ .08 \end{bmatrix} \quad &; \quad C = \begin{bmatrix} .20 & .05 & -.01 \\ .05 & .30 & .015 \\ -.01 & .015 & .1 \end{bmatrix} \\ L = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad &; \quad U = \begin{bmatrix} \infty \\ \infty \\ \infty \end{bmatrix} \end{aligned} \tag{11.20}$$

We have restricted this portfolio to be long only. For a given value of the standard deviation of the portfolio return ($sd(R_p)$), then any point below the curve is not efficient, since there is another portfolio with the same risk (standard deviation) and higher expected return. Only points on the curve are efficient

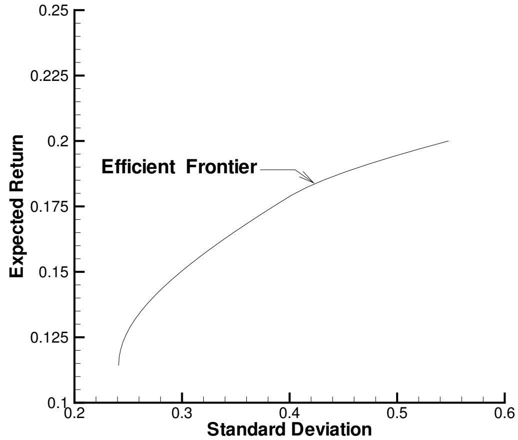



Figure 11.1: A typical efficient frontier. This curve shows, for each value of portfolio standard deviation $SD(R_p)$, the maximum possible expected portfolio return $R_p$. Data in equation (11.20).

in this manner. In general, a linear combination of portfolios at two points along the efficient frontier will be feasible, i.e. satisfy the constraints. This feasible region will be convex along the efficient frontier. Another way of saying this is that a straight line joining any two points along the curve does not intersect the curve except at the given two points. Why is this the case? If this was not true, then the efficient frontier would not really be efficient. (see Portfolio Theory and Capital Markets, W. Sharpe, McGraw Hill, 1970, reprinted in 2000).

Figure 11.2 shows results if we allow the portfolio to hold up to .25 short positions in each asset. In other words, the data is the same as in (11.20) except that

$$L = \begin{bmatrix} -.25 \\ -.25 \\ -.25 \end{bmatrix} . \tag{11.21}$$

In general, long-short portfolios are more efficient than long-only portfolios. This is the advertised advantage of long-short hedge funds.

Since the feasible region is convex, we can actually proceed in a different manner when constructing the efficient frontier. First of all, we can determine the maximum possible expected return ($\alpha = \infty$ in equation (11.16)),

$$\begin{aligned} &\min_{\bar{w}} -\bar{w}^t \bar{\mu} \\ &\qquad \sum_i w_i = 1 \\ &\qquad L_i \le w_i \le U_i \ \ ; \ \ i = 1, ..., N \end{aligned} \tag{11.22}$$

which is simply a linear programming problem. If the solution weight vector to this problem is $(\bar{w})_{max}$, then the maximum possible expected return is $(\overline{R_p})^{max} = \bar{w}^t_{max} \bar{\mu}$.

Then determine the portfolio with the smallest possible risk, ($\alpha = 0$ in equation (11.16) )

$$\begin{aligned} &\min_{\bar{w}} \bar{w}^t C \bar{w} \\ &\qquad \sum_i w_i = 1 \\ &\qquad L_i \le w_i \le U_i \ \ ; \ \ i = 1, ..., N \ . \end{aligned} \tag{11.23}$$

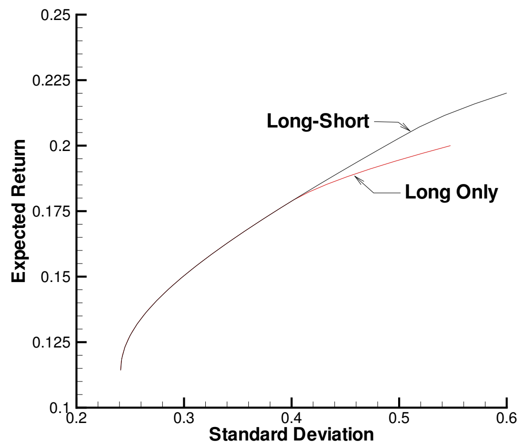



Figure 11.2: Efficient frontier, comparing results for long-only portfolio (11.20) and a long-short portfolio (same data except that lower bound constraint is replaced by equation (11.21).

If the solution weight vector to this quadratic program is given by $\bar{w}_{min}$, then the minimum possible portfolio return is $(\overline{R_p})^{min} = \bar{w}^t_{min}\bar{\mu}$. We then divide up the range $[(\overline{R_p})^{min}, (\overline{R_p})^{max}]$ into a large number of discrete portfolio returns $(\overline{R_p})_k; k = 1, ..., N_{pts}$. Let $e = [1, 1, ..., 1]^t$, and

$$A = \begin{bmatrix} \bar{\mu}^t \\ e^t \end{bmatrix} \quad ; \quad B^k = \begin{bmatrix} (R_p)_k \\ 1 \end{bmatrix} \tag{11.24}$$

then, for given $(\overline{R_p})_k$ we solve the quadratic program

$$\begin{aligned} &\min_{\bar{w}} \bar{w}^t C \bar{w} \\ &\qquad A\bar{w} = B^k \\ &\qquad L_i \le w_i \le U_i \quad ; \quad i = 1, ..., N \quad , \end{aligned} \tag{11.25}$$

with solution vector $(\bar{w})_k$ and hence portfolio standard deviation $sd((R_p)_k) = \sqrt{(\bar{w})^t_k C(\bar{w})_k}$. This gives us a set of pairs $(sd((R_p)_k), (\overline{R_p})_k), k = 1, ..., N_{pts}$.

## 11.3 Adding a Risk-free asset

Up to now, we have assumed that each asset is risky, i.e. $\sigma'_i > 0, \forall i$. However, what happens if we add a risk free asset to our portfolio? This risk-free asset must earn the risk free rate $r' = r\Delta t$, and its standard deviation is zero. The data for this case is (the risk-free asset is added to the end of the weight vector, with $r' = .03$).

$$\begin{aligned} \bar{\mu} &= \begin{bmatrix} .15 \\ .20 \\ .08 \\ .03 \end{bmatrix} \quad ; \quad C = \begin{bmatrix} .20 & .05 & -.01 & 0.0 \\ .05 & .30 & .015 & 0.0 \\ -.01 & .015 & .1 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \\ L &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\infty \end{bmatrix} \quad ; \quad U = \begin{bmatrix} \infty \\ \infty \\ \infty \\ \infty \end{bmatrix} \end{aligned} \tag{11.26}$$

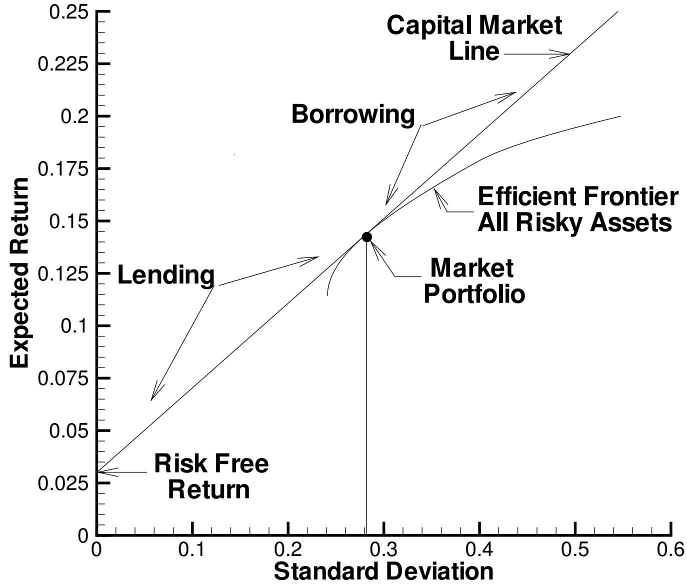



Figure 11.3: The efficient frontier from Figure 11.1 (all risky assets), and the efficient frontier with the same assets as in Figure 11.1, except that we include a risk free asset. In this case, the efficient frontier becomes a straight line, shown as the capital market line.

where we have assumed that we can borrow any amount at the risk-free rate (a dubious assumption).

If we compute the efficient frontier with a portfolio of risky assets and include one risk-free asset, we get the result labeled *capital market line* in Figure 11.3. In other words, in this case the efficient frontier is a straight line. Note that this straight line is always above the efficient frontier for the portfolio consisting of all risky assets (as in Figure 11.1). In fact, given the efficient frontier from Figure 11.1, we can construct the efficient frontier for a portfolio of the same risky assets plus a risk free asset in the following way. First of all, we start at the point $(0, r')$ in the $(sd(R_p), \overline{R_p})$ plane, corresponding to a portfolio which consists entirely of the risk free asset. We then draw a straight line passing through $(0, r')$, which touches the all-risky-asset efficient frontier at a single point (the straight line is tangent the all-risky-asset efficient frontier). Let the portfolio weights at this single point be denoted by $\bar{w}_M$. The portfolio corresponding to the weights $\bar{w}_M$ is termed the *market portfolio*. Let $(\overline{R_p})_M = \bar{w}_M^t \bar{\mu}$ be the expected return on this market portfolio, with corresponding standard deviation $sd((R_p)_M)$. Let $w_r$ be the fraction invested in the risk free asset. Then, any point along the capital market line has

$$\begin{aligned} \overline{R_p} &= w_r r' + (1 - w_r)(\overline{R_p})_M \\ sd(R_p) &= (1 - w_r)\ sd((R_p)_M)\ . \end{aligned} \tag{11.27}$$

If $w_r \geq 0$, then we are lending at the risk-free rate. If $w_r < 0$, we are borrowing at the risk-free rate.

Consequently, given a portfolio of risky assets, and a risk-free asset, then all investors should divide their assets between the risk-free asset and the market portfolio. Any other choice for the portfolio is not efficient. Note that the actual fraction selected for investment in the market portfolio depends on the risk preferences of the investor.

The capital market line is so important, that the equation of this line is written as $\overline{R_p} = r' + \lambda_M\ sd((R_p))$, where $\lambda_M$ is the market price of risk. In other words, all diversified investors, at any particular point in time, should have diversified portfolios which plot along the capital market line. All portfolios should have the same Sharp ratio

$$\lambda_M = \frac{\overline{R_p} - r'}{sd(R_p)} . \tag{11.28}$$

## 11.4 Criticism

Is mean-variance portfolio optimization the solution to all our problems? Not exactly. We have assumed that $\mu', \sigma'$ are independent of time. This is not likely. Even if these parameters are reasonably constant, they are difficult to estimate. In particular, $\mu'$ is hard to determine if the time series of returns is not very long. Remember that for short time series, the noise term (Brownian motion) will dominate. If we have a long time series, we can get a better estimate for $\mu'$, but why do we think $\mu'$ for a particular firm will be constant for long periods? Probably, stock analysts should be estimating $\mu'$ from company balance sheets, sales data, etc. However, for the past few years, analysts have been too busy hyping stocks and going to lunch to do any real work. So, there will be lots of different estimates of $\mu', C$, and hence many different *optimal* portfolios.

In fact, some recent studies have suggested that if investors simply use the $1/N$ rule, whereby initial wealth is allocated equally between $N$ assets, that this does a pretty good job, assuming that there is uncertainty in the estimates of $\mu', C$.

We have also assumed that risk is measured by standard deviation of portfolio return. Actually, if I am long an asset, I like it when the asset goes up, and I don't like it when the asset goes down. In other words, volatility which makes the price increase is good. This suggests that perhaps it may be more appropriate to minimize downside risk only (assuming a long position).

Perhaps one of the most useful ideas that come from mean-variance portfolio optimization is that diversified investors (at any point in time) expect that any optimal portfolio will produce a return

$$\begin{aligned}
\overline{R_p} &= r' + \lambda_M \sigma'_p \\
&\overline{R_p} = \text{ Expected portfolio return} \\
&r' = \text{ risk-free return in period } \Delta t \\
&\lambda_M = \text{ market price of risk} \\
&\sigma'_p = \text{ Portfolio volatility} ,
\end{aligned} \tag{11.29}$$

where different investors will choose portfolios with different $\sigma'_p$ (volatility), depending on their risk preferences, but $\lambda_M$ is the same for all investors. Of course, we also have

$$\overline{R_M} = r' + \lambda_M \sigma'_M . \tag{11.30}$$

Note: there is a whole field called *Behavioural Finance*, whose adherents don't think much of mean-variance portfolio optimization.

Another recent approach is to compute the optimal portfolio weights using using many different perturbed input data sets. The input data (expected returns, and covariances) are determined by *resampling*, i.e. assuming that the observed values have some observational errors. In this way, we can get an some sort of optimal portfolio weights which have some effect of data errors incorporated in the result. This gives us an average efficient frontier, which, it is claimed, is less sensitive to data errors.

## 11.5 Individual Securities

Equation (11.30) refers to an efficient portfolio. What is the relationship between risk and reward for individual securities? Consider the following portfolio: divide all wealth between the market portfolio, with weight $w_M$ and security $i$, with weight $w_i$. By definition

$$w_M + w_i = 1 , \tag{11.31}$$

and we define

$$
\begin{aligned}
\overline{R^M} &= \text{expected return on the market portfolio} \\
\overline{R^i} &= \text{expected return on asset } i \\
\sigma'_M &= \text{s.d. of return on market portfolio} \\
\sigma'_i &= \text{s.d. of return on asset } i \\
C_{i,M} &= \sigma'_M \sigma'_i \rho_{i,M} \\
&= \text{Covariance between } i \text{ and } M
\end{aligned}
\tag{11.32}
$$

Now, the expected return on this portfolio is

$$
\begin{aligned}
\overline{R_p} = E[R_p] &= w_i \overline{R^i} + w_M \overline{R^M} \\
&= w_i R^i + (1 - w_i) \overline{R^M}
\end{aligned}
\tag{11.33}
$$

and the variance is

$$
\begin{aligned}
Var(R_p) &= (\sigma'_p)^2 = w_i^2 (\sigma'_i)^2 + 2 w_i w_M C_{i,M} + w_M^2 (\sigma'_M)^2 \\
&= w_i^2 (\sigma'_i)^2 + 2 w_i (1 - w_i) C_{i,M} + (1 - w_i)^2 (\sigma'_M)^2
\end{aligned}
\tag{11.34}
$$

For a set of values $\{w_i\}$, equations (11.33-11.34) will plot a curve in expected return-standard deviation plane $(\overline{R_p}, \sigma'_p)$ (e.g. Figure 11.3). Let's determine the slope of this curve when $w_i \to 0$, i.e. when this curve intersects the capital market line at the market portfolio.

$$
\begin{aligned}
2(\sigma'_p) \frac{\partial(\sigma'_p)}{\partial w_i} &= 2 w_i (\sigma'_i)^2 + 2(1 - 2w_i) C_{i,M} + 2(w_i - 1)(\sigma'_M)^2 \\
\frac{\partial \overline{R_p}}{\partial w_i} &= \overline{R^i} - \overline{R^M} \ .
\end{aligned}
\tag{11.35}
$$

Now,

$$
\begin{aligned}
\frac{\partial \overline{R_p}}{\partial (\sigma'_p)} &= \frac{\frac{\partial \overline{R_p}}{\partial w_i}}{\frac{\partial (\sigma'_p)}{\partial w_i}} \\
&= \frac{(\overline{R^i} - \overline{R^M})(\sigma'_p)}{w_i (\sigma'_i)^2 + (1 - 2w_i) C_{i,M} + (w_i - 1)(\sigma'_M)^2} \quad .
\end{aligned}
\tag{11.36}
$$

Now, let $w_i \to 0$ in equation (11.36), then we obtain

$$
\frac{\partial \overline{R_p}}{\partial (\sigma'_p)} = \frac{(\overline{R^i} - \overline{R^M})(\sigma'_M)}{C_{i,M} - (\sigma'_M)^2}
\tag{11.37}
$$

But this curve should be tangent to the capital market line, equation (11.30) at the point where the capital market line touches the efficient frontier. If this curve is not tangent to the capital market line, then this implies that if we choose $w_i = \pm\epsilon$, then the curve would be above the capital market line, which should not be possible (the capital market line is the most efficient possible portfolio). This assumes that positions with $w_i < 0$ in asset $i$ are possible.

