1

BASIC CONCEPTS IN FINANCE

Aims

- To consider different methods of measuring returns for pure discount bonds, coupon-paying bonds and stocks.
- Use discounted present value techniques, DPV, to price assets.
- Show how utility functions can be used to incorporate risk aversion, and derive asset demand functions from one-period utility maximisation.
- Illustrate the optimal level of physical investment and consumption for a two-period horizon problem.

The aim of this chapter is to quickly run through some of the basic tools of analysis used in finance literature. The topics covered are not exhaustive and they are discussed at a fairly intuitive level.

1.1 Returns on Stocks, Bonds and Real Assets

Much of the theoretical work in finance is conducted in terms of compound rates of return or interest rates, even though rates quoted in the market use 'simple interest'. For example, an interest rate of 5 percent payable every six months will be quoted as a simple interest rate of 10 percent per annum in the market. However, if an investor rolled over two six-month bills and the interest rate remained constant, he could actually earn a 'compound' or 'true' or 'effective' annual rate of $(1.05)^2 = 1.1025$ or 10.25 percent. The effective annual rate of return exceeds the simple rate because in the former case the investor earns 'interest-on-interest'.

We now examine how we calculate the terminal value of an investment when the frequency with which interest rates are compounded alters. Clearly, a quoted interest rate of 10 percent per annum when interest is calculated monthly will amount to more at the end of the year than if interest accrues only at the end of the year.

Consider an amount $\$A$ invested for n years at a rate of R per annum (where R is expressed as a decimal). If compounding takes place only at the end of the year, the future value after n years is FV_n, where

$$FV_n = \$A(1 + R)^n \tag{1}$$

However, if interest is paid m times per annum, then the terminal value at the end of n years is

$$FV_n^m = \$A(1 + R/m)^{mn} \tag{2}$$

R/m is often referred to as the periodic interest rate. As m, the frequency of compounding, increases, the rate becomes 'continuously compounded', and it may be shown that the investment accrues to

$$FV_n^c = \$Ae^{R_c n} \tag{3}$$

where $R_c =$ the continuously compounded rate per annum. For example, if the quoted (simple) interest rate is 10 percent per annum, then the value of \$100 at the end of one year ($n = 1$) for different values of m is given in Table 1. For *daily* compounding, with $R = 10\%$ p.a., the terminal value after one year using (2) is \$110.5155. Assuming $R_c = 10\%$ gives $FV_n^c = \$100e^{0.10(1)} = \$100.5171$. So, daily compounding is almost equivalent to using a continuously compounded rate (see the last two entries in Table 1).

We now consider how to switch between simple interest rates, periodic rates, effective annual rates and continuously compounded rates. Suppose an investment pays a periodic interest rate of 2 percent each quarter. This will usually be quoted in the market as 8 percent per annum, that is, as a simple annual rate. At the end of the year, $\$A = \$100$ accrues to

$$\$A(1 + R/m)^m = 100(1 + 0.08/4)^4 = \$108.24 \tag{4}$$

The effective annual rate R_e is 8.24% since $\$100(1 + R_e) = 108.24$. R_e exceeds the simple rate because of the payment of interest-on-interest. The relationship between

Table 1 Compounding frequency

Compounding Frequency	Value of \$100 at End of Year ($R = 10\%$ p.a.)
Annually ($m = 1$)	110.00
Quarterly ($m = 4$)	110.38
Weekly ($m = 52$)	110.51
Daily ($m = 365$)	110.5155
Continuous ($n = 1$)	110.5171

the quoted simple rate R with payments m times per year and the *effective annual rate* R_e is

$$(1 + R_e) = (1 + R/m)^m \tag{5}$$

We can use (5) to move from periodic interest rates to effective rates and vice versa. For example, an interest rate with quarterly payments that would produce an effective annual rate of 12 percent is given by $1.12 = (1 + R/4)^4$, and hence,

$$R = [(1.12)^{1/4} - 1]4 = 0.0287(4) = 11.48\% \tag{6}$$

So, with interest compounded quarterly, a simple interest rate of 11.48 percent per annum is equivalent to a 12 percent effective rate.

We can use a similar procedure to switch between a simple interest rate R, which applies to compounding that takes place over m periods, and an equivalent continuously compounded rate R_c. One reason for doing this calculation is that much of the advanced theory of bond pricing (and the pricing of futures and options) uses continuously compounded rates.

Suppose we wish to calculate a value for R_c when we know the m-period rate R. Since the terminal value after n years of an investment of $\$A$ must be equal when using either interest rate we have

$$Ae^{R_c n} = A(1 + R/m)^{mn} \tag{7}$$

and therefore,

$$R_c = m \ln[1 + R/m] \tag{8}$$

Also, if we are given the continuously compounded rate R_c, we can use the above equation to calculate the simple rate R, which applies when interest is calculated m times per year:

$$R = m(e^{R_c/m} - 1) \tag{9}$$

We can perhaps best summarise the above array of alternative interest rates by using one final illustrative example. Suppose an investment pays a periodic interest rate of 5 percent every six months ($m = 2$, $R/2 = 0.05$). In the market, this might be quoted as a 'simple rate' of 10 percent per annum. An investment of \$100 would yield $100[1 + (0.10/2)]^2 = \$110.25$ after one year (using equation 2). Clearly, the effective annual rate is 10.25% p.a. Suppose we wish to convert the simple annual rate of $R = 0.10$ to an equivalent continuously compounded rate. Using (8), with $m = 2$, we see that this is given by $R_c = 2\ln(1 + 0.10/2) = 0.09758$ (9.758% p.a.). Of course, if interest is continuously compounded at an annual rate of 9.758 percent, then \$100 invested today would accrue to $100\ e^{R_c \cdot n} = \$110.25$ in $n = 1$ year's time.

Arithmetic and Geometric Averages

Suppose prices in successive periods are $P_0 = 1$, $P_1 = 0.7$ and $P_2 = 1$, which correspond to (periodic) returns of $R_1 = -0.30$ (-30%) and $R_2 = 0.42857$ (42.857%). The *arithmetic average* return is $\overline{R} = (R_1 + R_2)/2 = 6.4285\%$. However, it would be incorrect to assume that if you have an initial wealth $W_0 = \$100$, then your final wealth after 2 periods will be $W_2 = (1 + \overline{R})W_0 = \$106.4285$. Looking at the price series it is clear that your wealth is unchanged between $t = 0$ and $t = 2$:

$$W_2 = W_0[(1 + R_1)(1 + R_2)] = \$100\ (0.70)(1.42857) = \$100$$

Now define the *geometric average* return as

$$(1 + \overline{R}_g)^2 = (1 + R_1)(1 + R_2) = 1$$

Here $\overline{R}_g = 0$, and it correctly indicates that the return on your 'wealth portfolio' $R_w(0 \to 2) = (W_2/W_0) - 1 = 0$ between $t = 0$ and $t = 2$. Generalising, the geometric average return is defined as

$$(1 + \overline{R}_g)^n = (1 + R_1)(1 + R_2) \cdots (1 + R_n) \tag{10}$$

and we can always write

$$W_n = W_0(1 + \overline{R}_g)^n$$

Unless (periodic) returns R_t are constant, the geometric average return is always less than the arithmetic average return. For example, using one-year returns R_t, the geometric average return on a US equity value weighted index over the period 1802–1997 is 7% p.a., considerably lower than the arithmetic average of 8.5% p.a. (Siegel 1998).

If returns are serially uncorrelated, $R_t = \mu + \varepsilon_t$ with $\varepsilon_t \sim iid(0, \sigma^2)$, then the arithmetic average is the best return forecast for any *randomly selected* future year. Over long holding periods, the best *forecast* would also use the arithmetic average return compounded, that is, $(1 + \overline{R})^n$. Unfortunately, the latter clear simple result does not apply in practice over long horizons, since stock returns are not *iid.*

In our simple example, if the sequence is repeated, returns are negatively serially correlated (i.e. −30%, +42.8%, alternating in each period). In this case, forecasting over long horizons requires the use of the geometric average return compounded, $(1 + \overline{R}_g)^n$. There is evidence that over long horizons stock returns are 'mildly' mean reverting (i.e. exhibit some negative serial correlation) so that the arithmetic average overstates *expected* future returns, and it may be better to use the geometric average as a *forecast* of future average returns.

Long Horizons

The (periodic) return is $(1 + R_1) = P_1/P_0$. In intertemporal models, we often require an expression for terminal wealth:

$$W_n = W_0(1 + R_1)(1 + R_2) \cdots (1 + R_n)$$

Alternatively, this can be expressed as

$$\begin{aligned} \ln(W_n/W_0) &= \ln(1 + R_1) + \ln(1 + R_2) + \cdots + \ln(1 + R_n) \\ &= (R_{c1} + R_{c2} + \cdots + R_{cn}) = \ln(P_n/P_0) \end{aligned}$$

where $R_{ct} \equiv \ln(1 + R_t)$ are the continuously compounded rates. Note that the term in parentheses is equal to $\ln(P_n/P_0)$. It follows that

$$W_n = W_0 \exp(R_{c1} + R_{c2} + \cdots + R_{cn}) = W_0(P_n/P_0)$$

Continuously compounded rates are additive, so we can *define* the (total continuously compounded) return over the whole period from $t = 0$ to $t = n$ as

$$R_c(0 \to n) \equiv (R_{c1} + R_{c2} + \cdots + R_{cn})$$
$$W_n = W_0 \exp[R_c(0 \to n)]$$

Let us now 'connect' the continuously compounded returns to the geometric average return. It follows from (10) that

$$\ln(1 + \overline{R}_g)^n = (R_{c1} + R_{c2} + \cdots + R_{cn}) \equiv R_c(0 \to n)$$

Hence

$$W_n = W_0 \exp[\ln(1 + \overline{R}_g)^n] = W_0(1 + \overline{R}_g)^n$$

as we found earlier.

Nominal and Real Returns

A number of asset pricing models focus on real rather than nominal returns. The real return is the (percent) rate of return from an investment, in terms of the purchasing power over goods and services. A real return of, say, 3% p.a. implies that your initial investment allows you to purchase 3% more of a fixed basket of domestic goods (e.g. Harrod's Hamper for a UK resident) at the end of the year.

