Chapter 2

Statistical foundations and dealing with data

The population and the sample

- The population is the total collection of all objects to be studied.
- The population may be either finite or infinite, while a sample is a selection of just some items from the population.
- A population is finite if it contains a fixed number of elements.
- In general, either all of the observations for the entire population will not be available, or they may be so many in number that it is infeasible to work with them, in which case a sample of data is taken for analysis.
- The sample is usually random, and it should be representative of the population of interest.
- A random sample is one in which each individual item in the population is equally likely to be drawn.

Probability and probability distributions – some definitions

- A random variable can take any value from a given set
- A discrete random variable can take on only certain specific values (e.g., the sum of two dice thrown)
- A probability is the likelihood of a particular event happening
- A probability distribution function shows the outcomes that are possible from a random process and how likely each one is to occur
- A continuous random variable can take any value (possibly only within a fixed range), and the probabilities associated with each range of outcomes is shown in a probability density function (pdf)

Probability and probability distributions – some definitions (Cont'd)

- The probability that a continuous variable takes on a specific value is always zero, since the variable could be defined to any arbitrary degree of accuracy (0.1 vs 0.1000001 etc.) and thus we can only calculate the probability that the variable lies within a particular range.
- There are many continuous distributions, including the uniform and the normal.

The normal distribution

- The normal (Gaussian) distribution is the most commonly used in statistics
- It has many desirable properties and is easy to work with
- It is unimodal (has only one peak) and symmetric
- The moments of a distribution describe its properties.
- The first two moments of a distribution are its mean and variance respectively
- Only knowledge of the mean and variance are required to completely describe the distribution
- A normal distribution has a skewness of zero and a kurtosis of 3 (excess kurtosis of zero)
- Skewness and kurtosis are the (standardised) third and fourth moments of the distribution respectively.

The normal distribution 2

- The formula for the pdf of a normal distribution is given by:

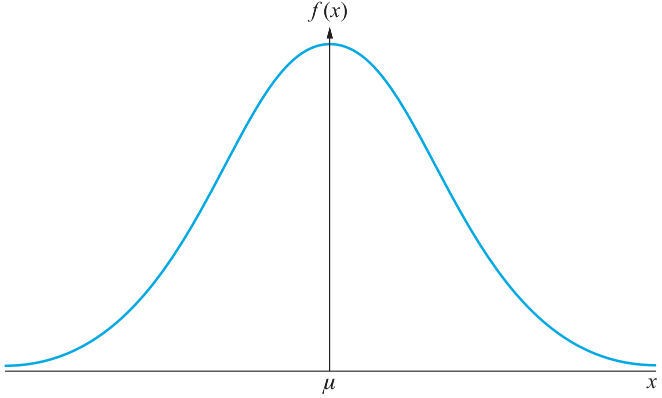
$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$$

- A standard normally distributed variable can be constructed from any normal random variable by subtracting its mean (μ) and dividing by its standard deviation (σ):

$$Z = \frac{y-\mu}{\sigma} \sim N(0,1)$$

- The probability that a continuous random variable lies above a certain value (or below a certain value) is given by the cumulative density function (cdf)
- The cdf for a normal distribution has a sigmoid shape

A plot of the pdf for a normal distribution



The central limit theorem

- If we take N draws from a normally distributed random variable with population mean μ and variance σ^2 then the sample mean will also be normally distributed with mean μ and variance σ^2/N
- In fact, the sample mean of any random variable (whether normally distributed or not) will tend towards a normal distribution as the sample size tends to infinity.
- This is known as the *central limit theorem*.

Other important distributions

- Three other important continuous distributions are the chi-squared (χ^2), the F and the t (sometimes known as Student's t)
- These distributions all relate to the normal and to each other
- The sum of squares of n independent normal distributions will be a χ^2 distribution with n degrees of freedom
- The ratio of two independent χ^2 distributions divided by their respective degrees of freedom n_1 and n_2 will be an F-distribution with n_1 and n_2 degrees of freedom
- The t distribution tends to the normal as its degrees of freedom increase towards infinity
- Each of these distributions will be discussed later as it is used.

