



Basic Econometric Foundation to Modeling Agricultural Technology Adoption

**CIMMYT-ALP
Training
Manual
No. 1/2005
(Part One)**

Augustine Langyintuo

**Presented at a Training Course Organized by
CIMMYT-ALP for its NARS Partners in Southern
Africa on: “Econometric Application to Modeling
the Adoption of Agricultural Technologies”.
21 – 25 February, 2005, Harare, Zimbabwe**

**International Maize and Wheat Improvement Center
(CIMMYT)
PO Box MP 163, Mount Pleasant, Harare, Zimbabwe**

Table of Contents

Table of Contents.....	i
1. Introduction to the Training Course	1
1.1 Objectives of the course.....	1
1.2 Outline of the course in Part One.....	1
1.3 Expectations of the course	1
2. Basic Econometric Approach to Modeling.....	2
2.1 Definitions of Relevant Terms.....	2
2.1.1 Definitions of Selected Econometric Terminologies.....	2
2.1.2 Relevant Rudimentary Matrix Algebra.....	6
2.2 The Linear Statistical Model 1.....	14
2.3 The Linear Statistical Model 2.....	15
2.3.1 Estimating the Location Parameters	16
2.3.2 Sampling Properties.....	17
2.4 General Linear Statistical Model with Non-scalar Identity Covariance Matrix	17
2.4.1 The Statistical Model and Estimators	18
2.5 General Linear Statistical Model with an Unknown Covariance Matrix	19
2.5.1 Heteroskedasticity.....	20
2.5.2 Testing for Heteroskedasticity	21
2.5.3 Autocorrelation	22
3 Application of econometrics in modelling the adoption of agricultural technologies.....	23
3.1 Regression Analysis in Adoption Studies.....	23
3.2 Linear Probability Model (LPM).....	25
3.3 Non-linear Probability Models	26
Bibliography/Further Readings	28

1. Introduction to the Training Course

1.1 Objectives of the course

As part of an effort to improve socio-economic capacity within the NARS, CIMMYT plans on organizing a one-week course on the “**Econometric Application to Modeling Agricultural Technology Adoption**” scheduled for February 21 – 25, 2005 in Harare, Zimbabwe under the auspices of the Rockefeller Foundation Funded seed project “*Strengthening Seed Marketing Incentives in Southern Africa to Increase the Impact of Maize Breeding Research*”. The main objective of the course is to expose participants to basic econometric methods required to build effective agricultural technology adoption models and modeling using the software package STATA. The training course in itself is divided into two complementary parts. Part One deals with the basic foundations of econometrics applied to technology adoption models. Part Two takes the concepts forward and applied to the setting up technology adoption models and implementing them using the software STATA.

1.2 Outline of the course in Part One

The outline of the course in Part One is as follows: (i) Definitions of selected econometric terminologies: To provide a good appreciation of the demands for modeling, terminologies commonly used and contained in the text are defined and in some cases illustrated for clarity. (ii) Rudimentary matrix algebra: This section presents rudimentary algebra essential for understanding the regression modeling approach presented. The elements of econometrics are then presented. (iii) Elements of basic econometrics: Participants are taken through the development of regression modeling. This is followed by modifications necessary in situations where the assumptions of ordinary least squares are violated. The necessary conditions under which the linear probability models are called into play especially for modeling technology adoption decisions is clearly presented.

Part Two of the course complements Part One in its discussions on determinants of agricultural technologies, specification of agricultural technology adoption models and implementation of the models using the software STATA.

1.3 Expectations of the course

It is expected that participants of the workshop would leave with some confidence in the conduct of adoption studies. They would also have some experience in the use of STATA.

2. Basic Econometric Approach to Modeling

2.1 Definitions of Relevant Terms

2.1.1 Definitions of Selected Econometric Terminologies

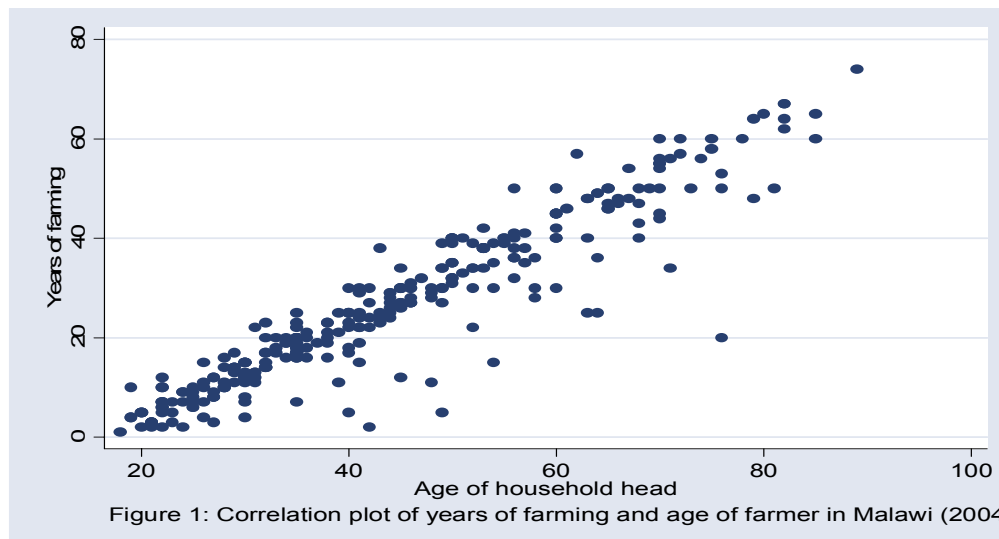
Autocorrelation: Autocorrelation means that for a given variable the observations are not independent. Each observation will tend to be close in value to the next.

Continuous/quantitative data: Data that can be measured on a continuum or scale are quantitative or continuous data. They can have almost any numeric value and can be meaningfully subdivided into finer increments, depending upon the precision of the measurement system e.g., age, length, size, width, cost, etc.

Correlation: Correlation is the degree or extent of the relationship between two variables. If the value of one variable increases as the other increases, they are said to be positively correlated. On the other hand, if the value of one variable decreases when the other variable is increasing it is said to be negatively correlated. If one variable does not affect the other they are considered to be uncorrelated. In Table 1, for example, years of farming is highly, positively correlated with age (See also Figure 1) while education negatively correlated with age. Distance to market and education are uncorrelated.

Table 1: Correlation matrix of selected socio-economic variables of farmers in Malawi

	Age	Years of farming	Education	Distance to market
Age	1.0000			
Years of farming	0.9376	1.0000		
Education	-0.2350	-0.2362	1.0000	
Distance to market	-0.0572	-0.0890	0.0006	1.0000



Discrete/qualitative (or attribute) data: Data that can be categorized into a classification based on counts are qualitative or discrete data. Only a finite number of values is possible, and the values cannot be subdivided meaningfully. For example, whether or not a farmer adopts an improved variety (yes or no) or the gender of a farmer (male or female) are qualitative data.

Homoskedastic (and heteroskedasticity): A univariate stochastic process X is said to be homoskedastic if the standard deviations are constant for all observations. Otherwise, it is said to be heteroskedastic. This is illustrated with realizations of two stochastic processes in Figure 2.

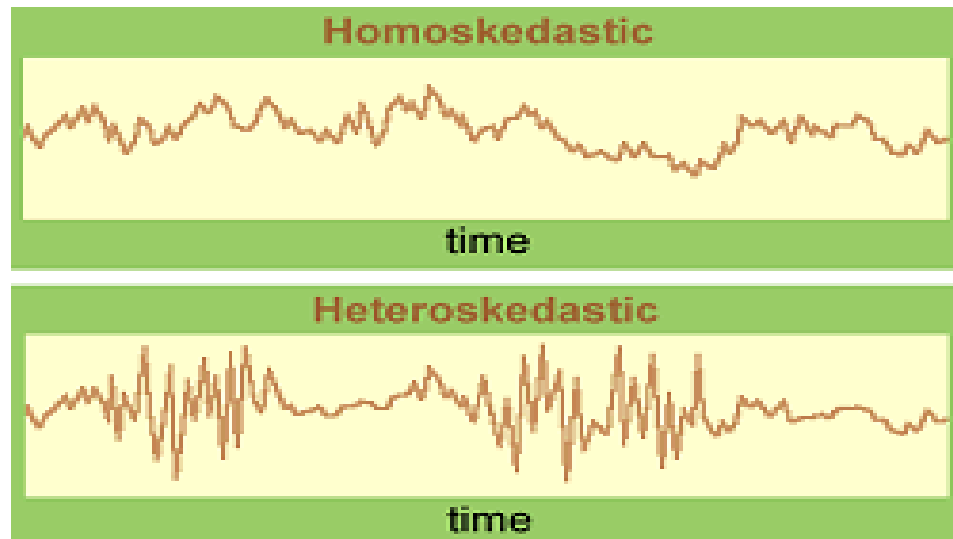


Figure 2: Graphs of homoskedastic versus heteroskedastic distributions

Source: <http://www>.

