Financial Economics I

Shino Takayama[1]

Faculty of Economics and Business

The University of Sydney

July, 2005

[1]This note is for the lecture of financial economics in the University of Sydney. This is incomplete and please do not cite for any purpose.

Contents

1 Foundations for financial Economics

The purpose of the first chapter is to provide the foundations for the study of modern financial economics. Financial economics is the study to price risky asset, insurance or real estate and so forth. In one word, it is about how rational agents should make a decision on holding of these assets. Therefore, we have to analyze individuals' decision making under uncertainty in the future. In the world where everything in the future is "certain", agents just need to choose one out of some alternatives. On the other hand, in the world where some uncertainty in the future is involved, agents choose lottery or gamble.

1.1 Choice under uncertainty

1.1.1 Expected utility

The concept of utility enters economic analysis typically via the concept of a utility function which itself is just a mathematical representation of an individuals preferences over alternative bundles of consumption goods (or, more generally, over goods, services, and leisure). Expected utility is an axiomatic extension of the ordinal concept of utility to uncertain payoffs. An agent possesses a von Neumann-Morgenstern utility function if she ranks uncertain payoffs according to (higher) expected value of her utility of the individual outcomes that may occur.

Risk averse In economics, people are often supposed to be "risk averse." This means that people want to avoid risk unless adequately compensated for it. For example, if two investments have the same expected return, the one with lower risk will be preferred. If investors just care about the total wealth, people are called "risk neutral." The concave shape of utility functions explains agents' risk averse attribute.

Certainty equivalent The amount of payoff (e.g. money or utility) that an agent would have to receive to be indifferent between that payoff and a given gamble is called that gamble's 'certainty equivalent'. For a risk averse agent, the certainty equivalent is less than the expected value of the gamble because the agent prefers to reduce uncertainty.

Risk premium We can use the expected utility framework in order to work out how much extra return an investor requires to take extra risk. Risk premium measures this extra return.

1.1.2 Measures of risk aversion

How does one measure the "degree" of risk aversion of an agent? Our first instincts may be to appeal immediately to the concavity of the elementary utility functions. However, as utility functions are not unique, second derivatives of utility functions are not unique, and thus will not serve to compare the degrees of risk aversion in any pair of utility functions. However, the risk premiums, expressed in terms of "wealth," might be a better magnitude.

If these can be connected then to the "concavity" of utility curves - adjusted to control for non-uniqueness - so much the better. The most famous measures of risk-aversion were introduced by John W. Pratt and Kenneth J. Arrow.

The Arrow-Pratt measure is an attribute of a utility function. Denote a utility function by $u(W)$ for some alternative W. The Arrow-Pratt measure of absolute risk aversion is defined by:

$$ARA = -u''(W)/u'(W).$$

In simple terms, what we are measuring above is the actual dollar amount an individual will choose to hold in risky assets, given a certain wealth level W. For this reason, the measure described above is referred to as a measure of absolute risk-aversion.

If we want to measure the percentage of wealth held in risky assets, for a given wealth level W, we simply multiply the Arrow-Pratt measure of absolute risk-aversion by the wealth W, to get a measure of relative risk-aversion The Arrow-Pratt measure of relative risk aversion is defined by:

$$RRA = -W \cdot u''(W)/u'(W).$$

If $ARA_A(W) > ARA_B(W)$, then A accepts less gambles expressed in monetary terms than B. If $RRA_A(W) > RRA_B(W)$, then A accepts less gambles expressed in fractions of wealth than B.

1.2 The value of insurance

Now we consider the value of insurance. By buying insurance, people can avoid exposure to risk. The question here is: how much a person is willing to pay for this purpose. This defines the value of insurance.

Suppose that you want to buy a car insurance. Suppose that insurance pays q for coverage S. Obviously, there are two cases. One is that the car could be stolen and the other is that nothing happens. Suppose that the probability that nothing happens is π and the total wealth that you have is W and loss of being stolen the car is L. Then, the amount that you can spend on consumption if nothing happens is $x_1 = W$ and the amount that you can spend on consumption if the car is stolen is $x_2 = W - L$. Suppose that you buy a car insurance. Then, the amount that you can spend on consumption if nothing happens is $x_1 = W - q$ and the amount that you can spend on consumption if the car is stolen is $x_2 = W - q + S - L$. Under a fair price, the following must hold:

$$\pi W + (1-\pi)(W-L) = \pi x_1 + (1-\pi)x_2.$$

Thus, we obtain:

$$x_2(x_1) = \frac{\pi(L-x_1) + W - L}{1-\pi}.$$

Thus, we obtain:

$$\frac{\partial x_2}{\partial x_1} = -\frac{\pi}{1-\pi}.$$

1.3 Trading of simple securities

Financial securities are state contingent claims. For example, in state 1 where economy is boom, the security pays \$1, while in state 2 where economy is bust, the security does not pay anything. Here we consider a simple security that pays payoff in exactly one state of the world. In the above example, actually instead of insuring the case, you could buy the security that pays off in the state in which your car gets stolen and sell security that pays off when the car is not stolen.

Now we consider the following situation. Agent A has payoffs of \$1000 if state is boom and \$250 if state is bust. Agent B has payoffs of \$500 in both states. How should these agents trade securities? Agents trade securities before they learn the state.

We denote units of boom bond by s_1 and those of bust bonds by s_2. We denote price of one unit of boom (bust) bond by $p_1(p_2)$. The questions that we are asking here are: what are prices of the bonds? And what trades occur and what degree of insurance is achieved?

We use the following equilibrium condition called "tangency condition for equilibrium":

$$\frac{p_1}{p_2} = \frac{\pi U'(f_1)}{(1-\pi)U'(f_2)} \tag{1}$$

When both agents are risk averse, the utility functions are concave. Therefore, we have: $U'(f_1) > U'(f_2)$ if and only if $f_1 < f_2$. When $\frac{p_1}{p_2} < \frac{\pi}{1-\pi}$, $f_2 < f_1$ because we need $U'(f_1) > U'(f_2)$ in order to have the tangency condition for equilibrium.

2 Fixed income securities

The first part of this course deals with asset pricing theory. The theory has on one hand positive implications, answering the question of how are the asset prices determined? On the other hand we can view our theory as providing normative recommendations. What should the asset price be? How to take advantage of apparent mispricings? The central idea in modern asset pricing theory is that the price of a security is the discounted present value of the assets payments (dividends). Therefore securities with a fixed, predetermined payments structure, called fixed income securities (bonds), are a natural starting point. The fixed income securities represent roughly a \$11.2 trillion market in US. By comparison, the much more talked about equity markets amount to \$12 trillions.

2.1 Bond prices and yields (Chapter 14)

A bond is described by:

- maturity (expiration date)

  We will denote by T the number of periods until expiration. A period is taken to be the time interval between two consecutive coupon payments.

- coupon payment, C, representing the payment made in each period until expiration.

  Thus there are T coupon payments.

- face value, FV , also called par-value, representing a last payment, in period T.

  We define the coupon rate as the ratio $\frac{C}{FV}$.

2.1.1 Bond pricing

If the interest in the world would be constant, at a level r, then no arbitrage implies that the price of the asset would be

$$p_0 = \sum_{t=1}^{T} \frac{C}{(1+r)^t} + \frac{FV}{(1+r)^T}. \tag{2}$$

Hidden in this formula is the fact that coupon payments are made at the end of each period (because we discount the first coupon payment). The price of the bond, p_0, should be viewed as the beginning of period 1, end of period 0 ex-dividend price. When we will talk about p_1, this will be the end of period 1 (after the coupon payment in period 1), or if you want, immediately at the beginning of period 2 etc. This is the way how bond prices are also quoted, and sometimes referred to as flat price. However since most of the bonds are traded between coupon payments, to the flat price one has to add the accrued interest in order to arrive at the actual invoice price. I will recall now the formula for computing the sum of a finite geometric progression, which is extremely useful in computing present values in the presence of constant interest rates:

$$1 + x + x^2 + ... + x^n = \frac{1 - x^{n+1}}{1 - x}, \tag{3}$$

and notice that by taking $n \to \infty$ and assuming $x \in (0, 1)$, we get the formula for the sum of a geometric progression:

$$1 + x + x^2 + ... = \frac{1}{1 - x} \tag{4}$$

for $|x| < 1$.

