NATIONAL OPEN UNIVERSITY OF NIGERIA

APPLIED ECONOMETRICS I

ECO 453

FACULTY OF SOCIAL SCIENCES

COURSE GUIDE

Course Writer/Developer **Dr Likita J. Ogba**
Department of Economics
Faculty of Social Sciences
University of Jos
likiogba@gmail.com

Course Content Editor Dr, Ganiyat A. Adesina-Uthman, acma,mnes,fifp
Department of Economics
Faculty of Social Sciences
National Open University of Nigeria

NATIONAL OPEN UNIVERSITY OF NIGERIA
Headquarters
Plot 91, Cadastral Zone, University Village,
Nnamdi Azikiwe Expressway, Jabi, Abuja.

Table of Contents Page

Introduction:……..2
What you will learn in this course:……3
Course Contents:……3
Course Aims:…….3
Course Objectives:…….3
Working through the Course:……4
Study Units:……..5
Assignment:…….....6
Tutor Marked Assignments (TMAs):……7
Final Written Examination:……..7
How to get the most from this course:…….7
Facilitators/Tutors and Tutorials ……8
Conclusion:……..12

INTRODUCTION

Welcome to ECO 453 Applied Econometrics I. The course is available for students in undergraduate Economics. The Course Applied Econometrics I (ECO453) 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 Faculty of 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 demonstrate 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:

- Apply simple and multiple regression to solve economics policy and theory.
- Evaluate Nonlinear Regression Models.
- Discuss Qualitative Response Regression Models.

- Discuss Panel Data Regression Models
- Examine Econometric Models
- Evaluate Autoregressive and Distributed-Lag Models and their applications in the Economy.
- 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 Eviews software this require that you have access to 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 fifteen units in this course, which should be studied carefully. Such units are as follows:

MODULE 1 INTRODUCTION TO ECONOMETRIC RESEARCH USING SOFTWARE

Unit 1 Meaning of Applied Econometric Research

Unit 2: Simple Linear Regression Model.

Unit 3: How To Run Time Series Data Using Eviews Software

Unit 4: Nonlinear Regression

Unit 5: Qualitative Response Regressions

MODULE 2 PANEL DATA AND PROBIT MODEL REGRESSION

Unit 1. The Probit Model

Unit 2: Autoregressive Process

Unit 3: Stationarity

Unit 4: Panel Data Regression Model

Unit 5: Fixed Versus Random Effects Panel Data

MODULE 3 DYNAMIC MODELS REGRESSION

Unit 1: Testing Fixed and Random Effects

Unit 2: Panel Data Estimation in Eviews.

Unit 3: Dynamic Models

Unit 4: Autoregressive Distributed Lag (ARDL) Model

Unit 5: ARDL level Relation

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.	Meaning of Applied Econometric Research	1	1
2.	Simple Linear Regression Model.		
3.	How To Run Time Series	1	1

	Data Using Eviews Software		
4.	Nonlinear Regression	1	1
5.	Qualitative Response Regressions	1	1
6.	The Probit Model	1	1
7.	Autoregressive Process	1	1
8.	Stationarity	1	1
9.	Panel Data Regression Model	1	1
10.	Fixed Versus Random Effects Panel Data	1	1
11.	Testing Fixed and Random Effects	1	1
12.	Panel Data Estimation in Eviews	1	1
13.	Dynamic Models	1	1
14.	Autoregressive Distributed Lag (ARDL) Model	1	1
15.	ARDL level Relation	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 sides 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 question which are 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 I (ECO45) exposes you to time series data utilization, model estimation and the issues involved in Applied Economics, such as application of simple and multiple regression to solve economics policy and theory, Nonlinear Regression Models, Qualitative Response Regression Models, Panel Data Regression Models, Dynamic Econometric Models: Autoregressive and Distributed-Lag Models and their applications in the Economy.. 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.

MODULE 1 INTRODUCTION TO ECONOMETRIC RESEARCH USING SOFTWARE

Unit 1 Meaning of Applied Econometric Research
Unit 2: Simple Linear Regression Model.
Unit 3: How To Run Time Series Data Using Eviews Software
Unit 4: Nonlinear Regression
Unit 5: Qualitative Response Regressions

UNIT 1 MEANING OF APPLIED ECONOMETRIC RESEARCH

CONTENTS

1. INTRODUCTION
2. OBJECTIVES
3.0 MAIN CONTENT
    3.1Meaning of Econometrics
    3.2 The Basic Tool for Econometrics
    3.3Methodology of Econometrics
    3.4 Stages of Applied Econometric Research.
    3.5 Concept of Economics and Econometrics Model.
    3.6 Properties of a good Econometric Model.
    3.7 Limitations and criticisms of Econometrics Research
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

Applied econometrics uses theoretical econometrics and real-world data for assessing economic theories, developing econometric models, analyzing economic history, and forecasting. 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 an applied research. In this unit you will be exposed to the basic assumptions with respect to the independent variables to be estimated.

2.0 OBJECTIVES

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

- Discuss meaning of Econometrics
- Analyze the basic tools for econometrics
- State methodology of econometrics
- Discuss the basic stages of econometric research
- Explain the assumptions of econometric models
- Make a critique of econometrics research

3.0 MAIN CONTENT

3.1Meaning of Econometrics

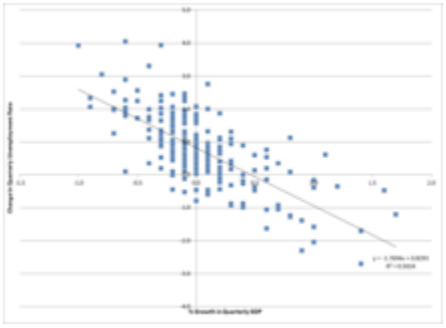
Econometrics is the application of statistical methods to economic data and is described as the branch of economics that aims to give empirical content to economic relations. More precisely, it is "the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference". An introductory economics textbook describes econometrics as allowing economists "to sift through mountains of data to extract simple relationships". The first known use of the term "econometrics" (in cognate form) was by Polish economist Paweł Ciompa in 1910. Jan Tinbergen is considered by many to be one of the founding fathers of econometrics. Ragnar Frisch is credited with coining the term in the sense in which it is used today.

The basic tool for econometrics is the multiple linear regression model. Econometric theory uses statistical theory and mathematical statistics to evaluate and develop econometric methods. Econometricians try to find estimators that have desirable statistical properties including unbiasedness, efficiency, and consistency. Applied

econometrics uses theoretical econometrics and real-world data for assessing economic theories, developing econometric models, analysing economic history, and forecasting.

3.2 The Basic Tool for Econometrics

The basic tool for econometrics is the multiple linear regression model. In modern econometrics, other statistical tools are frequently used, but linear regression is still the most frequently used starting point for an analysis. Estimating a linear regression on two variables can be visualised as fitting a line through data points representing paired values of the independent and dependent variables.



Okun's law representing the relationship between GDP growth and the unemployment rate. The fitted line is found using regression analysis.

The basic tool for econometrics is the multiple linear regression model. In modern econometrics, other statistical tools are frequently used, but linear regression is still the most frequently used starting point for an analysis. Estimating a linear regression on two variables can be visualised as fitting a line through data points representing paired values of the independent and dependent variables.

For example, consider Okun's law, which relates GDP growth to the unemployment rate. This relationship is represented in a linear regression where the change in unemployment rate (^ Unemployment) is a function of an intercept (β_0) a given value of GDP growth

multiplied by a slope coefficient β_1 and an error term, U:

$\hat{}$ Unemployment = $\beta_0 + \beta_1$ Growth + U.

The unknown parameters β_0 and β_1 can be estimated. Here β_1 is estimated to be −1.77 and β_0 is estimated to be 0.83. This means that if GDP growth increased by one percentage point, the unemployment rate would be predicted to drop by 1.77 points. The model could then be tested for statistical significance as to whether an increase in growth is associated with a decrease in the unemployment, as hypothesized. If the estimate of β_1 were not significantly different from 0, the test would fail to find evidence that changes in the growth rate and unemployment rate were related. The variance in a prediction of the dependent variable (unemployment) as a function of the independent variable (GDP growth) is given in polynomial least squares.

3.3 Methodology of Econometrics

Applied econometrics uses theoretical econometrics and real-world data for assessing economic theories, developing econometric models, analysing economic history, and forecasting. Econometrics may use standard statistical models to study economic questions, but most often they are with observational data, rather than in controlled experiments. In this, the design of observational studies in econometrics is similar to the design of studies in other observational disciplines, such as astronomy, epidemiology, sociology and political science. Analysis of data from an observational study is guided by the study protocol, although exploratory data analysis may be useful for generating new hypotheses. Economics often analyses systems of equations and inequalities, such as supply and demand hypothesized to be in equilibrium. Consequently, the field of econometrics has developed methods for identification and estimation of simultaneous-equation models. These methods are analogous to methods used in other areas of science, such as the field of system identification in systems analysis and control theory. Such

methods may allow researchers to estimate models and investigate their empirical consequences, without directly manipulating the system.

One of the fundamental statistical methods used by econometricians is regression analysis. Regression methods are important in econometrics because economists typically cannot use controlled experiments. Econometricians often seek illuminating natural experiments in the absence of evidence from controlled experiments. Observational data may be subject to omitted-variable bias and a list of other problems that must be addressed using causal analysis of simultaneous-equation models.

3.4 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 on the stability of the estimates in term of their sensitivity to changes in the size of the model even outside the given

sample data within the period.

SELF ASSESSMENT EXERCISE

Itemize the basic stages of an econometric research

3.5 Concepts of Economic and Econometric Model

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. 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.

3.6 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. 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. Example a model should predict with accuracy price, demand, supply exchange rates, market behaviour, and any other practical situation. This feature often help 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, over identified or under identified.

SELF ASSESSMENT EXERCISE

Discuss the properties of a good econometric model

3.7 Limitations and criticisms of Econometrics Research

Like other forms of statistical analysis, badly specified econometric models may show a spurious relationship where two variables are correlated but causally unrelated. In a study of the use of econometrics in major economics journals, McCloskey concluded that some economists report p-values (following the Fisherian tradition of tests of significance of point null-hypotheses) and neglect concerns of type II errors; some economists fail to report estimates of the size of effects (apart from statistical significance) and to discuss

their economic importance. She also argues that some economists also fail to use economic reasoning for model selection, especially for deciding which variables to include in a regression.

In some cases, economic variables cannot be experimentally manipulated as treatments randomly assigned to subjects. In such cases, economists rely on observational studies, often using data sets with many strongly associated covariates, resulting in enormous numbers of models with similar explanatory ability but different covariates and regression estimates. Regarding the plurality of models compatible with observational data-sets, Edward Leamer urged that "professionals ... properly withhold belief until an inference can be shown to be adequately insensitive to the choice of assumptions".

4.0 CONCLUSION

In this unit which is the first unit in Module 1 of this course, you learnt the meaning and basic stages of applied econometric research. These stages of econometric model research are necessary for effective analysis of any model building. You need to thoroughly master these steps because their violation leads to several econometric problems that we shall study in this course.

5.0 SUMMARY

Econometric theory uses statistical theory and mathematical statistics to evaluate and develop econometric methods. Econometricians try to find estimators that have desirable statistical properties including unbiasedness, efficiency, and consistency. An estimator is unbiased if its expected value is the true value of the parameter; it is consistent if it converges to the true value as sample size gets larger, and it is efficient if the estimator has lower standard error than other unbiased estimators for a given sample size. Ordinary least squares (OLS) is often used for estimation since it provides the BLUE or "best linear unbiased estimator" (where "best" means most efficient, unbiased estimator) given

the Gauss-Markov assumptions. When these assumptions are violated or other statistical properties are desired, other estimation techniques such as maximum likelihood estimation, generalised method of moments, or generalised least squares are used. Estimators that incorporate prior beliefs are advocated by those who favour Bayesian statistics over traditional, classical or "frequentist" approaches

6.0 TUTOR MARKED ASSIGNMENT

1 What do you understand by the phrase “econometric research”?.

2 Carefully discuss the properties of an econometric model.

3. Make a critique of econometric research

7.0 REFERNCES/ FURTHER READING

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

Berndt, Ernst R.(1991) The practice of Econometrics: Classic and contemporary, Addison-Wesley,
Goldberger, Arhur S. (1998) Introductory Econometrics, Harvard University Press.

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

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

Oosterbaan, R.J. (1994), Frequency and Regression Analysis. In: H.P.Ritzema (ed.), Drainage Principles and Applications, Publ. 16, pp. 175-224, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. ISBN 90-70754-33-9 . Download as PDF

Oosterbaan, R.J. (2002). Drainage research in farmers' fields: analysis of data. Part of project “Liquid Gold” of the International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. Download as PDF

Wooldridge, J. M. (2009) Introductory Econometrics A modern Approach, Cengage Learning Singapore 4th Edition

UNIT 2: SIMPLE LINEAR REGRESSION MODEL.

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
    3.1 Linear Regression Approach
    3.2 Parameter Estimation Strategies
        3.2.1 Sources of Deviations in Parameters and Models
        3.2.2 The uses of Random Variable in Models
    3.3 Assumptions of Linear Stochastic Regression Model
        3.3.1 Assumptions with respect to the random variable
        3.3.2 Assumption of error term with respect to the explanatory Variable
        3.3.3 Assumption in Relation to the Explanatory Variable.
    3.3 Numerical Estimation of Parameters
        3.3.1 Algebraic Method.
        3.3.2 Quantitative Method Illustration.
    3.4 Eviews Software Applications.
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the preceding unit you learnt the meaning of econometrics research approach. That laid the foundation for the present unit in which you will learn linear regression model. 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:

$$Y = bo + b1x \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1).$$

This function shows the positive linear relationship between quantity supply, Y and price of the commodity, X. The dependent variable in this model is the quantity supply, 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 supply, but not the other way round. From the above equation (…1), the parameters are bo and b1, and we need to obtain numerical value of these parameters. The left hand variable Y is variously referred to as the endogenous variables, the regressand, the dependent variable or the explained variable. Similarly, the right hand variable X is variously described as exogenous variables, the regressor, the independent variable or the explanatory variable.

2.0 OBJECTIVES

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

- Explain linear regression approach
- Discuss parameter estimation procedures.
- Evaluate the assumptions of the stochastic variable.
- Analyze the assumptions of the explanatory variables.
- Attempt algebraic and software estimation of simple regression.

3.0 MAIN CONTENT

3.1 Linear Regression Approach

Linear regression is a special case of regression analysis, which tries to explain the relationship between a dependent variable and one or more explanatory variables. Mathematical functions are used to predict or estimate the value of the dependent variables. In linear regression, these functions are linear. Linear regression was the first

type of regression analysis to be studied rigorously. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters. What is more, the statistical properties of the resulting estimators are easier to determine. Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- Linear regression can be used to fit a predictive model to a set of observed values (data). This is useful, if the goal is prediction, or forecasting, or reduction. After developing such a model, if an additional value of X is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.
- Given a variable y and a number of variables X_1, ..., X_p that may be related to y, linear regression analysis can be applied to quantify the strength of the relationship between y and the X_j, to assess which X_j has no relationship with y at all, and to identify which subsets of the X_j contain redundant information about y.

Linear regression models are often fitted using the least squares approach. Other ways of fitting exist; they include minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or minimizing a penalized version of the least squares loss function as in ridge regression. The least squares approach can also be used to fit models that are not linear. As outlined above, the terms "least squares" and "linear model" are closely linked, but they are not synonymous

3.2 Parameter Estimation Strategies

The parameters of this model are to be estimated using ordinary least square (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 O.L.S. 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 = bo + b1x 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) responsible for the change. When this is represented on a graph, the pairs of observation (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.

3.2.1 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 ground 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 behavioral 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 behavior 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 central bank of Nigeria and 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

Outline sources of deviations in parameters estimation.

3.2.2 The Uses of Random variable in Models

The inclusion of a random variable usually denoted by U, into the econometric function help in overcoming the above stated sources of errors. The U's 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 the X. Thus, the variation in Y could be explained in terms of explanatory variable X and the random disturbance term U.

That is $Y = bo + b1x + ui$ …… (ii)

Where Y = variation in Y; bo + b1x = systematic variation, Ui = random variation. Simply put, variation in Y = explained variation plus unexplained variation. Thus, Y = bo + b1x + Ui 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:

3.3 Assumptions of the Linear Stochastic Regression Model

3.3.1 Assumptions With Respect To 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) Ui is a random variable, this means that the value which Ui 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 (ui) denoted by U is zero. By this assumption, we may express our regression in (ii) above as Yi = bo + b1x.

(iii) (iii) The variance of ui is constant in each period. That is, Var (ui) = E (ui)2 = $\delta(ui)2 = \delta 2u$ which is constant. This implies that for all values of *x*, the *U's* will show the same dispersion about their mean. Violation of this assumption makes the *Us* heteroscedastic.

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

(v) (v) The covariance of *ui* and uj = 0. i ≠ j. This assumes the absence of autocorrelation among the *ui*. 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

3.3.2 Assumptions in terms of the relationship between '*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, c*ov*. *X*u = 0.

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

3.3.3 Assumptions in relation to the explanatory variable(s) alone.

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. Xi Xj = 0. i ≠ 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.

3.4 Numerical Estimation of Parameters

The following procedures are used in finding numerical values of the parameters bo and bi.

From the true relationship $Yi = bo + bi\ x + u$, …….(i)

and the estimated relationship $\hat{Y} = bo + b1\ x + ei$,

the residual $ei = Yi - \hat{Y}$ ……..(ii)

Squaring the residual and summing over n, gives:

$\sum ei2 \sum (Yi-\hat{Y})2$ or $\sum(Yi - bo + b1\ x)^2$ …… (iii)

The expression in (2) is to be minimized with respect to bo and b1 respectively.

Thus $\frac{\delta \sum e^2}{\delta \sum bo} \sum(Yi-bo-b1X)(-1) = 2\sum[Yi-bo-b1X]$

Thus $\frac{\delta \sum e^2}{\delta \sum b1} = -2 \sum(Yi-bo-b1X)Xi$

Setting each of the partial derivatives to Zero, and dividing each term by -2 the OLS estimate of bo and b1 could be written in the form:

$\sum Yi - boN - b1\sum xi = 0$ …… (V)

$\sum YXi - bo \sum X - b1\sum xi^2 = 0$ ……. (Vi)

From which

$\sum Yi = boN + bi\sum xi$ …… (Vii)

$\sum YiXi = bo\sum X+b1\sum xi^2$ …… (Viii)

The two equations (vii) and (viii) are the normal equation of the regression model.

Using Crammer's rule, the values of the parameters bo and b1 are respectively:

$bo = \sum Y\sum X2 - \sum X\sum YX$

$N\sum X^2 - (\sum X)^2$ ……(iX)

$b1 = \sum Y\sum X - \sum X \sum Y$

$N\sum X^2 - (\sum X)^2$ …..(X)

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

$bo = Y - b1X$ …… (Xi)

$b1 = \sum XY$ ……..(Xii)

$\sum X2$

3.4.2 Quantitative Method Illustration

Example 1

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

No	Yi (Quantity)	Xi (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 expression for b_o and b_i in ix and x as well as that of (x_1) and (x_{11}) lead us to reproduce the above Table as seen in table 2.

Table 1 excel

NO	Y1 QUANTITY	X1 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

Table 2: Worksheet for the estimation of the supply function for Rice (R).

	Y	X	X^2	XY	y=Y-Y	x=X-X	x^2	Xy	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	∑Y=0	∑X=0	$\sum X_2$ = 18	∑XY =81	∑e = 0	∑e= 281.5	

Where e = residual, $Y = b_o + b_{1x}$

From the Table 2.

$$\bar{Y} = \sum y/n = 55.\ X = \sum x/n = 8$$

Therefore, using upper case letters,

$$b_o = \frac{440\,(530) - (64)\,(3607)}{8(530) - (64)^2} = 19$$

$$b_1 = \frac{8\,(3601) - (64)\,(440)}{8(530) - (64)^2} = 4.5$$

Similarly, using lower case letters

$$b_1 = \frac{\sum XY = 81}{\sum X_2 \quad 18} = 4.5$$

3.5 EVIEWS SOFTWARE APPLICATIONS

The linear regression can be solved using the software package using the following steps. Create an excel worksheet type the data for quantity (QUA)) and price (PRI). Open your Eviews software, go to file, create worksheet, copy the data on excel worksheet paste it on the Eviews worksheet already created go to file, import data. Go to 'Quick' on the tool bar scroll to estimate equation and click on it a dialogue box opens, type the respective quantity and price click ok, the output is as follows:

Dependent Variable: QUA
Method: Least Squares
Date: 11/03/17 Time: 08:43
Sample: 1 8
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

To copy the estimated result from Eviews to words format, highlight the output (result), click copy and select HTML from the drop down menu. The software estimated model for the data can be presented in a line form as follows: Qua = 19 + 4.5Pri, Qua represent the quantity while Pri represent the price. This 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 b0 and b_1, the estimated regression lines or equation is got by substituting these values into $Y = bo + b_1X$ and this gives $Y = 19 + 4.5X$.

Thus, given the values of x_1 (1 = 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. $Y = b0 + b_1xi$

This is also the equation of the line with intercept b0 and slope b_1. Note that $b_1 = \delta Y/\delta X$. Therefore, price elasticity = $b_1 = Xi/Yi$.

Taking the mean of Xi and Yi, we have average elasticity, $ep = b_1. Xi/Y$. Therefore $ep = 4.5(8)/55 = 0.65$.

5.0 CONCLUSION

In econometrics, linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. In linear regression, the relationships are modeled using linear predictor

functions whose unknown model parameters are estimated from the data. Such models are called linear models

5.0 SUMMARY

In this unit which is the second in our Module 1 you learnt the practical applications of econometrics as well as basic stages of applied econometric research, the assumptions of econometric model estimation. Of course, 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. The next unit showcases how to run time series data in Eviews software.

6.0 TUTOR MARKED ASSIGNMENT

1. Outline the basic assumptions of linear regression. Given the following data estimate the multiple regression equation using CPI as dependent variable and MS as independent variable.

Year	CPI	MS
2000	99	0.04
2001	96.3	3.7
2002	92.3	7.7
2003	90.1	9.9
2004	89.2	10.8
2005	89.1	10.7
2006	89.2	10.4
2007	89.9	10.1
2008	89.9	9.5
2009	90.4	9.6
2010	90.9	9.2

Estimate the regression equation and interprete your findings.

7.0 REFERNCES/ FURTHER READING

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

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

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

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

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

Oosterbaan, R.J. (1994), Frequency and Regression Analysis. In: H.P.Ritzema (ed.), Drainage Principles and Applications, Publ. 16, pp. 175-224, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. ISBN 90-70754-33-9 . Download as PDF

Oosterbaan, R.J. (2002). Drainage research in farmers' fields: analysis of data. Part of project “Liquid Gold” of the International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. Download as PDF

Wooldridge, J. M. (2009) Introductory Econometrics A modern Approach, Cengage Learning Singapore 4th Edition

UNIT 3: HOW TO RUN TIME SERIES DATA USING EVIEWS SOFTWARE

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 History of Time Series Data Analysis
- 3.2 Stochastic Process
- 3.3 Stationary and Nonstationary Variables
- 3.4 Weakly Stationarity and Strict Stationarity
- 3.5 Illustrative Example of How to Run Time Series Data in Eviews 9
  - 3.5.1 Unit Root Test
  - 3.5.2 Co-integration
  - 3.5.3 Impulse Response

4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the preceding unit, you learnt simple linear regression. The stage is now set for you to learn the processes of running time series data in Eviews software. The history of methodological developments in econometrics appears to be broadly classified into Traditional Econometrics and Modern Econometrics in literature. The former often refers to the use of economic theory and the study of contemporaneous relationships to explain relationships among dependent variables. It is concerned with building structural models, understanding of the structure of an economy and making statistical inference. In contrast, Modern Econometrics is based on exploiting the information that can be gotten from a variable that is available through the variable itself. It is concerned with building efficient models which forecasts the time path of a variable very well. Essentially, the term “modern econometrics” refers to Time Series Analysis.

2.0 OBJECTIVES

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

- Discuss history of Time Series Data Analysis
- Explain the Stochastic Process
- Evaluate Stationary and Nonstationary Variables
- Determine weakly Stationarity and Strict Stationarity
- Demonstrate how to run Time Series Data in Eviews 10.0 or 11.0
- Calculate Unit Root Test
- Calculate Co-integration
- Plot Impulse Response

3.0 MAIN CONTENT

3.1 History of Time Series Data Analysis

The analysis of time-series is of particular interest to many groups, such as;

(a) *Macroeconomists*: studying the behaviour of national and international economies.

(b) *Finance economists*: analyzing the stock market.

(c) *Agricultural economists*: predicting supplies and demands of agricultural products.

Regression models with time series data often exhibit some special characteristics designed to capture their dynamic nature. For instance, Including lagged values of the dependent variable or explanatory variables as regressors, or considering lags in the errors, can be used to model dynamic relationship. Regression of the current value of series on its past values can be used in forecasting. An important assumption for using time series data in regression analysis is that the series have a property called Stationarity. However, many economic variables are nonstationary and the consequences of nonstationary variables for regression modeling are profound. Therefore, the aim of this section is to examine the various data generating processes, the concept of stationarity and how to graphically examine the properties of a time series.

3.2 Stochastic Process

A random or stochastic process is a collection of random variables ordered in time. A Stochastic random variable could be ***continuous in time*** or ***discrete in time***. Most economic data are collected at discrete points in time, for instance, GDP

Notation

Let y denote the random variable at time t.

Example

If we let *y* represents GDP, then denotes the third observation on GDP. The economic variable observed over time is random because we cannot perfectly predict it. The econometric model generating is called a **Stochastic** or **random process**. A sample of observed values is called a particular **realization** of the stochastic process. It is one of many possible paths that the stochastic process could have taken.

3.3 Stationary and Nonstationary Variables

A time series is stationary if its ***mean*** and ***variance*** are constant over time and if the ***covariance*** between two values from the series depends only on the length of time separating the two values, and not on the actual times at which the variables are observed. This can be summarized as follows:

(i) $E(Y_t) = U$ ---------- (ia

The mean is constant over time. It exhibits ***mean reversion*** in that it fluctuates around a constant long-run mean. Shocks are temporary; over time, the effect of the shocks will dissipate and the series will revert to its long-run mean.

(ii) $V(Y_t) = q^2$ ---------- (ib)

The variance is constant over time. It has a finite variance that is time-invariant.

(iii) $Cov\ (Y_t,\ Y_{t+s}) = Cov\ (Y_t,\ Y_{t-s})$ ----------(ic)

The covariances are not constant over time. The covariances depend on the lag length, not time.

3.4 Weakly Stationarity and Strict Stationarity

A process is defined to be weakly stationary (or covariance stationary) if for all t, equations (1a), (1b), and (1c) holds. That is, the mean, variance and autocovariances are independent of time. Unlike weakly stationarity, strict stationarity is stronger because it requires that the whole distribution is unaffected by a change in time horizon, not just first and second order moments. Under joint normality assumption, the distribution is ompletely characterized by first and second order moments, and strict stationarity and weak stationarity are equivalent.

SELF – ASSESSMENT EXERCISE

Outline stochastic process

3.5 Illustrative Example of How to Run Time Series Data in Eviews 9

The following table contains time series data for;

(a) Real U.S. gross domestic product (**GDP**) – a measure of aggregate economic production.

(b) Annual inflation rate (**inf**) – a measure of changes in the aggregate price level.

(c) Federal funds rate (**f**) – Interest rate on overnight loans between banks.

(d) 3-year Bond rate (**b**) – Interest rate on functional asset to be held for three years.

Date	gdp	inf	f	b	date	gdp	inf	f	b
1984q1	3807.4	9.47	9.69	11.19	1990q1	5708.1	13.35	8.25	8.38
1984q2	3906.3	10.03	10.56	12.64	1990q2	5797.4	13.55	8.24	8.62
1984q3	3976	10.83	11.39	12.64	1990q3	5850.6	12.1	8.16	8.25
1984q4	4034	11.51	9.27	11.1	1990q4	5846	11.91	7.74	7.76
1985q1	4117.2	10.51	8.48	10.68	1991q1	5880.2	10.65	6.43	7.27
1985q2	4175.7	9.24	7.92	9.76	1991q2	5962	9.33	5.86	7.25
1985q3	4258.3	8.37	7.9	9.29	1991q3	6033.7	10.29	5.64	6.89
1985q4	4318.7	7	8.1	8.84	1991q4	6092.5	9.12	4.82	5.84
1986q1	4382.4	6.16	7.83	7.94	1992q1	6190.7	7.32	4.02	5.77
1986q2	4423.2	5.9	6.92	7.18	1992q2	6295.2	6.53	3.77	5.78
1986q3	4491.3	5.37	6.21	6.66	1992q3	6389.7	5.61	3.26	4.68
1986q4	4543.3	4.95	6.27	6.48	1992q4	6493.6	4.42	3.04	5
1987q1	4611.1	5.69	6.22	6.52	1993q1	6544.5	3.54	3.04	4.64
1987q2	4686.7	6.62	6.65	7.72	1993q2	6622.7	3.28	3	4.41

1987q3	4764.5	6.47	6.84	8.15	1993q3	6688.3	2.59	3.06	4.32
1987q4	4883.1	6.4	6.92	8.29	1993q4	6813.8	3.25	2.99	4.41
1988q1	4948.6	6.4	6.66	7.58	1994q1	6916.3	4.43	3.21	4.9
1988q2	5059.3	6.73	7.16	8.1	1994q2	7044.3	4.25	3.94	6.2
1988q3	5142.8	7.65	7.98	8.59	1994q3	7131.8	4.17	4.49	6.56
1988q4	5251	8.57	8.47	8.75	1994q4	7248.2	4	5.17	7.4
1989q1	5360.3	9.3	9.44	9.38	1995q1	7307.7	3.52	5.81	7.27
1989q2	5453.6	10.2	9.73	8.92	1995q2	7355.8	3.67	6.02	6.25
1989q3	5532.9	11.11	9.08	8.07	1995q3	7452.5	3.29	5.8	5.96
1989q4	5581.7	11.92	8.61	7.86	1995q4	7542.5	3.45 5.	72	5.58
1996q1	7638.2	3.04	5.36	5.38	2002q1	10498.7	2.83	1.73	3.75
1996q2	7800	1.6	5.24	6.29	2002q2	10601.9	3.05	1.75	3.77
1996q3	7892.7	1.62	5.31	6.36	2002q3	10701.7	3.05	1.74	2.62
1996q4	8023	1.28	5.28	5.94	2002q4	10766.9	3	1.44	2.27
1997q1	8137	2.17	5.28	6.19	2003q1	10888.4	3.15	1.25	2.07
1997q2	8276.8	3.69	5.52	6.42	2003q2	11008.1	3.1	1.25	1.77
1997q3	8409.9	4.1	5.53	6.01	2003q3	11255.7	2.71	1.02	2.2
1997q4	8505.7	4.4	5.51	5.78	2003q4	11416.5	2.69	1	2.38
1998q1	8600.6	3.89	5.52	5.46	2004q1	11597.2	2.48	1	2.17
1998q2	8698.6	3.84	5.5	5.57	2004q2	11778.4	2.35	1.01	2.98
1998q3	8847.2	4.03	5.53	5.11	2004q3	11950.5	2.84	1.43	2.92
1998q4	9027.5	4.22	4.86	4.41	2004q4	12144.9	2.62	1.95	3.05
1999q1	9148.6	4.71	4.73	4.87	2005q1	12379.5	2.8	2.47	3.61
1999q2	9252.6	5.09	4.75	5.35	2005q2	12516.8	3.05	2.94	3.73
1999q3	9405.1	4.57	5.09	5.71	2005q3	12741.6	2.61	3.46	3.98
1999q4	9607.7	4.5	5.31	6	2005q4	12915.6	2.62	3.98	4.37
2000q1	9709.5	5.1	5.68	6.56	2006q1	13183.5	2.7	4.46	4.58
2000q2	9949.1	4.48	6.27	6.52	2006q2	13347.8	2.81	4.91	4.98
2000q3	10017.5	5.39	6.52	6.16	2006q3	13452.9	2.9	5.25	4.87
2000q4	10129.8	6.04	6.47	5.63	2006q4	13611.5	3.14	5.25	4.65
2001q1	10165.1	5.15	5.59	4.64	2007q1	13789.5	2.9	5.26	4.68
2001q2	10301.3	4.73	4.33	4.43	2007q2	14008.2	2.32	5.25	4.76
2001q3	10305.2	3.8	3.5	3.93	2007q3	14158.2	2.18	5.07	4.41
2001q4	10373.1	2.95	2.13	3.33	2007q4	14291.3	1.85	4.5	3.5

date	**gdp**	**inf**	**f**	**b**
2008q1	14328.4	1.45	3.18	2.17
2008q2	14471.8	1.59	2.09	2.67
2008q3	14484.9	1.58	1.94	2.63
2008q4	14191.2	1.54	0.51	1.48
2009q1	14049.7	1.65	0.18	1.27
2009q2	14034.5	2.09	0.18	1.49
2009q3	14114.7	2.32	0.16	1.56
2009q4	14277.3	2.59	0.12	1.39

To Upload the Data Above

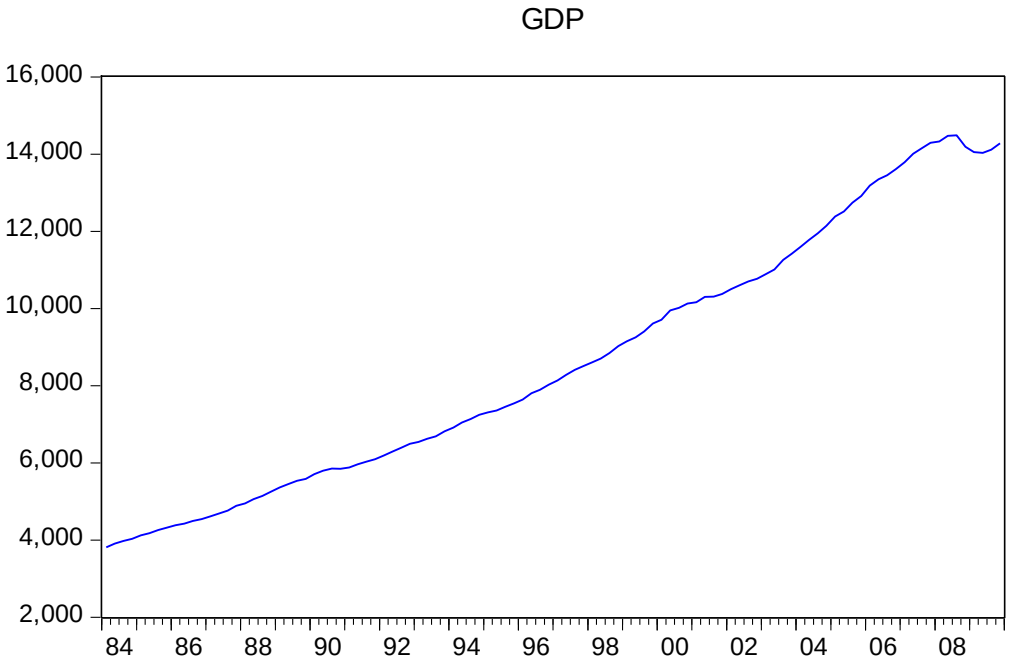
- Open E-view 9
- Open new workbook
- Set dates and frequency (i.e. annually, quarterly, etc)
- Cllick ok
- Enter **data** on the space below the **menu bar**

- Upload your data (by **copy & paste** or by **import**)
- Select your variables either **individually** or as a **group**

Individual Selection

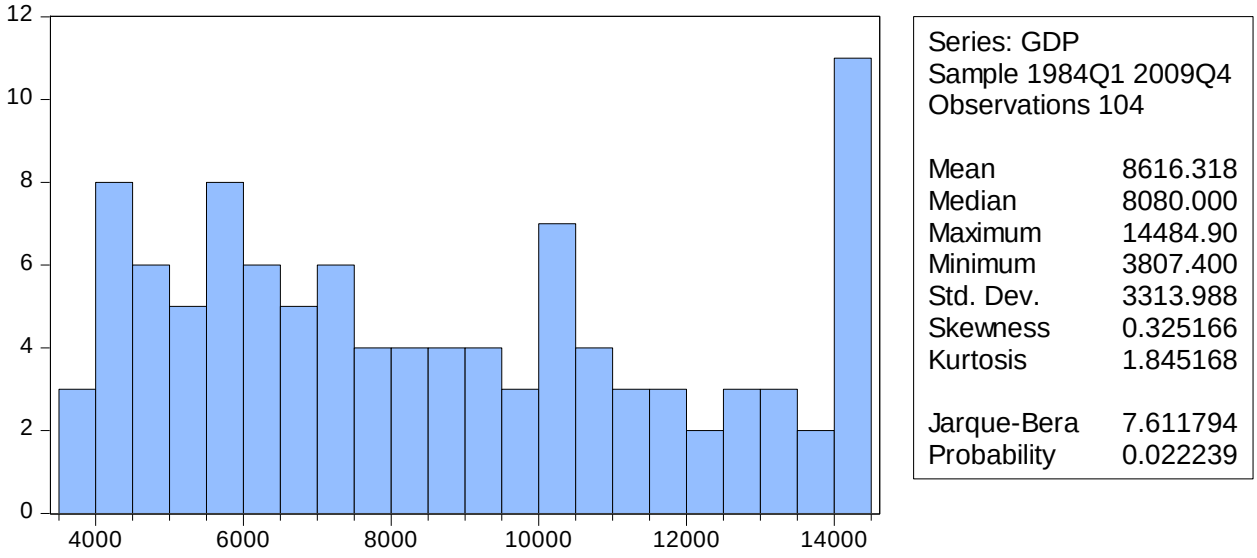
- Double-Click on **GDP**
- From view, click on **graph** or, on **Descriptive Statistics & Tests**, then on **Histogram & Statistics**, as the case may be

  Graph Result is shown below:



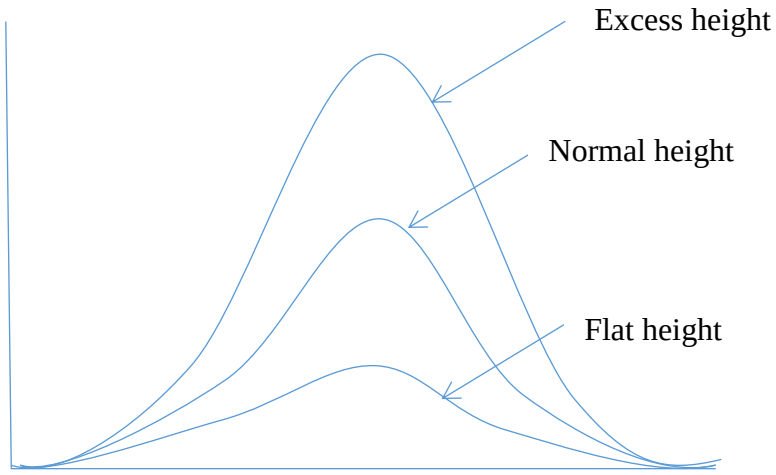
3.2.1 Descriptive Statistics & Tests

The descriptive statistics are shown below:



Interpretation of the Result

- Look out for the values of:
  - (i) **Mean:** Remember that the **mean** is normally surposed to be zero. But it will not be so at this junction.
  - (ii) **Skewness:** By the rule, it should not be significantly far from the mean. Ideally, it should be symmetry (i.e. it should be close to zero, though not exactly equal to zero
  - (iii) **Kurtosis:** Kurtosis shows the height of the graph.



For a normal distribution, we except the Kurtosis to be around **3.0**.

If ***K*> 3** → excess height, above average height

If ***K*< 3** → flat height, not of average height

(iv) **Jarque-Bera:** J-Bera is a **perfect test** for normality. It is a combination of both Skewness and Kutosis. The normal standard or **Decision Rule**:

If ***J-B < 5.99*** → We do not reject the $\#_o$ (i.e. There is normality)

If ***J-B > 5.99*** → We reject the $\#_o$ (i.e. there is no normality)

(v) **Probability:**The rule of thumb is tested against a particular level of significance of interest (i.e. at 10%, 5% or 1% significant levels). You must pick and stick to one.

- **Note:**Remember that all that have been done for **GDP** above will be repeated for all the variables in the model, one-by-one.

3.5.1 Unit Root Test

Again, the **unit root test** has to be conducted individually for each variable as we did above. For example:

- Go back to **workfile**
- Double-click on GDP, to activate the GDP data
- Go to **view**
- Go to **Unit Root Test:** To determine whether to test at **level** or not and whether to include **trend**,or **trend and intercept** or not, first include **trend** in your original data and upload it or you **generatetrend** directly on the **workfile.**

(i) To include trend in your original data, simply create a new column named **trend** and list, starting from 1 to the end of the data point (e.g. since we are having 104 observations in our example, the new column will have 1 to 104 listed numerically). Then, upload your data again including the new column (**trend**).

- Double-Click on **GDP** again to bring up the data
- Click on **process**
- Click on **generate by equation** and specify the equation:

  GDP c trend
- Click ok

(ii) You can as well generate the trend directly inside the **workfile** by:

- Go back to the **workfile**
- Click **Quick → Estimate Equation** and specify the following:

  GDP c @trend
- Make your decision on the results generated as shown below:

Dependent Variable: GDP
Method: Least Squares
Date: 11/04/17 Time: 01:03
Sample: 1984Q1 2009Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3011.630	88.52972	34.01829	0.0000
@TREND	108.8289	1.485118	73.27964	0.0000

R-squared	0.981359	Mean dependent var	8616.318
Adjusted R-squared	0.981177	S.D. dependent var	3313.988
S.E. of regression	454.6741	Akaike info criterion	15.09608
Sum squared resid	21086314	Schwarz criterion	15.14694
Log likelihood	-782.9963	Hannan-Quinn criter.	15.11668
F-statistic	5369.905	Durbin-Watson stat	0.024903
Prob(F-statistic)	0.000000		

From the above regression results, both **constant** and **trend** are highly significant going by the P-values. This indicates that you must run your **Unit Root test** with a constant and a trend.

