I.1
Basic Calculus for Finance

I.1.1 INTRODUCTION

This chapter introduces the functions that are commonly used in finance and discusses their properties and applications. For instance, the *exponential function* is used to discount forward prices to their present value and the inverse of the exponential function, the natural *logarithmic function* or 'log' for short, is used to compute returns in continuous time. We shall encounter numerous other financial applications of these functions in the subsequent volumes. For instance, the fair price of a *futures* contract is an exponential function of an interest rate multiplied by the spot price of the underlying asset. A standard futures contract is a contract to buy or sell a *tradable asset* at some specified time in the future at a price that is agreed today. The four main types of tradable assets are stocks, bonds, commodities and currencies.

The futures price is a linear function of the underlying asset price. That is, if we draw the graph of the futures price against the price of the underlying we obtain a straight line. But *non-linear functions*, which have graphs that are not straight lines, are also used in every branch of finance. For instance, the price of a bond is a non-linear function of its yield. A *bond* is a financial asset that periodically pays the bearer a fixed *coupon* and is redeemed at maturity at *par* value (usually 100). The *yield* (also called *yield to maturity*) on a bond is the fixed rate of interest that, if used to discount the payments (i.e. the *cash flow*) to their present value, makes the net present value of the cash flow equal to the price of the bond.

Functions of several variables are also very common in finance. A typical example of a function of several variables in finance is the price of an *option*. An option is the right to buy or sell an underlying asset on (or before) a certain time in the future at a price that is fixed today. An option to buy is called a *call option* and an option to sell is called a *put option*. The difference between a futures contract and an option is that the holder of a futures contract is bound to buy or sell at the agreed price, but the holder of the option has the right to buy or sell the asset if he chooses. That is, the holder has the 'option' of exercising their right or not.[1] Early exercise of the option is not possible with a *European option*, but an *American option* or a *Bermudan option* may be exercised before the expiry of the contract. Americans can be exercised at any time and Bermudans can be exercised on selected dates.

After an informal description of the concepts of continuous and differentiable functions we focus on standard techniques for differentiation and integration. Differentiation is a core concept, and in finance the derivatives of a function are commonly referred to as *sensitivities* rather than 'derivatives'.[2] This is because the term *derivative* when used in a financial

[1] Hence buying options is not nearly as risky as selling (or *writing*) options.
[2] Derivatives of a function are not *always* called sensitivities in finance. For instance in Chapter I.6 we introduce *utility functions*, which are used to assess the optimal trade-off between risk and return. The first and second derivatives of an investor's utility function tell us whether the investor is risk averse, risk loving or risk neutral.

context refers to a financial instrument that is a contract on a contract, such as a futures contract on an interest rate, or an option on an interest rate swap.[3]

We shall employ numerous sensitivities throughout these volumes. For instance, the first order yield sensitivity of a bond is called the *modified duration*. This is the first partial derivative of the bond price with respect to the yield, expressed as a percentage of the price; and the second order yield sensitivity of a bond is called the *convexity*. This is the second derivative of the bond price with respect to its yield, again expressed as a percentage of the price.

The first and second derivatives of a function of several variables are especially important in finance. To give only one of many practical examples where market practitioners use derivatives (in the mathematical sense) in their everyday work, consider the role of the *market makers* who operate in exchanges by buying and selling assets and derivatives. Market makers make money from the *bid–ask spread* because their *bid price* (the price they will buy at) is lower than their *ask price* (the price they will sell at, also called the *offer price*). Market makers take their profits from the spread that they earn, not from taking risks. In fact they will *hedge* their risks almost completely by forming a *risk free portfolio*.[4] The *hedge ratios* determine the quantities of the underlying, and of other options on the same underlying, to buy or sell to make a risk free portfolio. And hedge ratios for options are found by taking the first and second partial derivatives of the option price.

The pricing of financial derivatives is based on a *no arbitrage* argument. No arbitrage means that we cannot make an instantaneous profit from an investment that has no uncertainty. An investment with no uncertainty about the outcome is called a *risk free investment*. With a risk free investment it is impossible to make a 'quick turn' or an instantaneous profit. However, profits will be made over a period of time. In fact no arbitrage implies that all risk free investments must earn the same rate of return, which we call the *risk free return*.

To price an option we apply a no arbitrage argument to derive a *partial differential equation* that is satisfied by the option price. In some special circumstances we can solve this equation to obtain a formula for the option price. A famous example of this is the *Black–Scholes–Merton formula* for the price of a standard European option.[5] The model price of an option depends on two main variables: the current price of the underlying asset and its *volatility*. Volatility represents the uncertainty surrounding the expected price of the underlying at the time when the option expires. A standard European call or put, which is often termed a *plain vanilla option*, only has value because of volatility. Otherwise we would trade the corresponding futures contract because futures are much cheaper to trade than the corresponding options. In other words, the bid–ask spread on futures is much smaller than the bid–ask spread on options.

Taylor expansions are used to approximate values of non-linear differentiable functions in terms of only the first few derivatives of the function. Common financial applications include the *duration–convexity approximation* to the change in value of a bond and the *delta–gamma approximation* to the change in value of an options portfolio. We also use Taylor expansion to simplify numerous expressions, from the adjustment of Black–Scholes–Merton options

[3] *Financial instrument* is a very general term that includes tradable assets, interest rates, credit spreads and all derivatives.
[4] A *portfolio* is a collection of financial instruments, i.e. a collection of assets, or of positions on interest rates or of derivative contracts.
[5] Liquid options are not actually priced using this formula: the prices of all liquid assets are determined by market makers responding to demand and supply.

prices to account for uncertainty in volatility to the stochastic differential equations that we use for modelling continuous time price processes.

Integration is the opposite process to differentiation. In other words, if f is the derivative of another function F we can obtain F by integrating f. Integration is essential for understanding the relationship between continuous probability distributions and their *density functions*, if they exist. Differentiating a probability distribution gives the density function, and integrating the density function gives the distribution function.

This chapter also introduces the reader to the basic analytical tools used in portfolio mathematics. Here we provide formal definitions of the *return* and the *profit and loss* on a single investment and on a portfolio of investments in both discrete and continuous time. The *portfolio weights* are the proportions of the total capital invested in each instrument. If the weight is positive we have a *long position* on the instrument (e.g. we have bought an asset) and if it is negative we have a *short position* on the instrument (e.g. we have 'short sold' an asset or written an option). We take care to distinguish between portfolios that have constant holdings in each asset and those that are rebalanced continually so that portfolio weights are kept constant. The latter is unrealistic in practice, but a constant weights assumption allows one to represent the return on a linear portfolio as a weighted sum of the returns on its constituent assets. This result forms the basis of portfolio theory, and will be used extensively in Chapter I.6.

Another application of differentiation is to the *optimal allocation* problem for an investor who faces certain constraints, such as no short sales and/or at least 30% of his funds must be invested in US equities. The investor's problem is to choose his portfolio weights to optimize his objective whilst respecting his constraints. This falls into the class of constrained optimization problems, problems that are solved using differentiation.

Risk is the uncertainty about an expected value, and a *risk-averse investor* wants to achieve the maximum possible return with the minimum possible risk. Standard measures of portfolio risk are the *variance* of a portfolio and its square root which is called the *portfolio volatility*.[6] The portfolio variance is a quadratic function of the portfolio weights. By differentiating the variance function and imposing any constraints we can solve the optimal allocation problem and find the *minimum variance portfolio*.

Very little prior knowledge of mathematics is required to understand this chapter, although the reader must be well motivated and keen to learn. It is entirely self-contained and all but the most trivial of the examples are contained in the accompanying Excel spreadsheet. Recall that in all the spreadsheets readers may change the values of inputs (marked in red) to compute a new output (in blue).

I.1.2 FUNCTIONS AND GRAPHS, EQUATIONS AND ROOTS

The value of a function of a single variable is written $f(x)$, where f is the *function* and x is the *variable*. We assume that both x and $f(x)$ are real numbers and that for each value of x there is only one value $f(x)$. Technically speaking this makes f a 'single real-valued function of a single real variable', but we shall try to avoid such technical vocabulary where possible. Basically, it means that we can draw a *graph* of the function, with the values x along the horizontal axis and the corresponding values $f(x)$ along the vertical axis, and that this graph

[6] Volatility is the annualized standard deviation, i.e. the standard deviation expressed as a percentage per annum.

has no 'loops'. Setting the value of a function $f(x)$ equal to 0 gives the values of x where the function crosses or touches the x-axis. These values of x satisfy the equation $f(x)=0$ and any values of x for which this equation holds are called the *roots* of the equation.

I.1.2.1 Linear and Quadratic Functions

A *linear function* is one whose graph is a straight line. For instance, the function $f(x)=3x+2$ is linear because its graph is a straight line, shown in Figure I.1.1. A linear function defines a linear equation, i.e. $3x+2=0$ in this example. This has a root when $x=-2/3$. Readers may use the spreadsheet to graph other linear functions by changing the value of the coefficients a and b in the function $f(x)=ax+b$.

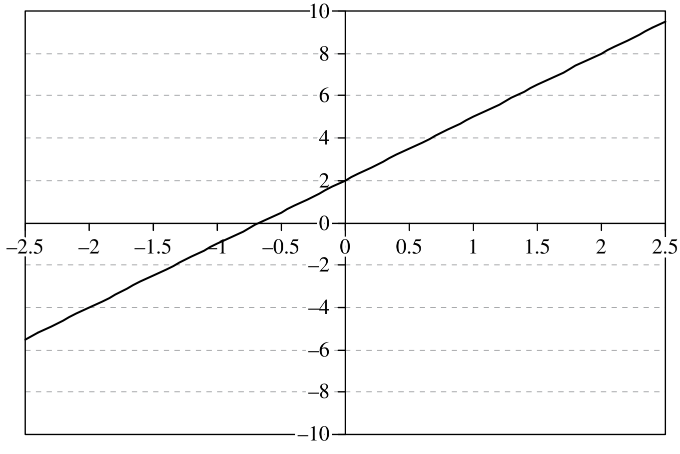


Figure I.1.1 A linear function

By contrast, the function $f(x)=4x^2+3x+2$ shown in Figure I.1.2 defines an equation with no real roots because the function value never crosses the x-axis. The graph of a general *quadratic function* $f(x)=ax^2+bx+c$ has a '∩' or '∪' shape that is called a *parabola*:

- If the coefficient $a>0$ then it has a ∪ shape, and if $a<0$ then it has a ∩ shape. The size of a determines the steepness of the curve.
- The coefficient b determines its horizontal location: for $b>0$ the graph is shifted to the left of the vertical axis at $x=0$, otherwise it is shifted to the right. The size of b determines the extent of the shift.
- The coefficient c determines its vertical location: the greater the value of c the higher the graph is on the vertical scale.

Readers may play with the values of a, b and c in the spreadsheet for Figure I.1.2 to see the effect they have on the graph. At any point that the graph crosses or touches the x-axis we have a real root of the quadratic equation $f(x)=0$.

A well-known formula gives the *roots of a quadratic equation* $ax^2+bx+c=0$ where a, b and c are real numbers. This is:

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}. \tag{I.1.1}$$

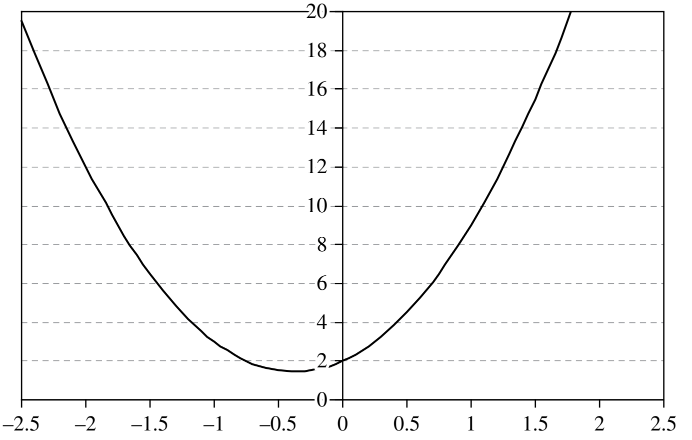


Figure I.1.2 The quadratic function $f(x) = 4x^2 + 3x + 2$

The term inside the square root, $b^2 - 4ac$, is called the *discriminant* of the equation. If the discriminant is negative, i.e. $b^2 < 4ac$, the quadratic equation has no real roots: the roots are a pair of complex numbers.[7] But if $b^2 > 4ac$ there are two distinct real roots, given by taking first '+' and then '−' in (I.1.1). If $b^2 = 4ac$ the equation has two identical roots.

