COURSE GUIDE

NATIONAL OPEN UNIVERSITY OF NIGERIA

ECO 454
APPLIED ECONOMETRICS II

Course Team Dr Likita J. Ogba (Course Writer/Developer) – UNIJOS
Dr Adesina-Uthman, Ganiyat Adejoke, ACMA, FMNES, FCE (Course Editor) – NOUN
Dr Gadong T. Dalyop (Course Reviewer) – UNIJOS

© 2018 by NOUN Press
National Open University of Nigeria Headquarters
University Village, 91 Cadastral Zone
Nnamdi Azikiwe Expressway, Jabi
Abuja

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng
Re-Printed 2023
ISBN:
All Rights Reserved Printed by ……………..
For
National Open University of Nigeria *Multimedia Technology in Teaching and Learning*

Printed 2018
Re-Printed 2023

ISBN: 978-978-8521-43-3

CONTENTS **PAGE**

INTRODUCTION

Welcome to ECO 454 Applied Econometrics II. The course is available for students in undergraduate Economics. The Course Applied Econometrics II

(ECO454) is a core course which carries two credit units. It is prepared and made available to all the students who are taking Economics; a programme tenable in the School of Arts and Social Sciences. The course provides an opportunity for students to acquire a detailed knowledge and understanding of theory and applications of econometrics in data for policy interpretations. The course is a useful material to in your academic pursuit as well as in your workplace as economists, managers and administrators.

This Course Guide is meant to provide you with the necessary information about the use of data run regression and provide policy interpretations. The course demonstrates the nature of the materials you will be using and how to make the best use of the materials towards ensuring adequate success in your programme as well as the practice of policy analysis. Also included in this course guide are information on how to make use of your time and information on how to tackle the tutor-marked assignment (TMA) questions. There will be tutorial sessions during which your instructional facilitator will take you through your difficult areas and at the same time have meaningful interaction with your fellow learners.

Overall, this module will fill an important niche in the study of applied economics which has been missing on the pathway of Economics. Students will acquire an understanding of the method of practical data estimation and the skills to evaluate and discuss econometric literature.

WHAT YOU WILL LEARN IN THIS COURSE

Applied Econometrics provides you with the opportunity to gain mastery and an in-depth understanding of application of economics quantitatively. If you devote yourself to practice of the software you will become used to quantitative analysis using different types of data. The data used will form part of the practical focus of real application implementation.

COURSE AIM

This course will introduce you to the major aspects of applied econometrics.

This begins with knowing the basic assumptions of econometric variables that will be estimated. The practical estimation of models using real life data and identify deviations from models. The course will give you an opportunity to determine model variables stationarity and the use of other options in estimation of econometric models. In this course you will learn about advanced applications of econometrics and how to use such in policy analysis.

COURSE OBJECTIVES

By the end of this course, you should be able to

- Run a regression using a time series data
- Determine data stationary using Dickey-Fuller method
- Evaluate the identification of an equation model
- Discuss simultaneous equation model
- Evaluate and discuss other methods of model estimation

WORKING THROUGH THE COURSE

To complete this course, you are required to read the study units, read the set text books and read other materials that would be provided to you by the National Open University of Nigeria (NOUN). You will also need to undertake practical exercise using Econometric Views (EViews) software. This requires that you have access to a personal computer, purchase and install EViews for practical. Each unit contains self-assessment exercise; and at certain points during the course, you will be expected to submit assignments. At the end of the course, you will be expected to write a final examination. The course will take you about 12 weeks to complete.

Below are the components of the course. What you should do and how to allocate your time to each unit so as to complete the course successfully and on time.

COURSE MATERIALS

Major components of the course are:

1. Course Guide
2. Study Units
3. Textbooks
4. Assignment Guide

STUDY UNITS

There are twelve units in this course, which should be studied carefully. Such units are as follows:

Module 1

Module 2

Module 3

TEXTBOOKS AND REFERENCES

There are certain textbooks that have been recommended for this course. You should ensure that you read them where you are so directed before attempting the exercise.

ASSIGNMENT

There are many assignments on this course and you are expected to do all of them by following the schedule prescribed for them in terms of when to attempt them and submit same for grading by your tutor. The marks you obtain for these assignments will count towards the final score.

TUTOR-MARKED ASSIGNMENT

In doing the tutor-marked assignment, you are to apply your transfer knowledge and what you have learnt in the contents of the study units. These assignments which are many in number are expected to be turned in to your Tutor for grading. They constitute 30% of the total score for the course.

FINAL WRITTEN EXAMINATION

At the end of the course, you will write the final examination. It will attract the remaining 70%. This makes the total final score to be 100%.

COURSE OVERVIEW

The table below brings together the units and the number of weeks you should take to complete them and the assignment that follow them.

Unit	Title of work	Weekly activity	Assignment (end of Unit)
1	Stages of Applied Econometric Research	1	1
2	Model Estimation	1	1
3	Statistical Test of Significance of OLS Estimates	1	1
4	Confidence interval of Econometric Estimates	1	1
5	Stationarity of Time Series Data	1	1
6	General Forecasting	1	1
7	Forecasting and Seasonal Variation	1	1
8	Identification as an Econometric Problem	1	1
9	Two Stage Least Square (2SLQ)	1	1
10	Simultaneous Equation	1	1
11	Instrumental Variable Method	1	1
12	Full Information Maximum likelihood Method	1	1
13	Three Stage Least Square (3SLS)	1	1

HOW TO GET THE MOST FROM THIS COURSE

In distance learning, the study units replace the lecturer. There is the advantage of reading and working through the course material at the pace that suits the learner best. You are advised to think of it as reading the lecture as against listening to the lecturer. The study units provide exercises for you to do at appropriate periods instead of receiving exercises in the class.

Each unit has common features which are designed, purposely, to facilitate your reading. The first feature being an introduction to the unit, the manner in which each unit is integrated with other units and the entire course. The second feature is a set of learning objectives. These objectives should guide your study. After completing the unit, you should go back and check whether you have achieved the objectives or not. The next feature is self-assessment exercises; study questions, which are found throughout each unit.

The exercises are designed basically to help you recall what you have studied and to assess your learning by yourself. You should do each self-assessment exercise and features are conclusion and summary at the end of each unit. These help you to recall all the main topics discussed in the main content of each unit. These are also tutor-marked assignments at the end of appropriate units. Working on these questions will help you to achieve the objectives of the unit and to prepare for the assignments which you will submit and the final examination.

It should take you a couple of hours to complete a study unit, including the exercises and assignments. Upon completion of the first unit, you are advised to note the length of time it took you, and then use this information to draw up a timetable to guide your study of the remaining units. The margins on either side of each page are meant for you to make notes on main ideas or key points for your usage when revising the course. These features are for your usage to significantly increase your chances of passing the course.

FACILITATORS/TUTORS AND TUTORIALS

There are 13 hours of tutorials provided in support of this course. You will be notified of the dates, times and location of these tutorials, together with the names and phone number of your tutor, as soon as you are allocated a tutorial group.

Your tutor will mark and comment on your assignments; keep a close watch on your progress and on any difficulties you may encounter, as this will be of help to you during the course. You must mail your tutor-marked assignments to your tutor- well before the due date (at least two working days are required). They will be marked by your tutor and returned to you as soon as possible. Do not hesitate to contact your tutor by telephone, e-mail, or discussion board if you need help. The following may be circumstances in which you would find help necessary when:

- You do not understand any part of the study units or the assigned readings.

- You have difficulty with the self-assessment with your tutor's comment on an assignment or with the grading of an assignment.

You should try your best to attend tutorials. This is the only chance to have face-to-face contact with your tutor and to ask questions which are in the course of your study. To gain maximum benefit from course tutorials, prepare your list of questions ahead of time. You will learn a lot from participating in the discussions.

CONCLUSION

The course, Applied Econometrics (ECO454), exposes you to time series data utilization, model estimation and the issues involved in Applied Economics, such as the problem of simultaneous equation models, and how to manage and solve them. On the successful completion of the course, you would have been armed with the materials necessary for efficient and effective management of applications of econometric related matters in any organization, policy institution and the country.

MAIN COURSE

CONTENTS **PAGE**

Module 1

Unit 1: Stages of Applied Econometric Research

Contents

1.1 Introduction

In econometric research there are different stages and one stage of the research leads to another, you need to learn these stages and know them in a chronological order. The stages of econometric research will help to give the basic foundational knowledge of what it takes to carry out applied research. In this unit you will be exposed to the basic assumptions with respect to the parameters to be estimated.

1.2 Learning Outcomes

- Learn the basic stages of econometric research
- Learn the assumptions of econometric models

1.3 Stages of Applied Econometric Research

In econometric research there are four main stages. These stages follow a chronological order, a good knowledge of economic theory will assist in identifying the econometric research structure and the characteristics associated with it, some of these basic stages of applied econometric research include:

(i) Model Formulation: This stage involves expressing economic relationships between the given variables in mathematical form. Here, one needs to determine the dependent variable as well as the explanatory variable(s) which will be included in the model. Also expressed here is a prior theoretical expectation regarding the sign and size of the parameters of the function, as well as the nature of the mathematical form the model will take such that the model is theoretically meaningful and mathematically useful.

Model specification or formulation therefore, presupposes knowledge of economic theory and the familiarity with the particular phenomenon under investigation, the theoretical knowledge allows the researcher to have an idea of the interdependence of the variables under study.

(ii) Model Estimation:

Model estimation entails obtaining numerical estimates (values) of the coefficients of the specified model by means of appropriate econometrics techniques. This gives the model a precise form with appropriate signs of the parameters for easy analysis. In estimating the specified model, the following steps are important.

- Data collection based on the variables included in the model.
- Examining the identification conditions of the model to ensure that the function that is being estimated is the real function in question.
- Examining aggregation problems of the function to avoid biased estimates.
- Ensuring that the explanatory variables are not collinear, the situation which always results in misleading results.
- Appropriate methods should be adopted on the basis of the specified model.

(iii) Evaluation of Model Estimates:

Evaluation entails assessing the results of the calculation in order to test their reliability. The results from the evaluation enable us to judge whether the estimates of the parameters are theoretically meaningful and statistically satisfactory for the econometric research.

(iv) Testing the Forecasting Power of the Estimated Model

Before the estimated model can be put to use, it is necessary to test its forecasting power. This will enable one to be assured of the stability of the estimates in terms of their sensitivity to changes in the size of the model even outside the given sample data within the period.

Self-Assessment Exercise

Outline the basic stages of an econometric research.

1.4 Concepts of Economic and Econometric Models

A model is a simplified representation of a real-world process. That is, it is a prototype of reality, and so describes the way in which variables are interrelated. These models exhibit the power of deductive reasoning in drawing conclusions relevant to economic policy.

Economic model describes the way in which economic variables are interrelated. Such model is built from the various relationships between the given variables.

In examining these concepts Bergstron (1966) defined model as any set of assumption and relationships which approximately describe the behaviour of an economy or a sector of an economy. In this way, an economic model guides economic analysis.

Econometric model, on the other hand, consists of a system of equations which relate observable variables and unobservable random variables using a set of assumptions about the statistical properties of the random variables. In this respect, econometric model is built on the basis of economic theory.

Econometric model differs from economic model in the following ways:

i. For an econometric model, its parameters can be estimated using appropriate econometric techniques.

ii. In formulating econometric model, it is usually necessary to decide the variables to be included or not. Thus, the variables here are selective, depending on the available statistical data.

iii. Because of the specific nature of econometric model, it allows fitting in line of best fit, and this is not possible with economic model.

iv. The formulation of an econometric model involves the introduction of random disturbance term. This will enable random element that are not accounted for to be taken care of in the sample.

Self-Assessment Exercise

Distinguish between an economic model and an econometric model.

1.5 Properties of a Good Econometric Model.

The "goodness" of an econometric model is judged on the basis of some basic fundamental properties that are universal in nature. The following are some of these properties:

i. *Conformity with economic theory*. A good model should agree with the postulate of economic theory. It should describe precisely the economic phenomena to which it relates.

ii. *Accuracy* – the estimate of the co-efficient should be accurate. They should approximate as best as possible the true parameters of the structural model.

iii. *The model should possess explanatory ability*. That is, it should be able to explain the observations of the real world. For example, a model should explain price, demand, supply, exchange rates, market behaviour, and any other practical situation.

iv. *Prediction*. The model should be able to correctly predict future values of the dependent variable. For example, a model should predict with accuracy price, demand, supply, exchange rates, market behaviour, and any other practical situation. This feature often helps strengthen the validity of an econometric model within a given period.

v. *Mathematical form*. The mathematical form of the model should be simple with fewer equations. Such model should represent economic relationships with maximum simplicity.

vi. *Identification*. The equations of the model should be easily identified that is, it must have a unique mathematical form. This means the model should either be exactly identified, overidentified or under identified.

Self-Assessment Exercise

Discuss the properties of a good econometric model.

1.6 Summary

In this unit, you have learned the basic stages of applied econometric research that could be followed to carry out acceptable and standard econometric model estimation. These stages of econometric model research are necessary for effective analysis of any model building. You need to learn these steps well because their violation leads to several econometric problems that we shall study in this course. You have also learned the properties of econometric model estimation.

1.7 References/ Further Reading

Asteriou, D., & Hall, S. G. (2007). *Applied Econometrics: A Modern Approach using Eviews and Microfit*. New York: Macmillan Palgrave.

Berndt, Ernst R. (1991). *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley.

Bridge, J. I. (1971). *Applied Econometrics*. North-Holland, Amsterdam.

Chow, Gregory C. (1983). *Econometric Methods*. New York: McGraw-Hill.

Christ, C. F. (1966). *Econometric Models and Methods*. New York: John Wiley & Sons.

Cramer, J. S. (1969). *Empirical Econometrics*. North-Holland, Amsterdam.

Davidson, James (2000). *Econometric Theory*. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). *Applied Econometrics*. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). *Introductory Econometrics*. New York: Springer-Verlag.

Dutta, M. (1975). *Econometric Methods*. Cincinnati: South-Western Publishing Company.

Greene, William H. (2000). *Econometric Analysis* (4th ed.). Engle-wood cliffs, N. J.: Prentice Hall.

Goldberger, Arthur S. (1998). *Introductory Econometrics*. Harvard University Press.

Gujarati, D. N. (2003). *Basic Econometrics*. New-Delhi: Tata Mc-Graw-Hill Publishing Company Ltd.

Koutsoyiannis, A. (1977). Theory of Econometrics: An Introductory Exposition Econometric Methods Macmillan

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Singapore: Cengage Learning.

1.8 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Outline the basic stages of an econometric research.

   a. Model formulation.

   b. Model estimation,

   c. Evaluation of model estimates.

   d. Testing the forecasting power of the estimated model.

2. Distinguish between an economic model and an econometric model.

   An economic model describes the way in which economic variables are interrelated. Such model is built from the various relationships between the given variables and approximately describe the behaviour of an economy or a sector of an economy. An econometric model, on the other hand, consists of a system of equations which relate observable variables and unobservable random variables using a set of assumptions about the statistical properties of the random variables.

3. Discuss the properties of a good econometric model.

   A good econometric model should have the following properties (Note that you need to discuss them in detail):

   a. Conformity with economic theory.

   b. Accuracy.

   c. The model should possess explanatory ability.

   d. Prediction.

   e. Simple mathematical form.

   f. Identification.

Unit 2: Simple Linear Regression Model

Contents

2.1 Introduction

Linear model shows the relationship between two variables. In this relationship, one variable is depending on the other variable. The model consists of the independent variables and the constant term, with their respective coefficient and we need to estimate the parameters of the model in order to know the magnitude of their relationship. Consider a familiar supply function of the form

$Y = b_0 + b_1X$ …… (i).

This function shows the positive linear relationship between quantity supplied, Y, and price of the commodity, X. The dependent variable in this model is the quantity supplied, denoted by Y, while the independent variable (explanatory variable) is the price, X. This is a two-variable case with two parameters representing the intercept and the slope of the function. This supply-price relationship, Y = f(X) is a one-way causation between the variables Y and X: price is the cause of changes in the quantity supplied, but not the other way around. From the above equation (i), the parameters are b_0 and b_1, and we need to obtain numerical value of these parameters.

The left-hand variable, Y, is variously referred to as the endogenous variable, the regressand, the dependent variable or the explained variable. Similarly, the right-hand variable, X, is variously described as the exogenous variable, the regressor, the independent variable or the explanatory variable.

2.2 Learning Outcomes

At the end of this unit, you should be able to:

- Discuss parameter estimation procedures.
- Discuss the assumptions of the stochastic variable.
- Discuss the assumptions of the explanatory variables.
- Attempt algebraic and software estimation of simple regression.

2.3 Parameter Estimation Strategies

The parameters of this model are to be estimated using ordinary least squares (OLS) method. We shall employ this method for a start due to the following reasons.

i. The computational procedure using this method is easy and straight forward.

ii. The mechanics of the OLS method are simple to understand.

iii. This method always produces satisfactory results.

iv. The parameter estimates using the OLS method are Best, Linear and Unbiased. This makes the estimates to be more accurate compared with the estimates obtained using other methods.

v. The OLS method is an essential component of most econometric techniques.

Note that the model $Y = b_0 + b_1X$ implies an exact relationship between Y and X, that is, all the variation in Y is due to changes in X only, and no other factor(s) are responsible for the change. When this is represented on a graph, the pairs of observations (Y and X) would all lie on a straight line

Ideally, if we gather observations on the quantity actually supplied in the market at various prices and plot them on a diagram, we will notice that they do not really lie on a straight line.

2.3.1 Sources of deviations in parameter model estimation

There are deviations of observations from the line. These deviations are attributable to the following factors:

- *Omission of variable(s)* from the function on grounds that some of these variables may not be known to be relevant.
- *Random behaviour of human beings* – Human reactions at times are unpredictable and may cause deviation from the normal behavioural pattern depicted by the line.
- *Imperfect specification of the mathematical form of the model* – A linear

model, for instance, may mistakenly be formulated as a non-linear model. It is also possible that some equations might have been left out in the model.

- *Error of aggregation* – Usually, in model specification, we use aggregate data in which we add magnitudes relating to individuals whose behaviour differs. The additions and approximations could lead to the existence of errors in econometric models.
- *Error of measurement* – This error arises in the course of data collection, especially in the methods used in the collection of data. Data on the same subject collected from the Central Bank of Nigeria and the National Bureau of Statistics could vary in magnitude and units of measurements. Therefore, when you use different sources, you could get different results.

Self-Assessment Exercise

In econometric analysis, there are several sources of deviations from the actual model estimates. Discuss some of the sources of deviation.

2.3.2 The uses of random variable in models

The inclusion of a random variable, usually denoted by "u", into the econometric function helps in overcoming the above stated sources of errors. The "u" is variously termed the error term, the random disturbance term, or the stochastic term.

This is so called because its introduction into the system disturbs the exact relationship which is assumed to exist between Y and X. Thus, the variations in Y could be explained in terms of explanatory variable X and the random disturbance term U.

That is, $Y = b_0 + b_1X + u_i$ ………………… (ii)

Where Y = variation in Y; $b_0 + b_1X$ = systematic variation, u_i = random variation. Simply put, variation in Y = explained variation plus unexplained variation. Thus, $Y = b_0 + b_1X + u_i$ is the true relationship that connects the variable Y and X and this is our regression model which we need to estimate its parameters using OLS method. To achieve this, we need observations on X, Y and u. However, u is not observed directly like any other variables, thus, the following assumptions hold:

2.4 Assumptions of The Linear Stochastic Regression Model

2.4.1 Assumptions with respect to the random variable

In respect to the random variable, u, the following assumptions apply to any given econometric model that is used in prediction of any economic phenomenon:

(i) u_i is a random variable. This means that the value which u_i takes in any one period depends on chance. Such values may be positive, negative or zero. For this assumption to hold, the omitted variables should be numerous and should change in different directions.

(ii) The mean value of "u" in any particular period is zero. That is, $E(u_i)$ denoted by u is zero. By this assumption, we may express our regression in (ii) above as $Y_i = b_0 + b_1 X$.

(iii) The variance of u_i is constant in each period. That is,

$Var(u_i) = E(u_i)^2 = \delta(u_i)^2 = \delta^2 u$

which is constant. This implies that for all values of X, the "u"s will show the same dispersion about their mean (homoscedastic). Violation of this assumption makes the "u"s heteroscedastic.

(iv) u has a normal distribution. That is, a bell-shaped symmetrical distribution about their zero mean. Thus, $u = N(0, 1)$.

(v) The covariance of u_i and u_j is zero ($Cov_{u_i, u_j} = 0$), for $i \neq j$. This assumes the absence of autocorrelation among the u_i. In this respect, the value of u in one period is not related to its value in another period.

Self-Assessment Exercise

Discuss the assumptions with respect to the random variable.

2.4.2 Assumptions in terms of the relationship between the error term 'u' and the explanatory variables.

The following assumptions also hold when you conduct a regression analysis in terms of the relationship between the explanatory variable and the stochastic variable:

i. u and X do not covary. This means that there is no correlation between the disturbance term and the explanatory variable. Therefore, $Cov_{X,u} = 0$.

ii. The explanatory variables are measured without error. This is because the u absorbs any error of omission in the model.

2.4.3 Assumptions in relation to the explanatory variable

The following assumptions are made.

(i) The explanatory variables are not linearly correlated. That is, there is absence of multicollinearity among the explanatory variables. This means that $Cov_{X_i,X_j} = 0$, for $i \neq j$. (This assumption applies to multiple regression model).

(ii) The explanatory variables are correctly aggregated. It is assumed that the correct procedures for such aggregate explanatory variables are used.

(iii) The coefficients of the relationships to be estimated are assumed to have a unique mathematical form. That is, the variables are easily identified.

(iv) The relationships to be estimated are correctly specified.

Self-Assessment Exercise

Discuss the assumptions of the random variable and the assumptions of u with respect to the other explanatory variables.

2.5 Numerical Estimation of Parameters

2.5.1 Algebraic method

The following procedures are used in finding numerical values of the parameters $\hat{b}_0$ and $\hat{b}_1$.