- Go back to **workfile**
- Activate **GDP** data
- Go to **View → Unit Root Test**
- Specify appropriately:
  - Test Type = ADF;
  - Test for Unit Root in = level **or** 1st difference **o**r 2nd difference
  - Include in Test Equation: none **or** intercept **or** trend & intercept. In our example, we must include a constant and a trend.
  - Automatic criterion = Schwartz Information Criterion
  - Maximum lag = specify appropriately. Here we use 4 because we are using quarterly data. We can as well increase it to 8 or leave it at maximum

length of 12. If we use annual data, it will be sufficient if we use 1 or 2 lag length.

Result at 1st Difference

Null Hypothesis: GDP has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=4)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.095714	0.5417
Test critical values:	1% level	-4.050509	
	5% level	-3.454471	
	10% level	-3.152909	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GDP)
Method: Least Squares
Date: 11/04/17 Time: 01:27
Sample (adjusted): 1984Q3 2009Q4
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP(-1)	-0.026978	0.012873	-2.095714	0.0387
D(GDP(-1))	0.547567	0.082538	6.634088	0.0000
C	110.9831	39.45962	2.812575	0.0059
@TREND(1984Q1)	3.192185	1.421521	2.245612	0.0270

R-squared	0.362195	Mean dependent var	101.6765
Adjusted R-squared	0.342671	S.D. dependent var	71.73841
S.E. of regression	58.16253	Akaike info criterion	11.00279
Sum squared resid	331522.2	Schwarz criterion	11.10573
Log likelihood	-557.1421	Hannan-Quinn criter.	11.04447
F-statistic	18.55068	Durbin-Watson stat	2.145990
Prob(F-statistic)	0.000000		

Interpretation

(i) The value of ***t-statistics*** must be greater than a **specified significant level** you have pre-selected, either at 1%, 5% or 10% levels respectively (in absolute terms). You must pick and stick to one.

(ii) Probability value must be significant – very close to zero.

From above result, the *t*-statistics is not greater than any of the levels of significance, and, at the same time, not significant, giving the probability value of 0.03then we can conclude that the variable **GDP** is not stationary at level and therefore, we must include 1st difference in our specification.

- Go back to **workfile**
- Activate **GDP** data
- Go to Unit Root Test
- Include 1st difference in your specification

Result

Null Hypothesis: D(GDP) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=4)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.339944	0.0001
Test critical values:	1% level	-4.050509	
	5% level	-3.454471	
	10% level	-3.152909	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(GDP,2)
Method: Least Squares
Date: 11/04/17 Time: 01:38
Sample (adjusted): 1984Q3 2009Q4
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(GDP(-1))	-0.448096	0.083914	-5.339944	0.0000
C	33.17336	13.58997	2.441019	0.0164
@TREND(1984Q1)	0.242518	0.202716	1.196340	0.2344

R-squared	0.223798	Mean dependent var	0.624510
Adjusted R-squared	0.208117	S.D. dependent var	66.47037
S.E. of regression	59.15054	Akaike info criterion	11.02702
Sum squared resid	346379.9	Schwarz criterion	11.10422
Log likelihood	-559.3780	Hannan-Quinn criter.	11.05828
F-statistic	14.27202	Durbin-Watson stat	2.115713
Prob(F-statistic)	0.000004		

Interpretation

The result shows that *t*-statistic is greater than any of the levels of significance and is equally significant. We can then conclude that the variable **GDP** is stationary at 1st difference.

Determination of the lag length

After noting that the variable is not stationary at level but at 1st difference, you then determine the lag length. Note that you can only determine the lag length at point of stationarity. Determination of lag length can be done either by using:

(i) Schwarz Criterion or

(ii) Akaike information Criterion **Just stick to one.**

With Schwarz Criterion, you determine the maximum lag length by increasing the length to generate new results and compare it with the value generated at the point the variable became stationary. Any lag length included that generates a value less than Schwarz Criterion (11.10422 in this case) is an indication that we have not got the optimal lag length.When you increase the lag length, you want to determine whether the Schwarz Criterion is improving or not.

- On the result, click **view**
- Go back to **Unit Root Test**
- Increase the lag length to say, 8
- **Run** the test again

Note that the result is not significantly different from the last one because it is already stationary. When dealing with annual series, the rule of thumb is that we use a lag length of say, 1 or 2. But, when using quarterly data set, we use 4 or 8 or 12, etc. but note that when you include more lag length, you are simply reducing your data points and hence, you lose more information.

Remember that to de-trend, simply means that you do not include trend in your specification or delete it from your data set.

Unit Root Test - Report

Variable	Level	Difference	Order of stationarity
GDP	-2.09514	-5.3339	I(1)
INF	-2.472095	-5.122112	I(1)

3.5.2 Co-integration

The idea of Co-integration is just to see whether there is long-run relationship among the variables in a model (e.g. if 2 variables, say, consumption (*C*) and income (*Y*), are I(1), it means that one cannot predict a long-run relationship between them because the two variables cannot converge to their mean – no mean reversal.

Question: If I have variables with different behaviours (one is I(0) and the other trending), can I use them to predict a long-run relationship?. You'll need to difference the trended variable, to make it stationary.

- When 2 variables are trending together in the same direction they may not necessarily have long-run relationship; it may be their trend that is moving together. This is why we must subject them to some diagnostic tests.

Co-integration Test

There two ways of performing the co-integration test:

1. Augmented Dickey-Fuller Test
2. Engle-Granger Co-integrated Test (Johansen Integrated Test)

1. ADF Test

To perform the Residual Test – ADF Test:

Residual (Error term):

- Once you open **E-view,** it will automatically generate a default **constant** and default **Residual**. However, any regression afterwards will automatically update the default **residual**.

1. Run the long-run regression as it is.

- From the **workfile**, open all the variables as a group, then Go to **process**
- Make equation (i.e. b f gdpinf c)
- Run the regression by clicking **ok**
- Don't bother about the result

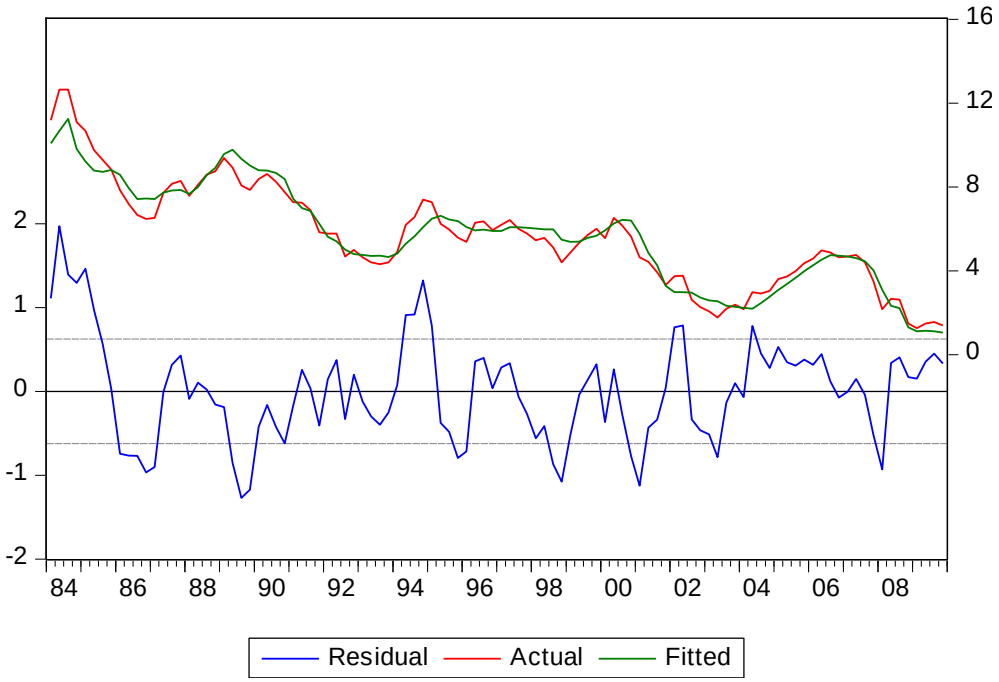
Dependent Variable: B
Method: Least Squares
Date: 11/04/17 Time: 20:43
Sample: 1984Q1 2009Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
F	0.682934	0.039882	17.12406	0.0000
GDP	-0.000218	2.89E-05	-7.516770	0.0000
INF	0.030568	0.033165	0.921681	0.3589
C	4.002034	0.412455	9.702952	0.0000

R-squared	0.938640	Mean dependent var	5.686538
Adjusted R-squared	0.936799	S.D. dependent var	2.480898
S.E. of regression	0.623691	Akaike info criterion	1.931379
Sum squared resid	38.89905	Schwarz criterion	2.033086
Log likelihood	-96.43171	Hannan-Quinn criter.	1.972584
F-statistic	509.9104	Durbin-Watson stat	0.516191
Prob(F-statistic)	0.000000		

From the **result**: Only INF is **not** significant.

- Note that the updated **residual** will automatically generate from the regression
  Click on the **residual** from the **workfile**, a graph will come up



In order not to lose the **updated residual**, do the following:

- From the **workfile, double-**Click on **residual.**
- From **process**, click on **generate by equation**
- Specify: **residual1=resid**
- **Residual1** will be saved on the workfile (i.e. Residual1 will appear among the variables on the workfile)

- It is only **resid** that will be updated henceforth while leaving the new **residual1** untouched.

.To perform Unit Root Test on the Residual

- Go back to the workfile
- Go to **Quick → Estimate Equation**
- Specify equation as: **residual1 c trend** to generate the results below:

Dependent Variable: RESIDUAL1
Method: Least Squares
Date: 11/04/17 Time: 21:44
Sample: 1984Q1 2009Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.035244	0.121923	0.289065	0.7731
TREND	-0.000671	0.002016	-0.332988	0.7398

R-squared	0.001086	Mean dependent var	3.92E-15
Adjusted R-squared	-0.008707	S.D. dependent var	0.614541
S.E. of regression	0.617211	Akaike info criterion	1.891831
Sum squared resid	38.85681	Schwarz criterion	1.942685
Log likelihood	-96.37522	Hannan-Quinn criter.	1.912433
F-statistic	0.110881	Durbin-Watson stat	0.516727
Prob(F-statistic)	0.739827		

- Perform the **Unit Root Test**

Note:

This can be done:

(i) At **level** without intercept & trend; with intercept; and with intercept & trend

(ii) At **first difference** or **2nd difference** without intercept & trend; with intercept; and with intercept & trend

If the Unit Root Test shows significance at level [i.e. I(0)], then you reject the Null Hypothesis (i.e. you reject "there is no co-integration", meaning that there is co-integration or long-run relationship among the variables). The reverse will be the case if they are **differenced stationary**, they are not co-integrated.

To perform the **Unit Root Test**, do the following:

- Go back to the **workfile**
- Double-click on **residual1**, the raw data for **residual1** is activated
- Go to **view**
- Go to **Unit Root Test**
- Specify accordingly, change the lag length from a maximum of 12 to 4 (remember you are using even no because you are dealing with quarterly data; it is sufficient if you use 1 or 2 if it is an annual data).

Null Hypothesis: RESIDUAL1 has a unit root
Exogenous: None
Lag Length: 4 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.427735	0.0000
Test critical values:	1% level	-2.588530	
	5% level	-1.944105	
	10% level	-1.614596	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RESIDUAL1)
Method: Least Squares
Date: 11/04/17 Time: 22:03
Sample (adjusted): 1985Q2 2009Q4
Included observations: 99 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESIDUAL1(-1)	-0.518120	0.095458	-5.427735	0.0000
D(RESIDUAL1(-1))	0.335447	0.099224	3.380700	0.0011
D(RESIDUAL1(-2))	0.072722	0.102328	0.710681	0.4790
D(RESIDUAL1(-3))	0.210125	0.093040	2.258444	0.0262
D(RESIDUAL1(-4))	0.029088	0.095005	0.306178	0.7601

R-squared	0.292689	Mean dependent var	-0.011395
Adjusted R-squared	0.262591	S.D. dependent var	0.439807
S.E. of regression	0.377674	Akaike info criterion	0.939612
Sum squared resid	13.40791	Schwarz criterion	1.070679
Log likelihood	-41.51081	Hannan-Quinn criter.	0.992642
Durbin-Watson stat	2.026434		

- Interpretation: Since the absolute value of *t*-statistic (5.427735) is greater than the Test critical values (at 10%, 5% and 1% - anyone of interest must be so specified), we cannot reject the null hypothesis (Ho) i.e. residual1 has a unit root; meaning that there the variables are co-integrated of order 0 (i.e. I(0)].

Engle-Granger Co-integration Test

- Note that co-integration does not look for specific pattern but general pattern of behaviour among the variables (i.e. it makes use of moving averages).
- When dealing with ADF, you do not bother with co-integrating vectors, whereas, Engle-Granger test bothers on how many integrating factors exist, which may not be unique.

To perform the Engle-Granger Test:

- Go to workfile
- Open your variables as a group (i.e. b, f, gdp, inf). Note, assuming each of these variables is I(1), running the regression may generate spurious regression because they are showing trend relationship.

  When performing the Engle-Granger Test, you are dealing with three things at the same time:

  (i) Long-run model (a model that is specified without **lead** or **lag** or **difference** is a pure long-run model)

  (ii) Spurious regression

  (iii) Co-integration
- Go to **View**
- Go to **Co-integration Test**
- Go to **Johansen System Co-integration Test**
- Make your specification:

  (i) Deterministic trend assumption

    - No intercept or trend

- Lag length already specified (btw 1 and 4)

- Click ok

Date: 11/04/17 Time: 22:42
Sample (adjusted): 1985Q2 2009Q4
Included observations: 99 after adjustments
Trend assumption: No deterministic trend
Series: B F GDP INF
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.276525	63.66021	40.17493	0.0001
At most 1 *	0.203461	31.61501	24.27596	0.0050
At most 2	0.076300	9.094618	12.32090	0.1636
At most 3	0.012420	1.237232	4.129906	0.3104

Trace test indicates 2 cointegratingeqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.276525	32.04520	24.15921	0.0035
At most 1 *	0.203461	22.52040	17.79730	0.0090
At most 2	0.076300	7.857386	11.22480	0.1837
At most 3	0.012420	1.237232	4.129906	0.3104

Max-eigenvalue test indicates 2 cointegratingeqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

B	F	GDP	INF
-1.637036	1.686503	0.000169	0.105793
0.533886	-0.470061	0.000211	0.074388
0.428735	0.292002	-5.54E-05	-0.626750
0.144081	-0.281327	0.000143	-0.141016

Unrestricted Adjustment Coefficients (alpha):

D(B)	-0.101865	-0.058118	-0.033308	0.044881
D(F)	-0.145636	0.009789	0.025133	0.017101

D(GDP)	-6.633601	18.43955	-5.544440	3.902193
D(INF)	0.107955	-0.006908	0.100768	0.029126

1 Cointegrating Equation(s): Log likelihood -655.2999

Normalized cointegrating coefficients (standard error in parentheses)

B	F	GDP	INF
1.000000	-1.030217	-0.000103	-0.064624
	(0.07916)	(3.2E-05)	(0.06665)

Adjustment coefficients (standard error in parentheses)

D(B)	0.166756
	(0.08489)
D(F)	0.238411
	(0.05356)
D(GDP)	10.85945
	(10.5648)
D(INF)	-0.176726
	(0.08710)

2 Cointegrating Equation(s): Log likelihood -644.0397

Normalized cointegrating coefficients (standard error in parentheses)

B	F	GDP	INF
1.000000	0.000000	0.003329	1.338368
		(0.00085)	(0.83676)
0.000000	1.000000	0.003332	1.361841
		(0.00083)	(0.81958)

Adjustment coefficients (standard error in parentheses)

D(B)	0.135728	-0.144476
	(0.08860)	(0.09009)
D(F)	0.243638	-0.250217
	(0.05631)	(0.05725)
D(GDP)	20.70407	-19.85531
	(10.5448)	(10.7217)
D(INF)	-0.180414	0.185313
	(0.09160)	(0.09314)

3 Cointegrating Equation(s): Log likelihood -640.1110

Normalized cointegrating coefficients (standard error in parentheses)

B	F	GDP	INF
1.000000	0.000000	0.000000	-0.828418
			(0.15309)
0.000000	1.000000	0.000000	-0.806635
			(0.15178)
0.000000	0.000000	1.000000	650.7850
			(246.742)

Adjustment coefficients (standard error in parentheses)

D(B)	0.121447	-0.154202	-2.77E-05
	(0.09107)	(0.09110)	(1.4E-05)
D(F)	0.254413	-0.242878	-2.39E-05

	(0.05782)	(0.05784)	(9.0E-06)
D(GDP)	18.32697	-21.47430	0.003081
	(10.8123)	(10.8154)	(0.00168)
D(INF)	-0.137211	0.214738	1.12E-05
	(0.09231)	(0.09234)	(1.4E-05)

Decision Criteria: Our decision criterion is based on the values of **Trace Statistic**

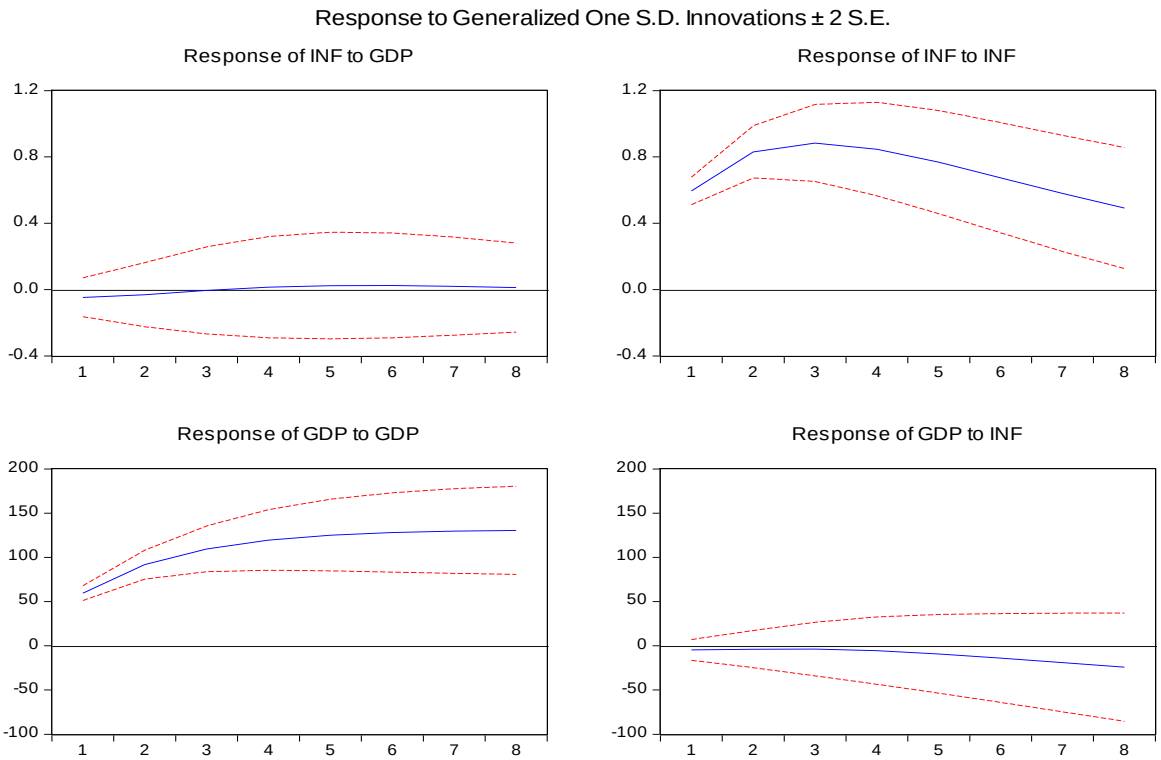
Viz.: Compare **Trace Statistic** with **Critical Value**

Interpretation:

- **None*** → No co-integration at all between the variables.

  If TS > CV, reject the null hypothesis that there is no co-integration
- **At most 1*→** There is at most one co-integrating vector among the variables i.e. It cannot be more than one vectors.
- **At most 2** → since our result shows that **TS < CV**, the co-integrating vectors are not more than just two.

3.5.3 Impulse Response



The graph represents the response of each variable to one another.

Remember, we specified 8 quarters, that is why we are having 8 on the x-axis.

4.0 CONCLUSION

Time Series data analysis and practical applications and the conduct of the stochastic process, stationary and nonstationary variables unit root test, co-integration as well as impulse response are important skills to be acquired by Economists.

5.0 SUMMARY

In this unit you have learnt the stages that could be followed in order to carry out acceptable and standard econometric model estimation. You have also learnt how to conduct econometric model estimation. The time is now ripe for you to try your hands on nonlinear regression models in the next unit, that is Unit 4 of our Module 1.

6.0 TUTOR MARKED ASSIGNMENT

Given the following data estimate the multiple regression equation using Y as dependent variable and the rest as independent variables.

year	I	Y	M	CN	PC	IR
1980	5.29	11.03	5.25	5.91	5.68	9.75
1981	5.46	10.86	5.28	6.14	5.91	10.00
1982	5.37	10.83	5.35	6.19	5.95	11.75
1983	5.10	10.75	5.45	6.29	6.07	11.50
1984	4.69	10.68	5.54	6.42	6.24	13.00
1985	4.63	10.75	5.60	6.53	6.35	11.75
1986	4.86	10.75	5.59	6.54	6.37	12.00
1987	5.13	10.71	5.77	6.84	6.73	19.20
1988	5.24	10.78	6.03	7.15	7.05	17.60
1989	5.64	10.83	6.12	7.28	7.19	24.60
1990	6.02	10.88	6.38	7.58	7.50	27.70
1991	6.11	10.90	6.67	7.71	7.63	20.80
1992	6.54	10.91	7.14	8.29	8.20	31.20
1993	6.83	10.90	7.55	8.56	8.47	36.09
1994	6.89	10.88	7.82	8.89	8.64	21.00

1995	7.16	10.88	7.97	9.60	9.44	20.79
1996	7.50	10.90	8.10	9.97	9.84	20.86
1997	7.65	10.90	8.22	9.97	9.81	23.32
1998	7.63	10.90	8.40	10.06	9.91	21.34
1999	7.56	10.89	8.66	10.01	9.92	27.19
2000	7.89	10.92	9.03	10.05	9.91	21.55
2001	7.99	10.92	9.24	10.38	10.28	21.34
2002	8.26	10.91	9.39	10.74	10.66	30.19
2003	8.78	10.99	9.49	10.94	10.88	22.88
2004	8.75	11.06	9.66	11.09	10.99	20.82
2005	8.66	11.09	9.84	11.28	11.19	19.49
2006	9.29	11.13	10.12	11.46	11.36	18.70
2007	9.47	11.16	10.59	11.69	11.59	18.36
2008	9.65	11.20	10.99	11.93	11.83	18.70
2009	9.83	11.24	11.10	12.17	12.07	22.90
2010	10.01	11.29	11.16	12.40	12.30	22.51

Estimate the regression equation, then interprete the findings.

7.0 REFERNCES/ FURTHER READING

Olusanya E. Olubusoye (2014). Introduction to time series econometrics. *CEAR*, IUinversity of Ibadan.

.

UNIT 4: NONLINEAR REGRESSION

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
3.1Meaning of Nonlinear Regression
3.2 Examples of Nonlinear Functions
3.3Assumptions of Nonlinear Regression
3.4 Nonlinear Regression Equations
3.4.1 Transformation of Nonlinear to linear Model
3.4.2 Segmentation
3.4.3 Why Should We Use Nonlinear Models?
3.4.3 Fitting Nonlinear Models
3.4.4Choosing Starting Values
3.4.5 Model Selection Criteria
3.5 Practical Application of Non-linear Regression Model in Eviews
3.5.1 Specifying Nonlinear Least Squares
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
6.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the preceding unit, you learnt about how to run time series data in Eviews. The focus of the present unit is nonlinear regression model. In econometrics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. The data consist of error-free independent variables (explanatory variables), x, and their associated observed dependent variables (response variables), y. Each y is modeled as a

random variable with a mean given by a nonlinear function f(x,β). Systematic error may be present but its treatment is outside the scope of regression analysis. If the independent variables are not error-free, this is an errors-in-variables model, also outside this scope.

2.0 OBJECTIVES

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

* Discuss the meaning of nonlinear regression
* Conduct practical application of non-linear regression model in Eviews
* Specify nonlinear least squares

3.0 MAIN CONTENTS

3.1 Meaning of Nonlinear Regression

Nonlinear regression is a regression in which the dependent or criterion variables are modeled as a non-linear function of model parameters and one or more independent variables. There are several common models, such as Asymptotic Regression/Growth Model, which is given by:

b1 + b2 * exp(b3 * x)

Logistic Population Growth Model, which is given by:

b1 / (1 + exp(b2 + b3 * x)), and

Asymptotic Regression/Decay Model, which is given by:

b1 – (b2 * (b3 * x)) etc.

The reason that these models are called nonlinear regression is because the relationships between the dependent and independent parameters are not linear. This test in SPSS is done by selecting “analyze” from the menu. Then, select “regression” from

analyze. After this, select “linear from regression,” and then click on “perform nonlinear regression.”There are certain terminologies in nonlinear regression which will help in understanding nonlinear regression in a much better manner. These terminologies are:

i. **Model Expression** is the model used, the first task is to create a model. The selection of the model in is based on theory and past experience in the field. For example, in demographics, for the study of population growth, logistic nonlinear regression growth model is useful.
ii. **Parameters** are those which are estimated. For example, in logistic nonlinear regression growth model, the parameters are b1, b2 and b3.
iii. **Segmented model** is required for those models which have multiple different equations of different ranges, equations are then specified as a term in multiple conditional logic statements.
iv. **Loss function** is a function which is required to be minimized. This is done by nonlinear regression.

3.2 Examples of Nonlinear Functions

Examples of nonlinear functions include exponential functions, logarithmic functions, trigonometric functions, power functions, Gaussian function, and Lorenz curves. Some functions, such as the exponential or logarithmic functions, can be transformed so that they are linear. When so transformed, standard linear regression can be performed but must be applied with caution. See Linearization, below, for more details.

In general, there is no closed-form expression for the best-fitting parameters, as there is in linear regression. Usually numerical optimization algorithms are applied to determine the best-fitting parameters. Again in contrast to linear regression, there may be many local minima of the function to be optimized and even the global minimum may produce a biased estimate. In practice, estimated values of the parameters are used, in conjunction with the optimization algorithm, to attempt to find the global minimum of a sum of

squares. For details concerning nonlinear data modeling see least squares and non-linear least squares.

3.3Assumptions of Nonlinear Regression

i. The data level in must be quantitative, the categorical variables must be coded as binary variables.

ii. The value of the coefficients can be correctly interpreted, only if the correct model has been fitted, therefore it is important to identify useful models.

iii. It's important to note that R-squared is invalid for nonlinear models and statistical software can't calculate p-values for the terms.

iv. The defining characteristic for both types of models (linear and nonlinear) are the functional forms. If you can focus on the form that represents a linear model, it's easy enough to remember that anything else must be a nonlinear. Now that you understand the differences between the two types of regression models, learn more about fitting curves and choosing between them in the following section. 3.4

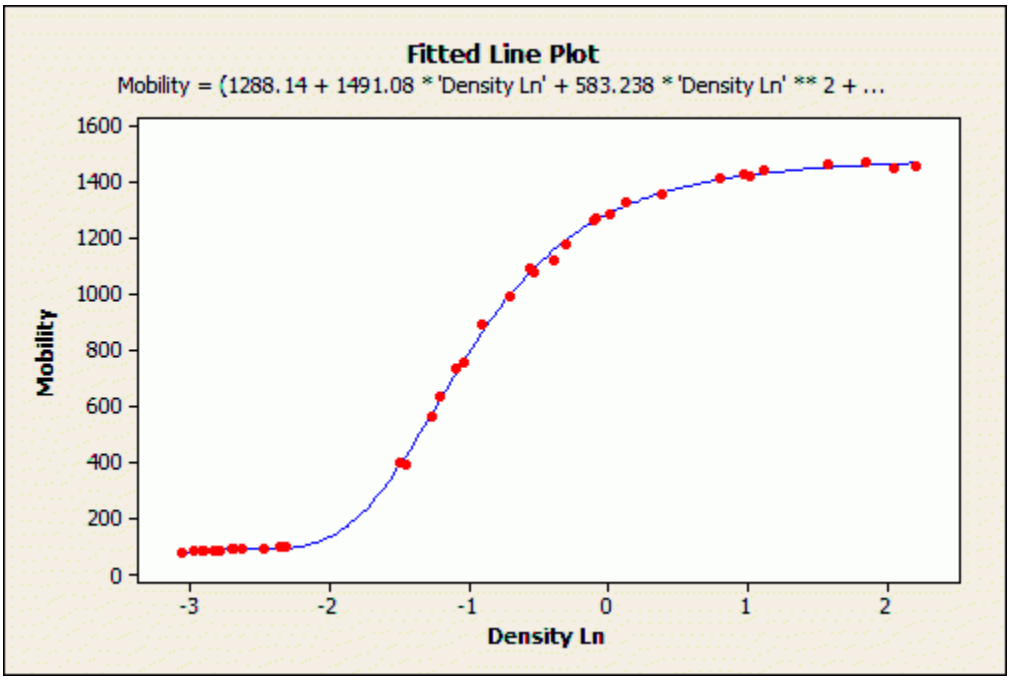
3.4 Nonlinear Regression Equations

I showed how linear regression models have one basic configuration. Now, we'll focus on the "non" in nonlinear! If a regression equation doesn't follow the rules for a linear model, then it must be a nonlinear model. It's that simple! A nonlinear model is literally not linear. The added flexibility opens the door to a huge number of possible forms. Consequently, nonlinear regression can fit an enormous variety of curves. However, because there are so many candidates, you may need to conduct some research to determine which functional form provides the best fit for your data.

Below, I present a handful of examples that illustrate the diversity of nonlinear regression models. Keep in mind that each function can fit a variety of shapes, and there are many nonlinear functions. Also, notice how nonlinear regression equations are not

comprised of only addition and multiplication! In the table, thetas are the parameters, and Xs are the independent variables.

The nonlinear regression example below models the relationship between density and electron mobility, Fujii and Konishi (2006).



The equation for the nonlinear regression analysis is too long for the fitted line plot:

Electron Mobility = (1288.14 + 1491.08 * Density Ln + 583.238 * Density Ln^2 + 75.4167 * Density Ln^3) / (1 + 0.966295 * Density Ln + 0.397973 * Density Ln^2 + 0.0497273 * Density Ln^3)

SELF – ASSESSMENT EXERCISE

List and explain examples of nonlinear regression models

3.4.1 Transformation of Nonlinear to linear Model

If the model is non-linear (but the dependent variable is still continuous), we can transform or add variables to get the equation to be linear. For instance, We can take logs of Y and/or the X's depending on the source(s) of nonlinearity. We can also add squared terms as well as interaction terms

Some nonlinear regression problems can be moved to a linear domain by a suitable transformation of the model formulation. For example, consider the nonlinear regression problem below:

$$Y = ae^{bx}U$$

with parameters a and b and with multiplicative error term U. If we take the logarithm of both sides, this becomes:

$$\ln(y) = \ln(a) + bx + u,$$

where u = ln(U), suggesting estimation of the unknown parameters by a linear regression of ln(y) on x, a computation that does not require iterative optimization. However, use of a nonlinear transformation requires caution. The influences of the data values will change, as will the error structure of the model and the interpretation of any inferential results. These may not be desired effects. On the other hand, depending on what the largest source of error is, a nonlinear transformation may distribute the errors in a Gaussian fashion, so the choice to perform a nonlinear transformation must be informed by modeling considerations.

3.4.2 Segmentation

The independent or explanatory variable (say X) can be split up into classes or segments and linear regression can be performed per segment. Segmented regression with confidence analysis may yield the result that the dependent or response variable (say Y) behaves differently in the various segments.[1]

3.4.3 Why Should We Use Nonlinear Models?

The main advantages of nonlinear models are parsimony, interpretability, and prediction (Bates and Watts, 2007). In general, nonlinear models are capable of accommodating a vast variety of mean functions, although each individual nonlinear model can be less flexible than linear models (i.e., polynomials) in terms of the variety of data they can describe; however, nonlinear models appropriate for a given application can be more parsimonious (i.e., there will be fewer parameters involved) and more easily interpretable. Interpretability comes from the fact that the parameters can be associated with an economically meaningful process. For example, one of the most widely used nonlinear models is the logistic equation.. The parameters have a clear meaning and units associated with their definition. Assuming the asymptotic parameter (Y_{asym}) has units equal to the response variable (Y), the inflection point (t_m) has units equal to the independent variable (t), and the parameter that determines the steepness of the curves (k) has units equal to t. This last parameter can be interpreted as the time (when t is time) that it takes to move from the inflection point to approximately 0.73 of the asymptotic value. A competing polynomial model used to describe the same data would have the disadvantages that more parameters would be needed (more than just three) and that the parameters would not be easily interpretable (Pinheiro and Bates, 2000). For example, what would be the interpretation of the parameters in a five degree polynomial?

The final advantage of using nonlinear regression models is that their predictions tend to be more robust than competing polynomials, especially outside the range of observed data (i.e., extrapolation). Nonlinear regression models, however, come at a cost. Their main disadvantages are that they can be less flexible than competing linear models and that generally there is no analytical solution for estimating the parameters. The first point has as a consequence that the choice of model is crucial

3.4.3 FITTING NONLINEAR MODELS

Presently there are many statistical software packages available for fitting nonlinear models (e.g., SAS, R, JMP, GenStat, MatLab, Sigmaplot, OriginLab, Eviews and SPSS).

Nonlinear parameter estimates can be obtained using different methods (Bates and Watts, 2007); the most common are: (i) ordinary least squares (OLS), which minimize the sum of squared error between observations and predictions, and (ii) the maximum likelihood method (MLM), which seeks the probability distribution that makes the observed data most likely. For non-normal data such as binomial or counts, generalized (non)linear models should be used (Lindsey, 2001; Huet e., 2003; Gbur., 2012). Most problems encountered during the use of standard nonlinear regression software functions are due to a poor choice of competing models or an incorrect equation or starting values. The choice of estimation method can affect the parameter estimates (Ruppert, 1989), but in general, estimates from least squares and maximum likelihood methods tend to differ only when the data are not normally distributed and are approximately identical when the data follow a normal distribution (Myung, 2003).

3.4.4Choosing Starting Values

All the procedures for nonlinear parameter estimation require initial values. The choice of values will influence the convergence of the estimation algorithm, in the worst case yielding no convergence and in the best case convergence in a few iterations (Ritz and Streibig, 2005); however, there is no standard procedure for getting initial estimates. We indicate five practical methods:

i. If the model has parameters with economics meaning, then use information from the literature.

ii. Use graphical exploration

iii. Transform the nonlinear model into a linear model. For instance logarithmic transformation of $Y = Y_o\exp(-kt)$ yields a linear equation (viz. $\ln Y = Y_o - kt$) in which rough estimates of the parameter values can be easily obtained by linear regression.

This method is recommended for getting initial estimates and to detect deviations from linearity, but these estimates may also be used as the final estimates.

3.4.5 Model Selection Criteria

When we are dealing with multiple models, the question is how to find the best model among competing models. Depending on the structure of the models, different statistical criteria can be used to find the best model: *F* -test, Alkaike information criterion (AIC), Bayesian information criterion (BIC), or the likelihood ratio test .When models are *nested* (one model is a special case of another), any of these criteria are applicable. When models are *non-nested* (models having different structures, typically the AIC and the BIC criteria are used. From a practical point of view, however, one model might be preferred over another based on interpretability and specific objectives. There needs to be a balance between statistical model performance and how effectively the model answers research questions.

3.5 Practical Application of Non-linear Regression Model in Eviews

Suppose that we have the regression specification:

$$y_t = f(x_t, \beta) + \epsilon_t$$

where f is a general function of the explanatory variables x_t and the parameters β . Least squares estimation chooses the parameter values that minimize the sum of squared residuals:

We say that a model is linear in parameters if the derivatives of f with respect to the parameters do not depend upon β ; if the derivatives are functions of β , we say that the model is nonlinear in parameters (Ritz and Streibig).

For example, consider the model given by:

$$y_t = \beta_1 + \beta_2 \log L_t + \beta_3 \log K_t + \epsilon_t$$

It is easy to see that this model is linear in its parameters, implying that it can be estimated using ordinary least squares.

In contrast, the equation specification:

$$y_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} + \epsilon_t$$

has derivatives that depend upon the elements of β. There is no way to rearrange the terms in this model so that ordinary least squares can be used to minimize the sum-of-squared residuals. We must use nonlinear least squares techniques to estimate the parameters of the model.

Nonlinear least squares minimizes the sum-of-squared residuals with respect to the choice of parameters β. While there is no closed form solution for the parameters, estimates may be obtained from iterative methods as described in "Optimization Algorithms".

3.5.1 Specifying Nonlinear Least Squares

For nonlinear regression models, you will have to enter your specification in equation form using EViews expressions that contain direct references to coefficients. You may use elements of the default coefficient vector C (e.g. C(1), C(2), C(34), C(87)), or you can define and use other coefficient vectors Crainiceanu and Rupprt(2004). For example:

y = c(1) + c(2)*(k^c(3)+l^c(4))

is a nonlinear specification that uses the first through the fourth elements of the default coefficient vector, C.

To create a new coefficient vector, select Object/New Object.../Matrix-Vector-Coef in the main menu and provide a name. You may now use this coefficient vector in your specification. For example, if you create a coefficient vector named CF, you can rewrite the specification above as:

y = cf(11) + cf(12)*(k^cf(13)+l^cf(14))

which uses the eleventh through the fourteenth elements of CF.

You can also use multiple coefficient vectors in your specification:

y = c(11) + c(12)*(k^cf(1)+l^cf(2))

which uses both C and CF in the specification.

It is worth noting that EViews implicitly adds an additive disturbance to your specification. For example, the input

y = (c(1)*x + c(2)*z + 4)^2

is interpreted as $y_t = (c(1)x_t + c(2)z_t + 4)^2 + \epsilon_t$, and EViews will minimize:

If you wish, the equation specification may be given by a simple expression that does not include a dependent variable. For example, the input,

(c(1)*x + c(2)*z + 4)^2

is interpreted by EViews as $-(c(1)x_t + c(2)z_t + 4)^2 = \epsilon_t$, and EViews will minimize:

While EViews will estimate the parameters of this last specification, the equation cannot be used for forecasting and cannot be included in a model. This restriction also holds for any equation that includes coefficients to the left of the equal sign. For example, if you specify,

x + c(1)*y = z^c(2)

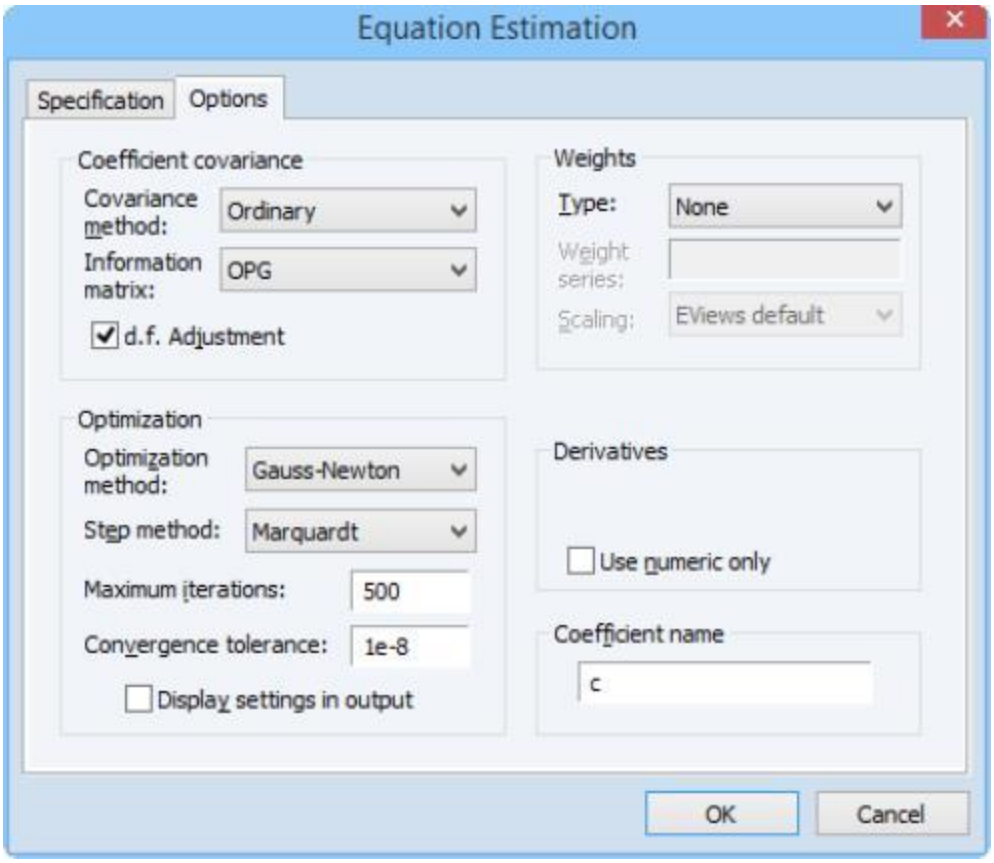
EViews will find the values of C(1) and C(2) that minimize the sum of squares of the implicit equation:

$$x_t + c(1)y_t - z_t^{c(2)} = \epsilon_t$$

The estimated equation cannot be used in forecasting or included in a model, since there is no dependent variable.

Estimation Options

Clicking on the Options tab displays the nonlinear least squares estimation options:



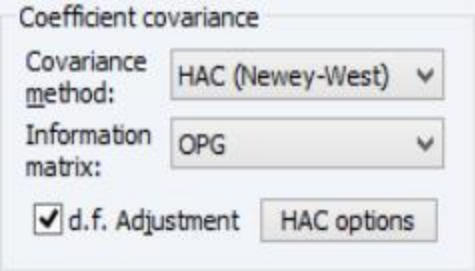
Coefficient Covariance

EViews allows you to compute ordinary coefficient covariances using the inverse of either the OPG of the mean function or the observed Hessian of the objective function, or to compute robust sandwich estimators for the covariance matrix using White or HAC (Newey-West) estimators.

• The topmost Covariance method dropdown menu should be used to choose between the default Ordinary or the robust Huber-White or HAC (Newey-West) methods.

• In the Information matrix menu you should choose between the OPG and the Hessian - observed estimators for the information.