Descriptive Statistics

Measures of central tendency

- The average value of a series is its measure of location or measure of central tendency, capturing its typical behaviour
- There are three broad methods to calculate the average value of a series: the *mean*, *median* and *mode*
- The mean is the very familiar sum of all N observations divided by N
- More strictly, this is known as the arithmetic mean
- The mode is the most frequently occurring value in a set of observations
- The median is the middle value in a series when the observations are arranged in ascending order
- Each of the three methods of calculating an average has advantages and disadvantages

The geometric mean

- The geometric mean involves calculating the Nth root of the product of the N observations
- So the geometric mean of six numbers in a series would be obtained by multiplying them together and taking the sixth root
- In finance, since the numbers in the series can be less than one or negative, we use a slightly different method to calculate the geometric mean compared to the one reported in many non-finance statistics books

The geometric mean (Cont'd)

- Here we add one to each data point, then multiply together, take the Nth root and then subtract one at the end:

$$\bar{R}_G = [(1+r_1)(1+r_2)\ldots(1+r_N)]^{1/N} - 1$$

  where r_1, r_2, etc. are the data points that we wish to take the geometric mean of

- The geometric mean will always be smaller than the arithmetic mean unless all of the data points are the same.

Measures of spread

- The spread of a series about its mean value can be measured using the *variance* or *standard deviation* (which is the square root of the variance)
- This quantity is an important measure of risk in finance
- The standard deviation scales with the data, whereas the variance scales with the square of the data. S
- So, for example, if the units of the data points are US dollars, the standard deviation will also be measured in dollars whereas the variance will be in dollars squared
- Other measures of spread include the *range* (the difference between the largest and smallest of the data points) and the *semi-interquartile range* (the difference between the first and third quartile points in the series

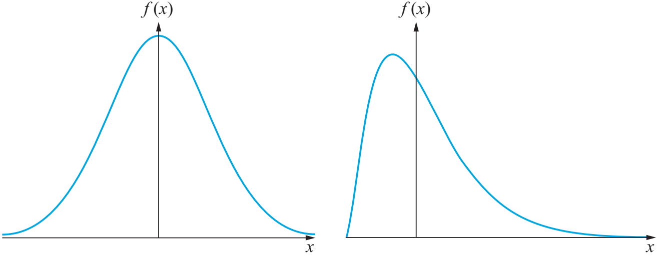
Measures of spread (Cont'd)

- The *coefficient of variation* divides the standard deviation by the sample mean to obtain a unit-free measure of spread that can be compared across series with different scales.

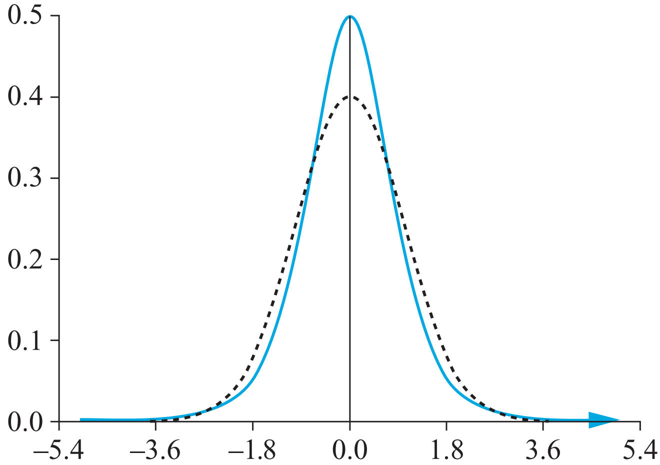
Higher moments

- The higher moments of a data sample give further indications of its features and shape.
- Skewness is the standardised third moment of a distribution and indicates the extent to which it is asymmetric
- Kurtosis is the standardised fourth moment and measures whether a series is fat or thin tailed
- Skewness can be positive or negative while kurtosis can only be positive
- The formulae for skewness and kurtosis calculate the quantities from the sample data in the same way that the variance is calculated.