From the true relationship,

$Y_i = b_0 + b_1 X_i + u$ … (i)

and the estimated relationship

$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i + e_i$ … (ii)

With the residual being

$e_i = Y_i - \hat{Y}_i$ … (iii)

Squaring the residual and summing over n, gives:

$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N}(Y_i - \hat{b}_0 - \hat{b}_1 X_i)^2$ … (iv)

The expression in (iv) is to be minimized with respect to $\hat{b}_0$ and $\hat{b}_1$, respectively.

Thus, $\frac{\partial \sum e_i^2}{\partial \sum \hat{b}_0} = 2\sum_{i=1}^{N}\left[Y_i - \hat{b}_0 - \hat{b}_1 X_i\right](-1) = -2\sum_{i=1}^{N}[Y_i - \hat{b}_0 - \hat{b}_1 X_i]$ … (v)

and $\frac{\partial \sum e_i^2}{\partial \sum \hat{b}_1} = 2X\sum_{i=1}^{N}\left[Y_i - \hat{b}_0 - \hat{b}_1 X_i\right](-1) = -2\sum_{i=1}^{N}[Y_i - \hat{b}_0 - \hat{b}_1 X_i]X_i$… (vi)

Setting each of the partial derivatives to zero, and dividing each term by -2, the OLS estimate of $\hat{b}_0$ and $\hat{b}_1$ could be written in the form:

$\sum_{i=1}^{N} Y_i - N\hat{b}_0 - \hat{b}_1 \sum_{i=1}^{N} X_i = 0$ … (vii) (Note that $N\hat{b}_0 = \sum_{i=1}^{N} \hat{b}_0$)

$\sum_{i=1}^{N} X_i Y_i - \hat{b}_0 \sum_{i=1}^{N} X_i - \hat{b}_1 \sum_{i=1}^{N} X_i^2 = 0$ … (viii)

From which,

$\sum_{i=1}^{N} Y_i = N\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{N} X_i$ … (ix)

$\sum_{i=1}^{N} X_i Y_i = \hat{b}_0 \sum_{i=1}^{N} X_i + \hat{b}_1 \sum_{i=1}^{N} X_i^2$ … (x)

The two equations (ix) and (x) are the normal equation of the regression model.

Using Crammer's rule, the values of the parameters $\hat{b}_0$ and $\hat{b}_1$ are respectively: $\hat{b}_0 =$

$$\frac{\sum X^2 \sum Y - \sum X \sum XY}{\mathrm{N}\sum X^2 - (\sum \mathrm{X})^2} \qquad \ldots(xi)$$

$$\hat{b}_1 = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} \qquad \ldots(xii)$$

Using lower case letters (i.e., deviation of the observations from their means). It can be shown that:

$$\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X} \qquad \ldots(xiii)$$

$$\hat{b}_1 = \frac{\sum xy}{\sum x^2} \qquad \ldots(xiv)$$

2.5.2 Quantitative method illustration

Example 1

Given the following data on the supply of commodity W, find the estimated supply function (Table 1).

Table 1

No	Y_i (Quantity)	X_i (Price)
1	64	8
2	68	10
3	44	6
4	48	9
5	50	6
6	65	10
7	45	7
8	56	8

The given expressions for $\hat{b}_0$ and $\hat{b}_1$ in equations (xi) and (xii) as well as those of (xiii) and (xiv) lead us to reproduce the above Table as seen in Table 2.

Table 2: Worksheet for the estimation of the supply function for W.

	Y	X	X^2	XY	$y=$ $Y-\bar{Y}$	$x=$ $X-\bar{X}$	x^2	xy	$\hat{Y}$	e	e^2
1	64	8	64	572	9	0	0	0	55	9	81
2	68	10	100	680	13	2	4	26	64	4	16
3	44	6	36	264	-11	-2	4	22	46	-2	4
4	48	9	81	432	-7	1	1	-7	59	-11.5	132.25
5	50	6	36	300	-5	-2	4	10	46	4	16
6	65	10	100	650	10	2	4	20	64	4	16
7	45	7	49	315	-10	-1	1	10	50.5	-5.5	32.25
8	56	8	64	448	1	0	0	0	55	1	1
n= 8	∑Y= 440	∑X= 64	$\sum X^2$= 530	∑XY= 3601	$\sum y$ =0	$\sum x$ =0	$\sum x^2$ =18	$\sum xy$ =81		∑e=0	$\sum e^2$=28 1.5

Where e = residual, $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i$

From the Table 2,

$\bar{Y} = \frac{\sum Y}{n} = 55, \bar{X} = \frac{\sum X}{n} = 8$

Therefore, using uppercase letters,

$$\hat{b}_0 = \frac{440\ (530)-(64)\ (3607)}{8(530)-(64)^2} = 19$$

$$\hat{b}_1 = \frac{8\ (3601)-(64)\ (440)}{8(530)-(64)^2} = 4.5$$

Similarly, using lowercase letters,

$$\hat{b}_1 = \frac{\sum xy}{\sum x^2} = \frac{81}{18} = 4.5$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} = 55 - 4.5(8) = 19$$

2.6 EViews Software Applications

The linear regression can be solved using the software package through the following steps. Create an excel worksheet and type in the data for quantity (QUA) and price (PRI). There are a number of methods for creating a new workfile in EViews and inputting data.

Method 1: Copy the data on excel worksheet, open your EViews software, right click and paste the data onto EViews. Select the number of Header lines and the Header type under "Column headers" and the Data type for each column in the preview under "Column info" on the dialogue box that appears. Also click on the check box beside "Read series by row (transpose incoming data)" if the titles of the variables in the data are presented in rows and not columns. When satisfied with the selections made, click "Next" at the bottom of the dialogue box. On the next dialogue box that appears, choose from the dropdown list under "Basic structure" to select the type of data. (Note that EViews normally selects the appropriate type of data automatically). You may also click on "Rename Series" under "Import options" to rename the data. Click "Finish" when satisfied with the selections made. You may also click "Back" to return to the previous dialogue box.

Method 2: Go to the "File" tab in EViews and select "Import", then "Import from file" to select the file to be imported. On the first dialogue box that appears, choose the Predefined range (to select the Excel Worksheet to be analysed) or Custom range under "Cell range" (to select the Excel Worksheet and data items in the sheet to be included for analysis), then click "Next" at the bottom of the dialogue box. Follow the same procedure for entering data in the subsequent dialogue boxes as in the case of the Method 1.

Having prepared the data, go to "Quick" tab on the tool bar and scroll to "Estimate Equation" and click on it. A dialogue box opens where you can type the labels for quantity (QUA) and price (PRI), separated by a single space using the space bar on your keyboard, under "Equations specification". You can then select the method of analysis under "Estimation settings" and click OK at the bottom of the dialogue box. The output is as follows:

Dependent Variable: QUA
Method: Least Squares
Date: 04/19/15 Time: 16:28
Sample: 2000 2007
Included observations: 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.00000	13.14075	1.445884	0.1983
PRI	4.500000	1.614460	2.787309	0.0317

R-squared	0.564241	Mean dependent var	55.00000
Adjusted R-squared	0.491615	S.D. dependent var	9.606545
S.E. of regression	6.849574	Akaike info criterion	6.898568
Sum squared resid	281.5000	Schwarz criterion	6.918428
Log likelihood	-25.59427	Hannan-Quinn criter.	6.764618
F-statistic	7.769094	Durbin-Watson stat	1.722913
Prob(F-statistic)	0.031690		

The software estimated model for the data can be presented in a line form as follows:

$$QUA = 19 + 4.5PRI$$

QUA represents the quantity, while PRI represents the price. This is similar to the one computed using the manual method in the regression.

Note:

(i) Only one of the two methods is to be used, and each gives the same result.

(ii) Unless specified, one is free to use any of the methods.

From the values of $\hat{b}_0$ and $\hat{b}_1$, the estimated regression line or equation is got by substituting these values into $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i$ and this gives $Y = 19 + 4.5X$. Thus, given the values of X_i (i = 1, 2, …, N), the estimated values of Y can be obtained using the regression equation.

From the estimated regression line, one can estimate **price elasticity.** Recall the estimated regression equation, $\hat{Y} = \hat{b}_0 + \hat{b}_1 X_i$, this is also the equation of the line with intercept $\hat{b}_0$ and slope $\hat{b}_1$. Note that $\hat{b}_1 = \partial Y / \partial X$. Therefore, price elasticity is $\frac{\partial Y}{\partial X}\frac{\bar{X}}{\bar{Y}} = \hat{b}_1 \frac{\bar{X}}{\bar{Y}}$. Taking the mean of X_i and Y_i, we have average elasticity, $ep = \hat{b}_1 \frac{\bar{X}}{\bar{Y}}$. Therefore, $ep = 4.5 \times \frac{8}{55} = 0.65$.

Self-Assessment Exercise

1. Outline the basic assumptions of linear regression.
2. Given the following data estimate the multiple regression equation using Realgdp as dependent variable and Privcons or Govcons as independent variable.

Realgdp	Privcons	Govcons
31546.8	27463.8	6473.2
205222.1	28574.9	7584.3
199685.3	30411.4	8406.3
185598.1	35215.1	8889.2
183563	42858.7	8459.2
201036.3	49302.9	9363.2
205971.4	51537.5	9424.5
204806.5	75981.1	8064.4
219875.6	106678.6	11311
236729.6	128186.2	12443.7
267550	177234.6	13977.4
265379.1	206813.5	15904.8
271365.5	373526.7	33115.1
274833.3	502775.2	46796.5
275450.6	610340.2	169669.2
281407.4	1387445.6	242737.5
293745.4	2124271.4	280384.4
302022.5	2091069.1	377779
310890.1	2371328.2	393547.2

Estimate the regression equation and interpret your findings.

2.7 Summary

In this unit you have learned the basic stages of applied econometric research and the assumptions of econometric model estimation. You need to learn these assumptions well because their violation leads to several econometric problems that we shall study in this course such as simultaneous equation bias. You have also learned how to conduct an econometric model estimation.

2.8 References/ Further Reading

Asteriou, D., & Hall, S. G. (2007). Applied Econometrics: A Modern Approach using Eviews and Microfit. New York: Macmillan Palgrave.

Berndt, Ernst R. (1991). The Practice of Econometrics: Classic and contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics. North-Holland: Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods. Cincinnati: South-Western Publishing Company.

Greene, William H. (2000). Econometric Analysis (4^{th} ed.). Engle-wood cliffs, N. J., Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Gujarati, D. N. (2003). Basic Econometrics. New-Delhi: Tata Mc-Graw –Hill Publishing Company Ltd.

Koutsoyiannis, A. (1977). Theory of Econometrics: An Introductory Exposition. Econometric Methods Macmillan

Wooldridge, J. M. (2009) Introductory Econometrics: A Modern Approach (4^{th} ed.). Singapore: Cengage Learning

2.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. In econometric analysis, there are several sources of deviations from the actual model estimates. Discuss some of the sources of deviation.

   The sources of deviation from the actual model estimates include (Note that you need to discuss them in detail):

   a. Omission of variable(s).

   b. Random behaviour of human beings.

   c. Imperfect specification of the mathematical form of the model.

   d. Error of aggregation.

   e. Error of measurement.

2. Discuss the assumptions with respect to the random variable.

   The assumptions with respect to the random variable include (Note that you need to discuss them in detail):

   a. u_i is a random variable.

   b. The mean value of "u" in any period is zero.

   c. The variance of u_i is constant in each period.

   d. u has a normal distribution.

   e. The covariance of u_i and u_j is zero ($Cov_{u_i,u_j} = 0$), for $i \neq j$.

3. Discuss the assumptions of the random variable and the assumptions of u with respect to the other explanatory variables.

   The assumptions with respect to the random variable include (Note that you need to discuss them in detail):

a. u_i is a random variable.

b. The mean value of "u" in any period is zero.

c. The variance of u_i is constant in each period.

d. u has a normal distribution.

e. The covariance of u_i and u_j is zero ($Cov_{u_i,u_j} = 0$), for $i \neq j$.

With respect to the relationship between u and other explanatory variables,

f. u and X do not covary.

g. The explanatory variables are measured without error.

4. Outline the basic assumptions of linear regression.

The basic assumptions of linear regression include:

a. u_i is a random variable.

b. The mean value of "u" in any period is zero.

c. The variance of u_i is constant in each period.

d. u has a normal distribution.

e. The covariance of u_i and u_j is zero ($Cov_{u_i,u_j} = 0$), for $i \neq j$.

With respect to the relationship between u and other explanatory variables,

f. u and X do not covary.

g. The explanatory variables are measured without error.

With respect to the relationship between the explanatory variables,

h. The explanatory variables are not linearly correlated.

i. The explanatory variables are correctly aggregated.

j. The coefficients of the relationships to be estimated are assumed to have a unique mathematical form.

k. The relationships to be estimated are correctly specified.

Unit 3 Statistical Tests of Significance of The OLS Estimates

Contents

3.1 Introduction

After the estimation of the model in the previous section, we need to test the explanatory power or goodness of fit of the model, as well as the statistical reliability/significance at a given level in respect of $\hat{b}_i$. There are a number of tests that are used to measure and evaluate the econometric efficacy of the estimated equation. In this unit you will be taken through the rudiments of test of statistical significance so as to be able to evaluate and use models.

3.2 Learning Outcomes

At the end of this unit, you should be able to conduct an:

- R^2 test
- *Z* test
- *T* test

We need to test the explanatory power or goodness of fit of the model, as well as the statistical reliability/significance at a given level in respect of $\hat{b}_i$. This is achieved using the following tests:

i. Coefficient of determination, R^2
ii. Standard error test
iii. *Z* and *t*-statistics (tests)

3.3 R^2 and the Regression Line

The coefficient of determination, R^2, is used in determining goodness of fit of the regression line obtained using the OLS method. That is, it is used in testing the explanatory power of the linear regression of Y on X. Thus, in order to determine the degree to which the explanatory variable is able to explain the variation in the dependent variable Y, the R^2 provides a useful guide.

If we measure the dispersion of observations around the regression line some may be closer to the line while others may be far away from it. Our argument is that the **closer**

these observed values are to the line, the better the goodness of fit. On the basis of this, we may turn out to state that a change in the dependent variable, *Y*, is explained by changes in the explanatory (independent) variable *X*. To know precisely the extent to which the total variation in *Y* is explained by the independent variable, *X*, we compute the value of R^2 as the ratio of explained variation to total variation. That is, R^2 = Explained variation/Total variation = $\frac{\sum \hat{y}^2}{\sum y^2}$.

However, the coefficient of determination, R^2 determines the proportion of the variation in Y which is explained by variation in *X*. If, for instance, $R^2 = 0.75$, it means that 75% of the variation in *Y* is due to the variation in *X*, while 25% of the variation is explained by the disturbance term, *u*. Thus, the regression line gives a good fit to the observed data.

If, however, $R^2 = 0.45$, it means that only 45% of the variation in *Y* is as a result of variation in *X*, while 55% of the variation is due to the disturbance term. This is a poor indication and the regression line does not give a good fit to the observed data. Generally, if R^2 is 0.5 and above, it shows a good fit while a value of R^2 less than 0.5 shows a bad fit.

Note that the R^2 is much of relevance when the estimated model is used for forecasting. Note also that the value of $R^2 = 0$ signifies that the independent variable cannot explain any change in the dependent variable, hence variation in the independent variable has no effect on the dependent variable.

Self-Assessment Exercise

1. Carefully discuss the econometric significance of the R^2.
2. Attempt a simple regression analysis and discuss the explanatory power of R^2.

3.4 Standard Error Test of the Statistical Significance of $\hat{b}_i$

To use this test, it is important to know the mean and variance of the parameter estimates $\hat{b}_0$ and $\hat{b}_1$. It has been established that the mean and variance of $\hat{b}_0$ and $\hat{b}_1$ are, respectively,

$$E(\hat{b}_0) = b_0; Var\ \hat{b}_0 = E(\hat{b}_0 - b_0)^2 = \frac{\sigma_u^2 \Sigma X_i^2}{n \Sigma x_i^2}$$

$$E(\hat{b}_1) = b_1; Var\ \hat{b}_1 = E(\hat{b}_1 - b_1)^2 = \frac{\sigma_u^2}{\Sigma x_i^2}$$

The standard error test enables us to determine the degree of confidence in the validity of the estimate, that is, from the test, we are able to know whether the estimates $\hat{b}_0$ and $\hat{b}_1$ are significantly different from zero. The test is mainly useful when the purpose of the research is the explanatory (analysis) of economic phenomena and the estimation of reliable values.

We formally start by stating the null hypothesis,

$H_0: \hat{b}_i = 0,$

against the alternative hypothesis,

$H_1: \hat{b}_i \neq 0$

The standard error for the parameter estimates $\hat{b}_0$ and $\hat{b}_1$ are respectively computed as shown:

i. $$S(\hat{b}_0) = \sqrt{Var(\hat{b}_0)} = \sqrt{\frac{\sigma_u^2 \Sigma X_i^2}{n \Sigma x_i^2}}$$

But $\sigma_u^2 = \frac{\Sigma e^2}{n-2}$

$$S(\hat{b}_0) = \sqrt{\frac{(\Sigma e^2) \Sigma X^2}{(n-2) n \Sigma x_i^2}}$$

ii. $$S(\hat{b}_1) = \sqrt{Var(\hat{b}_1)} = \sqrt{\frac{\sigma_u^2}{\Sigma x_i^2}} = \sqrt{\frac{\Sigma e^2}{(n-2) \Sigma x_i^2}}$$

When the numerical values for the $S(\hat{b}_0)$ and $S(\hat{b}_1)$ are each compared with the numerical values of $\hat{b}_0$ and $\hat{b}_1$, the following decision rules apply.

(i) If $S(\hat{b}_i) < \hat{b}_i/2$, we reject the null hypothesis and conclude that $\hat{b}_i$ is statistically significant.

(ii) If on the other hand, $S(\hat{b}_i) > \hat{b}_i/2$ we accept the null hypothesis that the true population parameter $\hat{b}_i = 0$. This concludes that the estimate is not statistically significant. Therefore, change in *X* has no effect on the values of *Y*.

The acceptance of the null hypothesis has economic implication. Thus, the acceptance of the null hypothesis, $\hat{b}_i = 0$, implies that the explanatory variable to which this estimate relates does not influence the dependent variable *Y*, and should not be included in the function. This situation renders the relationship between *Y* and *X*; hence the regression equation is parallel to axis of the explanatory variable *X* in the case, $Y_i = b_0 + 0X_i$ or $Y_i = b_0$. The zero slope indicates that no relationship exists between *Y* and *X*.

Similarly, if the null hypothesis $\hat{b}_0 = 0$ is accepted, on the basis that $S(\hat{b}_0) > ½\hat{b}_0$, it implies that the intercept of this regression line is zero. Therefore, the line passes through the origin. In this case, the relationship between Y and X will be $Y_i = 0 + b_1X_i$, or $Y_i = b_1X_i$. The two situations are diagrammatically presented as shown:

Example II

Refer to the example 1 on the regression analysis, the R^2 and the standard errors, $S(\hat{b}_0)$ and $S(\hat{b}_1)$ are as computed:

From the computed values of the Table.

$\sum xy = 81, \sum x^2 = 18, \sum y^2 = 646, \sum e^2 = 281.5, n = 8, K = 2$; where *K* = degrees of freedom.

$$R^2 = \frac{\sum(xy)^2}{\sum x^2 \sum y^2} = \frac{6561}{11628} = 0.56$$

Hence, *X* is a fairly good predicator of *Y*; about 56% of the variation in *Y* is explained by variation in the explanatory variable *X*, while 44% is due to disturbance term u. Thus, the regression line fairly gives a good fit to the observed data. On the statistical significance of the parameter estimate we compute:

$$S(\hat{b}_0) = \sqrt{\frac{(\sum e^2)\sum X^2}{(n-2)n\sum x_i^2}}$$

$$S(\hat{b}_0) = \sqrt{\frac{281.5(530)}{6(8)(18)}} = 13.14$$

$$S(\hat{b}_1) = \sqrt{\frac{\sum e^2}{(n-2)\sum x_i^2}}$$

$$S(\hat{b}_1) = \sqrt{\frac{281.5}{6(18)}} = 1.61$$

Thus, the result of our regression may be presented formally as:

$\hat{Y}$ = 19 + 4.5X

(13.14) (1.61), R^2 = 0.56, n = 8

From the above estimated result,

$\hat{b}_0 = 19$; $\hat{b}_1 = 4.5$; $S(\hat{b}_0) = 13.14$; $S(\hat{b}_1) = 1.61$

Therefore, $S(\hat{b}_0) > ½\hat{b}_0$ and we accept the null hypothesis that $H_0\!: \hat{b}_0 = 0$. This shows that $\hat{b}_0$ is not statistically significant. Similarly, $S(\hat{b}_1) < ½\hat{b}_1$ and we reject the null hypothesis. Thus, the estimate $\hat{b}_1$ is statistically significant. Note: the acceptance and meaningful interpretation of any econometric result entails a combination of high R^2 and low standard errors.

Self-Assessment Exercise

Given an estimated model,

$\hat{Y}$ = 27 + 14X

(13.44) (8.61), R^2 = 0.63, n = 8

Test the reliability of the estimates $\hat{b}_0$ and $\hat{b}_1$, respectively using the standard error test.

3.5 *Z* Test of the Statistical Significance of $\hat{b}_i$

The *z* test is employed when the sample size is sufficiently large (i.e., n ≥ 30). It could be applied whether the population variance is known or not. The z test is applied using the formula, $z_i^* = \frac{\hat{b}_i}{S(\hat{b}_i)}$ where the z_i^* is the calculated z_i, which is compared with the *z* table (theoretical value of z) at a given level of significance, say 5%. If $-z<z^*<+z$ at 0.025, we accept the null hypothesis that $H_0: \hat{b}_i = 0$; and conclude that the estimate is not statistically significant. If, however, $z^*>z$, then we accept that $H_1: \hat{b}_i \neq 0$ and the estimate is statistically significant.