Mean: The mean is the average data point value within a data set. To calculate the mean, add all of the individual data points and then divide that figure by the total number of data points.

Median: The median is the middle point of a data set; 50% of the values are below this point, and 50% are above this point when all possible values are listed in order.

Normal Distribution: Normal distribution is the spread of information (such as demographics) where the most frequently occurring value is in the middle of the range and other probabilities tail off symmetrically in both directions (Figure 3). Normal distribution sometimes referred to as Gaussian distribution is graphically categorized by a bell-shaped curve. For normally distributed data, the mean and median are very close and may be identical.

Null Hypothesis (H0): A null hypothesis (H0) is a stated assumption that there is no difference in parameters (mean, variance, etc) for two or more populations. According to the null hypothesis, any observed difference in samples is due to chance or sampling error.

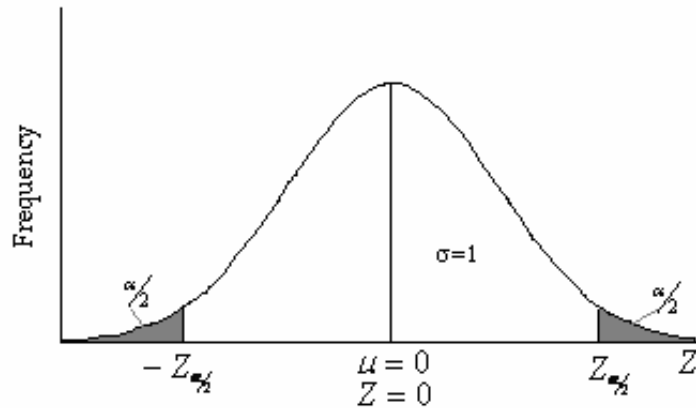


Figure 3: A plot of normal distribution

Probability: Probability refers to the chance of something happening or the fraction of occurrences over a large number of trials. Probability can range from 0 (no chance) to 1 (full certainty).

Random Sample: By definition, a sample of size n is random if the probability of selecting the sample is the same as the probability of selecting every other sample of size n . If the sample is not random, a bias is introduced which causes a statistical sampling or testing error by systematically favoring some outcomes over others.

Regression analysis: Regression analysis is a method of analysis that enables you to quantify the relationship between two or more variables (X) and (Y) by fitting a line through all the points such that they are evenly distributed about the line.

Sample: A Sample is a portion or subset of units taken from the population whose characteristics that would be used for analysis are considered to be identical with a notion that any unit can represent the population.

Sample size determination: In empirical research, it is always important to determine sample size because samples that are too large may waste time, resources and money, while samples that are too small may lead to inaccurate results. In many cases, we can easily determine the minimum sample size needed to estimate a process parameter, such as the population mean μ .

It is usually the case that when sample data are collected and the sample mean \bar{x} calculated, it is typically different from the population mean μ due to sampling error. The margin of error, E , is the maximum difference between the observed sample mean \bar{x} and the true value of the population mean μ given as:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is the critical value or positive z value that is at the vertical boundary for the area of $\alpha/2$ in the right tail of the standard normal distribution (Figure 3). σ is the population standard deviation and n the sample size. Using the formula above, one can solve for the value of n that would produce results accurate to a specified confidence and margin of error, E for known σ or small samples. That is,

$$n = \left[\frac{z_{\alpha/2} \sigma}{E} \right]^2$$

It is not possible to know σ without knowing μ , however, σ can be generated from a pilot test or simulation.

For example, assume we want to estimate the mean farm size of households in Malawi, how many households must we randomly select to be 95% sure that the sample mean is within 0.5 ha of the population mean μ ? Assume that a previous survey of households has shown $\sigma = 2.6$ ha. We need to solve for the sample size n .

A 95% degree confidence corresponds to $\alpha = 0.05$. Each of the shaded tails in the following figure has an area of $\alpha/2 = 0.025$. The region to the left of $z_{\alpha/2}$ and to the right of $z = 0$ is $0.5 - 0.025$, or 0.475 (See Figure 4). In the Table of the Standard Normal (z) Distribution, an area of 0.475 corresponds to a z value of 1.96. The critical value is therefore $z_{\alpha/2} = 1.96$. Given $E = 0.5$ and $\sigma = 2.6$, n can be estimated as:

$$n = \left[\frac{z_{\alpha/2} \sigma}{E} \right]^2 = \left[\frac{1.96 \times 2.6}{0.5} \right]^2 = [10.192]^2 = 103.89 \cong 104$$

So we will need to sample at least 104 randomly selected households to be 95% confident that the sample mean farm size \bar{x} will be within 0.5 ha of the true population farm size.

Standard Deviation: Standard deviation is a statistic used to measure the variation in a distribution or spread of data in relation to the mean. Sample standard deviation, s , is equal to the square root of the sum of the squared deviations of the mean divided by the sample size minus 1 as below.

$$s = \sqrt{\frac{\sum (\bar{x} - \mu)^2}{n - 1}}$$

Where the whole population is known, the minus 1 should be omitted.

t Statistic: The t statistic is used to determine whether two means are statistically different. The formula uses the means of the two samples, their standard deviation and sample size. The t value is then evaluated against your alpha to determine if your null hypothesis can be rejected or not.

t Test: The t test employs the statistic (t), with n-1 degrees of freedom, to test a given statistical hypothesis about a population parameter. Usually used with small sample sizes (<30) population standard deviation unknown.

Variance: The sum of the squared deviations of n measurements from their mean divided by (n-1).

$$v = \frac{\sum (\bar{x} - \mu)^2}{n - 1}$$

Z: A Z value is a data point's position between the mean and another location as measured by the number of standard deviations. Z is a universal measurement because it can be applied to any unit of measure.

2.1.2 Relevant Rudimentary Matrix Algebra

To motivate our discussion on matrices, consider the following simple linear-equation system:

$$\begin{aligned} 6x_1 + 3x_2 + x_3 &= 33 \\ 2x_1 + 4x_2 - 3x_3 &= 12 \\ 4x_1 - x_2 + 5x_3 &= 10 \end{aligned}$$

This can be re-written as:

$$A = \begin{bmatrix} 6 & 3 & 1 \\ 2 & 4 & -3 \\ 4 & -1 & 5 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad d = \begin{bmatrix} 33 \\ 12 \\ 10 \end{bmatrix}$$

Each of the three arrays constitutes a matrix.

Definition of a matrix: A matrix is defined as a rectangular array of numbers, parameters or elements arranged in rows and columns. The members in the array are usually enclosed in brackets and sometimes in parenthesis or double vertical lines, $\| \|$. A matrix of order, or dimension, M by N (usually written as M x N) is a set of M x N elements arranged in M rows and N columns. Thus, letting boldface letters denote matrices, an (M x N) matrix **A** may be expressed as

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ a_{M1} & a_{M2} & a_{M3} & \dots & a_{MN} \end{bmatrix}$$

Where a_{ij} is the element appearing in the i^{th} row and the j^{th} column of \mathbf{A} and where $[a_{ij}]$ is a shorthand expression for the matrix \mathbf{A} whose typical element is a_{ij} . The order, or dimension, of a matrix (i.e., the number of rows and columns) is often written underneath the matrix for easy reference as shown below.

$$\mathbf{A}_{2 \times 3} = \begin{bmatrix} 2 & 1 & 4 \\ 6 & 0 & 2 \end{bmatrix} \quad \mathbf{B}_{3 \times 3} = \begin{bmatrix} 1 & 5 & 7 \\ -1 & 1 & 2 \\ 6 & 0 & 10 \end{bmatrix}$$

Scalar: A scalar is a single (real) number. Alternatively, a scalar is a 1x1 matrix. For example, 2; 3; 10; etc.