• If you select HAC (Newey-West), you will be presented with a HAC options button that, if pressed, brings up a dialog to allow you to control the long-run variance computation.

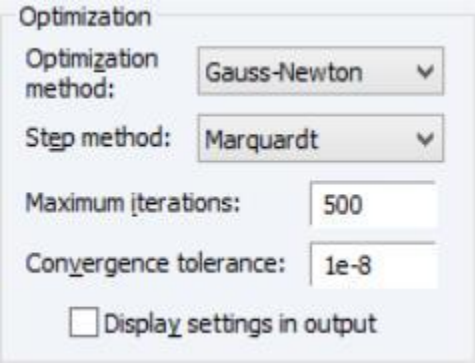


See "Robust Standard Errors" for a discussion of White and HAC standard errors.

You may use the d.f. Adjustment checkbox to enable or disable the degree-of-freedom correction for the coefficient covariance. For the Ordinary method, this setting amounts to determining whether the residual variance estimator is or is not degree-of-freedom corrected. For the sandwich estimators, the degree-of-freedom correction is applied to the entire matrix.

Optimization

You may control the iterative process by specifying the optimization method, convergence criterion, and maximum number of iterations.



The Optimization method dropdown menu lets you choose between the default Gauss-Newton and BFGS, Newton-Raphson, and EViews legacy methods.

In general, the differences between the estimates should be small for well-behaved nonlinear specifications, but if you are experiencing trouble, you may wish to experiment with methods. Note that EViews legacy is a particular implementation of Gauss-Newton

with Marquardt or line search steps, and is provided for backward estimation compatibility.

The Step method allow you to choose the approach for choosing candidate iterative steps. The default method is Marquardt, but you may instead select Dogleg or Line Search.

See “Optimization Method”, and “Optimization Algorithms” for related discussion.

EViews will report that the estimation procedure has converged if the convergence test value is below your convergence tolerance. See for details. While there is no best choice of convergence tolerance, and the choice is somewhat individual, as a guideline note that we generally set ours something on the order of 1e-8 or so and then adjust it upward if necessary for models with difficult to compute numeric derivatives.

See “Iteration and Convergence” for additional discussion.

In most cases, you need not change the maximum number of iterations. However, for some difficult to estimate models, the iterative procedure may not converge within the maximum number of iterations. If your model does not converge within the allotted number of iterations, simply click on the Estimate button, and, if desired, increase the maximum number of iterations. Click on OK to accept the options, and click on OK to begin estimation. EViews will start estimation using the last set of parameter values as starting values.

These options may also be set from the global options dialog. See Appendix A, “Estimation Defaults” for details.

Derivative Methods

Estimation in EViews requires computation of the derivatives of the regression function with respect to the parameters.

In most cases, you need not worry about the settings for the derivative computation. The EViews estimation engine will employ analytic expressions for the derivatives, if possible, or will compute high numeric derivatives, switching between lower precision computation early in the iterative procedure and higher precision computation for later iterations and final computation. You may elect to use only numeric derivatives.

See “Derivative Computation” for additional discussion.

Starting Values

Iterative estimation procedures require starting values for the coefficients of the model. The closer to the true values the better, so if you have reasonable guesses for parameter values, these can be useful. In some cases, you can obtain good starting values by estimating a restricted version of the model using least squares. In general, however, you may need to experiment in order to find starting values.

There are no general rules for selecting starting values for parameters so there are no settings in this page for choosing values. EViews uses the values in the coefficient vector at the time you begin the estimation procedure as starting values for the iterative procedure. It is easy to examine and change these coefficient starting values. To see the current starting values, double click on the coefficient vector in the workfile directory. If the values appear to be reasonable, you can close the window and proceed with estimating your model.

If you wish to change the starting values, first make certain that the spreadsheet view of your coefficients is in edit mode, then enter the coefficient values. When you are finished setting the initial values, close the coefficient vector window and estimate your model.

You may also set starting coefficient values from the command window using the PARAM command. Simply enter the PARAM keyword, following by each coefficient and desired value. For example, if your default coefficient vector is C, the statement:

param c(1) 153 c(2) .68 c(3) .15

sets C(1)=153, C(2)=.68, and C(3)=.15.

See Appendix C, “Estimation and Solution Options”, for further details.

Output from NLS

Once your model has been estimated, EViews displays an equation output screen showing the results of the nonlinear least squares procedure. Below is the output from a regression of LOG(CS) on C, and the Box-Cox transform of GDP using the data in the workfile “Chow_var.WF1”:

Dependent Variable: LOG(CS)
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 03/09/15 Time: 11:25
Sample: 1947Q1 1994Q4
Included observations: 192
Convergence achieved after 68 iterations
Coefficient covariance computed using outer product of gradients
LOG(CS)=C(1)+C(2)*(GDP^C(3)-1)/C(3)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.839332	0.281733	10.07810	0.0000
C(2)	0.259119	0.041680	6.216837	0.0000
C(3)	0.182315	0.020335	8.965475	0.0000

R-squared	0.997260	Mean dependent var	7.472280
Adjusted R-squared	0.997231	S.D. dependent var	0.463744
S.E. of regression	0.024403	Akaike info criterion	-4.572707
Sum squared resid	0.112552	Schwarz criterion	-4.521808
Log likelihood	441.9798	Hannan-Quinn criter.	-4.552093
F-statistic	34393.45	Durbin-Watson stat	0.136871
Prob(F-statistic)	0.000000		

If the estimation procedure has converged, EViews will report this fact, along with the number of iterations that were required. If the iterative procedure did not converge, EViews will report “Convergence not achieved after” followed by the number of iterations attempted.

Below the line describing convergence, and a description of the method employed in computing the coefficient covariances, EViews will repeat the nonlinear specification so that you can easily interpret the estimated coefficients of your model.

EViews provides you with all of the usual summary statistics for regression models. Provided that your model has converged, the standard statistical results and tests are asymptotically valid.

The estimation is done using the method of Maximum Likelihood Estimation

4.0 CONCLUSION

Nonlinear models are very useful because of their properties of parsimony, interpretability, and prediction. In addition, nonlinear models are capable of accepting a vast variety of mean functions. Indeed, using nonlinear regression models ensures that their predictions tend to be more robust than competing polynomials, especially outside the range of observed data (i.e., extrapolation). Nonlinear regression models, however, come at a cost. Their main disadvantages are that they can be less flexible than competing linear models and that generally there is no analytical solution for estimating the parameters. The first point has as a concern that the choice of model is crucial

5.0 SUMMARY

In this unit you learnt the meaning of nonlinear regression, examples of nonlinear functions, assumptions of nonlinear regression, nonlinear regression equations, and transformation of nonlinear to linear model and practical application of non -linear regression model in Eviews. The next unit treatsqualitative response regressions.

6.0TUTOR MARKED ASSIGNMENT

What is nonlinear regression?

7.0 REFERENCES/FURTHER READING

http://www.eviews.com/help/helpintro.html#page/content%2FRegress2-Nonlinear_Least_Squares.html%23

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York: John Wiley & Sons.

Crainiceanu, C. M., & Ruppert, D. (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis, 91*(1), 35-52.

Fujii, T., & Konishi, S. (2006). Nonlinear regression modeling via regularized wavelets and smoothing parameter selection. *Journal of Multivariate Analysis, 97*(9), 2023-2033.

Gross, A. L., & Fleishman, L. E. (1987). The correction for restriction of range and nonlinear regressions: An analytic study. *Applied Psychological Measurement, 11*(2), 211-217.

Hanson, S. J. (1978). Confidence intervals for nonlinear regression: A BASIC program. *Behavior Research Methods & Instrumentation, 10*(3), 437-441.

Huet, S., Bouvier, A., Poursat, M. -A., & Jolivet, E. (2004). *Statistical tools for nonlinear regression: A practical guide with S-PLUS and R examples* (2nd ed.). New York: Springer.

McGwin, G., Jr., Jackson, G. R., & Owsley, C. (1999). Using nonlinear regression to estimate parameters of dark adaptation. *Behavior Research Methods, Instruments & Computers, 31*(4), 712-717.

Rao, B. L. S. P. (2004). Estimation of cusp in nonregular nonlinear regression models. *Journal of Multivariate Analysis, 88*(2), 243-251.

Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear regression*. New York: John Wiley & Sons.

Sheu, C. -F., & Heathcote, A. (2001). A nonlinear regression approach to estimating signal detection models for rating data. *Behavior Research Methods, Instruments & Computers, 33*(2), 108-114.

Verboon, P. (1993). Robust nonlinear regression analysis. *British Journal of Mathematical and Statistical Psychology, 46*(1), 77-94.

Wang, J. (1995). Asymptotic normality of L-sub-1-estimators in nonlinear regression. *Journal of Multivariate Analysis, 54*(2), 227-238.

UNIT 5: QUALITATIVE RESPONSE REGRESSIONS

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
3.1 Qualitative Response Regress
3.2 Linear Probability Model (LPM)
3.3 Problems of estimating binary choice models using LPM
3.4 The Logit Model
3.5 Estimating the Logit model *(A practical example)*
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the preceding unit, you learnt nonlinear regression models. This present unit which is Unit 5 and the last of our Module 1, you will learn qualitative response regressions. When we talk of Qualitative Response Regression, we are referring to a regression which is estimated from a cross sectional data set, in which case, the regressand becomes a qualitative variable. These data sets are usually generated through a survey instrument and pose quite a few estimation and interpretation challenges due to the non-linear nature of such. Ideally, certain techniques which are adopted for non-linear regression models with continuous endogenous variables are usually adopted for this kind of regressions.

2.0 OBJECTIVES

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

*Discuss qualitative response regression

*Explain linear probability model (LPM)

*Analyze problems of estimating binary choice models using LPM

*Do practical on the Logit model, estimating the Logit model

3.0 MAIN CONTENTS

3.1 Qualitative Response Regression

Regressions are easily estimated with variables that are continuous and fully observable (Time series data) but in real life situation, people also make choices that cannot be measured by continuous outcome variables but discrete variables. Thus, an investigation into the determinants of such choices gives rise to specialsed modelling called discrete choice models. Discrete choice models involve scenarios where the range values of the dependent variable is limited and as such, people are constrained to choose from finite countable number of distinct outcomes.

A limited dependent variable can be binary (dichotomous), ordered or multinomial category (trichotomous or polychotomous). It becomes **dichotomous** when the dependent variable has only two response category such as "Yes" or "No". *Example:* when a survey is conducted to determine house ownership in a particular geographical area, say, Jos metropolis in Nigeria, each of the families sampled in the study must either own a house or not. Thus, the response is coded "1" for families that own a house or "0" for families that do not own a house**. Ordered category** refers to cases where the dependent variable has more than two responses (trichotomous or polychotomous) but with natural ordering. In this case, coding follows the natural ordering such that, the highest rank gets the biggest value. *Example:* A survey which seeks to determine a subjective health rating will follow thus; (5) Excellent (4) Very Good (3) Good (2) Fair (1) Poor. This follows the conventional likert scale format occasioned by a statement, e.g. how would you rate your current health status? Lastly, it becomes a **Multinomial category** when the dependent variable is a multiple response category (trichotomous or polychotomous) but has no natural ordering. In this case, coding does not follow a predefined order and options must be mutually exclusive. *Example*: Making a decision whether to buy soft drinks such as Pepsi, Mirinda, Seven Up or Coke does not follow any ordering.

It is important to note the peculiarities of these qualitative dependent variables as that will determine the technique for their estimation and interpretation. Also, note should be taken of the fundamental difference between a regression where the endogenous variable *(Y)* is quantitative and where it is qualitative; When *Y* is quantitative, the thrust is to

estimate its mean value but where it is qualitative, the conditional probability is estimated. In this sense, qualitative response models are termed probability models.

In modeling discrete choice models, we shall concern ourselves first with binary models. Consider the binary outcome of a variable, ***y***, which usually takes either of the following values;

Y =

1 with probability *p* 0 with probability *1 - p*

Since we are interested in modelling ***p*** as a function of regressors ***x***, the outcome values are set to 1 or 0. The probability function for the observed outcome, ***y***, is given as;

y, *is* $p^y(1- p)^{1-y}$, *with E(y) = p and Var(y) = p(1-p)…………………. (1)*

A regression model is formed by parameterizing p to depend on an index function x'β, where x is a K *1 regressor vector and β is a vector of unknown parameters. In standard binary outcome models, the conditional probability has the form;

$$\rho i \equiv \Pr\left(yi = \frac{1}{x}\right) = F(Xi'\beta) \text{ ………………………………………… } (2)$$

Where $F(.)$ serves as the parametric function of $x'\beta,$ usually called a Cumulative Distribution Function (CDF) on $(-\infty, \sigma 2)$ ensuring that the bounds $0 \le p \le 1$ are satisfied.

However, the general approach to developing these models include; Linear Probability Model (LPM), Logit Model, Probit Model and Tobit Model (Beyond the scope of the course).

3.2 Linear Probability Model (LPM)

A Linear Probability Model (LPM) specifies a functional form for the probability ***p*** as a function of some regressors and the models are fitted by Ordinary Least Squares (OLS) estimation technique. It is required therefore, that LPM satisfies the basic assumptions of OLS (BLUE) when estimating a discrete choice model.

If for example, $F(.)$ is assumed to be a linear function and thus, $p = x'\beta$, then the linear conditional mean function explains the LPM. In this wise, the LPM is estimated using OLS regression of y on x by typing the command regress or reg (when using Stata). Remember that probability models are designed to estimate conditional probabilities and not conditional mean. Thus, when LPM is set to find a conditional mean in a discrete choice model such as the above hypothetical survey on house ownership in Jos Metropolis-Nigeria, the fitted values $x'\beta$ will not fall within 0 and 1 interval. And, since $\text{Var}\left(\frac{y}{x}\right) = (x'\beta)(1 - x'\beta)$ for the LPM, the regression intrinsically suffers from heteroscedasticity.

SELF-ASSESEMENT EXERCISE

1. Give practical example of a Logit problem.

3.3 Problems of estimating binary choice models using LPM

The scenario described above shows that LPM is inherently unstable as an estimation technique of binary choice models. However, the basic problems of using LPM include;

i. Non-normality of the error term V_i

The normality of the error term is not usually achieved using LPM and this is so because, just like y_i, V_i is set within the boundaries of 0 and 1.Thus, V_i is not normally distributed instead, it follows a binomial distribution. However, it should be noted that the normality criteria of OLS estimators are achieved with increases

in a sample size and thus, LPM automatically achieves this normality assumption as the sample size becomes increased, just like the OLS.

ii. Errors are not homoscedastic

One of the assumptions of OLS is achieving minimum variance (homoscedasticity) for the error term V_i in order to ensure that its estimators are unbiased and efficient. However, $pi = E\left(\frac{Yi}{Xi}\right) = x'\beta$, the values of V_i depends on the values of x, which makes it heteroscedastic. Thus, even if $E(Vi) = 0\ and\ cov(ui, uj) = 0\ so\ that, i \neq j$ in which case we conclude that there is no serial correlation, we cannot conclude that V_i is homoscedastic in the LPM.

iii. Non-fulfillment of $0 \leq E(Yi/Xi) \leq 1$

Predicted values of conditional probabilities E(Y) can fall outside (0,1) interval, whereas, they are meant to be bounded between 0 and 1. This fundamentally is the problem of fitting OLS in a Linear Probability Model. It is so because OLS does not consider such inequality restrictions. However, based on the foregoing limitations of LPM, probability models which overcome these short-comings have been developed. These are the Logit and Probit models. These models will form the remaining part of the discussion in this chapter.

3.4 The Logit Model

The above limitations of LPM can be overcome by the use of more complex binary response models such as the binomial Logit and Probit.

Now supposing we are interested in investigating the factors which determine whether or not a woman works and we estimate the regression equation below;

$P(y=1/x) = P(y=1/x_1, x_2 \ldots\ldots x_k)$ …… *(3)*

In eqn. 3 above, ***y*** is the employment indicator (a response variable showing whether or not a woman works) whereas, ***x*** is used to represent the explanatory variables. In this case, ***x*** will constitute of variables such as marital status, age, educational level, number of children, etc.

To illustrate the binomial logit, consider the presentation of the employment indicator below;

$P_i = 1/1 + e^{-(\beta 1 + \beta 2Xi)}$ …………………………………………………………………… *(4)*

Let $(\beta_1 + \beta_2 X_i) = Z_i$ *and thus,*

$P_i = 1/1 + e^{-Zi} = e^{Z}/1 + e^{Z}$ …………………………………………….. *(5)*

The above equation is called the logistic distribution function (cumulative) and is symmetric around zero. It can thus be shown that as Z_i ranges from -∞ to +∞, P_i ranges from 0 and 1 and also, P_i is non-linearly related to Z_i (X_i). However, since P_i is non-linearly related to *X and β,* the OLS is no longer suitable for estimating the parameters of the equation. To linearize the equation, consider the probability of a woman not working to be $(1 - P_i)$ since the probability of a woman working is $P_{i.}$ Thus,

$(1 - P_i) = 1/1 + e^{Zi}$ ……………………………………………………… *(6)*

Putting the two probabilities together;

$P_i / 1 - P_i = 1 + e^{Zi} / 1 + e^{-Zi} = e^{Zi}$ …………………………………… *(7)*

However, $P_i / 1 - P_i$ is the odd ratio showing that a woman will work. Finding the natural log of the equation 7 gives you the following transformation;

$L_i = ln(P_i/1 - P_i) = Z_i$ …………………………………………………….. *(8)*

L symbolizes the logit model and shows that the log of the odds ratio is non-linear but linear in the parameters (when estimated). The logit and probit models rely on maximum likelihood methods of estimation as illustrated below.

3.5 Estimating the Logit model *(A practical example)*

Table 1. Factors that determine whether a woman will work or not

Age	Education	marital status	children	Work
22	10	1	0	0
36	10	1	0	1
28	10	1	0	0
37	10	1	0	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
39	12	1	0	0
25	10	0	3	1
26	16	0	0	1
28	10	1	1	0

Task: *Using the hypothetical data in Table 1 above, estimate a logistic regression and attempt an interpretation of your result.*

- Load the Stata software
- Import the binary data and select the sheet that contains the binary data (where you have data sets in more than one sheet)
- Use the command ***logit work children maritalstatus education age*** ↲
  Observe that the above stata command is preceded by the logit command, the dependent variable (work) and then the other explanatory variables.
- Once the enter key is hit, the result in Fig. 1 is displayed; Asteriou and Hall (2007)

```
Iteration 0:   log likelihood = -1266.2225
Iteration 1:   log likelihood = -1040.6658
Iteration 2:   log likelihood = -1027.9567
Iteration 3:   log likelihood = -1027.9145
Iteration 4:   log likelihood = -1027.9144

Logistic regression                               Number of obs   =       2000
                                                  LR chi2(4)      =     476.62
                                                  Prob > chi2     =     0.0000
Log likelihood = -1027.9144                       Pseudo R2       =     0.1882
```

work	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
children	.7644882	.0515289	14.84	0.000	.6634935	.865483
maritalstatus	.7417775	.1264705	5.87	0.000	.4938998	.9896552
education	.0982513	.0186522	5.27	0.000	.0616936	.134809
age	.0579303	.007221	8.02	0.000	.0437773	.0720833
_cons	-4.159247	.3320401	-12.53	0.000	-4.810034	-3.508461

Figure. 1

The above result is a logistic regression from Stata 10 software. Explanatory variables such as number of children, marital status, education and age where tested on the dependent variable, work. The result shows that all the explanatory variables are significant at one percent and have a positive relationship with the dependent variable (work). It follows therefore that there is more likelihood that a woman with children, married, educated and with advancement in age will work. LR chi^2 of 476.62 shows that the overall model is statistically significant, this is further buttressed by its probability of 0.000 showing statistical significance at one percent. Stata 10.0

Mfx Result

```
Marginal effects after logit
      y  = Pr(work) (predict)
         =  .72678588
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
children	.151803	.00938	16.19	0.000	.133425	.170181	1.6445
marita~s*	.1545671	.02703	5.72	0.000	.101592	.207542	.6705
educat~n	.0195096	.0037	5.27	0.000	.01226	.02676	13.084
age	.0115031	.00142	8.08	0.000	.008713	.014293	36.208

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Figure. 2

Fig. 2 above shows marginal effect result. It measures the magnitude of impact arising from each of the explanatory variables on the dependent variable. Using the variable children as an example, it implies after having two children, there is a 0.15 probability that a woman will work. Also, after an average education of 13 years, a further increase in education will increase the likelihood of a woman working by 0.19.

4.0 CONCLUSSION

Logit model deals with discrete variables and not continuous variable. Discrete choice models involve scenarios where the range values of the dependent variable is limited and as such, people are constrained to choose from finite countable number of distinct outcomes. When dealing with Logistic regressions, the following parameters are usually estimated: odds ratios, Log of odds, Marginal effects and Conditional Probability.

5.0 SUMMARY

In this unit you learnt qualitative response regress with linear probability model (LPM) and problems of estimating binary choice models using LPM. You also learnt how to estimating the logit model. In the next unit we will discuss probit model.

6.0 TUTOR-MARKED ASSIGNMENT

Design a questionnaire to obtain discrete variable.

7.0 REFERENCES AND FURTHER READING

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

MODULE 2

Unit 1. The Probit Model
Unit 2: Autoregressive Process
Unit 3: Stationarity
Unit 4: Panel Data Regression Model
Unit 5: Fixed Versus Random Effects Panel Data

UNIT 1. THE PROBIT MODEL

CONTENTS

1. INTRODUCTION
2. OBJECTIVES
3.0 MAIN CONTENT
    3.1 Probit Model
    3.2 Estimating the Probit model *(A practical example)*
    3.3 Ordered Logit and Probit models
    3.4 Multinomial Logit and Probit models
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

3.0 INTRODUCTION

In the previous unit you learnt qualitative response regression with Logit model, which we hope has prepared you for the Probit model we will discuss in this unit. The Probit model is similar to the Logit model discussed in the Unit 5 of Module 1 except that it uses a normal cumulative distributive function while the Logit model uses a cumulative logistic function in its estimation of a binary model.

2.0 OBJECTIVES

* Discuss Probit Model
* Estimate the Probit model
* Explain ordered Logit and Probit models

- Discuss Multinomial Logic and Probit models

3.0 MAIN CONTENTS

3.1 Probit Model

To illustrate the probit model, assume that in the example on employment indicator (factors that determine if a woman works or not), we have an index I which is determined by a group of explanatory variables represented by X_i. Such an index can be expressed thus;

$I_i = \beta_1 + \beta_2 X_i$ …… *(1)*

In the equation above, Y = 1 (if a woman works) and Y = 0 (if she does not work). We thereafter assume a set point of the index such as I_i^* so that, I_i becomes 1 when it exceeds this point otherwise it becomes 0. Though the set point (I_i^*) and the index (I_i) are unobservable, they are normally distributed and as such are assumed to have the same mean and variance. However, going by the normality assumption, the probability that $I_i^* \leq I_i$ is calculated using the standard normal CDF as shown below;

$P_i = P(Y = 1/X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i)$…… *(10)*

Where, $P(Y = 1/X)$ showing the probability that a woman works given X explanatory variables, Z_i is the standard normal variable and F is the standard normal CDF.

Estimating the Probit model *(A practical example)*

```
Iteration 0:   log likelihood = -1266.2225
Iteration 1:   log likelihood = -1031.4962
Iteration 2:   log likelihood = -1027.0625
Iteration 3:   log likelihood = -1027.0616
Iteration 4:   log likelihood = -1027.0616

Probit regression                                 Number of obs   =       2000
                                                  LR chi2(4)      =     478.32
                                                  Prob > chi2     =     0.0000
Log likelihood = -1027.0616                       Pseudo R2       =     0.1889
```

work	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
children	.4473249	.0287417	15.56	0.000	.3909922	.5036576
maritalstatus	.4308575	.074208	5.81	0.000	.2854125	.5763025
education	.0583645	.0109742	5.32	0.000	.0368555	.0798735
age	.0347211	.0042293	8.21	0.000	.0264318	.0430105
_cons	-2.467365	.1925635	-12.81	0.000	-2.844782	-2.089948

.

Figure. 3

The same data set was used to run the probit regression in Fig. 3 and logistic regression in Fig. 1. The idea is to check for similarity or otherwise in the regression results. Howbeit, the regressions are similar except for minor differences in the parameter estimates and z statistics. The result can be interpreted the same way as the logit interpretation. Consequently, logit and probit models are often used interchangeably although probit model is recommended for estimating regressions with multiple continuous exogenous variables, while logit can be used in situations where the dominant exogenous variables are discrete.

Mfx result

```
Marginal effects after probit
      y  = Pr(work) (predict)
         =  .71835948
```

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
children	.1510059	.00922	16.38	0.000	.132939	.169073		1.6445
marita~s*	.150478	.02641	5.70	0.000	.098716	.20224		.6705
educat~n	.0197024	.0037	5.32	0.000	.012442	.026963		13.084
age	.011721	.00142	8.25	0.000	.008935	.014507		36.208

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Figure 4.

3.2 Ordered Logit and Probit models

In the introductory part of the chapter, it was established that ordered logit or probit models are concerned with multiple response variables with natural ordering. In this example (Table 2), the effect of income level, number of disease and age were tested on the health status of an individual. The result reveals that health status rises with increase in income and reduces with increases in disease and age. The output /cut1 compares the first two health outcomes that were initially coded in stata (i.e. poor and good) while /cut2 compares the second and third outcomes (good and excellent) in this order of ranking. Their coefficients show that they are significant and as such, justify their usage.

Table 2. **Determinants of health status**

age	Ndisease	Linc	hlthstat
43.8775	13.7319	9.52878	2
17.5914	13.7319	9.52878	3
15.4997	13.7319	9.52878	3
44.1431	13.7319	9.52878	2
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
61.4873	4.3	8.68286	2
15.59	13	8.68286	2
60.2772	13	8.68286	2
15.2231	4.3	8.1879	3
46.4292	0	8.1879	3
38.1116	8.7	8.1879	3

```
. ologit hlthstat linc ndisease age

Iteration 0:   log likelihood = -5140.0463
Iteration 1:   log likelihood = -4776.0079
Iteration 2:   log likelihood = -4769.8692
Iteration 3:   log likelihood = -4769.8524
Iteration 4:   log likelihood = -4769.8524

Ordered logistic regression                     Number of obs   =       5574
                                                LR chi2(3)      =     740.39
                                                Prob > chi2     =     0.0000
Log likelihood = -4769.8524                     Pseudo R2       =     0.0720
```

hlthstat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
linc	.2836537	.0231098	12.27	0.000	.2383593	.3289482
ndisease	-.0549905	.0040692	-13.51	0.000	-.0629661	-.047015
age	-.0292944	.001681	-17.43	0.000	-.0325891	-.0259996
/cut1	-1.395981	.2061301			-1.799988	-.9919731
/cut2	.9513089	.2054301			.5486732	1.353944

Figure. 5

Mfx result

```
. mfx, predict(outcome(3))

Marginal effects after ologit
      y  = Pr(hlthstat==3) (predict, outcome(3))
         =  .53747615
```

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
linc	.0705151	.00575	12.26	0.000	.05924	.08179		8.69693
ndisease	-.0136704	.00101	-13.50	0.000	-.015655	-.011686		11.2053
age	-.0072825	.00042	-17.43	0.000	-.008101	-.006463		25.5761

Figure 6

The marginal effect result shows that after an average income level of nine thousand naira, a unit increase in income will boost health status by 0.70 and after an average number of 11 diseases, a further increase in disease will reduce the health status of the individual by 0.13.

3.3 Multinomial Logit and Probit models

Multinomial models were also introduced and the emphasis was on multiple response variables without natural ordering. In this example (Table 3), regressands such as price, crate and income were used to test choice of fishing mode. The result in Fig. 7 shows that the choice of selecting beach (mode1) instead of charter (base outcome) increases with income and reduces with price and crate. Whereas, in Fig.8 where the base outcome was set to 3, the choice of beach as a fishing mode instead of the base outcome (pier) increases with all three exogenous variables.

Table 3. Determinants of fishing mode

Mode	Price	Crate	Income
Charter	182.93	0.5391	7.083332
Charter	34.534	0.4671	1.25
Private	24.334	0.2413	3.75
Pier	15.134	0.0789	2.083333
private	41.514	0.1082	4.583332
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
private	16.722	0.2413	7.083332
Pier	17.862	0.0451	1.25
Pier	15.134	0.0789	2.083333
beach	74.514	0.2537	8.750001
Pier	33.534	0.0789	5.416667
beach	48.114	0.1049	5.416667

```
. mlogit mode1 price crate income

Iteration 0:   log likelihood = -1497.7229
Iteration 1:   log likelihood = -1275.5552
Iteration 2:   log likelihood = -1266.1896
Iteration 3:   log likelihood = -1265.9942
Iteration 4:   log likelihood = -1265.9939
Iteration 5:   log likelihood = -1265.9939

Multinomial logistic regression                 Number of obs   =       1182
                                                LR chi2(9)      =     463.46
                                                Prob > chi2     =     0.0000
Log likelihood = -1265.9939                     Pseudo R2       =     0.1547
```

mode1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
price	-.0227022	.0033333	-6.81	0.000	-.0292353	-.0161691
crate	-1.411544	.2759567	-5.12	0.000	-1.952409	-.8706784
income	.1788935	.0489054	3.66	0.000	.0830407	.2747462
_cons	-.1588518	.2367121	-0.67	0.502	-.6227989	.3050953
2	(base outcome)					
3						
price	-.0257278	.0033493	-7.68	0.000	-.0322924	-.0191633
crate	-2.360833	.3501235	-6.74	0.000	-3.047062	-1.674603
income	.050279	.0510976	0.98	0.325	-.0498704	.1504284
_cons	.9248018	.2267411	4.08	0.000	.4803973	1.369206
4						
price	-.0191449	.0021019	-9.11	0.000	-.0232645	-.0150253
crate	-2.711124	.2793064	-9.71	0.000	-3.258554	-2.163693
income	.2482001	.0382806	6.48	0.000	.1731715	.3232287
_cons	.829194	.183051	4.53	0.000	.4704207	1.187967

.

Fig. 7

- To change the base outcome, use the command:
- ***Mlogit mode price crate income, baseoutcome(3)*** ↲

```
. mlogit mode1 price crate income, baseoutcome(3)

Iteration 0:   log likelihood = -1497.7229
Iteration 1:   log likelihood = -1275.5552
Iteration 2:   log likelihood = -1266.1896
Iteration 3:   log likelihood = -1265.9942
Iteration 4:   log likelihood = -1265.9939
Iteration 5:   log likelihood = -1265.9939

Multinomial logistic regression                 Number of obs   =       1182
                                                LR chi2(9)      =     463.46
                                                Prob > chi2     =     0.0000
Log likelihood = -1265.9939                     Pseudo R2       =     0.1547
```

mode1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
price	.0030256	.0041797	0.72	0.469	-.0051664	.0112176
crate	.9492889	.4023005	2.36	0.018	.1607944	1.737783
income	.1286145	.0551442	2.33	0.020	.0205339	.2366951
_cons	-1.083654	.2584121	-4.19	0.000	-1.590132	-.5771752
2						
price	.0257278	.0033493	7.68	0.000	.0191633	.0322924
crate	2.360833	.3501235	6.74	0.000	1.674603	3.047062
income	-.050279	.0510976	-0.98	0.325	-.1504284	.0498704
_cons	-.9248018	.2267411	-4.08	0.000	-1.369206	-.4803973
3	(base outcome)					
4						
price	.0065829	.0033194	1.98	0.047	.0000771	.0130888
crate	-.3502911	.3796472	-0.92	0.356	-1.094386	.3938037
income	.1979211	.0453163	4.37	0.000	.1091028	.2867395
_cons	-.0956078	.2068841	-0.46	0.644	-.5010933	.3098777

Figure 8

Mfx result