EXAMPLE I.1.1: ROOTS OF A QUADRATIC EQUATION

Find the roots of $x^2 - 3x + 2 = 0$.

SOLUTION We can use formula (I.1.1), or simply note that the function can be factorized as

$$x^2 - 3x + 2 = (x - 1)(x - 2),$$

and this immediately gives the two roots $x = 1$ and $x = 2$. Readers can use the spreadsheet for this example to find the roots of other quadratic equations, if they exist in the real numbers.

I.1.2.2 Continuous and Differentiable Real-Valued Functions

Loosely speaking, if the graph of a function $f(x)$ has no jumps then $f(x)$ is *continuous function*. A *discontinuity* is a jump in value. For instance the *reciprocal function*,

$$f(x) = x^{-1}, \text{ also written } f(x) = \frac{1}{x},$$

has a graph that has a shape called a *hyperbola*. It has a discontinuity at the point $x = 0$, where its value jumps from $-\infty$ to $+\infty$, as shown in Figure I.1.3. But the reciprocal function is continuous at all other points.

Loosely speaking, if the graph of a continuous function $f(x)$ has no corners then $f(x)$ is a *differentiable function*. If a function is not differentiable it can still be continuous, but if it is not continuous it cannot be differentiable. A differentiable function has a unique *tangent*

[7] The *square root function* $\sqrt{x}$ or $x^{1/2}$ is only a real number if $x \geq 0$. If $x < 0$ then $\sqrt{x} = i\sqrt{-x}$, where $i = \sqrt{-1}$ is an *imaginary* or *complex* number.

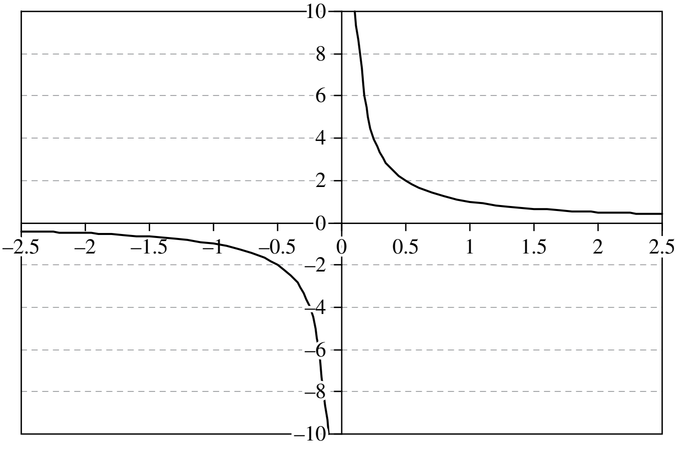


Figure I.1.3 The reciprocal function

line. That is, f is differentiable at x if we can draw only one straight line that touches the graph at x.

Functions often have points at which they are not continuous, or at least not differentiable. Examples in finance include:

- **The pay-off to a simple call option**. This is the function $\max(S-K,0)$, also written $(S-K)^+$, where S is the variable and K is a constant, called the strike of the option. The graph of this function is shown in Figure III.3.1 and it is clearly not differentiable at the point $S=K$.
- **Other option pay-offs**. Likewise, the pay-off to a simple put option is not differentiable at the point $S=K$; more complex options, such as barrier options may have other points where the pay-off is not differentiable.
- **The indicator function**. This is given by

$$1_{\{\text{condition}\}} = \begin{cases} 1 \text{ if the condition is met,} \\ 0 \text{ otherwise.} \end{cases} \tag{I.1.2}$$

  For instance, $1_{\{x>0\}}$ is 0 for non-positive x and 1 for positive x. There is a discontinuity at $x=0$, so the function cannot be differentiable there.
- **The absolute value function**. This is written $|x|$ and is equal to x if x is positive, $-x$ if x is negative and 0 if $x=0$. Clearly there is a corner at $x=0$, so $|x|$ is not differentiable there.

I.1.2.3 Inverse Functions

The inverse function of any real-valued function $f(x)$ is *not* its reciprocal value at any value of x:[8]

$$f^{-1}(x) \neq \frac{1}{f(x)}.$$

[8] The reciprocal of a real number a is $1/a$, also written a^{-1}.

Instead

$$f^{-1}(x) = g(x) \Leftrightarrow f(g(x)) = x.$$

That is, we obtain the inverse of a function by reflecting the function in the 45% line, as shown in Figure I.1.4.

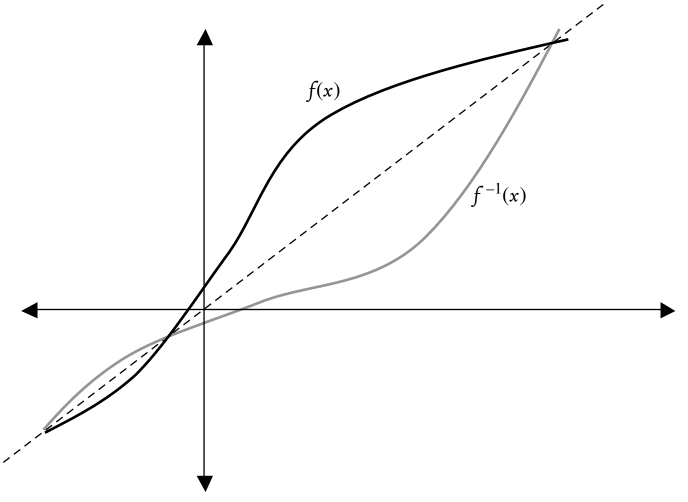


Figure I.1.4 The inverse of a function

I.1.2.4 The Exponential Function

An *irrational number* is a real number that has a decimal expansion that continues indefinitely without ever repeating itself, i.e. without ending up in a cycle, and most irrational numbers are transcendental numbers.[9] Even though there are infinitely many transcendental numbers we only know a handful of them. The ancient Greeks were the first to discover a transcendental number $\pi = 3.14159\ldots$. The next transcendental number was discovered only many centuries later. This is the number $e = 2.7182818285\ldots$.

Just as π is a real number between 3 and 4, e is simply a (very special) real number between 2 and 3. Mathematicians arrived at e by computing the limit

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n. \tag{I.1.3}$$

[9] According to Professor Walter Ledermann: Every *periodic* decimal expansion, that is, an expansion of the form $N.a_1a_2\ldots a_n(b_1b_2\ldots b_s)(b_1b_2\ldots b_s)$ is equal to a *rational* number m/n where m and n are integers and $n \neq 0$. Conversely, every rational number has a periodic decimal expansion. Hence a real number is irrational if and only if its decimal expansion is non-periodic. For example the expansion $\sqrt{2} = 1.414213562\ldots$ is non-periodic. There are two types of irrational number. A solution of a polynomial equation

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \quad (a_0 \neq 0)$$

with integral coefficients a_i is called an *algebraic number*. For example $\sqrt{2}$ is an algebraic number because it is a solution of the equation $x^2 - 2 = 0$. An irrational number which is not the solution of any polynomial equation with integral coefficients is called a *transcendental number*. It is obviously very difficult to prove that a particular number is transcendental because all such polynomials would have to be considered.

We can consider the functions 2^x and 3^x where x is any real number, so we can also consider e^x. This is called the exponential function and it is also denoted $\exp(x)$:

$$e^x = \exp(x) = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n. \quad \text{(I.1.4)}$$

Thus $\exp(1) = e$ and $\exp(0) = 1$. Indeed, any number raised to the power zero is 1.

Figure I.1.5 shows the graph of the exponential function. It has the unique property that at every point x its slope is equal to the value of $\exp(x)$. For instance, the slope of the exponential function at the point $x = 0$ (i.e. when it crosses the vertical axis) is 1.

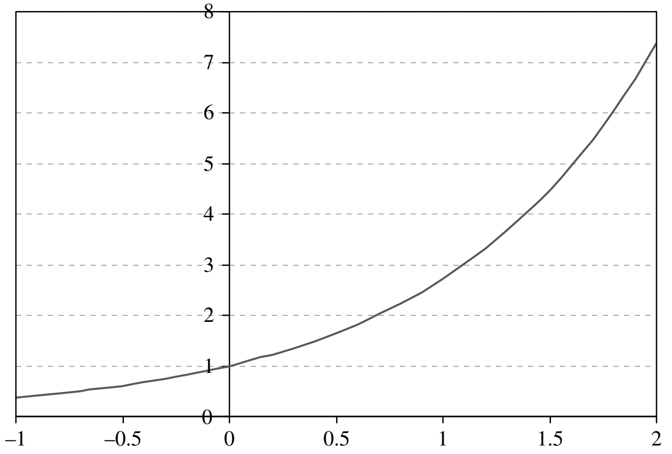


Figure I.1.5 The exponential function

Since e is a real number, $\exp(x)$ obeys the usual *laws of indices* such as[10]

$$\exp(x+y) = \exp(x)\exp(y), \quad \text{(I.1.5)}$$

$$\exp(x-y) = \frac{\exp(x)}{\exp(y)}. \quad \text{(I.1.6)}$$

The exponential function may be represented as a *power series*, viz.,

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots, \quad \text{(I.1.7)}$$

where $n!$ denotes the *factorial function*:

$$\begin{aligned} n! &= n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1, \\ 0! &= 1. \end{aligned} \quad \text{(I.1.8)}$$

Power series expansions are useful for approximating the value of a function when x is not too large. Thus

$$e = \exp(1) \approx 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} = 2.708$$

[10] For any real numbers a, b and c we have $a^b a^c = a^{b+c}$.

and, for instance,

$$7.389 = \exp(2) \approx 1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} + \frac{4}{15} = 7.267.$$

I.1.2.5 The Natural Logarithm

The inverse of the exponential function is the natural logarithm function, abbreviated in the text to 'log' and in symbols to 'ln'. This function is illustrated in Figure I.1.6. It is only defined for a positive real number x.[11]

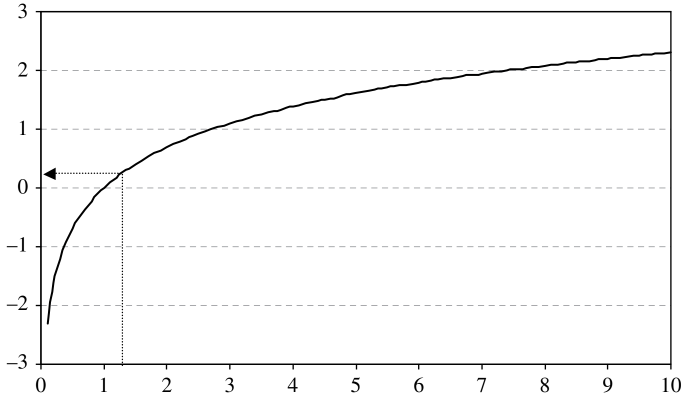


Figure I.1.6 The natural logarithmic function

Notice that ln $1 = 0$, and ln $x < 0$ for $0 < x < 1$. The dotted arrow on the graph shows that the natural logarithm function is approximately linear in the region of $x = 1$. That is,

$$\ln(1 + x) \approx x \text{ when } x \text{ is small.} \tag{I.1.9}$$

More generally, we have the following *power series expansion* for the log function:

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \text{ provided } -1 < x. \tag{I.1.10}$$

The logarithm function is useful for a number of reasons. One property that is important because it often makes calculations easier is that the log of a product is the sum of the logs:

$$\ln(xy) = \ln(x) + \ln(y)\,. \tag{I.1.11}$$

Also,

$$\ln\left(x^{-1}\right) = -\ln(x)\,. \tag{I.1.12}$$

Putting these together gives

$$\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)\,. \tag{I.1.13}$$

[11] This is not strictly true, since we can extend the definition beyond positive real numbers into complex numbers by writing, for negative real x, $\ln(x) = \ln(-x\exp(\pi \mathrm{i})) = \ln(-x) + \pi\mathrm{i}$, where the imaginary number i is the square root of -1. You do not need to worry about this for the purposes of the present volumes.