Column Vector: A matrix consisting of M rows and only one column is called a column vector. Letting the boldface lowercase letters denote vectors, an example of a vector is:

$$\mathbf{x}_{4 \times 1} = \begin{bmatrix} 3 \\ 5 \\ 2 \\ 8 \end{bmatrix}$$

Row Vector: A matrix consisting of only one row and N columns is called a row vector.

$$\mathbf{y}_{1 \times 4} = [3 \quad 5 \quad 2 \quad 8]$$

Transposition: The transpose of an M x N matrix \mathbf{A} , denoted by A prime, \mathbf{A}' , is an N x M matrix obtained by interchanging the rows and columns of \mathbf{A} . That is, the i^{th} row of \mathbf{A} becomes the i^{th} column of \mathbf{A}' . For example,

$$\mathbf{A}_{3 \times 2} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \\ 0 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{A}'_{2 \times 3} = \begin{bmatrix} 3 & 1 & 0 \\ 2 & 4 & 5 \end{bmatrix}$$

This means that the transpose of a row vector is a column vector and the transpose of a column vector is a row vector. For example:

$$\mathbf{X} = [2 \quad 4 \quad 5] \quad \text{and} \quad \mathbf{X}' = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

2.1.2.1 Types of Matrices

Square Matrix: A matrix that has the same number of rows as columns is called a square matrix.

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 7 & 6 & 5 \\ 4 & 2 & 1 \end{bmatrix}$$

Diagonal Matrix: A square matrix with at least one nonzero element on the main diagonal (running from the upper-left-hand corner to the lower-right-hand corner) and zeros elsewhere.

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scalar Matrix: A diagonal matrix whose diagonal elements are all equal is called a scalar matrix. An example is the variance-covariance matrix of classical linear regression model given as:

$$\text{var-cov}(u) = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

Identity (or Unit Matrix): An identity (or unit) matrix is a special kind of scalar matrix whose diagonal elements are all 1's.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Symmetric Matrix: A square matrix whose elements above the main diagonal are mirror images of the elements below the main diagonal is called a symmetric matrix. The transpose of a symmetric matrix is itself, i.e., $\mathbf{A} = \mathbf{A}'$. The element a_{ij} of \mathbf{A} is equal to the element a_{ji} of \mathbf{A}' .

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 9 \\ 2 & 6 & 3 \\ 9 & 3 & 8 \end{bmatrix}$$

Null Matrix: A matrix whose elements are all zeros is called a null matrix usually denoted by $\mathbf{0}$.

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Null Vector: A row or column vector whose elements are all zero is called a null vector usually denoted by $\mathbf{0}$.

$$\mathbf{X} = [0 \ 0 \ 0] \quad \mathbf{Y} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Equal Matrices: Two matrices \mathbf{A} and \mathbf{B} are said to be equal if they are of the same order and their corresponding elements are equal. That is, $a_{ij} = b_{ij}$ for all i and j as shown the matrices below.

$$\mathbf{A}_{3 \times 3} = \begin{bmatrix} 1 & 2 & 9 \\ 2 & -6 & 3 \\ 9 & 3 & 8 \end{bmatrix} \quad \mathbf{B}_{3 \times 3} = \begin{bmatrix} 1 & 2 & 9 \\ 2 & -6 & 3 \\ 9 & 3 & 8 \end{bmatrix}$$

2.1.2.2 Addition, Subtraction and Multiplication Properties of Matrices

Matrix Addition: The addition of matrices \mathbf{A} and \mathbf{B} is only possible if \mathbf{A} and \mathbf{B} are conformable for addition (i.e., of the same order). Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$. If \mathbf{A} and \mathbf{B} are of the same order, we define matrix addition as: $\mathbf{A} + \mathbf{B} = \mathbf{C}$ where \mathbf{C} is of the same order as \mathbf{A} and \mathbf{B} . \mathbf{C} is obtained by adding the corresponding elements of \mathbf{A} and \mathbf{B} such that $c_{ij} = a_{ij} + b_{ij}$ for all i and j as in the example below:

$$\text{Let } \mathbf{A} = [a_{ij}] = \begin{bmatrix} 2 & 3 & 4 \\ 1 & -3 & 5 \end{bmatrix} \text{ and } \mathbf{B} = [b_{ij}] = \begin{bmatrix} 1 & 2 & -1 \\ 6 & 4 & 0 \end{bmatrix}$$

This means that \mathbf{C} will be given by:

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 5 & 4 \\ 7 & 1 & 5 \end{bmatrix}$$

Matrix Subtraction: Matrix subtraction follows the same principle as matrix addition provided the matrices to be subtracted are of the same order. For $\mathbf{C} = \mathbf{A} - \mathbf{B}$, we subtract the elements of \mathbf{B} from the corresponding elements of \mathbf{A} .

$$\text{Let } \mathbf{A} = [a_{ij}] = \begin{bmatrix} 2 & 3 & 4 \\ 1 & -3 & 5 \end{bmatrix} \text{ and } \mathbf{B} = [b_{ij}] = \begin{bmatrix} 1 & 2 & -1 \\ 6 & 4 & 0 \end{bmatrix}$$

This means that \mathbf{C} will be given by:

$$\mathbf{C} = \mathbf{A} - \mathbf{B} = \begin{bmatrix} 1 & 1 & 5 \\ -5 & -7 & 5 \end{bmatrix}$$

Scalar Multiplication: To multiply a matrix \mathbf{A} by a scalar λ (a real number), we multiply each element of the matrix by λ : For $\lambda = 2$ and $\mathbf{A} = [a_{ij}] = \begin{bmatrix} 2 & 3 & 4 \\ 1 & -3 & 5 \end{bmatrix}$, $\lambda\mathbf{A}$ will be given by:

$$\lambda\mathbf{A} = [\lambda a_{ij}] = \begin{bmatrix} 4 & 6 & 8 \\ 2 & -6 & 10 \end{bmatrix}$$

Matrix Multiplication: For a given matrix \mathbf{A} of order $M \times N$ and \mathbf{B} of order $N \times P$, their product \mathbf{AB} defined as a new matrix \mathbf{C} will be of order $M \times P$. In matrix multiplication one follows the the “row by column” rule of multiplication whereby one multiplies the elements of the i th row of \mathbf{A} by the corresponding elements of the j th column of \mathbf{B} and summing over all terms. For instance, to obtain c_{11} , the element in the first row and the first column of \mathbf{C} , we multiply the elements in the first row of \mathbf{A} by the corresponding elements in the first column of \mathbf{B} and sum over all terms. Similarly, to obtain c_{12} , we multiply the elements in the first row of \mathbf{A} by the corresponding elements in the second column of \mathbf{B} and sum over all terms, and so on (See example below).

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$\mathbf{C} = \mathbf{AB} = [c_{ij}] = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \\ a_{31}b_{11} + a_{32}b_{21} & a_{31}b_{12} + a_{32}b_{22} & a_{31}b_{13} + a_{32}b_{23} \end{bmatrix}$$

Using figures, we have:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & -2 \\ 1 & 4 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 4 & 3 & 1 \\ -2 & 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 2x^4 + 1x - 2 & 2x^3 + 1x^0 & 2x^1 + 1x^1 \\ 3x^4 + -2x - 2 & 3x^3 + -2x^0 & 3x^1 + -2x^1 \\ 1x^4 + 4x - 2 & 1x^3 + 4x^0 & 1x^1 + 4x^1 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 3 \\ 16 & 9 & 1 \\ -4 & 3 & 5 \end{bmatrix}$$

Note the multiplication of matrices **A** and **B** is only possible if they are conformable with respect to multiplication. That is, the number of **columns** in **A** must be equal to the number of **rows** in **B**.

Properties of Matrix Multiplication:

1. Matrix multiplication is not necessarily commutative. That is, $\mathbf{AB} \neq \mathbf{BA}$. Therefore, the order in which matrices are multiplied is very important. \mathbf{AB} means that **A** is postmultiplied by **B** or **B** is premultiplied by **A**.
2. Even if \mathbf{AB} and \mathbf{BA} exist, the resulting matrices may not be of the same order. Thus, if **A** is $M \times N$ and **B** is $N \times M$, \mathbf{AB} is $M \times M$ whereas \mathbf{BA} is $N \times N$ and hence of different order.
3. Even if **A** and **B** are both square matrices, so that \mathbf{AB} and \mathbf{BA} are both defined, the resulting matrices will not necessarily be equal as shown below.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$

Then

$$AB = \begin{bmatrix} 5 & 10 \\ 11 & 24 \end{bmatrix} \text{ and } BA = \begin{bmatrix} 13 & 18 \\ 8 & 14 \end{bmatrix}$$

4. A row vector postmultiplied by a column vector is a scalar.

$$\mathbf{X} = [2 \quad 4 \quad 5] \text{ and } \mathbf{Y} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}, \text{ then } \mathbf{XY} = [55]$$

5. A column vector postmultiplied by a row vector produces a symmetric matrix.

$$\mathbf{Y} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} \text{ and } \mathbf{X} = [2 \quad 4 \quad 5] \text{ then } \mathbf{YX} = \begin{bmatrix} 4 & 8 & 10 \\ 8 & 16 & 20 \\ 10 & 20 & 25 \end{bmatrix}$$

6. A matrix postmultiplied by a column vector is a column vector. (*Try out an example*)
7. A row vector postmultiplied by a matrix is a row vector. (*Try out an example*)
8. Matrix multiplication is associative, i.e., $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, where **A** is $M \times N$, **B** is $N \times P$, and **C** is $P \times K$.
9. Matrix multiplication is distributive with respect to addition, i.e., $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ and $(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$.