Plot of a skewed series versus a normal distribution



Plot of a leptokurtic series versus a normal distribution



Some useful algebra for working with means, variances and covariances

Means

- The mean of a random variable y is also known as its expected value, written $\mathrm{E}(y)$
- The expected value of a constant is the constant, e.g. $\mathrm{E}(c) = c$
- The expected value of a constant multiplied by a random variable is equal to the constant multiplied by the expected value of the variable: $\mathrm{E}(cy) = c\,\mathrm{E}(y)$
- It can also be stated that $\mathrm{E}(cy + d) = c\mathrm{E}(y) + d$, where d is also a constant
- For two independent random variables, y_1 and y_2, $\mathrm{E}(y_1 y_2) = \mathrm{E}(y_1)\,\mathrm{E}(y_2)$.

Some useful algebra for working with means, variances and covariances 2

<u>Variances</u>

- The variance of a random variable y is usually written var(y)
- The variance of a random variable y is given by $\text{var}(y) = \text{E}(y - \text{E}(y))^2$
- The variance of a constant is zero: $\text{var}(c) = 0$
- For c and d constants, $\text{var}(cy + d) = c^2\text{var}(y)$
- For two independent random variables, y_1 and y_2, $\text{var}(cy_1 + dy_2) = c^2\text{var}(y_1) + d^2\text{var}(y_2)$

Some useful algebra for working with means, variances and covariances 3

<u>Covariances</u>

- The covariance between two random variables, y_1 and y_2 may be expressed as $\text{cov}(y_1, y_2)$
- $\text{cov}(y_1, y_2) = \text{E}[(y_1 - \text{E}(y_1))(y_2 - \text{E}(y_2))]$
- For two independent random variables, y_1 and y_2, $\text{cov}(y_1, y_2) = 0$
- For four constants, c, d, e, and f, $\text{cov}(c + dy_1, e + fy_2) = df\,\text{cov}(y_1, y_2)$.

Data types

Types of Data and Notation

- There are 3 types of data which econometricians might use for analysis:
  1. Time series data
  2. Cross-sectional data
  3. Panel data, a combination of 1. & 2.

- The data may be quantitative (e.g. exchange rates, stock prices, number of shares outstanding), or qualitative (e.g. day of the week).

- Examples of *time series data*

Series	*Frequency*
GNP or unemployment	monthly, or quarterly
government budget deficit	annually
money supply	weekly
value of a stock market index	as transactions occur

Time Series versus Cross-sectional Data

- *Examples of Problems that Could be Tackled Using a Time Series Regression*
  - How the value of a country's stock index has varied with that country's macroeconomic fundamentals.
  - How the value of a company's stock price has varied when it announced the value of its dividend payment.
  - The effect on a country's currency of an increase in its interest rate

- Cross-sectional data are data on one or more variables collected at a single point in time, e.g.
  - A poll of usage of internet stock broking services
  - Cross-section of stock returns on the New York Stock Exchange
  - A sample of bond credit ratings for UK banks

Cross-sectional and Panel Data

- *Examples of Problems that Could be Tackled Using a Cross-Sectional Regression*
  - The relationship between company size and the return to investing in its shares
  - The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

- Panel Data has the dimensions of both time series and cross-sections, e.g. the daily prices of a number of blue chip stocks over two years.

- It is common to denote each observation by the letter t and the total number of observations by T for time series data, and to denote each observation by the letter i and the total number of observations by N for cross-sectional data.

Continuous and Discrete Data

- Continuous data can take on any value and are not confined to take specific numbers.
- Their values are limited only by precision.
  - For example, the rental yield on a property could be 6.2%, 6.24%, or 6.238%.
- On the other hand, discrete data can only take on certain values, which are usually integers
  - For instance, the number of people in a particular underground carriage or the number of shares traded during a day.
- They do not necessarily have to be integers (whole numbers) though, and are often defined to be count numbers.
  - For example, until recently when they became 'decimalised', many financial asset prices were quoted to the nearest 1/16 or 1/32 of a dollar.