If at $t = 0$ you have a nominal wealth W_0, then your real wealth is $W_0^r = W_0/P_o^g$, where P^g = price index for goods and services. If R = nominal (proportionate) return on your wealth, then at the end of year-1 you have nominal wealth of $W_0(1 + R)$ and real wealth of

$$W_1^r \equiv \frac{W_1}{P_1^g} = \frac{(W_0^r P_o^g)(1 + R)}{P_1^g}$$

Hence, the increase in your real wealth or, equivalently, your (proportionate) real return is

$$(1 + R^r) \equiv W_1^r/W_0^r = (1 + R)/(1 + \pi) \tag{11}$$

$$R^r \equiv \frac{\Delta W_1^r}{W_0^r} = \frac{R - \pi}{1 + \pi} \approx R - \pi \tag{12}$$

where $1 + \pi \equiv (P_1^g/P_0^g)$. The proportionate change in real wealth is your real return R^r, which is *approximately* equal to the nominal return R minus the rate of goods price inflation, π. In terms of continuously compounded returns,

$$\ln(W_1^r/W_0^r) \equiv R_c^r = \ln(1 + R) - \ln(P_1^g/P_o^g) = R_c - \pi_c \tag{13}$$

where $R_c =$ (continuously compounded) nominal return and $\pi_c =$ continuously compounded rate of inflation. Using continuously compounded returns has the advantage that the log real return over a horizon $t = 0$ to $t = n$ is additive:

$$\begin{aligned} R_c^r(0 \to n) &= (R_{c1} - \pi_{c1}) + (R_{c2} - \pi_{c2}) + \cdots + (R_{cn} - \pi_{cn}) \\ &= (R_{c1}^r + R_{c2}^r + \cdots + R_{cn}^r) \end{aligned} \tag{14}$$

Using the above, if initial real wealth is W_0^r, then the level of real wealth at $t = n$ is $W_n^r = W_0^r e^{R_c^n(0 \to n)} = W_0^r e^{(R_{c1}^r + R_{c2}^r + \cdots + R_{cn}^r)}$. Alternatively, if we use proportionate changes, then

$$W_n^r = W_0^r(1 + R_1^r)(1 + R_2^r) \cdots (1 + R_n^r) \tag{15}$$

and the *annual average geometric real return* from $t = 0$ to $t = n$, denoted $\overline{R}_{r,g}$ is given by

$$(1 + \overline{R}_{r,g}) = \sqrt[n]{(1 + R_1^r)(1 + R_2^r) \cdots (1 + R_n)^r}$$

and $W_n^r = W_0^r(1 + \overline{R}_{r,g})^n$

Foreign Investment

Suppose you are considering investing abroad. The nominal *return measured in terms of your domestic currency* can be shown to equal the foreign currency return (sometimes called the *local currency return*) plus the appreciation in the foreign currency. By investing abroad, you can gain (or lose) either from holding the foreign asset or from changes in the exchange rate. For example, consider a UK resident with initial nominal wealth W_0 who exchanges (the UK pound) sterling for USDs at a rate S_0 (£s per \$) and invests in the United States with a nominal (proportionate) return R^{us}. Nominal wealth in Sterling at $t = 1$ is

$$W_1 = \frac{W_0(1 + R^{us})S_1}{S_0} \tag{16}$$

Hence, using $S_1 = S_0 + \Delta S_1$, the (proportionate) nominal return to foreign investment for a UK investor is

$$R(UK \to US) \equiv (W_1/W_0) - 1 = R^{us} + \Delta S_1/S_0 + R^{us}(\Delta S_1/S_0) \approx R^{US} + R^{FX} \tag{17}$$

where $R^{FX} = \Delta S_1/S_0$ is the (proportionate) appreciation of FX rate of the USD against sterling, and we have assumed that $R^{us}(\Delta S_1/S_0)$ is negligible. The nominal return to foreign investment is obviously

Nominal return(UK resident) = local currency(US)return + appreciation of USD

In terms of continuously compound returns, the equation is exact:

$$R_c(UK \to US) \equiv \ln(W_1/W_0) = R_c^{us} + \Delta s \tag{18}$$

where $R_c^{us} \equiv \ln(1 + R^{us})$ and $\Delta s \equiv \ln(S_1/S_0)$. Now suppose you are concerned about the *real* return of your foreign investment, in terms of purchasing power *over domestic goods*. The real return to foreign investment is just the nominal return less the domestic rate of price inflation. To demonstrate this, take a UK resident investing in the United States, but ultimately using any profits to spend on UK goods. Real wealth at $t = 1$, in terms of purchasing power over UK goods is

$$W_1^r = \frac{(W_0^r P_o^g)(1 + R^{us})S_1}{P_1^g S_0} \tag{19}$$

It follows that the continuously compounded and proportionate real return to foreign investment is

$$R_c^r(UK \to US) \equiv \ln(W_1^r/W_0^r) = R_c^{us} + \Delta s - \pi_c^{uk} \tag{20}$$

$$R^r(UK \to US) \equiv \Delta W_1^r/W_0^r \approx R^{us} + R^{FX} - \pi^{uk} \tag{21}$$

where $\Delta s = \ln(S_1/S_0)$. Hence, the real return $R^r(UK \to US)$ to a UK resident in terms of UK purchasing power from a round-trip investment in US assets is

$$\begin{aligned}\text{Real return (UK resident)} &= \text{nominal `local currency' return in US} \\ &\quad + \text{appreciation of USD} - \text{inflation in UK}\end{aligned}$$

From (20) it is interesting to note that the *real* return to foreign investment for a UK resident $R_c^r(UK \to US)$ would equal the real return to a US resident investing in the US, $(R_c^{us} - \pi_c^{us})$ if

$$\pi_c^{uk} - \pi_c^{us} = \Delta s \tag{22}$$

As we shall see in Chapter 24, equation (22) is the relative purchasing power parity (PPP) condition. Hence, if relative PPP holds, the *real* return to foreign investment is equal to the real local currency return $R_c^{us} - \pi_c^{us}$, and the change in the exchange rate is immaterial. This is because, under relative PPP, the exchange rate alters to just offset the differential inflation rate between the two countries. As relative PPP holds only over horizons of 5–10 years, the real return to foreign investment over shorter horizons will depend on exchange rate changes.

1.2 Discounted Present Value, DPV

Let the quoted annual rate of interest on a completely safe investment over n years be denoted as r_n. The future value of $\$A$ in n years' time with interest calculated annually is

$$FV_n = \$A(1 + r_n)^n \tag{23}$$

It follows that if you were given the opportunity to receive with certainty $\$FV_n$ in n years' time, then you would be willing to give up $\$A$ today. The value *today* of a certain payment of FV_n in n years' time is \$$A$. In a more technical language, the *discounted present value* DPV of FV_n is

$$DPV = FV_n/(1 + r_n)^n \tag{24}$$

We now make the assumption that the safe interest rate applicable to $1, 2, 3, \ldots, n$ year horizons is *constant* and equal to r. We are assuming that the term structure of interest rates is flat. The DPV of a *stream* of receipts FV_i ($i = 1$ to n) that carry no default risk is then given by

$$DPV = \sum_{i=1}^{n} FV_i/(1 + r)^i \tag{25}$$

Annuities

If the future payments are *constant* in each year ($FV_i = \$C$) and the first payment is at the end of the first year, then we have an *ordinary annuity*. The DPV of these payments is

$$DPV = C \sum_{i=1}^{n} 1/(1 + r)^i \tag{26}$$

Using the formula for the sum of a geometric progression, we can write the DPV of an ordinary annuity as

$$DPV = C \cdot A_{n,r} \quad \text{where } A_{n,r} = (1/r)[1 - 1/(1 + r)^n] \tag{27}$$

and $\quad DPV = C/r \quad$ as $n \to \infty$

The term $A_{n,r}$ is called the *annuity factor*, and its numerical value is given in annuity tables for various values of n and r. A special case of the annuity formula is when n approaches infinity, then $A_{n,r} = 1/r$ and $DPV = C/r$. This formula is used to price a bond called a perpetuity or console, which pays a coupon \$$C$ (but is never redeemed by the issuers). The annuity formula can be used in calculations involving constant payments such as mortgages, pensions and for pricing a coupon-paying bond (see below).

Physical Investment Project

Consider a physical investment project such as building a new factory, which has a set of prospective net receipts (profits) of FV_i. Suppose the capital cost of the project which we assume all accrues today (i.e. at time $t = 0$) is \$$KC$. Then the entrepreneur should invest in the project if

$$DPV \geq KC \tag{28}$$

or, equivalently, if the net present value NPV satisfies

$$NPV = DPV - KC \geq 0 \tag{29}$$

If $NPV = 0$, then it can be shown that the net receipts (profits) from the investment project are just sufficient to pay back both the principal (\$$KC$) and the interest on the

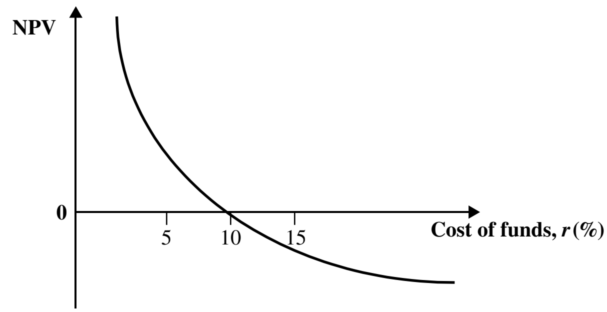


Figure 1 NPV and the discount rate

loan, which was taken out to finance the project. If $NPV > 0$, then there are surplus funds available even after these loan repayments.

As the cost of funds r increases, then the NPV falls for any given stream of profits FV_i from the project (Figure 1). There is a value of r (= 10% in Figure 1) for which the $NPV = 0$. This value of r is known as the internal rate of return IRR of the investment project. Given a stream of net receipts FV_i and the capital cost KC for a project, one can always calculate a project's IRR. It is that constant value of y for which

$$KC = \sum_{i=1}^{n} FV_i/(1+y)^i \tag{30}$$

An equivalent investment rule to the NPV condition (28) is to invest in the project if

$$IRR(= y) \geq \text{cost of borrowing } (= r) \tag{31}$$

There are some technical problems with IRR (which luckily are often not problematic in practice). First, a meaningful solution for IRR assumes all the $FV_i > 0$, and hence do not alternate in sign, because otherwise there may be more than one solution for the IRR. Second, the IRR should not be used to compare two projects as it may not give the same decision rule as NPV (see Cuthbertson and Nitzsche 2001a).

We will use these investment rules throughout the book, beginning in this chapter, with the derivation of the yield on bills and bonds and the optimal scale of physical investment projects for the economy. Note that in the calculation of the DPV, we assumed that the interest rate used for discounting the future receipts FV_i was constant for all horizons. Suppose that 'one-year money' carries an interest rate of r_1, two-year money costs r_2, and so on, then the DPV is given by

$$DPV = FV_1/(1+r_1) + FV_2/(1+r_2)^2 + \cdots + FV_n/(1+r_n)^n = \sum \delta_i FV_i \tag{32}$$

where $\delta_i = 1/(1+r_i)^i$. The r_i are known as *spot rates* of interest since they are the rates that apply to money that you lend over the periods $r_1 = 0$ to 1 year, $r_2 = 0$ to 2 years, and so on (expressed as annual compound rates). At any point in time, the relationship between the spot rates, r_i, on default-free assets and their maturity is known as the yield curve. For example, if $r_1 < r_2 < r_3$ and so on, then the yield curve is said to be upward sloping. The relationship between changes in short rates over time and changes in long rates is the subject of the term structure of interest rates.