Matrix Division: It is not possible to divide one matrix by another!!

2.1.2.3 Evaluation of a Determinant of a Matrix

Definition of determinant: To every square matrix, \mathbf{A} , there corresponds a number (scalar) known as the determinant of the matrix, which is denoted by $\det \mathbf{A}$ or by $|\mathbf{A}|$, where $|\mathbf{A}|$ means "the determinant of."

Evaluation of a Determinant: The process of finding the value of a determinant is known as expansion, or reduction of the determinant. This is done by entries of the matrix in a well-defined manner.

$$\text{If } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \text{ then } |\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\text{If } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \text{ then}$$

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Determinant of a 3 x 3 matrix can also be evaluated using the expansion method.

Properties of Determinants:

1. A matrix whose determinant is zero is called a singular matrix, whereas a matrix with a nonzero determinant is called a nonsingular matrix.
2. If all the elements of any row of \mathbf{A} are zero, its determinant is zero. For example:

$$|\mathbf{A}| = \begin{vmatrix} 0 & 0 & 0 \\ 1 & 4 & 6 \\ 8 & -2 & 5 \end{vmatrix} = 0$$

3. If two rows or columns of a matrix are identical, its determinant is zero.
4. If one row or a column of a matrix is a multiple of another row or column of that matrix, its determinant is zero. More generally, if any row (column) of a matrix is a linear combination of other rows (columns), its determinant is zero.
5. The determinants of \mathbf{A} and \mathbf{A} transpose are the same.
6. Interchanging any two rows or any two columns of a matrix \mathbf{A} changes the sign of \mathbf{A} .
7. If every element of a row or a column of \mathbf{A} is multiplied by a scalar λ , then $|\mathbf{A}|$ is multiplied by λ .

Minor: If the i th row and j th column of an $N \times N$ matrix \mathbf{A} are deleted, the determinant of the resulting submatrix is called the minor of the element a_{ij} denoted by $|\mathbf{M}_{ij}|$.

Example

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The minor of a_{11} is

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32}$$

Cofactor: The cofactor of the element a_{ij} of an $N \times N$ matrix \mathbf{A} , denoted by c_{ij} , defined as

$$c_{ij} = (-1)^{i+j} |\mathbf{M}_{ij}|$$

is a signed minor, the sign being positive if $i + j$ is even and being negative if $i + j$ is odd. For example, the cofactor of the element a_{11} of the 3×3 matrix \mathbf{A} given above is $a_{22}a_{33} - a_{23}a_{32}$ (a positive number).

Cofactor matrix: Replacing the elements a_{ij} of a matrix \mathbf{A} by their cofactors, we obtain a matrix known as the cofactor matrix of \mathbf{A} , denoted by $(\text{cof } \mathbf{A})$.

Adjoint matrix: The adjoint matrix, written as $(\text{adj } \mathbf{A})$, is the transpose of the cofactor matrix; that is $(\text{adj } \mathbf{A}) = (\text{cof } \mathbf{A})'$.

Matrix Inversion: An inverse of a square matrix \mathbf{A} , denoted by \mathbf{A}^{-1} (read as \mathbf{A} inverse), if it exists, is a unique square matrix such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ as shown below. (Note that \mathbf{I} is an identity matrix whose order is the same as that of \mathbf{A}).

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Finding the Inverse of a Square Matrix: If \mathbf{A} is square and nonsingular matrix (that is, $|\mathbf{A}| \neq 0$), its inverse \mathbf{A}^{-1} can be found as follows:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} (\text{adj } \mathbf{A})$$

The steps to follow are:

1. Find the determinant of \mathbf{A} . If it is nonzero, proceed to step 2.
2. Replace each element a_{ij} of \mathbf{A} by its cofactor to obtain the cofactor matrix.
3. Transpose the cofactor matrix to obtain the adjoint matrix $(\text{adj } \mathbf{A})$.
4. Divide each element of the adjoint matrix by $|\mathbf{A}|$.

2.2 The Linear Statistical Model 1

Consider a sample y of T observations that we assume have been drawn from a distribution that has a mean (location parameter) of β and a variance (scale parameter) of σ^2 . Also assume that the T outcomes are independent from drawing to drawing. Thus, the random variable $y_t \sim (\beta, \sigma^2)$ and independence implies that the covariance $E[(y_t - \beta)(y_s - \beta)] = 0$ for $s \neq t$. We can thus model each sample outcome y_t as composed of its mean β and an unobservable random component e_t . Consequently,

$$y_t = \beta + e_t \quad \dots (1)$$

where the error term, $e_t = y_t - \beta$. The mean of e_t , $E[e_t] = 0$. The variance, $E[e_t - E[e_t]]^2 = E[e_t^2] = \sigma^2$ for all $t = 1, 2, \dots, T$. And e_t is an iid (*independently and identically distributed*) random variable.

The T equations, one for each y_t , may be written as:

$$\begin{aligned} y_1 &= \beta + e_1 \\ y_2 &= \beta + e_2 \\ &\vdots \\ y_T &= \beta + e_T \end{aligned} \quad \dots (2)$$

This may be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} = \begin{bmatrix} \beta \\ \beta \\ \vdots \\ \beta \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} \quad \dots (3)$$

Letting \mathbf{x} represent a vector of ones and $\mathbf{e} = (e_1, e_2, \dots, e_T)'$, we may write (3) as

$$y = \mathbf{x}\beta + \mathbf{e} \quad \dots (4)$$

where,

$$E[\mathbf{y}] = \mathbf{x}\beta \text{ and } E[\mathbf{e}] = \mathbf{0}.$$

In terms of the covariance, we have assumed that the random variables y_t , and e_t are statistically independent. Given that $E[e] = 0$, for the vector of random variables \mathbf{e} , the corresponding matrix of covariances may be represented as:

$$E\{[\mathbf{e} - E[\mathbf{e}]][\mathbf{e} - E[\mathbf{e}]]'\} = E[\mathbf{e}\mathbf{e}'] \quad \dots (5)$$

Thus,

$$E[\mathbf{ee}'] = E \left[\begin{array}{c} e_1 \\ e_2 \\ \vdots \\ e_T \end{array} \right]_{(Tx1)} \left[e_1 \ e_2 \ \cdots \ e \right]_{(1xT)} \quad \dots (6)$$

or

$$E[\mathbf{ee}'] = \begin{bmatrix} E[e_1^2] & E[e_1 e_2] & \cdots & E[e_1 e_T] \\ E[e_2 e_1] & E[e_2^2] & \cdots & E[e_2 e_T] \\ \vdots & \vdots & \ddots & \vdots \\ E[e_T e_1] & E[e_T e_2] & \cdots & E[e_T^2] \end{bmatrix} \quad \dots (7)$$

By assumption, e_i and e_j are independent random variables, and thus the covariance between e_i and e_j , for $i \neq j$, is 0. Hence, $E[e_i e_j] = 0$, for $i \neq j$. We may thus write (7) as:

$$E[\mathbf{ee}'] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 \mathbf{I}_T \quad \dots (8)$$

In (8), \mathbf{I}_T denotes a T th order identity matrix and $\sigma^2 \mathbf{I}_T$ is a scalar diagonal matrix.

The key assumption here is that elements of the random error vector \mathbf{e} are uncorrelated and the variance is identical (i.e., $E[\mathbf{ee}'] = \sigma^2 \mathbf{I}_T$).

