INTRODUTION TO ECONOMETRIC RESEARCH METHOD

SALIHU H. MUHAMMED-GANI

INTRODUTION TO ECONOMETRIC RESEARCH METHOD

Copyright (a) 2011 - Salihu H. Muhammed-Gani

ISBN 978 - 978 – 49389-7-6

First Published 2011

ALL RIGHTS RESERVED

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, or by any means, electronic. Mechanical, electrostatic, magnetic tape, photocopying, recording or otherwise, without the prior written consent of the author who is the copyright owner.

Printed in Nigeria by
Azaben Press,
David Mark By-Pass, High Level Makurdi
08069313617, 08026731891

DEDICATION

To My Students

ACKNOWLDGEMENT

The effort of my teacher is the most important element in the development and the writing of this book, most especially Professor Joseph Odama who thought me the art of econometric method in such a crafty manner that have ever been developed in my world.

The family which forms the root system in personality development social orientation is also important in this project because they developed the capacity and the conviction in the development and the writing of this text.

As the saying goes" tell me your friends so that I will tell you the kind of persons you are" the effort of my friends is the pivot on which the wheel of this movement revolve. Notable amongs them are Ahmad Jibril Khalid , Jubrin Ahmed Ibrahim, Osoki Muhammed, Ismaila Ewa(Ajiga), Usman Aliyu Ardo, Joel Odonye, Sagir El-Nawami, Abdulahi Omama, yakubu Ramat Muhammed, Abdulahi Usman Umbamas, Abdullkavim chindo, Ibrahim Dauda, Bala Zakari Yusuf Tanko Kano and indeed the entire people of Keana and Alago race.

To my colleagues in Economics Department Nasarawa State University Keffi who are also my teachers this text would have been without their challenge, and moral support. Notable among them is the Head of Department Professor Mrs. Ebele Amali, Professor N. Chuta (the daddy of the department)Dr. Rabo Dakare, Dr. Ibrahim G. Umar, Dr. Duru(Sabatical) Dr. Akawu Francis, Dr. Ibi, Dr. Anzaku, Mr. Angbas Jonathan, Mr. Ahmed hery Eggon, Hon. Ugeh, Mr. Pius and of course my big ogah the one and only Yusuf Dauda Bulus.

The effort of Abubakar Balarabe Adamu, Abdullrazak Itari and Felicia Com is important since they are the people that type the first draft of the text.

The work of My printer Mr. Aberagi Peter of Azaben Press makurdi is overwhelming. Azaben press has made this test a reality. To them I owed great respect.

FORWARD

Having gone through this book which is an effort made by the author whom I know as a colleague in the department of Economics for some years, I can not but appreciate his zeal in the book to add to knowledge. The book has particularly captured a myriad of problems faced by students of economics, economist and researchers alike.

Introduction to Econometric Research has addressed problem areas like research in econometric, choice of models in research, specification in mathematical and econometric model, data transformation, and estimation, hypothesis testing, inferences and predictions and also violation of assumptions in research study among other problems.

The author using his wealth of experience he has garnered over the years in teaching the course Econometrics has simplified some of this areas in this book using simple language and a some simple worked examples to make students and other users of this book acquire the desired intentions of the book.

Aside from this, the author has also dwelled in this fifteen chaptered book models and new

techniques that are compatible with modern research undertakings.

While commending the efforts of the author for this work, I encourage all who want to acquire knowledge especially in the area of econometrics to have and use this book.

Dr. Ibrahim Gerard Umar
Economic Department
Nasarawa State university,
Keffi

PREFACE

The development of any economy is promoted by researches, especially in the field of economics. This is made possible via econometric method which is developed to investigate the nature and magnitude of economic relationship in empirical manner and to verify economic theory.

This text has been written to promote the understanding of econometric method particularly among students, economists and researchers; hence, it is presented in s simplified but detailed manner consistent with the approved syllabus. The book is divided into fifteen chapters.

The chapters were arranged in such a manner that will enable the learner to develop the knowledge of method in a gradual and adaptive fashion. Most of the areas of difficulties to students has been simplified via worked examples. For readers that are deficient in statistical method, a review of statistical concept and method is made handy to facilitate the understanding of econometric method.

It is my hope that this text will go along way in improving student understanding of econometric method. And as a student myself I stand to be molded.

Salihu H. Muhammed Gani
June 2011

CONTENT

Dedication - - - - - - iii
Acknowledgements - - - - iv
Forward - - - - - - vi
Preface - - - - - - viii

CHAPTER ONE

1.1 Introduction - - - - 1
1.2 What is Econometrics - - - 1
1.3 Scope of Econometrics - - - 3
1.4 Uses of Econometric research - - 4

CHAPTER TWO

2.1 Methodology of Econometric Research - - - - - 7
2.2 Definition of the Research Problem - 7
2.3 Specification of the Mathematical form of the model - - - 8
2.4 Specification of the Econometric form of the model - - - - 9
2.5 Desirable properties of a good model - 10
2.6 Data gathering and confronting the model with data - - - - 11
2.7 Estimation of parameters in the Model - - - - - 11
2.8 Testing of hypothesis or verification of the model - - - - 12

2.9 Forecasting - - - - 15

2.10 The Use of Estimation Model for Control or Policy Purpose - - - 16
2.11 The language of econometric research - - - - - 17
2.12 Causation - - - - - 18

CHAPTER THREE
3.1 Correlation - - - - - 20
3.2 Rank correlation - - - - 27
3.3 Limitation of correlation analysis - 29

CHAAPTER FOUR
4.1 Bivariate Regression analysis (Simple regression) - - - 30
4.2 Conditional Model - - - - 30
4.3 OLS Method and the underlying assumptions - - - - - 33
4.4 The Normal equations - - - 36
4.5 Total, Explain and the Residual sums of square - - - - - - 46
4.6 The mean and the variance of OLS estimates - - - - 52
4.7 Test of significance - - - - 59
4.8 Gauss Markov theorem - - - 61

CHAPTER FIVE

5.1 Time as a regressor, (the linear trend model) - - - 69

5.2 Constant growth model (the log-line model) - - - 70

5.3 The lin-log model - - - - 81

5.4 The log-log model/the constant elasticity model - - - - 82

CHAPTER SIX

6.1 Multiple Regression Analysis - - 87

6.2 The multiple coefficient of Determination - - - - 94

6.3 Hypothesis Testing in Multiple Regression - - - 100

6.4 Hypothesis Testing about Individual regression coefficient - - 100

6.5 Test for parameter restriction - - 102

6.6 Test of overall significance of the sample regression - - - - 103

6.7 Test of equality of two regression coefficients - - - - - 104

6.8 Chow Test - - - - - 105

CHAPTER SEVEN

Dynamic Econometric Model

7.1 Distributed Lag Model - - - 107

7.2 Sort run multiplier Interim/Intermediate Multiplier - - - - - 114
7.3 Long-Run Multiplier - - - 115
7.4 The Lag-Distribution - - - 115
7.5 The Mean-Lag - - - - 116
7.6 Estimation of Distributed Lag Model - 117
7.7 The Koyck Distributed Lag Model - 118
7.8 Nerlove partial Adjustment Model - 120
7.9 Auto-Regressive Model - - - 122

CHAPTER EIGHT
8.1 Univariate Analysis (Time series Econometrics) - - 126
8.2 Stationarity/Non-stationary Process - - - - - 126
8.3 Unit root stockastic process - - 129
8.4 Spurious Regression - - - 130
8.5 Integrated Stochastic Process - - 132
8.6 Test of Stationarity – The unite root test - - - - - 133
8.7 Dickey-Fuller (DF) test - - - 134
8.8 Augmented Dickey-Fuller (ADF) Test - 138
8.9 The Philiph-Perron (PP) test - - 139
8.10 Cointegration - - - - 141
8.11 Test for Cointegration - - - 142
8.12 Engel-Granger (EG) test - - - 142
8.13 Augmented (EG) test - - - 144
8.14 Durbin Watson for cointergration - 145

8.15 Error Correction Mechanicsm/Model - 147
8.16 Worked Examples - - - - 148
8.17 Causality - - - - - 155
8.18 Granger Causality test - - - 155
8.19 Causality and Exogenerity - - - 157

CHAPTER NINE
9.1 Dummy variable regression model - 159

CHAPTER TEN
10.1 Violation of the classical assumptions - 166
10.2 Non randomness of the error term Ut - 171
10.3 Non normality of Ut - - - 171
10.4 Non Zero mean - - - - 172
10.5 Autocorrelation - - - - 174
10.6 Test for Autocorrelation - - - 179
10.7 Heteroscedasticity - - - 183
10.8 Multicollinearity - - - - 184

CHAPTER ELEVEN
11.1 Statistical Bias - - - - 193

CHAPTER TWELVE
ANOVA - - - - - - 198

CHAPTER THIRTEEN
13.1 The (X2) Chi-Square - - - 206
13.2 X2 test of goodness of fit - - 206

13.3 X2 test of independence of event - 212

CHAPTER FOURTEEN

14.1 Test of Hypothesis - - - - 216

14.2 Step involve in the lest of hypothesis - 218

CHAPTER FIFTEEN

Review of statistical concepts - - - 228

Bibliography - - - - - - 306

Indexces - - - - - - 313

CHAPTER ONE

NATURE AND SCOPE OF ECONOMETRICS

1.1 Introduction

Econometrics is born out of the desire to explain the behavior of economic variables in a more simplified and empirical manner that can be acceptable by economists and policy makers alike. This is made possible by bringing together economic theory; mathematical economic method and the statistical method in order to adequately explain the ever changing behavior of economic variable. Thus, a good understanding of economic theory, and economic environment; relevant mathematical concepts and methods as well as a great deal of statistical methods are particularly important and are blended to undertake a comprehensive study in econometrics. Hence, econometrics is a research instrument that presupposes a solid understanding and the application of mathematical and statistical methods, coupled with economic theory to investigate the nature and magnitude of economic relationship.

1.2 What is Econometrics?

Econometrics is concern with the investigation of the nature and magnitude of economic relationship. It provides numerical values for the parameters of economic relationship and verifies economic theory. Hence, econometrics can be describe as the science that deals with quantification of economic relationship, testing of hypothesis, making inferences

about these relationship and forecasting the future of economic phenomenon.

This suggests that the scope of econometrics is much more than economic measurement. In the opening chapter of his book, Gujarati D. N. (2004), relates the following description of econometrics:

(a) Econometrics connote the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economists and to obtain numerical result, this means that econometrics is use to measure the economy. (Gerhard T. 1968 p. 74),

(b) Econometric serves as the quantitative analysis of actual economic phenomena based on concurrent development of theory and observation, related by appropriate methods of inference. (Samuelson, P. A. 1954, P,141)

(c) Econometric galvanizes the tools of econometrics theory, mathematical and statistical inferences to analyze economic phenomena. (Arthur, S. G. 1964, P. 1).

(d) Econometrics is concern with the empirical determinations of economic laws. (Thiel, H. 1971, p. 1).

(e) Econometrician tries to find the sets of assumption that are both sufficiently specific and efficiently realistic to allow him to take the best possible advantage of the data available to him. (Malinvaud, E. 1966 p514)

(f) The method of econometrics research aims, essentially at a conjunction of economic theory and actual measurement, using the theory and technique of statistical inference as a bridge pier. (Haavelmo, T.1990 p.54).

The above definition of econometrics suggest that,

(1) Econometrics is more than just an "economic measurement".

(2) Econometrics is distinct discipline from economic theory mathematical economics and statistics.

(3) Econometrics is more of a research method, instrument or a tool. In which case it requires for mastery of every different skills of modality, analytical technique as well as familiarity with the nature and structure of data in questions

1.3 The Scope of Econometrics

Econometrics as a research discipline which encompasses many discipline, it galvanizes or combine economic theory, mathematical method and statistical inferences. Economic theory formulates hypothesis that expresses the relation between economic variable for instance, economic theory states that all other things being equal, an increase in input will create more output. The hypothesis may define a positive or negative relation between input and output. This relationship can be represented thus:. output = f (input).

However, theory does not provide numerical value or tell by how much output will increase or decreases resulting from any change in input.

Econometrics provides the empirical content to most economic theory. Mathematical economics expresses the relationship postulated by economic theory in the form of an equation. It however, does not provide value or measure the relationship, and it does not provide empirical verification of economic theory, which is the basic concern of econometrics. Econometrics is interested in providing values, measure for economic relation and verification of economic theory.

Econometrics uses the mathematical economics equation in such a way that is empirically testable. For instance,

$C = f(y)$	..... Economic theory
$C = \beta_0 + \beta_1 y$	..... Mathematical economics
$C = \beta_0 + \beta_1 y + U_1$	.... Econometric model

The error term U_1 inserted by econometrics lend the equation to empirical testing. The error term U_1 is the random variables which takes care of or captures other variables that are not specified in the equation.

The economic statistician is saddled with the responsibility of collecting, processing and presentation of economic data without going further by using the data collected to test economic theory.

`1.4 The Uses of Econometrics Research

From the foregoing discussions it is clear that econometrics is a mere instrument for investigating and or verifying economic theory, hence it is designed

1. To look at the past and verify the interrelatedness of economic variables over time.
2. For testing economic theory,
3. For policy decision making
4. For forecasting the future value of economic magnitude.

1- *Looking at the past.*
Econometrics is a tool which is used to look at magnitude of the interrelatedness of economic variables in the past

and compare them to the present so as to give a clear picture of the future.

2- *Testing of economic theory*

Econometrics is important in the area of testing and or verifying economic theory. This involves subjecting the hypothesis to empirical evidence and consequent upon the investigation, the claim can be accepted or rejected. Accepting the theory implies that, the researcher has come to a conclusion that with the evidence before him, the hypothesis has passed the taste of time and can be applied universally, and vice vassal.

3- *Policy decision making*

The application of econometrics method is important in the estimation of parameters of economic relationship. The values obtain are pointers to progress or set back to certain economic indicators. These values of the coefficient are important for policy decision making in government, firms, etc. Therefore, quantitative measure of the magnitude of economic relationship is important in making policy decision. For instance a firm that intends to increase its output has to evaluate its market potentials. It must have empirical evidence before it to enable it takes the right policy decision.

4- *Forecasting the future value of economic magnitude,*

It is not enough to estimate the parameters economic relationship for policy makers. These estimates are also used to forecast the future of economic magnitude which will provide the basis for planning. Forecasting also enable policy makers in the area of their policy decision making. The ability to forecast based on relevant

information guarantees the regulation of developed economies as well as planning of the economic development of underdeveloped economies.

CHAPTER TWO

2.1 Methodology of Econometric Research

The methodology of economic research is refers to the steps in econometrics research. It is the paths through which econometricians follow in their effort of investigate economic theory. We can classify the methodology of econometric research into several stages. Therefore econometricians proceed in their analysis of economic problems in the following way.

(1) Define the problem by stating the theory or hypothesis

(2) Specify the mathematical model of the theory

(3) Specify the econometric model

(4) Obtain and confront the model with data

(5) Estimate the parameters of the econometric model.

(6) Evaluate the estimate or verify the model by testing hypothesis and other tests.

(7) Forecasting and prediction

(8) Using the model for control or policy purpose

2.2 Definition of the research Problems

A research is associated with problem, therefore econometricians is always out to provide solutions to certain economic problem. Hence the first thing he does is to define the

problem by stating a hypothesis or bringing fort a theoretical statement which is expected to be verified.

For instance if we intend to study the consumption patterns of student of Nasarawa State University Keffi we may bring fort the theoretical statement. We all know from theory that consumption is a function of income i.e. C = f(y). We also know that consumption can be influenced by other factors such as taste, weather etc.

We may decide to state that, as the income of a consumer increase he tends to consume less than proportionate to the change in income. In this way, we are trying state clearly the research problem in such a way that available evidence will assert or disprove it and this will make our research work much easier.

2.3 Specification of the Mathematical Form of the Model

After defining the research problem, the next thing to do is to specify the mathematical model of the theory or statement of hypothesis earlier stated. This we can specify the statement of the theory stated in mathematical form in such a way that could be better understood.

For instance, scholars' postulates that consumption has a positive relationship with income, they did not specify it in mathematical form. Hence, mathematical economists suggests that $C = f(y)$ and $C = \beta_0 + \beta_1 y$ where $0 < \beta 1 < 1$

Where C= consumption expenditure,

y= income,

β_0 and $\beta 1$= are parameter of the model.

A model is an abstraction from reality, it is a set of mathematical equation with dependent variable at the left hand side and independent variable at the right hand side. Consumption is influenced not only by income alone but by other factors like taste, weather, etc. and this can be specify thus

$$C = f(Y, T, W)$$

$$C = \beta 0 + \beta 1Y + \beta 2T + \beta 3W$$

Where C = consumption expenditure

Y = income

T = taste

W = weather

B_0, β_1, β_2, and β_3 are parameters.

The variables at the right hand side are the independent variables while the only one at the left hand side, (i.e. consumption) is a dependent variable which depends on those on the right. It should be noted that the number of explanatory or independent variables to be included in the model depends on the nature of the problem under study and the availability of data in respect of such variables.

2.4 Specification Of The Econometric Form Of The Model

In econometric research, only important variables are included in a study, especially those that are easily measurable. The less important and immeasurable variables are taken care of or captured by the introduction of a random variable, usually denoted by U in the model. The value of the random variable cannot be observed like the ones already in the model.

E.g. $C = \beta 0 + \beta_1 Y + \beta_2 T + \beta_3 W + U$.

the introduction of the random error term make the model an econometric model because mathematical economist specify a mathematical form of economic theory without regards to other factors that are less relevant but have influence on the dependent variable or without regards for variables that are not easily measurable but could have influence on the dependent variable.

B_0, β_1, β_2, β_3, otherwise called parameters tells us the relative influence of their respective variables on the dependent variable. Hence in specifying our model, we should expect that the estimates of these parameters should conform to theoretical or a priori relationship between their respective variables and the dependent variable. That is to say that we expect them to be positive if the respective variables are positively related to the dependent variable.

2.5 The desirable properties of good model

1. The model must be relevant to the problem at hand. That is how relevant is the model to the problem at hand.
2. The theoretical justification of the model. That is how the model can be justified theoretically.
3. The model must be simple to understand. That is how simple is the model. However, simplicity must not be at the expense of reality.
4. The model must conform to real life situation and the data must be consistent.
5. The model must be stable and consistent.
6. The model must have a good forecasting ability. That is, how powerful is the forecasting ability of the model.

2.6 Data gathering and confronting the model with data

After specifying a model, the next stage is to gather data required and confront the model with the data required for estimation. Three types of data are commonly used in an empirical research, the time series, the cross sectional and the pooled data.

Time series data is the data that is collected over time; such data may be collected daily, weekly, monthly, quarterly or yearly. It may be quantitative, such as output, income, price, or qualitative such as religion, sex etc. The cross sectional data are data that are collected at a given point in time, such as opinion pool, population census etc.

Both time series and the cross sectional data can be brought together depending on the requirement of the research. Such data is called pooled data. These types of data are usually designed or classify by their sources.

Two sources of data are prominent in empirical research. The primary source of data, which is collected through the use of questionnaire or personal interview and the secondary data, is collected by a source of data which are not collected directly by the researcher himself. Such data are collected by certain government agencies of private organizations.

2.7 Estimation of the parameters in the model.

Estimation involves confronting the model with the data gathered so as to obtain numerical values for the estimates of the parameters. The processes of obtaining this value are as follows

(1) Examination of the identification condition of the function in which we are interested, so as to know whether the estimated parameters have a unique value for a given data.

(2) Examination of the problem of aggregation of variable included in the model

(3) Checking the degree of correlation between the explanatory variables, that is the degree of multicollinearity between the explanatory variables

(4) Choosing appropriate econometric technique for the estimation of the function and the critical examination of the underlying assumptions of the chosen technique.

This stage of econometric research requires the knowledge of various econometric methods, their assumptions and the economic implication for the estimates of the parameter. For now, the statistical technique of regression analysis is the main tool use to obtain estimates of the parameters. These estimates give the empirical content to the model in question.

2.8 Testing Hypothesis Or Verification Of The Model

The process of verification of a model is concerned with the determination of the statistical significance of the estimates of parameters and the entire model.

Assuming that a particular model is a reasonable approximation of reality, we have to develop suitable criteria to find out whether the estimates obtained in the model have conform to the expectation of theory that is been tested.

Before we accept that a finding is a confirmation of the theory, we must enquire whether this estimate is sufficiently below unity to convince us that this has not occurred by chance

or peculiar to the particular data we use. Such confirmation or refutation of economic theory on the bases sample evidence is base on a branch of statistical inference (hypothesis testing).

There are three major criteria by which we can verify our model. The a priori criteria, base on economic theory, the statistical criteria base on statistical theory, and the econometric criteria base on econometric method.

The a priori Criteria

Econometric theory prescribes, or defines the signs and magnitude of the coefficients of variables. For instance economic theory prescribe that consumption is a positive function of disposable income.

$C = f(Y_d)$ ..... Economic theory

$C = \beta_0 + \beta_1 Y_d$ ..... Mathematical economics

$C = \beta_0 + \beta_1 Y_d + U_1$ ..... Econometrics model

On a priori bases, we expect that $\beta_0 > 0$ and $\beta_1 > 0$ but less than 1 (i.e. $0 < \beta > 1$) since economic theory prescribed a positive relationship between consumption and disposable income. Unless otherwise stated, these restrictions must always hold.

If the signs and magnitude do not conform to our hypothesis, then we should be able to state the reason for the abnormality. Therefore the a priori criteria are based on theoretical expectation.

Statistical Criteria

The statistical criteria are referring to as the first order tests. The most widely used statistical criteria are the covariance,

correlation and standard deviation or standard error. The covariance shows how two or more variables vary together. If two variables vary together, assuming they are dependent, the covariance is the extended value of the product of the variables from the means. Correlation measures the degree of linear relationship between two or more variables.

It may be positive, negative or zero. The coefficient of correlation lies between -1 to +1. If two variables do not correlate, correlation is zero. If two variables are perfectly positively correlated the value is 1. The square of correlation coefficient is known as R^2. Thus, the standard deviation or the standard error measures the dispersion of the variable from their means. Higher standard error implies inefficient estimates while low standard error implies efficient estimates.

Econometric Criteria

The econometric criteria are known as the second-order test and are aimed at the investigating whether or not the assumptions of the econometric method are satisfied. These criteria assist us to determine whether the estimates of our parameter are unbiased, efficient, consistent and sufficient. The criteria help us to detect the violation of the ordinary least square, such as zero mean, constant variance, absence of multicollinearity, serial correlation and the normality of the disturbance term. Econometric criteria make use of F-test in testing the overall significance of our model, the stability of coefficient etc. the use of Durbin-Watson test for serial correlation, the Von-Newman ratio test have proved valuable in econometric research.

2.9 Forecasting

The ability to use the estimated model for the purpose of forecasting is one of the main objectives of econometric technique. If the model does not refute the hypothesis or theory, we may use it to predict the future value of the dependent variable on the basis of known or expected future value(s) of the explanatory variables.

A researcher can attempt to forecast the future by using the estimate of the model for period not included in our sample. It should be noted that the parameter estimate may be economically meaningful, and statistically as well as economically correct for a sample period under consideration, yet may not be well suitable for forecasting due, for example, to rapid change in the structural parameters of the relationship in the real world.

One way of using the forecasting power of a model is to apply the estimates of the model to a period not included in the sample. The resultant value is usually compared with the actual value and difference between the actual value and the forecast value is subjected to significance test. If the difference is statistically significant on examination, we conclude that the forecasting power of the model is poor.

Another way to establishing the stability of the estimates and the performance of the model in terms of its forecasting power is to estimate the function with an expanded sample. That is the sample includes additional observations. The original estimates will normally differ from the new estimates. The difference is tested for statistical significance, with appropriate technique. This will inform the decision whether to accept or reject the model for lack of forecasting ability.

Consider another measure of the forecasting ability of a model in which the government is contemplating on increasing the level of investment in order to increase the level of investment so as to increase income and consumption. Assume

that the marginal propensity to consume (MPC) is 0.75, which is 75% of the increase income goes to consumption. The multiplier is 1/1-0.75 = 4. In other word, N1million increase in investment will increase the level of income by N4million.

Reasons for poor perfomance of a model

(1) The value of the explanatory variable used in the forecast may be in accurate.

(2) The estimates of the coefficient may be poor due to deficiencies in the sample data.

(3) Sporadic changes in the economic environment may render the estimates unnecessarily fit as basis for forecasting into the future that has different economic outlook from that one in which the initial estimates were obtained.

2.10 The use of the estimated model for control or policy purpose

The model of an economy is important to a firm in term of the ability of the firm to forecast the future of its product, but government uses the model to control the economy and to actually make policy decision. The primary objective of any model is controls. Suppose that an economist estimates the consumption function as C = 10 + 0.8Yd, where C is consumption and Yd = the disposable income. Assume that government want wishes to reduce unemployment rate to about 6% by increasing expenditure to 4,00billion Naira. What level of disposable inceome will achieve the desired consumption expenditure? This gives

$$400 = 10 + 0.8Yd$$

$$400 - 10 = 0.8Yd$$

$$390 = 0.8Yd$$

$$Yd = 390/0.8$$

$$= 487.50\text{billion}$$

That is an income level of 487.50billion will produce a consumption expenditure of 400billion. Thus an estimated model can be used for control and policy decision making.

2.11 The Language Of Econometric Measurement

In econometrics measurement as may be found in statistical measurement. We can classify measurement technique into two. The parametric measurement technique and the non parametric measurement technique.

Parametric measures are measures that depend on a particular type of distribution of variables Example, the z, t, x^2, f, h, statistics, etc. In parametric measurement, it is assumed that data are normally distributed. That is the assumption of normality is adequately observed.

A non parametric measure does not depend on a particular type of distribution. Both parametric and the non parametric measures are used to measure economic variable and to determine the extent and magnitude of relationship between economic variables.

In this relationship between economic variable, one variable depends on the other. That is one is a dependent variable that depends on the other. In other words it value is determine within the equation which defined the relationship, while the other variable is an independent variable whose value

is not determine within the equation, rather, it is determined outside the equation and hence it becomes an exogenous variables.

Therefore, casual relation exist between exogenous (i.e. the independent variable) and endogenous variable (i.e. the dependent variable), for example, $C= f(Y^d)$ as stated by economic theory only tell us that there is a relationship between two variables. One is causing the other to change. Mathematical economics carries further explanation on the variables by explaining the exact type of relationships that is either positive or negative, whether the relationship starts from the origin or not. E.g. $C = B_0 + B_1Yd$. This equation tells us that there is a linear but positive relationship between consumption and disposable income. Econometrician will like to go further than this, by attempting to quantify the relationship and test the variable to establish if there is a relationship between such variables. E.g $C = B_0 + B_1Y^d + U$, where U stands error term which is introduced by econometricians to capture other variables that are not captured in the model, but are likely to affect the test of the hypothesis given that they have certain influence on the dependent variable.

2.12 Causation

Economic theory posits that consumption depends on income. That is, a change in income Y causes consumption to change in a particular way. This implies that there is a causal relationship between these variables. Therefore causation implies association (correlation) between two variables.

However, correlation or association does not imply causation. That is to say that causation is much more than association or correlation.

Conditions under Which Causation Occurs

1. There must be correlation (that is there must exist an association between the variables)
2. The timing must be right. That is the causer must occur before the causee. In other words the causer must appear before the causee.
3. There must be the absence of a third party or third factor. That is, the likely cause that may have caused the causee to change other than the first causer.

For instance, if a person who is known for witchcraft walks past the frontage of a house in the evening, and the following day a child died. People may say that it is the witch that killed the child because.

(1) There is an association between witchcraft and death. That is people associate witchcraft with death
(2) The witch passes before the child died.

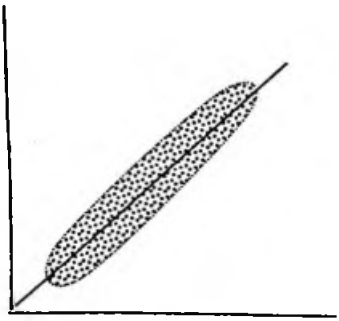
But the child may have died of convulsion or malaria. That is, there is the presence of a third party.

CHAPTER THREE

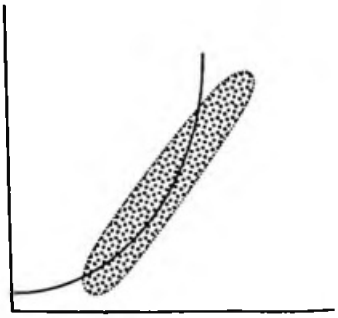
3.1 CORRELATION

Correlation is the technique use to measure the relationship between economic variables. It measures the strength of a linear relationship between economic variables. It is define as the degree of relationship existing between two or more variables. A correlation between two variables is called simple correlation, while correlation between three or more variables is called multiple correlations. Correlation can be positive or negative.

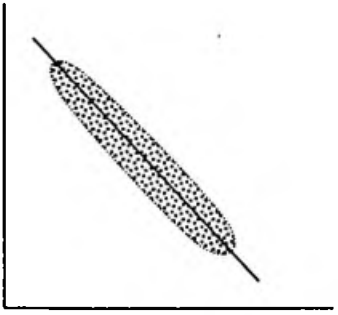
A positive correlation occurs when the variables tend to change together in the same direction. That is when they tend to increase or decrease or decrease together. Example the higher the price, the higher the quality demanded and the lower the price the lower quantity demand.



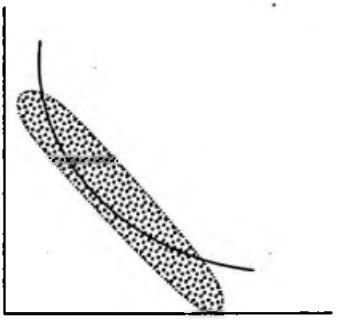
Positive Linear Correlation



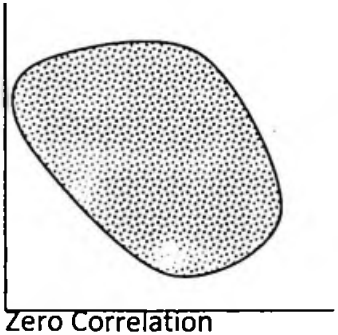
PositiveNon–Linear Correlation



Negative Linear Correlation



Negative Non-linear Correlation



Zero Correlation

Correlation can be linear, non linear or there may be no relationship at all. Correlation has serious limitation, because no matter the value of the correlation coefficient, it does not establish causation. However, it is relevant in establishing linear and non linear relationship. Correlation measures only data gotten from a normal distribution. That is correlation conficient is only meaningfull if data are normally distributed.

For correlation between X and Y the population correlation coefficient is represented by Pxy and its sample is estimated by r_{xy} hence the sample coefficient is given by

$$Rxy = \frac{\Sigma xiyi}{\sqrt{\Sigma x_i^2}\sqrt{\Sigma y_i^2}}$$

Where xi =Xi – Xi, and y= Yi – Yi these can be refers to as the variation of x and y. the sum of product of x and y is positive for instance

(Xi - Xi) (Yi - Yi) = Σxiyi the correlation between x and y can be said to be positive. If the sum of the product of deviation between x and y is negative that is (Xi - Xi)(Yi - Yi) = Σxiyi < 0

The correlation between x and y can be said to be negative.

However, this measure of correlation has two defects.

(1) It is affected by the number of observation. The greater the number of observation, the greater the number of product will be, and therefore it effect the sum of x and y that is Σxy will be different. The higher the number of observation will make the correlation appear stronger even if it were not supposed to be strong.

(2) The sum of x and y is affected by the unit of measurement. Supposing x represent height and y represent weight, because of the unit of measurement, the correlation may be higher when it not true or lower when is not true since height is measured in meter and weight in kilometers.

To correct the sum of product of the deviation is divided by the number of observation i.e. $\frac{\Sigma xiyi}{n} = S_{xy}$

Where Sxy = covariance of x and y and it is a better measures of correlation because it will not change or vary directly with the number of observation.

The covariance can still be affected by unit of measurement of x and y as stated in the second defect this can be correlated by dividing the covariance by the standard deviations of the variables. That is

$$\frac{\Sigma xy}{nS_xS_y} = \frac{Sxy}{S_xS_y}$$

Where S_{xy} = covariance of X and Y = $\Sigma x_i y_i$

S_x = Standard deviation of X_i = $\sqrt{\Sigma(x_i - \bar{x})^2/n} = \sqrt{\Sigma x_i^2/n}$

S_y = Standard deviation of Y_i = $\sqrt{\Sigma(y_i - \bar{y})^2/n} = \sqrt{\Sigma y_i^2/n}$

Therefore R_{xy} $\frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2/n\ \Sigma y_i^2/n}} = \frac{\Sigma x_i y_i}{\sqrt{(\Sigma x_i^2)\ (\Sigma y_i^2)}}$

For actual values of observations

$$R_{xy} = \frac{n\Sigma(X_iY_i) - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)\ n\Sigma y_i^2 - (\Sigma y_i)}}$$

For deviation

$$R_{xy} = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2}\ \sqrt{\Sigma y_i^2}}$$

Worked example 3.1

Given the following data

X = 1, 2, 3, 4, 5, 7.

Y = 8, 10, 12, 11, 12.

Where

X = price of a commodity

Y = quantity of commodity

Determine the correlation coefficient and test its statistical significance at 5% level of significance

X	Y	X^2	Y^2	XY
1	8	1	64	8
2	10	4	100	20
4	12	16	144	48
5	11	25	121	55
7	12	49	144	84
19	53	95	573	215

Table 3.1 price and quantity of commodity

$$Rxy = \frac{N\Sigma(XY) - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$Rxy = \frac{5(215) - (19)(53)}{\sqrt{5(95) - (19)}\ \sqrt{5(73) - (53)^2}}$$

$$= 68/79.89 = 0.85$$

The student t test statistics is used to test the significance or otherwise of sample correlation if the simple size is less than thirty (30)

$$t^* = \frac{r}{\theta_r} = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = r\sqrt{\frac{n-2}{1-r^2}}$$

$$r \approx N(0, \delta) = \sqrt{(1-r^2)/(n-2)}$$

H_0: $t^* < t_T$ or $r = 0$: r is not significant

H_1: $t^* > t_T$ or $r \neq 0$: r is significant

Level of significant = $t_{0.025}$ (two – tailed test)

Critical region = $- t_{0.025} \leq t^* \geq + t_{0.025}$

Degree of Freedom (V) = n -2 = 5 – 2 =

$$\text{Test statistic} = t^* = r\sqrt{\frac{n-2}{1-r^2}} = \frac{0.85\sqrt{5-2}}{\sqrt{1-(0.85)^2}}$$

$$= \frac{0.85\sqrt{3}}{\sqrt{1-0.7225}}$$

$$t^* = \frac{0.85\,(175)}{0.2775} = \frac{1.47}{0.53} = 2.77$$

Decision Rule

If $t^* > t_T$, reject H_0

If $t^* < t_T$, accept H_0

$t_T = 3.182$

since t^* 2.77 < t_T = 3.182, we reject H_0 and conclude that P is significant.

X	Y	x $(X - \bar{X})$	x^2 $(X - \bar{X})^2$	y^2 $(Y - \bar{Y})$	y^2 $(Y - \bar{Y})^2$	xy
1	8	-2.8	7.84	-2.6	6.76	7.28
2	10	-1.8	3.24	-0.6	0.36	1.08
4	12	0.2	0.04	1.4	1.96	0.28
5	11	1.2	1.44	0.4	0.16	0.48
7	12	3.2	10.24	1.4	1.96	0.48
19	53		22.8		11.2	13.6

Table 3.2 Variance and Covariance of price and quantity

$\bar{X} = \frac{19}{5} = 3.8$

$\Sigma x_i^2 = 22.8$

$\Sigma y^2 = 11.2$

$\bar{Y} = \frac{53}{5} = 10.6$ $\Sigma xy = 13.6$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x_i^2}\ \sqrt{\Sigma y^2}} = \frac{13.6}{\sqrt{22.8}\ \sqrt{11.2}} = \frac{13.6}{478\ (3.35)}$$

$$= \frac{13.6}{15.996} = 0.85$$

3.2 The Rank Correlation Coefficient

The correlation coefficient discussed was base on the assumption that the variables are quantitative and in most cases they have accurate data for their measurement and such data is normally distributed. In a situation where the variables are qualitative such that it cannot be measured numerically such as religion political affiliation preference for a commodity profession etc. these variable are categorical variable that may not have a precise values, hence can only be measure using another statistic. The rank correlation coefficient (developed by Spearman) is used in this case by ranking the variables in a specific sequence. That is either in order of importance or size using number 1,2,3,...,etc and measure the relationship between the ranks instead of their numerical values.

