UNIVERSITY OF AGRICULTURE, ABEOKUTA

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

STS 471 – ECONOMETRIC THEORY

(LECTURE NOTE)

COURSE SYNOPSES

1. Economic Models and Econometrics
2. The Regression Model
3. Heteroscedasticity
4. Serial Correlation
5. Multicollinearity
6. Non linear Regression
7. Errors in Variables
8. Simultaneous Equation Models

INTRODUCTION

Literally interpreted, econometrics means “economic measurement”. Although, measurement is an important part of econometrics, the scope of econometrics is much broader.

Suitable Definitions

1. Quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference (Samuelson, 1954).
2. The social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena (Goldberger, 1964).
3. The integration of economic theory, mathematics, and statistical techniques for the purpose of testing hypotheses about economic phenomena, estimating coefficients of economic relationships and forecasting or predicting future values of economic variables or phenomena.

Why Econometrics is a separate discipline

Econometrics is the integration of economic theory, mathematical economics, economic statistics and mathematical statistics. Each of these branches of knowledge is deficient in the measurement of economic phenomena.

Economic Theory - economic theory makes statements or hypotheses that are mostly qualitative in nature e.g. microeconomic theory states that ceteris paribus, a reduction in the price of a commodity is expected to increase the quantity demanded. No numerical measure of the relationship between price and quantity demanded is provided by economic theory. Econometrics provides the numerical value, and gives empirical content to most economic theory.

Mathematical Economics - the concern of mathematical economics is to express economic theory in mathematical form (equations) without regard to measurability or empirical verification of the theory. Econometrics is interested in the empirical verification of economic theory.

Economic Statistics - economic statistics is mainly concerned with collecting, processing, and presenting economic data in the form of charts and tables. Econometrics uses the collected economic statistics to test economic theories.

Mathematical Statistics - mathematical statistics provides many tools used in econometrics, especially special methods in view of the unique nature of most economic data, namely, that the data are not generated as a result of controlled experiment.

Traditional or Classical Methodology of Econometrics

1. Statement of theory or hypothesis
2. Specification of the mathematical model of the theory
3. Specification of the statistical, or econometric model
4. Obtaining the data
5. Estimation of the parameters of the econometric model
6. Hypothesis testing
7. Forecasting or prediction
8. Using the model for control or policy purposes.

Illustration of Econometric Methodology

Statement of Theory

Men and woman on the average are disposed to increase their consumption as their income increases, but not as much as the increase in their income, i.e. Marginal Propensity to Consume (MPC), the rate of change of consumption for a unit change in income (per #) is greater than zero but less than 1, (John Maynard Keynes).

Specification of the Mathematical Model of consumption

$Y = B_1 + B_2x$, $o < B_2 < 1$

Y = Consumption expenditure.

X = Income, B_1 and B_2 are parameters of the model

Specification of the Econometric Model of consumption

$Y = B_1+B_2x$ exact or deterministic model

$Y = B_1+B_2x + U$ econometric/statistical model

Obtaining data

Year	Y (PCE)	X (GDP)
2005	3081	4620
2006	3240	4803
2007	3407	5140
2008	3566	5323
2009	3708	5487

Estimation of the Econometric Model

$$\hat{y} = -184.08 + .07064x$$

Hypothesis Testing

$B_2 = 0.7064$, positive, less than 1, genuine or chance occurrence?

Is 0.7064 statistically less than 1?

Forecasting or prediction

If model support hypothesis, we use it to predict the future values of y

e.g y_{2011} = -184.0779 + 0.7064 (7269.8)

= 4951. 3167

Use of the Model for control or policy purposes

Fiscal and monetary policy mix

Increase income tax?

Introduce VAT?

Divisions of Econometrics

Theoretical Econometrics

Theoretical econometrics is concerned with the development of appropriate methods for measuring economic relationships specified by econometric models.

Econometric methods may be classified into two groups:

Single - equation technique

Methods applied to one relationship at a time.

Simultaneous equations technique

Methods applied to all the relationships of a model simultaneously.

Applied Econometrics

In applied econometric, the tools of theoretical econometrics are used to study some special fields of economic and business such as the production function, investment functions, demand and supply function, portfolio theory, e.t.c.

Functions of Econometrics

1. To test economic theories or hypotheses
2. To provide numerical estimates of the coefficients of economic relationships
3. Forecasting of economic events.