The DPV formula can also be expressed in real terms. In this case, future receipts FV_i are deflated by the aggregate goods price index and the discount factors are calculated using real rates of interest.

In general, physical investment projects are not riskless since the future receipts are uncertain. There are a number of alternative methods of dealing with uncertainty in the DPV calculation. Perhaps, the simplest method, and the one we shall adopt, has the discount rate δ_i consisting of the risk-free spot rate r_i plus a risk premium rp_i.

$$\delta_i = (1 + r_i + rp_i)^{-1} \tag{33}$$

Equation (33) is an identity and is not operational until we have a model of the risk premium. We examine alternative models for risk premia in Chapter 3.

Stocks

The difficulty with direct application of the DPV concept to stocks is that future dividends are uncertain and the discount factor may be time varying. It can be shown (see Chapter 4) that the *fundamental value* V_t is the expected DPV of future dividends:

$$V_t = E_t\left[\frac{D_{t+1}}{(1+q_1)} + \frac{D_{t+2}}{(1+q_1)(1+q_2)} + \cdots\right] \tag{34}$$

where q_i is the one-period return between time period $t+i-1$ and $t+i$. If there are to be no systematic profitable opportunities to be made from buying and selling shares between well-informed rational traders, then the actual market price of the stock P_t must equal the fundamental value V_i. For example, if $P_t < V_t$, then investors should purchase the undervalued stock and hence make a capital gain as P_t rises towards V_t. In an efficient market, such profitable opportunities should be immediately eliminated.

Clearly, one cannot directly calculate V_t to see if it does equal P_t because expected dividends (and discount rates) are unobservable. However, in later chapters, we discuss methods for overcoming this problem and examine whether the stock market is efficient in the sense that $P_t = V_t$. If we add some simplifying assumptions to the DPV formula (e.g. future dividends are expected to grow at a constant rate g and the discount rate $q = R$ is constant each period), then (34) becomes

$$V_0 = D_o(1+g)/(R-g) \tag{35}$$

which is known as the *Gordon Growth Model*. Using this equation, we can calculate the ‘fair value’ of the stock and compare it to the quoted market price P_0 to see whether the share is over- or undervalued. These models are usually referred to as dividend valuation models and are dealt with in Chapter 10.

Pure Discount Bonds and Spot Yields

Instead of a physical investment project, consider investing in a pure discount bond (zero coupon bond). In the market, these are usually referred to as ‘zeros’. A pure discount bond has a fixed redemption price M, a known maturity period and pays no coupons. The yield on the bond if held to maturity is determined by the fact that it is purchased at a market price P_t below its redemption price M. For a one-year bond, it seems sensible to calculate the yield or interest rate as

$$r_{1t} = (M_1 - P_{1t})/P_{1t} \tag{36}$$

where r_{1t} is measured as a proportion. However, when viewing the problem in terms of DPV, we see that the one-year bond promises a future payment of M_1 at the end of the year in exchange for a capital cost of P_{1t} paid out today. Hence the IRR, y_{1t}, of the bond can be calculated from

$$P_{1t} = M_1/(1 + y_{1t}) \tag{37}$$

But on rearrangement, we have $y_{1t} = (M_1 - P_{1t})/P_{1t}$, and hence the one-year spot yield r_{1t} is simply the IRR of the bill. Applying the above principle to a two-year bill with redemption price M_2, the annual (compound) interest rate r_{2t} on the bill is the solution to

$$P_{2t} = M_2/(1 + r_{2t})^2 \tag{38}$$

which implies

$$r_{2t} = (M_2/P_{2t})^{1/2} - 1 \tag{39}$$

If spot rates are continuously compounded, then

$$P_{nt} = M_n e^{-r_{nt}n} \tag{40}$$

where r_{nt} is now the continuously compounded rate for a bond of maturity n at time t. We now see how we can, in principle, calculate a set of (compound) spot rates at t for different maturities from the market prices at time t of pure discount bonds (bills).

Coupon-Paying Bonds

A *level coupon* (non-callable) bond pays a fixed coupon \$$C$ at known fixed intervals (which we take to be every year) and has a fixed redemption price M_n payable when the bond matures in year n. For a bond with n years left to maturity, the current market price is P_{nt}. The question is how do we measure the return on the bond if it is held to maturity?

The bond is analogous to our physical investment project with the capital outlay today being P_{nt} and the future receipts being \$$C$ each year (plus the redemption price). The internal rate of return on the bond, which is called the yield to maturity y_t, can be calculated from

$$P_{nt} = C/(1 + y_t) + C/(1 + y_t)^2 + \cdots + (C + M_n)/(1 + y_t)^n \tag{41}$$

The yield to maturity is that *constant* rate of discount that at a point in time equates the DPV of future payments with the current market price. Since P_{nt}, M_n and C are the known values in the market, (41) has to be solved to give the quoted rate for the yield to maturity y_t. There is a subscript 't' on y_t because as the market price falls, the yield to maturity rises (and vice versa) as a matter of 'actuarial arithmetic'. Although widely used in the market and in the financial press, there are some theoretical/conceptual problems in using the yield to maturity as an unambiguous measure of the return on a bond even when it is held to maturity. We deal with some of these issues in Part III.

In the market, coupon payments C are usually paid every six months and the interest rate from (41) is then the periodic six-month rate. If this periodic yield to maturity is calculated as, say, 6 percent, then in the market the quoted yield to maturity will be the simple annual rate of 12 percent per annum (known as the bond-equivalent yield in the United States).

A *perpetuity* is a level coupon bond that is never redeemed by the primary issuer (i.e. $n \to \infty$). If the coupon is \$$C$ per annum and the current market price of the bond is $P_{\infty,t}$, then from (41) the yield to maturity on a perpetuity is

$$y_{\infty,t} = C/P_{\infty,t} \tag{42}$$

It is immediately obvious from (42) that for small changes, the percentage change in the price of a perpetuity equals the percentage change in the yield to maturity. The flat yield or interest yield or running yield $y_{rt} = (C/P_{nt})100$ and is quoted in the financial press, but it is not a particularly theoretically useful concept in analysing the pricing and return on bonds.

Although compound rates of interest (or yields) are quoted in the markets, we often find it more convenient to express bond prices in terms of continuously compounded spot interest rates/yields. If the continuously compounded spot yield is r_{nt}, then a coupon-paying bond may be considered as a portfolio of 'zeros', and the price is (see Cuthbertson and Nitzsche 2001a)

$$P_{nt} = \sum_{k=1}^{n} C_k e^{-r_{kt}k} + M_n e^{-r_{nt}n} = \sum_{k=1}^{n} P_{kt}^* + P_{nt}^* \tag{43}$$

where $P_k^* = C_k e^{-r_k k}$ and P_n^* are the prices of zero coupon bonds paying C_k at time $t + k$ and M_n at time $t + n$, respectively.

Holding Period Return

Much empirical work on stocks deals with the one-period holding period return H_{t+1}, which is defined as

$$H_{t+1} = \frac{P_{t+1} - P_t}{P_t} + \frac{D_{t+1}}{P_t} \tag{44}$$

The first term is the proportionate capital gain or loss (over one period) and the second term is the (proportionate) dividend yield. H_{t+1} can be calculated *ex-post* but, of course, viewed from time t, P_{t+1} and (perhaps) D_{t+1} are uncertain, and investors can only try and forecast these elements. It also follows that

$$1 + H_{t+i+1} = [(P_{t+i+1} + D_{t+i+1})/P_{t+i}] \tag{45}$$

where H_{t+i} is the one-period return between $t+i$ and $t+i+1$. Hence ex-post if \$$A$ is invested in the stock (and all dividend payments are reinvested in the stock), then the \$$Y$ payout after n periods is

$$Y = A[1 + H_{t+1}][1 + H_{t+2}] \cdots [1 + H_{t+n}] \tag{46}$$

The continuously compounded holding period return (or 'log-return') is defined as

$$h_{t+1} = \ln(P_{t+1}/P_t) = p_{t+1} - p_t \tag{47}$$

The continuously compounded return over period t to $t+n$ is

$$h_{t+n} = p_{t+n} - p_t = h_t + h_{t+1} + \cdots + h_{t+n} \tag{48}$$

Throughout the book, we will demonstrate how expected one-period returns H_{t+1} can be directly related to the DPV formula. Much of the early empirical work on whether the stock market is efficient centres on trying to establish whether one-period returns H_{t+1} are predictable. Later empirical work concentrated on whether the stock price equalled the DPV of future dividends, and the most recent empirical work brings together these two strands in the literature.

With slight modifications, the one-period holding period return can be defined for any asset. For a coupon-paying bond with initial maturity of n periods and coupon payment of C, we have

$$H_{n,t+1} = [(P_{n-1,t+1} - P_{n,t})/P_{n,t}] + C/P_{n,t} \tag{49}$$

and is also often referred to as the (one-period) holding period yield HPY. Note that the n-period bond becomes an $n-1$ period bond at $t+1$. The first term is the capital gain on the bond and the second is the coupon (or running) yield. For a zero coupon bond, $C = 0$. In models of the term structure, we usually use *continuously compounded* returns on *zero coupon* bonds, and hence h_{t+1} is given by

$$h_{t+1} = p_{n-1,t} - p_{n,t} \tag{50}$$

Often, we can apply the same type of economic model to explain movements in holding period returns for both stock and bonds (and other speculative assets), and we begin this analysis with the Capital Asset Pricing Model (CAPM) in the next chapter.

1.3 Utility and Indifference Curves

In this section, we briefly discuss the concept of utility but only to a level such that the reader can follow the subsequent material on portfolio choice and stochastic discount factor (SDF) models. Economists frequently set up portfolio models in which the individual chooses a set of assets in order to maximise some function of terminal wealth or portfolio return or consumption. For example, a certain level of wealth will imply a certain level of satisfaction for the individual as he contemplates the goods and services he could purchase with the wealth. If we double his wealth, we may not double his level of satisfaction. Also, for example, if the individual consumes one bottle of wine per night, the additional satisfaction from consuming an extra bottle may not be as great as from the first. This is the assumption of diminishing marginal utility. Utility theory can also be applied to decisions involving uncertain (strictly 'risky') outcomes. In fact, we can classify investors as 'risk averters', 'risk lovers' or 'risk neutral' in terms of the shape of their utility function. Finally, we can also examine how individuals might evaluate 'utility', which arises at different points in time, that is, the concept of discounted utility in a multiperiod or intertemporal framework.

Fair Lottery

A fair lottery (game) is defined as one that has an expected value of zero (e.g. tossing a coin with \$1 for a win (heads) and $-\$1$ for a loss (tails)). Risk aversion implies that the individual would not accept a 'fair' lottery, and it can be shown that this implies a concave utility function over wealth. Consider the random payoff x:

$$x = \begin{cases} k_1 \text{ with probability } p \\ k_2 \text{ with probability } 1-p \end{cases} \tag{51}$$

A fair lottery must have an expected value of zero

$$E(x) = pk_1 + (1-p)k_2 = 0 \tag{52}$$

which implies $k_1/k_2 = -(1-p)/p$ or $p = -k_2/(k_1 - k_2)$. For our 'coin toss', $p = 1/2, k_1 = -k_2 = \$1$.