Assuming that the slope of the $\overline{R_p}$ portfolio is tangent to the capital market line gives (from equations (11.30,11.37))

$$
\frac{\overline{R^M} - r'}{(\sigma'_M)} = \frac{(\overline{R^i} - \overline{R^M})(\sigma'_M)}{C_{i,M} - (\sigma'_M)^2}
\tag{11.38}
$$

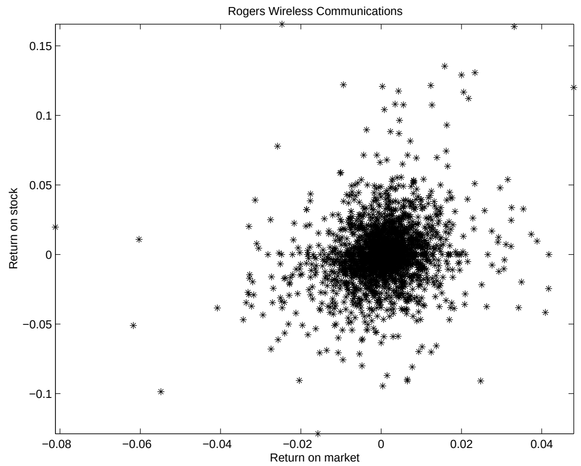



Figure 11.4: Return on Rogers Wireless Communications versus return on TSE 300. Each point represents pairs of daily returns. The vertical axis measures the daily return on the stock and the horizontal axis that of the TSE300.

or

$$\begin{aligned} \overline{R^i} &= r' + \beta_i(\overline{R^M} - r') \\ & \beta_i = \frac{C_{i,M}}{(\sigma'_M)^2} \ . \end{aligned} \tag{11.39}$$

The coefficient $\beta_i$ in equation (11.39) has a nice intuitive definition. Suppose we have a time series of returns

$$\begin{aligned} (R^i)_k &= \text{Return on asset i, in period k} \\ (R^M)_k &= \text{Return on market portfolio in period k} \ . \end{aligned} \tag{11.40}$$

Typically, we assume that the market portfolio is a broad index, such as the TSX 300. Now, suppose we try to obtain a least squares fit to the above data, using the equation

$$R^i \simeq \alpha_i + b_i R^M \ . \tag{11.41}$$

Carrying out the usual least squares analysis (e.g. do a linear regression of $R^i$ vs. $R^M$), we find that

$$b_i = \frac{C_{i,M}}{(\sigma'_M)^2} \tag{11.42}$$

so that we can write

$$R^i \simeq \alpha_i + \beta_i R^M \ . \tag{11.43}$$

This means that $\beta_i$ is the slope of the best fit straight line to a $((R^i)_k, (R^M)_k)$ scatter plot. An example is shown in Figure 11.4. Now, from equation (11.39) we have that

$$\overline{R^i} = r' + \beta_i(\overline{R^M} - r') \tag{11.44}$$

which is consistent with equation (11.43) if

$$\begin{aligned} R^i &= \alpha_i + \beta_i R^M + \epsilon_i \\ & E[\epsilon_i] = 0 \\ & \alpha_i = r'(1 - \beta_i) \\ & E[\epsilon_i, R^M] = 0 \ , \end{aligned} \tag{11.45}$$

since

$$E[R^i] = \overline{R^i} \quad = \quad \alpha_i + \beta_i \overline{R^M} \ . \tag{11.46}$$

Equation (11.46) has the interpretation that the return on asset $i$ can be decomposed into a drift component, a part which is correlated to the market portfolio (the broad index), and a random part uncorrelated with the index. Make the following assumptions

$$\begin{aligned} E[\epsilon_i \epsilon_j] &= 0 \ ; \ \ i \neq j \\ &= e_i^2 \ ; \ \ i = j \end{aligned} \tag{11.47}$$

e.g. that returns on each each asset are correlated only through their correlation with the index. Consider once again a portfolio where the wealth is divided amongst $N$ assets, each asset receiving a fraction $w_i$ of the initial wealth. In this case, the return on the portfolio is

$$\begin{aligned} R_p &= \sum_{i=1}^{i=N} w_i R^i \\ \overline{R_p} &= \sum_{i=1}^{i=N} w_i \alpha_i + \overline{R^M} \sum_{i=1}^{i=N} w_i \beta_i \end{aligned} \tag{11.48}$$

and

$$\begin{aligned} s.d.(R_p) &= \sqrt{(\sigma'_M)^2 \sum_{i=1}^{i=N} \sum_{j=1}^{j=N} w_i w_j \beta_i \beta_j + \sum_{i=1}^{i=N} w_i^2 e_i^2} \\ &= \sqrt{(\sigma'_M)^2 \left( \sum_{i=1}^{i=N} w_i \beta_i \right)^2 + \sum_{i=1}^{i=N} w_i^2 e_i^2} \ . \end{aligned} \tag{11.49}$$

Now, if $w_i = O(1/N)$, then

$$\sum_{i=1}^{i=N} w_i^2 e_i^2 \tag{11.50}$$

is $O(1/N)$ as $N$ becomes large, hence equation (11.49) becomes

$$s.d.(R_p) \quad \simeq \quad \sigma'_M \left| \sum_{i=1}^{i=N} w_i \beta_i \right| \ . \tag{11.51}$$

Note that if we write

$$\overline{R^i} \quad = \quad r' + \lambda_i \sigma'_i \tag{11.52}$$

then we also have that

$$\overline{R^i} \quad = \quad r' + \beta_i (\overline{R^M} - r') \tag{11.53}$$

so that the market price of risk of security $i$ is

$$\lambda_i \quad = \quad \frac{\beta_i (\overline{R^M} - r')}{\sigma'_i} \tag{11.54}$$

which is useful in real options analysis.

| 1 year | 2 years | 5 years | 10 years | 20 years | 30 years | 30 year bond yield |
|---|---|---|---|---|---|---|
| -2% | -5% | 10% | 8% | 7% | 6% | 3% |

Table 12.1: Historical annualized compound return, XYZ Mutual Equity Funds. Also shown is the current yield on a long term government bond.

# 12 Some Investing Facts

## 12.1 Stocks for the Long Run?

Conventional wisdom states that investment in a diversified portfolio of equities has a low risk for a long term investor. However, in a recent article ("Irrational Optimism," Fin. Anal. J. E. Simson, P.Marsh, M. Staunton, vol 60 (January, 2004) 25-35) an extensive analysis of historical data of equity returns was carried out. Projecting this information forward, the authors conclude that the probability of a negative real return over a twenty year period, for a investor holding a diversified portfolio, is about 14 per cent. In fact, most individuals in defined contribution pension plans have poorly diversified portfolios. Making more realistic assumptions for defined contribution pension plans, the authors find that the probability of a negative real return over twenty years is about 25 per cent.

Let's see if we can explain why there is this common misconception about the riskiness of long term equity investing. Table 12.1 shows a typical table in a Mutual Fund advertisement. From this table, we are supposed to conclude that

- Long term equity investment is not very risky, with an annualized compound return about 3% higher than the current yield on government bonds.
- If $S$ is the value of the mutual fund, and $B$ is the value of the government bond, then

$$\begin{aligned} B(T) &= B(0)e^{rT} \\ &\quad r = .03 \\ S(T) &\simeq S(0)e^{\alpha T} \\ &\quad \alpha = .06, \end{aligned} \tag{12.1}$$

  for $T$ large, which gives

$$\frac{\frac{S(T=30)}{S(0)}}{\frac{B(T=30)}{B(0)}} = e^{1.8-.9} = e^{.9} \simeq 2.46, \tag{12.2}$$

  indicating that you more than double your return by investing in equities compared to bonds (over the long term).

A convenient way to measure the relative returns on these two investments (bonds and stocks) is to compare the total compound return

$$\begin{aligned} \text{Compound return: stocks} &= \log\left[\frac{S(T)}{S(0)}\right] = \alpha T \\ \text{Compound return: bonds} &= \log\left[\frac{B(T)}{B(0)}\right] = rT \ , \end{aligned} \tag{12.3}$$

or the annualized compound returns

$$\begin{aligned} \text{Annualized compound return: stocks} &= \frac{1}{T}\log\left[\frac{S(T)}{S(0)}\right] = \alpha \\ \text{Annualized compound return: bonds} &= \frac{1}{T}\log\left[\frac{B(T)}{B(0)}\right] = r \ . \end{aligned} \tag{12.4}$$

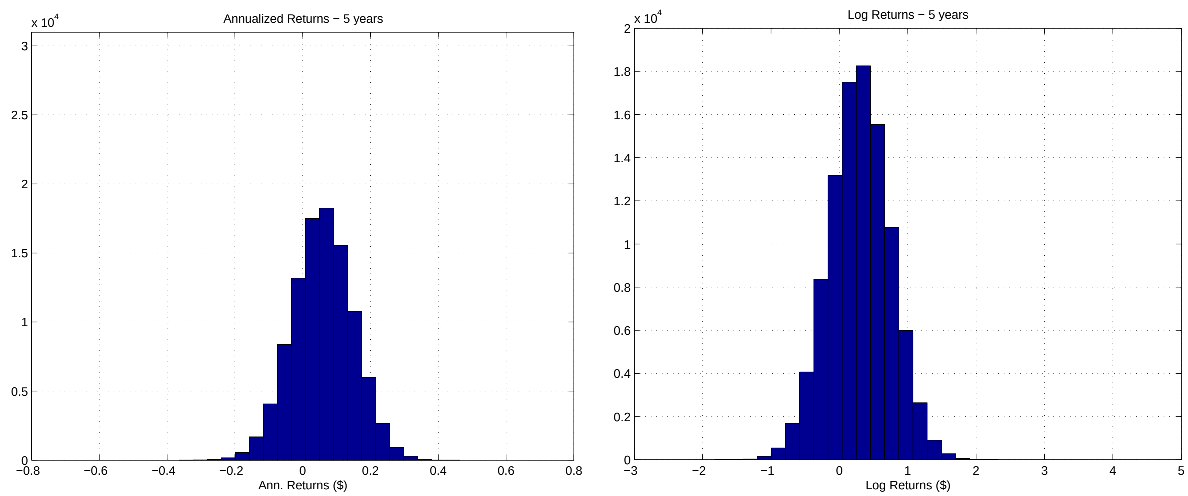



Figure 12.1: *Histogram of distribution of returns $T = 5$ years. $\mu = .08, \sigma = .2$, 100,000 simulations. Left: annualized return $1/T \log[S(T)/S(0)]$. Right: return $\log[S(T)/S(0)]$.*

If we assume that the value of the equity portfolio $S$ follows a Geometric Brownian Motion

$$dS = \mu S \, dt + \sigma S \, dZ \tag{12.5}$$

then from equation (2.57) we have that

$$\log\left(\frac{S(T)}{S(0)}\right) \sim N((\mu - \frac{\sigma^2}{2})T, \sigma^2 T) \ , \tag{12.6}$$

i.e. the compound return in is normally distributed with mean $(\mu - \frac{\sigma^2}{2})T$ and variance $\sigma^2 T$, so that the variance of the total compound return increases as $T$ becomes large.

Since $var(aX) = a^2 var(X)$, it follows that

$$\frac{1}{T}\log\left(\frac{S(T)}{S(0)}\right) \sim N((\mu - \frac{\sigma^2}{2}), \sigma^2/T) \ , \tag{12.7}$$

so that the the variance of the annualized return tends to zero at $T$ becomes large.

Of course, what we really care about is the total compound return (that's how much we actually have at $t = T$, relative to what we invested at $t = 0$) at the end of the investment horizon. This is why Table 12.1 is misleading. There is significant risk in equities, even over the long term (30 years would be long-term for most investors).

Figure 12.1 shows the results of 100,000 simulations of asset prices assuming that the asset follows equation (12.5), with $\mu = .08, \sigma = .2$. The investment horizon is 5 years. The results are given in terms of histograms of the annualized compound return (equation (12.4)) and the total compound return ((equation (12.3)).

Figure 12.2 shows similar results for an investment horizon of 30 years. Note how the variance of the annualized return has decreased, while the variance of the total return has increased (verifying equations (12.6-12.7)).

Assuming long term bonds yield 3%, this gives a total compound return over 30 years of .90, for bonds. Looking at the right hand panel of Figure 12.2 shows that there are many possible scenarios where the return on equities will be less than risk free bonds after 30 years. The number of scenarios with return less than risk free bonds is given by the area to the left of .9 in the histogram.

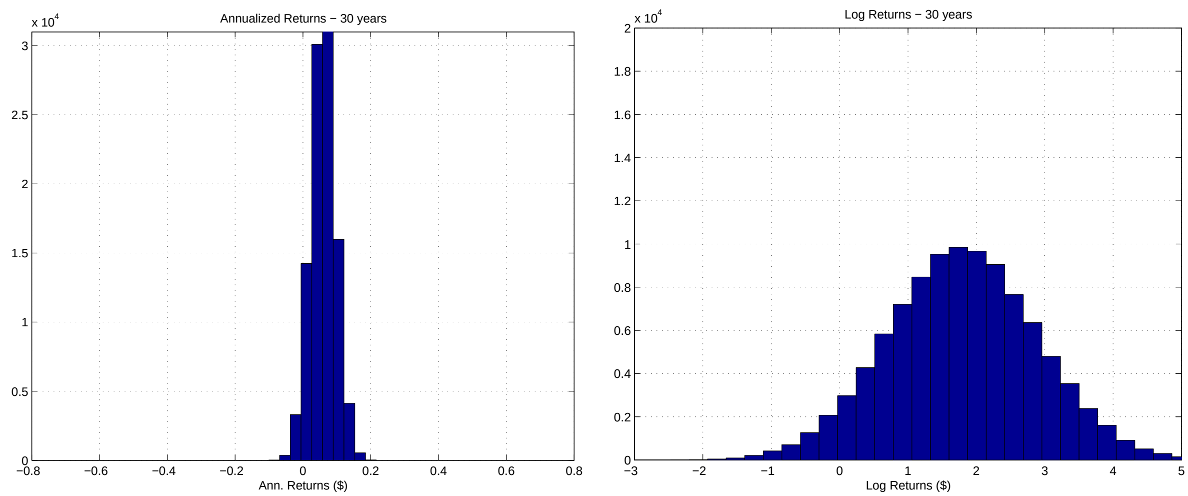



Figure 12.2: *Histogram of distribution of returns* $T = 30$ *years.* $\mu = .08, \sigma = .2$, $100,000$ *simulations. Left: annualized return* $1/T \log[S(T)/S(0)]$. *Right: return* $\log[S(T)/S(0)]$,

| $\alpha$ | $Pr(S(T)) < \alpha E[S(T)]$ |
|---|---|
| 1.0 | 0.61 |
| 0.9 | 0.53 |
| 0.8 | 0.45 |
| 0.7 | 0.35 |
| 0.6 | 0.25 |
| 0.5 | 0.16 |

Table 12.2: $T = 30$ years, $\sigma = 0.10$, $\mu = 0.04$. For example, there is a 25% chance that the realized value of $S(T)$ is less than 60% of the expected value.

### 12.1.1 GBM is Risky

Consider an investment which follows Geometric Brownian Motion

$$\frac{dS}{S} = \mu \, dt + \sigma \, dZ \; . \tag{12.8}$$

GBM has a very large relative standard deviation for long times, even for moderate volatilities. From equations (2.67 - 2.70), we have

$$\begin{aligned} E[S] &= S_0 e^{\mu t} \\ \frac{standard\ deviation[S(t)]}{E[S(t)]} &= \sqrt{e^{\sigma^2 t} - 1} \; . \end{aligned} \tag{12.9}$$

Figure 12.3 shows some Monte Carlo simulations (30 realizations) along with the mean path. You can see that there is a large spread compared to the mean path. Figure 12.4 shows the probability density of $S(T)/E[S(T)]$.

Table 12.2 shows the probabilities of underperforming the mean path. This is something which financial planners never tell you. As well, there is standard method for determining the financial health of a pension plan called *going concern* valuation. Essentially, this method assumes that the plan investments will follow the mean path. You can see from table 12.2 that this is an extremely risky idea.

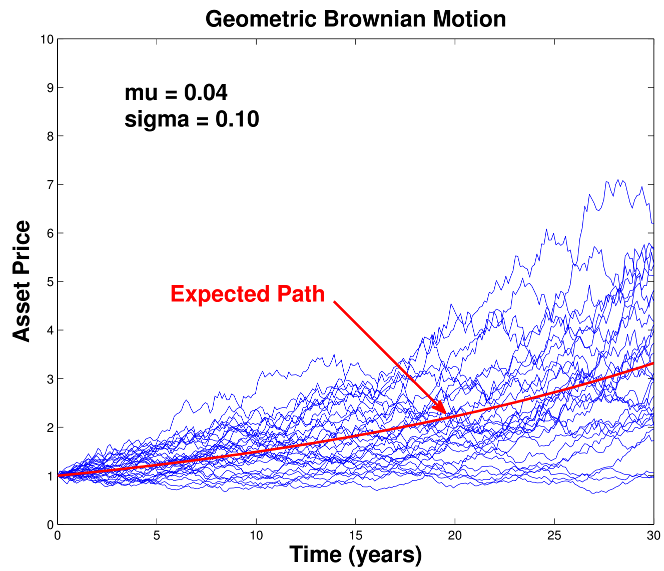



Figure 12.3: Thirty Monte Carlo realized Geometric Brownian Motion paths, compared with mean path. $\sigma = .10$, $\mu = .04$, $T = 30$ years.