Roles of an econometrician

1. Remodel an existing economic structure and fit in a new adequate economic model on the new economic structure
2. Estimate the parameters of the econometric model and test their significance
3. Formulate economic policy and take decisions using the parameter so obtained
4. Forecast future values of the variables of the models.

The Regression Model

Multiple Regression (Model with two explanatory variables)

The Model

$Yi = B_0 + B_1x_1 + B_2x_2 + U_i$, structured form

$Yi = b_0+b_1x_1 + b_2x_2 + e_i$, estimated form

Assumptions

1. Randomness of U

   The error (u) is a real random variable

2. Zero mean of U

   The random variable U has a zero mean value for each X_i

   $E(u_i) = 0$

3. Homoscedasticity

   The variance of each u_i is the same for all the X_i values

   $E(U_i^2) = \sigma_u^2 (Cons\tan ts)$

4. Normality of U
   The values of each u_i are normally distributed
   $U_i \rightarrow U(0, \sigma_u^2)$

5. Non autocorrelation or serial independence of the U's

   Covariance of u_i and u_j is zero

   $E(u_i u_j) = 0$ for $i \neq j$

6. Independence of U_i and X_i

   Every disturbance term u_i is independent of the explanatory variables

   $E(U_i X_{1i}) = E(U_iX_{2i}) = 0$

7. No errors of measurement in the explanatory variables

   The explanatory variables are measured without error.

8. No perfect multicollinear explanatory variables

   The explanatory variables are not perfectly linearly correlated

9. Correct aggregation of the macro variables

10. Identifiability of the function

    The relationship being studied is identified.

11. Correct specification of the model

Estimation of parameters

$$\hat{b}_1 = \frac{\sum x_1 y \sum x_2^2 - \sum x_2 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$\hat{b}_2 = \frac{\sum x_2 y \sum x_1^2 - \sum x_1 y \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2$$

Model Diagnosis

1. ***Coefficient of Multiple Determination***

$$R^2_{Y.x1x2} = \frac{\hat{b}_1 \sum x_1 y + \hat{b}_2 \sum x_2 y}{\sum y^2}$$

(a) The value of R^2 lies between 0 and 1.

(b) The higher R^2 the greater the percentage of the variation of Y explained by the regression plane (equation).

(c) The closer R^2 to zero, the worst the fit

Test for Significance of R^2

Test Statistic

$$F_c = \frac{R^2 / K}{\left(1 - R^2\right) / \left(n - k - 1\right)}$$

The Adjusted Coefficient of Determination

The inclusion of additional explanatory variable in the model increases the coefficient of determination, R^2. By introducing a new regressor, we increase the value of the numerator of the expression for R^2, while the denominator remains the same. To correct for this defect, we adjust R^2 by considering the degrees of freedom which deereses as new regressors are introduced in the function. The adjusted R^2 is computed as:

Adjusted $R^2 = 1 - (1 - R^2\ )\frac{n-1}{n-k}$

2. ***Hypothesis Testing***

$Ho: \beta_i = 0$ (Not statistically significant)

H1: at least one β_i is non zero (Statistically significant)

$\alpha = 0.05$

Reject Ho if /tc/ > $t_{\alpha/2,, n-k-1}$

Test Statistic

$$t_c = \frac{\hat{b}_i}{\sqrt{V(\hat{b}_i)}}$$

where:

$$V(\hat{b}_1) = \frac{\hat{\sigma}_u^2 \sum x_2^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$V(\hat{b}_2) = \frac{\sigma_u^2 \sum x_1^2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2}$$

$$V(\hat{b}_0) = \hat{\sigma}_u^2 \left| \frac{1}{n} + \frac{\bar{X}_1^2 \sum x_2^2 + \bar{X}_2^2 \sum x_1^2 - 2\bar{X}_1 \bar{X}_2 \sum x_1 x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1 x_2\right)^2} \right|$$

Where:

Common Variance: $\hat{\sigma}_u^2 = \frac{\sum \ell^2}{n-k-1} = \frac{(1-R^2)\sum y^2}{n-k-1}$

Standard error $\sqrt{V(\hat{b}_i)}$

The significance of the slopes can be verified by using partial correlation coefficients.

