Chapter 2

Some Basic Theory of Finance

Introduction to Pricing: Single Period Models

Let us begin with a very simple example designed to illustrate the no-arbitrage approach to pricing derivatives. Consider a stock whose price at present is $\$s$. Over a given period, the stock may move either up or down, up to a value su where $u > 1$ with probability p or down to the value sd where $d < 1$ with probability $1-p$. In this model, these are the only moves possible for the stock in a single period. Over a longer period, of course, many other values are possible. In this market, we also assume that there is a so-called risk-free bond available returning a guaranteed rate of $r\%$ per period. Such a bond cannot default; there is no random mechanism governing its return which is known upon purchase. An investment of $\$1$ at the beginning of the period returns a guaranteed $\$(1+r)$ at the end. Then a portfolio purchased at the beginning of a period consisting of y stocks and x bonds will return at the end of the period an amount $\$x(1+r) + ysZ$ where Z is a random variable taking

values u or d with probabilities p and $1-p$ respectively. We permit owning a negative amount of a stock or bond, corresponding to shorting or borrowing the correspond asset for immediate sale.

An ambitious investor might seek a portfolio whose initial cost is zero (i.e. $x+ys=0$) such that the return is greater than or equal to zero with positive probability. Such a strategy is called an *arbitrage.* This means that the investor is able to achieve a positive probability of future profits with no down-side risk with a net investment of \$0. In mathematical terms, the investor seeks a point (x,y) such that $x+ys=0$ (net cost of the portfolio is zero) and

$$x(1+r)+ysu \geq 0,$$
$$x(1+r)+ysd \geq 0$$

with at least one of the two inequalities strict (so there is never a loss and a non-zero chance of a positive return). Alternatively, is there a point on the line $y=-\frac{1}{s}x$ which lies *above both* of the two lines

$$y=-\frac{1+r}{su}x$$
$$y=-\frac{1+r}{sd}x$$

and strictly above one of them? Since all three lines pass through the origin, we need only compare the slopes; an arbitrage will NOT be possible if

$$-\frac{1+r}{sd} \leq -\frac{1}{s} \leq -\frac{1+r}{su} \tag{2.1}$$

and otherwise there is a point (x,y) permitting an arbitrage. The condition for no arbitrage (2.1) reduces to

$$\frac{d}{1+r} < 1 < \frac{u}{1+r} \tag{2.2}$$

So the condition for no arbitrage demands that $(1+r-u)$ and $(1+r-d)$ have opposite sign or $d \leq (1+r) \leq u$. Unless this occurs, the stock *always* has either better or worse returns than the bond, which makes no sense in a

free market where both are traded without compulsion. Under a no arbitrage assumption since $d \leq (1+r) \leq u$, the bond payoff is a *convex combination* or a weighted average of the two possible stock payoffs; i.e. there are probabilities $0 \leq q \leq 1$ and $(1-q)$ such that $(1+r) = qu + (1-q)d$. In fact it is easy to solve this equation to determine the values of q and $1-q$.

$$q = \frac{(1+r)-d}{u-d}, \qquad \text{and} \quad 1-q = \frac{u-(1+r)}{u-d}.$$

Denote by Q the probability distribution which puts probabilities q and $1-q$ on these points su, sd. Then if S_1 is the value of the stock at the end of the period, note that

$$\frac{1}{1+r}E_Q(S_1) = \frac{1}{1+r}(qsu + (1-q)sd) = \frac{1}{1+r}s(1+r) = s$$

where E_Q denotes the expectation assuming that Q describes the probabilities of the two outcomes.

In other words, *if there is to be no arbitrage, there exists a probability measure Q such that the expected price of future value of the stock S_1 discounted to the present using the return from a risk-free bond is exactly the present value of the stock.* The measure Q is called the *risk-neutral* measure and the probabilities that it assigns to the possible outcomes of S are not necessarily those that determine the future behaviour of the stock. The risk neutral measure embodies both the current consensus beliefs in the future value of the stock and the consensus investors' attitude to risk avoidance. It is not usually true that $\frac{1}{1+r}E_P(S_1) = s$ with P denoting the actual probability distribution describing the future probabilities of the stock. Indeed it is highly unlikely that an investor would wish to purchase a risky stock if he or she could achieve exactly the same expected return with no risk at all using a bond. We generally expect that to make a risky investment attractive, its expected return should be greater than that of a risk-free investment. Notice in this example that the risk-neutral measure Q did not use the probabilities p, and $1-p$ that the stock would go

up or down and this seems contrary to intuition. Surely if a stock is more likely to go up, then a call option on the stock should be valued higher!

Let us suppose for example that we have a friend willing, in a private transaction with me, to buy or sell a stock at a price determined from his subjectively assigned distribution P, different from Q. The friend believes that the stock is presently worth

$$\frac{1}{1+r}E_P S_1 = \frac{psu + (1-p)sd}{1+r} \neq s \quad \text{since } p \neq q.$$

Such a friend offers their assets as a sacrifice to the gods of arbitrage. If the friend's assessed price is greater than the current market price, we can buy on the open market and sell to the friend. Otherwise, one can do the reverse. Either way one is enriched monetarily (and perhaps impoverished socially)!

So why should we use the Q measure to determine the price of a given asset in a market (assuming, of course, there is a risk-neutral Q measure and we are able to determine it)? Not because it precisely describes the future behaviour of the stock, *but because if we use any other distribution, we offer an intelligent investor (there are many!) an arbitrage opportunity, or an opportunity to make money at no risk and at our expense.*

Derivatives are investments which derive their value from that of a corresponding asset, such as a stock. A *European call option* is an option which permits you (but does not compel you) to purchase the stock at a fixed future date (the *maturity date*) or for a given predetermined price, the *exercise price* of the option). For example a call option with exercise price \$10 on a stock whose future value is denoted S_1, is worth on expiry $S_1 - 10$ if $S_1 > 10$ but nothing at all if $S_1 < 10$. The difference $S_1 - 10$ between the value of the stock on expiry and the exercise price of the option is your profit if you exercises the option, purchasing the stock for \$10 and sell it on the open market at $\$S_1$. However, if $S_1 < 10$, there is no point in exercising your option as you are not compelled to do so and your return is \$0. In general, your payoff from pur-

chasing the option is a simple function of the future price of the stock, such as $V(S_1) = \max(S_1 - 10,\ 0)$. We denote this by $(S_1 - 10)^+$. The future value of the option is a random variable but it derives its value from that of the stock, hence it is called a *derivative* and the stock is the *underlying.*

A function of the stock price $V(S_1)$ which may represent the return from a portfolio of stocks and derivatives is called a *contingent claim.* $V(S_1)$ represents the payoff to an investor from a certain financial instrument or derivative when the stock price at the end of the period is S_1. In our simple binomial example above, the random variable takes only two possible values $V(su)$ and $V(sd)$. We will show that there is a portfolio, called a *replicating* portfolio, consisting of an investment solely in the above stock and bond which reproduces these values $V(su)$ and $V(sd)$ exactly. We can determine the corresponding weights on the bond and stocks (x, y) simply by solving the two equations in two unknowns

$$x(1+r) + ysu = V(su)$$
$$x(1+r) + ysd = V(sd)$$

Solving: $y^* = \frac{V(su)-V(sd)}{su-sd}$ and $x^* = \frac{V(su)-y^*su}{1+r}$. By buying y^* units of stock and x^* units of bond, we are able to replicate the contingent claim $V(S_1)$ exactly- i.e. produce a portfolio of stocks and bonds with exactly the same return as the contingent claim. So in this case at least, there can be only one possible present value for the contingent claim and that is the present value of the replicating portfolio $x^* + y^*s$. If the market placed any other value on the contingent claim, then a trader could guarantee a positive return by a simple trade, shorting the contingent claim and buying the equivalent portfolio or buying the contingent claim and shorting the replicating portfolio. Thus this is the only price that precludes an arbitrage opportunity. There is a simpler

expression for the current price of the contingent claim in this case: Note that

$$\begin{aligned}\frac{1}{1+r}E_Q V(S_1) &= \frac{1}{1+r}(qV(su)+(1-q)V(sd)) \\ &= \frac{1}{1+r}(\frac{1+r-d}{u-d}V(su)+\frac{u-(1+r)}{u-d}V(sd)) \\ &= x^* + y^* s.\end{aligned}$$

In words, *the discounted expected value of the contingent claim is equal to the no-arbitrage price of the derivative where the expectation is taken using the Q-measure*. Indeed any contingent claim that is attainable must have its price determined in this way. While we have developed this only in an extremely simple case, it extends much more generally.

Suppose we have a total of N risky assets whose prices at times $t = 0, 1$, are given by $(S_0^j, S_1^j), j = 1, 2, ..., N$. We denote by S_0, S_1 the column vector of initial and final prices

$$S_0 = \begin{pmatrix} S_0^1 \\ S_0^2 \\ . \\ . \\ . \\ S_0^N \end{pmatrix}, S_1 = \begin{pmatrix} S_1^1 \\ S_1^2 \\ . \\ . \\ . \\ S_1^N \end{pmatrix}$$

where at time 0, S_0 is known and S_1 is random. Assume also there is a riskless asset (a bond) paying interest rate r over one unit of time. Suppose we borrow money (this is the same as shorting bonds) at the risk-free rate to buy w_j units of stock j at time 0 for a total cost of $\sum w_j S_0^j$. The value of this portfolio at time $t = 1$ is $T(w) = \sum w_j(S_1^j - (1+r)S_0^j)$. If there are weights w_j so that this sum is always non-negative, and $P(T(w) > 0) > 0$, then this is an arbitrage opportunity. Similarly, by replacing the weights w_j by their negative $-w_j$, there is an arbitrage opportunity if for some weights the sum is non-positive and negative with positive probability. In summary, there are *no arbitrage op-*

portunities if for all weights w_j $P(T(w) > 0) > 0$ and $P(T(w) < 0) > 0$ so $T(w)$ takes both positive and negative values. We assume that the moment generating function $M(w) = E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))]$ exists and is an analytic function of w.Roughly the condition that the moment generating function is analytic assures that we can expand the function in a series expansion in w. This is the case, for example, if the values of S_1, S_0 are bounded. The following theorem provides a general proof, due to Chris Rogers, of the equivalence of the no-arbitrage condition and the existence of an equivalent measure Q. Refer to the appendix for the technical definitions of an equivalent probability measure and the existence and properties of a moment generating function $M(w)$.

Theorem 2 *A necessary and sufficient condition that there be no arbitrage opportunities is that there exists a measure Q equivalent to P such that $E_Q(S_1^j) = \frac{1}{1+r}S_0^j$ for all $j = 1, ..., N$.*

Proof. *Define $M(w) = E\exp(T(w)) = E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))]$ and consider the problem*

$$\min_w \ln(M(w)).$$

The no-arbitrage condition implies that for each j there exists $\varepsilon > 0$,

$$P[S_1^j - (1+r)S_0^j > \varepsilon] > 0$$

and therefore as $w_j \to \infty$ while the other weights $w_k, k \neq j$ remain fixed,

$$M(w) = E[\exp(\sum w_j(S_1^j - (1+r)S_0^j))] > C\exp(w_j\varepsilon)P[S_1^j - (1+r)S_0^j > \varepsilon] \to \infty \quad \text{as } w_j \to \infty.$$

Similarly, $M(w) \to \infty$ as $w_j \to -\infty$. From the properties of a moment generating function (see the appendix) $M(w)$ is convex, continuous, analytic and $M(0) = 1$. Therefore the function $M(w)$ has a minimum w^* satisfying $\frac{\partial M}{\partial w_j} = 0$ or

$$\frac{\partial M(w)}{\partial w_j} = 0 \quad \text{or} \tag{2.3}$$

$$E[S_1^j \exp(T(w))] = (1+r)S_0^j E[\exp(T(w))]$$

or

$$S_0^j = \frac{E[\exp(T(w))S_1^j]}{(1+r)E[\exp(T(w))]}.$$

Define a distribution or probability measure Q as follows; for any event A,

$$Q(A) = \frac{E_P[I_A \exp(w'S_1)]}{E_P[\exp(w'S_1)]}.$$

The Radon-Nikodym derivative (see the appendix) is

$$\frac{dQ}{dP} = \frac{\exp(w'S_1)]}{E_P[\exp(w'S_1)]}.$$

Since $\infty > \frac{dQ}{dP} > 0$, the measure Q is equivalent to the original probability measure P (in the intuitive sense that it has the same support). When we calculate expected values under this new measure, note that for each j,

$$\begin{aligned} E_Q(S_1^j) &= E_P[\frac{dQ}{dP}S_1^j] \\ &= \frac{E_P[S_1^j \exp(w'S_1)]}{E_P[\exp(w'S_1)]} \\ &= (1+r)S_0^j. \end{aligned}$$

or

$$S_0^j = \frac{1}{1+r}E_Q(S_1^j).$$

Therefore, the current price of each stock is the discounted expected value of the future price under this "risk-neutral" measure Q.

Conversely if

$$E_Q(S_1^j) = \frac{1}{1+r}S_0^j, \quad \textit{for all } j \tag{2.4}$$

holds for some measure Q then $E_Q[T(w)] = 0$ for all w and this implies that the random variable $T(w)$ is either identically 0 or admits both positive and negative values. Therefore the existence of the measure Q satisfying (2.4) implies that there are no arbitrage opportunities. ■

The so-called risk-neutral measure Q is constructed to minimize the cross-entropy between Q and P subject to the constraints $E(S_1 - (1+r)S_0) = 0$

where cross-entropy is defined in Section 1.5. If there N possible values of the random variables S_1 and S_0 then (2.3) consists of N equations in N unknowns and so it is reasonable to expect a unique solution. In this case, the Q measure is unique and we call the market *complete*.

The theory of pricing derivatives in a complete market is rooted in a rather trivial observation because in a complete market, the derivative can be replicated with a portfolio of other marketable securities. If we can reproduce exactly the same (random) returns as the derivative provides using a linear combination of other marketable securities (which have prices assigned by the market) then the derivative must have the same price as the linear combination of other securities. Any other price would provide arbitrage opportunities.

Of course in the real world, there are costs associated with trading, these costs usually related to a bid-ask spread. There is essentially a different price for buying a security and for selling it. The argument above assumes a frictionless market with no trading costs, with borrowing any amount at the risk-free bond rate possible, and a completely liquid market- any amount of any security can be bought or sold. Moreover it is usually assumed that the market is complete and it is questionable whether complete markets exist. For example if a derivative security can be perfectly replicated using other marketable instruments, then what is the purpose of the derivative security in the market? All models, excepting those on *Fashion File,* have deficiencies and critics. The merit of the frictionless trading assumption is that it provides an accurate approximation to increasingly liquid real-world markets. Like all useful models, this permits tentative conclusions that should be subject to constant study and improvement.

Multiperiod Models.

When an asset price evolves over time, the investor normally makes decisions about the investment at various periods during its life. Such decisions are made

with the benefit of current information, and this information, whether used or not, includes the price of the asset and any related assets at all previous time periods, beginning at some time $t = 0$ when we began observation of the process. We denote this information available for use at time t as H_t. Formally, H_t is what is called a *sigma-field* (see the appendix) generated by the past, and there are two fundamental properties of this sigma-field that will use. The first is that the sigma-fields increase over time. In other words, our information about this and related processes increases over time because we have observed more of the relevant history. In the mathematical model, we do not "forget" relevant information: this model fits better the behaviour of youthful traders than aging professors. The second property of H_t is that it includes the value of the asset price $S_\tau, \tau \leq t$ at all times $\tau \leq t$. In measure-theoretic language, S_t *is adapted to or measurable with respect to* H_t. Now the analysis above shows that when our investment life began at time $t = 0$ and we were planning for the next period of time, absence of arbitrage implies a risk-neutral measure Q such that $E_Q(\frac{1}{1+r}S_1) = S_0$. Imagine now that we are in a similar position at time t, planning our investment for the next unit time. All expected values should be taken in the light of our current knowledge, i.e. given the information H_t. An identical analysis to that above shows that under the risk neutral measure Q, if S_t represents the price of the stock after t periods, and r_t the risk-free one-period interest rate offered that time, then

$$E_Q(\frac{1}{1+r_t}S_{t+1}|H_t) = S_t. \tag{2.5}$$

Suppose we let B_t be the value of \$1 invested at time $t = 0$ after a total of t periods. Then $B_1 = (1 + r_0)$, $B_2 = (1 + r_0)(1 + r_1)$, and in general $B_t = (1+r_0)(1+r_1)...(1+r_{t-1})$. Since the interest rate per period is announced at the beginning of this period, the value B_t is known at time $t - 1$. If you owe exactly \$1.00 payable at time t, then to cover this debt you should have an

investment at time $t = 0$ of $\$E(1/B_t)$, which we might call the *present value* of the promise. In general, at time t, the present value of a certain amount $\$V_T$ promised at time T (i.e. the present value or the value discounted to the present of this payment) is

$$E(V_T \frac{B_t}{B_T}|H_t).$$

Now suppose we divide (2.5) above by B_t. We obtain

$$E_Q(\frac{S_{t+1}}{B_{t+1}}|H_t) = E_Q(\frac{1}{B_t(1+r_t)}S_{t+1}|H_t) = \frac{1}{B_t}E_Q(\frac{1}{1+r_t}S_{t+1}|H_t) = \frac{S_t}{B_t}. \tag{2.6}$$

Notice that we are able to take the divisor B_t outside the expectation since B_t is known at time t (in the language of Appendix 1, B_t is measurable with respect to H_{t+1}). This equation (2.6) describes an elegant mathematical property shared by all marketable securities in a complete market. Under the risk-neutral measure, the discounted price $Y_t = S_t/B_t$ forms a *martingale*. A *martingale* is a process Y_t for which the expectation of a future value given the present is equal to the present i.e.

$$E(Y_{t+1}|H_t) = Y_t.\text{for all } t. \tag{2.7}$$

Properties of a martingale are given in the appendix and it is easy to show that for such a process, when $T > t$,

$$E(Y_T|H_t) = E[...E[E(Y_T|H_{T-1})|H_{T-2}]...|H_t] = Y_t. \tag{2.8}$$

A martingale is a fair game in a world with no inflation, no need to consume and no mortality. Your future fortune if you play the game is a random variable whose expectation, given everything you know at present, is your present fortune.

Thus, under a risk-neutral measure Q in a complete market, all marketable securities discounted to the present form martingales. For this reason, we often refer to the risk-neutral measure as a martingale measure. The fact that prices of

marketable commodities must be martingales under the risk neutral measure has many consequences for the canny investor. Suppose, for example, you believe that you are able to model the history of the price process nearly perfectly, and it tells you that the price of a share of XXX computer systems increases on average 20% per year. Should you use this $P-$measure in valuing a derivative, even if you are confident it is absolutely correct, in pricing a call option on XXX computer systems with maturity one year from now? If you do so, you are offering some arbitrager another free lunch at your expense. The measure Q, not the measure P, determines derivative prices in a no-arbitrage market. This also means that there is no advantage, when pricing derivatives, in using some elaborate statistical method to estimate the expected rate of return because this is a property of P not Q.