I.1.3 DIFFERENTIATION AND INTEGRATION

The *first derivative* of a function at the point x is the slope of the tangent line at x. All linear functions have a constant derivative because the tangent at every point is the line itself. For instance, the linear function $f(x) = 3x + 2$ shown in Figure I.1.1 has first derivative 3. But non-linear functions have derivatives whose value depends on the point x at which it is measured. For instance, the quadratic function $f(x) = 2x^2 + 4x + 1$ has a first derivative that is increasing with x. It has value 0 at the point $x = -1$, a positive value when $x > -1$, and a negative value when $x < -1$.

This section defines the derivatives of a function and states the basic rules that we use to differentiate functions. We then use the first and second derivatives to define properties that are shared by many of the functions that we use later in this book, i.e. the monotonic and convexity properties. Finally we show how to identify the stationary points of a differentiable function.

I.1.3.1 Definitions

Technically speaking, we can define the derivative of a function f as:

$$f'(x) = \lim_{\Delta x \downarrow 0} \left(\frac{f(x + \Delta x) - f(x)}{\Delta x} \right). \tag{I.1.14}$$

That is, we take the slope of the chord between two points a distance Δx apart and see what happens as the two points get closer and closer together. When they touch, so the distance between them becomes zero, the slope of the chord becomes the slope of the tangent, i.e. the derivative.

This is illustrated in Figure I.1.7. The graph of the function is shown by the black curve. The chord from P to Q is the dark grey line. By definition of the slope of a line (i.e. the vertical height divided by the horizontal distance), the slope of the chord is:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The tangent line is drawn in light grey and its slope is the derivative $f'(x)$, by definition. Now we let the increment in x, i.e. Δx, get smaller. Then the point Q moves closer to P and

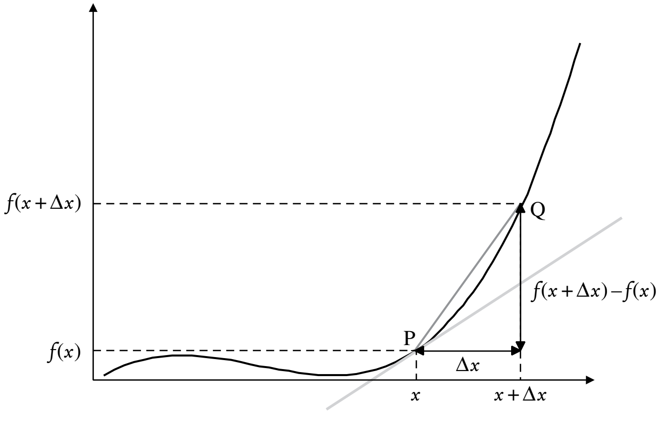


Figure I.1.7 Definition of the first derivative the slope of the chord gets closer to the slope of the tangent, i.e. the derivative. In the limit, when $\Delta x = 0$, the points P and Q coincide and the slope of the chord, i.e. the right-hand side of (I.1.14), is equal to the slope of the tangent, i.e. the left-hand side of (I.1.14).

Notation and Terminology

- The *second derivative* is the derivative of the derivative, given by

$$f''(x) = \lim_{\Delta x \downarrow 0} \left[\left(\frac{f'(x+\Delta x) - f'(x)}{\Delta x} \right) \right], \tag{I.1.15}$$

  and higher derivatives are defined analogously. If we differentiate the function m times, we denote the mth derivative by $f^{(m)}(x)$.
- We can use the alternative notation $\frac{df}{dx}$ for the first derivative $f'(x)$, and more generally the notation $\frac{d^m f}{dx^m}$ also stands for the mth derivative.
- Associated with the definition of derivative is the *differential operator d*. The *differential* or *total derivative* of a function $f(x)$ is given by:

$$df(x) = f'(x)\,dx. \tag{I.1.16}$$

  The differential operator is used, for instance, in Section I.1.4.5 below to describe the dynamics of financial asset returns in continuous time.

I.1.3.2 Rules for Differentiation

A number of simple rules for differentiation may be proved using the methodology depicted in Figure I.1.7 to calculate the derivatives of certain functions from first principles. These are summarized as follows:

1. **Power**. The derivative of ax^n is nax^{n-1} for any constant a and any (positive or negative) real number n: in other words,
$$\frac{d}{dx}(ax^n) = nax^{n-1}.$$
2. **Exponential**. The derivative of $\exp(x)$ is $\exp(x)$: in other words,
$$\frac{d}{dx}(e^x) = e^x.$$
3. **Logarithm**. The derivative of $\ln x$ is x^{-1}: in other words,
$$\frac{d}{dx}(\ln x) = \frac{1}{x}.$$
4. **Chain rule**. The derivative of a function of another function $f(g(x))$ is the product of the two derivatives, i.e.
$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).$$
5. **Sum**. The derivative of a sum is the sum of the derivatives, i.e.
$$\frac{d}{dx}(f(x) + g(x)) = f'(x) + g'(x)$$
6. **Product**. For the derivative of $f(x)g(x)$:
$$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + g'(x)f(x)$$

7. **Quotient**. The derivative of the reciprocal of $f(x)$:

$$\frac{d}{dx}\left(\frac{1}{f(x)}\right) = -\frac{f'(x)}{f(x)^2}.$$

More generally, the derivative of a quotient of functions is:

$$\frac{d}{dx}\left(\frac{g(x)}{f(x)}\right) = \frac{f'(x)g(x) - g'(x)f(x)}{f(x)^2}.$$

These rules allow the derivatives of certain functions to be easily computed. For instance, we can use rule 1 to find successive derivatives of $f(x) = x^4$, that is:

$$f'(x) = 4x^3, f''(x) = 12x^2, f'''(x) = f^{(3)}(x) = 24x, f^{(4)}(x) = 24 \text{ and } f^{(5)}(x) = 0,$$

Similarly, rules 1 and 5 together show that the first derivative of $f(x) = 2x^2 + 4x + 1$ is $4x + 4$.[12]

Example I.1.2: Calculating derivatives

Find the first derivative of the functions whose graphs are shown in Figure I.1.8, viz.

(a) $x^3 - 7x^2 + 14x - 8$

(b) $10 - 0.5x^2 + \sqrt{\ln(1 + x^2)} - \dfrac{4\ln(x)}{\exp(x)}$

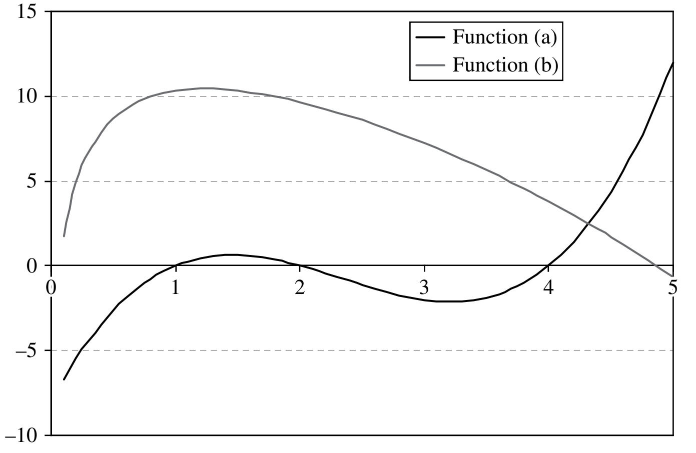


Figure I.1.8 Two functions

Solution

(a) This function is a cubic polynomial, and its graph crosses the x-axis at the points $x = 1$, $x = 2$ and $x = 4$. We know this because the function may also be written

$$f(x) = x^3 - 7x^2 + 14x - 8 = (x-1)(x-2)(x-4).$$

[12] So the first derivative is zero when $x = -1$.

By the summation rule (5) we just differentiate each term in $x^3 - 7x^2 + 14x - 8$ separately using rule 1, and then sum the results:

$$\frac{d}{dx}\left(x^3 - 7x^2 + 14x - 8\right) = 3x^2 - 14x + 14.$$

(b) We again differentiate term by term, and the first two terms are easy:

$$\frac{d}{dx}\left(10 - 0.5x^2\right) = -x.$$

For the next term use the chain rule (4):

$$\frac{d}{dx}\left(\sqrt{\ln(1+x^2)}\right) = \frac{d}{dx}\left(\ln\left(1+x^2\right)\right)^{1/2} = \frac{1}{2}\left(\ln\left(1+x^2\right)\right)^{-1/2} \times \frac{d}{dx}\left(\ln\left(1+x^2\right)\right).$$

Then by the chain rule again and rule 3 for the derivative of the log,

$$\frac{d}{dx}\left(\ln\left(1+x^2\right)\right) = 2x\left(1+x^2\right)^{-1}.$$

Hence

$$\frac{d}{dx}\left(\sqrt{\ln(1+x^2)}\right) = x\left(\ln\left(1+x^2\right)\right)^{-1/2}\left(1+x^2\right)^{-1}.$$

For the last term we use the quotient rule (7) and rules 2 and 3 for the derivatives of the exponential and log functions:

$$\frac{d}{dx}\left(\frac{-4\ln x}{\exp(x)}\right) = 4\frac{\ln x\exp(x) - \exp(x)x^{-1}}{\exp(x)^2} = 4\frac{\ln x - x^{-1}}{\exp(x)}.$$

Finally, summing these derivatives give the result:

$$\frac{d}{dx}\left(10 - 0.5x^2 + \sqrt{\ln(1+x^2)} - \frac{4\ln(x)}{\exp(x)}\right) = -x + x\left(\ln\left(1+x^2\right)\right)^{-1/2}\left(1+x^2\right)^{-1} + 4\frac{\ln x - x^{-1}}{\exp(x)}.$$

Of course, this example is not a standard function. It has merely been provided for readers to practice using the rules for differentiation.

I.1.3.3 Monotonic, Concave and Convex Functions

A differentiable function is *strictly monotonic increasing* if $f'(x) > 0$ for all x. That is, the function always increases in value as x increases. Similarly, we say that f is *strictly monotonic decreasing* if $f'(x) < 0$ for all x. Sometimes we drop the 'strictly', in which case we allow the function to be flat at some points, i.e. to have zero derivative for some x. *Concavity* and *convexity* are related to the second derivative, $f''(x)$:

$$f \text{ is strictly concave if } f''(x) < 0 \text{ and strictly convex if } f''(x) > 0. \qquad \text{(I.1.17)}$$

We drop the 'strictly' if the inequalities are not strict inequalities. Hence:

$$f \text{ is concave if } f''(x) \leq 0 \text{ and convex if } f''(x) \geq 0. \qquad \text{(I.1.18)}$$

For instance, the logarithmic function is a strictly monotonic increasing concave function and the exponential function is a strictly monotonic increasing convex function. Other functions may be concave on some ranges of x and convex on other ranges of x. Any point at which a function changes from being concave to convex, or from convex to concave, is called a *point of inflexion*. At a point of inflexion $f''(x) = 0$. For instance, the cubic polynomial function (a) shown in Figure I.1.8 is concave for $x < 7/3$ and convex for $x > 7/3$. We find this inflexion point 7/3 by setting $f''(x) = 0$, i.e. $6x - 14 = 0$.