Partial Correlation Coefficients – are used to determine the relative importance of the different explanatory variables in a multiple regression. It measures the net correlation between the dependent variable (Y) and one explanatory variable while holding constant the other explanatory variable. The partial correlation coefficients are calculated as shown below:

$$r_{yx1.x2} = \frac{r_{yx1} - r_{yx2} r_{x1x2}}{\sqrt{1 - r_{x1x2}^2}\sqrt{1 - r_{yx2}^2}}$$

$$r_{yx2.x1} = \frac{r_{yx2} - r_{yx1} r_{x1x2}}{\sqrt{1 - r_{x1x2}^2}\sqrt{1 - r_{yx1}^2}}$$

To obtain these coefficients, we first obtain the correlation coefficients as follows:

$$r_{yx1} = \frac{\sum x_1 y}{\sqrt{\sum x_1^2}\sqrt{\sum y^2}}$$

$$r_{yx2} = \frac{\sum x_2 y}{\sqrt{\sum x_2^2}\sqrt{\sum y^2}}$$

$$r_{x1x2} = \frac{\sum x_2 x_1}{\sqrt{\sum x_2^2}\sqrt{\sum x_1^2}}$$

Example

The following table shows the imports, the national income and relative price of imports (all measured in index form) of a certain country:

Year	95	96	97	98	99	00	01	02	03
Imports(Y)	100	106	107	120	110	116	123	133	137
NI(X_1)	100	104	106	111	111	115	120	124	126
RP of imports (X_2)	100	99	110	126	113	103	102	103	98

<u>Intermediate result</u>

$$\bar{Y} = 117, \bar{X}_1 = 113, \bar{X}_2 = 106$$
$$\sum y^2 = 1260.89, \sum x_1^2 = 650, \sum x_2^2 = 648$$
$$\sum x_1 y = 874, \sum x_2 y = -79, \sum x_1 x_2 = -112$$

1. Find the least squares regression equation of Y on X_1 and X_2
2. What percentage of the total variation in imports is explained by both national income and relative price of imports?
3. Test R^2 for statistical significance at the 5% level.
4. Test the statistical significance of the individual coefficients at the 5% level.

Solution

1. $\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \ell$

$$\hat{b}_1 = \frac{(874)(648) - (-79)(-112)}{(650)(648) - (-112)^2} = \frac{557504}{408656} = 1.3642$$

$$\hat{b}_2 = \frac{(-79)(650) - (874)(-112)}{408656} = \frac{46538}{408656} = 0.1139$$

$$\hat{b}_0 = 117 - [(1.3642)(113)] - [(0.1139)(106)] = -49.228$$

Hence, $\hat{y} = -49.228 + 1.3642x_1 + 0.1139x_2$

2. $R^2 = \dfrac{(1.3642)(874) + (0.1139)(-79)}{1260.89} = 0.9385 = 93.85\%$

3. *<u>Test for Significance of R^2</u>*

Test Procedure

$H_0 : \beta_i = 0$

H_1 : at least one β_i is non zero

$\alpha = 0.05$

Reject H_o if $F_c > F_{\alpha}^{K,n-k-1} (5.14)$

Test Statistic

$$\frac{0.9385/2}{(1-.9385)/(9-2-1)} = 45.8$$

Statistical Decision

Reject H_0

Conclusion

R^2 is statistically significant.

4. *Variance and Standard errors of Regression Coefficients*

$$\hat{\sigma}_u = \frac{(1-.9385)(1260.89)}{9-2-1} = 12.9241$$

$$V(\hat{b}_1) = \frac{(12.9241)(648)}{(650)(648)-(-112)^2} = 0.0205$$

$$s(\hat{b}_1) = \sqrt{0.0205} = 0.1432$$

$$V(\hat{b}_2) = \frac{12.9241(650)}{408656} = 0.0206$$

$$s(\hat{b}_2) = \sqrt{0.0206} = 0.1435$$

$$V(\hat{b}_0) = 12.9241\left[\frac{1}{9} + \frac{(113)^2(648) + [(106)^2(650)] - 2(113)(106)(-112)}{408656}\right] = 578.95$$