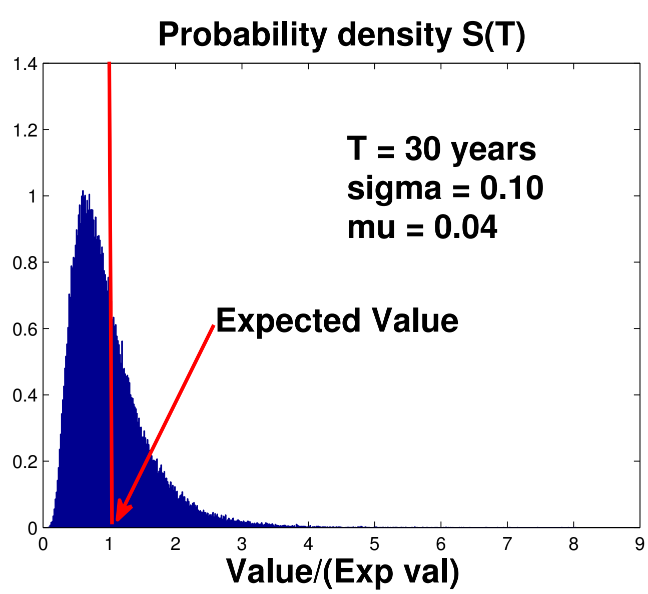



Figure 12.4: Probability density of $S(T)/E[S(T)]$. This is a log-normal density, which is highly skewed. $\sigma = .10$, $\mu = .04$, $T = 30$ years. The median of this distribution is always less than the mean.

## 12.2 Volatility Pumping

Consider an investment with value $S(t_i) = S_i$. We consider some discrete sets of times $t_1, .., t_n$. Let the value of the the investment at $t_{i+1}$ be $S_{i+1} = X_i S_i$ The total relative gain $G^{tot}$ is

$$\begin{aligned} G^{tot} &= S_n/S_1 \\ &= X_1 X_2 ... X_{n-1} \ . \end{aligned} \tag{12.10}$$

The average compound return $A_{ret}$ is given by

$$\begin{aligned} A_{ret} &= \frac{1}{n-1} \log(S_n/S_1) \\ &= \frac{1}{n-1} \sum_{i=1}^{i=n-1} \log X_i \end{aligned} \tag{12.11}$$

Suppose the $X_i$ are random, independent, and identically distributed. Then the expected average compound return is

$$\begin{aligned} E[A_{ret}] &= \frac{1}{n-1} \sum_{i=1}^{i=n-1} E[\log(X_i)] \\ &= E[\log(X)] \end{aligned} \tag{12.12}$$

where $E[\log(X)]$ is the expected value of $\log(X_i)$, which is the same for any period, since we have assumed that the gains $X_i$ are independent and identically distributed. Suppose we have an investment $a$ which has the gain distribution

$$\begin{aligned} X &= 2 \ ; \quad \text{with probability } \frac{1}{2} \\ &= \frac{1}{2} \ ; \quad \text{with probability } \frac{1}{2} \end{aligned} \tag{12.13}$$

The expected average compound return of $a$ is is

$$\begin{aligned} E[(A_{ret})^a] &= \frac{1}{2} \log(2) + \frac{1}{2} \log(\frac{1}{2}) \\ &= 0 \end{aligned} \tag{12.14}$$

Suppose we have investment $b$, has the gain distribution

$$X = 1 \ ; \quad \text{with probability } 1 \ . \tag{12.15}$$

Clearly $b$ has expected average compound return of zero as well, i.e.

$$E[(A_{ret})^b] = 0 \tag{12.16}$$

.

However, consider the following strategy. At the beginning of each period, we divide our total wealth equally into investment $a$ and investment $b$. This is a rebalancing strategy. Now, let's work out our expected average compound return $E[(A_{ret})^{Rebalance}]$

$$\begin{aligned} E[(A_{ret})^{Rebalance}] &= \frac{1}{2} \log(\frac{1}{2} + 1) + \frac{1}{2} \log(\frac{1}{2} + \frac{1}{4}) \\ &\simeq 6\% \ . \end{aligned} \tag{12.17}$$

Let's summarize this result. We have constructed a portfolio consisting of two assets, $a$ and $b$. Each of these assets has zero average compound return. However, the rebalancing strategy generates an average compound return of 6%. This is a bit surprising. This is the whole idea behind the well known strategy of frequently rebalancing a portfolio. This allows us to *buy low, sell high*, and generate a positive average compound return, even though the components of the portfolio may individually have zero expected average compound return.

We can formalize this idea in continuous time in the next section.

### 12.2.1 Constant Proportions Strategy

Suppose we have $n$ assets $S_i; , i = 1, .., n$ which follow

$$dS_i = \mu S_i \ dt + \sigma S_i \ dZ_i. \tag{12.18}$$

Suppose our investment strategy at time $t$ is to invest $n_i(t)$ in each asset. Note that $n_i(t)$ is a function of time in general. As usual, we cannot peek into the future, i.e. we have to pick $n_i(t)$, and then let the asset evolve randomly in $t \rightarrow t + dt$. At $t + dt$, we can then rebalance our portfolio.

Let

$$P(t) = \sum_i n_i(t) S_i(t) \tag{12.19}$$

be the value of our portfolio at $t$. Since $P = P(S_1, ..., S_n)$, we will need the multidimensional version of Ito's Lemma. To keep things general here, suppose that

$$\begin{aligned} dS_i &= a_i \ dt + b_i \ dZ_i \\ G &= G(t, S_1, .., S_n) \\ E[dZ_i dZ_k] &= \rho_{ik} \ dt \end{aligned} \tag{12.20}$$

where $\rho_{ik}$ is the correlation between $dZ_i$ and $dZ_k$. Then

$$dG = \left[ G_t + \sum_k \frac{\partial G}{\partial S_k} a_k + \frac{1}{2} \sum_{k,m} \frac{\partial^2 G}{\partial S_k \partial S_m} b_k b_m \rho_{km} \right] dt + \sum_k \frac{\partial G}{\partial S_k} b_k \ dZ_k \ . \tag{12.21}$$

We will be interested in the log return on our portfolio. Consider $G = \log P(t)$. Using Ito's Lemma (12.21) for the case where $a_i = \mu_i S_i$ and $b_i = \sigma_i S_i$, then

$$\begin{aligned} G_t &= 0 \\ G_{S_k} &= \frac{n_k}{P} \\ G_{S_k S_m} &= -\frac{n_k n_m}{P^2} \ . \end{aligned} \tag{12.22}$$

Combining equations (12.21-12.22), gives

$$dG = \left[ \sum_k \frac{n_k \mu_k S_k}{P} - \frac{1}{2} \sum_{k,m} \frac{n_k n_m S_k S_m \sigma_k \sigma_m \rho_{km}}{P^2} \right] dt + \sum_k \frac{n_k n_k S_k \sigma_k}{P} \ dZ_k \ . \tag{12.23}$$

Now, suppose we choose a constant proportions strategy, i.e we rebalance at every time so that we have a constant weight $w_i$ in each asset,

$$\begin{aligned} w_i &= \frac{n_i S_i}{P(t)} \\ & \sum_i w_i = 1 \ . \end{aligned} \tag{12.24}$$

Note that $w_i$ is independent of $t$, since we rebalance at every opportunity. We can then write equation (12.23) as (using equation (12.24))

$$dG = \left[\sum_k w_k \mu_k - \frac{1}{2} \sum_{k,m} w_k w_m \sigma_k \sigma_m \rho_{km}\right] dt + \sum_k w_k \sigma_k \ dZ_k \ . \tag{12.25}$$

Equation (12.25) has the exact solution

$$\begin{aligned} G(t) &= G(0) + \\ &\left[\sum_k w_k \mu_k - \frac{1}{2} \sum_{k,m} w_k w_m \sigma_k \sigma_m \rho_{km}\right] t + \sum_k w_k \sigma_k (Z_k(t) - Z_k(0)) \ . \end{aligned} \tag{12.26}$$

Now, let's look at what happens for the special case

$$\begin{aligned} \mu_i &= \mu \ ; \forall i \\ \sigma_i &= \sigma \ ; \forall i \\ w_i &= \frac{1}{n} \ ; \forall i \ . \end{aligned} \tag{12.27}$$

Recall that $G(t) = \log(P(t)$, so that, using equation (12.27), equation (12.26) becomes

$$\log\left(\frac{P(t)}{P(0)}\right) = \left[\mu - \sigma^2 \frac{1}{2n^2} \sum_{k,m} \rho_{km}\right] t + \frac{\sigma}{n} \sum_k (Z_k(t) - Z_k(0)) \tag{12.28}$$

so that

$$E\left[\log\left(\frac{P(t)}{P(0)}\right)\right] = \left[\mu - \frac{\sigma^2}{2}\left(\frac{1}{n^2} \sum_{k,m} \rho_{km}\right)\right] t \ . \tag{12.29}$$

As well,

$$var\left[\log\left(\frac{P(t)}{P(0)}\right)\right] = \left[\sigma^2 \left(\frac{1}{n^2} \sum_{k,m} \rho_{km}\right)\right] t \ . \tag{12.30}$$

Recall that for a single asset which follows Geometric Brownian Motion

$$dS = \mu S \ dt + \sigma S \ dZ \tag{12.31}$$

with constant $\mu, \sigma$, then (from equation (2.58))

$$\log\left(\frac{S(t)}{S(0)}\right) = (\mu - \frac{\sigma^2}{2})t + \sigma(Z(t) - Z(0)) \tag{12.32}$$

so that the mean and variance of the log return are

$$\begin{aligned} E\left[\log\left(\frac{S(t)}{S(0)}\right)\right] &= (\mu - \frac{\sigma^2}{2})t \\ var\left[\log\left(\frac{S(t)}{S(0)}\right)\right] &= \sigma^2 t \ . \end{aligned} \tag{12.33}$$

Since

$$\frac{1}{n^2} \sum_{k,m} \rho_{km} \leq 1 \tag{12.34}$$

then, in this special case,

$$
\begin{aligned}
E\left[\log\left(\frac{P(t)}{P(0)}\right)\right] &\geq E\left[\log\left(\frac{S(t)}{S(0)}\right)\right] \\
var\left[\log\left(\frac{P(t)}{P(0)}\right)\right] &\leq var\left[\log\left(\frac{S(t)}{S(0)}\right)\right]
\end{aligned}
\tag{12.35}
$$

so that in this case, using a constant weight portfolio, even if all the assets in the portfolio have the same drift, then the portfolio has a larger drift. Note that the variance of the portfolio is also usually reduced as well (the usual diversification result).

This enhanced drift is sometimes called *volatility pumping*, since the effect is large if the volatility is large. Note that for a single asset, the drift of the log return is reduced to $\mu - \frac{\sigma^2}{2}$ by the volatility. The constant weight portfolio reduces the effect of the volatility on the drift term.

### 12.2.2 Leveraged Two Times Bull/Bear ETFs

Recently, new ETFs (exchange traded fund) have been marketed which allow investors to receive

- Twice the daily return of the index (Bull version).
- Twice the daily negative return of the index (Bear version).

Are these ETFs a good idea? Here is an interesting fact. Suppose you hold equal amounts of both Bull and Bear versions. Then, after a finite period of time (say one year), what is your expected gain? You might think it would be zero. However, if the market is very volatile, you can end up losing on both ETFs.

We can use the analysis of Section 12.2.1 to examine these ETFs. The ETFs are actually an investment strategy which works as follows. Suppose you invest \$100 in the Bull version. The manager of the fund then borrows another \$100, and invests \$200 in the index, at the start of the trading day. The total position of the manager is $(200 - 100) = 100$. At the end the day, suppose the market goes up by 10%, this gives the manager a total position of $(220 - 100) = 120$. Assume you leave your money in the ETF for the following day. At the start of the next trading day, the manager rebalances, so that he has \$240 in the index, and a total borrowing of \$120, so that his total position is $(240 - 120) = 120$. Let's approximate this daily rebalancing by a continuous rebalancing, so we can use equation (12.26).

In this example, we have two assets, the risky index and a risk free bank account, which we assume grows at rate $r$. (This is unrealistic, no bank is going to loan the fund manager at the risk-free rate).

In this case

$$
\begin{aligned}
\mu_1 &= \mu \\
\mu_2 &= r \\
\sigma_1 &= \sigma \\
\sigma_2 &= 0 \\
\rho_{12} = \rho_{21} &= 0 \ .
\end{aligned}
\tag{12.36}
$$

From equations (12.26) and (12.36) we have

$$
\begin{aligned}
E[\log(Return)] &= \big(w_1\mu + w_2 r - \frac{|w_1|^2\sigma^2}{2}\big)t \\
std[\log(Return)] &= \big(|w_1|\sigma\big)\sqrt{(t)} \ ,
\end{aligned}
\tag{12.37}
$$

where $w_1, w_2$ are the weights in each asset, and we have

$$
w_1 + w_2 = 1 \ . \tag{12.38}
$$

For simplicity, let's let $t = 1$ in equation (12.37).

Now, let's suppose that

$$\mu = r + \lambda\sigma \tag{12.39}$$

where $\lambda$ is the market price of risk. Let's examine the *Sharpe ratio* of this investment strategy. We will look at the excess return here, i.e. expected return in excess of the risk-free rate. Our Sharpe ratio in terms of excess return is (using equations (12.38 and 12.39) )

$$\begin{aligned} S_{ratio} &= \frac{w_1\mu + w_2 r - \frac{|w_1|^2\sigma^2}{2} - r}{|w_1|\sigma} \\ &= \frac{w_1\lambda\sigma - \frac{|w_1|^2\sigma^2}{2}}{|w_1|\sigma} \\ &= \mathrm{sgn(w_1)}\lambda - \frac{|\mathrm{w}_1|\sigma}{2} \ . \end{aligned} \tag{12.40}$$

Note that the Bear ETF ($w_1 = -2$) only makes sense if we assume that the drift is less than the risk-free rate, i.e. $\lambda < 0$. Now, suppose $w_1 = 2$, which is the Bull ETF strategy. This gives us

$$S_{ratio}^{Bull} = \lambda - \sigma \ . \tag{12.41}$$

Now, suppose we had an ETF (the Forsyth fund) which rebalanced daily with $w_1 = 1/2$. In this case, we have

$$S_{ratio}^{Forsyth} = \lambda - \frac{\sigma}{4} \tag{12.42}$$

which is always better than equation (12.41), especially if the volatility is high. In other words, you could achieve the same expected gain with a smaller standard deviation by borrowing to invest in the Forsyth Fund, compared to the Bull ETF.

Why to these Bull ETFs exist? Good question. I am sure it generates lots of fees for the fund managers.

## 12.3 More on Volatility Pumping

In the last section we considered optimality in terms of the continuously compounded return, i.e. $\log(P(t)/P(0))$. This is really a type of log utility function. This section is based on the paper *"A dual approach to portfolio evaluation: a comparison of the static, myopic and generalized buy and hold strategies,"* Quantitative Finance Vol 11 (2011) pages 81-99.

You can get further insights here by looking at the actual values of $P(t)/P(0)$. Let's look at a simple case for equation (12.26), with

$$\mu_1 = \mu \ ; \ \mu_2 = r \ ; \ \sigma_1 = \sigma \ ; \ \sigma_2 = 0 \ ; \ \rho_{12} = \rho_{21} = 0 \ . \tag{12.43}$$

From equations (12.26) and (12.43) we get ($G(t) = \log(P(t))$)

$$G(t) = G(0) + (1 - w_1)rt + w_1\mu t - \frac{1}{2}w_1^2\sigma^2 t + w_1\sigma(Z_1(t) - Z_1(0)) \ . \tag{12.44}$$

Some Ito calculus also shows that

$$\frac{dP}{P} = \Big((1 - w_1)r + w_1\mu\Big)dt + w_1\sigma dZ_1 \ . \tag{12.45}$$

The exact solution to equation (12.44) is

$$\frac{P(t)}{P(0)} = e^{(1-w_1)rt + w_1\mu t} \ e^{-w_1^2\sigma^2 t/2} \ e^{w_1\sigma(Z_1(t)-Z_1(0))} \ . \tag{12.46}$$

Now

$$S_1(t) = S_1(0)e^{\mu t-\sigma^2/2t}e^{\sigma(Z_1(t)-Z_1(0))} \ . \tag{12.47}$$

or

$$\left(\frac{S_1(t)}{S_1(0)}\right)^{w_1} = e^{w_1\mu t-w_1\sigma^2/2t}e^{w_1\sigma(Z_1(t)-Z(t))} \ . \tag{12.48}$$

Substituting equation (12.48) into equation (12.46) gives

$$\frac{P(t)}{P(0)} = e^{(1-w_1)rt+w_1(1-w_1)\sigma^2 t/2)}\left(\frac{S_1(t)}{S_1(0)}\right)^{w_1} \ . \tag{12.49}$$

Suppose $S_1(t) = S_1(0)$ (i.e. a sideways market), and $r = 0$, then equation (12.49) gives

$$\frac{P(t)}{P(0)} = e^{\sigma^2 t/2w_1(1-w_1)} \ . \tag{12.50}$$

This means that in a sideways market, we will get a positive return if $0 < w_1 < 1$ (the pumping effect). On the other hand, if $w_1 < 0$ or $w_1 > 1$, we will get a negative return (e.g. leveraged ETFs).

