Chapter 2

Mathematical and Statistical Foundations

Functions

- A function is a mapping or relationship between an input or set of inputs and an output
- We write that y, the output, is a function $f(x)$, the input, or $y = f(x)$
- y could be a linear function of x where the relationship can be expressed on a straight line
- Or it could be non-linear where it would be expressed graphically as a curve
- If the equation is linear, we would write the relationship as

$$y = a + bx$$

  where y and x are called variables and a and b are parameters
- a is the intercept and b is the slope or gradient

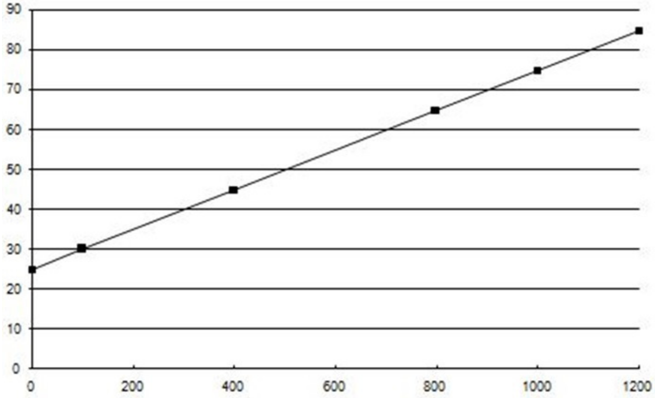
Straight Lines

- The intercept is the point at which the line crosses the y-axis
- Example: suppose that we were modelling the relationship between a student's average mark, y (in percent), and the number of hours studied per year, x
- Suppose that the relationship can be written as a linear function

$$y = 25 + 0.05x$$

- The intercept, a, is 25 and the slope, b, is 0.05
- This means that with no study (x=0), the student could expect to earn a mark of 25%
- For every hour of study, the grade would on average improve by 0.05%, so another 100 hours of study would lead to a 5% increase in the mark

Plot of Hours Studied Against Mark Obtained



Straight Lines

- In the graph above, the slope is positive
  - i.e. the line slopes upwards from left to right
- But in other examples the gradient could be zero or negative
- For a straight line the slope is constant – i.e. the same along the whole line
- In general, we can calculate the slope of a straight line by taking any two points on the line and dividing the change in y by the change in x
- Δ (Delta) denotes the change in a variable
- For example, take two points $x = 100$, $y = 30$ and $x = 1000$, $y = 75$

Straight Lines (Cont'd)

- We can write these using coordinate notation (x,y) as (100,30) and (1000,75)

- We would calculate the slope as

$$\frac{\Delta y}{\Delta x} = \frac{75 - 30}{1000 - 100} = 0.05$$

Roots

- The point at which a line crosses the x-axis is known as the root
- A straight line will have one root (except for a horizontal line such as y=4 which has no roots)
- To find the root of an equation set y to zero and rearrange

$$0 = 25 + 0.05x$$

- So the root is $x = -500$
- In this case it does not have a sensible interpretation: the number of hours of study required to obtain a mark of zero!

Quadratic Functions

- A linear function is often not sufficiently flexible to accurately describe the relationship between two series
- We could use a quadratic function instead. We would write it as

$$y = a + bx + cx^2$$

  where a, b, c are the parameters that describe the shape of the function
- Quadratics have an additional parameter compared with linear functions
- The linear function is a special case of a quadratic where c=0
- a still represents where the function crosses the y-axis
- As x becomes very large, the x^2 term will come to dominate
- Thus if c is positive, the function will be $\cup$-shaped, while if c is negative it will be $\cap$-shaped.

The Roots of Quadratic Functions

- A quadratic equation has two roots
- The roots may be distinct (i.e., different from one another), or they may be the same (repeated roots); they may be real numbers (e.g., 1.7, -2.357, 4, etc.) or what are known as complex numbers
- The roots can be obtained either by factorising the equation (contracting it into parentheses), by "completing the square", or by using the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

- If $b^2 > 4ac$, the function will have two unique roots and it will cross the x-axis in two separate places

The Roots of Quadratic Functions (Cont'd)

- If $b^2 = 4ac$, the function will have two equal roots and it will only cross the x-axis in one place
- If $b^2 < 4ac$, the function will have no real roots (only complex roots), it will not cross the x-axis at all and thus the function will always be above the x-axis.