$$s(\hat{b}_0) = \sqrt{578.95} = 24.06$$

Test Procedure

H_o: $\beta_i = 0$

$H_1 : \beta_i \neq 0$

$\alpha = 0.05$

Reject H_0 if /t_c/ > $t_{\alpha/2}^{n-k-1}(2.447)$

Test Statistic

$$t_c = \frac{\hat{b}_i}{\sqrt{V(\hat{b}_i)}}$$

For $\hat{b}_1 \rightarrow t_c = \frac{1.3642}{0.1432} = 9.53$

For $\hat{b}_2 \rightarrow t_c = \frac{0.1139}{0.1435} = 0.794$

For $\hat{b}_0 \rightarrow t_c = \frac{-49.228}{24.06} = -2.05$

Decision

$\hat{b}_1 \rightarrow \text{Re}\ jectH_0$

$\hat{b}_2 \rightarrow AcceptH_0$

$\hat{b}_0 \rightarrow AcceptH_0$

Conclusion

Only $\hat{b}_1$ is statistically significant.

HETEROSCEDASTICITY

Meaning

The assumption of constant variance is also known as the assumption of Homoscedasticity i.e. the variance of error is assumed to be constant at all level Xi, of the independent variable [E(u_i^2) = σ_u^2. But when this assumption fails to hold, there is problem of Heteroscedasticity. Better still, we say that the disturbances (u_i^2) are heteroscedastic. This means that, rather than remaining constant, the variance of error varies with Xi.

Effects of Heteroscedasticity

1. Regression parameters estimates, b_i, are unbiased.
2. Ordinary least squares formulae for computing the variance of parameter estimates become inapplicable. If applied, it will give a wrong variance for b.
3. The variance of parameter estimates will tend to be very large. This is likely to cause a wrongful acceptance of the null hypothesis that B = 0
4. Predictions based on the least squares estimates will be inefficient.

Testing for Heteroscedasticity

1. Graphical Method - Scatter Diagram. This gives only a visual impression of Heteroscedasticity which cannot be statistically tested. One may plot either the values of residual error or the values of dependent variable, Y, against the values of an independent variable Xi suspected to be affected.

2. Spearman's Rank Correlation Test

The degree of co variation between the values of the error term and the values of an independent variable provides an evidence of Heteroscedasticity. Hence, the correlation coefficient for the two variables, e and x can serve as an index of the degree of Heteroscedasticity. The test is defined as:

$$r_{e.x}^{1} = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

3. The Glejser Test

This test provides an evidence of the presence or absence of Heteroscedasticity, and gives an indication of the pattern of heteroscedastic relationship. The procedure first calls for the regression of the dependent variable on all explanatory variables and the estimation of residual errors. Next the absolute values of residual error are regressed on the actual values of affected explanatory variable.

4. Goldfield Quandit Test

This is a large sample test which statistically evaluates the ratio of the variance of error of two equal sub samples. The existence or otherwise of Heteroscedasticity is mainly determined by hypothesis testing. This test is usually apply when the number of observations (T) is more than twice the number of explanatory variables i.e. $T > 2k$.

Correcting Heteroscedasticity

1. Transforming the values of the variable in such a way that the variance of error becomes constant for all levels of the independent variable. Techniques that may be used for such transformation includes: semi logarithm, double logarithm or reciprocal transformations.
2. Reduce the variables by some measure of size.

Illustrations

Spearman's Rank Correlation Test

Recall data and Regression equation on imports(Y), National income(X_1) and Relative price of imports (X_2),

$Y = 49.228 + 1.342x_1 + .1139x_2$

Substitute actual values of $X_1$and X_2 into regression equation to obtain $\hat{y}$

Y	X_1	X_2	$\hat{y}$	e_i	$/e_i/$	R_{X1}	$R_{/e/}$	d	d^2
100	100	100	98.6	1.4	1.4	1	2	-1	1
106	104	99	103.9	2.1	2.1	2	3	-1	1
107	106	110	107.9	- 6.9	6.9	3	9	-6	36
120	111	126	116.6	3.4	3.4	4.5	6.5	-2	4
110	111	113	115.1	-5.1	5.1	4.5	8	-3.5	12.25
116	115	103	119.4	-3.4	3.4	6	6.5	-0.5	0.25
123	120	102	126.1	-3.1	3.1	7	4	3	9
133	124	103	137.1	1.3	1.3	8	1	7	49
137	126	98	133.8	3.2	3.2	9	5	4	16
									128.5

For $X_1 \rightarrow r_{e.x}^{1} = 1 - \frac{6(128.5)}{9(9^2 - 1)} = -.0708$

Test Procedure

Ho:$e_I = 0$ (There is no Heteroscedasticity)

H_1:$e_i \neq 0$ (There is Heteroscedasticity)

$\alpha = 0.05$

Reject Ho if $/t_c/ > t_{\alpha/2}^{n-2}(2.365)$

Test statistic

$$t_c = \frac{-0.0708\sqrt{9-2}}{\sqrt{1-(-0.0708)^2}} = -0.1878$$

Decision

Do not reject Ho

Conclusion

There is no Heteroscedasticity at 5% level of significance.