It is also interesting to compare equation (12.49) with a static portfolio, where we take our initial wealth $P(0)$ and buy $w_1 P(0)/S_1(0)$ shares of asset 1, and put $(1-w_1)P(0)$ into the risk free bond (asset 2). In this case, the static or buy and hold portfolio will yield

$$\frac{P(t)}{P(0)} = (1-w_1)e^{rt} + w_1\frac{S_1(t)}{S_1(0)} \ . \tag{12.51}$$

### 12.3.1 Constant Proportion Portfolio Insurance

Another idea for providing some sort of insurance on a portfolio is Constant Proportion Portfolio Insurance (CPPI). In its simplest form, we can consider the possibility of investing in a risky asset (with price $S$) and a risk free bond (price $B$). If $B < 0$, this represents borrowing. Consider a set of rebalancing times $t_i$, with $\Delta t = t_{i+1} - t_i$. Let

$$\begin{aligned}
B(t_i) = B_i &= \text{Amount in risk free asset at } \ t_i \\
S(t_i) = S_i &= \text{Price of risky asset at } \ t_i \\
\alpha(t_i) = \alpha_i &= \text{Number of units of risky asset at } \ t_i \\
P(t_i) = P_i = &= \text{total portfolio at } t_i \\
M(t_i) = M_i &= \text{CPPI multiplier at } \ t_i \\
F(t_i) = F_i &= \text{CPPI floor at } \ t_i \ .
\end{aligned} \tag{12.52}$$

A CPPI strategy is determined by specifying a floor $F_i$ and a multiplier $M_i$. The state of the strategy is given by $\alpha_i$ and $B_i$. At each rebalancing date $t_{i+1}$ the following reallocation takes place

$$\begin{aligned}
\alpha_{i+1} &= M_{i+1}\left[\frac{\max(0, B_i e^{r\Delta t} + \alpha_i S_{i+1} - F_{i+1})}{S_{i+1}}\right] \\
B_{i+1} &= B_i e^{r\Delta t} - (\alpha_{i+1} - \alpha_i)S_{i+1}
\end{aligned} \tag{12.53}$$

Usually $M_i > 1$, so that the strategy becomes more invested in the risky asset as the total wealth $(B_i e^{r\Delta t} + \alpha_i S_{i+1})$ increases above the floor. As the wealth decreases, more assets are diverted to the risk free bond, hence the floor is the lowest possible value of the portfolio (if the rebalancing interval is infinitesimal).

| | |
|---|---|
| $T$ | 1.0 |
| $\sigma$ | .4 |
| $\mu$ (real world drift) | .10 |
| $P_0$ (initial wealth) | 100 |
| $r$ | .02 |
| rebalancing frequency | 250 |
| initial cash | 100 |
| initial risky asset position $\alpha_0$ | 0 |

Table 12.3: Data used in the CPPI strategy experiments.

| F | M | Mean | Standard Deviation | VAR (95%) | CVAR (95%) |
|---|---|---|---|---|---|
| 0 | 1 | 0.021876 | 0.39973 | -0.63697 | -0.80533 |
| 0 | .5 | 0.04094 | 0.19986 | -0.28857 | -0.3724 |
| 0 | 2 | -0.13662 | 0.80046 | -1.4576 | -1.7951 |
| 80 | 2 | .037598 | 0.1809 | -0.1563 | -0.17154 |
| 80 | 4 | 0.0083732 | .33117 | -0.21038 | -0.2142 |

Table 12.4: Results in terms of the log return (12.54), data in Table 12.3. 80,000 Monte Carlo simulations.

Consider the case where $M_i = M = const.$, $F_i = F = const.$. Then $M = 2, F = 0$ corresponds to the Bull ETF. A more conservative constant proportions strategy might be $M = .5, F = 0$. A typical CPPI strategy would be $F = .8S_0, M = 2$. Consider the data in Table 12.3.

We will measure the performance of these strategies by examining the statistics of the log return

$$R = \log(P(T)/P(0)) \ . \tag{12.54}$$

Table 12.4 gives the results for different choices of $M, F$. Table 12.4 indicates that $F = 80, M = 2$ is a good strategy. The leveraged Bull ETF is very bad.

Figure 12.5 shows the density of the log return for the strategy with $F = 80$, $M = 2$. Note the highly skewed distribution.

### 12.3.2 Covered Call Writing

Another popular strategy is covered call writing. There are now some ETFs which follow this strategy.

Basically, you own the stock and sell a call on the stock. The optimal strategy turns out to be to sell a call with a strike just above the current stock price. Using the parameters in Table 12.3, this strategy was simulated with a strike of $K = 100.1$. The results are shown in Table 12.5. The probability density of the log return is shown in Figure 12.6. Note the highly skewed distribution.

If $C(t)$ is the call price, and $P(t)$ is the put price, then Put-Call parity for European options is

$$C(t) - P(t) + Ke^{-r(T-t)} = S \ . \tag{12.55}$$

A covered call is $S - C(t)$, which is equivalent to

$$S - C(t) = Ke^{-r(T-t)} - P(t) \ , \tag{12.56}$$

which is long the risk free account $Ke^{-r(T-t)}$ and short a put. In other words, owning stock worth $K$ and selling a call with strike $K$ is equivalent to investing $K$ in a risk free account and selling a put. This is a bet that the market won't go down, and hence the investor gains the insurance premium (the put). This results in the highly skewed return distribution.

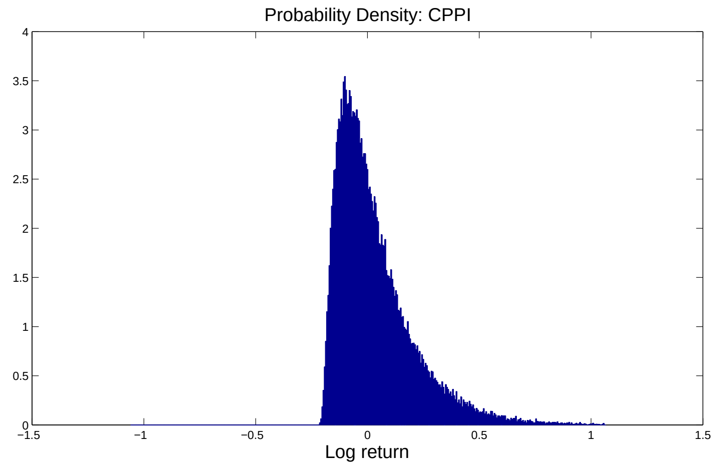



Figure 12.5: Probability density (y-axis) of the log return (12.54), $F = 80, M = 2$.

| Mean | Standard Deviation | VAR (95%) | CVAR (95%) |
|---|---|---|---|
| 0.034853 | 0.18239 | -0.3574 | -0.48052 |

Table 12.5: Results in terms of the log return (12.54), data in Table 12.3. $80,000$ Monte Carlo simulations. Covered call, $K = 100.1$.

Covered call writing is often justified by *back testing*, i.e. seeing what would have happened by applying the strategy to historical market data. This amounts to examining a single stochastic price path. From the probability density shown in Figure 12.6, we can see that the most frequently observed outcome (the highest value on the graph) is quite good. This is, of course, not the mean outcome, and does not take into account the tail of the plot. This explains why backtesting strategies which are very skewed often suggests that these strategies work well (since the historically observed single stochastic path is most likely one where nothing bad happens).

In *"Derivatives in portfolio management: Why beating the market is easy"* (F-S Lhabitant, EDHEC Business School, 2000), the author suggests that use of covered call writing allows fund managers to pretend they are providing superior returns in terms of Sharpe ratios. This paper suggests that this is simply due to the inadequacy of the Sharpe ratio in this situation, since the return distribution is far from normal.

### 12.3.3 Stop Loss, Start Gain

A simple idea for providing some sort of portfolio insurance is to use *stop loss* triggers. This can be set automatically through an online broker. The idea is to set a floor price $K$, and sell the asset if the stock price $S$ falls below $K$. Of course, there is no guarantee that the order will be executed at $K$. What happens is that a *market order* is submitted if the price drops below $K$ and the stock is then sold at whatever bid price appears on the order book. If the stop-loss is triggered, we assume that the investor simply holds the cash in a risk free account until the end of the investment period. We call this the pure stop-loss strategy.

Similarly, if one believes that this is a good investment, one could couple this with a *start-gain*. Once the stock has dropped below $K$ and has been sold, the investor could then set a start-gain order. If the stock rises above $K$, then a market order is submitted, and the stock is purchased at whatever ask price appears on the order book.

Alternatively, one could simple buy a European put option a with strike $K$. This would guarantee that the terminal value of the stock holding would never drop below $K$. In other words, we take out initial investment, buy a put, and then invest whatever is left over in the stock. However, this strategy is

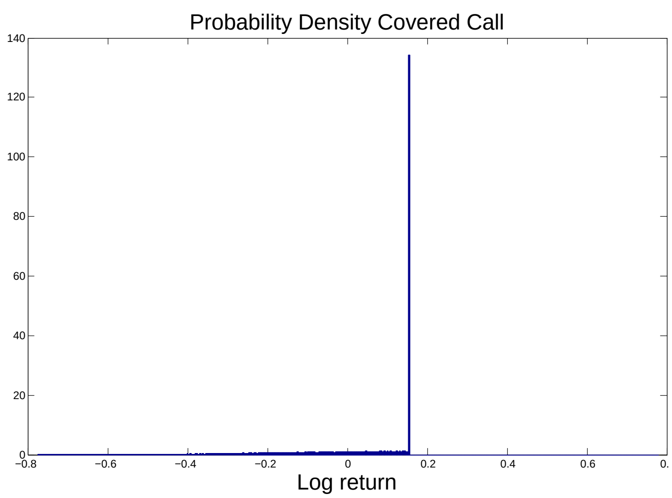



Figure 12.6: Probability density (y-axis) of the log return (12.54), covered call strategy. Data in Table 12.3. $80,000$ Monte Carlo simulations. Covered call, $K = 100.1$.

| | |
|---|---|
| $T$ | 1.0 |
| $P_0$ (initial wealth) | 100 |
| $r$ | .02 |
| rebalancing frequency | Every 3 hours |
| initial cash | 100 |
| initial risky asset position | 0 |
| Stop loss price trigger $K$ | 90 |

Table 12.6: Data used in the stop-loss start-gain strategy experiments. $K$ is the stock price stop-loss trigger. $K$ is the strike price of the European put if the Buy-Put strategy is used.

| $\sigma$ | $\mu$ | Mean | Standard Deviation | VAR (95%) | CVAR (95%) |
|---|---|---|---|---|---|
| Pure Stop-loss | | | | | |
| .2 | .10 | 0.062856 | 0.18242 | -0.10453 | -0.10807 |
| .15 | .10 | 0.077945 | 0.14964 | -0.10252 | -0.10585 |
| Stop-loss start-gain | | | | | |
| .2 | .10 | 0.073376 | 0.19646 | -0.21016 | -0.26076 |
| .15 | .10 | 0.085093 | 0.15126 | -0.15222 | -0.18997 |
| Buy put | | | | | |
| .2 | .10 | 0.068787 | 0.16958 | -0.13479 | -0.13545 |
| .15 | .10 | 0.079694 | 0.13723 | -0.12048 | -0.12095 |

Table 12.7: Comparison of stop-loss start-gain with buy put strategies. Results in terms of the log return of the portfolio. Data in Table 12.6. 80,000 Monte Carlo simulations. GBM with volatility $\sigma$, real world drift $\mu$. Buy-Put strategy includes the cost of buying the put in the portfolio return.

psychologically unappealing, since we have to pay for the put, which represents an immediately realized cost, and this would reduce the return on the portfolio.

Of course, the stop-loss start-gain strategy accumulates costs since we always sell low $(K - \epsilon)$ and buy high $(K + \epsilon)$.

As an example, we will assume GBM, with the parameters given in Table 12.6. The stop-loss start-gain strategy is compared with a buy put strategy in Table 12.7. The density of this strategy is highly skewed, but the buy-put strategy definitely protects better against extreme losses. Stop-loss start-gain strategies are negatively affected by volatility.

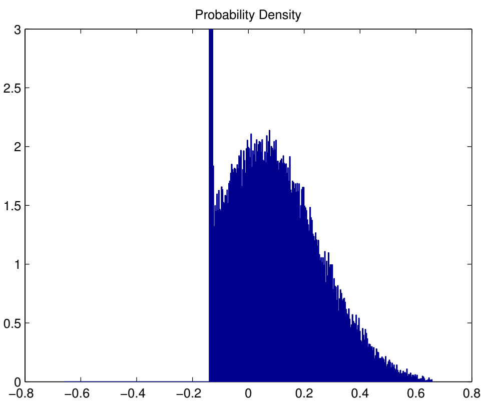
## 12.4 Target Date: Ineffectiveness of glide path strategies

It is often suggested that investors should purchase Target Date (Lifecycle) funds. These funds are for investors saving for retirement in $20 - 30$ years. The idea here is to start off with high proportion in equities (i.e. 100%), and then gradually rebalance to a large proportion in bonds (i.e. 80%) as the target date is approached. We will show here that this strategy does not make much sense.

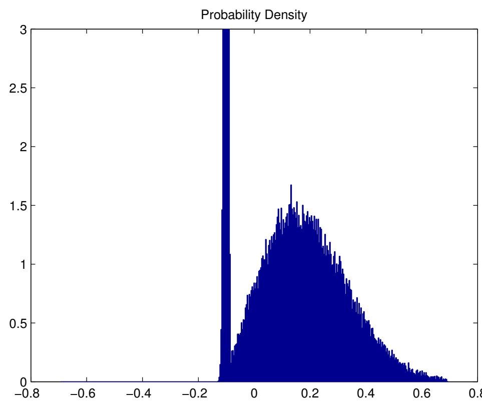
Assume we have two assets, the bond $B$ and risky asset $S$

$$\begin{aligned} dS &= \mu S \; dt + \sigma S \; dZ \\ dB &= rB \; dt \; . \end{aligned} \tag{12.57}$$

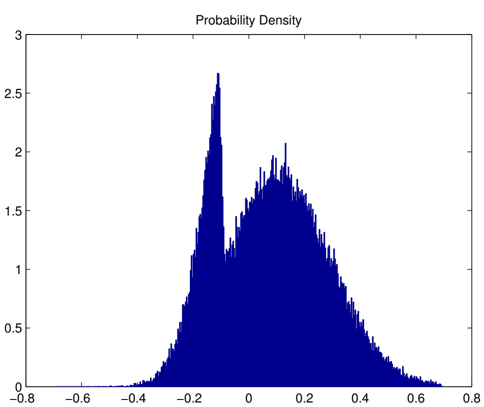
Let $W = S + B$ be the total wealth. Initially, $W = W_0$ at $t = 0$. We will evaluate the performance of different strategies by examining the mean and standard deviation at $t = T$.




(a) Buy Put




(b) Pure Stop-Loss




(c) Stop-Loss Start-Gain

Figure 12.7: Densities of the log return, Buy Put, Pure Stop-Loss and Stop-Loss Start-Gain. Data in Table 12.6. $\sigma = .20$.

Suppose we specify a *deterministic* glide path, i.e.

$$\begin{aligned} p &= \frac{S}{W} \\ &= p(t) \end{aligned} \tag{12.58}$$

with continuous rebalancing, then

$$\begin{aligned} dW &= \mu' W \ dt + \sigma' W \ dZ \\ & \quad \mu' = p\mu + (1-p)r \ \ ; \ \ \sigma' = p\sigma \end{aligned} \tag{12.59}$$

Let $\bar{W}_t = E[W_t]$, then from equation (12.59) we obtain (see Section 2.6.3 for details)

$$d\bar{W}_t = \mu' \bar{W}_t \ dt \tag{12.60}$$

which gives

$$\begin{aligned} \bar{W}_T = E[W_T] &= W_0 e^{\mu^* T} \\ & \quad \mu^* = p^* \mu + (1-p^*) r \\ & \quad p^* = \frac{1}{T} \int_0^T p(s) \ ds \end{aligned} \tag{12.61}$$

Let $G = W^2$. From Ito's Lemma

$$\begin{aligned} dG &= \left[ 2\mu' + (\sigma')^2 \right] W^2 \ dt + 2W^2 \sigma' \ dZ \\ &= \left[ 2\mu' + (\sigma')^2 \right] G \ dt + 2G\sigma' \ dZ \end{aligned} \tag{12.62}$$

Let

$$\bar{G}_t = E[G_t] = E[W_t^2] \tag{12.63}$$

then

$$\begin{aligned} d\bar{G}_t &= \left[ 2\mu' + (\sigma')^2 \right] \bar{G}_t \ dt + 2\bar{G}_t \sigma' \underbrace{E[dZ]}_{=0} \\ &= \left[ 2\mu' + (\sigma')^2 \right] \bar{G}_t \ dt \end{aligned} \tag{12.64}$$

so that

$$\begin{aligned} \bar{G}_T = E[G_T] = E[W_T^2] &= \bar{G}_0 e^{\int_0^T 2\mu' \ dt + \int_0^T (\sigma')^2 \ dt} \\ &= W_0^2 e^{2\mu^* T} e^{\int_0^T (\sigma')^2 \ dt} \end{aligned} \tag{12.65}$$

Combining equations (12.61) and (12.65) gives

$$\begin{aligned} Var[W] &= E[W_T^2] - (E[W_T])^2 \\ &= W_0^2 e^{2\mu^* T} \left( e^{\sigma^2 \int_0^T p(s)^2 \ ds} - 1 \right) \end{aligned} \tag{12.66}$$

or

$$Stndrd \ dev[W_T] = E[W_T] \sqrt{e^{\sigma^2 \int_0^T p(s)^2 \ ds} - 1} \tag{12.67}$$

Now, consider two strategies

- Glide Path: choose a glide path strategy $p = p(t)$.
- Constant Mix: choose a fixed constant fraction $p = p^*$ to invest in equities, where

$$p^* = \frac{1}{T}\int_0^T p(s)\ ds \tag{12.68}$$

From equation (12.61), both these strategies have the same expected value. From equation (12.67), the Constant Mix strategy has standard deviation

$$Stndrd\ dev[W_T] \quad = \quad E[W_T]\sqrt{e^{\sigma^2 (p^*)^2 T} - 1} \tag{12.69}$$

Now,

$$\begin{aligned}
(p^*)^2 T &= \left[\frac{1}{T}\int_0^T p(s)\ ds\right]^2 T \\
&= \left[\int_0^T p(s)\ ds\right]^2 \cdot \frac{1}{T} \\
&\leq \int_0^T p(s)^2\ ds
\end{aligned} \tag{12.70}$$

where the last step follows from the Cauchy-Schwartz inequality.