```
Marginal effects after mlogit
      y  = Pr(mode1==4) (predict, outcome(4))
         =  .33192907
```

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
price	-.0021523	.00041	-5.27	0.000	-.002953	-.001352		52.082
crate	-.4377057	.05231	-8.37	0.000	-.540233	-.335178		.389368
income	.0452243	.00678	6.67	0.000	.031935	.058513		4.09934

Fig. 9

The marginal effect shows that after an average price of 52 naira, an additional increase in price reduces the chance of choosing fishing modes such as: beach, charter and pier instead of private (mode 4). Ideally, marginal effects should be computed for each of the modes.

- Also use the following command to compute the outcome for the marginal elasticity;
- ***Mfx, predict(outcome(4))*** ↲

4.0 CONCLUSSION

When dealing with Probit regressions, the following parameters are usually estimated: Z-scores, Marginal effects Conditional Probability.

5.0 SUMMARY

In this unit you learnt ordered Logit and Probit mod and Multinomial Logic and Probit models. In the next unit we shall discuss the autoregressive process.

6.0 TUTOR-MARKED ASSIGNMENT

A researcher is interested in how variables, such as GPA (grade point average) and GQES (graduate qualifying exam scores) affect admission into National Open University of Nigeria's graduate school (GSA). The response variable (GSA), admit/don't admit, is a binary variable. Design a questionnaire and Probit equation for this problem.

6.0 REFERENCES AND FURTHER READING

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

UNIT 2. AUTOREGRESSIVE (AR) PROCESS

CONTENTS

1. INTRODUCTION

2. OBJECTIVES

3.0 MAIN CONTENT

3.1 Meaning of the Term Autoregressive (AR)

3.2 Estimation of an Autoregressive Model (AR)

3.3 Autocorrelation or Serial Correlation

3.4 Consequences of Serial Correlation

3.5Testing for serial correlation

3.6 Steps for Carrying Out the LM Test

4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

7.0 INTRODUCTION

In the preceding unit, you learnt Probit models. In the present unit, we will discuss autoregressive models. In regression analysis involving time series data, if the regression model includes one or more lagged values of the dependent variable among its explanatory variables, it is called autoregressive model.

2.0 OBJECTIVES

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

*Discuss the meaning of the term Autoregressive (AR)

*Explain estimation of an Autoregressive Model (AR)

* State autocorrelation or Serial Correlation

* Analyze consequences of Serial Correlation

* State steps for carrying out the LM Test

3.0 MAIN CONTENTS

3.1 Meaning of the Term Autoregressive (AR)

The term autoregressive (AR) describes a random or stochastic process used in econometrics through which future values are estimated based upon a weighted sum of previous or past values. The “auto” signals autoregressive models are regression of variable in question against itself.

If y_t is univariate for instance $y_{t-1}, y_{t-2,} \cdots y_{t-p}$ the model is called autoregressive AR(P) on the other hand if y_t is multivariate, i.e $y_t = [y_{1t}, y_{2t}, \cdots y_{Nt}]$, the model is called vector autoregressive model VAR (P)

The AR model has proven to be useful for describing the dynamic behaviour of economic and financial time series and forecasting. It often provides superior forecast and are quite flexible because they can be made conditional on the potential future paths of specified variables in the model.

The AR is also used for structural inference and policy analysis. In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the model are summarized Wooldridge (2013)

3.2 Estimation of an Autoregressive Model (AR)

Consider the $AR(1)$ model for y_t

$$y_t = \beta y_{t-1} + \varepsilon_t \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

The model in equation (1) above is the first order autoregressive

Where $E(\varepsilon_t|y_{t-1}) = 0, E(\varepsilon_t^2|y_{t-1}) = \sigma^2$ and $E(\varepsilon_t\varepsilon_s|y_{t-1}, y_{s-1}) = 0$ for $t \neq s$

$$\varepsilon_t \sim IID(0, \sigma^2)$$

Here

$$b = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2} = \beta + \frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

b is not unbiased because

$$E\left(\frac{\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}\right) = E\left(\sum_{t=1}^{T} \frac{y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}\varepsilon_t\right) \neq \left(\frac{E\sum_{t=1}^{T} y_{t-1}\varepsilon_t}{E\sum_{t=1}^{T} y_{t-1}^2}\right) = 0$$

However, provided that $plim T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = 0$, then b is consistent because

$$plim\ b = \beta + \left(\frac{plim T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t}{plim T^{-1} \sum_{t=1}^{T} y_{t-1}^2}\right) = \beta$$

The important feature here is that y_{t-1} and ε_t are uncorrelated. And the OLS estimator would not be consistent if for example ε_t was a MA(1) process because y_{t-1} and ε_t would then be correlated. If OLS is used in the model with lagged dependent variables, then it is important to test for serial correlation in the disturbances (Wooldridge 2013)

We should proceed as follows to estimate the model. Consider the model below

$$y_t = x_t'\beta + \varepsilon_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (2)$$

Where the vector x_t may include lagged values of y_t, we find evidence of first order serial correlation of the form

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (3)$$

We could attempt to estimate the model allowing for serially correlated disturbance. However, we need to ask why is there autocorrelation in the disturbances? This can often arise because of:

- Misspecification (fitting a linear model when y and x are related non-linearly)
- Neglected dynamics (not including enough lags of y_t and x_t. The problem can be eliminated by modelling the dynamics by including lagged variables as regressors, rather than simply confirming dynamics to the error term. Wooldridge (2013)

For example, notice that the model in equation (2) and (3) can be rewritten as

$$y_t = \rho y_{t-1} + x_t'\beta - x_{t-1}'\rho\beta + u_t$$

More generally, this approach leads to models of the form

$$y_t = \rho_1 y_{t-1} + \cdots\cdots + \rho_p y_{t-p} + x_t'\beta + \varepsilon_t$$

Where x_t can contain current and lagged values of the other regressors.

In general, the ultimate purpose is to specify a model that incorporates all the relevant information available at period t. That is a model such that

$$E[\varepsilon_t|x_t, y_{t-1}, x_{t-1}, y_{t-2}, \cdots] = 0$$

3.3 Autocorrelation or Serial Correlation

It is now a common practice to treat the terms autocorrelation and serial correlation synonymously although there may be technical differences. But for this course, we will use both concepts interchangeably.

We shall consider the linear regression model

$$y_t = x_t' + \varepsilon_t, \qquad t = 1, \cdots\cdots, T$$

We are using a t subscript to index observations in time, a superscript on x(‘) indicating a vector of explanatory variables and T denotes sample size. We shall assume the following:

$E(\varepsilon|X) = 0$ with $E(\varepsilon\varepsilon'|X) = \sigma^2\Omega$

Where Ω is a matrix with the following elements

$$E(\varepsilon\varepsilon'|X) = \sigma^2\Omega = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 \end{pmatrix}$$

Where the autocovariances are defined as $cov(\varepsilon_t, \varepsilon_{t-s}|X) = cov(\varepsilon_{t+s}, \varepsilon_t|X) = \gamma_s$

Note that $\gamma_0 = cov(\varepsilon_t, \varepsilon_t|X) = var(\varepsilon_t|X) = \sigma^2$ (Wooldridge (2013)

The autocovariances depend on the unit of measurement, so we cannot tell whether a value of 0.1 is large or small without further information. Autocorrelations do not depend on the units of measurement, and are defined as

$$\rho_s = \frac{\gamma_s}{\gamma_o}$$

Note that $\rho_0 = 1$ and that $-1 \leq \rho_s \leq 1$ for all s

It is obviously impractical to try to estimate all T autocovariances $(\gamma_0, \cdots, \gamma_{T-1})$ or autocorrelations from a sample of T observations. Therefore we try to model the autocovariances or autocorrelations in terms of a small number of parameters.

The first order autoregressive process, or AR(1) given below

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, |\rho| < 1,$$

Where $E(u_t|X) = 0, E(u_t^2|X) = \sigma_u^2,\ \ E(u_t u_s|X) = 0$ for $t \neq s$

By repeated backward substitution:

$$\begin{aligned}
\varepsilon_t &= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t \\
&= \rho^2\varepsilon_{t-2} + u_t + \rho u_{t-1} \\
&= \rho^2(\rho\varepsilon_{t-3} + u_{t-2}) + u_t + \rho u_{t-1} \\
&= \rho^3\varepsilon_{t-3} + u_t + \rho u_{t-1} + \rho^2 u_{t-2} \\
&\vdots \\
&= \rho^s\varepsilon_{t-s} + u_t + \rho u_{t-1} + \cdots + \rho^{s-2}u_{t-s+2} + \rho^{s-1}u_{t-s+1}
\end{aligned}$$

$$i.e, \varepsilon_t = \rho^s\varepsilon_{t-s} + \sum_{i=0}^{s-1}\rho^i u_{t-i} (s > 0)$$

It is sometimes convenient to let $s \longrightarrow \infty$ to obtain

$$\varepsilon_t = \sum_{i=0}^{\infty}\rho^i u_{t-i},$$

because $\lim_{s\to\infty}\rho^s = 0$ due to $|\rho| < 1$ (Wooldridge 2013)

Note that ε_t depends on current and all lagged values of u_t

We can show that:

$$E(\varepsilon_t|X) = 0,$$

$$\gamma_0 = var(\varepsilon_t|X) = \frac{\sigma_u^2}{1-\rho^2}$$

$$\gamma_s = cov(\varepsilon_t, \varepsilon_{t-s}|X) = \frac{\rho^s\sigma_u^2}{1-\rho^2}$$

hence $\rho_s = \frac{\gamma_s}{\gamma_0} = \rho^s$

The AR(1) process can be extended to the AR(p) process which includes p lags of ε_t.

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \cdots + \rho_p \varepsilon_{t-p} + u_t$$

3.4 Consequences of Serial Correlation

1. The OLS estimators are still unbiased and consistent.
2. The OLS estimators will be inefficient and therefore no longer best linear unbiased estimator (BLUE)
3. The estimated variances of the regression coefficients will be biased and inconsistent, and therefore hypothesis testing is no longer valid. In most of the cases, the R^2 will be overestimated and the t-statistics will tend to be higher

SELF-ASSESEMENT EXERCISE

Discuss the consequences of serial correlation

3.5 Testing for serial correlation

Usually we estimate a linear model by ordinary least squares (OLS) assuming that the classical assumptions holds, and then attempt to test whether those assumptions appear to be satisfied for the estimated model.

We therefore need to be able to test whether the assumption of zero covariance of the disturbances appears to hold, based on the estimated residuals.

One method commonly used in applied econometric research for detecting autocorrelation is to plot the regression residuals, e, against time. If the residuals in successive periods show a regular time pattern (for example, a saw tooth pattern, or a cyclical pattern) we conclude that there is autocorrelation.

Until the 1990s, the most commonly used test for serial correlation was the **Durbin-Watson test** for first order autocorrelation. Nowadays, the **Lagrange Multiplier (LM)** test is more popular because it can be applied in a wider set of circumstances and can test for higher-order serial correlation such as AR(1), AR(2), AR(3), etc.

The null and alternative hypotheses for the Breusch-Godfrey LM test are as follows:

H_0 : no serial correlation in ε_t,

H_1 : ε_t is a AR(p)

Notice that because an LM test is used, we do not need to be specific about the nature of the serial correlation process under the alternative. The test statistic is mostly easily calculated by estimating the following auxiliary regression:

$$e_t = x_t'\gamma + \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p} + u$$

The test statistic is simply $(T - P)\ R^2$ and under H_0, we have as $T \longrightarrow \infty$,

$$LM = (T - P)R^2 \sim \chi_p^2$$

note that the R^2 is from the auxiliary regression above

3.6 Steps for Carrying Out the LM Test

Step 1. Estimate the regression model by OLS and compute its estimated residuals, ε_t.

Step 2. The LM statistic can be calculated by $(T - P)R^2$. Where R^2 is the R-squared from the auxiliary regression. $(T - P)$ is used because the efficient number of observations is $(T - P)$. Where P is the number of lagged residuals

Step 3. Reject the null hypothesis of zero autocorrelation in favour of the alternative that $\rho \neq 0$ if $(T - P)R^2 > \chi^2_{1,(1-\alpha)}$, the value of χ_1^2, in the chi-square distribution with 1 d.f such that the area to the right of it is $(1 - \alpha)$ and α is the significance level.

Note: the LM test is a large sample test and would need at least 30 d.f to be meaningful.

Example

1. A wage (y_t) productivity (x_t) model $y_t = \beta_1 + \beta_2 \mathrm{x}_t + \varepsilon_t$, and an annual data from 1959 to 1998 was used in the regression. To test for serial correlation, we regress the variables using OLS, obtain the residuals and then regress the residuals on lagged residuals going back six periods i.e $AR(6)$. Our auxiliary regression becomes

$$e_t = \alpha + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4} + \rho_5 e_{t-5} + \rho_6 e_{t-6} + u$$

The data and the corresponding stata command are presented. Students are adviced to try the example personally. The outcome of the auxiliary regression is given below.

Source	SS	df	MS
Model	157.931273	6	26.3218789
Residual	30.8666819	27	1.14321044
Total	188.797955	33	5.72115016

Number of obs	= 34
F(6, 27)	= 23.02
Prob > F	= 0.0000
R-squared	= 0.8365
Adj R-squared	= 0.8002
Root MSE	= 1.0692

r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lag1	1.003894	.1914144	5.24	0.000	.6111443	1.396644
lag2	-.0614921	.2703019	-0.23	0.822	-.6161059	.4931216
lag3	-.1186325	.2703319	-0.44	0.664	-.6733077	.4360427
lag4	.2545361	.270836	0.94	0.356	-.3011734	.8102457
lag5	-.0276549	.2751347	-0.10	0.921	-.5921848	.536875
lag6	-.1313361	.2017945	-0.65	0.521	-.5453842	.282712
_cons	-.0375935	.1976633	-0.19	0.851	-.4431652	.3679782

Total number of observation $= T = 40$, number of lagged residuals $= P = 6, DF = 34$

$$LM = (40 - 36)0.84 = 28.56$$

The critical value of X^2 at 5% is 43.77. We therefore do not reject the null hypothesis of no serial correlation

observation	Y	X
1959	58.5	47.2
1960	59.9	48.0
1961	61.7	49.8
1962	63.9	52.1
1963	65.3	54.1
1964	67.8	54.6

1965	69.3	58.6
1966	71.8	61.0
1967	73.7	62.3
1968	76.5	64.5
1969	77.6	64.8
1970	79.0	66.2
1971	80.5	68.8
1972	82.9	71.0
1973	84.7	73.1
1974	83.7	72.2
1975	84.5	74.8
1976	87.0	77.2
1977	88.1	78.4
1978	89.7	79.5
1979	90.0	79.7
1980	89.7	79.8
1981	89.8	81.4
1982	91.1	81.2
1983	91.2	84.0
1984	91.5	86.4
1985	92.8	88.1
1986	95.9	90.7
1987	96.3	91.3
1988	97.3	92.4
1989	95.8	93.3
1990	96.4	94.5
1991	97.4	95.9
1992	100.0	100.0
1993	99.9	100.1
1994	99.7	101.4
1995	99.1	102.2
1996	99.6	105.2
1997	101.1	107.5
1998	105.1	110.5

Declare data time series	tsset observation, yearly
OLS regression of wage and productivity	regress Y X
Obtain residuals	Predict r, residuals
Generate the lag residual for first period	gen lag1 = r[_n-1]
Generate the lag residual for the second period	gen lag2 = r[_n-2]
Lag third period	gen lag3 = r[_n-3]
Lag fourth period	gen lag4 = r[_n-4]
Lag fifth period	gen lag5 = r[_n-5]
Lag sixth period	gen lag6 = r[_n-6]

Auxiliary regression regress r lag1 lag2 lag3 lag4 lag5 lag6

4.0 CONCLUSION

Autoregressive (AR) models describe a random or stochastic process used in econometrics through which future values are estimated based upon a weighted sum of previous or past values. The “auto” signals autoregressive models are regression of variable in question against itself.

5.0 SUMMARY

In this unit you have learn the meaning of the term Autoregressive (AR) and how to estimate of an Autoregressive Model (AR). You also learn autocorrelation or Serial Correlation and the consequences of Serial Correlation as well as Testing for serial correlation and steps for carrying out the LM Test. In the next unit which is Unit 3 of our Module 2 we shall discuss the concept of stationarity.

6.0 TUTOR MARKED ASSIGNMENT

The annual time series data below is for inventories and sales in the US manufacturing, 1950 – 1991

Year	Sales	inventories
1950	38596	59822
1951	43356	70242
1952	44840	72377
1953	47987	76122
1954	46443	73175
1955	51694	79516
1956	54063	87304
1957	55879	89052
1958	54201	87055
1959	59729	92097
1960	60827	94719
1961	61159	95580
1962	65662	101049

1963	68995	105463
1964	73682	111504
1965	80283	120929
1966	87187	136824
1967	90918	145681
1968	98794	156611
1969	105812	170400
1970	108352	178594
1971	117023	188991
1972	131227	203227
1973	153881	234406
1974	178201	287144
1975	182412	288992
1976	204386	318345
1977	229786	350706
1978	260755	400929
1979	298328	452636
1980	328112	510124
1981	356909	547169
1982	348771	575486
1983	370501	591858
1984	411427	651527
1985	423940	665837
1986	431786	664654
1987	459107	711745
1988	496334	767387
1989	522344	813018
1990	540788	835985
1991	533838	828184

a. Estimate the regression using OLS and find the residuals
b. Derive the auxiliary regression equation using a lag residuals of 4 period i.e AR(4)
c. Regress the residuals with the lag residuals of the respective periods and find out if serial correlation exists using the Lagrange Multiplier test.

7.0 REFERENCES/FURTHER READING

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Ilorin, Nigeria.

UNIT 3: CONCEPT OF STATIONARITY

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 What is Stationarity?
- 3.2 Unit Roots In The Ar(1) Model
- 3.3 Testing For Unit Roots
- 3.4 Cointegration
- 3.4.1Testing for Cointegration
- 3.5 Test For Stationarity Of The Variables Using Eviews Software

4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
REFERENCES/FURTHER READING

1.0 INTRODUCTION

In econometric analysis the use of stationarity is an inevitable way of ensuring that that the data used for regression is reliable and can be used for prediction of future performance. It help to ensure that the data that will be used for regression is mean reverting and stable.

2. OBJECTIVES

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

* Discuss stationarity
* Conduct unit roots test in the Ar(1) Model
* Conduct cointegration
* Estimate stationarity of variables in Eviews software

3.0 MAIN CONTENT

3.1 What is Stationarity

From the assumptions below

$E(\varepsilon_t|y_{t-1}) = 0, E(\varepsilon_t^2|y_{t-1}) = \sigma^2$ and $E(\varepsilon_t\varepsilon_s|y_{t-1}, y_{s-1}) = 0$ for $t \neq s$

There has been some notion of stationarity. But many time series are not stationary but contain deterministic/stochastic trends. By definition, a time series is stationary if it is not explosive or trending and not wandering aimlessly without returning to its mean. Mathematically,

Woodridge (2003) stated that A time series $\{y_t\}$ $(t = 1, \cdots, T)$ is said to be **stationary** if:

i. $E(y_t) = \mu$ (constant does not depend on t)

ii. $var(y_t) = \sigma^2 < \infty$ (constant – does not depend on t)

iii. $cov(y_t, y_s) = \lambda_{|t-s|}$ (only depends on $|t - s|$ and not on t or s alone)

The most fundamental stationary process is **white noise**, which we shall denote ε_t and which has the properties:

$$E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t\varepsilon_s) = 0 \ (t \neq s)$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

3.2 Unit Roots in the AR(1) model

Consider the AR(1) process

$$y_t = \gamma y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim W\ N(0, \sigma^2) \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \qquad (4)$$

More generally, for AR(p) process of the form $y_t = \gamma_1 y_{t1} + \cdots + \gamma_p y_{tp} + \varepsilon_t$, we need to consider the roots of the equation

$$\gamma(z) = 1 - \gamma_1 z - \cdots - \gamma_p z^p = 0$$

If the roots lie outside the unit circle (modulus greater than one) the process is **stationary**; if is equal to one, there is **unit root**; while if inside the unit circle (modulus less than one) the process is **explosive**.

The roots of this autoregressive process are nothing but the roots obtained from.

$$\gamma(z) = 0$$

In order to determine whether a process is stationary, nonstationary/explosive or has a unit root, the following rules apply:

- If any of the roots of $\gamma(z) = 0$ is less than 1 in modulus, then we have a nonstationary process.
- If any of the roots of $\gamma(z) = 0$ is equal to 1 in modulus, then we have a process that contains a unit root.
- If all roots of $\gamma(z) = 0$ are greater than one in modulus, then we have a stationary process.

Note: obviously for the quadratic case the roots are obtain from

$$z^* = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Examples

Determine if the following AR processes are stationary, unit root or explosive

i. $y_t = 0.5y_{t-1} + \varepsilon_t$

ii. $y_t = 1.5y_{t-1} + \varepsilon_t$

iii. $y_t = y_{t-1} - 0.25y_{t-2} + \varepsilon_t$

iv. $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \varepsilon_t$

Solution

i. $y_t = 0.5y_{t-1} + \varepsilon_t \Rightarrow \varepsilon_t = y_t - 0.5y_{t-1}$

$\varepsilon_t = \gamma(z)y_t$ Where $\gamma(z) = 1 - 0.5z$

$$1z^0 - 0.5z = 0 \Rightarrow z = \frac{1}{0.5} = 2$$

$$|2| > 0$$

This process has only one root which is greater than 1 in modulus $\Longrightarrow y_t$is a stationary process

ii. $y_t = 1.5y_{t-1} + \varepsilon_t \Longrightarrow \varepsilon_t = y_t - 1.5y_{t-1}$

$\varepsilon_t = \gamma(z)y_t \quad \Longrightarrow \quad \gamma(z) = 1z^0 - 1.5z$

$z = \frac{2}{3}$; $\left|\frac{2}{3}\right| < 1$

y_t is a non-stationary process

iii. $y_t = y_{t-1} - 0.25y_{t-2} + \varepsilon_t \Longrightarrow \varepsilon_t = y_t - y_{t-1} + 0.25y_{t-2}$

$\gamma(z) = z^0 - z^1 + 0.25z^2$

$\gamma(z) = 0 \Longrightarrow 0.25z^2 - z + 1 = 0$

$z = 2$; $|2| > 1$

Solving for the roots shows that y_t is a stationary process.

iv. $y_t = 1.5y_{t-1} - 0.5y_{t-2} + \varepsilon_t \Longrightarrow \varepsilon_t = y_t - 1.5y_{t-1} + 0.5y_{t-2}$

$\gamma(z) = 0.5z^2 - 1.5z + 1 = 0$

Solving for the roots gives $z_1 = |2| > 1;\ z_2 = 1$

This process has two roots one of which is equal to 1 in modulus $\Longrightarrow y_t$ is a process with one unit root (Random walk), therefore it is nonstationary.

SELF-ASSESEMENT EXERCISE

Find out if the following AR processes are stationary, unit root or nonstationary/explosive.

i. $y_t = 0.25y_{t-1} + \varepsilon_t$

ii. $y_t = y_{t-1} - 0.5y_{t-2} + \varepsilon_t$

iii. $y_t = 0.5y_{t-1} - 0.8y_{t-2} + \varepsilon_t$

Note: in regressing a time series variable on another time series variable(s), sometimes we expect no relationship between variables yet it often shows a significant relationship. This is a situation of **spurious regression**. It usually occurs when we regress time series data that are non stationary.

Explosive series are not frequent in economics and consequently we focus on the case $\gamma = 1$, which leads to $y_t = y_{t-1} + \varepsilon_t$ being a **random walk (RW)**. By repeated substitution, we find that

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= (y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &\vdots \\ \Longrightarrow\ y_t &= y_0 + \textstyle\sum_{i=1}^{t} \varepsilon_i \end{aligned}$$

Therefore, $E(y_t) = y_0$ (which we assume to be fixed) and $var(y_t) = t\sigma^2 \rightarrow \infty\ as\ t \rightarrow \infty$

Often it is appropriate to include an intercept in the model, so that equation (4) becomes

$$y_t = \mu + \gamma y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim W\ N(0, \sigma^2)$$

If $\gamma = 1$ then $\Delta y_t = \mu + \varepsilon_t$ **(RW with drift)** and changes in y_t are equal to a constant μ plus a random component ε_t (and they are stationary). Sometimes a linear trend is also appropriate, in which equation (4) becomes

$$y_t = \mu + \beta t + \gamma y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim W\ N(0, \sigma^2)$$

If $\gamma = 1$then $\Delta y_t = \mu + \beta t + \varepsilon_t$ **(RW with drift and trend)** and changes in y_t are equal to a linear trend $\mu + \beta t$ plus a random component ε_t (stationary around the trend).

A process with a unit root is often called **integrated to order one**, or **I(1)** because it requires differencing once to become stationary. In this terminology, a stationary series is I(0) because it does not need differencing to become stationary.

3.3 Testing for Unit Roots

For the AR(1) model $y_t = \gamma y_{t-1} + \varepsilon_t$, subtracting y_{t-1} from both sides leads to $\Delta y_t \equiv y_t - y_{t-1} = \gamma^* y_{t-1} + \varepsilon_t,\ \gamma^* = \gamma - 1$

Note that $-1 < \gamma < 1 \Longrightarrow\ -1 < 1 + \gamma^* < 1 \Longrightarrow\ -2 < \gamma^* < 0$

Testing for a unit root is therefore equivalent to testing

$H_o: \gamma^* = 0$ unit root

Against $H_1: \gamma^* < 0$ stationary AR(1)

This is most easily carried out using a t-test. But y_t is nonstationary under H_o, and so the usual limiting (normal) distribution does not apply for the t-ratio. The appropriate distribution is the **Dickey-Fuller** distribution, for which critical values are tabulated.

The test statistic is

$$DF = \frac{\acute{\gamma}^*}{se(\acute{\gamma}^*)}$$

Where se( ) denotes the standard error.

Note that this is one sided test – we are looking for significantly negative values of DF in order to reject H_o.

Let $\overline{DF}$ denote the critical value (where $\overline{DF} < 0$), the decision rule for the test is:

If $DF\ \leq\ \overline{DF}$ reject H_o ($y_t$is stationary)

If $DF\ >\ \overline{DF}$ do not reject H_o ($y_t$is nonstationary)

This test procedure can be adapted for the case in which the model has a drift or a trend. For the case with a drift, the test regression becomes

$$\Delta y_t = \mu + \gamma^* y_{t-1} + \varepsilon_t$$

and the same testing procedure applies (but with different critical values)

With a drift and a trend, the test regression becomes

$$\Delta y_t = \mu + \beta t + \gamma^* y_{t-1} + \varepsilon_t$$

again the same testing procedure applies (but with different critical values)

if an AR(p) model is more appropriate than an AR(1), we can augment the test regression with $p-1$ lags of Δy_t

$$\Delta y_t = \gamma^* y_{t-1} + \sum_{i=1}^{p=1} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

$$\Delta y_t = \mu + \gamma^* y_{t-1} + \sum_{i=1}^{p=1} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

$$\Delta y_t = \mu + \beta t + \gamma^* y_{t-1} + \sum_{i=1}^{p=1} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

This is the **augmented Dickey-Fuller**, or **ADF** test and can also have the effect of removing serial correlation from the residuals (desirable because the critical values depend on ε_t being the white noise). Augmenting the test does not change the critical values to use.

Examples

A researcher with a sample size of 100 observations performed an ADF test and obtained the following results (standard errors in parentheses). What can you conclude about the stationarity of y_t? The 5% critical value for the test is

-2.89

1. $\hat{\Delta} y_t = 5.16 - 0.78 y_{t-1} - 0.02 \Delta y_{t-1}$
   (0.71) (0.10) (0.08)
2. $\hat{\Delta} y_t = 1.80 - 0.65 y_{t-1}$
   (0.29) (0.65)
3. $\hat{\Delta} y_t = 0.53 - 0.94 y_{t-1} + 0.14 \Delta y_{t-1} - 0.1 \Delta y_{t-2}$
   (0.43) (0.30) (0.04) (0.04)

Solution

1. $ADF = \frac{\hat{\gamma}^*}{Se\,(\hat{\gamma})} = \frac{-0.78}{0.10} = -7.8$

   Since $DF \leq \overline{DF}$, at 5% critical value the process is stationary

2. $ADF = \frac{\hat{\gamma}^*}{Se\,(\hat{\gamma})} = \frac{-0.65}{0.11} = -5.91$

   The process is stationary

3. $ADF = \frac{\hat{\gamma}^*}{Se\,(\hat{\gamma})} = \frac{-0.94}{0.3} = -3.1$

   The process is stationary

Exercises

i. Using the Dickey-Fuller test, establish whether x_t is stationary or random walk at 1% critical and standard errors in parentheses.

$$\Delta x_t = 1.8 - 0.65x_{t-1}$$
$$(0.29) \quad (0.11)$$

ii. Establish whether z_t is stationary or a random walk at 5% critical

$$\Delta z_t = 0.51 + 0.45z_{t-1} + 0.2\Delta z_{t-1}$$
$$(0.42) \qquad (0.09) \qquad (0.03)$$

iii. $x_t = 1.4 + 0.75_t - 0.56x_{t-1}$

$$(0.40) \qquad (0.19) \qquad (0.08)$$

At 1% critical, what can you conclude about the stationarity of x_t?

iv. $x_t = 0.07 - 0.91x_{t-1} + 0.64\Delta x_{t-1}$

At 5% critical, determine whether x_t is stationary or not?

3.4 Cointegration

Suppose we have two independent random walks:

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2)$$
$$x_t = x_{t-1} + u_{t,} \qquad u_t \sim WN(0, \omega^2)$$

Where $E(\varepsilon_t u_s) = 0$ for all t and s

Suppose we regress y on x, obtaining

$$y_t = b_1 + b_2 x_t + e_t$$

We would then have:

a. $t_2 = \frac{b_2}{se(b_2)}$ to be small, thereby not rejecting $H_0: \beta_2 = 0$
b. R^2 to be low (y and x are independent)

However, because both y_t and x_t are $I(1)$ (random walks), the results we have seen so far are invalid. This leads to what is known as **spurious regression** which we earlier mentioned.

We find that

a. $t_2 \longrightarrow \infty \; as \; T \longrightarrow \infty$, i.e, the larger the sample, the higher the probability of rejecting $H_0: \beta_2 = 0$
b. R^2 will be acceptable, suggesting a reasonable fit

Therefore, we have to be extremely careful when estimating regressions with $I(1)$ data. A characteristics of spurious regression is that e_t will also be $I(1)$.

Suppose we run a regression involving only $I(1)$ regressors and find e_t to be stationary, i.e, e_t is $I(0)$. What this suggest about the relationship between y_t and x_t is that the linear combination $y_t - b_1 - b_2 x_t$ of the two $I(0)$ series is stationary. In this case, y_t and x_t are said to be **cointegrated**.

Many time series in economics have been found to be $I(1)$, and so cointegration is important for economic theory. In particular, the fact that two series are cointegrated suggests that there is a long-run relationship between them (Green, 2012).

3.4.1Testing for Cointegration

To test for the presence of cointegration between a pair of variables y_t and x_t, the following procedure can be used:

a. Test for the orders of integration of y_t and x_t
b. If both variables are $I(1)$, regress y on x to obtain $y_t = b_1 + b_2 x_t + e_t$
c. Apply the DF/ADF test (without drift or trend) to e_t. If $e_t \sim I(1)$ there is no cointegration, but if $e_t \sim I(0)$, there is evidence of cointegration

Note: a different set of critical values is required for the tests of cointegration (different from the ones for the DF/ADF test without a drift or trend).

This test is called the Engle-Granger test (EG), and it can also be "augmented" if there are signs of serial correlation in the residuals of the auxiliary regression used to perform the test.

3.5 Test For Stationarity of The Variables Using Eviews

Given the following demand for loans by business firms:

$$Qt = \beta_0 + \beta_1 R_t + \beta_2 RD + \beta_3 X_t + \mu_t \text{ ------1}$$

And the supply by banks of commercial loans:

$$Qt = \alpha_0 + \alpha_1 R_t + \alpha_2 RS + \alpha_3 y_t + v_t \text{ ------2}$$

Where:

Qt= Total commercial loans

Rt= Average prime rate charged by the banks

RSt= 3-month Treasury bill rate

RDt= AAA corporate bond rate

Xt= Industrial Production index

Yt= Total bank deposits

Test for stationarity of the variable **Qt**

Solutions:

Time series analysis are said to be stationary if the constant, variance and autocovariance are stable over time. That is they are time invariant. Before conducting stationarity test, there are three forms of non-stationarity that exists:

Random walk with drift: $Yt = \delta + Yt - 1 + \mu t$

Random walk without drift: $Yt = Yo + \mu t$

Random walk with drift and deterministic trend: $Yt = \beta_1 + \beta_2 t + Y_{t-1} + \mu t$

The major test for nonstationarity here shall be conducted using ADF- test statistics.

Before performing the ADF test for unit root, it imperative to ascertain what kind of non-stationarity exist so as to enable us perform the ADF test.

Unit root test for Qt

Step one:

Table: Trend and intercept test

Dependent Variable: Qt

Method: Least Squares

Date: 06/09/15 Time: 14:19

Sample: 1941 2012

Included observations: 72

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	258.5803	1.822135	141.9106	0.0000
@TREND	2.865540	0.044295	64.69152	0.0000

R-squared	0.983549	Mean dependent var	360.3069
Adjusted R-squared	0.983314	S.D. dependent var	60.47077
S.E. of regression	7.811335	Akaike info criterion	6.976414

Sum squared resid	4271.187	Schwarz criterion	7.039654
		Hannan-Quinn	
Log likelihood	-249.1509	criter.	7.001590
F-statistic	4184.993	Durbin-Watson stat	0.215090
Prob(F-statistic)	0.000000		

The random walk that exist here shows that the trend and constant (or intercept) are significant at 5% level (that is, the probability values for both, which are 0.0000 and 0.0000 are less than 0.05), and as such we shall select constant and trend when conducting unit root test for Qt.

Step two:

Table 2; ADF test of Qt at levels

Null Hypothesis: QT has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.593552	0.7859
Test critical values:	1% level	-4.092547	
	5% level	-3.474363	
	10% level	-3.164499	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(QT)

Method: Least Squares

Date: 06/09/15 Time: 14:22

Sample (adjusted): 1942 2012

Included observations: 71 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
QT(-1)	-0.090012	0.056485	-1.593552	0.1157
C	25.87837	14.49420	1.785429	0.0787
@TREND(1941)	0.265970	0.162192	1.639846	0.1057

R-squared	0.038929	Mean dependent var	3.169014
Adjusted R-squared	0.010662	S.D. dependent var	3.609812
S.E. of regression	3.590515	Akaike info criterion	5.435803
Sum squared resid	876.6425	Schwarz criterion	5.531410
Log likelihood	-189.9710	Hannan-Quinn criter.	5.473823
F-statistic	1.377207	Durbin-Watson stat	1.748816
Prob(F-statistic)	0.259228		

The ADF test of Qt at levels shows that the variable has unit root, that is the ADF test statistic is less than the critical values at 1, 5 and 10%.

Table 3; ADF test of Qt at first difference

Null Hypothesis: D(QT) has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.546457	0.0000
Test critical values:	1% level	-4.094550	
	5% level	-3.475305	
	10% level	-3.165046	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(QT,2)

Method: Least Squares

Date: 06/09/15 Time: 14:24

Sample (adjusted): 1943 2012

Included observations: 70 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(QT(-1))	-0.930785	0.123341	-7.546457	0.0000
C	2.559662	0.977096	2.619661	0.0109
@TREND(1941)	0.010572	0.021736	0.486379	0.6283

R-squared	0.459896	Mean dependent var	0.061429
Adjusted R-	0.443773	S.D. dependent var	4.925541

squared

		Akaike info	
S.E. of regression	3.673499	criterion	5.482078
Sum squared resid	904.1378	Schwarz criterion	5.578442
		Hannan-Quinn	
Log likelihood	-188.8727	criter.	5.520355
F-statistic	28.52505	Durbin-Watson stat	1.985956
Prob(F-statistic)	0.000000		

The ADF test of Qt at first difference shows that it is now stationary. That is, the ADF statistic of -7.546457 is greater than the critical values at 5%(that is -7.546457 is greater than the critical value of -3.475305)

4.0 CONCLUSION

The Augmented Dickey Fuller unit root test has been widely accepted and used in econometric analysis. It allows the data to become valid for econometric analysis and reliable for forecasting

5.0 SUMMARY

In this unit you learnt the answer to the question “What is Stationarity?”. You also learnt how to conduct unit roots test in the AR(1) Models as well as testing for cointegration and stationarity of the variables using Eviews software. The mastery of theses concepts now prepared you for the study of panel data in the next unit.

6.0 TUTOR MARKED ASSIGNMENT

Null Hypothesis: QT has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.593552	0.7859
Test critical			
values:	1% level	-2.092547	
	5% level	-3.474363	
	10% level	-3.164499	

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

Interpret the table for the Augmented Dickey Fuller statistics

7.0 REFERENCES/FURTHER READING

Green, W.H. (2012). Econometric Analysis. (7th edition). Pearson Education Limited. England

Gujarati, D.N. (2005). Basic Econometrics. (4th edition). Tata McGraw-Hill Publishing Company Limited. New Delhi

UNIT 4: PANEL DATA REGRESSION MODEL

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 Meaning of Panel Data Regression Model
- 3.2 Panel Data Examples
- 3.3 Advantages of Panel Data
- 4.4 Importance of Panel Data3.5 Format of a Panel Data

3.5 Format of a Panel Data
4.6 Types of Panel Data
4.6.1 Long Versus Short Panel Data
4.6.2 Balanced Versus Unbalanced Panel Data
4.6.3 Fixed Versus Rotating Panel Data
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
8.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the former unit you learnt the answer to the question "What is Stationarity?" You also learnt how to conduct unit roots test in the AR(1) Models as well as testing for cointegration and stationarity of the variables using Eviews software. The mastery of these concepts now prepared you for the study of panel data in this present unit. Panel data is a model which comprises variables that vary across time and cross section, in this paper we will describe the techniques used with this model including a pooled regression, a fixed effect and a random effect.

2.0 OBJECTIVES

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

* Know the meaning of Panel Data Regression Model
* Explain panel Data Examples
* Discuss the advantages of Panel Data
* List the importance of Panel Data3.
* Design the format of a Panel Data
* List types of Panel Data

3.1 Meaning of Panel Data Regression Model

Panel (data) analysis is a statistical method, widely used in social science, epidemiology, and econometrics, which deals with two and "*n*"-dimensional (in and by the - cross sectional/times series time) panel data. The data are usually collected over time and over

the same individuals and then a regression is run over these two dimensions. Multidimensional analysis is an econometric method in which data are collected over more than two dimensions (typically, time, individuals, and some third dimension). Panel data are repeated cross-sections over time, in essence there will be space as well as time dimensions. Panel data are a type of longitudinal data, or data collected at different points in time. To put succinctly, panel data is a combination of both cross sectional data and time series data.

That is, Cross-sectional data (data collected on several individuals/units at one point in time) and time series data (data collected on one individual/unit over several time periods). Panel data are repeated cross-sections over time, in essence there will be space as well as time dimensions. Other names are pooled data, micropanel data, longitudinal data, event history analysis and cohort analysis.

Panel data allows you to control for variables you cannot observe or measure like cultural factors or difference in business practices across companies; or variables that change over time but not across entities (i.e. national policies, federal regulations, international agreements, etc.). This is, it accounts for individual heterogeneity. With panel data you can include variables at different levels of analysis (i.e. students, schools, districts, states) suitable for multilevel or hierarchical modeling.

It is important to note: Time series data: Many observations (large t) on as few as one unit (small N). Examples: stock price trends, aggregate national statistics. Cross sectional data: analysis of data on n, several distinct entities at a given point of time (cross sectional data). Panel data: A data set with both a cross section and a time dimension.

Taken together, the repeated observation of one unit constitutes a “panel”. Other names of panel data are pooled data, micropanel data, longitudinal data, event history analysis and cohort analysis.

3.2 Panel Data Examples

The individuals/units can for example be workers, firms, states or countries. Annual unemployment rates of each state over several years. Quarterly sales of individual stores over several quarters. Wages for the same worker, working at several different jobs.

3.3 Advantages of Panel Data

