E-NOTE TEMPLATE

COURSE CODE: AEM 507

COURSE TITLE: Econometrics

NUMBER OF UNITS: 2 Units

COURSE DURATION: Two hours per week

COURSE DETAILS:

Course Coordinator: **Dr. S.A. Adewuyi,** *B.Sc., M.Sc., PhD*
Email: samwuyi@yahoo.com
Office Location: **Agric. Econs & Farm Mgt., COLAMRUD**

COURSE CONTENT:

Needs and Econometrics Models, Basic Linear Regression Models, The Method of Least Square; Properties and assumptions of Least Square Estimators; Coefficient of Correlation; Coefficient of Determination Statistical Test of significance of the estimates, Violations of Basic Assumption; Multi-collinearity, Heteroscedascity,auto-correlation; Multiple Regression, lagged and Omitted Variables.

COURSE REQUIREMENTS:

This is a compulsory course for 500 level students in the university. In view of this, students are expected to participate in all the course activities and have minimum of 75% attendance to be able to write the final examination.

READING LIST:

1. Olayemi, J.K. and S.O. Olayode (1981). Elements of Applied Econometrics, Published by CARD & Printed by Leshyraden Nigeria Limited.
2. Koutsoyiannis, A. Theory of Econometrics, The Macmillan Press Ltd, London and Basingstoke

LECTURE NOTES

MEANING AND CONCEPT OF ECONOMETRICS

Econometrics is the scientific application of statistics and mathematics to economic theory.

Steps in Econometrics

1. Selection of variables
2. Construction of mathematical model
3. Data collection- cross-sectional data, time series data and pooled data
4. Estimation of data
5. Data estimation.

Criteria for evaluating econometric data

1. Economic criterion
2. Statistics criterion

3. Econometric criterion

SIMPLE CORRELATION

Correlation is a measure of association that exists between 2 or more variables. Correlation analysis attempts to find out the degree or extent to which variables tend to move together, for instance, in demand theory. Correlation analysis is used to show the degree of relationship between the prize and the quantity.

Types of correlation

1) Positive, negative and zero
2) Simple, partial and multiple
3) Linear and Non-Linear

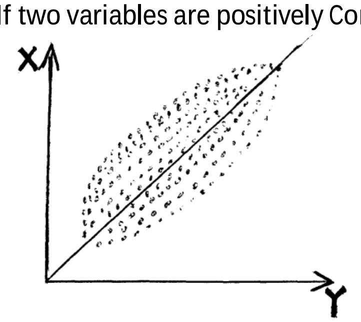
If 'r' is the correlation coefficient between 2 variables X and Y then,

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}}$$

'r' always lies between – 1 and + 1. Correlation may be positive, negative or zero. When r = 1, the 2 variables are said to be positively correlated. When r = – 1, they are said to be negatively correlated. When r = o. there is no correlation between the 2 variables.

Positive Correlation.

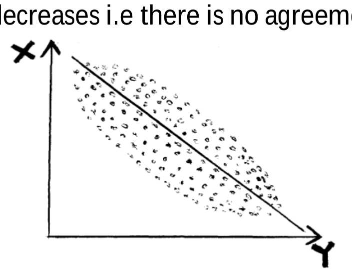
If two variables are positively Correlated, their values tend to rise fall together.



As shown in the graph, when X increases, Y also increases implies that is an agreement between X and Y.

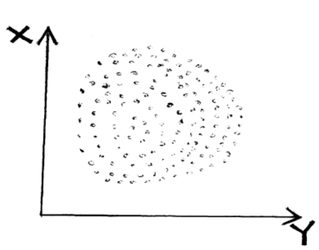
Negative Correlation

Under negative Correlation, 2 variables move in opposite direction. As one increases, the other one decreases i.e there is no agreement between them. This is depicted in the graph below.



Zero Correlation

This occurs when there is no joint movement between two variables. A situation of zero correlation between X and Y is shown below.



Zero Correlation

The diagram is referred to as the scattered diagram.