2.3 The Linear Statistical Model 2

We have considered the specification of a model that could be used to describe the data-generating process for a sample of data with a single location parameter β and the scale parameter σ^2 . We introduce additional information and recognize that economic variables are interrelated and that the value that one variable takes on may condition the mean outcome for another variable. Thus, for example, from economic theory we are led to visualize that price conditions the level of consumption of a particular commodity, the level of an input conditions the level of output, and the level of consumption is affected by the level of income. To model this type of situation, let y_t be the outcome or response variable and x_{12} be the explanatory, instrument, or conditioning variable. Thus, y_t is a random variable and x_{12} is fixed or nonstochastic. The observed random variable y_t , may therefore be modeled as:

$$y_t = E[y_t] + e_t = \beta_1 + x_{12} \beta_2 + e_t \quad \dots (9)$$

where β_1 reflects the level and β_2 reflects the slope of the relationship that is linear in terms of the parameters. The parameters β_1 , β_2 and the random variable e_t , are unobserved and

unobservable. e_t is an *iid* random variable with mean $E[e_t] = 0$, variance $E[(e_t)^2] = \sigma^2$ and covariance $E[(e_t e_s)] = 0$ for $t \neq s$ implying that the random vector $\mathbf{e} \sim (0, \sigma^2 \mathbf{I}_T)$.

Given (9), we may write the statistical model for the sample y_1, y_2, \dots, y_T as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_1 + \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{T2} \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} \quad \dots (10)$$

or, compactly, as

$$y = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \mathbf{e} = (\mathbf{x}_1 \quad \mathbf{x}_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \mathbf{e} = \mathbf{X} \boldsymbol{\beta} + \mathbf{e} \quad \dots (11)$$

where \mathbf{x}_1 is a $(T \times 1)$ vector of ones and \mathbf{x}_2 is a $(T \times 1)$ vector of values of x_{12} the control variable; \mathbf{X} is a $(T \times 2)$ known matrix, and $\boldsymbol{\beta}$ is a (2×1) vector of unknown location parameters. \mathbf{y} has mean $E[\mathbf{y}] = \mathbf{X} \boldsymbol{\beta}$ and covariance matrix

$$E[(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})'] = E[\mathbf{e} \mathbf{e}'] = \sigma^2 \mathbf{I}_T \quad \dots (12)$$

\mathbf{y} and \mathbf{X} are observed and $\boldsymbol{\beta}$, σ^2 and \mathbf{e} are unobserved and unobservable.

2.3.1 Estimating the Location Parameters

If we have a theoretical model explaining how the sample data are generated and assuming that the scale parameter σ^2 is known, we are then confronted with the problem of how best to use our sample information \mathbf{y} , in conjunction with the known vector \mathbf{x}_2 , to estimate the unknowns (β_1 and β_2) representing the unknown level and slope coefficients for the economic relationship under study.

We need a rule (estimator) that will specify how the sample data should be used to estimate $\boldsymbol{\beta}$. We may use the least squares criterion to obtain such a rule. By this criterion, given a sample of observed values of the random variables y_1, y_2, \dots, y_T to obtain an estimate \mathbf{b} for the unknown parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)$, an estimator is chosen that minimizes the error sum of squares $\sum e_i^2 = \mathbf{e}' \mathbf{e}$.

More formally,

$$\begin{aligned} S &= \sum (y_t - x_{t1} \beta_1 - x_{t2} \beta_2)^2 \\ &= \mathbf{e}' \mathbf{e} = (\mathbf{y} - \mathbf{x}_1 \beta_1 - \mathbf{x}_2 \beta_2)' (\mathbf{y} - \mathbf{x}_1 \beta_1 - \mathbf{x}_2 \beta_2) \\ &= (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})' = \mathbf{y}' \mathbf{y} - 2 \boldsymbol{\beta}' \mathbf{X}' \mathbf{y} + \boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta} \end{aligned} \quad \dots (13)$$

Using rules of differentiation with respect to a vector and algebraic manipulations, the following least squares estimator can be obtained:

$$\mathbf{b} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \quad \dots (14)$$

2.3.2 Sampling Properties

Since \mathbf{b} is a linear function of the sample observations \mathbf{y} , which is a random vector, the least squares estimator of the location vector \mathbf{b} is also a random vector. Consequently the estimate of $\boldsymbol{\beta}$ would vary from sample to sample and hence would have a mean and exhibit sample variability. Using expectations, the mean is estimated thus:

$$\begin{aligned} E[\mathbf{b}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}] = E[\mathbf{1}\boldsymbol{\beta}] + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \boldsymbol{\beta} \end{aligned} \quad \dots (15)$$

Since $E[e_i] = 0$.

This means that using the least squares estimator leads to unbiased rule to estimate β_1 and β_2 . Note that this does not mean that β_1 and β_2 are unbiased, but rather the rule or estimator.

From above,

$$\begin{aligned} \mathbf{b} &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \text{ and hence the covariance of } \mathbf{b} \text{ can be written as:} \\ E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] & \\ &= E[(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} - \boldsymbol{\beta})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} - \boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}\mathbf{e}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \quad \dots (16) \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

(since $E[\mathbf{e}\mathbf{e}'] = \sigma^2\mathbf{I}_T$)

2.4 General Linear Statistical Model with Non-scalar Identity Covariance Matrix

The simplifying assumptions of the simple linear statistical model, $y = \mathbf{X}\boldsymbol{\beta} + e$, of uncorrelated elements of the error vector and identical variance may not always hold true. For example, when using cross-section data to estimate a farmer relationship it is likely that large disturbances (e_i 's) will tend to be associated with large farmers and small errors with small farmers implying that the e_i 's have different variances (*heteroskedasticity*). With time-series data, the impact of e_t on the dependent variable will not always be completely instantaneous, therefore implying that the e_t 's are correlated (*autocorrelation*). Equations on different farms may be written as one single model in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ whereby it is possible that errors for different farm communities at the same point in time are likely to be correlated (*contemporaneous correlation*). These three cases violate the scalar identity covariance assumption of the simple OLS. In this case, the covariance matrix is specified as $\sigma^2\boldsymbol{\Psi}$ instead of $\sigma^2\mathbf{I}_T$ as in the earlier case.

2.4.1 The Statistical Model and Estimators

Consider the statistical model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad \dots (17)$$

where \mathbf{y} is a $(T \times 1)$ vector of observations, \mathbf{X} is a known $(T \times K)$ design matrix, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of unknown coefficients, and \mathbf{e} is a $(T \times 1)$ random vector with mean vector $E[\mathbf{e}] = 0$ and covariance matrix $E[\mathbf{e}\mathbf{e}'] = \boldsymbol{\Phi} = \sigma^2 \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a $(T \times T)$ known positive definite symmetric matrix and σ^2 is an unknown scalar. The variances of the random variables e_t are not all identical and/or there may be a nonzero covariance between e_t and e_τ if $t \neq \tau$.

2.4.1.1 The Least Squares Estimator of $\boldsymbol{\beta}$

The random variable \mathbf{y} has a mean $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ and covariance $E[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'] = \boldsymbol{\Phi} = \sigma^2 \boldsymbol{\Psi}$
The linear estimator

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}\boldsymbol{\beta} + \mathbf{e}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \quad \dots (18)$$

with a mean of

$$\begin{aligned} E[\mathbf{b}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}] \\ &= E[\mathbf{1}\boldsymbol{\beta}] + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \boldsymbol{\beta} \end{aligned} \quad \dots (19)$$

since $E[\mathbf{e}] = 0$.

The covariance is given by

$$\begin{aligned} E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})'] &= E[(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} - \boldsymbol{\beta})(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} - \boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Psi}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad \dots (20)$$

But because $\boldsymbol{\Psi} \neq \mathbf{I}$, this covariance matrix is different from $\Sigma\mathbf{b} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

2.4.1.2 The Generalized Least Squares Estimator

To develop a best linear unbiased estimator for the case when the error process has the covariance matrix $E[\mathbf{e}\mathbf{e}'] = \sigma^2 \boldsymbol{\Psi}$, we first need to transform the statistical model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ by multiplying it by a $(T \times T)$ matrix \mathbf{P} which has the property that $\mathbf{P}\boldsymbol{\Psi}\mathbf{P}' = \mathbf{I}_T$. Since $\boldsymbol{\Psi}$ is positive definite, a matrix \mathbf{P} with the property $\mathbf{P}\boldsymbol{\Psi}\mathbf{P}' = \mathbf{I}_T$ always exists.

Using the matrix \mathbf{P} to transform the model yields

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{e} \quad \dots (21)$$

Or

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{e}^* \quad \dots (22)$$

where $\mathbf{y}^* = \mathbf{P}\mathbf{y}$, $\mathbf{X}^* = \mathbf{P}\mathbf{X}$, and $\mathbf{e}^* = \mathbf{P}\mathbf{e}$.