It is easy to see that the using formula (2) we can write the price of the bond in (1) simply as

$$p_0 = \frac{C}{r} \cdot (1 - \frac{1}{(1+r)^T}) + \frac{FV}{(1+r)^T}. \tag{5}$$

Notice that if our bond is a perpetuity (consol), which is a bond without expiration date, then the price of such a consol is simply

$$p_0 = \frac{C}{r}. \tag{6}$$

Excercise 1 *Using the formula (5) for the price of a consol, obtain formula (4) for the price of a coupon bond by rewriting its discounted payments as the difference between the discounted payments of two consols.*

2.1.2 Bond yields

The value of r which produces the equality (1) is called the yield to maturity (YTM).

Remark The price of a bond is equal to its par-value if and only if the coupon rate equals the yield to maturity. Just notice that if $C = r \cdot FV$,

$$\frac{C}{(1+r)^T} + \frac{FV}{(1+r)^T} = \frac{FV}{(1+r)^{T-1}}, \tag{7}$$

and continue to iterate until you get $p_0 = FV$. Bonds that sell above par-value are called premium bonds, and have an YTM smaller than the coupon rate (why?). Bonds selling for less than the par-value are called discount bonds and have an YTM bigger than the coupon rate. If for example the YTM of a bond with a semiannual coupon payment is 3%, then the bond equivalent yield is simply the annual percentage rate (APR) computed using simple interest techniques, that is 6%. However the effective annual yield of the bond accounts for compound interest and is $(1.03^2 - 1) \times 100\% = 6.09\%$. The YTM is the most used proxy for the average return of the bond. In practice it is sometimes used also the current yield of a bond, defined as the annual coupon payment divided by the bond price,

$$CY = \frac{C}{p_0}. \tag{8}$$

Excercise 2 *Show that for premium bonds the coupon rate is greater than the current yield which in turn is greater than the YTM.*

The opposite inequalities hold for discount bounds. The proof follows from the following:

$$\sum_{t=1}^{T} \frac{C}{(1+\frac{C}{p_0})^t} + \frac{p_0}{(1+\frac{C}{p_0})^T} = p_0 < \sum_{t=1}^{T} \frac{C}{(1+YTM)^t} + \frac{p_0}{(1+YTM)^T}. \tag{9}$$

Alternatively one could have used (4). Therefore by computing (easily) the coupon rate and the current yield, we can get some bounds for the YTM, which can be computed usually only using a financial calculator, or Excel (or a nonlinear solver, etc.). Another concept is the realized compound yield. It is computed after the expiration of the bond, because it requires a knowledge of the realized interest rates. It assumes that the coupon payments are reinvested using the realized market rates, computing the value of the bond at the expiration. The realized yield is simply the (constant) rate at which the future value of your initial investment equals the future value obtained through the reinvestment strategy. For example, assume that you buy at par value a two period coupon bond, with coupon payments of \$100 and face value of \$1000. The realized interest rates 11% for the first period and 15% for the second period. Therefore the reinvested coupons and face value in period two terms will be

$$(1+.15) \times 100 + 100 + 1000 = 1215.0 \tag{10}$$

and the realized yield is computed from

$$1000 \times (1 + y_{realized})^2 = 1215, \tag{11}$$

which gives

$$y_{realized} = (1215/1000)^{1/2} - 1 = 0.10227, \tag{12}$$

or 10.227%.

The concept of yield to call plays an important role for callable bonds. It is easy to compute the "yield to first call" which assumes that the bond will be called as soon as it becomes callable. Actually it can be shown that the yield from a callable bond is at least the minimum between the yield to first call or the yield from not calling the bond at all. In other words, the lowest yield strategy for an investor is if the firm calls the bond as soon as it becomes callable, or if does not call it at all. If it calls the bond any time between the first time it becomes callable and the expiration, the yield for investor is higher. We will discuss it in class and analyze the implied yield on such a bond.

2.1.3 Bond prices over time and holding-period returns

One can show that if the interest rates do not move, then the price of the bond would converge monotonically toward the face value, in other words if the initial price is below par-value, then $p_0 < p_1 < ... < p_T = FV$, while if $p_0 > FV$, then $p_0 > p_1 > \cdots > p_T = FV$.

The idea is that if, e.g., $p_0 < FV$, then clearly $r > C/FV$ and thus

$$\frac{C}{(1+r)^T} + \frac{FV}{(1+r)^T} < \frac{FV}{(1+r)^{T-1}}, \tag{13}$$

which proves that

$$p_0 = \sum_{t=1}^{T} \frac{C}{(1+r)^t} + \frac{FV}{(1+r)^T} < \sum_{t=1}^{T-1} \frac{C}{(1+r)^t} + \frac{FV}{(1+r)^{T-1}} = p_1. \tag{14}$$

Another important concept is the holding-period return, or simply return. For any asset in general, the return in period 1, or the return from period 0 to 1 is calculated as

$$R_1 = \frac{D_1 + p_1 - p_0}{p_0}, \tag{15}$$

where D_1 is the dividend paid by the asset in the period under consideration (period 1 here). Clearly for a coupon bond, $D_1 = C$.

Remark If the interest rates are constant, then the return on the bond (in each period) is equal with the yield to maturity (interest rate). Indeed, notice that

$$(1+r) \times p_0 = p_1 + C, \tag{16}$$

hence the conclusion. However when there are fluctuations in the market interest rates, it means that the YTM must fluctuate, which means that returns fluctuate, due to capital gain or losses.

2.1.4 Default risk and bond pricing

We will discuss the spread in yields between different bonds due to default risk. Default risk is estimated by rating agencies etc. (slides will be used for this part).

2.2 The term structure of interest rates (Chapter 15)

In a world of constant interest rates, we saw that bonds can have different yields (returns) only if they have different default risk. However, even if we restrict our attention to US government bonds, which are default free, we notice that the yields on different maturity bonds are very different. This is due to the fact that interest rates fluctuate a lot in time. But if interest rates fluctuate, two coupon bonds with the same maturity but different coupons (or face values) command different yields, so again for comparison reasons we should look only at the yields of zero-coupon bonds with different maturities. The graph plotting the yields on (zero-coupon government) bonds versus their maturity is called the yield curve. Even if zero-coupon government bonds are not available for all maturities, bond dealers might ask the Treasury to break down the cash flows paid by a coupon bond into independent stand-alone zero-coupon bonds. This stripping process is administered by a Treasury program called STRIPS (Separate Trading of Registered Interest and Principal of Securities), and these zero coupon securities are called strips and carry their own CUSIP identification number (CUSIP = Committee on Uniform Securities Identification Procedures). In the data we observe that usually the yields on longer-term bonds are higher the bigger is the maturity of the bond. Intuition tells us that this indeed should be the case, since longer-term bonds are riskier (subject to changes in market interest rates) and therefore should offer higher yields to compensate for interest rate risk. Another possibility is that investors might expect future interest rates to be higher and therefore higher yields are simply a reflection of the increased interest rates in the later period of the bond. We will build some theories based on the intuition just mentioned, to explained the following stylized empirical facts:

- Interest rates for different maturities move together over time.
- Yield curves tend to have steep upward slope when short rates are low and downward slope when short rates are high.
- Yield curve is typically upward sloping.