In ranking the variables, it does not matter if they are ranked in ascending or descending order so long as they apply the same on both sides. If two variables have the same value we assign their mean rank. This is a non-parametric measure of correlation or association between two economic variables; it is also used when the data are feared not to be normally distributed. That is to say that this measure does not depend on any kind of distribution. The spearman rank correlation coefficient is given by

$$r = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

where;

D = difference between ranks of corresponding paves of X and Y

n = number of observation

Given the following preference on the commodity X and commodity Y

Time Periods Bays	Qty Y_i	Price X_i	Rank of Y_i	Rank of X_i	D	D^2
1	10	2	10	10	0	0
2	20	4	9	9	0	0
3	50	6	6.5	8	-1.5	2.25
4	40	8	8	7	0.5	0.25
5	50	10	6.5	6	0	0
6	6	12	5	5	0	0
7	80	14	4	4	0	0
8	90	16	2.5	2	-0.5	0.25
9	90	18	2.5	2	0.5	0.25
10	120	20	1	1	0	0
						4

Figure 3:1 Rank Correlation table

Therefore, $r = 1 - \frac{6(4)}{10(10^2-1)} = 1 - \frac{24}{990} = 1 - 0.00242$

$= 0.98$

3.3 Limitations of Analysis of Correlation

1. Correlation is only a measure of linear relationship. That is, it can only be applied when the relationship is linear. Two variables can be strongly related but may be non-linear
2. Although, it is a measure of association, but it does imply a functional relationship between the variables. That is, correlation does not imply causal relationship. It does not suggest whether one variable cause the other to change.

CHAPTER FOUR

4.1 Bivariate Regression Model (Simple Regression)

This is the simplest possible regression analysis of two variables. This is where the dependent variable is related to a single independent variable. Other names for the dependent and independent variables are:

- Dependent → Independent
- Regressand → Regressor
- Explained → Explanatory

For example, change in the price of a commodity (independent variable) can explain fluctuations in the quantity demanded for a particular commodity (dependent variable) at a particular time.

PURPOSE: The main purpose of regression analysis is to estimate and/or predict the population mean value of the dependent variable, given some values of the explanatory variables.

In the Bivariate linear model, the parameters (β's) are raised to the first power (power 1) and the explanatory variable may or may not be raised to that power.

4.2 Conditional Model

Suppose we have 50 families who are divided into 10 income groups, say N100, N200, N300, N400, N500, N600, N700, N800, N900 and N1,000 per week.

Suppose under each income group, we have five different consumption expenditures that deepends on the respective incomes.

- Hence, income is the independent variable denoted by X
- Consumption expenditure is the dependent variable denoted by Y

Clearly $Y = f(X)$

When we get the mean consumption expenditure corresponding to the weekly level of income of N100, we call this value the conditional expected value, also for income level of N200, N300, N400, etc.

These mean values are called conditional expected values as they depend on the given values of the (conditional) variables X, i.e. N100, N200, N300, etc.

They are denoted as

$E(Y/X)$ that is, the expected value of Y given the value of X.

However, when we get the mean consumption expenditure for all the 50 families, it is unconditional mean since the income levels; N100, N200, etc are disregarded.

It is denoted as

$E(Y)$ that is the expected value of Y. Thus, when we say expected value of Y, we mean '**Mean**' value of Y.

When we deal with conditional expected values $E(Y/X)$, we are dealing with conditional model.

It is clear that $E(Y/X) = f(X)$

This equation is called conditional model, conditional expectation function (CEF) or population regression function (PRF).

The general form of CEF is

$$E(Y/X_t) = \alpha + \beta X_t$$

Which is a linear function of X_i, where α, and β are given parameters. α = intercept while β = slope of the linear CEF.

The regression model can be written as

$$Y_t = E(Y/ X_t) + U_t$$

Substituting for E(Y/X), we have:

$Y_t = \alpha + \beta X_t + U_t$ = this is the actual or true relationship.

Where

Y_t = dependent variable at period t

X_t = independent variable at period t

U_t = the deviation or the unobservable random variable (positive or negative) at period t

It is technically called the stochastic disturbance or error term.

We can see that

$$Y_t = (\alpha + \beta x_t) + (U_t)$$

Systematic or Deterministic component

Non-systematic or Stochastic component

It then follows that

$$U_t = Y_t - E(Y/X)$$

The estimated relationship is given as

$$Y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{e}$$

Where

$\hat{\alpha}$ = estimate of the true intercept α

$\hat{\beta}$ = estimate of the true parameter β

e = estimate of the true value of U_t

The estimated regression line is given as

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta}x_t$$

Since e stand for stochastic error term, it is the difference between the actual value of Y_t and its estimated value, $\hat{Y}_t$, then:

$$e = Y_t - \hat{Y}_t$$

We shall use this later to determine the OLS equations

4.3 The OLS Model And Its Assumption

We can see that our linear classical estimation function CEF is divided into two components. We have two forms of assumption. the stochastic assumptions (features of the random variable, u) and the other assumptions

Stochastic Assumptions

- U_t is a random variable , that is, for each value X_t it can assume any value, positive, negative or zero

- Zero mean value of U_t : $E(U_t) = 0$ technically, the conditional mean value of $U_t = 0$, which means U_t does not systematically affect the mean value of Y. this is

because the positive values of U_t offset the negative ones so that their average or mean of Y is zero

- U_t is homoscedastic or the variance of U_t is constant. This means that for all values of X_t, the U_t will show the same dispensation round their mean.

Symbolically,

$$\text{Var}\,(U_t/X_t) = E\,(U_t - E(U_t/X_t)^2$$

$$= E(U_t^2/X_t)$$

$$= \delta^2$$

$$\text{Or } \delta^2 = K$$

- U_t is normally distributed, we can symbolically summarize the above four assumptions as

$$U_t \approx N\,(0, \delta^2)$$

- U_t is a normally distributed random variable with zero mean and constant variance δ^2.

- There is zero covariance between U_t and X_t of different periods. Ie Zero relationship between U_t and X_t

Symbolically,

$$\text{Cov}\,(U_t/X_t) = E\{(X - E(X)\,(U - E(U)\}$$

$$= E\{X - E(X)\}\{U$$

Since E(U) = 0

$$= E\{X - E(X)\,(U)\}$$

$$= E\{XU - E\,X(U)\}$$

$$= E(XU) - EX.\ E(U)$$

$$= E(XU) - EX.\ E(0)$$

$$= X\ E(U) - 0$$

$$= X\ (0) - 0$$

$$= 0, \text{ since X is fixed.}$$

- There is also zero Covariance between U_t's of different period.

That is they do not vary together, they are independent of one another.

Formally, is the same way

$$Cov\ (U_i\ U_j) = E\ \{(\ U_i - E(U_i)\ (U_j - E(U_j)\}$$

$$= 0$$

- All errors are only accumulated by U_t, meaning that the explanatory variables are measured without errors.

Non – Stochastic Assumptions

- The regression model is linear in the parameter
  E.g. $Y_t = \alpha_t + \beta X_t + U_t$
- X_t is assumed to be non-stochastic that is, they are fixed in repeated sampling.
- The number of observations n must be greater than the number of explanatory variables or parameters.
- X_t is variable in a given sample.
- The regression model or relationship is correctly specified – no specification bias or error in the model

- No perfect multicollinearity among the explanatory variables – no perfect linear correlation.
- The macro-variables are correctly aggregated – individual units are correctly summed up.

Having reviewed the assumptions, we consider the normal equation of OLS.

4.4 The Normal Ols Equation

First: let us establish the expected value of the dependent variable Y_t and its variance.

Recall: $Y_t = \alpha_t + \beta x_t + U_t$

Then: $E(Y_t) = E(\alpha_t + \beta x_t + U_t)$

$= E(\alpha_t + \beta x_t) + E(U_t)$

But by our assumption above:

$E(U_t) = 0$

Therefore, $E(Y_t) = \alpha_t + \beta X_t + 0$

Hence, $E(Y_t) = \alpha_t + \beta X_t$

This is the mean value of Y

For the variance

$E(Y_t) = \alpha_t + \beta X_t + U_t$

Var (Y_t) $= E\{Y_t - E(Y_t)\}^2$

But, $E(Y_t)$ $= E(\alpha_t + \beta X_t + U_t - \alpha_t + \beta X_1)^2$

$= E(U_t)^2$

In our assumption $E(U_t)^2 = \delta^2$

Hence, $Var\ (Y_t) = \delta^2$

Now let's find the OLS equations:

Recall:

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta}x_t$$

And $$e = Y_t - \hat{Y}_t$$

Substituting

$$e = Y_t - \hat{Y}_t$$

$$= Y_t - (\hat{\alpha} - \hat{\beta}x_t)$$

$$e = Y_t - \hat{\alpha} + \hat{\beta}X_t$$

Square both sides

$$e^2 = (Y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

Take their sum

$$\Sigma e^2 = \Sigma(Y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

According to one of our stochastic assumption

$$\Sigma e^2 = 0$$

Or rather, Σe^2 is the deviation associated with the influence of U_t. the larger its value (variance), the less reliable the value of the parameters or coefficients. So, we have to minimize Σe^2.

To do so, we differentiate it with respect to α and β.

$$\Sigma e^2 = \Sigma(Y_t - \hat{\alpha} - \hat{\beta}X_t)^2$$

$$\text{Let } U = (Y_t - \hat{\alpha} - \hat{\beta}X_t)^2$$

Since it is a composite function, (note $U \neq$ our U_t), then

$$\Sigma e^2 = \Sigma(U)^2$$

More precisely: $U = Y_t - \hat{\alpha} - \hat{\beta}X_t$ ...(i)

$$\Sigma e^2 = \Sigma(U)^2 \quad \text{...(ii)}$$

From (i)

$$\delta U/\delta\alpha = -1$$

From (ii)

$$\delta\Sigma e^2/\delta U = 2\Sigma U$$

Then, $e^2/\delta\hat{\alpha} = \delta\Sigma e^2/\delta\alpha \times \delta U/\delta Q$

$$= -1 \times 2\Sigma U$$

Therefore,

$$\delta\Sigma e^2/\delta\hat{\alpha} = \delta\Sigma e^2/\delta\alpha \times \delta U/\delta Q$$

$$= -2\Sigma (Y_t - \hat{\alpha} - \hat{\beta}X_t)$$

But according to first-order condition

$$\delta\Sigma e^2/\delta\alpha = 0$$

$$-2\Sigma(Y_t - \hat{\alpha} - \hat{\beta}X_t) = 0$$

Dividing by -2

$$\Sigma (Y_t - \hat{\alpha} - \hat{\beta}X_t) = 0$$

Opening bracket,

$$\Sigma Y_t - \Sigma\hat{\alpha} - \Sigma\hat{\beta}X_t = 0$$

Applying the usual summation rule $\Sigma = n$ to α

$$\Sigma Y_t - n\hat{\alpha} - \Sigma\hat{\beta}X_t = 0$$

Re-arranging

$$\Sigma Y_t - \Sigma\hat{\beta}X_t = n\hat{\alpha} \quad \text{... Normal Equation (1)}$$

$$\Sigma Y_t - \hat{\beta}\Sigma X_t = n\hat{\alpha}$$

Dividing through by n

$$\Sigma Y_t/n - \hat{\beta}\Sigma X_t/n = n/n\hat{\alpha}$$

Therefore

$$\hat{\alpha} = \Sigma Y_t/n - \hat{\beta}\Sigma x_t/n$$

But, $\Sigma Y_t/n = \overline{Y}_t$ and $\Sigma X_t/n = \overline{X}_t$:

Then

$$\alpha = Y_t - \beta \Sigma x_t$$

In general

$$\hat{\alpha} = \bar{Y} - \beta \bar{X}$$

We also differentiate Σe^2 with respect to $\hat{\beta}$

Given $U = Yt - \hat{\alpha} - \hat{\beta} X_t$

$$\delta U / \delta \hat{\beta} = -X_t$$

And $\Sigma e^2 = \Sigma U^2$

$$\delta \Sigma e^2 / \delta U = 2 \Sigma U$$

Similarly, $\delta \Sigma e^2 / \delta \beta = \delta \hat{U}^2 / \delta \beta \times \delta \hat{\Sigma} e^2 / \delta u$

$$= -2 X_t \Sigma U$$

$$\delta \Sigma e^2 / \delta \beta = -2 X_t \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t)$$

and

$$\delta \Sigma e^2 / \delta \hat{\beta} = 0$$

$$-2 X_t \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$$

Dividing by -2

$$X_t \Sigma (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$$

Opening the bracket,

$$\Sigma X_t Y_t - X_t \Sigma \hat{\alpha} - \Sigma \hat{\beta} X_t^2 = 0$$

substituting $\alpha = (\overline{Y_t} - \hat{\beta}\overline{X_t})$

$$\Sigma Y_t - n - \Sigma \beta X_t = 0$$

Then: $\Sigma X_t Y_t - \Sigma X_t(\overline{Y_t} - \hat{\beta}\overline{X_t}) - \hat{\beta}\Sigma X_t^2 = 0$

$$\Sigma X_t Y_t - \Sigma X_t \overline{Y_t} - \beta \Sigma X_t \overline{X_t} - \beta \Sigma X_t^2 = 0$$

$$\Sigma X_t Y_t - Y_t \overline{\Sigma X_t} - \overline{X_t} \beta \Sigma X_t - \beta \Sigma X_t^2 = 0$$

recall that, $\Sigma \overline{Y_t} = nY_t$ and $\Sigma X_t = n\overline{X_t}$: then

then, $\Sigma X_t Y_t - n\overline{Y_t}\overline{X_t} + \hat{\beta} n \overline{X_t}^2 - \beta \Sigma X_t^2 = 0$

collecting like terms in $\hat{\beta}$.

$$\Sigma X_t Y_t - n\overline{Y_t}\overline{X_t} = \beta \overline{X_t}^2 - \beta \Sigma n \overline{X_t}^2$$

$$\Sigma X_t Y_t - n\overline{Y_t}\overline{X_t} = \beta(X_t^{\overline{2}} - \Sigma n X_t^{\overline{2}})$$

Therefore,

$$\beta = \frac{\Sigma X_t Y_t - n\overline{Y_t}\overline{X_t}}{\Sigma X_t^2 - \Sigma \overline{X_t}^2}$$

Now substituting $\overline{Y_t} = \Sigma Y_t / n$ and $\overline{X_t} = \Sigma X_t / n$,

$$\hat{\beta} = \frac{N\Sigma X_t Y_t - (\Sigma X_t)(\Sigma Y_t)}{N\Sigma X_t^2 - (\Sigma X_t)^2}$$

Therefore

$$\hat{\beta} = \frac{N\Sigma X_t Y_t - (\Sigma X_t)(\Sigma Y)}{N \Sigma X_t^2 - (\Sigma X_t)^2}$$

Which is the same thing as, the second normal equation.
In deviations

$$\hat{\beta} = \frac{\sum xy}{\sum x^2}$$

And

$$\hat{\alpha} = \bar{Y}_t - \beta\bar{X}_t$$

It should also be noted that in demand function, the elasticity can be derived from the estimated regression line.

Given

$$\hat{Y}_t = \hat{\alpha} - \hat{\beta}X_t$$

$$\delta\hat{Y}_t / \delta X_t = \hat{\beta}$$

By definition, price elasticity is given as

$$S/p = \frac{\delta Y}{\delta X} . X/Y$$

Where X = price, Y = quantity and $\delta Y / \delta X = \hat{\beta}$

Then s/p $= \hat{\beta} . \bar{X}_t / \bar{Y}_t$

$\bar{\hat{Y}}_t$ = mean value of the estimated quantity

$\bar{X}_t$ = the average price in the sample.

But $\bar{\hat{Y}}_t = \bar{Y}_t$

Therefore

$$S/p = \hat{\beta} . \bar{X}_t / \bar{Y}_t$$

More precisely, according to the OLS method, the best way to fit a curve in a set of points is to find the line whose sum of square deviation is minimum. And we can do this by the following ways;

(a) Find the error; if $Y = \beta o + \beta_1 X_1 + U$, then the error $U = Y - \beta o - \beta_1 X_1$

(b) Sum the error, that is $\sum U = \sum(Y - \beta o - \beta_1 X_1)$

(c) Square the sum of the error, that is $\sum U^2 = \sum(Y - \beta o - \beta_1 X_1)^2$

(d) Then minimize the error, by taking partial derivative in respect of βo and β_1.

These estimate or coefficient can be obtained by minimizing the sum of squared residuals. i.e $\sum e_i^2 = \Sigma(Y_i - Y_i)^2 = \Sigma(Y - \beta_0 - \beta_1 X_1)^2$

We obtain by

(1) Finding the error, (in our model)

$$U_1 \hat{} = Y \hat{} - \beta_0 - \beta_1 X_1$$

(2) Squaring the error

$$U_i^2 = (Y - \beta_0 - \beta_1 X_1)^2$$

(3) Summing the squared error

$$\Sigma U_i^2 = \Sigma (Y - \beta_0 - \beta_1 X_1)^2$$

To minimize we take partial derivation in respect of β_o, β_1 and β_2 respectively so as to get our normal equation and subsequently, the coefficient β_o and β_1. i.e take partial derivation and equate it to zero.

$$\frac{d\Sigma Ui^2}{d\beta_0} = (-1)(2)\,\Sigma(Y - \beta_0 - \beta_1 X_1) = 0$$

$$= -2\,\Sigma\,(Y - \beta_0 - \beta_1 X_1) = 0$$

Divide through by -2

$$= \frac{-2\,\Sigma\,(Y - \beta_0 - \beta_1 X_1)}{-2} = \frac{0}{-2}$$

$$= \Sigma\,(Y - \beta_0 - \beta_1 X_1) = 0$$

Where $\frac{0}{-2} = 0$

Open the bracket

$= \Sigma Y - \Sigma\beta_0 - \Sigma\beta_1 X_1 = 0$

$= \Sigma Y - N\beta_0 - \beta_1 \Sigma X_1 = 0$ ..... (1) Normal equation

Where $\Sigma\ \beta_0 = \beta_0 + \beta_0 + \beta_0 + ... = N\ \beta_0$ i.e. Number of $\beta_0 \times \beta_0$

To find our β_0, we multiply through by N and make β_0 the subject of the formula.

$$\frac{\Sigma Y}{N} = \frac{N\beta_0}{N} + \frac{\beta_1 \Sigma X_1}{N}$$

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}_1$$

Therefore

$$\beta_0 = \bar{Y} - \beta_1 \bar{X_1}$$

To get the 2nd normal equation we take partial derivation in respect of the β_1

$$\frac{d\Sigma Ui^2}{d\beta_1} = (-X_1)\ (2)\ \Sigma(Y - \beta_0 - \beta_1 X_1) = 0$$

$$= (-2X)\ \Sigma\ (Y - \beta_0 - \beta_1 X_1) = 0$$

To eliminate -2 we divide through by -2

$$\frac{(-2)\Sigma(Y - \beta_0 - \beta_1 X_1)}{-2} = \frac{0}{-2}$$

$$X_1 \Sigma\ (Y - \beta_0 - \beta_1 X_1) = 0$$

$$\Sigma X_1 Y - \beta_0 \Sigma X_1 - \beta_1 \Sigma X_1^2 = 0$$

$$\Sigma X_1 Y - \beta_0 \Sigma X_1 - \beta_1 \Sigma X_1^2 \quad \text{....(2) normal equation}$$

$$\beta_0 = \Sigma Y = N\beta_0 + \beta_1 \Sigma X_1 \quad \text{...(1)}$$

$$\beta_1 = \Sigma X_1 Y = \beta_0 \Sigma X_1 + \beta_1 \Sigma X_1^2 \quad \text{...(2)}$$

These equations are called normal equations and are used to obtain the coefficient β_0, and β_1 in other words they are used to fit the regression line. To do that we use crammer rule matrix as follows

Given the above OLS estimators we need to also obtain the standard error of these estimates which are important to establish confidence interval and to test the statistical hypothesis. The relevant formulas are as follows

$$[A] = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \end{bmatrix} = \begin{pmatrix} N & \Sigma X_1 \\ \Sigma X_1 & \Sigma X_1^2 \end{pmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$[A] = N \Sigma X_2^2 - (\Sigma X_{1i})^2$$

To determine β_1 we can replace the middle column with the sums on the right hand of the equations and divide by the determinants

$$\hat{\beta}_1 = \frac{\begin{vmatrix} N & \Sigma Y \\ \Sigma X_1 & \Sigma X_1 Y \end{vmatrix}}{N\Sigma X_1^2 - (\Sigma X_1)^2}$$

$$= \frac{N\Sigma X_1 Y - \Sigma X_1 \Sigma Y}{N\Sigma X_1^2 - (\Sigma X_1)^2}$$

Given the above OLS estimates, we need to also obtain the standard error of these estimates which are important to establish confidence interval and to test the statistical hypothesis. The relevant formulas are as follows

4.5 Total, Explained And Residual Sum Of Squares

Recall:

$Y_t = \alpha + \beta X_t + U_t \rightarrow$ Actual relationship

$Y_t = \hat{\alpha} + \hat{\beta} X_t + e \rightarrow$ estimated relationship

$\hat{Y}_t = \hat{\alpha} + \beta X_t \rightarrow$ estimated regression line

- Total deviation in Y_t from its mean is given as $\bar{Y}_t$
  $y = Y_t - \bar{Y}_t$
  This is decomposed into
- Explained Deviation/variation, which is the deviation of the regressed value from the mean.
  It is given as
  $\hat{y} = \hat{Y}_t - \bar{Y}_t$
- Residual/unexplained variation which is the deviation of the observation Y_t from the mean.

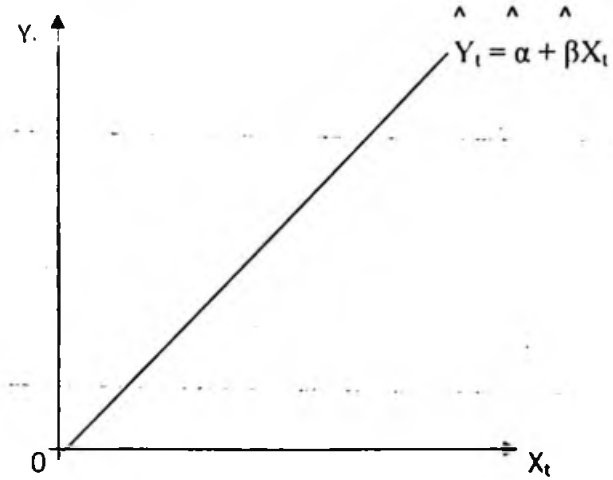
It is given as

$e = Y_t - \hat{Y}_t$

$\hat{y}$ is the part of total variation which is the derivation of the observation Y_t which is explained by the regression line.

e is that part of total variation in Y_t which is attributed to the random variable U_t.

Graphically:



Points above and below the regression line are expected by e.

It is clear that:

$y_t = Y_t - e$

We now find the value of $\hat{y}_t$

recall that:

$\hat{y} = Y_t - \bar{Y}_t$

but $\hat{Y}_t = \hat{\alpha} + \hat{\beta}X_t$

and $\bar{Y}_t = \hat{\alpha} + \hat{\beta}\bar{X}_t$

(From the first normal OLS equation)

Substituting:

$\hat{y}_t = \hat{\alpha} + \hat{\beta}X_t \quad - (\hat{\alpha} + \hat{\beta}\bar{X}_t)$

$\hat{y}_t = \hat{\alpha} + \hat{\beta}X_t \quad - \hat{\alpha} - \hat{\beta}\bar{X}_t$

$\hat{y}_t = \quad \hat{\beta}X_t - \hat{\beta}\bar{X}_t$

Factorizing

$\hat{y}_t = \quad \hat{\beta}(X_t - \bar{X}_t)$

But, $\quad X_t - \bar{X}_t = \bar{x}$

Therefore

$\hat{y}_t = \quad \hat{\beta}\bar{X}_1$

Now, the total deviation is

$y = \hat{y}_t + e$

Taking the sum of their squares

$$\underset{\text{TSS}}{\sum y_t^2} = \underset{\text{ESS}}{\sum \hat{y}_t^2} + \underset{\text{RSS}}{\sum e} \quad \ldots (1)$$

This can also be written as

$$\sum(Y_t - \bar{Y}_t)^2 = \sum(\hat{Y}_t - \bar{Y}_t)^2 + \sum(Y_t - \hat{Y}_t)$$

Where; TSS = Total Sum of Square

ESS = Explained Sum of Square

RSS = Residual Sum of Square

Now substitute the value $\hat{y}_t = \hat{\beta}X_1$ in equation (1)

$$\sum Y_t^2 = \sum(\hat{\beta}X_t)^2 + \sum e \quad \ldots (2)$$

We want to know the ESS and RSS as a percentage of TSS

ESS/TSS = Explained Variation/Total Variation

$$= \frac{\sum(\hat{\beta}X_t)^2}{\sum Y_t^2}$$

$$= \frac{\sum(\beta)^2 (X_t)^2}{\sum Y_t^2}$$

But recall that $\beta = \frac{\sum xy}{\sum x^2}$

Therefore $\frac{ESS}{TSS} = \frac{(\beta)^2 \sum(X_t)^2}{\sum x^2}$

$$\frac{[\sum xy]^2}{\sum x} \cdot \frac{\sum(X_t)^2}{\sum Y_t^2}$$

$$\frac{ESS}{TSS} = \frac{\sum(xy)^2}{\sum x^2 . \sum y^2}$$

Therefore $\frac{\sum y^2}{\sum y^2} = \frac{ESS}{TSS} = \frac{\sum(xy)^2}{\sum x^2 . \sum y^2}$

Recall that,

The $r^2 = \frac{\sum(xy)^2}{\sum x^2 . \sum y^2} = \frac{\sum \hat{y}^2}{\sum y^2}$

therefore r^2 = is the explained variation expressed as a percentage of total variation

r^2 = co-efficient of determination, which is the test of the goodness of fit of the regression line. The better the goodness of fit, the closer the observation to the line.

r^2 shows the percentage change in Y_t that is brought about by the explanatory variable X_t. The closes it is to 1, the better the goodness of fit and vice versa.

Symbolically $0 \leq r^2 \leq 1$

From (2)

$$\sum Y_t^2 = \sum(\hat{\beta}X_t)^2 + \sum e$$

To find the remaining/residual/unexplained variation in terms of r, divide through by $\sum y_t^2$

$$\frac{\sum y_t^2}{\sum y_t^2} = \frac{\sum(\beta x_t)^2}{\sum y_t^2} + \frac{\sum e^2}{\sum y_t^2}$$

$$1 = r^2 + \sum e^2/\sum Y_t^2$$

Re arranging

$$\sum e^2/\sum y_t^2 = 1 - r^2$$

$$\sum e^2 = \sum y_t^2 (1 - r^2)$$

Hence, $\sum e^2 = \sum y_t^2 (1 - r^2)$

Conversely, if we know $\sum e^2/\sum y_t^2$, we can find r^2 as follows

$$r^2 = 1 - \sum e^2/\sum y_t^2$$

Based on this, we deduce that:

- If $\sum e^2/\sum y_t^2 = 0$, $r^2 = 1$; this implies that all the observations lie on the regression line and there will be no scatter of points. This means that Y_t is perfectly explained by X_t
- If $r^2 = 0$, $\sum e^2/\sum Y_t^2 = 1$; this means that variations in Y_t are not at all explained by X_t
- If $r^2 < 1$, $\sum e^2/\sum y_t^2 > 0$ but < 1, this means that some parts of the variations is explained by the regression line and some part is unexplained.

It should be noted that r^2 is very much related to the slope of the regression line, β

Recall that:

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2}$$

$$= \frac{(\sum xy)(\sum xy)}{\sum x^2 \sum y^2}$$

$$= \frac{\sum xy}{\sum x} \quad \frac{\sum xy}{\sum y}$$

But,

$$\beta = \frac{\sum xy}{\sum y^2}$$

Therefore,

$$r^2 = \beta_1 \frac{\sum xy}{\sum y^2}$$

The r^2 is the coefficient of determination of a simple regression

Given the above OLS estimates, we need to also obtain the mean, variances, and standard error of these estimates which are important to establish confidence interval and to test the statistical significance. The relevant formulas are as follows

4.6 Mean And Variance Of Ols Estimates

Mean of $\hat{\beta}$

If all the possible sample of size n are taken from the population of X_t and $\hat{Y}_t$, then the mean value of β will be its expected value:
Mean $\beta_h = E(\beta)$

Recall: $\beta = \frac{\sum xy}{\sum x^2}$

But, $y = Y - \bar{Y}$

Therefore, $\hat{\beta} = \frac{\sum (Y - \bar{Y})x}{\sum x^2}$

$\hat{\beta} = \frac{\sum x(Y - \bar{Y})}{\sum x^2}$

$\hat{\beta} = \frac{Y\sum x - \bar{Y}\sum x)}{\sum x^2}$

$= \frac{Y\sum x}{\sum x^2} - \frac{\bar{Y}\sum x}{\sum x^2}$

Taking the second part $\frac{\bar{Y}\sum x}{\sum x^2}$ and by definition $\sum x = 0$

Therefore, $\frac{\bar{Y}\sum x}{\sum x^2} = 0$

Or rather $\frac{\bar{Y}\sum x}{\sum x^2} = \frac{\bar{Y}\sum (X - \bar{X})}{\sum X^2} = \frac{\sum \bar{X}Y - \sum \bar{X}\bar{Y})}{\sum X^2}$

But since $n = \sum$ and $\sum X = n\bar{X}$,

$$\frac{\bar{Y}\sum x}{\sum x^2} = \frac{n\bar{X}\bar{Y} - n\bar{X}\bar{Y})}{\sum X^2} = \frac{0}{\sum X^2}$$

Therefore, $\frac{\nabla\sum x}{\sum x^2} = 0$

We are left with

$$\frac{\sum xy}{\sum X^2} = \sum K_i Y$$

We now take the expected value of $\hat{\beta}$

$E(\hat{\beta}) = E(\sum K_i Y_i)$

But $Y_t = \alpha + \hat{\beta} X_t - U_t$

$E(\hat{\beta}) = E\{(\sum K_i)(\sum Y_i)\}$

$= E\{(\sum K_i)(\alpha + \beta X_t - U_t)\}$

$= E(\alpha \sum K_i + \sum K_i \beta X_t - \sum K_i U_t)$

$= E(\alpha \sum K_i + \beta \sum K_i X_t - \sum K_i U_t)$

But originally, $\sum K_i = 0$

Since, $\frac{\sum x}{\sum x^2} = 0$

I.e. $\frac{\sum x}{\sum x^2} = \frac{\sum(X - \bar{X})}{\sum x^2} = \frac{(\sum X - \sum \bar{X})}{\sum x^2}$

$\frac{(\bar{X} - n\bar{X})}{\sum x^2} = 0$

But $\sum y_i x_i = 1$, since $\sum y_i x_i = \frac{\sum x}{\sum x^2} \cdot X$

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

$$= \frac{\sum X^2 - n\bar{X}\bar{X}}{\sum X^2}$$

$$= \frac{\sum X^2 - nX^2}{\sum X^2 - nX^2} = 1$$

Therefore,

$$E(\beta) = E[\beta(0) + \beta(1) + \sum k_i U]$$

$$= E[\beta + \sum k_i E(U)]$$

But by one of our assumption $E(U) = 0$

Therefore

$$E(\hat{\beta}) = E(\beta) + \sum k_i(0)$$

$$E(\hat{\beta}) = E(\beta)$$

Hence

$$E(\hat{\beta}) = \beta$$

This means that the OLS estimate $\hat{\beta}$ is equal to true value of the population parameter β. Hence the OLS estimate is unbiased.

The Variance of $\hat{\beta}$

By definition Var $(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2$

But since OLS is unbiased, we use.

$\text{Var}(\hat{\beta}) = \text{Var}(\sum k_i Y)$

$= \sum k_i^2 \text{Var}(Y)$

And $\text{Var}(Y) = E[Y - E(Y)]^2$

And also $Y = \alpha + \beta X - U$, while

$E(Y) = \alpha + \beta X$

Therefore, $\text{Var}(Y) = E[\alpha + \beta X + U - (\alpha + \beta X)]^2$

$= E[\alpha + \beta X + U - \alpha - \beta X)]^2$

$= E(U)^2$

Therefore, $\text{Var}(\beta) = \sum k_i^2 \text{Var}(Y)$

$= \sum k_i^2 E(U)^2$

But $E(U)^2$ is given as $6^2 = \sum e^2/n\text{-}K$

And $\sum k_i^2 = \frac{(\sum x)^2}{\sum x^2} = \frac{(\sum x.)^2}{\sum x.\sum x}$

$= (1/\sum x)^2$

$= 1/\sum x^2$

Therefore,

$\text{Var}(\hat{\beta}) = \sum k_i^2 . E(U)^2$

$= 1/\sum x^2 . \sum e^2/n\text{-}K$

Hence,

$\text{Var}(\hat{\beta}) = 1/\sum x^2 . \sum e^2/n\text{-}K$

The Mean of α

Recall $\hat{\alpha} = \overline{Y_t} - \hat{\beta}\overline{X_t}$

subs $\hat{\beta} = \sum k_i^2 Y$

$\hat{\alpha} = \overline{Y_t} - \sum k_i^2 \overline{Y}\overline{X}$

we have $\hat{\alpha} = Y_t - X_t \sum k_i^2 Y$

$\hat{\alpha} = \sum [1/n - \overline{X}k_i^2]Y$ since $\overline{Y} = \sum y/n$

by factorizing ∑ and Y.

Taking the expected value

$E(\hat{\alpha}) = E(Y). \sum [1/n - \overline{X}k_i^2]$

$= \sum [1/n - \overline{X}k_i^2] (\alpha + \beta x)$ since $E(Y) = \alpha + \beta x$

Opening the bracket

$E(\hat{\alpha}) = \sum [\alpha/n - \alpha\overline{X}\sum k_i^2 + \beta x/n - \beta\overline{X}\sum k_i^2]$

$E(\hat{\alpha}) = \sum \alpha/n - \alpha\sum\overline{X}\sum k_i^2 + \sum\beta x - \beta \sum\overline{X}\sum k_i^2 X$

$E(\hat{\alpha}) = \sum \alpha/n - \alpha\sum\overline{X}\sum k_i^2 + \beta\sum x - \beta \sum\overline{X}\sum k_i^2 X$

Taking $\sum = n$, $\sum x = n\overline{X}$ and knowing that $\sum k_i = 0$ and $\sum k_i X = 1$

$E(\hat{\alpha}) = \frac{n\alpha}{n} - \alpha\sum \overline{X}(0) + \beta n\overline{X} - \beta n\overline{X}(1)$

$= \alpha - 0 + \beta n\overline{X} - \beta n\overline{X}$

$= \alpha - 0 + 0$

Hence,

$E(\hat{\alpha}) = \alpha$

This means that the OLS estimate α is exactly equal to the true value of the population parameter α

The Variance of α

$Var(\hat{\alpha}) = E\,[\hat{\alpha} - E(\alpha)]^2$

But $E(\alpha) = \alpha$

Therefore

$Var\,(\hat{\alpha}) = E\,[\hat{\alpha} - \alpha]^2$

Recall that:

$\hat{\alpha} = \overline{Y} - \beta\overline{X}$, and $\beta = \sum k_i Y$

Therefore,

$\hat{\alpha} = \overline{Y} - \sum k_i Y\overline{X}$

$= \frac{\sum Y}{n} - \frac{\sum k_i \overline{Y}\overline{X}}{1}$

Factorizing Y and $\sum$

$= \sum[1/n - \sum X]\overline{Y}$

Taking the variance

$$\text{Var}(\hat{\alpha}) = \text{Var}[\sum(1/n - kX)Y]$$

$$\text{Var}(\hat{\alpha}) = [\sum(1/n - k_iX)]^2 \text{Var}(Y)$$

$$= \sum(1/n - k_i\bar{X})(1/n - k_i\bar{X}).\text{Var}(Y)$$

$$= \sum[1/n^2 - k_i\bar{X}) - k_i\bar{X} + k_i^2\bar{X}^2].\text{Var}(Y)$$

$$= \sum/n^2 - 2\bar{X}\sum k_i + \sum k_i^2\bar{X}^2].\text{Var}(Y)$$

$$= n/n^2 - 2\bar{X}(0) + (\sum x/\sum x^2)\bar{X}^2].\text{Var}(Y)$$

$$= 1/n - (1/\sum x^2)\bar{X}^2 . \text{Var}(Y)$$

$$= 1/n - (\bar{X}^2/\sum x^2) . \text{Var}(Y)$$

$$= \frac{\sum X^2 + n\bar{X}^2}{n\sum X^2} . \text{Var}(Y)$$

But, recall also that, $\sum X^2 = \sum X - n\bar{X}^2$

$$= \frac{\sum X^2 - nX^2 + n\bar{X}^2}{n\sum X^2} . \text{Var}(Y)$$

$$= \frac{\sum X}{n\sum X^2} . \text{Var}(Y)$$

But $\text{Var}(Y) = \delta^2 = \frac{\sum e^2}{n-K}$

Therefore,

$$\text{Var}(\alpha) = \frac{\sum X}{n\sum X^2} . \frac{\sum e^2}{n-K}$$

Hence,

$$\text{Var}(\hat{\alpha}) = \frac{\sum X^2}{n\sum x^2} \cdot \frac{\sum e^2}{n-K}$$

These mean values and variances lead us to another aspect of econometrics

4.7 Test Of Significance

We can test hypothesis by

- **Simple Test: the rule of thumb or the standard error test.**

We compare $s(\hat{\alpha})$ and $s(\hat{\beta})$ with $\hat{\alpha}$ and $\hat{\beta}$. In terms of $s(\hat{\alpha}) \rightarrow \hat{\alpha}/2$ and $s(\beta) \rightarrow \beta/2$.