What have we discovered? In general, prices in a market are determined as expected values, but expected values with respect to the measure Q. This is true in any complete market, regardless of the number of assets traded in the market. For any future time $T > t$, and for any derivative defined on the traded assets in a market whose value at time t is given by V_t, $E_Q(\frac{B_t}{B_T}V_T|H_t] = V_t =$ the market price of the derivative at time t. So in theory, determining a reasonable price of a derivative should be a simple task, one that could be easily handled by simulation. Suppose we wish to determine a suitable price for a derivative whose value is determined by some stock price process S_t. Suppose that at time $T > t$, the value of the derivative is a simple function of the stock price at that time $V_T = V(S_T)$. We may simply generate many simulations of the future value of the stock and corresponding value of the derivative $S_T, V(S_T)$ given the current store of information H_t. These simulations must be conducted under the measure Q. In order to determine a fair price for the derivative, we then average the discounted values of the derivatives, discounted to the present, over all the simulations. The catch is that the Q measure is often neither obvious from the present market prices nor statistically estimable from its past. It is given

implicitly by the fact that the expected value of the discounted future value of traded assets must produce the present market price. In other words, a first step in valuing any asset is to determine a measure Q for which this holds. Now in some simple models involving a single stock, this is fairly simple, and there is a unique such measure Q. This is the case, for example, for the stock model above in which the stock moves in simple steps, either increasing or decreasing at each step. But as the number of traded assets increases, and as the number of possible jumps per period changes, a measure Q which completely describes the stock dynamics and which has the necessary properties for a risk neutral measure becomes potentially much more complicated as the following example shows.

Solving for the Q Measure.

Let us consider the following simple example. Over each period, a stock price provides a return greater than, less than, or the same as that of a risk free investment like a bond. Assume for simplicity that the stock changes by the factor $u(1+r)$ (greater) or $(1+r)$ (the same) $d(1+r)$(less) where $u > 1 > d = 1/u$. The Q probability of increases and decreases is unknown, and may vary from one period to the next. Over two periods, the possible paths executed by this stock price process are displayed below assuming that the stock begins at time $t = 0$ with price $S_0 = 1$.

[FIGURE 2.1 ABOUT HERE]

In general in such a tree there are three branches from each of the nodes at times $t = 0, 1$ and there are a total of $1 + 3 = 4$ such nodes. Thus, even if we assume that probabilities of up and down movements do not depend on how the process arrived at a given node, there is a total of $3 \times 4 = 12$ unknown parameters. Of course there are constraints; for example the sum of the three probabilities on branches exiting a given node must add to one and the price

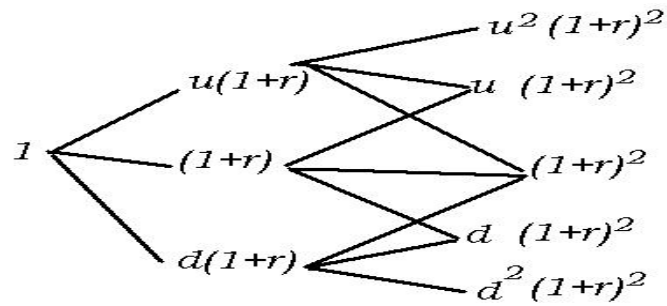

Figure 2.1: A Trinomial Tree for Stock Prices

process must form a martingale. For each of the four nodes, this provides two constraints for a total of 8 constraints, leaving 4 parameters to be estimated. We would need the market price of 4 different derivatives or other contingent claims to be able to generate 4 equations in these 4 unknowns and solve for them. Provided we are able to obtain prices of four such derivatives, then we can solve these equations. If we denote the risk-neutral probability of 'up' at each of the four nodes by p_1, p_2, p_3, p_4 then the conditional distribution of S_{t+1} given $S_t = s$ is:

Stock value	$su(1+r)$	$s(1+r)$	$sd(1+r)$
Probability	p_i	$1 - \frac{u-d}{1-d}p_i = 1 - kp_i$	$\frac{u-1}{1-d}p_i = cp_i$

Consider the following special case, with the risk-free interest rate per period r, $u = 1.089$, $S_0 = \$1.00$. We also assume that we are given the price of four call options expiring at time $T = 2$. The possible values of the price at time $T = 2$ corresponding to two steps up, one step up and one constant, one up one down, etc. are the values of $S(T)$ in the set

$$\{1.1859, 1.0890, 1.0000, 0.9183, 0.8432\}.$$

Now consider a "call option" on this stock expiring at time $T = 2$ with strike

price K. Such an option has value at time $t = 2$ equal to $(S_2 - K)$ if this is positive, or zero otherwise. For brevity we denote this by $(S_2 - K)^+$. The present value of the option is $E_Q(S_2 - K)^+$ discounted to the present, where K is the exercise price of the option and S_2 is the price of the stock at time 2. Thus the price of the call option at time 0 is given by

$$V_0 = E_Q(S_2 - K)^+/(1+r)^2$$

Assuming interest rate $r = 1\%$ per period, suppose we have market prices of four call options with the same expiry and different exercise prices in the following table;

K =Exercise Price	T =Maturity	V_0 =Call Option Price
0.867	2	0.154
0.969	2	.0675
1.071	2	.0155
1.173	2	.0016

If we can observe the prices of these options only, then the equations to be solved for the probabilities associated with the measure Q equate the observed price of the options to their theoretical price $V_0 = E(S_2 - K)^+/(1+r)^2$.

$$\begin{aligned}
0.0016 &= \frac{1}{(1.01)^2}(1.186 - 1.173)p_1p_2 \\
0.0155 &= \frac{1}{(1.01)^2}[(1.186 - 1.071)p_1p_2 + (1.089 - 1.071)\{p_1(1 - kp_2) + (1 - kp_1)p_2\}] \\
0.0675 &= \frac{1}{(1.01)^2}[0.217p_1p_2 + 0.12\{p_1(1 - kp_2) + (1 - kp_1)p_2\} \\
&\quad + 0.031\{(1 - kp_1)(1 - kp_2) + cp_1p_2 + cp_1p_4)\} \\
0.154 &= \frac{1}{(1.01)^2}[0.319p_1p_2 + 0.222\{p_1(1 - kp_2) + (1 - kp_1)p_2\} \\
&\quad + 0.133\{(1 - kp_1)(1 - kp_2) + cp_1p_2 + cp_1p_4)\} \\
&\quad + 0.051\{\{cp_1(1 - kp_4) + (1 - kp_1)cp_3\}].
\end{aligned}$$

While it is not too difficult to solve this system in this case one can see that with more branches and more derivatives, this non-linear system of equations becomes difficult very quickly. What do we do if we observe market prices for only two derivatives defined on this stock, and only two parameters can be obtained from the market information? This is an example of what is called an incomplete market, a market in which the risk neutral distribution is not uniquely specified by market information. In general when we have fewer equations than parameters in a model, there are really only two choices

(a) Simplify the model so that the number of unknown parameters and the number of equations match.

(b) Determine additional natural criteria or constraints that the parameters must satisfy.

In this case, for example, one might prefer a model in which the probability of a step up or down depends on the time, but not on the current price of the stock. This assumption would force equal all of $p_2 = p_3 = p_4$ and simplify the system of equations above. For example using only the prices of the first two derivatives, we obtain equations, which, when solved, determine the probabilities on the other branches as well.

$$0.0016 = \frac{1}{(1.01)^2}(1.186 - 1.173)p_1p_2$$
$$0.0155 = \frac{1}{(1.01)^2}[(1.186 - 1.071)p_1p_2 + (1.089 - 1.071)\{p_1(1 - kp_2) + (1 - kp_1)p_2\}]$$

This example reflects a basic problem which occurs often when we build a reasonable and flexible model in finance. Frequently there are more parameters than there are marketable securities from which we can estimate these parameters. It is quite common to react by simplifying the model. For example, it is for this reason that binomial trees (with only two branches emanating from each node) are often preferred to the trinomial tree example we use above, even though they provide a worse approximation to the actual distribution of stock

returns.

In general if there are n different securities (excluding derivatives whose value is a function of one or more of these) and if each security can take any one of m different values, then there are a total of m^n possible states of nature at time $t = 1$. The Q measure must assign a probability to each of them. This results in a total of m^n unknown probability values, which, of course must add to one, and result in the right expectation for each of n marketable securities. To uniquely determine Q we would require a total of $m^n - n - 1$ equations or $m^n - n - 1$ different derivatives. For example for $m = 10$, $n = 100$, approximately one with a hundred zeros, a prohibitive number, are required to uniquely determine Q. In a complete market, Q is uniquely determined by marketed securities, but in effect no real market can be complete. In real markets, one asset is not perfectly replicated by a combination of other assets because there is no value in duplication. Whether an asset is a derivative whose value is determined by another marketed security, together with interest rates and volatilities, markets rarely permit exact replication. The most we can probably hope for in practice is to find a model or measure Q in a subclass of measures with desirable features under which

$$E_Q[\frac{B_t}{B_T}V(S_T)|H_t] \approx V_t \quad \text{for all marketable } V. \tag{2.9}$$

Even if we had equalities in (2.9), this would represent typically fewer equations than the number of unknown Q probabilities so some simplification of the model is required before settling on a measure Q. One could, at one's peril, ignore the fact that certain factors in the market depend on others. Similar stocks behave similarly, and none may be actually independent. Can we, with any reasonable level of confidence, accurately predict the effect that a lowering of interest rates will have on a given bank stock? Perhaps the best model for the future behaviour of most processes is the past, except that as we have seen the historical distribution of stocks do not generally produce a risk-neutral

measure. Even if historical information provided a flawless guide to the future, there is too little of it to accurately estimate the large number of parameters required for a simulation of a market of reasonable size. Some simplification of the model is clearly necessary. Are some baskets of stocks independent of other combinations? What independence can we reasonably assume over time?

As a first step in simplifying a model, consider some of the common measures of behaviour. Stocks can go up, or down. The drift of a stock is a tendency in one or other of these two directions. But it can also go *up and down*- by a lot or a little. The measure of this, the variance or variability in the stock returns is called the *volatility* of the stock. Our model should have as ingredients these two quantities. It should also have as much dependence over time and among different asset prices as we have evidence to support.

Determining the Process B_t.

We have seen in the last section that given the Q or risk-neutral measure, we can, at least in theory, determine the price of a derivative if we are given the price B_t of a risk-free investment at time t (in finance such a yardstick for measuring and discounting prices is often called a "numeraire"). Unfortunately no completely liquid risk-free investment is traded on the open market. There are government treasury bills which, depending on the government, one might wish to assume are almost risk-free, and there are government bonds, usually with longer terms, which complicate matters by paying dividends periodically. The question dealt with in this section is whether we can estimate or approximate an approximate risk-free process B_t given information on the prices of these bonds. There are typically too few genuinely risk-free bonds to get a detailed picture of the process $B_s, s > 0$. We might use government bonds for this purpose, but are these genuinely risk-free? Might not the additional use of bonds issued by other large corporations provide a more detailed picture of the bank account process B_s?

Can we incorporate information on bond prices from lower grade debt? To do so, we need a simple model linking the debt rating of a given bond and the probability of default and payoff to the bond-holders in the event of default. To begin with, let us assume that a given basket of companies, say those with a common debt rating from one of the major bond rating organisations, have a common distribution of default time. The thesis of this section is that even if no totally risk-free investment existed, we might still be able to use bond prices to estimate what interest rate such an investment would offer.

We begin with what we know. Presumably we know the current prices of marketable securities. This may include prices of certain low-risk bonds with face value F, the value of the bond on maturity at time T. Typically such a bond pays certain payments of value d_t at certain times $t < T$ and then the face value of the bond F at maturity time T, *unless the bond-holder defaults.* Let us assume for simplicity that the current time is 0. The current bond prices P_0 provide some information on B_t as well as the possibility of default. Suppose we let τ denote the random time at which default or bankruptcy would occur. Assume that the effect of possible default is to render the payments at various times random so for example d_t is paid provided that default has not yet occurred, i.e. if $\tau > t$, and similarly the payment on maturity is the face value of the Bond F if default has not yet occurred and if it has, some fraction of the face value pF is paid. When a real bond defaults, the payout to bondholders is a complicated function of the hierarchy of the bond and may occur before maturity, but we choose this model with payout at maturity in any case for simplicity. Then the current price of the bond is the expected discounted value of all future payments, so

$$\begin{aligned} P_0 &= E_Q\Big(\sum_{\{s;0<s<T\}} \frac{1}{B_s} d_s I(\tau > s) + \frac{pF}{B_T} I(\tau \le T) + \frac{F}{B_T} I(\tau > T)\Big) \\ &= \sum_{\{s;0<s<T\}} d_s E_Q[B_s^{-1} I(\tau > s)] + F E_Q[B_T^{-1}(p + (1-p) I(\tau > T))] \end{aligned}$$

The bank account process B_t that we considered is the compounded value at time of an investment of \$1 deposited at time 0. This value might be random but the interest rate is declared at the beginning of each period so, for example, B_t is completely determined at time $t-1$. In measure-theoretical language, B_t is H_{t-1} measurable for each t. With Q is the risk-neutral distribution

$$P_0 = E_Q\{\sum_{\{s;0<s<T\}} d_s B_s^{-1} Q(\tau > s|H_{s-1}) + FB_T^{-1}(p + (1-p)Q(\tau > T|H_{T-1}))\}.$$

This takes a form very similar to the price of a bond which does not default but with a different bank account process. Suppose we define a new bank account process $\widetilde{B_s}$, equivalent in expectation to the risk-free account, but that only pays if default does not occur in the interval. Such a process must satisfy

$$E_Q(\widetilde{B_s} I(\tau > s)|H_{s-1}) = B_s.$$

From this we see that the process $\widetilde{B_s}$ is defined by

$$\widetilde{B_s} = \frac{B_s}{Q[\tau > s|H_{s-1}]} \quad \text{on the set } Q[\tau > s|H_{s-1}] > 0.$$

In terms of this new bank account process, the price of the bond can be rewritten as

$$P_0 = E_Q\{\sum_{\{s;0<s<T\}} d_s \widetilde{B_s^{-1}} + (1-p)F\widetilde{B_T^{-1}} + pFB_T^{-1}\}.$$

If we subtract from the current bond price the present value of the guaranteed payment of pF, the result is

$$P_0 - pFE_Q(B_T^{-1}) = E_Q\{\sum_{\{s;0<s<T\}} d_s \widetilde{B_s^{-1}} + (1-p)F\widetilde{B_T^{-1}}\}.$$

This equation has a simple interpretation. The left side is the price of the bond reduced by the present value of the guaranteed payment on maturity Fp. The right hand side is the current value of a risk-free bond paying the same dividends, with interest rates increased by replacing B_s by $\widetilde{B_s}$ and with face value $F(1-p)$ all discounted to the present using the bank account process

$\widetilde{B_s}$. In words, *to value a defaultable bond, augment the interest rate using the probability of default in intervals, change the face value to the potential loss of face value on default and then add the present value of the guaranteed payment on maturity.*

Typically we might expect to be able to obtain prices of a variety of bonds issued on one firm, or firms with similar credit ratings. If we are willing to assume that such firms share the same conditional distribution of default time $Q[\tau > s|H_{s-1}]$ then they must all share the same process $\widetilde{B_s}$ and so each observed bond price P_0 leads to an equation of the form

$$P_0 = \sum_{\{s;0<s<T\}} d_s \widetilde{v_s} + (1-p)F\widetilde{B_T^{-1}} + pFv_T.$$

in the unknowns $\widetilde{v}_s = E_Q(\widetilde{B_s^{-1}}), ...s \leq T.$ and $v_T = E_Q(B_T^{-1})$. If we assume that the coupon dates of the bonds match, then k bonds of a given maturity T and credit rating will allow us to estimate the k unknown values of $\widetilde{v}_s$. Since the term v_T is included in all bonds, it can be estimated from all of the bond prices, but most accurately from bonds with very low risk.

Unfortunately, this model still has too many unknown parameters to be generally useful. We now consider a particular case that is considerably simpler. While it seems unreasonable to assume that default of a bond or bankruptcy of a firm is unrelated to interest rates, one might suppose some simple model which allows a form of dependence. For most firms, one might expect that the probability of survival another unit time is negatively associated with the interest rate. For example we might suppose that the probability of default in the next time interval conditional on surviving to the present is a function of the current interest rate, for example

$$h_t = Q(\tau = t|\tau \geq t, r_t) = \frac{a + (b-1)r_t}{1 + a + br_t}.$$

The quantity h_t is a more natural measure of the risk at time t than are other measures of the distribution of τ and the function h_t is called the hazard

function. If the constant $b > 1+a$, then the "hazard" h_t increases with increasing interest rates, otherwise it decreases. In case the default is independent of the interest rates, we may put $b = 1+a$ in which case the hazard is $a/(1+a)$. Then on the set $[\tau \geq s]$

$$\widetilde{B_s} = \frac{1+r_s}{1-h_s}\widetilde{B}_{s-1} = (1+a+br_s)\widetilde{B}_{s-1}$$

which means that the bond is priced using a similar bank account process but one for which the effective interest rate is not r_s but $a + br_s$. The difference $a + (b-1)r_s$ between the effective interest rate and r_s is usually referred to as the *spread* and this model justifies using a linear function to model this spread. Now suppose that default is assumed independent of the past history of interest rates under the risk-neutral measure Q. In this case, $b = 1+a$ and the spread is $a(1+r_s) \simeq a \simeq a/(1+a)$ provided both a and r_s is small. So in this case the spread gives an approximate risk-neutral probability of default in a given time interval, conditional on survival to that time.

We might hope that the probabilities of default are very small and follow a relatively simple pattern. If the pattern is not perfect, then little harm results provided that indeed the default probabilities are small. Suppose for example that the time of default follows a geometric distribution so that the hazard is constant $h_t = h = a/(1+a)$. Then

$$\widetilde{B_s} = (1+a)^s B_s \text{ for } s > 0.$$

$\widetilde{B_s}$ grows faster than B_s and it grows even faster as the probability of default h increases. The effective interest rate on this account is approximately a units per period higher.

Given only three bond prices with the same default characteristics, for example, and assuming constant interest rates so that $B_s = (1+r)^s$, we may solve for the values of the three unknown parameters (r, a, p) equations of the form

$$P_0 - pF(1+r)^{-T} = \sum_{0<s<T} (1+a+r+ar)^{-s} d_s + (1-p)F(1+a+r+ar)^{-T}.$$

Market prices for a minimum of three different bonds would allow us to solve for the unknowns (r, a, p) and these are obtainable from three different bonds.