Consider an estimated function from a sample of 50 observations in the form

$\hat{Y} = 15 + 6.8X$

(2.8) (1.05)

To conduct a z test for the estimates $\hat{b}_0$ and $\hat{b}_1$ at 5% level of significance, we proceed as follows: $H_0: \hat{b}_i = 0$ (Null hypothesis)

$H_1: \hat{b}_i \neq 0$ (Alternative hypothesis)

$z_0^* = \frac{\hat{b}_0}{s(\hat{b}_0)} = \frac{15}{2.8} = 5.36$ (for $\hat{b}_0$)

$z_1^* = \frac{\hat{b}_1}{s(\hat{b}_1)} = \frac{6.8}{1.05} = 6.48$ (for $\hat{b}_1$)

z table at 5% level of significance = 1.96

For $\hat{b}_0$, $z_0^* > Z$ Table; *for* $\hat{b}_1$, $z_1^* > Z$ Table

Therefore, the null hypothesis is rejected and we conclude that estimates $\hat{b}_0$ and $\hat{b}_1$ are statistically significant at 0.05 level.

3.6 *T* Test of the Statistical Significance of $\hat{b}_i$

The student t-test is used when the sample size is small (i.e., n < 30) provided that the population of the parameter follows a normal distribution. With this in view, and taking degrees of freedom (d.f.) into consideration, we need to compare this with the theoretical t, at given level of significance, say 5%.

Our null and alternative hypotheses are respectively formulated thus

$H_0: \hat{b}_i = 0$

$H_1: \hat{b}_i \neq 0$

Following a normal distribution, our t is computed as follows:

$t_i^* = \frac{\hat{b}_i}{s(\hat{b}_i)}$

As stated previously, the empirical t (t) value is compared with the table t with n - k degrees of freedom, given a 5% level of significance.

Decision rule: if $-t^*_{0.025} < t < t_{0.025}$ (with n - k degree of freedom), we accept the null hypothesis, and conclude that our estimate $\hat{b}_1$is not statistically significant at 0.05 level of significance.

If, however, $t^* > +t_{0.025}$, we reject the null hypothesis, and accept the alternative hypothesis. This concludes that the estimate $\hat{b}_1$ is statistically significant.

Example

Given a sample size of n=18, the model was estimated to be:

$\hat{Y}$ = 21 + 0.75 X

(10.2) (1.4)

We wish to test the reliability of the estimates $\hat{b}_0$ and $\hat{b}_1$, respectively. From the estimated model,

$\hat{b}_0 = 21; S(\hat{b}_0) = 10.2; \hat{b}_1 = 0.75; S(\hat{b}_1) = 1.4$

Therefore, $t_0^* = \frac{\hat{b}_0}{S(\hat{b}_0)} = \frac{21}{10.2} = 2.06$ (for $\hat{b}_0$)

$t_1^* = \frac{\hat{b}_1}{S(\hat{b}_1)} = \frac{0.75}{1.4} = 0.54$ (for $\hat{b}_1$)

The critical values of t for (n – k) or 18 – 2 = 16 d.f. are:

$-t_{0.025} = -2.120$ and $+t_{0.025} = 2.120$.

For the estimate $\hat{b}_0$, we can see that $t^* < t_{0.025}$. We, therefore, accept the null hypothesis and conclude that the estimate $\hat{b}_0$ is not statistically significant at 5% significance level.

In the case of the estimate $\hat{b}_1$ since $t^* < t_{0.025}$, we also accept the null hypothesis and conclude that $\hat{b}_1$ is not statistically significant at 5% significance level.

Self-Assessment Exercise

Given a sample size of n=20, the model was estimated to be:

$\hat{Y} = 48 + 0.55X$

(12.2) (0.4)

Test the reliability of the estimates, $\hat{b}_0$ and $\hat{b}_1$, respectively.

3.7 Summary

R^2, T and Z tests are important components of econometric test that are used in evaluating the strength of econometric variables. In this unit you have learned how to determine and interpret the coefficient of multiple determination, R^2, the Z test and the T test for evaluating an estimated model. Evaluating the model helps in the assessment of reliability; how the model estimates can be used for policy and applications. A good and continuous learning of these methods will enable you to conduct an effective and useful econometric research.

3.8 References/ Further Reading

Asteriou, D., & Hall S. G. (2007). Applied Econometrics: A Modern Approach using EViews and Microfit. New York: Macmillan Palgrave.

Berndt, Ernst R. (1991). The Practice of Econometrics: Classic and Contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics. North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods. Cincinnati: South-Western Publishing Company.

Gujarati, D. N. (2003). Basic Econometrics. New-Delhi: Tata McGraw-Hill Publishing Company Ltd.

Koutsoyiannis, A. (1977). Theory of Econometrics: An Introductory Exposition

Econometric Methods Macmillan

Wooldridge, J. M. (2009). Introductory Econometrics: A Modern Approach (4th ed.). Singapore: Cengage Learning.

3.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Carefully discuss the econometric significance of the R^2.

The coefficient of determination, R^2, is used in determining goodness of fit of the regression line obtained using the OLS method. The R^2 gives the degree to which the explanatory variable can explain the variation in the dependent variable *Y*. Thus, it is used in testing the explanatory power of the linear regression of *Y* on *X*.

2. Given an estimated model,

$\hat{Y}$ = 27 + 14X

(13.44) (8.61), R^2 = 0.63, n = 8

Test the reliability of the estimates $\hat{b}_0$ and $\hat{b}_1$, respectively, using the standard error test.

Solution

$\hat{b}_0 = 27; \hat{b}_1 = 14; S(\hat{b}_0) = 13.44; S(\hat{b}_1) = 8.61$

Therefore, $S(\hat{b}_0) = 13.44 < ½\hat{b}_0 = 13.5$ and we reject the null hypothesis that $H_0: \hat{b}_0 = 0$. This shows that $\hat{b}_0$ is statistically significant. Similarly, $S(\hat{b}_1) = 8.61 > ½\hat{b}_1 = 7$ and we accept the null hypothesis. Thus, the estimate $\hat{b}_1$ is not statistically significant.

3. Given a sample size of n=20, the model was estimated to be:

$\hat{Y}$ = 48+0.55X

(12.2) (0.4)

Test the reliability of the estimates, $\hat{b}_0$ and $\hat{b}_1$, respectively.

Solution

Applying the t test since n = 20 < 30,

$$t_0^* = \frac{\hat{b}_0}{S(\hat{b}_0)} = \frac{48}{12.2} = 3.934$$

$$t_1^* = \frac{\hat{b}_1}{S(\hat{b}_1)} = \frac{0.55}{0.4} = 1.375$$

The critical values of t for $(n - k)$ or 20 – 2 = 18 d.f. are:

$-t_{0.025} = -2.101$ and $+t_{0.025} = 2.101$.

Since $t_0^* = 3.934 > t_{0.025} = 2.101$, reject the null hypothesis and conclude that the estimate $\hat{b}_0$ is statistically significant at 5% significance level.

However, since $t_1^* = 1.375 < t_{0.025} = 2.101$, accept the null hypothesis and conclude that the estimate $\hat{b}_0$ is not statistically significant at 5% significance level.

Unit 4: Confidence Interval of Econometric Estimates

Contents

4.1 Introduction

One of the basic requirements in carrying out a regression analysis is an understanding of the level of confidence at which model variables are estimated. It serves as a control in model estimation. It helps in the measurement and reporting of model variables that are estimated at any given point in time. This is a dynamic process requiring deliberate and purposeful actions in order to ensure model compliance between the theoretical and the empirical model.

4.2 Learning Outcomes

At the end of this unit, you should be able to

- Discuss and apply confidence interval for $\hat{b}_0$.
- Discuss and apply confidence interval for $\hat{b}_1$.
- Understand the goodness of estimators.

4.3 Confidence Intervals of Parameter Estimates

The construction of confidence intervals for the estimates $\hat{b}_0$ and $\hat{b}_1$, enables us to state how close to these estimates the true parameter lies. It shows the limiting values within which the true parameter is expected to lie within a certain degree of confidence. In econometrics, we usually choose 95%. This implies that the confidence limit computed from a given sample would include from a population parameter in 95% of the cases. That is, we are 95% confident that our parameter estimates represent the true population parameter.

4.3.1 Confidence interval for the *Z*-statistic

Using the z distribution, the 95% confidence interval for the parameters $\hat{b}_i$ is constructed as follows:

$$\hat{b}_i - 1.96(S(\hat{b}_i)) < \hat{b}_i < \hat{b}_i + 1.96(S(\hat{b}_i))$$

This means that the unknown population parameter $\hat{b}_i$ will lie within the limit 95 times out of 100.

Thus, from the previous regression equation

$\hat{Y}$ = 19 + 4.5X

(13.14) (1.61)

and choosing 95% for the level of confidence, our confidence interval for $\hat{b}_0$ is

$\hat{b}_0 - 1.96(S(\hat{b}_0)) < \hat{b}_0 < \hat{b}_0 + 1.96(S(\hat{b}_0))$

$= 19 - 1.96(13.14) < \hat{b}_0 < 19 + 1.96(13.14)$

$= -6.75 < \hat{b}_0 < 44.75$

This implies that the true population parameter b_0 will lie between -6.75 and 44.75 with a probability of 0.95.

Similarly, the confidence interval for $\hat{b}_0$ is constructed as shown:

$\hat{b}_1 - 1.96(S(\hat{b}_1)) < \hat{b}_1 < \hat{b}_1 + 1.96(S(\hat{b}_1))$

$= 4.5 - 1.96(1.61) < \hat{b}_1 < 4.5 + 1.96(1.61)$

$= 1.34 < \hat{b}_1 < 7.66$

We also conclude that the true population b_1 will lie between 1.34 and 7.66 with a probability of 0.95.

4.3.2 Confidence interval for the *t*-statistic

This is also constructed the same way as for the *z* distribution. The difference is that the *t*-distribution uses *n* – *k* degrees of freedom. Thus, the 95% confidence interval for the parameter estimates $\hat{b}_i$ is constructed as follows;

$\hat{b}_i - t_{0.025}S(\hat{b}_i) < \hat{b}_i < \hat{b}_i + t_{0.025}S(\hat{b}_i)$ with *n* – *k* degrees of freedom.

Example:

From a sample of 20, if the estimated model was $\hat{Y}$= 13.25 + 1.8X

(6.04) (0.35)

The 95% confidence interval for the parameter estimates is constructed as shown: df = 20 – 2 = 18.

For $\hat{b}_0$: $\hat{b}_0 - 2.101(S(\hat{b}_0)) < \hat{b}_0 < \hat{b}_0 + 2.101(S(\hat{b}_0))$

$$= 13.25 - 2.101\ (6.04) < \hat{b}_0 < 13.25 + 2.101(6.04)$$

$$= 0.560 < \hat{b}_0 < 25.94$$

This shows that the value of the true population parameter $\hat{b}_0$ (or the intercept) will lie between 0.56 and 25.96 with a probability of 0.95.

Also, for $\hat{b}_1$: $\hat{b}_1 - 2.101\big(S(\hat{b}_1)\big) < \hat{b}_1 < \hat{b}_1 + 2.101\big(S(\hat{b}_1)\big)$

$$= 1.8 - 2.101(0.35) < \hat{b}_1 < 1.8 + 2.101(0.35)$$

$$= 1.095 < \hat{b}_1 < 2.535$$

This implies that the value of true parameter b_1 will lie between 1.095 and 2.535, given a probability of 0.95.

Self-Assessment Exercise

1) A sample of 20 was taken. If the estimated model was $\hat{Y} = \underset{(4.01)}{17.20} + \underset{(0.65)}{2.81}X$

   Determine the confidence interval for the parameter estimates.

2) A sample of 32 was taken. If the estimated model was $\hat{Y} = \underset{(3.01)}{15.20} + \underset{(0.60)}{0.81}X$

   Determine the confidence interval for the parameter estimates.

3) A sample of 40 was taken. If the estimated model was $\hat{Y} = \underset{(2.01)}{19.20} + \underset{(0.25)}{1.81}X$

   Determine the confidence interval for the parameter estimates.

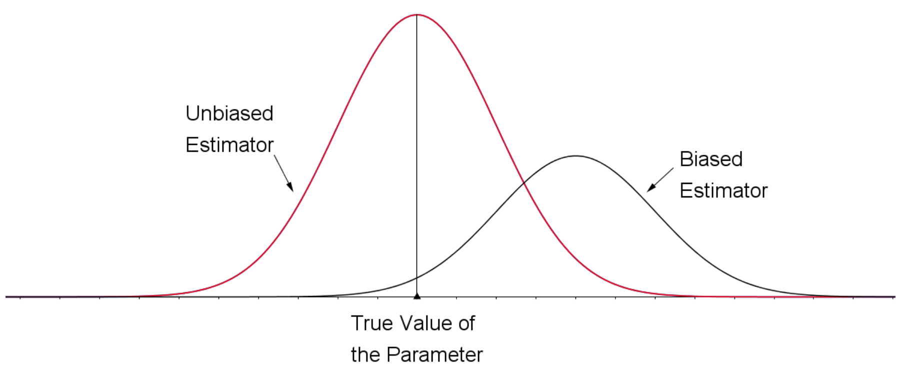
4.4 Ordinary Least Squares Estimators

The ordinary least squares (OLS) method is among the best methods used in obtaining estimates of the parameter $\hat{b}_0$ and $\hat{b}_1$. The use of OLS method in estimating economic relationship is based on the fact that the estimates of the parameters have some optimal properties.

Generally, in choosing a particular estimation method, one should aim at such method that gives an estimate, which is close to the value of the true population parameter; the variation of which (if at all it exists) will be within only a small range around the true parameter.

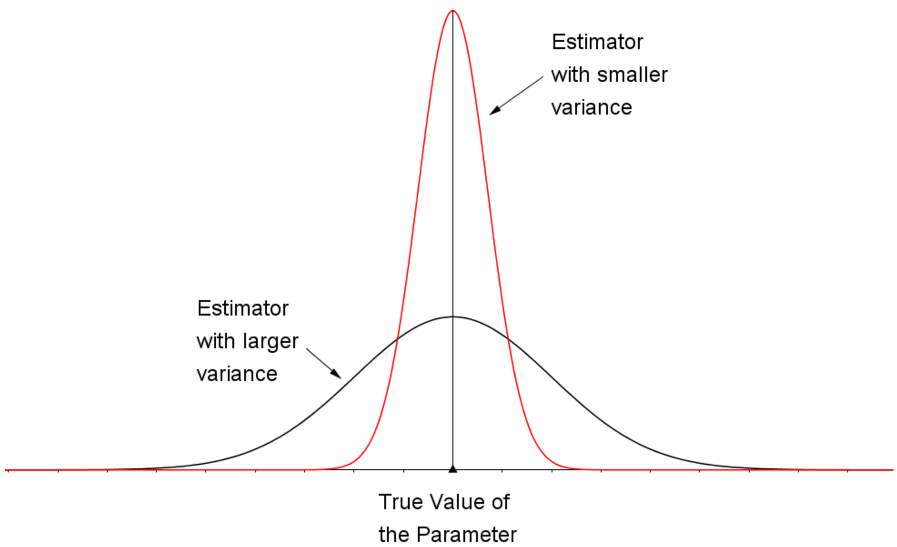
For any estimation method, the goodness of the estimator is judged on the basis of the following desirable properties.

(i) **Unbiasedness:** An estimator is unbiased if its mean is zero, i.e., $E(\hat{b}_i) - b_i = 0$. In such case, the unbiased estimator changes to the true value of the parameter as the number of samples increases. An unbiased estimator always gives, on the average, the true value of the parameter. The cases of biased and unbiased estimator of the true value are as illustrated diagrammatically.



Source: Statistics (http://timbusken.com/estimation.html)

(ii) **Minimum Variance:** An estimator is best if it has the smallest variance compared with any other estimate obtained using other methods. By minimum variance, we mean that the values of the parameter $\hat{b}_i$ cluster very closely around the true parameter b_i, as seen in the diagram. In the diagram, the narrower curve has minimum variance and, therefore, it is best.



Source: Statistics (http://timbusken.com/estimation.html)

(iii) **Efficient estimator:** An estimator is efficient when it combines the property of unbiasedness with minimum variance property. Symbolically, $\hat{b}_i$ is efficient if the following two conditions are fulfilled.

(a) $E(\hat{b}_i) = b_i$

(b) $E[\hat{b}_i - E(\hat{b}_i)]^2 < E[(b_i^*)]^2$

Where b_i^* is another unbiased estimate of the true b_i. This means that in the class of unbiased estimators, such estimator has a minimum variance.

(iv) **Linear Estimator:** An estimator is linear if it is a linear combination of the given sample data. Thus, with the sample observation, $Y_1, Y_2, \ldots, Y_n$, a linear estimator takes the form:

$K_1Y_1 + K_2Y_2 + \ldots + K_nY_n$, where the K_is are some constants.

(v) **BLUE (Best, Linear, Unbiased Estimator)**

This property is abbreviated to BLUE meaning that the estimator is best (having minimum variance), linear, and unbiased as compared with all other linear unbiased estimators.

Thus, all the properties (i-iv) are included in the BLUE property.

(vi) Minimum mean square error (MSE) Estimators.

This property combines unbiasedness and the minimum variance properties. An estimator, therefore, is a minimum MSE estimator if it has the smallest mean square error, defined as the expected value of the squared difference of the estimator around the true population parameter b_i. That is, $MSE(\hat{b}_i) = E[(\hat{b}_i - b_i)^2]$.

(viii) Sufficiency:

This property implies that the estimator uses all the available information a sample contains about the true parameter. For this property to hold, the estimator should accommodate all the observations of the sample and should not give room for any additional information in connection with the true population parameter.

The ordinary least squares method (OLS) satisfies the above stated properties. For this reason, the method seems to be the best and most widely used of all the estimation methods. In a nutshell, the OLS has the BLUE (best, linear, unbiased properties among the class of linear and unbiased estimators). The linearity property as previously discussed implies that the parameter estimates are linear functions of the observed Y_i. That is, the estimators $\hat{b}_0$ and $\hat{b}_1$ include the variable Y and X in the first power.

Thus, $\hat{b}_i = f(Y)$. This property enables one to compute the values of the parameter estimates with ease.

The unbiased property of the OLS estimates implies that the expected value of the estimated parameter is equal to the true value of the parameter, that is, $E(\hat{b}_i = b_i)$. The importance of this property lies in the fact that for large samples, the parameter estimates obtained will on the average give a true value of the bs.

The minimum variance property becomes desirable when combined with unbiasedness. The importance of this property is obvious when we want to apply the standard tests of significance for $\hat{b}_0$ and $\hat{b}_1$ and to construct confidence intervals for these estimates. Because of the minimum variance they have, their respective confidence intervals will be narrower than for other estimates obtained using any other econometric procedures.

The smaller confidence interval obtained is interpreted to mean in effect that we are extracting more information from our sample than we would be, if we were to use any other methods which yielded the same unbiased estimates.

Self-Assessment Exercise

Discuss the properties of ordinary least squares estimators.

4.5 Summary

Confidence interval of estimate and estimators of parameters are two basic requirements that are considered in the analysis of models in econometrics. In this unit you have learned how to calculate confidence interval of the parameter estimates of models. This can be applied to model interpretation and the implication for the purpose of determining the use of the model. Your ability to assess a model will depend on a combination of these econometric tests.

4.6 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The Practice of Econometrics: Classic and contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics. North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods, John Wiley & Sons, New York.

Cramer, J. S. (1969) Empirical Econometric, North-Holland, Amsterdam. Davidson,

James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969) Applied Econometrics, McGraw-Hill, New York,

Dhrymes, Phoebus J. (1978) Introductory Econometrics, Springer-Verlag, New York

Dutta, M. (1975) Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis (4th ed.). Prentice Hall, Engle-wood cliffs, N. J.

Goldberger, Arthur S. (1998) Introductory Econometrics, Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). Cambridge, Massachusetts: MIT Press.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, and George Judge (2001). Undergraduate Econometrics, John Wiley & Sons, New York.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics, John Wiley & Sons, New York.

Katz, David A. (1962). Econometric Theory and Applications, Prentice Hall, Englewood Cliffs, N.J.,

Klein, Lawrence R.: A (1974). Textbook of Econometrics, 2nd ed., Prentice Hall, Englewood Cliffs, N.J.

Kmenta, Jan (1986). Elements of Econometrics, 2nd ed., Macmillan, New York.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., & Kenneth F. Wallis (1981). Introductory Econometrics, 2nd ed., John Wiley & Sons, New York. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. New York: Routledge.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin's Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics, Wadsworth, Belmont,

Califonia.
Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.
Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).
Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, London.
Wonnacott, R. J., & Wonnacott, T. H. (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.
Wooldridge, Jeffrey M. (2000). Introductory Econometrics. South-Western College Publishing.

4.7 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. A sample of 20 was taken. If the estimated model was $\hat{Y} = 17.20 + 2.81X$
$$\hat{Y} = \underset{(4.01)}{17.20} + \underset{(0.65)}{2.81}X$$

Determine the confidence interval for the parameter estimates.

Solution

Using the t-test, the 95% confidence interval for the parameter estimates is constructed as shown: df = 20 – 2 = 18.

For $\hat{b}_0$: $\hat{b}_0 - 2.101(S(\hat{b}_0)) < \hat{b}_0 < \hat{b}_0 + 2.101(S(\hat{b}_0))$

$$= 17.20 - 2.101\ (4.01) < \hat{b}_0 < 17.20 + 2.101(4.01)$$

$$= 8.77499 < \hat{b}_0 < 25.62501$$

This shows that the value of the true population parameter b_0 (or the intercept) will lie between 8.775 and 25.625 with a probability of 0.95.

Also, for $\hat{b}_1$: $\hat{b}_1 - 2.101\big(S(\hat{b}_1)\big) < \hat{b}_1 < \hat{b}_1 + 2.101\big(S(\hat{b}_1)\big)$

$$= 2.81 - 2.101(0.65) < \hat{b}_1 < 2.8 + 2.101(0.65)$$

$$= 1.44435 < \hat{b}_1 < 4.16565$$

This implies that the value of true parameter b_1 will lie between 1.444 and 4.166, given a probability of 0.95.

2. A sample of 32 was taken. If the estimated model was $\hat{Y} = 15.20 + 0.81X$

$$\hat{Y} = \underset{(3.01)}{15.20} + \underset{(0.60)}{0.81}X$$

Determine the confidence interval for the parameter estimates.