The transformed vector \mathbf{e}^* has mean
 $E[\mathbf{e}^*] = E[P\mathbf{e}] = PE[\mathbf{e}] = 0$... (23)

and covariance matrix
 $E[\mathbf{e}^*\mathbf{e}^{*\prime}] = E[Pe\mathbf{e}'P'] = PE[ee']P' = \sigma^2 P\Psi P' = \sigma^2 \mathbf{I}_T$... (24)

The transformed error vector \mathbf{e}^* has the same properties as assumed previously; its elements are uncorrelated and have identical variances. Therefore the least squares estimator

$\hat{\beta} = (X^*{}' X^*)^{-1} X^*{}' y^*$ or $\hat{\beta} = (X' P' P X)^{-1} X' P' P y$ is best linear unbiased. Since $P\Psi P' = \mathbf{I}_T$ it follows that $\Psi = P^{-1}P'^{-1}$ and $\Psi^{-1} = P'P$. Thus, the least squares estimator applied to the transformed observations is given by

$$\hat{\beta} = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}y \quad \dots (25)$$

This estimator is known as the *generalized least squares estimator* which is the best linear unbiased estimator for β in the model $y = X\beta + \mathbf{e}$, when $E[\mathbf{e}] = 0$ and $E[\mathbf{e}\mathbf{e}'] = \sigma^2 \Psi$.

The mean of $\hat{\beta}$ is given as

$$E[\hat{\beta}] = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}E[y] = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}X\beta = \beta \quad \dots (26)$$

The covariance matrix is

$$\Sigma_{\hat{\beta}} = \sigma^2 (X^*{}' X^*)^{-1} = \sigma^2 (X'\Psi^{-1}X)^{-1} \quad \dots (27)$$

This covariance matrix differs from the covariance matrix obtained by using the least squares rule directly with the original untransformed model; that is,

$$\sigma^2 (X'\Psi^{-1}X)^{-1} \neq \sigma^2 (X'X)^{-1} X'\Psi X (X'X)^{-1} \quad \dots (28)$$

2.5 General Linear Statistical Model with an Unknown Covariance Matrix

In the previous sections we looked at the general linear statistical model

$$y = X\beta + \mathbf{e} \quad \dots (29)$$

where X is a $(T \times K)$ observable nonstochastic matrix, β is a $(K \times 1)$ vector of parameters to be estimated, y is a $(T \times 1)$ observable random vector, \mathbf{e} is a $(T \times 1)$ unobservable random vector with properties $E[\mathbf{e}] = 0$ and $E[\mathbf{e}\mathbf{e}'] = \Phi = \sigma^2 \Psi$ and Ψ is a known positive definite symmetric matrix but the scale parameter σ^2 assumed unknown implying that the disturbance covariance matrix Φ is known except for a factor of proportionality. Under these conditions, the generalized least squares estimator $\hat{\beta} = (X'\Psi^{-1}X)^{-1} X'\Psi^{-1}y$ is best linear unbiased and has covariance matrix $\sigma^2 (X'\Psi^{-1}X)^{-1}$. But the least squares estimator $\mathbf{b} = (X'X)^{-1} X'y$ is unbiased, but it is inefficient because its covariance matrix $\sigma^2 (X'X)^{-1} X'\Psi X (X'X)^{-1}$ exceeds

$$\sigma^2(X'\Psi^{-1}X)^{-1}$$

Thus if we assume that $E[ee'] = \sigma^2 I$ and use the least squares estimator \mathbf{b} , when $E[ee'] = \sigma^2 \Psi$, we will not be obtaining the best out of the class of possible linear estimators. Additionally, we are likely to obtain a biased estimator of the matrix for \mathbf{b} , eventually leading to misleading inferences about β through hypothesis tests and interval estimation. Under these conditions, the conventional estimator for the covariance matrix of \mathbf{b} is $\sigma^2(X'X)^{-1}$, where $\sigma^2 = (y - X\mathbf{b})'(y - X\mathbf{b})/(T - K)$, and this is a biased estimator when $E[ee'] = \sigma^2 \Psi$. As a result, the natural questions when approaching an applied problem are: How do we tell whether or not $\Psi = I$ and, if $\Psi = I$, how are the elements Ψ determined?

One usually assumes the structure of Ψ depending on the nature of the data and the economic environment from which they were generated. Common assumptions are that Ψ is a diagonal matrix (heteroskedasticity) or the covariance matrix corresponding to what is known as a first-order autoregressive error (autocorrelation). Whatever be the assumption, the elements of Ψ will usually be unknown and must thus be estimated. Consequently, using $\hat{\Psi}$ in place of Ψ in the the expression for the generalized least squares estimator leads to what is known as *estimated* or a *feasible* generalized least squares estimator $\hat{\beta} = (X'\hat{\Psi}^{-1}X)^{-1}X'\hat{\Psi}^{-1}y$.

2.5.1 Heteroskedasticity

For the general linear statistical model $y = X\beta + \mathbf{e}$, where $E[\mathbf{e}] = 0$ and $E[\mathbf{e}\mathbf{e}'] = \Phi = \sigma^2 \Psi$, *heteroskedasticity* exists when the diagonal elements of Ψ are not identical. Assume that we want to quantify the relationship between food expenditure and income using the following equation:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_t^2 + e \quad \dots (30)$$

In this case cross-section data would consist of data on a sample of households, and y_t and x_t would be food expenditure and income, respectively, of the t th household. It is reasonable to expect that when x is large households observations on y are more likely to deviate farther from the mean function $E[y_t] = \beta_1 + \beta_2 x_t + \beta_3 x_t^2$ than when x is small. For example, we would expect food expenditure for all low-income households to be much the same in the sense that it would be almost totally explained by income. In high-income households, however, food expenditure is likely to depend on many other factors, so that part of expenditure that is not explained by income is more likely to vary considerably from household to household. In other words, the values of the errors e_t are more likely to be small for small x_t and large for large x_t . A disturbance with these properties is called *heteroskedastic*. Conversely, if σ_t^2 is constant for all observations, the disturbance is *homoskedastic*.

2.5.2 Testing for Heteroskedasticity

If heteroskedasticity does not exist, then least squares will yield a best linear unbiased estimator of the coefficients as well as unbiased variance estimates. However, if heteroskedasticity goes undetected, the least squares estimator will not be the best obtainable, and biased variance estimator is likely to lead to misleading inferences. Consequently, it is of interest to test for heteroskedasticity. Two commonly used tests for heteroskedasticity are Goldfeld-Quandt and Breusch-Pagan tests

2.5.2.1 The Goldfeld-Quandt Test

This is suggested for circumstances in which, assuming that heteroskedasticity does exist, it is possible to order the observations according to increasing variance. This test is based on the statistic $(\hat{\sigma}_1^2 / \hat{\sigma}_2^2)$ derived under the assumption that, under H_0 , the sample could be partitioned into two subsets of observations where the error variance is different for each subset, but is constant within a subset. The estimates $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ were computed from separate regressions on each of the subsets.

The alternative hypothesis is: $H_1 : \sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_T^2$. If H_1 is true, then an estimated variance $\hat{\sigma}_2^2$ based on a separate regression on the last half of the observations is likely to be higher than an estimated variance $\hat{\sigma}_1^2$ that is based on a regression on the first half of the observations. To make the test more powerful, omit some of the central observations that contribute little towards making $\hat{\sigma}_2^2$ big or $\hat{\sigma}_1^2$ small. The following are the steps required for implementing the Goldfeld-Quandt test:

- 1) Assuming H_1 is true, order the observations according to increasing error variance.
- 2) Omit r central observations. [The optimum choice of r is not obvious. Large values are likely to increase the power of the test through an increase in the value of the F statistic but decrease the power through a reduction in degrees of freedom. Reasonably satisfactory values are $r = 4$ for $T = 30$ and $r = 10$ for $T = 60$. It is not essential that the two regressions be based on the same number of observations. If the number of observations does differ, then the degrees of freedom and the t statistic must be changed accordingly.]
- 3) Run two separate regressions, one using the first $(T - r)/2$ observations and the other using the last $(T - r)/2$ observations.
- 4) Compare the statistic $\lambda = S_2/S_1$ where S_1 and S_2 are the residual sums of squares from the first and second regressions, respectively. Under the null hypothesis of homoskedasticity, λ has an F-distribution with $[(T - r - 2K)/2, (T - r - 2K)/2]$ degrees of freedom.
- 5) Compare the computed value for λ with a relevant critical value from the F-distribution and accept or reject the null hypothesis accordingly.