The first step in understanding the yield curve is to assume a world with no uncertainty, in other worlds we assume that investors have perfect foresight about the future short term interest rates. Indeed assume that the short interest rates for the next n periods are $r_1, \cdots, r_n$. Clearly, by no arbitrage, the price of a zero-coupon bond with face value FV and expiring after n periods is $P = \frac{FV}{(1+r_1)(1+r_2)\cdots(1+r_n)}$. The yield to maturity y_n for the n-period zero coupon bond, also called the spot rate can be computed from $P = \frac{FV}{(1+y_n)^n}$. This shows that the gross yield $1 + y_n$ is simply the geometric average of the short rates: $(1+y_n)^n = (1+r_1)(1+r_2)\cdots(1+r_n)$, (of course taking logs we can approximate the above relation as $y_n \simeq (r_1 + \cdots + r_n)/n$, but we will prefer to work with the exact formula).

Also notice that we can express the future short rates as a function of the current spot rates for two different maturity bonds, $1 + r_n = \frac{(1+y_n)^n}{(1+y_{n-1})^{n-1}}$.

Excercise 3 *In this world with no uncertainty we can show that the returns on all bonds are equal. This includes coupon bonds, which as we mentioned, have usually very different yields. Specifically, one can show that period* *one return is r_1, period two return is r_2, etc. Try to do the proof yourself. Start with a general coupon bond, write the formula for p_0 and p_1 and as in the constant interest rate world, write the relation between $p_1/(1+r_1)$ and p_0 etc.*

Unfortunately, investors do not have perfect foresight and do not know the future short rates. However if they use the same formula for inferring future short rates as in a world without uncertainty, they will obtain the so-called forward rates:

$$1 + f_n = \frac{(1+y_n)^n}{(1+y_{n-1})^{n-1}}. \tag{17}$$

Notice also that

$$(1+y_n)^n = (1+r_1)(1+f_2)\cdots(1+f_n), \tag{18}$$

The forward rates exactly represent the market interest rate at which you would be able to arrange now to make a loan in the future (a forward loan). We are now ready to understand the different theories of the term structure:

2.2.1 The expectations hypothesis

This theory postulates that the forward rates equal the expected future short rates, $f_2 = \mathrm{E}(r_2)$, or more generally,

$$(1+f_2)\cdots(1+f_n) = \mathrm{E}_0[(1+r_2)\cdots(1+r_n)]. \tag{19}$$

This implies that the (gross) spot rates are indeed a geometric average of the (gross) expected future rates. This explain the first two stylized facts about the yield curve. Indeed, yields of different maturities move together since they average the interest rates for the common period the bonds are outstanding (that is, the overlapping period until their corresponding maturity). Secondly, when short term interest rates are low it means that people expect higher future short term interest rates, which averaged will provide higher yields for longer maturity bonds. The opposite holds true when current interest rates are high. However, since there is equal probability for the future rates to be low or high, expectations hypothesis cannot satisfactorily explain the third fact, that yield curves are typically sloping upward.

2.2.2 Liquidity Preference

The previous theory relies on the idea that investors are risk-neutral. However, investors are risk averse. This observation, combined with the reasonable fact that short-term investors are predominant in the market, will explain the fact that typically the yield curve is upward sloping. First, let us see what happens if the market is dominated by long term investors, who care about their return over longer periods, in our case over two periods. Now the actual return of such an investor is $(1+r_1)(1+f_2)$ if he buys \$1 worth of a two-period bond, and it is $(1+r_1)(1+r_2)$ if he buys \$1 worth of a one period bond and reinvest the proceeds into another one period bond. Notice however that now the future short term rate r_2 is a random variable. Therefore in order for a risk averse investor with a two-period investing horizon to be indifferent between the two strategies, we must have that the risky investment strategy offers higher expected return $\mathrm{E}[(1+r_1)(1+r_2)] > (1+r_1)(1+f_2)$ which is equivalent to $\mathrm{E}(r_2) > f_2$.

Therefore if long term investors dominate the market, one would expect to see that the forward rates are below the expected future short rates. However, there are good reasons that the investors prefer shorter investing horizon, in order to increase the liquidity of their portfolio. Therefore the short term investors are the predominant ones in the market, and they will drive the forward rates above the expected future rates[1].

Therefore longer maturity bonds have to pay a higher yield, a liquidity premium, in order to entice investors. This explains why typically the yield curve is sloping upward.

3 Portfolio Theory

Portfolio theory strives to understand the optimal proportion of wealth to be allocated among different classes of assets. The theme for this part of the course is that investors face a trade-off between the return and the risk embedded in different classes of assets.

3.1 History of interest rates and risk premiums (Ch. 5)

Clearly investors prefer, everything else equal, assets with higher return (or offering higher interest rates). We have to understand, however, what is embedded in the "everything else equal" phrase. In the rest of this chapter we will investigate what are the factors influencing the desirability of different assets.

3.1.1 Nominal versus real returns (interest rates)

Investors are interested in the purchasing power of their nominal returns. For example, faced with the choice between a US T-Bill with a 3% interest rate and a EU T-Bill with a 3% return, the investors will the bill subject to lower expected inflation. If the expected inflation in US is lower than that expected for the EU, the investor will choose the US T-Bill and vice versa. What is the relationship between nominal interest rates (R), real interest rates (r) and inflation rate (i)? If you deposit \$1 into a bank you will receive next period $1+R$ dollars, therefore your change in purchasing power, which is how real interest rate is defined, is simply

$$r = \frac{(1+R)/P_{t+1}}{1/P_{t-1}}, \tag{20}$$

where P_t represents the price level at t, and P_{t+1} represents the price level in period $t+1$. Since inflation is defined as the relative change in prices,

$$1+i = \frac{P_{t+1}}{P_t}, \tag{21}$$

[1] Strictly speaking, forward prices should be smaller than expected future prices of the bond. Indeed, a short term investor can get $1+r_1$ for sure by investing one dollar in the short term bond, or he can get p_1/p_0 by investing in the two-period bond. Thus $1+r_1 < \mathrm{E}[p_1/p_0]$ and since $p_1 = 1/(1+r_2)$ and $p_0 = 1/[(1+r_1)(1+f_2)]$ we get $\mathrm{E}[1/(1+f_2)] < E[1/(1+r_2)]$. However we will be sloppy on purpose to simplify the exposition.

we obtain

$$r = \frac{1+R}{1+i} - 1 = \frac{R-i}{1+i}. \tag{22}$$

Most of the time you will encounter an approximation of the previous formula,

$$r \simeq R - i \tag{23}$$

which overstates slightly the real interest rate (assuming positive inflation rates). Therefore real returns fluctuate even on riskless nominal securities, due to unexpected inflation. This implies that even the riskless nominal securities are risky in real terms.

3.1.2 Taxes

Investors also care about the net (of taxes) return, as opposed to gross returns. Notice that tax liabilities are based on nominal income, which implies that if there is substantial inflation, an investor can be pushed in a higher tax bracket and have therefore a reduced after-tax return. However this problem was remedied in the tax code by using index-linked tax brackets (bracket creep). However, inflation hurts investors even if they do not move into a higher tax bracket, and even if it is perfectly anticipated. Let us analyze the following example. Consider an investor that is taxed at a 40% marginal rate. Assume that he invests into a treasury inflation protected security (TIPS) which guarantees him a 2 % real return. If there is no inflation, he will be left with $(1-.4)\times 2 = 1.2\%$ real return. If however inflation turns out to be positive, let us say 2%, the investor receives an after-tax nominal return of $(1-.4)\times 4\% = 2.4\%$, and in real terms this represents only a net $2.4-2 = .4\%$ real return. Investors suffer an inflation penalty equal to the tax rate times the inflation rate, because the tax code does not recognize that part of your income is just compensation for inflation.