Solution

Using Z-Test and choosing 95% for the level of confidence, our confidence interval for $\hat{b}_0$ is

$$\hat{b}_0 - 1.96(S(\hat{b}_0)) < \hat{b}_0 < \hat{b}_0 + 1.96(S(\hat{b}_0))$$

$$= 15.20 - 1.96(3.01) < \hat{b}_0 < 15.20 + 1.96(3.01)$$

$$= 9.3004 < \hat{b}_0 < 21.0996$$

This implies that the true population parameter b_0 will lie between 9.3004 and 21.0996 with a probability of 0.95.

Similarly, the confidence interval for $\hat{b}_1$ is constructed as shown:

$$\hat{b}_1 - 1.96(S(\hat{b}_1)) < \hat{b}_1 < \hat{b}_1 + 1.96(S(\hat{b}_1))$$

$$= 0.81 - 1.96(0.60) < \hat{b}_1 < 0.81 + 1.96(0.60)$$

$$= -0.366 < \hat{b}_1 < 1.986$$

We also conclude that the true population parameter b_1 will lie between -0.366 and 1.986 with a probability of 0.95.

3. Discuss the properties of ordinary least squares estimators.

The properties of ordinary least squares estimators include (Note that you need to discuss them in detail):

a. Unbiasedness
b. Minimum variance
c. Efficient estimator
d. Linear estimator
e. Minimum mean square error estimator
f. Sufficiency

Unit 5: Stationarity of Time Series Data

Contents

5.1 Introduction

Stationarity of data is considered to be the most important and most effective aspect of data management in order to minimize the occurrence of spurious results in econometric model estimation. Stationarity is fundamental to all Modern Econometric Estimation.

5.2 Learning Outcomes

At the end of this unit, you should be able to:

- Define stationarity.
- State the steps for determining stationarity in data management.
- Determine and discuss data stationarity using Dickey–Fuller method.

5.3 Meaning of Stationarity

In econometric analysis, data is stationary when its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed (Gujarati, 2005). Generally, if a time series data is stationary, its mean, variance and auto covariance at various lags remain the same no matter at what point they are measured, that is, they are time invariant. The time series data will trend to return to its mean. This is called mean reversion. The fluctuations around this mean measured by its variance will have a constant.

According to Wooldridge (2009), a stationary time series process is one whose probability distribution are stable over time. If a collection of random variables in the sequence is taken and then that sequence is shifted ahead some given time periods, the joint probability distribution should remain unchanged. However, it does not require that the nature of any correlation between adjacent terms is the same across all time periods.

Generally, a key concept underlying time series processes is that of stationarity. A time series is covariance stationary when it has the following three key characteristics:

a) A stationary time series exhibits mean reversion in that fluctuates around a constant long run mean $E(Y_t) = Constant$, for all t.

b) A stationary time series data has a finite variance that is time invariant. $Var(Y_t) = Constant$, for all t.

c) A stationary time series data has a theoretical correlogram that diminishes as the lag length increases. $Cov(Y_t, Y_{t+k}) = Constant$, for all t and $k \neq 0$.

Stationarity is important because if the series is non-stationary, then all the typical results of the classical regression analysis are not valid. Regression with non-stationary series may have no meaning and are, therefore, called spurious.

It is also important to note that in stationary data, shocks to such time series are necessarily temporary overtime. The effects of shocks will dissipate, and the series will revert to its long run mean level. As such long-term forecast of a stationary series will converge to the unconditional mean of the series. On the other hand, non-stationary time series will necessarily contain permanent components so the mean and variance of non-stationary series will depend on time, which leads to a case where a series has no long-run mean to which the series returns, and the variance will depend on time and will approach infinity as time goes to infinity (Asteriou & Hall, 2007).

Self-Assessment Exercise

What is stationarity?

5.4 Test for Stationarity

There are several tests of stationarity but the most commonly used in applied econometrics is the Dickey–Fuller augmented unit root test.

$$Y_t = \rho Y_{t-1} + \mu_t \qquad -1 \leq \rho \leq 1 \qquad ---- (1)$$

Where μ_t is a white noise error term. If $\rho = 1$ (in the case of unit root), equation (1) becomes a random walk model without drift which we know is a nonstationary stochastic process.

Using theoretic underpinning subtract Y_{t-1} from both sides of equation (1)

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \mu_t \qquad ----(2)$$

$$= (\rho - 1)Y_{t-1} + \mu_t \qquad ----(3)$$

Equation (3) can be written alternatively as

$$\triangle Y_t = \sigma Y_{t-1} + \mu_t \qquad ----(4)$$

Where $\sigma = (1 - \rho)$ and $\triangle$ is the first difference operator. Equation (4) is what is estimated and test the null hypothesis that σ = 0. If $\sigma = 0$, then $\rho = 1$, that is, there is unit roots; meaning the time series is nonstationary.

Similarly, when $\sigma = 0$, equation (4) becomes:

$$\triangle Y_t = (Y_t - Y_{t-1}) = \mu_t \qquad ----(5)$$

Since μ_t is a white noise error term, it is stationary, which means that the first difference of a random walk time series is stationary. In estimation, take the first differences of Y_t and regress them on Y_{t-1} to see if the estimated slope coefficient in the regression ($= \hat{\sigma}$) is zero or not. If it is zero, it shows that Y_t is nonstationary. But if it is negative, we can conclude that Y_t is stationary.

5.4.1 Dickey–Fuller Test

The Dickey–Fuller test evaluates the null hypothesis that $\sigma = 0$; that is, there is a unit root, meaning that the time series is nonstationary. The alternative hypothesis is that σ is less than zero; that is, the time series is stationary. When the null hypothesis is rejected, it means that Y_t is a stationary time series.

5.5 Overcoming Non-Stationary Data Problem

When data is tested for stationarity and they show nonstationarity, then do transform the nonstationary time series to make them stationary. The process of transformation varies depending on whether the time series are difference stationary or trend stationary.

In difference stationarity, it involves taking the first difference of the time series in order to make them stationary. Similarly, if the time series are trend nonstationary then we take the trend.

Self-Assessment Exercise

Outline the processes in transforming nonstationary time series to make them stationary.

5.6 Econometric Applications

Given the following data determine the stationarity using Dickey–Fuller method.

Varasset	Fixasset
3820.1	1318.42
3757.9	1418.9
5382.8	1147.74
5949.5	1408.06
8418.3	1585.88
8804	1640.44
9313.8	1703.82
9993.6	1844.39
11339.2	2045.32
10899.6	2239.51
104368.1	4974.5
12243.8	5499.78
20512.7	6339.28
66787	7243.61
70714.6	8096.27
119391.6	112408.1
122800.9	157477.9
128331.8	158998.9
152409.6	162423
154188.6	162458.2
157835.4	162531.5
162343.4	182541.2
168831.6	182637
178478	183418.3
249229.6	184946.5
269844.7	188468.6
302848.3	189312.6
317571.2	188183.9
329381	187706.9
380339.7	191203.2
438201.3	194322.2

The procedure is as follows: open EViews software create a worksheet, import your set of data by browsing and clicking to upload your data to the workbook; go to "Quick" icon on the software toolbar, select "Unit Root Test" under "Series Statistics" and a dialogue box opens where you input the variable name. Select “Augmented Dickey–Fuller” under “Test type” then click “OK”.

Null Hypothesis: VARASSET has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=7)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		2.164124	0.9999
Test critical values:	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

VARASSET is nonstationary because it failed to meet the stationarity condition, with absolute value of the test statistic (2.164124) being less than the absolute values of all the levels of significance and a probability value of 0.9999. It can be concluded that it is nonstationary (having unit root) at “level”. We, therefore, repeat the procedure and select “1st difference” in the Unit Root Test dialogue box.

Null Hypothesis: D(VARASSET) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.488224	0.0000
Test critical values:	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

VARASSET is now stationary because it satisfies the stationarity condition, with absolute value of the test statistic (6.488224) being greater than the absolute values of all the levels of significance and a probability value of 0.0000.

For FIXASSET (shown below), it is also nonstationary (having unit root) at level, with absolute value of the test statistic (0.492480) being less than the absolute values of all the levels of significance and a probability value of 0.8793. We, therefore, repeat the procedure and select “1st difference” in the Unit Root Test dialogue box.

Null Hypothesis: FIXASSET has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.492480	0.8793
Test critical values:	1% level	-3.670170	
	5% level	-2.963972	
	10% level	-2.621007	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(FIXASSET) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.821668	0.0071
Test critical values:	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

VARASSET is now stationary because it satisfies the stationarity condition, with absolute value of the test statistic (3.821668) being greater than the absolute values of all the levels of significance and a probability value of 0.0071.

Self-Assessment Exercise

Given the following data, test for the stationarity of gdpl and reerl using your EViews software.

Gdpl	Reerl
11.83957	2.075806
11.78133	2.055386
11.75842	2.03529
11.73871	2.018888
11.71619	2.011659
11.73485	2.022474
11.88075	2.037689
11.9649	2.042883
12.00148	2.033174
12.00353	2.023693
12.09498	2.038501
12.09484	2.039128
12.13759	2.045691
12.11095	2.050573
12.13491	2.049547
12.19589	2.060913
12.19689	2.059162
12.15363	2.036549
12.16782	2.037144
12.16359	2.02435
12.12319	2.033106
12.12703	1.998442
12.16356	2.005447
12.25528	2.028002
12.31416	2.035662
12.33174	2
12.35529	2.031445
12.41397	2.03442
12.45549	2.04023
12.42315	2.040994

5.7 Summary

Stationarity of time series data is very important. If it is effectively achieved, then the regression result will not be spurious. The result emerging after stationary test has been carried out can be used for policy analysis. In this unit, you have been taken through the process of knowing stationarity in econometrics. You have also learnt about how to achieve stationarity in data management in econometric analysis.

5.8 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley,

Bridge, J. I. (1971). Applied Econometrics. North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric, North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods. South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis (4th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). Cambridge, Massachusetts: MIT Press.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, & George Judge (2001). Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973) Econometrics: An Introductory Analysis. Baltimore: University Park Press.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, & Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics. New York: John Wiley & Sons.

Katz, David A. (1962). Econometric Theory and Applications. Englewood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R.: A (1974). Textbook of Econometrics (2nd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Kmenta, Jan (1986). Elements of Econometrics (2nd ed.). New York: Macmillan.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., and Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). New York: John Wiley & Sons. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series. Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. Routledge, New York.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin’s Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics. Wadsworth, Belmont, California.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics. London: Macmillan.

Wonnacott, R. J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, J. M. (2000). Introductory Econometrics. South-Western College Publishing.

5.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. What is stationarity?

Data is stationary when its mean and variance are constant over time and the value of the covariance between the two periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. If a time series data is stationary, its mean, variance and auto covariance at various lags remain the same no matter at what point they are measured, that is, they are time invariant.

2. Outline the processes in transforming nonstationary time series to make them stationary.

The process of transformation varies depending on whether the time series are difference stationary or trend stationary as follows:

For difference stationarity, take the first difference of the time series in order to make them stationary.

For trend non-stationarity, take the trend.

3. Given the data on GDPL and REERL, test for the stationarity of gdpl and reerl using your EViews software.

The results below show that GDPL is nonstationary at level, while REERL is stationary at level.

Null Hypothesis: GDPL has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

	t-Statistic	Prob.*

Augmented Dickey-Fuller test statistic		-0.167251	0.9322
Test critical values:	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: REERL has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=7)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.463441	0.0166
Test critical values:	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

*MacKinnon (1996) one-sided p-values.

Module 2

Unit 1: General Forecasting

Contents

1.1 Introduction

Forecasting is often seen as bridge between the present economic events occurring and the future. It is needed for effective planning and future projections based on the present economic phenomenon. Forecasting is particularly important because resources are scarce, coupled with the environmental uncertainties. It is, therefore, good to forecast and minimize the degree of uncertainty in the economy.

1.2 Learning Outcomes

At the end of this unit, you should be able to:

- Explain the meaning of forecasting.
- Discuss moving average seasonal variations.

1.3 Meaning of Forecasting

Forecasting is generally concerned with data associated with a time period of less than one year (i.e. calendar months, accounting periods in a year – usually 12 or 13, quarters, weeks, etc.); situations where forecasts are associated with a particular item and for which forecasts are updated every time period; situations where forecasts are required for a large number of items; situations where the forecasts produced for a particular item or product and are used on a period to period basis to:

1.2.1 Analyse demand and to assess appropriate inventory levels and production schedules, and

1.2.2 Analyse sales to help assess cash flows and to ascertain marketing procedures.

It is apparent that for these types of application, the most appropriate forecasting model (or series of models) needs to be:

(i) cheap to operate - in terms of implementation, routine updating and storage requirement costs;

(ii) flexible and hence able to offer a variety of different, but closely related, model types suitable for a wide variety of items and situations for which forecasts could be required;

(iii) largely automatic, such that a minimum of manual interruption is necessary;

(iv) Well proven, and hence, readily available in both the literature and computer software.

The range of forecasting models based on the exponentially weighted average (collectively referred to as exponential smoothing methods), which were introduced in the early 1960's, have been shown to fulfil most of the above requirements.

The superiority of exponential smoothing methods over the traditional moving average concept is such that today most manufacturing organizations of any size use them and no industrial or business computer software is complete without them.

A stationary situation is one in which, although observed values fluctuate from one time period to the next, the average value remains steady over a reasonably long period of time. To illustrate this, this, a series of demand values plotted against time. It can be seen that the average value per month over the one-year period is about 100 items per month, and that this average figure is neither increasing nor declining significantly with time. This is a typical stationary situation, with individual values fluctuating above and below a reasonably steady average.

Examining nonstationary situation, however, reveals a completely different situation. Here the average value is definitely not stationary but increasing with time. Such a situation could arise with a product maintaining its share of a rapidly expanding market or a product gaining an increasing share of a static (or even declining) market. Exactly the types of market condition that cause such values can only be identified by sales and marketing intelligence, but whatever the reason, this forecasting situation is obviously a more complicated one than the stationary.

Before discussing the mathematical techniques involved in forecasting, it might be the best to discuss initially just what one should expect of forecasting system. First, it should be appreciated that because one's forecasts are based on past information, there will always be some degree of forecasting error. Accepting this, it is apparent that the most logical aim of a forecasting system should be to minimize such errors over a longish period of time. Because forecasting errors can be both positive and negative (i.e., the forecast can be below or above the actual value that occurs) a simple cumulative sum of the errors alone will not be a good indication of whether forecasting errors are being minimized or not, as this sum will tend towards zero irrespective of the forecasting system used. A more effective measure is the cumulative sum of the squared errors since squaring always produces a positive result irrespective of whether the original figure was positive or negative. Thus, all errors contribute towards the cumulative sum of squared errors.

Once it has been accepted that some degree of error is present in any forecasting system, it is evident that the forecast can be only an average value of what is expected to occur, with errors distributed evenly on either side of that average. In practice, it is generally assumed that these errors are distributed according to a probability distribution known as the Gaussian or Normal distribution. This point will not be explained further here except to say that the assumption is reasonably true so long as the value of the average per time period is not too low, which is generally true of the items we are concerned with in short-term forecasting if this average value is generally greater than ten.

Assuming that errors are distributed normally, we need some measure of the spread or the degree of dispersion of errors around the average. The usual measure of dispersion is known as the **standard deviation**, represented universally by the **Greek symbol $\boldsymbol{\sigma}$**, which indicates a situation in which the standard deviation suddenly increases owing to a change in the underlying data pattern.

For this situation, the forecast would not change appreciably as the average value remains at approximately 75 per month, even after the change. However, it is evident that the spread of values, or variation, has changed after the sixth month and this change, in an efficient forecasting system, would be noted by a change in the value of the standard deviation.

Thus, there are two basic parameters we wish to estimate in any forecasting system. The **first is the actual forecast** which predicts what the expected or average value in the future is likely to be. The **second is the standard deviation** which measures the spread or dispersion of individual values about that average.

Self-Assessment Exercise

1) What is forecasting?

2) Discuss forecasting in stationary and non-stationary situation.

1.4 Forecasting Process-Linked Period

A short-term forecasting system treats the total of individual values in each time period as a single item of data, i.e., demand per day, sales per week, production per month, etc. Increasing the length of the time period increases the sample size and hence reduces the variability of successive individual values per period thus enabling more accurate forecasts to be made. At the same time, however, the speed of response of the forecasting system to real or actual changes in the data is obviously reduced with longer time periods. A balance between these two effects can be achieved only by the selection of a suitable forecast time period.

It can be shown that the minimum value of the forecast time period should be of a duration to ensure that at least one non-zero value occurs in two time periods, that is, that there is a 50 per cent probability that a non-zero value will occur during one time period. This is a minimum; overall considerations may require a longer interval.

This criterion of having one non-zero value occurring within two periods is normally met for industrial and business data if the time period is greater than or equal to one week, and a calendar month (or more practically a planning or accounting period of that order) is most typical. The number of time periods ahead for which forecasts are produced is known as the forecast horizon.

1.5 The Moving or Rolling Average*

One traditional method of forecasting the future average expected value per time period is to average the past individual values over the last n time periods. Such a moving average could be defined as

$$m_t = \frac{1}{n}\sum_{i=t}^{t-n+1} d_i$$

Or, alternatively,

$$m_t = m_{t-1} + \frac{1}{n}(d_t - d_{t-n})$$

This latter version simply means that we put the current value of the moving average as being equal to the immediate past value plus $1/n$ times the current value less the value now n periods old.

Having calculated m_t for a stationary situation this then becomes the forecast of what one expects to occur, not only in the next time period but for any future time period.

This does not mean, however, that if one makes a forecast, say, for six months, hence, this estimate cannot be modified next month, when an additional month's information can be used to improve the estimate, which is now only five months away. This concept may be slightly difficult to follow but corresponds very much with planning schedules, which are definite for the first and second months but only tentative for the third and fourth months. The tentative plans then become definite at the next two-monthly review in the same way that a vague forecast can become more definite as further information is received as time progresses.

The moving average does, however, in practice have several drawbacks which are discussed here:

(i) Starting a moving average. When beginning the calculation of moving average from new data, because it is necessary to have individual values available for the previous $n - 1$ periods, no true forecast can he made until at least n periods have passed.

(ii) With a moving average, all the data included within the average are 'weighted' equally and data too old to be included are obviously given zero weighting. The weight of an item indicates the proportion of its value that an item contributes to an average, which in the case of the moving average is $1/n$ for all items included in the average, and zero for those not included.

Self-Assessment Exercise

1. Using appropriate examples discuss moving average in forecasting.
2. Discuss some of the drawbacks of moving averages.

A method using unequal weights could be proposed to resolve this feature, and two possible weighted averages based on either fractional or decimal concepts are shown here as equations respectively. Note that in both, the sum of the weights is one; this, by definition, is always necessary of a true average.

$$m_t = {}^1\!/_{2}d_t + {}^1\!/_{4}d_{t-1} + {}^3\!/_{16}d_{t-2} + {}^1\!/_{16}d_{t-3}$$

Or

$$m_t = 0.4d_t + 0.3d_{t-1} + 0.2d_{t-2} + 0.1d_{t-3}$$

(iii) When dealing with a moving average, the amount of past data that is retained can become excessive. A six-period moving average has been discussed but, in practice, to obtain an average which is not excessively sensitive, one can require data from up to 20 periods.

(iv) As the sensitivity or the speed of response of a moving average is inversely proportional to n, the number of time periods included in the average, it is difficult to change this sensitivity since it is also most difficult to change the value of n, as already illustrated by the initializing average situation.

Most of the disadvantages presented by the moving average can be overcome by a moving average of a special type in which the weighting series is exponential.

1.6 Summary

Forecasting is important in economics analysis and daily life. We often forecast changes in demand, prices, supply, exchange rates and other market situations. It is therefore important to note that forecasting is a dynamic component of our daily life. In this unit you have learn what forecasting entails, the need to forecast a stationary variable in an economic situation. You have also learned about forecasting of weighted variables.

1.7 References/ Further Reading

Aigner, D. J. (1971) Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of econometrics: Classic and contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory, Blackwell Publishers, Oxford, U.K.

Desai, Meghnad. (1969). Applied Econometrics, McGraw-Hill, New York,

Dhrymes, Phoebus J. (1978). Introductory Econometrics, Springer-Verlag, New York.

Dutta, M. (1975). Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis (4th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.).

Cambridge, Massachusetts: MIT Press.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, and George Judge (2001). Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis. Baltimore: University Park Press.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of econometrics. New York: John Wiley & Sons.

Katz, David A. (1962). Econometric Theory and Applications. Engle-wood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R. A. (1974). Textbook of Econometrics (2nd ed.). Engle-wood Cliffs, N.J.: Prentice Hall.

Kmenta, Jan (1986). Elements of Econometrics (2nd ed.). New York: Macmillan.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., & Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). New York: John Wiley & Sons, A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. New York: Routledge.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin’s Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics, Wadsworth, Belmont, California.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons. (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, London.

Wonnacott, R. J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics. South-Western College Publishing.

1.8 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. What is forecasting?

   Forecasting is the process of predicting future events or values from events or values recorded in the past.

2. Discuss forecasting in stationary and non-stationary situation.

   A stationary situation is one in which, although observed values fluctuate from one period to the next, the average value remains steady over a reasonably long period of time. For the non-stationary situation, average value is not stationary but increasing or decreasing with time. (Note that you need to discuss this in more detail.)

3. Discuss some of the drawbacks of moving averages.

   Some of the drawbacks of moving averages include (Note that you need to discuss them in detail):

   a. No true forecast can he made until at least n periods have passed.

   b. All the data included within the average are 'weighted' equally and data too old to be included are given zero weighting.

Unit 2: Forecasting and Seasonal Variations

Contents

2.1 Introduction

A linear trend is one in which the average value either increases or decreases linearly with time. A product with a linearly increasing trend could be typified by a new product with ever-increasing sales, or by a product in an expanding market retaining its percentage of that market. A product with a linearly decreasing trend could be typified by a product rapidly becoming obsolete.