i. Panel data takes heterogeneity into account, get individual-specific estimates.

ii. Panel data is especially suitable to study dynamics of change.

iii. Studies more sophisticated behavioral models.

iv. Minimize bias due to aggregation.

3.4 Importance of Panel Data

i. We are interested in describing change over time; social change, e.g. changing attitudes, behaviors, social relationships. Individual growth or development, e.g. life-course studies, child development, career trajectories, school achievement. Occurrence (or non-occurrence) of events.

ii. We want *superior estimates* trends in social phenomena. Panel models can be used to inform policy – e.g. health, obesity. Multiple observations on each unit can provide superior estimates as compared to cross-sectional models of association.

iii. We want to estimate causal models, e.g Policy evaluation and Estimation of treatment effects.

SELF-ASSESSEMENT EXERCISE

Discuss the importance of panal data

Some drawbacks are data collection issues (i.e. sampling design, coverage), non-response in the case of micro panels or cross-country dependency in the case of macro panels (i.e. correlation between countries).

3.5 Format of a Panel Data

Below is a typical example of the format a Panel data usually takes.

country	year	Y	X1	X2	X3
1	2000	6.0	7.8	5.8	1.3
1	2001	4.6	0.6	7.9	7.8
1	2002	9.4	2.1	5.4	1.1
2	2000	9.1	1.3	6.7	4.1
2	2001	8.3	0.9	6.6	5.0
2	2002	0.6	9.8	0.4	7.2
3	2000	9.1	0.2	2.6	6.4
3	2001	4.8	5.9	3.2	6.4
3	2002	9.1	5.2	6.9	2.1

Below is a typical example of a panel data equation.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}$$

$$y_{it} = \mathbf{x_{it}}\boldsymbol{\beta} + c_i + u_{it} \qquad i = 1...N \quad t = 1...T$$

4.6 Types of Panel Data

A panel data set contains *n* entities or subjects, each of which includes *T* observations measured at 1 through *t* time period. Thus, the total number of observations in the panel data is *nT*. Ideally, panel data are measured at regular time intervals (e.g., year, quarter, and month). Otherwise, panel data should be analyzed with caution. A panel may be long or short, balanced or unbalanced, and fixed or rotating.

4.6.1 Long Versus Short Panel Data

A short panel has many entities (large n) but few time periods (small T), while a long panel has many time periods (large T) but few entities (Cameron and Trivedi, 2009: 230). Accordingly, a short panel data set is wide in width (cross-sectional) and short in length (time-series), whereas a long panel is narrow in width. Both too small N (Type I error) and too large N (Type II error) problems matter. Researchers should be very careful especially when examining either short or long panel.

4.6.2 Balanced Versus Unbalanced Panel Data

In a balanced panel, all entities have measurements in all time periods. In a contingency table (or cross-table) of cross-sectional and time-series variables, each cell should have only one frequency. Therefore, the total number of observations is nT. This tutorial document assumes

that we have a well-organized balanced panel data set. When each entity in a data set has different numbers of observations, the panel data are not balanced. Some cells in the contingency table have zero frequency. Accordingly, the total number of observations is not nT in an unbalanced panel. Unbalanced panel data entail some computation and estimation issues although most software packages are able to handle both balanced and unbalanced data.

4.6.3 Fixed Versus Rotating Panel Data

If the same individuals (or entities) are observed for each period, the panel data set is called a *fixed panel* (Greene 2008: 184). If a set of individuals changes from one period to the next, the data set is a *rotating panel*. This document assumes a fixed panel.

4.0 CONCLUSION

Panel data can be viewed as a finite set of time-series data

5.0 SUMMARY

In this unit you learnt panel data model which comprises variables that vary across time and cross section, and described the techniques used with this model including a pooled regression, a fixed effect and a random effect. In the next unit which is Unit 5 of our Module 2 we shall continue with our discussion on panel data by examining fixed versus random effects panel data.

6.0 TUTOR MARKED ASSIGNMENT

Discuss panel data models

7.0 REFERENCES/FURTHER READING

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

UNIT 5: FIXED VERSUS RANDOM EFFECTS PANEL DATA

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 Fixed Versus Random Effects
- 3.2 Fixed-Effects Model
  - 3.2.1 The least squares dummy variable model (LSDV)
  - 3.2.2 The "within" estimation
  - 3.2.4 The "Between"Estimation
  - 3.2.1 A Note on Fixed-Effects…
- 3.3 Random-Effects Model (Random Intercept, Partial Pooling Model)

4.0 CONCLUSION
5.0 SUMMARY
6.0TUTOR MARKED ASSIGNMENT

7.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the preceding unit you learnt panel data models. In the present unit which is Unit 5 and last of our Module 2 we shall continue with our discussion on panel data by examining fixed versus random effects panel data

2.0 OBJECTIVES

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

* Discuss fixed versus random effects model
* Analyze the least squares dummy variable model (LSDV)
* Explain the "within" estimation
* Discuss the "between"estimation

3.0 MAIN CONTENTS

3.1 Fixed Versus Random Effects

Panel data models examine group (individual-specific) effects, time effects, or both in order to deal with heterogeneity or individual effect that may or may not be observed. These effects are either fixed or random effect. A fixed effect model examines if intercepts vary across group or time period, whereas a random effect model explores differences in error variance components across individual or time period. A one-way model includes only one set of dummy variables (e.g., firm1, firm2, ...), while a two-way model considers two sets of dummy variables (e.g., city1, city2, … and year1, year2, …). Panel data models examine fixed and/or random effects of individual or time. The core difference between fixed and random effect models lies in the role of dummy variables.

A parameter estimate of a dummy variable is a part of the intercept in a fixed effect model and an error component in a random effect model. Slopes remain the same across group or time period in either fixed or random effect model. The functional forms of one-way fixed and random effect models are,

Fixed effect model: $y_{it} = (\alpha + u_i) + X_{it}^{'}\beta + v_{it}$

Random effect model: $y_{it} = \alpha + X_{it}^{'}\beta + (u_i + v_{it})$,

where u_i is a fixed or random effect specific to individual (group) or time period that is not included in the regression, and errors are independent identically distributed, $v_{it} \sim IID(0, \sigma_v^2)$.

A fixed group effect model examines individual differences in intercepts, assuming the same slopes and constant variance across individual (group and entity). Since an individual specific effect is time invariant and considered a part of the intercept, u is allowed to be correlated with other regressors; That is, OLS assumption 2 is not violated. This fixed effect model is effect estimated by least squares dummy variable (LSDV) regression (OLS with a set of dummies) and within effect estimation methods (Green, 2012).

Table 3.1 Fixed Effect and Random Effect Models

	Fixed Effect Model	Random Effect Model
Functional form	$y_{it} = (\alpha + u_i) + X_{it}^{'}\beta + v_{it}$	$y_{it} = \alpha + X_{it}^{'}\beta + (u_i + v_{it})$
Assumption	-	Individual effects are not correlated with regressors
Intercepts	Varying across group and/or time	Constant
Error variances	Constant	Randomly distributed across group and/or time
Slopes	Constant	Constant
Estimation	LSDV, within effect estimation	GLS, FGLS (EGLS)
Hypothesis test	F test	Breusch-Pagan LM test

3.2Fixed-Effects Model

(Covariance Model, Within Estimator, Individual Dummy Variable Model, Least Squares Dummy Variable Model)

Use fixed-effects (FE) whenever you are only interested in analyzing the impact of variables that vary over time. FE explore the relationship between predictor and outcome variables within an

entity (country, person, company, etc.). Each entity has its own individual characteristics that may or may not influence the predictor variables (for example, being a male or female could influence the opinion toward certain issue; or the political system of a particular country could have some effect on trade or GDP; or the business practices of a company may influence its stock price).

When using FE we assume that something within the individual may impact or bias the predictor or outcome variables and we need to control for this. This is the rationale behind the assumption of the correlation between entity's error term and predictor variables. FE removes the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable.

Another important assumption of the FE model is that those time-invariant characteristics are unique to the individual and should not be correlated with other individual characteristics. Each entity is different therefore the entity's error term and the constant (which captures individual characteristics) should not be correlated with the others. If the error terms are correlated, then FE is no suitable since inferences may not be correct and you need to model that relationship (probably using random-effects).

The equation for the fixed effects model becomes:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \qquad \text{[eq.1]}$$

Where

- α_i (i=1….n) is the unknown intercept for each entity (n entity-specific intercepts).
- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- X_{it} represents one independent variable (IV),
- β_1 is the coefficient for that IV,
- u_{it} is the error term

Fixed-effects will not work well with data for which within-cluster variation is minimal or for slow changing variables over time.

Another way to see the fixed effects model is by using binary variables. So the equation for the fixed effects model becomes:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + u_{it} \quad \text{[eq.2]}$$

Where

- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- $X_{k,it}$ represents independent variables (IV),
- β_k is the coefficient for the IVs,
- u_{it} is the error term
- E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.
- γ_2 Is the coefficient for the binary repressors (entities)

Both eq.1 and eq.2 are equivalents:

You could add time effects to the entity effects model to have a *time and entity fixed effects regression model*:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + \delta_2 T_2 + \ldots + \delta_t T_t + u_{it} \quad \text{[eq.3]}$$

Where

- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- $X_{k,it}$ represents independent variables (IV),
- β_k is the coefficient for the IVs,
- u_{it} is the error term
- E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.
- γ_2 is the coefficient for the binary regressors (entities)
- T_t is time as binary variable (dummy), so we have t-1 time periods.
- δ_t is the coefficient for the binary time regressors .

Control for time effects whenever unexpected variation or special events my affect the outcome variable. There are several strategies for estimating a fixed effect model**. The least squares dummy variable model (LSDV)** uses dummy variables, whereas the **"within" estimation** does not. These strategies, of course, produce the identical parameter estimates of regressors

(nondummy Independent variables). **The "between"estimation** fits a model using individual or time means of dependent and independent variables without dummies.

3.2.1 The least squares dummy variable model (LSDV)

LSDV with a dummy dropped out of a set of dummies is widely used because it is relatively easy to estimate and interpret substantively. This LSDV, however, becomes problematic when there are many individuals (or groups) in panel data. If T is fixed and $n \rightarrow 8$ (n is the number of groups or firms and T is the number of time periods), parameter estimates of regressors are consistent but the coefficients of individual effects, $a + u$, are not (Baltagi, 2001: 14). In this short panel, LSDV includes a large number of dummy variables; the number of these parameters to be estimated increases as n increases (*incidental parameter problem*); therefore, LSDV loses n degrees of freedom but returns less efficient estimators (p.14). Under this circumstance, LSDV is useless and thus calls for another strategy, the within effect estimation.

3.2.2 The "within" estimation

Unlike LSDV, the "within" estimation does not need dummy variables, but it uses deviations from group (or time period) means. That is, "within" estimation uses variation within each individual or entity instead of a large number of dummies. The "within" estimation is,

$$(y_{it} - \bar{y}_{i\bullet}) = (x_{it} - \bar{x}_{i\bullet})'\beta + (\varepsilon_{it} - \bar{\varepsilon}_{i\bullet}),$$

Where $\overline{y}_i$ is the mean of dependent variable (DV) of individual (group) i, $\overline{x}_I$ represent the means of independent variables (IVs) of group i, and ϵ_i is the mean of errors of group i.

In this "within" estimation, the incidental parameter problem is no longer an issue. The parameter estimates of regressors in the "within" estimation are identical to those of LSDV.

The "within" estimation reports correct the *sum of squared errors* (SSE). The "within" estimation, however, has several disadvantages.

First, data transformation for "within" estimation wipes out all time-invariant variables (e.g., gender, citizenship, and ethnic group) that do not vary within an entity since deviations of time-invariant variables from their average are all zero, it is not possible to estimate coefficients of such variables in "within" estimation. As a consequence, we have to fit LSDV when a model has time-invariant independent variables.

$$se_k^* = se_k \sqrt{\frac{df_{error}^{within}}{df_{error}^{LSDV}}} = se_k \sqrt{\frac{nT-k}{nT-n-k}}$$

of the "within" estimation is not correct because the intercept term is suppressed. Finally, the "within" estimation does not report dummy coefficients. We have to compute them, if really needed, using the formula $d_i^* = \bar{y}_I - \bar{x}_i\beta$

Table 3.2 Comparison of Three Estimation Methods

	LSDV	Within Estimation	Between Estimation
Functional form	$y_i = i\alpha_i + X_i\beta + \varepsilon_i$	$y_{it} - \bar{y}_{i\bullet} = x_{it} - \bar{x}_{i\bullet} + \varepsilon_{it} - \bar{\varepsilon}_{i\bullet}$	$\bar{y}_{i\bullet} = \alpha + \bar{x}_{i\bullet} + \varepsilon_i$
Time invariant variables	Yes	No	No
Dummy variables	Yes	No	No
Dummy coefficients	Presented	Need to be computed	N/A
Transformation	No	Deviation from the group means	Group means
Intercept estimated	Yes	No	Yes
R^2	Correct	Incorrect	
SSE	Correct	Correct	
MSE/SEE (SRMSE)	Correct	Incorrect (smaller)	
Standard errors	Correct	Incorrect (smaller)	
DF_{error}	nT-n-k^*	nT-k (n larger)	n-k-1
Observations	nT	nT	n

* It means that the LSDV estimation loses n degrees of freedom because of dummy variables included.

3.2.4 The "Between"Estimation

The "between group" estimation, so called the group mean regression, uses variation between

individual entities (groups). Specifically, this estimation calculates group means of the dependent and independent variables and thus reduces the number of observations down to *n*. Then, run OLS on these transformed aggregated data: $\overline{y}_I = \alpha + \overline{x}_I + \epsilon_i$ Table 3.2 contrasts LSDV, "within group" estimation, and "between group" estimation.

3.2.5 A Note on Fixed-Effects…

"…The fixed-effects model controls for all time-invariant differences between the individuals, so the estimated coefficients of the fixed-effects models cannot be biased because of omitted time-invariant characteristics…[like culture, religion, gender, race, etc]

One side effect of the features of fixed-effects models is that they cannot be used to investigate time-invariant causes of the dependent variables. Technically, time-invariant characteristics of the individuals are perfectly collinear with the person [or entity] dummies. Substantively, fixed-effects models are designed to study the causes of changes within a person [or entity]. A time-invariant characteristic cannot cause such a change, because it is constant for each person.

3.3 Random-Effects Model (Random Intercept, Partial Pooling Model)

The rationale behind random effects model is that, unlike the fixed effects model,the variation across entities is assumed to be random and uncorrelated with the predictor or independent variables included in the model:

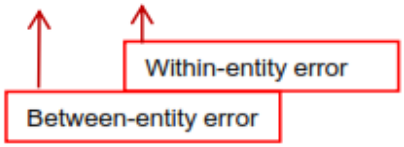
"…the crucial distinction between fixed and random effects is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not"

If you have reason to believe that differences across entities have some influence on your dependent variable then you should use random effects. An advantage of random effects is that

you can include time invariant variables (i.e. gender). In the fixed effects model these variables are absorbed by the intercept.

The random effects model is:

$$Y_{it} = \beta X_{it} + \alpha + u_{it} + \varepsilon_{it} \qquad [eq.4]$$

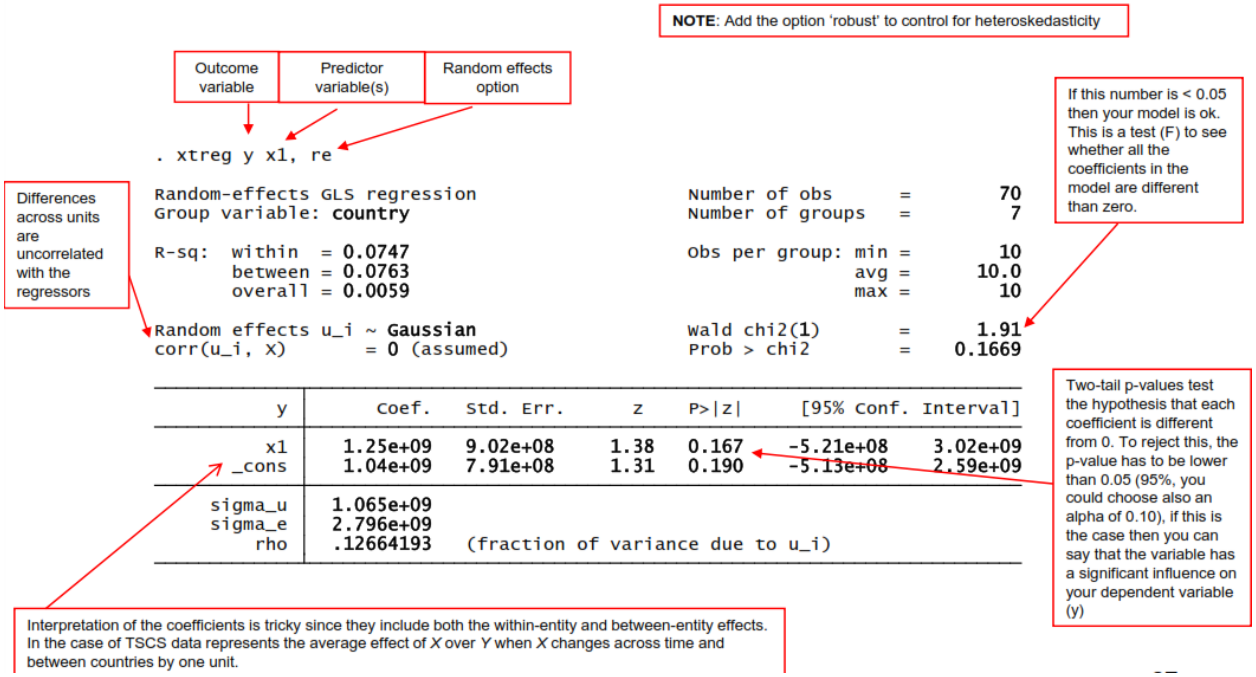


Random effects assume that the entity's error term is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables.

In random-effects you need to specify those individual characteristics that may or may not influence the predictor variables. The problem with this is that some variables may not be available therefore leading to omitted variable bias in the model.

RE allows to generalize the inferences beyond the sample used in the model.

You can estimate a random effects model using `xtreg` and the option `re`.



4.0 CONCLUSION

You also learned how to analyze the least squares dummy variable model (LSDV), the "within" estimation and the "between" estimation model. The next module that is Module 3 and last ushers us into the testing of panel data models.

5.0 SUMMARY

In this unit 5, which is the last unit in our module 1, you have learnt two techniques use to analyze panel data. They are fixed effects random effects fixed versus random effects. You also learned how to analyze the least squares dummy variable model (LSDV), the "within" estimation and the "between" estimation model. The next module that is Module 3 and last ushers us into the testing of panel data models.

6.0 TUTOR MARKED ASSIGNMENT

Distinguish using examples between fixed effects and random effects in econometric analysis

7.0 REFERENCES/FURTHER READING

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

MODULE 3

Unit 1: Testing Fixed and Random Effects

Unit 2: Panel Data Estimation in Eviews.

Unit 3: Dynamic Models
Unit 4: Autoregressive Distributed Lag (ARDL) Model
Unit 5: ARDL level Relation

UNIT 1: TESTING FIXED AND RANDOM EFFECTS

CONTENTS

1. INTRODUCTION
2. OBJECTIVES
3.0 MAIN CONTENT
3.1 Testing Fixed and Random Effects
3.1.2 Breusch-Pagan LM Test for Random Effects
3.1.3 Hausman Test for Comparing Fixed and Random Effects
3.2 Model Selection: Fixed or Random Effect?
3.3 Substantive Meanings of Fixed and Random Effects
3.4 Two Recommendations for Panel Data Modeling
3.5 Guidelines of Model Selection
4.0 CONCLUSION
5.0 SUMMARY
6.0 TUTOR MARKED ASSIGNMENT
7.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

This is the first unit of our Module 3 and last module of this study. It focus on testing fixed and random effects of panel data.

2.0 OBJECTIVES

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

* Test fixed and random effects
* Discuss Breusch-Pagan LM Test for Random Effects
* Discuss Hausman Test for Comparing Fixed and Random Effects
* Understand guidelines of Model Selection

3.0 MAIN CONTENTS

3.1 Testing Fixed and Random Effects

How do we know if fixed and/or random effects exist in panel data in hand? A fixed effect is tested by F-test, while a random effect is examined by Breusch and Pagan's (1980) Lagrange

multiplier (LM) test. The former compares a fixed effect model and OLS to see how much the fixed effect model can improve the goodness-of-fit, whereas the latter contrast a random effect model with OLS. The similarity between random and fixed effect estimators is tested by a Hausman test.

3.1.1 F-Test for Fixed Effects

In a regression of $Y_{it} = \alpha + \mu_i + X_{it}$ ‘$\beta + \epsilon_{it}$ the null hypothesis is that all dummy parameters except for one for the dropped are all zero, $H_0 : \mu_1 = \ldots = \mu_{n-1} = 0$ The alternative hypothesis is that at least one dummy parameter is not zero. This hypothesis is tested by an F test, which is based on loss of goodness-of-fit. This test contrasts LSDV (robust model) with the pooled OLS (efficient model) and examines the extent that the goodness-of-fit measures (SSE or R^2) changed.

$$F(n-1, nT-n-k) = \frac{(e'e_{pooled} - e'e_{LSDV})/(n-1)}{(e'e_{LSDV})/(nT-n-k)} = \frac{(R^2_{LSDV} - R^2_{pooled})/(n-1)}{(1-R^2_{LSDV})/(nT-n-k)}$$

If the null hypothesis is rejected (at least one group/time specific intercept *u* is not zero), you may conclude that there is a significant fixed effect or significant increase in goodness-of-fit in the fixed effect model; therefore, the fixed effect model is better than the pooled OLS.

SELF-ASSESEMENT EXERCISE

Examine F-test for fixed effects of a panel data

3.1.2 Breusch-Pagan LM Test for Random Effects

Breusch and Pagan’s (1980) Lagrange multiplier (LM) test examines if individual (or time) specific variance components are zero, $H_0 : \sigma^2_U = 0.$ The LM statistic follows the chi-squared

$$LM_u = \frac{nT}{2(T-1)}\left[\frac{T^2\bar{e}'\bar{e}}{e'e} - 1\right]^2 \sim \chi^2(1),$$

Where $\overline{e}$ is the $n \times 1$ vector of the group means of pooled regression residuals, and $e'e$ is the SSE of the pooled OLS regression.

Baltagi (2001) presents the same LM test in a different way.

$$LM_u = \frac{nT}{2(T-1)}\left[\frac{\sum\left(\sum e_{it}\right)^2}{\sum\sum e_{it}^2} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{\sum\left(T\overline{e}_{i\bullet}\right)^2}{\sum\sum e_{it}^2} - 1\right]^2 \sim \chi^2(1).$$

If the null hypothesis is rejected, you can conclude that there is a significant random effect in the panel data, and that the random effect model is able to deal with heterogeneity better than the pooled OLS.

3.1.3 Hausman Test for Comparing Fixed and Random Effects

How do we know which effect (fixed effect or random effect) is more relevant and significant in the panel data? The Hausman specification test compares fixed and random effect models under the null hypothesis that individual effects are uncorrelated with any regressor in the model (Hausman, 1978). If the null hypothesis of no correlation is not violated, LSDV and GLS are consistent, but LSDV is inefficient; otherwise, LSDV is consistent but GLS is inconsistent and biased (Greene, 2008: 208). The estimates of LSDV and GLS should not differ systematically under the null hypothesis. The Hausman test uses that "the covariance of an efficient estimator with its difference from an inefficient estimator is zero".

$$LM = \left(b_{LSDV} - b_{random}\right)'\hat{W}^{-1}\left(b_{LSDV} - b_{random}\right) \sim \chi^2(k),$$

where $\hat{W} = Var[b_{LSDV} - b_{random}] = Var(b_{LSDV}) - Var(b_{random})$ is the difference in the estimated covariance matrices of LSDV (robust model) and GLS (efficient model). Keep in mind that an intercept and dummy variables SHOULD be excluded in computation. This test statistic follows the chi-squared distribution with k degrees of freedom.

The formula says that a Hausman test examines if "the random effects estimate is insignificantly different from the unbiased fixed effect estimate" (Kennedy, 2008: 286). If the null hypothesis of no correlation is rejected, you may conclude that individual effects *u* are Significantly correlated with at least one regressors in the model and thus the random effect model is problematic. Therefore, you need to go for a fixed effect model rather than the random effect counterpart. A drawback of this Hausman test is, however, that the difference of covariance matrices *W* may not be positive definite; Then, we may conclude that the null is not rejected assuming similarity of the covariance matrices renders such a problem (Greene, 2008: 209).

3.2 Model Selection: Fixed or Random Effect?

When combining fixed vs. random effects, group vs. time effects, and one-way vs. two-way effects, we get 12 possible panel data models as shown in Table 3.3. In general, one-way models are often used mainly due to their parsimony, and a fixed effect model is easier than a random counterpart to estimate the model and interpret its result. It is not, however, easy to sort out the best one out of the following 12 models Nymoen (1991).

Table 3.3 Classification of Panel Data Analysis

	Type	Fixed Effect	Random Effect
One-way	Group	One-way fixed group effect	One-way random group effect
	Time	One-way fixed time effect	One-way random time effect
Two-way	Two groups*	Two-way fixed group effect	Two-way random group effect
	Two times*	Two-way fixed time effect	Two-way random time effect
	Mixed	Two-way fixed group & time effect	Two-way random group & time effect
		Two-way fixed time and random group effect	
		Two-way fixed group and random time effect	

* These models need two group (or time) variables (e.g., country and airline).

3.3 Substantive Meanings of Fixed and Random Effects

Specifically, the F-test compares a fixed effect model and (pooled) OLS, whereas the LM test contrasts a random effect model with OLS. The Hausman specification test compares fixed and

random effect models. However, these tests do not provide substantive meanings of fixed and random effects. What does a fixed effect mean? How do we interpret a random effect substantively?

Here is a simple and rough answer. Suppose we are regressing the production of firms such as Apple, IBM, LG, and Sony on their R&D investment. A fixed effect might be interpreted as initial production capacities of these companies when no R&D investment is made; each firm has its own initial production capacity. A random effect might be viewed as a kind of consistency or stability of production. If the production of a company fluctuates up and down significantly, for example, its production is not stable (or its variance component is larger than those of other firms) even when its productivity (slope of R&D) remains the same across company.

Kennedy (2008: 282-286) provides theoretical and insightful explanation of fixed and random effects. Either fixed or random effect is an issue of unmeasured variables or omitted relevance variables, which renders the pooled OLS biased. This heterogeneity is handled by either putting in dummy variables to estimate individual intercepts of groups (entities) or viewing "the different intercepts as having been drawn from a bowl of possible intercepts, so they may be interpreted as random … and treated as though they were a part of the error term" (p. 284); they are fixed effect model and random effect model, respectively. A random effect model has a "composite error term" that consists of the traditional random error and a "random intercept" measuring the extent to which individual's intercept differs from the overall intercept (p. 284). He argues that the key difference between fixed and random effects is not whether unobserved heterogeneity is attributed to the intercept or variance components, but whether the individual specific error component is related to regressors.

3.4 Two Recommendations for Panel Data Modeling

The first recommendation, as in other data analysis processes, is to describe the data of interest carefully before analysis. Although often ignored in many data analyses, this data description is very important and useful for researchers to get ideas about data and analysis strategies. In panel data analysis, properties and quality of panel data influence model section significantly.

i. Clean the data by examining if they were measured in reliable and consistent manners. If different time periods were used in a long panel, for example, try to rearrange (aggregate) data to improve consistency. If there are many missing values, decide whether you go for a balanced panel by throwing away some pieces of usable information or keep all usable observations in an unbalanced panel at the expense of methodological and computational complication.

ii. Examine the properties of the panel data including the number of entities (individuals), the number of time periods, balanced versus unbalanced panel, and fixed versus rotating panel. Then, try to find models appropriate for those properties.

iii. Be careful if you have "long" or "short" panel data. Imagine a long panel that has 10

thousand time periods but 3 individuals or a short panel of 2 (years) × 9,000 (firms).

If *n* and/or *T* are too large, try to reclassify individuals and/or time periods and get some manageable *n'* and *T'*. The null hypothesis of $u1 = u2 = \ldots = U_{999.999} = 0$ in a

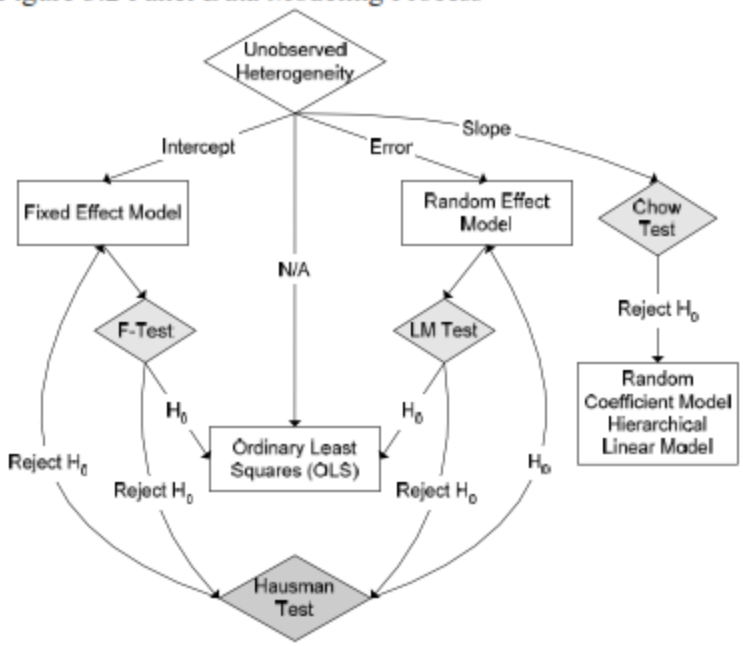
fixed effect model, for instance, is almost useless. This is just as you are seriously arguing that at least one citizen looks different from other 999,999 people! Didn't you know that before? Try to use yearly data rather than weekly data or monthly data rather than daily data.

Second recommendation is to begin with a simpler model. Try a pooled OLS rather than a fixed or random effect model; a one-way effect model rather than a two-way model; a fixed or random effect model rather than a hierarchical linear model; and so on. Do not try a fancy, of course, complicated, model that your panel data do not support enough (e.g., poorly organized panel and long/short panel).

3.5 Guidelines of Model Selection

On the modeling stage, let us begin with pooled OLS and then think critically about its potential problems if observed and unobserved heterogeneity (a set of missing relevant variables) is not taken into account. Also think about the source of heterogeneity (i.e., cross sectional or time series variables) to determine individual (entity Or group) effect or time effect. Figure 3.2 provides a big picture of the panel data modeling process Nymoen (1991).

Figure 3.2 Panel Data Modeling Process



If you think that the individual heterogeneity is captured in the disturbance term and the individual (group or time) effect is not correlated with any regressors, try a random effect model. If the heterogeneity can be dealt with individual specific intercepts and the individual effect may possibly be correlated with any regressors, try a fixed effect model. If each individual (group) has its own initial capacity and shares the same disturbance variance with other individuals, a fixed effect model is favored. If each individual has its own disturbance, a random effect will be better at figuring out heteroskedestic disturbances.

Next, conduct appropriate formal tests to examine individual group and/or time effects. If the null hypothesis of the LM test is rejected, a random effect model is better than the pooled OLS. If the null hypothesis of the F-test is rejected, a fixed effect model is favored over OLS. If both hypotheses are not rejected, fit the pooled OLS.

Conduct the Hausman test when both hypotheses of the F-test and LM test are all rejected. If the null hypothesis of uncorrelation between an individual effect and regressors is rejected, go for the robust fixed effect model; otherwise, stick to the efficient random effect model.

If you have a strong belief that the heterogeneity involves two cross-sectional, two time series, or one cross-section and one time series variables, try two-way effect models. Double-check if your panel data are well-organized, and *n* and *T* are large enough; do not try a two-way model for a poorly organized, badly unbalanced, and/or too long/short panel. Conduct appropriate F-test and LM test to examine the presence of two-way effects. Stata does not provide direct ways to fit two-way panel data models but it is not impossible. In Stata, two-way fixed effect models seem easier than two-way random effect models.

Finally, if you think that the heterogeneity entails slops (parameter estimates of regressors) varying across individual and/or time. Conduct a Chow test or equivalent to examine the poolability of the panel data. If the null hypothesis of poolable data is rejected, try a random coefficient model or hierarchical linear model.

4.0 CONCLUSION

Fixed-effects models and alternatives are part of panel data model. The limitation or hidden assumption of each fixed-effects model, such as the individual fixed-effects model, time fixed-effects model, and the individual-time fixed-effects model, are discussed in detail.

5.0 SUMMARY

In this first unit of our Module 3, you learned how to conduct tests for fixed and random effects, Breusch-Pagan LM Test for Random Effect, Test for Comparing Fixed and Random Effect and guidelines of Model Selection. The next unit is about panel data estimation in Eviews.

6.0 TUTOR MARKED ASSIGNMENT

What is fixed effect?

7.0 REFERENCES/FURTHER READING

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

UNIT2: PANEL DATA ESTIMATION IN EVIEWS

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
    3.1 Steps in Estimating a Panel Equation in Eviews
    3.2 Least Squares Estimation
        3.2.1 Least Squares Panel Options
    3.3 Instrumental Variables Estimation
    3.4 Generalized Method of Moments (GMM) Estimation
        3.4.1 GMM Panel Options
        3.4.2 GMM Instruments
        3.4.3 GMM Options
4.0 CONCLUSION
5.0SUMMARY

6.0 TUTOR MARKED ASSIGNMENT
7.0 REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the previous unit of you learnt how to conduct tests for fixed and random effects, Breusch-Pagan LM Test for Random Effect, Test for Comparing Fixed and Random Effect and guidelines of Model Selection. In the present unit we shall discuss panel data estimation in Eviews.

2.0 OBJECTIVES

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

* List the steps in estimating a Panel Equation in Eviews
* Discuss Least Squares Estimation
* Discuss Instrumental Variables Estimation
* Analyze Generalized Method of Moments (GMM) Estimation

3.0 MAIN CONTENT

3.1 Steps in Estimating a Panel Equation in Eviews 10.0

The first step in estimating a panel equation is to call up an equation dialog by clicking on **Object/New Object.../Equation or Quick/Estimate Equation…** from the main menu of **Eviews**, or typing the keyword equation in the command window. You should make certain that your **workfile** is structured as a panel **workfile**. EViews will detect the presence of your panel structure and in place of the standard equation dialog will open the panel Equation Estimation dialog.

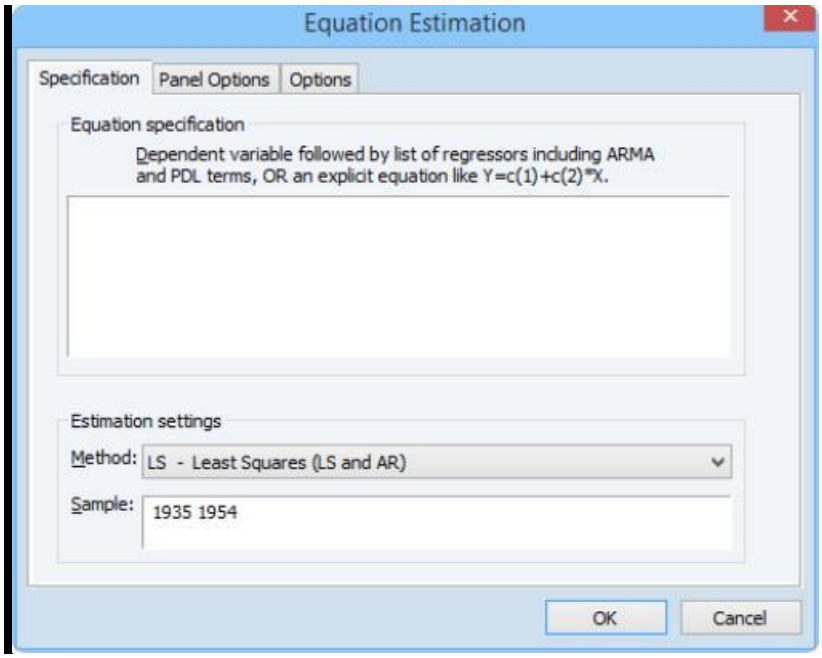
You should use the Method dropdown menu to choose between LS - Least Squares (LS and AR), TSLS - Two-Stage Least Squares (TSLS and AR), and GMM / DPD - Generalized Method of Moments / Dynamic Panel Data techniques. If you select the either of the latter two methods, the dialog will be updated to provide you with an additional page for specifying instruments (see "Instrumental Variables Estimation").

The remaining estimation supported estimation techniques do not account for the panel structure of your **workfile**, save for lags not crossing the boundaries between cross-section units.

3.2 Least Squares Estimation

The basic least squares estimation dialog is a multi-page dialog with pages for the basic specification, panel estimation options, and general estimation options.

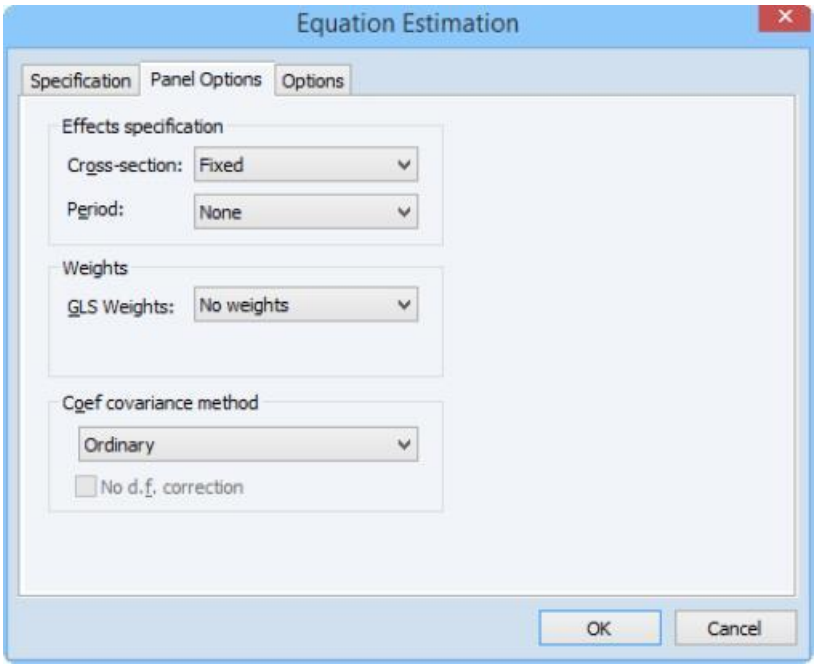
Least Squares Specification



You should provide an equation specification in the upper Equation specification edit box, and an estimation sample in the Sample edit box. The equation may be specified by list or by expression as described in "Specifying an Equation in EViews".

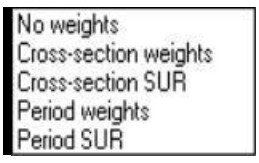
In general, most of the specifications allowed in non-panel equation settings may also be specified here. You may, for example, include AR terms in both linear and nonlinear specifications, and may include PDL terms in equations specified by list. You may not, however, include MA terms in a panel setting.

3.2.1 Least Squares Panel Options



Next, click on the Panel Options tab to specify additional panel specific estimation settings. First, you should account for individual and period effects using the Effects specification dropdown menus. By default, EViews assumes that there are no effects so that both dropdown menus are set to None. You may change the default settings to allow for either Fixed or Random effects in either the cross-section or period dimension, or both. See the pool discussion of "Fixed and Random Effects" for details.

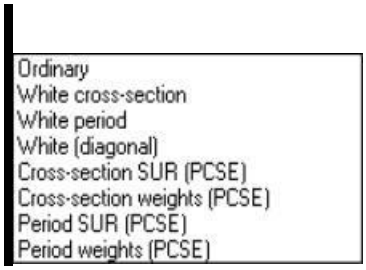
You should be aware that when you select a fixed or random effects specification, EViews will automatically add a constant to the common coefficients portion of the specification if necessary, to ensure that the effects sum to zero.



Next, you should specify settings for GLS Weights. You may choose to estimate with no weighting, or with Cross-section weights, Cross-section SUR, Period weights, Period SUR. The Cross-section SUR setting allows for contemporaneous correlation between

cross-sections (clustering by period), while the Period SUR allows for general correlation of residuals across periods for a specific cross-section (clustering by individual). Cross-section weights and Period weights allow for heteroskedasticity in the relevant dimension.

For example, if you select Cross section weights, EViews will estimate a feasible GLS specification assuming the presence of cross-section heteroskedasticity. If you select Cross-section SUR, EViews estimates a feasible GLS specification correcting for heteroskedasticity and contemporaneous correlation. Similarly, Period weights allows for period heteroskedasticity, while Period SUR corrects for heteroskedasticity and general correlation of observations within a cross-section. Note that the SUR specifications are both examples of what is sometimes referred to as the Parks estimator. See the pool discussion of "Generalized Least Squares" for additional details.

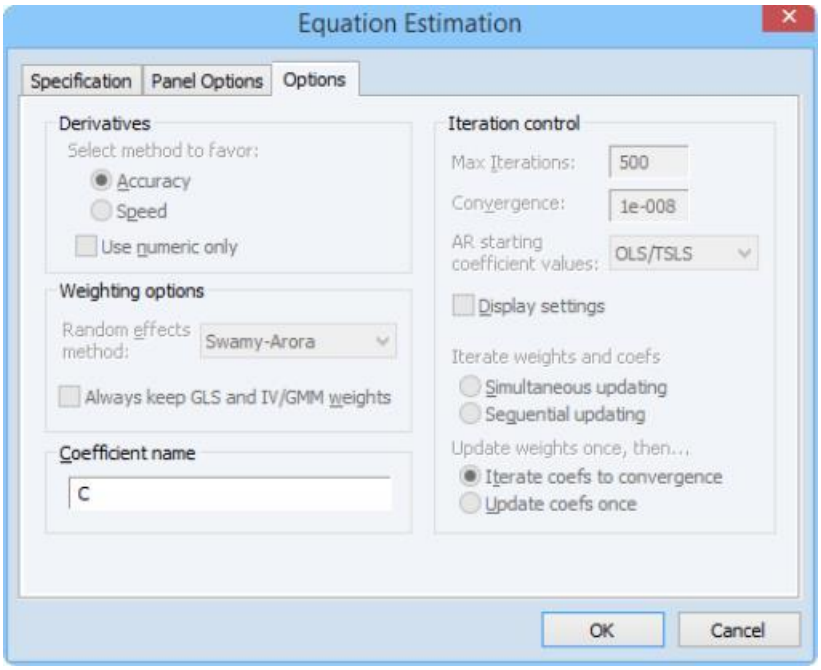


Lastly, you should specify a method for computing coefficient covariances. You may use the dropdown menu labeled Coef covariance method to select from the various robust methods available for computing the coefficient standard errors. The covariance calculations may be chosen to be robust under various assumptions, for example, general correlation of observations within a cross-section, or perhaps cross-section heteroskedasticity. Click on the checkbox No d.f. correction to perform the calculations without the leading degree of freedom correction term. Each of the coefficient covariance methods is described in greater detail in "Robust Coefficient Covariances" of the pool chapter.

You should note that some combinations of specifications and estimation settings are not currently supported. You may not, for example, estimate random effects models with cross-section specific coefficients, AR terms, or weighting. Furthermore, while two-way

random effects specifications are supported for balanced data, they may not be estimated in unbalanced designs.

LS Options



Lastly, clicking on the Options tab in the dialog brings up a page displaying computational options for panel estimation. Settings that are not currently applicable will be grayed out. These options control settings for derivative taking, random effects component variance calculation, coefficient usage, iteration control, and the saving of estimation weights with the equation object.

These options are identical to those found in pool equation estimation, and are described in considerable detail in “Options”.

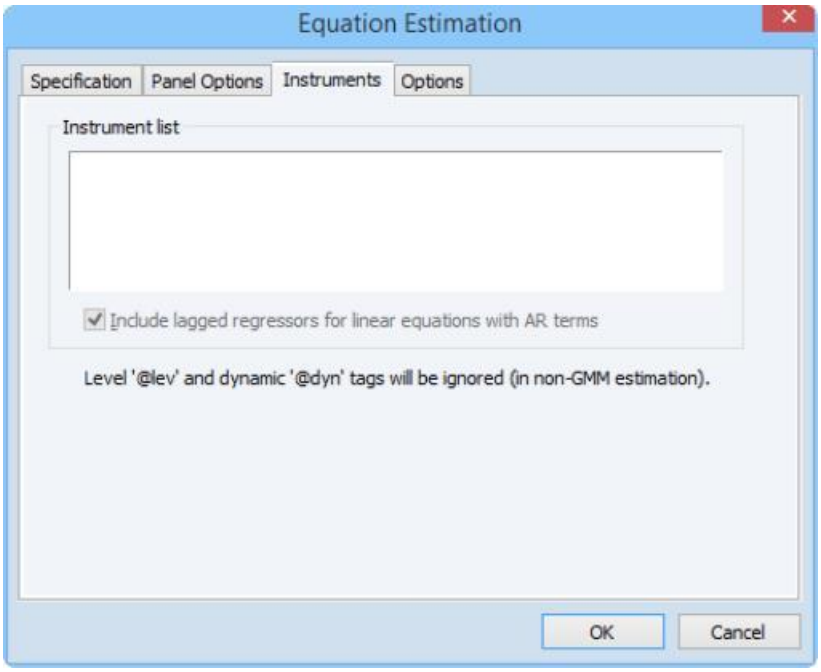
SEFE-ASSESEMENT EXERCISE

What are least squares panel options?

3.3 Instrumental Variables Estimation