3.1.3 Risk and risk premiums

Finally, assets subject to identical inflation and tax rates can have very different returns due to their different risk characteristics. The excess return of an asset is defined as the difference between the return on that asset and the risk-free rate. The expected excess return on an asset is called the risk premium on the asset. To quantify the risk in return distributions, we use, in addition to variance or standard deviation, the lower partial standard deviation, which represents the standard deviation computed separately for the negative deviations. Another way to check how important the large negative outcomes are is to compute the skewness of the distribution, which is the third centered moment normalized by the cubed standard deviation. Professional investors and regulators use also the concept of value at risk (VaR)(as a measure of potential losses), representing the 5% - quantile of the distribution, which is the value below which lie 5% of the distribution. It indicates that one can expect a loss greater than VaR with a 5% probability. For a normal random variable VaR always lies 1.65 standard deviations below the mean. The historical data for average returns and standard deviations for different asset classes is given below:

Series	Geom. Mean	Arith. Mean	Stan. Dev.
Sm Stk	11.6	17.7	39.3
Lg Stk	10.0	12.0	20.6
LT Gov	5.4	5.7	8.2
T-Bills	3.8	3.8	3.2
Inflation	3.1	3.1	4.4

The risk premiums therefore are:

Series	Risk Premiums	Real Returns
Sm Stk	13.9	14.6
Lg Stk	9.3	8.9
LT Gov	1.9	2.6
T-Bills	–	0.7

3.2 Risk and risk aversion (Ch. 6)

The historic data on risk premiums shows that investors also care about volatility (risk) of returns. To eliminate any doubt that investors care not only about expected return, we discussed the idea of leverage, and the St. Peterburg Paradox.

Example 1 *(Leverage). Given a stock with positive expected return $E(R) > 0$, you can achieve any expected return you desire by leveraging the stock. Suppose that you have 1 dollar. Indeed, if you invest 1 dollar in the stock, you get expected gross return of* $\mathrm{E}(1 + R) = 1 + \mathrm{E}(R)$. *However, if you borrow $n - 1$ dollars, and invest everything in the stock, you get* $\mathrm{E}[n(1 + R) - (n - 1)] = 1 + nE(R)$ *as a gross return. In other words you increased your return by a factor n. Therefore one can "leverage" returns by borrowing and increasing by any factor the expected return. What is the downside? Well, since your risky returns are described by the random variable nR, clearly the volatility of your returns are higher (check that the standard deviation increases by a factor n). QUESTION: Since you could always invest in T-Bills, which have a guaranteed (nominal) return, that is zero-volatility, leveraging a T-bill as in the discussion above implies that you can achieve unlimited return for sure, at zero volatility. What is the problem in this argument?*

Example 2 *(St. Petersburg Paradox) Assume that somebody offers the following lottery (gamble) to you. Suppose that an unbiased coin is tossed until the first head appears. If the head appears on the n-th trial, you win 2^n dollars. How much will you be willing to pay to be allowed into this lottery? It turns out that the expected return of the gamble is infinite, however most persons would be willing to pay only very small amounts for it. We explain the paradox by pointing out that people display decreasing marginal utility of wealth. Mathematically, this means that their utility function $u : \mathbb{R} \to \mathbb{R}$ derived from wealth is concave. In the face of uncertainty,* *investors compute their expected utility from wealth, as postulated by Von Neumann and Morgenstern. It turns out that risk averse investors (with concave Bernoulli utilities) will derive a finite utility from the St. Petersburg gamble. We assume $u(w) = \ln w$ and show that if R denotes the risky gamble described above, the expected utility of an investor is*

$$\mathrm{E}u(R) = \sum_{t=1}^{\infty} \frac{1}{2^t} \ln(2^t) = 2 \ln 2.$$

Thus the certainty equivalent CE of the St. Petersburg gamble for log utility investor is simply the solution to

$$\mathrm{E} \ln(R) = \ln(CE), \tag{24}$$

or

$$CE = e^{2 \ln 2} = 4. \tag{25}$$

This shows that an investor with log utility of wealth will be willing to pay only $4 to play the lottery.

From now on, the concept of expected utility will be our tool for understanding the trade-off between higher return and higher risk faced by investors. In practice, the most used preferences in modeling the behavior of the agents are the mean-variance preferences. Thus, the utility of an investor derived from a portfolio R is simply

$$U(R) = \mathrm{E}(R) - b \cdot \mathrm{var}(R). \tag{26}$$

Here, a higher $b > 0$ reflects a more risk averse agent[2]. Clearly the mean variance utility is an expected utility, with quadratic Bernoulli utility. As one simple example, when we define the utility function by

$$u(x) = -(y - x)^2, \tag{27}$$

for $x \leq y$, then we see that $\mathrm{E}u(R)$ can be formulated by a function of mean utility of R, and variance of R.

Excercise 4 *What is the certainty equivalent of a risky asset R, for an investor with mean-variance preferences as in (26)?*

Investors that are judging risky prospects solely based on their expected return are called risk neutral. Clearly they have a linear Bernoulli utility. Investors that are inclined to take risk are called risk lovers, and they have a convex Bernoulli utility.

3.3 Optimal portfolios (Ch. 7 and 8)

We saw that individual assets are subject to considerable risk. However, investors can protect themselves against excessive volatility by hedging their risk or through portfolio diversification. Hedging risk means simply investing in assets with offsetting patterns of returns. The idea of diversification requires investments be made in a wide

[2] For example, AIMR (Association of Investment Management and Research) uses $U = \mathrm{E}(R) - .005A \cdot \sigma_R^2$) where A represents the risk aversion of the investor, and .005 is a scaling convention allowing the use of expected returns and standard deviations as percentages rather than decimals.

variety of assets so that exposure to the risk of any particular security is limited. Both of this ideas will be illustrated below. Assume that you can invest in two stocks, with returns R_1 and R_2. If you invest a fraction w_1 of your wealth in the first asset, and a fraction $w_2 = 1 - w_1$ in the second asset, the expected return and variance of your portfolio $R = w_1 R_1 + w_2 R_2$ are:

$$\mathrm{E}(R) = E(w_1 R_1 + w_2 R_2) = w_1 \mathrm{E}(R_1) + w_2 \mathrm{E}(R_2), \tag{28}$$

and [3]

$$\begin{aligned} \mathrm{var}(R) &= \mathrm{var}(w_1 R_1 + w_2 R_2) = w_1^2 \mathrm{var}(R_1) + w_2^2 \mathrm{var}(R_2) + 2 w_1 w_2 \mathrm{cov}(R_1, R_2) \\ &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_1 \sigma_2 (1 + \rho). \end{aligned}$$

Notice that

$$\mathrm{var}(R) = (w_1 \sigma_1 - w_2 \sigma_2)^2 + 2 w_1 w_2 \sigma_1 \sigma_2 (1 + \rho) \tag{29}$$

and

$$\mathrm{var}(R) = (w_1 \sigma_1 + w_2 \sigma_2)^2 - 2 w_1 w_2 \sigma_1 \sigma_2 (1 - \rho) \tag{30}$$

Since the correlation coefficient satisfies $-1 \leq \rho \leq 1$, equation (29) implies that if the stocks are perfectly negatively correlated, that is if $\rho = -1$, then each asset is a perfect hedge for the other, and one can get a portfolio a riskless portfolio (zero variance) by choosing the weights such that

$$w_1 \sigma_1 = w_2 \sigma_2,$$

or

$$w_1 = 1 - w_2 = \frac{\sigma_2}{\sigma_1 + \sigma_2}.$$

In general, if ρ is negative, the assets can still serve as an (imperfect) hedge for each other. However, even if $\rho \geq 0$, agents can benefit from investing in both assets, and this is called diversification. Indeed, if the two assets are not perfectly correlated, that is if $\rho < 1$, relation (30) shows that

$$\rho_R < w_1 \sigma_1 + w_2 \sigma_2,$$

thus the standard deviation of the mixed portfolio is strictly smaller than the average standard deviations of the assets included in the portfolio.