2.2 Learning Outcomes

At the end of this unit you should be able to:

- Discuss forecast trends.
- Explain reasons for forecasting.
- Understand the applications of forecasting and manpower needs.

2.3 Forecasting and Trends

Since planning is a systematic, economic and rational way of making decisions today that will affect tomorrow, then, forecasting becomes an integral part of the planning process in any economy. There are different types of trends that are used to describe forecasting.

(a) Seasonal trends

A seasonal trend is one in which the average varies in a cyclical fashion in sympathy with some imposed time cycle. In most practical situations, this time cycle is invariably a yearly one in which the average in some months is up, compared with the overall monthly average, and in some months it is down. Fashion goods such as clothes and shoes are naturally subject to such seasonal trends, but a large section of this country's engineering industry is similarly affected via the sales of cars, which decline with the changes in seasons.

(b) Combined linear and seasonal trends

As the name indicates, this type of trend is a direct combination of the two already mentioned. A good example of this can be seen in the bookings for airline seats. The increasing use of air transport over the years provides the linear element of the trend and the seasonal pattern of travel within each year (as influenced by peak demands during the Christmas, Easter and summer holidays) provides the seasonal element, based on an annual time cycle.

(ii) Types of trends

(A) Additive trend

An additive type of trend is one in which an approximately regular amount is added to or subtracted from each consecutive average value, as influenced by the character of the trend. For instance, with a linear additive trend, an average increase in demand of ten items for each consecutive month might be expected.

(B) Ratio trends (sometimes known as multiplicative trends)

A trend of a ratio type is subject to a percentage increase or decrease, as influenced by the character of the trend. Thus, the demand for an item with an increasing linear ratio trend might be expected to increase at, say, 2 per cent per month.

(C) Combined additive and ratio trends

This type of trend, which is of course a direct combination of the two already mentioned, tends to be rather complicated to analyse mathematically and is usually dealt with as one or the other.

When describing any demand trend, it is necessary to describe both its type and character. The more usual trends for which relatively simple forecasting models have been developed are:

(D) Linear additive trend

A product subject to this kind of trend would have an average value increasing (or decreasing) by an approximately fixed quantity each time period, which is distinguished from a linear ratio trend case by the fact that, although the average value is increasing, the spread or variation of individual values about that average remains virtually constant.

Self-Assessment Exercise

Discuss the types of trends.

2.4 Forecasting Techniques Using Regression Line

Forecasting using the regression line equation is simply a matter of substituting the correct value of a variable at time t in the regression equation:

$$\hat{y}_t = a + bt$$

and evaluating the results. Given expenditure data to produce a forecast for 2008, one recognizes that for that year t = 12 and, hence

$$y_{12} = 366.27 + 110.09w$$

We can estimate the forecast amount by substituting for "w" by the number in the appropriate year.

$$y_{12} = 366.27 + 110.09\,(12)$$
$$= 16687.37$$

If one were to adopt the normal forecasting convention, namely that the most recent observed value occurred at time t and forecasts were for τ periods ahead, it follows that in equation $t = n + \tau$.

2.5 Econometric Applications

Given the following data for an economy, estimate the regression equation and forecast the real gross domestic product using the end product year.

Gc	Pc	I	Y
4.3410	5.6809	5.2867	11.0291
4.5822	5.9086	5.4586	10.861
4.6594	5.9452	5.3721	10.833
4.6903	6.0669	5.0959	10.7541
4.6158	6.2385	4.6943	10.679
4.6921	6.3533	4.6300	10.747
4.6730	6.3720	4.8591	10.746
4.4913	6.7343	5.1270	10.713
4.8038	7.0478	5.2437	10.782
4.8736	7.1902	5.6418	10.826
4.9648	7.5048	6.0192	10.879
5.0693	7.6345	6.1136	10.9015
5.7784	8.2014	6.5384	10.906
6.1002	8.4745	6.8282	10.903
7.3645	8.6447	6.8901	10.881
7.6991	9.4423	7.1624	10.882
7.8199	9.8449	7.5021	10.900
8.0947	9.8059	7.6531	10.904
8.1124	9.9084	7.6272	10.899
7.5574	9.9195	7.5590	10.887
8.0651	9.9055	7.8922	10.915
8.0650	10.2781	7.9851	10.922
8.2117	10.6612	8.2554	10.913
8.1272	10.8766	8.7806	10.986
8.6588	10.9931	8.7526	11.062
8.8782	11.1877	8.6574	11.090
9.0998	11.3565	9.2863	11.125
9.3430	11.5929	9.4678	11.163
9.5862	11.8295	9.6493	11.196
9.8293	12.0660	9.8306	11.239
10.0723	12.3024	10.0119	11.289

The regression equation for the model can be presented in a line equation as follows:

using the procedure as follows: open EViews software, create a worksheet, import your set of data by browsing and clicking to upload your data to the workbook; go to the "Quick" icon on the software toolbar, select "Estimate Equation" and a dialogue box opens in which you input the variable name selected, then click OK.

Dependent Variable: Y
Method: Least Squares
Date: 02/15/16 Time: 12:53
Sample (adjusted): 1980 2010
Included observations: 31 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.49375	0.061958	169.3686	0.0000
GC	0.063434	0.008686	7.302817	0.0000

R-squared	0.647764	Mean dependent var	10.92944
Adjusted R-squared	0.635618	S.D. dependent var	0.154170
S.E. of regression	0.093064	Akaike info criterion	-1.848726
Sum squared resid	0.251164	Schwarz criterion	-1.756211
Log likelihood	30.65526	Hannan-Quinn criter.	-1.818569
F-statistic	53.33113	Durbin-Watson stat	0.329547
Prob(F-statistic)	0.000000		

Y = 10.494 + 0.063GC.

If the value of Y is given we can forecast the GC by substituting it into the equation. Assume GC is given as 24.61110. We can substitute as follows.

Y = 10.494 + 0.063GC

Y = 10.494 + 0.063(24.61110)

Y = 10.494 + 1.5506

Y = 12.0445

Self-Assessment Exercise

Give the following data for an economy, estimate a regression equation $Y = f(I)$ and forecast for the economy using the econometric software.

Pc	I	Y
5.6809	5.2867	11.0291
5.9086	5.4586	10.861
5.9452	5.3721	10.833
6.0669	5.0959	10.7541
6.2385	4.6943	10.679
6.3533	4.6300	10.747
6.3720	4.8591	10.746
6.7343	5.1270	10.713
7.0478	5.2437	10.782
7.1902	5.6418	10.826
7.5048	6.0192	10.879
7.6345	6.1136	10.9015
8.2014	6.5384	10.906
8.4745	6.8282	10.903
8.6447	6.8901	10.881
9.4423	7.1624	10.882
9.8449	7.5021	10.900
9.8059	7.6531	10.904
9.9084	7.6272	10.899
9.9195	7.5590	10.887
9.9055	7.8922	10.915
10.2781	7.9851	10.922
10.6612	8.2554	10.913
10.8766	8.7806	10.986
10.9931	8.7526	11.062
11.1877	8.6574	11.090
11.3565	9.2863	11.125
11.5929	9.4678	11.163
11.8295	9.6493	11.196
12.0660	9.8306	11.239
12.3024	10.0119	11.289

2.5.1 Some areas of forecasting

(a) Economic development: The economic conditions of the country as well as those of the whole world will have significant effect on the operations of an organization. This will include perditions relating to Gross Domestic Product (GDP), currency strength and foreign exchange market, industrial expansion, labour market, inflation rate, interest rate, balance of payments, etc. Good economic trends assist the growth of a company. On the other hand, the depressions of 1929 – 1930, for instance, pushed many companies out of business. Knowledge of economic trends will assist management in making plans for its organization.

(b) Technological Forecasts: these forecasts predict the new technological developments that may change the operations of an organization. The advent of the transistors put the vacuum tube totally out of business. The age of the electronic calculator totally wiped out the market for the slide rules. An aggressive entrepreneur keeps up to date with new technological developments and readily adopts new methods to improve performance of the company or business

(c) Competition Forecasts: It is equally necessary to predict as what strategies your competitors will be employing to gain market share, perhaps at the cost of your market share. This is constantly happening in the airline business, electrical and electronics business. The competitor may be working to employ a different marketing strategy for a product or brining out a substitute for the product which can be cheaper and easily acceptable by the consumers.

(d) Social Forecasts: these forecasts involve predicting changes in consumer tastes, demand and attitude. Consumers have already established a trend for convenience, comforts and for products that are easy to use and easily manageable. In the 1970's, the trend was to buy small economic cars. In the 1980's, the trend moved back to luxury and comfort. Partly, it may depend upon the general economic trends and partly on the consumer taste. Diesel cars were very popular in 1979 – 1980 and they totally lost popularity by 1984 – 1985. Jacket, highly popular in the 1960's, is no longer fashionable today.

(e) Other forecasts: In addition to the above major aspects, there are other critical areas with impacts on the planning process, where forecasting is equally important. These areas include the following:

i. New laws and regulations: Organizations are on a constant watch for any new laws passed relating to consumer protection or dumping of chemical wastes, and anti-trust laws. As an example, laws about truth in advertising have had a major impact on the operations of many organizations which had to make extensive changes in their operational style.

ii. Political Events: In politically stable countries, there may not be any dramatic political changes except that a different party in power may be more socialistic in nature, as it may interfere in private business matters. In more volatile political situations, organizations must be prepared for the worst. The cases of Kennecot Copper and ITT in Chile, during the communist takeover are classical examples of being caught unawares. Additionally, most countries have become so interdependent that political events in one country affect the organizational events in others. The oil embargo by some Middle East countries in 1973 affected other countries consumption, demand and other market variables.

2.5.2 Uses of forecasting

1. It improves the resource utilization; for example, human resources. Forecasting helps the management of a company forecast the recruitment needs in terms of both numbers as well as the types of skills required and develop ways to avoid or correct problems before they become serious and disrupt operations.
2. It helps focus the management efforts on the most likely sources of supply. This will

cut down the total costs of idle resources and reduce costs associated with market mistakes.

3. It makes provisions for replacement or back-up stock of goods from either inside or outside the organization whenever the need arises either on a temporary or permanent basis. These available sources of supply are important to identify specially in the case of any emergencies that may occur.
4. It helps achieve an integration of business plans with other operating as well as strategic plans by making available the required information base on other organizational activities.

Self-Assessment Exercise

Discuss the uses of forecasting.

2.6 Forecasting Manpower Needs

Manpower forecasting is defined as the process of collecting and analysing information to determine the future supply of, and demand for, any given skill or job category. The primary purpose for manpower forecasting is to prepare for employment, training and development and proper use of human resources. Forecasting has become very important due to complexity of organizations requiring more skilled personnel which are in short supply.

In addition to assessing the current situation relevant to personnel in the organization, an effective human resource planning program takes into consideration the future needs and trends in both job and personnel availability. This will require reliable forecasting of both the demand and supply. The key to any forecasting procedure is to determine what determines the supply and demand for people, we can predict the needs, making the planning for changes in personnel easier. Now, let us consider some salient issues here.

a. **Opinion of expert forecasters**: Expert forecasters are those who are knowledgeable at estimating future human resource needs. These may be the personnel managers or external consultants who are asked to assess future needs. These forecasts can be taken further ahead by using the "Delphi technique", where the experts are anonymously surveyed, their opinions refined, and these opinions sent back to experts for further evaluation. The process is repeated until a consensus is reached.

b. **Trend projection forecasts-** past trends of employee needs can he projected into the future. One method used is known as extrapolation – where, if the past trend is linear in nature, it can simply be extended into the future. For example, if in the past the company has been consistently growing, requiring on the average 10 additional workers per month, then by extrapolation, the future needs will be 120 workers per year.

Another method of trend projection is known as "**indexation**". This method relies upon matching employment growth with some index. For example, growth in the number of employees in the production area may be tied in with the volume of sales, so that for a certain increase in sales, there is a corresponding increase in the number of employees in production.

Both extrapolation and indexation are simple and appropriate. However, in long range human resources projection, more sophisticated statistical analysis techniques are available and employed, taking into consideration external, organisational and workforce factors, which may not remain constant from time to time.

Some of the other forecasting methods include the ones listed below.

(a) Budget and planning analysis – short-term human resource needs can be estimated by looking into the departmental budgets and financial avocations for more employees as well as any anticipated changes in the workforce like terminations, resignations, promotions, transfers, etc.

(b) New ventures analysis – human resource needs for new ventures or organisational expansion can be calculated by comparison with similar operating functions. The growth of the new venture can also be established by past trends of similar operating companies.

(c) Computer models – these are most sophisticated approaches involving solutions of mathematical formulae and equations, taking into consideration the relationship and effect of all variables are extrapolation, indexation, survey results and other estimates.

Self-Assessment Exercise

1) Discuss the key issues in manpower forecasting.
2) Discuss some manpower forecasting techniques.

2.7 Summary

Forecasting is important in economic analysis and daily life. We often forecast changes in demand, prices, supply, exchange rates and other market situations. It is, therefore, important to note that forecasting is a dynamic component of our daily life. In this unit you have learned what forecasting entails, the need to forecast a stationary and non-stationary data. We have also learned how to forecast using a regression line.

2.8 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Engle-wood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers

Desai, Meghnad (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis (4th ed.). Engle-wood Cliffs, N.J.:

Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). Cambridge, Massachusetts: MIT Press.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, and George Judge (2001) Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung - Chao Lee (1982). Introduction to the theory and practice of econometrics. New York: John Wiley & Sons.

Katz, David A. (1962). Econometric Theory and Applications. Engle-wood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R. A. (1974). Textbook of Econometrics (2nd ed.). Engle-wood Cliffs, N.J.: Prentice Hall.

Kmenta, J. (1986). Elements of Econometrics, 2nd ed., Macmillan, New York.

Maddala, G. S. (2001). Introduction to Econometrics, John Wiley & Sons, 3rd ed., New York.

Mark, Stewart B., & Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). John Wiley & Sons, New York. A Halsted Press Book.

Madansky, A. (1976) Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. New York: Routledge.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin’s Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics. Wadsworth, Belmont, California.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, London.

Wonnacott, R. J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics, South-Western College Publishing.

2.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Discuss the types of trends.

   The types of trends include (Note that you need to discuss them in detail):

   a. Additive trends
   b. Ratio trends
   c. Combined additive and ratio trends
   d. Linear additive trends

2. Discuss the uses of forecasting.

   Forecasting can be used to (Note that you need to discuss them in detail):

   a. Improve resource utilization.
   b. Cut down the total cost of idle resources.
   c. Help in inventory management.
   d. Help in the integration of business plans with other operating and strategic plans.

3. Discuss the key issues in manpower forecasting.

   In manpower forecasting, the following issues need to be taken into cognisance (Note that you need to discuss them in detail):

   a. The opinion of expert forecasters.

b. Trend projection forecasts.

4. Discuss some manpower forecasting techniques.

Manpower forecasting techniques include (Note that you need to discuss them in detail):

a. Delphi technique
b. Extrapolation
c. Indexation
d. Budget and planning analysis
e. New venture analysis
f. Computer models

Unit 3: Identification Problem

Contents

3.1 Introduction

One of the main ways of finding solution to simultaneous equation is to determine the identification of the econometric model. The identification of the model helps the researcher to focus on the method of estimating the model.

3.2 Learning Outcomes

At the end of this unit you should be able to:

- Discuss the concept of identification.
- Discuss the rank method of identification.
- Discuss the order method of identification.

3.3 Meaning of Identification

Identification is a problem associated with model formulation. Identification problem arises if a given model is not correctly specified. This is sequel to the violations of the assumption that the relationship to be estimated must have a unique mathematical form. If this is violated truly, then the relationship to be estimated will contain the same variable as any other equation one is estimating. In the face of an unidentified model, the estimates of the parameters of the relationship between the variable measured in samples relate to the model in question or to another model or to a mixture of models.

Let us illustrate this concept of non-identifiability by making reference to a model of market equilibrium.

$D = b_0 + b_1P + u$ (i) Demand function

$S = a_0 + a_1P + v$ (ii) Supply function

$D = S$ (iii) Clearance equation

Here, we cannot identify the two models in equation (i) and (ii) since in both, quantities $Q = f(P)$ Hence, we may not know exactly which function one is estimating – is it the demand or the supply function?

To be sure of the function one is estimating, there is the need to examine identification to enable us judge the precise equation. To achieve this objective, we need further information on factors affecting the demand and the supply function separately (shift factors), this will help us to state clearly the function one is dealing with.

In the demand function, other determining factors such as consumers income, taste, price of other commodities, fashion, etc, should be included to differentiate this function from the supply function determined by factors such as weather, technological changes etc. The stability of each of these functions depend, to a large extent, on the aforementioned factors.

We can state here that a model is identified if it is in unique mathematical form such that its estimates can be uniquely obtained from the given data. As a corollary, if a model is not identified, its parameter estimates relate to the model in question or to another model or both. Thus, for a set of simultaneous equations, the identification of the entire set requires that the model be complete. The completeness of such models implies that it should contain at least as many independent equations as the endogenous variable.

Referring to the model of market equilibrium previously indicated, the given model is complete, because it contained 3 equations and 3 endogenous variables, D, S, and P. However, they are not identified, as previously highlighted, with reasons).

We generalize by stating that a function belonging to a system of simultaneous equation is identified, if it has a unique statistical form, meaning that there is no other equation in the system which contains the same variable as the function in question.

Thus, we can state that:

(i) The identification of a system requires that each of the equations in the system be identified.

(ii) The parameters of the given equation be identified (note that in the previous example, the parameters of D (b_0 and b_1) and S (a_0 and a_0) cannot be statistically identified since one is not really sure of the function he is dealing with.

To establish identification, therefore, two rules are adhered to, namely,

(a) Order condition

(b) Rank condition

In econometrics, there are two possible situations of identifiability of any possible situation or equation:

(i) Under-identified equation

(ii) Identified equation

An under-identified equation exists if the mathematical form of the equation is not unique. A system is under-identified when one or more of its equations is not unique. A system is under-identified when one or more of its equations are under-identified. For such an equation or model, it is not possible to estimate all its parameters using any econometric technique.

An identified equation on the other hand has a unique statistical form. An identified equation may either come in the form of exact identification or overidentified condition. A given system, therefore, is identified if all its equations are identified.

For exact identification, we estimate such model using indirect least squares. In the case of overidentification, we use two stage least squares (2SLS), maximum likelihood methods.

Note that identification problem is peculiar to only those equations which contained a coefficient that needs to be estimated (From the given set of data).

3.4 Rules (Conditions) For Identification

Identification of a model can best be established by the examination of the specification of the structural model. In applying the identification rules, we either ignore the constant term or retain it, and include in the set of variables, a dummy variable which takes the value I. In this constant, we shall ignore the constant term.

For an equation (or model) to be identified the conditions are:

3.4.1 The Order Condition for Identification

This condition is based on the number of variables included and those excluded from a particular equation. This is a necessary condition, although not sufficient for complete identification. This implies that the order condition for identification may be satisfied if the equation is not identified. The condition states that: for an equation to be identified, the number of variables (endogenous and exogenous) excluded from it must be equal to or greater than the number of endogenous variables in the model less one.

Since the number of endogenous variables equal the number of equation in a complete model, the above condition may also be equivalently stated in this form: an equation is identified, if the total number of variables excluded from it but included in other equations is at least as great as the number of equations of the system less one.

Symbolically, the order condition for identification is given as:

$$K - M \geq G - 1$$

Where:

G = Total number of equations (Total number of endogenous and exogenous variables)

M = number of variables included in a particular equation.

So that $K - M$ = excluded variables; $G - 1$ = total number of equations less one.

Note that the order condition for identification is necessary for a relation to be identified but it is not sufficient for identification.

3.4.2 The Rank Condition for Identification

This condition maintains that in a system of G equations, any particular equation is identified if and only if it is possible to construct at least one non-zero determinant of order $G - 1$ from the coefficient of the variables excluded from that particular equation but contained in the other equation of the model.

In order to identify a particular equation, therefore, the following steps should be followed.

I. Write out the parameters of all the equations of the model in a separate table, bearing in mind to assign zero to each parameter of a variable excluded from a given equation. Make a table of coefficients for each of the equations.

II. Strike out the row of the coefficients of the equations which is being examined for identification. Also, strike out the columns in which a non-zero coefficient of the equation being examined appears. (By deleting the relevant row and columns, we are left with the coefficients of variables not in the other equations of the model).

III. Form the determinant(s) of order $(G - 1)$ and examine their values.

- If at least one of these determinants is non-zero, the equation is identified.
- If, however, all the determinants are zero, the equation is unidentified, and we make conclusion.

IV. Once the equation is identified, we then move to the order condition to determine the nature of identification.

Examples

(1) Examine the identification state of the following model.

$C_t = a_0 + a_1 Y_t - a_2 T_t + u_t$ ……………………………………….. (i)

$I_t = b_0 + b_1 Y_t + b_2 Y_{t-1} + b_3 R_t + u_2$ ………………………… (ii)

$Y_t = C_t + I_t + G_t$ …………………………………………………… (iii)

Solution

Re-writing the above equations, we have:

$-C_t + a_0 + a_1 Y_t - a_2 T_t + u_t = 0$ …………………………………… (i)

$-I_t + b_0 + b_1 Y_t + b_2 Y_{t-1} + b_3 R_t + u_2 = 0$ ………………………… (ii)

$-Y_t + C_t + I_t + G_t = 0$ ……………………………………………… (iii)

I. Expressing the coefficients of the variables in a tabular form, we have:

Equation	$\boldsymbol{C_t}$	$\boldsymbol{Y_t}$	$\boldsymbol{I_t}$	$\boldsymbol{T_t}$	$\boldsymbol{Y_{t-1}}$	$\boldsymbol{R_t}$	$\boldsymbol{G_t}$
(i)	-1	a_1	0	$-a_2$	0	0	0
(ii)	0	b_1	-1	0	b_2	b_3	0
(iii)	1	-1	1	0	0	0	1

II. To identify equation (i) delete the first row containing the coefficient of the equation and also delete the column of the table with non-zero coefficients of the equation to be identified. Thus, we are left with,

$$\begin{matrix} -1 & b_2 & b_3 & 0 \\ 1 & 0 & 0 & 1 \end{matrix}$$

i.e., coefficient of the excluded variables (in tabular form).