Consequently, it follows from equation (12.67), (12.69) and (12.70) that

$$\underbrace{Stndrd\ dev[W_T]}_{Constant\ Mix} \quad \leq \quad \underbrace{Stndrd\ dev[W_T]}_{Glide\ Path} \tag{12.71}$$

while from the definition of $p^*$ in equation (12.68) and equation (12.61) we have

$$\underbrace{E[W_T]}_{Constant\ Mix} \quad = \quad \underbrace{E[W_T]}_{Glide\ Path} \tag{12.72}$$

We can summarize this as follows

**Theorem 1 (Ineffectiveness of Glide Path Strategies)** *Consider a market with two assets following the processes (12.57). For any continuously rebalanced deterministic glide path strategy $p = p(t)$ where $p$ is the fraction of total wealth invested in the risky asset, there exists a constant mix strategy $p = p^*$,*

$$p^* = \frac{1}{T}\int_0^T p(s)\ ds \tag{12.73}$$

*such that*

- *The expected value of the terminal wealth is the same for both strategies.*
- *The standard deviation of the glide path strategy cannot be less than the standard deviation of the constant mix strategy*

**Remark 1** *We can get the same result if we assume a jump diffusion process.*

**Example 1 (Linear glide path)** *Suppose we use a linear glide path, i.e.*

$$p(t) \quad = \quad p_{\max} + \frac{t}{T}(p_{\min} - p_{\max}) \quad ; \quad t \in [0, T] \tag{12.74}$$

*which gives us*

$$
\begin{aligned}
p^* &= \frac{1}{T}\int_0^T p(s)\ ds \\
&= \frac{p_{\max}+p_{\min}}{2}\ . \qquad (12.75)
\end{aligned}
$$

*and*

$$
\begin{aligned}
\sigma^2 \int_0^T p(s)^2\ ds &= \sigma^2 \left[ \frac{[p_{\max} + \frac{t}{T}(p_{\min}-p_{\max})]^3}{\frac{3}{T}(p_{\min}-p_{\max})} \right]_0^T \\
&= \sigma^2 T \frac{1}{3} \frac{(p_{\max}^3 - p_{\min}^3)}{p_{\max}-p_{\min}} \qquad (12.76)
\end{aligned}
$$

*Suppose* $p_{\min}=0$*, then from equation (12.75) and equation (12.69), we obtain*

$$
\underbrace{Stndrd\ dev[W_T]}_{Constant\ Mix} = E[W_T]\left( \exp\left( \frac{\sigma^2 p_{\max}^2 T}{4}\right) - 1\right)^{1/2} \qquad (12.77)
$$

*while from equation (12.76) and equation (12.67) we obtain*

$$
\underbrace{Stndrd\ dev[W_T]}_{Linear\ Glide\ Path} = E[W_T]\left( \exp\left( \frac{\sigma^2 p_{\max}^2 T}{3}\right) - 1\right)^{1/2} \qquad (12.78)
$$

The takeaway message here is that you need to use an adaptive strategy (i.e. one that responds to your current wealth) in order to beat a constant mix strategy. This requires solution of an *optimal stochastic control problem.*

### 12.4.1 Extension to jump diffusion case

Let the risky asset $S$ follow the process

$$
\frac{dS_t}{S_{t^-}} = (\mu - \lambda E[J-1])\ dt + \sigma\ dZ + (J-1)\ dq\ \ , \qquad (12.79)
$$

where $\lambda$ is the jump intensity representing the mean arrival rate of the Poisson process

$$
dq = \begin{cases} 0 & \text{with probability } 1-\lambda\ dt \\ 1 & \text{with probability } \lambda\ dt \end{cases} \qquad (12.80)
$$

and when a jump occurs, $S \rightarrow JS$, with $J$ being a random jump size. The Poisson process, the Brownian motion, and the jump size process are all independent.

We assume that the bond process follows

$$
dB = rB\ dt\ . \qquad (12.81)
$$

Let $W = B+S$, then, if $p = S/W$, with continuous rebalancing, then

$$
\begin{aligned}
\frac{dW_t}{W_{t^-}} &= \mu'\ dt - \lambda\left[pE[J] + (1-p) - 1\right]\ dt + \sigma'\ dZ + \left[pJ + (1-p) - 1\right]\ dq \\
&= \mu'\ dt - \lambda p E[J-1]\ dt + \sigma'\ dZ + p(J-1)\ dq \\
&\qquad \mu' = p\mu + (1-p)r \quad ; \quad \sigma' = p\sigma\ . \qquad (12.82)
\end{aligned}
$$

Let $\bar{W}_t = E[W_t]$, then

$$d\bar{W}_t = \mu' \bar{W}_t \ dt \tag{12.83}$$

and

$$\begin{aligned} \bar{W}_T = E[W_T] &= W_0 e^{\mu^* T} \\ & \quad \mu^* = p^* \mu + (1 - p^*) r \\ & \quad p^* = \frac{1}{T} \int_0^T p(s) \ ds \end{aligned} \tag{12.84}$$

Write equation (12.82) as

$$\begin{aligned} \frac{dW_t}{W_{t^-}} &= \hat{\mu} \ dt + \sigma' \ dZ + p(J-1) \ dq \\ & \quad \hat{\mu} = [p\mu + (1-p)r] - \lambda p E[J-1] \ . \end{aligned} \tag{12.85}$$

Now, let $G_t = W_t^2$. From equation (12.85) and Ito's Lemma for jump processes, we obtain

$$\frac{dG_t}{W_{t^-}^2} = [2\hat{\mu} + (\sigma')^2] \ dt + 2\sigma' \ dZ + [(pJ + (1-p))^2 - 1] \ dq \tag{12.86}$$

or

$$\begin{aligned} \frac{dG_t}{G_{t^-}} &= [2\hat{\mu} + (\sigma')^2] \ dt + 2\sigma' \ dZ + [(p(J-1)+1)^2 - 1] dq \\ &= [2\hat{\mu} + (\sigma')^2] \ dt + 2\sigma' \ dZ + [(p^2 (J-1)^2 + 2p(J-1)] dq \end{aligned} \tag{12.87}$$

Let $\bar{G}_t = E[G_t] = E[W_t^2]$, so that, from equation (12.87) we obtain

$$\frac{d\bar{G}_t}{\bar{G}_t} = [2\hat{\mu} + (\sigma')^2] \ dt + \left( \lambda p^2 E[(J-1)^2] + 2\lambda p E[(J-1)] \right) \ dt \tag{12.88}$$

which gives

$$\begin{aligned} \bar{G}_T &= G_0 \exp \Bigg[ \Big( 2(p^* \mu + (1-p^*) r) \ - \ 2\lambda p^* E[J-1] \Big) T \\ & \qquad + \ (\sigma^2 + \lambda E[(J-1)^2]) \int_0^T p^2 \ dt \ + \ \Big( 2\lambda p^* E[J-1] \Big) T \Bigg] \end{aligned} \tag{12.89}$$

or

$$E[W_T^2] = (E[W_T])^2 \exp \left[ (\sigma^2 + \lambda E[(J-1)^2]) \int_0^T p^2 \ dt \right] \tag{12.90}$$

From $Var[W_T] = E[W_T^2] - (E[W_T])^2$, we obtain

$$Stndrd \ Dev[W_T] = E[W_T] \left( \exp \left[ (\sigma^2 + \lambda E[(J-1)^2]) \int_0^T p^2 \ dt \right] - 1 \right)^{1/2} \tag{12.91}$$

Now, using equation (12.70), we can easily prove the following result.

**Theorem 2 (Ineffectiveness of glide path strategies: jump diffusion case)** *Consider a market with two assets following the processes (12.79) and (12.81). For any continuously rebalanced deterministic glide path strategy $p = p(t)$ where $p$ is the fraction of total wealth invested in the risky asset, there exists a constant mix strategy, with $p^*$ being the constant fraction invested in the risky asset, where*

$$p^* = \frac{1}{T} \int_0^T p(s)\ ds \tag{12.92}$$

*such that*

- *The expected value of the terminal wealth is the same for both strategies.*
- *The standard deviation of the glide path strategy cannot be less than the standard deviation of the constant mix strategy*

### 12.4.2 Dollar cost averaging

Suppose you have a lump sum $W_0$ to invest. You have a target asset mix of $\hat{p}$ fraction in the risky asset. Is it better to gradually buy stocks? This is often suggested, since this way you are *dollar cost averaging.*

Let $t^*$ be the period over which you dollar cost average, and $T$ be the target date of your investment. Then, dollar cost averaging amounts to this deterministic strategy

$$p(t) \quad = \quad \min\left(\frac{t\hat{p}}{t^*}, \hat{p}\right) . \tag{12.93}$$

However, from Theorem 1, this strategy cannot be better than buying $p^* W_0$ right away ($p^*$ defined in equation (12.73)).

## 12.5 Bootstrap Resampling

An alternative to using Monte Carlo simulation to evaluate investment strategies is *bootstrap resampling.* This method samples historical returns directly, thus avoiding the need to estimate parameters for a parametric stochastic model. This is also considered to be superior to the the traditional *rolling period* method. [1]

As a first step, the historical index would be converted to a list of returns, i.e. if $S_i$ is the stock index at time $t_i = i\Delta t$, then the log return $\Delta X_i$ for period $(t_i, t_{i+1})$ would be

$$\Delta X_i \quad = \quad \log(S_{i+1}/S_i) \tag{12.94}$$

This set of returns $\{\Delta X_i\}$ would be the input data for the bootstrap resampling algorithm. As an example, we will assume that the data is converted to monthly returns, i.e. $\Delta t = 1/12$ years. Of course, any other sampling period could be used.

A single bootstrap resampled path can be constructed as follows. Divide the total investment horizon of $T$ years into $k$ blocks of size $b$ years, so that $T = kb$. We then select $k$ blocks at random (with replacement) from the historical return data. Each block starts at a random month. A single path is formed by concatenating these blocks. The historical data is wrapped around to avoid end effects, as in the circular block bootstrap [2]. This procedure is then repeated for many paths.

The sampling is done in blocks in order to account for possible serial dependence effects in the historical time series. The choice of blocksize is crucial and can have a large impact on the results. If we were interested in an investment strategy where the portfolio was a stock index and a bond index, then we would simultaneously sample stock and bond returns from the historical data. This would preserve and correlations between stock and bond returns.

[1] H. Dichtl, W. Drobetz and M. Wambach, "Testing rebalancing strategies for stock-bond portfolios across different asset allocations," Applied Economics, Vol 48 (2016) 772-788.

[2] D. Politis and J. Romano, "The Stationary Bootstrap", Journal of the American Statistical Association, Vol 89 (1994) 1303-1313.

To reduce the impact of a fixed blocksize and to mitigate the edge effects at each block end, we use the stationary block bootstrap. The blocksize is randomly sampled from a geometric distribution with an expected blocksize $\hat{b}$. The probability that the blocksize $b$, drawn from a geometric distribution, is $k$, is

$$\begin{aligned} Pr[b = k] &= (1-p)^{k-1} p \ ; \ k = 1, 2, \ldots \\ p &= \frac{1}{\hat{b}} \end{aligned} \tag{12.95}$$

The pseudo-code for block bootstrap resampling is given in Algorithm 12.1. This algorithm returns a single bootstrapped path. We would generate many such paths. In the case of a stock and bond portfolio, then we would this algorithm to simultaneously draw returns from various stock and bond indexes to evaluate an investment strategy.

```
Require: Function geo( b̂ ) ;  returns draw from a shifted geometric distribution with mean b̂
Require: Function rand_int( N ) ;  returns a uniformly distributed draw from {1, 2, ..., N}

  input: Vector of returns ΔX_i ; i = 1, ..., N ;  Expected block size b̂ ;  Number of samples N_s
  initialize: actual_block_size = 0 ; total_samples = 1  ;  sub_block_total = N

  while (total_samples ≤ N_s) do
    if (sub_block_total > actual_block_size) then
      actual_block_size = geo( b̂ )   {restart subblock}
      index = rand_int( N )          {choose random starting index}
      sub_block_total = 1
    end if
    index = mod( index, N)           {circular bootstrap}
    out_array( total_samples ) = ΔX_index
    index += 1 ; total_samples += 1 ;  sub_block_total += 1
  end while

  return out_array(i) ; i = 1, ..., N_s
```

**Algorithm 12.1:** A single stationary block bootstrapped sequence of returns.

### 12.5.1 Data and Investment Portfolio

We assume that the investment portfolio consists of only two assets. One of the assets is a broad stock market index, and the other asset is a risk-free bond.

The data used in this work was obtained from Dimensional Returns 2.0 under license from Dimensional Fund Advisors Canada. In particular, we use the Center for Research in Security Prices (CRSP) Deciles (1-10) index. This is a total return value-weighted index of US stocks. We also use one month Treasury bill (T-bill) returns for the risk-free asset.[3] Both the equity returns and the Treasury bill returns are in nominal terms, so we adjust them for inflation by using the US CPI index. We use real indexes since long-term retirement saving should be attempting to achieve real (not nominal) wealth goals. All of the data used was at the monthly frequency, with a sample period of 1926:1 to 2017:12.

A crucial parameter for block bootstrap resampling is the expected blocksize. We have carried out our tests using a range of expected blocksizes. Although the absolute performance of variance strategies is mildly sensitive to the choice of blocksize, the relative performance of the various strategies appears to be insensitive to blocksize. We show results for a blocksize of two years in the following.

[3]We have also carried out tests using a 10 year US treasury as the bond asset. The results are qualitatively similar to those reported in this section

| | |
|---|---|
| Investment horizon (years) | 45 |
| Equity market index | Value-weighted CRSP deciles 1-10 US market index |
| Risk-free asset index | 1-month T-bill |
| Historical Period | 1926:1 - 2017:12 |
| Initial investment $W_0$ | 500 |
| Real investment each year | 20.0 ($0 \le t_i \le 15$), $-40.0$ ($16 \le t_i \le 45$) |
| Rebalancing interval (years) | 1 |

Table 12.8: Input data for examples. Cash is invested at $t_i = 0, 1, \ldots, 15$ years, and withdrawn at $t_i = 16, 17, \ldots, 45$ years. Units for real investment: thousands of dollars. All indexes are deflated using the US historical CPI index.

### 12.5.2 Investment scenario

Table 12.8 shows the parameters for our investment scenario. Let's call our investor Bob, whose cumulative saving and investments totals \$500,000 by at age 50. We assume that Bob has with a constant salary of \$100,000 per year (real), and saves 20% of his salary for 15 years (i.e. this would include both employee and employer contributions to a tax advantaged registered account), until retirement at age 65. Upon retirement, Bob withdraws \$40,000 for 30 years (adjusted for inflation) from his investment account. We assume that government benefits for Bob (CPP and OAS), will total about \$20,000 (in constant dollars), which will mean that Bob collects a total of \$60,000 per year in retirement, which is about 60% of pre-retirement income. Note that these amounts are all real (inflation adjusted). We do not consider escalating the (real) contribution during the accumulation phase (which also impacts the desired replacement ratio), although this is arguably more realistic.

Assuming flat contributions and withdrawals, we can interpret the above scenario as an investment strategy which allows real withdrawals of twice as much as real contributions. We shall see that this rather modest objective still entails significant risk. As indicated in Table 12.8, we assume yearly rebalancing.[4]

### 12.5.3 Deterministic Strategies

Defining the fraction invested in the stock index as $p$, then a common strategy is to simply rebalance to a constant weight at regular intervals (in this case we are assuming yearly).