Cardinal, Ordinal and Nominal Numbers

- Another way in which we could classify numbers is according to whether they are cardinal, ordinal, or nominal.
- *Cardinal numbers* are those where the actual numerical values that a particular variable takes have meaning, and where there is an equal distance between the numerical values.
  - Examples of cardinal numbers would be the price of a share or of a building, and the number of houses in a street.
- *Ordinal numbers* can only be interpreted as providing a position or an ordering.
  - Thus, for cardinal numbers, a figure of 12 implies a measure that is 'twice as good' as a figure of 6. On the other hand, for an ordinal scale, a figure of 12 may be viewed as 'better' than a figure of 6, but could not be considered twice as good. Examples of ordinal numbers would be the position of a runner in a race.

Cardinal, Ordinal and Nominal Numbers (Cont'd)

- Nominal numbers occur where there is no natural ordering of the values at all.
  - Such data often arise when numerical values are arbitrarily assigned, such as telephone numbers or when codings are assigned to qualitative data (e.g. when describing the exchange that a US stock is traded on).

- Cardinal, ordinal and nominal variables may require different modelling approaches or at least different treatments, as should become evident in the subsequent chapters.

Arithmetic and geometric series, present and future values

Arithmetic series

- An arithmetic progression is a sequence where a specific entry in that series is formed by adding a fixed number, known as the common difference, to the previous one.
- For example: 2, 5, 8, 11, . . . or –10, –30, –50, . . .
- The first of these is an arithmetic series with an initial value of 2 and adds 3 each time we move from one entry in the sequence to the next
- The second set is an arithmetic series with an initial value of –10 and a common difference of –20

Geometric series

- A geometric progression is a series where instead of adding a fixed amount to move from one entry to the next, we multiply by a fixed amount (the common ratio).
- For example: 4, 8, 16, 32, . . . or 2, 1, 0.5, 0.25, . . .
- The first of these is a geometric series with an initial value of 4 and a common ratio of 2
- The second set is a geometric series with an initial value of 2 and a common ratio of 0.5
- Geometric series are very useful in finance as they describe the situation where a sum of money is invested and earns a certain percentage of interest in each time period.

Sums of geometric series

- To develop some notation, let a denote the initial value of a geometric series (starting with the term numbered 0 and ending with term numbered $n-1$), and let d denote the common ratio
- We could write a geometric series containing n terms as $a, ad, ad^2, ad^3, \ldots, ad^{n-1}$
- There is an expression that can be used to calculate the sum of the first n terms, denoted S_n, of the series (running from a to ad^{n-1}):

$$S_n = \frac{a(1-d^n)}{1-d}$$

- For instance, if a geometric series begins with 2 and has a common ratio of 3, the sum of the first 8 terms would be: $Sn = 2 \times (1-3^8)/(1-3) = 6560$.

Sums of infinite geometric series

- Of particular use in financial applications is the infinite sum of a geometric progression, denoted S_∞
- We can see what will happen to the sum of n terms as n tends to infinity, since the term d^n will tend towards zero (so long as $0 < d < 1$), in which case the expression can be simplified as:

$$S_\infty = \frac{a}{1-d}$$

- Here, even though there is an infinite number of terms in the series, their sum is finite
- Note that if $d \geq 1$, the series would not converge (i.e., successive terms would not become smaller and smaller) and therefore the sum would be infinite.

Present values and future values

- Money has time value
- This means that receipt of a given amount of money is worth a different amount depending on when it is received
- In general, money has positive time value
- As a result of the time value of money, we cannot simply combine cashflows in their raw form into financial calculations if they are received at different points in time
- The way that we ensure we are comparing like-with-like is to transform the cashflows to what they would be worth if they were all received at the same point in time

Present values and future values (Cont'd)

- So we either transform all cashflows to the amount that they would be worth at some given point in the future (the *future value*) or we transform all future cashflows into the equivalent amount that they would be worth if received today (the *present value*).