The benefits of diversification can be illustrated even more dramatically if we assume that we have a large number of assets.

We will see that all the idiosyncratic risk (unsystematic, firm-specific) can be diversified away, while the market risk (aggregate, systematic) cannot be diversified away. Coming back to the case of two assets, one sees that $\mathrm{var}(R)$ is a quadratic function of $\mathrm{E}(R)$ as the weights w_1, w_2 are taking all possible values. For more than two assets, one has to plot the lowest possible variance that can be attained for a given portfolio expected return.

[3] We use the notation $\sigma_1^2 = \mathrm{var}(R_1)$, $\sigma_2^2 = \mathrm{var}(R_2)$, $\rho = \mathrm{corr}(R_1, R_2)$.

This is called the minimum-variance frontier or the efficient frontier[4] and it is again a (convex) parabola if the expected return is interpreted as the dependent variable. We can get to any point on the mean-variance frontier by starting with two returns on the frontier and forming portfolios, in other words the efficient frontier is spanned by any two frontier returns. Notice also that portfolios of a risky asset and risk-free asset R^f give rise to straight lines in mean-standard deviation space, called capital allocation line (CAL). For the two asset case, using simple calculus, one can even compute the minimum variance portfolio (the left-most point on the efficient frontier, if expected returns are on the vertical axis), which is obtained for $w_1^{min} = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2 - 2\rho\sigma_1\sigma_2}$.

However, in general the investors will not choose the minimum variance portfolio, since they might want higher expected return. Once the agents preferences are given by a Bernoulli utility u, in order to compute the optimal portfolio, we just solve the maximization problem

$$\begin{aligned} \max_{w_1,w_2} \quad & \mathrm{E}u(w_1R_1 + w_2R_2) \\ s.t. \quad & w_1 + w_2 = 1. \end{aligned}$$

If there is a risk free asset and an arbitrary number n of risky assets, one derives first the efficient frontier for the n risky assets and then draws the CAL tangent to the upper part of the frontier (and passing through the risk free return R^f). This is called capital market line (CML). A mean-variance agent will just pick the risky portfolio where his indifference curves are tangent to the CML. In other words, we have an important separation property. The agents always choose the same risky portfolio, and allocate their wealth between the risk free asset and the risky portfolio according to their preferences. The risky portfolio, being the tangency point of the CML to the efficient frontier, is the portfolio that maximizes the slope of the CAL passing through him, therefore the optimal risky portfolio P is the portfolio with the highest reward to variability ratio, called Sharpe ratio:

$$\frac{\mathrm{E}(P) - R^f}{\sigma(P)} = \max_R \frac{\mathrm{E}(R) - R^f}{\sigma(R)} \tag{31}$$

where the maximization is across all feasible portfolios (sum of weighted individual returns). To understand better the separation property, assume that there are n risky assets $R_1, \cdots, R_n$ and a riskless asset R^f which is not in the asset span of the risky assets. A mean variance agent will solve the problem

$$\begin{aligned} \max_{w,R} \quad & \mathrm{E}((1-w)R^f + wR) - bw^2\mathrm{var}(R) \\ s.t. \quad & R = w_1R_1 + w_2R_2 + \cdots + w_nR_n \\ & w_1 + w_2 + \cdots + w_n = 1. \end{aligned}$$

We see that given R, w is optimal only if (take the derivative of the objective w.r.t w):

$$w = \frac{E(R) - R^f}{2b\mathrm{var}(R)}.$$

Notice that with this anticipated best response, the objective becomes

$$\max_R \frac{R^f + (E(R) - R^f)^2}{2b\mathrm{var}(R)} - b\left(\frac{\mathrm{E}(R) - R^f}{4b\mathrm{var}(R)}\right)^2 \mathrm{var}(R) = R^f + \frac{(E(R) - R^f)^2}{4b\sigma^2(R)}.$$

[4]Efficient frontier usually refers only to "half" of the mean-variance frontier, the part with high expected returns.

which shows why the optimal risky portfolio is maximizing the Sharpe ratio and why it does not depend on the risk aversion coefficient b.

4 Capital asset pricing model (CAPM)

CAPM extends Markowitz' portfolio analysis to an equilibrium model. In the optimal portfolio problem analyzed in the previous chapter, asset returns (or prices) are taken as given, and therefore no answer as to what the price of an asset should be is offered. The previous analysis is incomplete, in that it analyzes only the demand side of the market. However, it offered the crucial insight that investors with mean-variance preferences[5] will choose risky assets in the same proportion in their optimal portfolio[6]. CAPM determines the price of an asset so that the investors are content in holding the current market portfolio.

4.1 Formal derivation of CAPM

What do we understand by market portfolio? Assume that there are n risky assets $(\{1, 2, \cdots, n\})$ available for trade, and a risk free asset. Denote the return on asset i by R_i and the risk free return by R^f. Assume also that the value of the outstanding shares of risky asset i is a fraction w_M^i of total market capitalization (the value of all outstanding equity). Then the market portfolio consists of all outstanding equity, and the market return is

$$R^M = \sum_{i=1}^{n} w_i^M R_i.$$

Assume that we have a finite number m of investors, with mean-variance preferences, but allowed to differ in their risk aversion characteristic. Concretely, assume that the utility of investor j is given by

$$U_j(R) = \mathrm{E}(R) - A_j \mathrm{var}(R).$$

The investor has to decide how much to invest in the risk free asset and in the risky assets. Denoting by k_j the fraction of wealth agent j puts into the risky assets, and by w_i the percentage invested in risky asset i of the part dedicated to risky investments, we have to solve the maximization problem:

$$\max_{k_j, w_1^j, \cdots, w_n^j} R^f + k_j \sum_{i=1}^{n} w_i^j [\mathrm{E}(R_i) - R_f] - A_j k_j^2 \mathrm{var}(\sum_{i=1}^{n} w_i^j R_i).$$

The first order conditions for investor j are (letting $R^j = \sum_{i=1}^n w_i^j R_i$):

$$\begin{aligned} \mathrm{E}(R^j) - R^f &= 2k_j A_j \mathrm{var}(R^j) \\ \mathrm{E}(R_l) - R^f &= 2A_j k_j \sum_{i=1}^{n} w_i^j \mathrm{cov}(R_l, R_i) = 2A_j k_j \mathrm{cov}(R_l, R^j), \ \ \forall \ \ l \in \{1, 2, \cdots, n\}. \end{aligned}$$

[5]Paul Samuelson (Review of Economic Studies 37, 1970) proves that in many important circumstances, disregarding moments higher than the variance will not affect portfolio choice, which is the major theoretical justification for mean-variance analysis. Essentially what is needed is the continuity of stock prices (non-existence of sudden jumps), which implies that the uncertainty of stock returns over increasingly small periods vanishes. By frequently revising their portfolio, investors will make higher moments of the stock return distribution negligible.

[6]They will choose a different combination of the risky portfolio and the risk-free asset, depending on their coefficient of risk aversion.