Expected Utility

Suppose a random variable end-of-period wealth W can have n possible values W_i with probability p_i $\left(\sum_{i=1}^{n} p_i = 1\right)$. The utility from any wealth outcome W_i is denoted $U(W_i)$, and the expected utility from the risky outcomes is

$$E[U(W)] = \sum_{i=1}^{n} p_i U(W_i) \tag{53}$$

Uncertainty and Risk

The first restriction placed on utility functions is that more is always preferred to less so that $U'(W) > 0$, where $U'(W) = \partial U(W)/\partial W$. Now, consider a simple gamble of receiving \$16 for a 'head' on the toss of a coin and \$4 for tails. Given a fair coin, the probability of a head is $p = 1/2$ and the expected monetary value of the risky outcome is \$10:

$$EW = pW_H + (1-p)W_T = (1/2)16 + (1/2)4 = \$10 \tag{54}$$

We can see that the game is a fair bet when it costs $c = \$10$ to enter, because then $E(x) = EW - c = 0$. How much would an individual pay to play this game? This depends on the individual's attitude to risk. If the individual is willing to pay \$10 to play the game, so that she accepts a fair bet, we say she is *risk neutral*. If you dislike risky outcomes, then you would prefer to keep your \$10 rather than gamble on a fair game (with *expected value* of \$10) – you are then said to be *risk averse*. A risk lover would pay more than \$10 to play the game.

Risk aversion implies that the second derivative of the utility function is negative, $U''(W) < 0$. To see this, note that the utility from keeping your \$10 and not gambling is $U(10)$ and this must exceed the expected utility from the gamble:

$$U(10) > 0.5U(16) + 0.5U(4) \quad \text{or} \quad U(10) - U(4) > U(16) - U(10) \tag{55}$$

so that the utility function has the concave shape, marked 'risk averter', as given in Figure 2. An example of a utility function for a risk-averse person is $U(W) = W^{1/2}$. Note that the above example fits into our earlier notation of a fair bet if x is the risky outcome with $k_1 = W_H - c = 6$ and $k_2 = W_T - c = -6$, because then $E(x) = 0$.

We can demonstrate the concavity proposition in reverse, namely, that concavity implies an unwillingness to accept a fair bet. If z is a random variable and $U(z)$ is concave, then from Jensen's inequality:

$$E\{U(z)\} < U[E(z)] \tag{56}$$

Let $z = W + x$ where W is now the *initial* wealth, then for a fair gamble, $E(x) = 0$ so that

$$E\{U(W + x)\} < U[E(W + x)] = U(W) \tag{57}$$

and hence you would not accept the fair bet.

It is easy to deduce that for a risk lover the utility function over wealth is convex (e.g. $U = W^2$), while for a risk-neutral investor who is just indifferent between the gamble or the certain outcome, the utility function is linear (i.e. $U(W) = bW$, with $b > 0$). Hence, we have

$$U''(W) < 0 \text{ risk averse}; U''(W) = 0 \text{ risk neutral}; U''(W) > 0 \text{ risk lover}$$

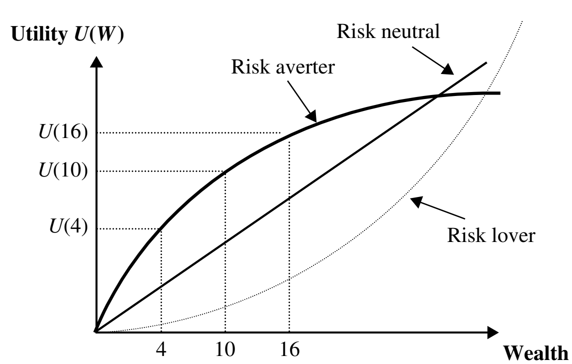


Figure 2 Utility functions

A risk-averse investor is also said to have diminishing marginal utility of wealth: each additional unit of wealth adds less to utility, the higher the initial level of wealth (i.e. $U''(W) < 0$). The degree of risk aversion is given by the concavity of the utility function in Figure 2 and equivalently by the absolute size of $U''(W)$. Note that the degree of risk aversion even for a specific individual may depend on initial wealth and on the size of the bet. An individual may be risk-neutral for the small bet above and would be willing to pay \$10 to play the game. However, a bet of \$1 million for 'heads' and \$0 for tails has an expected value of \$500,000, but this same individual may not be prepared to pay \$499,000 to avoid the bet, even though the game is in his or her favour – this same person is risk-averse over large gambles. Of course, if the person we are talking about is Bill Gates of Microsoft, who has rather a lot of initial wealth, he may be willing to pay up to \$500,000 to take on the second gamble.

Risk aversion implies concavity of the utility function, over-risky gambles. But how do we quantify this risk aversion in monetary terms, rather than in terms of utility? The answer lies in Figure 3, where the distance π is the *known* maximum amount you would be willing to pay to avoid a fair bet. If you pay π, then you will receive the *expected value* of the bet of \$10 for certain and end up with $\$(10 - \pi)$. Suppose the utility function of our risk-averse investor is $U(W) = W^{1/2}$. The expected utility from the gamble is

$$E[U(W)] = 0.5U(W_{\mathrm{H}}) + 0.5U(W_{\mathrm{T}}) = 0.5(16)^{1/2} + 0.5(4)^{1/2} = 3$$

Note that the expected utility from the gamble $E[U(W)]$ is less than the utility from the certain outcome of not playing the game $U(EW) = 10^{1/2} = 3.162$. Would our risk-averse investor be willing to pay $\pi = \$0.75$ to avoid playing the game? If she does so, then her certain utility would be $U = (10 - 0.75)^{1/2} = 3.04$, which exceeds the expected utility from the bet $E[U(W)] = 3$, so she would pay \$0.75 to avoid playing. What is the maximum insurance premium π that she would pay? This occurs when the certain utility $U(W - \pi)$ from her lower wealth $(W - \pi)$ just equals $E[U(W)] = 3$, the expected utility from the gamble:

$$U(W - \pi) = (10 - \pi)^{1/2} = E[U(W)] = 3$$

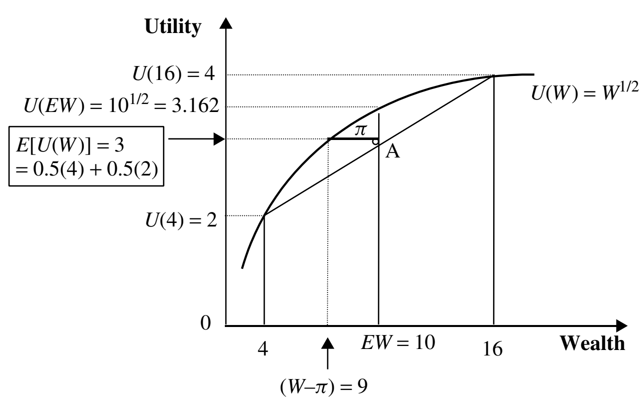


Figure 3 Monetary risk premium

which gives the maximum amount $\pi = \$1$ that you would pay to avoid playing the game. The amount of money π is known as the risk premium and is the maximum insurance payment you would make to avoid the bet (note that 'risk premium' has another meaning in finance, which we meet later – namely, the expected return on a risky asset in excess of the risk-free rate). A *fair bet* of plus or minus $\$x = 6$ gives you expected utility at point A. If your wealth is reduced to $(W - \pi)$, then the level of utility is $U(W - \pi)$. The risk premium π is therefore defined as

$$U(W - \pi) = E\{U(W + x)\} \tag{58}$$

where W is *initial* wealth. To see how π is related to the curvature of the utility function, take a Taylor series approximation of (58) around the point $x = 0$ (i.e. the probability density is concentrated around the mean of zero) and the point $\pi = 0$:

$$\begin{aligned} U(W - \pi) &\approx U(W) - \pi U'(W) \\ &= E\{U(W + x)\} \approx E\{U(W) + xU'(W) + (1/2)x^2 U''(W)\} \\ &= U(W) + (1/2)\sigma_x^2 U''(W) \end{aligned} \tag{59}$$

Because $E(x) = 0$, we require three terms in the expansion of $U(W + x)$. From (59), the risk premium is

$$\pi = -\frac{1}{2}\sigma_x^2 \frac{U''(W)}{U'(W)} = \frac{1}{2}\sigma_x^2 R_A(W) \tag{60}$$

where $R_A(W) = -U''(W)/U'(W)$ is the Arrow (1970)–Pratt (1964) measure of absolute (local) risk aversion. The measure of risk aversion is 'local' because it is a function of the initial level of wealth.

Since σ_x^2 and $U'(W)$ are positive, $U''(W) < 0$ implies that π is positive. Note that the amount you will pay to avoid a fair bet depends on the riskiness of the outcome σ_x^2 as well as *both* $U''(W)$ and $U'(W)$. For example, you may be very risk-averse ($-U''(W)$ is large) but you may not be willing to pay a high premium π, if you are also very poor, because then $U'(W)$ will also be high. In fact, two measures of the degree of risk aversion are commonly used:

$$R_A(W) = -U''(W)/U'(W) \tag{61}$$

$$R_R(W) = R_A(W)W \tag{62}$$

$R_A(W)$ is the Arrow–Pratt measure of (local) absolute risk aversion, the larger $R_A(W)$ is, the greater the degree of risk aversion. $R_R(W)$ is the coefficient of relative risk aversion. R_A and R_R are measures of how the investor's risk preferences change with a change in wealth around the initial ('local') level of wealth.

Different mathematical functions give rise to different implications for the form of risk aversion. For example, the function $U(W) = \ln(W)$ exhibits diminishing absolute risk aversion and constant relative risk aversion (see below). Now, we list some of the 'standard' utility functions that are often used in the asset pricing and portfolio literature.

Power (Constant Relative Risk Aversion)

With an initial (safe) level of wealth W_0, a utility function, which relative to the starting point has the property $U(W)/U(W_0) = f(W/W_0)$ so that utility reacts to the *relative* difference in wealth, is of the relative risk aversion type. The latter condition is met by power utility, where the response of utility to W/W_0 is constant, hence the equivalent term *constant relative risk aversion CRRA* utility function:

$$
\begin{aligned}
U(W) &= \frac{W^{(1-\gamma)}}{1-\gamma} \qquad \gamma > 0, \gamma \neq 1 \\
U'(W) &= W^{-\gamma} \qquad U''(W) = -\gamma W^{-\gamma-1} \\
R_{\mathrm{A}}(W) &= \gamma/W \quad \text{and} \quad R_{\mathrm{R}}(W) = \gamma (\text{a constant})
\end{aligned} \tag{63}
$$

Since $\ln\,[U'(W)] = -\gamma \ln W$, then γ is also the elasticity of *marginal* utility with respect to wealth.