Future values

- Suppose that we place £100 in a bank savings account for five years, paying an annual interest rate of 2%
- The sum of money in the account at the end of the period would be given by
  $P_T = P_0 \times (1 + r)^T$
- where P_T denotes the terminal (future) value of the account, r is the interest rate, P_0 is the amount placed in the account now, and T is the number of time periods for which the money is invested
- In this example, the future value of the investment at the end of the first year would be $P_T = £100 \times (1 + 0.02) =$ £102,while at the end of the second year it would have grown to $P_T = £100 \times (1 + 0.02)^2 = £102 \times (1 + 0.02) = £104.04$

Future values (Cont'd)

- The savings balance would continue to grow in this way until, at the end of the fifth year, it would have reached $P_T = £100 \times (1 + 0.02)^5 = £110.41$
- In this case the interest is compounded annually – interest is paid this year on the total value of this year's end savings, which will comprise both last year's savings value and last year's interest.

Future values – calculating the interest rate

- We can rearrange the future value formula on the previous slide to make r the subject
- This would enable us to calculate the rate of interest required to secure a specific future value, P_T, given the initial investment, P_0:

$$r = (\frac{P_T}{P_0})^{1/T} - 1$$

- For example, if we make an initial investment of £1000 and no further investments, and we leave the funds for ten years, what rate of interest is required to enable us to achieve a sum of £1500 by the end of the decade?
- The calculation is $r = [1500/1000]^{1/10} - 1 = 0.0414$
- So an annual interest rate of about 4.14% is required.

Future values - calculating the time to achieve a certain sum

- A further rearrangement of the equations on the previous slides enables us to make the term of the investment, T, the subject of the formula:

$$T = \frac{\ln(P_T/P_0)}{\ln(1+r)}$$

- So, for instance, if we can invest £1000 initially and wish it to grow to £2000, assuming an interest rate of 10%, how many years do we need to wait?
- We would have $T = \ln(2000/1000)/\ln(1+0.1) = 7.273$

Future values - calculating the time to achieve a certain sum (Cont'd)

- Notice that we can use the formula above to determine how many years it would take for an investment to grow by a factor of Z

$$T = \frac{\ln(Z)}{\ln(1+r)}$$

where $P_T = ZP_0$

- So, for example, if we wanted to triple the initial investment with $r = 10\%$, we would set $Z = 3$ and we would have $T = \ln(3)/\ln(1+0.1) = 11.527$ years.

Compounding frequency

- The above examples assumed that interest was received annually but in many cases it will be received more frequently
- The future value of an investment (P_T) when interest of r per year in total is paid n times per year for T years on the original amount P_0 is given by

$$P_T = P_0[1 + \frac{r}{n}]^{nT}$$

- In the limit, as the compounding frequency increases and so we have more and more shorter and shorter time periods (i.e., we move from annual to monthly to weekly to daily to hourly compounding and so on), we would eventually reach a situation where the time period was infinitesimally small

Compounding frequency (Cont'd)

- We would term this continuous compounding. If interest is compounded continuously at an annual equivalent rate r, we would write
  $P_T = P_0 e^{rT}$
- If $T = 5$ and $r = 2\%$, then the terminal value if interest is continuously compounded is $P_T = e^{0.02 \times 5} = £110.52$.

Present value

- The reverse of calculating the future value would be where we calculate the present value of an amount of money to be received at some point in the future
- Instead of an interest rate as we would have for a future value calculation, in the case of present values we use a discount rate, which is the rate at which we would reduce the future payment into today's terms
- We would write

$$P_0 = \frac{P_T}{(1+r)^T}$$

where P_0 is the present value, r is the discount rate, P_T is the sum to be paid or received in the future, and T is the number of periods into the future that it will be paid or received

Present value (Cont'd)

- Example: What is the present value of £100 to be received in five years' time if the discount rate is 2%?

$$P_0 = \frac{\$100}{(1 + 0.02)^5} = \$90.57$$

- This shows that £100 in five years' time is worth £90.57 in today's money terms.