For X_2, the same procedure is repeated.

SERIAL CORRELATION OR AUTO CORRELATION

The assumption of serial correlation states that the error term should be independent of one another. If this assumption fails to hold, the result is serial correlation or auto correlation.

Sources

1. Omitted explanatory variables.
2. Mispecification of the mathematical form of the model.
3. Interpolations in the statistical observations.

4. Mispecification of the true random term, u.

i.e random errors caused by flood, strike, fire disaster, e.t.c.

Consequences

1. OLS estimates of parameters remain unbiased.
2. Variance of the random error term, U, may be seriously under estimated.

3. Variance of the parameters estimate tend to be larger than those of other econometric methods.

4. Predictions based on OLS estimate will be inefficient.

Test for Autocorrelation

1. The scatter Diagram method
2. Durbin - Watson method
3. Von Neumann ratio method for large samples
4. Auto correlation coefficient method

The Durbin - Watson method

Test procedure

Ho:e=o

Hi:e # o

Test Statistic.

D=

Decision role 2 regions of rejection.

1 regions of acceptance.
2 regions of inconeclusiveness.

i.e

reject Ho

A if dl dl Reject Ho +
B if d>4-dl Reject Ho
C if du ld <d<4-du Acccept Ho
D if dl l du or 4-du<d<4-dl
Inconclusive.

Example Given dl=82, du=1.75, =1%

Y	x1	x2	et	et-1	$(et-_{t-1})^2$	et^2
100	100	100	1.4	-	-	1.96
106	104	99	2.1	1.4	49	4.41
107	106	110	.9	2.1	1.44	.81
120	111	126	3.4	.9	6.25	11.56
110	111	113	-5.1	3.4	72.25	2601
116	115	103	3.4	5.1	2.89	11.56
123	120	102	3.1	3.4	.09	9.61
133	124	103	1.3	3.1	19.36	1.69
137	126	98	3.2	1.3	1.9	10.24
					104.67	77.85

$$D=\frac{104.67}{77.85}= 1.34$$

MULTICOLLINEARITY

The presence of linear relationship or near linear relationship among explanatory variables.

Cases

1. If explanatory variables are perfectly linearly correlated
2. If explanatory variables are not inter correlated at all
3. Existences of some degree of inter correlation among the explanatory variables.

Reasons

1. Influence of same factors
2. use of lagged values

Consequences

If rxj xj =, estimation of coefficients becomes indeterminate and standard errors of these estimates becomes infinitely large.

If rxjrj = 0, no need to perform a multiple regression analysis and each parameter can be estimated by simple regression of y on corresponding regressor.

If the xs are not perfectly collinear i.e ($0<r_{XIXI,}$li), the effects of collinearity are uncertain.

Test for Multicollinearity

1. Frisch Confluence Analysis.
2. Bunch Map Analysis.
3. Farrer - Glauber Test.

The Farrer - Glauber Tests

Test1

Chi Square test for presence and severity of Multicollinearity.

Test 2

F - test for the location of Multicollinearity.

Test 3

T - test for the pattern of Multicollinearity.

ERRORS IN VARIABLES

Errors in variables refer to the in the case in which the variables in the regression modle include measurement errors. The assumption of absence of errors of measurement in the explanatory variables does not seem plausible in most cases. The x's, like the y's, may well be expected to include errors of measurement.

Consequences of the violation of the assumption of no errors of measurement

The presence of errors of measurement in the variables renders the estimates of the coefficients both biased and inconsistent i.e the value of the slope b1 of the regression line is under estimated, while the value of the constant intercept b0 is over estimated.

Testing for errors in the regressors

There is no formal test for assessing the validity or the violation of assumption of no errors of measurement. Only economic theory and knowledge of how the data were gathered can sometimes give some indication of the seriousness of the problem.