Minimum Variance Portfolios and the Capital Asset Pricing Model.

Let us begin by building a model for portfolios of securities that captures many of the features of market movements. We assume that by using the methods of the previous section and the prices of low-risk bonds, we are able to determine the value B_t of a risk-free investment at time t in the future. Normally these values might be used to discount future stock prices to the present. However for much of this section we will consider only a single period and the analysis will be essentially the same with our without this discounting.

Suppose we have a number n of potential investments or securities, each risky in the sense that prices at future dates are random. Suppose we denote the price of these securities at time t by $S_i(t), i = 1, 2, ..., n$. There is a better measure of the value of an investment than the price of a security or even the change in the price of a security $S_i(t) - S_i(t-1)$ over a period because this does not reflect the cost of our initial investment. A common measure on investments that allows to obtain prices, but is more stable over time and between securities is the *return*. For a security that has prices $S_i(t)$ and $S_i(t+1)$ at times t and $t+1$, we define the return $R_i(t+1)$ on the security over this time interval by

$$R_i(t+1) = \frac{S_i(t+1) - S_i(t)}{S_i(t)}.$$

For example a stock that moved in price from \$10 per share to \$11 per share over a period of time corresponds to a return of 10%. Returns can be measured

in units that are easily understood (for example 5% or 10% per unit time) and are independent of the amount invested. Obviously the \$1 profit obtained on the above stock could has easily been obtained by purchasing 10 shares of a stock whose value per share changed from \$1.00 to \$1.10 in the same period of time, and the return in both cases is 10%. Given a sequence of returns and the initial value of a stock $S_i(0)$, it is easy to obtain the stock price at time t from the initial price at time 0 and the sequence of returns.

$$\begin{aligned} S_i(t) &= S_i(0)(1+R_i(1))(1+R_i(2))...(1+R_i(t)) \\ &= S_i(0)\Pi_{s=1}^{t}(1+R_i(s)). \end{aligned}$$

Returns are not added over time they are multiplied as above. A 10% return followed by a 20% return is not a 30% return but a return equal to $(1+.1)(1+.2)-1$ or 32%. When we buy a portfolio of stocks, the individual stock returns combine in a simple fashion to give the return on the whole portfolio. For example suppose that we wish to invest a total amount $\$I(t)$ at time t. The amounts will change from period to period because we may wish to reinvest gains or withdraw sums from the account. Suppose the proportion of our total investment in stock i at time t is $w_i(t)$ so that the amount invested in stock i is $w_i(t)I(t)$. Note that since $w_i(t)$ are proportions, $\sum_{i=1}^{n} w_i(t) = 1$. What is the return on this investment over the time interval from t to $t+1$? At the end of this period of time, the value of our investment is

$$I(t)\sum_{i=1}^{n} w_i(t)S_i(t+1).$$

If we now subtract the value invested at the beginning of the period and divide by the value at the beginning, we obtain

$$\frac{I(t)\sum_{i=1}^{n} w_i(t)S_i(t+1) - I(t)\sum_{i=1}^{n} w_i(t)S_i(t)}{I(t)\sum_{i=1}^{n} w_i(t)S_i(t)} = \sum_{i=1}^{n} w_i(t)R_i(t+1)$$

which is just a weighted average of the individual stock returns. Note that it does not depend on the initial price of the stocks or the total amount that we

invested at time t. The advantage in using returns instead of stock prices to assess investments is that the *return of a portfolio over a period is a value-weighted average of the returns of the individual investments.*

When time is measured continuously, we might consider defining returns by using the definition above for a period of length h and then reducing h. In other words we could define the instantaneous returns process as

$$\lim_{h\to 0} \frac{S_i(t+h) - S_i(t)}{S_i(t)}.$$

In most cases, the returns over shorter and shorter periods are smaller and smaller, and approach the limit zero so some renormalization is required above. It seems more sensible to consider returns per unit time and then take a limit i.e.

$$R_i(t) = \lim_{h\to 0} \frac{S_i(t+h) - S_i(t)}{hS_i(t)}.$$

Notice that by the definition of the derivative of a logarithm and assuming that this derivative is well-defined,

$$\begin{aligned} \frac{d\ln(S_i(t))}{dt} &= \frac{1}{S_i(t)}\frac{d}{dt}S_i(t) \\ &= \lim_{h\to 0} \frac{S_i(t+h) - S_i(t)}{hS_i(t)} \\ &= R_i(t) \end{aligned}$$

In continuous time, if the stock price process $S_i(t)$ is differentiable, the natural definition of the returns process is the derivative of the logarithm of the stock price. This definition needs some adjustment later because the most common continuous time models for asset prices does not result in a differentiable process $S_i(t)$. The solution we will use then will be to adopt a new concept of an integral and recast the above in terms of this integral.

The Capital Asset Pricing Model (CAPM)

We now consider a simplified model for building a portfolio based on quite basic properties of the potential investments. Let us begin by assuming a single period so that we are planning at time $t = 0$ investments over a period ending at time $t = 1$. We also assume that investors are interested in only two characteristics of a potential investment, the expected value and the variance of the return over this period. We have seen that the return of a portfolio is the value-weighted average of the returns of the individual investments so let us denote the return on stock i by

$$R_i = \frac{S_i(1) - S_i(0)}{S_i(0)},$$

and define $\mu_i = E(R_i)$ and w_i the proportion of my total investment in stock i at the beginning of the period. For brevity of notation, let $\mathbf{R}, \mathbf{w}$ and μ denote the column vectors

$$\mathbf{R} = \begin{pmatrix} R_1 \\ R_2 \\ . \\ . \\ . \\ R_n \end{pmatrix}, \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ . \\ . \\ . \\ w_n \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_n \end{pmatrix}.$$

Then the return on the portfolio is $\sum_i w_i R_i$ or in matrix notation $\mathbf{w}'\mathbf{R}$. Let us suppose that the covariance matrix of returns is the $n \times n$ matrix $\mathbf{\Sigma}$ so that

$$cov(R_i, R_j) = \Sigma_{ij}.$$

We will frequently use the following properties of expected value and covariance.

Lemma 3 *Suppose*

$$\mathbf{R} = \begin{pmatrix} R_1 \\ R_2 \\ . \\ . \\ . \\ R_n \end{pmatrix}$$

is a column vector of random variables R_i *with* $E(R_i) = \mu_i, i = 1, ..., n$ *and suppose* $\mathbf{R}$ *has covariance matrix* Σ. *Suppose* A *is a non-random vector or matrix with exactly* n *columns so that* $A\mathbf{R}$ *is a vector of random variables. Then* $A\mathbf{R}$ *has mean* $A\mu$ *and covariance matrix* $A\Sigma A'$.

Then it is easy to see that the expected return from the portfolio with weights w_i is $\sum_i w_i E(R_i) = \sum_i w_i \mu_i = \mathbf{w}'\mu$ and the variance is

$$var(\mathbf{w}'\mathbf{R}) = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}.$$

We will need to assume that the covariance matrix $\mathbf{\Sigma}$ is non-singular, that is it has a matrix inverse $\mathbf{\Sigma}^{-1}$. This means, at least for the present, that our model covers only risky stocks for which the variance of returns is positive. If a risk-free investment is available (for example a secure bond whose return is known exactly in advance), this will be handled later.

In the Capital Asset Pricing model it is assumed at the outset that investors concentrate on two measures of return from a portfolio, the expected value and standard deviation. These expected values and variances are computed under the real-world probability distribution P not under some risk-neutral Q measure. Clearly investors prefer high expected return, wherever possible, associated with small standard deviation of return. As a first step in this direction suppose we plot the standard deviation and expected return for the n stocks, i.e. the n points $\{(\sigma_i, \mu_i), i = 1, 2, ..., n\}$ where $\mu_i = E(R_i)$ and $\sigma_i = \sqrt{var(R_i)} = \sqrt{\Sigma_{ii}}$. These n points do not consist of the set of all achievable values of mean and

standard of return, since we are able to construct a portfolio with a certain proportion of our wealth w_i invested in stock i.In fact the set of possible points consists of

$$\{(\sqrt{\mathbf{w}'\Sigma\mathbf{w}}, \mathbf{w}'\mu) \text{ as the vector } \mathbf{w} \text{ ranges over all possible weights such that } \sum w_i = 1\}.$$

The resulting set has a boundary as in Figure 2.2.

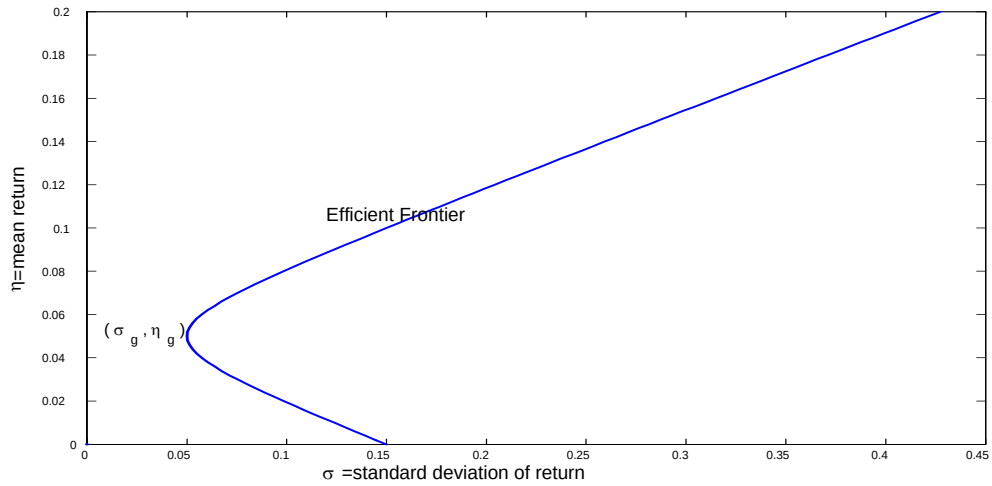

Figure 2.2: The Efficient Frontier

[FIGURE 2.2 ABOUT HERE]

Exactly what form this figure takes depends in part on the assumptions applied to the weights. Since they represent the proportion of our total investment in each of n stocks they must add to one. Negative weights correspond to selling short one stock so as to be able to invest more in another, and we may assume no limit on our ability to do so. In this case the only constraint on $\mathbf{w}$ is the constraint $\sum w_i = 1$. With this constraint alone, we can determine the boundary of the admissible set by fixing the vertical component (the mean return) of a portfolio at some value say η and then finding the minimum possible standard

deviation corresponding to that mean. This allows us to determine the leading edge or left boundary of the region. The optimisation problem is as follows

$$\min \sqrt{\mathbf{w}'\Sigma\mathbf{w}} \text{ subject to}$$

subject to the two constraints on the weights

$$\mathbf{w}'\mathbf{1} = 1$$

$$\mathbf{w}'\mu = \eta.$$

where $\mathbf{1}$ is the column vector of n ones. Since we will often make use of the method of Lagrange multipliers for constrained problems such as this one, we interject a lemma justifying the method. For details, consult Apostol (1973), Section 13.7 or any advanced calculus text.

Lemma 4 *Consider the optimisation problem*

$$\min\{f(w); w \in R^n\} \textit{ subject to } p \textit{ constraints} \tag{2.10}$$

$$\textit{of the form } g_1(w) = 0, g_2(w) = 0, ..., g_p(w) = 0.$$

Then provided the functions $f, g_1, ..., g_p$ are continuously differentiable, a necessary solution for a solution to (2.10) is that there is a solution in the $n + p$ variables $(w_1, ...w_n, \lambda_1, ..., \lambda_p)$ of the equations

$$\frac{\partial}{\partial w_i}\{f(w) + \lambda_1 g_1(w) + ... + \lambda_p g_p(w)\} = 0, i = 1, 2, ..., n$$
$$\frac{\partial}{\partial \lambda_j}\{f(w) + \lambda_1 g_1(w) + ... + \lambda_p g_p(w)\} = 0, j = 1, 2, ..., p.$$

This constants $\lambda_i$are called the Lagrange multipliers and the function that is differentiated, $\{f(w) + \lambda_1 g_1(w) + ... + \lambda_p g_p(w)\}$ is the Lagrangian.

Let us return to our original minimization problem with one small simplification. Since minimizing $\sqrt{\mathbf{w}'\Sigma\mathbf{w}}$ results in the same weight vector $\mathbf{w}$ as does minimizing $\mathbf{w}'\Sigma\mathbf{w}$ we choose the latter as our objective function.

We introduce Lagrange multipliers λ_1, λ_2 and we wish to solve

$$\frac{\partial}{\partial w_i}\{\mathbf{w}^{'}\Sigma\mathbf{w} + \lambda_1(w'\mathbf{1} - 1) + \lambda_2(w'\mu - \eta)\} = 0, i = 1, 2, ..., n$$
$$\frac{\partial}{\partial \lambda_j}\{\mathbf{w}^{'}\Sigma\mathbf{w} + \lambda_1(w'\mathbf{1} - 1) + \lambda_2(w'\mu - \eta)\} = 0, j = 1, 2.$$

The solution is obtained from the simple differentiation rule

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} = 2\mathbf{\Sigma}\mathbf{w} \text{ and } \frac{\partial}{\partial \mathbf{w}}\mu'\mathbf{w} = \mathbf{w}$$

and is of the form

$$\mathbf{w} = \lambda_1\Sigma^{-1}\mathbf{1} + \lambda_2\Sigma^{-1}\mu$$

with the Lagrange multipliers λ_1, λ_2chosen to satisfy the two constraints, i.e.

$$\lambda_1\mathbf{1}'\Sigma^{-1}\mu + \lambda_2\mathbf{1}'\Sigma^{-1}\mathbf{1} = 1$$
$$\lambda_1\mu'\Sigma^{-1}\mu + \lambda_2\mu'\Sigma^{-1}\mathbf{1} = \eta.$$

Suppose we define an $n \times 2$ matrix M with columns $\mathbf{1}$ and μ,

$$M = [\ \mathbf{1}\ \ \mu\ ]$$

and the 2×2 matrix $A = (M'\Sigma^{-1}M)^{-1}$, then the Lagrange multipliers are given by the vector

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = A \begin{bmatrix} 1 \\ \eta \end{bmatrix}$$

and the weights by the vector

$$\mathbf{w} = \Sigma^{-1}MA \begin{bmatrix} 1 \\ \eta \end{bmatrix}. \tag{2.11}$$

We are now in a position to identify the boundary or the curve in Figure 2.2. As the mean of the portfolio η changes, the point takes the form $(\sqrt{\mathbf{w}'\Sigma\mathbf{w}}, \eta)$

with $\mathbf{w}$ given by (2.11). Notice that

$$\begin{aligned}
\mathbf{w}'\Sigma\mathbf{w} &= [\ 1 \quad \eta\ ]A'M'\Sigma^{-1}\Sigma\Sigma^{-1}MA\begin{bmatrix} 1 \\ \eta \end{bmatrix} \\
&= [\ 1 \quad \eta\ ]A'M'\Sigma^{-1}MA\begin{bmatrix} 1 \\ \eta \end{bmatrix} \\
&= [\ 1 \quad \eta\ ]A\begin{bmatrix} 1 \\ \eta \end{bmatrix} \\
&= A_{11} + 2A_{12}\eta + A_{22}\eta^2.
\end{aligned}$$

Therefore a point on the boundary $(\sigma, \eta) = (\sqrt{\mathbf{w}'\Sigma\mathbf{w}}, \eta)$ satisfies

$$\sigma^2 - A_{22}\eta^2 - 2A_{12}\eta - A_{11} = 0$$

or

$$\begin{aligned}
\sigma^2 &= A_{22}\eta^2 + 2A_{12}\eta + A_{11} \\
&= \sigma_g^2 + A_{22}(\eta - \eta_g)^2
\end{aligned}$$

where

$$\eta_g = -\frac{A_{12}}{A_{22}} = \frac{\mathbf{1}'\Sigma^{-1}\mu}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{2.12}$$

$$\begin{aligned}
\sigma_g^2 &= A_{11} - \frac{A_{12}^2}{A_{22}} = \frac{|A|}{A_{22}} \\
&= \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}.
\end{aligned} \tag{2.13}$$

and the point (σ_g, μ_g) represents the point in the region corresponding to the minimum possible standard deviation over all portfolios. This is the most conservative investment portfolio available with this class of securities. What weights to do we need to put on the individual stocks to achieve this conservative portfolio? It is easy to see that the weight vector is given by

$$\mathbf{w}_g' = \frac{\mathbf{1}'\Sigma^{-1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{2.14}$$

and since the quantity $\mathbf{1}'\Sigma^{-1}\mathbf{1}$ in the denominator is just a scale factor to insure that the weights add to one, the amount invested in stock i is proportional to the sum of the elements of the i'th row of the inverse covariance matrix $\mathbf{\Sigma}^{-1}$.

An equation of the form

$$\sigma^2 - A_{22}(\eta - \eta_g)^2 = \sigma_g^2$$

represents a hyperbola since $A_{22} > 0$. Of course investors are presumed to prefer higher returns for a given value of the standard deviation of portfolio so it is only the upper boundary of this curve in Figure 2.2 that is efficient in the sense that there is no portfolio that is strictly better (better in the sense of higher return combined with standard deviation that is not larger).

Now let us return to a portfolio whose standard deviation and mean return lie on the efficient frontier. Let us call these *efficient portfolios*. It turns out that any portfolio on this efficient frontier has the same covariance with the minimum variance portfolio $\mathbf{w}_g'\mathbf{R}$ derived above.

Proposition 5 *Every efficient portfolio has the same covariance* $\frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$ *with the conservative portfolio* $\mathbf{w}_g'\mathbf{R}$.