The correlation coefficient can also be obtained as follows:

1. Pearson (product – moment) correlation coefficient.
(a) Using the standard deviation method.

$r = \frac{Sxy}{Sx\,Sy}$; $Sxy = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N}$ = Standards deviation of X and Y

$Sx = \sqrt{\frac{\sum (X-\bar{X})^2}{N}}$ = Standard deviation of X.

$Sy = \sqrt{\frac{\sum (Y-\bar{Y})^2}{N}}$ = Standards deviation of Y.

∴ The correlation coefficient 'r' is given as

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y}) / N}{\sqrt{\frac{\sum (X-\bar{X})^2}{N}}\sqrt{\frac{\sum (Y-\bar{Y})^2}{N}}}$$

$$= \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2 \sum (Y-\bar{Y})^2}}$$

SIMPLE LINEAR REGRESSION

Regression is the amount of change in value of one variable associated with a unit change in the value o another variable. It shows the dependence of a random variable, on another variable X, which is not necessarily a random variable; an equation, which relates Y to X, is usually called a regression equation. Two variables are involved in regression analysis. They are the dependent or the endogenous variable and the independent or the exogenous variable.

Regression analysis is simply when only two variable (one independent and dependent) are involved. It is linear in the sense that we want to fit a straight line to a set of data involving the 2 variables.

Assumptions of the simple linear model

1. The regression model of the linear in the unknown coefficient
2. The explanatory variables are not perfectly linearly correlated.
3. The error term (e_i) is a random real variable
4. The mean of e_i in any particular period is zero i.e. E (e_i)
5. The variable of e1 is constant in each period, i.e. Var (e_i^2) = S^2
6. The variance has a normal distribution
7. The explanatory variance are measured without any error
8. The error term (e_i) is normally distributed so that U_i ---- N(O, &^{2})
9. Not all independent variables (X_s) are the same; at least one of them is different.
10. The random term of different observation (U_i, U_j) are independently distributed so that Cov (U_i,U_j) = E(U_i, U_j) = 0

Suffice is to say, all these assumptions must hold before we use the ordinary least square model.

<u>**Normal equation and the estimation** of **regression parameters**</u>

Given a regression line $Y_i = a + b_i X_1 + e_i$

$$e_i = Y_i — b_o — b_o X_i$$

Min $\sum e_i^2 = \min \sum (Y_i — b_o — b_o X_i)^2$

$$\frac{\partial \sum e_i^2}{\partial bo} = -2 (Y_i - b_o - b_i X_i) = 0$$

$$\frac{\partial \sum e_i^2}{\partial b_i} = -2 (Y_i - b_o - b_i X_i) = 0$$

$\sum y_i - \sum b_0 - b_i \sum \mathbf{X}_{ii} = 0$...(1)

$\sum_{xi} y_i = b_0 \sum \mathbf{X}_i - b_1 \sum x_i^2 =$ ....(2)

$\sum y_i = \sum b_0 + b_1 \sum x_i$

$\sum_{xi} y_i = b_0 \sum x_i + b_1 \sum x_i^2$

$\sum y_i = nb_0 + b_i \sum x_i$ ....(3)

Equations 3 and 4 are normally least square equations. From equation 3

$b_0 = \sum y_i / n \qquad b_i \sum x_i / n$ ....(4)

$b_0 = \bar{y} - b_i x_i$ ....(5)

Substituting equation (5) in (4)

$\sum x_i y = (y - b_i x) \sum x_i + b_i x_i^2$
$\sum x_i y = (\sum x_i - b_i x \sum x_i + b_i \sum x_i^2$

$$b_i = \frac{\sum x_i y_i - y \sum x_i}{\sum x_i^2 y_i - y \sum x_i}$$

$$= \sum x_i y_i - \sum x_i \sum y_i / n$$

$$\sum x_i = \sum x_i \sum x_i / n$$

$$n \sum x_i y_i - \sum x_i y_i$$

$$n \sum x^2_i - \sum x_i^2$$

$$b_i = \frac{n \sum x_i y_i}{\sum x_i^2}$$

Example

The Table below shows the respective salaries of a father (X) and his oldest Son (Y) of a sample of 12