Calculating the Roots of Quadratics - Examples

Determine the roots of the following quadratic equations:

1. $y = x^2 + x - 6$
2. $y = 9x^2 + 6x + 1$
3. $y = x^2 - 3x + 1$
4. $y = x^2 - 4x$

Calculating the Roots of Quadratics - Solutions

- We solve these equations by setting them in turn to zero
- We could use the quadratic formula in each case, although it is usually quicker to determine first whether they factorise

1. $x^2 + x - 6 = 0$ factorises to $(x - 2)(x + 3) = 0$ and thus the roots are 2 and -3, which are the values of x that set the function to zero. In other words, the function will cross the x-axis at $x = 2$ and $x = -3$.
2. $9x^2 + 6x + 1 = 0$ factorises to $(3x + 1)(3x + 1) = 0$ and thus the roots are $-1/3$ and $-1/3$. This is known as repeated roots – since this is a quadratic equation there will always be two roots but in this case they are both the same.

Calculating the Roots of Quadratics - Solutions (Cont'd)

3. $x^2 - 3x + 1 = 0$ does not factorise and so the formula must be used with $a = 1, b = -3, c = 1$ and the roots are 0.38 and 2.62 to two decimal places.
4. $x^2 - 4x = 0$ factorises to $x(x - 4) = 0$ and so the roots are 0 and 4.

- All of these equations have two real roots
- But if we had an equation such as $y = 3x^2 - 2x + 4$, this would not factorise and would have complex roots since $b^2 - 4ac < 0$ in the quadratic formula.

Powers of Number or of Variables

- A number or variable raised to a power is simply a way of writing repeated multiplication
- So for example, raising x to the power 2 means squaring it (i.e., $x^2 = x \times x$).
- Raising it to the power 3 means cubing it ($x^3 = x \times x \times x$), and so on
- The number that we are raising the number or variable to is called the index, so for x^3, the index would be 3

Manipulating Powers and their Indices

- Any number or variable raised to the power one is simply that number or variable, e.g., $3^1 = 3$, $x^1 = x$, and so on
- Any number or variable raised to the power zero is one, e.g., $5^0 = 1$, $x^0 = 1$, etc., except that 0^0 is not defined (i.e., it does not exist)
- If the index is a negative number, this means that we divide one by that number – for example, $x^{-3} = \frac{1}{x^3} = \frac{1}{x \times x \times x}$.
- If we want to multiply together a given number raised to more than one power, we would add the corresponding indices together – for example, $x^2 \times x^3 = x^2 x^3 = x^{2+3} = x^5$.

Manipulating Powers and their Indices (Cont'd)

- If we want to calculate the power of a variable raised to a power (i.e., the power of a power), we would multiply the indices together – for example, $x^{2^3} = x^{2\times 3} = x^6$.

- If we want to divide a variable raised to a power by the same variable raised to another power, we subtract the second index from the first – for example, $\frac{x^3}{x^2} = x^{3-2} = x$.

- If we want to divide a variable raised to a power by a different variable raised to the same power, the following result applies:

$$\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$$

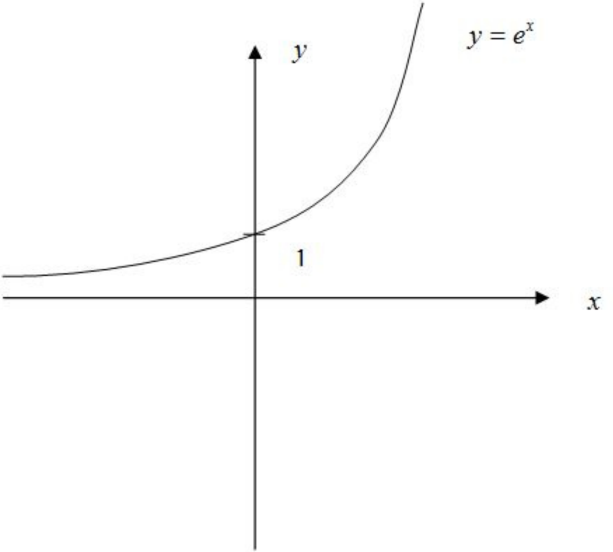
Manipulating Powers and their Indices (Cont'd)