If a function is strictly concave then the value of the function at any two points x_1 and x_2 is greater than the corresponding weighted average of the function's values. In other words, the graph of a function that is strictly concave (convex) in the range $[x_1, x_2]$ always lies above (below) the chord joining two points $f(x_1)$ and $f(x_2)$. For instance, the function (b) in Figure I.1.8 is strictly concave. Formally, f is strictly concave if and only if

$$f(px_1 + (1-p)x_2) > pf(x_1) + (1-p)f(x_2) \text{ for every } p \in [0, 1]. \tag{I.1.19}$$

Likewise, if a function is strictly convex then

$$f(px_1 + (1-p)x_2) < pf(x_1) + (1-p)f(x_2) \text{ for every } p \in [0, 1]. \tag{I.1.20}$$

We shall encounter many examples of concave and convex functions in this book. For instance, the price of a bond is a convex monotonic decreasing function of its yield. For this reason we call the second derivative of the bond price with respect to its yield the convexity of the bond; and the Black–Scholes–Merton price of an in-the-money or an out-of-the money option is a convex monotonic increasing function of the implied volatility.

I.1.3.4 Stationary Points and Optimization

When $f'(x) = 0$, x is called a *stationary point* of f. Thus the tangent to the function is horizontal at a stationary point. For instance, the cubic polynomial function (a) in Figure I.1.8 has two stationary points, found by solving

$$\frac{d}{dx}\left(x^3 - 7x^2 + 14x - 8\right) = 3x^2 - 14x + 14 = 0.$$

Using (I.1.1) gives the stationary points $x = 1.451$ and $x = 3.215$, and we can see from the figure that the tangent is indeed horizontal at these two points.

A function f has a *local maximum* at a stationary point x if $f''(x) < 0$ and a *local minimum* if $f''(x) > 0$. For instance, the cubic polynomial function (a) in Figure I.1.8 has second derivative $6x - 14$, which is negative at $x = 1.451$ and positive at $x = 3.215$. Hence the function has a local maximum at $x = 1.451$ and a local minimum at $x = 3.215$. A stationary point that is neither a local maximum nor a local minimum is called a *saddle point*.

Example I.1.3: Identifying stationary points

Find the stationary point of the function $f(x) = x^2 \ln x$, $x > 0$, and identify whether it is a maximum, minimum or saddle point.

Solution We have

$$f'(x) = 2x \ln x + x,$$

$$f''(x) = 2 \ln x + 3.$$

Hence a stationary point in the region $x > 0$ is the point x where $2x \ln x + x = 0$. That is, $\ln x = -1/2$ so $x = \exp(-1/2) = 0.6065$. At this point $f''(x) = 2$ so it is a local minimum.

If, as well as $f'(x) = 0$, we have $f''(x) = 0$ then the point *could* be neither a maximum nor a minimum but a point of inflexion at which $f'(x)$ also happens to be 0. But the point might also be a maximum, or a minimum. The only way we can tell which type of point it is, to find the value of the derivative either side of x. If the derivative is positive (negative) just below x and negative (positive) just above x, then x is a maximum (minimum). If the derivative has the same sign just above and just below x, then x is a saddle point.

I.1.3.5 Integration

The integral of a function is the area between the curve and the x-axis. If the area is above the axis it is positive and if it is below the axis it is negative. When we place limits on the integral sign then the area is calculated between these limits, as depicted in Figure I.1.9. We call this the *definite integral* because the result is a number, as opposed to *indefinite integral*, which is a function.

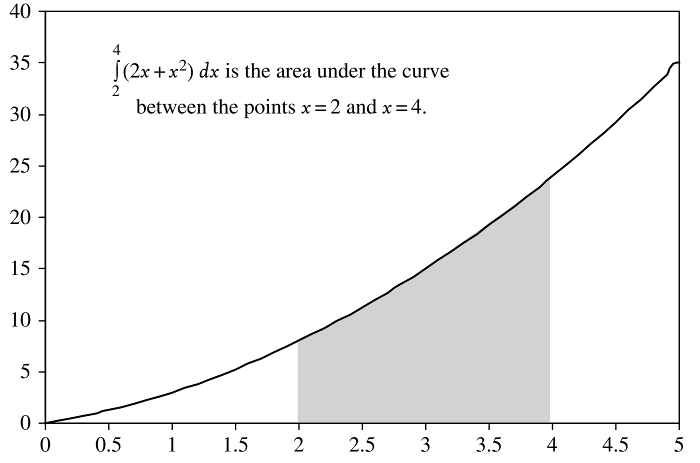


Figure I.1.9 The definite integral

Integration is the opposite process to differentiation:

$$\text{if } f(x) = F'(x) \text{ then } F(x) = \int f(x)\,dx.$$

Rules for integration are the opposite of the rules for differentiation. For instance,[13]

$$\int x^{-1} dx = \ln x + c \text{ and } \int x\,dx = \tfrac{1}{2}x^2 + c,$$

where c is an arbitrary constant. Therefore if we know the rules for differentiation, we can find the integral of a function.

For the definite integral we write

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

[13] The indefinite integral is determined only up to an additive constant because $f(x) = F'(x) \Rightarrow f(x) = (F(x) + c)'$.

where a is the lower limit and b is the upper limit of the range over which the area is calculated. Then, if $f(x) = \int f(x)dx$ we calculate the area as:

$$\int_a^b f(x)\ dx = F(b) - F(a). \tag{I.1.21}$$

EXAMPLE I.1.4: A DEFINITE INTEGRAL

Find the area under the curve defined by the function $f(x) = 2x + x^2$ between the points $x = 2$ and $x = 4$ as depicted in Figure I.1.9.

SOLUTION We have

$$\int_2^4 \left(2x + x^2\right)\ dx = \left[x^2 + \frac{x^3}{3}\right]_2^4 = \left(4^2 + \frac{4^3}{3}\right) - \left(2^2 + \frac{2^3}{3}\right) = 12 + \frac{56}{3} = 30.67.$$

In this example the function value always lies above the x-axis so the area was positive. But an area below the axis is negative. For instance,

$$\int_{-1}^1 x\ dx = \left[\frac{x^2}{2}\right]_{-1}^1 = 1/2 - 1/2 = 0$$

because the area under the line $y = x$ between -1 and 0 is below the axis and equal in size to the area between 0 and 1, which is above the axis.

I.1.4 ANALYSIS OF FINANCIAL RETURNS

This section introduces the concept of the return on a financial asset and the return on a portfolio. We provide notation and definitions in both discrete and continuous time and explain the connection between the two. The *profit and loss* on an asset is the change in price of an asset over a period of time. It is henceforth usually abbreviated to P&L.[14] When the price of the asset or the value of the portfolio is always positive we may also define the *percentage return* as the change in price divided by the current price. We shall see that over short time periods the percentage return may be well approximated by the *log return*.

I.1.4.1 Discrete and Continuous Time Notation

It is usually easier to derive theoretical pricing models for bonds and derivatives assuming that interest rates and the prices of assets can be measured in continuous time, and assuming we know the value of an instrument at any point in time. But in reality we only accrue interest over discrete time intervals and we only observe prices at discrete points in time.

The definition of a return requires the distinction between discrete and continuous time modelling. This is a source of confusion for many students. The continuous and discrete time approaches to financial models were developed independently, and as a result different

[14] The P&L on an asset is the change in value, and this should be distinguished from the change in an interest rate. The latter corresponds to the percentage return on a bond and does *not* represent a profit and loss.

notations are often used for the same thing, or one notation can mean two different things depending on whether the analysis is in discrete or continuous time.

In continuous time we often use $S(t)$ to denote the value of an investment, e.g. the price of a single asset or the value of a portfolio at time t. But in discrete time the notations P_t or p_t are standard. In each case we mean that the asset price or portfolio value is a function of time. And in both cases it is common to drop the time dependence notation, i.e. denote the price or value by S or P depending on whether time is continuous or discrete; their implicit dependence on time is assumed, because the price or value at time t is a random variable.

The use of the Δ notation in discrete and continuous time must also be distinguished. In discrete time it denotes the *first difference operator*, i.e. the difference between two consecutive values, but in continuous time it is used to denote an *increment* in a variable; that is, a small change in the value of a variable. That is, in continuous time ΔX denotes an increment in a random variable X, which is not necessarily linked to a change happening over time. However, Δt is a small amount of time.[15]

We also distinguish between whether one is looking forward or backward in time. Hence the change in price is denoted

$$\Delta P_t = P_t - P_{t-1} \tag{I.1.22}$$

when looking backward, but

$$\Delta P_t = P_{t+1} - P_t \tag{I.1.23}$$

when looking forward.

I.1.4.2 Portfolio Holdings and Portfolio Weights

Consider a simple portfolio with positions on k assets and denote the ith asset price at time t by p_{it}. At time 0 a certain amount is invested in each of these assets, so that the portfolio contains a unit amount n_i of asset i, for $i = 1, 2, \ldots, k$. The set $\{n_1, n_2, \ldots, n_k\}$ is called the vector of *portfolio holdings* at time 0. If we *rebalance* the portfolio at some point in time the holdings in certain assets may change, but without rebalancing the portfolio holdings will be constant over time.

In a *long portfolio* all the holdings are positive; in a *short portfolio* all the holdings are negative; and a portfolio with some negative and some positive holdings is a *long-short portfolio*. Negative holdings may be achieved by either taking a short position in a futures or forward contract or by making a *short sale*. The latter requires using the *repo market* to borrow a security that you do not own, with an agreement to return this security at a later date.[16]

Suppose there is no rebalancing over time, and that there are no dividends if the asset is an equity and no coupons if the asset is a bond. Then at any time $t > 0$, the *value of the portfolio* is the sum of the product of asset prices and holdings,

$$P_t = \sum_{i=1}^{k} n_i p_{it}. \tag{I.1.24}$$

This is positive for a long portfolio, negative for a short portfolio and it may be positive, negative or zero for a long-short portfolio.

---

[15] We should mention that the reason why we do not use capital letters to denote the 'Greeks' of option portfolios in this book is to avoid confusion between the option 'delta' and these other usages of the Δ notation.

[16] A *security* is a tradable financial claim that is usually listed on an exchange, such as a stock or a bond.

The proportion of capital invested in a certain asset i at time t is called the *portfolio weight* on this asset. The portfolio weight on asset i is

$$w_{it} = \frac{n_i p_{it}}{P_t}. \tag{I.1.25}$$

This is positive for a long portfolio, negative for a short portfolio and it may be positive, negative or zero for a long-short portfolio. For a *self-financing portfolio* the sum of the weights is 0 and in a *fully-funded* portfolio the sum of the weights is 1.

When the portfolio's holding in each asset remains constant, i.e. there is *no rebalancing* over time, the proportion of capital invested in each asset changes whenever the price of one of the assets changes. Hence, the portfolio weights change over time unless the portfolio is continually rebalanced whenever an asset price moves to keep the weights constant. The following example illustrates:

Example I.1.5: Portfolio weights

Suppose we invest in two assets whose dollar prices at the beginning of four consecutive years are as shown in Table I.1.1. At the beginning of 2003 we buy 600 units of asset 1 and 200 units of asset 2. Find the portfolio weights and the portfolio value over the rest of the period when the portfolio is not rebalanced.

Table I.1.1 Asset prices

Year	Price of asset 1	Price of asset 2
2003	100	200
2004	125	500
2005	80	250
2006	120	400

Solution In 2003 the portfolio value is $600 \times 100 + 200 \times 200 = \$100{,}000$ and the weight on asset 1 is $60{,}000/100{,}000 = 0.6$, so the weight on asset 2 is 0.4. In 2004 the portfolio value is $600 \times 125 + 200 \times 500 = \$175{,}000$ and the weight on asset 1 is $75{,}000/175{,}000 = 0.429$, so the weight on asset 2 is 0.571. Continuing in this way, the results for all four years are shown in Table I.1.2.