Our hypothesis

H_0: $\hat{\alpha}\ \hat{\beta} = 0$

H_1: $\hat{\alpha}\ \hat{\beta} \neq 0$

We accept H_0 if:

$\hat{\alpha}/2 < s(\hat{\alpha})$ or $s(\hat{\alpha}) > \alpha/2$

and

$\hat{\beta}/2 < s(\hat{\beta})$ or $s(\hat{\beta}) > \beta/2$

And conclude that and $\hat{\alpha}$ and β are statistically insignificant in explaining the relationship between X and Y.

We reject H_0 if:

$\hat{\alpha}/2 > s(\hat{\alpha})$ or $s(\hat{\alpha}) < \hat{\alpha}/2$

. And if,

$\hat{\beta}/2 > s(\hat{\beta})$ or $s(\hat{\beta}) < \hat{\beta}/2$

We conclude that $\hat{\alpha}$ and $\hat{\beta}$ are significant.

Here the t-statistic (calculated) is reduced to

$t_{cal}(\hat{\alpha}) = \hat{\alpha}/s(\alpha)$ note that's $(\hat{\alpha}) = \sqrt{var(\hat{\alpha})}$

And

$t_{cal}(\hat{\beta}) = \hat{\beta}/s(\hat{\beta})$

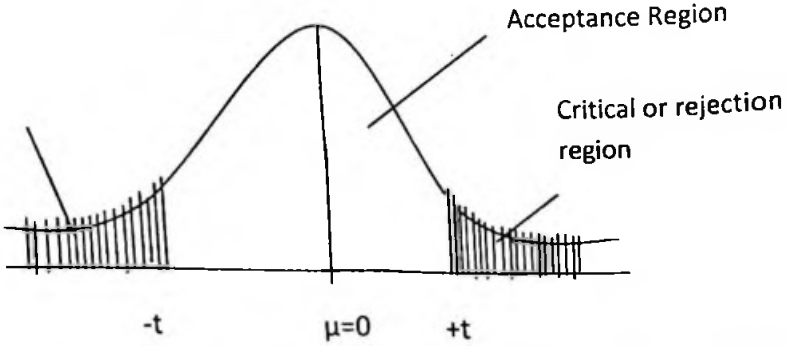
And the null hypothesis

$H_0: \hat{\alpha}\ \hat{\beta} = 0$

Is tested against the alternative hypothesis

$H_1: \hat{\alpha}\ \hat{\beta} \neq 0$

Standard error test is a two-tailed test graphically shown as



The calculated t is compared with the theoretical t which is found from the t-table at n-k degree of freedom
Where n = size of the sample

K = number of parameters

The confidence interval is usually 99% or 95% which is 1% or 5% significance level.

Since it is two-tailed lost 1% and 5% have to be divided by 2.

i.e. 1% = 1%/2 = 0.005

or 5% = 5%/2 = 0.025

i.e. $t_{0.005}$ and $t_{0.025}$ respectively.

- **The Decision Rule**
  According to rule of thumb, we
  - Reject H_0: $\hat{\alpha}$ $\hat{\beta}$ = 0 if $t_{cal} > t$
  - Accept H_0: α β = 0 if $t_{cal} < t$

4.8 The Gauss Markov Theorem

The Gauss Markov Theorem captures the basic properties of OLS estimates. The theorem states that given the assumption of the classical linear regression model, the least –square estimators, in the class of unbiased linear estimators, have minimum variance, that is, they are BLUE.

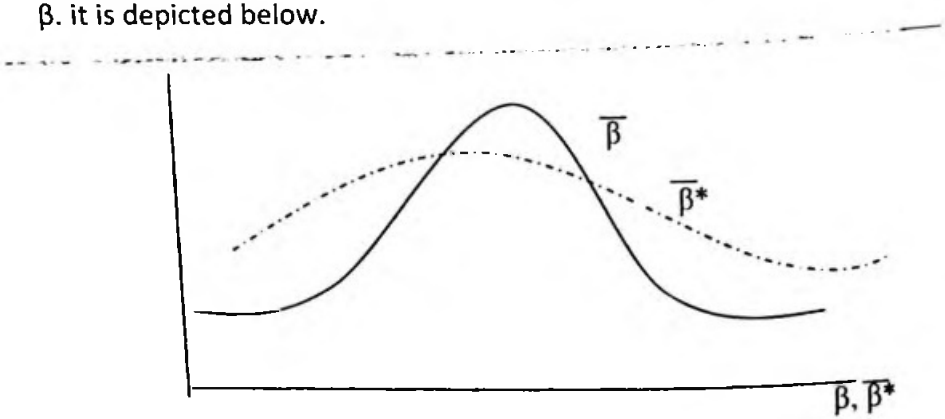
This means that estimators α, β, etc are said to be BLUE of α, β, etc if the following hold

- It is linear → linear function is a random variable, such as the dependent variable Y_t in the regression model.
- It is unbiased, that is, its average or expected value $E(\beta)$, is equal to the BLUE value of β
- It has minimum variance in the class of such linear unbiased estimators.

- unbiased estimators with the least variance is known as an efficient estimator

Note that

Minimum variance of β means that β is more likely to be closer to β. it is depicted below.



It shows that the distribution of β^* is more widespread around the mean (longer variance) than that of β.

An estimator is sufficient if it utilizes all the information a sample contains about the true parameters. An OLS estimator is also sufficient.

Alternatively, the assumptions or properties of the OLS can be stated as follows:

1. It is a conditional model such that $E(Y/X) = \alpha + \beta T + U$
   This means that the expected value of Y is conditioning upon the value of X
2. U's are iind $(0,\delta^2)$
   Where I = identitical
   I = independent
   Nd = normally distributed

Implying

$E(U_t) = 0$
$E(U_t)^2 = \delta^2$
$E(U_i U_j) = 0$

} "White Noise assumptions"

All of which have already been stated under the stochastic assumptions.

3. $E(U_i\ U_j) = 0$ This is also called the fixed regressor assumption, which means that U_t and X_t are correlated.

When these assumptions hold also, $\hat{\alpha}$ and $\hat{\beta}$ are said to be best, linear unbiased estimators (BLUE) as envisaged by Gauss Markov Theorem.

4.9 Summary of Ols Bivariate Model

Let us summarize the procedures of OLS estimation

Obtain α and β using:

$$\hat{\alpha} = Y_t - \hat{\beta}\bar{X}_t$$

$$\hat{\beta} = \Sigma x_t y_t / \Sigma x_t^2$$

- Obtain r^2 as

$$r^2 = \hat{\beta}\Sigma xy/\Sigma y^2 \quad = (\Sigma xy)^2/\Sigma x^2.\Sigma y^2$$

- Obtain Σe^2 as

$$\Sigma e^2 = \Sigma y^2 (1-r^2)$$

- Obtain δ^2 as

$$\delta^2 = \Sigma e^2 / n-K \text{ where } K = 2$$

- Obtain Standard errors of $\hat{\alpha}$ and $\hat{\beta}$ as

$$s(\hat{\alpha}) = \sqrt{\Sigma x^2/n\Sigma x^2 . \delta^2}$$

$$s(\hat{\beta}) = \sqrt{1/\Sigma x^2 . \delta^2}$$

- Obtain t calculated as

$t_{cal(\hat{\alpha})} = \hat{\alpha}/s(\hat{\alpha})$

$t_{cal(\hat{\beta})} = \hat{\beta}/s(\hat{\beta})$

Which are compared with t theoretical at a given lev
significance with

$H_0: \alpha \quad \beta = 0$

$H_1: \hat{\alpha} \quad \hat{\beta} \neq 0$

H_0 is accepted when $t_{cal(\hat{\alpha}\ \hat{\beta})} < t_{\bullet}$

H_0 is rejected when $t_{cal(\hat{\alpha}\ \hat{\beta})} < t_{\bullet}$

The following computational expressions are also useful

$\sum x^2 = \sum X^2 - n\bar{X}^2$

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

$\sum xy = \sum XY - \overline{nXY}$

Worked Example:4.1

Ex(1): Given the price and quantity (y) of Bread

Y = 3,5,3,4,6,8,8,6,9,10,11

X =2,4,3,5,5,7,6,4,7,9,8

Estimate the linear regression model of the above information and conduct the r^2 and standard error test

Solution

X	Y	X^2	Y^2	XY
2	3	4	9	6
4	5	16	25	20
3	3	9	9	9
5	4	25	16	20
5	6	25	36	30
7	8	49	64	56
6	8	36	64	48
4	6	16	36	24
7	9	49	81	63
9	10	81	100	90
8	11	64	121	88
60	73	374	561	454
ΣX	ΣY	ΣX^2	ΣY^2	ΣXY

Table 4:1 Price X and quantity supply of Bread

$\bar{X} = \Sigma X/n = 60/11 = 5.43$

$\bar{Y} = \Sigma Y/n = 73/11 = 6.64$

Recall that

$Y = \alpha + \beta X + U$

$\hat{Y} = \hat{\alpha} + \hat{\beta}\hat{X}$

$$\hat{\beta} = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$$

$$= \frac{459 - 11(5.45)(6.64)}{374 - 11(5.45)^2}$$

$$= 1.183$$

$$\hat{\beta} = 1.183$$

$$\hat{\alpha} = \bar{Y} + \hat{\beta}\bar{X}$$

$$= 6.64 - (1.183)(5.45)$$

Therefore,

$$\hat{\alpha} = 0.1913$$

The regression line is, therefore

$$\hat{Y} = 0.1913 + 1.18X$$

To find r^2

$$r^2 = (\Sigma xy)^2/\Sigma x^2.\Sigma y^2)$$

$$= \hat{\beta} \cdot \Sigma xy^2/\Sigma y^2$$

But $\Sigma xy = \Sigma XY - n\bar{X}\bar{Y} = 484 - 11(5.45)(6.69) = 55.93$

$$\Sigma y^2 = \Sigma y^2 - \Sigma \bar{Y}^2$$

$$= 561 - 11(6.64)^2 = 76.01$$

Therefore,

$$r^2 = \hat{\beta} \cdot \Sigma xy^2/\Sigma y^2$$

$$= 1.183 \times \frac{55.93}{76.01}$$

$$= 0.8706$$

Therefore $r^2 = 0.87$

This implies that has a strong power of explaining the behaviour of (Y) Qty ss, since 87% variation in Y is brought about by change in (X) price of bread, while 13% is due to erro term

To obtain Σe^2

$\Sigma e^2 = \Sigma y^2 (1 - r^2)$

$= 76.01 (1 - 0.87)$

Therefore $\Sigma e^2 = 9.88$

To find δ^2

$$\delta^2 = \frac{\Sigma e^2}{n-K} = \frac{9.88}{11-2} = 1.098$$

Therefore, $\delta^2 = 1.1$

The standard error

$$s(\alpha) = \sqrt{\Sigma X^2 / n\Sigma x^2 \ \ 5^2} = \sqrt{\frac{374}{(11 \times 47.27)} \times 1.098} = 0.89$$

$$s(\beta) = \hat{} \sqrt{1/\Sigma x^2 \cdot \delta^2} = \sqrt{1/47.27 \times 1.098} = 0.15$$

We can summarize the results as below

$\hat{Y} = 0.1913 + 1.18X$
(0.89) (0.15)
n = 11, k =2

On a prior I Bases a and β have conform to a prior I base because :hery are positively signed.

Testing at 99% confidence interval:

$$t_{cal(\hat{\alpha})} = \hat{\alpha}/s(\hat{\alpha}) = 0.1913/0.89 = 0.215$$

$$t_{cal(\hat{\beta})} = \hat{\beta}/s(\hat{\beta}) = 1.18/0.15 = 7.87$$

The degree of freedom is n-k = 11-2 = 9

From the t-table, $t_{0.005} = 3.25$

Now:

Reject H_0: $\hat{\alpha}, \beta \triangleq 0$

Accept H_0: $\hat{\alpha}, \hat{\beta} \neq 0$

We can see that $t_{cal(\hat{\alpha})} < t_{0.005}$

We therefore accept H_0 and conclude that α is statistically insignificant in explaining Y and

$t_{cal(\beta)} > t_{0.005}$

We therefore reject H_0 and conclude that β is significant in explaining Y

CHAPTER FIVE

OTHER ASPECTS IF BIVARIATE MODEL

5.1 Time as a Regressor; the Linear Trend Model

Most economic variables exhibit trend, that is, they are affect by the passage of time, or they change with tine. In this situation, time is an important variable. It has becomes regressor, that determine the expected value of the dependent variable.

When Y is regress on time, T, the model is called a **Linear Trend Model**, and the time variable T is known as the **Trend Variable**. The linear trend model can be formulated as

$Y_t = \alpha + \beta T - U_t$

As in the CEF $\hat{\beta}$ stands for the slope which determines the direction of the trend.

- If $\hat{\beta}$ is positive, there is an upward trend in Y then Y increases as time goes on
- If $\hat{\beta}$ is negative, there is a downward trend in Y then Y decreases as time goes on.

The time variable (t) can be used as describe time period → 1, 2, 3, 4, 5,.....n. It can also be used as years eg: 1990, 1991, 1992,....., 2010

What is important is that, in linear trend model, time is considered as the regressor – the main determinant or cause of change in the regressand Y.

Linear trend model is also like CEF, so we can write it as

$$E(Y_t/T) = \alpha + \beta T - U_t$$

Since time is a regressor.

The estimated linear trend model 1

$$\hat{Y}_t = \hat{\alpha} + \hat{\beta}\hat{T} + U_t$$

Where:

$$\hat{\alpha} = \overline{Y_t} - \beta\overline{T}$$

$$\beta = \sum ty/\sum t^2$$

The linear Trend Model is very useful in determining or finding the role of growth of certain economic variables such as

- Population
- GNP/GNP per capita
- Employment
- Productivity
- Money Supply
- Trade Surplus/Deficit
- Government expenditure, etc

5.2 Constant Growth Models/The Log-Lin Model

Non-Constant Growth Rule:

The linear Trend Model involves non-constant

For example, at period 1, Y will be Y_t. To find change in Y, we:

$\Delta Y = Y_t - Y_{t-1}$ as it is in difference equation.

But in $E(Y_t/T) = \alpha + \beta T - U_t$ We say that it represents the slope of the trend line, which shows the change in Y as T changes.

Then $\Delta Y = \beta = Y_t - Y_{t-1}$ which is the growth (positive or negative) in Y. The role of growth is given as:

$$\frac{Y_t - Y_{t-1}}{Y_t} = \frac{\Delta Y_t}{Y_t} = \frac{\Delta Y}{Y} = \frac{\beta}{Y}$$

Hence the rate of growth is $\frac{\beta}{Y}$ which implies that the growth in Y cannot be constant so:

(a) When $\beta > 0$ i.e. positive Y will be an increasing series, which will increases at a decreasing rate since the rate of growth will be declining as implied by the fraction β/Y.

(b) When $\beta < 0$ i.e. it is negative , Y will be a declining series and the rate of growth will be increasing very fast. Y will be declining at an increasing rate.

The linear trend model does not therefore exhibit constant growth rate.

- **Constant Growth Rate:**

Suppose we want to find out the growth rate of productivity in firm A for the period 1990 – 2010.

Y_t the productivity in ('000 N)

t the time from 1990 – 2010

We can use the compound interest formula to determine this

Recall: $A = P(1+r)^n$

Where A = TV (trend value) = Amount

n = years

r = rate of interest

P = initial deposit/principal

We can substitute these variables with our concerned variables

$A = Y_t$ = productivity

$P = Y_o$ = Initial productivity

n = T = the time

r = g = the compound (overtime) rate of growth of Y

Our model becomes: $Y_t = Y_o (1+g)^T$

We can see that the model is non-linear, as it is raised to the power of T.

To transform it to linear, we take the natural logarithm of both sides

$$\ln Y_t = \ln Y_o + T\ln(1+g)$$

Recall that in log, $\log_c A^b$ can be written as:

$$\log_c A^b = b.\log_c A$$

Therefore, $\ln Y_t = \ln Y_o + T.\ln(1+g)$

This is a linear model, which can also be reduced to

$$\ln Y_t = \alpha + \beta T - U_t$$

By assuming or letting

$$\ln Y_o = \alpha$$

$$\ln(1+g) = \beta$$

Therefore

$$\ln Y_t = \alpha + \beta T$$

Since the compound interest formula will not hold exactly, we add the error term.

Therefore,

$$\ln Y_t = \alpha + \beta T - U_t$$

Which is like any other linear regression model such that the parameters α and β are linear.

The difference between this model and other linear model is that the regressant is the logarithm of Y and the regressor is time which take the values, 1,2, 3, ... etc.
This model is called semi-log model or log-lin model. The model gives a constant proportional or relative change in Y for a given absolute change in the regresor T, which is measured by the slope β.
Therefore, β = relative change in regressand/absolute change in regressor

$$\beta = \frac{Y_t - Y_{t-1} / Y_{t-1}}{T_t - T_{t-1}}$$

$$\beta = \frac{Y_t - Y_{t-1} / Y_{t-1}}{T_t - T_{t-1}}$$

This is the relative (growth rate) change in Y for an absolute change in T.
Note that, β x 100 = growth rate in Y = Semi-elasticity of Y with respect to T.

- **Types of Growth Rate**
  - Instantaneous rate of growth = the rate of growth in Y at a period in time. It is given by the slope, β
  - Compound rate of growth = the rate of growth of Y over a period of time.

  It is given: $e^{\beta} - 1$

  i.e. (antilog β-1) x 100

  This is found from our compound rate of growth, g, in our original representation.

Recall that, $\beta = \ln(1+g)$

So, $1+g = -\beta.1/\ln$

$$g = \frac{\beta.1}{\ln -1}$$

Therefore, $g = e^{\beta} - 1$, which is the compound rate of growth

Computational Procedure

Given the model:

$$\ln Y_t = \alpha + \beta T$$

$$\hat{\alpha} = \ln\overline{Y_t} - \hat{\beta}\overline{T}$$

$$\hat{\ln Y_t} = \sum \ln Y_t / n$$

$$\beta = \frac{\sum y_t}{\sum t^2}$$

Example:5.1

From the quarterly expenditure on services of a particular country for six years, the following growth model is estimated

$$\ln Y_t = 3.2890 + 0.00345t$$

$$(0.0033)\ (0.00012) \qquad r^2 = 0.99$$

Note that:

- Y = expenditure on service
- The period is between 1993:1 1.

1993:3

I.e. From the first quarter of 1994, to the third quarter of 1995.

We can give the following interpretations

$$*\beta = 00345$$

The instantaneous rate of growth is (0.00345) x100 = 0.345. This means that between the period 1994:1 to 1995:3, there has been a quarterly increase in expenditure on service by 0.345 percent.

Yearly, there is roughly a (3.2840 x 4) = 138 percent increase in expenditure on services. This value is elastically significant.

The intercept 3.2840 gives the beginning value of service expenditure (at the end of the fourth quarter of 1992) when we take its antilog, since

Log of Y = 3.2890 at the period.

Therefore, antilog of 3.2890 = 2413.90 at the beginning of 1993 is 2413.90 million or billion – depending on the unit.

The compound rate of growth is

$$e^{0.0345} - 1$$

$$= (\text{Antilog } 0.00345 - 1) \times 100$$

$$= 0.00345 \times 100$$

$$= 0.345$$

Hence, the quarterly compound rate of growth in expenditure on services is 0.345 percent. This value is slightly greater than the instantaneous rate of growth because of the compounding effect.

To differentiate the log-lin model from the linear trend model, we gave the estimated linear trend model of the same data (without using log) as follows:

$Y_t = 2405.848 + 19.6920t$

$r^2 = 0.9843$

This is a linear model;

The interpretation is as follows:

the slope 19.6920 implies that over the period 1994:1 to 1995:3, on average, there is an absolute increase in the expenditure on service (absolute rate of growth) of 19.69 (approx. 20) million or billion per quarter. There was an upward trend, thus, in the expenditure on services.

Hence the main difference between the log-lin model and linear model is that: log-lin gives the relative rate of growth.

Example 5.1

The per capita incomes of Nigeria between 1996 to 2005 (hypothetical) are given below.

Yt:	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
PCI:	484	504	520	560	591	632	685	750	794	866

Estimate the log-lin and linear trend models and interpret them.

Solution

T	Y	Y^2	T^2	YT	LnY	lnY^2	lnYT
1	484	234356	1	484	6.182	38.218	6.182
2	504	254016	4	1008	6.223	38.720	12.445
3	520	270400	9	1560	6.254	39.100	18.761
4	560	313600	16	2240	6.328	40.043	25.312
5	591	349281	25	2955	6.382	40.728	31.909
6	632	399424	36	3792	6.449	41.588	38.693
7	685	469225	49	4795	6.529	42.633	45.706
8	750	562500	84	6000	6.620	43.825	52.961
9	794	630436	81	7146	6.677	44.583	60.639
10	866	749956	100	8660	6.764	45.750	67.639
55	6,386	4,233,094	385	38,640	64,408	415.198	359.702

Table 5.1: Log lin transformation table

$\sum T = 55$ $\sum YT = 38640$ $\sum lnYT = 359.702$

$\sum Y = 6386$ $\sum lnY = 64.408$ $\sum Y^2 = 4,233,094$

$\sum lnY^2 = 415.198$

Linear Trend Model

$$\hat{Y} = \hat{\alpha} + \hat{\beta}\bar{T}$$

$$\hat{\beta} = \frac{\sum YT - n\bar{Y}\bar{T}}{\sum Y^2 - n\bar{T}} = \frac{\sum Yt}{\sum t^2}$$

$$= \frac{38640 - (10)(6386/10)\,(55/10)}{385 - 10(55/10)^2}$$

$$\hat{\beta} = \frac{38640 - (10)(638.6)\,(5.5)}{385 - 10(5.5)^2}$$

$$= \frac{38640 - 33123}{385 - 302.5} \qquad = \frac{3517}{82.5}$$

$$\hat{\beta} = 42.63$$

$$\hat{\alpha} = \hat{Y} + \hat{\beta}\bar{T}$$

$$\hat{\alpha} = 638.6 - (42.63)(5.5)$$

$$= 638.6 - 234.467$$

$$= 404.133$$

$$\approx 404$$

Therefore, the estimated regression line is

$$\hat{Y} = 404.133 + 42.63T$$

To determine the r^2

$$r^2 = \frac{\beta . \Sigma yt}{\Sigma y^2}$$

$\Sigma yt = 3517$ above

$\Sigma y^2 = \Sigma y^2 - n\bar{Y}^2$

$$= 4{,}233094 - (10)(638.6)^2$$

$$= 154994.4\ (1\text{-}0.967)$$

$$= 5063.62$$

$\delta^2 = \Sigma e^2/n-K = 5063.62/10-2 = 5063.62/8$

$s(\hat{\alpha}) = \sqrt{\Sigma T^2/\Sigma t^2 \cdot \delta^2}$

$s(\hat{\alpha}) = \sqrt{\frac{385}{10(82.5)} \times 632.9525}$

$\delta^2 = \Sigma e^2/n-K = 5063.62/10-2 = 5063.62/8$

$s(\hat{\alpha}) = \sqrt{295.3778} = 17.187$

$s(\hat{\beta}) = \sqrt{1/\Sigma t^2 \cdot \delta}$

$= \sqrt{1/82.5 \times 632.9525} = \sqrt{7.672}$

$s(\hat{\beta}) = 2.77$

$t_{cal}\ (\hat{\beta})/s(\hat{\beta}) = 15.37$

$t_{cal}\ (\hat{\alpha})/s(\hat{\alpha}) = 404.133/17.187 = 23.514$

We can summarize our results as follows:

$\hat{Y}_t = 404.133 + 42.63T$

s.e. = (17.187) (2.77)

t = (23.514) (15.39) $r^2 = 0.969$

At 95% our theoretical = 2.306 (8d.f)

Hence, both $\hat{\alpha}$ and $\hat{\beta}$ are statistically significant.

- The intercept 404.133 implies that the per capita income in 1995 was 409.133

- The slope, β = 42.63 implies that on average there is a yearly absolute growth in per capita income of 42.63
- r^2 = 0.967 implies that time explains about 97% variation in per capita income.

Hence, there was an upward trend in the per capita income.

LOG-LIN MODEL:

$\ln\hat{Y}_t = \hat{\alpha} + \hat{\beta}\hat{T}$

$$\hat{\beta} = \frac{\Sigma \ln YT - n\bar{T}\ln\bar{Y}}{\Sigma T^2 - nT^2}$$

$$= \frac{359.702 - (10)(5.5)(64.408/10)}{385 - 10(5.5)^2}$$

$$= \frac{359.702 - 534.244}{82.5}$$

$$= \frac{5.458}{82.5}$$

$$= 0.0662$$

$$\hat{\alpha} = \ln\hat{Y}_t + \hat{\beta}\hat{T}$$

$$= 6.4408 - (0.0662)(5.5)$$

$$= 6.4408 - 0.364$$

$$\hat{\alpha} = 6.077$$

$$\Sigma \ln Y^2 = \ln Y_t^2 - \ln\bar{Y}_t^2$$

$$= 415.198 - (10)(6.4408)^2$$

= 418.198 – 414.839

= 0.389

Therefore $r^2 = \hat{\beta}. \frac{(\Sigma \ln Yt)}{\Sigma \ln Y^2}$

$= 0.0662 \times \frac{5.458}{0.369}$

$= \frac{0.3613}{0.369}$

= 0.199

The model is:

$\ln Y_t = 6.077 + 0.66T$

- The antilog of the intercept gives the per capita income in 1993 i.e.
  Antilog 6.077 = 435.12
- The slope B gives the instance rate of growth
  i.e. there is annual rate of growth of
  0.0662 x 100 = 6.62%.
  The compound rate of growth is
  Antilog 0.0662 – 1
  = 1.0684 – 1
  = 0.0684 x 100
  = 6.84% per annum

5.3 The Lin-Log Model

In the log-lin model, we find the percentage growth in Y for an absolute change in X. In the lin-log model, we find the absolute

change in Y for a percentage change in X, in this case, the regressand Y is not in log, but our regresor X is in Log.

$$Y_t = \alpha + \beta \ln X + U$$

The slope β gives the absolute change in Y for a percentage change in X. Therefore, β can be written

$$\beta = \frac{\text{Change in Y}}{\text{Relative change in X}}$$

$$\beta = \frac{\Delta Y}{\Delta X/X} = \frac{Y_t - Y_{t-1}}{X_t - X_{t-1}/ X_{t-1s}}$$

In making interpretation β is divided by 100% just as it is multiplied by it in log-lin model.

Again, the absolute change in Y (ΔY) is β.(ΔX/X).

Note that our regresor X is no longer time. The lin-log model is commonly used in estimating the relationship between a particular type of expenditure and the total (whole) expenditure. For example, it is used in the Engel expenditure model to estimate or show that the total expenditure that is devoted to food tends to increase in (AP) Arithematic progestion as total expenditure increases in (G.P)Geometric progestion.

5.4 The Log Linear Model: The Log – Log Model

A model in which both the regressor and regressand are in log is called double-log, log – log, log-linear or constant elasticity model.

The original model is the exponential regression model of the form

$$Y_t = \alpha X_t^{\beta} e^{ut}$$

This can be written in log-linear form by taking both sides with ln.

$$\ln Y_t = \ln \alpha X_t^{\beta} e^{ut}$$

$$\ln Y_t = \ln \alpha + \beta \ln X_t^{\beta} + e^{ut}$$

$$\ln Y_t = \ln \alpha + \beta \ln X_t + e^{ut}$$

$$\ln Y_t = \ln \alpha + \beta \ln X_t + U_t(1)$$

Therefore,

$$\ln Y_t = \ln \alpha + \beta \ln X_t + U$$

Taking $\ln \alpha$ as α^* we have

$$\ln Y_t = \alpha^* + \beta \ln X_t + U$$

To estimate the regression line, we can set

$$\alpha^* = \alpha$$

$$\ln Y_t = Y^*$$

$$\ln X_t = X^*$$

So that we have

$$Y_t^* = \alpha + \beta X^* + U$$

This can be estimated using OLS method, since the model is linear in the parameters.

α and β are unbiased.

The slope β, measures the elasticity of Y with respect to X. accordingly, the elasticity coefficient is constant. i.e. change in

lnY per unit. Change in lnX remains the same no matter at which lnX we measure it.

Note that the simplest way to determine whether bivariate model fits log-log model is to plot the scatter diagram of lnY against lnX and see if the graph is approximately a straight line.

Example:5.3 Given the following data, estimate the demand function

$$Y = \alpha X^{\beta} e^{u}$$

Where Y = demand for rice

X = price of rice in arbitrary units

Y	X	LnY	lnX
543	61	6.2971	4.1109
580	54	6.3631	3.0890
618	50	6.7265	3.9120
695	43	6.5939	3.7612
724	38	6.5848	3.6376
812	36	6:6995	3.5835
887	28	6.7879	3.3322
991	23	6.8987	3.1353
1186	19	7.0685	2.9445
1940	1-	7.5785	2.3026

Table 5.2: log – log Transformation table

Using OLS estimation, we have estimated the model as follows:

$$Y = \alpha X^{\beta} e^{u}$$

$$Log_e Y = Log_e \alpha + Log_e \beta + U$$

Or

$$\ln Y = \ln \alpha + \beta \ln X$$

Setting $\ln \alpha = \theta$

$$\ln Y = \theta + \beta \ln X$$

$$\ln Y = \hat{\theta} + \hat{\beta} \ln X$$

From our estimation we got

$$Log_e Y = 9.121 - 0.69 \log X \quad \text{or}$$

$$\ln Y = 9.121 - 0.69 \ln x$$

$$(0.07)\ (0.02)\ r = 0.992$$

Writing the model in exponential regression form we have:

$$Y = \alpha X^{\beta} \quad \text{or}$$

$$Q = \alpha P$$

$$Q = \alpha P^{-0.69}$$

Where $\beta = -0.69$ which is the elasticity of demand. The demand for rice is inelastic.

The intercept $\alpha = \text{Antilog}\theta$

$$= \text{Antilog}9.121$$

=9145.34

≈9145

Meaning that at zero prices the consumers are willing to demand for 9145 units of rice.

Note: the simple interpretation (non technical) is that 1 units change (increase) in price of rice will lead to 0.69 decreases in the quantity of rice demanded.

CHAPTER SIX

6.1 MULTIPLE REGRESSIONS

In the real world, the dependent variable can be determine by two or more variable for instance consumption is determined not only by income alone but by other variable such as taste, weather and so on therefore to determine the influence of these variables on consumption a researcher may have to estimate a partial regression coefficient. Economic theory states that consumption is a function of income, taste, etc.

$$Y = f(X_1, X_2, \ldots X_n)$$

Where

Y = consumption

X_1= income

X_2 = Taste

Mathematical economist will say that

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

This relationship means that, the variations in consumption Y are fully explained by changes in Income X_1 and taste X_2.

An econometrician will go further than the relationship because there are other variables that may influence consumption that are not specified in this equation. To take care of such variables, therefore, we introduce the random error term.

$$Y = (\beta_0 + \beta_1 X_1 + \beta_2 X_2) + U$$

Systematic Component — random components

Where β_1, β_2 are partial regression coefficients while the intercept β_0 gives the mean or average effect on Y of the entire variable excluded from the model. Although, its methematically interpreted as the average value of Y, when X_1 and X_2 are set equal to zero.

To make this model more meaningful it is important to operate within the frame work of the assumptions underlying the estimation of the parameters involve. Most especially

- The randomness of U
- The zero means value of U i.e. E (Ui/X1i X2i) = 0 for each I
  No serial correlation i.e. cov (Ui, Uj) = 0 $i \neq j$
  Homoscedasticity i.e. var (Ui) = δ^2

Normality of U i.e. Ui NN $(0_1 \delta^2)$

- Zero covariance between Ui and each Xi variable. i.e cov (Ui X1i) = cov (Ui, X2i) =0
- No specification bias
- No multicollinearity
- Linearity in parameter assumption

Having specify the model, the next thing is to use sample observation of Y, X_1 and X_2 to estimate the true parameter β_0, β_1, β_2

$$Y= \hat{\beta}_0 + \hat{\beta}_1\hat{X}_1 + \hat{\beta}_2X_2 + U$$

Where $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ are estimates of the true parameters β_0, β_1, β_2 of the consumption relationship.

These estimate or coefficient can be obtained by minimizing the sum of squared residuals. i.e $\hat{\Sigma}\hat{e}_i^2 = \Sigma(Y_i - \hat{Y}_i)^2 = \Sigma(Y - \beta_0 - \beta_1X_1 - \beta_2X_2)^2$

We obtain by

1. Finding the error, (in our model)
   $U_1 = Y - \hat{\beta_0} - \beta_1\hat{X}_1 - \beta_2X_2$
2. Squaring the error
   $U_i^2 = (Y - \beta_0 - \beta_1X_1 - \beta_2X_2)^2$
3. Summing the squared error
   $\Sigma U_i^2 = \Sigma (Y - \beta_0 - \beta_1X_1 - \beta_2X_2)^2$
4. To minimize we take partial derivation in respect of β_0, β_1 and β_2 respectively so as to get our normal equation and subsequently, the coefficient β_0, β_1 and β_2. i.e take partial derivation and equate it to zero.

   $$\frac{d\Sigma Ui^2}{d\beta_0} = (-1)(2)\,\Sigma(Y - \beta_0 - \beta_1X_1 - \beta_2X_2) = 0$$

   $$= -2\,\Sigma\,(Y - \beta_0 - \beta_1X_1 - \beta_2X_2) = 0$$

   Divide through by -2

   $$= \frac{-2\,\Sigma\,(Y - \beta_0 - \beta_1X_1 - \beta_2X_2)}{-2} = \frac{0}{-2}$$

   $$= \Sigma\,(Y - \beta_0 - \beta_1X_1 - \beta_2X_2) = 0$$

Where $\frac{0}{-2} = 0$

Open the bracket

$$= \Sigma Y - \Sigma\beta_0 - \Sigma\beta_1 X_1 - \Sigma\beta_2 X_2 = 0$$

$$= \Sigma Y - N\beta_0 - \beta_1\Sigma X_1 - \beta_2\Sigma X_2 = 0 \quad \ldots\ldots (1)$$

Where $\Sigma\beta_0 = \beta_0 + \beta_0 + \beta_0 + \ldots\ldots = N\beta_0$ ie Number of $\beta_0 \times \beta_0$

To find our β_0, we multiply through by N and make β_0 the subject of the formula.

$$\frac{\Sigma Y}{N} = \frac{N\beta_0}{N} + \frac{\beta_1 \Sigma X_1}{N} + \frac{\beta_2 \Sigma X_2}{N}$$

$$\bar{Y} = \beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2$$

Therefore

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2$$

To get the 2nd normal equation we take partial derivation in respect of the β_1

$$\frac{d\Sigma Ui^2}{d\beta_1} = (-X_1)(2)\ \Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$$

$$= (-2)\ \Sigma\ (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$$

To eliminate -2 we divide through by -2

$$\frac{(-2X_1)\Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)}{-2} = \frac{0}{-2}$$

$$X1\Sigma\ (Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = 0$$

$$\Sigma X_1 Y - \beta_0\Sigma X_1 - \beta_1\Sigma\ X_1^2 - \beta_2\Sigma X_2 X_1 = 0$$

$$\Sigma X_1 Y - \beta_0 \Sigma X_1 - \beta_1 \Sigma X_1^2 - \beta_2 \Sigma X_2 X_1 \quad ....(2)$$

To determine the third normal equation differentiate $\Sigma U1^2$ in respect of β_2

$$\frac{d\Sigma U^2}{d\beta_2} = (-X_2)(2)\,\Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 \Sigma X_2) = 0$$

$$= (-2\,X_2)\Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 \Sigma X_2) = 0$$

Divide through by -2

$$\frac{= (-2X_2)\,\Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 \Sigma X_2)}{-2} = \frac{0}{-2}$$

$$X_2\,\Sigma(Y - \beta_0 - \beta_1 X_1 - \beta_2 \Sigma X_2) = 0$$

$$\Sigma X_2 Y - \beta_0 \Sigma X_2 - \beta_1 \Sigma X_1 X_2 - \beta_2 \Sigma X_2^2 = 0$$

$$\Sigma X_2 Y = \beta_0 \Sigma X_2 + \beta_1 \Sigma X_1 X_2 + \beta_2 \Sigma X_2^2$$

$$\beta_0 = \Sigma Y = N\beta_0 + \beta_1 \Sigma X_1 + \beta_2 \Sigma X_1 \quad ...(1)$$

$$\beta_1 = \Sigma X_1 Y = \beta_0 \Sigma X_1 + \beta_1 \Sigma X_1^2 + \beta_2 \Sigma X_1 X_2 \quad ...(2)$$

$$\beta_2 = \Sigma X_2 Y = \beta_1 \Sigma X_2 + \beta_1 \Sigma X_1 X_2 + \Sigma X_2^2 \quad ...(3)$$

These equations are called normal equations and are used to obtain the coefficient β_0, β_1, and β_2 in other words they are used to fit the regression line. To do that we use crammer rule matrix as follows

$$\begin{bmatrix} A \end{bmatrix} = \begin{pmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \end{pmatrix} = \begin{pmatrix} N & \Sigma X_1 & \Sigma X_1 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 \\ \Sigma X_2 & X_1 X_2 & \Sigma X_2^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\begin{bmatrix} A \end{bmatrix} = \Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2$$

To determine β_1 we can replace the middle column with the sums on the right hand of the equations and divide by the determinants

$$\hat{\beta}_1 = \frac{\begin{pmatrix} N & \Sigma Y & \Sigma X_1 \\ \Sigma X_1 & \Sigma X_1 Y & \Sigma X_1 X_2 \\ \Sigma X_2 & \Sigma X_2 Y & \Sigma X_2^2 \end{pmatrix}}{\Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2}$$

$$= \frac{\Sigma X_1^2 Y (\Sigma X_2^2) - \Sigma X_1^2 Y (\Sigma X_1 X_2)}{\Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2}$$

$$\hat{B}_2 = \frac{\begin{pmatrix} N & \Sigma X_1 & \Sigma Y \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 Y \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2 Y \end{pmatrix}}{\Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2}$$

$$= \frac{\Sigma X_2^2 Y (\Sigma X_1^2) - \Sigma X_2^2 Y (\Sigma X_1 X_2)}{\Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2}$$

Given the above OLS estimates, we need to also obtain the standard error of these estimates which are important to establish confidence interval and to test the statistical hypothesis. The relevant formulas are as follows

$$\text{Var}(\hat{\beta}_0) = [1/n + X_1^2 \Sigma X_2^2 + X_2^2 \Sigma \bar{X}_1^2 - 2X_1X_2 \Sigma \bar{X}_1 \bar{X}_2]\theta U^2 \over \Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_2)^2$$

$$\text{Se}(\hat{\beta}_0) = + \sqrt{\text{var}(\beta 0)}$$

$$\text{Var}(\beta_1) = \frac{\Sigma x_2^2 \quad OU_1^2}{\Sigma x_1^2 \Sigma x_2^2 - (\Sigma x_1 x_2)}$$

Or equivalently

$$(\beta_1) = \frac{OUi^2}{\Sigma X_1^2 (1-r^2)}$$

$$\text{Se}(\beta_1) = + \sqrt{\text{var}(\beta 1)}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\Sigma X_1^2}{\Sigma X_1^2 \Sigma X_2^2 - (\Sigma X_1 X_1)^2} . \theta Ui^2$$

Or equivalently

$$\text{Var}(\hat{\beta}_2) = \frac{\theta Ui^2}{\Sigma x_2^2 (1 - r^2)}$$

$$\text{Se}(\beta_2) = + \sqrt{\text{var}(\beta 1)}$$

and it is given by

$\theta u^2 = \frac{\Sigma Ui^2}{n-k} = \frac{\Sigma Ui^2}{n-k}$ while the covariance (β_1, β_2)

$$= \frac{-r23\theta^2}{\sqrt{(1-r^2)\ X_1^2\ X_2^2}}$$

Where, r23 = correlation between X_1 and X_2.