- The power of a product is equal to each component raised to that power – for example, $(x \times y)^3 = x^3 \times y^3$.
- The indices for powers do not have to be integers, so $x^{1/2}$ is the notation we would use for taking the square root of x, sometimes written $\sqrt{x}$
- Other, non-integer powers are also possible, but are harder to calculate by hand (e.g. $x^{0.76}$, $x^{-0.27}$, etc.)
- In general, $x^{1/n} = \sqrt[n]{x}$.

The Exponential Function, e

- It is sometimes the case that the relationship between two variables is best described by an exponential function
- For example, when a variable grows (or reduces) at a rate in proportion to its current value, we would write $y = e^x$
- e is a simply number: 2.71828. . .
- It is also useful for capturing the increase in value of an amount of money that is subject to compound interest
- The exponential function can never be negative, so when x is negative, y is close to zero but positive
- It crosses the y-axis at one and the slope increases at an increasing rate from left to right.

A Plot of the Exponential Function



Logarithms

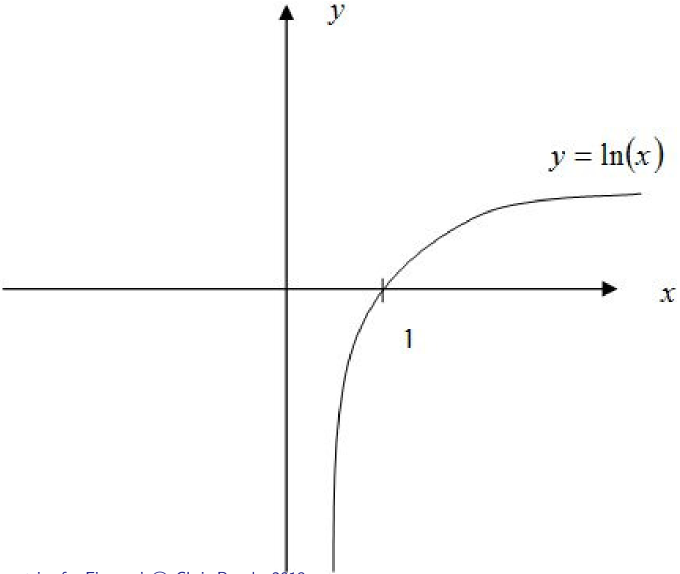
- Logarithms were invented to simplify cumbersome calculations, since exponents can then be added or subtracted, which is easier than multiplying or dividing the original numbers

- There are at least three reasons why log transforms may be useful.
  1. Taking a logarithm can often help to rescale the data so that their variance is more constant, which overcomes a common statistical problem known as heteroscedasticity.
  2. Logarithmic transforms can help to make a positively skewed distribution closer to a normal distribution.
  3. Taking logarithms can also be a way to make a non-linear, multiplicative relationship between variables into a linear, additive one.

How do Logs Work?

- Consider the power relationship $2^3 = 8$
- Using logarithms, we would write this as $\log_2 8 = 3$, or 'the log to the base 2 of 8 is 3'
- Hence we could say that a logarithm is defined as the power to which the base must be raised to obtain the given number
- More generally, if $a^b = c$, then we can also write $\log_a c = b$
- If we plot a log function, $y = \log x$, it would cross the x-axis at one – see the following slide
- It can be seen that as x increases, y increases at a slower rate, which is the opposite to an exponential function where $y$increases at a faster rate as x increases.

A Graph of a Log Function



How do Logs Work?

- Natural logarithms, also known as logs to base e, are more commonly used and more useful mathematically than logs to any other base
- A log to base e is known as a natural or Naperian logarithm, denoted interchangeably by $\ln(y)$ or $\log(y)$
- Taking a natural logarithm is the inverse of a taking an exponential, so sometimes the exponential function is called the antilog
- The log of a number less than one will be negative, e.g. $ln(0.5) \approx -0.69$
- We cannot take the log of a negative number
  - So ln(-0.6), for example, does not exist.