Present value calculations for pricing bonds

- A bond pays a £10 coupon annually with the next coupon due immediately.
- The bond has exactly five years left to maturity when it will be redeemed at its par value of £100 and an appropriate discount rate is 10%.
- What would be a fair price to pay today for the bond?
- We would calculate the fair price as the discounted sum of the six coupon payments (one now and one at the end of each of the next five years) and plus the discounted value of the par amount of the bond.

Present value calculations for pricing bonds (Cont'd)

- If we let the fair price in pounds be denoted by P_0, the calculation would be

$$P_0 = 5 + \frac{5}{(1+0.1)} + \frac{5}{(1+0.1)^2} + \frac{5}{(1+0.1)^3} + \frac{5}{(1+0.1)^4} + \frac{5}{(1+0.1)^5} + \frac{100}{(1+0.1)^5}$$

- The fair price to pay for the bond is $P_0 =$£86.04.

Present value calculations for pricing irredeemable bonds

- If the bond under study was irredeemable (so that it had an infinite lifetime like a stock), we would need to use the S_∞ formula to calculate its present value.
- As an illustration, what should an investor pay today for a perpetual (irredeemable) bond that pays a coupon of £5 every six months if the appropriate rate to discount future cashflows is 4% per annum and the next coupon will be paid immediately?
- We could discount each cashflow with the discount rate for six months, which would be 2%
- We could write this as an infinite series beginning with £5:

$$5, \frac{5}{(1+0.02)}, \frac{5}{(1+0.02)^2}, \frac{5}{(1+0.02)^3}, \frac{5}{(1+0.02)^4}, \ldots$$

Present value calculations for pricing irredeemable bonds (Cont'd)

- We need to be slightly careful as the common difference here is $1/(1+0.02)$
- Then we would have the value as $S_\infty = £5/(1 - [1/(1+0.02)]) = £255$.

Internal rate of return

- It is sometimes the case that we know both the present value of a particular set of cashflows, and all of the future cashflows, but we do not know the discount rate
- The value of r that would equate the amount to be paid today P_0 to the present value of all of the cashflows we would receive if we purchased the asset is known as the internal rate of return or IRR
- We could calculate that by solving the annuity formula above for r
- More generally, if the future cashflow payments were not fixed but varied over time, we would have a more flexible formula:

$$P_0 = a_0 + \frac{a_1}{(1+r)} + \frac{a_2}{(1+r)^2} + \frac{a_3}{(1+r)^3} + \frac{a_4}{(1+r)^4} + \ldots + \frac{a_T}{(1+r)}$$

Internal rate of return (Cont'd)

- If there are only one or two time-periods, we could solve the equation by hand; for more we would need to use a spreadsheet.

Returns in Financial Modelling

- It is preferable not to work directly with asset prices, so we usually convert the raw prices into a series of returns. There are two ways to do this:

Simple returns	or	log returns
$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} * 100\%$		$R_t = \ln \frac{p_t}{p_{t-1}} * 100\%$

where, R_t denotes the return at time t, P_t denotes the asset price at time t and ln denotes the natural logarithm

- We also ignore any dividend payments, or alternatively assume that the price series have been already adjusted to account for them.

Log Returns

- The returns are also known as log price relatives, which will be used throughout this book. There are a number of reasons for this:
  1. They have the nice property that they can be interpreted as continuously compounded returns.
  2. Can add them up, e.g. if we want a weekly return and we have calculated daily log returns:

$$
\begin{aligned}
r_1 &= \ln p_1/p_0 &&= \ln p_1 - \ln p_0 \\
r_2 &= \ln p_2/p_1 &&= \ln p_2 - \ln p_1 \\
r_3 &= \ln p_3/p_2 &&= \ln p_3 - \ln p_2 \\
r_4 &= \ln p_4/p_3 &&= \ln p_4 - \ln p_3 \\
r_5 &= \ln p_5/p_4 &&= \ln p_5 - \ln p_4 \\
\hline
& &&\ln p_5 - \ln p_0 = \ln p_5/p_0
\end{aligned}
$$

A Disadvantage of using Log Returns

- There is a disadvantage of using the log-returns. The simple return on a portfolio of assets is a weighted average of the simple returns on the individual assets:

$$R_{pt} = \sum_{i=1}^{N} w_i R_{it}$$

- But this does not work for the continuously compounded returns.