Alternatively, target date funds specify a time dependent fraction in stocks $p(t)$, with a high equity fraction in the early years, and then reducing the equity fraction to and through retirement. The model glide path (MGP) below is characteristic of a typical target date fund (TDF) path, where we assume that $t = 0$ corresponds to age 50.

$$
\begin{aligned}
p(t) \quad &= \quad \max(p_{\min}, p^*) \\
& \quad\quad p^* = p_{\max} + \frac{t}{t_{const}}(p_{\min} - p_{\max}) \\
& \quad\quad p_{\max} = .75 \ ; \ p_{\min} = .30 \ ; \ t_{const} = 22
\end{aligned} \tag{12.96}
$$

This glide path is not the path used by any particular fund, but has the characteristics of a typical TDF.

### 12.5.4 Criteria for Success

In practice, it is usually the case the investors saving for retirement are most concerned with running out of savings during retirement. In our tests of various strategies, we will focus on the risk of shortfall, as measured by

[4]More frequent rebalancing has little effect for long-term (> 20 years) investors.

| Method | $Median[W_T]$ | $std[W_T]$ | $Pr[W_T < 0]$ | 5% CVAR | TimeWtd Equity Fraction |
|---|---|---|---|---|---|
| $p = .35$ | 1100 | 1358 | .092 | -324 | .35 |
| $p = .45$ | 1852 | 2370 | .060 | -275 | .45 |
| $p = .55$ | 2821 | 4077 | .046 | -255 | .55 |
| $p = .65$ | 4036 | 6895 | .044 | -265 | .65 |
| $p = .75$ | 5512 | 11476 | .044 | -303 | .75 |
| $p = .85$ | 7260 | 18823 | .047 | -366 | .85 |
| MGP (12.96) | 1661 | 2105 | .073 | -335 | .42 |

Table 12.9: Scenario in Table 12.8. $p$ denotes the fraction in the stock index, with yearly rebalancing. $W_T$ denotes real terminal wealth after 45 years, measured in thousands of dollars. Statistics are based on $100,000$ bootstrap resamples of the historical data, blocksize 2 years. The historical data is from the CRSP stock index $1926 : 1 - 2017 : 12$. The bond index is the the 30-day Tbill returns over the same period. Both indexes are deflated using the CPI index, so all results are in real dollars. *"MGP"* refers to the model glide path, in equation (12.96).

- Probability of exhausting savings, i.e. $Pr[W_T < 0]$. Actuaries call this the probability of ruin.
- Conditional value at risk, denoted by CVAR(x). This is the mean of the worst x% of outcomes. In our case, we will consider a CVAR at the 5% level. This tells us, in the event that the strategy has poor results, how bad the result is, on average. This is a bit more informative than simply looking at the probability of ruin.

### 12.5.5 Bootstrap results

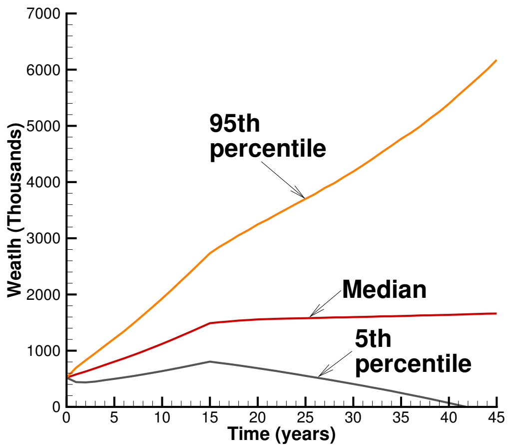
Table 12.9 compares the bootstrapped results for various strategies. Note that the glide path (MGP) strategy is not better than a constant weight strategy. This is consistent with the theoretical results in Section 12.4. Note that CVAR at the 5% level is negative, which means that if Bob lives for 30 years, then there is a non-negligible chance that he will run out of savings. However, if Bob has other assets, i.e. a residential property, then this risk can be hedged by using (if necessary) a reverse mortgage on this asset. A reasonable risk management approach here is to set the withdrawal rate so that the CVAR risk is less than the value of Bob's real estate assets.

Figure 12.8 shows the percentiles of wealth (as a function of time) for the model glide path (MGP) as in equation (12.96), and for the $p = .45$ case. Both strategies fail (run out of savings) before the end of the retirement phase at the fifth percentile.

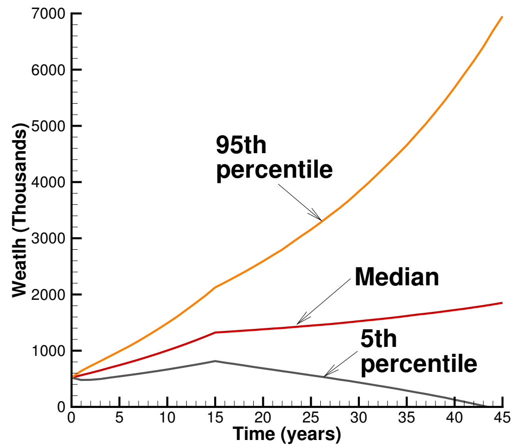
Figure 12.9 shows the percentiles of wealth for the aggressive strategy of a constant weight $p = .75$ to and through retirement. Rather surprisingly, in this case the strategy does not fail at the fifth percentile. However, the CVAR(5%) risk is worse in this case, compared to the $p = .45$ case. This indicates that simply looking at probability of ruin does not give an entire picture of the strategy. From a probability of ruin metric, $p = .75$ looks superior to $p = .45$, but this is not reflected in the CVAR(5%) statistic (see Table 12.9).

## 12.6 Maximizing Sharpe ratios

The ubiquitous Sharpe ratio is a commonly used measure of investment performance. However, the Sharpe ratio is prone to manipulation. Any strategy which includes non-linear payoffs (e.g. options) can produce an apparent outperformance (Dybvig and Ingersoll, 1982; Lhabitant, 2000; Goetzmann et al., 2002). As noted in (Spurgon, 2001), selling off the right side of the terminal wealth distribution can improve the Sharpe ratio.




(a) Percentiles of wealth: MGP (12.96)




(b) Percentiles of wealth: $p = .45$

Figure 12.8: Scenario in Table 12.8. Statistics based on $100,000$ bootstrap resamples of the historical market returns. Blocksize 2 years.

Such a strategy is easily implemented by owning the underlying asset and selling out of the money calls on the asset (covered call writing).

Of course, in a complete market, options can be replicated by dynamically trading stocks and bonds. Consequently any portfolio containing options is equivalent to a dynamic trading strategy. Hence Sharpe ratios can be maximized by using optimal stochastic control techniques, coupled with a suitable objective function. An interesting corollary to this observation is the use of stochastic control in fraud detection (Bernard and Vanduffel, 2014).

Let $r$ be the risk-free return, and $T$ be the investment horizon. $W_t$ is the wealth of a portfolio at time $t$. The continuously compounded Sharpe ratio is then defined as

$$\mathcal{S} = \frac{E[W_T] - W_0 e^{rT}}{std[W_T]}, \tag{12.97}$$

where $E[\cdot]$ is the expectation, and $std[\cdot]$ is the standard deviation. Note that $\mathcal{S}$ is defined in terms of the terminal wealth and standard deviation at time $T$ (Lhabitant, 2000; Goetzmann et al., 2002; Bernard and Vanduffel, 2014), in contrast to the *instantaneous* Sharpe ratio, which is defined in terms of averaging short period returns. Clearly, if we are examining dynamic trading strategies, it does not make sense to average short period returns in $[0, T]$.

Consider a market containing a stock index and a risk-free bond. Let the amount invested in the stock index be $S_t$, and the amount in the risk-free bond be $B_t$. We assume that

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu \, dt + \sigma \, dZ \\ \frac{dB_t}{B_t} &= r \, dt \, , \end{aligned} \tag{12.98}$$

where $\mu$ is the stock drift rate, $\sigma$ is the volatility, and $dZ$ is the increment of a Wiener process. Let $p$ be the fraction of the total portfolio $W_t$ invested in the stock. Assuming continuous rebalancing, then the process

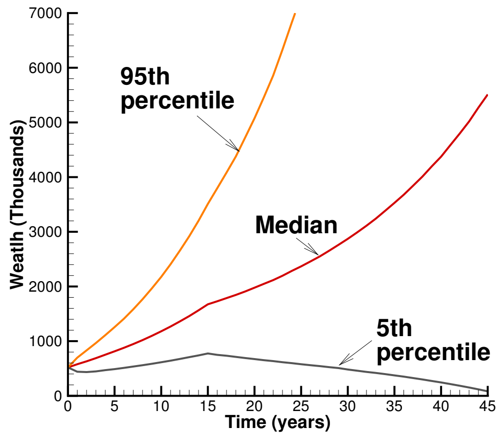



Figure 12.9: Scenario in Table 12.8. Percentiles of wealth: $p = .75$. Statistics based on 100,000 bootstrap resamples of the historical market returns. Blocksize 2 years. Source: author computations.

for $W_t$ is

$$\begin{aligned} \frac{dW_t}{W_t} &= p\frac{dS_t}{S_t} + (1-p)\frac{dB_t}{B_t} \\ &= (r + p(\mu - r))\ dt + p\sigma dZ\ . \end{aligned} \tag{12.99}$$

Given an initial investment $W_0$ at $t = 0$, with terminal wealth $W_T$, we can pose the problem of determining the optimal control $p(W_t, t), t \in [0, T]$ in terms of a mean variance objective. Defining a scalarization parameter $\kappa > 0$, then the mean variance problem can be formulated as

$$\sup_{p(\cdot)} E[W_T] - \kappa \text{Var}[W_T]\ . \tag{12.100}$$

Varying $\kappa$ in equation (12.100) traces out the efficient frontier. Problem (12.100) cannot be solved directly using dynamic programming. From (Zhou and Li, 2000; Li and Ng, 2000), we learn that we can determine the control $p(\cdot)$ which maximizes objective function (12.100) by solving the alternative problem

$$\inf_{p(\cdot)} E[(W^* - W_T)^2]\ , \tag{12.101}$$

where, by varying $W^*$, we trace out the efficient frontier. Note that Problem 12.101 can be solved using dynamic programming.[5]

The optimal control for Problem 12.101 is given by (Zhou and Li, 2000; Li and Ng, 2000; Vigna, 2014)

$$\begin{aligned} p &= \frac{\xi}{\sigma W_t}\left(W^* e^{-r(T-t)} - W_t\right) \\ &\qquad \xi = \frac{\mu - r}{\sigma}\ . \end{aligned} \tag{12.102}$$

Let $W_T^{opt}$ denote the terminal wealth under strategy (12.102). The efficient frontier is then given by the straight line (Zhou and Li, 2000; Li and Ng, 2000)

$$E[W_T^{opt}] = W_0 e^{rT} + \left(e^{\xi^2 T} - 1\right)^{1/2} \text{std}(W_T^{opt})\ . \tag{12.103}$$

[5] In Vigna (2014), it is shown that $W_t < W^*, \forall t$ under the optimal control.

Recall that varying $W^*$ will move us along the efficient frontier. From equations (12.97) and ( 12.103), the optimal Sharpe ratio is

$$\mathcal{S}^{opt} = \left( e^{\xi^2 T} - 1 \right)^{1/2} . \tag{12.104}$$

On the other hand, suppose we rebalance to a constant weight, i.e. $p = const.$ in equation (12.99). Let $W_T^p$ denote the terminal wealth under a constant weight strategy $p$. Then we have

$$\begin{aligned} E[W_T^p] &= W_0 e^{(p(\mu - r) + r)T} \\ \mathrm{std}[W_T^p] &= E[W_T^p] \left( e^{(\sigma p)^2 T} - 1 \right)^{1/2} , \end{aligned} \tag{12.105}$$

with Sharpe ratio

$$\begin{aligned} \mathcal{S}^p &= \frac{1 - e^{-p(\mu - r)T}}{\left( e^{(\sigma p)^2 T} - 1 \right)^{1/2}} \\ &\simeq \xi \sqrt{T} = \frac{(\mu - r)\sqrt{T}}{\sigma} \ ; \ T \to 0 \text{ or } \ p \to 0 . \end{aligned} \tag{12.106}$$

Note that the continuously compounded Sharpe ratio is a function of $p$ in general, and approaches the instantaneous Sharpe ratio only in the limit as $T \to 0$ or $p \to 0$.

Let

$$E[W_T^{opt,p}] = W_0 e^{rT} + \left( e^{\xi^2 T} - 1 \right)^{1/2} \mathrm{std}(W_T^p) . \tag{12.107}$$

This can be interpreted as follows. Given a constant weight strategy with equity fraction $p$, which generates $\mathrm{std}(W_T^p)$, then $E[W_T^{opt,p}]$ is the expected terminal wealth under control (12.102) which has the same the standard deviation $\mathrm{std}(W_T^p)$ (this follows from equation (12.103) ).

A convenient way to compare these strategies is through the *apparent* annualized $\alpha$, which we define as

$$\alpha^p = \frac{\log(E[W_T^{opt,p}]) - \log(E[W_T^p])}{T} , \tag{12.108}$$

which is the extra annualized expected return generated by strategy (12.102) compared to the constant weight strategy with equity fraction $p$, given that both strategies have the same risk, as measured by standard deviation. Consistent with Goetzmann et al. (2002), we term $\alpha^p$ the apparent $\alpha$, since there is no stock-picking skill involved here, merely use of a dynamic control.

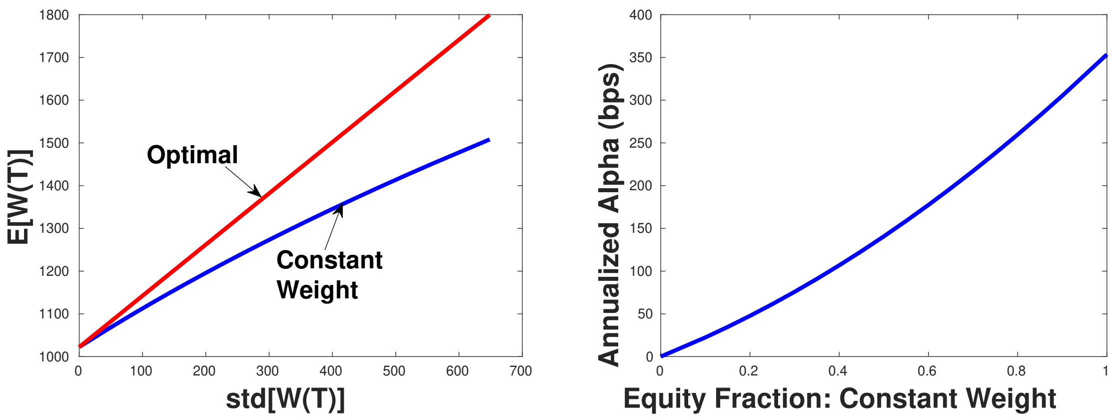
### 12.6.1 Numerical Examples

We use data from the Center for Research in Security Prices (CRSP) on a monthly basis over the 1926:1-2019:12 period.[6] Our base case tests use the CRSP 30 day T-bill for the bond asset and the CRSP value-weighted total return index for the stock asset. This latter index includes all distributions for all domestic stocks trading on major U.S. exchanges. All of these various indexes are in nominal terms, so we adjust them for inflation by using the U.S. CPI index, also supplied by CRSP. Maximum likelihood fits to the data are given in Table 12.10.

[6]More specifically, results presented here were calculated based on data from Historical Indexes, ©2020 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Wharton Research Data Services was used in preparing this article. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

| | | | |
|---|---|---|---|
| $\mu$ | .0822 | $W_0$ | 1000 |
| $\sigma$ | .1842 | T | 5.0 years |
| r | .0044 | Data | CRSP |

Table 12.10: Estimated annualized parameters for processes (12.98), annualized. Value-weighted CRSP index, 30 day T-bills, deflated by the CPI. Sample period 1926:1 to 2019:12.




(a) Efficient frontiers, constant weight compared to optimal control.

(b) Annualized $\alpha$, optimal vs. constant weight strategy.

Figure 12.10: $W_0 = 1000$, $T = 5$ years. Parameters based on CRSP inflation adjusted data, 1926:1-2019:12.

In Figure 12.10 we compare the efficient frontiers for the optimal strategy (12.102) compared to a constant weight strategy, and the annualized $\alpha$ as determined by equation (12.108). We determine points on the efficient frontier by varying $W^*$ in equation (12.101) and varying the constant weight $p$ in equation (12.105. The frontiers are shown using the data in Table 12.10 with $W_0 = 1000$ and $T = 5$ years.

Figure 12.10 shows that the dynamic strategy impressively outperforms a constant weight strategy. However, this is somewhat misleading. The optimal control (12.102) is unconstrained, which allows for infinite leverage and borrowing if insolvent.