Table I.1.2 Portfolio weights and portfolio value

Year	Price 1	Price 2	Weight 1	Weight 2	Value 1	Value 2	Portfolio Value
2003	100	200	0.600	0.400	60,000	40,000	100,000
2004	125	500	0.429	0.571	75,000	100,000	175,000
2005	80	250	0.490	0.510	48,000	50,000	98,000
2006	120	400	0.474	0.526	72,000	80,000	152,000

The above example shows that if the portfolio is not rebalanced then the portfolio weights will change when the prices of the assets change. That is, constant holdings imply variable weights and constant weights imply variable holdings over time.

I.1.4.3 Profit and Loss

Discrete Time

Consider first the discrete time case, letting P_t denote the value of a portfolio at time t. In a long-short portfolio this can be positive, negative or indeed zero, as we have seen above. The *profit and loss* is the change in the value of the portfolio between two consecutive times. Thus the P&L at time t is either (I.1.22) or (I.1.23) depending on whether we are looking backward or forward in time. Here the subscript t denotes the price at that time and Δ denotes either the:

- *backward difference operator*, in (I.1.22), for instance when we are dealing with historical data; or the
- *forward difference operator*, in (I.1.23), for instance when we are making forecasts of what could happen in the future.

Continuous Time

The forward-looking P&L over a very small 'infinitesimal' time interval of length Δt is $S(t+\Delta t)-S(t)$ and the backward looking P&L is $S(t)-S(t-\Delta t)$. But by definition of the derivative,

$$\lim_{\Delta t \downarrow 0}\left[\frac{S(t+\Delta t)-S(t)}{\Delta t}\right]=\lim_{\Delta t \downarrow 0}\left[\frac{S(t)-S(t+\Delta t)}{\Delta t}\right]=\frac{dS(t)}{dt}. \quad \text{(I.1.26)}$$

In other words, we can use the differential, $dS(t)$ to denote the P&L in continuous time.

Note that P&L is measured in *value terms*, i.e. in the same units as the investment. For instance, if the investment is in a hedge fund with a net asset value measured in US dollars then the P&L is also measured in US dollars. This can sometimes present a problem because if prices have been trending then a P&L of \$1 million today has a different economic significance than a P&L of \$1 million some years ago. For this reason we often prefer to analyse returns, if possible.

I.1.4.4 Percentage and Log Returns

Discrete Time

We shall phrase our discussion in terms of backward-looking returns but the main results are the same for forward-looking returns. Suppose that:

- the portfolio value P_t is always positive;
- there are no interim payments such as dividends on stocks or coupons on bonds.[17]

Then the one-period percentage return on the investment is

$$R_t=\frac{P_t-P_{t-1}}{P_{t-1}}=\frac{\Delta P_t}{P_{t-1}}. \quad \text{(I.1.27)}$$

Whereas the P&L is measured in value terms, the percentage return (often called just the *return*) is a relative change in value, and it is normally quoted as a percentage.

[17] If a dividend or coupon is received between time t and time $t+1$ then this, and any interest accruing on this, should be added to the numerator. For commodities the carry cost can be treated like a negative dividend or coupon. See Section III.2.3.6 for further details on carry costs.

But what is the return if the value of the investment is negative or zero? If the investment *always* has a negative value, for instance when one goes short every asset, then (I.1.27) still gives a meaningful definition. But it is very difficult to define a return on an investment in a long-short portfolio, although the P&L *is* well defined for any type of portfolio. A return must be defined as the P&L relative to some price level. We cannot take the current price because this could be zero! And if it is small but not zero the return can be enormous. For this reason we often analyse long-short portfolios by their P&L rather than by their returns.

If it is absolutely essential to work with returns then some convention must be agreed. For instance we might assume that the P&L is measured relative to the *size* of investment that is necessary to make the trade. If the value can be either positive or negative and we really have to define a return, provided that the price is never zero we could define a one-period return as

$$R_t = \frac{\Delta P_t}{|P_{t-1}|}. \tag{I.1.28}$$

Example I.1.6: Returns on a long-short portfolio

A British bank sells short £2 million of US dollar currency and purchases sterling. The trade was made when the exchange rate was 1.65 US dollars per British pound. Later, when the bank closes out the position, the exchange rate has moved to 1.75. What is (a) the P&L (measured in £s since this is from the British bank's perspective) and (b) the corresponding return? You may assume, for simplicity, that interest rates are zero on both sides of the Atlantic.[18]

Solution When the trade is made the bank has a long position of £2 million in sterling and a short position of $2 \times 1.65 = \$3.3$ million in US dollars. The position has a net value of zero, but the size of the investment is £2 million. When the position is closed the exchange rate has changed to 1.75 US dollars per British pound. So we only need to use

$$\frac{3{,}300{,}000}{1.75} = £1{,}885{,}714$$

to close the position. Hence the bank has made a profit of $£2{,}000{,}000 - £1{,}885{,}714 = £114{,}286$. If this is measured relative to the initial size of the investment, the return is

$$\frac{114{,}286}{2{,}000{,}000} = 5.7143\%.$$

This example calculates the return over only one period. When there are many periods we may still use the same convention, i.e. that the return on a long-short portfolio is the P&L divided by the investment required to take the position. This investment would be the *initial* outlay of funds (i.e. £2 million in the above example) which is fixed over time, plus the funding costs of this outlay, which will change over time.

[18] Hence there is zero interest on the sterling position and the US dollar forward rates are equal to the spot exchange rates used in the question. In reality, we would need to include changes in interest rates to define the total return.

Continuous Time

If the value of a portfolio is strictly positive then the forward-looking percentage return over a time interval of length Δt is

$$R(t) = \frac{(S(t+\Delta t) - S(t))}{S(t)}.$$

Note that

$$1 + R(t) = \frac{S(t+\Delta t)}{S(t)}.$$

If the increment Δt is a very small interval of time then the percentage return is small. Now recall the property (I.1.9) that

$$\ln(1+x) \approx x \quad \text{if } x \text{ is small.}$$

Hence, for small Δt,

$$R(t) \approx \ln(1 + R(t));$$

in other words,

$$R(t) \approx \ln S(t+\Delta t) - \ln S(t).$$

So over small time periods the percentage return is very close to the *log return*. By (I.1.16) we know that, in the limit as the time increment tends to zero, the log return is given by the differential of the log price, $d\ln S(t)$. Hence the log return is the *increment in the log price of the asset*.

I.1.4.5 Geometric Brownian Motion

We use the differential operator to describe the evolution of prices of financial assets or interest rates in continuous time. Let $S(t)$ denote the price of an asset at some future time t, assuming the current time is $t = 0$. Consider the price $S(t)$ at $t > 0$. If there were no uncertainty about this price we might assume that its growth rate is constant. The growth rate is the *proportional* change per unit of time. So if the growth rate is a constant μ, we can write

$$\frac{dS(t)}{dt} = \mu S(t). \tag{I.1.29}$$

Now by the chain rule,

$$\frac{d\ln S(t)}{dt} = \frac{d\ln S(t)}{dS(t)}\frac{dS(t)}{dt} = S(t)^{-1}\frac{dS(t)}{dt}.$$

Thus an equivalent form of (I.1.29) is

$$\frac{d\ln S(t)}{dt} = \mu. \tag{I.1.30}$$

Integrating (I.1.30) gives the solution

$$S(t) = S(0)\exp(\mu t). \tag{I.1.31}$$

Hence, the asset price path would be an exponential if there were no uncertainty about the future price.

However, there *is* uncertainty about the price of the asset in the future, and to model this we add a *stochastic differential* term $dW(t)$ to (I.1.29) or to (I.1.30). The process $W(t)$ is called a *Wiener process*, also called a *Brownian motion*. It is a continuous process that has independent increments $dW(t)$ and each *increment* has a normal distribution with mean 0 and variance dt.[19] On adding uncertainty to the exponential price path (I.1.31) the price process (I.1.29) becomes

$$\frac{dS(t)}{S(t)} = \mu\, dt + \sigma dW(t). \tag{I.1.32}$$

This is an example of a *diffusion process*. Since the left-hand side has the proportional change in the price at time t, rather than the absolute change, we call (I.1.32) *geometric Brownian motion*. If the left-hand side variable were $dS(t)$ instead, the process would be called *arithmetic Brownian motion*. The *diffusion coefficient* is the coefficient of $dW(t)$, which is a constant σ in the case of geometric Brownian motion. This constant is called the *volatility* of the process. By definition $dW(t)$ has a normal distribution with mean 0 and variance dt, so $dS(t)/S(t)$ has a normal distribution with mean μdt and variance $\sigma^2 dt$.[20]

I.1.4.6 Discrete and Continuous Compounding in Discrete Time

Another way of expressing the return (I.1.27) is:

$$1 + R_t = \frac{P_t}{P_{t-1}}. \tag{I.1.33}$$

On the left-hand side we have the *discrete compounding factor*, so called because the above is also equivalent to

$$P_t = (1 + R_t)\, P_{t-1}. \tag{I.1.34}$$

For example, if $P_0 = 100$ and $P_1 = 105$ then by (I.1.33) $R_1 = 5\%$ and (I.1.34) becomes $105 = 1.05 \times 100$.

Alternatively, using Δ to denote the *forward* difference operator, the one-period *forward-looking* return is

$$R_t = \frac{P_{t+1} - P_t}{P_t} = \frac{\Delta P_t}{P_t}, \tag{I.1.35}$$

and turning this equation around now gives

$$P_{t+1} = (1 + R_t)\, P_t. \tag{I.1.36}$$

Definition (I.1.34) applies when returns are discretely compounded, but under continuous compounding we use *log returns*. The one-period historical log return is defined as

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} = \Delta \ln P_t. \tag{I.1.37}$$

Another way of expressing (I.1.37) is:

$$\exp(r_t) = \frac{P_t}{P_{t-1}}. \tag{I.1.38}$$

[19] See Section I.3.3.4 for an introduction to the normal distribution.
[20] See Section I.3.7 for further details on stochastic processes in discrete and continuous time.

On the left-hand side we have the *continuous compounding factor*, so called because the above is also equivalent to

$$P_t = \exp(r_t)\, P_{t-1}. \tag{I.1.39}$$

For example, if $P_0 = 100$ and $P_1 = 105$ then by (I.1.37) $r_1 = \ln(1.05) = 4.879\%$ and (I.1.39) becomes:

$$105 = \exp(0.04879) \times 100 = 1.05 \times 100.$$

Our numerical examples above have shown that for the same prices at the beginning and the end of the period the continuously compounded return (i.e. the log return) is *less* than the discretely compounded return. Other examples of discrete and continuous compounding are given in Section III.1.2.

I.1.4.7 Period Log Returns in Discrete Time

Using (I.1.9) we can write the log return over one period as

$$\ln\left(\frac{P_t}{P_{t-1}}\right) = \ln\left(\frac{P_t - P_{t-1}}{P_{t-1}} + 1\right) \approx \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{I.1.40}$$

Similarly, the forward-looking one-period log return is

$$r_t = \ln\left(\frac{P_{t+1}}{P_t}\right) = \ln P_{t+1} - \ln P_t = \Delta \ln P_t \tag{I.1.41}$$

and

$$\ln\left(\frac{P_{t+1}}{P_t}\right) \approx \frac{P_{t+1} - P_t}{P_t}. \tag{I.1.42}$$

So the log return is approximately equal to the return but, as shown in Section I.1.4.4, the approximation is good *only* when the return is small. It is often used in practice to measure returns at the daily frequency.