2.5.2.2 The Breusch-Pagan Test

This is used for cases in which we assume that variance may be a function of more than one explanatory variable but we do not necessarily wish to impose the multiplicative specification. In other words, we may wish to entertain the alternative hypothesis that the variance is some function (but not necessarily multiplicative) of more than one explanatory variable. A Goldfeld-Quandt test is not possible because it would not be possible to order observations according to increasing variance. Assume that under H_1 .

$$\sigma_t^2 = h(\mathbf{z}_t' \boldsymbol{\alpha}) = h(\alpha_1 + \mathbf{z}_t^* \boldsymbol{\alpha}^*) \quad \dots (31)$$

where h is any function independent of t , $\mathbf{z}' = (1, z_{t2}, z_{t3}, \dots, z_{ts})$ is a vector of observable explanatory variables, and $\boldsymbol{\alpha}' = (\alpha_1, \boldsymbol{\alpha}^*) = (\alpha_1, \alpha_2, \dots, \alpha_s)$ is a vector of unknown coefficients. Under the null hypothesis $\boldsymbol{\alpha}^* = 0$ and the assumption that the e_t are normally distributed, it can be shown that one-half the difference between the total sum of squares and the residual sum of squares from regression

$$\frac{\hat{e}_t^2}{\bar{\sigma}^2} = \mathbf{z}_t' \boldsymbol{\alpha} + v_t \quad \dots (32)$$

is distributed asymptotically as $\chi_{(s-1)}^2$. The $\hat{e}_t = y_t - \mathbf{x}'_t \mathbf{b}$ are the least squares residuals, and

$$\bar{\sigma}^2 = \sum_{t=1}^T \hat{e}_t^2 / T. \quad \dots (33)$$

2.5.3 Autocorrelation

Here we are concerned with the situation where the sample observations y and the control variables X occur at different points in time and, consequently, are known as time-series observations. Such observation could be collected for a single economic unit, such as a firm or a household, or they could be aggregate quantities for a whole region or an economy. For example time-series data on consumption, investment, and income, among other variables are frequently used to estimate macroeconomic relationships, such as consumption and investment functions.

Using the general linear model $y = X\beta + e$ to describe the generation of the data over time, it is common for the relationship to be dynamic in nature. When effects of changes in economic variables are not instantaneous, this can be captured by specifying a model where

- (1) y_t depends on past values of itself,
- (2) X contains current and lagged values of some explanatory variables, and/or
- (3) the equation error e_t depends on the values of previous errors.

The first condition implies that X is stochastic and falls under Distributed Lag Models. The second condition has similar assumptions as those made for least squares estimation earlier. But if there are too many lagged variables, it can be advantageous to place restrictions on the coefficients. The third condition is known as *autocorrelation*. Its existence implies that the total effect of a random error is not instantaneous, but is also felt in future periods, a reasonable assumption for many economic relationships.

There are many possible forms of autocorrelation, and each one leads to a different structure for the error covariance matrix. The most popular form of autocorrelation and one that has proved to be useful in many applications is *the first-order autoregressive process*.

We can write the t th observation on the general linear model as

$$y_t = X_t' \beta + e_t \quad \dots (34)$$

where

$$e_t = \rho e_{t-1} + v_t \quad \dots (35)$$

$\mathbf{x}'_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ is a $(1 \times K)$ vector containing the t th observation on K nonstochastic explanatory variables and β is a $(K \times 1)$ vector of coefficients to be estimated. It is assumed that the current disturbance e_t depends on the error in the previous period, e_{t-1} , and on another error, v_t , which is assumed to have 0 mean and constant variance and to be uncorrelated over time. That is, $E[v_t] = 0$, $E[v_t^2] = \sigma_v^2$, and $E[v_t v_s] = 0$ for $t \neq s$, the assumptions of the general linear statistical model with first-order autoregressive disturbance. ρ and β are unknown and must be estimated.

2.5. 3.1 The Durbin-Watson Test For First-Order Autoregressive Errors

To establish whether or not autocorrelation is likely to exist, the null hypothesis $H_0: \rho = 0$ can be tested against the alternative $H_1: \rho \neq 0$. Rejection of H_0 suggests the presence of autocorrelation. For a finite sample, the Durbin-Watson test, d ($0 < d < 4$), is appropriate. A value of d close to 0 indicates positive autocorrelation ($\rho > 0$), a value close to 4 indicates negative autocorrelation ($\rho < 0$), and a value around 2 suggests no autocorrelation ($\rho = 0$).

3 Application of econometrics in modelling the adoption of agricultural technologies

3.1 Regression Analysis in Adoption Studies

Social scientists have traditionally applied regression analysis as a standard tool in research for several reasons. In the main, regression analysis offers a high explanatory power because of its multivariate nature, easy to interpret the results, gives reasonable estimates even when some of its underlying assumptions are violated, and the Gauss Markov Theorem (discussed in the previous Part) ensures that the estimated parameters have desirable statistical properties. Unfortunately, most researchers rarely pay attention to the underlying assumptions and to the effects of violation of these assumptions in empirical studies thus rendering regression analysis one of the most abused methodologies in social science research. Several problems in social science research involve qualitative dependent variable but not continuous. For example, if one is interested in the decision to adopt or not to adopt an improved variety, the dependent variable is dichotomous; taking a value of 0 or 1. Or, if the interest is to understand why a farmer increases, decreases or does not change his/her

farm size, the dependent variable takes on three different possible options, which is a polytomous case. The use of multiple regression analysis for such qualitative choice decisions faces some serious statistical problems such as (1) the parameter estimates from the regression analysis are invalid and the magnitude of the effects of the explanatory variables on the dependent variable is wrong, (2) the statistical hypothesis testing (t, F, etc) is invalid, and (3) the estimated regression parameter estimates become very sensitive to changes in the levels of the explanatory variables. Let's review the assumptions of multiple regression and see what happens if it is used to investigate dichotomous adoption decisions.

The general multiple regression model is given by:

$$y_i = \sum \hat{\beta}_k X_{ik} + e_i \quad \dots (36)$$

where y_i is the dependent variable, X 's the explanatory variables, and β 's are the parameters to be estimated. e_i is the error term. Given the K parameters to be estimated, it is necessary that we have as many observations as there are parameters (i.e, $N \geq K$). To estimate this model we need to estimate the parameters (i.e, β 's). Suppose that the estimate is $\hat{\beta}$, substituting it into (37) yields an estimate of the error, e_i as:

$$e_i = y_i - \sum \hat{\beta}_k X_{ik} \quad \dots (38)$$

The Ordinary Least Squares (OLS) estimates of β are those that minimize the sum of error squares from $\sum (e_i^2)$ obtained from (2). The main assumptions of the OLS are:

1. The K equations are not linear combinations of one another (i.e, no linear dependency).
2. The e_i is normally distributed with a zero mean and finite variance σ^2 .
3. The expected value (or mean) of the error term is zero, i.e, $E(e_i) = 0$.
4. There does not exist any correlation between the error term and the explanatory variables, i.e, $E(e_i X_{ik}) = 0$.
5. Variance of the error term is constant and not correlated from observation to observation, i.e, "homoskedasticity".

These assumptions constitute what is often referred to as the Gauss Markov Theorem. If these assumptions hold, OLS method is said to be the best linear unbiased estimator of B . This means that, of all available potential unbiased linear estimators of B , OLS will give us parameter estimates (i.e, B) with the least variance. The estimate of the variance σ^2 is given

as: $S_{ek}^2 = \sum e_i^2 / N - k$, where N is the number of observations and k the number of estimated parameters. This statistic is related to t and F distributions and we can get an estimate of the explanatory power of the regression model using the R^2 .

3.2 Linear Probability Model (LPM)

The fundamental problem with the regression model in (35) is that it does not restrict the range of values that the dependent and independent variables, that is, they can take values between 0 and infinity. Clearly, one would expect problems if the dependent variable is dichotomous as in the case of modeling technology adoption decisions where the dependent variable takes a value of 1 if the farmer adopted the technology and 0 if the farmer did not adopt the technology. Under such conditions, using the multiple regression analysis has serious econometric implications for the validity of the results.

From assumption 3 above, the expected value of y (i.e., $E(y)$) for $E(e) = 0$ can be written as:

$$E(y) = \sum \hat{\beta}_k X_{ik} \quad \dots (39)$$

If y_i can only take on a value of 0 or 1, then equation (39) could be interpreted as a Linear Probability Model (LPM). However, that would mean no restrictions on the predicted values of y . That is, y can take on values between negative infinity to positive infinity and consequently interpreting values of the dependent variable outside $[0,1]$ as probability does not make sense.