A rigorous solution of Problem 12.100 with no-shorting and no-leverage constraints requires numerical solution of a Hamilton-Jacobi-Bellman (HJB) equation Wang and Forsyth (2010). However, it is more instructive for our purposes to approximate the constrained control using the approach in Vigna (2014). We constrain the unconstrained control so that there is no-shorting and no-leverage.

$$
\begin{aligned}
p^* &= \frac{\xi}{\sigma W_t}\left( W^* e^{-r(T-t)} - W_t \right) \\
p &= \max(0.0, \min(p^*, 1.0)) \ . \qquad (12.109)
\end{aligned}
$$

The apparent $\alpha$, using equation (12.109) for various values of the constant weight control is shown in Figure 12.11. This curve was computed using Monte Carlo simulation of the wealth process (12.99), using 300 timesteps and $64 \times 10^3$ simulations. The constraints force $\alpha$ to zero for $p = 0$ and $p = 1.0$. The maximum $\alpha$ occurs when the benchmark strategy has constant weight $p \simeq 0.6$.

**Remark 2 (Pre-commitment policy)** *Strategy 12.102 is the pre-commitment solution, which is not necessarily time consistent. In other words, the investor may have an incentive to deviate from the optimal*

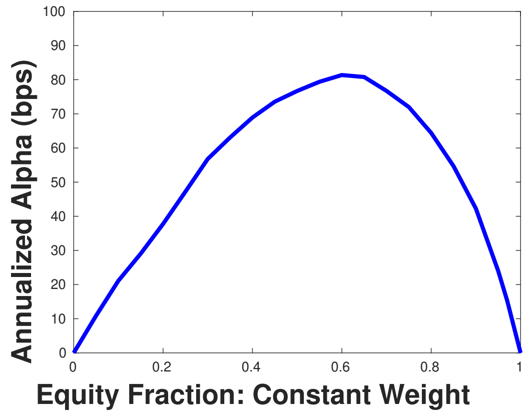



Figure 12.11: Constrained control (12.109), no-shorting, no leverage. $W_0 = 1000$, $T = 5$ years. Parameters based on CRSP inflation adjusted data, 1926:1-2019:12.

*policy determined at time zero at later times. Hence, some authors have labelled pre-commitment polices as non-implementable. However, consider the the case of a retail investor, who purchases a financial product from a financial institution. The investor does not trade herself (in the assets underlying the product) during the lifetime of the contract. Performance of the is product is evaluated in terms of the attainment of goals devised at the inception of the contract. Hence the pre-commitment policy is appropriate in this case.(Bernard and Vanduffel, 2014)*

**Remark 3 (Pre-commitment strategies equivalence to induced time consistent strategy)** *The control (12.102) is formally the pre-commitment policy. However, the the time zero strategy based on the pre-commitment policy solution of Problem 12.100 identical to the strategy for an induced time consistent policy, hence is implementable.*[7] *The induced time consistent strategy in this case is a target based shortfall, Problem 12.101, with a fixed value of $W^*$ $\forall t > 0$. The concept of induced time consistent strategies is discussed in Strub et al. (2019). The relationship between pre-commitment and implementable target based schemes in the mean-variance context is discussed in Vigna (2014) and Menoncin and Vigna (2017).*

In the complete market case, dynamic trading in the stock and bond is equivalent to using options in the trading strategy. Hence, even if options are not directly including in, for example, strategy (12.102), this is clearly equivalent to use of derivatives. Hence, we can think of any financial product which employs a dynamic trading strategy as a structured product. In the presence of constraints, the market may not be complete. However, with some abuse of common terminology, we will refer to such packaged investment vehicles as structured products, even if the market is incomplete.

### 12.6.2 Deficiencies of mean-variance (Sharpe ratio) criteria

The apparent $\alpha$ generated by a mean-variance optimal strategy comes about from skewing the terminal wealth distribution. The right side of the distribution is cut-off (eliminating very large gains), and at the same time, an increase in left tail risk occurs Lhabitant (2000); Goetzmann et al. (2002); Forsyth and Vetzal (2017a,b, 2019). This can also be viewed as a natural consequence of the control (12.102), which increase the weight in stocks when wealth decreases, and decreases wealth in stocks when wealth increases, i.e. buy when the market goes down, sell when the market goes up. This has the implication that the investor is fully invested in bonds after stocks do well, and will not participate in further gains. On the other hand, the investor increases holdings in stocks when stocks perform poorly. This has the consequence that poor results

[7] An implementable strategy has the property that the investor has no incentive to deviate from the strategy computed at time zero at later times (Forsyth, 2020).

can be expected if the market trends downward over the entire investment horizon. This will generate a worse result than a constant weight strategy, which has at least some proportion of wealth always invested in bonds. The opposite is true at large values of wealth. In this case, the optimal strategy is always in invested in bonds, while the constant weight strategy has some investment in stocks, and can participate in further stock gains.

In summary, we can see that a major problem with dynamic mean-variance (Sharpe ratio maximizing) strategies is that variance is a symmetric risk measure, which penalizes upside as well as downside. An easy way to improve Sharpe ratios is to sell off the upside, which trivially reduces variance. On the other hand, somewhat counterintuitively, the extreme left tail of the distribution is worse compared to the benchmark constant weight strategy.

Note that in some cases, an asset allocation strategy based on objective function (12.100), see for example Forsyth and Vetzal (2019), may be desirable, based on the CDF of the final wealth distribution. We emphasize that mean-variance (i.e. Sharpe ratio) criteria are not particularly informative when dealing with dynamic trading strategies, or, equivalently, portfolios containing options.

## 12.7 Capitalization Weight vs. Equal Weighted Indexes

Nowadays, most retail customers buy index exchange traded funds (ETFs). A good portfolio for a Canadian investor, for example, could be constructed using

- A domestic index portfolio (TSX 60)
- A US stock index (S&P 500)
- An *ex-North America* index
- A domestic bond index

Of course, the trick is to try to optimally allocate wealth between these four indexes. A simple strategy is to choose constant weights, and rebalance to those weights at yearly or quarterly intervals.

Most standard ETFs use a capitalization (cap) weighted index. In other words, the proportion of each stock in the index is just the market capitalization of that stock divided by the total stock market capitalization. If there are no complications such as companies going bankrupt, or new companies being added to the index, then an advantage of a cap weighted index is that once it is set up, very little trading is required (e.g. reinvestment of dividends).

But there are many other possible ways to construct an index. We are going to look at an equal weighted index, compared to the usual cap weighted index.

There are two interesting papers on this topic:

1. *"Why does an equal-weighted portfolio outperform value-and price-weighted portfolios?"* Y. Plyakha, R. Uppal, G. Vilkov, (2012) SSRN 2724535.
2. *"Why has the equal weight portfolio underperformed and what can we do about it?* B. H. Tljaard and E. Mare, *Quantitative Finance* 21:11 (2021)

Paper [1] (Plyakha et al, 2012) suggests that equal weight indexes outperform cap weighted indexes over the long term. However, paper [2] (Tljaard et al, 2021) does not disagree, but suggests that the equal weight index may underperform for short periods, in particular, the last decade.

### 12.7.1 Stochastic Dominance

How do we compare the performance of various investment strategies, or indexes? Simplistic measures are things like mean, variance and Sharpe Ratio. However, these are just a few summary statistics of investment performance. We would like to look at the entire probability distribution function for each strategy.

Suppose we have an investment strategy, which, starting at time zero, generates wealth $W$ at time $T$. Of course, in the real world, $W$ is a random variable, with probability density $p(W)$. The cumulative distribution function $F(W)$ is given by

$$F(W) \quad = \quad \int_{-\infty}^{W} p(W')\ dW'\ . \tag{12.110}$$

If $W_T$ is a possible value of wealth at time $T$, then we can interpret the CDF as

$$Prob(W_T < W) \quad = \quad F(W), \tag{12.111}$$

that is, $F(W)$ is the probability that we end up with less than $W$ dollars. Suppose we would like to obtain a final wealth of $W^*$ dollars. Then, we would like $F(W^*)$ to be as small as possible, i.e. the probability of ending up with less than $W^*$ is very small, which is what we want.

Figure 12.12 shows the cumulative distribution functions of the final wealth $W$, for two investment strategies, A and B, which we denote by $F_A(W)$ and $F_B(W)$. Consider the point on the x-axis $W = 10,000$. We can see that for strategy A, the probability of obtaining less than $10,000$ is 0.6, while for strategy B, the probability of obtaining less than $10,000$ is 0.86. Hence, if we want to have a strategy which minimizes the probability of obtaining less than $10,000$, we would prefer strategy A. However, note that for Figure 12.12 we have that

$$F_A(W) \leq F_B(W)\ ;\ \ -\infty \leq W \leq \infty \tag{12.112}$$

and there is at least one point $\widetilde{W}$ such that $F_A(\widetilde{W}) < F_B(\widetilde{W})$, i.e. a point where equation (12.112) holds with strict inequality. In fact, in Figure 12.12, the strict inequality holds for many points. So, we can repeat the argument we went through for $W = 10,000$ for every value of $W$ along the x-axis. In other words, for every value of $W$, strategy A has a smaller probability of ending up with less than $W$ compared to strategy B.

In this case, we say that strategy A stochastically dominates (in the first order sense) strategy B. Any reasonable investor (i.e. with any reasonable utility function) would always prefer strategy A to strategy B. Note that this criteria is based on the entire distribution function, not just a few summary statistics. First order stochastic dominance is easy to spot from the CDFs. If the CDF of strategy A always plots at or below the CDF of strategy B, then A dominates B.

However, in practice, given two reasonable strategies, it is rare to find that one strategy dominates another. Often the CDFs cross at various points. For example, Figure 12.13 shows CDFs for various strategies. Each of these strategies is *reasonable*, yet no strategy strictly stochastically dominates another strategy, which would be typical.

This leads us to the definition of partial stochastic dominance [8] We say that strategy A partially dominates strategy B if

$$F_A(W) \quad \leq \quad F_B(W)\ ;\ \ -\infty \leq W \leq W^*\ . \tag{12.113}$$

This a practical criteria: if $W^*$ is quite large (i.e. we would be fabulously wealthy if $W_T > W^*$), then we really don't care if strategy A underperforms B for these large wealth values, as long as A outperforms B for all values of $W_T \leq W^*$. We are very happy with any amount larger than $W^*$.

We can generalize this a bit more. We can say that A partially dominates B if

$$F_A(W) \quad \leq \quad F_B(W)\ ;\ \hat{W} \leq W \leq W^*\ . \tag{12.114}$$

We have just explained why the upper bound can be a reasonable criteria for partial stochastic dominance. However, at first sight it seems foolhardy to also apply a lower bound criteria $\hat{W}$. This means that we allow A to have a worse performance than B in the left tail, where results are bad.

[8] see *On the Distribution of Terminal Wealth under Dynamic Mean-Variance Optimal Investment Strategies*, P. M. van Staden, D-M. Dang, P. A. Forsyth, SIAM Journal on Financial Mathematics 12:2 (2021) 566–603.

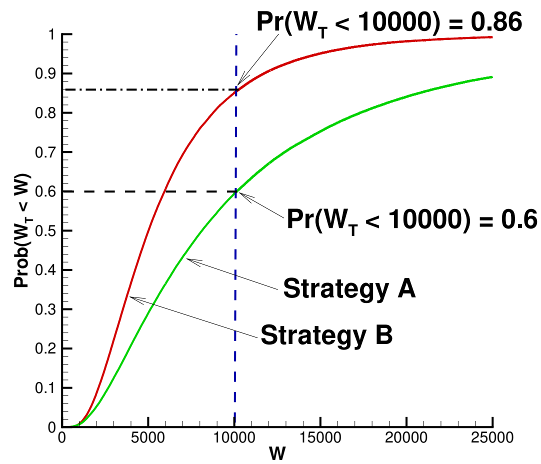



Figure 12.12: Illustration of first order stochastic dominance. Cumulative distribution functions for two strategies, in terms of final wealth $W$. Strategy A stochastically dominates strategy B (in the first order sense). Any reasonable investor would always prefer strategy A.

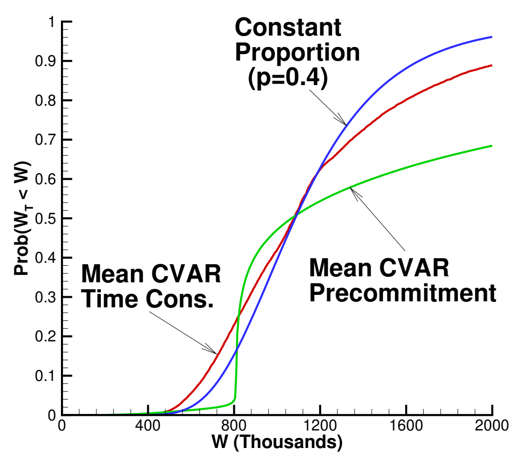



Figure 12.13: Comparison of CDFs for various strategies. Note that in general, the CDFs cross, and we do not observe strict stochastic dominance. See P. A. Forsyth, "Multi-period mean-CVAR asset allocation: is it advantageous to be time consistent?," SIAM Journal on Financial Mathematics 11:2 (2020) 358-384.

However, sometimes this can be reasonable. Suppose we start off with an initial wealth of 1000, and that strategy A has $Med[W_T] = 10,000$, which looks quite good. Suppose condition 12.114 is satisfied with $\hat{W} = 10$ and $W^* = 50,000$. We don't care what happens if we start off with 1000 and end up with more than $50,000$.

However, A underperforms B in the left tail where $W_T < 10$. These are the scenarios where essentially everything has turned bad. We started with 1000, and after years of investing, we are left with only 10. Basically, we are bankrupt. Under strategy A, perhaps our probability of having, say, five dollars or less, is twice the probability of strategy B having five dollars or less. So, strategy A has twice the probability of being in this extreme left tail compared to strategy B. This sounds bad. But this is all peanuts compared to our original stake of 1000. So, perhaps in this case, we don't care about the extreme left tail either. The fact that in these bad cases, we are more likely to end up with two cents in our pocket from strategy B compared to one cent from strategy A is cold comfort.

### 12.7.2 Expected Shortfall

If $F(W)$ is the cumulative distribution function of $W_T$, with density $p(W)$, then suppose we select a value $\mathbb{W}$ such that

$$F(\mathbb{W}) = \int_{-\infty}^{\mathbb{W}} p(W')\ dW' = \alpha\ ;\ 0 < \alpha < 1 \tag{12.115}$$

We define the expected shortfall at level $\alpha$ ($ES(\alpha)$) as

$$ES(\alpha) = \frac{\int_{-\infty}^{\mathbb{W}} W'\ p(W')\ dW'}{\alpha}\ , \tag{12.116}$$

which is just the mean of the worst $\alpha$ fraction of outcomes. Note that this is essentially the negative of the Conditional Value at Risk (CVAR). Since $W$ is the final wealth, a larger value of $ES(\alpha)$ is better. This is measure of left tail risk.

### 12.7.3 Data

We are going to compare an equal weight index with a capitalization based index. We will use the Center for Research in Securities (CRSP) capitalization weighted total return index (includes all dividends and distributions) [9] Similarly, we will also use the CRSP equal weighted index. The equal weighted index has an equal amount invested in all stocks in the index. This index is rebalanced back to equal weights monthly. The CRSP data covers the range 1926.00 - 2022.00. We use the monthly data for both series. For the bond indexes, we consider two cases: a constant maturity 10-year US treasury index [10], and a short term 30 day US T-bill index [11]. These indexes are also for monthly data. We adjust all the indexes for inflation, by dividing by the CPI (also from CRSP). In other words, all investments are in real dollars.

Note that the 10-year Treasury index was constructed by (i) buying a 10 year treasury at the start of every month, (ii) selling the 10-year treasury at the start of the next month, and then (iii) immediately buying a fresh 10 year treasury. The return over the month includes interest and capital gains/losses, all in constant dollars.

### 12.7.4 Scenario

In order to model a realistic scenario, we consider an investor who has a portfolio of 60% in the equity index, and 40% in a bond index. The investor rebalances the investments in the stock and bond index, back to the 60 : 40 ratio once a year. The investor starts with 1000, with no injection or withdrawals of cash over the investment horizon, and we examine the statistics of the terminal wealth $W_T$ at $T = 30$ years. We will consider a long term 30 year investment strategy, since this would be typical of an investor saving for retirement.

[9] More specifically, results presented here were calculated based on data from Historical Indexes, ©2022 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Wharton Research Data Services (WRDS) was used in preparing this article. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

[10] The 10-year Treasury index was constructed from monthly returns from CRSP back to 1941. The data for 1926-1941 were interpolated from annual returns in Homer and Sylla, "A history of interest rates," (2005).

[11] The 30 day T-bill index was obtained from CRSP

### 12.7.5 Bootstrap results

We will compare the equal weighted index to the captilization weighted index, using stationary block bootstrap resampling of the data for the period 1926.00-2022.00 [12] This approach for evaluating investment policies is entirely data driven [13]. We use the standard algorithm[14] for estimating the optimal expected blocksize for each series. The bootstrap procedure concatenates randomly selected blocks of data to account for possible serial correlation. We use the average of the optimal expected blocksize for each individual time series, and then simultaneously draw samples from the stock and bond indexes.