Real Versus Nominal Series

- The general level of prices has a tendency to rise most of the time because of inflation
- We may wish to transform nominal series into real ones to adjust them for inflation
- This is called *deflating* a series or displaying a series *at constant prices*
- We do this by taking the nominal series and dividing it by a price deflator:

  $real\ series_t = nominal\ series_t * 100 / deflator_t$
  (assuming that the base figure is 100)
- We only deflate series that are in nominal price terms, not quantity terms.

Deflating a Series

- If we wanted to convert a series into a particular year's figures (e.g. house prices in 2010 figures), we would use:

  $real\ series_t = nominal\ series_t * deflator_{reference\ year} / deflator_t$

- This is the same equation as the previous slide except with the deflator for the reference year replacing the assumed deflator base figure of 100

- Often the consumer price index, CPI, is used as the deflator series.

Portfolio Theory using Matrix Algebra

Portfolio Theory and Matrix Algebra - Basics

- Probably the most important application of matrix algebra in finance is to solving portfolio allocation problems
- Suppose that we have a set of N stocks that are included in a portfolio P with weights $w_1, w_2, \ldots, w_N$ and suppose that their expected returns are written as $E(r_1), E(r_2), \ldots, E(r_N)$. We could write the $N \times 1$ vectors of weights, w, and of expected returns, $E(r)$, as

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \ldots \\ w_N \end{pmatrix} \quad E(r) = \begin{pmatrix} E(r_1) \\ E(r_2) \\ \ldots \\ E(r_N) \end{pmatrix}$$

- The expected return on the portfolio, $E(r_P)$ can be calculated as $E(r)'w$.

The Variance-Covariance Matrix

- The variance-covariance matrix of the returns, denoted V includes all of the variances of the components of the portfolio returns on the leading diagonal and the covariances between them as the off-diagonal elements.
- The variance-covariance matrix of the returns may be written

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2N} \\ \vdots & & & \vdots & \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \cdots & \sigma_{NN} \end{pmatrix}$$

- For example:
  - σ_{11} is the variance of the returns on stock one, σ_{22} is the variance of returns on stock two, etc.
  - σ_{12} is the covariance between the returns on stock one and those on stock two, etc.

Constructing the Variance-Covariance Matrix

- In order to construct a variance-covariance matrix we would need to first set up a matrix containing observations on the actual returns , R (not the expected returns) for each stock where the mean, $\bar{r}_i$ $(i = 1, \ldots, N)$, has been subtracted away from each series i.

- We would write

$$R = \begin{pmatrix} r_{11} - \bar{r}_1 & r_{21} - \bar{r}_2 & r_{31} - \bar{r}_3 & \ldots & r_{N1} - \bar{r}_N \\ r_{12} - \bar{r}_1 & r_{22} - \bar{r}_2 & r_{32} - \bar{r}_3 & \ldots & r_{N2} - \bar{r}_N \\ \vdots & & & \vdots & \\ r_{1T} - \bar{r}_1 & r_{2T} - \bar{r}_2 & r_{3T} - \bar{r}_3 & \ldots & r_{NT} - \bar{r}_N \end{pmatrix}$$

- r_{ij}, is the jth time-series observation on the ith stock. The variance-covariance matrix would then simply be calculated as $V = (R'R)/(T - 1)$.

The Variance of Portfolio Returns

- Suppose that we wanted to calculate the variance of returns on the portfolio P
  - A scalar which we might call V_P
- We would do this by calculating

$$V_P = w'Vw$$

- Checking the dimension of V_P, w' is $(1 \times N)$, V is $(N \times N)$ and w is $(N \times 1)$ so V_P is $(1 \times N \times N \times N \times N \times 1)$, which is (1×1) as required.