6.2 The Multiple Coefficient Of DeterminatioN.

As in the two variables case the R^2 shows the percentage of total variable of $\hat{Y}$ that is explained by the regression plane. The notation r^2 is extended to the two variable models while in three variable model we would like to know the proportion of the variation in Y that is explained by variable X_1 and X_2 jointly and therefore we may extend the capital letter R^2 with sub script, of the variables whose relationship is being studied. In the three or more variable model, the square multiple correlation coefficient (which is another name for R^2) is denoted by $R^2_{Y\,x_1x_2}$. It measures the goodness of fit of the regression equation. To derive $R^2_{Y\,x_1x_2}$, we establish in the previous chapter that TSS = ESS + RSS

i.e. $$\frac{\Sigma y^2}{\Sigma y^2} = \frac{\Sigma y^2}{\Sigma y^2} + \frac{\Sigma ei^2}{\Sigma y^2}$$

$$1 = R^2 + \frac{\Sigma e_i^2}{\Sigma y^2}$$

Therefore $R^2 yx_1x_2 = \frac{\Sigma \hat{y}^2}{\Sigma y^2} = \frac{\Sigma(\bar{y} - \hat{y})^2}{\Sigma(y - \bar{y})^2} = 1 - \frac{\Sigma ei^2}{\Sigma y^2}$

$$= \frac{\Sigma y^2 - \Sigma e_i^2}{\Sigma y2}$$

We have also established that

$e_i = y - \hat{y}$ and

$$y = \beta_1 X_1 + \beta_2 X_2$$

The square residuals are

$\Sigma e_i^2 = \Sigma e_i \ (y - \hat{y})$ Since $\Sigma ei^2 = \Sigma eiy$

$= \Sigma e_i \ (yi - \beta_1 X_1 + \beta_2 X_2)$

$= \Sigma e_i yi - \beta_1 \Sigma e_i x_1 + \beta_2 \Sigma e_i x_2$

$\Sigma e_i^2 = \Sigma e_i yi$

Since, $\Sigma eiyi = 0$, $\Sigma eix_2 = 0$ and $\Sigma ei = yi - yi$

Therefore, $\Sigma ei^2 = \Sigma(y - y)y$

$= \Sigma y \ (y - \beta_1 X_1 + \beta_2 X_2)$

Where $y = \beta_1 X_1 + \beta_2 X_2$

When we open the bracket, we have

$$\Sigma ei^2 = \Sigma yi^2 - \beta_1 \Sigma yx_1 + \beta_2 \Sigma yx_2$$

Therefore, $Ryx_1x_2, = \frac{\Sigma yi^2 - (\Sigma yi^2 - \beta_1 \Sigma yx_1 + \beta_2 \Sigma yx_2)}{\Sigma y2}$

$$= \frac{\beta_1 \Sigma yx_1 + \beta_2 \Sigma yx_2}{\Sigma y2}$$

Note that R^2 like the r^2 lies between zero and 1. That is $0<R^2<1$ If it is =1, the fitted regression line explain 100 percent of the variability in Y. if it is zero, the model did not explain any of about the variation. However, the fit of the model is better when R closer to 1.

If $R^2 = 0.9$ for instance, it is closer to 1, but this value must be tested to establish its statistical significance. For example we first of all set our own hypothesis

$H_0: R^2 = 0$

$H_1: R^2 \neq 0$

At $\alpha = \alpha_0$

The statistical for testing R^2 is the F statistics given by

$$F^* = \frac{R^2/k-1}{(1-R^2)/n-k}$$

This is another way of testing the overall significance of the model.

Where R^2 = the variability in Y that is explained by the model.

$1 - R^2$ = the proportion or that fraction of the variation that is not explain by model.

That is R^2 = explained proportion %

$1 - R^2$ = unexplained proportion %

Not also that

$R^2 \sim X^2$ with $K - 1$ degree of freedom

$1 - R^2 \sim X^2$ with $n - k$ degree of freedom

This means that R^2 tends towards or is equivalent to chi-square with k – 1 degree of freedom

Decision rule

If $F_c > F_T$ reject H_0

If $F_c < F_T$ accept H_0

Note that, if the purpose of estimating your model is for policy analysis or decision making, the parameters becomes the focus of the model. If the parameters are significant, no matter the value of R^2, the relevance of the model is justified.

However, if the purpose of the model is to forecast or for prediction, what is important is the R^2, not the parameters. Therefore, the value or importance of your estimated model depends on the problem it inttend to solve.

Worked example 6.1

Given that model $Y = B_0 + B_1X_1 + B_2X_2 + U$, supposing $R^2 = 0.95$, n = 10 $\alpha = 0.05$, test the significance of R^2.

Solution

$H_0 : R^2 = 0$

$H_1 : R^2 > 0$

$$Fc = \frac{R^2/(K-1)}{(1-12^2)/n-k} \quad \text{where } v_1 = K-1 = 3-1 = 2, \; v_2 = n-k = 10-3 = 7$$

$$= \frac{0.95/3-1}{(1-0.95)/10-3}$$

$$= \frac{0.95/2}{0.05/7}$$

$$= \frac{0.48}{0.007}$$

$$= 68.57$$

Decision rule

If $Fc > F_T$ rejects H_0 otherwise accept H_1

since Fc 68.57 > F_T 3.46 we reject H_0 and conclude that R^2 is significantly different from zero.

The adjusted $\bar{R}^2$

R^2 vary directly with increase in the number of explanatory variable. That is to say that R^2 increases a the number of explanatory variable increases. R^2 statistic never decrease. If a researcher increase the quantity of his model by simply adding additional regression. The value of the numerator of the R^2 statistics is increased by that amount while the denominator remain the same.

R^2 can be adjusted by the value of degree of freedom to connect the likely defect. The adjusted coefficient of determination is denoted by R^2 with a bar, i.e

$$\bar{R}^2 = 1 - \frac{\Sigma ei/n - k}{\Sigma y^2 i/(n-1)}$$

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

From the above it is apparent that $R^2 < R^2$ if k · 1 which implies that, if the X variables increase, the adjusted R^2 increase less than R^2

Therefore, if K = 1, then ($R^2 = \bar{R}^2$) the coefficient of determination will be equal to the adjusted coefficient of determination.

For example since

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

If k = 1 then

$$\bar{R}^2 = 1 - (1 - R^2)$$

$$= 1 - (1 - R^2) \text{ by opening the bracket}$$

Therefore, $\bar{R}^2 = R^2$

For researcher to compare two models on the bases of their coefficient of determination, the sample size and the dependent variable must be the same, no matter the form of the explanatory variable and whether it is adjusted or not.

For example, if you have estimated two models one about consumption and the other about production as follows

$Yi = 0.25 + 0.67X_1 + 0.26X_2$ with $R^2 = 0.86$

$Y2 = 0.12 + 0.57X_1 + 0.34X_2$ with $R^2 = 0.95$

As long as the dependent are not the same as above you cannot compare R^2 but, when you have a dependent variable that are the same, then R^2 can be compared. That is comparism becomes relevant when the dependent variables are the same.

e.g. $Y_1 = 0.25 + 0.67X_1 + 0.26X_2 \quad R^2 = 0.86$

$Y_2 = 0.12 + 0.57X_1 + 0.34X_2 \quad R^2 = 0.95$

6.3 Hypothesis Testing In Multiple Regressions

To adequately analyze a multiple regression estimates a researcher is expected to test the following

(1) Test hypothesis about the individual regression estimates
(2) Test the overall significance of the estimated regression model, to find out if all the slopes are simultaneously equal to or greater than, or at least one of them is less than or equal to zero.
(3) Test that two or more coefficients are equal to one another
(4) Test that the partial regression coefficients satisfy certain restrictions.
(5) Test for the stability of the regression model overtime or in different cross sectional units
(6) Test the functional function of regressing the model

6.4 Hypothesis Testing About Individual Regression Coefficients

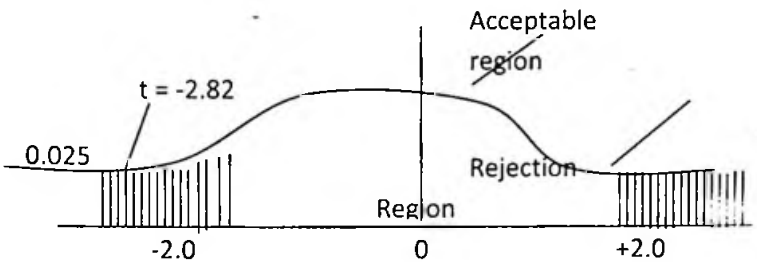
Given the normality assumption, we can use the t – statistic to test a hypothesis about any individual partial regression coefficient. To do so, a researcher must set his hypothesis

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Choose the level of significance. That is either to0.5 or to.025

Choose critical region depending on whether the test is a two tailed or one tailed test. For instance if our H_1 says $\beta_1 \neq 0$, then it is a two tailed. If it is either > or <, it is a one tailed test.



After that the researcher determine the value of the t- statistic which is given by

$$t^* = \frac{\beta_1 - \beta_1}{\hat{\delta\beta_1}} = \frac{\beta_1}{\hat{\delta\beta_1}}$$

The researcher will then compare the value of t* and the table value of t. if t*> it, we reject H_0 otherwise we accept H_0.

6.5 Test for Parameter Restriction

Another type of test is the test for parameter restriction which is different from the test of significance of the partial regression coefficients. In this type of test the hypothesis are set as follow

$H_0: \beta_1 + \beta_2 = 0$

$H_1: \beta_1 + \beta_2 \neq 1$

There are occasions where economic theory may suggest that the coefficients in a regression model satisfy some linear equality restriction. For instance, consider the following equation

$$Y = \beta_0 X_1^{\beta 1} X_2^{\beta 2} e^{Ui}$$

Where Y = output, X_i = labour inputs, X_2 = capital inputs, if the above equation is transform using log, we have

$$\ln Y = \alpha_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + e^{Ui}$$

Where $\alpha = \ln \beta_0$

Hence if there are constant return to scale (equiproportional change in output for an equiproportionate change in the inputs) economic theory hypothesis would suggest that

$$\beta_1 + \beta_2 = 1$$

This $\beta_1 + \beta_2$ is an example of a linear equality restriction.

This hypothesis can be tested by t statistic

$$to = \frac{(\hat{\beta}_1 + \hat{\beta}_2) - (\beta_1 + \beta_2)}{Se\,(\beta_1 + \beta_2)}$$

$$\text{Then } t^{+} = \frac{(\hat{\beta}_1 + \hat{\beta}_2) - 1}{\sqrt{ar(\hat{\beta}_1) + var(\hat{\beta}_2) + 2\,cov\,(\hat{\beta}_1 + \hat{\beta}_2)}}$$

Since $(\hat{B}_1 + B_2) = 1$ according to H_0

It can also be tested using the F statistics

Where $F = \dfrac{R^2_{UR} - R^2_R / M}{T - R^2_{UR} / n - k}$

where $R^2_{UR} = R^2$ unrestricted

$R^2_R = R^2$ restricted

M = number of linear restriction

K = number of parameters in the unrestricted regression

n = number of observation

6.6 Test of Overall Significance of the Sample Regression.

This is the test of joint significance of B_1 and B_2 and the hypothesis is set as follows

$H_0 : \beta_1 = \beta_2 = 0$

$H_1 : \beta_1 \neq \beta_2 \neq 0$

$$F^* = \frac{(\beta_1 \Sigma xy + \beta_2 \Sigma y) / K - 1}{\Sigma U_i^2 / n - k} = \frac{ESS/df}{RSS/df}$$

That is the analysis of variance approach or

$$F^* = \frac{R^2 / k - 1}{(1 - R^2)/(n - k)}$$

If $F^* > F_T$ reject H_0 otherwise, accept H_0.

6.7 Test of equality of two regression coefficients.

In this type of test the researcher will want to test the following hypothesis

$H_0 : \beta_2 = \beta_3$ or $(\beta_2 - \beta_3) = 0$

$H_1 : \beta_2 = \beta_3$ or $(\beta_2 - \beta_3) \neq 0$

Give that $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U$

The above hypothesis, state that, the two slope coefficient B_2 and B_3 are equal. Such hypothesis is important because it expresses

(1) The similarity of the two slopes
(2) It also express how closely related the two slopes are, e.g X_1 = income of consumer and X_2 wealth of the consumer, the hypothesis try to shows that the income and wealth coefficients are the same. If the Y and the Xs are expressed in logarithm, the hypotheses mean that the elasticity of wealth and income are the same.

$$t = \frac{(\beta_2 - \beta_3) - (\beta_2 - \beta_3)}{Se\,(\beta_2 - \beta_3)}$$

Where $se(B_2 - B_3) = \sqrt{\beta_2) + var(\beta_3) + 2\,cov\,(\beta_2 + \beta_3)}$

If H_o is true

$$t = \frac{\beta_2 - \beta_3}{\sqrt{\beta_2) + var(\beta_3) + 2\,cov\,(\beta_2 + \beta_3)}}$$

The test procedure involves the following

(1) Estimate β_2 and β_3
(2) Obtain the variance, covariance and the standard error of the estimates.
(3) Obtain the t ratio, Note that null hypothesis is $(\beta_2 - \beta_3) = 0$
(4) If the t variable computed is greater than the critical t value, at a given level of significant, then reject H_0.

6 8 Chow Test.

This is a test for structural or parameter stability of regression models, since the value of parameters of a model do not remain the same through the entire time period, when one is dealing with time series data. This may be due to structural changes emanating from external forces. Or due to policy changes. Thus if $Y = \alpha + B_1X_1 + B_2X_2 + B_3X_3 + U$ and the researcher collect data from say 1970 – 1990. The explicit assumption is that $B_1\ B_2\ B_3$ has remained all the same over the years. If they change then there has been a structural break.

Now how do the researcher find out that a structural change has occur? The researcher can find out by obtaining the OLS regression of Y and the Xs and save the residuals. By so doing he is maintaining or assuming that, the relationship by Y and the Xs has not change much over the span of (21) twenty one years that is from 1970 to 1990.

After that the researcher can divide the data into two, may be from 1970 to 1979 that is (10) ten years and from 1980 to 1990 on another side. Then run an OLS regression of the two data separately and save their residuals.

Time period 1970 – 1979: $Y = \alpha_0 + \alpha_1X_1 + \alpha_2X_2 + \alpha_3X_3 + U : n = 10$ (1)

Time period 1980 – 1990: $Y = \alpha_0 + \alpha_1X_1 + \alpha_2X_2 + \alpha_3X_3 + U : n = 11$ (2)

Time period 1970 – 1990: $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + U$: $n = n_1 + n_2 = 21$ (3)

Regression equation 3 assumes that there is no difference between the two time period and therefore estimates the relationship between Y and the Xs for the entire time period consisting of 21 observations. In other words this regression assumes that the intercepts as well as the slope coefficients remain the same over the period. That is there is no structural change. If that happened then $\alpha_1 = n_1$ $\alpha_2 = X_2$ and $\alpha_3 = X_3$ (i.e. restriction).

However, equation (1) and (2) assumes otherwise, that the intercept and the slope coefficient are not the same, over that period under study.

$$F = \frac{R^2_{UR} - R^2_R / M}{1 - R^2_{UR} / n - k}$$

CHAPTER SEVEN

DYNAMIC ECONOMETRIC MODEL

7.1 Distributed – Lag Model:

When we are dealing with time series data we will face two problems which we don't face when we deal with cross- section data. These problems are:

One time series variable can influence another with a time lag. If the variable is non- stationary a problem known as superior regression may arise.

Let consider the first problem for now, in time series data in most cases the effect of the explanatory variables on the dependant variables may take time to manifest itself. In this case, we say there is lag in the explanatory variables.

A model in which the lagged (post) values of the independent variables are included among the set of the current values of the independent variables is called Distributed lag-model.

It takes the form:

$$Yt = \alpha+\beta_1X_t+ ;\cdot_1X_{t-1}+I\beta_3X_{t-2}+.... + \beta_qX_{t-q}+U_t$$

Where q = maximum lag period.

Lag (time) or lagged values of the variables are important explanatory variables in most economic relationship because economic behavior in any one period is to a great extent determined by post experience and post patterns of behavior.

For example:

The impact of compatenising a particular business undertaking on productivity efficiency involves lag.

. it takes time to buy and install the systems.

. it takes time to train the workers how to use them.

The effect of this action will be felt in the future time.

For instance, the current consumption levels depend on the current income, past levels of consumption, due to habit persistence past levels of income and other factors.

$$C_t = f(Y_t, Y_{t-1}, C_{t-1}, \ldots\ldots\ldots)$$

It takes time before government securities are bought by the CBN and the fed into the general economy. The expansion takes time, one stage after the other

The link between inflation and increase in money supply is not instantaneous, studies have shown that the 1 percent increase in M1 fall over a period of 20 quarters it has a long run effect –in the inflation rate.

The demand per durables depend on current income and past levels of income (saving) Y_{t-1}

The investment function/projects depend on past output or proportional change in output. This is the accelerator model of investment which is written as

$$I_t = \alpha (X_t - X_{t-1})$$

Where X_t= output at time t

X_{t-1}=past level of output.

α=capital-output ratio

I_t= investment at period t.

So, there is lag in the function.

The impact of adjusting interest rate on the macroeconomic variables is not instantaneously felt or fed to the economy until after some period of time (lag). Reduced interest rate may not bring instantaneous increase in investment.

Demand for nondurable at micro or individual levels because of human or characteristics of human behavior of habit persistence. For example, demand for food, tobacco, etc. depends on past levels of consumption of these goods.

There are therefore lags everywhere in economic relationships. Their refinement has made possible the handling of expectations about future event, which offers much flexibility to the formulation of models of economic behavior.

There are mainly three reasons for these lags in economics

- Psychological reasons → habit persistence
- Technological reasons → technological adjustment as well as expectation of future changes
- Institutional reasons → contractual obligation, insurance policy, long-term saving or investments.

We face some problems when dealing with distributed lag-model

- Lag length
- Data efficiency
- Estimation of lags

Lag Length involves the determination of the maximum lag.

What is the maximum lag?

What or how do we determine the maximum lag?

The issue of maximum lag varies between institutions, individuals, policies and another. The lag order therefore depends on experience of the particular individual or organization and policy concerned. There is no economic theory of determining that.

The determination of lag length can be done, adopting the following general procedure.

The steps are outlined as follows:

- specify as many lag as possible: beginning with a fairly large length, say q_{max}
- Use standards error test to test whether the coefficient of the maximum lag variable is statistically significant i.e. $H_0: \beta q_{max} = 0$
- If the null hypothesis is accepted, we drop the highest lag (βq) and then re-estimate the model with maximum lag equal to q-1 i.e. $\beta X q-1$.
- We test its statistical significance using standard error test.
  $H_0: \beta q_{max-1} = 0$
- If the null hypothesis is accepted, we drop the maximum lag and repeat the above steps at which H_0 is rejected gives the required maximum lag to be included in the models.
- If in step two, three or any other step we reject hypothesis H_0, especially in the first step of statistical test of the hypothesis, we have to take another test.
- We need to test $H_0: \beta q_{max+1} \ldots \beta q_{max+2}\ \beta q_{max+3} \ldots$ until we reflect the H_0 hypothesis. i.e. $\beta q_{max+1} = 0$, $\beta q_{max+2} = 0$, etc. In this case we are moving forward, so we stop where we are accept H_0. Suppose we reject H_0 : $\beta q_{max+5} = 0$

Our maximum lag should be β_{max+4} since q_{t+5} is statistically insignificant.

- **Data sufficiency**: when we are dealing with lag model we tend to lose some observations. Our model should therefore include as much observations as possible, as required by the central limit theorem, the more efficient our observations are the more consistent and reliable our findings and generalization become.
- **Estimation of lags using OLS**. Two problems are involved here:
    - If the number of lags in large and the sample is small, which is usually the case in time series data, we may be unable to estimate the parameter because there will be no adequate degree of freedom to carry all t-test.
    - It is almost certain that we will have multicollinearity problem, since we might expect a strong correlation between successive values of the same variable.

As for data sufficiency and losing of some observations in lag-model, lets take 10 observations on which we want to say a regression model using X_t , X_{t-1}, X_{t-2} and X_{t-3} i.e. maximum lag of three period our data look like this

Y_t	X_t	X_{t-1}	X_{t-2}	X_{t-3}
Y_4	X_4	X_3	X_2	X_1
Y_5	X_5	X_4	X_3	X_2
Y_6	X_6	X_5	X_4	X_3
Y_7	X_7	X_6	X_5	X_4
Y_8	X_8	X_7	X_6	X_5
Y_9	X_9	X_8	X_7	X_6
Y_{10}	X_{10}	X_9	X_8	X_7

We can see that:

- On Y_t and X_t we have to start with period four (4)
- We lose Y_1, Y_2, Y_3 and X_1, X_2, X_3 in the first two columns
- We lose X_1, X_2 and X_{10} in the third column
- We lose X_1, X_9 and X_{10} in the fourth column
- We lose X_8, X_9 and X_{10} in the last

All together we lose 15 observations.

Example 7.1

Consider the model

$$Y_t = \beta + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + U_t$$

Where: Y_t = loses due to accident (as per month) in a company, X_1 = hours of safety training provided for each worker per month. The estimated distributed lag model is found to be:

$$Y_t = 92001.52 - 145X_t - 462X_{t-1} - 424X_{t-2} - 199X_{t-3} - 36X_{t-4}$$

(2001.71) (476) (47.66) (46.21) (97.8) (47.45)

This is summarized in the following table

Table 7.1 sumarmary of estimated distributed lag model

	Coefficient	S.E.	t-table
Intercept	92001.52	2001.71	45.96
X_t	- 145	476	-3.04
X_{t-1}	- 462	47.66	-9.70
X_{t-2}	- 424	46.21	-9.19
X_{t-3}	- 199	97.8	-4.18
X_{t-4}	- 36	47.45	0.78

Interprete the model giving clear policy implications and give your recommendations.

Solution:

- The model implies that there is a negative relationship between losses due to the accidents and safety training of workers. The training reduces losses due to accident. It also implies that current and previous month safety training affect current rate of losses due to accident.

  Reading between the lines (from the analysis of the model estimates) the company can conclude the following that:

- An increase in safety training of each worker by 1 hour per month in any given month is associated with
- An immediate reduction in loses due to accident by N45, all other things being equal. This is the instantaneous effect of the safety training.
- A reduction in loses due to accident by N462 one month later, all other things being equal.
- A reduction in loses due to accident by N424, two months later, ceteris paribus.
- A reduction in loses due to accident by N199 three month later, all other things being equal.
- A reduction in loses due to accident by N36 four month later, all other things being equal.
- The statistical test of significance of the co-efficients shows that all the coefficients are statistically significant at 1% to 5, except, $\beta_4 = -36$.

  This means that X_{t-4} is not significant in explaining the impact of safety training due to accident.

The implication is that workers forget their safety training after 3 months. Hence, after every three months safety training should be offered to the workers.

- The intercept i.e. value 92001.52 implies that the company incures a total loss of N92001.51 due to accident when no safety training is offered to the workers.
- The reduction in losses due to accident is declining from N462, to N434, to N199 and then to N36, which means that the workers are forgetting their training after every months – getting involved in little accidents after every month.

In general, the effect of safety training on accident exhibits a hump-shaped pattern over time, the immediate effect is fairly small. The effect then increases before it fails again three month later.

7.2 Short Run Multiplier

The short run multiplier or impact multiplier gives the effect of X on Y on the same period. It is given by

$\delta Y_t / \delta X_t = \beta_0$

In other words, it gives the change in X in the same period of time. It is thus a partial derivations of Y with respect to X. in our example, the S.R.M. β_0 = (N195)

7.3 Interim/Intermediate Multiplier

This gives the multiplier at a period next to the short-run or period after the next. For example, in the next three period or 2nd period

2^{nd} period → interim = $\beta_0 + \beta_1 + \beta_2$

3^{rd} period → interim = $\beta_0 + \beta_1 + \beta_2 + \beta_3$

For example, in our example, the 3^{rd} interim multiplier is

$\beta_0 + \beta_1 + \beta_2 + \beta_3$ = (N145 + N462 + N424 + N199) = (N1230)

7.4 longrun multiplier

Measures the effect of a sustained change in X on Y i.e a change in X from a steady state of equilibrium to another which captures the dynamics of the model. It is given by the total derivative of Y with respect to X →

L.R.M. = $\Sigma\beta_i$

It is also called total or distributed – lag multiplier. In our example, the long-run multiplier,

$\Sigma\beta = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4$ = (N145 + N462 + N424 + N199+ N35) = (N1266).

Hence, the long run impact of an increase in safety training of each worker by 1 hour per month in any given month is reduction in loses due to accident by N 1266.

7.5 The Lag Distribution

The lag distribution gives the weight of β_i in the long run relationship. It is simply found by expressing the coefficient of β_i as a multiplier. It is symbolically WRITTEN AS:

$w_i = \beta_i / \Sigma\beta_i$, where I = 1,2,3,...k

for example, $w_1 = \beta_i / \Sigma\beta_i$, =145/1266 = 0.11

$w_2 = \beta_2 / \Sigma\beta_i$, =462/1266 = 0.36

$w_3 = \beta_3 / \Sigma\beta_i$, =424/1266 = 0.33

$w_4 = \beta_4 / \Sigma\beta_i$, =199/1266 = 0.16

$w_5 = \beta_5 / \Sigma\beta_i$, =36/1266 = 0.03

0.99

This indicates that 11% of the total impact of 1 hour safety training in any given month on loss due to accident is felt immediately 36% after one month, 33% after two month and son on... and 100% by the end of the fifth month.

MEAN LAG

It is given by $\mu = \Sigma_i(\beta_i)$. it shows the extent to which Y responds to X as it change. i.e. The response of the regressand to the regressor.

Symbolically

$\mu = \Sigma_i (W_i)$

$= 1(W_1) + 2(W_2) + 3(W_3) + ... + K(W_k)$

In our example above

$\mu = 1(0.11) + 2(0.36) + 3(0.33) + 4(0.16) + 5(0.03)$

$\mu = 2.61$

Therefore $\mu = 2.61$

a higher value of μ indicates slow response of Y to X i.e. response of loss due to accident to safety training.

A low value, on the other hand, indicates fast response of Y to X.

Note: X is the explanatory variables, while Y is the regressand, as usual.

μ is therefore the average speed of adjustment.

7.6 Estimation of Distributed Lag Model

Distributed lag model is very important such that it helps economists to go into distant past in order to look at preset and the future behavior of economic variable. However, the estimation of this type of model becomes a problem because of the following:

a) The problem of multicollinearity resulting from the number of lags of the same economic variable. Economic variables are naturally related even when they are different. In the case of distributed lag we look at the effect of a particular variable over time.

b) Insufficient degree of freedom to test the significance of the parameter as the lags may deplete the degree of freedom, since df = n-k and k = number of (parameters) lags. If the lags are many the degree of freedom df will be small.

As a result of the above problems economists came up with prescriptions to make estimation easier. Prominent among them are;

a) The Koyck model

b) Nerlove's partial adjustment model

c) Lagan's Adaptive expectation model

Etc.

.7 The Koyck Distributed Lags Model (Transformation)

oyck, (1) assumed that the β's are the same sign (2) they decline eometrically as follows:

$\beta k = XkB_0$ where K = 0, 1, 2, ------n, i.e. the β's have a constant geometric relationship.

λ =such that $0 \leq \lambda \leq 1$, is the rate of decline or decay of the distributed lag.

Where $1 - \lambda$ the speed of adjustment.

What this means is that, each successive β coefficient is numerically less than each preceding β.

This implies that as one goes back into the distant past, the effect of that lag on Y becomes progressively smaller – quite plausible assumption.

Suppose $Yt = \alpha + \beta_0 X_t + \beta_1 X_{t-1} = \beta_2 X_{t-2} + Ut$ .....(1)

Applying the constant geometric relationship

$\beta_0 = \beta_0$

$\beta_1 = \lambda (\beta_0) = \lambda \beta_0$

$\beta_2 = \lambda (\lambda\beta_0) = \lambda^2 \beta_0$

$\beta_3 = \lambda (\lambda^2 \beta_0) = \lambda^3 \beta_0$

Substitute into equation 1.

$Y_t = \alpha + \beta_0 X_t + \lambda\beta_0 X_{t-1} + \lambda^2\beta_2 X_{t-2} + U_t$ .....(2)

Then,

$$Y_{t-1} = \alpha + \beta_0 X_{t-1} + \lambda\beta_0 X_{t-2} + \lambda^2\beta_2 X_{t-3} + \text{------} + U_{t-1} \quad \text{.....(3)}$$

Multiply equation 3 by λ

$$\lambda Y_{t-1} = \lambda\alpha + \lambda\beta_0 X_{t-1} + \lambda^2\beta_0 X_{t-2} + \lambda^3\beta_0 X_{t-3} + \text{------} + \lambda U_{t-1} \quad \text{.....(4)}$$

Subtract equation 4 from 2

$$Y_t - \lambda Y_{t-1} = \alpha - \lambda\alpha + \beta_0 X_t - \lambda\beta_0 X_{t-1} + \lambda\beta_0 X_{t-1} - \lambda^2\beta_0 X_{t-2} + \lambda^2\beta_0 X_{t-2} - \lambda^3\beta_0 X_{t-3} + \ldots + U_t - \lambda U_{t-1} \quad \text{.....(5)}$$

$$Y_t - \lambda Y_{t-1} = \alpha - \lambda\alpha + \beta_0 X_{t-1} + U_t - \lambda U_{t-1} \quad \text{.....(6)}$$

Factorizing,

$$Y_t - \lambda Y_{t-1} = \alpha(1 - \lambda) + \beta_0 X_t + U_t - \lambda U_{t-1}$$

Rearranging,

$$Y_t = \alpha(1 - \lambda) + \beta_0 X_t + \lambda Y_{t-1} + V_t \quad \text{.....(7)}$$

Where $U_t - \lambda U_{t-1} = V_t$

Koyck transformation has been able to solve the dual problem of multicollinearity of the lag variables and the problem of low degree of freedom. It however created other problems as a result.

1) It leads to an autoregressive scheme such that Yt_{-1} which is a dependent variable appeared as an explanatory variable. Therefore the estimates become bias.

2) It creates the problem of serial correlation because Y_{t-1} has a relationship with the error term $U_t - \lambda U_{t-1}$, (i.e. E $(U_1 x_1) \neq 0$). Hence the variance will be unnecessarily large.

3) $E(u_t, u_{t-1}) \neq 0$ hence , aurocorrelation problem.

7.8 Nerlove's Partial Adjustment Model

$Y^*_t = \alpha + \beta X_t + U_t$ - - - - - (1)

Y^* = expected value, X_t = actual value

e.g. supposing we have a model explaining growth of capital as a function of output.

Y^*_t is an expected value which we don't have values or data for it, and this is going to create a problem.

To run round this problem we make an assumption: that the actual change observed will partially adjust itself to the desired change.

Assumption where δ is the adjustment ratio

Actual change is a function of the desired change

$Y_t - Y_{t-1} = \delta(Y^*_t - Y_{t-1})$ $0 \leq \delta \leq 1$ - - - (2)

Supposing $\delta = 0$

$Y_t - Y_{t-1} = 0\ (Y^*_t - Y_{t-1})$

$Y_t - Y_{t-1} = 0$ $= Y_t = Y_{t-1}$

Therefore, the actual change partially adjust itself to the desired change

The actual change is equal to expected or desired change

Hence, if change is equal to zero, then output in this period is equal to output in the previous period.

Supposing $\delta = 1$

$Y_t - Y_{t-1} = \delta(Y^*_t - Y_{t-1})$

$Y_t - Y_{t-1} = 1(Y^*_t - Y_{t-1})$

$Y_t - Y_{t-1} = Y^*_t - Y_{t-1}$

$Y_t = Y^*_t - Y_{t-1} + Y_{t-1}$

$Y_t = Y^*_t$

If adjustment ratio (δ) = 1, then actual output in the current period will be equal to the expected output.

Interpretation

Actual level fully adjusts itself to the desired level.

$Y^*_t = \alpha + \beta X_t + U_t$ . . . . . (1)

$Y_t - Y_{t-1} = \delta(Y^*_t - Y_{t-1})$ . . . . . (2)

Since $Y^*_t = \alpha + \beta X_t + U_t$ in equation (1) substitute the equation for Y^*_t in equation (2) and simplify

$Y_t - Y_{t-1} = \delta(\alpha + \beta X_t + U_t - Y_{t-1})$

$Y_t - Y_{t-1} = \delta\alpha + \delta\beta X_t + \delta U_t - \delta Y_{t-1}$

$Y_t = \delta\alpha + \delta\beta X_t + \delta U_t - \delta Y_{t-1} + Y_{t-1}$

Rearranging

$$Y_t = \delta\alpha + \delta\beta X_t + \delta Y_{t-1} - \delta Y_{t-1} + \delta U_t$$

$$Y_t = \delta\alpha + \delta\beta X_t - (1-\delta)Y_{t-1} + \delta U_t \quad (3)$$

Equation (3) can be written as:

$$Yt = \alpha^* + \beta^* X_t + (1-\delta)Y_{t-1} + \delta U_t$$

Where

$\alpha^* = \delta\alpha$ and $\beta^* = \delta\beta$

This partial adjustment model has helped to solve the problems of autocorrelation and serial correlation which the Koyck's model created in trying to solve multicollinearity.

The only problem left is the presence of a dependent variable as part of the explanatory variable thereby rendering the estimates bias.