The Laws of Logs

For variables x and y:

- $\ln(x\,y) = \ln(x) + \ln(y)$
- $\ln(x/y) = \ln(x) - \ln(y)$
- $\ln(y^c) = c\,\ln(y)$
- $\ln(1) = 0$
- $\ln(1/y) = \ln(1) - \ln(y) = -\ln(y)$.
- $\ln(e^x) = e^{\ln(x)} = x$

Sigma Notation

- If we wish to add together several numbers (or observations from variables), the sigma or summation operator can be very useful

- Σ means 'add up all of the following elements.' For example, $\Sigma(1+2+3)=6$

- In the context of adding the observations on a variable, it is helpful to add 'limits' to the summation

- For instance, we might write $\sum_{i=1}^{4} x_i$

  where the i subscript is an index, 1 is the lower limit and 4 is the upper limit of the sum

- This would mean adding all of the values of x from x_1 to x_4.

Properties of the Sigma Operator

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} z_i = \sum_{i=1}^{n} (x_i + z_i)$$

$$\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i.$$

$$\sum_{i=1}^{n} x_i z_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} z_i$$

$$\sum_{i=1}^{n} x = x + x + \ldots + x = nx$$

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n = n\bar{x}.$$

Pi Notation

- Similar to the use of sigma to denote sums, the pi operator ($\prod$) is used to denote repeated multiplications.

- For example

$$\prod_{i=1}^{n} x_i = x_1 x_2 \ldots x_n$$

means 'multiply together all of the x i for each value of i between the lower and upper limits.'

- It also follows that

$$\prod_{i=1}^{n} (cx_i) = c^n \prod_{i=1}^{n} x_i$$

Differential Calculus

- The effect of the rate of change of one variable on the rate of change of another is measured by a mathematical derivative
- If the relationship between the two variables can be represented by a curve, the gradient of the curve will be this rate of change
- Consider a variable y that is a function f of another variable x, i.e. $y = f(x)$: the derivative of y with respect to x is written

$$\frac{dy}{dx} = \frac{df(x)}{dx}$$

  or sometimes $f'(x)$.
- This term measures the instantaneous rate of change of y with respect to x, or in other words, the impact of an infinitesimally small change in x
- Notice the difference between the notations Δy and dy

Differentiation: The Basics

1. The derivative of a constant is zero – e.g. if

$$\text{e.g. if } y = 10, \frac{dy}{dx} = 0.$$

   This is because $y = 10$ would be a horizontal straight line on a graph of y against x, and therefore the gradient of this function is zero

2. The derivative of a linear function is simply its slope

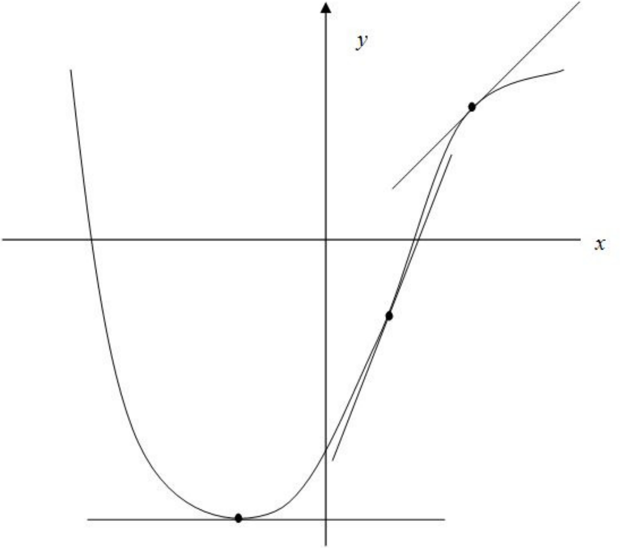
$$\text{e.g. if } y = 3x + 2, \frac{dy}{dx} = 3.$$

- But non-linear functions will have different gradients at each point along the curve

Differentiation: The Basics (Cont'd)

- In effect, the gradient at each point is equal to the gradient of the tangent at that point
- The gradient will be zero at the point where the curve changes direction from positive to negative or from negative to positive – this is known as a turning point.