Salary of father X (1,000)	**5**	**3**	**7**	**4**	**8**	**2**	**10**	**6**	**8**	**7**	**9**	**11**

Salary of son Y (1,000)	**8 6 8 5 9 6 8 5 11 7 8 10**

(a) Find the least square regression line of Y on X
(b) Estimate the salary of the oldest son if the salary of the father is N9,500
Solution:

X	**Y**	**X^2**	**XY**
5	8	25	40
3	6	9	18
7	8	49	56
4	5	16	20
8	9	64	72
2	6	4	12
10	8	100	80
6	5	36	30
8	11	64	88
7	7	49	49
9	8	81	72
11	10	121	110

$\sum X = 80$, $\sum Y = 91$, $\sum XY = 647$, $\sum X2 = 618$

Let the regression line of Y on X be
$Y = a + bX$

$$B = \frac{N\sum Xy - \sum X\sum Y}{N\sum X^2 - (\sum X)^2}$$

$$= \frac{12 \times 647 - 80 \times 91}{12 \times 618 - (80)^2}$$

$$= \frac{7764 - 7280}{7416 - 6400}$$

$$= \frac{484}{1016}$$

$$= 0.4764$$

$a = Y - b_i X$

$= \sum Y - b\sum X$

N N

= 91 = 0.4764 X 80

12 12

= 9.5833 – 3. 176

= 4. 4073

Substituting a and b in the equation

Y = a +b_i X

Y = 4.4073 + 0.4764X

This is the required regression line.

(b). If X = # 9, 500, from the equation

Y = 4.4073 + 0.4764X

= 4.4073 + 0.4764 (9520)

= 453.20

The regression line shows that there is a positive relationship between the salary of the father (X) and that of the oldest son (Y). It further implies that if the salary of the son increases by #1, the father's salary will also increase by #0.48.

MULTIPLE REGRESSION

Regression can be defined as the amount of change in the value of one variable associated with a unit change in the value of another variable. There are two types of variables involved in regression analysis. These are independent and dependent variables. Multiple regressions analysis used involves three or more variables.

Multiple regression analysis is used for testing the relationship between dependent variable, Y and two or more independent variables, Xs, and for prediction. The variable linear regression models can be written as

$Y = b_0 + b_i X_i + b_2 X_2$ ……………………………………………… equation 6

Ordinary least square parameter estimates for equation can be obtained by minimizing the sum of the square residuals

$\sum ei^2 = \sum (Y - \hat{Y})^2 = (Y - b_o + b_1 X_1 + b_2 X_2)^2$ …………………………..equation 7

The normal equations are.

$\sum Y = b_o + b_1 \sum X_1 + b_2 \sum X_2$

$\sum X_1Y = b_o \sum X_1^2 + b_2 \sum X_1X_2$ ……………………………………….equation 8

$\sum X_2Y = b_o \sum X_1^2 + b_1 \sum X_1X_2 + b_2 \sum X_2$

In deviation for the normal equations become

$\sum X_1Y = b_o \sum X_1^2 + b_2 \sum X_1X_2$

$\sum X_2Y = b_o \sum X^2_2 + b_1 \sum \underline{X_1X_2} + b_2 \sum X^2_2$…………………………….equation 9

$\sum X_1^2 = \sum X_1^2 - n(\bar{X})^2$

$\sum X_1Y = \sum X_1Y - n(\bar{Y})^2$

$\sum X_1Y^2 = \sum X_1Y^2 - n \bar{X_1} \bar{Y}$

$\sum X_2^2 = \sum X_2^2 - n(X_2)^2$

$\sum X_2Y = \sum X_2Y - n \bar{X_2} \bar{Y}$

$$\begin{pmatrix} \sum X_1^2 & \sum X_1X_2 \\ \sum X_1X_2 & \sum X_2^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sum X_1Y \\ \sum X_2Y \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sum X_1^2 & \sum X_1X_2 \\ \sum X_1X_2 & \sum X_1^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_1Y \\ \sum X_2Y \end{pmatrix}$$