Logarithmic

As $\gamma \to 1$ in (63), it can be shown that the limiting case of power utility is logarithmic.

$$U(W) = \ln(W) \quad \text{and} \quad R_{\mathrm{R}}(W) = 1 \tag{64}$$

This has the nice simple intuitive property that your satisfaction (utility) doubles each time you double your wealth.

Quadratic

$$
\begin{aligned}
U(W) &= W - \frac{b}{2} W^2 \qquad b > 0 \\
U'(W) &= 1 - bW \qquad U''(W) = -b \\
R_{\mathrm{A}}(W) &= b/(1-bW) \quad \text{and} \quad R_{\mathrm{R}}(W) = bW/(1-bW)
\end{aligned} \tag{65}
$$

Since $U'(W)$ must be positive, the quadratic is only defined for $W < 1/b$, which is known as the 'bliss point'. Marginal utility is linear in wealth and this can sometimes be a useful property. Note that both R_{R} and R_{A} are not constant but functions of wealth.

Negative Exponential (Constant Absolute Risk Aversion)

With an initial (safe) level of wealth W_0, a utility function, which relative to the starting point has the property $U(W)/U(W_0) = f(W - W_0)$ so that utility reacts to the *absolute* difference in wealth, is of the absolute risk aversion type. The only (acceptable) function meeting this requirement is the (negative) exponential, where the response of utility to changes in $W - W_0$ is constant, hence the term *constant absolute risk aversion CARA* utility function:

$$U(W) = a - be^{-cW} \quad c > 0$$
$$R_A(W) = c \qquad \text{and } R_R(W) = cW \tag{66}$$

It can be shown that the negative exponential utility function plus the assumption of normally distributed asset returns allows one to reduce the problem of maximising expected utility to a problem involving only the maximisation of *a linear function* of expected portfolio return ER_p and risk, that is (unambiguously) represented by the variance σ_p^2. Then, maximising the above CARA utility function $E[U(W)]$ is equivalent to maximising

$$ER_p - (c/2)\sigma_p^2 \tag{67}$$

where $c =$ the constant coefficient of absolute risk aversion. Equation (67) depends only on the mean and variance of the return on the portfolio: hence the term mean-variance criterion. However, the reader should note that, in general, maximising $E[U(W)]$ cannot be reduced to a maximisation problem in terms of a *general function* ER_p and σ_p^2 only (see Appendix), and only for the negative exponential can it be reduced to maximising a *linear* function. Some portfolio models assume at the outset that investors are only concerned with the mean-variance maximand and they, therefore, discard any direct link with a specific utility function.

HARA (Hyperbolic Absolute Risk Aversion)

$$U(W) = \frac{1-\gamma}{\gamma}\left(\frac{\alpha W}{1-\gamma} + \beta\right)^{\gamma} \tag{68}$$

$$R_A(W) = \left(\frac{W}{1-\gamma} + \frac{\beta}{\alpha}\right)^{-1} \tag{69}$$

$$R_A(W) > 0 \quad \text{when} \quad \gamma > 1, \beta > 0$$

The restrictions are $\gamma \neq 1$, $[\alpha W/(1-\gamma)] + \beta > 0$, and $\alpha > 0$. Also $\beta = 1$ if $\gamma = -\infty$. HARA (Hyperbolic Absolute Risk Aversion) is of interest because it nests constant absolute risk aversion ($\beta = 1, \gamma = -\infty$), constant relative risk aversion ($\gamma < 1, \beta = 0$) and quadratic ($\gamma = 2$), but it is usually these special cases that are used in the literature.

1.4 Asset Demands

Frequently, we want to know what determines the optimal demand for stocks, bonds and other assets, in investors' portfolios. Not surprisingly, the answer depends on how we set up the maximisation problem and the constraints facing the investor. Here we concentrate on one-period models with relatively simple solutions – later chapters deal with more complex cases.

Mean-Variance Optimisation

The simplest model is to assume investors care only about one-period expected portfolio returns and the standard deviation (risk) of these portfolio returns. Let $\alpha =$ proportion of initial wealth W_0 held in the *single* risky asset with return R and $(1 - \alpha) =$ amount held in the risk-free asset with return r. The budget constraint (with zero labour income) is

$$W_1 = (\alpha W_0)(1 + R) + [(1 - \alpha)W_0](1 + r)$$

and therefore the return and variance of the portfolio are

$$R_{\mathrm{p}} \equiv \frac{W_1}{W_0} - 1 = \alpha(R - r) + r$$
$$\sigma_{\mathrm{p}} = \alpha\sigma_{\mathrm{R}}$$

where σ_{R} is the standard deviation of the only stochastic variable R. Investors are assumed to maximise

$$\max_{\alpha} \theta = ER_{\mathrm{p}} - \frac{c}{2}\sigma_{\mathrm{p}}^2$$

where $c > 0$ is a measure of risk aversion (more precisely, the trade-off between expected portfolio return and the variance of portfolio returns). The first-order condition FOC is

$$ER - r - \alpha c\sigma_{\mathrm{R}} = 0$$

so that the optimal share of the risky asset is independent of wealth:

$$\alpha^* = \frac{(ER - r)}{c\sigma_{\mathrm{R}}}$$

Hence, the *absolute* (dollar) amount held in the risky asset $A_0 = \alpha^* W_0$ is proportional to initial wealth, and is positively related to the excess return on the risky asset and inversely related to the degree of risk aversion and the volatility of the risky asset. The share of the risk-free asset is simply $(1 - \alpha^*) \equiv A_{of}/W_0$. The above is Tobin's (1956) mean-variance model of asset demands, and the reason for the simple closed form solution is that the maximand is quadratic in α (because $\sigma_{\mathrm{p}}^2 = \alpha^2\sigma_{\mathrm{R}}^2$). If we had included *known* non-stochastic labour income y in the budget constraint, this would not alter the solution. This one-period model is sometimes used in the theoretical literature because it is linear in expected returns, which provides analytic tractability.

The mean-variance approach is easily extended to n-risky assets $R = (R_1, R_2, \ldots, R_n)'$, and the maximand is

$$\max_{\alpha} \theta = \alpha'(R - r.e) + r - \frac{c}{2}\alpha'\Omega\alpha$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$, e is an $n \times 1$ column vector of ones and $\Omega = (n \times n)$ variance–covariance matrix of returns. The FOCs give

$$\alpha^* = (c\Omega)^{-1}(ER - r.e)$$

and the share of the risk-free asset is $\alpha_{\mathrm{f}}^* = 1 - \sum_{i=1}^{n} \alpha_i^*$. For two risky assets,

$$\Omega^{-1} = (\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21})^{-1} \begin{pmatrix} \sigma_{22} & -\sigma_{21} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix}$$

and therefore the relative weights attached to the expected returns $(ER_1 - r)$ and $(ER_2 - r)$ depend on the individual elements of the variance–covariance matrix of returns. The second-order conditions guarantee that $\partial\alpha_i^*/\partial ER_i > 0$ $(i = 1$ *or* $2)$.

Negative Exponential Utility

It is worth noting that the maximand θ does not, in general, arise from a second-order Taylor series expansion of an arbitrary utility function depending only on terminal wealth $U(W_1)$. The latter usually gives rise to a *non-linear function* $EU(W) = U(ER_{\mathrm{p}}) + \frac{1}{2}\sigma_{\mathrm{p}}^2 U''(ER_{\mathrm{p}})$, whereas the mean-variance approach is *linear* in expected return and variance. However, there is one (very special) case where the maximand θ can be directly linked to a specific utility function, namely,

$$\max_{\alpha} E[U(W)] = -E\{\exp(-bW_1)\} = -E\{\exp(-bW_0(1 + R_{\mathrm{p}}))\}$$
$$\text{subject to } R_{\mathrm{p}} \equiv (W_1/W_0) - 1 = \alpha'(R - r.e) + r$$

where b is the constant coefficient of *absolute* risk aversion. Thus, the utility function must be the negative exponential (in end-of-period wealth, W_1), and as we see below, asset returns must also be multivariate normal. If a random variable x is normally distributed, $x \sim N(\mu, \sigma^2)$, then $z = \exp(x)$ is lognormal. The expected value of z is

$$Ez = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

In our case, $\mu \equiv R_{\mathrm{p}}$ and $\sigma^2 \equiv \mathrm{var}(R_{\mathrm{p}})$. The maximand is monotonic in its exponent, therefore, max $E[U(W)]$ is equivalent to

$$\max_{\alpha} E[U(W_1)] = \alpha'(ER - r.e) - \frac{1}{2}bW_0\alpha'\Omega\alpha$$

where we have discarded the non-stochastic term $\exp(-bW_0)$. The maximand is now linearly related to expected portfolio return and variance. The solution to the FOCs is

$$\alpha^* = (bW_0\,\Omega)^{-1}(ER - r.e)$$

This is the same form of solution as for the mean-variance case and is equivalent if $c = bW_0$. Note, however, that the asset demand functions derived from the negative exponential utility function imply that the absolute dollar amount $A = \alpha^* W_0$ invested in the risky asset is independent of initial wealth. Therefore, if an individual obtains additional wealth next period, then she will put all of the extra wealth into the risk-free asset – a somewhat implausible result.

Quadratic Utility

In this section, we will outline how asset shares can be derived when the utility function is quadratic – the math gets rather messy (but not difficult) and therefore we simplify the notation (as in Cerny 2004). We assume one risky asset and a risk-free asset. The budget constraint is

$$W = \tilde{\alpha} W_0 R^* + (1 - \tilde{\alpha}) W_0 R_{\mathrm{f}}^* + y$$

where $\tilde{\alpha} = A/W_0$ is the risky asset share, y is the known labour income and $R^* = 1 + R$ is the *gross* return on the risky asset. The risk-free asset share is $\tilde{\alpha}_{\mathrm{f}} = 1 - \tilde{\alpha}$. After some rearrangement, the budget constraint becomes

$$W = W_{\mathrm{safe}}(1 + \alpha X)$$

where $\quad W_{\mathrm{safe}} \equiv R_{\mathrm{f}}^* W_0 + y$, $X = R - R_f$ and $\alpha = \tilde{\alpha} W_0 / W_{\mathrm{safe}}$

Hence, α is just a scaled version of $\tilde{\alpha}$. The utility function is quadratic with a fixed 'bliss level' of wealth W_{bliss}:

$$U(W) = -\frac{1}{2}(W - W_{\mathrm{bliss}})^2 = -\frac{1}{2}(W^2 - 2W W_{\mathrm{bliss}} + W_{\mathrm{bliss}}^2)$$

We assume investors are always below the bliss point. It is easy to see from the above that $E[U(W)]$ depends on α, α^2, EX and $E(X^2)$. The FOCs will therefore be linear in α. In addition, $E(X^2) \equiv \mathrm{var}(X) - (EX)^2$ so that the optimal α will depend only on expected excess returns EX and the variance of returns on the risky assets (but the relationship is not linear). Substituting the budget constraint in $E[U(W)]$ and solving the FOC with respect to α gives (after tedious algebra)

$$\alpha^* = q_k \frac{EX}{E(X^2)}$$

where $q_k = 2k(1-k)/2k^2$ and $k = W_{\mathrm{safe}}/W_{\mathrm{bliss}}$.