The Tangent to a Curve



The Derivative of a Power Function or of a Sum

- The derivative of a power function n of x,

$$\text{i.e. } y = cx^n \text{ is given by } \frac{dy}{dx} = cnx^{n-1}.$$

- For example:

$$y = 4x^3, \frac{dy}{dx} = (4 \times 3)x^2 = 12x^2$$

$$y = \frac{3}{x} = 3x^{-1}, \frac{dy}{dx} = (3 \times -1)x^{-2} = -3x^{-2} = \frac{-3}{x^2}.$$

- The derivative of a sum is equal to the sum of the derivatives of the individual parts: e.g., if

$$\text{e.g. if } y = f(x) + g(x), \frac{dy}{dx} = f'(x) + g'(x)$$

- The derivative of a difference is equal to the difference of the derivatives of the individual parts: e.g.,

$$\text{if } y = f(x) - g(x), \frac{dy}{dx} = f'(x) - g'(x).$$

The Derivatives of Logs and Exponentials

- The derivative of the log of The derivative of the log of x is given by $1/x$

$$\text{i.e. } \frac{d(\log(x))}{dx} = \frac{1}{x}.$$

- The derivative of the log of a function of x is the derivative of the function divided by the function, i.e.

$$\text{i.e. } \frac{d(\log(f(x)))}{dx} = \frac{f'(x)}{f(x)}.$$

The Derivatives of Logs and Exponentials (Cont'd)

E.g., the derivative of $\log(x^3 + 2x - 1)$ is given by

$$\frac{3x^2 + 2}{x^3 + 2x - 1}.$$

- The derivative of e^x is e^x.
- The derivative of $e^{f(x)}$ is given by $f'(x)e^{f(x)}$.
  E.g., if $y = e^{3x^2}$, $\frac{dy}{dx} = 6xe^{3x^2}$.

Higher Order Derivatives

- It is possible to differentiate a function more than once to calculate the second order, third order, . . ., n^{th} order derivatives

- The notation for the second order derivative, which is usually just termed the second derivative, is

$$\frac{d^2y}{dx^2} = f''(x) = \frac{d(\frac{dy}{dx})}{dx}$$

- To calculate second order derivatives, differentiate the function with respect to x and then differentiate it again

Higher Order Derivatives (Cont'd)

- For example, suppose that we have the function $y = 4x^5 + 3x^3 + 2x + 6$, the first order derivative is

$$\frac{dy}{dx} = \frac{d(4x^5 + 3x^3 + 2x + 6)}{dx} = f'(x) = 20x^4 + 9x^2 + 2.$$

- The second order derivative is

$$\begin{aligned} \frac{d^2y}{dx^2} &= f''(x) = \frac{d(\frac{d(4x^5+3x^3+2x+6)}{dx})}{dx} \\ &= \frac{d(20x^4 + 9x^2 + 2)}{dx} \\ &= 80x^3 + 18x. \end{aligned}$$

Higher Order Derivatives (Cont'd)

- The second order derivative can be interpreted as the gradient of the gradient of a function – i.e., the rate of change of the gradient
- How can we tell whether a particular turning point is a maximum or a minimum?
- The answer is that we would look at the second derivative
- When a function reaches a maximum, its second derivative is negative, while it is positive for a minimum.

Maxima and Minima of Functions

- Consider the quadratic function $y = 5x^2 + 3x - 6$.

- Since the squared term in the equation has a positive sign (i.e., it is 5 rather than, say, -5), the function will have a $\cup$-shape rather than an $\cap$-shape, and thus it will have a minimum rather than a maximum:

$$\frac{dy}{dx} = 10x + 3, \frac{d^2y}{dx^2} = 10.$$

- Since the second derivative is positive, the function indeed has a minimum

- To find where this minimum is located, take the first derivative, set it to zero and solve it for x

Maxima and Minima of Functions (Cont'd)

- So we have $10x + 3 = 0$, and $x = -3/10 = -0.3$. If $x = -0.3$, y is found by substituting -0.3 into $y = 5x^2 + 3x - 6 = 5 \times (-0.3)^2 + (3 \times -0.3) - 6 = -6.45$. Therefore, the minimum of this function is found at $(-0.3, -6.45)$.