7.9 Autoregressive Model

Recall that if the regression model includes not only the current but also the lagged values of the explanatory variables (the X^2 S^2), it is called Distributed Lag Mode (DLM).

If, however, the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called an Autoregressive model or dynamic model, for example

$$Y_t = \alpha + \beta X_t + Y_{t-1} + U_t$$

Or the simplest case:

$$Y = \alpha + \beta y_{t-1} - U$$

$$Yt=a+By_{t-1}+y_{t-1}+u$$

Or the complex case

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + ... + \beta_k X_{t-k} + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + ... + \alpha_m Y_{t-m}$$

The important question is: are the OLS assumptions satisfied in an autoregressive model?

Let's recall the stochastic assumption

$E(Y/X) = \alpha + \beta X_t + U$ Which means that our model is a conditional one

U_i are iind $(0,\delta^2)$ i.e. identical, independent, normally distributed, which implies that,

$E(U_i) = 0$
$E(U_i^2) = \delta^2$ which is constant
$E(U_i U_j) = 0 \quad i = j$
Which collectively are referred to as the **White noise** assumption.

$E(U_t X_t) = 0$ which is also called the fixed regressor assumption

When these assumption hold, our estimators $\hat{\alpha}$ and $\hat{\beta}$ are best, linear, unbiased estimator - BLUE

The case of autoregressive scheme (AR), these properties, particularly, the fixed regressor assumption does not hold. Therefore, our estimate will be invalid for small sample, because of consistency. As the sample becomes large it moves closer to 1. The population and the estimates became closer to the population parameters.

Let us show this

Recall that,

$$Y_t = \alpha + \beta Y_{t-1} + U_t$$

Which is the simplest case of AR(1) i.e. autoregressive of the first order.

Now:

At t = 1,

$$Y_1 = \alpha + \beta Y_0 + U_1$$

At t =2

$$Y_2 = \alpha + \beta Y_1 + U_2$$

And substituting for Y_1

$$Y_2 = \alpha + \beta(\alpha + \beta Y_0 + U_1) + U_2$$

$$= \alpha + \alpha\beta + \beta^2 Y_0 + \beta U_1 + U_2$$

Rearranging

$$Y_2 = \alpha (1+\beta) + \beta^2 Y_0 + \beta U_1 + U_2$$

At t =3

$$Y_3 = \alpha + \beta Y_2 + U_3$$

$$= \alpha + \beta[\alpha (1+\beta) + \beta^2 Y_0 + \beta U_1 + U_2] + U_3$$

$$= \alpha [1+\beta + \beta^2] + \beta^3 Y_0 + \beta^2 U_1 + \beta U_2 + U_3$$

At t = 4

$$Y_4 = \alpha [1+\beta + \beta^2 Y_0 + \beta^3] + \beta^4 Y_0 + \beta^3 U_1 + \beta^2 U_3 + U_4$$

Observing this pattern we generalize

In general,

$Y_t = \alpha [1+\beta + \beta^2...+ \beta^{t-1}] + \beta^t Y_0 + U_t + \beta U_{t-1} + \beta^2 U_{t-2} + \beta^3 U_{t-3} +... + \beta^{t-1} U_t$

As we can see, there is an interaction between U and B which is increasing. But we are worried about $E(Y_t, U_t) = 0$

That is, is there any correlation between Y_t, U_t,U_{t-1}, U_{t-2}, etc?

To check this, we multiply our general model by U_t

$U_t Y_t = U_t[\alpha(1+\beta + \beta^2...+ \beta^{k-1}] + U_t\beta + \beta U_t U_{t-1} + \beta^2 U_t U_{t-2} + \beta^3 U_t U_{t-3} +...$
$+ U_t\beta^{t-1} U_t$

Checking for co-variance between Y_t and U_t and the assumption of fixed regressor is violated

The consequences of this are:

- $\hat{\alpha}$ and $\hat{\beta}$ will cease to be valid for small sample. They are not BLUE.
- Inference procedure using $\hat{\alpha}$ and $\hat{\beta}$ are also invalid to test of hypothesis and confidence interval inference.
- Since the sample size is large as a result of AR scheme.

CHAPTERPTER EIGTH

8.1 UNIVARIATE ANALYSIS (TIME SERIES ECONOMETRICS)

Basic concepts:

- Stochastic/random process
- Stationary and non-stationary process
- Cointegration
- Unit Root
- Nonsense or Spurious Regression

The most important type of data used in empirical analysis is **time series data**. This type of data usually follows a **stochastic** or **random** process. A stochastic process is a collection of random variables ordered in time. For example, the GDP of Nigeria between, 1991 – 2009, etc. The GDP of each year is a random variable because; it is a particular realization of many possible (possibilities) values which depend on the economic and political climate prevailing at the period. Other examples of stochastic process are unemployment rate, money supply, consumer expenditure, stock prices, population, etc. Which can be recorded daily, weekly, monthly, quarterly, yearly, etc

8.2 Stationarity/Non-Stationary Process

Empirical work based on time series data assumes that the underlying time series data is stationary. A time series is stationary if its mean and variance do not vary .

A stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or lag between the two periods and not the actual time at which the covariance is computed.

Other names for stationary stochastic process are:

- Weakly stationary
- Covariance stationary
- Second-order stationary or
- Wide sense stochastic process

In this way, let us show what stationary is:

Let Y_t be a stochastic variable with the following properties

Mean: $E(Y_t) = \mu$

Variance: $Var(Y_t) = E(Y_t - \mu^2) = \delta^2$

Covariance: $Y_k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$

Y_k = autocovariance at lag K, which is the covaraince between Y_t and Y_{t+k}. K = 0,1,2,...,m.

Now, if we move from the origin of Y from Y_t to say Y_{t+3}. If Y is to be stationary, the mean, variance and auto covariance (at various lags) remain the same no matter at what point we measure them, they are time invariant.

Note that, if a stochastic process has zero mean; constant variance δ^2, and is serially uncorrelated, it is called **purely random** or **white noise process**. If in addition, it is independent, we call it **strictly white noise.**

Recall that, μ_t is a white noise process i.e. $\mu_t \approx iidn(d, \delta)$.

Important of Stationarity

If a time series is nonstationary the following consequences are met:

- We can only study its behavior only for the time period under consideration. In this case, each time series will be for a particular episode.
- It is not possible to generalize it to other time periods
- For the purpose of forecasting such nonstationary time series may be of little practical value.
- Sometimes **autocorrelation** results because the underlying time series is nonstationary.
- Non-stationarity may lead to the problem of **Spurious Regression**

8.3 Nonstationarity

When a time series is nonstationary it classifiably, called Random Walk Model (RWM). Example is stock prices or exchange rates

RWM divides into two:

- Random Walk without Drift
- Random Walk with Drift

In the former, there is no constant or intercept term. Given that $\mu_t \approx iidn(d, \delta)$, Y_t is said to be a random walk if it depends the random variable U_t

e.g $Y_1 = Y_0 + U_t$

$$Y_2 = Y_1 + U_2 = Y_0 + U_1 + U_2$$

$$Y_3 = Y_2 + U_3 = Y_0 + U_1 + U_2 + U_3$$

This is called a random walk without drift, which is of course, an AR(1) model.

In the latter, there is a constant term, such that

$$Y_t = \delta + Y_{t-1} + U_t$$

Where δ is known as the drift parameter

This can be re-written as

$$Y_t - Y_{t-1} = \delta + U_t,$$

This shows that Y_t drifts upward or downward, depending on δ being positive or negative.

The RWM is an example of what is popularly known as **Unit Root Process.**

8.4 Unit Root Stochastic Process.

Unit root is the same as nonstationarity or random walk. Suppose our RWM is written as

$$Y_t = PY_{t-1} + U_t \qquad -1 \leq P \leq 1.$$

If P is in fact 1, we face what is known as unit root problem, that is, a situation of nonstationarity. The variance of Y_t as we know is non-stationary.

The name Unit Root is due to that fact that P = 1. If P = 1, we can write our RWM above as,

$$Y_t - Y_{t-1} = U_t$$

Now, using the lag operator L so that

$$LY_t = Y_{t-1}$$

$L^2 = Y_{t-2}$, etc. the above RWM becomes

$$(1-L)Y_t = U_t$$

The term unit root refers to the root of the polynomial in the lag operator. If you set (1-L) = 0, we obtain L = 1, hence the name unit root.

It is important in practice to find out if a time series possesses a unit root. The essence of testing unit root is to be able to combine X and Y in model and regress them.

8.5 SPURIOUS REGRESSION

Spurious regression is the regression of two or more variables that are not related on a priori consideration.

Consider the following RWMs

$$Y_t = Y_{t-1} + U_t$$

$$X_t = X_{t-1} + V_t$$

Given that, $U_t \sim N(0,1)$. And, $V_t \sim N(0,1)$.

We can see that the variance of U_t and V_t are non-stationary, as we assume that U_t and V_t are serially and mutually uncorrelated.

For instance, if we take 500 observations regress Y_t on X_t and obtain the following result.

Note that, Y_t and X_t are uncorrelated, the R^2 should be zero – there should not be any relationship between the two variables.

The results are:

Variable	Coefficient	Std. Error	t-statistic
Y	-13.26	0.62	-21.37
X	0.34	0.09	7.61
	$R^2 = 0.104$	d = 0.012	

Table 8.1: Spurious regression result

We can see that the coefficients of X is statistical significant and R^2 is also statistically significant i.e. $\neq 0$.

From these results we may be tempted to conclude that there is a significant statistical relationship between X and Y, whereas on a priori bases, there should be none.

This is what is called Spurious or Nonsense regression. This problem may persist in nonstationary time series, even if the sample is large.

The extremely low Durbin-Watson d value, which suggests very strong first-order autocorrelation, means that something is wrong!

According to Granger and Newhold, an $R^2 > d$ is a good value of thumb to suspect that the estimated regression is spurious, as in the above example.

The above results are thus meaningless. But when we regress the first differences of Y_t (ΔY_t) on the first difference of X_t (ΔX_t), R^2 will be found practically zero, as it should be and Durbin-Watson, d is about 2.

Note: although Y_t and X_t are nonstationary, their first difference are stationary.

8.6 INTERGRATED STOCHASTIC PROCESS

If a time series is stationary, it is said to be integrated of order zero, denoted by $Y_t \sim I(0)$.

We know that RWM without drift is nonstationary, but its first difference is stationary. So we call it integrated of order 1, denoted by $Y_t \sim I(1)$.

Similarly, if a time series has to be differentiated twice to make it stationary, we call it integrated of order 2, and so on.

Most economic time series are generally, I(1). Hence, integration is one way of transforming nonstationary time series to stationary, i.e. if the time series has a unit root. We call this **DIFFERENCE STATIONARY (DS)** stochastic process.

But, a **TREND STATIONARY PROCESS (TSP)** is found to be stationary around the trend line. Hence, the simplest way to make it stationary is to regress it on time and the residuals from this regression will then be stationary.

8.7 Test Of Stationarity

THE UNIT ROOT TEST

A test of Stationarity and nonstationarity that has become popular is the **Unit Root Test**. The test is outline as follows:

Suppose we have

$$Y_t = PY_{t-1} + U_t \qquad -1 \leq P \leq 1$$

Where U_t = white noise error term.

Now, if P=1, i.e. in the case of unit root, our RWM without drift is nonstationarity stochastic process.

We thus try to regress Y_t on its one-period lagged value Y_{t-1} to find out if the estimated P is statistically equal to 1? If it is, then Y_t is nonstationary. This is the general idea!

But for theoretical reasons, we use the following method

DICKEY-FULLER (DF) TEST

Given Y = f(X), where there is only one explanatory variable, the DF is used to test for unit root as follow

Given $$Y_t = PY_{t-1} + U_t$$

We subtract Y_{t-1} from both sides

$$Y_t - Y_{t-1} = PY_{t-1} - Y_{t-1} + U_t$$

$$\Delta Y_t = Y_{t-1}(P-1) + U_t$$

$$\Delta Y_t = (P-1)Y_{t-1} + U_t$$

Take P-1 as δ

Thus,

$$\Delta Y_t = \delta Y_{t-1} + U_t$$

We estimate $\Delta Y_t = \delta Y_{t-1} + U_t$ and test the null hypothesis that

$$H_0: \delta = 0 \quad \text{and}$$

$$H_1: \delta \neq 0$$

If it is accepted (that $\delta = 0$), then P=1, so we have a unit root. We can thus conclude the time series is stationary.

If δ is negative, we conclude that Y_t is stationary.

Use statistical test of significance, i.e. to test the H_0, we find τ (tau) statistic and check the DF or Mackinnon critical values of tau (τ).

In this case:

τ (tau) statistic = $\delta / s(\hat{\delta})$

i.e. τ statistic = the estimated coefficient of Y_{t-1} divided by its standard error.

Note: τ = tau is developed by Dickey and Fuller. For critical tau values, we refer to DF table.

If $|\tau|$ statistic > the DF critical tau values, we reject $H_0: \delta = 0$ and conclude that Y_t is nonstationary.

But, we should make sure that we use the appropriate critical values.

We repeat the same operation for X_t i.e.

$$\Delta X_t = \delta X_{t-1} + U_t$$

Hence, we will always begin with generating values of Y_t

Deriving values of Y_{t-1}

Finding ΔY_t i.e. $Y_t - Y_{t-1}$

Regressing Y_t on Y_{t-1} to find δ

The same will be done to X when we find X_t and Y_t to be stationary. We can safely conduct our regression analysis and the estimates will be valid and reliable!. DF test is also called UNIT – VARIATE ANALYSIS

The implications of unit roots in macro economic data are enormous and dangerous. For example if a structural variable, say, real output is truly 1(1) i.e. integrated of order one, then, shocks to it will have permanent effects. If confirmed, then this observation would rather generate some serious reconsideration of the analysis of macroeconomic policy.

To be more precise, recall that, differencing or detrending is required to achieve Stationarity of time series. Let us consider the first order autoregressive model AR (1) as:

$$Y_t = \rho Y_{t-1} + V_t$$

If we fit (i) above to an I (1) series, we will observe that the conventional measures will tend to hide the true value of ρ; the sample estimate is biased downward and by dint of the very small true sampling variance, the conventional t and F tests will be misleading. Thus, Dickey and Fuller have devised through Monte-Carlo methods an appropriate set of critical values for testing the hypothesis that $\rho = 1$ in an AR(1) regression when there truly is a unit root. One of their general results is that the test may be carried out using 3 conventional t-statistic, but the

critical values for the test must be revised; the standard t-table is inappropriate. The simple idea behind Dickey and Fuller experiment is that they want to find out if the estimated coefficient of Y_{t-1} i.e. p in the above equation is zero or not. Unfortunately, under the null hypothesis that $\rho = 1$, the t value of the estimated coefficient of Y_{t-1} does not follow the t-distribution even in large samples i.e. it does not have an asymptotic normal distribution. Dickey and Fuller observed that under the null hypothesis that $\rho = 1$, the estimated t value of the coefficient of Y_{t-1} in the above equation follows the ( ) tau statistic, i.e. t = ( )

The simplest model to consider is the random walk model (RWM) without drift given in equation (i) as:

$$Y_t = \rho Y_{t-1} + V_t$$

Where $V_t \sim N(0, \sigma^2)$ and cov $(V_t, V_s) = O$, $t \neq s$ Under the null hypothesis that ρ"="1. There are two formulas to carry out the DF - test.

(1) DF_t = the conventional t-ratio $= \left(\frac{\rho - 1}{\sigma p} \right)$

(2) $DF_y = N(\rho - 1)$

N = number of sample observation

p = estimated coefficient of Y_{t-1}

The tau statistic is also known as the Dickey and Fuller (DF) test. Apart from the two formulas given above, the actual procedure of implementing the DF test involves several decisions. The unit root stochastic processes can exhibit three natural behaviors which in turn will make the DF - test to be estimated in three

different possible forms or under three different null hypotheses. We may have:

$\Delta Y_t = \lambda Y_{t-1} + V_t$ ... (ii) RWM without drift

$\Delta Y_t = \alpha, + \lambda Y_{t-1} + V_t$ ... (iii) RWM with drift

$\Delta Y_t = \alpha_1 + \alpha_2 + \lambda Y_{t-1} + v_t$ ... (iv) RWM drift around a stochastic trend

Recall that,

Given equation (i) $Y_t = pY_{t-1} + V_t$

Subtract Y_{t-1} from both sides

$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + V_t$ Where $(Y_t - Y_{t-1}) = \Delta Y_t$

$\Delta Y_t = (\rho-1) Y_{t-1} + V_t$

Let $(p -1)$ be represented by λ

We have $\Delta Y_t = \lambda Y_{t-1} + V_{t\lambda}$

This is how the abov equations (ii - iv) are obtained. In each case the null hypothesis is that $\lambda = 0$ or $(p =1)$ that is, there is a unit root – the time series is non-stationary. The alternative hypothesis is that $\lambda < 0$ $(\rho < 1)$ that is, the time series is stationary.

It is very important to understand that the critical values of the tau (λ) test to test the hypothesis that $\lambda = 0$ or $\rho = 1$ are different for each of the structures of the RWM as shown in equations (ii - iv) for DF test.

Recall the procedure for DF test

a) Estimate equation (ii) or (iii) or (iv) by OLS

b) Divide the estimated coefficient of Y_{t-1} in each case by its standard error to get the computed (τ) tau statistic

c) Refer to the DF t-able to obtain the critical values (or Mackinnon critical tau) at a given level of significance.

d) If the absolute value of the DF calculated > tau (τ), we reject the null hypothesis that $\lambda = 0$ or $p = 1$ this means that the time series is stationary, Also, if the absolute value of the DF calculate does not exceed the tau (τ), we accept the null hypothesis and conclude that the time series is non-stationary.

However, the major task before the researcher is to determine which of the three models (ii - iv) – may be appropriate. In this case, Peter Kennedy rightly said that we shall use common sense and economic theory (Plausible a priori theoretical analysis).

Augumented Dickey Fuller (Adf) Test

The test is used where:

- There are more than one explanatory variables
- There are lags.
- The error term U_t is correlated.

Dickey-Fuller (DF) test assume that the disturbances in the model are white noise (uncorrelated). An extension which will accommodate some forms of serial correlation is the Augmented Dickey-Fuller test (ADF). The ADF test is the same as the simple (DF) test above. This test is carried out by augmenting the preceding three equations by introducing the lagged values of the dependent variable ΔY_t For instance, suppose we use

equation (iv), the ADF test here consists of estimating the following regression. .

$$(\Delta Y_t = \alpha_1 + \alpha_{2t} + \lambda Y_{t-1} + \alpha_1 \Sigma \Delta Y_{t-1} + V_t \ldots\ldots$$

Where $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2})$, $Y_{t-2} = (Y_{t-2} - Y_{t-3})$ etc and V_t = a pure white noise error term. The number of lagged difference terms to be included in the model is often determined empirically, the idea is to include enough terms so that the disturbance term in (v) is serially uncorrelated. In ADF we still test whether $\lambda = 0$ or ($p = 1$) and it follows the same asymptotic distribution as the DF statistic, hence the same critical values can be used. What the researcher should know is that, the ADF test adjusts the DF test to take care of any possible serial correlation in the error terms by adding the lagged difference terms of the regressand.

The Philips - Perron (PP) tests

The Philips (1987) and Philips and Perron (1988) statistic are very general but appear to have less than optimal small sample properties. It is known as pp test and may be computed for the same three functional forms of RWM (ii - iv). The standard formula to compute the Philips -Perron test is given as:

$$Z = \left[\frac{C_o}{a}\right] ½ \left[\frac{p-1}{v}\right] -½ (\alpha - Co) \frac{NV}{\sqrt{as^2}}$$

$$\text{Where } C_o = \left[\frac{N-K}{N}\right] S^2$$

$$\alpha = C_o + 2 \Sigma \frac{1-l}{L+1}$$

$$Cj = \frac{1}{} \Sigma e_t e_{t-s} \quad j = o.., p = j \text{ th}$$

N (auto-covariance of residuals)

V = estimated standard error of ρ

$S^2 = \Sigma e^2_t$

$\underline{t=1}$

N-K

N = number of sample observations

K = number of estimated parameters

L = length of lag

We will observe that Philips and Perron made use of non-parametic statistical tools to take care of autocorrelation of the error terms without adding lagged difference forms. Since the asymptotic distribution of the PP test is the same as the ADF test statistic, we will always have the same result.

Example

The following results were obtained from the regression analysis of the determinants of the federal government expenditure in Nigeria (1970-2005).

$\Delta(TFGE)_t = 0.368 + 0.0038GDP_t + 0862\ (TFGE)_{t-1} + 0.250\ (\Delta TFGE)_{t-1} + e_t = (0.124)\ (0000121)\ (0.0157) - (00652)$

Where Δ is the difference operator

$S = 0.00910$, $R^2 = 0.9996$, adjusted $R^2 = 0.9812$

$d = 2.032$, $n = 121$

We are required to test for unit roots:

$DF_t = \frac{0.8620}{0.0157} - 1 = -\frac{0.138}{} = -8.790$

$= -0.015'$

$DF = 120\ \frac{(0.8620 - 1)}{1 - 0.250} = -\frac{16.56}{0.750} = -2208$

From the DF of table, the critical values at 5% level are:

DF_t (critical) = -3.60

DF (critical)= -17.9

We can observe that neither statistic is greater than the critical values, so we conclude that there is no unit root in the process. Thus we reject the hypothesis of a unit root.
To apply the same data to Philips and Perron tests, we are obliged with further information as follows:
L = 4 length of lag
N = 120, p = 0.8620,82 =.0.0000828
V = 0.01572, C_o = 0.00008152
a = 0.0000740731
Substituting the above information to equation (a) above we obtain.
Z = - 9.1299 < 3. 60
We therefore reach the same conclusion that there is no unit root - we reject the null hypothesis.

8.8 Cointegration

We know that if we regress a nonstationary time series, the problem of spurious regression may result, but there is a special case where a unit root in Y_t cancels out the unit root in X_t. In this case, X_t and Y_t are said to be cointegrated.

Cointegration is therefore a situation where a linear combination of nonstationary time series data cancels out the stochastic trend in the two series.

More precisely, if X_t and Y_t have unit root, but same linear combination of them is stationary, we can say that X_t and Y_t are cointegrated. In this case, the error will not get too large (explode over time) and X_t and Y_t will not diverge from one another. In order word, X_t and Y_t will trend together. We say that,

AEG test procedures

Suppose we are to consider the federal government expenditure (TFGE) and GDP time series data, subjecting these time series individually to unit root analysis, we find that both are I (1), that is, they contain a unit root. Suppose then that we regress TFGE on GPP as follows:

$(TFGE)_t = \alpha + \beta GDP_t + \mu$ ..... (i)

This implies that:

$\mu_t = (TFGE)_t - \alpha - \beta GDP_t$ ..... (ii)

Applying relevant data, we obtain the following results.

$(TFGE)_t = -20.116 + 0.8763GDP_t$ .....(iii)

t = (-5.3601) (101.282)

$R^2 = 0.9940$. d = 0.5213

$R^2 = 0.9625$

Since TFGE and GDP are individually nonstationary, there is the possibility that the regression is spurious. But when we perform a unit root test on the residuals obtained from (iii) above, we obtained the following results.

$\Delta\mu = -0.3421\mu_{t-1}$

t = (- 4.671), $R^2 = 0.1421$, d = 2.2110

Note that we still make use of the DF or ADF table (τ) tau. In this case at 1 % level τ value = - 2.5899. Since the computed τ (= t) = - 4.671 is less than the critical (tau), our conclusion is that the residuals from the regression of TFGE on GDP are I (0) i.e. they are integrated of order zero which implies stationarity. Thus, equation (iii) is a cointegrating regression and this regression is not spurious even when the two variables are individually nonstationary.

We can refer to equation (iii) as the long run federal government

expenditure function and interpret its parameters as long-run parameters. Therefore, 0.8763 represents the long-run GDP slope.

Durbin-Watson Test For Cointegration

This is a quicker method of testing Cointegration between Y_t and X_t.

1st compute the Durbin-Watson d from the cointegrating regression of Y_t and X_t.

2nd conduct the Durbin-Watson test: H_0: d = 0 since $d \approx 2(1-\hat{P})$ and with unit root, P=1 so that d = 0. Critical values of d at 1, 5 or 10 percent.

3rd Decision Rule:

If the computed d is greater than the d at critical value (1, 5 or 10%), we accept the null hypothesis of Cointegration, and conclude that Y_t and X_t are cointegrated.

If the computed is less than the d*, we reject the null hypothesis of Cointegration and conclude that Y_t and X_t are not cointegrated.

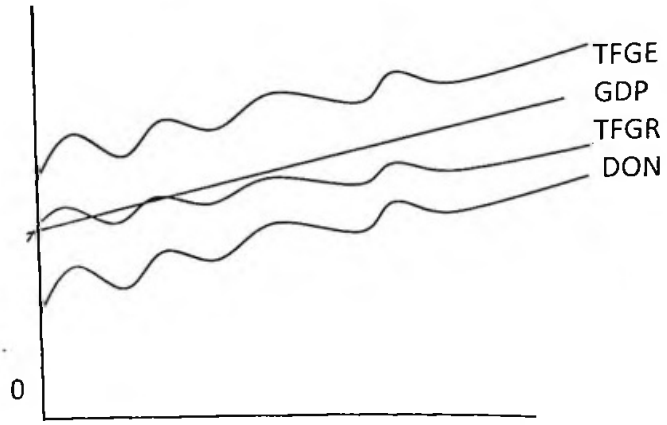
Cointergrating Durbin-Wastson test is the shortest and simplest method of finding out whether two variables are cointegrated. The critical- values were first developed by SARGAN and BHARGAVA. The procedure is that we use the Durbin-Watson d obtained from the cointegrating regression, such as d = 0.5213 given in our example.
We then test the null hypothesis that:

H_o: d = 0
H_1: d≠ 0

Rather than the standard d = 2. This is because since d = 2 (1 -

p), if there is to be a unit root, the estimated ρ will be about 1, which implies that d = will be about zero. From the work of J.D SARGAN the 1%, 5% and 10% critical values to test the hypothesis that the true d = 0 are 0.5110, 0.3681 and 0.3220 respectively. These critical values are contained in modern software statistical packages. However, if the computed d value is less than the critical value, we reject the null hypothesis of cointegration, but if the computed d value is greater than the critical value, we accept the null hypothesis of cointegration. In the present example of TFGE and GDP, the computed d 0.5213 which is above the critical values at 1 %, 5% and 10% suggesting that TFGE and GDP are cointegrated, thus supporting the finding on the basis of AEG test. Our finding in general is that although the TFGE and GDP individually exhibit random walk, but there seem to exist a stable long-run relationship between them such that they will not wander a far from each other. See below diagram for clarity

Fig. (A)



DON = degree of openness
TFGR = Total federal government revenue
GDP = Gross domestic" product
TFGE = Total federal government expenditure

We can observe in fig. (A) That the GDP does not depart much from the TFGE.

8.10 Error Correction Mechanism/Model

According to Granger Representation Theorem, if two variables X and Y are cointegrated, then the relationship between the two can be expressed as the error correction model, through the error correction mechanism. The fact that X and Y have long run equilibrium does not mean that they have short run equilibrium – there may be disequilibrium. The equilibrium error is therefore used to tie the short run behavior of Y to its long run value.

It should be noted that in so far as we are interested in understanding the long-run relationship between the two variables, the estimation of cointegration between them is necessary. This is because the coefficient from the cointegrating regression model is the LONG RUN MULTIPLIER.

Problems associated with the use of time series data

(1) Non Stationarity
(2) Spurious regression
(3) Random shocks
(4) Difficulties in prediction
(5) Auto regressive scheme. Etc.

CONCLUSION

Time series regression analysis begin with the

- Accumulation of the data
- We determine the reliability and validity of the data by conducting:
    - Unit root (Stationarity) test
    - Cointegration test

  Thereby eliminating the problem of nonsense (spurious) regression and probably autocorrelation disturbances

- We regress the dependent variable Y_t on the explanatory variable, X_t.
- Carry out statistical test of significance of the coefficient of the regression line or model.
- We check for econometric disturbances which include:
    - Autocorrelation: violating the assumption $E(U_iU_j) = 0$, ie $E(U_iU_j) \neq 0$. The overall consequence of which is that predictions based on the OLS estimates will be inefficient.
    - Heteroscedasticity: Violation of the Homoscedasticity assumption i.e. assumption of constant variance $E(U_i^2) \neq \delta^2$ which has the same overall consequence as that of autocorrelation.
    - Multicollinearity: (in case of more than one explanatory variable) i.e. presence of perfect linear relationship among the explanatory variables which leads to intermediate values of the parameters. When all these are done with no disturbances are found, the estimated regression model is COMPRESIVE for prediction and results obtained from it are reliable and valid.

Worked example 8.2

ΣX(1) Suppose Y_t is inflation and we find that

$$\Delta Y_t = 0.51 - 0.11\ Y_{t-1}$$

$$(0.21) \quad (0.04)$$

$$t = -2.69$$

$$t = -2.86$$

$$H_0: P = 1.\ \text{And}\ P=0$$

Based on t-test, we can see that:

$$t_t > t^* \text{ i.e.}$$

$$\text{or } t_{cal} < t_t\ (-2.86 < -2.69)$$

We can not therefore reject the null hypothesis. So we concluded that inflation has a unit root – non stationary.

Note that: t-test can be used when the ΔY_t are computed

EXAMPLE 8.3

Given

$$Y_t = 20.666 + 0.996\ X_t$$

Where: Y_t = price of aganic orange

X_t = price of regular orange

If X_t and Y_t are cointegrated, how can we interpret this?

Solution:

The intercept 20.666 gives the premium price of organic orange. This means that when price of regular orange in zero, organic orange is 20.66 per unit. At this point, supply is zero.

The slope 0.996 implies that 1 unit increase in the price of regular orange will lead to 0.996 increase in the price of organic orange in the long run versa.

Example 8.4 Given

$$Y_t = 0.2618 + 0.0159t - 0.0044Y_{t-1}$$

Where: Y_t = money supply in a country and

t = number of years

where:

$$t= (0.7919) \quad (4.4227) \qquad (-3.0016)$$

$$R^2=0.067 \quad d=0.7172$$

Given that at 1, 5 and 10 percent τ(tau) critical values are -3.9811, -3.421 and

-3.1328 is m1 stationary?

Solution

$$H_0: \delta=0$$

$$H_0: \delta\neq0$$

(tau) statistic = $\hat{\delta}/s(\delta)$ = 0.0044/-3.0046

$|\tau|$ = 0.0146 or τ = -3.0046. Critical τ at 1,5 and 10 are 3.9811, 3.421 and 3.1329 respectively in absolute value.

Hence $|\tau|$ < critical values

I.e. 0.0146 < 3.421, < 3.1329 and 3.9811. We conclude that money supply is nonstationary – it is = (1), it has a unit root.

Alternatively, we can use t-value of -3.0046 to test for the unit root.

We compare 1 with any of these three tau critical values.

We can see that 1 is less (negative) than any of these critical values, 1.

3.0096 < 3.9811< 3.0096 < 3.4329 < 3.921 < 3.9811

Conversely,

-3.0046 > -3.1329 > -3.9811. We therefore conclude that money supply time series is nonstationary.

Example 8.5 Given $\hat{Y}_t = -171,4412 + 0.9672X_t$

$t = (-7.4808) \quad (119.8712)$

$R^2 = 0.994 \qquad d = 0.5316$

Where: Y_t = personal consumption expenditure

X_t = personal disposable income

Given that Y_t and X_t are individually nonstationary, the following results are obtained by conducting a unit root test on the residuals.

$\Delta U_t = -0.2753U_{t-1}$

$t = (-3.7791)$

$R^2 = 0.9422$ $d = 2.2775$

The Engle-Granger 1 present critical value of τ(tau) is -2.5899 (critical d=0.51, 0.388 and 0.32) from 100 observations

(a) Test for Cointegration

(b) What does the value 0.9672 represent, based on your results of the test?

(c) What do you call the regression equation $Y_t = -171, 4412 + 0.9672X_t$ based on your result in (a) above?

Solution:

a) Test for Cointegration

Engle-Granger Test

Given $\Delta\hat{U}_t = -0.2753U_{t-1}$

$P=0.2753$

The computed E.G x (tau)=

$\Sigma = p = 0.2753$

$S(p) \quad -3.7791$

The Engel-Granger 1 percent critical value is -2.5899 since Y value is also represented by the computed t i.e. Y= -3.7791, the Y computed is used as

-3.7791, not -0.0728

t computed = -3.7791

Y statistic = -2, 5899

$H_0 : \delta=0$ unit root hypothesis

$H_1 : \delta\neq0$

Since Y computed is much more negative than T statistic (i.e. 3.7791>2.589) we reject the unit root (null) hypothesis and conclude that the residuals from the regression of Y_t on X_t are stationary i.e. they are I (0), hence

The model $Y_t = -171.4412 + 0.9672X_1$ is a cointegrating regression and X_1 and Y_t are co-integrated.

Durbin – Watson Cointegration test we can alternatively use the d-test to check for Cointegration.

Given d computed = 0.5316.

Critical values of d at 1,5 and 10 percent are 0.511,0.386 and 0.322.

H: d = 0 co-integration hypothesis

Hi: $d \neq 0$

Since d computed > d statistic, we accept the null hypothesis and conclude that X_t and Y_t are co-integrated.

i.e. 0.5316>0.511>0.386>0.322

Note: if d can < d statistic we reject the null hypothesis and conclude that X_t and Y_t are not co integrated.

Conclusively both method indicates and proves the existence of cointegrated between X_1 and Y_t.

(b) Since X_t and Y_t are cointegrated, the equation $\hat{Y}_t = -171.4412+0.9672X_t$ and Y_t

The coefficient 0.9672 hence represents the long un or equilibrium, marginal propensity to consume (MPC).

(d) The equation is called co-integrating regression model, since X_t and Y_t are co-integrated.

It should be noted that in testing for unit root and co integration, t =τ, instead of using Y= p!/S(p)

Example 8.6: from the Nigerian private sector housing starts (x) for the period 1960 to 1996, BKT and co- obtain the following regression result.

$\Delta X_t = 31.03 - 0.185X_{t-1}$

(12.5) (0.08t)

T= (τ) (-2.35)

Given that the 5 period critical value of τ is -2.95 and the 10 percent value is -2.60.

On the basis of this results is the housing starts time series stationary or non stationary?

What will be your conclusion if the usual t test is used?

Solution:

H_0: $\delta=0$ unit root hypothesis

H_1: $\delta\neq0$

T_{st} = -2.95 or 2.6

T_{cal} = -2.35

Since $\tau_{cal} < \tau_{sta}$ (less negative), we accept the null hypothesis and conclude that housing starts time series is non stationary.

i.e $2.35 < 2.60 < 2.95$

Using the usual t-test, and applying the rate of thumb:

H_0: $\delta=0$

H_1: $\delta\neq0$

- we accept H_0 if $\hat{\delta}/2 < s(\hat{\delta})$
- we reject H_0 if $\delta/2 > s(\delta)$

In our case, $\delta = 0.188$ and $\delta/2 = 0.188/2 = 0.094$ and $3(\delta) = 0.080$.

We can see that

$\hat{\delta}/2$ (0.094) $> s(\hat{\delta})$ (0.08). Base on this, we reject the null hypothesis of unit root and conclude that housing starts time series is stationary.

However, it is unfortunate tat under the null hypothesis that $\delta= 0$ (i.e p=1) the t value of estimated coefficient X_{t-1} does not follow the t- distribution even in large samples; i.e d does not have an asymptotic normal distribution. Hence our conclusion that X_t is non stationary still stands- we reject the conclusion from the t-test!

8.11 Causality

Although, regression analysis is concern with the dependence of one variable on other variables, it does not actually imply causal-relations. In other words, the existence of a relation between variables does not prove causality or direction of influence. In recent time, the regression involving time series data has shown that we need to establish causal relationships. As Damodar Gujarati rightly puts it, "... time does not run backward. That is if event A happens before event B, then it is possible that A is causing B, However it is not possible that B is causing A. This implies that events in the past can cause events to happen in the present." This is the doctrine or the underlying principles of the so called GRANGER CAUSALITY.

8.12 Granger Causality Test

One may raise a question like this: Is it investment that causes the level of profit or is it the level of profit that causes investment?