Note that no *explicit* measure of risk aversion appears in the equation for α^* but it is implicit in the squared term '2' in the utility function, and is therefore a scaling factor in the solution for α^*. (Also see below for the solution with power utility that collapses to quadratic utility for $\gamma = -1$.)

In order that we do not exceed the bliss point, we require $k \ll 1$ and to simplify the algebra, take $k = 1/2$ so that $q_k = 1$ (Equation 3.53, p. 68 in Cerny 2004). Hence,

$$\alpha^* = \frac{EX}{E(X^2)} = \frac{\mu_x}{\sigma_x^2 + \mu_x^2} = \frac{1}{\mu_x(1 + 1/SR_x^2)}$$

where $SR_x = \mu_x/\sigma_x$ is known as the Sharpe ratio, and appears through the book as a measure of return per unit of risk (reward-to-risk ratio). Here the optimal α (and $\tilde{\alpha}$) is directly related to the Sharpe ratio. It can be shown that α^* for quadratic utility (and for $W(\alpha^*) < W_{\mathrm{bliss}}$) is also that value that gives the maximum Sharpe ratio (and this generalises when there are many risky assets, (see Cerny 2004, Chapter 4)).

Choosing a portfolio (i.e. $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$) to maximise the Sharpe ratio can therefore be directly linked to maximising expected quadratic utility (for $W(\alpha^*) < W_{\text{bliss}}$) – although the Sharpe ratio criterion would not be valid if the optimal α implies that $W(\alpha^*) > W_{\text{bliss}}$. The link between the (basic) Sharpe ratio and utility cannot be established for other types of utility function. Nevertheless, the Sharpe ratio is often used in practice to rank alternative risky portfolios without trying to link the decision to any specific utility function – as we see in later chapters. Also, if the 'basic' Sharpe ratio above can be generalised to link it to a constant CARA utility function, then it can be referred to as the *Hodges ratio* (see Cerny 2004, Hodges 1998).

Power Utility

A closed form solution for asset shares for most utility functions is not possible. We then have to use numerical techniques. We demonstrate this outcome for power utility over (one period) final wealth for one risky and one risk-free asset. The budget constraint is

$$W(\tilde{\alpha}) = \tilde{\alpha} W_0 R^* + (1 - \tilde{\alpha}) W_0 R_f^* + y = W_{\text{safe}}(1 + \alpha X)$$

Suppose we have a simple set-up where $R_u^* = 1.20$ and $R_D^* = 0.90$ so that the risky asset has only two possible outcomes, up or down (i.e. 20% or −10%), with equal probability of 1/2. Let $\gamma = 5$, $r = 0.03$, $W_0 = \$1\,m$ and $y = \$200{,}000$. The maximisation problem with power utility is then

$$\max_{\alpha} \theta = E\left[\frac{W^{1-\gamma}}{(1-\gamma)}\right] = \left\{\frac{1}{2}\frac{W_u^{1-\gamma}}{(1-\gamma)} + \frac{1}{2}\frac{W_D^{1-\gamma}}{(1-\gamma)}\right\}$$

$$= W_{\text{safe}}^{1-\gamma}\left\{\frac{1}{2}\frac{(1+\alpha X_u)}{(1-\gamma)} + \frac{1}{2}\frac{(1+\alpha X_u)^{1-\gamma}}{(1-\gamma)}\right\}$$

Everything in θ is known except for α. Actually, an analytic solution for this case is possible. We would set $\partial\theta/\partial\alpha = 0$, and the resulting equation is solved for α^* and then for $\tilde{\alpha}^* = \alpha^* W_{\text{safe}}/W_0 = 0.3323$ (see Cerny 2004, p. 60). Alternatively, any numerical optimiser would also directly give the solution. Having obtained α^*, we can substitute this in the above equation to give the expected utility at the optimum.

$$\theta^* = E[U[W(\alpha^*)]] = -1.110397(10^{-25})$$

The certainty equivalent level of wealth W_{cert} can be calculated from

$$U(W_{\text{cert}}) = E\{U[W(\alpha^*)]\} \quad \text{that is,} \quad \frac{W_{\text{cert}}^{1-\gamma}}{1-\gamma} = -1.110397(10^{-25})$$

which for $(\gamma = 5)$ gives $W_{\text{cert}} = \$1{,}224{,}942$. We have $W_{\text{safe}} = R_f^* W_0 + y = \$1{,}220{,}000$, so the investor is better off by \$4,942 compared to holding only the risk-free asset.

The above is easily extended to the case where we still have only one risky asset but there are m possible outcomes ('states') for the risky asset excess return X with probabilities p_i. The maximand θ summed over all states is

$$\theta = \max_{\alpha} \sum_{i=1}^{m} p_i U(W_i)$$

where $U(W_i) = W_i^{1-\gamma}/(1-\gamma)$ and $W_i = W_{\text{safe}}(1 + \alpha X_i)$. Again, the only unknown in θ is the risky asset share α (*or* $\tilde{\alpha}$), and the optimal α can be obtained from a numerical optimiser (or 'by hand' if you have numerous sheets of paper).

Now let us be a little more adventurous and assume we choose a set of risky assets $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$, but for the moment assume there are only $m = 4$ states of nature for each excess return with an associated *joint* probability distribution:

$$\begin{aligned} X^{(i)} &= (X_1^{(i)}, \ldots, X_n^{(i)})' && (n \times 1) \\ W_i &= W_{\text{safe}}(1 + \alpha' X^{(i)}) && (i = 1 - 4) \\ p_i &= (p_1, p_2, p_3, p_4) && (1 \times 4) \end{aligned}$$

W_i is a scalar, and each outcome for the n-vector $X^{(i)}$ has an associated joint probability p_i. There are only four possible outcomes for the four vectors $X^{(i)}$ with probabilities p_i and hence four outcomes for W_i and $U(W_i)$. So, θ contains a sum over four states and is easily calculated but it now depends on the n values of α_i. The FOCs are $\partial\theta/\partial\alpha_i = 0$ (for $i = 1, 2, \ldots, n$) and, in general, these non-linear equations cannot be solved analytically and hence an optimiser must be used to obtain the α_i^* $(i = 1, 2, \ldots, n)$, which maximises θ.

Continuous Distribution

Suppose we have n assets but the distribution of returns R is continuous and for expositional purposes, assume the $(n \times 1)$ vector $X = R - r.e$ (where $e = n \times 1$ vector of ones) is multivariate normal with conditional density $f(x|\Lambda)$, where Λ is information at $t - 1$ or earlier. We have

$$\begin{aligned} W &= W_{\text{safe}}(1 + \alpha' X) \\ U(W) &= W^{1-\gamma}/(1-\gamma) \end{aligned}$$

Hence:

$$\theta = E\{U[W(\alpha)]\} = \frac{W_{\text{safe}}^{1-\gamma}}{(1-\gamma)} \cdot \int_{-\infty}^{\infty} (1 + \alpha' x)^{1-\gamma} f(x/\Lambda)\, dx$$

For illustrative purposes, assume $n = 2$ and $X|\Lambda \sim N(\mu, \Omega)$, then

$$f(x) = \frac{1}{(2\pi)^n |\det \Omega|} \exp\left[-\frac{1}{2}(x - \mu)\Omega^{-1}(x - \mu)\right]$$

The conditional mean and covariance matrix are assumed 'known' (i.e. estimated by a statistician from historic data), and the term in square brackets is a scalar function of

x. In the simplest case where each x_{it} is *iid* over time but there is cross-correlation at time t

$$x_i = \mu + \varepsilon_i \quad \Omega = E(\varepsilon' \varepsilon)$$

and the conditional and unconditional moments are equal. The (multiple) integral may be solvable analytically, but usually has to be evaluated numerically. In effect, the optimiser chooses alternative trial values for $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, calculates $\theta(\alpha)$ and chooses that value $\alpha = \alpha^*$ that achieves the largest value of $\theta(\alpha^*)$. A clear exposition of optimisation, including some useful GAUSS programmes, can be seen in Chapter 4 of Cerny (2004).

Risk Aversion and Portfolio Choice

When investors have a one-period horizon and maximise $U(W)$, there is a little more we can say about the response of the demands for risky assets to a change in initial wealth. For simplicity, we assume only one risky asset. We state these results without proof, but they are based on an analysis of the FOC, for any concave utility function:

$$E[U'(W)(R - R_f)] = 0$$

where $W = W_0(1 + R_f) + A(R - R_f)$ and W_0 = initial wealth, A = \$-amount invested in the risky asset (and $(W_0 - A)$ is invested in the risk-free asset). Our first result is easily seen from the FOC. If $ER = R_f$, then $A = 0$ satisfies the FOC, $W = W_0(1 + R_f)$ is non-stochastic, so $E[U'(W)(R - R_f)] = U'(W)E(R - R_f) = 0$. Hence, a risk-averse individual would not accept a fair gamble, namely, $ER - R_f = 0$, since the latter implies $A = 0$. Other results for any concave utility function are:

(i) If $ER > R_f$, then $A > 0$, the investor holds a *positive* amount of the risky asset – she is willing to accept a small gamble with positive expected return.

(ii) Declining *absolute* risk aversion (i.e. $\partial R_A/\partial W_0 < 0$, where R_A is the coefficient of absolute risk aversion) implies that $\partial A/\partial W_0 > 0$, that is, the individual invests more 'dollars' in the risky asset if initial wealth is higher (and vice versa).

(iii) If the coefficient of *relative* risk aversion R_R is decreasing in wealth (i.e. $\partial R_R/\partial W_0 < 0$), then $(\partial A/A_0)/(\partial W/W_0) > 1$, so the individual invests a greater *proportion* in the risky asset as wealth increases. The opposite applies for an investor with a coefficient of relative risk aversion that increases in wealth.

The above complement our earlier results that for *constant* absolute risk aversion, CARA (e.g. negative exponential), $\partial A/\partial W_0 = 0$ and for *constant* relative risk aversion, $\partial A/A = \partial W/W$, so optimal asset *shares* remain constant.

1.5 Indifference Curves and Intertemporal Utility

Although it is only the case under somewhat restrictive circumstances, let us assume that the utility function in Figure 2 for the risk averter can be represented solely in terms of the expected return and the variance of the return on the portfolio. The link between end-of-period wealth W and investment in a portfolio of assets yielding an expected return ER_p is $W = (1 + ER_p)W_0$, where W_0 is initial wealth. However, we now do assume that the utility function can be represented as

$$U = U(ER_p, \sigma_p^2) \qquad U_1 > 0, U_2 < 0, U_{11}, U_{22} < 0 \tag{70}$$

The sign of the first-order partial derivatives (U_1, U_2) imply that expected return adds to utility, while more 'risk' reduces utility. The second-order partial derivatives indicate diminishing marginal utility to additional expected 'returns' and increasing marginal disutility with respect to additional risk. The indifference curves for the above utility function are shown in Figure 4.