In market risk analysis we often need to predict risk over several forward-looking periods such as 1 day, 10 days or more. The forward-looking *h-period return* is

$$R_{ht} = \frac{P_{t+h} - P_t}{P_t} = \frac{\Delta_h P_t}{P_t},$$

where Δ_h denotes the h-period forward difference operator. The *h-period log return* is

$$r_{ht} = \ln\left(\frac{P_{t+h}}{P_t}\right) = \ln P_{t+h} - \ln P_t = \Delta_h \ln P_t.$$

Note that

$$\begin{aligned}\ln P_{t+h} - \ln P_t = {} & \ln P_{t+h} + \left[-\ln P_{t+h-1} + \ln P_{t+h-1}\right] + \left[-\ln P_{t+h-2} + \ln P_{t+h-2}\right] + \ldots + \\ & \left[-\ln P_{t+1} + \ln P_{t+1}\right] - \ln P_t \\ = {} & \left[\ln P_{t+h} - \ln P_{t+h-1}\right] + \left[\ln P_{t+h-1} - \ln P_{t+h-2}\right] + \ldots + \left[\ln P_{t+1} - \ln P_t\right].\end{aligned}$$

That is,

$$r_{ht} = \sum_{i=0}^{h-1} r_{t+i}, \tag{I.1.43}$$

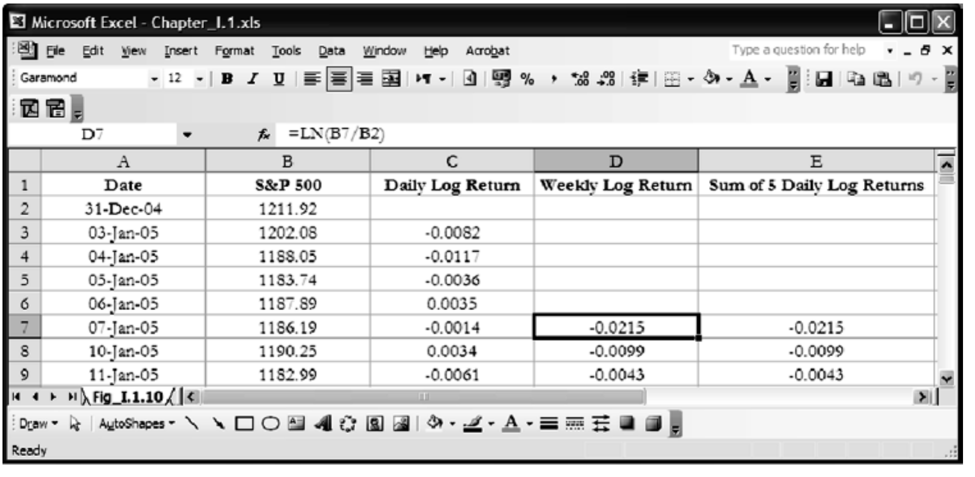
or equivalently,

$$\Delta_h \ln P_t = \sum_{i=0}^{h-1} \Delta \ln P_{t+i}.$$

Hence the h-period log return is the sum of h consecutive one-period log returns. This property makes log returns very easy to analyse and it is one of the main reasons we prefer to work with log returns whenever possible.

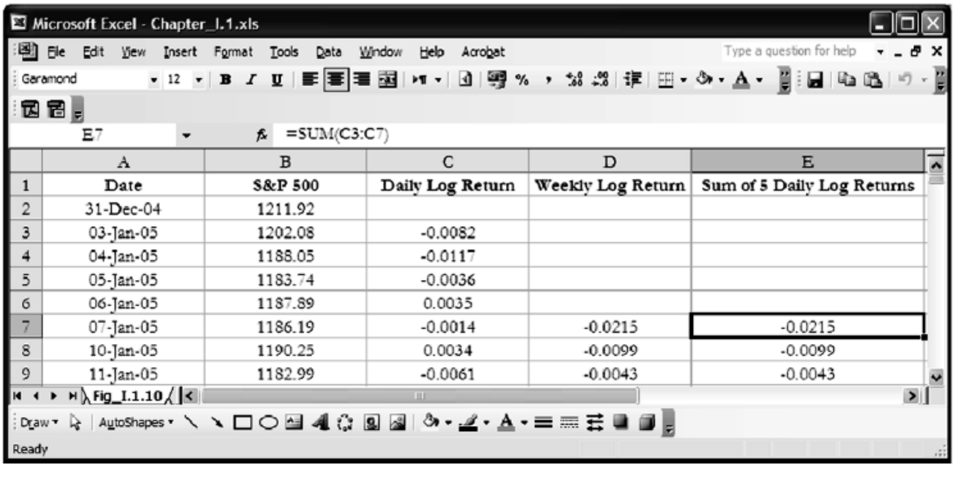
When analysing historical time series data we often convert the historical prices into a return series. If the data are sampled at a monthly frequency, the one-period return is a monthly return; similarly, weekly data give weekly returns, when measured over one time period, and daily data give daily returns. We can also create weekly or monthly log returns from daily log returns, using a result similar to (I.1.43), viz.

$$r_{ht} = \sum_{i=1}^{h} r_{t-i}, \tag{I.1.44}$$



D7 =LN(B7/B2)

	A	B	C	D	E
1	Date	S&P 500	Daily Log Return	Weekly Log Return	Sum of 5 Daily Log Returns
2	31-Dec-04	1211.92			
3	03-Jan-05	1202.08	-0.0082		
4	04-Jan-05	1188.05	-0.0117		
5	05-Jan-05	1183.74	-0.0036		
6	06-Jan-05	1187.89	0.0035		
7	07-Jan-05	1186.19	-0.0014	-0.0215	-0.0215
8	10-Jan-05	1190.25	0.0034	-0.0099	-0.0099
9	11-Jan-05	1182.99	-0.0061	-0.0043	-0.0043



E7 =SUM(C3:C7)

	A	B	C	D	E
1	Date	S&P 500	Daily Log Return	Weekly Log Return	Sum of 5 Daily Log Returns
2	31-Dec-04	1211.92			
3	03-Jan-05	1202.08	-0.0082		
4	04-Jan-05	1188.05	-0.0117		
5	05-Jan-05	1183.74	-0.0036		
6	06-Jan-05	1187.89	0.0035		
7	07-Jan-05	1186.19	-0.0014	-0.0215	-0.0215
8	10-Jan-05	1190.25	0.0034	-0.0099	-0.0099
9	11-Jan-05	1182.99	-0.0061	-0.0043	-0.0043

Figure I.1.10 The h-period log return is the sum of h consecutive one-period log returns or equivalently, $\Delta_h \ln P_t = \sum_{i=1}^{h} \Delta \ln P_{t-i}$. For instance, the Excel spreadsheet for Figure I.1.10 takes daily data on the S&P 500 index from 31 December 2004 to 30 December 2005 and computes the 5-day log return by (a) taking the difference of the log prices 5 days apart and (b) summing five daily log returns. The answers are the same.

I.1.4.8 Return on a Linear Portfolio

We now prove one of the most fundamental relationships in portfolio mathematics: that the percentage return on a long-only portfolio may be written as a weighted sum of the asset returns, with weights given by the portfolio's weights at the *beginning* of the period.

To prove this, we denote the one-period percentage return on the portfolio from time 0 to time 1 by

$$R = \frac{P_1 - P_0}{P_0}.$$

Suppose there are k assets in the portfolio. Then, by definition,

$$1 + R = \frac{P_1}{P_0} = \frac{\sum_{i=1}^{k} n_i p_{i1}}{P_0} = \sum_{i=1}^{k} \frac{n_i p_{i0}}{P_0} \frac{p_{i1}}{p_{i0}}.$$

In other words, letting R_i denote the one-period return on asset i and w_i denote the portfolio weight on asset i at the beginning of the period, we have

$$1 + R = \sum_{i=1}^{k} w_i (1 + R_i) = \sum_{i=1}^{k} w_i + \sum_{i=1}^{k} w_i R_i = 1 + \sum_{i=1}^{k} w_i R_i.$$

Hence,

$$R = \sum_{i=1}^{k} w_i R_i. \tag{I.1.45}$$

I.1.4.9 Sources of Returns

Since (I.1.45) defines the portfolio as a linear function of the returns on its constituent assets we call such a portfolio a *linear portfolio*. The relationship (I.1.45) shows that the return, and therefore also the risk, of the portfolio may be attributed to two sources:

- changes in the prices of the individual assets;
- changes in portfolio weights.

The following example illustrates the effect of these two sources of risk and return. In the first case the portfolio return is due to both changes in asset prices and portfolio weights, but in the second case the portfolio returns only comes from changes in asset prices because we rebalance the portfolio continually so that its weights are held constant.

EXAMPLE I.1.7: PORTFOLIO RETURNS

Continuing Example I.1.5, find the portfolio's annual returns under the assumption that:

(i) portfolio holdings are held constant, i.e. there is no rebalancing; and
(ii) portfolio weights are held constant, i.e. the portfolio is rebalanced every quarter so that 60% of the portfolio's value is invested in asset 1 and 40% is invested in asset 2.

Solution

(i) The values of the constant holdings portfolio in each year are given in the solution to Example I.1.5. The returns are summarized as follows:

$$2003–2004: 75/100 = 75\%,$$

$$2004–2005: -77/175 = -44\%,$$

$$2005–2006: 54/98 = 55.1\%.$$

(ii) For the constant weights portfolio we use (I.1.45) to find the portfolio return as the weighted sum of the individual asset returns. The returns are different from those obtained in part (i). For comparison with Table I.1.2 we compute the portfolio value using (I.1.36) and the beginning-of-year holdings in each asset that keep the weights constant. The results are summarized in Table I.1.3.

Table I.1.3 Portfolio returns

Year	Price 1	Price 2	Return 1	Return 2	Portfolio return	Portfolio value	Holding 1	Holding 2
2003	100	200				100,000	600	200
2004	125	500	25%	150%	75%	175,000	840	140
2005	80	250	−36%	−50%	−41.6%	102,200	767	164
2006	120	400	50%	60%	54%	157,388	787	157

In practice neither portfolio weights nor portfolio holdings remain constant. To rebalance a portfolio continually so that the portfolio weights are the same at the beginning of every day or at the beginning of every week is difficult, if not impossible. Even at the monthly frequency where constant weights rebalancing may be feasible, it would incur very high transactions costs. Of course portfolios are rebalanced. The holdings do change over time, but this is normally in accordance with the manager's expectations of risk and return.

However, in theoretical models of portfolio risk and return it greatly simplifies the analysis to make the constant weights assumption. It can also make sense. Ex post, risk and return can be calculated using the actual historical data on asset values and holdings. For ex post analysis a constant weights assumption is not necessary, but it is sometimes made for simplification. But when we are forecasting risk and return we want to capture the risk and returns arising from the risks and returns of the assets, not from portfolio rebalancing. So for ex ante analysis it is standard to keep portfolio weights constant. This way, the risk and return on the current portfolio can be forecast using changes in asset returns as the *only* influence on portfolio value changes.

I.1.5 FUNCTIONS OF SEVERAL VARIABLES

A function $f(x, y)$ of two variables x and y has a three-dimensional graph, such as that shown in Figure I.1.11 below. We can think of it like a landscape, with a mountaintop at a local maximum and a valley at a local minimum. More generally, a real-valued function of

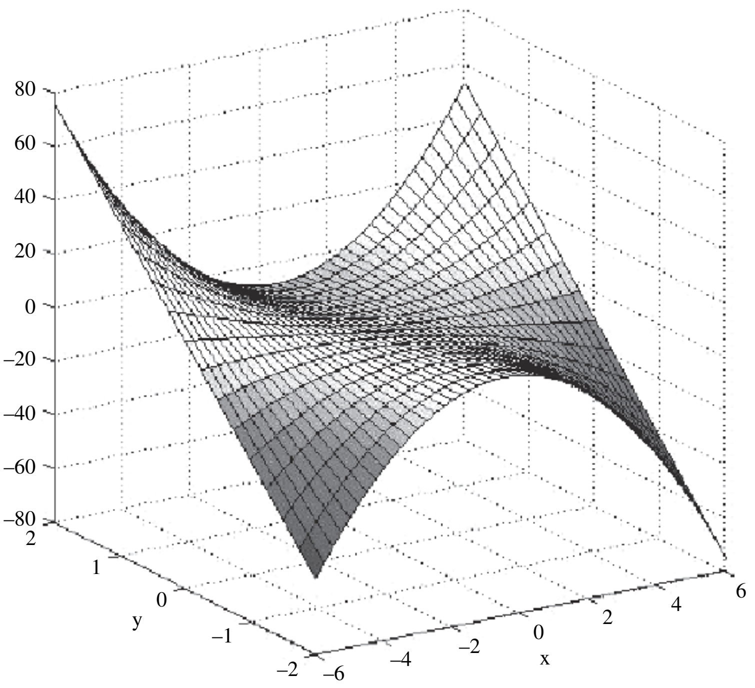


Figure I.1.11 Graph of the function in Example I.1.8

several real variables $f(x_1, \ldots, x_n)$ is an object in $(n+1)$-dimensional space. This section introduces the two types of derivatives of functions of several variables, partial derivatives and total derivatives. We characterize the stationary points of a function of several variables and show how to optimize such a function.