Let Y_t = investment

X_t = level of profit

Note: the question in words can be symbolically written thus:

$(Y_t \longrightarrow X_t)$ $(X_t \longrightarrow Y_t)$

The arrow points to the direction of causality. One relevant assumption of causality is that the information regarding the prediction of the respective variables Y_t and X_t is contained solely in the time series data on these variables. The Granger test involves the estimation of the following pair of regression:

$Y_t = \Sigma\beta_1 X_{t-1} + \Sigma k_1 Y_{t-j} + \mu_{1t}$ ..... (i)

$X_t = \Sigma\alpha_i X_{t-1} + \Sigma\alpha_j Y_{t-1} + \mu_{2t}$ ..... (i)

Where we assume that the error terms μ_{1t} and μ_{2t} are uncorrelated i.e. $E(\mu_{1t}, \mu_{2t}) = 0$. In this case, we are dealing with only two variables, Y_t and X_t This is known as BILATERAL CAUSALITY. Above two variables is called MUL TIVARIABLE CAUSALITY

Generally, since the future cannot predict the past, if variable X granger causes variable Y, then changes in X should precede changes in Y. Therefore, in regressing Y on other variables, (including Y_{t-1}), if we include past values of x i.e. (X_{t-i}) and it significantly improves the prediction of Y, then we can say that X granger causes Y. A similar situation applies if Y granger causes X

Steps in Granger Causality Test

1. Regress Y_t on all lagged Y terms and other variables but not the lagged X variables to obtain the restricted regression. This will help us to derive the Restricted Residual Sum of Squares (RSS_R)

2. Run another regression as in (1) but include the lagged X variables to obtain the unrestricted regression. From this we will derive the unrestricted residual sum of squares (RSS_{UR})

3. Establish the hypothesis

   $H_o: \Sigma\beta_i = 0$

   $H_1: \Sigma\beta_i \neq 0$

Where H_o implies that lagged X terms do not belong in the regression.

4. To test the null hypothesis, we use the F statistic:

$$F = \frac{(RSS_R - RSS_{UR}) / N}{RSS_{UR}/(n-K)}$$

Where N = number of lagged X terms

K = number of parameters estimated in the unrestricted regression.

n = sample size

UR = unrestricted

R = restricted.

5) If $F^*_t > F_t$ at a chosen level of significance, we reject the null hypothesis (H_o) and conclude that the lagged X term belong in the regression and invariably X granger causes Y_t, simply put, X causes Y.

8.13 Causality Vs Exogeneity

As we have studied in the previous chapter, economic variables are classified into two broad categories, ENDOGENOUS and EXOGENOUS variables.

From a layman's view, endogenous variable are treated as dependent variable in a single-equation model while exogenous variable are the equivalent of independent variable or regressors (X), provided the X variables are uncorrelated with the error turn in the model.

Assuming that in a granger causality test, we find that an X variable granger causes Y variable without feedback effect (No

bilateral causality) can we now treat X variable as exogenous? The question implies, can we use granger causality or non-causality to establish exogeneity of a variable? This question borders on the distinction between the three types of exogeneity.

1) WEAK EXOGENEITY - suppose we consider only two variable X_t and Y_t, the X_t is said a be weakly exogenous if Y does not explain X_t in a regression model. In this case, X_t will be conditionally used to estimate and test the regression model.

2) STRONG EXOGENEITY: X_t is said to be strongly exogenous if the present Y and lagged values of Y do not explain X_t. That is observes of feedback relationship.

3) SUPER EXOGENEITY:- X_t is said to be super exogenous if the parameters in the regression of Y and X do not change even if the X values change. That is, the parameter values are invariant to changes in the value (s) of X.

We can now observe from the foregoing discussions that weak exogeneity is for estimation and testing strong exogeneity is for forecasting (prediction) and super exogeneity is for ... Analysis. As for granger causality, if a variable, say, Y does not cause another variable say X we cannot just conclude that X is exogenous. Thus, granger causality is neither necessary nor sufficient condition to establish exogeneity.

CHAPTER NINE

9.1 DUMMY VARIABLE REGRESSION MODEL

Regression analysis may involve qualitative variables such, socio-economic condition, political condition, location, political instability, etc. either as dependent or independent variable. Others include sex, race, color, religion, nationality, geographical region and party affiliation.

These variables usually, indicate the presence or absence of quality or an attribute such as

- male or female
- black or white
- happiness or sorrow
- Educated or uneducated; and so on.

Thus, they take either 1 or 0 as their value, with 1 indicating the presence of that attribute and 0 indicating its absence. For instance, 1 may indicate that the person is happy while 0 indicates that he/she is unhappy.

These variables are referred to as:

- dummy variables
- indicator variables
- switching variables
- binary variables
- categorical variables
- dichotomous variables or
- qualitative variables

The dependent variable can be a dummy variable for instance

$$Y = \alpha + \beta_1 LOC + \beta_2 Fin + \beta_3 FLE + \beta_1 MLE$$

Where y = enrolment into western school, LOC = location, Fin = Father's income, FLE = father's level of education, MLE = Mother's level of education.

The dependent variables as well as the first und last two independent variables are all qualitative. Suppose we obtain

$$\hat{Y} = \alpha + \beta_1 LOC + 0.76 Fin$$

The slope, 0.76 means that propensity or probability of enrolling children is 0.76, thus, the model is a probability model. Hence, the β will capture the probability of attendance.

On the other hand, dummy variable can be the independence variables (as a way of turning qualitative variables). Once the variables are made qualitative then the formulation and estimation can be applied directly.

It should be noted that since the dummy variable can take only one of the two values (1 or 2), they are called Binary.

When all the regressors in a regression model are dummy variables, such model is called Analysis of variance (ANOVA) model.

Example

$Y = \alpha + \beta$, loc + FLE 1 β2 MLE as in the above example or:

$$Y = \alpha + \beta_1 E + \alpha_2 L$$

Where Y = model school teacher's salary

E = level of vocational leaning

L = location, whether in the metropolis or in the village.

When the regressors are a mixture of quantitative and qualitative variables, the model is called Analysis of covariance (ANOVA) model for example

$$Y = \alpha + \beta_1 \times_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Where Y = price of house

X1 = plot size(quantitative variable)

X2 = No of bedroom (Q11 variable)

X3 = location (Q1 variable)

X4 = proximity to good school (Ql variable)

X5 = presence of drive-in guage (QL v)

This is an ANCOVA model.

Another example is

$$Y = \alpha + \beta 1 \times 1 + \beta 2\ D + \beta 2W$$

Where Y = hourly wage in dollar

X = education (years of schooling) [QN]

D = sex, male or female [QL]

W = nationality, Nigerian or non Nigerian (quantitative variable)

In conclusion, we should note that dummy variables are useful in season analysis.

With I = season and 0 =off-seas.

EXAMPLES

Ex(9.1) given the model

$$Y = \alpha + \beta D_1 + u$$

Where Y = price of house

D = location = 1 = low density

= 0 = high density

The estimated regression line is found to be Y= 590,885 + 250,996 Di

Interpret this model.

Solution:1

- the intercept 590,885 implies that a house which is located at a high density area is sold at #590,885 (in this case, D= D)
- the slope 250,996 implies the difference between the price of a house in a low-population density area and house in a high density area. It implies that a house located in low density area tends to be worth #250,996 more than houses located in high density. (in this case D=1)
  Note that, it is a statement of tendency!

Example 9.2 given the regression model

$$Y = 26158.6 - 1734.47X1 - 3264.62X2$$

Se = (1128.52) (1435.95) (1499.62)

T = (23.1759) (-1.2018) (-2.1776)

R = 0.0901

Where Y = average salary of public school teachers in Nigeria (in naira).

X1 = dummy variable = 1 if state is NE or NC

= 0 other region i.e otherwise

X2 = = 1 if the state is in south

= 0 otherwise i.e. other regions.

Interpret your/this model.

Solutions:

NC = north central

NE = north east.

- Significant the intercept 26,158.6 implies that the mean salary of teachers in the west is about #26158 (X1 = 0 and X2 = 0).
- the mean salary of teachers in the NE or NC tends to be lower than that of the west by #1734. i.e. it is about # 24424 (X1 =1)

- The coefficient -3264.62 implies that the mean salary of teachers in the NC tends to be lower than that of the west by #3264, i.e. it is about #22894. (in this case x2 = 1) base on statistical test of significance:

- The difference between the salary of teachers in the NE & C and those in west is not statistically significant. By rule of thumb:

  173447/2 < 1435.95 i.e 867.235<1435.95 so we can conclude that there may be other qualitative factor that are responsible for this difference such as gender, cost of living, educational level etc not only location!
- The difference in salary between the south and the west teachers is statistically significant. By rule of thumb.

- 3264.62/2 >1499.62 i.e 1632 >1499.6
  However, using p test we can see that this value is also statistically significant, since it is 3.5% i.e p = 3.5% the p test for NE & NC difference shows 23.3%.

  The overall conclusion is that statistically the mean salaries of public school teachers in the west and the NE and NC are about the same, but the mean salary of teachers in the south is statistically lower by about #3265.

Exercise 8.1: the following dummy variable regression models.

(a) $Y = -0.2610 - 2.3606\,M - 1.7327\,N + 10.8028\,X$

$t^* = (-0.2357)\ (-5.4873)\ (-2.1803)\ (9.9094)$

$R = 0.2032 \quad n = 528$

Where Y = hourly wages in dollars

M = = 1 female

= 0 male

N = = 1 non while

= 0 otherwise

X= education (years of schooling)

(b) Y =13269.11-1673.51X1-1144,16X2+ 3,2889M
Se = (1395.056) (801.1703) (861.118) (0.3176)
t= (9.5115) (-2.0889) (-1.3786) (10.3174)
Where Y= public school teachers salary in a particular state in #.

9.2 Caution in the use of Dummy variable (dummy variable trap or pitfalls)

(1) Given that a qualitative variable has n category, a researcher is expected to introduce (n-1) dummy variable to enable him to differentiate between them. If the entire n category dummy variables are introduced that will be a case of (perfect collinieality) exact linear relationships among the variables. This will render the estimates indeterminate.

(2) The first category for which no dummy variable is introduced is known as the base, benchmark, control, comparism, reference, or omitted category. And all references or comparism are made in relation to the benchmark category.

(3) The value of the intercept represents the mean value of the benchmark category.

(4) The coefficients attached to the dummy variables are known as the differential intercept coefficients, because they tell by how much the value of the intercept that receive the value of 1 differ from the intercept coefficient of the category.

(5) The choice of the benchmark category is strictly up to the researcher. Although it is sometime dictated by the particular problem at hand.

CHAPTER TEN

10.1 VIOLATION OF THE CLASSICAL ASSUMPTIONS

Review of the classical assumptions

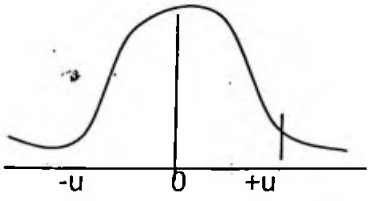
Stochastic assumptions

1) U_t is a random variable, that is for each value of X_t, U_t can assume any value, (positive, negative or zero). This means that the error U_t occur by chance.

2) U_t has a zero mean. That is, the expectation of error term U_t is equal to zero (i.e. $E(U_t) = 0$. Meaning that in the long run the negative and the positive errors will cancel out. This means that, it does not systematically affect the mean value of Y. since the negative value of offset the positive values such that the mean effect on Y is zero.

3) The variance of the error term is constant, (i.e. Var $(U_t) = 6^2$), for each explanatory variable and from sample to sample. This is called the assumption of Homoscedasticity (common variance). That is to say that the variance of the error term does not change.

4) The error term is normally distributed with zero mean and constant variance i.e. U_t N $(0, 6^2)$. This means that, the probability of observing smaller value of the error U_t is higher than the probability of observing larger value of error U_t.



There is no relationship between errors of two different period ie. E $(U_{ti}, U_{tj}) = 0$ where, $i \neq j$. this is the assumption of non autocorrelation.

5) All errors are only accounted for by U_t, meaning that, the explanatory variables are measured without errors.

Non-Stochastic assumption (Data Problems)

1) X_t is assumed to be non-stochastic. That is to say that U_t is fixed in repeated sampling. This means that it does not assume random values as the error term.

2) The number of observation n must be greater than the number of explanatory variables.

3) X_t must vary in a given sample. That is, there must be sufficient variability in the value taken by the regressors.

4) The regression model is correctly specified that is, there is no specification bias in the model.

5) No perfect multicollinearity among the explanatory variables. That is no perfect linear correlation between the explanatory variables.

6) The error term should not be related with the explanatory variable i.e $E(U_t, X_t) = 0$

7) The macro-variables are correctly aggregated. i.e. individual units are correctly summed up.

The Ordinary Least Square OLS is premised on the above assumptions hence, if the assumptions are fulfilled, the estimates of the parameters have BLUE properties. The failure of any of the assumptions, leads to certain consequences.

Now, the questions are (i.e. for stochastic assumptions).

1) How severe must the violation of the assumption be before the estimates of the parameters looses their BLUE properties?

2) How do we find out whether a particular assumption is infact violated?

3) What are the measures to be taken if one or more assumptions are violated?

For the non-stochastic assumptions

1) How serious a particular problem? For example to what extent of severity of multicollinearity that will make estimation difficult?

2) How do we find out the severity of the data problem?

3) Can some of the data problems be easily remedied?

To answer these questions, and to solve the problem of violation, the researcher can use the following ways.

1) **Definition of the problems**: - That is to define the problem properly to be certain that, that type of problem really occurred.

2) **Identification of the problem:** - That is to find out whether that particular assumption is actually violated.

3) **Consequences of the problems**: - That is to find out the implication of the violation, whether it is severe or not

4) **Source of sources of the problem:** - That is to find out where the problem is coming from.

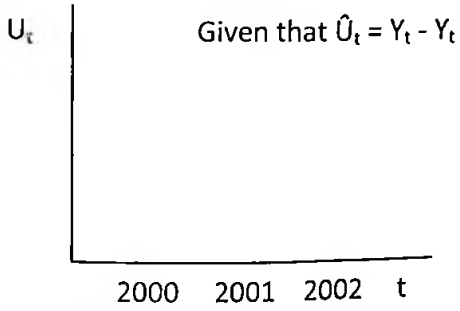
5) **Correction of the problem**

10.2 Non-randomness of the error term U_t

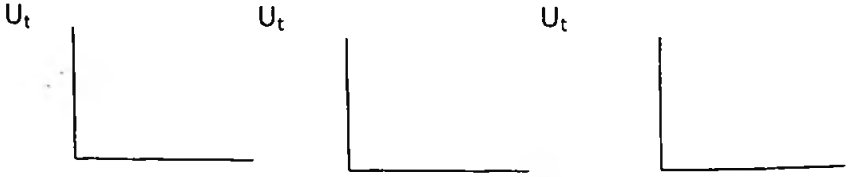
This problem occurs when there is a pattern in the occurrence of the error term (by definition)

Ways of Identification

1. Plot U_t against time e.g.



If the presence shows that U_t and time are moving in the same direction (positive or negative), or there is a pattern of movement in the error we conclude that such problem had actually occurred e.g.



Not random Not random Random (no pattern)

2. Use correlation coefficient to determine the patterns.

3. One can also use the Rank Correlation to determine the presence or otherwise of a pattern.

Consequence of the Problem

The estimates of the parameter will not have the BLUE properties. When U_t is not random it affects the unbiasness of the estimates. It also affects the linearity of the estimates.

Sources of the problem

The problem usually occurs as a result of one of the following:

1) Specification error

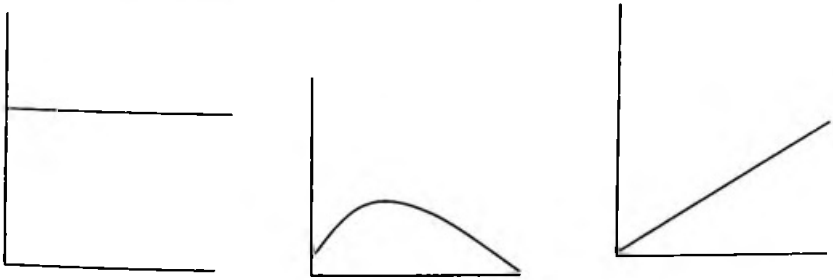
2) Measurement error

3) Omission of data

10.3 Non-normality of U

If for any value U_t, the probability of that occurrence is uniform, positively skewed or negatively skewed then,

there is the problem of non-normality of U_t. e.g.



Uniformit +vely skewed -vely skewed

Ways of Identification of Non-normality of U

1) Plot U_t against time

2) Use mathematical or econometric programs like the SPSS Enviers etc t obtain shapes of skeweness.

Carry out a normality test either $t = \frac{u-u}{\delta U^2}$. J.B., X^2 normality test.

Consequence of the Problem

1) The test of hypothesis relies on the normality assumption. Hence the occurrence of this problem invalidates the test of hypothesis, because the β that is to be tested is expected to be normally distributed i.e. β N $(\beta, \delta u^2/\Sigma x_i^2)$.

2) We cannot find the confidence interval within which to locate the values of α and β.

Sources of the Problem

This problem may be as a result of one or more of the following:

1) Aggregation error
2) Measurement error
3) Omission of data
4) Specification error

Correction

The solution to this problem can be found in the likely sources mentioned above.

The solution can also be found in the central limit theorem.

10.4 The expectation of error term is not equal to zero i.e. $E(Ui) \neq 0$

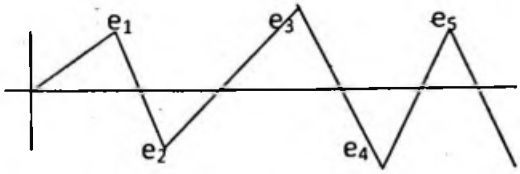
This problem occurs when the both the negative and the positive errors do not cancel out, by definition.

Identification

1) Find the estimates of Ut and get the value of e_1 e_2 e_3----- e_n and find out whether their sum is no zero i.e. whether $\Sigma ei \neq 0$.

2) Plot the e_s against time

i.e.



Check the number of positive e_s and the negative e_s, whether they cancel out or not.

Consequences of the problem of non zero mean

It affects the unbiasness of the estimates. This is because the mean effect of the error term on Y is no longer zero. Since the expectation of error term is no longer zero in $E(u) \neq 0$ then the $E(\beta) \neq \beta$ which mean that y is systematically not only affected by β only but by U which is not equal; to zero.

Sources of the Problem

1) Aggregation error
2) Omission of data
3) Measurement error
4) Misspecification of the mathematical form of the model

Correction

Check the sources to identify where the problem is coming from

1) Increase sample size

10.5 E (Ui, Uj) ≠ 0 where i ≠ j (autocorrelation)

The problem of autocorrelation occurs when the expectations of two error terms of different periods is not equal to zero.

Identification

1) Plot e_t against e_{t-1}

e_t	e_{t-1}
e_{t1}	-
e_{t2}	e_1
e_{t3}	e_2
e_{t4}	e_3
e_{tn}	e_{n-1}



-tive autocorrelation -tive autocorrelation non-autocorrelation

2) If there is a pattern, whether negative or positive then we conclude that there is autocorrelation.

Find the coefficient of correlation between et and e_{t-1}

Find the rank correlation between et and e_{t-1} and test the significance of the correlation.

3) Find the Durbin-Watson statistic

DW statistic is a follow up from correlation coefficient given by:

$$\frac{\Sigma(et - e_{t-1})^2}{\Sigma e_t^2} \approx 2\,(1-P)$$

First find the correlation $P = \frac{\Sigma(et - e_{t-1})^2}{\Sigma e_t^2\, e_{t-1}^2}$ or $P = 1 - \frac{6D^2}{n(n^2-1)}$

then find the DW = 2(1-P)

If P= -1 then DW = 2 (1-(-1))

= 2(1+1)

= 2 x 2

= 4 perfect negative correlation

If p = 0 then DW = 2(1-0)

= 2(1)

= 2 no autocorrelation

If P = +1 then DW = 2(1 – (+1))

$= 2(1-1)$

$= 2(0)$

$= 0$ perfect positive autocorrelation

These demonstration shows that DW lies between zero and 4,

i.e. $0 < DW < 4$

Consequence of the Problem of autocorrelation

1) If there is autocorrelation, the variance of the error term will be too large i.e. δu^2 will be large. Hence the minimum variance property is no longer there.

2) Since the variance of U is large then it affects the variance of the estimates. That is the variance of β will also get larger i.e. var $\beta = \delta u^2 / \Sigma x_i^2$

3) This will affect the test significance and the confidence interval.

$t = \beta - \beta/\delta\beta = \beta/\delta\beta$

Since it affects the standard error, it will affect the t value which is going to be smaller and as a result, in greater times, the null hypothesis will be accepted, even when it is not true, since the decision rule states that:

If t^*_c calculated is greater than t_t tabulated, reject Ho otherwise accept it.

Sources of the problem of Autocorrelation

1) Omission of data

2) Aggregation error

3) Misspecification of the model

4) Measurement error

Correction of the Problem of Autocorrelation

First the researcher must reexamine his model and the data used. At the level of data, he can assume that e_t is correlated with e_{t-1} in the first order correlation.

i.e. $Ut = PU_{t-1} + V_t$ first order correlation

when U_t = present error

u_{t-1} = previous error

P = coefficient of correlation

V_t = the error term

If he assumed this kind of relationship, it is suggested that he internalize the relationship and find the value (i.e. the magnitude) and adjust the data. That is using the value of the correlation coefficient to transform the data, and by so doing, the autocorrelation is taken care of.

The data can be adjusted as follows:

$Y^*_t = Y_t - PY_{t-1}$ i.e. Y_t is adjusted to Y^*_t

$X^*_t = X_t - PX_{t-1}$ i.e. X_t is adjusted to X^*_t

$U^*_t = U_t - PU_{t-1}$ i.e. U_t is adjusted to U^*_t

In essence the information gotten from the relationship between the successive errors (i.e. e_t and e_{t-1}) is used to transform the data, as a means of correcting autocorrelation and hence, the following model $Y^*_t = \alpha + BX^*_t + U^*_t$

After transforming the data, the researcher can go ahead to estimate the above model using OLS method which is devoid of autocorrelation.

The econometric test of statistical assumptions of linear regression model involved the examination of the regression residuals (i.e. the e's) which are the estimates of the U's in order to establish whether the U's conform, to their assumed patterns of behavior. The econometric test of non-stochastic assumptions (i.e. correct aggregation, identification, correct specification, degree of colliniearity of the regressors etc.) involves the examination of the observable explanatory variables and their interrelationships.

Once autocorrelation is detected y applying any relevant technique or test, the appropriate procedure is to obtain an estimate of P's and apply OLS to a set of transformed data.

Since we have to lose one observation of the e's in et-1 i.e. the process f n-1 analysis, Khadiyala suggested the fooling transformation of the first order scheme to avoid the n-1 losses of the e's.

$$Y^*_t = Y_i \sqrt{1-P^2} \qquad X^*_t = X_i \sqrt{1-P^2}$$

TEST FOR AUTOCORRELATION

There are two test techniques prescribed by scholars

1) Von-Neuman test

2) Durbn-Watson test

The Von Neuman ratio is given by

$$\frac{\theta_2 = \Sigma (x_t - x_{t-1})^2/n-1}{x^2 \quad \Sigma(x_t - x)^2/n}$$

The ratio could be applied approximately (with e = 0 by definition), and it is applicable to large sample only.

However, Durbin and Watson suggested a test which can be applied to small samples and it appropriate for only the first order auto-correlation test e.g.

Ho: P = 0

Ho: P ≠ 0

To test the null hypothesis

$$DW = \frac{\Sigma(e_t - e_{t-1})^2}{\Sigma e_t^2}$$

$$= \frac{\Sigma (e_t^2 + e^2_{t-1} - 2 e_t e_{t-1})}{\Sigma e_t^2}$$

$$= \frac{\Sigma (e_t^2 + e^2_{t-1} - 2 \Sigma e_t e_{t-1})}{\Sigma e_t^2}$$

If the sample is large e_t^2-1and are approximately equal, therefore:

$$DW \approx \frac{2\Sigma e_{t-1} - 2\Sigma e_t\, e_{t-1}}{\Sigma et^2_{t-1}}$$

$$= \frac{2\Sigma e_{t-1}}{\Sigma e_{t-1}} - \frac{2\Sigma e_t\, e_{t-1}}{\Sigma e_{t-1}}$$

$$= 2 - \frac{2\Sigma e_t\, e_{t-1}}{\Sigma e_{t-1}}$$

$$= 2\left(1 - \frac{\Sigma e_t\, e_{t-1}}{\Sigma e_{t-1}}\right)$$

$$\text{Since } \frac{\Sigma e_t\, e_{t-1}}{\Sigma e_{t-1}} = P$$

Then $DW \approx 2(1-P)$ or $2 - 2P$

Suppose that

$$\Sigma e_t^2 = 386.98$$

$$\Sigma e^2_{t-1} = 368.92$$

$$\Sigma e_t\, e_{t-1} = 73.00$$

$$\Sigma e_t - e_{t-1} = 4.00$$

$$\Sigma (e_t - e_{t-1})^2 = 614.48$$

$$n = 12$$

$$DW = \frac{\Sigma(e_t - e_{t-1})^2}{\Sigma e^2_t}$$

$$= \frac{614.48}{286} = 1.59 \approx 1.60 \text{ AND}$$

$$P = \frac{\Sigma(e_t\ e_{t-1})^2}{\Sigma\ e_{t-1}^2}$$

$$= \frac{(73)^2}{368} = 0.19$$

If the sample size is equal to or greater than thirty, then we apply the following DW = 2 (1 -P) or DW = 2 – 2P. For example if P = 0.19.

Then DW ≈2 (1 – 0.19)

≈ 2 (0.81)

≈ 1.62

The value of the DW lies between 0 and 4, therefore,

if (1) there is no autocorrelation, P = 0 and we accept that there is no autocorrelation.

(2) P = +1 and d*= 0, we have perfect positive autocorrelation. Therefore if 0 < d* < 2 there is some degree of positive autocorrelation and it is even stronger, when d is close to zero.

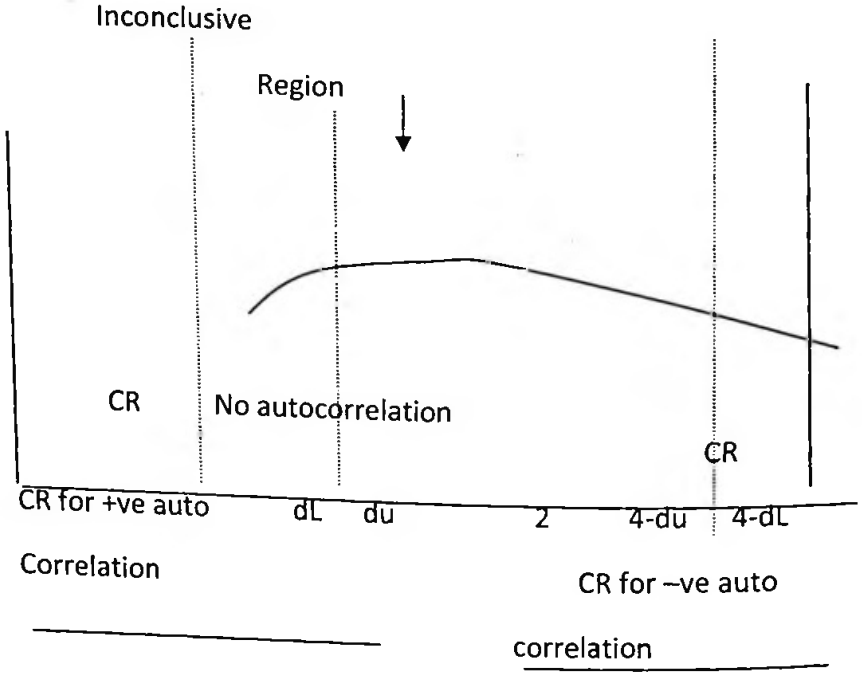
(3) P = -1. d = 4, there is negative autocorrelation. Therefore, if 2 < d* < 4, there is some degree of negative autocorrelation, which is stronger if the value of d is higher.

Next Step

Compute d* and compare it with the table d

That is, the value of d which defines the critical region of the test

Since the exact distribution d is not known Durbin-Watson set out du upper limit and dl lower limit for significance level of d.



1) If d* > dl we reject Ho i.e. there is positive autocorrelation

2) If d* > (4-dl) we reject Ho i.e. there is negative autocorrelation

3) If du < d* < (4-du0 we accept Ho i.e. no autocorrelation

4) If du < d* < du or if (4-du) < d* < (4-dl), the test is inconclusive

Statistical bias is the difference between the expected value of the estimate and the true parameter.

Summary of Consequence of Autocorrelation

1) The estimate of the parameter will be statistical bias because the values, as well as, the standard errors of the estimates are affected.

2) With autocorrelated values of the disturbances term, the OLS variances of the parameter estimates are likely to be larger than those of other econometric methods.

3) The variance of the random term u may be seriously underestimated if the U's are autocorrelated.

4) Finally, if the value of u is autocorrelated, the prediction based on OLS estimates will be inefficient, in the sense that they will have a larger variance as compared with predictions based on estimates obtained from other econometric technique.

Note $re_t\, e_{t-1} = \frac{\Sigma e_t\, e_{t-1}}{\Sigma e^2_t \Sigma e^2_{t-1}} = PU_t\, U_{t-1}$

10.6 Heteroscedasticity

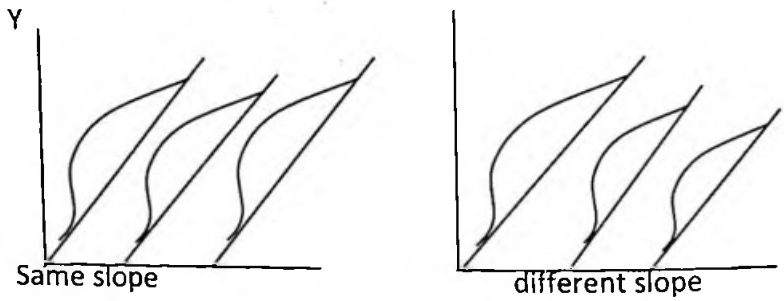
Heteroscedasticity occurs when the variance of the error term is not constant. That is when $E(Ui,Uj) \neq \delta^2$, $i \neq j$. While Homoscedasticity is when $E(U_i^2) = \delta^2$. Therefore, when the variance of the error term is not constant under every level of your explanatory variable, there is a problem of heteroscedasticity.

Definition of Problem

Def: The variance of the error term is expected to be constant

$E(U_i U_j) = \delta^2$at every level of the explanatory variable, where we are not looking at the same error term, meaning $\delta^2 x_i$, $\delta^2 x_2$....
Homoscedasticity

Heteroscedasticity



Identification of the Problem

i. For different levels of x_i, we can find the relationship between the estimate of the error term "e" and the explanatory variable "x', and finding the correlation between them, if the correlation is high we suspect Heteroscedasticity.

ii. Next, is to look at the variance of the error term at different levels of x_i, through testing of hypothesis.

iii. The next thing to do is to use the rank correlation which does not depend on any assumption.

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

In essence, we

a) Plot the graph of x_i and e_i

b) Find the relationship between x_i and e_i

Recalling that estimated equation $Y = \beta x_1 + u$
Clearly the steps are thus:

Given the model $Y = \beta_1 x_1 + \beta_2 x_2 + u$

And the miss-specified model as $Y = \beta_1 x_1 + u$

And $$\beta_1 = \frac{\sum x_1^2 Y}{\sum x_1^2}$$

Thus the question of is β as estimated unbiased? To answer the question, we substitute for the true value of Y and take the expected value. Simplified and taken the second part we arrive at

$$E(\beta_1) = \beta_1 + \beta_2 \frac{(\sum x_1 x_2)}{\sum x_1^2}$$

Note that, Unbiased means $E(\beta) = \beta$, hence, equation 5 shows clearly that β_1 is biased as the 2nd part is added and is not equal to zero, because x_2 is (relevant) an important variable. Thus if x_1 and x_2 are close substitutes and we know now that it is biased thus we may want to find out if it was over/under evaluated or biased.

Note that, β_2 is the co-efficient of the omitted variable and what is in the bracket is the slope of the co-efficient in the regression of x_2 and x_1. Thus our biased is the co-efficient of the omitted variables and the (explanation of bracket item).

The extent of biasness depends on $\beta_2 \frac{(\sum x_1 x_2)}{\sum x_1^2}$. Hence, if β_2 is

positive and

$\frac{(\sum x_1 x_2)}{\sum x_1^2}$ is also positive, then β_1 is over estimated and therefore,

positively bias.

Consider an earnings function

$W: = \alpha_1 + \alpha_2 S_1 + \alpha_3 A_1 + e$ - Correct equation

W_1 = wage, S_1 = years if schooling, A_1 = ability and our estimated equation is

$w_1 = \alpha_1 + 2 S_1 + e$

Therefore, α_2 will be biased thus

$$E(\alpha_2) = \alpha_2 + \alpha_3 \frac{(\Sigma S_1 A_1)}{ES_1^2}$$

To sign the bias (+) (+)

Thus we will end up over estimating (+) α_2. Thus the omission of relevant variable has implications.

11.3 Consequence of omission of important variable (under fitting or miss-specification)

(1) If the omitted variable (x_2 for instance) is related to the included variable x_1, the correlation coefficient between the two variables x_1 and x_2 will be non-zero. Hence, β_1 and β_2 will be bias and inconsistent. And the bias does not disappear as the sample size gets larger.

(2) If x_1 and x_2 are uncorrelated, only $\beta_1$will remain bias, while β_2 is unbiased.

(3) The variance of the disturbance term will be (too large) incorrectly estimated.

(4) The conventional measure of variance of β_1 estimate will be a bias estimator of the variance of the true β_1

(5) Consequently, the usual confident interval and the test of hypothesis procedure are likely to give misleading

$E(U_i, U_j)\ 6^2x_2, \qquad I = j$

Then $Y/\sqrt{x^2}_i = \alpha/\sqrt{x_i}^2 + \beta x/\sqrt{x_i}^2 + U/\sqrt{x_i}^2$

$Y/x = \alpha/x + \beta x/x + u/x$

$E(u/x)^2 = E(u^2)/x^2$

$= 6^2x_i^2/x^2_i$

$= 6^2$

(EXR) Given the form of Heteroscedasticity as $E(U_i\ U_j) = 6^2x^3$

Then let $z = x^3$, so that

$E(U_i, U_j) = 6^2_z$

$u/\sqrt{z} = \alpha/\sqrt{z} + \beta/\sqrt{z} + U/\sqrt{z}$

$E(U^2/\sqrt{z}) = \Sigma\ (U_2/(\sqrt{z})^2)$

$\Sigma\ (U2/(\sqrt{z})^2) = \Sigma(U^2)/z_i$

Replacing z_i for x3

$= 6^2z_i/z_i$ or $6^2x^3/x^3 = 6^2$

10.7 Multicollinearity

$$E(x_1, x_2) \neq 0$$

Definition of Problem

It occurs when $E(x_1\ x_2) \neq 0$, = multicollinearity which can also be written as $X_i = KX_2$.

Under normal circumstances, economic variables are related to each other.

Multicollinearity is a problem of multiple correlations between two or more explanatory variables.

$$Y = \alpha + \beta_1 x_1 = \beta_2 x_2 + U$$

Identification of the Problem

i. Plot the graph of X_1 and X_2

ii. Find the coefficient of correlation between X_1 and X_2 or the rank correlation of X_1 and X_2 or any measure of association between X_1 and X_2.

iii. If the coefficient of correlation is significant or high, there is the problem of multicollinearity.

Consequence of the Problem of Multicollinearity

Ex. 1; If $x_1 = kx_2$, when $Y = \alpha + \beta_1 x_1 = \beta_2 x_2 + U$

Then the new equation can be written as

$$Y = \alpha + \beta_1 k x_2 = \beta_2 x_2 + U$$

$$= \alpha + (\beta_1 k + \beta_2) x_2 + U$$

Where $\lambda = \beta_1 k + \beta_2$

Hence, $y = \alpha + \lambda x_2 + U$, the impact of x_1 and x_2 is diffused in to λ.

NB

a) If X_1 is a policy variable you cannot see the impact on Y.

b) Can β_1 and β_2 be estimated separately? The answer is no.

c) If β_1 and β_2 cannot be estimated separately, it follows that the impact of X_2 cannot be felt on Y also;

Summary of the consequences of Multicollinearity

i. One of the variables will disappear

ii. More difficult to estimate parameter separately

iii. It affects the unbiasness of the BLUE properties.

iv. Estimate of the coefficients are indeterminable.

v. The standard errors of the estimates become infinitely large

Further Explanation

Note if $x_i = kx_2$, where $x_i = x_i - x$. since $x_i = kx_2$, it follows that $x_i = x_1 - x_1 = k(x_2 - x_2)$.