Original Matrix

$$\begin{pmatrix} \sum X_1^2 & \sum X_1X_2 \\ \sum X_1X_2 & \sum X_2^2 \end{pmatrix}$$

Matrix of Minor

$$\begin{pmatrix} \sum X_2^2 & \sum X_1X_2 \\ \sum X_1X_2 & \sum X_1^2 \end{pmatrix}$$

Cofactor

$$\begin{pmatrix} \sum X_2^2 & - \sum X_1X_2 \\ \sum X_1 X_2 \; + & \sum X_1^2 \end{pmatrix}$$

Transpose

$$\begin{pmatrix} \sum X_2^2 & - \sum X_1X_2 \\ \sum X_1 X_2 & \sum X_1^2 \end{pmatrix}$$

Determinant

$$\sum X_1^2X_2^2 - \sum X_1X_2^{\,2}$$

$$\text{Inverse} = \frac{\text{Transpose}}{\text{Determinant}}$$

$$= \frac{\begin{pmatrix} \sum X_2^2 & - \sum X_1X_2 \\ - \sum X_1 X_2 & \sum X_1^2 \end{pmatrix}}{\sum X_1^2X_2^2 - \sum X_1X_2^{\,2}}$$

$$\begin{matrix} b_1 = \\ b_2 = \end{matrix} \frac{\begin{pmatrix} \sum X_2^2 & - \sum X_1X_2 \\ - \sum X_1 X_2 & \sum X_1^2 \end{pmatrix} \begin{pmatrix} \sum X_1Y \\ \sum X_2Y \end{pmatrix}}{\sum X_1^2 \sum X_2^2 - (\sum X_1X_2)^2}$$

$$b_1 = \frac{\sum X_2^2 \sum X_1^2 Y - \sum X_1^2 X_2 \sum X_1^2 Y}{\sum X_1^2 \sum X_2^2 - (X_1 \sum X_2^2)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 10$$

$$b_2 = \frac{\sum X_2Y \sum X_1^2 - \sum X_1Y \sum X_1X_2}{\sum X_1^2 \sum X_2^2 - (X_1 \sum X_2^2)^2} \quad \ldots\ldots\ldots\ldots 11$$

$$Bo = Y - b_1X_1 - b_2 X_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 12$$

Variance of error term (S^2)

$$S^2 = \frac{\sum e^2}{n-k} = \frac{\sum Y^2 - (\,bo(\sum Y + b_1 \sum X_1 Y + b_2 \sum X_2 Y)}{n-k} \quad ..13$$

Where S^2 = Variance of error
n = Sample Size
K = No of parameters

Variance of coefficients

$$\text{Var}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = S^2 (X^1X)^{-1}$$

$$\text{Var}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = S^2 \begin{pmatrix} \sum X_2^2X & \sum X_1X_2 \\ & \sum X_1X_2 \sum X_1^2 \end{pmatrix}^{-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{equation 14}$$

The variances of b1 and b2 are found in the principal diagonal of the matrix starting with var(bi) in the upper left corner and ending with var(b2) in the lower right corner Therefore

$$\text{Var(bi)} = S^2 \frac{\sum X_2^2}{\sum X_1^2X_2^2 - \sum X_1 X_2^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{equation 15}$$

$$\text{Var(b2)} = S^2 \frac{\sum X_1^2}{\sum X_1^2X_2^2 - (\sum X_1 X_2)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{equation 16}$$

Coefficient of determination (R^2)

This is used t measure how well the estimated regression equation fits the observed data on X and Y. The coefficient measures the "goodness of fit"

$$R^2 = \frac{SSR}{SST} = \frac{b^1X^1Y - nY^2}{Y^1Y - nY^2} \quad \text{.. equation 17}$$