Figure 12.14 compares using the cap weighted CRSP index compared to the equal weighted index, in terms of the CDFs for both strategies. The bond index is a constant maturity 10 year US treasury index. Figure 12.14 shows that the portfolio with the equal weighted stock index clearly stochastically dominates the portfolio with the capitalization weighted stock index. This is a somewhat surprising result, but is consistent with some previous work.

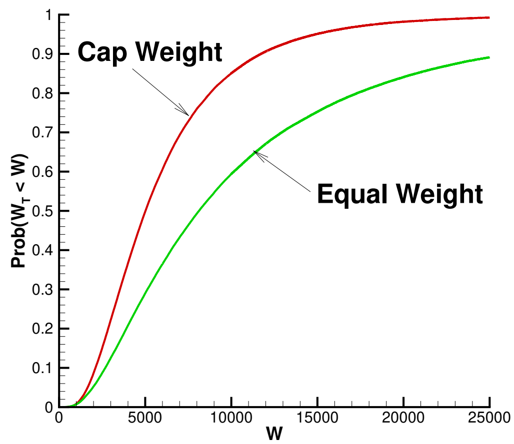



Figure 12.14: Initial stake $W_0 = 1000$, no cash injections for withdrawals, $T = 30$ years. Block bootstrap resampling, expected blocksize 0.5 years. 60% stocks, 40% bonds, rebalanced annually. Bond index: constant maturity 10 year US treasuries. Stock index: CRSP capitalization weighted or CRSP equal weighted index. Data range 1926.00 - 2022.00. All indexes are deflated by the CPI. $10^6$ resamples.

The cumulative, continuously compounded Sharpe ratio is defined as

$$\mathcal{S} = \frac{E[W_T] - W_0 e^{rT}}{std[W_T]} \ . \tag{12.117}$$

where $r$ is the risk-free rate. Note that equation (12.117) should not be confused with the usual *instantaneous* Sharpe ratio, which is estimated using average short term arithmetic returns. In our case, since we carry out annual rebalancing, the equity fraction between rebalancing dates is not constant, so the usual procedure does not make sense.

Since we are examining real (adjusted for inflation) quantities, there is no real risk-free asset. However, the annualized real return of a 30-day T-bill over the entire historical period is approximately zero. Using

[12]D. Politis and J. Romano, "The Stationary Bootstrap", Journal of the American Statistical Association, 89 (1994) 1303-1313.

[13]H. Dichtl, W. Drobetz, M. Wambach, "Testing rebalancing strategies for stock-bond portfolios across different asset allocations," Applied Economics 48 (2016) 772-788.

[14]A. Patton, D. Politis, H. White, "Correction to: automatic block-length selection for the dependent bootstrap," Econometric Reviews 28 (2009) 372-375.

| | Equal Weight | Cap Weight |
|---|---|---|
| $E[W_T]$ | 12546 | 6237 |
| $std[W_T]$ | 15664 | 4639 |
| $Med[W_T]$ | 8122 | 5016 |
| ES(5%) | 1431 | 1280 |
| $\mathcal{S}$ | 0.73 | 1.13 |

Table 12.11: Initial stake $W_0 = 1000$, no cash injections for withdrawals, $T = 30$ years. Block bootstrap resampling, expected blocksize 0.5 years. 60% stocks, 40% bonds, rebalanced annually. Bond index: constant maturity 10 year US treasuries. Stock index: CRSP capitalization weighted or CRSP equal weighted index. Data range 1926.00 - 2022.00. All indexes are deflated by the CPI. ES(5%) is the mean of the worst 5% of the outcomes. $\mathcal{S}$ is the continuously compounded Sharpe ratio, as defined in equation (12.117).

this value of $r = 0$ in equation (12.117) we obtain the results shown in Table 12.11. Even though we have seen that the equal weight portfolio dominates the cap weighted portfolio, the continuously compounded Sharpe ratio for the equal weight 60:40 portfolio is actually worse (smaller) than for the 60:40 cap weighted portfolio.

This is essentially because standard deviation is a poor measure of risk, since it penalizes upside as well as downside. The equal weighted index generates a much bigger right skew compared with using the cap weighted index. Most investors would prefer a large right skew, which is penalized by the standard deviation. The left tail risk, as measured by ES(5%) is larger (better) for the equal weighted portfolio. This is consistent with our intuition from Figure 12.14.

Figure 12.15 and Table 12.12 show similar results for 60:40 portfolios, this time the bond index is based on 30-day T-bills. The results are qualitatively similar to the 10 year treasury case.

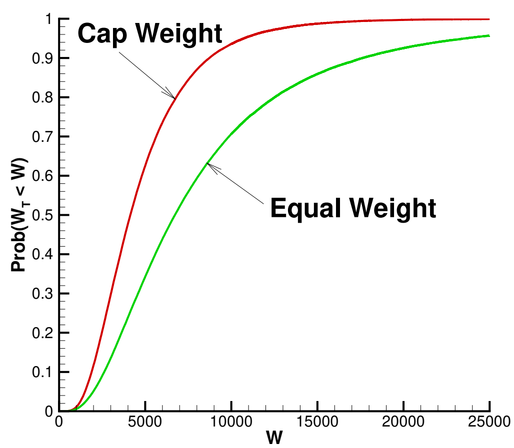



Figure 12.15: Initial stake $W_0 = 1000$, no cash injections for withdrawals, $T = 30$ years. Block bootstrap resampling, expected blocksize 2.0 years. 60% stocks, 40% bonds, rebalanced annually. Bond index: 30 day US T-bills. Stock index: CRSP capitalization weighted or CRSP equal weighted index. Data range 1926.00 - 2022.00. All indexes are deflated by the CPI. $10^6$ resamples.

However, more recent papers suggest that equal weighted portfolios have underperformed. Figure 12.16 shows the bootstrap results, for the 60:40 equal and cap weighted portfolios, using data in the range 1980.00-2022.00. The bond index is based on 30 day T-Bills. In this case, the use of the equal weighted portfolio

|  | Equal Weight | Cap Weight |
|---|---|---|
| $E[W_T]$ | 9058 | 4877 |
| $std[W_T]$ | 8653 | 3091 |
| $Med[W_T]$ | 6726 | 4135 |
| ES(5%) | 1484 | 1184 |
| $\mathcal{S}$ | .93 | 1.25 |

Table 12.12: Initial stake $W_0 = 1000$, no cash injections for withdrawals, $T = 30$ years. Block bootstrap resampling, expected blocksize 2.0 years. 60% stocks, 40% bonds, rebalanced annually. Bond index: 30 day US T-bills. Stock index: CRSP capitalization weighted or CRSP equal weighted index. Data range 1926.00 - 2022.00. All indexes are deflated by the CPI. ES(5%) is the mean of the worst 5% of the outcomes. $\mathcal{S}$ is the continuously compounded Sharpe ratio, as defined in equation (12.117).

still stochastically dominates the cap weighted portfolio, but the effect is very small. Table 12.13 gives some additional statistics for the past two decades. It appears that the equal weighted portfolio underperforms for the past decade.

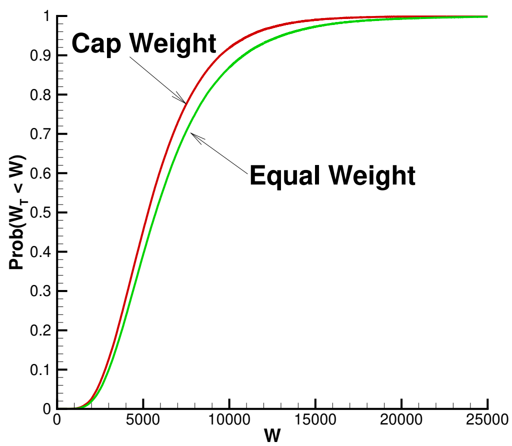



Figure 12.16: Initial stake $W_0 = 1000$, no cash injections for withdrawals, $T = 30$ years. Block bootstrap resampling, expected blocksize 2.0 years. 60% stocks, 40% bonds, rebalanced annually. Bond index: 30 day US T-bills. Stock index: CRSP capitalization weighted or CRSP equal weighted index. Data range 1980.00 - 2022.00. All indexes are deflated by the CPI. $10^6$ resamples.

How can we explain this? The Tljaard et al (2021) paper suggests that this is a short term effect. However, an equal weighted index will obviously put a lot more weight on small cap stocks, compared to a capitalization weighted index. Years ago, it was noted that small cap stocks seemed to perform better than you would expect. In fact, small cap portfolios were one of the original factors in the Fama-French three factor model of stock returns [15]

A reasonable explanation for the small cap effect would be that analysts ignored small cap stocks, so nobody was really following them. This allowed skilled stock pickers to do better than you would expect (a market imperfection). This was first noted in [16]. The usual assumption in academic finance is the *no*

[15] E. F. Fama, K. R. French, "The cross section of expected stock returns," Journal of Finance 47:2 (1992) 427-465.

[16] see *"The relationship between return and market value of common stocks"*, R. Banz, Journal of Financial Economics 9:1

| Data series | Annualized Return 2012.00-2022.00 | Annualized Return 2002.00-2022.00 |
|---|---|---|
| Real CRSP cap weighted index | 11.79% | 6.72% |
| Real CRSP equal-weighted index | 8.78% | 7.58% |
| Real 10 year US treasury index | 1.61% | 2.33% |
| Real 30-day US T-bill index | -1.58% | -1.33% |

Table 12.13: Recent historical real returns of CRSP equal and cap weighted stock indexes (100% stocks, no bonds). Also, the return of the 10 year treasury index and the 30-day T-bill index (0% stocks, 100% bonds).

*arbitrage* principle, i.e. no free lunch. If the equal weight index outperformance is simply due to the small cap effect, then, once everyone knows about it, the effect will disappear (i.e. arbitraged away). So, if we look at the bootstrap results from 1980.00, we see a much smaller effect than on the entire data set from 1926.00. The classic small cap effect paper was published in 1981. Is this just a coincidence?

### 12.7.6 What about taxes and distributions?

Note that the above indexes assumed that any distributions (e.g. dividends) were immediately reinvested. In addition, the equal weighted portfolios are re-balanced monthly (back to equal weights). The dividends and capital gains are usually subject to taxation. This is clearly not a problem if the portfolio is held in a tax advantaged account, such as an RRSP or TFSA in Canada, or a 401(k) in the US.

However, in a taxable account, how realistic is the assumption that all taxes on gains are deferred? In Canada, there are swap based index ETFs, which do not distribute any dividends. Effectively, all dividends are reinvested and tax deferred until the ETF is sold.

It is also interesting to note that in the US, it is perfectly legal for ETFs to use *heartbeat transactions*, which essentially defer taxes on any gains from rebalancing. [17].

### 12.7.7 Summary: equal weight vs. cap weight

Bootstrap resampling using long term data 1926.00 - 2022.00, shows that equal weighted indexes stochastically dominate cap weighted indexes. However, if we repeat the experiment, this time using data in the range 1980:00 - 2022:00, the effect almost disappears. Is this a permanent effect, or are the last 40 years anomalous? Certainly, the last 40 years have had very low inflation, and declining interest rates, which is historically unusual, and perhaps provided a tailwind for large cap tech stocks. On the other hand, maybe the small cap effect has actually disappeared, now that there is a large literature on this.

Maybe the answer is to equal weight the cap weight and equal weighted indexes?

(1981) 3-18.

[17] `https://www.bloomberg.com/graphics/2019-etf-tax-dodge-lets-investors-save-big/`

# 13 Further Reading

## 13.1 General Interest


- Peter Bernstein, *Capital Ideas: the improbable origins of modern Wall street,* The Free Press, New York, 1992.
- Peter Bernstein, *Against the Gods: the remarkable story of risk,* John Wiley, New York, 1998, ISBN 0-471-29563-9.
- Benjamin Graham, *The intelligent investor,* various editions. Warren Buffet's bible. There is a Collins edition, of the 1973 version, with commentary by Jason Zweig.
- Burton Malkeil, *A random walk down Wall Street,* W.W. Norton, New York, various editions.
- N. Taleb, *Fooled by Randomness*, Texere, 2001, ISBN 1-58799-071-7.
- S. Das, *Traders, Guns and Money*, Prentice Hall, 2006.
- N. Taleb, *The black swan*, Random House, 2007.


## 13.2 More Background


- A. Dixit and R. Pindyck, *Investment under uncertainty,* Princeton University Press, 1994.
- John Hull, *Options, futures and other derivatives,* Prentice-Hall, 1997, ISBN 0-13-186479-3.
- S. Ross, R. Westerfield, J. Jaffe, *Corporate Finance,* McGraw-Hill Irwin, 2002, ISBN 0-07-283137-5.
- W. Sharpe, *Portfolio Theory and Capital Markets,* Wiley, 1970, reprinted in 2000, ISBN 0-07-135320-8. (Still a classic).
- Lenos Triegeorgis, *Real Options: Managerial Flexibility and Strategy for Resource Allocation,* MIT Press, 1996, ISBN 0-262-20102-X.


## 13.3 More Technical


- P. Brandimarte, *Numerical Methods in Finance: A Matlab Introduction,* Wiley, 2002, ISBN 0-471-39686-9.
- Boyle, Broadie, Glassermman, *Monte Carlo methods for security pricing,* J. Econ. Dyn. Con., 21:1267-1321 (1997)
- P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer (2004) ISBN 0-387-00451-3.
- D. Higham, *An Introduction to Financial Option Valuation*, Cambridge (2004) ISBN 0-521-83884-3.
- P. Jackel, *Monte Carlo Methods in Finance,* Wiley, 2002, ISBN 0-471-49741-X.
- Y.K. Kwok, *Mathematical Models of Finance,* Springer Singapore, 1998, ISBN 981-3083-565.
- S. Neftci, *An Introduction to the Mathematics of Financial Derivatives,* Academic Press (2000) ISBN 0-12-515392-9.
- R. Seydel, *Tools for Computational Finance,* Springer, 2002, ISBN 3-540-43609-X.
- D. Tavella and C. Randall, *Pricing Financial Instruments: the Finite Difference Method,* Wiley, 2000, ISBN 0-471-19760-2.

- D. Tavella, *Quantitative Methods in Derivatives Pricing: An Introduction to Computational Finance,* Wiley (2002) ISBN 0-471-39447-5.
- N. Taleb, *Dynamic Hedging*, Wiley, 1997, ISBN 0-471-15280-3.
- P. Wilmott, S. Howison, J. Dewynne, *The mathematics of financial derivatives: A student introduction,* Cambridge, 1997, ISBN 0-521-49789-2.
- Paul Wilmott, *Paul Wilmott on Quantitative Finance,* Wiley, 2000, ISBN 0-471-87438-8.

clearpage

# References


Bernard, C. and S. Vanduffel (2014). Mean-variance optimal portfolios in the presence of a benchmark with applications to fraud detection. *European Journal of Operational Research 234*, 469–480.

Dybvig, P. H. and J. E. Ingersoll (1982). Mean-variance theory in complete markets. *Journal of Business 55:2*, 233–251.

Forsyth, P. A. (2020). Multi-period mean CVAR asset allocation: Is it advantageous to be time consistent? *SIAM Journal on Financial Mathematics 11:2*, 358–384.

Forsyth, P. A. and K. R. Vetzal (2017a). Dynamic mean variance asset allocation: Tests for robustness. *International Journal of Financial Engineering 4*, 1750021:1–1750021:37. DOI: 10.1142/S2424786317500219.

Forsyth, P. A. and K. R. Vetzal (2017b). Robust asset allocation for long-term target-based investing. *International Journal of Theoretical and Applied Finance 20:3*. 1750017 (electronic).

Forsyth, P. A. and K. R. Vetzal (2019). Optimal asset allocation for retirement savings: deterministic vs. time consistent adaptive strategies. *Applied Mathematical Finance 26:1*, 1–37.

Goetzmann, W., J. E. Ingersoll, M. Spiegel, and I. Welch (2002). Sharpening Sharpe ratios. NBER Working paper.

Lhabitant, F.-S. (2000). Derivatives in portfolio management: why beating the market is easy. Working paper, EDHEC Business School.

Li, D. and W.-L. Ng (2000). Optimal dynamic portfolio selection: multiperiod mean-variance formulation. *Mathematical Finance 10*, 387–406.

Menoncin, F. and E. Vigna (2017). Mean-variance target based optimisation for defined contribution pension schemes in a stochastic framework. *Insurance: Mathematics and Economics 76*, 172–184.

Spurgon, R. B. (2001). How to game your Sharpe ratio. *Journal of Alternative Investments 4:3 (Winter)*, 38–36.

Strub, M., D. Li, and X. Cui (2019). An enhanced mean-variance framework for robo-advising applications. SSRN 3302111.

Vigna, E. (2014). On efficiency of mean-variance based portfolio selection in defined contribution pension schemes. *Quantitative Finance 14*, 237–258.

Wang, J. and P. A. Forsyth (2010). Numerical solution of the Hamilton-Jacobi-Bellman formulation for continuous time mean variance asset allocation. *Journal of Economic Dynamics and Control 34*, 207–230.

Zhou, X. Y. and D. Li (2000). Continuous-time mean-variance portfolio selection: a stochastic LQ framework. *Applied Mathematics and Optimization 42*, 19–33.