Proof. We noted before that such a portfolio has mean return η and standard deviation σ which satisfy the relation

$$\sigma^2 - A_{22}\eta^2 - 2A_{12}\eta - A_{11} = 0.$$

Moreover the weights for this portfolio are described by

$$\mathbf{w} = \Sigma^{-1}MA \begin{bmatrix} 1 \\ \eta \end{bmatrix}. \tag{2.15}$$

so the returns vector from this portfolio can be written as

$$\mathbf{w}'\mathbf{R} = [\ 1 \quad \eta\ ]AM'\mathbf{\Sigma}^{-1}\mathbf{R}.$$

It is interesting to observe that the covariance of returns between this efficient portfolio and the conservative portfolio $\mathbf{w}_g^{'}\mathbf{R}$ is given by

$$\begin{aligned}
cov(\mathbf{w}_g^{'}\mathbf{R}, [\ 1 \quad \eta\ ]AM'\mathbf{\Sigma}^{-1}\mathbf{R}) &= [\ 1 \quad \eta\ ]AM'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{w}_g \\
&= [\ 1 \quad \eta\ ]A \begin{bmatrix} \mathbf{1}' \\ \mu' \end{bmatrix} \Sigma^{-1}\mathbf{1}\frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \\
&= [\ 1 \quad \eta\ ]A \begin{bmatrix} \mathbf{1}'\Sigma^{-1}\mathbf{1} \\ \mu'\Sigma^{-1}\mathbf{1} \end{bmatrix} \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \\
&= [\ 1 \quad \eta\ ] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \\
&= \frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}
\end{aligned}$$

where we use the fact that, by the definition of A,

$$A \begin{bmatrix} \mathbf{1}'\Sigma^{-1}\mathbf{1} & \mu'\Sigma^{-1}\mathbf{1} \\ \mu'\Sigma^{-1}\mathbf{1} & \mu'\Sigma^{-1}\mu \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

■

Now consider two portfolios on the boundary in Figure 2.2. For each the weights are of the same form, say

$$\mathbf{w}_p = \Sigma^{-1}MA \begin{bmatrix} 1 \\ \eta_p \end{bmatrix} \quad \text{and} \quad \mathbf{w}_q = \Sigma^{-1}MA \begin{bmatrix} 1 \\ \eta_q \end{bmatrix} \tag{2.16}$$

where the mean returns are η_p and η_q respectively. Consider the covariance between these two portfolios

$$\begin{aligned}
cov(\mathbf{w}_p'\mathbf{R}, \mathbf{w}_q'\mathbf{R}) &= \mathbf{w}_p'\mathbf{\Sigma}\mathbf{w}_q \\
&= [\ 1 \quad \eta_p\ ](M'\Sigma^{-1}M)^{-1} \begin{bmatrix} 1 \\ \eta_q \end{bmatrix} \\
&= A_{11} + A_{12}(\eta_p + \eta_q) + A_{22}\eta_p\eta_q \\
&= var(\mathbf{w}_p'\mathbf{R}) - [\ 1 \quad \eta_p\ ]A \begin{bmatrix} 0 \\ \eta_p - \eta_q \end{bmatrix}
\end{aligned}$$

An interesting special portfolio that is a "zero-beta" portfolio, one that is perfectly uncorrelated with the portfolio with weights $\mathbf{w}_p'\mathbf{R}$. This is obtained by setting the above covariance equal to 0 and solving we obtain

$$\begin{aligned}\eta_q &= -\frac{A_{11} + A_{12}\eta_p}{A_{12} + A_{22}\eta_p} \\ &= \frac{\mu'\Sigma^{-1}\mu - (\mu'\Sigma^{-1}\mathbf{1})\eta_p}{\mu'\Sigma^{-1}\mathbf{1} - (\mathbf{1}'\Sigma^{-1}\mathbf{1})\eta_p}.\end{aligned}$$

There is a simple method for determining the point $(, \eta_q)$ graphically indicated in Figure **??**. From the equation relating points on the boundary,

$$\sigma^2 - A_{22}(\eta - \eta_g)^2 = \sigma_g^2$$

we obtain

$$\frac{\partial \eta}{\partial \sigma} = \frac{\sigma}{A_{22}(\eta - \eta_g)}$$

and so the tangent line at the point (σ_p, η_p) strikes the $\sigma = 0$ axis at a point η_q which satisfies

$$\frac{\eta_p - \eta_q}{\sigma_p} = \frac{\sigma_p}{A_{22}(\eta_p - \eta_g)}$$

or

$$\begin{aligned}\eta_q &= \eta_p - \frac{\sigma_p^2}{A_{22}(\eta_p - \eta_g)} \\ &= \eta_p - \frac{A_{22}\eta_p^2 + 2A_{12}\eta_p + A_{11}}{A_{22}\eta_p + A_{12}} \\ &= -\frac{A_{11} + A_{12}\eta_p}{A_{12} + A_{22}\eta_p}. \qquad (2.17)\end{aligned}$$

Note that this is exactly the same mean return obtained earlier for the portfolio which has zero covariance with $\mathbf{w}_p'\mathbf{R}$. This shows that we can find the standard deviation and mean of this uncorrelated portfolio by constructing the tangent line at the point (σ_p, η_p) and then setting η_q to be the y-coordinate of the point where this tangent line strikes the $\sigma = 0$ axis as in Figure 2.3.

[FIGURE 2.3 ABOUT HERE]

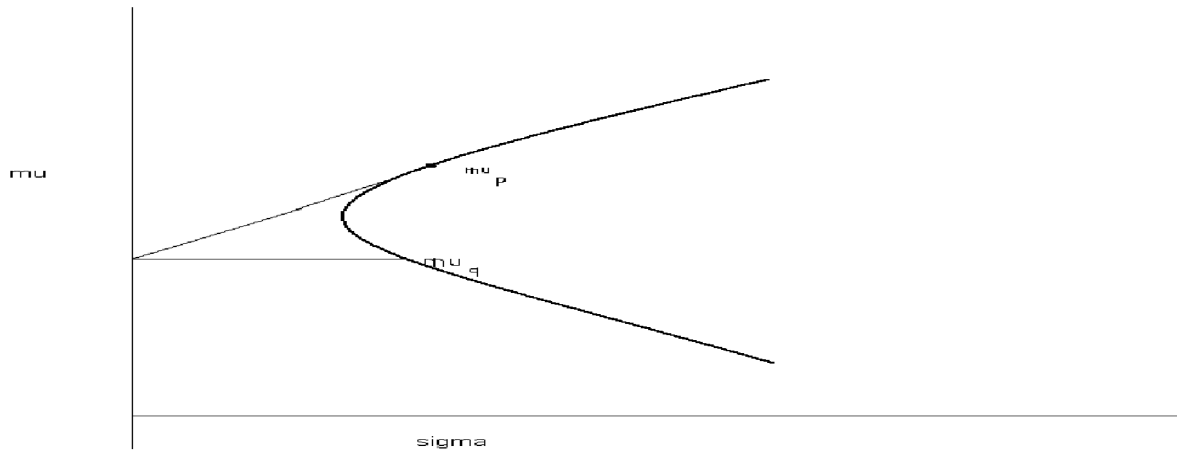

Figure 2.3: The tangent line at the point (σ_p, η_p)

Now suppose that there is available to all investors a risk-free investment. Such an investment typically has smaller return than those on the efficient frontier but since there is no risk associated with the investment, its standard deviation is 0. It may be a government bond or treasury bill yielding interest rate r so it corresponds to a point in Figure 2.4 at $(0, r)$. Since all investors are able to include this in their portfolio, the efficient frontier changes. In fact if an investor invests an amount β in this risk-free investment and amount $1-\beta$ (this may be negative) in the risky portfolio with standard deviation and mean return (σ_p, η_p) then the resulting investment has mean return

$$E(\beta r + (1-\beta)\mathbf{w}_p'\mathbf{R}) = \beta r + (\mathbf{1}-\beta)\eta_p$$

and standard deviation of return

$$\sqrt{Var(\beta r + (1-\beta)\mathbf{w}_p'\mathbf{R})} = (1-\beta)\sigma_p.$$

This means that every point on a line joining $(0, r)$ to points in the risky portfolio are now attainable and so the new set of attainable values of (σ, η) consists of a cone with vertex at $(0, r)$,the region shaded in Figure 2.4. The efficient frontier

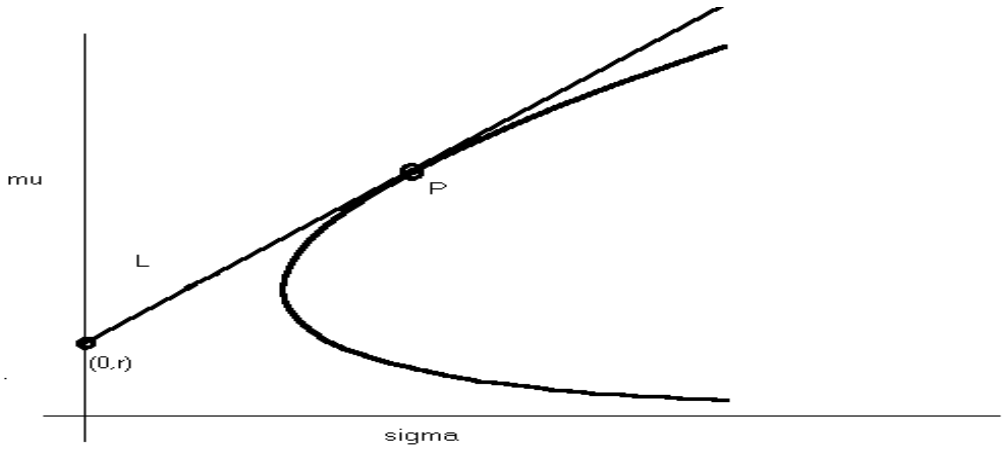

Figure 2.4: _____

is now the line L in Figure 2.4. The point m is the point at which this line is tangent to the efficient frontier determined from the risky investments. Under this theory, this point has great significance.

[FIGURE 2.4 ABOUT HERE]

Lemma 6 *The value-weighted market average corresponds to the point of tangency m of the line to the risky portfolio efficient frontier.*

From (2.17) the point m has standard deviation, mean return η_m which solves

$$\begin{aligned} r &= -\frac{A_{11} + A_{12}\eta_m}{A_{12} + A_{22}\eta_m} \\ &= \frac{\mu'\Sigma^{-1}\mu - (\mu'\Sigma^{-1}\mathbf{1})\eta_m}{\mu'\Sigma^{-1}\mathbf{1} - (\mathbf{1}'\Sigma^{-1}\mathbf{1})\eta_m} \end{aligned}$$

and this gives

$$\eta_m = \frac{\mu'\Sigma^{-1}\mu - r(\mu'\Sigma^{-1}\mathbf{1})}{\mu'\Sigma^{-1}\mathbf{1} - r(\mathbf{1}'\Sigma^{-1}\mathbf{1})}.$$

The corresponding weights on individual stocks are given by

$$\begin{aligned}
\mathbf{w}_m &= \Sigma^{-1} M A \begin{bmatrix} 1 \\ \eta_m \end{bmatrix}. \\
&= \Sigma^{-1}[\mathbf{1}\ \mu] \begin{bmatrix} A_{11} + A_{12}\eta_m \\ A_{12} + A_{22}\eta_m \end{bmatrix} \\
&= c\Sigma^{-1}[\mathbf{1}\ \mu] \begin{bmatrix} -r \\ 1 \end{bmatrix}, \quad \text{where } c = A_{12} + A_{22}\eta_m \\
&= c\Sigma^{-1}(\mu - r\mathbf{1}).
\end{aligned}$$

These market weights depend essentially on two quantities. If R denotes the correlation matrix

$$R_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$$

where $\sigma_i = \sqrt{\Sigma_{ii}}$ is the standard deviation of the returns from stock i, and

$$\lambda_i = \frac{\mu_i - r}{\sigma_i}$$

is the standardized excess return or the price of risk, then the weight w_i on stock i is such that

$$w_i \sigma_i \propto R^{-1}\lambda \tag{2.18}$$

with λ the column vector of values of λ_i. For the purpose of comparison, recall that the conservative portfolio, one minimizing the variance over all portfolios of risky stocks, has weights

$$\mathbf{w}_g \propto \Sigma^{-1}\mathbf{1}$$

which means that the weight on stock i satisfies a relation exactly like (2.18) except that the mean returns μ_i have all been replaced by the same constant.

Let us suppose that stocks, weighed by their total capitalization in the market result in some weight vector $\mathbf{w} \neq \mathbf{w_m}$. When there is a risk-free investment, m is the only point in the risky stock portfolio that lies in the efficient frontier and so evidently if we are able to trade in a market index (a stock whose value

depends on the total market), we can find an investment which is a combination of the risk-free investment with that corresponding to m which has the same standard deviation as $\mathbf{w}'\mathbf{R}$ but higher expected return. By selling short the market index and buying this new portfolio, an arbitrage is possible. In other words, the market will not stay in this state for long.

If the market portfolio m has standard deviation σ_m and mean η_m, then the line L is described by the relation

$$\eta = r + \frac{\eta_m - r}{\sigma_m}\sigma.$$

For any investment with mean return η and standard deviation of return σ to be competitive, it must lie on this efficient frontier, i.e. it must satisfy the relation

$$\begin{aligned} &\eta - r = \beta(\eta_m - r), \quad \text{where} \quad \beta = \frac{\sigma}{\sigma_m} \text{ or equivalently} \\ &\frac{\eta - r}{\sigma} = \frac{(\eta_m - r)}{\sigma_m}. \end{aligned} \tag{2.19}$$

This is the most important result in the capital asset pricing model. The excess return of a stock $\eta - r$ divided by its standard deviation σ is supposed constant, and is called the *Sharpe ratio* or the *market price of risk*. The constant β called the *beta* of the stock or portfolio and represents the change in the expected portfolio return for each unit change in the market. It is also the ratio of the standard deviations of return of the stock and the market. Values of $\beta > 1$ indicate a stock that is more variable than the market and tends to have higher positive and negative returns, whereas values of $\beta < 1$ are investments that are more conservative and less volatile than the market as a whole.

We might attempt to use this model to simplify the assumed structure of the joint distribution of stock returns. One simple model in which (2.19) holds is one in which all stocks are linearly related to the market index through a simple linear regression. In particular, suppose the return from stock i, R_i, is

related to the return from the market portfolio R_m by

$$R_i - r = \beta_i(R_m - r) + \epsilon_i, \text{ where } \beta_i = \frac{\sigma_i}{\sigma_m}, \text{ and } \sigma_i^2 = \Sigma_{ii}.$$

The "errors" ϵ_i are assumed to be random variables, uncorrelated with the market returns R_m. This model is called the *single-index model* relating the returns from the stock R_i and from the market portfolio R_m.It has the merit that the relationship (2.19) follows immediately.

Taking variance on both sides, we obtain

$$var(R_i) = \beta_i^2 var(R_m) + var(\epsilon_i) = \sigma_i^2 + var(\epsilon) > \sigma_i^2$$

which contradicts the assumption that $var(R_i) = \sigma_i^2$. What is the cause of this contradiction? The relationship (2.19) assumes that the investment lies on the efficient frontier. Is this not a sufficient condition for investors to choose this investment? All that is required for rational investors to choose a particular stock is that it forms part of a portfolio which does lie on the efficient frontier.

Is every risk in an efficient market rewarded with additional expected return? We cannot expect the market to compensate us with a higher rate of return for additional risks that could be diversified away. Suppose, for example, we have two stocks with identical values of β. Suppose their returns $R_1$and R_2 both satisfy a linear regression relation above

$$R_i - r = \beta(R_m - r) + \epsilon_i, \ i = 1, 2,$$

where $cov(\epsilon_1, \epsilon_2) = 0$. Consider an investment of equal amounts in both stocks so that the return is

$$\frac{R_1 + R_2}{2} = \beta(R_m - r) + \frac{\epsilon_1 + \epsilon_2}{2}.$$

For simplicity assume that $\sigma_1 \leq \sigma_2$ and notice that the variance of this new investment is

$$\beta^2\sigma_m^2 + \frac{1}{4}[var(\epsilon_1) + var(\epsilon_2)] < var(R_2).$$

The diversified investment consisting of the average of the two results in the same mean return with smaller variance. Investors should not compensated for the additional risk in stock 2 above the level that we can achieve by sensible diversification. In general, by averaging or diversifying, we are able to provide an investment with the same average return characteristics but smaller variance than the original stock. We say that the risk (i.e. $var(\epsilon_i)$) associated with stock i which can be diversified away is the *specific risk*, and this risk is not rewarded with increased expected return. Only the so-called systematic risk σ_i which cannot by removed by diversification is rewarded with increased expected return with a relation like (2.19).

The covariance matrix of stock returns is one of the most difficult parameters to estimate in practice form historical data. If there are n stocks in a market (and normally n is large), then there are $n(n+1)/2$ elements of Σ that need to be estimated. For example if we assume all stocks in the TSE 300 index are correlated this results in a total of $(300)(301)/2 = 45,150$ parameters to estimate. We might use historical data to estimate these parameters but variances and covariances among stocks change over time and it is not clear over what period of time we can safely use to estimate these parameters. In spite of its defects, the single index model can be used to provide a simple approximate form for the covariance matrix Σ of the vector of stock returns. Notice that under the model, assuming uncorrelated random errors ϵ_i with $var(\epsilon_i) = \delta_i$,

$$R_i - r = \beta_i(R_m - r) + \epsilon_i,$$

we have

$$cov(R_i, R_j) = \beta_i\beta_j\sigma_m^2, i \neq j, \qquad var(R_i) = \beta_i^2\sigma_m^2 + \delta_i.$$

Whereas n stocks would otherwise require a total of $n(n+1)/2$ parameters in the covariance matrix Σ of returns, the single index model allows us to reduce this to the $n+1$ parameters σ_m^2, and δ_i, $i = 1, ..., n$. There is the disadvantage

in this formula however that every pair of stocks in the same market must be *positively correlated,* a feature that contradicts some observations of real market returns.

Suppose we use this form $\Sigma = \beta\beta'\sigma_m^2 + \Delta$, to estimate weights on individual stocks, where Δ is the diagonal matrix with the δ_i along the diagonal and β is the column vector of individual stock betas. In this case $\Sigma^{-1} = \Delta^{-1} + c\Delta^{-1}\beta\beta'\Delta^{-1}$ where

$$c = \frac{-1}{\sigma_m^{-2} + \sum_i \beta_i^2/\delta_i} = -\sigma_m^2 \frac{1}{1 + \sum_i \beta_i^2 \sigma_m^2/\delta_i}$$

and consequently the conservative investor by (2.14) invests in stock i proportionally to the components of $\Sigma^{-1}\mathbf{1}$

$$\text{or to } \frac{1}{\delta_i} + c\beta_i(\sum_j \beta_j/\delta_j)$$

$$\text{or proportional to } \beta_i + \frac{1}{c\delta_i(\sum_j \beta_j/\delta_j)}$$

The conditional variance of R_i given the market return R_m is δ_i. Let us call this the *excess volatility for stock i. Then the weights for the conservative portfolio are linear in the beta for the stock and the reciprocal of the excess volatility.*

The weights in the market portfolio are given by

$$\mathbf{w}_m = \Sigma^{-1}MA \begin{bmatrix} 1 \\ \eta_p \end{bmatrix} = (\Delta^{-1} + c\Delta^{-1}\beta\beta'\Delta^{-1})[\ \mathbf{1}\ \ \mu\ ](M'\Sigma^{-1}M)^{-1} \begin{bmatrix} 1 \\ \eta_p \end{bmatrix}$$

Minimum Variance under Q.