Adjusted Coefficient of determination (R2)

$$R^{-2} = 1 - \left(\frac{(1 - R^2)}{n} \quad \frac{n - 1}{} \right) \quad \overline{} \quad - \quad k$$

PROBLEMS OF SINGLE EQUATION REGRESSION MODEL

1. Autocorrelation: The problems are arises when the assumption that error are independently fails to hold.

$$Cov(e_i, e_j) = 0$$

The first order autocorrelation can be obtained from simple residual error as:

$$et = \beta e_t - 1 + v_t$$

Where:

β = Auto residual error

e_t = Autocorrelation coefficient

v_t =Random autocorrelation error.

TYPES OF AUTOCORRELATION

1. 1st Order
2. 2nd Order
3. Higher Order.

CAUSES OF AUTOCORRELATION

1. Omission of important variable
2. Misspecification of function
3. Nature of economic data

EFFECT OF AUTOCORRELATION

1. Large standard error
2. Universal estimators
3. Ordinary Least Square becomes inefficient.

DEFECTION OF AUTOCORRELATION

1. Use of scattered diagram
2. Trial estimation of autocorrelation equation by testing the significance of the autocorrelation coefficient.
3. Use of Durbin Watson Statistics.

MULTICOLLINEARITY

One of the basic assumptions of ordinary Least Square is that explanatory variables are independent. The violation of this assumption causes multicollinearity.

$E(X_iX_j) = 0$

SOURCES

1. Small Sample Size
2. Nature of economic data
3. Incorporation of lagged independent variable

EFFECTS

1. Large standard error
2. Unbiased value of regression coefficient**.**

METHODS OF DETECTION

1. Klein's rule
2. Beaton and Glamber method
3. Fairrar-Glauber Test
4. Uses of coefficient of determination (R^2)
5. Use of Statistics

HETEROSCEDASTICITY

Heteroscedasticity occurs when the assumption of a mean value and a constant variance of the Ordinary Least Square fail to hold. When the assumption holds there is Heteroscedasticity. The problem of Heteroscedasticity is more common with cross sectional data than with time series data.

CAUSES

1. Use of Heteroscedasticity Samples

CONSEQUENCES

1. Regression parameters estimates are unbiased
2. Large variance of parameter estimate
3. Inefficient predicted estimates of the Least Square.

CORRECTIONS

1. Weighed regression method
2. Use inverse Least Square Method
3. Method of instrumental variance

STEPS IN ECOMOMETRIC STUDY

1. Selection of variable
2. Construction of mathematical model
3. Data collection viz: Cross sectional and pooled data.

CRITERIA OF EVALUATING ECONOMETRIC DATA

1. Economic criterion
2. Statistical criterion
3. Econometric criterion.

FUNCTIONAL FORMS

1. Quadratic function

   $Y = b_o + b_iX_i + b_2X_i^2 + b_3X_3 + b_4X_2^2$

   Used for production and cost equations

2. Hyperbolic functions

   $Y = b_o + b_1 \frac{}{X_1} + b_2 \frac{}{1}$

   It is an isoquant equation.

3. Reciprocal function

   $Y = \frac{1}{b_o + b_iX_i + b_2X_2}$

   It is an isoquant equation

4. Square root function

   $Y = b_1X^{½} – b_2 X_2^{½}$

   Used for production and demand equation

5. Logarithm function

   $\ln Y = \ln b_o X + bi \ln X_i + b_2 \ln X_2$

   It is used for production function.

The regression coefficients reports represent elasticity coefficient s when the sum of the coefficient is > 1, it implies decreasing return to scale.

When the sum of the coefficient = 1, it implies constant return to scale.

6. Semi logarithmic function

   $E^Y = boX_i^{bi} X_2^{b2}$

   Used for cost and supply equation

7. Exponential function

   $Y = e^{bo + biXi + b2\ X2 + e}$

   It is good for cost and supply equation.

8. Linear function

   $Y = b_o + b_iX_i + b_2 X_2 + e$

   Used for demand and supply function.