Ex. 1: Choose any of the formulae for var (β_1) or var (β_2) and substitute $x_i = kx_2$ and see what happens.

$$\text{Var}(\beta_l) = \frac{\sigma^2_u \Sigma X^2_2}{\Sigma(kx_2)^2 \Sigma x^2_2 - \Sigma(kx_2x_2)^2}$$

where, $\Sigma(xi)^2 = \Sigma(kx_2)^2$

$= \Sigma k^2 x^2_2$

$X^2_i = (kx_2)^2$

$= k_2 x^2_2$

$$= \frac{\sigma^2_u \Sigma x^2_2}{0}$$

= undefined or indeterminable

Ex. 2: Given $\hat{\beta}i = \dfrac{\Sigma x_i y \, \Sigma x^2_2 - \Sigma x_2 y \, \Sigma x_i x_2}{\Sigma x^2_1 \, \Sigma x^2_2 - \Sigma (x_i x_2)^2}$

Substitute $x_i = kx_2$ and see what happens

$$\hat{\beta}_i = \frac{\Sigma k x_2 y \, \Sigma x^2_2 - \Sigma x_2 y \, \Sigma k x_2 x_2}{\Sigma (k x_2)^2 \, \Sigma x^2_2 - \Sigma (k x_2 x_2)^2}$$

$$= \frac{\Sigma k x_2 y \Sigma x^2_2 - \Sigma x_2 y \, \Sigma k x_2 x_2}{\Sigma (k x_2)^2 \, \Sigma x^2_2 - \Sigma (k x_2 x_2)^2}$$

$$= \frac{\Sigma k x_2 y \Sigma x^2_2 - \Sigma x_2 y \, \Sigma k x_2 x_2}{0} \quad = \text{undefined}$$

For $\beta_2 = \dfrac{\Sigma x_2 y \, \Sigma x^2_1 - \Sigma x_1 y \, \Sigma x_i x_2}{\Sigma x^2_1 \, \Sigma x^2_2 - \Sigma (x_i x_2)^2}$

Given that $x_2 = kx_i$

Substitute $x_2 = kx_1$ and see what happens

$$\hat{\beta}_2 = \frac{\Sigma (k x_1 y) \, \Sigma x^2_1 - \Sigma x_1 y \, \Sigma (x_i k x_1)}{\Sigma x^2_2 \, \Sigma (k x_1)^2 - \Sigma (x_i k x_1)^2}$$

$$= \frac{\Sigma k x_1 y) \Sigma x^2_1 - \Sigma x_i \, y \Sigma (x_i k x_1)}{\Sigma x^2_1 k^2 \, \Sigma x^2_1 - \Sigma x^2_1 k^2 \Sigma x^2_1}$$

$$\beta_2 = \frac{\Sigma (k x_1 y) \Sigma x^2_1 - \Sigma x_1 y \, \Sigma x_1 k x_1}{0}$$

$$= \frac{k(\Sigma x_1 y)(\Sigma x^2_1) - \Sigma x_1 y \, k(\Sigma x_1 x_1}{0}$$

$$= \frac{k(\Sigma x_1 y)(\Sigma x^2_1) - \Sigma (x_1 y) \, (\Sigma x^2_1)^2}$$

$$\beta_2 = \frac{0}{0/0}$$

Sources of the Problem of Multicollinearity

i. Mis-specification of the mathematical form of the model

ii. Omission of variable

iii. Aggregation error

iv. Measurement error

v. Erratic element inherent in human behavior

vi. Misspecification of macro-variables

vii. Interpretation in the statistical observation

viii. Misspecification of the true random term (u)

Supposing we have $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ----+ U$

The solution is to look at the specification of the model e.g. C= f (GDP, IN, W) where

GDP = Gross Domestic Product

IN = Net Investment and

W = Nation wealth

Solution to the problem of Multicollinearity

i. Run sequential regression analysis. This mean that, after regressing the entire model, the researcher will omit the last variable in the model and regress and on and on.

ii. Find R, note if it improves as you regress sequentially, then the problem of multicollinearity is solved, and vice versa.

$Y = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + ----+ U$

$Y = \alpha + \beta_1X_1 + U, \qquad R^2_1$

$Y = \alpha + \beta_2X_2 + U, \qquad R^2_2$

$Y = \alpha + \beta_3X_3 + U, \qquad R^3_3$

NB: X_i could mean GDP, X_2 – Net investment and X_3 – Nation's wealth

If X_1 and X_2 are regressed, and the R_2, is not improved upon , then GDP in trying to explain consumption is removed or omitted, such that, the other two IN and "W' remain and vice versa.

The same procedure follows for "IN' and "W".

CHAPTER ELEVEN

11.1 STATISTICAL BIAS

Statistical bias is the difference between the expected value of the estimate and the true parameter. When a multiple regression model is estimated it is assumed that all relevant variables have been included. Thus the question; what happens when a relevant variable is omitted.

11.2 OMMISSION OF RELEVANT VARIABLE OR MODEL MIS-SPECIFICATION

Consider $Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$ - (1) and we estimated

$$Y = \alpha + \beta_1 x_1 \quad - (2)$$

What is the implication, looking at the difference between the correct model (1) and the estimated model (2).

The question to ask is will the $E(\beta) = \beta$ or will β remain an unbiased estimate? The problem is alternatively called model miss-specification.

To check this, we will consider the equations without the intercept

i.e $Y = \beta_1 x_1 + \beta_2 x_2 + u$ - (3)

$$Y = \beta_1 x_1 \quad - (4)$$

Unbiased as the $E(\beta_1) = \beta_1$

From the estimated equation (3) $\beta_1 = \frac{\Sigma x_1 Y_1}{\Sigma x_1^2}$

We need to find $E(\beta)$, if $(E(\beta) = \beta)$ the relationship holds.

$E(\beta) = \beta$ = unbiased

STEPS TO ESTABLISH THE ABOVE

$\beta_1 = \frac{\Sigma x_1 Y}{\Sigma x_1^2}$ = Co-efficient of x_1 in the estimated equation

But recall that the true value of $Y_1 = \beta_1 x_1 + \beta_2 x_2 + U$, thus substituting we have

$$\beta_1 = \frac{\Sigma x_1 (\beta x_1 + \beta x_2 + U)}{\Sigma x_1^2}$$

Taking the expected value thus we arrive at

NB. Taking the (E) of the formula applies only to V=0

$$E(\beta) = \frac{E(\Sigma x_1 (\beta_1 x_1 + \beta_2 x_2) + E(u_i)}{\Sigma x_1^2}$$

solve the bracket

$$= \frac{\Sigma x_1^2 \beta_1 + \Sigma \beta_2 x_1 x_2}{\Sigma x_1^2} \quad \text{or} = \frac{\beta_1 \Sigma x_1 2 + \beta_2 \Sigma x_1 x^2}{\Sigma x_1^2}$$

$$= \frac{\beta_1 \Sigma x_1^2 + \beta_2 \Sigma_1 x_1 x^2}{\Sigma x_1^2}$$

$$= \frac{\beta_1 \Sigma x_1^2}{\Sigma x_1^2} + \frac{\beta_2 \Sigma_1 x_1 x^2}{\Sigma x_1^2}$$

$E(\beta) = \beta_1 + \frac{\beta_2 (\Sigma x_1 x_2)}{\Sigma x_1^2}$ = The scope of coefficient of the

regression of x_2 on x_1

c) Suspect Heteroscedasticity if the relationship is high

d) Find the correlation of x_i and e_i or using rank correlation, if it is high suspect Heteroscedasticity.

Consequences of Heteroscedasticity

Once you start having a varying variance, it affects the Best minimum variance of the BLUE properties.

The variance of the estimate (var (β)) becomes indeterminable because it s changing, it means that the variance of the parameter will be changing. This implies that the best minimum variance in the Blue property is invalid.

Sources of Heteroscedasticity

1) Misspecification of the mathematical form of the model

2) Aggregation problem

3) Omission problem or omission of the variable

4) Measurement error

Solution to Heteroscedasticity

Assuming we have $Y_i = \alpha + \beta x_i + U_i$, the var $(\beta) = 6^2$

Meaning U is constant, if it is not constant we have var (ui uj)$\neq 6^2 x_i$ $i \neq j$

The solution is to take the square root of x, or deflate the model. I.e. dividing the entire model through by the square root of x_i

Divide $Y_i = \alpha + \beta x_i + u_i$, by $\sqrt{x_i}$

Suppose $\dot{Y}_i = \alpha + \beta x_i + U_i$,

Then $\frac{Yi}{\sqrt{xi}} = \frac{\alpha}{\sqrt{xi}} + \frac{\beta}{\sqrt{xi}} + \frac{u}{\sqrt{xi}}$ hence $\frac{u}{\sqrt{xi}} = 6^2$

Which is constant for instance

$E\frac{(U)^2}{\sqrt{x}} = E\frac{(U_i^2)}{x}$ Since $E(U^2) = 6^2x_i$, substitute

$= \frac{6^2x_i}{x_i}$

$= 6^2$

Y	X	Y*	X*	U*
.	.	.	.	.
.	.			
Y_n	X_n			

$Y/\sqrt{x_i} = \alpha/\sqrt{x_i} + Bx_i/\sqrt{xi} + U_i/\sqrt{x_i}$

$Y^* = \alpha^* + \beta x^* + U^*$

Where $Y^* = Y/\sqrt{x_i}$

$\alpha^* = x/\sqrt{x_i}$

$U^* = U_i/\sqrt{x_i}$

$E(U^*) = E(u/\sqrt{x_i})^2$

$= (u^2)/x_i$ But $E(U^2) = 6^2x_t$

Therefore, $= 6^2x_i/x_i = 6^2$

Suppose the form of Heteroscedasticity

conclusions about the statistical significance of the estimated parameter.

(6) The forecast and its confident interval will be unreliable.

CHAPTER TWELVE

12.1 ANOVA

The basic aim is to split total variation of a variable (around its mean) into component which may be artributed to specific (additive) causes.

It is use in its statistical form to answer the question.

1. does a significant difference exist any where among three or more sample?
2. did all samples come from the same pop?
3. did at least one sample come from different pop?

(ANOVA)Analysis of Variation is sometimes used alongside regression analysis in order to:

1. Test the ovral significant of regression.
2. To obtain the test of improvement in the fit by introducing additional explanatory variable in the function.
3. To test of equality of coeficient obtain from different samples.

It is significantly the same with regression analysis since both of them are aimed at spliting total variation into additive components.

Total variation = Explain variation + unexplained

$$TSS = ESS + RSS$$

However, the main difference between the two methods is that, apart from the above, regression provide numerical value for the influence of various explanatory factor.

ANOVA Test of difference between means.

$$TSS = ESS + RSS$$

To analyse these variation we first of all look at the causes and soruces of variation which is rooted in the mean.

Therefore we look at within group means and between group mean.

To do this we have to formulate our hypothesis

$Ho : \mu_1 = \mu_2 = \mu_3$ or $Ho : F = 0$

Hi :. μj not equal or $Hi : F \neq 0$

If the mean are same, i.e. if Ho is true, the entire population is considered normally distrributed with mean $\mu(=\mu_1, \mu_2, \mu_3)$ and standard deviation Ó i.e

$y \infty N(\mu, Ó)$,

therefore,we estimate the common mean

$$\mu = \frac{\sum Y}{N} = \frac{\sum\sum_{ji} Y_{ji}}{N} = y$$

Between $\sigma^2 = \frac{\sum nj\,(Y-Y)^2}{K-1} = MSSB$

With $\sigma^2 = \frac{\sum\sum (Y_{ji}-Y_j)^2}{N-K} = MSSW$

Hence $\Sigma_{nj}(Y-Y)^2 + \Sigma\Sigma(Y_{ji} - Y_{ji})2 = TSS = \Sigma\Sigma(Y_{ji} - Y)$

SSB + SSW = TSS

(Between variation) + (within variation) = Total variation

$$F^* = \frac{\Sigma nj(Yji - Y)^2/K-1}{\Sigma\Sigma(Y_{ji} - Y_{ji})^2} = \frac{MSSB}{MSSW}$$

Where nj = size of j sample

N = Σ nj= size of pooled (enlarge) sample

K = no of samples

$$F^* = \frac{\text{Estimate of variance from between the mean variation}}{\text{Estimate of variance from within the sample variance}}$$

ANOVA TABLE

(1)	Sum of square (2)	Degree of freedom (3)	Mean square (4) = (2): (3)	F (5)
Between the means	$\Sigma uj(Y_{ji}-Y)^2$	V =K-1	$\Sigma nj(Y-Y)^2$	$F = \frac{\Sigma nj(Y-Y)^2/K-1}{\Sigma\Sigma(Y_{ji}-Y_j)/N-K}$
Within the samples	$\Sigma\Sigma(Y_{ji} - Y_j)^2$	V=N-K	$\Sigma\Sigma(Y_{ji}-Y)^2$	
Total variation	$\Sigma\Sigma(Yji - Y)^2$	V= N-1		

Example 12.1

Suppose we want to find out whether or not there is a significant difference in performance in statistic among the five departments that made up the faculty of social science, suppose also you conducted a statistics test in all the debts and obtained the follwing results.

Soc.	Psy	Geo	Eco	Pol
8	7	5	7	3
3	9	4	6	3
5	5	4	8	6
9	4	6	9	3
9	6	2	5	5
8	6	9	7	4

$\Sigma x_1 = 42$	$\Sigma x_2 = 36$	$\Sigma x_3 = 30$	$\Sigma x_4 = 42$	$\Sigma x_5 = 24$
Σx_1 324	$\Sigma x_2 = 234$	$\Sigma x_3 = 178$	$\Sigma x_4 = 304$	$\Sigma x_5 = 104$
$x_1 = 7$	$x_2 = 6$	$x_3 = 5$	$x_4 = 7$	$x_5 = 4$

Ho : F = 0 = The five accademic departments are not likely to be different in terms of students statistics scores.

Ho : F ≠ 0 = The five academic departments are likely to be different in terms of student statistics scores.

ANOVA

SSW = $\Sigma\Sigma (x - x)^2$

$(x_1 - x_1)$	$(x_1 - x_1)^2$	$(x_2 - x_2)$	$(x_2 - x_2)^2$
8-7= 1	1	6 =1	1
3-7= -4	16	9-6 = 3	9
5-7= -2	4	4-6 =-2	4
5-7 = 2	4	4-6 =-2	4
9-7= 2	4	6-6 = 0	0
8-9 =1	1	6-6 = 0	0
	30		18

$(x_3 - x_3)$	$(x_3 - x_3)$	$(x_4 - x_4)$	$(x_4 - x_4)$	$(x_5 - x_5)$	$(x_5 - x_5)$
5-5 = 0	0	7-7 = 0	0	3-4 = -1	1
4-5 =-1	1	6-7 = 1	1	3-4 = -1	1
4-5 =-1	1	8-7 = 1	1	6-4 = 2	4
6-5 = 1	1	9-7 = 2	4	3-4 = -1	1
2-5 =-3	9	5-7 = -2	4	5-4 = 1	1
9-5 = 4	16	7-7 = 0	0	4-4 = 0	0
	28		10		08

SSw = 30 + 18 + 29 + 10 + 08 = 94

MSSw = $\frac{\Sigma\Sigma(x - x)}{N - k} = \frac{94}{}$

SSB = $\Sigma n(x-x)^2$

$6(7-5.8)^2 = 6(1.2)^2 = 6(1.44) = 8.64$

$6(6-5.8)^2 = 6(0.2)^2 = 6(0.04) = 0.24$

$6(5-5.8)^2 = 6(-0.8)^2 = 6(0.64) = 3.84$

$6(7-5.8)^2 = 6(1-2)^2 = 6(1.44) = 8.64$

$6(4-5.8)^2 = 6(-1.8)^2 = 6(3.24) = \underline{19.44}$

40.8

$$MSSB = \frac{SSB}{k-1} = \frac{\Sigma n(x-x_T)^2}{k-1}$$

Therefore, $F^* = \frac{MSS\ B}{MSS\ W} = \frac{40.8}{94} = 2.71$

Total variable $\Sigma (x-xT) = \Sigma ni(x-xT) + \Sigma \Sigma (x-x)2 = 134.80$

Short method but not necessarily easy mehod.

It start by calculating the following:

1. $\frac{(\Sigma\Sigma x)^2}{\Sigma N} = \frac{(40+36+30+42+24)^2}{30} = \frac{(174)^2}{30} = \frac{30276}{30}$

= 10092

$\Sigma\Sigma x$ = the $\Sigma$of all sample accross all groups

ΣN = sum of sample sizes

II. $\Sigma\Sigma x^2 = (\Sigma x^2{}_1 + \Sigma x^2{}_2 + \ldots\ldots + \Sigma x^2{}_5) = (324+234+\ldots\ldots+104)$
$= 1144$

$\Sigma\Sigma x^2$ = the sum of sum of x squares

III. $\dfrac{\Sigma(\Sigma x)^2}{n} = ((\Sigma x)^2 + (\Sigma x2)^2 + \ldots\ldots + (\Sigma x)^2) = \dfrac{6300}{6}$
$= 1050$

$$SSB = \frac{\Sigma(\Sigma x)^2}{N} - \frac{(\Sigma\Sigma x)^2}{\Sigma n}$$

$$SSW = \Sigma\Sigma x^2 - \frac{\Sigma(\Sigma x)^2}{N}$$

$$SST = \Sigma\Sigma x^2 - \frac{(\Sigma\Sigma x)^2}{\Sigma N}$$

Having obtained the values of SSB, SSW, and SST, we procceed to set up the ANOVA table

12.2 ANOVA SUMMARY TABLE

Sources of variation	df	ss	ms	F*
Between groups	4	40.8	10.20	2.71
Within groups	25	94.00	3.76	
Total	29	134.80		

NB: df between = k-1 where k = no of groups under comparism

Df Within = N-k where K = no of groups, N = no of cases across all samples (i.e. 6x5 = 30).

Since, we have now obtained our F* to be 2.71, we then measure F* against F_T

In the table, the difference between groups are horizontal, while degree of freedom (df) and the sums of squares (ss) can be located on the vertical axis.

Df between groups = 4, the df within groups = 25, P = 0.05 F_T = 2.76. Hence we accept Ho since $F^* < F_T$ and conclude that there are not different among the five departments in terms of the students statistics scores. Therefore, the observed difference between groups is attributed to chance and or sample error.

CHAPTER THIRTEEN

13.1 THE X^2 CHI-SQUARE

The chi-square X^2 is a probability distribution that tries to answer the question "How well does a model depict reality" In other words it is a test of goodness of fit of a model to a particular problem or situation. It is also used to test the independence or otherwise of an event.

It is a technique for comparing the theoretical or expected frequencies with actual or observed frequencies.

Properties of Chi- square x^2

1. X^2 is a random variable that cannot assume negative value.

   Eg we could let $u = x^2$ and call it u distribution. However, squaring 'chi" remained as that x^2 can only be zero or greater.

2. The distribution of chi- square (x^2) depends only on the degree of freedom v.
3. X^2 distribution is always skewed to the right. The larger the Y the less the skeweness.
4. The mean of x^2 is given by the degree of freedom $E(x^2) = v$.
5. The variance is twice the degree of freedom; var $(x^2) = 2v$.

13.2 X^2 TEST OF GOODNESS OF FIT

Steps

1. State the null and alternative hypothesis. e.g.

H_o : the sample size have not in any way describe the pops

H_i : the sample size have describe the pops

2. Determine the size of random sample.

3. Specify the number of frequency distribution

4. Calculate the theoretical frequencies.

5. Calculate the degree of freedom.

6. Choose the level of significance (i.e. the probability of type I error)

7. State the decision rule

8. Calculate the X^2

9. Compare the calculated value against the table value.

Worked example 13.1

When a coin activated dispenser releases a volume of liquid which is call a fill. Due to particular constructions of the dispenser, we know that the fill is normally distributed with standard deviation of 0.10 ounce regardless of the fill setting or average volume of liquid dispensed per fill. A random sample of 49 fill produces the following data:

7.85	7.98	8.06	8.11
7.86	8.99	8.06	8.11
7.87	8.99	8.07	8.12
7.87	8.01	8.07	8.16
7.88	8.03	8.07	8.16
7.89	8.03	8.08	8.17
7.92	8.04	8.09	8.19
7.94	8.05	8.09	8.21
7.95	8.05	8.09	8.21
7.96	8.05	8.10	8.22
7.97	8.05	8.10	8.24
7.97	8.06	8.10	8.26

Test that the pop is normally distributed with standard deviation of 0.10.

Step 1

Ho : Pop is not normally distributed

Hi : The pop is normal distributed.

Step II n = 49

Step III

	Observation X	frequency
	7.80 to less than 7.90 ounce	6
i.e	7.90 ---------- 8.00	9
	8.00 ---------- 810	19
	810 ---------- 820	10
	820 ---------- 830	5
	Observed frequency	49

Step IV

Less than 7.90 ounce

7.9 To less than 800

8.00 To less than 8.10

8.10 To less than 8.20

8.20 ounce or more

Lower class limit XL		upper class limit XL	standard normal deviations $Z_L = \frac{X_L - 8.05}{0.10}$	$Z_U = \frac{X_U - 8.05}{0.10}$	Area of class $P(Z_L < Z < Z_U)$	theoretical frequency (fi) 49 $p(Z_L < Z < Z_U)$
-		7.90	$-\infty$	-1.50 =	0.6669	11.8433
7.90	-	8.00	-1.50	-0.50 =	0.3830	18.7670
8.00	-	8.10	-0.50	+0.50 =	0.2417	11.8433
8.10	-	8.20	+0.50	+1.50 =	0.668	3.2732
8.20		-	+1.50	$+\infty$ =	10000 -	49.0000

Step 5

Calculate df v.

$v = k - 1 - m$

K = no. of class interval

M = no of part estimate

Therefore, 5 - 1 - 1 = 3

Step 6 $\alpha = 0.05$

Since at v = 3 and $\alpha = 0.05$ the critical value = 7.81

Step 7 Decision rule

If $X^2 > 7.81$ reject Ho

$X^2 < 7.81$ accept Ho

Step 8 Calculate $X^2 = \frac{\Sigma(fo - f1)^2}{F1}$

Event	Observed frequency fo	theoretical frequency fi	(fo – fi)	$(fo-fi)^2$	$(fo-fi)^2/fi$
1	6	3.2732	2.7268	7.4354	2.27
2.	9	11.8433	-2.8433	8.0844	0.68
3.	19	18.7670	0.2330	0.0543	0.00
4.	10	11.8433	-1.8433	3.3978	0.29
5.	5	3.2732	1.7268	2.9818	0.91
		49.0000	0.0000		4.15

Since 4.15< 7.81 we accept null hypothesis and conclude that the sample data supports the assumption that the pop is normally distributed.

If $X^2 = 0$, the observed and the theoretical frequency agree exactly

If $X^2 > 0$ they do not agree exactly

The larger the value of X^2, the greater is the discrepancy between the observed frequency and the theoretical frequency.

13.3 X^2 Test Of Independence Of Events

"Association" and "Independence" are two important concepts in Economics. If two event are independent, then the occurrence of one event does not affect the probability of the other event occurring.

X^2 test is also a technique that is used to test the independence or otherwise of event occurring.

Example 13.2

A statistic professor argues that the afternoon class do not perform as much as morning classes. While the department chairman argues that between 8am and 5pm it does not make any difference when the class meets. The instructor locates records for past semester. The instructor claimed that, the same lecture was delivered to each of the classes, the same homework and examinations were given and all students had met the course prerequisites. The grades are present thus:

Letter grade

Time the class met	A	B	C	D	F	Total
Morning	12	18	13	11	6	60
Afternoon	3	7	17	4	9	40
Total	15	25	30	15	15	100

Let r = raw and k = column, then we have r times k (r x k) contingencies table.

Since r = 2 and k = 5 therefore,

r x k = 2 x 5 = 10 = frequencies of occurrence of events.

H_o : Event A and B are independent

H_i : Event A and B are not independent

Or

H_o : The time class meet and letter grade are independent.

H_i : The time class meet and letter grade are dependant.

In the case of the test of independence, the null hypothesis does not suggest the theoretical distribution it only suggest the algebraic relationship with r x k.

Therefore to calculate the theoretical frequency,

$$\frac{\text{(Total for row i) (total for colume j)}}{\text{Grand total}} = \text{frequency for cell ij.}$$

Example,

Time		Letter grades					
Total	A	B	C	D	F		
	AM	60.15 =9	60.25 =15	60.30 =18	60.15 =9	60.15 = 9	60
		100	100	100	100	100	
	PM	40.15 =6	40.25 = 10	40.30 =12	40.15 = 6	40.15 =6	40
	100	100	100	100	100	100	
Total	15	25	30	15		15	100

$$X^2 = \Sigma\frac{(f_o - f_1)^2}{f1} \quad \text{or} \quad = \Sigma\Sigma\frac{(fo_{ij} - f1_{ij})^2}{f1_{ij}}$$

Time	letter groups	fo	fl	(fo-fl)	$(fo-fl)^2$	$(fo-fi)^2/fi$
PM	A	12	9	3	9	1.00
AM	B	18	15	3	9	0.60
AM	C	13	18	-5	25	1.39
AM	D	11	9	2	4	0.44
AM	F	6	9	-3	9	1.00
PM	A	3	6	-3	9	1.50
PM	B	7	10	-3	9	0.90
PM	C	17	12	5	25	2.08
PM	D	4	6	-2	4	0.67
PM Total	F	9 / 100	6 / 100	-3 / 0	9	1.50 / 11.08

Degree of freedom v = (r -1) (k-1)

(2-1) (5-1) = 4 degrees of freedom

With $v = 4$ and $\alpha = 0.05$, X^2 given by the table = 9.49

Hence if $X^2 = > 9.49$ reject Ho (and conclude that time and grade are dependent i.e. the instructors hypothesis.)

If $X^2 < 9.49$ accept Ho (and conclude that time and grade are independent.

(The chairman's hypothesis).

Since $X^2 = 11.08 > 9.49$: therefore, we reject the null hypothesis.

CHAPTER FOURTEEN

14.1 TEST OF HYPOTHESIS

A hypothesis is an assumption about the frequency distribution of a random variable. It is a tentative statement that can be subjected to test. Hypothesis testing is concern with the developing rules or procedure for deciding whether to reject or accept the hypothesis. It is a way of determining whether a given observation or finding is compatible with some stated hypothesis. That is whether the finding is sufficiently close to the hypothesized value so that we do not reject the stated hypothesis.

Hence, if on a priori, we believe that the time value of β is unity and the sample estimated value is 0.498 is this value consistent with the stated hypothesis?

The stated hypothesis is known as the null hypothesis; it is denoted by the symbol H_0. This hypothesis is tested against an alternative, and it is also called maintained hypothesis denoted by H_1.

The null hypothesis is set in such a way that available evidence will nullify it, and hence against the direction of the research or observation such that it can be nullify easily. Example, H_0: $\beta = 0$. The alternative hypothesis which is a rival hypothesis is set in terms of the theoretical postulation.

14.2 STEPS INVOLVE IN TEST OF HYPOTHESIS

There are about (6) six steps involve in testing of hypothesis.

1. Formullations of hypothesis

2. Choose the level of significance.

3. Choose the location of the cretical and or rejection region.

4. Choose the appropriate test statistic.

5. Compute the sample vlue of the choosen test statistic.

6. Compare the actual sample value with the critical or table value of the statistic.

STEP 1 Formulation of hypothesis.

(a) H_O: Null Hypothesis is set against a rival hypothesis in such a way that available evidence will nullify it.

(b) Hi: Altanative or rival hypothesis is set again Ho as a rival hypothesis in order to direct the researcher.

E.g. H_o: μ =30 and

H_1: $\mu = 30$ - then, it is a two tail test

The two tail is divied into two

If $\mu < 30$ - One tail test to the left

$\mu > 30$ - One tail test to the right

A hypothesis is an assumption which we make about a population parametre.

In the test of hypothesis, we begin by making some assumptions about the true value of the population or estimate to decide whether or not the assumption is acceptable.

The hypothesis which would react to test is called the null hypothesis, it is set in such a way that available evidence will nullify it.

It is called a null hypothesis because, it implies that there are no difference between the true pop paramere and the (sample information or estimate) hypothesized value.

The alternative hypothesis represent a rival assumption about the population parametre. In otherwords, it is a counter proposition to the null hypothesis.

The Null hypothesis is denoted by Ho while alternative hypothesis is denoted by Hi.

The alternative hypothesis implies that the population parametre is not equall to the specified (sample) value.

The alternative hypothesis is important in defining the location of the rejection or the critical region of the test.

For the purpose of conducting a test of hypothesis about a population parametre, the whole sets of the values of the population is devided into two regions.

-The acceptance region and the rejection or critical region.

The acceptance region includes the values of (x) the population which have high probabilities of been observed while the reject or critical region include the value of (x) which are not likely to occur.

These highly unlikely values of x are those values whose total probability is less than 5% or less than 1% i.e those that define an area of the probability distribution equall to 0.05 and 0.01.

These probabilities are known as the level of significance of the test.

Step 2. Choose level of significance of the test.

In making a decision, one can never be 100% sure that he/she will make (or has made) the right decission, therefore, in making any decision one is likely to commit the followidng types of errors.

1. Type one error – occurs when one reject the null hypothesis when it is actually true.

2. Type two error – occurs when one accept null hypothesis when it is actually not true.

Fact / Decission	Ho. is true	Ho is False
Accept Ho Ho is true	Correct decission $(1-\alpha)$	Type II (β)
Reject Ho Ho is False	Type I Error α	Correct decision $(1-\beta)$

The probability that these errors will occur is given by

P(Type I error) = P(reject Ho/Ho is true) = α

and

P(Type II error)= P(accept Ho/Ho is False) = β

Of course, we would like α and β to be as small as possible, hence, α and β are called significance level i.e. The likelyhood of making type one error.

It is obvious that we want these probabilities of been wrong to be small (ie to be minimum).

That is why we usually assign a low value to the probabilities (ie probabilities of been wrong) and we call it level of significance of the test.

Choosing a certain level of significance involves our specifying the probability of encountering a type one error and in most cases we specifies a level of significance of 5% and in rear cases 1%.

Step 3. Choosing the location of the critical or rejection region.

The value of the variable x which have a very low probability of been observed i.e. a total proberbility of 5% or 1%.

In other words, the critical region includes values that correspond to the level of significance of the test.

The critical region may be choosen either at right tail or left tail of the distribution.

Where we have the critical region entirely on the right tail or left tail we are said to conduct a one tail test where the critical divides into the two tails or ends of the distribution, then we have a two tail test.

The decision of whether to choose a one tail or two tail critical region depend on the form in which the altanative hypothesis is empressed.

Hi :$\mu \neq 30$	located in the two tails
Hi : $\mu > 30$	right tail
Hi : $\mu < 30$	left tail

Step 4. Choosing the apropriate test statistic

The most commonly used statistic is the t, z and f statistic (ie statistic which are used in testing hypothesis).

The choice among the test statistic depends on the type of test you wants to cary out, the size of the sample and the information which we have about the population variane.

Step 5 Compute the sample value of the choosen test statistic

Using the sample information, we compute the sample value or the observed value of the choosen statistic which is denoted with asteric (z* t* f*) to distinguish it from the critical value otherwise known as theoritical value or tabule value.

Step 6 Compare the computed sample value with the critical or tabular value of the statistics.

Worked Example 14.1

The average weight of hones was found to be 100kg from previous records. Their standred deviation was 20kg. A new diet or horse-feed was inovated and was given to a large number of horses. A random sample of 35 horses was selected and the average wight is found to be 110kg. Using a 5% level of significance, can you say that there is a significant gain in the weight of the horses?

Step 1

$Ho : \mu = 100$

$Hi : \mu > 100$

Step 2 Level of significance is 5%

Step 3 - $Z_{0.05} = 1.640$

Step 4 (Apropriate stat)

Z statistic Where $Z^* = \frac{x - \mu}{s/\sqrt{n}}$

Step 5 Comput sample value

$$Z^* = \frac{x - \mu}{\frac{s}{\sqrt{n}}} = \frac{110 - 100}{\frac{20}{\sqrt{35}}} = \frac{10}{\frac{20}{5.91}}$$

$$= \frac{10}{3.38} = 2.95$$

Step 6 $Z_{0.05} = 1.640$

$Z^* = 2.95$

Since the Z* computed is greater than the tables Z value (i.e. $Z^* > Z_{0.05}$), and hence we reject Ho

$Z^* < Z0.05$ accept Ho

Therefore, there is a significant gain in the weight of the horses.

Example 14.2

A random sample of 100 recorded death in a developed country during the first year shows an average life span of 71.8 years with standard diviation 8.9 years. Does this seemed to indicate that the average life span today is greater than 70 yeas?

Use a 0.05%level of significance.

Step 1

Ho : $\mu = 70$ years

Hi : $\mu > 70$ years

Step 2 $\alpha = 0.05$

Step 3 $Z_{0.05} = 1.64$

Step 4 $Z^* = \dfrac{x - \mu}{s/\sqrt{n}}$

Step 5

$$Z^* = \frac{71.8 - 70}{\frac{8.9}{\sqrt{100}}} = \frac{1.8}{\frac{8.9}{10}} = \frac{1.8}{0.89} = 2.02$$

Step 6 $Z^* = 2.02$

$Z_{0.05} = 1.64$

Since $Z^* > Z$ reject Ho and conclude that the lifespan of today is more than 70 years.

Example 14.3

The average lenght of time for studens to revice their lecture in certain university is 50 min. A new proceedure using mordern computing machine is been tried. If a random sample is selected of 12 students had a revission time of 42 with a standard deviation of 11.9 minutes under the new system.

Test the hypothesis that the average revission time is now less than 50 minutes. Using 0.01% level of significance.

Step 1. Ho : $\mu = 50$ min

Hi : $\mu < 50$

Step 2 $\alpha = 0.01$

Step 3

Step 4 t table

Step 5 $t^* = \dfrac{x - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{42 - 50}{\frac{11.9}{\sqrt{12}}}$

$t^* = \dfrac{-8}{\frac{11.9}{3.46}} = \dfrac{-2.33}{3.44} = -2.72$

Step 6 Discussion

Since -2.72 < - 2.33, we accept Ho and conclude that the average revision time is not less than 50 minues.

Example 14.4

A couse in statisstics is thought to 2 seperate groups of students by two different method. In the first group 12 student were though by the conventional classroom method. The second group consisting of 10 is though by a new method.

At the end of the semester, the same examination was given to each group. The 12 students in the first group made an average grade of 85 with St. Div of 4 while the 2nd of 10 students made an average grade of 81 with st. Div of 5.

Test the hypothesis that the two methods are equal using 5% degree level of significance. Assuming that the population is normally distributed.

Step 1 Ho : $\mu = \mu_2$ or $\mu_i - \mu_2 = 0$

$H_i : \mu \neq \mu_2$ or $\mu_1 - \mu_2 \neq 0$

Step 2 level of sing $\alpha = 0.05\%$

Step 3 two tail test

Step 4 $t_\alpha = t_{0.025}$ with degree of freedom 2

Where df $v = \dfrac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{\left[\dfrac{s_1^2}{n_1}\right]^2}{n_1 - 1} + \dfrac{\left[\dfrac{s_2^2}{n_2}\right]^2}{n_2 - 1}} = \dfrac{\left[\dfrac{16}{12} + \dfrac{25}{10}\right]^2}{\dfrac{\left[\dfrac{16}{12}\right]^2}{12-1} + \dfrac{\left[\dfrac{25}{10}\right]^2}{10-1}}$

$$= \frac{(1.33 + 2.2)^2}{\frac{1.78}{11} + \frac{2.25}{9}} = \frac{(3.83)^2}{0.16 + 0.69}$$

$$= \frac{14.67}{0.85} = 17.26$$

$t^* < -t_{0.025}(v = 17)$ or $t^* > +t_{0.025}(v = 17)$

$t^* < -2.110$ and $t^* > +2.110$

$$= \frac{(1.33 + 2.2)^2}{\frac{1.78}{11} + \frac{2.25}{9}} = \frac{(3.83)^2}{0.16 + 0.69}$$

$$= \frac{14.67}{0.85} = 17.26$$

$t^* < -t_{0.025}(v = 17)$ or $t^* > +t_{0.025}(v = 17)$

$t^* < -2.110$ and $t^* > +2.110$

Step 5 Computation

$$t^* = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(s_1^2)}{n_1} + \frac{(s_2^2)}{n_2}}}$$

$$= \frac{85 - 81 - \mu 1 - \mu 2}{\frac{(16)}{(12)} + \frac{(25)}{(10)}}$$

Since according to our hypothesis $\mu_1 - \mu_2 = 0$

$$= \frac{85 - 81 - 0}{\sqrt{\frac{16}{12} + \frac{25}{10}}}$$

$$\frac{4}{\sqrt{1.33 + 2.5}} = \frac{4}{3.83} = \frac{4}{1.96} = 2.04$$

Step 6 Accept Ho and conclude that the two method of teaching are basically the same.

Example 14.5

Using eg 14.5 test the hypothesis that the 2 methods of learning are equally at 01/0 than at 10% level of significance. Assuming that the pop is aproximately normally distributed with equall variances.

$$Z = \frac{s = \mu s}{\sigma s}$$

For Means Here s = x, the sample mean

$\mu s = \mu x = \mu$ - Pop mean and

$\sigma s = \sigma x = \sigma/\sqrt{N}$

Where σ = pop st div α N = sample

Hence, $Z = \frac{x - \mu}{\sigma/\sqrt{N}}$

CHAPTER FIFTEEN

15.1 REVIEW OF STATISTICAL CONCEPTS

Sets -Properties And Notation

Set concepts and notation not only help us talk about certain mathematical ideas with greater clarity and precision, but are indispensable to a clear understanding of probability.