To estimate a pool specification using instrumental variables techniques, you should select TSLS - Two-Stage Least Squares (and AR) in the Method dropdown menu at the

bottom of the main (Specification) dialog page. EViews will respond by creating a four page dialog in which the third page is used to specify your instruments.



While the three original pages are unaffected by this choice of estimation method, note the presence of the new third dialog page labeled Instruments, which you will use to specify your instruments. Click on the Instruments tab to display the new page.

IV Instrument Specification There are only two parts to the instrumental variables page. First, in the edit box labeled Instrument list, you will list the names of the series or groups of series you wish to use as instruments.

Next, if your specification contains AR terms, you should use the checkbox to indicate whether EViews should automatically create instruments to be used in estimation from lags of the dependent and regressor variables in the original specification. When estimating an equation specified by list that contains AR terms, EViews transforms the linear model and estimates the nonlinear differenced specification. By default, EViews will add lagged values of the dependent and independent regressors to the corresponding

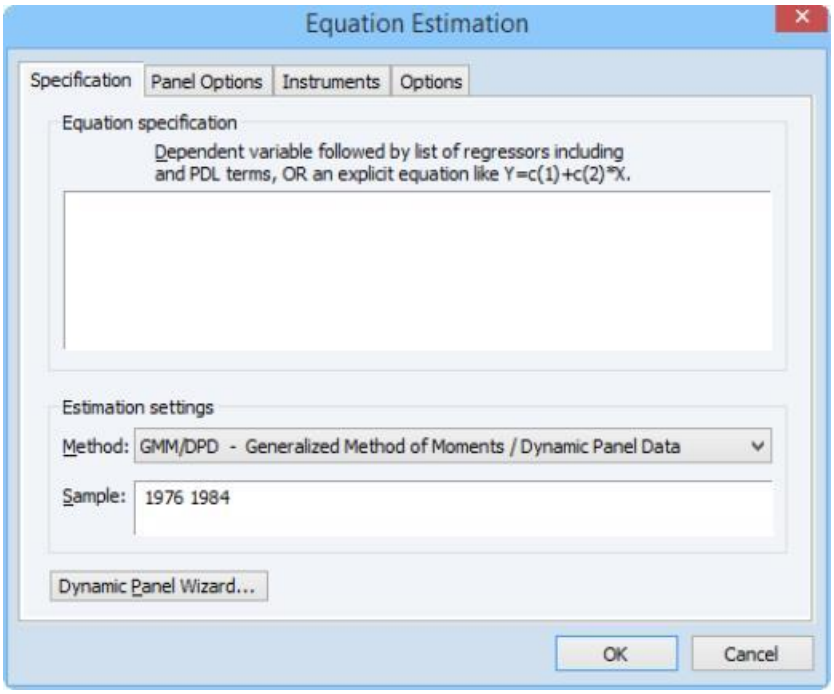
lists of instrumental variables to account for the modified specification, but if you wish, you may uncheck this option.

See the pool chapter discussion of “Instrumental Variables” for additional detail.

3.4 Generalized Method of Moments (GMM) Estimation

To estimate a panel specification using GMM techniques, you should select GMM / DPD - Generalized Method of Moments / Dynamic Panel Data in the Method dropdown menu at the bottom of the main (Specification) dialog page. Again, you should make certain that your workfile has a panel structure. EViews will respond by displaying a four page dialog that differs significantly from the previous dialogs.

GMM Specification

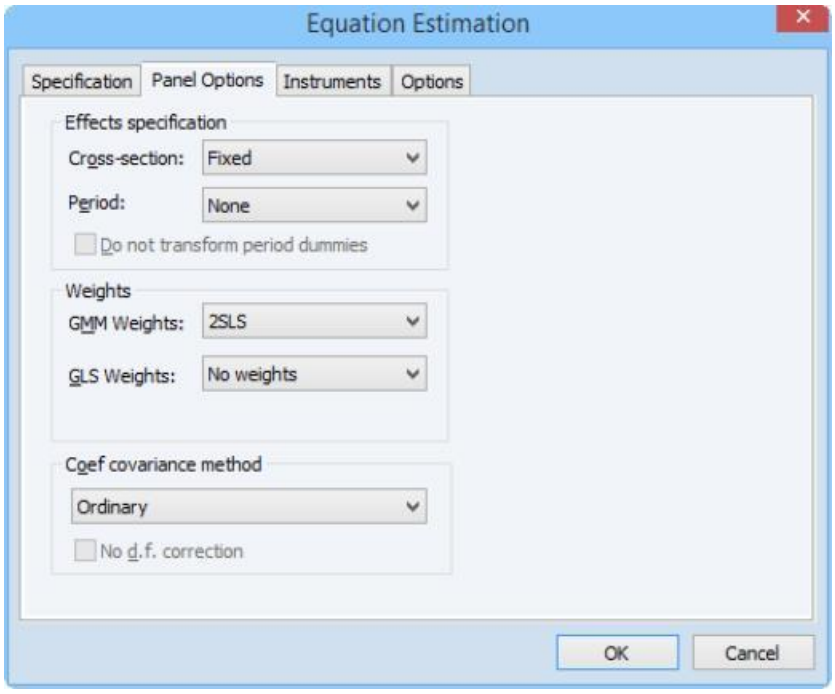


The specification page is similar to the earlier dialogs. As in the earlier dialogs, you will enter your equation specification in the upper edit box and your sample in the lower edit box.

Note, however, the presence of the Dynamic Panel Wizard... button on the bottom of the dialog. Pressing this button opens a wizard that will aid you in filling out the dialog so that you may employ dynamic panel data techniques such as the Arellano-Bond 1-step estimator for models with lagged endogenous variables and cross-section fixed effects. We will return to this wizard shortly ("GMM Example").

3.4.1 GMM Panel Options

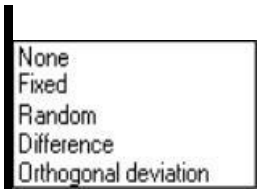
Next, click on the Panel Options dialog to specify additional settings for your estimation procedure.



As before, the dialog allows you to indicate the presence of cross-section or period fixed and random effects, to specify GLS weighting, and coefficient covariance calculation methods.

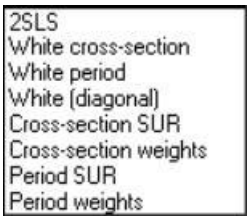
There are, however, notable changes in the available settings.

First, when estimating with GMM, there are two additional choices for handling cross-section fixed effects. These choices allow you to indicate a transformation method for eliminating the effect from the specification.

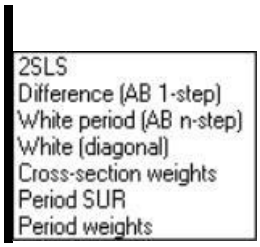


You may select Difference to indicate that the estimation procedure should use first differenced data (as in Arellano and Bond, 1991), and you may use Orthogonal Deviations (Arellano and Bover, 1995) to perform an alternative method of removing the individual effects.

Second, the dialog presents you with a new dropdown menu so that you may specify weighting matrices that may provide for additional efficiency of GMM estimation under appropriate assumptions. Here, the available options depend on other settings in the dialog.



In most cases, you may select a method that computes weights under one of the assumptions associated with the robust covariance calculation methods (see “Least Squares Panel Options”). If you select White cross-section, for example, EViews uses GMM weights that are formed assuming that there is contemporaneous correlation between cross-sections.



If, however, you account for cross-section fixed effects by performing first difference estimation, EViews provides you with a modified set of GMM weights choices. In particular, the Difference (AB 1-step) weights are those associated with the difference transformation. Selecting these weights allows you to estimate the GMM specification typically referred to as Arellano-Bond 1-step estimation. Similarly, you may choose the White period (AB 1-step) weights if you wish to compute Arellano-Bond 2-step or multi-step estimation. Note that the White period weights have been relabeled to indicate that they are typically associated with a specific estimation technique.

Note also that if you estimate your model using difference or orthogonal deviation methods, some GMM weighting methods will no longer be available.

3.4.2 Generalized Moment Method (GMM) Instruments

Instrument specification in GMM estimation follows the discussion above with a few additional complications.

First, you may enter your instrumental variables as usual by providing the names of series or groups in the edit field. In addition, you may tag instruments as period-specific predetermined instruments, using the @dyn keyword, to indicate that the number of implied instruments expands dynamically over time as additional predetermined variables become available.

To specify a set of dynamic instruments associated with the series X, simply enter "@DYN(X)" as an instrument in the list. EViews will, by default, use the series X(-2), X(-3), ..., X(-T), as instruments for each period (where available). Note that the default set of instruments grows very quickly as the number of periods increases. With 20 periods, for example, there are 171 implicit instruments associated with a single dynamic instrument. To limit the number of implied instruments, you may use only a subset of the

instruments by specifying additional arguments to @dyn describing a range of lags to be used.

For example, you may limit the maximum number of lags to be used by specifying both a minimum and maximum number of lags as additional arguments. The instrument specification:

@dyn(x, -2, -5)

instructs EViews to inclu

de lags of X from 2 to 5 as instruments for each period.

If a single argument is provided, EViews will use it as the minimum number of lags to be considered, and will include all higher ordered lags. For example:

@dyn(x, -5)

includes available lags of X from 5 to the number of periods in the sample.

Second, in specifications estimated using transformations to remove the cross-section fixed effects (first differences or orthogonal deviations), use may use the @lev keyword to instruct EViews to use the instrument in untransformed, or level form. Tagging an instrument with "@LEV" indicates that the instrument is for the transformed equation If @lev is not provided, EViews will transform the instrument to match the equation transformation.

If, for example, you estimate an equation that uses orthogonal deviations to remove a cross-section fixed effect, EViews will, by default, compute orthogonal deviations of the instruments provided prior to their use. Thus, the instrument list:

z1 z2 @lev(z3)

will use the transformed Z1 and Z2, and the original Z3 as the instruments for the specification.

Note that in specifications where @dyn and @lev keywords are not relevant, they will be ignored. If, for example, you first estimate a GMM specification using first differences with both dynamic and level instruments, and then re-estimate the equation using LS, EViews will ignore the keywords, and use the instruments in their original forms.

3.4.3 GMM Options

Lastly, clicking on the Options tab in the dialog brings up a page displaying computational options for GMM estimation. These options are virtually identical to those for both LS and IV estimation (see "LS Options"). The one difference is in the option for saving estimation weights with the object. In the GMM context, this option applies to both the saving of GLS as well as GMM weights.

4.0 CONCLUSION

In estimating panel data model in Eviews, certain steps must be followed. You should make certain that your workfile is structured as a panel workfile. EViews will detect the presence of your panel structure and in place of the standard equation dialog will open the panel Equation Estimation dialog.

5.0 SUMMARY

In this unit, you learned the steps in estimating a Panel Equation in Eviews. In which Least Squares Estimation, Instrumental Variables Estimation and Generalized Method of Moments (GMM) Estimation were discussed. In the next unit, which is Unit 3, we shall discuss dynamic models.

5.0 TUTOR MARKED ASSIGNMENT

7.0 REFERENCES/FURTHER READING

Adewara, S. O. & Kilishi, A. A. (2015). *Analysis of survey data using stata*. A workshop lecture presented on 27th – 30th April, 2015 in University of Illorin, Nigeria.

Cameron, A. C. & Trivedi, P. K. (2009). *Microeconometrics using stata*. Texas, USA: Stata Press.

Gujarati, D. N. & Porter, D. C. (2009). *Basic econometrics* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). OH, USA: Cengage.

UNIT3: DYNAMIC MODELS

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 Static and Dynamic Models
- 3.2 Dynamic Multipliers
  - 3.2.1 Dynamic Effects of Increased Income on Consumption
- 3.3 General Notation
- 3.4 Multipliers in the Text Books to This Course

4.0 CONCLUSION
5.0 SUMMARY
7.0 TUTOR MARKED ASSIGNMENT

7.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

In many areas of economics, time plays an important role: firms and households do not react instantly to changes in for example taxes, wages and business prospects but take time to adjust their decisions and habits. Moreover, because of information and processing lags, time goes by before changes in circumstances are even recognized.

There are also institutional arrangements, social and legal agreements and norms that hinder continuous adjustments of economic variables. Annual (or even biannual) wage bargaining rounds is one important example. The manufacturing of goods is usually not instantaneous but takes time, often several years in the case of projects with huge capital investments. Dynamic behaviour is also induced by the fact that many economic decisions are heavily influenced by what firms, households and the government anticipate. Often expectation formation will attribute a large weight to past developments, since anticipations usually have to build on past experience. Because dynamics is a fundamental feature of the macroeconomy, all serious policy analysis is based on a dynamic approach.

2.0 OBJECTIVES

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

* Discuss static and dynamic models
* Calculate dynamic multipliers
* Explain dynamic effects of increased income on consumption
* Write general notation of dynamic models
* Compare multipliers in the text books with this course

3.0 MAIN CONTENTS

3.1 Static and Dynamic Models

A variable yt is called a time series if we observe it over a sequence of time periods represented by the subscript t., i.e., {yT , yT -1, ..., y1} if we have observations from period 1 to T. Usually, we use the simpler notation yt, t = 1, ..., T, and if the observation period is of no substantive interest, that too is omitted. The interpretation of the time subscript varies from case to case, it can represent a year, a quarter or a month. In macroeconomics other periods are also considered, such as 5-year or 10 year averages of historical data, and daily or even hourly data at the other extreme (e.g., exchange rates, stock prices, money market interest rates). An example where yt is (the

logarithm) of private consumption, and we consider both static and dynamic models of consumption (consumption functions).

The following statement is typical of many central banks' view:

> "A substantial share of the effects on inflation of an interest rate change will occur within two years. Two years is therefore a reasonable time horizon for achieving the inflation target of 2.5 percent"

The quotation above is interesting because it is a clear statement about the time lag between a policy change and the effect on the target variable. Formally, response lags correspond to the concept of the dynamic multiplier which is introduced in section 3.2. The dynamic multiplier is a key concept in this course, and once you get a good grip on it, you also have a powerful tool which allows you to calculate the dynamic effects of policy changes (and of other exogenous shocks for that matter) on important variables like consumption, unemployment, inflation or other variables of your interest. When we consider economic models to be used in an analysis of real world macro data, care must be taken to distinguish between static and dynamic models. The well-known textbook consumption function, i.e., the relationship between private consumption expenditure (C) and households' disposable income (Y) is an example of a static equation

$$C_t = f(INC_t),\ f^1 > 0. \quad \text{(1)}$$

Consumption in any period t is strictly increasing in income, hence the positive signed first order derivative f^1–the marginal propensity to consume. To be able to apply the theory to observations of the real economy we have to specify the function f(INCt). Two of the most popular functional forms are the so called linear and log-linear specifications:

$$C_t = \beta_0 + \beta_1 INC_t + e_t,\ (linear) \quad \text{(2)}$$

$$\ln C_t = \beta_0 + \beta_1 \ln INC_t + e_t,\ (log\text{-}linear) \quad \text{(3)}$$

For simplicity, we use the same symbols for the coefficients in the two equations but it is important to note that the slope coefficient β_1 has a different economic

interpretation in the two cases. In equation (2), β_1 it is the marginal propensity to consume (MPC for short), and is assumed to be a constant parameter. In the log linear model equation (3) β_1 is the elasticity of consumption in period t with respect to income, thus β_1 measures the percentage increase in C_t following a 1% increase in INC. Hence the log-linear specification in (3) implies that the marginal propensity to consume is itself a function of income. In that sense, the log-linear model is the least restrictive of the two, and in the rest of this example we use that specification.

SELF-ASSESSEMENT EXERCISE

Show that, after setting et = 0 (for convenience), MPC $\equiv \partial C_t/\partial INC_t = k.\ \beta_1 INCt^{\beta_1^{-1}}$, where $k = \exp(\beta_0)$.

You should note that macroeconomic textbooks usually omit the term e_t in equation (3), but for applications of the theory to actual data it is a necessary to get an intuitive grip on this disturbance term in the static consumption function. So: let us consider real data corresponding to C_t and Y_t, and assume that we have really good way of quantifying the intercept β_0 and the marginal propensity to consume β_1. You have learned about so called least-squares estimation in courses in econometrics, but intuitively, least-squares estimation is a way of finding the numbers for β_0 and β_1 that give the on average best prediction of C_t for a given value of Y_t. Using quarterly data for Nigeria, for the period 1967(1)-2002(4)–the number in brackets denotes the quarter–we obtain by using the least squares method in PcGive:

$$\ln \hat{C}_t = 0.02 + 0.99 \ln INC_t \quad \text{(4)}$$

where the "hat" in $\hat{C}_t$ is used to symbolize the fitted value of consumption given the income level INC_t. Next, use (3) and (4) to define the residual:

$$\hat{e}_t\text{:(2.4) } \hat{e}_t = \ln Ct - \ln \hat{C}_t, \quad \text{(5)}$$

which is the empirical counterpart to e_t.

A cross-plot of the 140 observations of consumption and income (in logarithmic scale), can be plotted in a graph. The straight line in the plotted graph represents the linear function in equation (4), and for each observation the distance up (or down) to the line

can be drawn. These "projections" are the graphical representation of the residuals $\hat{e}_t$. Clearly, if we are right in our arguments about how pervasive dynamic behaviour is in economics, (3) is a very restrictive formulation. All adjustments to a change in income are assumed to be completed in a single period, and if income suddenly changes next period, consumer's expenditure changes suddenly too. A dynamic model is obtained if we allow for the possibility that also period t- 1 income affects consumption, and that e.g., habit formation induces a positive relationship between period t - 1 and period t consumption:

$$ln\ C_t = \beta_0 + \beta_1\ ln\ INC_t + \beta_2\ ln\ INC_{t-1} + \alpha\ ln\ C_t\text{-}1 + \varepsilon_t \quad \text{(6)}$$

The literature refers to this type of model as an autoregressive distributed lag model, ARDL model for short. "Autoregressive" is due to the presence of ln C_{t-1} on the right hand side of the equation, so that consumption today depends on its own past. "Distributed lag" refers to the presence of lagged as well as current income in the model.

How can we investigate whether equation (6) is indeed a better description of the data than the static model? The answer to that question brings us in the direction of econometrics, but intuitively, one indication would be if the empirical counterpart to the disturbance of (6) are smaller and less systematic than the errors of equation (3). To test this, we obtain the residual $\hat{\varepsilon}_t$, again using the method of least squares to find the best fit of ln C_t according to the dynamic model:

$$ln\ \hat{C}_t = 0.04 + 0.13\ ln\ INC_t + 0.08\ ln\ INC_{t-1} + 0.79\ ln\ C_{t-1} \quad \text{(7)}$$

When plotted as graph the two residual series $\hat{e}_t$ and $\hat{\varepsilon}_t$, will immediately show clearly that the dynamic model in equation (7) is a much better description of the behaviour of private consumption than the static model in equation (4). As already stated, this is a typical finding with macroeconomic data.

Judging from the estimated coefficients in equation (7), one main reason for the improved fit of the dynamic model is the lag of consumption itself. That the lagged value of the endogenous variable is an important explanatory variable is also a typical

finding, and just goes to show that dynamic models represent essential tools for empirical macroeconomics. The rather low values of the income elasticities (0.130 and 0.08) may reflect that households find that a single quarterly change in income is “too little to build on” in their expenditure decisions. As we will see below, the results in equation (7) imply a much higher impact of a permanent change in income.

3.2 Dynamic Multipliers

The quotation from a typical central bank stated in sub-section 3.1 above on monetary policy shows that the central bank has formulated a view about the dynamic effects of a change in the interest rate on inflation. In the quotation, the central bank states that the effect will take place within two years, i.e., 8 quarters in a quarterly model of the relationship between the rate of inflation and the rate of interest. That statement may be taken to mean that the effect is building up gradually over 8 quarters and then dies away quite quickly, but other interpretations are also possible. In order to inform the public more fully about its view on the monetary policy transmission mechanism, the Bank would have to report a set of dynamic multipliers. Similar issues arise in almost all areas of applied macroeconomics, it is of vital interest to form an opinion on how fast an exogenous shock or policy change affects a variable of interest. The key concept needed to make progress on this is the dynamic multiplier. In order to explain the derivation and interpretation of dynamic multipliers, we first show what our estimated consumption function implies about the dynamic effect of a change in income. We then derive dynamic multipliers using a general notation for autoregressive distributed lag models.

3.2.1 Dynamic Effects of Increased Income on Consumption

We want to consider what the estimated model in equation (7) implies about the dynamic relationship between income and consumption. For this purpose there is no point to distinguish between fitted and actual values of consumption, so we drop the ˆ above Ct. Assume that income rises by 1% in period t, so instead of INC_t we have INC^1_t

t = INCt(1 + 0.01). Since income increases, consumption also has to rise. Using Equation (7) we have

$$\ln(C_t(1 + \delta c,0)) = 0.04 + 0.13 \ln(INC_t(1 + 0.01)) + 0.08 \ln INC_{t-1} + 0.79 \ln C_{t-1}$$

where δ c,0 denotes the relative increase in consumption in period $_t$, the first period of the income increase. Using the approximation $\ln(1+\delta c,0) = \delta c,0$ when $-1 < \delta c,0 < 1$,and noting that

$$\ln C_t - 0.04 - 0.13 \ln INC_t - 0.08 \ln INC_{t-1} - 0.79 \ln C_{t-1} = 0,$$

we obtain δ c,0 = 0.0013 as the relative increase in Ct. In other words, the immediate effect of a one percent increase in INC is a 0.13% rise in consumption.

The effect on consumption in the second period depends on whether the rise in income is permanent or only temporary. It is convenient to first consider the dynamic effects of a permanent shock to income. Note first that equation (7) holds also for period t + 1, i.e.,

$$\ln C_{t+1} = 0.04 + 0.13 \ln INC_{t+1} + 0.08 \ln INC_t + 0.79 \ln C_t$$

before the shock, and

$$\ln(Ct+1(1 + \delta c,1)) = 0.04 + 0.13 \ln(INCt+1(1 + 0.01)) + 0.08(\ln INCt(1 + 0.01)) + 0.79 \ln(Ct(1 + \delta c,0)),$$

after the shock. Remember that in period t + 1 not only INC_{t+1} have changed, but also INC_t and period t consumption (by δc,0). From this, the relative increase in Ct in period t + 1 is

$$\delta c,1 = 0.0013 + 0.0008 + 0.79 \times 0.0013 = 0.003125, \text{ or } 0.3\%.$$

By following the same way of reasoning, we find that the percentage increase in consumption in period t + 2 is 0.46% (formally δc,2 × 100).

Since δ c,0 measures the direct effect of a change in INC, it is usually called the **impact multiplier**, and can be defined directly by taking the partial derivative $\partial \ln Ct/\partial \ln INC_t$ in equation (7) (more on the relationship between derivative and multipliers in section 3.2 below). The dynamic multipliers δc,1, δc,2, ...δc,∞ are in their turn linked by exactly the same dynamics as in equation (7), namely:

$$\delta c,j = 0.13\delta inc,j + 0.08\delta inc,j-1 + 0.79\delta c,j-1, \text{ for } j = 1, 2, ...\infty. \quad (8)$$

For example, for j = 3, and setting $\delta inc,3 = \delta inc,2 = 0.01$ since we consider a permanent rise in income, we obtain:

$$\delta\ c,3 = 0.0013 + 0.0008 + 0.79 \times 0.0046 = 0.005734$$

or 0.57% in percentage terms. Clearly, the multipliers increase from period to period, but the increase is slowing down since in equation (8) the last multiplier is always multiplied by the coefficient of the autoregressive term, which is less than 1. Eventually, the sequences of multipliers are converging to what we refer to as the **long-run multiplier**. Hence, in equation (8) if we set $\delta_{c,j} = \delta_{c,j-1} = \delta_{c,\ long\text{-}run}$ we obtain

$$\delta_{c,long\text{-}run} = \frac{0.0013 + 0.00081}{-\ 0.79} = 0.01$$

meaning that according to the estimated model in equation (7), a 1% permanent increase in income has a 1% long-run effect on consumption. Remember that the set of multipliers we have considered so far represent the dynamic effects of a permanent rise in income, and they are shown for convenience in the first column of table 1. In contrast, a temporary rise in income (by 0.01) in equation (7) gives rise to another sequence of multipliers: The impact multiplier is again 0.0013, but the second multiplier becomes $0.13\times0+0.08\times0.01+0.79\times0.0013= 0.0018$, and the third is found to be $0.79 \times 0.0018 = 0.0014$, so these multipliers are rapidly approaching zero, which is also the long-run multiplier in this case.

Table 1: Dynamic multipliers of the estimated consumption function in Equation (7), percentage change in consumption after a 1 percent rise in income.

		Permanent1% Change	Temporary 1% change
	Impact period	0.13	0.13
1.	Period after shock	0.31	0.18
2.	Period after shock	0.46	0.14
	---	---	---
3.	Long-run multiplier	1.00	0.00

If you supplement the multipliers in the column to the right with a few more periods and then sum the whole sequence you find that the sum is close to 1, which is the long-run multiplier of the permanent change. A relationship like this always holds, no matter what the long-run effect of the permanent change is estimated to be. Heuristically, another way to think about the effect of a permanent change in an explanatory variable, is as the sum of the changes triggered by a temporary change. In this sense, the dynamic multipliers of a temporary change is the more fundamental of the two, since the dynamic effects of permanent shock can be calculated in a second step. Also, perhaps for this reason, many authors reserve the term dynamic multiplier for the effects of a temporary change and use a different term–cumulated multipliers–for the dynamic effects of a permanent change. However, as long as one is clear about which kind of shock we have in mind, no misunderstandings should occur by the term dynamic multipliers in both cases.

3.3 General Notation of Dynamic Models

As noted in the consumption function example, the impact multiplier is (after convenient scaling by 100) identical to the (partial) derivative of C_t with respect to INC_t. We now establish more formally that also the second, third and higher order multipliers can be interpreted as derivatives. At this stage it is also convenient to introduce the general notation for the autoregressive distributed lag model. In (3.8), yt is the endogenous variable while the xt and xt-1 make up the distributed lag part of the model:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \alpha y_{t-1} + \varepsilon_t. \quad \text{(9)}$$

In the same way as before, ε_t symbolizes a small and random part of y_t which is unexplained by x_t and x_{t-1} and the lagged endogenous variable y_{t-1}. In many applications, as in the consumption function example, y and x are in logarithmic scale, due to the specification of a log-linear functional form. However, in other applications, different units of measurement are the natural ones to use. Thus, frequently, y and x are measured in million kroner, in thousand persons or in percentage points. Mixtures of measurement are also frequently used in practice: for example in studies of labour demand, y_t may

denote the number of hours worked in the economy (or by an individual) while x_t denotes real wage costs per hour. The measurement scale does not affect the derivation of the multipliers, but care must be taken when interpreting and presenting the results. Specifically, only when both y and x are in logs, are the multipliers directly interpretable as percentage changes in y following a 1% increase in x, i.e., they are (dynamic) elasticities.

To establish the connection between dynamic multipliers and the derivatives of y_t, y_{t+1}, y_{t+2}, ..., it is convenient to define x_t, x_{t+1}, x_{t+2}, , ... as functions of a continuous variable h. When h changes permanently, starting in period t, we have $\partial xt/\partial h > 0$, but no change in x_{t-1} or in y_{t-1} since those variables are predetermined, in the period of the shock. Since xt is a function of h, so is yt, and the effect of yt of the change in *h* is founds as:

$$\frac{\partial yt}{\partial \mathrm{ht}} = \beta_1 \frac{\partial x}{\partial \mathrm{h}}$$

It is customary to consider "unit changes" in the explanatory variable (corresponding to the 1% change in income in the consumption function example), which means that we let $\partial \mathrm{xt}/\partial \mathrm{h} = 1$. Hence the first multiplier is:

$$\partial yt\ \partial h = \beta_1 \quad \text{(10)}$$

The second multiplier is found by considering the equation for period t + 1, i.e.,

$$y_{t+1} = \beta_0 + \beta_1 x_t + 1 + \beta_2 x_t + \alpha y_t + \varepsilon_{t+1}.$$

and calculating the derivative $\partial y_{t+1}/\partial \mathrm{h}$. Note that, due to the change in h occurring already in period t, both x_{t+1} and x_t have changed , i.e., $\partial x_{t+1}/\partial \mathrm{h} > 0$ and $\partial x_t/\partial \mathrm{h} > 0$. Finally, we need to keep in mind that also y_t is a function of *h*, hence:

$$\partial y_{t+1}\ \partial h = \beta_1\ \partial x\ \partial h_{t+1} + \beta_2\ \partial x\ \partial h_t + \alpha \partial y\ \partial h_t \quad \text{(11)}$$

Again, considering a unit change, and using equation (10), the second multiplier can be written as:

$$\frac{\partial y_{t+1}}{} = \beta_1 + \beta_2 + \alpha\beta_1 = \beta_1(1 + \alpha) + \beta_2 \quad \text{(12)}$$

∂h

To find the third derivative, consider

$$y_{t+2} = \beta_0 + \beta_1 x_{t+2} + \beta_2 x_{t+2} + \alpha y_{t+1} + \varepsilon_{t+2}$$

Using the same logic as above, we obtain;

$$\frac{\partial y_{t+2}}{\partial h} = \frac{\beta_1 \partial x}{\partial h}_{t+2} + + \beta_2 \frac{\partial x}{\partial h_{t+1}} + \alpha \frac{\partial y}{\partial h_{t+1}} \quad \text{(13)}$$

$$= \beta_1 + \beta_2 + \frac{\alpha \partial y_{t+1}}{\partial h}$$

$$= \beta_1(1 + \alpha + \alpha^2) + \beta_2(1 + \alpha)$$

where the unit-change, $\partial x_t/\partial h = \partial x_{t+1}/\partial h = 1$, is used in the second line, and the third line is the result of substituting $\partial y_{t+1}/\partial h$ out with the right hand side of equation (12).

Comparing, equation (11) and the first line of (13) there is a clear pattern: The third and second multipliers are linked by exactly the same form of dynamics that the govern the y variable itself. This also holds for higher order multipliers, and means that the multipliers can be computed recursively: Once we have found the second multiplier, the third can be found easily using the second line of equation (13). Table 2 shows summary of the results. In the table, we use the notation δ_j (j = 0, 1, 2, ...) for the multipliers. For, example δ_0 is identical to $\partial y_t/\partial h$, and δ_2 is identical to the third multiplier, $\partial y_{t+2}/\partial h$ above. In general, because the multipliers are linked recursively, multiplier number j + 1 is given as:

$$\delta_j = \beta_1 + \beta_2 + \alpha\delta_{j-1}, \text{ for } j = 1, 2, 3, \ldots \quad \text{(14)}$$

In the consumption function example, we saw that as long as the autoregressive parameter is less than one, the sequence of multipliers is converging towards a long run multiplier. In this more general case, the condition needed for the existence of a long run multiplier is that α is less than one in absolute value, formally $-1 < \alpha < 1$.

In the next section, this condition is explained in more detailed. section. For the present purpose we simply assume that the condition holds, and define the long run multiplier as $\delta_j = \delta_{j-1} = \delta_{long\text{-}run}$. Using equation (14), the expression for $\delta_{long\text{-}run}$ is found to be:

$$\delta_{long\text{-}run} = \frac{\beta_1 + \beta_2}{1 - \alpha}, \quad if\ -1 < \alpha < 1. \qquad (15)$$

Clearly, if $\alpha = 1$, the expression does not make sense mathematically, since the denominator is zero. Economically, it doesn't make sense either since the long run effect of a permanent unit change in x is an infinitely large increase in y (if $\beta_1 + \beta_2 > 0$). The case of $\alpha = -1$, may at first sight seem to be acceptable since the denominator is 2, not zero. However, as explained below, the dynamics is essentially unstable also in this case meaning that the long run multiplier is not well defined for the case of $\alpha = -1$.

Table 2: Dynamic multipliers of the general autoregressive distributed lag model.

ARDL model:	$y_t = \beta_0 + \beta_1 xt + \beta_2 x_{t-1} + \alpha y_{t-1} + \varepsilon_t$.	
Permanent unit change in x(1)	Temporary unit change in x(2)	
1. multiplier:	$\delta_0 = \beta_1$	$\delta_0 = \beta_1$
2. multiplier:	$\delta_1 = \beta_1 + \beta_2 + \alpha\delta_0$	$\delta_1 = \beta_2 + \alpha\delta_0$
3. multiplier:	$\delta_2 = \beta_1 + \beta_2 + \alpha\delta_1$	$\delta_2 = \alpha\delta_1$
.	.	.
j+1 multiplier	$\delta_j = \beta_1 + \beta_2 + \alpha\delta_{j-1}$	$\delta_j = \alpha\delta_{j-1}$
long-run	$\delta long\text{-}run = \beta_1+\beta_2$	0
notes:	(1) As explained in the text, $\partial x_{t+j}/\partial h = 1$, j = 0, 1, 2, ... (2) $\partial x_t/\partial h = 1$, $\partial x_{t+j}/\partial h = 0$, j = 1, 2, 3, ... If y and x are in logs, the multipliers are in percent.	

3.4 Multipliers in the Text Books to This Course

As already noted in the 3.1 above the distinction between short and long-run multipliers permeates modern macroeconomics, and so is not special to the consumption function example above! You are invited to be on the look-out for expressions like short and long-run effects/multipliers/elasticities in the textbook by Burda and Wyplosz (2001). One early example in the book is found in Chapter 8, on money demand, Table 8.4. Note the striking difference between the short-run and long-run multipliers for all countries, a direct parallel to the consumption function example we have worked through in this section. Hence, the precise interpretation of (log) linearized money demand function in B&W's Box 8.4 is as a so called steady state relationship, thus the parameter μ is a long-run elasticity with respect to income. In the next section of this note, the relationship between long-run multipliers and steady state relationships in explained. The money demand function also plays an important role in Chapter 9 and 10, and in later chapters in B&W. In the book by B&W, the distinction between short and long-run is a main issue in Chapter 12, where short and long-run supply curves are derived. For example, the slope of the short-run curves in figure 12.6 corresponds to the impact multipliers of the respective models, while the vertical long-run curve suggest that the long-run multipliers are infinite. In section 8 below, we go even deeper into wage-price dynamics, using the concepts that we have introduced here. You should also study the joint determination of inflation, output and the exchange rates which are also dynamic in nature, so a good understanding the logic of dynamic multipliers will prove very useful. Interest rate setting with the aim of controlling inflation is one specific example. However, we will not always need (or be able to) give a full account of the multipliers of this more complex model. Often we will concentrate on the impact and long-run multipliers.

3.0 CONCLUSION

As students of economics you are well acquainted to model-based analysis, graphical or algebraic. Presumably, most of the models you have used have been

static, since time has played no essential part in the model formulation or in the analysis. This note therefore started, in section 3.1 by contrasting static models with models which have a dynamic formulation. Typically, dynamic models give a better description of macroeconomic time series data than static models.

5.0 SUMMARY

In this unit you learnt static and dynamic models, dynamic multipliers and dynamic effects of increased income on consumption. You also learnt how to write general notation of dynamic models and how to compare multipliers in the text books with this course. In the next unit we shall discuss autoregressive distributed lag popularly called ARDL models.

6.0 TUTOR-MARKED ASSIGNMENT

1. Use the numbers from the estimated consumption function and check that by using the formulae of Table 2, you obtain the same numerical results as in section 3.1.
2. Check that you understand, and are able to derive the results in the column for a temporary change in x_t in Table 2

7.0 REFERENCES AND FURTHER READING

Aukrust, O. (1977). Inflation in the Open Economy. An Norwegian Model. In Klein, L. B. and W. S. Sâlant (eds.), World Wide Inflation. Theory and Recent Experience. Brookings, Washington D.C.

Bjerkholt, O. (1998). Interaction Between Model Builders and Policy Makers in the Norwegian Tradition. Economic Modelling, 15, 317—339.

Burda, M. and C. Wyplosz (2001). Macroeconomics. A European Text.. Oxford University Press, Oxford, 3rd edn.

Nymoen, R. .(2004). Dynamic models

Nymoen, R. (1989). Modelling Wages in the Small Open Economy: An ErrorCorrection Model of Norwegian Manufacturing Wages. Oxford Bulletin of Economics and Statistics, 51, 239—258.
Nymoen, R. (1991). A Small Linear Model of Wage- and Price-Inflation in the Norwegian Economy. Journal of Applied Econometrics, 6, 255—269.

UNIT4: AUTOREGRESSIVE DISTRIBUTED LAG (ARDL) MODEL

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
3.1 ARDL Cointegration Equations
3.2 ARDL Computation in Eviews
3.3 A practical Example on ARDL
3.3.1 The ARDL Bound cointegration test model:
3.3.2 Diagnostic Check for Serial Correlation
3.4 ARDL Cointegration Bounds test.
4.0 CONCLUSION
5.0 SUMMARY
8.0 TUTOR MARKED ASSIGNMENT
7.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the previous unit you learnt static and dynamic models, dynamic multipliers and dynamic effects of increased income on consumption. You also learnt how to write general notation of dynamic models and how to compare multipliers in the text books with this course. In the present unit we shall discuss autoregressive distributed lag commonly called ARDL models.

Autoregressive Distributive Lagged (ARDL)-Bounds testing approach. This test was developed by Pesaran and Shin (1999) and later extended by Pesaran, Shin and Smith (2001). The ARDL bound test has superiority over Johansen (1991) and Engle and Granger (1987) cointegration approaches because for a number of reasons The endogeneity problems and inability to test hypotheses on the limited coefficients in the long run associated with the Engle-Granger Method are avoided.

2.0 OBJECTIVES

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

* Discuss ARDL Cointegration Equations
* Compute ARDL in Eviews
* Design the ARDL Bound cointegration test model:
* Conduct a diagnostic check for Serial Correlation in ARDL
* Explain ARDL Cointegration Bounds test.

3.0 MAIN CONTENTS

3.1 ARDL Cointegration Equations

Pesaran et al. (2001) first conduct the bounds tests in the unrestricted model or namely an ARDL (p,p,p,p,p) model (see their paper, Equation 30), and secondly adopt the ARDL (p,q,r,s,v) approach to the estimation of the level relations.

The VAR(*p*) model can be rewritten in vector ECM form as:

$$\Delta z_t = a_0 + a_1 trend + \boldsymbol{\pi} z_{t-1} + \sum_{i=1}^{\rho-1} \forall_i \Delta z_{t-i} + \varepsilon_t$$

where

$\Delta = 1 - L$ is the difference operator,

$z_t = f(y_t, x_t)$

we now partition the long-run multiplier matrix π conformably with $z_t = (y_t, x'_t)'$ as

$$\pi = \begin{pmatrix} \pi_{yy} & \pi_{yx} \\ \pi_{xy} & \pi_{xx} \end{pmatrix}$$

Under the assumption 1, 3, and 4 (see Pesaran et al. 2001), π has rank *r* and is given by

$$\pi = \begin{pmatrix} \pi_{yy} & \pi_{yx} \\ 0 & \pi_{xx} \end{pmatrix}$$

Consequently, the *conditional* ECM can be written as following:

$$\Delta y_t = a_0 + a_1 trend + \boldsymbol{\pi}_{yy} y_{t-1} + \boldsymbol{\pi}_{yx.x} x_{t-1} + \sum_{i=1}^{\rho-1} \forall_i \Delta z_{t-i} + w' \Delta x_t + \varepsilon_t$$

If the $\pi_{yy} \neq 0$ and $\pi_{yx.x} = 0'$, the y_t is (trend) stationary or I(0), whatever the value *r*. Consequently, the differenced variable Δy_t depends only on its own lagged level y_{t-1} in the conditional ECM. Second, if $\pi_{yy} = 0$, the Δy_t depends only on the lagged level x_{t-1} in the conditional ECM model. Therefore, in order to test for the absence of level effects in the conditional ECM model and more crucially, the absence of a level relationship between y_t and x_t, the emphasis in this approach is a test of the joint hypothesis the $\pi_{yy} = 0$ and $\pi_{yx.x} = 0'$ in the above model.