Partial Differentiation

- In the case where y is a function of more than one variable (e.g. $y = f(x_1, x_2, \ldots, x_n)$), it may be of interest to determine the effect that changes in each of the individual x variables would have on y

- Differentiation of y with respect to only one of the variables, holding the others constant, is partial differentiation

- The partial derivative of y with respect to a variable x_1 is usually denoted
$$\frac{\partial y}{\partial x_1}$$

- All of the rules for differentiation explained above still apply and there will be one (first order) partial derivative for each variable on the right hand side of the equation.

How to do Partial Differentiation

- We calculate these partial derivatives one at a time, treating all of the other variables as if they were constants.
- To give an illustration, suppose $y = 3x_1^3 + 4x_1 - 2x_2^4 + 2x_2^2$, the partial derivative of y with respect to x_1 would be

$$\frac{\partial y}{\partial x_1} = 9x_1^2 + 4$$

, while the partial derivative of y with respect to x_2

$$\frac{\partial y}{\partial x_2} = -8x_2^3 + 4x_2$$

How to do Partial Differentiation (Cont'd)

- The ordinary least squares (OLS) estimator gives formulae for the values of the parameters that minimise the residual sum of squares, denoted by L
- The minimum of L is found by partially differentiating this function and setting the partial derivatives to zero
- Therefore, partial differentiation has a key role in deriving the main approach to parameter estimation that we use in econometrics.

Integration

- Integration is the opposite of differentiation
- If we integrate a function and then differentiate the result, we get back the original function
- Integration is used to calculate the area under a curve (between two specific points)
- Further details on the rules for integration are not given since the mathematical technique is not needed for any of the approaches used here.

Matrices - Background

- Some useful terminology:
  - A scalar is simply a single number (although it need not be a whole number – e.g. 3, -5, 0.5 are all scalars)
  - A vector is a one-dimensional array of numbers (see below for examples)
  - A matrix is a two-dimensional collection or array of numbers. The size of a matrix is given by its numbers of rows and columns
- Matrices are very useful and important ways for organising sets of data together, which make manipulating and transforming them easy
- Matrices are widely used in econometrics and finance for solving systems of linear equations, for deriving key results, and for expressing formulae.

Working with Matrices

- The dimensions of a matrix are quoted as $R \times C$, which is the number of rows by the number of columns
- Each element in a matrix is referred to using subscripts.
- For example, suppose a matrix M has two rows and four columns. The element in the second row and the third column of this matrix would be denoted m_{23}.
- More generally m_{ij} refers to the element in the ith row and the jth column.
- Thus a 2×4 matrix would have elements

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \end{pmatrix}$$

- If a matrix has only one row, it is known as a *row vector*, which will be of dimension $1 \times C$, where C is the number of columns

$$\text{e.g.} \qquad (2.7 \quad 3.0 \quad -1.5 \quad 0.3)$$

Working with Matrices

- A matrix having only one column is a column vector, which will be of dimension $R \times 1$, where R is the number of rows, e.g.

$$\text{e.g.} \begin{pmatrix} 1.3 \\ -0.1 \\ 0.0 \end{pmatrix}$$

- When the number of rows and columns is equal (i.e. $R = C$), it would be said that the matrix is square, e.g. the 2×2 matrix:

$$\begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix}$$

- A matrix in which all the elements are zero is a zero matrix.

Working with Matrices (Cont'd)

- A symmetric matrix is a special square matrix that is symmetric about the leading diagonal so that $m_{ij} = m_{ji} \quad \forall \quad i, j$

$$\text{e.g.} \qquad \begin{pmatrix} 1 & 2 & 4 & 7 \\ 2 & -3 & 6 & 9 \\ 4 & 6 & 2 & -8 \\ 7 & 9 & -8 & 0 \end{pmatrix}$$

- A diagonal matrix is a square matrix which has non-zero terms on the leading diagonal and zeros everywhere else,

$$\text{e.g.} \qquad \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Working with Matrices (Cont'd)

- A diagonal matrix with 1 in all places on the leading diagonal and zero everywhere else is known as the identity matrix, denoted by I, e.g.

$$\text{e.g.} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- The identity matrix is essentially the matrix equivalent of the number one
- Multiplying any matrix by the identity matrix of the appropriate size results in the original matrix being left unchanged

Working with Matrices (Cont'd)

- So for any matrix M, $MI=IM=M$
- In order to perform operations with matrices , they must be conformable
- The dimensions of matrices required for them to be conformable depend on the operation.