III. Determinants of order $G-1$ i.e., $3-1$ are:

$|D_1| = \begin{vmatrix} -1 & b_2 \\ 1 & 0 \end{vmatrix} = -b_2,$ $\qquad$ $|D_2| = \begin{vmatrix} -1 & b_3 \\ 1 & 0 \end{vmatrix} = -b_3,$ $\qquad$ etc.

Since we have at least a non-zero determinant, equation (i) is identified.

IV. For nature of identification, we use order condition as follows:

$K - M \geq G - 1$; where $K = 7$, $M = 3$, $G = 3$, so that,

$7 - 3 > 2 - 1 \Leftrightarrow 4 > 1$ (a case of overidentified equation)

Using similar procedure as for equation (i) done above, equations (ii) and (iii) are, respectively, identified as follows:

II. Coefficients of excluded variables are:

$$\begin{matrix} -1 & -a_2 & 0 \\ 1 & 0 & 1 \end{matrix}$$

III. and $G-1$ order determinants are:

$|D_1| = \begin{vmatrix} -1 & -a_2 \\ 1 & 0 \end{vmatrix} = a_2,$ $\qquad$ $|D_2| = \begin{vmatrix} -1 & 0 \\ 1 & 1 \end{vmatrix} = -1,$ $\qquad$ etc.

The equation (ii) is identified.

IV. Using order condition,

$$K - M \geq G - 1; \text{ where } K = 7, M = 4, G = 3, \text{ we have:}$$

$$7 - 4 > 3 - 1 \Leftrightarrow 3 > 2 \text{ (Overidentified equation).}$$

For equation (iii),

II. coefficients of excluded variables are:

$$\begin{matrix} -a_2 & 0 & 0 \\ 0 & b_2 & b_3 \end{matrix}$$

III. So that determinants of order $G - 1$ are:

$$|D_1| = \begin{vmatrix} -a_2 & 0 \\ 1 & b_2 \end{vmatrix} = -a_2 b_2, \qquad |D_2| = \begin{vmatrix} -a_2 & 0 \\ 1 & b_3 \end{vmatrix} = -a_2 b_3, \qquad \text{etc.}$$

Equation (iii) is identified.

IV. The nature of identification is given by:

$$K - M \geq G - 1; \text{ where } K = 7, M = 4, G = 3, \text{ so that,}$$

$$7 - 4 > 3 - 1 \Leftrightarrow 3 > 2 \text{ (Overidentified).}$$

2. Consider the structural model:

$$y_1 = 2y_2 - x_1 + 2x_3 - x_4 + z_t + u \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

$$y_2 = x_1 + 2y_1 - 2x_1 - y_1 + 2x_2 + v \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$y_3 = y_1 - y_2 + 3y_1 - 3x_2 - x_4 + w \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

Where, u, v and w are error terms. Examine the identification of the model.

By re-arranging, we have:

$$-y_1 + 2y_2 - x_1 + 2x_3 - x_4 + z_t + u = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

$$y_1 - y_2 - x_1 + 2x_2 \qquad\qquad + v = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$4y_1 - y_2 - y_3 - 3x_2 - x_4 \qquad + w = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

I. Expressing the coefficients of the variables in a tabular form, we have:

Equation	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{z_t}$
(i)	-1	2	0	-1	0	2	-1	1
(ii)	1	-1	0	-1	2	0	0	0
(iii)	4	-1	-1	0	-3	0	-1	0

Note: u, v and w are not variables, hence, they were excluded from the Table and subsequent computation.

For equation (i),

II. Coefficients of excluded variables after deleting the 1st row and columns with non-zero coefficients in the table, as in example 1, are:

$\boldsymbol{y_3}$	$\boldsymbol{x_2}$
0	2
-1	-3

III. Therefore, determinant of order $G - 1$ is:

$$|D| = \begin{vmatrix} 0 & 2 \\ -1 & -3 \end{vmatrix} = 2$$

Equation (i) is identified.

IV. From $K - M \geq G - 1$; $K = 8$, $M = 6$, $G = 3$. Therefore,

$8 - 6 = 3 - 1$, that is, $2 = 2$ (Exact identification)

For equation (ii),

II. coefficients of excluded variables after deleting the second row, and columns with non-zero coefficients in the table are:

$\boldsymbol{y_3}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{z_t}$
0	2	-1	1
-1	0	-1	0

III. Determinants of order $G - 1$ are:

$$|D_1| = \begin{vmatrix} 0 & 2 \\ -1 & 0 \end{vmatrix} = 2, \qquad |D_2| = \begin{vmatrix} 0 & -1 \\ -1 & -1 \end{vmatrix} = -1, \qquad \text{etc.}$$

The equation is identified.

IV. From $K - M \geq G - 1$; where $K = 8$, $M = 4$, $G = 3$. Therefore,

$8 - 4 > 3 - 1$, that is, $4 > 2$ and equation (ii) is, therefore, overidentified.

For equation (iii),

II. Coefficients of excluded variable as in (i) and (ii) are:

x_1	x_3	z_t
-1	2	1
-1	0	0

III. Determinants of order $G - 1$ are:

$$|D_1| = \begin{vmatrix} -1 & 2 \\ -1 & 0 \end{vmatrix} = 2, \qquad |D_2| = \begin{vmatrix} -1 & 1 \\ -1 & 0 \end{vmatrix} = 1, \qquad |D_3| = \begin{vmatrix} 2 & 1 \\ 0 & 0 \end{vmatrix} = 0.$$

Equation (iii) is identified.

IV. Using $K - M \geq G - 1$; $K = 8$, $M = 5$, $G = 3$, so that:

$8 - 5 > 3 - 1$, that is, $3 > 2$. Therefore, equation (iii) is overidentified.

Self-Assessment Exercise

Consider the following structural model, attempt the identification of the model:

$Y_1 = 2Y_2 – x_1 + 2x_3 – x_4 + z_t + u$ …………………………… (i)

$Y_2 = 2x_1 + 2Y_1 - 2X_1 – Y_t + 2x_2 + v$ …………………………… (ii)

$Y_3 = Y_1 – 2Y_2 + 3Y_1 – 3x_2 – x_4 + w$ …………………………… (iii)

3.5 Identification and Choice of Econometric Method

Identification determines basically the choice of econometric techniques used in finding the values of the parameters of our models.

- Thus, if a relationship or model is under identified, its parameters cannot be statistically estimated by any econometric technique.
- For an exactly identified model or relationship, the most appropriate method of estimation is the indirect least squares (ILS) method.
- If an equation is overidentified, two stage least squares (2SLS) method, or maximum likelihood method are used.

3.6 Significance of Identification

- Identification enables one to easily identify a given structural equation that is used in econometric analysis.
- It helps in the choice of estimation technique to be used in obtaining the values of the parameters.
- Identification gives the econometricians a focus. It enables him to know which of the relationships he is really dealing with.
- The nature of relationship between structural equations is revealed though identification of that model.
- Correct identification gives consistent and reliable estimates.

Self-Assessment Exercise

1. Discuss identification and the choice of econometric models.
2. List the uses of Identification in economics

3.7 Summary

In the unit, you have learned the meaning of identification and illustration on how to carry out identification of models. You have also learned identification in econometric models using different methods: the order condition, the rank condition and the importance of identification in applied econometrics.

3.8 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics, Prentice Hall, Englewood Cliffs, N.J.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods, McGraw-Hill, New York.

Christ, C. F. (1966). Econometric Models and Methods, John Wiley & Sons, New York.

Cramer, J. S. (1969). Empirical Econometric, North-Holland, Amsterdam. Davidson, James (2000). Econometric Theory, Blackwell Publishers, Oxford, U.K.

Desai, Meghnad. (1969) Applied Econometrics, McGraw-Hill, New York.

Dhrymes, Phoebus J. (1978) Introductory Econometrics, Springer-Verlag, New York.

Dutta, M. (1975). Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000) Econometric Analysis, 4th ed., Prentice Hall, Engle-wood cliffs, N. J.

Goldberger, Arthur S. (1998). Introductory Econometrics, Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series, 2nd ed., MIT Press, Cambridge, Massachusetts.

Hayashi, Fumio (1984) Econometric Methods, 3rd ed., McGraw-Hill, New York.

Hill, Carter, William Griffiths, and George Judge (2001) Undergraduate Econometrics, John Wiley & Sons, New York.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics, John Wiley & Sons, New York.

Katz, David A. (1962). Econometric Theory and Applications, Prentice Hall, Englewood Cliffs, N.J.

Klein, Lawrence R.: A (1974). Textbook of Econometrics, 2nd ed., Prentice Hall, Englewood Cliffs, N.J.

Kmenta, Jan (1986). Elements of Econometrics, 2nd ed., Macmillan, New York.

Maddala, G. S. (2001). Introduction to Econometrics, 3rd ed., John Wiley & Sons, New York.

Mark, Stewart B., and Kenneth F. Wallis (1981). Introductory Econometrics, 2nd ed., John Wiley & Sons, New York. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics, 2nd ed., North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, and Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries, Rutledge, New York.

Patterson, Kerry (2000) An Introduction to Applied Econometrics: A Time series Approach, St. Martin's Press, New York.

Rao, Potluri, and Roger leRoy Miller (1971) Applied Econometrics, Wadsworth, Belmont, California.

Theil, Henry (1971) Principles of Econometrics, John Wiley & Sons, New York.

Tintner, Gerhard (1965) Econometrics, John Wiley & Sons (Science ed.), New York.

Walters, A. A. (1968) An Introduction to Econometrics, Macmillan, London.

Wonnacott, R. J., and T. H. Wonnacott (1979) Econometrics, 2nd ed., John Wiley & Sons, New York.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics, South-Western College Publishing.

3.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Consider the following structural model, attempt the identification of the model:

$$y_1 = 2y_2 - x_1 + 2x_3 - x_4 + z_t + u \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

$$y_2 = 2x_1 + 2y_1 - 2x_1 - y_1 + 2x_2 + v \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$y_3 = y_1 - 2y_2 + 3y_1 - 3x_2 - x_4 + w \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

Solution

By re-arranging, we have:

$$-y_1 + 2y_2 - x_1 + 2x_3 - x_4 + z_t + u = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

$$y_1 - y_2 + 2x_2 \qquad\qquad\qquad + v = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$4y_1 - 2y_2 - y_3 - 3x_2 - x_4 \qquad + w = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

I. Expressing the coefficients of the variables in a tabular form, we have:

Equation	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{z_t}$
(i)	-1	2	0	-1	0	2	-1	1
(ii)	1	-1	0	0	2	0	0	0
(iii)	4	-2	-1	0	-3	0	-1	0

Note: u, v and w are not variables, hence, they were excluded from the Table and subsequent computation.

For equation (i),

II. Coefficients of excluded variables after deleting the 1st row and columns with non-zero coefficients in the table are:

$\boldsymbol{y_3} \quad \boldsymbol{x_2}$

0	2
-1	-3

III. Therefore, determinant of order $G-1$ is:

$$|D| = \begin{vmatrix} 0 & 2 \\ -1 & -3 \end{vmatrix} = 2$$

Equation (i) is identified.

IV. From $K - M \geq G - 1$; $K = 8$, $M = 6$, $G = 3$. Therefore,

$$8 - 6 = 3 - 1, \text{ that is, } 2 = 2 \text{ (Exact identification)}$$

For equation (ii),

II. coefficients of excluded variables after deleting the second row, and columns with non-zero coefficients in the table are:

y_3	x_1	x_3	x_4	z
0	-1	2	-1	1
-1	0	0	-1	0

III. Determinants of order $G-1$ are:

$|D_1| = \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} = -1$, $|D_2| = \begin{vmatrix} 0 & 2 \\ -1 & 0 \end{vmatrix} = 2$, $|D_2| = \begin{vmatrix} 0 & -1 \\ -1 & -1 \end{vmatrix} = -1$, etc.

The equation is identified.

IV. From $K - M \geq G - 1$; where $K = 8$, $M = 4$, $G = 3$. Therefore,

$8 - 3 > 3 - 1$, that is, $5 > 2$ and equation (ii) is, therefore, overidentified.

For equation (iii),

II. Coefficients of excluded variable as in (i) and (ii) are:

x_1	x_3	z
-1	2	1
0	0	0

III. Determinants of order $G - 1$ are:

$|D_1| = \begin{vmatrix} -1 & 2 \\ 0 & 0 \end{vmatrix} = 0$, $|D_2| = \begin{vmatrix} -1 & 1 \\ 0 & 0 \end{vmatrix} = 0$, $|D_3| = \begin{vmatrix} 2 & 1 \\ 0 & 0 \end{vmatrix} = 0$.

Equation (iii) is unidentified.

Therefore, the system of equations is under-identified.

2. Discuss identification and the choice of econometric models.

Identification is a problem associated with model formulation. Identification problem arises if a given model is not correctly specified. This is sequel to the violations of the assumption that the relationship to be estimated must have a unique mathematical form. If this is violated, the relationship to be estimated will contain the same variable as any other equation one is estimating. In the face of an unidentified model, the estimates of the parameters of the relationship between the variable measured in samples relate to the model in question or to another model or to a mixture of models.

For exact identification, we estimate such model using indirect least squares. In the case of overidentification, we use two-stage least squares (2SLS), maximum likelihood methods. For an under-identified system of equations, it is not possible to estimate all its parameters using any econometric technique.

3. List the uses of Identification in economics.

The uses of identification include:

a. Determining the uniqueness of the mathematical form of an equation.
b. Determining the econometric technique to be used the estimation of the parameters in a system of equations.

Unit 4: Two Stage Least Squares (2SLS)

Contents

4.1 Introduction

In this unit, we shall be treating the two stage least squares as an econometric way of solving simultaneous equation problems. The unit shall cover the meaning, the steps to take in solving two stage least squares, the assumptions, the properties, and applications of two stage least squares.

4.2 Learning Outcomes

The main objective of this unit is to expose you to the method of two stage least squares. Other specific objectives include:

- Assumptions of two stage least squares.
- Properties of two stage least squares.
- Applications of two stage least squares in econometrics.

4.3 Meaning of Two Stage Least Squares

This is a single equation method. It is applied to one equation of the system at a time. Two stage least squares provides satisfactory result for the estimates of the structural parameters in econometric analysis of models. 2SLS has been accepted widely as one of the most important methods of the single-equation technique for estimating over-identified models. This method is applied after a proper identification is carried out in a system of equations and clear establishment that the equation is over-identified.

Two stage least squares is one of the methods developed to solve problems of simultaneous equation bias in econometric estimation of parameters. This method is generally an extension of indirect least squares.

One of the sources of simultaneous equation bias is the existence of endogenous variable in the set of exogenous variables of the econometric models. The endogenous variables have a system component determined by the pre- determined exogenous variables in the equations and a random component. The random or stochastic component creates the dependence of the relative variables with the random term of the structural equation in simultaneous equation models. Two stage least squares method generally involves the application of ordinary least squares in two stages as follows:

1. Apply ordinary least squares to the reduced form equation so as to obtain estimates of the exact and random components of the endogenous variables in the equation.
2. The second stage requires that the endogenous variable appearing in the right hand side of the equation is replaced with their estimated value $Y_t = Y_t - U_t$ and OLS is applied to the transformed original equation to obtain the estimate of the structural parameters desired. All these can be done by the EViews software.

Self-Assessment Exercise

Discuss in full the meaning of two stage least squares in econometrics.

4.4 Assumptions of Two Stage Least Squares

1. According to Koutsoyiannis (1977), in two stage least squares, the random term of the reduced form equation should satisfy the stochastic assumption of the ordinary least squares. These are (a) zero mean, (b) constant variance (c) zero covariance. This allows the reduced form random term (vi) to possess the characteristic as stated above.
2. The error term of the reduced form equation (vi) should satisfy the OLS assumption. Beyond that it should be independent of the exogenous variables of the entire structural model.
3. The explanatory variables are not perfectly multicollinear and all macro- variables are properly and adequately aggregated.
4. It is assumed that the specification of the model is correct as far as the exogenous variables are concerned. That is, we assume knowledge of all the exogenous variables of the system, irrespective of the equations in which they appear in the

model. It is not necessary to know the model. It is not necessary to know the mathematical formulation of the whole system in all its ramifications. It is important to note that all the exogenous variables of the equation system are known.

5. The sample that would be used for econometric analysis is large enough. The number of observations is greater than the number of pre-determined variables in the structural equation system. If the sample size is small in comparison with the total number of the exogenous variables, it will not be possible to obtain significant estimates of the reduced coefficients when the ordinary least squares is applied in the first stage.

Self-Assessment Exercise

Discuss in detail the assumptions of two stage least squares in econometrics.

4.5 Properties of Two Stage Least Squares

1. The two stage least squares estimates are asymptotically efficient under the assumptions about the distribution of the disturbances.
2. Two stage least squares estimates are consistent, that is, their distribution collapses on the true parameter b and $n \to \infty$
3. In large samples, as $n \to \infty$, the bias tends to zero. That is, the two stage least squares estimates are asymptotically unbiased. This property can be established by deriving an explicit formula for the bias and examining the size of the bias as $n \to \infty$.
4. In two stage least squares small samples, the estimates are biased. This can be established using simple two variable model.

$$y_1 = b_2 y_2 + u$$

The transformed function becomes

$$y_1 = b_2 \hat{y}_2 + (u + b_2 v)$$

Where $\hat{y}_2 = x_1$. Now $\hat{y}_2$ is an exact function of x_1, which is uncorrelated with u. Furthermore, x_1 is, by assumption, uncorrelated with v. Therefore, $\hat{y}_2$ is uncorrelated with the composite random term $(u + b_2 v)$. However, $\hat{y}_2 = y_2 - v$, so that

$$y_1 = b_2(y_2 - v) + (u + b_2 v)$$

Despite the transformation, the simultaneous equation bias is not eliminated in small samples.

Self-Assessment Exercise

Using the necessary algebra discuss the properties of two stage least squares.

4.6 Merits and Demerits of Two Stage Least Squares

4.6.1 Merits of 2SLS

1. The method of 2SLS is simple in conception and in computation using EViews. It has yielded more satisfactory results than any of the other econometric methods. It is the most important technique for the estimation of over-identified models.
2. The 2SLS method is more general than the instrumental variables method, because it takes into account the influence on the dependent variables of all the predetermined variables of the system, while other methods choose only a small subset of predetermined variable and ignores the effects of the other exogenous variables in the model.
3. Two stage least squares is adequate for the estimation of over-identified equations.
4. Two stage least squares yields consistent estimates under conditions in which ordinary least squares method faces difficulty to provide result.

4.6.2 Demerits of 2SLS

1. Two stage least squares is extremely sensitive to specification errors. If we take into consideration the complexity of economic variable, consequently errors may easily

emerge in the estimation process.

2. Two stage least squares requires large number of observations. For models that are simultaneous in nature and have small observations, 2SLS may not produce efficient result.

Self-Assessment Exercise

Discuss the merits and demerits of two stage least squares.

4.7 Applications of Two Stage Least Squares in Model Estimation

The following data represent real gross domestic product and other macroeconomic variables use two stage least squares.

Gc	Pc	I	Y
4.3410	5.6809	5.2867	11.0291
4.5822	5.9086	5.4586	10.861
4.6594	5.9452	5.3721	10.833
4.6903	6.0669	5.0959	10.7541
4.6158	6.2385	4.6943	10.679
4.6921	6.3533	4.6300	10.747
4.6730	6.3720	4.8591	10.746
4.4913	6.7343	5.1270	10.713
4.8038	7.0478	5.2437	10.782
4.8736	7.1902	5.6418	10.826
4.9648	7.5048	6.0192	10.879
5.0693	7.6345	6.1136	10.9015
5.7784	8.2014	6.5384	10.906
6.1002	8.4745	6.8282	10.903
7.3645	8.6447	6.8901	10.881
7.6991	9.4423	7.1624	10.882
7.8199	9.8449	7.5021	10.900
8.0947	9.8059	7.6531	10.904
8.1124	9.9084	7.6272	10.899
7.5574	9.9195	7.5590	10.887
8.0651	9.9055	7.8922	10.915
8.0650	10.2781	7.9851	10.922
8.2117	10.6612	8.2554	10.913
8.1272	10.8766	8.7806	10.986
8.6588	10.9931	8.7526	11.062
8.8782	11.1877	8.6574	11.090
9.0998	11.3565	9.2863	11.125

9.3430	11.5929	9.4678	11.163
9.5862	11.8295	9.6493	11.196
9.8293	12.0660	9.8306	11.239
10.0723	12.3024	10.0119	11.289

The procedure is as follows, open EViews software, create a worksheet, import your set of data by browsing and clicking to upload your data to the workbook. Go to the "Quick" icon on the software toolbar and click "Estimate Equation" then input the variable names under "Estimate specification", and select "TSLS" under "Method" and click "OK".

Dependent Variable: Y
Method: Two-Stage Least Squares
Date: 02/15/16 Time: 15:10
Sample (adjusted): 1980 2010
Included observations: 31 after adjustments
Instrument specification: GC PC I
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.37425	0.043857	236.5449	0.0000
PC	-0.126292	0.025627	-4.928163	0.0000
I	0.236791	0.032429	7.301804	0.0000

R-squared	0.877318	Mean dependent var	10.92944
Adjusted R-squared	0.868555	S.D. dependent var	0.154170
S.E. of regression	0.055895	Sum squared resid	0.087479
F-statistic	100.1164	Durbin-Watson stat	1.077134
Prob(F-statistic)	0.000000	Second-Stage SSR	0.087479
J-statistic	0.493340	Instrument rank	4
Prob(J-statistic)	0.482441		

Interpretation

The model is a two-variable model with personal consumption (PC) and domestic investment (I). The coefficient for personal consumption (PC) is negative (-0.126292), while the coefficient of domestic investment (I) is positive (0.236791). The R^2 is 87% meaning that the variables have a strong relationship. It also means that 87% of the dependent variable is explained by the independent variables of the model.

Self-Assessment Exercise

A) Using appropriate examples, discuss two stage least squares, stating merits and demerits.

B) Using the data given as follows, attempt a two stage least squares regression and explain your findings.