Suppose we wish to find a portfolios of securities which has the smallest possible variance under the risk neutral distribution Q. For example for a given set of weights $w_i(t)$ representing the number of shares held in security i at time t, define the portfolio $\Pi(t) = \sum w_i(t)S_i(t)$. Recall from Section 2.1 that under a risk neutral distribution, all stocks have exactly the same expected return as the risk-free interest rate so the portfolio $\Pi(t)$ will have exactly the same

conditional expected rate of return under Q as all the constituent stocks,

$$E_Q[\Pi(t+1)|H_t] = \sum_i w_i(t) E_Q[S_i(t+1)|H_t] = \sum_i w_i(t) \frac{B(t+1)}{B(t)} S_i(t) = \frac{B(t+1)}{B(t)} \Pi(t).$$

Since all portfolios have the same conditional expected return under Q, we might attempt to minimize the (conditional) variance of the portfolio return of the portfolio. The natural constraint is that the cost of the portfolio is determined by the amount $c(t)$ that we presently have to invest. We might assume a constant investment over time, for example $c(t) = 1$ for all t. Alternatively, we might wish to study a *self-financing portfolio* $\Pi(t)$, one for which past gains (or perish the thought, past losses) only are available to pay for the current portfolio so we neither withdraw from nor add money to the portfolio over its lifetime. I this case $c(t) = \Pi(t)$. We wish to minimise

$$var_Q[\Pi(t+1)|H_t] \text{ subject to the constraint } \sum_i w_i(t) S_i(t) = c(t).$$

As before, the solution is quite easy to obtain, and in fact the weights are given by the vector

$$\mathbf{w}(t) = \begin{pmatrix} w_1(t) \\ w_2(t) \\ . \\ . \\ . \\ w_n(t) \end{pmatrix} = \frac{c(t)}{S'(t)\Sigma_t^{-1} S(t)} \Sigma_t^{-1} S(t).$$

where $\Sigma_t = var_Q(S(t+1)|H_t)$ is the instantaneous conditional covariance matrix of $S(t)$ under the measure Q. If my objective were to minimize risk under the Q measure, then this portfolio is optimal for fixed cost. The conditional variance of this portfolio is given by

$$var_Q(\Pi(t+1)|H_t) = \mathbf{w}'(t)\Sigma_t \mathbf{w}(t) = \frac{c^2(t)}{S'(t)\Sigma_t^{-1} S(t)}.$$

In terms of the portfolio return $R_\Pi(t+1) = \frac{\Pi(t+1)-\Pi(t)}{\Pi(t)}$, if the portfolio is self-financing so that $c(t) = \Pi(t)$, the above relation states that the conditional variance of the return $R_\Pi(t+1)$ given the past is simply

$$var_Q(R_\Pi(t+1)|H_t) = \frac{1}{S'(t)\Sigma_t^{-1}S(t)}$$

which is similar to the form of the variance of the conservative portfolio (2.13).

Similarly, covariances between returns for individual stocks and the return of the portfolio Π are given by exactly the same quantity, namely

$$cov(R_i(t+1), R_\Pi(t+1)|H_t) = \frac{1}{S'(t)\Sigma_t^{-1}S(t)}.$$

Let us summarize our findings so far. We assume that the conditional covariance matrix Σ_t of the vector of stock prices is non-singular. *Under the risk neutral measure, all stocks have exactly the same expected returns equal to the risk-free rate. There is a unique self-financing minimum-variance portfolio* $\Pi(t)$ *and all stocks have exactly the same conditional covariance* β *with* Π. *All stocks have exactly the same regression coefficient* β *when we regress on the minimum variance portfolio.*

Are other minimum variance portfolios conditionally uncorrelated with the portfolio we obtained above. Suppose we define $\Pi_2(t)$ similarly to minimize the variance subject to the condition that $Cov_Q(\Pi_2(t+1), \Pi(t+1)|H_t) = 0$. It is easy to see that this implies that the cost of such a portfolio at the beginning of each period is 0. This means that in this new portfolio, there is a perfect balance between long and short stocks, or that the value of the long and short stocks are equal.

The above analysis assumes that our objective is minimizing the variance of the portfolio under the risk-neutral distribution Q. Two objections could be made. First we argued earlier that the performance of an investment should be made through the *returns* , not through the stock *prices.* Since under the risk neutral measure Q, the expected return from every stock is the risk-free rate of

return, we are left with the problem of minimizing the variance of the *portfolio return.* By our earlier analysis, this is achieved when the proportion of our total investment at each time period in stock i is chosen as the corresponding component of the vector $\frac{\Sigma_t^{-1}1}{1'\Sigma_t^{-1}1}$ where now Σ_t is the conditional covariance matrix of the stock *returns.* This may appear to be a different criterion and hence a different solution, but because at each time step the stock price is a linear function of the return $S_i(t+1) = S_i(t)(1+R_i(t+1))$ the variance minimizing portfolios are essentially the same. There is another objection however to an analysis in the risk-neutral world of Q. This is a distribution which determines the value of options in order to avoid arbitrage in the system, not the actual distribution of stock prices. It is not clear what the relationship is between the covariance matrix of stock prices under the actual historical distribution and the risk neutral distribution Q, but observations seem to indicate a very considerable difference. Moreover, if this difference is large, there is very little information available for estimating the parameters of the covariance matrix under Q, since historical data on the fluctuations of stock prices will be of doubtful relevance.

Entropy: choosing a Q measure

Maximum Entropy

In 1948 in a fundamental paper on the transmission of information, C. E. Shannon proposed the following idea of *entropy.* The entropy of a distribution attempts to measure the expected number of steps required to determine a given outcome of a random variable with a given distribution when using a simple binary poll. For example suppose that a random variable X has distribution

given by

x	0	1	2
$P[X = x]$	.25	.25	.5

In this case, if we ask first whether the random variable is ≥ 2 and then, provided the answer is no, if it is ≥ 1, the expected number of queries to ascertain the value of the random variable is $1+1(1/2) = 1.5$. There is no more efficient scheme for designing this binary poll in this case so we will take 1.5 to be a measure of entropy of the distribution of X. For a discrete distribution, such that $P[X = x] = p(x)$, the entropy may be defined to be

$$H(p) = E\{-\ln(p(X))\} = -\sum_x p(x)\ln(p(x)).$$

More generally we define the entropy of an arbitrary distribution through the form for a discrete distribution. If P is a probability measure (see the appendix),

$$H(P) = \sup\{-\sum P(E_i)\ln(P(E_i))\}$$

where the supremum is taken over all finite partitions $(E_i\}$ of the space.

In the case of the above distribution, if we were to replace the natural logarithm by the log base 2, (ln and $\log_2$ differ only by a scale factor and are therefore the corresponding measures of entropy are equivalent up a constant multiple) notice that $-\sum_x p(x)\log_2(p(x)) = .5(1) + .5(2) = 1.5$, so this formula correctly measures the difficulty in ascertaining a random variable from a sequence of questions with yes-no or binary answers. This is true in general. The complexity of a distribution may be measured by the expected number of questions in a binary poll to determine the value of a random variable having that distribution, and such a measure results in the entropy $H(p)$ of the distribution.

Many statistical distributions have an interpretation in terms of maximizing entropy and it is often remarkable how well the maximum entropy principle reproduces observed distributions. For example, suppose we know that a discrete random variable takes values on a certain set of n points. What distribution p

on this set maximizes the entropy $H(p)$? First notice that if p is uniform on n points, $p(x) = 1/n$ for all x and so the entropy is $-\sum_x \frac{1}{n}\ln(\frac{1}{n}) = \ln(n)$. Now consider the problem of maximizing the entropy $H(p)$ for any distribution on n points subject to the constraint that the probabilities add to one. As in (2.10), the Lagrangian for this problem is $-\sum_x p(x)\ln(p(x)) - \lambda\{\sum_x p(x) - 1\}$ where λ is a Lagrange multiplier. Upon differentiating with respect to $p(x)$ for each x, we obtain $-\ln(p(x)) - 1 - \lambda = 0$ or $p(x) = e^{-(1+\lambda)}$. The probabilities evidently do not depend on x and the distribution is thus uniform. Applying the constraint that the sum of the probabilities is one results in $p(x) = 1/n$ for all x. *The discrete distribution on n points which has maximum entropy is the uniform distribution.* What if we repeat this analysis using additional constraints, for example on the moments of the distribution? Suppose for example that we require that the mean of the distribution is some fixed constant μ and the variance fixed at σ^2. The problem is similar to that treated above but with two more terms in the Lagrangian for each of the additional constraints. The Lagrangian becomes

$$-\sum_x p(x)\ln(p(x)) - \lambda_1\{\sum_x p(x) - 1\} - \lambda_2\{\sum_x xp(x) - \mu\} - \lambda_3\{\sum x^2 p(x) - \mu^2 - \sigma^2\}$$

whereupon setting the derivative with respect to $p(x)$ equal to zero and applying the constraints we obtain

$$p(x) = \exp\{-\lambda_1 - \lambda_2 x - \lambda_3 x^2\},$$

with constants $\lambda_1, \lambda_2, \lambda_3$ chosen to satisfy the three constraints. Since the exponent is a quadratic function of x, this is analogous to the normal distribution except that we have required that it be supported on a discrete set of points x. With more points, positioned more closely together, the distribution becomes closer to the normal. Let us call such a distribution the discrete normal distribution. For a simple example, suppose that we wish to use the maximum entropy principle to approximate the distribution of the sum of the values on

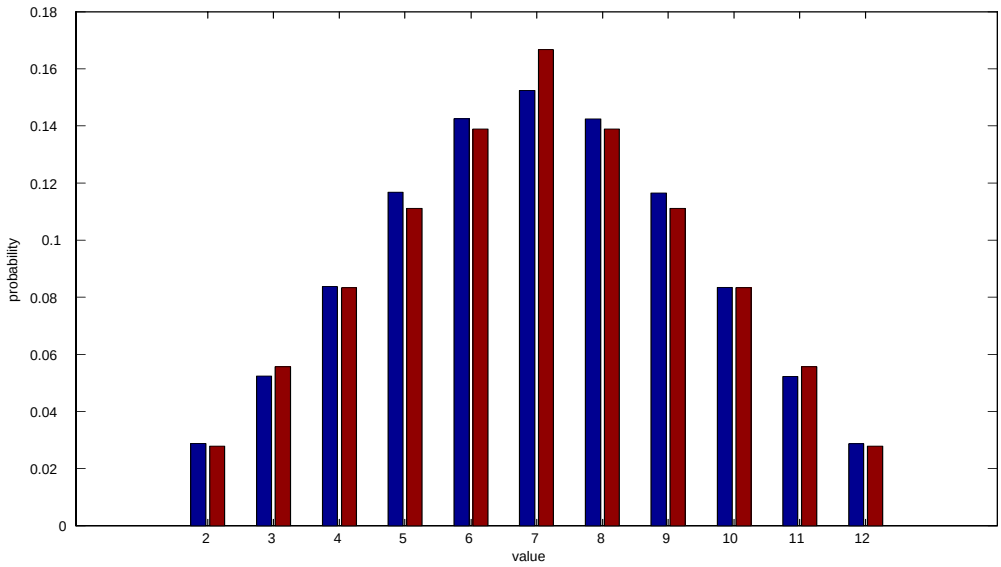

Figure 2.5: A discrete analogue of the normal distribution compared with the distribution of the sum of the values on two dice.

two dice. In this case the actual distribution is known to us as well as the mean and variance $E(X) = 7$, $var(X) = 35/6$;

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The maximum entropy distribution on these same points constrained to have the same mean and variance is very similar to this, the actual distribution. This can been seen in Figure 2.5.

[FIGURE 2.5 ABOUT HERE]

In fact if we drop the requirement that the distribution is discrete, or equivalently take a limit with an increasing number of discrete points closer and closer together, the same kind of argument shows that the maximum entropy distribution subject to a constraint on the mean and the variance is the normal distribution. So at least two well-known distributions arise out of maximum

entropy considerations. *The maximum entropy distribution on a discrete set of points is the uniform distribution. The maximum entropy subject to a constraint on the mean and the variance is a (discrete) normal distribution.* There are many other examples as well. In fact most common distributions in statistics have an interpretation as a maximum entropy distribution subject to some constraints.

Entropy has a number of properties that one would expect of a measure of the information content in a random variable. It is non-negative, and can in usual circumstances be infinite. We expect that the information in a function of X , say $g(X)$, is less than or equal to the information in X itself, equal if the function is one to one (which means in effect we can determine X from the value of $g(X)$). Entropy is a property of a distribution, not of a random variable. Nevertheless it is useful to be able to abuse the notation used earlier by referring to $H(X)$ as the entropy of the distribution of X. Then we have the following properties

Proposition 7 $H(X) \geq 0$

Proposition 8 $H(g(X)) \leq H(X)$ *for any function* $g(x)$..

The information or uncertainty in two random variables is clearly greater than that in one. The definition of entropy is defined in the same fashion as before, for discrete random variables (X, Y),

$$H(X, Y) = -E(\ln p(X, Y))$$

where $p(x, y)$ is the joint probability function

$$p(x, y) = P[X = x, Y = y].$$

If the two random variables are independent, then we expect that the uncertainty should add. If they are dependent, then the entropy of the pair (X, Y) is less than the sum of the individual entropies.

Proposition 9 $H(X,Y) \leq H(X) + H(Y)$ *with equality if and only if* X *and* Y *are independent.*

Let us now use the principle of maximum entropy to address an eminently practical problem, one of altering a distribution to accommodate a known mean value. Suppose we are interested in determining a risk-neutral distribution for pricing options at maturity T. Theorem 1 tells us that if there is to be no arbitrage, our distribution or measure Q must satisfy a relation of the form

$$E_Q(e^{-rT} S_T) = S_0$$

where r is the continuously compounded interest rate, S_0 is the initial (present) value of the underlying stock, and S_T is its value at maturity. Let us also suppose that we constraint the variance of the future stock price under the measure Q so that

$$var_Q(S_T) = \sigma^2 T.$$

Then from our earlier discussion, the maximum entropy distribution under constraints on the mean and variance is the normal distribution so that the probability density function of S_T is

$$f(s) = \frac{1}{\sigma\sqrt{2\pi T}} \exp\{-\frac{(s - e^{rT} S_0)^2}{2\sigma^2 T}\}.$$

If we wished a maximum entropy distribution which is compatible with a number of option prices, then we should impose these option prices as additional constraints. Again suppose the current time $t = 0$ and we know the prices $P_i, i = 1, ..., n$ of n different call options available on the market, all on the same security and with the same maturity T but with different strike prices K_i. The distribution Q we assign to S_T must satisfy the constraints

$$E(e^{-rT}(S_T - K_i)^+) = P_i, i = 1, ..., n \tag{2.20}$$

as well as the martingale constraint

$$E(e^{-rT} S_T) = S_0. \tag{2.21}$$

Once again introducing Lagrange multipliers, the probability density function of S_T will take the form

$$f(s) = k \exp\{e^{-rT} \sum_{i=1}^{n} \lambda_i (s - K_i)^+ + \lambda_0 s\}$$

where the parameters $\lambda_0, ..., \lambda_n$ are chosen to satisfy the constraints (2.20) and (2.21) and k so that the function integrates to 1. When fit to real option price data, these distributions typically resemble a normal density, usually however with some negative skewness and excess kurtosis. See for example Figure XXX. There are also"sawtooth" like appendages with teeth corresponding to each of the n options. Note too this density is strictly positive at the value $s = 0$, a feature that we may or may not wish to have. Because of the "teeth", a smoother version of the density is often used, one which may not perfectly reproduce option prices but is nevertheless appears to be more natural.

Minimum Cross-Entropy

Normally market information does not completely determine the risk-neutral measure Q . We will argue that while market data on derivative prices rather than historical data should determine the Q measure, historical asset prices can be used to fill in the information that is not dictated by no-arbitrage considerations. In order to relate the real world to the risk-free world, we need either sufficient market data to completely describe a risk-neutral measure Q (such a model is called a *complete market)* or we need to limit our candidate class of Q measures somewhat. We may either define the joint distributions of the stock prices or their returns, since from one we can pass to the other. For convenience, suppose we describe the joint distribution of the returns process. The conditions we impose on the martingale measure are the following;

1. Under Q, each normalized stock price $S_j(t)/B_t$ and derivative price V_t/B_t forms a martingale. Equivalently, $E_Q[S_i(t+1)|H_t] = S_i(t)(1+r(t))$

where $r(t)$ is the risk free interest rate over the interval $(t, t+1)$. (Recall that this risk-free interest rate $r(t)$ is defined by the equation $B(t+1) = (1+r(t))B(t)$.)

2. Q is a probability measure.

A slight revision of notation is necessary here. We will build our joint distributions conditionally on the past and if P denotes the joint distribution stock prices $S(1), S(2), ...S(T)$ over the whole period of observation $0 < t < T$ then P_{t+1} denotes the conditional distribution of $S(t+1)$ given H_t. Let us denote the conditional moment generating function of the vector $S(t+1)$ under the measure P_{t+1} by

$$m_t(u) = E_P[\exp(u'S(t+1)|H_t] = E_P[\exp(\sum_i u_i S_i(t+1))|H_t]$$

We implicitly assume, of course, that this moment generating function exists. Suppose, for some vector of parameters η we choose Q_{t+1} to be the exponential tilt of P_{t+1}, i.e.

$$dQ_{t+1}(s) = \frac{\exp(\eta' s)}{m_t(\eta)} dP_{t+1}(s)$$

The division by $m_t(\eta)$ is necessary to ensure that Q_{t+1} is a probability measure.

Why transform a density by multiplying by an exponential in this way? There are many reasons for such a transformation. Exponential families of distributions are built in exactly this fashion and enjoy properties of sufficiency, completeness and ease of estimation. This exponential tilt resulted from maximizing entropy subject to certain constraints on the distribution. But we also argue that the measure Q is the probability measure which is closest to P in a certain sense while still satisfying the required moment constraint. We first introduce cross-entropy which underlies considerable theory in Statistics and elsewhere in Science.

Cross Entropy

Consider two probability measures P and Q on the same space. Then the cross entropy or Kullbach-Leibler "distance" between the two measures is given by

$$H(Q,P) = \sup_{\{E_i\}} \sum Q(E_i) \log \frac{Q(E_i)}{P(E_i)}$$

where the supremum is over all finite partitions $\{E_i\}$ of the probability space. Various properties are immediate.

Proposition 10 $H(Q,P) \geq 0$ *with equality if and only if* P *and* Q *are identical.*

If Q is absolutely continuous with respect to P, that is if there is some density function $f(x)$ such that

$$Q(E) = \int_E f(x)dP \quad \text{for all } E$$

then provided that f is smooth, we can also write

$$H(Q,P) = E_Q \log(\frac{dQ}{dP}).$$

If Q is not absolutely continuous with respect to P then the cross entropy $H(Q,P)$ is infinite. We should also remark that the cross entropy is not really a distance in the usual sense (although we used the term "distance" in reference to it) because in general $H(Q,P) \neq H(P|Q)$. For a finite probability space, there is an easy relationship between entropy and cross entropy given by the following proposition. In effect the result tells us that maximizing entropy $H(Q)$ is equivalent to minimizing the cross-entropy $H(Q,P)$ where P is the uniform distribution.