Matrix Addition or Subtraction

- Addition and subtraction of matrices requires the matrices concerned to be of the same order (i.e. to have the same number of rows and the same number of columns as one another)
- The operations are then performed element by element

$$\text{e.g., if } A = \begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 0.2 & -0.1 \\ 0 & 0.3 \end{pmatrix},$$

$$A + B = \begin{pmatrix} 0.3+0.2 & 0.6-0.1 \\ -0.1+0 & 0.7+0.3 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ -0.1 & 1.0 \end{pmatrix},$$

$$A - B = \begin{pmatrix} 0.3-0.2 & 0.6--0.1 \\ -0.1-0 & 0.7-0.3 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.7 \\ -0.1 & 0.4 \end{pmatrix}$$

Matrix Multiplication

- Multiplying or dividing a matrix by a scalar (that is, a single number), implies that every element of the matrix is multiplied by that number

$$\text{e.g. } 2A = 2\begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.6 & 1.2 \\ -0.2 & 1.4 \end{pmatrix}$$

- More generally, for two matrices A and B of the same order and for c a scalar, the following results hold

$$A + B = B + A$$

$$A + 0 = 0 + A = A$$

$$cA = Ac$$

$$c(A + B) = cA + cB$$

$$A0 = 0A = 0$$

Matrix Multiplication

- Multiplying two matrices together requires the number of columns of the first matrix to be equal to the number of rows of the second matrix
- Note also that the ordering of the matrices is important, so in general, $AB \neq BA$
- When the matrices are multiplied together, the resulting matrix will be of size (number of rows of first matrix d́7 number of columns of second matrix), e.g. $(3 \times 2) \times (2 \times 4) = (3 \times 4)$.
- More generally,
  $(a \times b) \times (b \times c) \times (c \times d) \times (d \times e) = (a \times e)$, etc.
- In general, matrices cannot be divided by one another.
  - Instead, we multiply by the inverse.

Matrix Multiplication Example

- The actual multiplication of the elements of the two matrices is done by multiplying along the rows of the first matrix and down the columns of the second

Matrix Multiplication Example (Cont'd)

$$\text{e.g.} \begin{pmatrix} 1 & 2 \\ 7 & 3 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 0 & 2 & 4 & 9 \\ 6 & 3 & 0 & 2 \end{pmatrix}$$

$$(3 \times 2) \quad (2 \times 4)$$

$$= \begin{pmatrix} ((1\times 0)+(2\times 6)) & ((1\times 2)+(2\times 3)) & ((1\times 4)+(2\times 0)) & ((1\times 9)+(2\times 2)) \\ ((7\times 0)+(3\times 6)) & ((7\times 2)+(3\times 3)) & ((7\times 4)+(3\times 0)) & ((7\times 9)+(3\times 2)) \\ ((1\times 0)+(6\times 6)) & ((1\times 2)+(6\times 3)) & ((1\times 4)+(6\times 0)) & ((1\times 9)+(6\times 2)) \end{pmatrix}$$

$$(3 \times 4)$$

$$= \begin{pmatrix} 12 & 8 & 4 & 13 \\ 18 & 23 & 28 & 69 \\ 36 & 20 & 4 & 21 \end{pmatrix}$$

$$(3 \times 4)$$

The Transpose of a Matrix

- The transpose of a matrix, written A' or A^{T}, is the matrix obtained by transposing (switching) the rows and columns of a matrix

$$\text{e.g.if } A = \begin{pmatrix} 1 & 2 \\ 7 & 3 \\ 1 & 6 \end{pmatrix} \quad \text{then } A' = \begin{pmatrix} 1 & 7 & 1 \\ 2 & 3 & 6 \end{pmatrix}$$

- If A is of dimensions $R \times C$, A' will be $C \times R$.