Combining the two gives:

$$\mathrm{E}(R_l) - R^f = \frac{\mathrm{cov}(R_l, R^j)}{\mathrm{var}(R^j)}[\mathrm{E}(R^j) - R^f], \ \ \forall \ \ l \in \{1, \cdots, n\}.$$

The separation theorem derived in the section on optimal portfolio implies that $\forall i \in \{1, \cdots, n\}$,

$$w_i^j = w_i^{j'} = w_i^M, \ \ \forall \ \ j, j' \in \{1, \cdots, m\},$$

thus everybody holds a portfolio of risky assets that mimics the market portfolio. It follows that

$$R^j = R^M \ \ \ j \in \{1, \cdots, m\},$$

and therefore

$$\mathrm{E}(R_l) - R^f = \frac{\mathrm{cov}(R_l, R^M)}{\mathrm{var}(R^M)}[\mathrm{E}(R^M) - R^f] = \beta_l[\mathrm{E}(R^M) - R^f], \ \ \forall l \in \{1, \cdots, n\}, \tag{32}$$

where we defined the beta of the asset l to be

$$\beta_l = \frac{\mathrm{cov}(R_l, R^M)}{\mathrm{var}(R^M)}. \tag{33}$$

Beta measures how the return of an asset moves together with the return on the market portfolio[7].

Equation(32) summarizes CAPM, and states that the excess expected return on any asset is proportional with its beta, that is, proportional to its covariance with the market return. This shows that the assets that require a positive excess return are those that move together with the market, and hence are a poor hedge (since the market return is also the return on the risky part of your portfolio). On the contrary, assets that move negatively together with the market are a good hedge, and hence they are held even when they offer a low expected return. An asset that is uncorrelated to the market commands a zero expected excess return, since its risk can be completely diversified away.

4.2 Financial derivatives: Options

A derivative security is a financial contract written on an underlying asset from which it "derives" its value, hence the name. For example,

1. the value of a stock option depends upon the value of the stock on which is written;
2. the value of a foreign currency forward contract depends upon the foreign currency forward rate; similarly a swap option depends upon the value of the underlying swap contract;
3. the value of a Treasury bill futures contract depends upon the price of an underlying Treasury bill.

Derivatives are extremely useful for hedging purposes (and of course, for speculation). Futures and options are actively traded on organized exchanges, and hence are standardized with respect to description of the underlying

[7] Notice that β_l is the theoretical regression coefficient in a regression of the return on the asset and the return on the market portfolio.

asset, the rights of the owner and maturity. Forward contracts on the other hand, are customized to its owner (hence, non-standardized) and are traded in the inter-bank market. Our purpose is to understand how to price derivatives, given the characteristics of the underlying asset. The crucial idea is to replicate the payoff of the derivative with a portfolio of securities with known prices. Absence of arbitrage requires that the price of the derivative is exactly equal to the cost of the replicating portfolio. What is the meaning of "arbitrage" Arbitrage is a trading strategy requiring no cash input and that has a probability of making profits and zero probability of loss. If arbitrage opportunities exist, arbitrageurs will cause prices to adjust until arbitrage is no longer possible. The following example illustrates a forward contract and also the way to determine its price by arbitrage arguments.

Example 3 *(The price of a forward contract). Suppose at date 0 you enter into a forward contract to buy an asset at date T with a forward price of K. We denote by $V(t, K)$ the (re-trading) price (value) of the forward contract at time $t \in [0, T]$ and by S_t the price of the asset at date t, respectively. When the contract is initiated, it is simply an agreement between the two parties taking opposite sides in the deal (one commits on buying the asset at T paying K and the other selling the asset at T for K). In other words the forward contract is a bet between two parties, and its initial cost is $V(0, K) = 0$ (because the delivery price K is set such that this holds). However, the method below will clarify the price of the forward even if initially was not zero. Clearly the payoff of the forward contract at expiration is $V(T, K) = S_T - K$. Hence at period t one can replicate the payoff of the forward contract by buying the asset and borrowing the $PV_t(K) = K/(1 + R)^{T-t}$. Such a strategy is called "cash and carry." Then $V(t, K) = S_t - PV_t(K)$.*

4.3 Options

There are two types of option contracts, call options and put options, and each type appears in two varieties, American options and European options. An American call option gives the holder the right to buy the security at a predetermined price (called strike price or exercise price), on or before the expiration (or maturity) date. An American put option gives the holder the right to sell the security at a strike price, on or before the expiration date. European call and put options are defined as their American counterpart, except that they can be exercised only at maturity. Chicago Board of Trade (CBOT) was the most important player in the option market until 2003 when it was surpassed by the International Securities Exchange (NY). There are also non-standardized options traded in the over-the-counter market. The exchanges in which options are traded jointly own the Option Clearing Corporation (OCC). OCC acts as the middleman, becoming the effective buyer of the option form the writer and the writer of the option to the buyer. Hence, OCC guarantees contract performance and therefore option writers are required to maintain a margin with OCC to insure the fulfillment of their contractual obligations. The margins are marked to market, the margin required is determined in part by the amount by which the option is in the money, meaning the size of profits produced by exercising the option. A call option is in the money when the exercise price is below the asset's value [8]. Put options are in the money when the exercise price exceeds the

[8] In the language on the next subsection, an option is in the money if it has a positive intrinsic value.

asset's value. Out of money options[9] require less margin from the writer because expected payouts are lower.

4.4 Bounds for the price of options

The purchase price of an option is called the premium. Simple arbitrage arguments enables us to derive bounds for the price at which options are traded. Later, we will refine our arbitrage arguments to derive exact prices for options. We need the following notation:

- $C(S_0, T, X)$ = price of an American call option on a stock currently priced at S_0, with maturity T and exercise price X.
- $c(S_0, T, X)$ = price of an European call option.
- $P(S_0, T, X)$ = price of an American put option.
- $p(S_0, T, X)$ = price of an European put option.
- $PV_0(T, X)$ = discounted present value of X, defined as $PV_0(T, X) = \frac{X}{(1+R)^T}$, where R is the per-period interest rate.

We start by analyzing call options. We will define the intrinsic value of a call option (at period zero) as $\max\{S_0 - X, 0\}$, which is self-explanatory.

The expiration value of the option is simply its intrinsic value at expiration, $\max\{S_T - X, 0\}$.

The following result establishes the desired bounds, in the simpler case when the stock pays no dividends.

Fact

The price of the stock is always greater than the price of the American call, which is greater than the price of the European call, which in turn is greater than the price of a forward contract with delivery price X (if this is positive), which is also greater than the intrinsic value of the option:

$$S_0 > C(S_0, T, X) \geq c(S_0, T, X) > \max\{S_0 - PV_0(T, X), 0\} > \max\{S_0 - X, 0\}.$$

Proof

Since $S_t \geq \max\{S_t - X, 0\}$ for all $t \in [0, T]$, it follows that holding the stock dominates the payoff that an American call option can provide, even if it is exercised optimally. Thus it must be the case that the stock costs more than the American call. The fact that the American call costs more than the European call is obvious, since the American call can always be exercised only at maturity. An European call also must have a non-negative price (it is a limited liability security) and it guarantees at least a payoff of zero (if it ends up out of money). Now remember from (11) that $S_0 - PV_0(T, X)$ is the price of a forward contract with delivery price X.

Since the payoff of the European call dominates the payoff of the forward, and in addition the price of the call is nonnegative, we derive the desired inequality. Finally, the last inequality is obvious.

[9] Meaning options whose exercise would be unprofitable. A call option is out of money when the exercise price exceeds the asset value, and conversely for put options.

Remark (American calls are worth more “alive” than “dead”)

A very important conclusion emerges from the previous result. The price of an American or European call is greater than the intrinsic value of the option. Hence it is never optimal to exercise prematurely an American call, since selling the call will bring in more than the payoff resulting from exercising the option (intrinsic value). Therefore the American call will be priced identically to an European call. The argument breaks down if there are dividends paid on the stock over the life of the option. It may be optimal to exercise a call option just before the dividend is paid, because the benefits of receiving the dividend may outweigh the interest lost on the strike X. Now we will derive similar bounds for put options.

Fact

The following bounds holds for the prices of American and European put options on the same stock:

$$X > P(S_0, T, X) \geq p(S_0, T, X) > \max\{PV_0(T, X) - S_0, 0\} > \max\{X - S_0, 0\}.$$

Proof

Similar to the proof of the previous fact. Do it yourself.