According to Pesaran et al. (2001), there are 5 cases provided for testing the cointegrating bound test:

Case 1: (no intercepts; no trends) a_0 and $a_1 = 0$.

Case 2: (restricted intercepts; no trends) $a_0 = -(\pi_{yy}, \pi_{yx.x})\mu$ and $a_1 = 0$.

Case 3: (unrestricted intercepts; no trends) $a_0 \neq 0$ and $a_1 = 0$.

Case 4: (unrestricted intercepts; restricted trends) $a_0 \neq 0$ and $a_1 = -(\pi_{yy}, \pi_{yx.x})\mu$

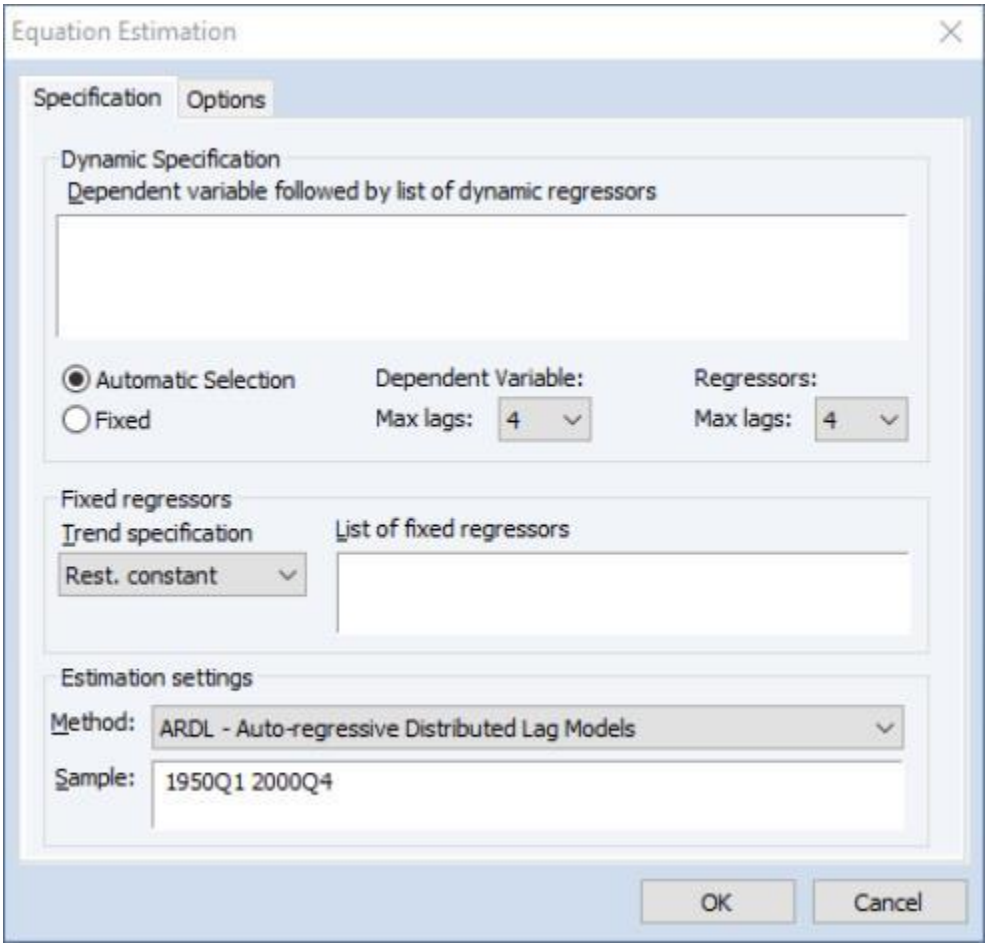
Case 5: (unrestricted intercepts; unrestricted trends) $a_0 \neq 0$ and $a_1 \neq 0$

If the Wald F-statistic fall	**Conclusion**
a. above the upper critical value	Cointegration
b. between the lower bound and upper bound critical value	Inconclusive
c. below the lower bound critical value	No Cointegration

3.2 ARDL Computation in Eviews

Since ARDL models are least squares regressions using lags of the dependent and independent variables as regressors, they can be estimated in EViews using an equation object with the Least Squares estimation method.

However, EViews also offers a specialized estimator for handling ARDL models. This estimator offers built-in lag-length selection methods, as well as post-estimation views. To estimate an ARDL model using the ARDL estimator, open the equation dialog by selecting Quick/Estimate Equation…, or by selecting Object/New Object…/Equation and then selecting ARDL from the Method dropdown menu. EViews will then display the ARDL estimation dialog:



The Specification tab allows you to specify the variables used in the regression, and whether to let EViews automatically detect the appropriate number of lags for each variable.

To begin, enter the name of the dependent variable, followed by a space delimited list of dynamic regressors (i.e., variables which will have lag terms in the model) in the Dynamic Specification edit box. You may then select whether you wish EViews to automatically select the number of lags for all variables by selecting the Automatic Selection radio button, fixing the independent variable and the regressors to a uniform fixed length by selecting the Fixed radio buttons, or by taking full control of granularity and specifying a specific lag for each of the independent and regressors variables. The latter can be specified via command in the Dynamic Specification edit box by replacing each variable by the Fixed Lag command @FL(VARIABLE). However, to enable you grasped the ARDL computational rudiments using Eviews, we are following the following steps in ARDL conitegration Bounds test instead of directly using ARDL in EViews drop down menu.

3.3 A practical Example on ARDL

The determinants of financial development

Data file: FD FDI TO.xls **(Annual Data from 1970 – 2009, 40 observations)**

Date	fd	fdi	to	pri	date	fd	fdi	to	pri
1970	3807.4	9.47	9.69	11.19	1990	5708.1	13.35	8.25	8.38
1971	3906.3	10.03	10.56	12.64	1991	5797.4	13.55	8.24	8.62
1972	3976	10.83	11.39	12.64	1992	5850.6	12.1	8.16	8.25
1973	4034	11.51	9.27	11.1	1993	5846	11.91	7.74	7.76
1974	4117.2	10.51	8.48	10.68	1994	5880.2	10.65	6.43	7.27
1975	4175.7	9.24	7.92	9.76	1995	5962	9.33	5.86	7.25
1976	4258.3	8.37	7.9	9.29	1996	6033.7	10.29	5.64	6.89
1977	4318.7	7	8.1	8.84	1997	6092.5	9.12	4.82	5.84
1978	4382.4	6.16	7.83	7.94	1998	6190.7	7.32	4.02	5.77
1979	4423.2	5.9	6.92	7.18	1999	6295.2	6.53	3.77	5.78
1980	4491.3	5.37	6.21	6.66	2000	6389.7	5.61	3.26	4.68
1981	4543.3	4.95	6.27	6.48	2001	6493.6	4.42	3.04	5
1982	4611.1	5.69	6.22	6.52	2002	6544.5	3.54	3.04	4.64
1983	4686.7	6.62	6.65	7.72	2003	6622.7	3.28	3	4.41
1984	4764.5	6.47	6.84	8.15	2004	6688.3	2.59	3.06	4.32
1985	4883.1	6.4	6.92	8.29	2005	6813.8	3.25	2.99	4.41
1986	4948.6	6.4	6.66	7.58	2006	6916.3	4.43	3.21	4.9
1987	5059.3	6.73	7.16	8.1	2007	7044.3	4.25	3.94	6.2
1988	5142.8	7.65	7.98	8.59	2008	7131.8	4.17	4.49	6.56
1989	5251	8.57	8.47	8.75	2009	7248.2	4	5.17	7.4

The financial development indicator is private sector credit (PRI), and three determinants are the real GDP per capita (GDPC), foreign direct investment (FDI) and trade openness (TO).

3.3.1 The ARDL Bound cointegration test model:

$$\Delta PRI_t = c + \beta_1 PRI_{t-1} + \beta_2 GDPC_{t-1} + \beta_3 FDI_{t-1} + \beta_4 TO_{t-1} + \sum_{i=1}^{p} \alpha_{1i} \Delta PRI_{t-i} + \sum_{i=1}^{p} \alpha_{2i} \Delta GDPC_{t-i} + \sum_{i=1}^{p} \alpha_{3i} \Delta FDI_{t-i} + \sum_{i=1}^{p} \alpha_{4i} \Delta TO_{t-i} + DUM$$

where

PRI = private sector credit (Financial Development), % of GDP

GDPC = real GDP per capita (Malaysian ringgit, RM)

C = constant

FDI = foreign direct investment (% of GDP)

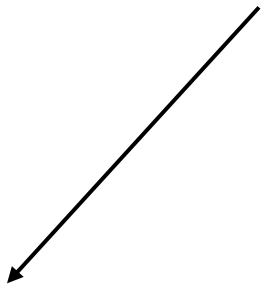
TO = total trade (% of GDP)

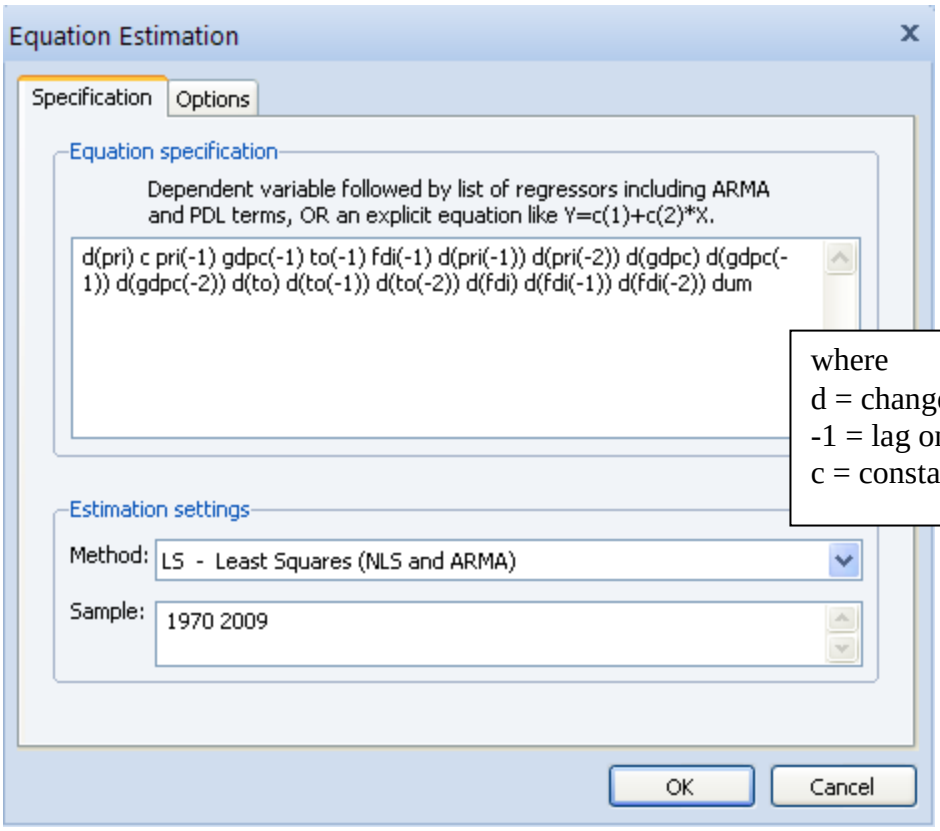
p = optimum lag length

DUM = 1 for crisis and 0 for otherwise; which refer to the Malaysian oil crisis in years 1973, 1974, 1980 and 1981; commodities crisis in years 1985 and 1986; and East Asian financial crisis in years 1997 and 1998.

First, we examine the Bounds test by selecting the higher lag length. In our example, the sample period is covering from 1970 – 2009 (40 observations). In order to avoid the over parameter problem, we start with the maximum lag order 2 and then reduce to lag 1:

After opening the data file, select “Quick” – “Estimate Equation” and type the model





where
d = change (First difference or Δ)
-1 = lag one variable or t –1
c = constant term

The estimated result is:

Dependent Variable: D(PRI)
Method: Least Squares
Sample (adjusted): 1973 2009
Included observations: 37 after adjustments

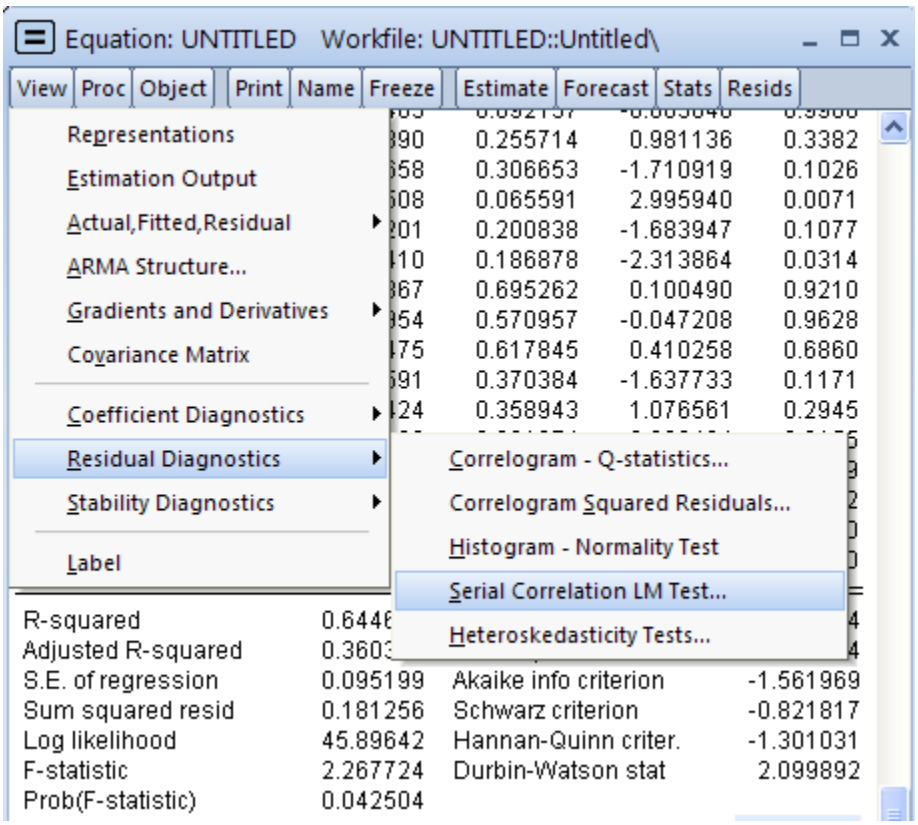
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.082957	1.075871	0.077107	0.9393
PRI(-1)	-0.000465	0.092137	-0.005048	0.9960
GDPC(-1)	0.250890	0.255714	0.981136	0.3382
TO(-1)	-0.524658	0.306653	-1.710919	0.1026
FDI(-1)	0.196508	0.065591	2.995940	0.0071
D(PRI(-1))	-0.338201	0.200838	-1.683947	0.1077
D(PRI(-2))	-0.432410	0.186878	-2.313864	0.0314
D(GDPC)	0.069867	0.695262	0.100490	0.9210
D(GDPC(-1))	-0.026954	0.570957	-0.047208	0.9628
D(GDPC(-2))	0.253475	0.617845	0.410258	0.6860
D(TO)	-0.606591	0.370384	-1.637733	0.1171
D(TO(-1))	0.386424	0.358943	1.076561	0.2945
D(TO(-2))	-0.090166	0.381374	-0.236424	0.8155
D(FDI)	0.029113	0.044021	0.661335	0.5159
D(FDI(-1))	-0.127029	0.048388	-2.625204	0.0162
D(FDI(-2))	-0.079244	0.042969	-1.844223	0.0800
DUM	0.103538	0.053661	1.929492	0.0680

R-squared	0.644657	Mean dependent var	0.044224
Adjusted R-squared	0.360382	S.D. dependent var	0.119034
S.E. of regression	0.095199	Akaike info criterion	-1.561969
Sum squared resid	0.181256	Schwarz criterion	-0.821817
Log likelihood	45.89642	Hannan-Quinn criter.	-1.301031
F-statistic	2.267724	Durbin-Watson stat	2.099892
Prob(F-statistic)	0.042504		

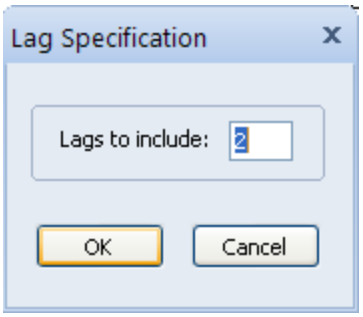
3.3.2 Diagnostic Check for Serial Correlation

Perform diagnostic check for serial correlation using the Breusch-Godfrey LM test

Select “View” – “Residual Diagnostics” – “Serial Correlation LM Test”:



Lag Specification: 2



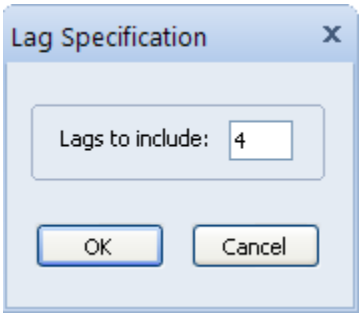
Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.456064	Prob. F(2,18)	0.2593
Obs*R-squared	5.152451	Prob. Chi-Square(2)	0.0761

The LM test indicates serial correlation problem since the p-value is significance at 10%.

Repeat the same process but now with

Lag Specification: 4

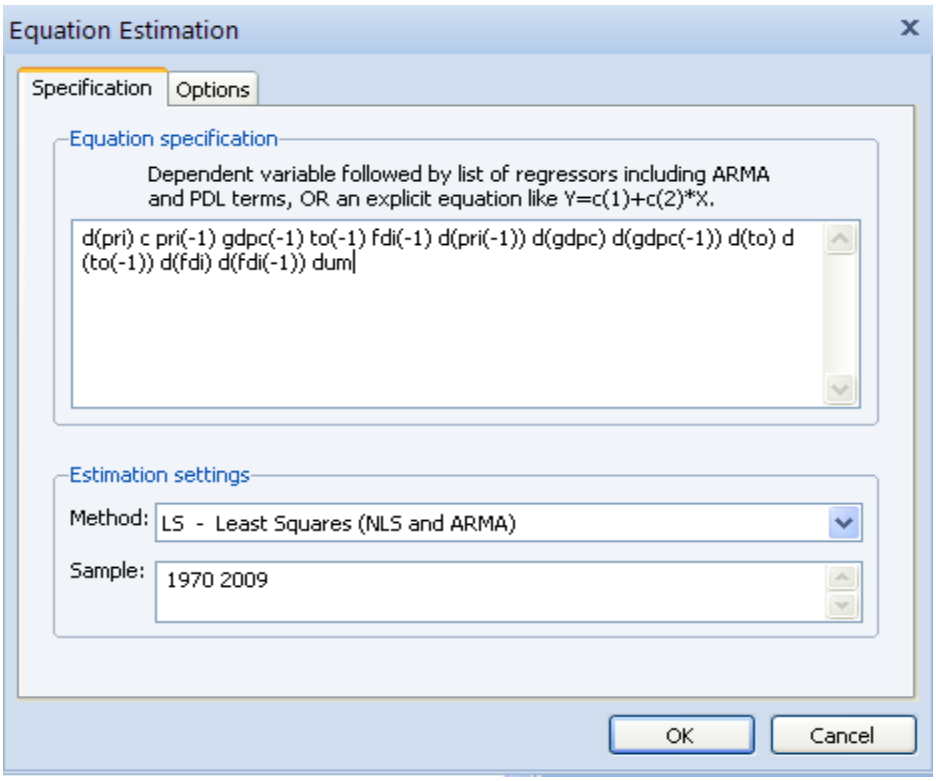


Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.231322	Prob. F(4,16)	0.3368
Obs*R-squared	8.708870	Prob. Chi-Square(4)	0.0688

The LM test indicates that serial correlation problem exist (the p-value is significance at 10%).

Since the lag length 2 model specification demonstrates serial correlation problem (10% significance level), we attempt to reduce the lag length to 1.



The empirical result is:

Dependent Variable: D(PRI)
Method: Least Squares
Sample (adjusted): 1972 2009
Included observations: 38 after adjustments

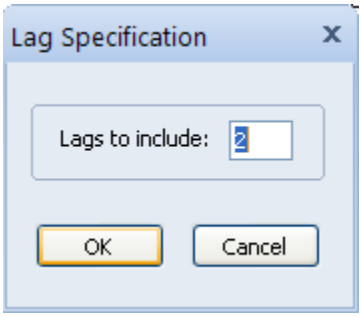
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.381073	0.944119	-0.403628	0.6899
PRI(-1)	-0.027029	0.084086	-0.321446	0.7505
GDPC(-1)	0.259557	0.226898	1.143938	0.2635
TO(-1)	-0.408034	0.286617	-1.423621	0.1669
FDI(-1)	0.110760	0.054081	2.048045	0.0512
D(PRI(-1))	-0.184509	0.198541	-0.929326	0.3616
D(GDPC)	0.138098	0.675542	0.204426	0.8397
D(GDPC(-1))	0.236633	0.599266	0.394871	0.6963
D(TO)	-0.892568	0.338281	-2.638541	0.0141
D(TO(-1))	0.485773	0.358980	1.353201	0.1881
D(FDI)	0.033571	0.044480	0.754751	0.4574
D(FDI(-1))	-0.065371	0.039495	-1.655152	0.1104
DUM	0.111442	0.054497	2.044930	0.0515

R-squared	0.495417	Mean dependent var	0.044951
Adjusted R-squared	0.253217	S.D. dependent var	0.117500
S.E. of regression	0.101539	Akaike info criterion	-1.471242
Sum squared resid	0.257756	Schwarz criterion	-0.911015
Log likelihood	40.95360	Hannan-Quinn criter.	-1.271918
F-statistic	2.045489	Durbin-Watson stat	2.218071
Prob(F-statistic)	0.063526		

Perform the diagnostic check for serial correlation LM test

Click “View” – “Residual Diagnostics” – “Serial Correlation LM Test”:

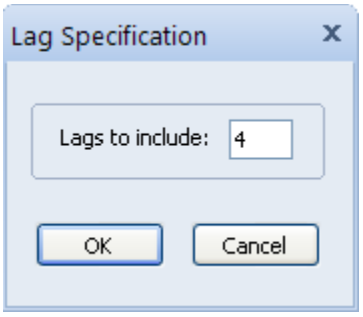
Lag Specification = 2



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.588804	Prob. F(2,23)	0.0440
Obs*R-squared	9.038129	Prob. Chi-Square(2)	0.0109

Lag Specification = 4



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.462778	Prob. F(4,21)	0.0767
Obs*R-squared	12.13383	Prob. Chi-Square(4)	0.0164

Again, both lag specifications indicate serial correlation problem (the p-values are less than 0.05)

The above empirical results can be summarized as follow:

Table 1: Optimal Lag-length Selection

P	AIC	SBC	$x_{SC}(2)$	$x_{SC}(4)$
2	-1.5619	-0.8218	5.15*	8.71*
1	-1.4712	-0.9110	9.038**	12.133**

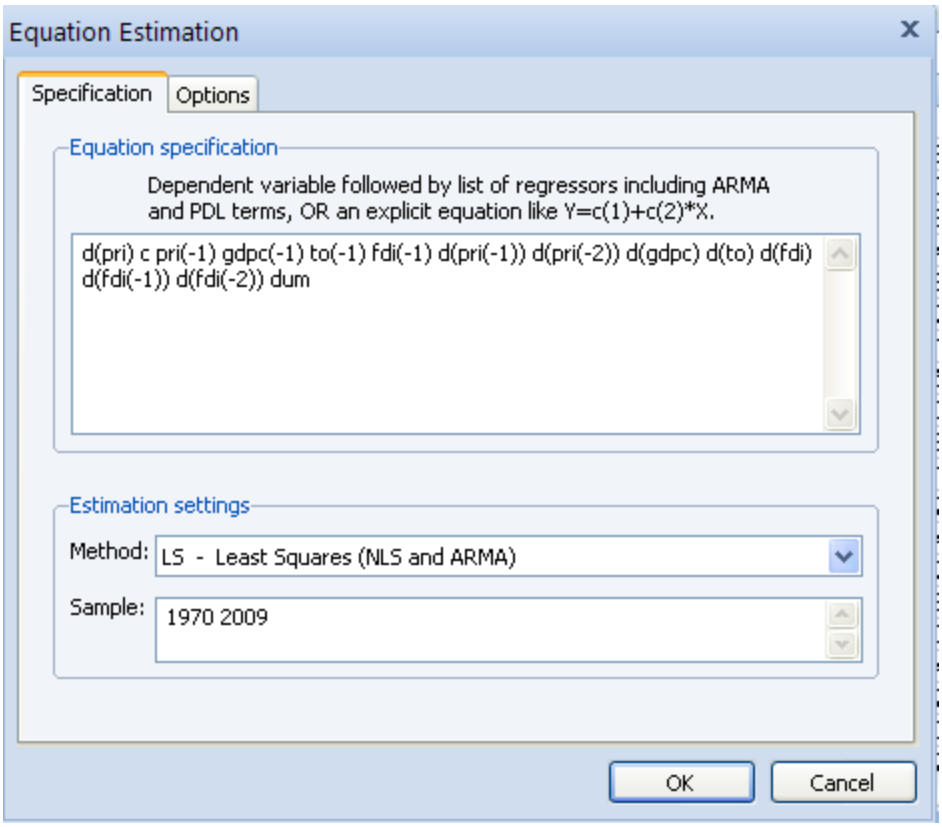
Note: *p* is the lag order of the underlying VAR model for the conditional ECM ( ), with zero restrictions on the coefficients of lagged changes in the independent variables. AICp = (-2*l* / T) + (2k / T) and SBCp = (-2*l* / T) + (k * *log*T / T) denote Akaike's and Schwarz's Bayesian Information Criteria for a given lag order p, where *l* is the maximized log-likelihood value of the model, k is the number of freely estimated coefficients and T is the sample size. The AIC and SBC are often used in model selection for non-nested alternatives—lowest values of the AIC and SBC are preferred (refer to Eviews Users Guide 4.0, pp. 279). X_{sc} (1) and X_{sc} (4) are LM statistics for testing no residual serial correlation against orders 2 and 4. The symbols ***, ** and * denote significance at 0.01, 0.05 and 0.10 levels, respectively.

Table 1 reveals that serial correlation problem exists in both models. Hence, we consider other approach namely Hendry's General to Specific method to select the optimum lag length. Again, we start the model with maximum lag 2, then eliminate the higher insignificant lag for the changes (Δ) variables.

For example: If D(TO(-2)) is insignificant, then it should be eliminated from the model first; follow by D(TO(-1)), and then other variable such as D(GDPC(-2)), etc.

At the end, the best model using the General to Specific criterion is as follows:

d(pri) c pri(-1) gdpc(-1) to(-1) fdi(-1) d(pri(-1)) d(pri(-2)) d(gdpc) d(to) d(fdi) d(fdi(-1)) d(fdi(-2)) dum



Dependent Variable: D(PRI)
Method: Least Squares
Sample (adjusted): 1973 2009
Included observations: 37 after adjustments

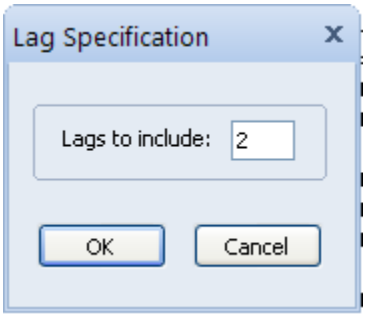
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.498144	0.729822	0.682556	0.5014
PRI(-1)	-0.020647	0.077317	-0.267045	0.7917
GDPC(-1)	0.148658	0.170390	0.872457	0.3916
TO(-1)	-0.397749	0.242923	-1.637349	0.1146
FDI(-1)	0.215580	0.053784	4.008287	0.0005
D(PRI(-1))	-0.371679	0.176409	-2.106914	0.0458
D(PRI(-2))	-0.437332	0.170173	-2.569917	0.0168
D(GDPC)	-0.341142	0.554339	-0.615403	0.5441
D(TO)	-0.466088	0.307703	-1.514733	0.1429
D(FDI)	0.043925	0.038780	1.132679	0.2685
D(FDI(-1))	-0.122975	0.042332	-2.905006	0.0078
D(FDI(-2))	-0.085573	0.036325	-2.355766	0.0270
DUM	0.092284	0.048225	1.913599	0.0677

R-squared	0.615621	Mean dependent var	0.044224
Adjusted R-squared	0.423432	S.D. dependent var	0.119034
S.E. of regression	0.090385	Akaike info criterion	-1.699641
Sum squared resid	0.196066	Schwarz criterion	-1.133643
Log likelihood	44.44337	Hannan-Quinn criter.	-1.500101

F-statistic	3.203201	Durbin-Watson stat	1.995463
Prob(F-statistic)	0.007335		

Perform the diagnostic check for serial correlation LM test:

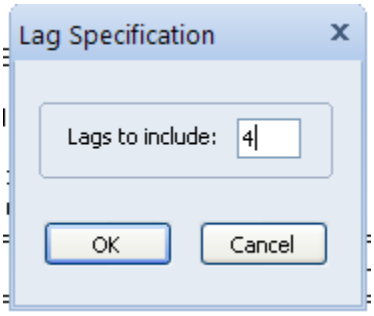
Lag 2



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.200978	Prob. F(2,22)	0.3199
Obs*R-squared	3.642018	Prob. Chi-Square(2)	0.1619

Lag 4



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.561262	Prob. F(4,20)	0.2232
Obs*R-squared	8.804207	Prob. Chi-Square(4)	0.0662

The Breusch-Godfrey LM test indicates that the residuals are homoscedasticity (no serial correlation). Even though the lag 4 LM test reveals serial correlation, but is not severe (since the p-value is significance at 10%).

SELF-ASSESEMENT EXERCISE

What is ARDL Bound Test?

3.4 ARDL Cointegration Bounds Test.

After specifying the optimum lag model, we proceed to the ARDL Cointegration Bounds test. However, we need to know the code used in Eviews for hypothesis testing – c(1) represents the first coefficient (constant); c(2) represents the second coefficient PRI(-1); c(3) represents the third coefficient GDPC(-1), etc.

Dependent Variable: D(PRI)
Method: Least Squares
Date: 06/29/11 Time: 14:04
Sample (adjusted): 1973 2009
Included observations: 37 after adjustments

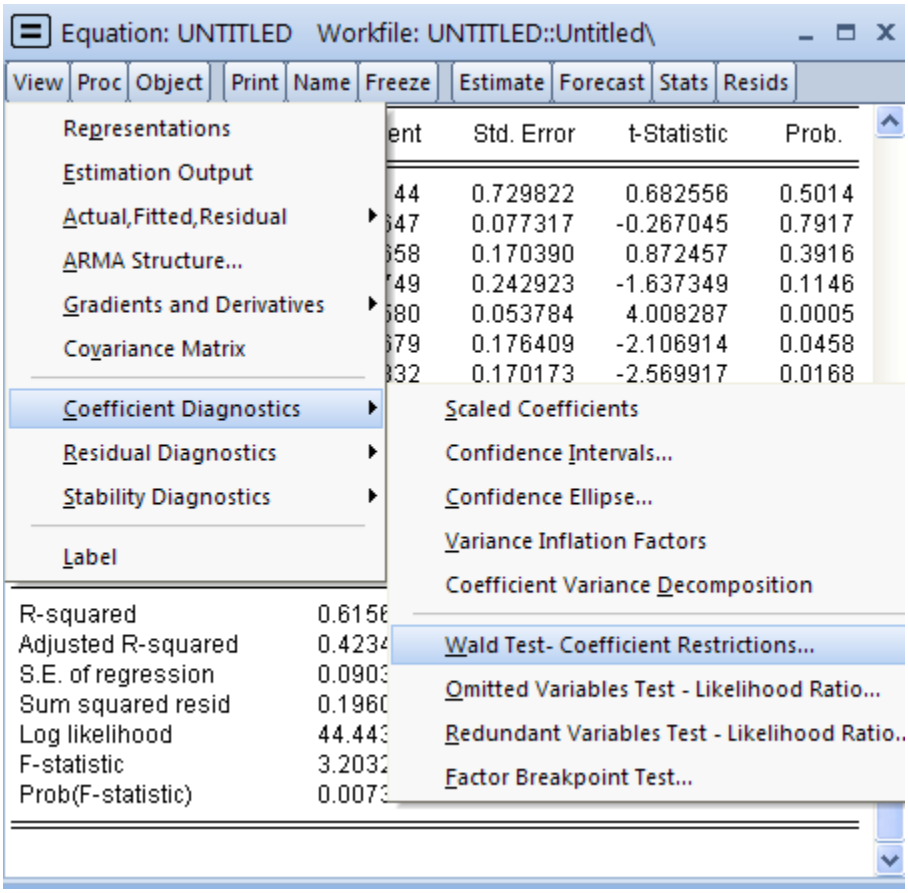
	Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	C	0.498144	0.729822	0.682556	0.5014
C(2)	PRI(-1)	-0.020647	0.077317	-0.267045	0.7917
C(3)	GDPC(-1)	0.148658	0.170390	0.872457	0.3916
C(4)	TO(-1)	-0.397749	0.242923	-1.637349	0.1146
C(5)	FDI(-1)	0.215580	0.053784	4.008287	0.0005
C(6)	D(PRI(-1))	-0.371679	0.176409	-2.106914	0.0458
C(7)	D(PRI(-2))	-0.437332	0.170173	-2.569917	0.0168
C(8)	D(GDPC)	-0.341142	0.554339	-0.615403	0.5441
c(9)	D(TO)	-0.466088	0.307703	-1.514733	0.1429
C(10)	D(FDI)	0.043925	0.038780	1.132679	0.2685
C(11)	D(FDI(-1))	-0.122975	0.042332	-2.905006	0.0078
C(12)	D(FDI(-2))	-0.085573	0.036325	-2.355766	0.0270
C(13)	DUM	0.092284	0.048225	1.913599	0.0677

R-squared	0.615621	Mean dependent var	0.044224
Adjusted R-squared	0.423432	S.D. dependent var	0.119034
S.E. of regression	0.090385	Akaike info criterion	-1.699641
Sum squared resid	0.196066	Schwarz criterion	-1.133643
Log likelihood	44.44337	Hannan-Quinn criter.	-1.500101
F-statistic	3.203201	Durbin-Watson stat	1.995463
Prob(F-statistic)	0.007335		

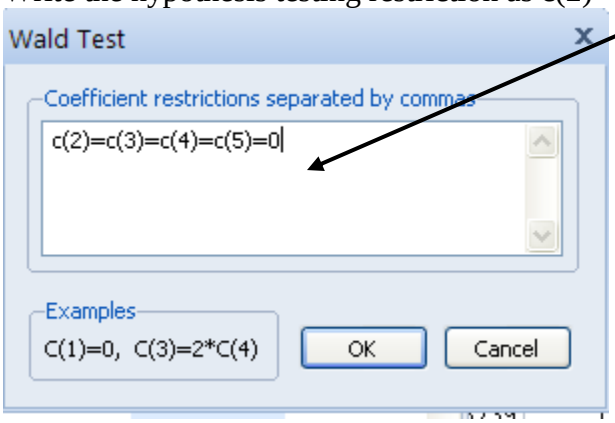
According to Pesaran et al. (2001), if the coefficients among the lag 1 variables (level) are jointly fall above the upper bound critical value, this implies that there is a long-run cointegration relationship among the variables.

In order to test this hypothesis, we need to restrict the coefficients PRI(-1) = GDPC(-1) = TO(-1) = FDI(-1) = 0.

Select “View” – “Coefficient Diagnostics” – “Wald Test - coefficient restrictions”.



Write the hypothesis testing restriction as c(2) = c(3) = c(4) = c(5) = 0.



Wald Test:
Equation: Untitled

Test Statistic	Value	Df	Probability
F-statistic	6.270894	(4, 24)	0.0013
Chi-square	25.08358	4	0.0000

Null Hypothesis: C(2)=C(3)=C(4)=C(5)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.020647	0.077317
C(3)	0.148658	0.170390
C(4)	-0.397749	0.242923
C(5)	0.215580	0.053784

Restrictions are linear in coefficients.

Compare the F-statistic with the Narayan (2005) Critical value (if the sample size is relative small, < 100 observations).

Compare the F-statistic value with critical value provided by Pesaran et al. (2001). However, if the sample size is small (< 100 observations), then compare with the critical value provided by Narayan (2005).

k = the dimension of
$x_t = (RGDP_t, TO_t, FDI_t)$ or 3

n = 40 (1970 – 2009)

Critical values for the bounds test: case III: unrestricted intercept and no trend

	$k=0$		$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
n	*I*(0)	*I*(1)	*I*(0)	*I*(1)	*I*(0)	*I*(1)	*I*(0)	*I*(1)	*I*(0)	*I*(1)	*I*(0)	*I*(1)
1 per cent												
30	13.680	13.680	8.170	9.285	6.183	7.873	5.333	7.063	4.768	6.670	4.537	6.370
35	13.290	13.290	7.870	8.960	6.140	7.607	5.198	6.845	4.590	6.368	4.257	6.040
40	13.070	13.070	7.625	8.825	5.893	7.337	5.018	6.610	4.428	6.250	4.045	5.898
45	12.930	12.930	7.740	8.650	5.920	7.197	4.983	6.423	4.394	5.914	4.030	5.598
50	12.730	12.730	7.560	8.685	5.817	7.303	4.865	6.360	4.306	5.874	3.955	5.583
55	12.700	12.700	7.435	8.460	5.707	6.977	4.828	6.195	4.244	5.726	3.928	5.408
60	12.490	12.490	7.400	8.510	5.697	6.987	4.748	6.188	4.176	5.676	3.783	5.338
65	12.400	12.400	7.320	8.435	5.583	6.853	4.690	6.143	4.188	5.694	3.783	5.300
70	12.240	12.240	7.170	8.405	5.487	6.880	4.635	6.055	4.098	5.570	3.747	5.285
75	12.540	12.540	7.225	8.300	5.513	6.860	4.725	6.080	4.168	5.548	3.772	5.213
80	12.120	12.120	7.095	8.260	5.407	6.783	4.568	5.960	4.096	5.512	3.725	5.163
5 per cent												
30	8.770	8.770	5.395	6.350	4.267	5.473	3.710	5.018	3.354	4.774	3.125	4.608
35	8.640	8.640	5.290	6.175	4.183	5.333	3.615	4.913	3.276	4.630	3.037	4.443
40	8.570	8.570	5.260	6.160	4.133	5.260	3.548	4.803	3.202	4.544	2.962	4.338
45	8.590	8.590	5.235	6.135	4.083	5.207	3.535	4.733	3.178	4.450	2.922	4.268
50	8.510	8.510	5.220	6.070	4.070	5.190	3.500	4.700	3.136	4.416	2.900	4.218
55	8.390	8.390	5.125	6.045	3.987	5.090	3.408	4.623	3.068	4.334	2.848	4.160
60	8.460	8.460	5.125	6.000	4.000	5.057	3.415	4.615	3.062	4.314	2.817	4.097
65	8.490	8.490	5.130	5.980	4.010	5.080	3.435	4.583	3.068	4.274	2.835	4.090
70	8.370	8.370	5.055	5.915	3.947	5.020	3.370	4.545	3.022	4.256	2.788	4.073
75	8.420	8.420	5.140	5.920	3.983	5.060	3.408	4.550	3.042	4.244	2.802	4.065
80	8.400	8.400	5.060	5.930	3.940	5.043	3.363	4.515	3.010	4.216	2.787	4.015
10 per cent												
30	6.840	6.840	4.290	5.080	3.437	4.470	3.008	4.150	2.752	3.994	2.578	3.858
35	6.810	6.810	4.225	5.050	3.393	4.410	2.958	4.100	2.696	3.898	2.508	3.763
40	6.760	6.760	4.235	5.000	3.373	4.377	2.933	4.020	2.660	3.838	2.483	3.708
45	6.760	6.760	4.225	5.020	3.330	4.347	2.893	3.983	2.638	3.772	2.458	3.647

The result can be summarized as follows:

Table 2. F-statistics for testing the existence of long-run cointegration

Model	**F-statistic**
Model 1: PRI = *f* (GDPC, FDI, TO)	**6.27****

Narayan (2005)	**k = 3, n=40**	
Critical Value	**Lower bound**	**Upper bound**
1%	5.018	6.610
5%	3.548	4.803
10%	2.933	4.020

Notes: *, **, and *** denote significant at 10%, 5%, and 1% levels, respectively. Critical values are obtained from Narayan (2005) (Table Case III: Unrestricted intercept and no trend; pg. 1988).

In this example, the model shows that there is a long-run cointegration relationship among financial development and it determinants, namely real GDP per capita, trade openness and FDI. The F-statistic is greater than 5% critical upper bound value at 5% significance level.

4.0 CONCLUSION

ARDL has superior statistical properties in small samples as it is relatively more efficient in small sample data sizes found mostly in studies on developing countries. Second, the long run and short run parameters of the model are estimated simultaneously. Third, it can be applied irrespective of whether the variables in the model are endogenous. Fourth, the econometric methodology is relieved of the burden of establishing the order of integration among the variables and of pre-testing of unit roots. Fifth, applying ARDL is helpful in data generating process through taking sufficient number of lags general-to-specific modeling framework Lastly, whereas all the other methods require that the variables in a time series regression are integrated of order one, I(1), only ARDL test could be used regardless of whether the underlying variables are I(0), I(1) or mixed or fractionally integrated. Nevertheless, the ARDL-bound testing approach requires that none of the variables in the model should be integrated in order two, and the variables should have long-run cointegration relationship.

5.0 SUMMARY

In this unit you learned ARDL cointegration equations as well as its computational procedures in Eviews. In addition, you learned a practical example on ARDL Bound cointegration test model and how to conduct diagnostic check for Serial Correlation in ARDL Cointegration Bounds test. The next unit which is our Unit 5 of Module 3 is the last in this course and it is a continuation of ARDL model in which ARDL level relation will be discussed.

6.0 TUTOR MARKED ASSIGNMENT

Use any data of your choice and conduct ARDL bound test.

7.0 REFERENCES/FURTHER READING

Narayan, P. K. (2005). The saving and investment nexus in China: evidence from cointegration tests. Applied Economics, 37, 1979 – 1990.

Pesaran, M., Shin, Y. & Smith, R.. (2001). Bound testing approaches to the analysis of level relationship. *J. Appl. Econ*. 16, 289–326.

Pesaran, H. and Shin, Y. (1999). An autoregressive distributed lag modeling approach to cointegration analysis. In: Strom, S. (Ed.), Econometrics and Economic Theory in 20th Century: The Ragnar–Frisch Centennial Symposium. Cambridge University Press: Cambridge.

UNIT5: ARDL LEVEL RELATION

CONTENTS

1.0 INTRODUCTION
2.0 OBJECTIVES
3.0 MAIN CONTENT
- 3.1 ARDL Level Relation
- 3.2 Autoregressive Distributed Lag Estimates
- 3.3 Diagnostic Test for ARDL
- 3.4 Stability Test for ARDL
- 3.5 ARDL Functional Form
- 3.6 Long-run Coefficients using the ARDL Approach
- 3.7 Error Correction Representation for the Selected ARDL Model

4.0 CONCLUSION
5.0 SUMMARY
7.0 TUTOR MARKED ASSIGNMENT

7.0REFERENCES/FURTHER READING

1.0 INTRODUCTION

In the unit you learnt ARDL cointegration equations as well as its computational procedures in Eviews. In addition, you learnt a practical example on ARDL Bound cointegration test model and how to conduct diagnostic check for Serial Correlation in ARDL Cointegration Bounds test. The prsent unit which is incidentally is our last in this course is a continuation of ARDL model in which ARDL level relation will be discussed.

2.0 OBJECTIVES

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

* Discuss ARDL level relation
*Conduct Autoregressive Distributed Lag Estimates
* Conduct diagnostic Test for ARDL
* Conduct stability Test for ARDL
* Design ARDL functional Form
* Estimate Long-run Coefficients using the ARDL Approach
* Examine Error Correction Representation for the Selected ARDL Model

3.0 MAIN CONTENTS

3.1 ARDL Level Relation

Example: In our model – **PRI = f(GDPC, TO, FDI)**

The ARDL model can be written as follows:

$$PRI_t = const + \sum_{i=1}^{p} \beta_1 PRI_{t-i} + \sum_{i=0}^{q} \beta_2 GDPC_{t-i} + \sum_{i=0}^{r} \beta_3 FDI_{t-i} + \sum_{i=0}^{s} \beta_{4,t} TO + \beta_{5,t} DUM + \varepsilon_t$$

where

PRI = private sector credit (financial development) (% of GDP)

const = constant

GDPC = real GDP per capita (RM)

FDI = foreign direct investment (% of GDP)

TO = trade openness (% of GDP)

p, q, r, s = optimum lag length

ε_t = residual

DUM = 1 for crisis and 0 for otherwise; which refer to the Malaysian oil crisis in years 1973, 1974, 1980 and 1981; commodities crisis in years 1985 and 1986; and East Asian financial crisis in years 1997 and 1998.

The below criteria can be used to select the optimum lag of the above ARDL modeling:

a) Akaike Information Criterion (AIC)
b) Schwarz Bayesian Criterion (SBC)
c) General to specific model

In our case, we use the general to specific approach to find the optimal lag length (where the model passed the misspecification tests, especially the serial correlation). Note: If we have two or more models selected by general to specific criterion, then the Schwarz Bayesian criterion or Akaike Information criterion can be used to select the best model (select the model with smallest value of AIC or SBC).

3.2Autoregressive Distributed Lag Estimates

1. Using the general to specific approach, the selected lag length of (p, q, r, s) is (1, 0, 0, 1). The long-run OLS output is as follows:

Autoregressive Distributed Lag Estimates
ARDL(1,0,0,1) Selected based on General to Specific Criterion

Dependent Variable: PRI
Method: Least Squares
Sample (adjusted): 1971 2009
Included observations: 39 after adjustments

Variable		Coefficient	Std. Error	t-Statistic	Prob.
C	C(1)	-0.007570	0.735818	-0.010288	0.9919
PRI(-1)	C(2)	0.942004	0.070699	13.32413	0.0000
GDPC	C(3)	0.087751	0.174634	0.502485	0.6188
FDI	C(4)	0.050689	0.032844	1.543297	0.1326
TO	C(5)	-0.732921	0.300590	-2.438271	0.0205
TO(-1)	C(6)	0.615725	0.304581	2.021548	0.0517
DUM	C(7)	0.100679	0.044096	2.283188	0.0292

R-squared	0.974912	Mean dependent var	4.325198
Adjusted R-squared	0.970207	S.D. dependent var	0.589545
S.E. of regression	0.101759	Akaike info criterion	-1.571279
Sum squared resid	0.331354	Schwarz criterion	-1.272691
Log likelihood	37.63994	Hannan-Quinn criter.	-1.464148
F-statistic	207.2476	Durbin-Watson stat	2.107965
Prob(F-statistic)	0.000000		

Dependent Variable: PRI
Method: Least Squares
Date: 07/12/11 Time: 16:15
Sample (adjusted): 1971 2009
Included observations: 39 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.523597	0.771148	0.678984	0.5020
PRI(-1)	0.962497	0.070987	13.55872	0.0000
GDPC	-0.026718	0.189089	-0.141300	0.8885
FDI	0.056278	0.033365	1.686742	0.1014
TO	-0.653958	0.300025	-2.179677	0.0368
TO(-1)	0.624007	0.311406	2.003842	0.0536
DUM	0.087902	0.041918	2.096980	0.0440

R-squared	0.974349	Mean dependent var	4.325198
Adjusted R-squared	0.969540	S.D. dependent var	0.589545

S.E. of regression	0.102892	Akaike info criterion	-1.549118
Sum squared resid	0.338779	Schwarz criterion	-1.250530
Log likelihood	37.20779	Hannan-Quinn criter.	-1.441987
F-statistic	202.5883	Durbin-Watson stat	2.156454
Prob(F-statistic)	0.000000		

3.3 Diagnostic Test for ARDL

1. Serial Correlation

The results of Breusch-Godfrey serial correlation LM test:

Lag 2

Breusch-Godfrey Serial Correlation LM Test: (Lag 2)

F-statistic	1.360282	Prob. F(2,30)	0.2720
Obs*R-squared	3.242669	Prob. Chi-Square(2)	0.1976

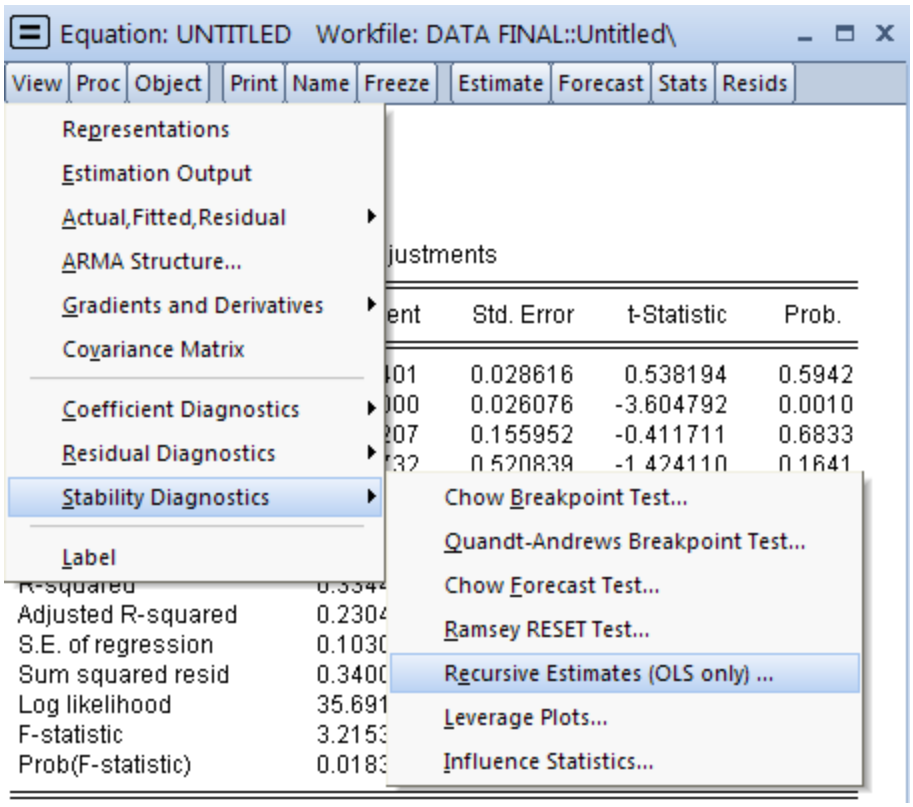
Lag 4

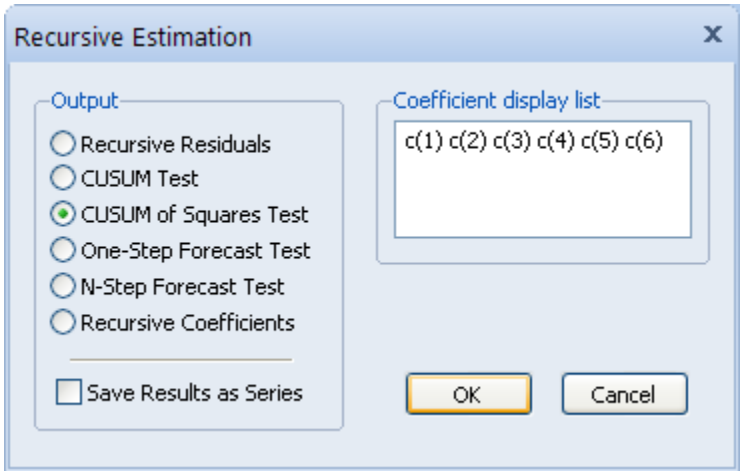
Breusch-Godfrey Serial Correlation LM Test: (Lag 4)

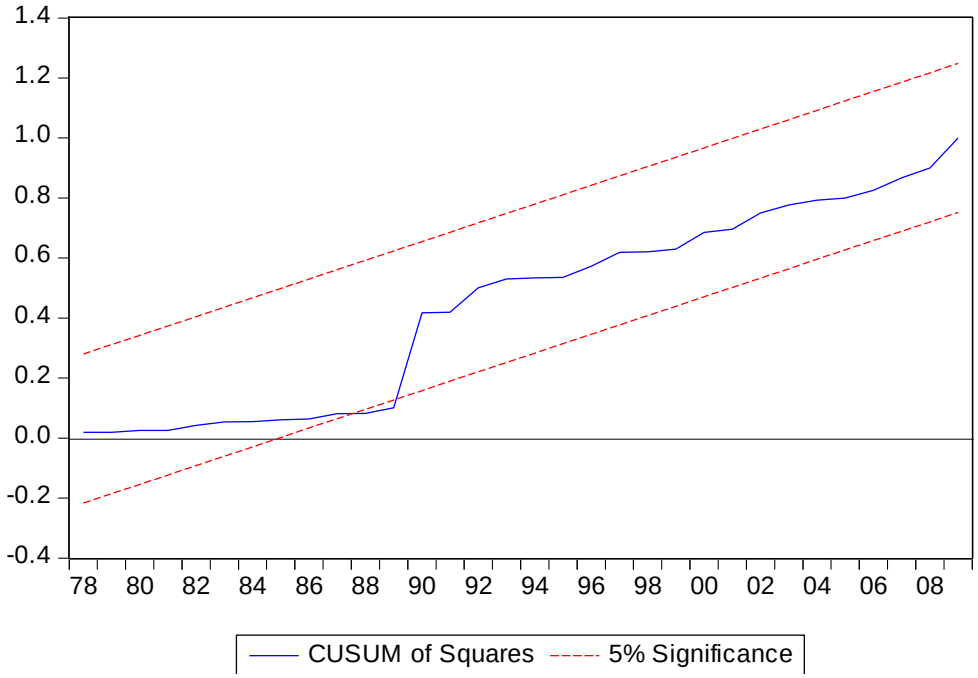
F-statistic	1.493188	Prob. F(4,28)	0.2310
Obs*R-squared	6.856592	Prob. Chi-Square(4)	0.1437

The above LM test results indicate that the residuals are homoskedasticity (no serial correlation) since the p-values are greater than 0.05.

3.4 Stability Test for ARDL


Equation: UNTITLED Workfile: DATA FINAL::Untitled\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
Representations
Estimation Output
Actual,Fitted,Residual
ARMA Structure...
Gradients and Derivatives
Covariance Matrix
Coefficient Diagnostics
Residual Diagnostics
Stability Diagnostics
Label
justments
ent Std. Error t-Statistic Prob.
401 0.028616 0.538194 0.5942
000 0.026076 -3.604792 0.0010
207 0.155952 -0.411711 0.6833
Adjusted R-squared 0.2304
S.E. of regression 0.1030
Sum squared resid 0.3400
Log likelihood 35.691
F-statistic 3.2153
Prob(F-statistic) 0.0183
Chow Breakpoint Test...
Quandt-Andrews Breakpoint Test...
Chow Forecast Test...
Ramsey RESET Test...
Recursive Estimates (OLS only) ...
Leverage Plots...
Influence Statistics...

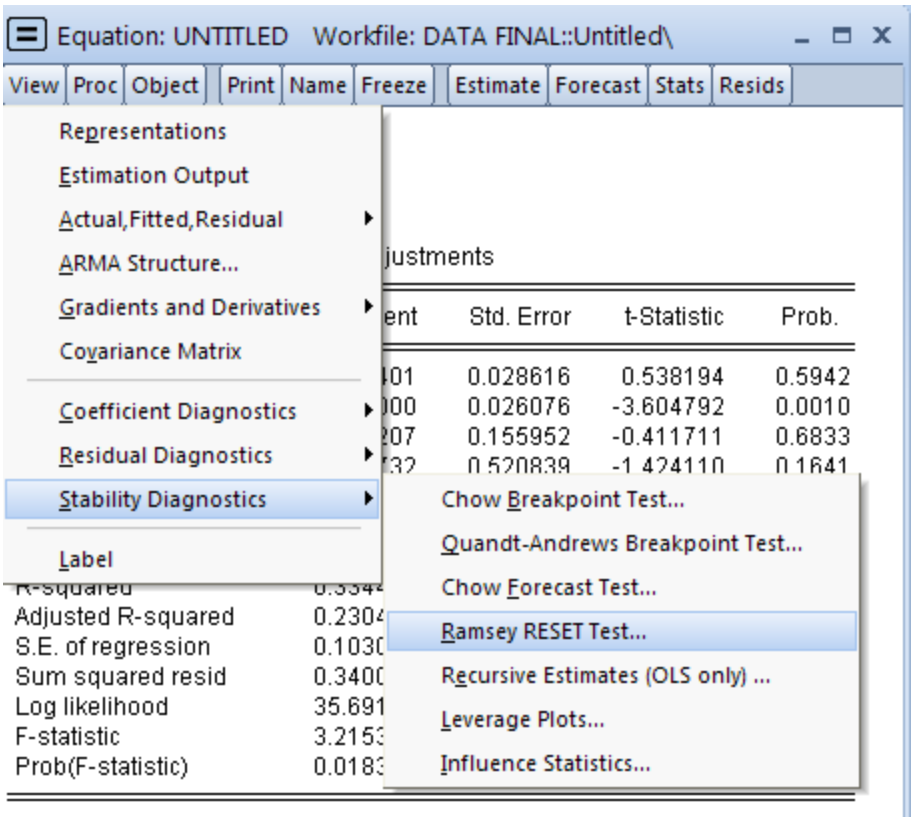

Recursive Estimation
Output
Recursive Residuals
CUSUM Test
CUSUM of Squares Test
One-Step Forecast Test
N-Step Forecast Test
Recursive Coefficients
Save Results as Series
Coefficient display list
c(1) c(2) c(3) c(4) c(5) c(6)
OK
Cancel



SELF-ASSESEMENT EXERCISE

Outline procedures for diagnostic tests in ARDL

3.5 ARDL Functional Form



Ramsey RESET Test
Equation: UNTITLED
Specification: LPRI CLPRI(-1)LGDPCLFDILTOLTO(-1)DUM
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	1.807314	31	0.0804
F-statistic	3.266384	(1, 31)	0.0804
Likelihood ratio	3.906927	1	0.0481

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.031586	1	0.031586
Restricted SSR	0.331354	32	0.010355
Unrestricted SSR	0.299768	31	0.009670
Unrestricted SSR	0.299768	31	0.009670

LR test summary:

	Value	df
Restricted LogL	37.63994	32
Unrestricted LogL	39.59340	31

Unrestricted Test Equation:
Dependent Variable: LPRI
Method: Least Squares
Sample: 1971 2009
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.868688	1.251417	-1.493258	0.1455
LPRI(-1)	1.850902	0.507520	3.646956	0.0010
LGDPC	0.049395	0.170090	0.290407	0.7734
LFDI	0.084038	0.036714	2.289014	0.0290
LTO	-1.395375	0.467686	-2.983570	0.0055
LTO(-1)	1.405543	0.526890	2.667620	0.0120
DUM	0.225680	0.081237	2.778035	0.0092
FITTED^2	-0.123432	0.068296	-1.807314	0.0804

R-squared	0.977303	Mean dependent var	4.325198
Adjusted R-squared	0.972178	S.D. dependent var	0.589545
S.E. of regression	0.098336	Akaike info criterion	-1.620175
Sum squared resid	0.299768	Schwarz criterion	-1.278931
Log likelihood	39.59340	Hannan-Quinn criter.	-1.497739
F-statistic	190.6887	Durbin-Watson stat	1.965471
Prob(F-statistic)	0.000000		

3.6 Long-run Coefficients using the ARDL Approach

4. After obtaining the ARDL (1,0,0,1) model, the next step is to find **the long run elasticities.**

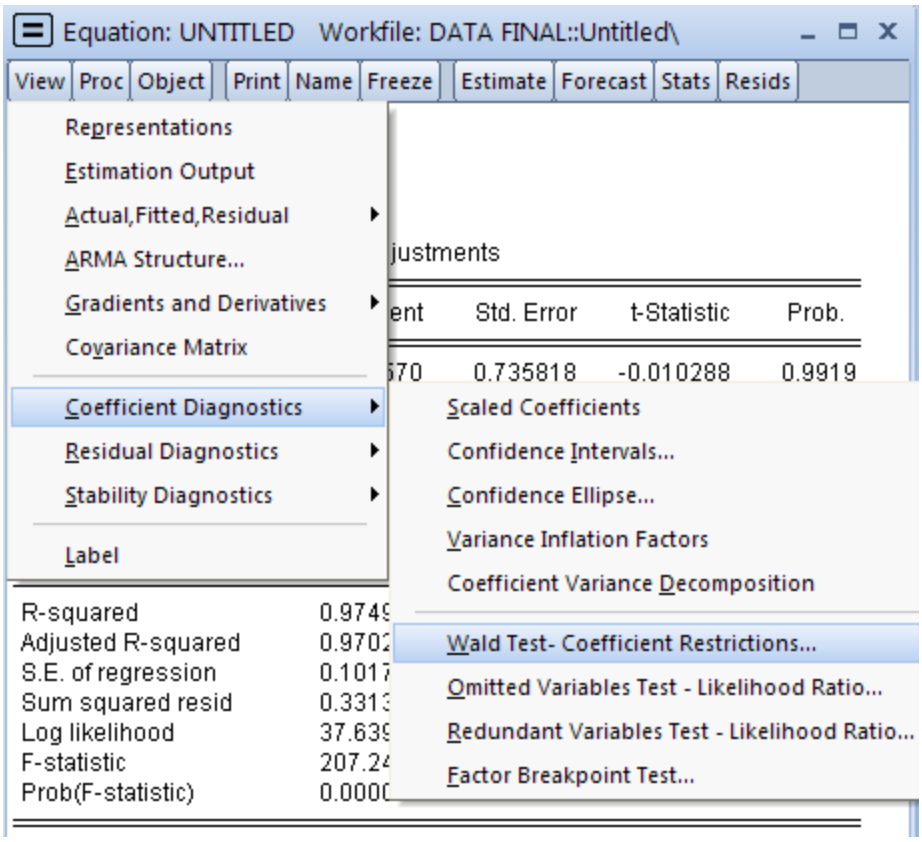
i. Elasticity of GDPC

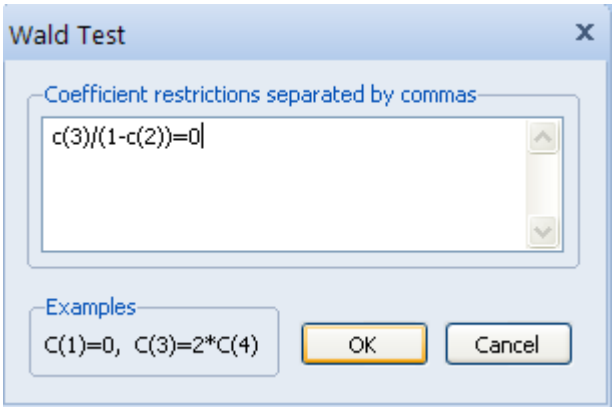
According to Pesaran et al. (2001), the long run elasticities can be obtained as follow:

$$\text{Elasticity GDPC} = \frac{\sum_{i=0}^{q} \beta_2}{1 - \sum_{i=0}^{p} \beta_1}$$

$$= \frac{\text{Sum of the independent coefficient(s) GDPC}}{1 - \text{sum of the dependent coefficient(s)}}$$

Go to "View" – "Coefficient Test" – "Wald Test"





Insert
c(3)/(1-c(2))=0
in the Wald Test window

Eviews Output:

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.461747	32	0.6474
F-statistic	0.213210	(1, 32)	0.6474
Chi-square	0.213210	1	0.6443

Null Hypothesis: C(3)/(1-C(2))=0
Null Hypothesis Summary:

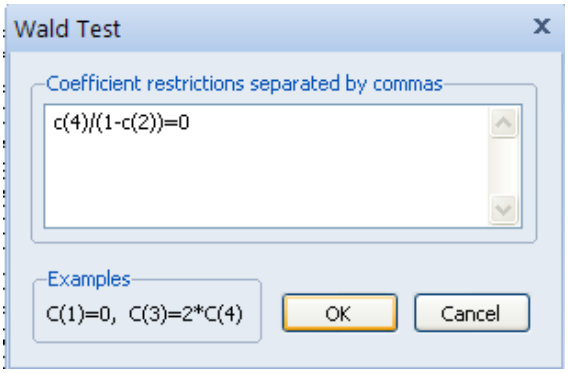
Normalized Restriction (= 0)	Value	Std. Err.
C(3) / (1 - C(2))	1.513043	3.276782

Delta method computed using analytic derivatives.

The elasticity is 1.5130 and the standard error is 3.2767.
The t-statistic can be computed as:
t-stat = Coefficient / Std. Err.
= 0.4617
The probability value (p-value) of 0.6474 (F-stat) is also served as a p-value for the computed t-statistic.

ii. Elasticity of FDI

$$\text{Elasticity FDI} = \frac{\sum_{i=0}^{q} \beta_3}{1 - \sum_{i=0}^{p} \beta_1}$$



Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.664552	32	0.5111
F-statistic	0.441630	(1, 32)	0.5111
Chi-square	0.441630	1	0.5063

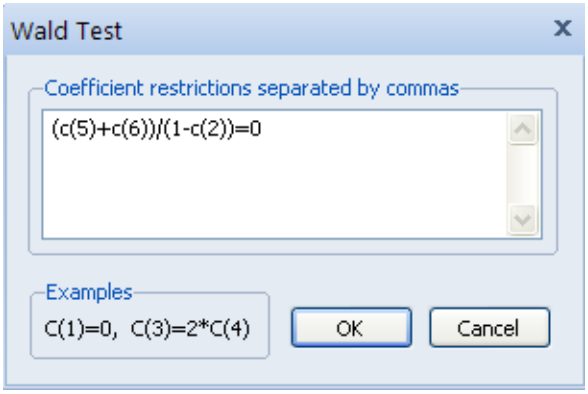
Null Hypothesis: C(4)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(4) / (1 - C(2))	0.873994	1.315162

Delta method computed using analytic derivatives.

iii. Elasticity of TO

$$\text{Elasticity TO} = \frac{\sum_{i=0}^{q} \beta_4}{1 - \sum_{i=0}^{p} \beta_1}$$



Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	-0.381448	32	0.7054
F-statistic	0.145503	(1, 32)	0.7054
Chi-square	0.145503	1	0.7029

Null Hypothesis: (C(5)+C(6))/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
(C(5) + C(6)) / (1 - C(2))	-2.020752	5.297581

Delta method computed using analytic derivatives.

iv. Long-run coefficient of Constant Term

Wald Test:
Equation: Untitled

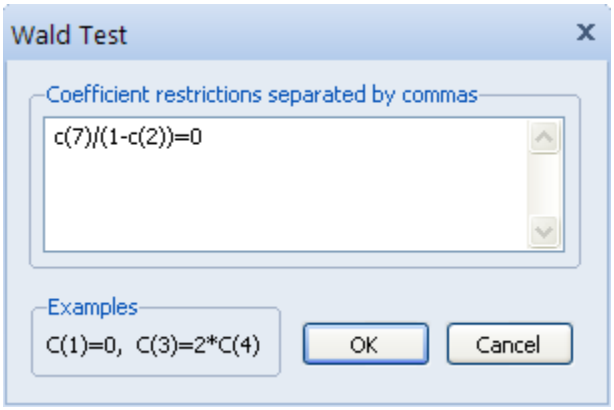
Test Statistic	Value	Df	Probability
t-statistic	-0.010349	32	0.9918
F-statistic	0.000107	(1, 32)	0.9918
Chi-square	0.000107	1	0.9917

Null Hypothesis: C(1)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1) / (1 - C(2))	-0.130530	12.61259

Delta method computed using analytic derivatives.

v. Coefficient of DUM



Wald Test:
Equation: Untitled

Test Statistic	Value	Df	Probability
t-statistic	0.802844	32	0.4280

F-statistic	0.644558	(1, 32)	0.4280
Chi-square	0.644558	1	0.4221

Null Hypothesis: C(7)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(7) / (1 - C(2))	1.735946	2.162246

Delta method computed using analytic derivatives.

The coefficients of all variables using the ARDL approach are:

GDPC = 1.5130
FDI = 0.8739
TO = -2.0207
Constant = -0.1305
DUM = 1.7359

Therefore, the long-run relation model can be written as follows:

PRI_t = -0.1305 + 1.5130 $GDPC_t$ + 0.8739 FDI_t – 2.0207 TO_t + 1.7359 DUM

t-stat (-0.0103) (0.4617) (0.6645) (-0.3814) (0.8028)

Summary

ARDL (1,0,0,1) Model:

$$PRI_t = c + \alpha_1 PRI_{t-1} + \beta_1 GDPC_t + \beta_2 FDI_t + \beta_3 TO_t + \beta_4 TO_{t-1} + \beta_5 DUM_t + \varepsilon_t$$

Based on the above ARDL model, we can estimate the long run model as following:

where

$$\varphi_{GDPC} = \frac{\beta_1}{1 - \alpha_1}$$

$$\varphi_{FDI} = \frac{\beta_2}{1 - \alpha_1}$$

$$\varphi_{TO} = \frac{\beta_3 + \beta_4}{1 - \alpha_1}$$

$$\varphi_{DUM} = \frac{\beta_5}{1 - \alpha_1}$$

$$\varphi_{Constant} = \frac{c}{1 - \alpha_1}$$

3.7Error Correction Representation for the Selected ARDL Model

After obtaining the long-run relation, the next step is to estimate the short-run Error- correction Model (ECM).

Compute the value of Error-correction Term (ECT), which represents the residuals from long-run cointegration model.

Recap the ARDL (1,0,01) Model

$$PRI_t = c + \alpha_1 PRI_{t-1} + \beta_1 GDPC_t + \beta_2 FDI_t + \beta_3 TO_t + \beta_4 TO_{t-1} + \beta_5 DUM_t + \varepsilon_t$$

The short run dynamic model can be transformed using the above ARDL model:
where

$PRI_t = \Delta PRI_t + PRI_{t-1}$; $\Delta PRI_{t-i} = PRI_{t-1} - \sum_{i=0}^{p-1} \Delta PRI_{t-i}$

$GDPC_t = \Delta GDPC_t + GDPC_{t-1}$; $\Delta GDPC_{t-j} = GDPC_{t-1} - \sum_{j=0}^{q-1} \Delta GDPC_{t-j}$

$FDI_t = \Delta FDI_t + FDI_{t-1}$; :

$TO_t = \Delta TO_t + TO_{t-1}$; :

$$PRI_t = c + \alpha_1 PRI_{t-1} + \beta_1 GDPC_t + \beta_2 FDI_t + \beta_3 TO_t + \beta_4 TO_{t-1} + \varepsilon_t$$

Transform:

$$\Delta PRI_t + PRI_{t-1} = c + \alpha_1 PRI_{t-1} + \beta_1(\Delta GDPC_t + GDPC_{t-1}) + \beta_2(\Delta FDI_t + FDI_{t-1}) + \\ + \beta_3(\Delta TO_t + TO_{t-1}) + \beta_4 TO_{t-1} + \varepsilon_t$$

$$\Delta PRI_t = c - PRI_{t-1} + \alpha_1 PRI_{t-1} + \beta_1 \Delta GDPC_t + \beta_1 GDPC_{t-1} + \beta_2 \Delta FDI_t + \beta_2 FDI_{t-1} + \\ \beta_3 \Delta TO_t + \beta_3 TO_{t-1} + \beta_4 TO_{t-1} + \varepsilon_t$$

$$\Delta PRI_t = c - (1-\alpha_1) PRI_{t-1} + \beta_1 \Delta GDPC_t + \beta_1 GDPC_{t-1} + \beta_2 \Delta FDI_t + \beta_2 FDI_{t-1} + \\ \beta_3 \Delta TO_t + (\beta_3 + \beta_4) TO_{t-1} + \varepsilon_t$$

$$\Delta PRI_t = c - (1-\alpha_1) PRI_{t-1} + \beta_1 GDPC_{t-1} + \beta_2 FDI_{t-1} + (\beta_3 + \beta_4) TO_{t-1} + \\ \beta_1 \Delta GDPC_t + \beta_2 \Delta FDI_t + \beta_3 \Delta TO_t + \varepsilon_t$$

$$\Delta PRI_t = c - (1-\alpha_1) \overbrace{\left(PRI_{t-1} - \frac{\beta_1}{1-\alpha_1} GDP_{t-1} - \frac{\beta_2}{1-\alpha_1} FDI_{t-1} - \frac{\beta_3 + \beta_4}{1-\alpha_1} TO_{t-1} \right)}^{\text{ECT}} + \\ \beta_1 \Delta GDPC_t + \beta_2 \Delta FDI_t + \beta_3 \Delta TO_t + \varepsilon_t$$

where

$$ECT_{t-1} = PRI_{t-1} - \frac{\beta_1}{1-\alpha_1} GDP_{t-1} - \frac{\beta_2}{1-\alpha_1} FDI_{t-1} - \frac{\beta_3 + \beta_4}{1-\alpha_1} TO_{t-1}$$

The ECT can be obtained as follows:

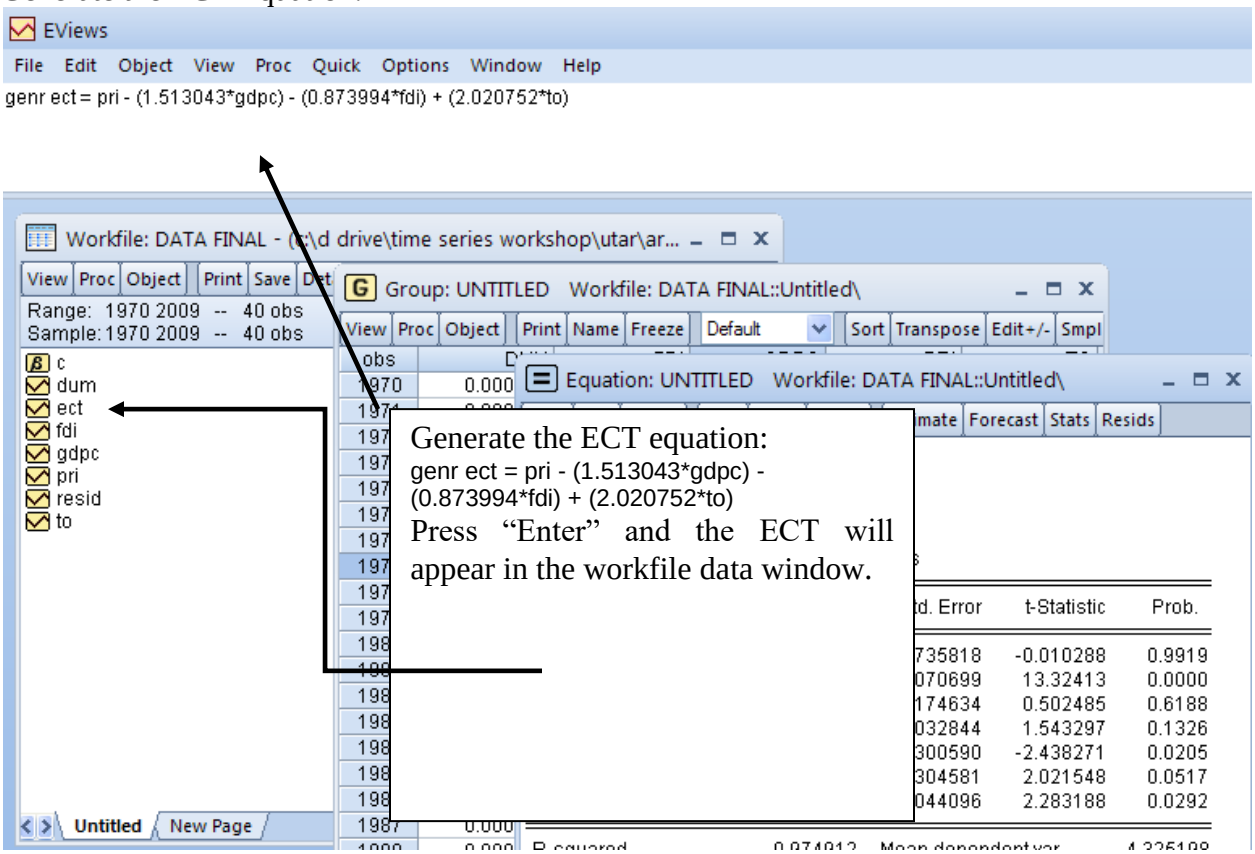
The long-run Equation is

PRI = -0.1305 + 1.5130 GDPC + 0.8739 FDI – 2.0207 TO + 1.7359 DUM

Hence, the ECT equation is:

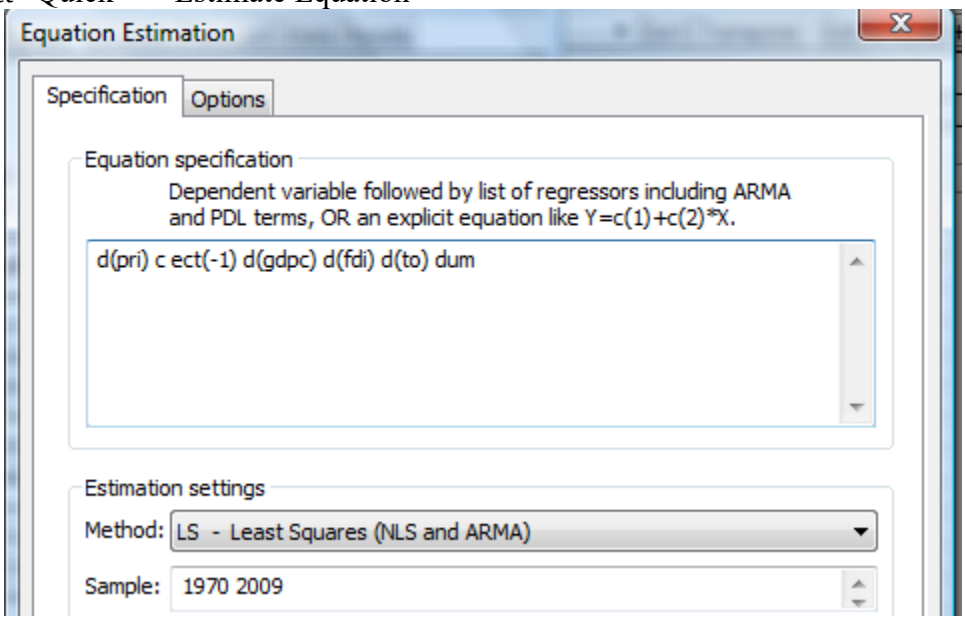
ECT = PRI – 1.5130 GDPC – 0.8739 FDI + 2.0207 TO

b. Generate the ECT Equation.



c. After generating the ECT series, estimate the **short-run dynamic equation**:

Select "Quick" – "Estimate Equation"



The short-run dynamic result is as follows:

Dependent Variable: D(PRI)
Method: Least Squares
Date: 07/04/11 Time: 22:28
Sample (adjusted): 1971 2009
Included observations: 39 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.009450	0.029983	-0.315184	0.7546
ECT(-1)	-0.073751	0.023777	-3.101804	0.0039
D(GDPC)	-0.194315	0.563271	-0.344977	0.7323
D(FDI)	0.029790	0.035645	0.835765	0.4093
D(TO)	-0.604492	0.263134	-2.297279	0.0281
DUM	0.085466	0.047206	1.810480	0.0793

R-squared	0.388518	Mean dependent var	0.046311
Adjusted R-squared	0.295869	S.D. dependent var	0.116254
S.E. of regression	0.097552	Akaike info criterion	-1.676229
Sum squared resid	0.314039	Schwarz criterion	-1.420296
Log likelihood	38.68646	Hannan-Quinn criter.	-1.584402
F-statistic	4.193449	Durbin-Watson stat	2.143115
Prob(F-statistic)	0.004641		

Dependent Variable: PRI
Method: Least Squares
Date: 07/06/11 Time: 17:43
Sample (adjusted): 1971 2009
Included observations: 39 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.523597	0.771148	0.678984	0.5020
PRI(-1)	0.962497	0.070987	13.55872	0.0000
GDPC	-0.026718	0.189089	-0.141300	0.8885
FDI	0.056278	0.033365	1.686742	0.1014
TO	-0.653958	0.300025	-2.179677	0.0368
TO(-1)	0.624007	0.311406	2.003842	0.0536
DUM	0.087902	0.041918	2.096980	0.0440

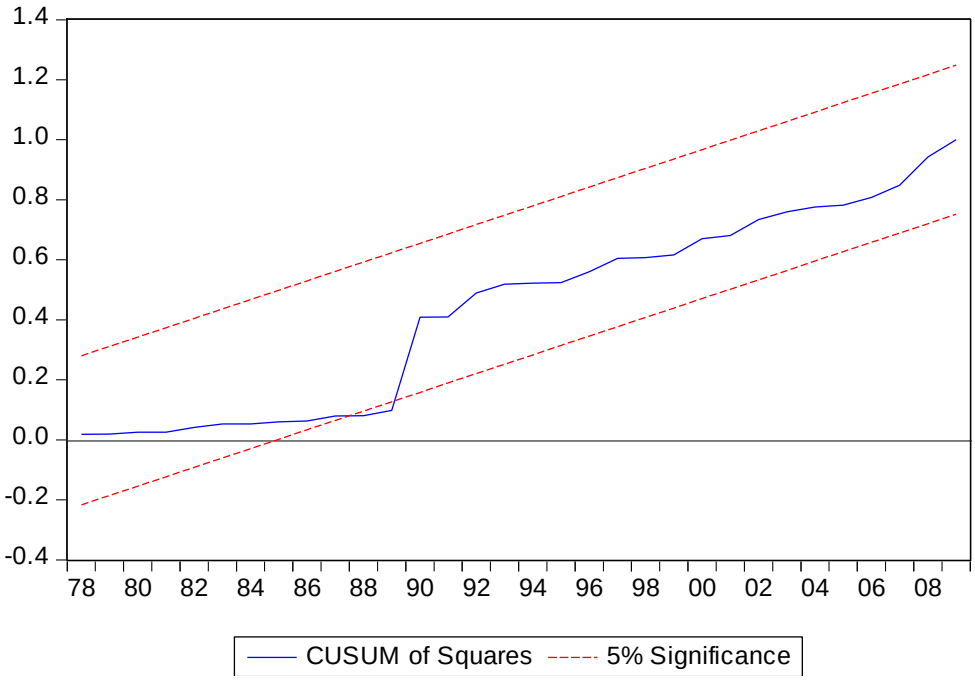
R-squared	0.974349	Mean dependent var	4.325198
Adjusted R-squared	0.969540	S.D. dependent var	0.589545
S.E. of regression	0.102892	Akaike info criterion	-1.549118
Sum squared resid	0.338779	Schwarz criterion	-1.250530
Log likelihood	37.20779	Hannan-Quinn criter.	-1.441987
F-statistic	202.5883	Durbin-Watson stat	2.156454
Prob(F-statistic)	0.000000		

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.374716	Prob. F(2,30)	0.2684
Obs*R-squared	3.274189	Prob. Chi-Square(2)	0.1945

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.358467	Prob. F(4,28)	0.2734
Obs*R-squared	6.338510	Prob. Chi-Square(4)	0.1753



Ramsey RESET Test
Equation: UNTITLED
Specification: PRI C PRI(-1) GDPC FDI TO TO(-1) DUM
Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	1.803185	31	0.0811
F-statistic	3.251477	(1, 31)	0.0811
Likelihood ratio	3.889956	1	0.0486

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.032160	1	0.032160
Restricted SSR	0.338779	32	0.010587
Unrestricted SSR	0.306619	31	0.009891

Unrestricted SSR	0.306619	31	0.009891

LR test summary:

	Value	df
Restricted LogL	37.20779	32
Unrestricted LogL	39.15277	31

Unrestricted Test Equation:
Dependent Variable: PRI
Method: Least Squares
Date: 07/06/11 Time: 17:59
Sample: 1971 2009
Included observations: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.692326	1.005130	-0.688793	0.4961
PRI(-1)	1.906796	0.528160	3.610265	0.0011
GDPC	-0.208620	0.208760	-0.999330	0.3254
FDI	0.097043	0.039384	2.463997	0.0195
TO	-1.223581	0.428824	-2.853342	0.0076
TO(-1)	1.430943	0.539315	2.653262	0.0125
DUM	0.200491	0.074433	2.693574	0.0113
FITTED^2	-0.124719	0.069166	-1.803185	0.0811

R-squared	0.976784	Mean dependent var	4.325198
Adjusted R-squared	0.971542	S.D. dependent var	0.589545
S.E. of regression	0.099453	Akaike info criterion	-1.597578
Sum squared resid	0.306619	Schwarz criterion	-1.256335
Log likelihood	39.15277	Hannan-Quinn criter.	-1.475143
F-statistic	186.3292	Durbin-Watson stat	2.017249
Prob(F-statistic)	0.000000		

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	-0.131406	32	0.8963
F-statistic	0.017267	(1, 32)	0.8963
Chi-square	0.017267	1	0.8955

Null Hypothesis: C(3)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3) / (1 - C(2))	-0.712435	5.421647

Delta method computed using analytic derivatives.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.473977	32	0.6387
F-statistic	0.224654	(1, 32)	0.6387
Chi-square	0.224654	1	0.6355

Null Hypothesis: C(4)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(4) / (1 - C(2))	1.500634	3.166049

Delta method computed using analytic derivatives.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	-0.113928	32	0.9100
F-statistic	0.012980	(1, 32)	0.9100
Chi-square	0.012980	1	0.9093

Null Hypothesis: (C(5)+C(6))/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
(C(5) + C(6)) / (1 - C(2))	-0.798630	7.009945

Delta method computed using analytic derivatives.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.347133	32	0.7308
F-statistic	0.120502	(1, 32)	0.7308
Chi-square	0.120502	1	0.7285

Null Hypothesis: C(1)/(1-C(2))=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.

C(1) / (1 - C(2))	13.96156	40.21961

Delta method computed using analytic derivatives.

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.511168	32	0.6127
F-statistic	0.261293	(1, 32)	0.6127
Chi-square	0.261293	1	0.6092

Null Hypothesis: C(7)/(1-C(2))=0
Null Hypothesis Summary:

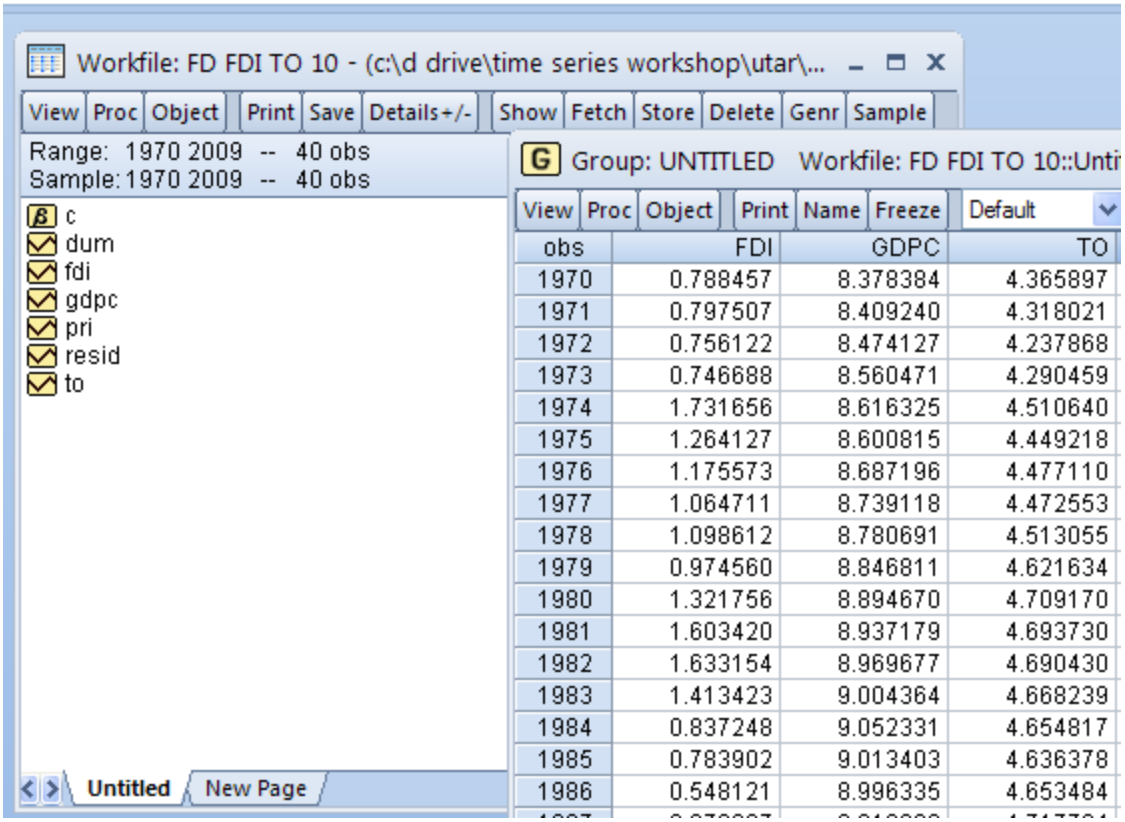
Normalized Restriction (= 0)	Value	Std. Err.
C(7) / (1 - C(2))	2.343869	4.585322

Delta method computed using analytic derivatives.

EViews

File Edit Object View Proc Quick Options Window Help