The Correlation between Returns Series

- We could define a correlation matrix of returns, C, which would be

$$C = \begin{pmatrix} 1 & C_{12} & C_{13} & \dots & C_{1N} \\ C_{21} & 1 & C_{23} & \dots & C_{2N} \\ \vdots & & & \vdots & \\ C_{N1} & C_{N2} & C_{N3} & \dots & 1 \end{pmatrix}$$

- This matrix would have ones on the leading diagonal and the off-diagonal elements would give the correlations between each pair of returns

- Note that the correlation matrix will always be symmetrical about the leading diagonal

The Correlation between Returns Series (Cont'd)

- Using the correlation matrix, the portfolio variance is

$$V_P = w'SCSw$$

where S is a diagonal matrix containing the standard deviations of the portfolio returns.

Selecting Weights for the Minimum Variance Portfolio

- Although in theory the optimal portfolio on the efficient frontier is better, a variance-minimising portfolio often performs well out-of-sample

- The portfolio weights w that minimise the portfolio variance, V_P is written

$$\min_{w} \; w'Vw$$

- We also need to be slightly careful to impose at least the restriction that all of the wealth has to be invested (weights sum to one)

- This restriction is written as $w' \cdot 1_N = 1$, where 1_N is a column vector of ones of length N.

Selecting Weights for the Minimum Variance Portfolio (Cont'd)

- The minimisation problem can be solved to

$$w_{MVP} = \frac{1_N \cdot V^{-1}}{1_N \cdot V^{-1} \cdot 1_N'}$$

where MVP stands for minimum variance portfolio

Selecting Optimal Portfolio Weights

- In order to trace out the mean-variance efficient frontier, we would repeatedly solve this minimisation problem but in each case set the portfolio's expected return equal to a different target value, $\bar{R}$

- We would write this as

$$\min_{w} \quad w'Vw \quad \text{subject to} \quad w' \cdot 1_N = 1, w'E(r) = \bar{R}$$

- This is sometimes called the Markowitz portfolio allocation problem
  - It can be solved analytically so we can derive an exact solution
- But it is often the case that we want to place additional constraints on the optimisation, e.g.
  - Restrict the weights so that none are greater than 10% of overall wealth

Selecting Optimal Portfolio Weights (Cont'd)

- Restrict them to all be positive (i.e. long positions only with no short selling)

- In such cases the Markowitz portfolio allocation problem cannot be solved analytically and thus a numerical procedure must be used

Selecting Optimal Portfolio Weights

- If the procedure above is followed repeatedly for different return targets, it will trace out the efficient frontier
- In order to find the tangency point where the efficient frontier touches the capital market line, we need to solve the following problem

$$\max_{w} \quad \frac{w'E(r) - r_f}{(w'Vw)^{\frac{1}{2}}} \quad \text{subject to} \quad w' \cdot 1_N = 1$$

- If no additional constraints are required on weights, this can be solved as

$$w = \frac{V^{-1}[E(r) - r_f \cdot 1_N]}{1_N' V^{-1}[E(r) - r_f \cdot 1_N]}$$

- Note that it is also possible to write the Markowitz problem where we select the portfolio weights that maximise the expected portfolio return subject to a target maximum variance level.

Bayesian versus Classical Statistics

- The philosophical approach to model-building used here throughout is based on 'classical statistics'
- This involves postulating a theory and then setting up a model and collecting data to test that theory
- Based on the results from the model, the theory is supported or refuted
- There is, however, an entirely different approach known as Bayesian statistics
- Here, the theory and model are developed together
- The researcher starts with an assessment of existing knowledge or beliefs formulated as probabilities, known as priors
- The priors are combined with the data into a model

Bayesian versus Classical Statistics (Cont'd)

- The beliefs are then updated after estimating the model to form a set of posterior probabilities
- Bayesian statistics is a well established and popular approach, although less so than the classical one
- Some classical researchers are uncomfortable with the Bayesian use of prior probabilities based on judgement
- If the priors are very strong, a great deal of evidence from the data would be required to overturn them
- So the researcher would end up with the conclusions that he/she wanted in the first place!
- In the classical case by contrast, judgement is not supposed to enter the process and thus it is argued to be more objective.