Remark (American put options can be exercised early)

Sometimes early exercise of an American put option can be optimal. Indeed, let the exercise price be \$105, the stock price be \$5 and the maturity be one year. Also the annual rate of interest is 10%.

Notice that by exercising the option immediately and depositing the proceeds, you will end up with $(105 - 5) \times 0.1 = 110$ at the end of the year. However the maximum possible value of the put at the expiration is $\$105$. In contrast to a call option that has an unlimited upside potential, an American put is limited upward by its strike price. Hence if this upper limit is close to the current intrinsic value of the put, it is better to exercise the option and earn interest. Finally we will present the relationship between the prices of the put and call options.

Fact(European Put-Call Parity).

If no dividends are paid on the stock until the expiration date, then $p(S_0, T, X) = c(S_0, T, X) + PV_0(T, X) - S_0$. If dividends D are paid on the stock, then denoting by $PV_0(T, D)$ the present value at 0 of the stream of dividends between 0 and T, we have

$$p(S_0, T, X) = c(S_0, T, X) + PV_0(T, X) + PV_0(T, D) - S_0.$$

Proof

The proof consists in simply noticing that the payoff at maturity of a protective put portfolio, which is composed of the asset and the put, is equal to the payoff at T of a portfolio formed by the call option and a bank deposit of $PV_0(T, X)$, and is equal to $\max\{S_T, X\}$. Similarly for the case of dividends. For American options, the parity theorem is not an exact relationship anymore. It amounts to two inequalities.

Fact(American Put-Call Parity)

If the stock pays no dividends, i) $P(S_0, T, X) = C(S_0, T, X) + PV_0(T, X) - S_0$ and ii) $P(S_0, T, X) =$

$C(S_0, T, X) + X - S_0$. If the stock pays dividends, i) remains unchanged, and in ii) we have to add $PV_0(T, D)$ to the right hand side of the inequality.

Proof

Do It Yourself.

4.5 Asset price dynamics

In the previous section we derived upper and lower bounds for option prices, which however, are not tight. Our goal is to provide a method of determining the price of an European call option on a stock. To price the call more precisely, we need to make assumptions concerning the probability distribution of the underlying stock, in other words we need to specify a model for the evolution of asset prices. The lognormal distribution provides a reasonable approximation to the actual evolution of stock price movements and provides us with a tractable framework for option pricing[10].

First we will need to explain the meaning of continuously compounded interest rates and returns. Many commercial banks use continuously compounded interest rates in quoting the interest rate on demand deposit accounts. Indeed, if the APR is given, and you know how often the interest is paid during the year (say n times), you can compute the compounded interest rate [11] as being $(1 + APR/n)n$. What happens when the frequency of compounding increases? Notice that in the limit (continuous compounding), we have

$$\lim_{n\to\infty}(1 + \frac{APR}{n})^n = e^{APR}.$$

Conversely, you can ask yourself what was the continuously compounded interest rate over the year if you observe at the end of the year an interest rate of R? As we saw, the continuously compounded interest rate r is defined implicitly as the solution to

$$1 + R = e^r,$$

or

$$r = \ln(1 + R).$$

More generally, if the interest rate over a period $[t, t + h]$ is given to be R, the continuously compounded interest rate r is defined implicitly as

$$1 + R = e^{rh}.$$

The same logic is used in defining the continuously compounded return over a given period. If the return on a stock is R from period t to $t + 1$, the continuously compounded return over period $[t, t + 1]$ will be defined as z that satisfies $e^z = 1 + R$.

We are now in a position to understand why the lognormal distribution has become the workhorse in asset pricing theory. Consider a period $[0, T]$ over which we track the price of a stock, denoted by S_t, with $t \in [0, T]$,

[10] The empirical distribution of continuously compounded returns has heavier tails than those of a normal distribution. This is inconsistent with a lognormal distribution for stock prices.

[11] Remember our discussion of the bond equivalent yield.

and assume for simplicity that the stock does not pay dividends over this period. Now divide the period into n subperiods of length h, in other words $T = nh$. Empirical studies suggest that stock returns appear to be independent and identically distributed over each subinterval. But notice that the gross stock return over the whole period is equal to the product of gross returns over each subinterval,

$$\frac{S_T}{S_0} = \frac{S_h}{S_0}\frac{S_{2h}}{S_h} \cdots \frac{S_T}{S_{T-h}}.$$

Using continuously compounded returns, that is letting

$$\begin{aligned}
\frac{S_T}{S_0} &= e^Z \\
\frac{S_h}{S_0} &= e^{z_1} \\
\frac{S_{2h}}{S_h} &= e^{z_2} \\
&\vdots \\
\frac{S_T}{S_{T-h}} &= e^{z_n},
\end{aligned}$$

we see that

$$Z = z_1 + z_2 + \cdots + z_n.$$

This is the reason why we prefer to work with continuously compounded returns, since we transfer products in sum, easier to handle. Based on the empirical considerations we mentioned above, we make the following assumptions:

A1 The returns $\{z_t\}$ are independent and identically distributed.

A2 The expected continuously compounded return can be written as

$$E[z_t] = \mu \cdot h$$

where μ is the expected continuously compounded return per unit of time. A3. The variance of the continuously compounded return is of the form $\text{var}(z_t) = \sigma^2 \cdot h$, where σ^2 is the variance of the continuously compounded return per unit of time. The last two assumptions require that the expected return and the variance of the return are proportional to the length of the time subinterval.

They insure that as the time subperiod decreases, or as n increases, the distribution of $Z = z_1 + \cdots + z_n$ does not degenerate or explode. Indeed, notice that

$$\begin{aligned}
\mathrm{E}(Z_n) &= \mathrm{E}(z_1) + \mathrm{E}(z_2) + \cdots + \mathrm{E}(z_n) = n\mu h = \mu T, \\
\text{var}(Z_n) &= \text{var}(z_1) + \text{var}(z_2) + \cdots + \text{var}(z_n) = n\sigma^2 h = \sigma^2 T.
\end{aligned}$$

Assumptions (A1)-(A3), made on any division of $[0, T]$, imply (by the Central Limit Theorem) that z_t has a normal distribution with mean μh and variance $\sigma^2 h$. Therefore also $Z = \log(S_T/S_0)$ is normally distributed with μh and variance $\sigma^2 h$. But this implies that S_T/S_0 is lognormally distributed, as desired. This imply that the expected value at 0 of the stock price at T is: [12]

$$\mathrm{E}[S_T] = S_0 e^{\mu T + \sigma^2 T/2}.$$

4.6 Binomial model of option pricing (Mathematically Advanced)

The assumption of lognormal distribution of stock prices can be used, using stochastic calculus, to derive the formula price of an European call option. To avoid technicalities, we will take a different route. We will approximate the lognormal distribution of asset prices by a discrete binomial lattice producing in the limit the lognormal distribution. We will derive by no arbitrage arguments the price of the European call in the binomial model, and see that it converges to the Black-Scholes formula when we enrich the tree progressively.