The Rank of a Matrix

- The rank of a matrix A is given by the maximum number of linearly independent rows (or columns). For example,

$$\mathrm{rank}\begin{pmatrix} 3 & 4 \\ 7 & 9 \end{pmatrix} = 2$$

$$\mathrm{rank}\begin{pmatrix} 3 & 6 \\ 2 & 4 \end{pmatrix} = 1$$

- In the first case, all rows and columns are (linearly) independent of one another, but in the second case, the second column is not independent of the first (the second column is simply twice the first)

- A matrix with a rank equal to its dimension is a matrix of full rank

The Rank of a Matrix (Cont'd)

- A matrix that is less than of full rank is known as a short rank matrix, and is singular
- Three important results: Rank(A) = Rank (A');

$$\text{Rank}(A) = \text{Rank}\ (A')$$
$$\text{Rank}(AB) \leq \min(\text{Rank}(A), \text{Rank}(B))$$
$$\text{Rank}\ (A'A) = \text{Rank}\ (AA') = \text{Rank}\ (A)$$

The Inverse of a Matrix

- The inverse of a matrix A, where defined and denoted A^{-1}, is that matrix which, when pre-multiplied or post multiplied by A, will result in the identity matrix,

$$\text{i.e.}\quad AA^{-1} = A^{-1}A = I.$$

- The inverse of a matrix exists only when the matrix is square and non-singular
- Properties of the inverse of a matrix include:
  - $I^{-1} = I$
  - $(A^{-1})^{-1} = A$
  - $(A')^{-1} = (A^{-1})'$
  - $(AB)^{-1} = B^{-1}A^{-1}$

Calculating Inverse of a 2 2 Matrix

- The inverse of a 2×2 non-singular matrix whose elements are

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

will be

$$\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- The expression in the denominator, $(ad - bc)$ is the determinant of the matrix, and will be a scalar

Calculating Inverse of a 2 2 Matrix (Cont'd)

- If the matrix is

$$\begin{pmatrix} 2 & 1 \\ 4 & 6 \end{pmatrix}$$

the inverse will be

$$\frac{1}{8}\begin{pmatrix} 6 & -1 \\ -4 & 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{8} \\ -\frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

- As a check, multiply the two matrices together and it should give the identity matrix I.

The Trace of a Matrix

- The trace of a square matrix is the sum of the terms on its leading diagonal
- For example, the trace of the matrix

$$A = \begin{pmatrix} 3 & 4 \\ 7 & 9 \end{pmatrix}$$

  $\mathrm{Tr}(A)$, is $3 + 9 = 12$

- Some important properties of the trace of a matrix are:
  - $\mathrm{Tr}(cA) = c\mathrm{Tr}(A)$
  - $\mathrm{Tr}(A') = \mathrm{Tr}(A)$
  - $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
  - $\mathrm{Tr}(I_N) = N$

The Eigenvalues of a Matrix

- Let Π denote a $p \times p$ square matrix, c denote a $p \times 1$ non-zero vector, and λ denote a set of scalars

  λ is called a characteristic root or set of roots of the matrix Π? if it is possible to

$$\Pi c = \lambda c$$

- This equation can also be written as $\Pi c = \lambda I_p c$ where $I_p c$ is an identity matrix, and hence $(\Pi - \lambda I_p)c = 0$

- Since $c \neq 0$ by definition, then for this system to have a non-zero solution, the matrix $(\Pi - \lambda I_p)c = 0)$ is required to be singular (i.e. to have a zero determinant), and thus

$$|\Pi - \lambda I_p| = 0$$

Calculating Eigenvalues: An Example

- Let Π be the 2×2 matrix

$$\Pi = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$$

- Then the characteristic equation is $|\Pi - \lambda I_p|$

$$= \left| \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

$$= \begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} = (5-\lambda)(4-\lambda) - 2 = \lambda^2 - 9\lambda + 18$$

- This gives the solutions $\lambda = 6$ and $\lambda = 3$.
- The characteristic roots are also known as eigenvalues
- The eigenvectors would be the values of c corresponding to the eigenvalues.

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

$$C = \left(\begin{array}{ccccc} 1 & C_{12} & C_{13} & \ldots & C_{1N} \\ C_{21} & 1 & C_{23} & \ldots & C_{2N} \\ \vdots & & & \vdots & \\ C_{N1} & C_{N2} & C_{N3} & \ldots & 1 \end{array} \right)$$

- 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

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

- 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.