I.1.5.1 Partial Derivatives: Function of Two Variables

The first partial derivative gives the slope of the tangent to the curve that is found by cutting through the graph of the function, holding one of the variables constant, so that we cut through the graph in a direction parallel to one axis. The partial derivative with respect to x is obtained by holding y constant and differentiating with respect to x, and we denote this by either f_x or $\dfrac{\partial f}{\partial x}$. Similarly, f_y or $\dfrac{\partial f}{\partial y}$ is obtained by holding x constant and differentiating with respect to y.

The second partial derivatives are obtained by partial differentiation of the first partial derivatives. There are three of them:

$$f_{xx}, f_{xy} \text{ and } f_{yy}, \text{ also denoted } \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y} \text{ and } \frac{\partial^2 f}{\partial y^2}.$$

Note that it does not matter in which order we differentiate the cross derivative, the result is the same: $f_{xy} = f_{yx}$.

I.1.5.2 Partial Derivatives: Function of Several Variables

Partial derivatives of a function of more than two variables are defined analogously. Sometimes we use number subscripts rather than letters to denote the derivatives. Hence for a function $f(x_1, \ldots, x_n)$ the first partial derivatives may be denoted $f_1, f_2, \ldots, f_n$ and the second partial derivatives by $f_{11}, f_{12}, \ldots$ and so on.

It is often more convenient to use *vector* and *matrix* notation when dealing with functions of several variables. A vector is a column and a matrix is a rectangular array, and we use bold lower-case letters to denote vectors and bold upper-case letters to denote matrices.[21] Hence we could write

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix},$$

so we may write the function $f(x_1, \ldots, x_n)$ as $f(\mathbf{x})$ for short. The vector $\mathbf{g}$ is the vector of first partial derivatives of f and is called the *gradient vector* of f. $\mathbf{H}$ is the matrix of second partial derivatives and it is called the *Hessian matrix* of f.

I.1.5.3 Stationary Points

The stationary points of a function of several variables are found by setting all the partial derivatives to zero, i.e. setting the gradient vector $\mathbf{g} = \mathbf{0}$, and solving the resulting set of equations. These equations are called the *first order conditions*. The *second order conditions* are to do with the Hessian. They tell us what type of stationary point we have found using the first order conditions: If the Hessian is *positive definite* (*negative definite*) at the stationary point we have a *local minimum* (*local maximum*).[22] Otherwise we could have any type of stationary point: we can only find out which type it is by plotting the graph in the small region around the point. Note that functions of two or more variables can take quite complex shapes. We have local maxima and minima, but we also have *saddle points* which are a maximum in the direction of some variables and a minimum in the direction of others.

EXAMPLE I.1.8: STATIONARY POINTS OF A FUNCTION OF TWO VARIABLES

Find the stationary points, if any, of the function $f(x, y) = x^2y - 2x - 4y$.

SOLUTION Taking the first and second partial derivatives gives the gradient vector and Hessian matrix:

$$\mathbf{g} = \begin{pmatrix} 2xy - 2 \\ x^2 - 4 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}.$$

Setting $\mathbf{g} = \mathbf{0}$, for the stationary points, gives two equations:

$$xy = 1 \quad \text{and} \quad x^2 = 4.$$

The second equation gives $x = \pm 2$ and using the first equation gives two stationary points:

$$(2, 0.5) \quad \text{and} \quad (-2, -0.5).$$

The Hessian matrix at each point is:

$$\begin{pmatrix} 1 & 4 \\ 4 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 & -4 \\ -4 & 0 \end{pmatrix}.$$

---

21 See Section I.2.1 for an introduction to vectors and matrices.

22 See Section I.2.4 for the definitions of positive and negative definite matrices and for methods for discerning whether a given matrix is positive or negative definite.

Neither matrix is positive definite or negative definite. They are not even semi-definite. So we need to consider the function itself to find out what type of stationary points we have found.

Since $x^2 = 4$ at the stationary points we have $f(x, y) = -2x$ so the function has the value 4 for all y when $x = -2$, or the value -4 for all y when $x = 2$. Thus in the y direction we have neither a maximum nor a minimum: the function is just a horizontal line at this point. In the x direction we have a minimum at the point (2, 0.5) and a maximum at the point (−2, −0.5), as can be seen from Figure I.1.11.

I.1.5.4 Optimization

An *unconstrained optimization* problem is to find the global maximum or minimum of a function of several variables. It is written

$$\max_{\mathbf{x}} f(\mathbf{x}), \tag{I.1.46}$$

which is shorthand for 'find the value of $\mathbf{x}$ for which $f(\mathbf{x})$ takes its maximum value'.[23] In this context the function $f(\mathbf{x})$ is called the *objective function*, because it is the object of the maximization. A common example of unconstrained optimization in finance is the maximization of a likelihood function, used to estimate the parameters of a distribution of asset prices or asset returns.[24]

We know from Section I.1.5.3 that (I.1.46) is solved by setting the gradient vector to zero, giving n equations in n unknowns, which we hope can be solved to find the local stationary points. Then examining the Hessian matrix for positive definiteness, or evaluating the function at and around these points, will tell us whether a given stationary point is a local maximum, a local minimum or neither. Once we have found all possible local maxima and minima, the global maximum (if it exists) is the one of the local maxima where the function takes the highest value and the global minimum (if it exists) is the one of the local minima where the function takes the lowest value.

More generally, a *constrained optimization* problem takes the form

$$\max_{\mathbf{x}} f(\mathbf{x}) \quad \text{such that} \quad \mathbf{h}(\mathbf{x}) \le \mathbf{0}, \tag{I.1.47}$$

where $\mathbf{h}(\mathbf{x}) \le 0$ is a set of linear or non-linear equality or inequality constraints on $\mathbf{x}$. Examples of constrained optimization in finance include the traditional portfolio allocation problem, i.e. how to allocate funds to different types of investments when the investor has constraints such as *no short sales* and/or at least 50% of his capital must be in UK bonds.

To find a constrained optimum we must introduce new variables, which are called *Lagrange multipliers* and denoted $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$, where k is the number of constraints on the objective. Then we augment the objective function by subtracting the product of the Lagrange multipliers and their constraints, thus forming the *Lagrangian function*. Next we solve (I.1.47) by maximizing the Lagrangian, i.e. we maximize

$$L(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_k) = f(x_1, \ldots, x_n) - \sum_{i=1}^{k} \lambda_i h_i(x_1, \ldots, x_n) \tag{I.1.48}$$

---

[23] Note that finding a minimum of $f(\mathbf{x})$ is equivalent to finding a maximum of $-f(\mathbf{x})$, so it is without loss of generality that we have used the maximum here.

[24] Maximum likelihood estimation is introduced in Section I.3.6.

or, in matrix notation,

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda}.\mathbf{h}(\mathbf{x}).$$

Now the first order conditions are obtained by setting all partial derivatives of the Lagrangian to zero, including the partial derivatives with respect to the Lagrange multipliers. Of course, taking the partial derivatives with respect to the Lagrange multipliers just returns us to the conditions $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. So if the original constraints have any inequalities then we need to introduce another set of conditions, called the *Kuhn–Tucker conditions*, which are added to the first order conditions as we identify the stationary points.

The second order conditions for maximizing or minimizing (I.1.48) are on the Hessian of second partial derivatives of the Lagrangian. If this is positive definite then we have a minimum and if this is negative definite then we have a maximum. If the Hessian is neither then we have to explore the *constrained* function's values in the region of the stationary point.

Example I.1.9: Constrained optimization[25]

Find the optimal values of x and y for the problem

$$\max_{x,y}\ x^2 + xy - y^2 \quad \text{such that} \quad x + y = 4.$$

Solution The Lagrangian is

$$L(x, y, \lambda) = x^2 + xy - y^2 - \lambda\,(x + y - 4)\,,$$

and differentiating this gives the first order conditions:

$$2x + y = \lambda,$$

$$x - 2y = \lambda,$$

$$x + y = 4.$$

Their solution is

$$x = 6,\ \ y = -2,\ \ \lambda = 10,$$

so the stationary point is at (6, −2) at which point $x^2 + xy - y^2 = 20$. The Hessian matrix is

$$\begin{pmatrix} 2 & 1 & -1 \\ 1 & -2 & -1 \\ -1 & -1 & 0 \end{pmatrix}.$$

In this case we can tell at a glance that this matrix is neither positive nor negative definite. Positive definite matrices must have positive diagonal elements, and negative definite matrices must have negative diagonal elements.[26] Since the diagonal elements are 2, −2 and 0, the Hessian is neither positive nor negative definite. Thus the point could be a maximum, a minimum or a saddle. We can only tell which type of point it is by examining the values of the constrained function in the region of the stationary point. The constrained function

[25] This problem is easily treated by elementary methods, since we may substitute $y = 4 - x$ into the objective function and thence maximize a function of a single variable. However, we want to illustrate the use of the Lagrangian function to solve constrained optimization problems, so we shall find the solution the long way in this example.

[26] See Section I.2.4.

is obtained by substituting in the constraint to the objective function. In this case the constrained function is $-x^2 + 12x - 16$ and it can easily be seen that when x is a little more or less than 6, the function has a value less than 20. Hence the point $(6, -2)$ is a constrained maximum.

I.1.5.5 Total Derivatives

The partial derivative examines what happens to the function when only one variable changes and the others are held fixed. The total derivative examines what happens to the function when all the variables change at once. We express this in terms of the differential operator, defined analogously to (I.1.16). For instance, consider a function of two variables. The *total differential* gives the *incremental change* in the function when each variable changes by a small amount. So if dx and dy are the increments in x and y, the total differential is defined as

$$df(x, y) = f(x + dx, y + dy) - f(x, y), \tag{I.1.49}$$

and the first order approximation to the total derivative is

$$df = f_x dx + f_y dy. \tag{I.1.50}$$

The extension to functions of more than two variables is obvious.

EXAMPLE I.1.10: TOTAL DERIVATIVE OF A FUNCTION OF THREE VARIABLES

Find an expression for the total derivative of the function

$$f(x, y, z) = x^2 y + 2xyz - z^3$$

and write down the total derivative at the point $(x, y, z) = (1, -1, 0)$.

SOLUTION

$$f_x = 2xy + 2yz, \quad f_y = x^2 + 2xz, \quad f_z = 2xy - 3z^2,$$

Hence

$$df = (2xy + 2yz)\, dx + \left(x^2 + 2xz\right) dy + \left(2xy - 3z^2\right) dz.$$

At the point $(x, y, z) = (1, -1, 0)$,

$$df = -2dx + dy - 2dz.$$

I.1.6 TAYLOR EXPANSION

This section introduces a mathematical technique that is very commonly applied to analyse the risk and return of portfolios whose value function is a non-linear function of the underlying asset prices (or, if the portfolio contains bonds, a non-linear function of interest rates). Taylor expansion techniques also provide useful approximations to the values of many theoretical functions, such as variance. For this reason the Taylor expansion technique will be relied on extensively throughout all four volumes.