Proposition 11 *If the probability space has a finite number* n *points, and* P *denotes the uniform distribution on these* n *points, then for any other probability measure* Q,

$$H(Q,P) = n - H(Q)$$

Now the following result asserts that the probability measure Q which is closest to P in the sense of cross-entropy but satisfies a constraint on its mean is generated by a so-called "exponential tilt" of the distribution of P.

Theorem 12 *: Minimizing cross-entropy.*

Let $f(X)$ be a vector valued function $f(X) = (f_1(X), f_2(X), ..., f_n(X))$ and $\mu = (\mu_1, ..., \mu_n)$. *Consider the problem*

$$\min_Q H(Q, P)$$

subject to the constraint $E_Q(f_i(X)) = \mu_i, i = 1, ..., n.$ *Then the solution, if it exists, is given by*

$$dQ = \frac{\exp(\eta' f(X))}{m(\eta)} dP = \frac{\exp(\sum_{i=1}^n \eta_i f_i(X))}{m(\eta)}$$

where $m(\eta) = E_P[\exp(\sum_{i=1}^n \eta_i f_i(X))]$ *and* η *is chosen so that* $\frac{\partial m}{\partial \eta_i} = \mu m(\eta)$.

The proof of this result, in the case of a discrete distribution P is a straightforward use of Lagrange multipliers (see Lemma 3). We leave it as a problem at the end of the chapter.

Now let us return to the constraints on the vector of stock prices. In order that the discounted stock price forms a martingale under the Q measure, we require that $E_Q[S(t+1)|H_t] = (1 + r(t))S(t)$. This is achieved if we define Q such that for any event $A \in H_t$,

$$Q(A) = \int_A Z_t dP \quad \text{where}$$

$$Z_s = k_t \exp(\sum_{t=1}^{s} \eta_t'(S_{t+1} - S_t)) \tag{2.22}$$

where k_t are H_t measurable random variables chosen so that Z_t forms a martingale

$$E(Z_{t+1}|H_t) = Z_t.$$

Theorem 9 shows that this exponentially tilted distribution has the property of being the closest to the original measure P while satisfying the condition that the normalized sequence of stock prices forms a martingale.

There is a considerable literature exploring the links between entropy and risk-neutral valuation of derivatives. See for example Gerber and Shiu (1994), Avellaneda et. al (1997), Gulko(1998), Samperi (1998). In a complete or incomplete market, risk-neutral valuation may be carried out using a martingale measure which maximizes entropy or minimizes cross-entropy subject to some natural constraints including the martingale constraint. For example it is easy to show that when interest rates r are constant, Q is the risk-neutral measure for pricing derivatives on a stock with stock price process $S_t, t = 0, 1, ...$ if and only if it is the probability measure minimizing $H(Q, P)$ subject to the martingale constraint

$$S_t = E_Q[\frac{1}{1+r} S_{t+1}]. \tag{2.23}$$

There is a continuous time analogue of (2.22) as well which we can anticipate by inspecting the form of the solution. Suppose that S_t denotes the stock price at time t where we now allow t to vary continuously in time. which we will discuss later but (2.22) can be used to anticipate it. Then an analogue of (2.22) could be written formally as

$$Z_s = \exp(\int_0^t \eta_t' dS_t - g_t)$$

where both processes η_t and g_t are "predictable" which loosely means that they are determined in advance of observing the increment $S_t, S_{t+\Delta t}$. Then the process Z_s is the analogue of the Radon-Nikodym derivative $\frac{dQ}{dP}$ of the processes restricted to the time interval $0 \leq t \leq s$. For a more formal definition, as well as an explanation of how we should interpret the integral, see the appendix. This process Z_s is, both in discrete and continuous time, a *martingale*.

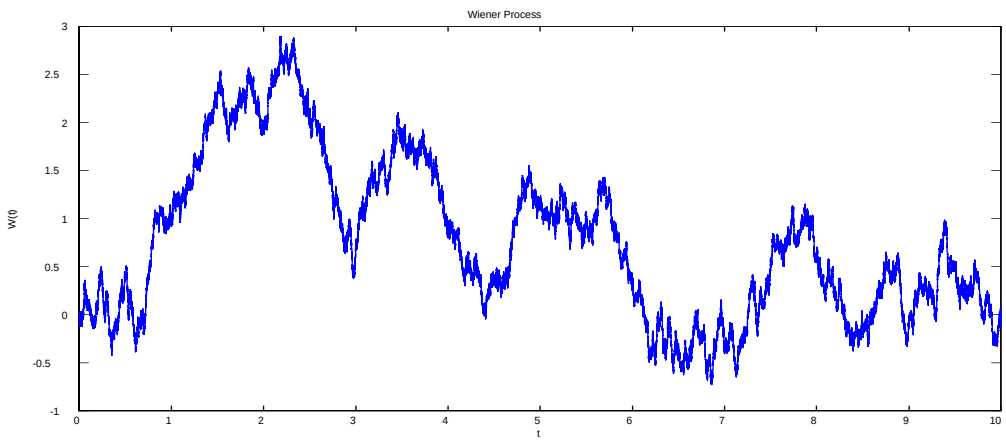

Figure 2.6: A sample path of the Wiener process

Models in Continuous Time

We begin with some oversimplified rules of stochastic calculus which can be omitted by those with a background in Brownian motion and diffusion. First, we define a stochastic process W_t called the *standard Brownian motion* or *Wiener process* having the following properties;

1. *For each* $h > 0$, *the increment* $W(t+h)-W(t)$ *has a* $N(0,h)$ *distribution and is independent of all preceding increments* $W(u)-W(v), t > u > v > 0$.

2. $W(0) = 0$.

[FIGURE 2.6 ABOUT HERE]

The fact that such a process exists is by no means easy to see. It has been an important part of the literature in Physics, Probability and Finance at least since the papers of Bachelier and Einstein, about 100 years ago. A Brownian motion process also has some interesting and remarkable theoretical properties; it is continuous with probability one but the probability that the process has finite

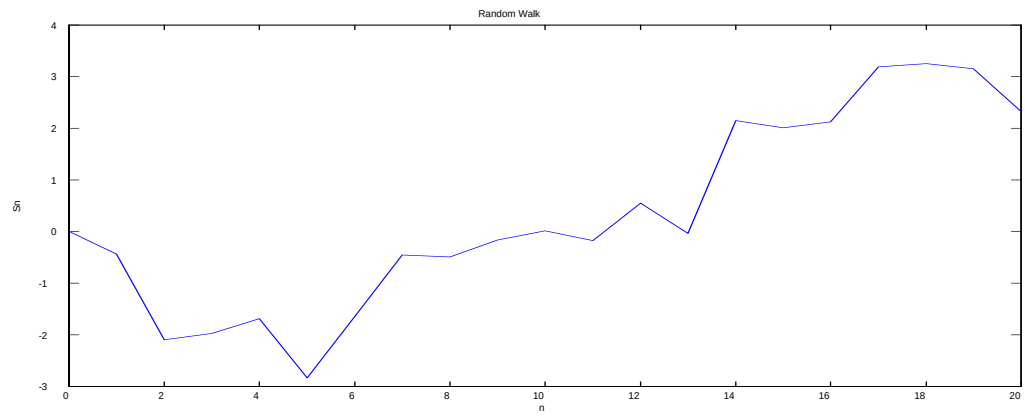

Figure 2.7: A sample path of a Random Walk

variation in any interval is 0. With probability one it is *nowhere differentiable*. Of course one might ask how a process with such apparently bizarre properties can be used to approximate real-world phenomena, where we expect functions to be built either from continuous and differentiable segments or jumps in the process. The answer is that a very wide class of functions constructed from those that are quite well-behaved (e.g. step functions) and that have independent increments converge as the scale on which they move is refined either to a Brownian motion process or to a process defined as an integral with respect to a Brownian motion process and so this is a useful approximation to a broad range of continuous time processes. For example, consider a random walk process $S_n = \sum_{i=1}^{n} X_i$ where the random variables X_i are independent identically distributed with expected value $E(X_i) = 0$ and $var(X_i) = 1$. Suppose we plot the graph of this random walk (n, S_n) as below. Notice that we have linearly interpolated the graph so that the function is defined for all n, whether integer or not.

[FIGURE 2.7 ABOUT HERE]

Now if we increase the sample size and decrease the scale appropriately on both axes, the result is, in the limit, a Brownian motion process. The vertical scale is to be decreased by a factor $1/\sqrt{n}$ and the horizontal scale by a factor n^{-1} . The theorem concludes that the sequence of processes

$$Y_n(t) = \frac{1}{\sqrt{n}} S_{nt}$$

converges weakly to a standard Brownian motion process as $n \to \infty$. In practice this means that a process with independent stationary increments tends to look like a Brownian motion process. As we shall see, there is also a wide variety of non-stationary processes that can be constructed from the Brownian motion process by integration. Let us use the above limiting result to render some of the properties of the Brownian motion more plausible, since a serious proof is beyond our scope. Consider the question of continuity, for example. Since $|Y_n(t+h) - Y_n(t)| \approx |\frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} X_i \mid$ and this is the absolute value of an asymptotically normally$(0, h)$ random variable by the central limit theorem, it is plausible that the limit as $h \to 0$ is zero so the function is continuous at t. On the other hand note that

$$\frac{Y_n(t+h) - Y_n(t)}{h} \approx \frac{1}{h} \frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} X_i$$

should by analogy behave like h^{-1} times a $N(0, h)$ random variable which blows up as $h \to 0$ so it would appear that the derivative at t does not exist. To obtain the total variation of the process in the interval $[t, t+h]$, consider the lengths of the segments in this interval, i.e.

$$\frac{1}{\sqrt{n}} \sum_{i=nt}^{n(t+h)} |X_i|$$

and notice that since the law of large numbers implies that $\frac{1}{nh} \sum_{i=nt}^{n(t+h)} |X_i|$ converges to a positive constant, namely $E|X_i|$, if we multiply by $\sqrt{n}h$ the limit must be infinite, so the total variation of the Brownian motion process is infinite.

Continuous time process are usually built one small increment at a time and defined to be the limit as the size of the time increment is reduced to zero. Let us consider for example how we might define a stochastic (*Ito*) integral of the form $\int_0^T h(t)dW_t$. An approximating sum takes the form

$$\int_0^T h(t)dW_t \approx \sum_{i=0}^{n-1} h(t_i)(W(t_{i+1}) - W(t_i)), 0 = t_0 < t_1 < ... < t_n = T.$$

Note that the function $h(t)$ is evaluated at the left hand end-point of the intervals $[t_i, t_{i+1}]$, and this is characteristic of the Ito calculus, and an important feature distinguishing it from the usual Riemann calculus studied in undergraduate mathematics courses. There are some simple reasons why evaluating the function at the left hand end-point is necessary for stochastic models in finance. For example let us suppose that the function $h(t)$ measures how many shares of a stock we possess and $W(t)$ is the price of one share of stock at time t. It is clear that we cannot predict precisely future stock prices and our decision about investment over a possibly short time interval $[t_i, t_{i+1}]$ must be made at the *beginning* of this interval, not at the end or in the middle. Second, in the case of a Brownian motion process $W(t)$, it *makes a difference* where in the interval $[t_i, t_{i+1}]$ we evaluate the function h to approximate the integral, whereas it makes no difference for Riemann integrals. As we refine the partition of the interval, the approximating sums $\sum_{i=0}^{n-1} h(t_{i+1})(W(t_{i+1}) - W(t_i))$, for example, approach a completely different limit. This difference is essentially due to the fact that $W(t)$, unlike those functions studied before in calculus, is of infinite variation. As a consequence, there are other important differences in the Ito calculus. Let us suppose that the increment dW is used to denote small increments $W(t_{i+1}) - W(t_i)$ involved in the construction of the integral. If we denote the interval of time $t_{i+1} - t_i$ by dt, we can loosely assert that dW has the normal distribution with mean 0 and variance dt. If we add up a large number of independent such increments, since the variances add, the sum has variance the sum of the values dt and standard deviation the square root. Very

roughly, we can assess the size of dW since its standard deviation is $(dt)^{1/2}$. Now consider defining a process as a function both of the Brownian motion and of time, say $V_t = g(W_t, t)$. If W_t represented the price of a stock or a bond, V_t might be the price of a derivative on this stock or bond. Expanding the increment dV using a Taylor series expansion gives

$$\begin{aligned} dV_t = & \frac{\partial}{\partial W} g(W_t, t) dW + \frac{\partial^2}{\partial W^2} g(W_t, t) \frac{dW^2}{2} + \frac{\partial}{\partial t} g(W_t, t) dt \\ & + (\text{stuff}) \times (dW)^3 + (\text{more stuff}) \times (dt)(dW)^2 + ... \end{aligned} \tag{2.24}$$

Loosely, dW is normal with mean 0 and standard deviation $(dt)^{1/2}$ and so dW is non-negligible compared with dt as $dt \to 0$. We can define each of the differentials dW and dt essentially by reference to the result when we integrate both sides of the equation. If I were to write an equation in differential form

$$dX_t = h(t) dW_t$$

then this only has real meaning through its integrated version

$$X_t = X_0 + \int_0^t h(t) dW_t.$$

What about the terms involving $(dW)^2$? What meaning should we assign to a term like $\int h(t)(dW)^2$? Consider the approximating function $\sum h(t_i)(W(t_{i+1}) - W(t_i))^2$. Notice that, at least in the case that the function h is non-random we are adding up independent random variables $h(t_i)(W(t_{i+1}) - W(t_i))^2$ each with expected value $h(t_i)(t_{i+1} - t_i)$ and when we add up these quantities the limit is $\int h(t) dt$ by the law of large numbers. Roughly speaking, as differentials, we should interpret $(dW)^2$ as dt because that is the way it acts in an integral. Subsequent terms such as $(dW)^3$ or $(dt)(dW)^2$ are all $o(dt)$, i.e. they all approach 0 faster than does dt as $dt \to 0$. So finally substituting for $(dW)^2$ in 2.24 and ignoring all terms that are $o(dt)$, we obtain a simple version of *Ito's lemma*

$$dg(W_t,t) = \frac{\partial}{\partial W}g(W_t,t)dW + \{\frac{1}{2}\frac{\partial^2}{\partial W^2}g(W_t,t) + \frac{\partial}{\partial t}g(W_t,t)\}dt.$$

This rule results, for example, when we put $g(W_t,t) = W_t^2$ in

$$d(W_t^2) = 2W_t dW_t + dt$$

or on integrating both sides and rearranging,

$$\int_a^b W_t dW_t = \frac{1}{2}(W_b^2 - W_a^2) - \frac{1}{2}\int_a^b dt.$$

The term $\int_a^b dt$ above is what distinguishes the Ito calculus from the Riemann calculus, and is a consequence of the nature of the Brownian motion process, a continuous function of infinite variation.

There is one more property of the stochastic integral that makes it a valuable tool in the construction of models in finance, and that is that a stochastic integral with respect to a Brownian motion process is *always a martingale*. To see this, note that in an approximating sum

$$\int_0^T h(t)dW_t \approx \sum_{i=0}^{n-1} h(t_i)(W(t_{i+1}) - W(t_i))$$

each of the summands has conditional expectation 0 given the past, i.e.

$$E[h(t_i)(W(t_{i+1}) - W(t_i))|H_{t_i}] = h(t_i)E[(W(t_{i+1}) - W(t_i))|H_{t_i}] = 0$$

since the Brownian increments have mean 0 given the past and since $h(t)$ is measurable with respect to H_t.

We begin with an attempt to construct the model for an Ito process or diffusion process in continuous time. We construct the price process one increment at a time and it seems reasonable to expect that both the mean and the variance of the increment in price may depend on the current price but does not depend on the process before it arrived at that price. This is a loose description of a Markov property. The conditional distribution of the future of the process

depends only on the current time t and the current price of the process. Let us suppose in addition that the increments in the process are, conditional on the past, normally distributed. Thus we assume that for small values of h, conditional on the current time t and the current value of the process X_t, the increment $X_{t+h} - X_t$ can be generated from a normal distribution with mean $a(X_t, t)h$ and with variance $\sigma^2(X_t, t)h$ for some functions a and σ^2 called the drift and diffusion coefficients respectively. Such a normal random variable can be formally written as $a(X_t, t\ )dt + \sigma^2(X_t, t)dW_t$. Since we could express X_T as an initial price X_0 plus the sum of such increments, $X_T = X_0 + \sum_i (X_{t_{i+1}} - X_{t_i})$.

The single most important model of this type is called the *Geometric Brownian motion or Black-Scholes model.* Since the actual value of stock, like the value of a currency or virtually any other asset is largely artificial, depending on such things as the number of shares issued, it is reasonable to suppose that the *changes in a stock price* should be modeled relative to the current price. For example rather than model the increments, it is perhaps more reasonable to model the relative change in the process. The simplest such model of this type is one in which both the mean and the standard deviation of the increment in the price are linear multiples of price itself; viz. dX_t is approximately normally distributed with mean $aX_t dt$ and variance $\sigma^2 X_t^2 dt$. In terms of stochastic differentials, we assume that

$$dX_t = aX_t dt + \sigma X_t dW_t. \tag{2.25}$$

Now consider the relative return from such a process over the increment $dY_t = dX_t/X_t$. Putting $Y_t = g(X_t) = \ln(X_t)$ note that analogous to our derivation of Ito's lemma

$$\begin{aligned} dg(X_t) &= g'(X_t)dX_t + \frac{1}{2}g''(X_t)(dX)^2 + ... \\ &= \frac{1}{X_t}\{aX_t dt + \sigma X_t dW_t.\} - \frac{1}{2X_t^2}\sigma^2 X_t^2 dt \\ &= (a - \frac{\sigma^2}{2})dt + \sigma dW_t. \end{aligned}$$

which is a description of a general Brownian motion process, a process with increments dY_t that are normally distributed with mean $(a - \frac{\sigma^2}{2})dt$ and with variance $\sigma^2 dt$. This process satisfying $dX_t = aX_t dt + \sigma X_t dW_t$ is called the *Geometric Brownian motion* process (because it can be written in the form $X_t = e^{Y_t}$ for a Brownian motion process Y_t) or a Black-Scholes model.

Many of the continuous time models used in finance are described as Markov diffusions or Ito processes which permits the mean and the variance of the increments to depend more generally on the present value of the process and the time. The integral version of this relation is of the form

$$X_T = X_0 + \int_0^T a(X_t, t)dt + \int_0^T \sigma(X_t, t)dW_t.$$

We often write such an equation with differential notation,

$$dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t. \tag{2.26}$$

but its meaning should always be sought in the above integral form. The coefficients $a(X_t, t)$ and $\sigma(X_t, t)$ vary with the choice of model. As usual, we interpret 2.26 as meaning that a small increment in the process, say $dX_t = X_{t+h} - X_t$ (h very small) is approximately distributed according to a normal distribution with conditional mean $a(X_t, t)dt$ and conditional variance given by $\sigma^2(X_t, t)var(dW_t) = \sigma^2(X_t, t)dt$. Here the mean and variance are conditional on H_t, the history of the process X_t up to time t.