Expt	Impt	Gdpl	Reerl
8.9817	8.836589	11.6626	1.961753
8.91746	8.834783	11.61802	1.951143
9.102807	8.676769	11.61455	1.957894
9.105067	8.558735	11.63029	1.983491
9.208795	8.542713	11.62521	1.988381
8.907678	8.403297	11.63919	1.98704
8.292528	8.682492	11.79029	2.015485
8.351442	8.64245	11.89003	2.024691
8.27505	8.479286	11.93402	2.02152
8.598038	8.410127	11.9517	2.033946
8.580214	8.410181	12.05439	2.050844
8.742171	8.574234	12.07746	2.052809
8.56821	8.703202	12.10237	2.045245
8.265851	8.560467	12.00902	1.971508
8.373988	8.388151	12.02281	1.959423
8.423975	8.339022	12.05155	1.927216
8.567049	8.417855	12.10023	1.971276
8.730424	8.562878	12.07639	1.972203
8.576768	8.51526	12.08532	1.978637
8.387281	8.519451	12.07948	1.974604
8.510927	8.616199	12.04034	1.957368
8.610895	8.692339	12.04819	1.963174
8.672677	8.717114	12.08598	1.973266
8.718001	8.863056	12.17816	1.997212
8.770742	8.867987	12.2375	2.004407
8.87736	8.847424	12.24986	2
9.007959	8.985416	12.2703	1.99987
9.078647	9.048413	12.32556	2.002166

9.205421	9.056459	12.36109	2.005781
9.035832	9.078515	12.32485	2.01013

4.8 Summary

In this unit you have learned what two stage least squares method is and the procedures for using it. You have also learned the assumptions, properties, applications, and the merits and demerits of using 2SLS in applied econometrics.

4.9 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric, North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics, Springer-Verlag, New York.

Dutta, M. (1975). Econometric Methods. Cincinnati: South-Western Publishing Company.

Greene, William H. (2000). Econometric Analysis (4th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Goldberger, Arthur S. (1998) Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). MIT Press, Cambridge, Massachusetts.

Hayashi, Fumio (1984) Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, & George Judge (2001). Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis. University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics. John Wiley & Sons, New York.

Katz, David A. (1962). Econometric Theory and Applications. Englewood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R. A. (1974). Textbook of Econometrics (2nd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Kmenta, Jan (1986). Elements of Econometrics (2nd ed.). Macmillan, New York.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., & Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). New York: John Wiley & Sons. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series. Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. New York: Routledge.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin's Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics, Wadsworth, Belmont, Califonia.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics. London: Macmillan.

Wonnacott, R. J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics. South-Western College Publishing.

4.10 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Discuss in detail the assumptions of two stage least squares in econometrics.

The assumptions of two stage least squares in econometrics include (Note that you need to discuss them in detail):

a. The random term of the reduced form equation should satisfy the stochastic assumptions of the ordinary least squares which are (a) zero mean, (b) constant variance (c) zero covariance.

b. The error term of the reduced form equation should satisfy the OLS assumption and should be independent of the exogenous variables of the entire structural model.

c. The explanatory variables are not perfectly multicollinear and all macro-variables are properly and adequately aggregated.

d. The specification of the model is correct as far as the exogenous variables are concerned.

e. The sample that would be used for econometric analysis is large enough.

2. Using the necessary algebra discuss the properties of two stage least squares.

a. The two stage least squares estimates are asymptotically efficient under the assumptions about the distribution of the disturbances.

b. Two stage least squares estimates are consistent, that is, their distribution collapses on the true parameter b and $n \to \infty$.

c. In large samples, as $n \to \infty$, the bias tends to zero. That is, the two stage least squares estimates are asymptotically unbiased. This property can be established by deriving an explicit formula for the bias and examining the size of the bias as $n \to \infty$.

d. In two stage least squares small samples, the estimates are biased. This can be established using simple two variable model.

$$y_1 = b_2 y_2 + u$$

The transformed function becomes:

$$y_1 = b_2 \hat{y}_2 + (u + b_2 v)$$

Where $\hat{y}_2 = x_1$. Now $\hat{y}_2$ is an exact function of x_1, which is uncorrelated with u. Furthermore, x_1 is, by assumption, uncorrelated with v. Therefore, $\hat{y}_2$ is uncorrelated with the composite random term $(u + b_2 v)$. However, $\hat{y}_2 = y_2 - v$, so that

$$y_1 = b_2(y_2 - v) + (u + b_2 v)$$

Despite the transformation, the simultaneous equation bias is not eliminated in small samples.

3. Discuss the merits and demerits of two stage least squares.

The merits of two stage least squares include:

a. Simplicity in conception and in computation using EViews.
b. It is more general than the instrumental variables method.
c. It is adequate for the estimation of over-identified equations.
d. It yields consistent estimates under conditions in which the OLS method faces difficulties.

The demerits include:

a. It is extremely sensitive to specification errors.
b. It requires a large number of observations.

Module 3

Unit 1: Simultaneous Equations

Contents

1.1 Introduction

These are models that exhibit two-way flow of influence among economic variables. It means an economic variable affects another economic variable or variables and in turn it is affected by other variables within the model. Generally, in simultaneous equation models there is more than one regression equation; one for each interdependent variable. This makes OLS inapplicable in estimation of equation parameters.

1.2 Learning Outcomes

At the end of this unit, you should be able to:

- Know the nature and characteristics of simultaneous equation models.
- Know about simultaneous equation bias.
- Find solutions to simultaneous equation problems.

1.3 Nature and Characteristics of Simultaneous Equation Models

In models that are estimated by OLS method, emphasis is on estimating the average value of Y subject to a fixed value of X variables. The process of determining cause – effect flows from X to Y. However, in simultaneous equation models, there is a two-way cause and effect. This take place when Y is determined by X and X is also determined by Y. There is a simultaneous relationship between Y and some of the X which makes the distinction between dependent and independent variables difficult to isolate. Therefore, in simultaneous equations, a set of variables are lumped together and determined simultaneously by the remaining set of variables in the model. Under such a situation, there is more than one equation – one for each of the mutually or jointly dependent variables in the equation system.

In simultaneous equation model estimation, the researcher cannot estimate the parameters of a single equation without considering the information provided by other equations in the model system. This is because of the joint dependence of the systems of equations. Similarly, if the equation systems are estimated using OLS by disregarding the dependence of the equation systems, the estimated variables in the model would be biased and inconsistent. As the sample size increases indefinitely, the estimators do not converge to their true value.

Consider the following simple Keynesian model of income determination.

$C_t = \beta_0 + \beta_1 Y_t + u_1$ $--------------$ (1)

$Y_t = C_t + I_t + u_2$ $----------------$ (2)

In equations (1) and (2), C_t and Y_t are mutually endogenous variables, u_1 and u_2 stochastic disturbance terms, the variables C_t and Y_t are both stochastic. It is, therefore, inconsistent to estimate the equations using OLS method.

Self-Assessment Exercise

Discuss the characteristics of simultaneous equation models.

1.4 Simultaneous Equation Bias

The methods of OLS cannot be applied to estimate single equation that is embedded in a system of simultaneous equations if one or more of the explanatory variables are correlated with the stochastic variable in the equation due to the fact that the estimators obtained would be inconsistent. Assume we are to estimate the parameters of the consumption function in equation (1).

Assume that $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_{t+j}) = 0$ (for $j \neq 0$) and $cov(I_t, u_t) = 0$ are the assumptions of the OLS regression model. It is important to show that Y_t and u_t in (2) are correlated and then prove that $\hat{\beta}_1$ is inconsistent estimator of β_1.

Self-Assessment Exercise 2

What is simultaneous equation bias?

1.5 Solution to Simultaneous Equation Problems

The correlation between Y_t and u_t can be proved as follows:

$C_t = \beta_0 + \beta_1 Y_t + u_t$ −−−−−−−−−−−−−− (3)

$Y_t = C_t + I_t$ −−−−−−−−−−−−−− (4)

Collect like terms after substituting equation (3) for C_t in equation (4).

$Y_t - \beta_1 Y_t = \beta_0 + I_t + u_t$ −−−−−−−−−−−−−−−− (5)

$Y_t(1 - \beta_1) = \beta_0 + I_t + u_t$ −−−−−−−−−−−−−−−−− (6)

$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t$ −−−−−−−−−−−−−−−−− (7)

$E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t$ −−−−−−−−−−−−−−−−−−−− (8)

Given that $E(u_t) = 0$ and I_t is exogenous since it is fixed in advance.

Subtract equation (8) from (7) and we then have:

$$Y_t - E(Y_t) = \frac{u_t}{1-\beta_1}$$

$$u_t - E(u_t) = u_t$$

$$cov(Y_t, u_t) = E\{Y_t - E(Y_t)\}[u_t - E(u_t)]$$

$$= \frac{E(u_t^2)}{1-\beta_1}$$

$$= \frac{\sigma^2}{1-\beta_1}$$

Because σ^2 is positive, the covariance between Y and u is different from zero. So Y_t and u_t are expected to be correlated which violates the assumption of the OLS that the disturbances are independent and uncorrelated with the explanatory variables (Gujarati, 2004).

Self-Assessment Exercise 3

Given the following hypothetical variables, attempt finding a solution using the simultaneous equation method.

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + u_{t1}$$

$$I_t = \rho_0 + \rho_1 r_t + u_{t2}$$

$$Y_t = C_t + I_t + G_t$$

where

C_t= consumption

Y_t= income

T_t= tax

r_t= interest rate

I_t= investment

G_t= government spending

1.6 Summary

In simultaneous equation models, the endogenous variable, otherwise known as regressand, in one equation usually appears as explanatory variable (regressor) in another equation. Because of this scenario the OLS method cannot be applied in solving the equation as estimates would become inconsistent and would not converge at their true mean. In this unit you have learn what simultaneous equation entails, simultaneous equation bias and the way to solve simultaneous equations in econometrics.

1.7 References/ Further Reading

Aigner, D. J, (1971). Basic Econometrics, Prentice Hall, Englewood Cliffs, N.J.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods, McGraw-Hill, New York.

Christ, C. F. (1966). Econometric Models and Methods, John Wiley & Sons, New York.

Cramer,J. S. (1969) Empirical Econometric, North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory, Blackwell Publishers, Oxford, U.K.

Desai, Meghnad. (1969). Applied Econometrics, McGraw-Hill, New York,

Dhrymes, Phoebus J. (1978). Introductory Econometrics, Springer-Verlag, New York.

Dutta, M. (1975) Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000) Econometric Analysis, 4th ed., Prentice Hall, Engle-wood cliffs, N. J.

Goldberger, Arthur S. (1998) Introductory Econometrics, Harvard University Press.

Harvey, A. C. (1990) The Econometrics Analysis of Time Series, 2nd ed., MIT Press, Cambridge, Massachusetts.

Hayashi, Fumio (1984) Econometric Methods, 3rd ed., McGraw-Hill, New York.

Hill, Carter, William Griffiths, and George Judge (2001) Undergraduate Econometrics, John Wiley & Sons, New York.

Hu, The-Wei (1973) Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics, John Wiley & Sons, New York.

Katz, David A. (1962). Econometric Theory and Applications, Prentice Hall, Engle-wood Cliffs, N.J.,

Klein, Lawrence R.: A (1974) Textbook of Econometrics, 2nd ed., Prentice Hall, Englewood Cliffs, N.J.

Kmenta, Jan (1986) Elements of Econometrics, 2nd ed., Macmillan, New York.

Maddala, G. S. (2001) Introduction to Econometrics, John Wiley & Sons, 3rd ed., New York.

Mark, Stewart B., and Kenneth F. Wallis (1981) Introductory Econometrics, 2nd ed., John Wiley & Sons, New York. A Halsted Press Book.

Madansky, A. (1976) Foundations of Econometrics, 2nd ed., North-Holland, Amsterdam,

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, and Mare Wuyts (1998) Econometrics and Data Analysis for Developing Countries, Routledge, New York.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach, St. Martin's Press, New York.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics, Wadsworth, Belmont, California.

Theil, Henry (1971). Principles of Econometrics. John Wiley & Sons, New York.

Tintner, Gerhard (1965). Econometrics, John Wiley & Sons (Science ed.), New York.

Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, London.

Wonnacott, R. J., & T. H. Wonnacott (1979) Econometrics (2nd ed.). John Wiley & Sons, New York.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics, South-Western College Publishing.

1.8 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Discuss the characteristics of simultaneous equation models.

The characteristics of simultaneous equation models include:

a. There is a two-way cause and effect, that is, Y is determined by X and X is also determined by Y.
b. The simultaneous relationship between Y and some of the X makes the distinction between dependent and independent variables difficult to isolate.
c. It is impossible to estimate the parameters of a single equation without considering the information provided by other equations in the model system.
d. If the equation systems are estimated using OLS by disregarding the dependence of the equation systems, the estimated variables in the model would be biased and inconsistent.

2. What is simultaneous equation bias?

Simultaneous equation bias occurs when an ordinary least squares regression is used to estimate an individual equation that is part of a system of simultaneous equations where one or more of the explanatory variables are correlated with the stochastic variable in the equation. The $\hat{\beta}_i$ found in this case is an inconsistent estimator of β_i and the bias will not disappear no matter how large the sample size is.

3. Given the following hypothetical variables, attempt finding a solution using the simultaneous equation method.

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + u_{t1}$$

$$I_t = \rho_0 + \rho_1 r_t + u_{t2}$$

$$Y_t = C_t + I_t + G_t$$

where

C_t= consumption

Y_t= income

T_t= tax

r_t= interest rate

I_t= investment

G_t= government spending

Solution

Let $C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + u_{t1}$ -------------------- (1)

$I_t = \rho_0 + \rho_1 r_t + u_{t2}$ ---------------------------- (2)

$Y_t = C_t + I_t + G_t$ ----------------------------- (3)

Collect like terms after substituting equation (1) for C_t and equation (2) for I_t in equation (3).

$$Y_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + \rho_0 + \rho_1 r_t + G_t + u_{t1} + u_{t2} \text{ -------- } (4)$$

$$Y_t - \beta_1 Y_t = \beta_0 - \beta_1 T_t + \beta_2 r_t + \rho_0 + \rho_1 r_t + G_t + u_{t1} + u_{t2} \text{ -------- } (5)$$

$$Y_t(1 - \beta_1) = \beta_0 - \beta_1 T_t + r_t(\beta_2 + \rho_1) + \rho_0 + G_t + u_{t1} + u_{t2} \text{ ------- } (6)$$

$$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{\rho_0}{1-\beta_1} - \frac{\beta_1}{1-\beta_1} T_t + \frac{(\beta_2+\rho_1)}{1-\beta_1} r_t + \frac{1}{1-\beta_1} G_t + \frac{1}{1-\beta_1} u_{t1} + \frac{1}{1-\beta_1} u_{t2} \text{ -- } (7)$$

$$E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{\rho_0}{1-\beta_1} - \frac{\beta_1}{1-\beta_1} T_t + \frac{(\beta_2+\rho_1)}{1-\beta_1} r_t + \frac{1}{1-\beta_1} G_t \text{ ------------- } (8)$$

Given that $E(u_t) = 0$ and T_t, r_t, and G_t are exogenous.

Subtract equation (8) from (7) and we then have:

$$Y_t - E(Y_t) = \frac{u_t}{1-\beta_1}$$

$$u_t - E(u_t) = u_t$$

$$cov(Y_t, u_t) = E\{Y_t - E(Y_t)\}[u_t - E(u_t)]$$

$$= \frac{E(u_t^2)}{1-\beta_1}$$

$$= \frac{\sigma^2}{1-\beta_1}$$

Because σ^2 is positive, the covariance between Y and u is different from zero. So Y_t and u_t are expected to be correlated which violates the assumption of the OLS that the disturbances are independent and uncorrelated with the explanatory variables.

Unit 2: Instrumental Variables Method

Contents

2.1 Introduction

The instrumental variables method was developed as a solution to the problem of simultaneous equation bias. It is applicable to simultaneous equation that is over-identified. This method attains the reduction of the dependence of the stochastic variable and the explanatory variables by using the appropriate exogenous variables as the instruments in model estimation. The model estimates using instrumental variables method are consistent when large samples are used, though inconsistent if small samples are used for estimation of simultaneous equation models.

2.2 Learning Outcomes

At the end of this unit you should be able to:

- ❖ Discuss the instrumental variables method as an alternative method of regression.
- ❖ Use the econometric software to estimate equations using the instrumental variables method.

2.3 Outline of Instrumental Variables Method

1. The researcher should choose the appropriate instrumental variables that will replace the endogenous variables that appear as explanatory variables in the model. The instrumental variable is an exogenous variable that is within the system of simultaneous equations. Such a variable should meet the following conditions.

a. It should be strongly correlated with the dependent variable that it will replace in the structural simultaneous equation system.

b. Instrumental variable should be truly exogenous hence uncorrelated with the stochastic variable (a) in the structural equation.

c. The instrumental variable should be correlated with exogenous variable already existing in the set of explanatory variables of the particular structural equation.

d. When more than one instrumental variable is to be used in the same structural equation, they must be the least correlated with each other. This is to avoid the problem of multicollinearity.

2. Multiply the structural equation through by the exogenous variables already existing in it, since these predetermined variables are their own instruments and sum the equation over all sample observations. This procedure will provide as many linear equations as there are unknown parameters. From the solution of these equations the estimates of the structural parameters are obtained (Koutsoyiannis, 1977).

Self-Assessment Exercise 1

Attempt an outline of the procedures of using the instrumental variables method in econometric analysis.

2.4 Assumptions of Instrumental Variables Method

The instrumental variables method uses the transformed OLS equation to provide solution to simultaneous equation problem. The instrumental variables method requires the multiplication of each of the OLS equations by an instrumental variable to determine the solution that is appropriate for the model equation. The assumptions of the instrumental variables method include the following.

1. The researcher should have a complete knowledge of some exogenous variables in other equations of the complete system that can be used as instruments in model estimation.
2. The new random term ZU should satisfy the ordinary least squares assumption of asymptotic behaviour.
3. The exogenous variable, that is, the instrumental variable used in estimation should not be perfectly multicollinear.
4. The structural function should be identified for it to be used effectively in estimation.
5. There should be proper aggregation of functional variables.

The instrumental variables method obeys the assumptions of the OLS method in addition to the additional assumptions stated above.

Self-Assessment Exercise 2

Discuss the assumptions of the instrumental variables method.

2.5 Properties of Instrumental Variables Method

1. When the sample used for research are small, the estimates of model equation by instrumental variables method will be biased.
2. Estimation of model equation with instrumental variables method using large sample size will produce estimates that would be consistent. However, for an estimate to be asymptotically consistent there are two conditions that should be met:
   a. The estimate should be asymptotically unbiased.
   b. The distribution of the values of the estimate should collapse on the true b, that is, it should be centred on the true structural parameter.
3. The estimates of instrumental variables method are inefficient. It means that the estimates do not possess minimum variance when compared with estimates obtained from other economic techniques of model estimation.

2.6 Merits and Demerits of Instrumental Variables Method

The merits of instrumental variables method include the following:

1. Instrumental variable method is a help to us to estimate models that are of simultaneous equation system.
2. Instrumental variable method provides a good alternative method of model estimation if the data is well aggregated.
3. Instrumental variable method gives a better result when data is quarterly and the

sample size is very large.

The demerits of instrumental variables method include the following:

1. The method of instrumental variables takes into account only part of the influence of the exogenous variables on the dependent variable. This is because the method includes only a certain exogenous variable and omission of other variables.
2. The selection of instrument in the estimation process is arbitrary. Therefore, the values of the estimates will differ depending on the choice of the model instrumental variable.
3. In practical estimation, the endogenous variable is related to more than one exogenous variable. The choice of a variable as the optimal instrument is quite challenging.

Self-Assessment Exercise 3

Discuss the properties, merits and demerits of the instrumental variables method.

2.7 Summary

In this unit you have learned about the instrumental variables method as another econometric method of estimating equations. This method requires that you can identify the required instruments when estimating the equations. You have also been taken through the subject matter from concept to the outline of instrumental variables, assumptions, properties, merits and demerits of instrumental variables method.

2.8 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics, Prentice Hall, Englewood Cliffs, N.J.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary, Addison-Wesley.

Bridge, J. I. (1971) Applied Econometrics, North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods, McGraw-Hill, New York.

Christ, C. F. (1966). Econometric Models and Methods, John Wiley & Sons, New York.

Cramer, J. S. (1969). Empirical Econometric, North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory, Blackwell Publishers, Oxford, U.K.

Desai, Meghnad. (1969). Applied Econometrics, McGraw-Hill, New York,

Dhrymes, Phoebus J. (1978). Introductory Econometrics, Springer-Verlag, New York.

Dutta, M. (1975). Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis, 4th ed., Prentice Hall, Engle-wood cliffs, N. J.

Goldberger, Arthur S. (1998). Introductory Econometrics, Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series, 2nd ed., MIT Press, Cambridge, Massachusetts.

Hayashi, Fumio (1984) Econometric Methods, 3rd ed., McGraw-Hill, New York.

Hill, Carter, William Griffiths, and George Judge (2001) Undergraduate Econometrics, John Wiley & Sons, New York,.

Hu, The-Wei (1973) Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics, John Wiley & Sons, New York.

Katz, David A. (1962). Econometric Theory and Applications, Prentice Hall, Engle-wood Cliffs, N.J.,

Klein, Lawrence R.: A (1974) Textbook of Econometrics, 2nd ed., Prentice Hall, Engle-wood Cliffs, N.J.

Kmenta, Jan (1986) Elements of Econometrics, 2nd ed., Macmillan, New York.

Maddala, G, S, (2001) Introduction to Econometrics, John Wiley & Sons, 3rd ed., New York.

Mark, Stewart B., and Kenneth F. Wallis (1981) Introductory Econometrics, 2nd ed., John Wiley & Sons, New York,. A Halsted Press Book.

Madansky, A. (1976) Foundations of Econometrics, 2nd ed., North-Holland, Amsterdam,

Mills, T.C. (1993) The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, and Mare Wuyts (1998) Econometrics and Data Analysis for Developing Countries, Routledge, New York.