Other problems of using the LPM are as follows:

Given the dichotomous dependent variable, the error terms can only take on two values:

$$e_i = -\sum \hat{\beta}_k X_{ik} \quad \text{when } y = 0; \text{ and}$$

$$e_i = 1 - \sum \hat{\beta}_k X_{ik} \quad \text{when } y = 1$$

From the above, the expected value of the error term, $E(e_i) = 0$ (as shown below) implying

that the OLS estimates of the parameters (i.e., $\hat{\beta}'s$) are still unbiased estimates of their true values $\beta's$.

$$\begin{aligned} E(e_i) &= p(y=0)(-\sum \hat{\beta}_k X_{ik}) + p(y=1)(1 - \sum \hat{\beta}_k X_{ik}) \\ &= -(1 - p(y=1))p(y=1) + p(y=1)(1 - p(y=1)) \\ &= 0 \end{aligned} \quad \dots (40)$$

However, the variance of e_i , $E(e_i)^2$ is no longer constant but now vary with levels of the explanatory variables (i.e., “heteroskedastic”) as demonstrated in equation (41), which is a violation of a critical assumption of OLS regression.

$$\begin{aligned} v(e_i) &= E(e_i)^2 = p(y=0)(-\sum \hat{\beta}_k X_{ik})^2 + p(y=1)(1 - \sum \hat{\beta}_k X_{ik})^2 \\ &= (1 - p(y=1))[p(y=1)]^2 + p(y=1)[1 - p(y=1)]^2 \\ &= p(y=1)[1 - p(y=1)] \end{aligned}$$

$$= [\sum \hat{\beta}_k X_{ik}] [1 - \sum \hat{\beta}_k X_{ik}] \quad \dots (41)$$

The violation of the OLS assumption (5) above leads to the following serious implications:

1. The estimates of the β 's are no longer valid,
2. All statistical tests (e.g., t, F etc) are invalid, and
3. No statistically valid prediction of adoption probabilities can be made.

Due to its ease of estimation, LPM is still used by some researchers in adoption studies. But, as shown above, results from such studies are statistically invalidated by the failure of a critical underlying assumption of OLS. To solve these problems, it has often recommended to use Weighted Least Squares (WLS). A weight (w_i) can be constructed using the inverse of the standard errors of e_i as demonstrated in equation (42) and subsequently used to weight all the variables by and then use OLS to estimate the parameter estimates (see equation (43)), which will give unbiased estimates and satisfy "homoskedasticity" assumption of OLS give. However, the predicted probabilities of the dependent variable could still be outside of the [0, 1] range due to the linearity assumption suggesting the consideration of non-linear models.

$$w_i = \{1/(-\sum \hat{\beta}_k X_{ik})^2 + p(y=1)(1 - \sum \hat{\beta}_k X_{ik})\}^{1/2} \quad \dots (42)$$

The OLS model can be estimated as:

$$(w_i y_i) = \sum (\hat{\beta}_k w_i X_{ik}) + (w_i e_i) \quad \dots (43)$$

3.3 Non-linear Probability Models

To be able to use the LPM, we will need to transform it by removing the upper constraint of $p=1$ using the ratio $p(y=1)/(1-p(y=1))$. This allows y to take values up positive infinity. That is, there is no more an upper bound. To remove the lower bound would require taking the logarithm of this ratio: i.e., $\log[p(y=1)/(1-p(y=1))]$. This ratio can now take on any real number value between negative infinity to positive infinity. We can relate this ratio linearly to the X 's as in equation 41.

$$\log(p(y=1)/(1-p(y=1))) = \sum \hat{\beta}_k X_{ik} \equiv Z \quad \dots (44)$$

We can use algebraic manipulations to obtain an expression for p . Please note that the natural log of an irrational number has the base of e , where $\log(e^x) = x$ and the anti-log of x is e^x . Using the common notation "exp(.)", equation (44) can be expressed in a logistic function shown in equation (45).

$$p_i = \exp(Z)/(1 + \exp(Z)) \quad \dots (45)$$

The logistic function gives a "sigmoid curve", which is continuous and can take on any real value between 0 and 1. It is nearer to 0 as Z goes to negative infinity, increases with Z , and is nearer to 1 as Z goes to positive infinity. In this way it would be possible to determine the

probability of adoption of an improved technology by farmers and be sure that the estimates of the probability will lie in the range of 0 and 1.

One important curve which approximates the “sigmoid curve” is the “normal curve”. There are several types of curves that can also approximate this "sigmoid curve" shape of the logistic function. These two curves have been demonstrated to be appropriate for examining dichotomous choice decisions due to their having highly desirable properties by allowing the interpretation of the dependent dichotomous variable as a probability. The two types of models that depend on these distributions are:

- (a) the “probit” model given by: $Y_{ik} = F(Z_{ik}) = \int_{-\infty}^{Z_{ik}} (2\pi)^{-\frac{1}{2}} \exp(-S^2 / 2) ds$, which depends on the normal distribution, and
- (b) the “logit” model given by: $P_{ik} F(Z_{ik}) = e^{Z_{ik}} / (1 + e^{Z_{ik}})$ for $Z_{ik} = X_{ik} \beta_k$ and $-\infty < Z_{ik} < \infty, \forall k = 1, 2$, which depends on the logistic distribution.

Note: P_{ik} is the probability that the i^{th} farmer will adopt the given technology or not, X_{ik} is a matrix of the explanatory variables affecting adoption, and β_k are the vector of parameters to be estimated. The probit and logit models are closely related and the choice of which one to use usually depends on the statistical properties of the estimated model results.

To be able to predict the probabilities that any farmer with a given set of characteristics will adopt or reject a technology could be derived from partial derivative of the logistic and normal density functions in the case of the probit and logit models, respectively. Note, however, that these derivatives are dependent not only on the corresponding parameter estimates, but also on the levels of all the variables in Z . This thus requires caution in the interpretation of the partial derivatives.

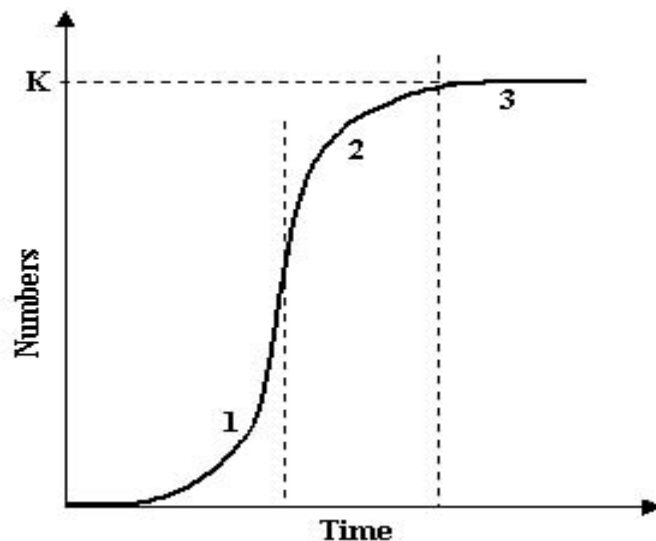


Figure x: Sample sigmoid curve

Source: <http://www.saburchill.com/IBbiology/chapters05/images/14060315.jpg>

Bibliography/Further Readings

- Adesina, A.A., Zinnah, M.M., 1993. Technology characteristics, farmers' perceptions and adoption decisions: A Tobit model application in Sierra Leone. *Agricultural Economics*, 9(4): 291-311.
- Brennan, J.P., and Byerlee, D., 1991. The rate of varietal replacement on farms: measures and empirical results for wheat. *Plant Varieties and Seeds*. 4, 99-106.
- Byerlee and Polanco (1986). Farmers step-wise adoption of technological packages: evidence from the Mexican Altiplano. *American J. Agricultural Economics*, 68: 519-527
- Chiang, A. C., 1984. *Fundamental Methods of Mathematical Economics*. Third Edition. McGraw-Hill, Inc. New York.
- Feder, G., Just, R.E., Zilberman. D., 1985. Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change* 33(2): 225-298.
- Hansen, H.G.P (1992). Inter-regional variation in the speed of adoption of modern cereal cultivars in India. *J. Agric. Economics.*, 43(1):88-95.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepoh and T. Lee. *Introduction to the theory and practice of econometrics*. Second edition. John Wiley and sons Inc. 1985.
- Novshek, W. 1993. *Mathematics for Economists*. Economic Theory, Econometrics and mathematics. Academic Press. New York.
- Zinnah, M.M, Compton, J.L, Adesina, A.A., 1993. Researcher-extension-farmer linkages within the context of the generation, transfer and adoption of improved mangrove swamp rice technology in West Africa. *Quarterly J. International Agriculture*, 32(2): 210-241.