With the notation of the previous section, we assume that

$$S_{(t+1)h} = \begin{cases} S_{th}U & : \textit{ with probability } p \\ S_{th}D & : \textit{ with probability } 1-p. \end{cases}$$

In order to approximate the lognormal distribution we assume

$$\begin{aligned} U &= e^{\mu h + \sigma\sqrt{h}} \\ D &= e^{\mu h - \sigma\sqrt{h}} \\ p &= \frac{1}{2}. \end{aligned}$$

Notice that with the previous notation for continuously compounded return, $z_t = \ln \frac{S_{t+1}}{S_t}$. Thus

$$\begin{aligned} \mathrm{E}(z_t) &= \mu h \\ \mathrm{var}(z_t) &= \sigma^2 h, \end{aligned}$$

where μ is the drift parameter and σ is the volatility parameter. Notice that Lindeberg's Central Limit Theorem implies that

$$\lim_{n \to 1} \frac{z_1 + z_2 + \cdots + z_n - \mu T}{\sigma T} \to N(0,1) \tag{34}$$

or as desired,

$$Z_n = z_1 + \cdots + z_n \to N(\mu T, \sigma^2 T), \tag{35}$$

where the convergence is in distribution (i.e., weak convergence). Therefore the binomial representation of stock prices approximates a stock price process that follows a lognormal distribution. At each node of the lattice (tree), we have two branches describing the direction of stock prices movements for the next period. By bringing another asset into the picture, a riskless bond, in addition to our stock, we will be able to dynamically "complete" the markets, that is to replicate the payoff of any asset. In particular we can create a "synthetic" option having exactly

[12] Remember the expression for the characteristic function of a normal distribution.

the payoff of an option. Of course, by arbitrage arguments, the price of the synthetic option must coincide with the price of the actual option. Assume that the continuously compounded interest rate on the bond is r, and let $R_h = e^{rh}$. Now we will derive the state prices associated with our asset structure and the absence of arbitrage. At any node in the tree, we need to solve:

$$\begin{cases} 1 & : \; = \lambda_u U + \lambda_d D \\ \frac{1}{R_h} & : \; = \lambda_u + \lambda_d. \end{cases}$$

which gives:

$$\begin{aligned} \lambda_u &= \frac{1}{R_h} \cdot \frac{R_h - D}{U - D}; \\ \lambda_d &= \frac{1}{R_h} \cdot \frac{U - R_h}{U - D}. \end{aligned}$$

Notice that $U > R_h > D$ in order not to have arbitrage opportunities (an asset will dominate the other otherwise), hence the state prices are strictly positive. Let

$$\begin{aligned} q &= \frac{R_h - D}{U - D}; \\ 1 - q &= \frac{U - R_h}{U - D}. \end{aligned}$$

the risk-neutral probabilities, or the equivalent martingale probabilities. They lead to the risk neutral valuation principle under which the value of any asset is simply the discounted expected value of its payoffs (dividends). Detailed explanations will be provided in class. Indeed, assume that you have an asset (e.g. a call option) which next period pays C_u or C_d depending on the state of the stock. Then clearly the current price of a synthetic option replicating its payoffs are

$$C = \lambda_u U + \lambda_d D = \frac{1}{R_h}[qC_q + (1-q)C_d] = H \cdot S_t + b, \tag{36}$$

where H represents the number of shares of stock and b the amount of bonds held in the replicating portfolio (S_t represents the current price of the stock, in other words we look at the stock evolution from $[t, t+h]$):

$$\begin{aligned} C_u &= H \cdot S_t U + bR_h \\ C_d &= H \cdot S_t D + bR_h, \end{aligned}$$

which gives

$$\begin{aligned} H &= \frac{1}{S_t} \frac{C_u - C_d}{U - D} \\ b &= \frac{i}{R_h} \frac{C_d U - C_u D}{U - D}. \end{aligned}$$

Here, H is called the hedge ratio, or the option's delta. It represents the sensitivity of the option's price to changes is the stock price. In the binomial model, the above formula shows that the hedge ratio is equal to the ratio of range in option and stock prices. In other words a perfect hedge requires that the option and stock be held in a fraction determined by relative volatility. If the investor writes one option and holds H shares of stock, the value of the portfolio will be unaffected by stock price movements. Knowing that the payoff of an European call with strike price X is $g(S_T) = \max\{S_T - X, 0\}$, by backwards iteration, one can obtain immediately that the current price of the call is

$$C_0 = \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) q^j q^{n-j} g(S_0 U^j D^{n-j}).$$

Denote by Q_n the product risk neutral measure on the space $\{U, D\}^n$, that is

$$Qn = \underbrace{qn \otimes qn \otimes qn}_{n-\text{th times}},$$

where $q_n(U) = q$ and $q_n(D) = 1 - q$.[13] Then we can rewrite the formula for the current price of the call as

$$C_0 = \frac{1}{e^{rT}} \mathrm{E}^{Q_n}[g(S_T)].$$

Hence if we can understand the limiting distribution of S_T under Q_n, we will be able to compute the continuous limit of the call price, and hence derive the Black-Scholes formula. Notice that as before,

$$S_T = S_0 \frac{S_h}{S_0} \cdots S_T S_{T-h} = Se^{Z_n},$$

where $Z_n = z_1 + z_2 + \cdots + z_n$ and $z_t = \ln(S_{th}/S_{(t-1)h}$. Hence it is enough to determine the limiting distribution of Z_n. But notice that

$$\mathrm{var}^{Q_n}(Z_n) = n[q(1-q)(2\sigma h)^2 + (1-q)q(-2\sigma ph)^2] = 4q(1-q)\sigma^2 T. \tag{37}$$

It is easy to see[14] that $\lim_{n\to\infty} q = 1/2$, more correctly, $\lim q_n(U) = 1/2$. It is an application of Lindeberg's central limit theorem to show that Z_n converges in distribution (under Q_n) to a normally distributed random variable. We will not pursue the details here, but the intuition should be clear, as Zn is a sum of independent and identically distributed variables. The previous the previous calculation shows that $Z_n \to_d N(\tilde{\mu}, \sigma^2 T)$. Our job is to calculate $\tilde{\mu}$. Fortunately there is a simple method. Notice that

$$\frac{S_T}{S_0} = e^{Z_n} \to e^{N(\tilde{\mu}, \sigma^2 T)},$$

and hence

$$\mathrm{E}^Q[S_T] = S_0 e^N(\tilde{\mu}, \sigma^2 T/2).$$

Risk neutral valuation implies that

$$S_0 = \frac{1}{e^{rT}} \mathrm{E}^Q[S_T].$$

Hence

$$e^{-rT + \tilde{\mu} + \sigma^2 T/2} = 1,$$

[13]Remember that q and U, D depend on n, even though the notation does not stress this out.

[14]Use L'Hopital rule.

Therefore we argued that S_T is lognormal

$$Z = \log(S_T/S_0) = N((r - \frac{\sigma^2}{2})T, \sigma^2 T).$$

To derive the Black-Scholes formula, we need the following simple lemma:

Lemma 1 *Suppose that $X = N(m, s^2)$. Then*

$$\mathrm{E}[e^{aX} \infty_{X \geq k}] = e^{am + a^2 s^2/2} N(d),$$

where a is a constant, $d = (-k + m + as^2)/s$ and $N(\cdot)$ represents the cumulative distribution function of a standard normal.

Thus the price of the call is

$$\begin{aligned}
C_0 &= \frac{1}{e^{rT}} \mathrm{E}^Q[g(S_T)] \\
&= \frac{1}{e^{rT}} \mathrm{E}^Q[\max S_T - X, 0] \\
&= \frac{1}{e^{rT}} \mathrm{E}^Q[(S_0 e^Z - X) \infty_{Z \geq \ln(X/S_0)}] \\
&= \frac{1}{e^{rT}} S_0 e^{(r - \frac{\sigma^2}{2})T + \frac{\sigma^2 T}{2}} N(d_1) - \frac{X}{e^{rT}} N(d_2).
\end{aligned}$$

where

$$\begin{aligned}
d_1 &= -\frac{\ln(X/S0) + (r - \frac{\sigma^2}{2})T + \sigma^2 T}{\sigma\sqrt{T}} \\
&= \frac{\ln S_0 X/e^{rT} + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} \\
d_2 &= -\frac{\ln(X/S_0) - \tilde{\mu}}{\sigma\sqrt{T}} \\
&= d_1 + \sigma\sqrt{T}.
\end{aligned}$$