Patterson, Kerry (2000) An Introduction to Applied Econometrics: A Time series Approach, St. Martin's Press, New York.

Rao, Potluri, and Roger leRoy Miller (1971) Applied Econometrics, Wadsworth, Belmont, Califonia.

Theil, Henry (1971) Principles of Econometrics, John Wiley & Sons, New York.

Tintner, Gerhard (1965) Econometrics, John Wiley & Sons (Science ed.), New York.

Walters, A.A, (1968) An Introduction to Econometrics, Macmillan, London.

Wonnacott, R.J., and T. H. Wonnacott (1979) Econometrics, 2nd ed., John Wiley & Sons, New York.

Wooldridge, Jeffrey M. (2000) Introductory Econometrics, South-Western College Publishing.

2.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Attempt an outline of the procedures of using the instrumental variables method in econometric analysis.

   The procedures of using the instrumental variables method include:

   a. The researcher should choose the appropriate instrumental variables that will replace the endogenous variables that appear as explanatory variables in the model.

   b. Multiply the structural equation through by the exogenous variables already existing in it, since these predetermined variables are their own instruments and sum the equation over all sample observations.

2. Discuss the assumptions of the instrumental variables method.

   The assumptions of the instrumental variables method include:

   a. The researcher should have a complete knowledge of some exogenous variables in other equations of the complete system that can be used as instruments in model estimation.

   b. The new random term ZU should satisfy the ordinary least squares assumption of asymptotic behaviour.

   c. The exogenous variable, that is, the instrumental variable used in estimation should not be perfectly multicollinear.

   d. The structural function should be identified for it to be used effectively in estimation.

   e. There should be proper aggregation of functional variables.

3. Discuss the properties, merits and demerits of the instrumental variables method.

The properties of the instrumental variables method include:

a. The estimates of model equation by instrumental variables method will be biased when the sample used for research are small.
b. Estimation of model equation with instrumental variables method using large sample size will produce estimates that would be consistent.
c. The estimates of instrumental variables method are inefficient.

The merits include:

a. Instrumental variable method is helpful in the estimation of models that are of simultaneous equation system.
b. It provides a good alternative method of model estimation if the data is well aggregated.
c. Instrumental variable method gives a better result when the sample size is very large.

The demerits are:

a. The method of instrumental variables takes into account only part of the influence of the exogenous variables on the dependent variable.
b. The selection of instrument in the estimation process is arbitrary.
c. In practical estimation, the endogenous variable is related to more than one exogenous variable.

Unit 3: Full Information Maximum Likelihood (FIML) Method

Contents

3.1 Introduction

According to Koutsoyiannis (1977), full information maximum likelihood method is a system method that is applied to all the equations of the model and yields estimates of all the structural parameters contemporaneously. It is a straightforward extension of the maximum likelihood method. In FIML, it is assumed that the complete specification of all the equations of the model is known. Similarly, the random variables of the structural equations are normally distributed with zero means. The random terms of the different equations are not intercorrelated. When they are independent their covariance will be zero.

3.2 Learning Outcomes

The objectives of this unit are:

- Know the estimation procedure of full information maximum likelihood method.
- Learn the properties and assumptions of full information maximum likelihood method.
- Learn the practical applications with on hand practical on EViews.

3.3 Estimation Procedure

1. Formulation of the likelihood function for the random variables of all the equations and for all sample values. Given the sample of size n, for each structural equation we have a set of n values for its random variable.
2. Estimation of the Jacobian which is the determinant for the error terms of all structural equations. The Jacobian is the determinant of partial derivatives of the transformation function solved for stochastic variable with respect to the endogenous variable.
3. Application of the transformation rule so as to obtain the likelihood function of the endogenous variable ($Y_1 - - - Y_G$) from the likelihood function of the random terms (μ_1, μ_2, μ_G).

4. Estimation of the partial derivatives of the likelihood function with respect to the structural parameters and the variances of all (G) random terms.
5. Equating of the partial derivatives of the likelihood function to zero and solution of the resulting system of equations for the unknown structural parameters and the variances of the random terms.
6. Estimation of the standard errors, variances of the parameter estimates.

Self-Assessment Exercise 1

Outline the estimation procedures of full information maximum likelihood method.

3.4 Properties of Full Information Maximum Likelihood Method

1. When the research samples are small, full information maximum likelihood method yield estimates that are biased.
2. If the research samples are large, full information maximum likelihood method estimates possess the property of efficiency.
3. When the research sample is large, FIML method estimates are consistent.

3.5 Assumptions of Full Information Maximum Likelihood Method

The basic assumptions of the full information maximum likelihood method of econometric estimates are as follows:

1. The full information maximum likelihood method assumes full information (full knowledge) of the complete specification of all the equations in the model. That is, information about the variables appearing in the model and the mathematical form of all the equation are well known to the researcher.
2. Full information maximum likelihood method assumes that the random disturbances of the various equations of the model are normally distributed with zero means and constant variances.
3. There should be the independence of the random terms of the various equations

that are adopted.

4. FIML assumes that the model is recursive.

Self-Assessment Exercise 2

Discuss the properties and the assumptions of full information maximum likelihood method in econometric analysis.

3.6 Merits and Demerits of Full Information Maximum Likelihood Method

3.6.1 Merits of FIML

1. Estimates of full information maximum likelihood method possess the characteristics of efficiency.
2. Full information maximum likelihood estimates are consistent.

3.6.2 Demerits of FIML

1. Estimates of full information maximum likelihood are biased for small samples.
2. The requirement of complete knowledge of the complete specification of the model is usually unattainable.
3. The method is cumbersome because of estimating equations which are nonlinear in the parameters; it is an expensive method in data, time, and money.
4. The method is sensitive to specification errors compared with other methods.

Self-Assessment Exercise 3

Discuss the merits and demerits of full information maximum likelihood method.

3.7 Applied Estimation

The data given below is derived from the performance of an economy. Estimate an equation using full information maximum likelihood method.

Gc	Pc	I	Y
4.3410	5.6809	5.2867	11.0291
4.5822	5.9086	5.4586	10.861
4.6594	5.9452	5.3721	10.833
4.6903	6.0669	5.0959	10.7541
4.6158	6.2385	4.6943	10.679
4.6921	6.3533	4.6300	10.747
4.6730	6.3720	4.8591	10.746
4.4913	6.7343	5.1270	10.713
4.8038	7.0478	5.2437	10.782
4.8736	7.1902	5.6418	10.826
4.9648	7.5048	6.0192	10.879
5.0693	7.6345	6.1136	10.9015
5.7784	8.2014	6.5384	10.906
6.1002	8.4745	6.8282	10.903
7.3645	8.6447	6.8901	10.881
7.6991	9.4423	7.1624	10.882
7.8199	9.8449	7.5021	10.900
8.0947	9.8059	7.6531	10.904
8.1124	9.9084	7.6272	10.899
7.5574	9.9195	7.5590	10.887
8.0651	9.9055	7.8922	10.915
8.0650	10.2781	7.9851	10.922
8.2117	10.6612	8.2554	10.913
8.1272	10.8766	8.7806	10.986
8.6588	10.9931	8.7526	11.062
8.8782	11.1877	8.6574	11.090
9.0998	11.3565	9.2863	11.125
9.3430	11.5929	9.4678	11.163
9.5862	11.8295	9.6493	11.196
9.8293	12.0660	9.8306	11.239
10.0723	12.3024	10.0119	11.289

The procedure is as follows, open EViews software, create a worksheet, import your set of data by browsing and clicking to upload your data to the workbook; go to "Quick" icon on the software toolbar, input the variable name, click method and a dialogue box opens from where you select “LILM”, then click OK.

Dependent Variable: Y
Method: LIML / K-Class
Date: 02/15/16 Time: 15:13
Sample (adjusted): 1980 2010
Included observations: 31 after adjustments
Covariance type: IV
Instrument specification: GC PC
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.38855	0.053305	194.8895	0.0000
I	0.157200	0.129927	1.209912	0.2364
PC	-0.064487	0.101029	-0.638301	0.5285

R-squared	0.850926	Mean dependent var	10.92944
Adjusted R-squared	0.840277	S.D. dependent var	0.154170
S.E. of regression	0.061615	Sum squared resid	0.106298
Durbin-Watson stat	0.673200	LIML min. eigenvalue	1.000000

Self-Assessment Exercise 4

Attempt a regression analysis using full information maximum likelihood method with data above.

3.8 Summary

In this unit you have learned about the full information maximum likelihood method, the estimation procedure, the basic assumptions, the properties of the full information maximum likelihood method, with an illustration on how to estimate equations using full information maximum likelihood method.

3.9 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics. North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometrics. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Blackwell Publishers, Oxford, U.K.

Desai, Meghnad. (1969). Applied Econometrics, McGraw-Hill, New York,

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods, South-Western Publishing Company, Cincinnati.

Greene, William H. (2000). Econometric Analysis (4th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). MIT Press, Cambridge, Massachusetts.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, and George Judge (2001). Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis, University Park Press, Baltimore.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics. New York: John Wiley & Sons.

Katz, David A. (1962). Econometric Theory and Applications. Englewood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R.: A (1974) Textbook of Econometrics (2nd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Kmenta, Jan (1986). Elements of Econometrics (2nd ed.). Macmillan, New York.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., and Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). New York: John Wiley & Sons. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam,

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series, Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries, Routledge, New York.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach, St. Martin's Press, New York.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics, Wadsworth, Belmont, Califonia.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics, Macmillan, London.

Wonnacott, R. J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, J. M. (2000). Introductory Econometrics, South-Western College Publishing.

Wooldridge, J. M. (2009). Introductory Econometrics: A Modern Approach (4th ed.). Singapore: Cengage Learning.

3.10 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Outline the estimation procedures of full information maximum likelihood method.

   The estimation procedures of full information maximum likelihood method include:

   a. Formulation of the likelihood function for the random variables of all the equations and for all sample values.
   b. Estimation of the Jacobian which is the determinant for the error terms of all structural equations.
   c. Application of the transformation rule so as to obtain the likelihood function of the endogenous variable ($Y_1 - - - Y_G$) from the likelihood function of the random terms (μ_1, μ_2, μ_G).
   d. Estimation of the partial derivatives of the likelihood function with respect to the structural parameters and the variances of all (G) random terms.
   e. Equating of the partial derivatives of the likelihood function to zero and solution of the resulting system of equations for the unknown structural parameters and the variances of the random terms.
   f. Estimation of the standard errors, variances of the parameter estimates.

2. Discuss the properties and the assumptions of full information maximum likelihood method in econometric analysis.

   The properties the full information maximum likelihood method include:

   a. The full information maximum likelihood method yields estimates that are biased when the research samples are small.
   b. If the research samples are large, full information maximum likelihood method estimates possess the property of efficiency.
   c. When the research sample is large, FIML method estimates are consistent.

The assumptions include:

a. The full information maximum likelihood method assumes full information (full knowledge) of the complete specification of all the equations in the model.
b. Full information maximum likelihood method assumes that the random disturbances of the various equations of the model are normally distributed with zero means and constant variances.
c. There is the independence of the random terms of the various equations that are adopted.
d. Full information maximum likelihood method assumes that the model is recursive.

3. Discuss the merits and demerits of full information maximum likelihood method.

The merits and of full information maximum likelihood method include:

a. Estimates of full information maximum likelihood method possess the characteristics of efficiency.
b. Full information maximum likelihood estimates are consistent.

The demerits include:

a. Estimates of full information maximum likelihood are biased for small samples.
b. The requirement of complete knowledge of the complete specification of the model is usually unattainable.
c. The method is cumbersome because of estimating equations which are nonlinear in the parameters; it is an expensive method in data, time, and money.
d. The method is sensitive to specification errors compared with other methods.

Unit 4: Three Stage Least Squares (3SLS)

4.1 Introduction

4.2 Learning Outcomes

4.3 The Meaning of Three Stage Least Squares

4.4 Assumptions of Three Stage Least Squares

4.5 Properties of Three Stage Least Squares Estimates

4.6 The Three Stages of Estimation

4.7 Summary

4.8 References/ Further Reading

4.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

4.1 Introduction

In this unit you will learn the meaning of three stage least squares, the assumptions and the stages of determining three stage least squares. The process is very simple, but you are required to go through the topic several times in order to have good knowledge of the three stage least squares.

4.2 Learning Outcomes

The objectives of this unit are:

- Know the meaning of three stage least squares.
- Discuss the assumptions and method of three stage least squares estimation.
- Learn the practical applications with on hand practical on EViews.

4.3 The Meaning of Three Stage Least Squares

Three stage least squares (3SLS) is an econometric method of estimation that is applied to all equations in the model at the same time and gives estimates of all the equations' estimates of parameters simultaneously. It is an extension of two stage least squares. Three stage least squares involves the application of least squares in three successive stages. It uses more information than the single equation methods, by taking into account the entire model structure with all restrictions that the structure imposes on the values of the equation parameters.

Usually, single equation techniques make use of only the variables appearing in the entire model, but ignore the restrictions set by the structure on the coefficients of other equations as well as the contemporaneous dependence of the random terms of the various equations (Koutsoyiannis, 1977). The random term of any equation u_i in simultaneous equation model is correlated with the error term of other equations.

4.4 Assumptions of Three Stage Least Squares

1. The model equation system is complete and correctly specified in all variables. This means the variables that appear in the model are completely known and their mathematical form well specified.
2. The error term of each equation is not serially correlated. The model specified should be free of autocorrelation.
3. The random variables of the various relations of the system are contemporaneously dependent.
4. Based on the nature of economic phenomena and the simplifications of economic models, the random variables become contemporaneously correlated, $E(u_i/u_j) \neq 0$. Because of restrictions, we include variables that are most important while those that are less useful in the model are taken care of by the random variable in the equation. When some variables are omitted from more relations of the equation system, it becomes inevitable that the random variable of these relations are correlated. That is why the application 3SLS becomes more appropriate.
5. The equation system is over-identified. When there are under-identified equations in the system, they can be made identifiable by changing their specification or they can be dropped from the system of equation. Similarly, identities and definitional equations can be dropped after using them to eliminate the relevant variables from the equation system.

Self-Assessment Exercise 1

Carefully discuss the assumptions of the three stage least squares in econometric analysis.

4.5 Properties of Three Stage Least Squares Estimates

1. The estimates of three stage least squares are biased but consistent.
2. Estimates of 3SLS are efficient. This is because more information is used in estimation.
3. 3SLS requires complete knowledge of the specification of the entire model.
4. 3SLS estimation process requires a large amount of data.
5. 3SLS is sensitive to specification errors in the equation system. A single specification error can affect all the equations in the model.
6. Though 3SLS is appropriate for econometric estimation of equation system, it is time consuming and tedious.

Self-Assessment Exercise 2

Outline the properties of three stage least squares.

4.6 The Three Stages of Estimation

Stage One

The first stage is to estimate the reduced form of all the equations of the system. This allows us to know the value of all the dependent variables in the system.

$$Y_1 = f(X_1, X_2, ---, X_n)$$
$$Y_2 = f(X_1, X_2, ---, X_n)$$
$$Y_Z = f(X_1, X_2, ---, X_n)$$

The first stage enables us to estimate the values of the endogenous variables Y_1, Y_2 and Y_Z.

Stage Two

The calculated values of the endogenous variables in stage one are substituted in the right hand side of the structural equation and apply least squares to the transformed equations. This give us the 2SLS of all the b's and Ys which are used for the estimation of the error terms of the various equations in the model. We then find the set of Z errors.

$$\ell_1, \ell_2, \ell_3, - - -, \ell_Z$$

Each corresponding to the error term (u) of the respective structural equation in the model.

The variance and covariance of the estimated error terms can be computed using the formula.

$$\sigma_{e_1 e_2} = \frac{\sum e_{1i} e_{2i}}{n} \qquad \sigma_{\ell_1 \ell_2} = \frac{\sum \ell_{1i} \ell_{2i}}{n}$$

$$\sigma_{e_1 e_3} = \frac{\sum e_{1i} e_{3i}}{n} \qquad \sigma_{\ell_1 \ell_3} = \frac{\sum \ell_{1i} \ell_{3i}}{n}$$

Stage Three

The variances and covariance of the error terms are used to obtain the transformation of the initial variables for the application of the generalized least squares.

Self-Assessment Exercise 3

Using algebraic method, discuss the three stage least squares.

4.7 Summary

In this unit you have learned about the three stage least squares method, the basic assumptions, the properties of the three stage least squares method, and the illustration on how to estimate equations using three stage least squares method.

4.8 References/ Further Reading

Aigner, D. J. (1971). Basic Econometrics. Englewood Cliffs, N.J.: Prentice Hall.

Berndt, Ernst R. (1991). The practice of Econometrics: Classic and contemporary. Addison-Wesley.

Bridge, J. I. (1971). Applied Econometrics. North-Holland, Amsterdam.

Chow, Gregory C. (1983). Econometric Methods. New York: McGraw-Hill.

Christ, C. F. (1966). Econometric Models and Methods. New York: John Wiley & Sons.

Cramer, J. S. (1969). Empirical Econometric. North-Holland, Amsterdam.

Davidson, James (2000). Econometric Theory. Oxford, U.K.: Blackwell Publishers.

Desai, Meghnad. (1969). Applied Econometrics. New York: McGraw-Hill.

Dhrymes, Phoebus J. (1978). Introductory Econometrics. New York: Springer-Verlag.

Dutta, M. (1975). Econometric Methods. Cincinnati: South-Western Publishing Company.

Greene, William H. (2000). Econometric Analysis (4th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Goldberger, Arthur S. (1998). Introductory Econometrics. Harvard University Press.

Harvey, A. C. (1990). The Econometrics Analysis of Time Series (2nd ed.). Cambridge, Massachusetts: MIT Press.

Hayashi, Fumio (1984). Econometric Methods (3rd ed.). New York: McGraw-Hill.

Hill, Carter, William Griffiths, & George Judge (2001). Undergraduate Econometrics. New York: John Wiley & Sons.

Hu, The-Wei (1973). Econometrics: An Introductory Analysis. Baltimore: University Park Press.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, & Tsoung-Chao Lee (1982). Introduction to the theory and practice of Econometrics. New York: John Wiley & Sons.

Katz, David A. (1962). Econometric Theory and Applications. Englewood Cliffs, N.J.: Prentice Hall.

Klein, Lawrence R.: A (1974). Textbook of Econometrics. (2nd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Kmenta, Jan (1986). Elements of Econometrics (2nd ed.). New York: Macmillan.

Maddala, G. S. (2001). Introduction to Econometrics (3rd ed.). New York: John Wiley & Sons.

Mark, Stewart B., & Kenneth F. Wallis (1981). Introductory Econometrics (2nd ed.). New York: John Wiley & Sons. A Halsted Press Book.

Madansky, A. (1976). Foundations of Econometrics (2nd ed.). North-Holland, Amsterdam.

Mills, T. C. (1993). The Econometric Modelling of Financial Time Series. Cambridge University Press.

Mukherjee, Chandan, Howard white, & Mare Wuyts (1998). Econometrics and Data Analysis for Developing Countries. New York: Routledge.

Patterson, Kerry (2000). An Introduction to Applied Econometrics: A Time series Approach. New York: St. Martin's Press.

Rao, Potluri, & Roger leRoy Miller (1971). Applied Econometrics. California: Wadsworth, Belmont.

Theil, Henry (1971). Principles of Econometrics. New York: John Wiley & Sons.

Tintner, Gerhard (1965). Econometrics. New York: John Wiley & Sons (Science ed.).

Walters, A. A. (1968). An Introduction to Econometrics. London: Macmillan.

Wonnacott, R.J., & T. H. Wonnacott (1979). Econometrics (2nd ed.). New York: John Wiley & Sons.

Wooldridge, Jeffrey M. (2000). Introductory Econometrics. South-Western College Publishing.

4.9 Possible Answers to Self-Assessment Exercise(s) Within the Content

1. Carefully discuss the assumptions of the three stage least squares in econometric analysis.

   The assumptions of the three stage least squares include:

   a. The model equation system is complete and correctly specified in all variables.
   b. The error term of each equation is not serially correlated.
   c. The random variables of the various relations of the system are contemporaneously dependent.
   d. The equation system is over-identified.

2. Outline the properties of three stage least squares.

   The properties of three stage least squares (3SLS) include:

   a. The estimates of three stage least squares are biased but consistent.
   b. Estimates of 3SLS are efficient.
   c. 3SLS requires complete knowledge of the specification of the entire model.
   d. 3SLS estimation process requires a large amount of data.
   e. 3SLS is sensitive to specification errors in the equation system.
   f. Though 3SLS is appropriate for econometric estimation of equation system, it is time consuming and tedious.

3. Using algebraic method, discuss the three stage least squares.

   The procedure is carried out in stages as follows:

   Stage One

   The first stage is to estimate the reduced-form of all the equations of the system. This allows us to know the value of all the dependent variables in the system.

$$Y_1 = f(X_1, X_2, ---, X_n)$$
$$Y_2 = f(X_1, X_2, ---, X_n)$$
$$Y_Z = f(X_1, X_2, ---, X_n)$$

The first stage enables us to estimate the values of the endogenous variables Y_1, Y_2 and Y_Z.

Stage Two

The calculated values of the endogenous variables in stage one are substituted in the right hand side of the structural equation and apply least squares to the transformed equations. This give us the 2SLS of all the b's and Ys which are used for the estimation of the error terms of the various equations in the model. We then find the set of Z errors.

$$\ell_1, \ell_2, \ell_3, ---, \ell_Z$$

Each corresponding to the error term (u) of the respective structural equation in the model.

The variance and covariance of the estimated error terms can be computed using the formula.

$$\sigma_{e_1 e_2} = \frac{\sum e_{1i} e_{2i}}{n} \qquad \sigma_{\ell_1 \ell_2} = \frac{\sum \ell_{1i} \ell_{2i}}{n}$$
$$\sigma_{e_1 e_3} = \frac{\sum e_{1i} e_{3i}}{n} \qquad \sigma_{\ell_1 \ell_3} = \frac{\sum \ell_{1i} \ell_{3i}}{n}$$

Stage Three

The variances and covariance of the error terms are used to obtain the transformation of the initial variables for the application of the generalized least squares.