Solutions for the case of errors in variables

1. Inverse least squares
2. The two - group method
3. The three - group method
4. Weighted regression
5. Durbin's ranking method
6. Instrumental variables
7. The maximum likelihood approach

SIMULTANEOUS EQUATION MODELS

So far, we have concerned ourselves with single equation model which assumes one - way causation between the dependent variable y and the independent variable(s) that is y depends on x. Occasions abounds where as y depends on x, x also depends on y so that y = f(x) and x = f(y) in which two way causatation is said to exist. If this exists, which typical of economics relationships a single equation model no longer suffices rather, we must use multi-equation model that would include separate equations in which y and x would appear as endogenous variables (variables determined within the model).

A system that describes the joint dependence of variables is called a system of simultaneous equations. If a system that belongs to a system of equations is presented or modelled by a single equation model, the OLS assumption of independence of x and u is violated and the application of OLS hence yields biased and in consistent estimates.

The bias that arises from the application of OLS to an equation belonging to a system of simultaneous equation is called simultaneous equation bias.

For example,

Suppose we want to estimate the demand for food from the function.

Q = bo +b1p+b2po+b3y+u---(1)

Where Q=qty demanded

P =price of food

Po =price of other commondities

Y =income

1 U =random variable

Application of OLS on the above function requires that cou (u,Q) =0. If this violated, biased and inconsistent estimates of the model parameters will result.

But just as the demand for a commodity is a function of its price, the price of the commodity is also influenced by the quatity demamded of that commodity. Therefore, the above function is only a subset of a system of simultameous equation to which it belongs. There should exist at least one more equation describing the relationship between p and Q, say;P =Co +C1Q+C2W+V.........(ii)

Where W= index of weather condition

V =random variable.

On substituting for Q in equation (ii), we have P=Co + C1(bo +b1p+b2p0+b3y+u)+C2W+V....(iii)

Equation (iii) shows that p depends on u, p in (i)is not truely exogenous.(exogenous variable is one determined outside the model), and hence cov (u,p) # 0. This is a violation of the OLS assumption, which automatically renders application of OLS in equation (i) invalid.

To obtain unbiased and consistent estimates of the parameters in equation (i), the equation must be combined with equation (ii) to form a system of simultaneous equations and parameters are then estimated by any of the simultaneous equation methods of parameter estimation.

Consequences of simultaneous equations

As ealier stated the consequences of appling OLS to a single equation that belongs to a system of equations are that the estimates of parameters are biased and inconsistent.in this section, we shall prove that cov (u,x) #o for a single equation that Y = f(x) that belongs to a system of equations and that consequenetly by estimate is biased.

Assume we have the model

Y =b0+b1x+N.............(i)

X=a+a1y+a2z+v............(2)

Where E(u)= E(v)=0

E (ui vj)=E (vi vj) =0

E (u2)=o

To prove that cov (ux) #o and E(b1) # b1.

Proof.

Substituting for y in (2), we have

$X=a_o+a,(bo+b1x+v)+9_2z+v$

$X=a_0+a_1bo+a_1b1x+9_1u+92z+v$

$X=a_1b_1x=a_o+a_1bo+a_{1u}+a_2z+v$

$X(1-a1b1)=a_o+a_1bo+a_{1u}+9_2z+v$

$X=\underline{a_o+a_1bo + a_2z +a_{1u}+v}$

1-a1b1 1-a1b1 1-a1b1

E cov(ux)=EuEX=E(ux)-o

=E ux=E[

Proof.

Writing eqn (1) in diviation form, we have

Y=b1x+u

Normally equation:

Dividing through by , we have

But

E(b1)=E(b1)=E(

But E(

Hence E(b1) # b1 and so b1 is biased.

We may also obtain the bias as

Bias=(E(b1)-b1)=E(#o.An estimator is unbiased if its bias is zero.

Solution to the Simultameous Equation Bias.

Since the application of OLS to an equation belonging to a system of simultameous equation would yield biased and inconsistent estimates, the obvious solution is to apply other methods which give better estimates of the parameters several such method exist, but the most common are;

A Indirect least square (ILS) otherwise known as the Reduced form method.

B The method of instrumentntal variables (1v).

C Limited information maximum likely hood (LIML).

D Two stage least squares (2SLS).

E The mixed estimation method.

F Three stage least squares (3SLS)

Full information maximum likelihood (FIML)

The first five methods are called single-equation methods because they are applied to one equation of the system at a time. The three –stage least squares and the full information maximum likelihood are called system methods because they are applied to all the equations of the system simultaneously.