At a point like A on indifference curve I_1, the individual requires a higher expected return (A‴–A″) as compensation for a higher level of risk (A–A″) if he is to maintain the level of satisfaction (utility) pertaining at A: the indifference curves have a positive slope in risk–return space. The indifference curves are convex to the 'risk axis', indicating that at higher levels of risk, say at C, the individual requires a higher expected return (C‴–C″ > A‴–A″) for each additional increment to the risk he undertakes, than he did at A: the individual is 'risk-averse'. The indifference curves in risk–return space will be used when analysing portfolio choice in a simple mean-variance model.

Intertemporal Utility

A number of economic models of individual behaviour assume that investors obtain utility solely from consumption goods. At any point in time, utility depends positively on consumption and exhibits diminishing marginal utility

$$U = U(C_t) \qquad U'(C_t) > 0, U''(C_t) < 0 \tag{71}$$

The utility function, therefore, has the same slope as the 'risk averter' in Figure 2 (with C replacing W). The only other issue is how we deal with consumption that accrues at different points in time. The most general form of such an intertemporal lifetime utility function is

$$U_N = U(C_t, C_{t+1}, C_{t+2}, \ldots, C_{t+N}) \tag{72}$$

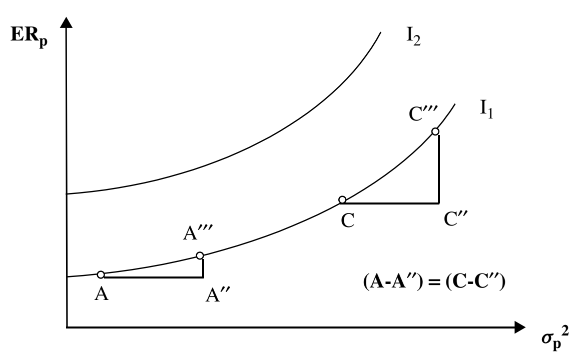


Figure 4 Risk–return: indifference curves

However, to make the mathematics tractable, some restrictions are usually placed on the form of U, the most common being additive separability with a constant subjective rate of discount, $0 < \theta < 1$:

$$U_N = U(C_t) + \theta U(C_{t+1}) + \theta^2 U(C_{t+2}) + \cdots + \theta^N U(C_{t+N}) \tag{73}$$

The lifetime utility function can be truncated at a finite value for N, or if $N \to \infty$, then the model is said to be an overlapping generations model since an individual's consumption stream is bequeathed to future generations.

The discount rate used in (73) depends on the 'tastes' of the individual between present and future consumption. If we define $\theta = 1/(1+d)$, then d is known as the subjective *rate of time preference*. It is the rate at which the individual will swap utility at time $t+j$ for utility at time $t+j+1$ and still keep lifetime utility constant. The additive separability in (73) implies that the marginal utility from extra consumption in year t is independent of the marginal utility obtained from extra consumption in any other year (suitably discounted).

For the two-period case, we can draw the indifference curves that follow from a simple utility function (e.g. $U = C_0^{\alpha_1} C_1^{\alpha_2}, 0 < \alpha_1, \alpha_2 < 1$) and these are given in Figure 5. Point A is on a higher indifference curve than point B since at A the individual has the same level of consumption in period 1, C_1 as at B, but at A, he has more consumption in period zero, C_0. At point H, if you reduce C_0 by x_0 units, then for the individual to maintain a constant level of lifetime utility he must be compensated by y_0 extra units of consumption in period 1, so he is then indifferent between points H and E. Diminishing marginal utility arises because at F, if you take away x_0 units of C_0, then he requires y_1 ($> y_0$) extra units of C_1 to compensate him. This is because at F he starts off with a lower initial level of C_0 than at H, so each unit of C_0 he gives up is relatively more valuable and requires more compensation in terms of extra C_1.

The intertemporal indifference curves in Figure 5 will be used in discussing investment decisions under certainty in the next section and again when discussing the consumption – CAPM model of portfolio choice and equilibrium asset returns under uncertainty.

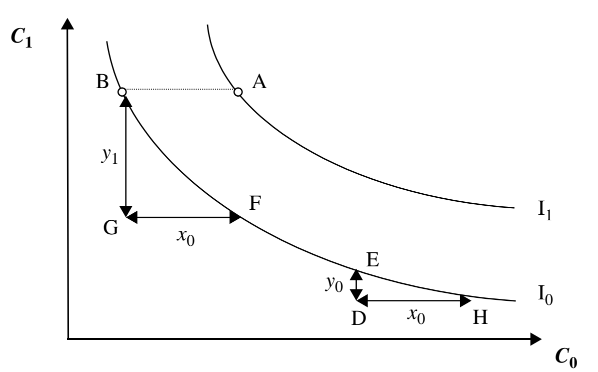


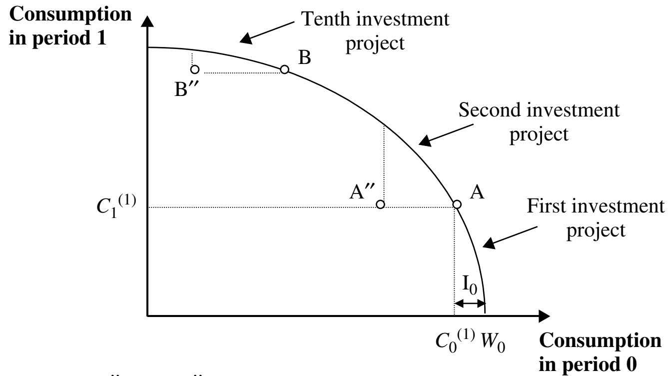
Figure 5 Intertemporal consumption: indifference curves

1.6 Investment Decisions and Optimal Consumption

Under conditions of certainty about future receipts, our investment decision rules indicate that managers should rank physical investment projects according to either their NPV or IRR. Investment projects should be undertaken until the NPV of the last project undertaken equals zero or equivalently if $IRR = r$, the risk-free rate of interest. Under these circumstances, the marginal (last) investment project undertaken just earns enough net returns (profits) to cover the loan interest and repayment of principal. For the economy as a whole, undertaking real investment requires a sacrifice in terms of lost current consumption output. Labour skills, man-hours and machines are, at $t = 0$, devoted to producing new machines or increased labour skills that will add to output and consumption but only in future periods. The consumption profile (i.e. less consumption goods today and more in the future) that results from the decisions of producers may not coincide with the consumption profile desired by individual consumers. For example, a high level of physical investment will drastically reduce resources available for current consumption and this may be viewed as undesirable by consumers who prefer at the margin, consumption today rather than tomorrow.

How can financial markets, through facilitating borrowing and lending, ensure that entrepreneurs produce the optimal level of physical investment (i.e. which yields high levels of future consumption goods) and also allow individuals to spread their consumption over time according to their preferences? Do the entrepreneurs have to know the preferences of individual consumers in order to choose the optimum level of physical investment? How can the consumers acting as shareholders ensure that the managers of firms undertake the ‘correct’ physical investment decisions, and can we assume that the financial markets (e.g. stock markets) ensure that funds are channelled to the most efficient investment projects?

Questions of the interaction between ‘finance’ and real investment decisions lie at the heart of the market system. The full answer to these questions involves complex issues. However, we can gain some useful insights if we consider a simple two-period model of the investment decision in which all outcomes are certain (i.e. riskless) in real terms (i.e. we assume zero price inflation). We shall see that under these assumptions, a *separation principle* applies. If managers ignore the preferences of individuals and simply invest in projects until the $NPV = 0$ or $IRR = r$, that is, maximise the value of the firm, then this policy will, given a capital market, allow each consumer to choose his desired consumption profile, namely, that which maximises his individual welfare. There is therefore a two-stage process or separation of decisions; yet, this still allows consumers to maximise their welfare by distributing their consumption over time according to their preferences. In step one, entrepreneurs decide the optimal level of physical investment, disregarding the preferences of consumers. In step two, consumers borrow or lend in the capital market to rearrange the time profile of their consumption to suit their individual preferences. In explaining this separation principle, we first deal with the production decision and then the consumers’ decision before combining these two into the complete model.



Note : (A – A″ = B – B″).

Figure 6 Production possibility curve

All output is either consumed or used for physical investment. The entrepreneur has an initial endowment W_0 at $t = 0$. He ranks projects in order of decreasing NPV, using the risk-free interest rate r as the discount factor. By abstaining from consumption of $C_0^{(1)}$, he obtains resources for his first investment project $I_0 = W_0 - C_0^{(1)}$. The physical investment in that project, which has the highest NPV (or IRR), yields consumption output at $t = 1$ of $C_1^{(1)}$, where $C_1^{(1)} > C_0^{(1)}$ (see Figure 6). The IRR of this project (in terms of consumption goods) is

$$1 + IRR^{(1)} = C_1^{(1)}/C_0^{(1)} \tag{74}$$

As he devotes more of his initial endowment W_0 to other investment projects with lower NPVs, the IRR (C_1/C_0) falls, which gives rise to the production opportunity curve with the shape given in Figure 6. The first and the most productive investment project has an NPV of

$$NPV^{(1)} = C_1^{(1)}/(1 + r) - I_0 > 0 \tag{75}$$

and

$$IRR^{(1)} = C_1^{(1)}/C_0^{(1)} > r \tag{76}$$

Let us now turn to the financing problem. In the capital market, any two consumption streams C_0 and C_1 have a present value PV given by

$$PV = C_0 + C_1/(1 + r) \qquad [77a]$$

and hence,

$$C_1 = PV(1 + r) - (1 + r)C_0 \qquad [77b]$$

For a given value of PV, this gives a straight line in Figure 7 with a slope equal to $-(1 + r)$. The above equation is referred to as the ‘money market line’ since it

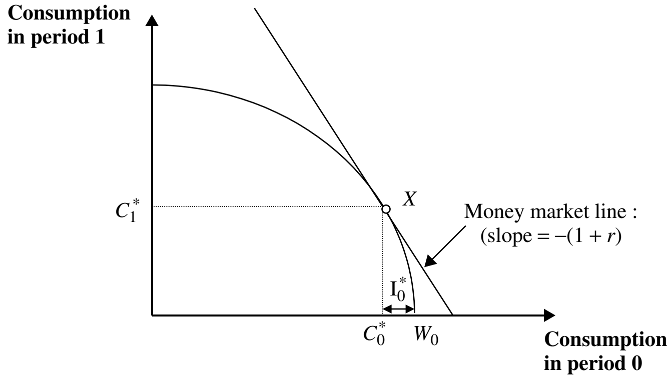


Figure 7 Money market line

represents the rate of return on lending and borrowing money. If you lend an amount C_0 today, you will receive $C_1 = (1 + r)\, C_0$ tomorrow.