```
genr ect=pri+(0.712435*gdpc)-(1.500634*fdi)+(0.798630*to)
```



obs	FDI	GDPC	TO
1970	0.788457	8.378384	4.365897
1971	0.797507	8.409240	4.318021
1972	0.756122	8.474127	4.237868
1973	0.746688	8.560471	4.290459
1974	1.731656	8.616325	4.510640
1975	1.264127	8.600815	4.449218
1976	1.175573	8.687196	4.477110
1977	1.064711	8.739118	4.472553
1978	1.098612	8.780691	4.513055
1979	0.974560	8.846811	4.621634
1980	1.321756	8.894670	4.709170
1981	1.603420	8.937179	4.693730
1982	1.633154	8.969677	4.690430
1983	1.413423	9.004364	4.668239
1984	0.837248	9.052331	4.654817
1985	0.783902	9.013403	4.636378
1986	0.548121	8.996335	4.653484

EViews

File Edit Object View Proc Quick Options Window Help

```
genr ect=pri+(0.712435*gdpc)-(1.500634*fdi)+(0.798630*to)
```

Workfile: FD FDI TO 10 - (c:\d drive\time series workshop\utar\...

View Proc Object Print Save Details+/- Show Fetch Store Delete Genr Sample

Range: 1970 2009 -- 40 obs
Sample: 1970 2009 -- 40 obs

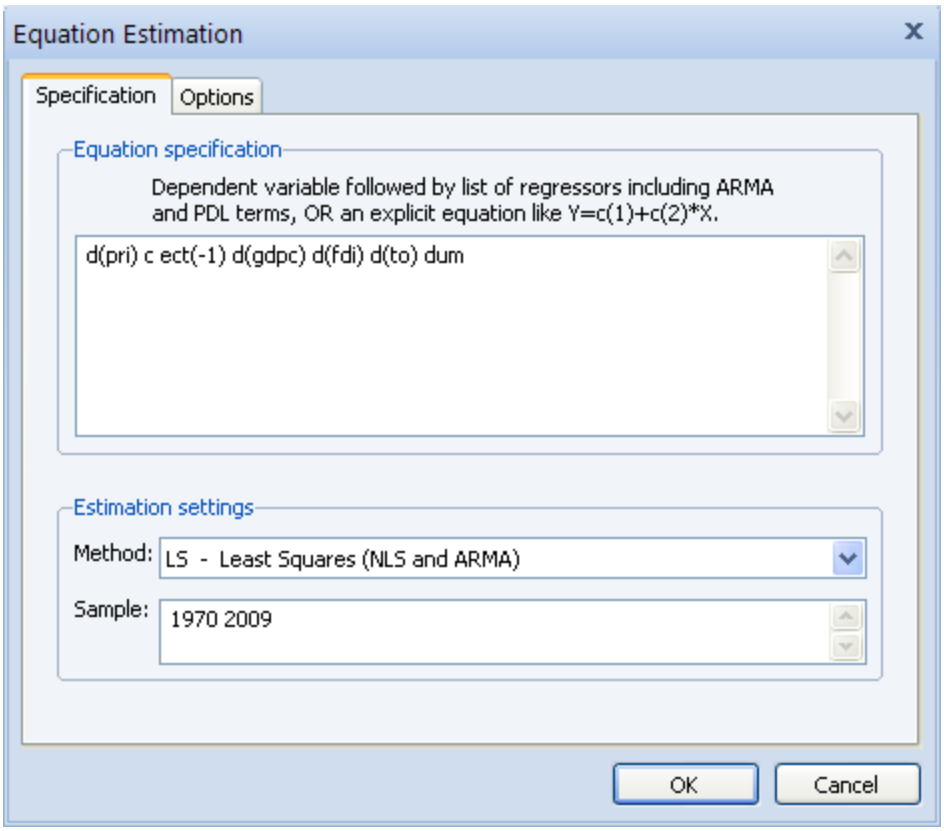
- c
- dum
- ect
- fdi
- gdpc
- pri
- resid
- to

Untitled New Page

Group: UNTITLED Workfile: FD FD

View Proc Object Print Name Freeze

obs	FDI	GDPC
1970	0.788457	8.378384
1971	0.797507	8.409240
1972	0.756122	8.474127
1973	0.746688	8.560471
1974	1.731656	8.616325
1975	1.264127	8.600815
1976	1.175573	8.687196
1977	1.064711	8.739118
1978	1.098612	8.780691
1979	0.974560	8.846811
1980	1.321756	8.894670
1981	1.603420	8.937179
1982	1.633154	8.969677
1983	1.413423	9.004364
1984	0.837248	9.052331
1985	0.783902	9.013403
1986	0.548121	8.996335
1987	0.270027	9.019629


Equation Estimation
Specification
Options
Equation specification
Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.
d(pri) c ect(-1) d(gdpc) d(fdi) d(to) dum
Estimation settings
Method:
LS - Least Squares (NLS and ARMA)
Sample:
1970 2009
OK
Cancel

Dependent Variable: D(PRI)
Method: Least Squares
Date: 07/06/11 Time: 17:56
Sample (adjusted): 1971 2009
Included observations: 39 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.635975	0.188492	3.374022	0.0019
ECT(-1)	-0.045285	0.013598	-3.330336	0.0021
D(GDPC)	-0.313815	0.568640	-0.551869	0.5848
D(FDI)	0.036565	0.034748	1.052279	0.3003
D(TO)	-0.551890	0.259209	-2.129134	0.0408
DUM	0.073582	0.041541	1.771329	0.0857

R-squared	0.367476	Mean dependent var	0.046311
Adjusted R-squared	0.271639	S.D. dependent var	0.116254
S.E. of regression	0.099216	Akaike info criterion	-1.642396
Sum squared resid	0.324846	Schwarz criterion	-1.386464
Log likelihood	38.02672	Hannan-Quinn criter.	-1.550570
F-statistic	3.834384	Durbin-Watson stat	2.163302
Prob(F-statistic)	0.007550		

4.0 CONCLUSION

Since ARDL models are least squares regressions using lags of the dependent and independent variables as regressors, they can be estimated in EViews using an equation object with the Least Squares estimation method. However, EViews also offers a specialized estimator for handling ARDL models. This estimator offers built-in lag-length selection methods, as well as post-estimation views.

8.0 SUMMARY

This Unit 5 of our Module 3 is the last unit in this course. In this unit you learned ARDL level relation estimates as well as how to conduct diagnostic test for ARDL. You also learned how to conduct stability test for ARDL, how to design ARDL functional Form, how to estimate Long-run Coefficients using the ARDL approach and error correction representation for ARDL Model.

You have to praise yourself for a job well done. Go over the course again to master all the key points and computational procedures. Make sure you buy Eviews 9 or 10 and install on your computer.

6.0 TUTOR MARKED ASSIGNMENT

Conduct ARDL model diagnostics tests using Eviews 9 or 10.

7.0 REFERENCES/FURTHER READING

Narayan, P. K. (2005) The saving and investment nexus in China: evidence from cointegration tests. Applied Economics, 37, 1979 – 1990.