SINGLE EQUATIONS AND NON LINEAR MODELS

Quantitative economic relationships cannot always be adequately expressed in linear forms.

Non linear models can be classified into two categories:

1. *Intrinsically linear models*

(a) Models which are non linear in variables only. They can be easily transformed into linear forms and solved by the well established methods.

(b) Models which are non linear in variables as well as in parameters but which can also be transformed into linear models.

2. *Intrinsically non linear models*

Models which are non linear in both variables and parameters and cannot be transformed into linear forms, and therefore cannot be solved by the usual linear regression models.

Approaches to estimating non linear regression models

There are several approaches or algorithms to NLRMs. These include;

(1) Direct search or trial and error

(2) Direct optimization

(3) Iterative linearization

Direct search or trial- and - error or derivative - free method

This approach does not require the use of calculus methods as the other methods do. This approach is not largely in used due to two basic problems: (a) if a NLRM involves several parameters, the method becomes very difficult and computationally expensive (b) there is no guarantee that the final set of parameter values selected will give the absolute minimum error sum of squares.

Direct optimization

This approach differentiate the error sum of squares with respect to each unknown coefficient, or parameter, set the resulting equation to zero, and solve the resulting normal equations simultaneously.

Iterative linearization method

In this method we linearize a non linear equation around some initial values of the parameters. The linearized equation is then estimated by OLS and the initially chosen values are adjusted. These adjusted values are used to re - linearize the model, and again we estimate it by OLS and readjust the estimated values. This process is continued until there is no substantial change in the estimated values from the last couple of iterations. The main technique used in linearizing a non-linear equation is the Taylor series expansion.

Estimating parameters of non linear regression models

Double logarithm Transformation

This is derivable from non linear model of the form

$y = ax^b$...............(1)

where a and b are its parameters.

Transform the model into the linear form and solve as follows:

(1) Take the log to base of e of both sides of model (1) to obtain:

$\log_e a + b\log_e x$......(2)

Observe that eqn (2) is linear in $\log_e$ y and $\log_e$ x. This is where the name double logarithm transformation is coined.

Equation (2) can further be simplified into

$P = \alpha + bq$.........(3)

Where

P =$\log_e {}^y$, q = $\log_e$ x and $\alpha = \log_e a$

(2) Apply the least squares estimation technique on eqn (3) to estimate b and α as:

$$\hat{b} = \frac{n\sum pq - \sum p \sum q}{n\sum q^2 - \left(\sum q\right)^2} \qquad .... (4a)$$

$$\hat{\alpha} = \frac{1}{n} = \left[\sum p - \hat{b}\sum q\right] \qquad .... (4b)$$

(3) $\hat{b}$ can be obtained directly from formula (4a). Substitute the value of a obtained in formula (4b) into the formula $\hat{\alpha} = \log_e a$ as earlier obtained and deduce the value of a there in.

Reciprocal Transformation

A form of this model is:

$$\frac{1}{y} = a + \frac{b}{x} \text{...........(1)}$$

Where the model is reciprocal in its variables x and y a, b are its parameters.

The model can be written as:

W = a + bz .... (2)

Where

$$z = \frac{1}{x} \text{ and } w = \frac{1}{y}$$

Model (2) is linear in w and z

The least squares estimation procedure can now be applied to model (2) to obtain the parameters as:

$$\hat{b} = \frac{n\sum zw - \sum w \sum z}{n\sum z^2 - \left(\sum z\right)^2}$$

$$\hat{a} = \frac{1}{n}\left[\sum w - \hat{b}\sum z\right]$$

REFERENCES

Gujarati, D. N. (2004). *Basic Econometrics (Fourth edition)*. New York: The McGraw – Hill Companies.

Koutsoyiannis, A. (1977). *Theory of Econometrics: an introductory exposition of econometrics methods*. (second edition). New York: Palgrave.

Johnson, J. (1991). *Econometric methods.* (Third edition). Singapare: Mcgraw - Hill Book company.

Omotoso, M. Y. (2000). *Econometrics: A Pratical Approach*. Ibadan: Yosode Book Publishers.

Salvatore, D. (1982). *Schaum Outline of theory and problems of Statistics and Econometrics*. New York: McGraw - Hill Book Company.