Our entrepreneur, with an initial endowment of W_0, will continue to invest in physical assets until the IRR on the nth project just equals the risk-free interest rate

$$IRR^{(n)} = r$$

which occurs at point (C_0^*, C_1^*). Hence, the investment strategy that maximises the (net present) value of the firm involves an investment of

$$I_0^* = W_0 - C_0^*$$

Current consumption is C_0^* and consumption at $t = 1$ is C_1^* (Figure 7). At any point to the right of X, the slope of the investment opportunity curve (= IRR) exceeds the market interest rate (= r) and at points to the left of X, the opposite applies. However, the optimal levels of consumption (C_0^*, C_1^*) from the production decision may not conform to those desired by an individual consumer's preferences. We now leave the production decision and turn exclusively to the consumer's decision.

Suppose the consumer has income accruing in both periods and this income stream has a present value of PV. The consumption possibilities that fully exhaust this income (after two periods) are given by (77a). Assume that lifetime utility (satisfaction) of the consumer depends on C_0 and C_1

$$U = U(C_0, C_1)$$

and there is diminishing marginal utility in both C_0 and C_1 (i.e. $\partial U/\partial C > 0$, $\partial^2 U/\partial C^2 < 0$). The indifference curves are shown in Figure 8. To give up one unit of C_0, the consumer must be compensated with additional units of C_1 if he is to maintain his initial level of utility. The consumer wishes to choose C_0 and C_1 to maximise lifetime utility, subject to his budget constraint. Given his endowment PV, his optimal consumption in the two periods is (C_0^{**}, C_1^{**}) – Figure 8. In general, the optimal production or physical investment plan that yields consumption (C_0^*, C_1^*) will not equal

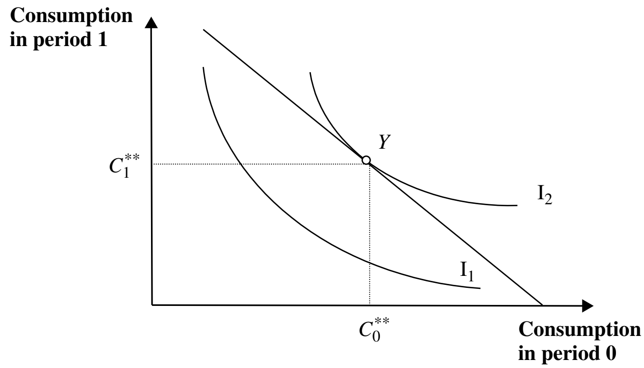


Figure 8 Consumer's maximisation

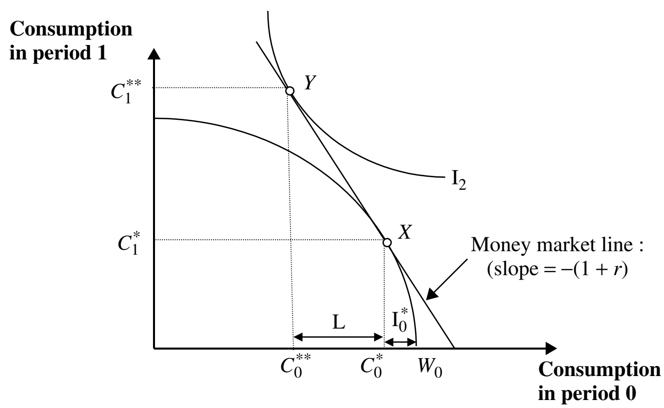


Figure 9 Maximisation with capital market

the consumer's optimal consumption profile (C_0^{**}, C_1^{**}). However, the existence of a capital market ensures that the consumer's optimal point can be attained. To see this, consider Figure 9.

The entrepreneur has produced a consumption profile (C_0^*, C_1^*) that maximises the value of the firm – Figure 9. We can envisage this consumption profile as being paid out to the owners of the firm in the form of (dividend) income. The present value of this 'cash flow' is PV*, where

$$PV^* = C_0^* + C_1^*/(1+r) \tag{78}$$

This is, of course, the 'income' given to our individual consumer as owner of the firm. But, under conditions of certainty, the consumer can 'swap' this amount PV* for any combination of consumption that satisfies

$$PV^* = C_0 + C_1/(1+r) \tag{79}$$

Given PV* and his indifference curve I_2 in Figure 9, he can then borrow or lend in the capital market at the riskless rate r to achieve that combination (C_0^{**}, C_1^{**}) that maximises his utility.

Thus, there is a separation of investment and financing (borrowing and lending) decisions. Optimal borrowing and lending take place independently of the physical investment decision. If the entrepreneur and consumer are the same person(s), the separation principle still applies. The investor (as we now call him) first decides how much of his own initial endowment W_0 to invest in physical assets, and this decision is independent of his own (subjective) preferences and tastes. This first-stage decision is an objective calculation based on comparing the IRR of his investment projects with the risk-free interest rate. His second-stage decision involves how much to borrow or lend in the capital market to 'smooth out' his desired consumption pattern over time. The latter decision is based on his preferences or tastes, at the margin, for consumption today versus (more) consumption tomorrow.

Much of the rest of this book is concerned with how financing decisions are taken when we have a risky environment. The issue of how shareholders ensure that managers act in the best interest of the shareholders, by maximising the value of the firm, comes under the heading of corporate control mechanisms (e.g. mergers, takeovers). The analysis of corporate control is not directly covered in this book. We only consider whether market prices provide correct signals for resource allocation (i.e. physical investment), but we do not look closely at issues involving the incentive structure within the firm based on these market signals: this is the principal–agent problem in corporate finance.

We can draw a parallel between the above results under certainty with those we shall be developing under a risky environment.

(i) In a risky environment, a somewhat different separation principle applies. Each investor, when choosing his portfolio of risky marketable assets (e.g. shares, bonds), will hold risky assets in the same proportion as all other investors, regardless of his preferences of risk versus return. Having undertaken this first-stage decision, each investor then decides how much to borrow or lend in the money market at the risk-free interest rate – it is at this point that his preferences influence the split between the risky assets and the risk-free asset. This separation principle is the basis of the mean-variance model of optimal portfolio choice and of the CAPM of equilibrium asset returns.

(ii) The optimal amount of borrowing and lending in the money market in the riskless case occurs where the individual's subjective marginal rate of substitution of future for current consumption [i.e. $(\partial C_1/\partial C_0)_u$] equals $-(1+r)$, where r is the 'price' or opportunity cost of money. Under uncertainty, an analogous condition applies, namely, that the individual's *subjective* trade-off between expected return and risk is equal to the market price of risk.

1.7 Summary

We have developed some basic tools for analysing behaviour in financial markets. There are many nuances on the topics discussed that we have not had time to elaborate in detail, and in future chapters, these omissions will be rectified. The main conclusions to emerge are:

- Market participants generally quote ‘simple’ annual interest rates but these can always be converted to effective annual (compound) rates or to continuously compounded rates.
- The concepts of DPV and IRR can be used to analyse physical investment projects and to calculate the fair price of bills, bonds and stocks.
- Theoretical models of asset demands and asset prices often use utility functions as their objective function. Utility functions and their associated indifference curves can be used to represent risk aversion, risk lovers and risk-neutral investors.
- Under conditions of certainty, a type of separation principle applies when deciding on (physical) investment projects. Managers can choose investment projects to maximise the value of the firm and disregard investor’s preferences. Then, investors are able to borrow and lend to allocate consumption between ‘today’ and ‘tomorrow’ in order to maximise their utility.
- One-period models in which utility is assumed to depend only on expected portfolio return and the variance of return give rise to asset shares that depend on expected excess returns, the covariance matrix of returns and ‘risk aversion’ of the investor.
- One-period models that depend on the expected utility of end-of-period wealth generally do not give closed-form solutions. The exceptions are quadratic utility or negative exponential utility plus multivariate normally distributed returns, where asset demands are linear in expected excess returns. For many other utility functions, the optimal solution for asset demands has to be calculated numerically.

Appendix: Mean-Variance Model and Utility Functions

If an investor maximises expected utility of end-of-period portfolio wealth, then it can be shown that this is equivalent to maximising *a function* of expected portfolio returns and portfolio variance providing

(a) either utility is quadratic, or

(b) portfolio returns are normally distributed (and utility is concave).

If initial wealth is W_0 and the stochastic portfolio return is R_p, then end-of-period wealth and utility are

$$W = W_0(1 + R_p) \tag{A1}$$

$$U(W) = U[W_0(1 + R_p)] \tag{A2}$$

Expanding $U(R_\mathrm{p})$ in a Taylor series around the mean of $R_\mathrm{p}(=\mu_\mathrm{p})$ gives

$$U(R_\mathrm{p}) = U(\mu_\mathrm{p}) + (R_\mathrm{p} - \mu_\mathrm{p})U'(\mu_\mathrm{p}) + (1/2)(R_\mathrm{p} - \mu_\mathrm{p})^2 U''(\mu_\mathrm{p}) + \text{higher order terms} \tag{A3}$$

Since $E(R_\mathrm{p} - \mu_\mathrm{p}) = 0$, and $E(R_\mathrm{p} - \mu_\mathrm{p})^2 = \sigma_\mathrm{p}^2$, taking expectations of (A3):

$$E[U(R_\mathrm{p})] = U(\mu_\mathrm{p}) + \frac{1}{2}\sigma_\mathrm{p}^2 U''(\mu_\mathrm{p}) + E(\text{higher} - \text{order terms}) \tag{A4}$$

If utility is quadratic, then higher-order terms other than U'' are zero. If returns are normally distributed, then $E[(R_\mathrm{p} - \mu_\mathrm{p})^n] = 0$ for n odd, and $E[(R_\mathrm{p} - \mu_\mathrm{p})^n]$ for n even is a function only of the variance σ_p^2. Hence for cases (a) and (b), $E[U(R_\mathrm{p})]$ is *a function* of only the mean μ_p and the variance σ_p^2. This result is moderately useful; for example, it can be shown that if utility is defined only in terms of μ_p and σ_p and is concave, then indifference curves in $(\mu_\mathrm{p}, \sigma_\mathrm{p})$ space are convex, as assumed in the text (see Figure 4). However, until we specify a specific utility function, we do not know the functional relationship between $E[U(R_\mathrm{p})]$ and $(\mu_\mathrm{p}, \sigma_\mathrm{p})$ and hence we cannot determine whether there is an analytic closed-form solution for asset demands.

Using quadratic utility has the problem that marginal utility is negative for levels of wealth above the bliss point. Assuming normality for returns may not be an accurate representation and also prices may become negative. Continuous time models assume returns that are *instantaneously* normally distributed, and this provides considerable tractability (see Cuthbertson and Nitzsche 2001a), which is widely used in derivatives pricing. Again, the empirical validity of normality even at high frequencies (e.g. tick-data) is debatable, but the usefulness of results depends on the problem at hand (e.g. it seems reasonable when pricing stock index options).