Various choices for the functions $a(X_t, t), \sigma(X_t, t)$ are possible. For the Black-Scholes model or geometric Brownian motion, $a(X_t, t) = aX_t$ and $\sigma(X_t, t) = \sigma X_t$ for constant drift and volatility parameters a, σ. The *Cox-Ingersoll-Ross model,* used to model spot interest rates, corresponds to $a(X_t, t) = A(b - X_t)$ and $\sigma(X_t, t) = c\sqrt{X_t}$ for constants A, b, c. The Vasicek model, also a model for interest rates, has $a(X_t, t) = A(b - X_t)$ and $\sigma(X_t, t) = c$. There is a large number of models for most continuous time processes observed in finance which can be written in the form 2.26. So called multi-factor models are of similar form

where X_t is a vector of financial time series and the coefficient functions $a(X_t, t)$ is vector valued, $\sigma(X_t, t)$ is replaced by a matrix-valued function and dW_t is interpreted as a vector of independent Brownian motion processes. For technical conditions on the coefficients under which a solution to 2.26 is guaranteed to exist and be unique, see Karatzas and Shreve, sections 5.2, 5.3.

As with any differential equation there may be initial or boundary conditions applied to 2.26 that restrict the choice of possible solutions. Solutions to the above equation are difficult to arrive at, and it is often even more difficult to obtain distributional properties of them. Among the key tools are the *Kolmogorov differential equations* (see Cox and Miller, p. 215). Consider the transition probability kernel

$$p(s, z, t, x) = P[X_t = x | X_s = z]$$

in the case of a *discrete Markov Chain.* If the Markov chain is continuous (as it is in the case of diffusions), that is if the conditional distribution of X_t given X_s is absolutely continuous with respect to Lebesgue measure, then we can define $p(s, z, t, x)$ to be the *conditional probability density function* of X_t given $X_s = z$. The two equations, for a diffusion of the above form, are:

Kolmogorov's backward equation

$$\frac{\partial}{\partial s} p = -a(z, s) \frac{\partial}{\partial z} p - \frac{1}{2} \sigma^2(z, s) \frac{\partial^2}{\partial z^2} p \tag{2.27}$$

and the *forward equation*

$$\frac{\partial}{\partial t} p = -\frac{\partial}{\partial x}(a(x, t)p) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(\sigma^2(x, t)p) \tag{2.28}$$

Note that if we were able to solve these equations, this would provide the transition density function p, giving the conditional distribution of the process. It does not immediately provide other characteristics of the diffusion, such as the distribution of the maximum or the minimum, important for valuing various exotic options such as look-back and barrier options. However for a European

option defined on this process, knowledge of the transition density would suffice at least theoretically for valuing the option. Unfortunately these equations are often very difficult to solve explicitly.

Besides the Kolmogorov equations, we can use simple ordinary differential equations to arrive at some of the basic properties of a diffusion. To illustrate, consider one of the simplest possible forms of a diffusion, where $a(X_t, t) = \alpha(t)+\beta(t)X_t$ where the coefficients $\alpha(t), \beta(t)$ are deterministic (i.e. non-random) functions of time. Note that the integral analogue of 2.26 is

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s \tag{2.29}$$

and by construction that last term $\int_0^t \sigma(X_s, s)dW_s$ is a zero-mean martingale. For example its small increments $\sigma(X_t, t)dW_s$ are approximately $N(0, \sigma(X_t, t)dt)$. Therefore, taking expectations on both sides conditional on the value of X_0, and letting $m(t) = E(X_t)$, we obtain:

$$m(t) = X_0 + \int_0^t [\alpha(s) + \beta(s)m(s)]ds \tag{2.30}$$

and therefore $m(t)$solves the ordinary differential equation

$$m'(t) = \alpha(t) + \beta(t)m(t). \tag{2.31}$$

$$m(0) = X_0 \tag{2.32}$$

Thus, in the case that the *drift term* a is a linear function of X_t, the mean or expected value of a diffusion process can be found by solving a similar ordinary differential equation, similar except that the diffusion term has been dropped.

These are only two of many reasons to wish to solve both ordinary and partial differential equations in finance. The solution to the Kolmogorov partial differential equations provides the conditional distribution of the increments of a process. And when the drift term $a(X_t, t\ )$ is linear in X_t, the solution of an ordinary differential equation will allow the calculation of the expected value of the process and this is the first and most basic description of its behaviour. The

appendix provides an elementary review of techniques for solving partial and ordinary differential equations.

However, that the information about a stochastic process obtained from a deterministic object such as a ordinary or partial differential equation is necessarily limited. For example, while we can sometimes obtain the marginal distribution of the process at time t it is more difficult to obtain quantities such as the joint distribution of variables which depending on the path of the process, and these are important in valuing certain types of exotic options such as lookback and barrier options. For such problems, we often use Monte Carlo methods.

The Black-Scholes Formula

Before discussing methods of solution in general, we develop the Black-Scholes equation in a general context. Suppose that a security price is an Ito process satisfying the equation

$$dS_t = a(S_t,\ t\ )\ dt\ +\ \sigma(S_t,\ t)\ dW_t \tag{2.33}$$

Assumed the market allows investment in the stock as well as a risk-free bond whose price at time t is B_t. It is necessary to make various other assumptions as well and strictly speaking all fail in the real world, but they are a reasonable approximation to a real, highly liquid and nearly frictionless market:

1. partial shares may be purchased
2. there are no dividends paid on the stock
3. There are no commissions paid on purchase or sale of the stock or bond
4. There is no possibility of default for the bond
5. Investors can borrow at the risk free rate governing the bond.
6. All investments are liquid- they can be bought or sold instantaneously.

Since bonds are assumed risk-free, they satisfy an equation

$$dB_t = r_t B_t dt$$

where r_t is the risk-free (spot) interest rate at time t.

We wish to determine $V(S_t, t)$, the value of an option on this security when the security price is S_t, at time t. Suppose the option has expiry date T and a general payoff function which depends only on S_T, the process at time T.

Ito's lemma provides the ability to translate an a relation governing the differential dS_t into a relation governing the differential of the process $dV(S_t, t)$. In this sense it is the stochastic calculus analogue of the chain rule in ordinary calculus. It is one of the most important single results of the twentieth century in finance and in science. The stochastic calculus and this mathematical result concerning it underlies the research leading to 1997 Nobel Prize to Merton and Black for their work on hedging in financial models. We saw one version of it at the beginning of this section and here we provide a more general version.

Ito's lemma.

Suppose S_t is a diffusion process satisfying

$$dS_t = a(S_t, t)dt + \sigma(S_t, t)dW_t$$

and suppose $V(S_t, t)$ is a smooth function of both arguments. Then $V(S_t, t)$ also satisfies a diffusion equation of the form

$$dV = [a(S_t, t)\frac{\partial V}{\partial S} + \frac{\sigma^2(S_t, t)}{2}\frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}]dt + \sigma(S_t, t)\frac{\partial V}{\partial S}dW_t. \tag{2.34}$$

Proof. The proof of this result is technical but the ideas behind it are simple. Suppose we expand an increment of the process $V(S_t, t)$ (we write V

in place of $V(S_t, t)$ omitting the arguments of the function and its derivatives. We will sometimes do the same with the coefficients a and σ.)

$$V(S_{t+h}, t+h) \approx V + \frac{\partial V}{\partial S}(S_{t+h} - S_t) + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(S_{t+h} - S_t)^2 + \frac{\partial V}{\partial t}h \qquad (2.35)$$

where we have ignored remainder terms that are $o(h)$. Note that substituting from 2.33 into 2.35, the increment $(S_{t+h} - S_t)$ is approximately normal with mean $a(S_t,\ t\ )\ h$ and variance $\sigma^2(S_t,\ t\ )\ h$. Consider the term $(S_{t+h} - S_t)^2$. Note that it is the square of the above normal random variable and has expected value $\sigma^2(S_t, t)h + a^2(S_t, t)h^2$. The variance of this random variable is $O(h^2)$ so if we ignore all terms of order $o(h)$ the increment $V(S_{t+h}, t+h) - V(S_t, t)$ is approximately normally distributed with mean

$$[a(S_t,\ t\ )\frac{\partial V}{\partial S} + \frac{\sigma^2(S_t,\ t)}{2}\ \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}]h$$

and standard deviation $\sigma(S_t,\ t)\frac{\partial V}{\partial S}\sqrt{h}$ justifying (but not proving!) the relation 2.34. ■

By Ito's lemma, provided $V$is smooth, it also satisfies a diffusion equation of the form 2.34. We should note that when V represents the price of an option, some lack of smoothness in the function V is inevitable. For example for a European call option with exercise price K, $V(S_T, T) = \max(S_T - K, 0)$ does not have a derivative with respect to S_T at $S_T = K$, the exercise price. Fortunately, such exceptional points can be worked around in the argument, since the derivative does exist at values of $t < T$.

The basic question in building a replicating portfolio is: for hedging purposes, is it possible to find a *self-financing* portfolio consisting only of the security and the bond which exactly replicates the option price process $V(S_t, t)$? The self-financing requirement is the analogue of the requirement that the net cost of a portfolio is zero that we employed when we introduced the notion of

arbitrage. The portfolio is such that no funds are needed to be added to (or removed from) the portfolio during its life, so for example any additional amounts required to purchase equity is obtained by borrowing at the risk free rate. Suppose the self-financing portfolio has value at time t equal to $V_t = u_t S_t + w_t B_t$ where the (predictable) functions u_t, w_t represent the number of shares of stock and bonds respectively owned at time t. Since the portfolio is assumed to be self-financing, all returns obtain from the changes in the value of the securities and bonds held, i.e. it is assumed that $dV_t = u_t dS_t + w_t dB_t$. Substituting from 2.33,

$$dV_t = u_t dS_t + w_t dB_t = [u_t a(S_t, t) + w_t r_t B_t]dt + u_t \sigma(S_t, t) dW_t \tag{2.36}$$

If V_t is to be exactly equal to the price $V(S_t, t\ )$ of an option, it follows on comparing the coefficients of dt and dW_t in 2.34 and 2.36, that $u_t = \frac{\partial V}{\partial S}$, called the *delta* corresponding to *delta hedging*. Consequently,

$$V_t = \frac{\partial V}{\partial S} S_t + w_t B_t$$

and solving for w_t we obtain:

$$w_t = \frac{1}{B_t}[V - \frac{\partial V}{\partial S} S_t].$$

The conclusion is that it is possible to dynamically choose a trading strategy, i.e. the weights w_t, u_t so that our portfolio of stocks and bonds perfectly replicates the value of the option. If we own the option, then by shorting (selling) delta= $\frac{\partial V}{\partial S}$ units of stock, we are **perfectly** hedged in the sense that our portfolio replicates a risk-free bond. Surprisingly, in this ideal word of continuous processes and continuous time trading commission-free trading, the perfect hedge is possible. In the real world, it is said to exist only in a Japanese garden. The equation we obtained by equating both coefficients in 2.34 and 2.36 is;

$$-r_t V + r_t S_t \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{\sigma^2(S_t, t)}{2} \frac{\partial^2 V}{\partial S^2} = 0. \tag{2.37}$$

Rewriting this allows an interpretation in terms of our hedged portfolio. If we own an option and are short delta units of stock our net investment at time t is given by $(V - S_t \frac{\partial V}{\partial S})$ where $V = V_t = V(S_t, t)$. Our return over the next time increment dt if the portfolio were liquidated and the identical amount invested in a risk-free bond would be $r_t(V_t - S_t \frac{\partial V}{\partial S})dt$. On the other hand if we keep this hedged portfolio, the return over an increment of time dt is

$$\begin{aligned} d(V - S_t \frac{\partial V}{\partial S}) &= dV - (\frac{\partial V}{\partial S})dS \\ &= (\frac{\partial V}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 V}{\partial S^2} + a\frac{\partial V}{\partial S})dt + \sigma \frac{\partial V}{\partial S} dW_t \\ &- \frac{\partial V}{\partial S}[adt + \sigma dW_t] \\ &= (\frac{\partial V}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 V}{\partial S^2})dt \end{aligned}$$

Therefore

$$r_t(V - S_t \frac{\partial V}{\partial S}) = \frac{\partial V}{\partial t} + \frac{\sigma^2(S_t, t)}{2}\frac{\partial^2 V}{\partial S^2}.$$

The left side $r_t(V - S_t \frac{\partial V}{\partial S})$ represents the amount made by the portion of our portfolio devoted to risk-free bonds. The right hand side represents the return on a hedged portfolio long one option and short delta stocks. Since these investments are at least in theory identical, so is their return. This fundamental equation is evidently satisfied by any option price process where the underlying security satisfies a diffusion equation and the option value at expiry depends only on the value of the security at that time. The type of option determines the terminal conditions and usually uniquely determines the solution.

It is extraordinary that this equation in no way depends on the drift coefficient $a(S_t, t)$. This is a remarkable feature of the arbitrage pricing theory. *Essentially, no matter what the drift term for the particular security is, in order to avoid arbitrage, all securities and their derivatives are priced as if they had as drift the spot interest rate. This is the effect of calculating the expected values under the martingale measure Q.*

This PDE governs most derivative products, European call options, puts,

futures or forwards. However, the boundary conditions and hence the solution depends on the particular derivative. The solution to such an equation is possible analytically in a few cases, while in many others, numerical techniques are necessary. One special case of this equation deserves particular attention. In the case of geometric Brownian motion, $a(S_t, t) = \mu S_t$ and $\sigma(S_t, t) = \sigma S_t$ for constants μ, σ. Assume that the spot interest rate is a constant $r$and that a constant rate of dividends D_0 is paid on the stock. In this case, the equation specializes to

$$-rV + \frac{\partial V}{\partial t} + (r - D_0)S\frac{\partial V}{\partial S} + \frac{\sigma^2 S^2}{2}\frac{\partial^2 V}{\partial S^2} = 0. \tag{2.38}$$

Note that we have not used *any* of the properties of the particular derivative product yet, nor does this differential equation involve the drift coefficient μ. The assumption that there are no transaction costs is essential to this analysis, as we have assumed that the portfolio is continually rebalanced.

We have now seen two derivations of parabolic partial differential equations, so-called because like the equation of a parabola, they are first order (derivatives) in one variable (t) and second order in the other (x). Usually the solution of such an equation requires reducing it to one of the most common partial differential equations, the heat or diffusion equation, which models the diffusion of heat along a rod. This equation takes the form

$$\frac{\partial}{\partial t}u = k\frac{\partial^2}{\partial x^2}u \tag{2.39}$$

A solution of 2.39 with appropriate boundary conditions can sometime be found by the separation of variables. We will later discuss in more detail the solution of parabolic equations, both by analytic and numerical means. First, however, when can we hope to find a solution of 2.39 of the form $u(x,t) = g(x/\sqrt{t})$. By differentiating and substituting above, we obtain an ordinary differential equation of the form

$$g''(\omega) + \frac{1}{2k}\omega g'(\omega) = 0, \omega = x/\sqrt{t} \tag{2.40}$$

Let us solve this using MAPLE.

```
eqn := diff(g(w),w,w)+(w/(2*k))*diff(g(w),w)=0;
dsolve(eqn,g(w));
```

and because the derivative of the solution is slightly easier (for a statistician) to identify than the solution itself,

```
> diff(%,w);
```

giving

$$\frac{\partial}{\partial w} g(\omega) = C_2 \exp\{-w^2/4k\} = C_2 \exp\{-x^2/4kt\} \tag{2.41}$$

showing that a constant plus a constant multiple of the Normal $(0, 2kt)$ cumulative distribution function or

$$u(x,t) = C_1 + C_2 \frac{1}{2\sqrt{\pi kt}} \int_{-\infty}^{x} \exp\{-z^2/4kt\} dz \tag{2.42}$$

is a solution of this, the heat equation for $t > 0$. The role of the two constants is simple. Clearly if a solution to 2.39 is found, then we may add a constant and/or multiply by a constant to obtain another solution. The constant in general is determined by initial and boundary conditions. Similarly the integral can be removed with a change in the initial condition for if u solves 2.39 then so does $\frac{\partial u}{\partial x}$. For example if we wish a solution for the half real $x > 0$ with initial condition $u(x,0) = 0, u(0,t) = 1$ all $t > 1$, we may use

$$u(x,t) = 2P(N(0,2kt) > x) = \frac{1}{\sqrt{\pi kt}} \int_{x}^{\infty} \exp\{-z^2/4kt\} dz, t > 0, x \geq 0.$$

Let us consider a basic solution to 2.39:

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \exp\{-x^2/4kt\} \tag{2.43}$$

This connection between the heat equation and the normal distributions is fundamental and the wealth of solutions depending on the initial and boundary conditions is considerable. We plot a fundamental solution of the equation as follows with the plot in Figure 2.8:

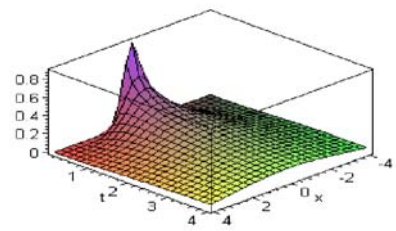

Figure 2.8: Fundamental solution of the heat equation

```
>u(x,t) := (.5/sqrt(Pi*t))*exp(-x^2/(4*t));

>plot3d(u(x,t),x=-4..4,t=.02..4,axes=boxed);
```

[FIGURE 2.8 ABOUT HERE]

As $t \to 0$, the function approaches a spike at $x = 0$, usually referred to as the "Dirac delta function" (although it is no function at all) and symbolically representing the derivative of the "Heaviside function". The Heaviside function is defined as $H(x) = 1, x \geq 0$ and is otherwise 0 and is the cumulative distribution function of a point mass at 0. Suppose we are given an initial condition of the form $u(x, 0) = u_0(x)$. To this end, it is helpful to look at the solution $u(x, t)$ and the initial condition $u_0(x)$ as a distribution or measure (in this case described by a density) over the space variable x. For example the density $u(x, t)$ corresponds to a measure for fixed t of the form $\nu_t(A) = \int_A u(x, t)dx$. Note that the initial condition compatible with the above solution 2.42 can be described somewhat clumsily as "$u(x, 0)$ corresponds to a measure placing all mass at $x = x_0 = 0$ ".In fact as $t \to 0$, we have in some sense the following convergence $u(x, t) \to \delta(x) = dH(x)$, the Dirac delta function. We could just as easily construct solve the heat equation with a more general initial condition of

the form $u(x,0) = dH(x - x_0)$ for arbitrary x_0 and the solution takes the form

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \exp\{-(x - x_0)^2/4kt\}. \tag{1.22}$$

Indeed sums of such solutions over different values of x_0, or weighted sums, or their limits, integrals will continue to be solutions to 2.39. In order to achieve the initial condition $u_0(x)$ we need only pick a suitable weight function. Note that

$$u_0(x) = \int u_0(z) dH(z - x)$$

Note that the function

$$u(x,t) = \frac{1}{2\sqrt{\pi kt}} \int_{-\infty}^{\infty} \exp\{-(z - x)^2/4kt\} u_0(z) dz \tag{1.22}$$

solves 2.39 subject to the required boundary condition.