I.1.6.1 Definition and Examples

Let $f(x)$ be a non-linear function with derivatives up to the nth order. We denote by $f'(x)$ the first derivative of f with respect to x and by $f''(x), f'''(x), \ldots, f^{(n)}(x)$ the second, third and higher order derivatives of f, assuming these exist. A *Taylor approximation* of $f(x)$ about a particular point x_0 gives a polynomial approximation to $f(x)$ in the region of x_0.

The nth order Taylor approximation of $f(x)$ about x_0 is

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) + \ldots + \frac{(x - x_0)^n}{n!}f^{(n)}(x_0), \qquad \text{(I.1.51)}$$

where $f^{(n)}(x_0)$ is shorthand for $f^{(n)}(x)\big|_{x=x_0}$. That is, we first take the derivative of the function to obtain another function of x, and then put in the value $x = x_0$. The expansion (I.1.51) gives an nth order approximation to the function that is valid only in a small region around $x = x_0$, that is, it is a *local approximation*.

This may look rather daunting at first, but a simple example of its application should clarify the idea. In the following we find a cubic polynomial to a function, which is a reasonably good approximation for values of x not too far from $x = 1$, by taking a third order Taylor expansion.

EXAMPLE I.1.11: TAYLOR APPROXIMATION

Find a third order Taylor approximation to the function

$$f(x) = x^3 - 2\ln x$$

about the point $x = 1$.

SOLUTION

$$\begin{aligned} f(1) &= 1, \\ f'(x) &= 3x^2 - 2x^{-1} \Rightarrow f'(1) = 1, \\ f''(x) &= 6x + 2x^{-2} \Rightarrow f''(1) = 8, \\ f'''(x) &= 6 - 4x^{-3} \Rightarrow f'''(1) = 2. \end{aligned}$$

Hence the third order Taylor expansion about $x = 1$ is

$$f(x) \approx 1 + (x - 1) + 4(x - 1)^2 + \frac{(x - 1)^3}{3},$$

which simplifies to the cubic polynomial

$$f(x) \approx \frac{x^3}{3} + 3x^2 - 6x + \frac{11}{3}.$$

For values of x that are very close to $x = 1$ this approximation is fairly accurate. For example, if $x = 1.02$ then the actual value of $f(x)$ is 1.021603 and the value of $f(x)$ approximated by the above cubic is 1.021597. But for values of x that are not very close to $x = 1$ this approximation is not accurate. For example, if $x = 1.5$ then the actual value of $f(x)$ is 2.564 but the value of $f(x)$ approximated by the above cubic is 2.458.

Again consider the nth order Taylor expansion (I.1.51) of an n times differentiable function $f(x)$ about a fixed point x_0, but now let us write x in place of x_0 and $x + \Delta x$ in place of x. This way we obtain the nth order Taylor approximation to the *change* in the function's value when x changes by a small amount Δx as

$$f(x + \Delta x) - f(x) \approx f'(x)\Delta x + \frac{1}{2!}f''(x)\,(\Delta x)^2 + \ldots + \frac{f^{(n)}}{n!}\,(\Delta x)^n. \qquad \text{(I.1.52)}$$

The form (I.1.52) is very commonly used in finance, as we shall see throughout the course of this text. But before providing a simple example of its application we must digress to introduce some financial terminology.

I.1.6.2 Risk Factors and their Sensitivities

A typical portfolio contains hundreds of positions on different risky assets, including stocks, bonds or commodities. To model the risk and return on a large portfolio we map the portfolio to its *risk factors*.[27] The risk factors include the prices of broad market indices, or futures on these indices, foreign exchange rates (for international portfolios) and zero coupon market interest rates of different maturities in the domestic currency and in each foreign currency to which the portfolio is exposed. We call a portfolio *linear* or *non-linear* depending on whether its price is a linear or a non-linear function of its risk factors. The standard example of a non-linear portfolio is one that contains options. Each option price is a non-linear function of the underlying asset price and the underlying asset volatility.

The *risk factor sensitivity* of an asset, or a portfolio of assets, measures the change in its price when a risk factor changes, holding all other risk factors constant. Risk factor sensitivities are given special names depending on the asset class. For instance, in a stock portfolio the risk factor sensitivities are called *factor betas*. The *market beta* of a stock portfolio measures the portfolio's price sensitivity to movements in the broad market index risk factor.

The measurement of market risk requires large portfolios to be 'mapped' to their risk factors. This consists of identifying the risk factors and then measuring the portfolio's sensitivity to each of these risk factors. For instance, in Chapter II.1 we introduce the *factor model* representation of a stock portfolio and show how regression is used to estimate the factor betas. But in a non-linear portfolio the mapping is based on Taylor expansion and the risk factor sensitivities are calculated using analytic formulae or using simple numerical differentiation.

I.1.6.3 Some Financial Applications of Taylor Expansion

For the moment let us regard the underlying asset price S as the only risk factor of an option, and so we denote its price by $g(S)$. We define the option *delta* and *gamma* as the first and second derivatives, $\delta = g'(S)$ and $\gamma = g''(S)$.[28] Taking $n = 2$ in (I.1.52), we obtain

[27] The mapping methodology depends on the type of assets in the portfolio, and is described in detail in Chapter III.5.
[28] In general the *delta* and the *gamma* are the first and second order price sensitivities of an option or of an options portfolio. They are calculated by taking the first and second partial derivative of the option price with respect to the price of the underlying. They measure the change in portfolio value for small changes in the underlying price. See Section III.3.4 for further details.

the following approximation to the change in the option price when there is a small change of an amount ΔS in S:

$$g(S+\Delta S)-g(S)\approx\delta\Delta S+\frac{1}{2}\gamma(\Delta S)^2. \tag{I.1.53}$$

The approximation (I.1.53) is called a *delta–gamma approximation*. It shows, amongst other things, that portfolios with positive gamma have less price risk than linear portfolios and that those with negative gamma have more price risk than linear portfolios. To see why, suppose we have a linear portfolio with $\delta=0.5$. If the underlying price increases by €1 then the portfolio price will increase by €0.50 and if the underlying price decreases by €1 then the portfolio price will decrease by €0.50. Now consider two options, Option A has $\delta=0.5$ and $\gamma=0.1$ and option B has $\delta=0.5$ and $\gamma=-0.1$. If the underlying price increases by €1 then option A price increases by €0.55 but option B price only increases by €0.45; and if the underlying price decreases by €1 then option A price decreases by €0.45 but option B price decreases by €0.55. Hence option A is preferable to the linear position but option B is not. In general, positive gamma reduces risk but negative gamma increases risk.

In Chapter III.1 we shall also use a second order Taylor approximation for the P&L of a bond portfolio: this is called the *duration–convexity approximation*. The duration is the coefficient on the linear term in the expansion and the convexity is the coefficient on the quadratic term. These coefficients are called the *interest rate sensitivities* of the portfolio. They are found by differentiating the portfolio price with respect to its yield.

I.1.6.4 Multivariate Taylor Expansion

In this subsection we consider Taylor approximations to the change in the value of a function of several variables, using a function of two variables $f(x,y)$ for illustration. The second order Taylor approximation to the change in the function's value when both x and y change by small amounts Δx and Δy is

$$f(x+\Delta x,y+\Delta y)-f(x,y)\approx f_x\Delta x+f_y\Delta y+\frac{1}{2}\left[f_{xx}(\Delta x)^2+f_{yy}(\Delta y)^2+2f_{xy}\Delta x\Delta y\right], \tag{I.1.54}$$

where we use subscripts to denote partial derivatives, as defined in Section I.1.5.1.

Higher order multivariate Taylor expansions are obtained using higher partial derivatives, for instance the multivariate Taylor approximation to the third order in x and the first order in y is, ignoring the cross derivatives,

$$f(x+\Delta x,y+\Delta y)-f(x,y)\approx f_x\Delta x+f_y\Delta y+\frac{1}{2}f_{xx}(\Delta x)^2+\frac{1}{3!}f_{xxx}(\Delta x)^3. \tag{I.1.55}$$

In Chapter III.5 we shall use multivariate Taylor expansions to approximate the P&L of an options portfolio. For instance, the *delta–gamma–vega approximation* for a single option is based on a Taylor approximation of a function of two variables, i.e. the underlying asset price and its volatility, with a second order approximation to price changes and first order approximation to volatility changes, again ignoring the cross derivatives. In addition to delta and gamma, this 'Greeks' approximation uses the *vega* of an option. The option vega is the sensitivity of an option price to volatility, i.e. it measures the change in the option price for small changes in the underlying volatility. It is calculated as the first partial derivative of the option price with respect to the volatility.

I.1.7 SUMMARY AND CONCLUSIONS

The *natural log function* is a strictly concave monotonic increasing function that is defined only for positive values of x and takes the value 0 when $x = 1$. The *exponential function* is a strictly convex monotonic increasing function whose value is always positive[29] and takes the value 1 when $x = 0$. They are inverse functions, that is

$$\ln(\exp(x)) = x \text{ and } \exp(\ln(x)) = x.$$

The exponential translates sums to products but the log function translates products to sums, which is often more useful. We usually prefer to use log usually returns rather than ordinary returns, because they are additive.

Loosely speaking, *continuous functions* have no jumps and *differentiable* functions have no corners. *Strictly monotonic functions* either always increase or always decrease, and their curvature determines their *concavity* or *convexity*. We have introduced the *derivative* of a function as the slope of the tangent line and the *integral* of a function as the area under the graph. Integration is the opposite of differentiation. Both the exponential and the natural log function are continuous and differentiable. Since the slope of the exponential function is equal to the value of exponential function at that point, the derivative of $\exp(x)$ is $\exp(x)$. However, the exponential is the only function that has this special property; in general we find derivatives using certain rules that can be derived from first principles.

Higher derivatives are derivatives of derivatives. Provided a function is sufficiently smooth we can find second, third, fourth, ... derivatives by successive differentiation. But if we hit a corner or a jump, no derivative exists at that point. *Stationary points* occur when the first derivative is zero, i.e. the tangent line is flat. These points can be *maxima, minima* or *points of inflexion*. We can usually find which type of point it is by examining the sign of the second derivative. Finding a stationary point is more generally known as an *optimization problem*.

Simple *portfolios* can be characterized by their positions in financial assets as: long-only, when portfolio weights are positive; short-only, when portfolio weights are negative; or long-short, when weights can be positive or negative. We are usually interested in the return on an investment portfolio, but in the case of a long-short portfolio, which may of course have a value of 0, this is difficult to define, so we usually work with the *profit and loss* instead, which can always be defined. Returns and P&L can be defined in both discrete and continuous time. In discrete time we distinguish between *discrete compounding* and *continuous compounding* of returns. The continuously compounded return is called the *log return* and in discrete time the log return is approximately equal to the return over very short time intervals. Profit and loss, returns and log returns can be forward-looking, as when we try to forecast their risk, or backward looking, as when we analyse historical data.

Functions of one or more variables are very common in financial analysis. For instance, the price of a portfolio may be approximated by a simple function of several risk factors. Common risk factors include equity indices, exchange rates and interest rates and, for options portfolios, volatility. The *partial derivatives* of the portfolio price with respect to the risk factors are called the risk factor sensitivities of the portfolio. A portfolio is classified as linear or non-linear according to whether it is a linear or a non-linear function of its risk

[29] Again (as for the natural logarithm function–see note 11) this is not strictly true.

factors. A typical example of a linear portfolio is a portfolio of cash positions on stocks, or a portfolio of futures contracts on equity indices. But as soon as we introduce options the portfolio becomes a non-linear function of its risk factors.

Taylor expansions have numerous applications to simplify complex functions in finance, such as the price of a portfolio of options or bonds. The coefficient of x^n in the Taylor expansion is given by the nth derivative, evaluated at the current value of x, and divided by $n!$. A Taylor approximation is only a *local approximation* to the change in the function's value, i.e. it is only a good approximation for *small* increments. The more terms used in the expansion the more accurate the approximation.