Solution of the Diffusion Equation.

We now consider the general solution to the diffusion equation of the form 2.37, rewritten as

$$\frac{\partial V}{\partial t} = r_t V - r_t S_t \frac{\partial V}{\partial S} - \frac{\sigma^2(S_t,t)}{2} \frac{\partial^2 V}{\partial S^2} \tag{2.44}$$

where $S_t$is an asset price driven by a diffusion equation

$$dS_t = a(S_t,t)dt + \sigma(S_t,t)dW_t, \tag{2.45}$$

$V(S_t,t)$ is the price of an option on that asset at time t, and $r_t = r(t)$ is the spot interest rate at time t. We assume that the price of the option at expiry $T$is a known function of the asset price

$$V(S_T,T) = V_0(S_T). \tag{2.46}$$

Somewhat strangely, the option is priced using a related but not identical process (or, equivalently, the same process under a different measure). Recall from the

backwards Kolmogorov equation 2.27 that if a related process $X_t$satisfies the stochastic differential equation

$$dX_t = r(X_t, t)X_t dt + \sigma(X_t, t)dW_t \tag{2.47}$$

then its transition kernel $p(t, s, T, z) = \frac{\partial}{\partial z}P[X_T \leq z|X_t = s]$ satisfies a partial differential equation similar to 2.44;

$$\frac{\partial p}{\partial t} = -r(s,t)s\frac{\partial p}{\partial s} - \frac{\sigma^2(s,t)}{2}\frac{\partial^2 p}{\partial s^2} \tag{2.48}$$

For a given process $X_t$this determines one solution. For simplicity, consider the case (natural in finance applications) when the spot interest rate is a function of time, not of the asset price; $r(s,t) = r(t)$. To obtain the solution so that terminal conditions is satisfied, consider a product

$$f(t, s, T, z) = p(t, s, T, z)q(t, T) \tag{2.49}$$

where

$$q(t, T) = exp\{-\int_t^T r(v)dv\}$$

is the discount function or the price of a zero-coupon bond at time t which pays 1\$ at maturity.

Let us try an application of one of the most common methods in solving PDE's, the "lucky guess" method. Consider a linear combination of terms of the form 2.49 with weight function $w(z)$. i.e. try a solution of the form

$$V(s,t) = \int p(t, s, T, z)q(t, T)w(z)dz \tag{2.50}$$

for suitable weight function $w(z)$. In view of the definition of $p$as a transition probability density, this integral can be rewritten as a conditional expectation:

$$V(t, s) = E[w(X_T)q(t, T)|X_t = s] \tag{2.51}$$

the discounted conditional expectation of the random variable $w(X_T)$ given the current state of the process, where the process is assumed to follow (2.18). Note

that in order to satisfy the terminal condition 2.46, we choose $w(x) = V_0(x)$. Now

$$\begin{aligned}
\frac{\partial V}{\partial t} &= \frac{\partial}{\partial t}\int p(t,s,T,z)q(t,T)w(z)dz \\
&= \int [-r(S_t,t)S_t\frac{\partial p}{\partial s} - \frac{\sigma^2(S_t,t)}{2}\frac{\partial^2 p}{\partial s}2]q(t,T)w(z)dz \\
&+ r(S_t,t)\int p(t,S_t,T,z)q(t,T)w(z)dz \text{ by 2.48} \\
&= -r(S_t,t)S_t\frac{\partial V}{\partial S} - \frac{\sigma^2(S_t,t)}{2}\frac{\partial^2 V}{\partial S^2} + r(S_t,t)V(S_t,t)
\end{aligned}$$

where we have assumed that we can pass the derivatives under the integral sign. Thus the process

$$V(t,s) = E[V_0(X_T)q(t,T)|X_t = s] \tag{2.52}$$

satisfies both the partial differential equation 2.44 and the terminal conditions 2.46 and is hence the solution. Indeed it is the unique solution satisfying certain regularity conditions. The result asserts that the value of any European option is simply the conditional expected value of the *discounted payoff* (discounted to the present) assuming that the distribution is that of the process 2.47. This result is a special case when the spot interest rates are functions only of time of the following more general theorem.

Theorem 13 *(Feynman-Kac)*

Suppose the conditions for a unique solution to (2.44,2.46) (see for example Duffie, appendix E) are satisfied. Then the general solution to (2.15) under the terminal condition 2.46 is given by

$$V(S,t) = E[V_0(X_T)exp\{-\int_t^T r(X_v,v)dv\}|\ X_t = S] \tag{2.53}$$

This represents the discounted return from the option under the distribution of the process X_t. The distribution induced by the process X_t is referred to as the *equivalent martingale measure* or *risk neutral measure*. Notice that when the original process is a diffusion, the equivalent martingale measure shares the same diffusion coefficient but has the drift replaced by $r(X_t, t)X_t$. The option is priced as if the drift were the same as that of a risk-free bond i.e. as if the instantaneous rate of return from the security if identical to that of bond. Of course, in practice, it is not. A risk premium must be paid to the stock-holder to compensate for the greater risk associated with the stock.

There are some cases in which the conditional expectation 2.53 can be determined explicitly. In general, these require that the process or a simple function of the process is Gaussian.

For example, suppose that both $r(t)$ and $\sigma(t)$ are deterministic functions of time only. Then we can solve the stochastic differential equation (2.22) to obtain

$$X_T = \frac{X_t}{q(t,T)} + \int_t^T \frac{\sigma(u)}{q(u,T)} dW_u \tag{2.54}$$

The first term above is the conditional expected value of X_T given X_t. The second is the random component, and since it is a weighted sum of the normally distributed increments of a Brownian motion with weights that are non-random, it is also a normal random variable. The mean is 0 and the (conditional) variance is $\int_t^T \frac{\sigma^2(u)}{q^2(u,T)} du$. Thus the conditional distribution of X_T given X_t is normal with conditional expectation $\frac{X_t}{q(t,T)}$ and conditional variance $\int_t^T \frac{\sigma^2(u)}{q^2(u,T)} du$.

The special case of 2.53 of most common usage is the Black-Scholes model: suppose that $\sigma(S,t) = S\sigma(t)$ for $\sigma(t)$ some deterministic function of t. Then the distribution of X_t is not Gaussian, but fortunately, its logarithm is. In this case we say that the distribution of X_t is lognormal.

Lognormal Distribution

Suppose Z is a normal random variable with mean μ and variance σ^2. Then we say that the distribution of $X = e^Z$ is lognormal with mean $\eta = \exp\{\mu + \sigma^2/2\}$ and volatility parameter σ. The lognormal probability density function with mean $\eta > 0$ and volatility parameter $\sigma > 0$ is given by the probability density function

$$g(x|\eta, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\{-(\log x - \log \eta - \sigma^2/2)^2/2\sigma^2\}. \tag{2.55}$$

The solution to (2.18) with non-random functions $\sigma(t), r(t)$ is now

$$X_T = X_t exp\{\int_t^T (r(u) - \sigma^2(u)/2)du + \int_t^T \sigma(u)dW_u\}. \tag{2.56}$$

Since the exponent is normal, the distribution of X_T is lognormal with mean $log(X_t) + \int_t^T (r(u) - \sigma^2(u)/2)du$ and variance $\int_t^T \sigma^2(u)du$. It follows that the conditional distribution is lognormal with mean $\eta = X_t q(t, T)$ and volatility parameter $\sqrt{\int_t^T \sigma^2(u)du}$.

We now derive the well-known Black-Scholes formula as a special case of 2.53. For a call option with exercise price E, the payoff function is $V_0(S_T) = max(S_T - E, 0)$. Now it is helpful to use the fact that for a standard normal random variable Z and arbitrary $\sigma > 0, -\infty < \mu < \infty$ we have the expected value of $max(e^{\sigma Z + \mu}, 0)$ is

$$e^{\mu + \sigma^2/2}\Phi(\frac{\mu}{\sigma} + \sigma) - \Phi(\frac{\mu}{\sigma}) \tag{2.57}$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function. As a result, in the special case that r and σ are constants, (2.53) results in the famous Black-Scholes formula which can be written in the form

$$V(S, t) = S\Phi(d_1) - Ee^{-r(T-t)}\Phi(d_2) \tag{2.58}$$

where

$$d_1 = \frac{\log(S/E) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}, d_2 = d_1 - \sigma\sqrt{T - t}$$

are the values $\pm\sigma^2(T-t)/2$ standardized by adding $log(S/E) + r(T-t)$ and dividing by $\sigma\sqrt{T-t}$. This may be derived by the following device; Assume (i.e. pretend) that, given current information, the distribution of $S(T)$ at expiry is lognormally distributed with the mean $\eta = S(t)e^{r(T-t)}$.

The mean of the log-normal in the risk neutral world $S(t)e^{r(T-t)}$ is exactly the future value of our current stocks $S(t)$ if we were to sell the stock and invest the cash in a bank deposit. Then the future value of an option with payoff function given by $V_0(S_T)$ is the expected value of this function against this lognormal probability density function, then discounted to present value

$$e^{-r(T-t)} \int_0^\infty V_0(x) g(x|S(t)e^{r(T-t)},\ \sigma\sqrt{T-t})dx. \tag{2.59}$$

Notice that the Black-Scholes derivation covers any diffusion process governing the underlying asset which is driven by a stochastic differential equation of the form

$$dS = a(S)dt + \sigma S dW_t \tag{2.60}$$

regardless of the nature of the drift term $a(S)$. For example a non-linear function $a(S)$ can lead to distributions that are not lognormal and yet the option price is determined as if it were.

Example: Pricing Call and Put options.

Consider pricing an index option on the S&P 500 index an January 11, 2000 (the index SPX closed at 1432.25 on this day). The option SXZ AE-A is a January call option with strike price 1425. The option matures (as do equity options in general) on the third Friday of the month or January 21, a total of 7 trading days later. Suppose we wish to price such an option using the Black-Scholes model. In this case, $T-t$ measured in years is $7/252 = 0.027778$. The annual volatility of the Standard and Poor 500 index is around 19.5 percent or 0.195 and assume the very short term interest rates approximately 3%. In *Matlab* we can value this option using

[CALL,PUT] = BLSPRICE(1432.25,1425,0.03,7/252,0.195,0)

CALL = 23.0381

PUT = 14.6011

Arguments of the function BLSPRICE are, in order, the current equity price, the strike price, the annual interest rate r, the time to maturity $T-t$ in years, the annual volatility σ and the last argument is the dividend yield in percent which we assumed 0. Thus the Black-Scholes price for a call option on SPX is around 23.03. Indeed this call option did sell on Jan 11 for \$23.00. and the put option for \$14 5/8. From the put call parity relation (see for example Wilmott, Howison, Dewynne, page 41) $S + P - C = Ee^{-r(T-t)}$ or in this case $1432.25 + 14.625 - 23 = 1425e^{-r(7/252)}$. We might solve this relation to obtain the spot interest rate r. In order to confirm that a different interest rate might apply over a longer term, we consider the September call and put options (SXZ) on the same day with exercise price 1400 which sold for \$152 and 71\$ respectively. In this case there are171 trading days to expiry and so we need to solve $1432.25 + 71 - 152 = 1400e^{-r(171/252)}$, whose solution is $r = 0.0522$. This is close to the six month interest rates at the time, but 3% is low for the very short term rates. The discrepancy with the actual interest rates is one of several modest failures of the Black-Scholes model to be discussed further later. The low implied interest rate is influenced by the cost of handling and executing an option, which are non-negligible fractions of the option prices, particularly with short term options such as this one. An analogous function to the Matlab function above which provides the Black-Scholes price in Splus or **R** is given below:

```
blsprice=function(So,strike,r,T,sigma,div){
d1<-(log(So/strike)+(r-div+(sigma^2)/2)*T)/(sigma*sqrt(T))
d2<-d1-sigma*sqrt(T)
call<-So*exp(-div*T)*pnorm(d1)-exp(-r*T)*strike*pnorm(d2)
put=call-So+strike*exp(-r*T)
```

```
c(call,put)}
```

Problems

1. It is common for a stock whose price has reached a high level to *split* or issue shares on a two-for-one or three-for-one basis. What is the effect of a stock split on the price of an option?

2. If a stock issues a dividend of exactly D (known in advance) on a certain date, provide a no-arbitrage argument for the change in price of the stock at this date. Is there a difference between deterministic D and the case when D is a random variable with known distribution but whose value is declared on the dividend date?

3. Suppose Σ is a positive definite covariance matrix and η a column vector. Show that the set of all possible pairs of standard deviation and mean return $(\sqrt{w^T \Sigma w}, \eta^T w)$ for weight vector w such that $\sum_i w_i = 1$ is a convex region with a hyperbolic boundary.

4. The current rate of interest is 5% per annum and you are offered a random bond which pays either \$210 or \$0 in one year. You believe that the probability of the bond paying \$210 is one half. How much would you pay now for such a bond? Suppose this bond is publicly traded and a large fraction of the population is risk averse so that it is selling now for \$80. Does your price offer an arbitrage to another trader? What is the risk-neutral measure for this bond?

5. Which would you prefer, a gift of \$100 or a 50-50 chance of making \$200? A fine of \$100 or a 50-50 chance of losing \$200? Are your preferences self-consistent and consistent with the principle that individuals are risk-averse?

6. Compute the stochastic differential dX_t (assuming W_t is a Wiener process) when

   (a) $X_t = \exp(rt)$

   (b) $X_t = \int_0^t h(t)dW_t$

   (c) $X_t = X_0 \exp\{at + bW_t\}$

   (d) $X_t = \exp(Y_t)$ where $dY_t = \mu dt + \sigma dW_t$.

7. Show that if X_t is a geometric Brownian motion, so is X_t^β for any real number β.

8. Suppose a stock price follows a geometric Brownian motion process

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

   Find the diffusion equation satisfied by the processes (a) $f(S_t) = S_t^n$,(b) $\log(S_t)$, (c) $1/S_t$. Find a combination of the processes S_t and $1/S_t$ that does not depend on the drift parameter μ. How does this allow constructing estimators of σ that do not require knowledge of the value of μ?

9. Consider an Ito process of the form

$$dS_t = a(S_t)dt + \sigma(S_t)dW_t$$

   Is it possible to find a function $f(S_t)$ which is also an Ito process but with zero drift?

10. Consider an Ito process of the form

$$dS_t = a(S_t)dt + \sigma(S_t)dW_t$$

   Is it possible to find a function $f(S_t)$ which has constant diffusion term?

11. Consider approximating an integral of the form $\int_0^T g(t)dW_t \approx \sum g(t)\{W(t+h) - W(t)\}$ where $g(t)$ is a non-random function and the sum is over values of $t = nh, n = 0, 1, 2, ...T/h - 1$. Show by considering the distribution

of the sum and taking limits that the random variable $\int_0^T g(t)dW_t$ has a normal distribution and find its mean and variance.

12. Consider two geometric Brownian motion processes X_t and Y_t both driven by the same Wiener process

$$dX_t = aX_t dt + bX_t dW_t$$
$$dY_t = \mu Y_t dt + \sigma Y_t dW_t.$$

Derive a stochastic differential equation for the ratio $Z_t = X_t/Y_t$. Suppose for example that X_t models the price of a commodity in \$C and Y_t is the exchange rate $(\$C/\$US)$ at time t. Then what is the process Z_t? Repeat in the more realistic situation in which

$$dX_t = aX_t dt + bX_t dW_t^{(1)}$$
$$dY_t = \mu Y_t dt + \sigma Y_t dW_t^{(2)}$$

and $W_t^{(1)}, W_t^{(2)}$ are correlated Brownian motion processes with correlation ρ.

13. Prove the *Shannon inequality* that

$$H(Q,P) = \sum q_i \log(\frac{q_i}{p_i}) \geq 0$$

for any probability distributions P and Q with equality if and only if all $p_i = q_i$.

14. Consider solving the problem

$$\min_q H(Q,P) = \sum q_i \log(\frac{q_i}{p_i})$$

subject to the constraints $\sum_i q_i = 1$ and $E_Q f(X) = \sum q_i f(i) = \mu$. Show that the solution, if it exists, is given by

$$q_i = \frac{\exp(\eta f(i))}{m(\eta)} p_i$$

where $m(\eta) = \sum_i p_i \exp(\eta f(i))]$ and η is chosen so that $\frac{m'(\eta)}{m(\eta)} = \mu$. (This shows that the closest distribution to P which satisfies the constraint is obtained by a simple "exponential tilt" or Esscher transform so that $\frac{dQ}{dP}(x)$ is proportional to $\exp(\eta f(x))$ for a suitable parameter η).

15. Let Q^* minimize $H(Q, P)$ subject to a constraint

$$E_Q g(X) = c. \tag{2.61}$$

Let Q be some other probability distribution satisfying the same constraint. Then prove that

$$H(Q, P) = H(Q, Q^*) + H(Q^*, P).$$

16. Let I_1, I_2 ,... be a set of constraints of the form

$$E_Q g_i(X) = c_i \tag{2.62}$$

and suppose we define P_n^* as the solution of

$$\max_P H(P)$$

subject to the constraints $I_1 \cap I_2 \cap ...I_n$. Then prove that

$$H(P_n^*, P_1^*) = H(P_n^*, P_{n-1}^*) + H(P_{n-1}^*, P_{n-2}^*) + ... + H(P_2^*, P_1^*).$$

17. Consider a defaultable bond which pays a fraction of its face value Fp on maturity in the event of default. Suppose the risk free interest rate continuously compounded is r so that $B_s = \exp(sr)$. Suppose also that a constant coupon $\$d$ is paid at the end of every period $s = t+1, ..., T-1$. Then show that the value of this bond at time t is

$$P_t = d\frac{\exp\{-(r+k)\} - \exp\{-(r+k)\{T-t)\}}{1 - \exp\{-(r+k)\}}$$
$$+ pF\exp\{-r(T-t)\} + (1-p)F\exp\{-(r+k)(T-t)\}$$

18. (a) Show that entropy is always positive and if $Y = g(X)$ is a function of X then Y has smaller entropy than X, i.e. $H(p_Y) \leq H(p_X)$.

    (b) Show that if X has any discrete distribution over n values, then its entropy is $\leq log(n)$.