We can think of a set of any collection of objects specified in such a way that we can tell whether any given object is or is not in the collection. Capital letters, such as A, B, and C, are often used to designate particular sets. Each object in a set is called a **member** or **element** of the set. Symbolically.

a ∈A	means	" a is an element of set A"
a ∈ A	means	" a is not an element of set A"

A set without any elements is called the **empty** or **null set.** For example, the set of all people over 10 feet tall is an empty set. Symbolically

Θ represents "the empty or null set"

A set is usually described either by listing all its elements between braces { }, or by enclosing a rule within braces that determines the ments of the set. Thus, if P(x) is a statement about x, then

S = {x P (x) } mqans "S is the set of all x such that P (x) is true"

Recall that the vertical bar in the symbolic form is read "such that" following example illustrates the rule and listing methods of representing sets.

Example 15.1 RULE LISTING

{x / x is a weekend day} = {Saturday, Sunday}

$\{x \mid x^2 = 4\}$ = {1,3,5, ...}

X/X is an oddpositive counting number = {1,3,5...}

The three dots ... in the last set in Example 1 indicate that the pattern established by the first three entries continues indefinitely. The first sets in Example 1 are finite sets (intuitively, the elements can be counted the last set is an infinite set (if we try to count the elements, there is an end). When listing the elements in a set we do not list an element more than once.

Problem 15.1 Let G be the set of all numbers such that $x^2 = 9$.

(A) Denote G by the rule method.

(B) Denote G by the listing method.

(C) Indicate whether the following are true or false: 3 G, 9 G.

Answers (A) $\{x \mid x^2 = 9\}$ (B) {-3,3} (C) True, True

If each element of a set A is also an element of set B, we say that a subset of B. For example, the set of all women students in a class subset of the whole class, Note that the definition allows a set to be a sub-set of itself. If set A and set B have exactly the same elements, then the sets are said to be equal. Symbolically.

A C B means "A is a subset of B"

A = B means "A and B have exactly the same elements"

It can be proved that

Θ is a subset of every set

Example 15.2 List all the subsets of the set {a,b,c),

Solution {a,b,c}, {a,b}, {a,c} {a}, {b}, {c}, Θ

Problem 15.2 List all the subsets of the set {1,2},

{1,2}, {1}, {2}, Θ

SET OPERATIONS The union of sets A and B, denoted by A U B, is the set of all elements formed by combining all the elements of A and all the elements of B into one set. Symbolically.

UNION

A U B = {x x $\in$ A or X/$\in$ B}

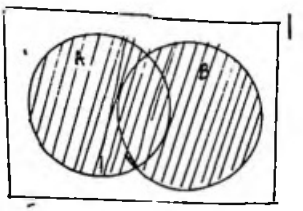


FIGURE 1: A U B is the shaded region

Here we use the word "or" in the way it is always used in mathematics: that is, x may be an element of set A or set B or both. Venn diagrams are useful aids in visualizing set relationships. The union of two sets can be illustrated as shown in Figure 1.

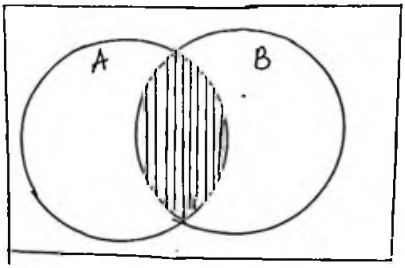


Figure 15.2: A ∩ B is the shaded region.

The **intersection** of sets A and B, denoted by A B, is set of elements in set A that are also in set B. Symbolically,

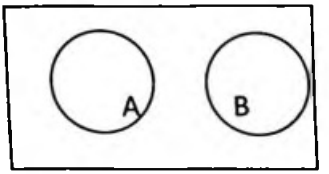
INTERSECTION

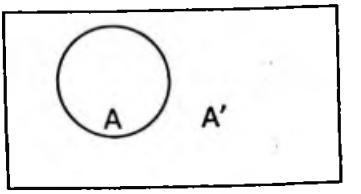
$$A \; b = \{X \; X \quad a \quad AND \quad X \quad b\}$$

This relationship is easily visualized in the Venn diagram shown in Figure 2.

If A B = Θ, then the sets A and B are said to be disjoint; this is illustrated in Figure 3.

The set of all elements under consideration is called the universal set U. Once the universal set is determined for a particular discussion, all other sets in that discussion must be subsets of U. We now define one more operation on sets called the complement. The complement of A (relative to U) denoted by A', is the set of elements in U that are not in A (see Fig.4). Symbolically.





igure:15.3aA B = Θ *fig. 13.3bThe complement of A is A'*

COMPLEMENT

$$A' = \{x \quad U \quad x \quad A\}$$

Example 15.3 If A = {3,6,9}, B = {3,4,5,6,7}, and C = {4,5,7}, then

A U B = {3,4,5,6,7,9}

A B = {3,6}

A C = Θ A and C are disjoint

C' relative to B is {3,6}

Problem 3 If R = {1,2,3,4}, S = {1,3,5,7), and T = {2,4}, find:

(A) {1,2,3,4,5,7} (B) {1,3} (C) Θ
(D) {1,3}

Example 15.4 From a survey involving 100 college students, a marketing research company found that seventy-five students owned stereos, forty-five owned cars, and thirty-five owned cars and stereos.

(A) How many students owned either a car or a stereo?

(B) How many students did not own wither a

car or a stereo?

Solutions Venn diagrams are very useful for this type of problem. If we let

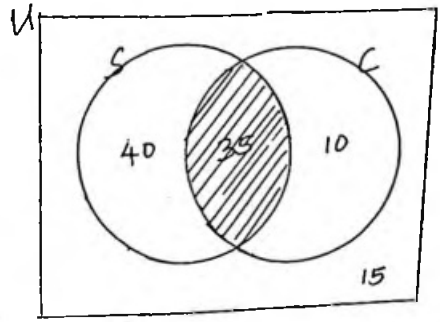
U = Set of students in sample (100)

S = Set of students who own stereos (75)

C = Set of students who own cars (45)

S C = Set of students who own cars and stereo (35)

Then



Place the number in the intersection first, then work outward.

(A) The number of students who own either a car or a stereo is the number of students in the set S U C. You might be tempted to say that this is just the number of students in S plus the number of students in C, 75 + 45 = 120, but this sum is larger than the sample we started with! What is wrong? We have actually counted the number in the intersection (35)

twice. The correct answer as seen in the Venn diagram, is 40 + 35 + 10 = 85

(B) The numbers of students who do not own wither a car or a stereo is the number of students in the set (S U C)': that is, 15.

Problem 4 Referring to Example 4:

(A) How many students owned a car but not a stereo?

(B) How many students did not own both a car and a stereo?

Answers (A) 10 [the number in S' C] (b) 65[the number in (S C)']

Note in Example 4 and Problem 4 that the word "and" is associated with "intersection" and the word "or" is associated with "union"

Sample Space, Sample Points, And Events

The set of all possible outcomes of a random, or chance, experiment is called the population, or sample space, and each member of this sample space is called a sample point. Thus, in the experiment of tossing two coins, the sample space consists of these four possible outcomes: HH, HT, TH, and IT, where HH means a head on the first toss and also a head on the second toss, HT means a head on the first toss and a tail on the second toss, and so on. Each of the preceding occurrences constitutes a sample point.

An event is a subset of the sample space, Thus, if we let A denote the occurrence of one head and one tail, then, of the preceding possible outcomes, only two belong to A, namely HT and TH. In this case A constitutes an event. Similarly, the occurrence of two heads in a toss of two coins is an event. Events are said to be mutually exclusive if the occurrence of one event precludes the occurrence of another event. If in the preceding example HH occurs, the occurrence of the event HT at the same time is not possible. Events are said to be (collectively) exhaustive if they exhaust all the possible outcomes of an experiment. Thus, in the example, the events (a) two heads, (b) two tails, and (c) one tail, one head exhaust all the outcomes; hence they are (collectively) exhaustive events.

PROBABILITY AND RANDOM VARIABLES

Probability

Let A be an event in a sample space. By P(A), the probability of the event A, we mean the proportion of times the event A will occur in repeated trials of an experiment. Alternatively, in a total of n possible equally likely outcomes of an experiment, if m of them are favorable to the occurrence of the event A. we define the ratio m/n as the relative frequency of A. For large values of n, this relative frequency will provide a very good approximation of the probability of A.

Properties of Probability. P(A) is a real-valued function[1] and has these properties:

1. $0 \leq P(A) \leq 1$ for every A.

2. If A, B, C.... Constitute an exhaustive set of events, then P(A + B +

C + . . .) = 1, where A + B + C means A or B or C, and so forth.

3. If A, B, C... are mutually exclusive events, then

P(A + B + C +...) = P(A) + P(B) + P(C) +..

Example 15.5

Consider the experiment of throwing a die numbered 1 through 6. The sample space consists of the outcomes 1, 2, 3, 4, 5, and 6. These six events therefore exhaust the entire sample space. The probability of anyone of these numbers showing up is 1/6 since there are six equally likely outcomes and anyone of them has an equal chance of showing up. Since 1, 2, 3, 4, 5, and 6 form an exhaustive set of events, P(1 + 2 + 3 + 4 + 5 + 6) = 1 where 1,2,3, .. means the probability of number 1 or number 2 or number 3, etc. And since 1, 2, , 6 are mutually exclusive events in that two numbers cannot occur simultaneously, P(1 + 2 + 3 + 4 + 5 + 6) = P(1) + P(2) + . .. + P(6) = 1 .

Random Variables

A variable whose value is determined by the outcome of a chance experiment is called a **random variable** (rv). Random variables are usually denoted by the capital letters X, Y, Z, and so on, and the values taken by then are denoted by small letters x, y, z, and so on.

A random variable may be either **discrete** or **continuous**. A discrete rv takes on only a finite (or countably infinite) number of value. For example in throwing two dice, each numbered 1 to 6, if we define the random variable X as the sum of the numbers

showing on the dice, then X will take one of these values: 2, 3,4,5,6, 7, 8, 9, 10, 11, or 12. Hence it is a discrete random variable. A continuous rv, on the other hand, is one that can take on any value in some interval of values. Thus, the height of an individual is continuous variable-in the range, say, 60 to 65 inches it can take any value, depending on the precision of measurement.

PROBABILITY DENSITY FUNCTION (PDF)

Probability Density Function of a Discrete Random Variable

Let X be a discrete rv taking distinct values $x_1, x_2. . . . . x_n$ .... Then the function

$f(x) = P(X = x_i)$ for i = 1. 2, . . . n

$= 0$ for $x \neq x_i$

is called the discrete probability density function (PDF) of X, where P(X = Xi) means the probability that the discrete rv X takes the value of x_i.

EXAMPLE 15.6

In a throw of two dice, the random variable X. the sum of the numbers shown on two dice, can take one of the 11 values shown. The PDF of this variable can be shown to be as follows (see also Figure A.I):

X = 2 3 4 5 6 7 8 9
10 11 12

f(x) = (1/36) (2/36) (3/36) (4/36) (5/36) (6/36) (5/36) (4/36) (3/36) (2/36) (1/36)

These probabilities can be easily verified. In all there are 36 possible outcomes, of which one is favorable to number 2. two are favorable to number 3 (since the sum 3 can occur either as 1 on the first die and 2 on the second die or 2 on the first die and 1 on the second die), and so on.

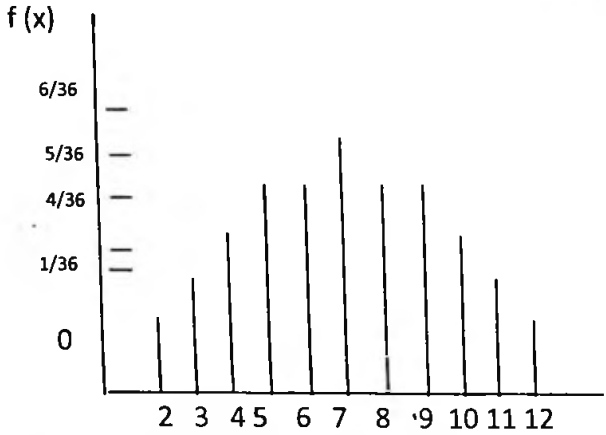


Figure 15.5: Density function of the discrete random variable of Example 15.6.

Probability Density Function of a Continuous Random Variable

Let X be a continuous rv. Then f(x) is said to be the PDF of X if the following conditions are satisfied:

$f(x) \geq 0$

$\int_{-\infty}^{\infty} f(x)dx = 1$

$\int_a^b f(x)dx = P(a \leq x \leq b)$

Where f(x) dx is known as the probability element (the probability associated with a small interval of a continuous variable) and where P(a ≤x≤ b) means the probability that X lies in the interval a to b. Geometrically, we have Figure A.2.

For a continuous rv, in contrast with a discrete rv, the probability that X takes a specific value is zero[3]; probability for such a variable is measurable only over a given range or interval, such as (a, b) shown in Figure A.2.

EXAMPLE 15.7

Consider the following density function:

$f(x) = \frac{1}{9}x^2 \qquad 0 \leq x \leq 3$

It can be readily verified that f(x) ≥ 0 for all x in the range 0 to 3 and that $\int_0^3 1/9x^2\,dx = 1$.

(Note: The integral is $(\frac{1}{9}x^3|^3_0) = 1$.) If we want to evaluate the above PDF between, say, 0 and 1, we obtain $\int_0^1 1/9x^2dx = (\frac{1}{9}x^3|^1_0)$; that is, the probability that x lies between 0 and 1 is 1/27.

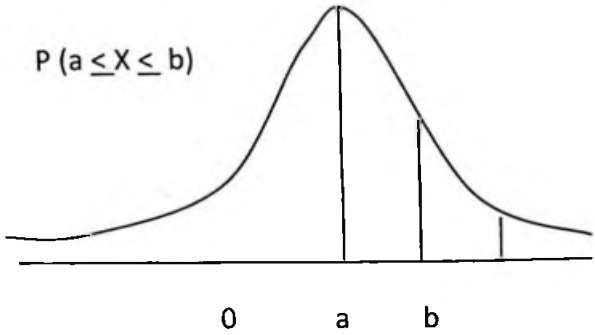


FIGURE 15.6 Density function of a continuous random variable.

Joint Probability Density Functions

Discrete Joint PDF. Let X and Y be two discrete random variables. Then the function

$f(x,y) = P(X = x \text{ and } Y = y)$

$= 0$ when $X \neq x$ and $Y \neq y$

is known as the discrete joint probability density function and gives the (joint) probability that X takes the value of x and Y takes the value of y.

EXAMPLE 15.8

The following table gives the joint PDF of the discrete variables X and Y.

	X			
	-2	0	2	3
3	0.27	0.08	0.16	0
Y				
6	0	0.04	0.10	0.35

This table tells us that the probability that X takes the value of - 2 while Y simultaneously takes the value of 3 is 0.27 and that the probability that X takes the value of 3 while Y takes the value of 6 is 0.35, and so on.

Marginal Probability Density Function

In relation to f(x, y), f(x) and f(y) are called individual, or marginal, probability density functions. These marginal PDFs are derived as follows:

$f(x) = \sum_{y} f(x, y)$ marginal PDF of X

$f(y) = \sum_{x} f(x, y)$ marginal PDF of Y

Where, for example, Σ_y means the sum over all values of Y and Σ_x means the sum over all values of X.

EXAMPLE 15.9

Consider the data given in Example 4. The marginal PDF of X is obtained as follows:

$$f(x = -2) = \sum_{y} f(x, y) = 0.27 + 0 = 0.27$$

$$f(x = 0) = \sum_{y} (x, y) = 0.08 + 0.04 = 0.12$$

$$f(x = 2) = \sum_{y} f(x, y) = 0.16+010 = 0.26$$

$$f(x = 3) = \sum_{y} f(x. y) = 0 + 0.35 = 0.35$$

Likewise, the marginal PDF of Y is obtained as

$$f(y = 3) = \sum_{x} f(x, y) = 0.27 + 0.08 + 0.16 + 0 = 0.51$$

$$f(y = 6) = \sum_{x} f(x, y) = 0 + 0.04 + 0.10 + 0.35 = 0.49$$

As this example shows, to obtain the marginal PDF of X we add the column numbers, and to obtain the marginal PDF of Y we add the row numbers. 654Notice that Σx f(x) over all values of X is 1, as is Σy f (y) over all values of Y (why?).

Conditional PDF. In regression analysis we are often interested in studying the behavior of one variable conditional upon the values

of another variable(s). This can be done by considering the conditional PDF. The function

$f(x \mid y) = P(X = x \mid Y = y)$

is known as the **conditional PDF** of X; it gives the probability that X takes on the value of x given that Y has assumed the value y. Similarly,

f(ylx)= P(Y=y/X=x) which gives the conditional PDF of Y.

The conditional PDFs may be obtained as follows:

$$f(x \mid y) = \frac{f(x,y)}{f(y)} \quad \text{conditional PDF of X}$$

$$f(y \mid x) = \frac{f(x,y)}{f(x)} \quad \text{conditional PDF of Y}$$

As the preceding expressions show, the conditional PDF of one variable can be expressed as the ratio of the joint PDF to the marginal PDF of another (conditioning) variable.

EXAMPLE 15.10

Continuing with Examples 4 and 5, let us compute the following conditional probabilities:

$$f(X=-2|Y=3) = \frac{f(X=-2,Y=3)}{f(Y=3)} = 0.27/0.51=053$$

Notice that the unconditional probability f(X = -2) is 0.27, but it Y has assumed the value of 3, the probability that X takes the value of -2 is 0.53.

$$f(X=2|Y=6) = \frac{f(X=2, Y=6)}{f(Y=6)} = 0.10/0.49=0.20$$

Again note that the unconditional probability that X takes the value of 2 is 0.26, which is different from 0.20, which is its value it Y assumes the value of 6.

Statistical Independence

Two random variables X and Yare statistically independent if and only if

$$f(x,y) = f(x)f(y)$$

that is, if the joint PDF can be expressed as the product of the marginal

PDFs.

EXAMPLE 11

A bag contains three balls numbered 1, 2, and 3. Two balls are drawn at random, with replacement, from the bag (i.e., the first ball drawn is replaced before the second is drawn). Let X denote the number of the first ball drawn and Y the number of the second ball drawn, The following table gives the joint PDF of X and Y.

	X		
	1	2	3
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
3	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Now $f(X = 1, Y = 1) = \frac{1}{9}$, $f(X = 1) = \frac{1}{3}$ (obtained by summing the first column), and $f(Y = 1) = \frac{1}{3}$ (obtained by summing the first row). Since $f(X, Y) = f(X)\, f(Y)$ in this example we can say that the two variables are statistically independent. It can be easily checked that for any other combination of X and Y values given in the above table the joint PDF factors into individual PDFs.

It can be shown that the X and Y variables given in Example 4 are not statistically independent since the product of the two marginal PDFs is not equal to the joint PDF. (Note:

$F(X, Y) = f(X)f(Y)$ must be true for all combinations of X and Y if the two variables are to be statistically independent.)

Continuous Joint PDF. The PDF f(x, y) of two continuous variables X and Y is such that

$f(x,y) \geq 0$

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dx\,dy = 1$

$\int_{c}^{d} \int_{a}^{b} f(x,y)dx\,dy = P(a \leq x \leq b,\ c < y \leq d)$

EXAMPLE 15.12

Consider the following PDF.

$F(x, y) = 2 - x - y \qquad 0 \leq x \leq 1;\ 0 \leq y \leq 1$

It is obvious that $f(x, y) \geq = 0$. Moreover

$\int_{0}^{1} \int_{0}^{1} (2-x-y)dxdy=1$

The marginal PDF of X and Y can be obtained as

$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$ marginal PDF of X

$f(y)) = \int_{-\infty}^{\infty} f(x, y)\,dx$ marginal PDF of Y

EXAMPLE 15.13

The two marginal PDFs of the joint PDF given in Example B are as follows:

$f(x) = \int_{0}^{2} f f(x, y)\,dy = \int_{0}^{2} (2 - x - y)dy$

$(2y-xy-y^2/2)|^1_0 = - x\ 0 \leq x \leq 1$

$f(y) = \int$ $(2 - x - y)dx$

$(2x - xy - x^2/2)|^1_0 =$ $- y \; 0 \leq y \leq 1$

To see if the two variables of Example B are statistically independent, we need to find out if $f(x, y) = f(x)\, f(y)$. Since $(2 - x - y) \neq (\frac{3}{2} - x)\,(\frac{1}{2} - y)$, we can say that the two variables are not statistically independent.

$$\int_0^{\frac{1}{2}} \left[\int_0^{\frac{1}{2}} (2 - x - y)dx\right]dy =$$

$$\int_0^{\frac{1}{2}} \left|(2x - \frac{x^2}{2} - xy?)\right|dy$$

$$= \int_0^{\frac{1}{2}} \left(\frac{3}{2} - y\right)dy$$

$$= \int_0^{\frac{1}{2}} \left(\frac{3}{2}y - y2/2\right) = 1$$

Note: The expression $(\frac{3}{2}y - y^2/2)$ means the expression in the parentheses is to be evaluated at the upper limit value of 1 and the lower limit value of 0: the latter value is subtracted from the former to obtain the value of the integral. Thus, in the preceding example the limits are $(\frac{1}{2} - \frac{1}{2})$ at y = 1 and 0 at y = 1 giving the value of the integral as 1.

Characteristics of Probability Distributions

A probability distribution can often be summarized in terms of a few of its characteristics, known as the moments of the distribution. Two of the most widely used moments are the mean, or expected value, and the variance.

Expected Value

The expected value of a discrete rv X, denoted by E(X), is defined as follows:

$$E(X) = \sum_x x f(x)$$

where Σ_x means the sum over all values of X and where f(x) is the (discrete) PDF of X.

EXAMPLE 15.14

Consider the probability distribution of the sum of two numbers in a throw of two dice given in Example 2. (See Figure A.1.) Multiplying the various X values given there by their probabilities and summing over all the observations, we obtain:

$$E(X)= 2(1/36)+3(2/26)+ 4(3/36)+12\ (1/36) =7$$

which is the average value of the sum of numbers observed in a throw of two dice.

EXAMPLE 15.15

Estimate E(X) and E(Y) for the data given in Example 4. We have seen that

x	-2	0	2	3
f(x)	0.27	0.12	0.26	0.35

Therefore,

$E(X) = \Sigma_x xf(x)$

$= (-2)(0.27) + (0)(0\ 12) + (2)(0.26) + (3)(0.35) = 1.03$

Similarly,

y	3	6
f(y)	0.51	0.49

$E(Y) = \Sigma_y yf(y)$

$= (3)(0.51) + (6)(0.49)$

$= 4.47$

The expected value of a continuous rv is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx$$

The only difference between this case and the expected value of a discrete rv is that we replace the summation symbol by the integral symbol.

EXAMPLE 15.16

Let us find out the expected value of the continuous PDF given in Example 3

$$E(X) = \int_{0}^{3} x(\frac{x^2}{9})dx$$

$$= \frac{1}{9}(\frac{x^4}{4})$$

$$= \frac{9}{4}$$

Properties of Expected Values

1. The expected value of a constant is the constant itself. Thus, if b is a constant, E(b) = b.

2. If a and b are constants,

$$E(aX + b) = aE(X) + b$$

This can be generalized If X_1, X_2, ..., X_N are N random variables and a_1, a_2,.....a_N and b are constants, then

$$E(a_1X_1 + a_2X_2 + ... + a_NX_N + b) = a_1E(X_1) + a_2E(X_2) + .,. + a_NE(X_N) + b$$

3. If X and Yare independent random variables, then

$$E(XY) = E(X)E(Y)$$

That is, the expectation of the product XY is the product of the (individual) expectations of X and Y.

4. If X is a random variable with PDF f(x) and if g(X) is any function of X, then

E[g(X)] = $\sum_{x} g(X)\, f(x)$ if X is discrete

$\int_{-\infty}^{\infty} g(X)\, f(x)\, dx$ if X is continuous

Thus, if $g(X) = X^2$,

$E(X^2) = \sum_{x} x2\, f(X)$ if x is discrete

$= \int_{-\infty}^{\infty} x2x^2\, f(X)\, dx$ if X is continuous

EXAMPLE 15. 17

Consider the following PDF:

x	-2	1	2
f(x)	5/8	1/8	2/8

Then

E(X) = -2(5/8) + 1(1/8) +2(2/8)

- 5/8

and

$E(x^2) = 4(5/8) + 1(1/8) +4(2/8)$

= 29/8

Variance

Let X be a random variable and let $E(X) = \mu$. The distribution, or spread, of the X values around the expected value can be measured by the variance, which is defined as

$$var(X) = \sigma_x^2 = E(X - \mu)^2$$

The positive square root of σ_x^2 , σ_x is defined as the **standard deviation** of X The variance or standard deviation gives an indication of how closely or widely the individual X values are spread around their mean value.

The variance defined previously is computed as follows:

$var(X) = \Sigma (X - \mu)^2 f(x)$ if X is a discrete rv

$= \int (X - \mu)^2 f(x)dx$ if X is a continuous rv

For computational convenience, the variance formula given above can also be expressed as

$$var(X) = \sigma^2_x = E(X - \mu)^2$$

$$= E(X^2) - \mu^2$$

$$= E(X^2) - [E(X)]^2$$

Applying this formula, it can be seen that the variance of the random variable given in Example 13 is 29/8 –

$(- 5/8)^2 = 207/64 = 3.23$

EXAMPLE 15. 18

Let us find the variance of the random variable given in Example 3.

Var (X) = $E(X^2) - [E(X)]^2$

Now $E(X^2) = \int_0^3 x^2 (\frac{x}{9})dx$

$= \int_0^3 x^4/9 dx$

$= \frac{1}{9}[\frac{x^5}{5}]_0^3$

$= \frac{243}{45}$

$= \frac{27}{5}$

Since $E(X) = \frac{9}{4}$ (see Example 12), we finally have

$var(X) = \frac{243}{45} - (\frac{9}{4})^2$

$= \frac{243}{720} = 0.34$

Properties of Variance

1. $E(X - \mu)^2 = E(X^2) - \mu^2$, as noted before.
2. The variance of a constant is zero.
3. If a and b are constants, then

$$\text{Var}(aX + b) = a^2 \text{ var}(X)$$

4. If X and Yare independent random variables, then

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y)$$

This can be generalized to more than two independent variables.

5. If X and Yare independent rv's and a and b are constants, then

$$\text{var}(aX + bY) = a\ 2 \text{ var}(X) + b2 \text{ var}(Y)$$

Covariance

Let X and Y be two rv's with means μ_x and μ_y respectively. Then the covariance between the two variables is defined as

$$\text{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - \mu_x\mu_y$$

It can be readily seen that the variance of a variable is the covariance of that variable with itself.

The covariance is computed as follows:

$$\text{cov}(X, Y) = \Sigma_y \Sigma_x (X - u_x)(Y - \mu_y) f(x,y)$$

$$= \Sigma_y \Sigma_x XYf(x, y) - \mu_x\mu_y$$

if X and Yare discrete random variables, and

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - \mu_x)(Y - {}_{\mu}y)\, f(x,y)dxdy$$

$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XYf(x, y)dxdy - \mu_x \mu_y$

if X and Y are continuous random variables.

Properties of Covariance

1. If X and Yare independent, their covariance is zero for

$cov(X, Y) = E(XY) - \mu_x\mu_y$

$= \mu_x\mu_y - \mu_x\mu_y$ since $E(XY) = E(X)E(Y) = \mu_x\mu_y$ when X and Yare

$= 0$ independent

$cov (a + bX, c + dY) = bd\ cov (X, Y)$

where a, b, c, and d are constants.

EXAMPLE 15. 19

Let us find out the covariance between discrete random variables X and Y whose joint PDF is as shown in Example 4. From Example 11 we already know that $\mu_x = E(X) = 1.03$ and $\mu_y = E(Y) = 4.47$.

$E(XY) = \sum_{x} \sum_{y} XY\ f(x, y)$

$= (-2)(3)(0.27) + (0)(3)(0.08) + (2)(3)(0.16) + (3)(3)(0)$

$+ (-2)(6)(0) + (0)(6)(0.04) + (2)(6)(0\ 10) + (3)(6)(035)$

$= 6.84$

Therefore,

$$\text{cov}(X, Y) = E(XY) - \mu_x\mu_y$$

$$= 6.84 - (103)(4.47)$$

$$= 2.24$$

Correlation Coefficient

The (population) correlation coefficient ρ(rho) is defined as

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\text{cov}(X,Y)}{\sigma_x\sigma_y}$$

Thus defined p is a measure of linear association between two variables and lies between -1 and + I, -1 indicating perfect negative association and + 1 indicating perfect positive association.

From the preceding formula, it can be seen that

$$\text{cov}(X, Y) = \rho\sigma_x\sigma_y$$

EXAMPLE 15.20

Estimate the coefficient of correlation for the data of Example 4.

From the PDFs given in Example 11 it can be easily shown that σ_x = 2.05 and σ_y = 1.50. We have already shown that cov(X, Y) = 2.24. Therefore, applying the preceding formula we estimate ρ as 2.24/(2.05)(1.50) = 0.73.

Variances of Correlated Variables. Let X and Y be two rv's. Then

var (X + Y) = var (X) + var (Y) + 2 cov (X,Y)

= var (X) + var (Y) + 2 $\rho\sigma_X\sigma_Y$

var (X – Y) = var (X) + var(Y) - 2 cov (X,Y)

= var (X) + var (Y) - 2 $\rho\sigma_X\sigma_Y$

If however, X and Y are independent, cov (X,Y) is zero, in which case the var (X + Y) and var (X - Y) are both equal to var (X) + var (Y) , as noted previously.

The preceding results can be generalized as follows. Let $\sum_{i=1}^{n} Xi = X_1 + X_2 + \ldots + Xn$, then the variance of the linear combination Σ Xi is

$$\text{var}\left(\sum_{i=1}^{n} Xi\right) = \sum_{i=1}^{n} var\, Xi + 2\sum \sum cov\,(Xi, Xj)$$

$$\sum_{i=1}^{n} var\, Xi + 2\sum \sum \rho ij\sigma i\sigma j$$

where ρ_{ij} is the correlation coefficient between X_i and X_j and where σ_i and σ_j are the standard deviations of X_i and X_j

Thus,

var ($X_1 + X_2 + X_3$) = var X_1 + var X_2 + var X_3 + 2cov(X_1, X_2)

+ 2 cov (X_1, X_3) + 2 cov (X_2, X_3)

= var X_1 + var X_2 + var X_3 + $2\rho_{12}\sigma_1\sigma_2$

+$2\rho_{13}\sigma_1\sigma_3$ + $2\rho_{23}\sigma_2\sigma_3$

where a_1, a_2, and a_3 are, respectively, the standard deviations of X_1, X_2, and X_3 and where ρ_{12} is the correlation coefficient

between X_1 and X_2, ρ_{13} that between X_1 and X_3, and ρ_{23} that between X_2 and X_3.

Conditional Expectation and Conditional Variance

Let f(x, y) be the joint PDF of random variables X and Y. The conditional expectation of X, given Y = y, is defined as

$E(X/Y=y) = \sum_x x f(x/Y=y)$ if X is discrete

$= \int_{-\infty}^{\infty} x f(x/Y=y)dx$ if X is continuous

where E(X/ Y = y) means the conditional expectation of X given Y = y and where f(x/ Y = y) is the conditional PDF of X. The conditional expectation of Y, E(Y/ X = x), is defined similarly.

Conditional Expectation. Note that E(X/Y) is a random variable because it is a function of the conditioning variable Y. However, E(X /Y = y), where y is a specific value of Y, is a constant.

Conditional Variance. The conditional variance of X given Y = y is defined as

$\text{var}(X/Y = y) = E\{[X - E(X/Y = y)]^2 / Y = y\}$

$= \sum_x [X - E(X|Y = y)]^2 f(x|Y = y)$ if X is discrete

$= \int_{-\infty}^{\infty} [X - E(X|Y = y)]^2 f(x/Y=y)dx$ if X is continuous

EXAMPLE 15. 21

Compute E(YIX = 2) and var (YIX = 2) for the data given in Example 4.

$E(Y| X= 2) = \sum_{Y} Y \cdot f(Y= y|X=2)$

$= 3f(Y = 3/ X = 2) + 6f(Y = 6/ X = 2)$

$= 3(0.16/026) + 6(0.10/0.26)$

$=4.15$

Note: $f(Y= 3/X = 2) = f(Y = 3, X = 2)/f(X = 2) = 0.16/0.26$, and $f(Y= 61 X = 2) =f(Y= 6, X = 2)/ f(X= 2) = 0.10/0.26$. so

$$\text{var} (Y | X = 2) = \sum_{Y} [Y - E (YIX = 2)]^2 \, f(Y | X = 2)$$

$= (3 - 415)^2 (0.16/0.26) + (6 - 415)^2(0.10/0.26)$

$=2.13$

Properties of Conditional Expectation and Conditional Variance

1. If $f(X)$ is a function of X, then $E(f(X) | X) = f(X)$, that is, the function of X behaves as a constant in computation of its expectation conditional on X. Thus, $[E(X^3 | X)] = E(X^3)$; this is because, if X is known, X^3 is also known.

2. If f(X) and g(X) are functions of X, then

$$E[f(X)Y + g(X) | X] = f(X)E(Y | X) + g(X)$$

For example, $E[XY + cX^2 | X] = XE(Y | X) + cX^2$, where c is a constant.

3. If X and Yare independent, E(Y I X) = E(Y). That is, if X and Yare independent random variables, then the conditional expectation of y, given X is the same as the unconditional expectation of Y.

4. **The law of iterated expectations.** It is interesting to note the following relation between the unconditional expectation of a random variable Y, E(Y), and its conditional expectation based on another random variable X, E(Y I X):

$$E(Y) = Ex[E(Y \mid X)]$$

This is known as the law of iterated expectations, which in the present context states that the marginal, or unconditional, expectation of Y is equal to the expectation of its conditional expectation, the symbol Ex denoting that the expectation is taken over the values of X. Put simply, this law states that if we first obtain E(Y I X) as a function of X and take its expected value over the distribution of X values, you wind up with E(Y), the unconditional expectation of Y. The reader can verify this using the data given in Example 4.

5. If X and Yare independent, then var(Y / X) = var(Y).

6. var (Y) = E[var(Y I X») + var[E(Y I X)]; that is, the (unconditional) variance of Y is equal to expectation of the conditional variance of Y plus the variance of the conditional expectation of Y.

Higher Moments of Probability Distributions

Although mean, variance, and covariance are the most frequently used summary measures of univariate and multivariate PDFs, we occasionally need to consider higher moments of the PDFs, such as the third and the fourth moments. The third and fourth moments of a univariate PDF (x) around its mean value (μ) are defined as

Third moment: $E(X - \mu.)^3$

Fourth moment: $E(X - \mu)^4$

In general, the rth moment about the mean is defined as

rth moment: $E(X - \mu)$

The third and fourth moments of a distribution are often used in studying the "shape" of a probability distribution, in particular, its skewness, S (i.e., lack of symmetry) and kurtosis, K (i.e., tallness or flatness), as shown in Figure A.3.

One measure of skewness is defined as

$$s = \frac{E(X - \mu)^3}{\sigma^3}$$

$$= \frac{\text{third moment about the mean}}{\text{Cube of the standard deviation}}$$

A commonly used measure of kurtosis is given by

$$K = \frac{E(X - \mu)^4}{[E(X - \mu)^2]^2}$$

$$= \frac{\text{fourth moment about mean}}{\text{Square of the second moment}}$$

PDFs with values of K less than 3 are called platykurtic (fat or short-tailed), and those with values greater than 3 are ca1led leptokurtic (slim or longtailed). See Figure A.3. A PDF with a kurtosis value of 3 is known as mesokurtic, of which the normal distribution is the prime example. (See the discussion of the normal distribution in Section A.6.)

We will show shortly how the measures of skewness and kurtosis can be combined to determine whether a random variable follows a normal distribution. Recall that our hypothesis-testing procedure, as in the t and

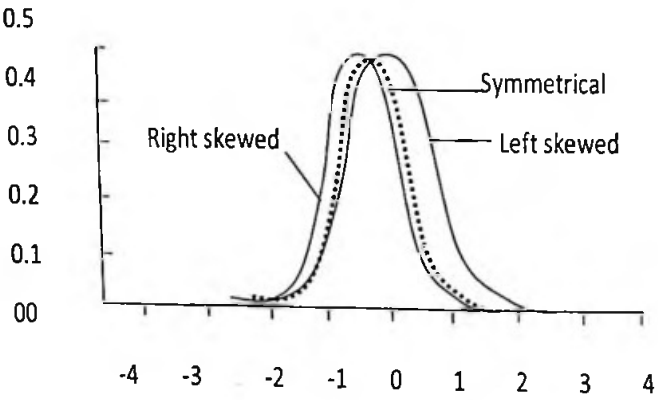
F tests, is based on the assumption (at least in small or finite samples) that the underlying distribution of the variable (or sample statistic) is normal. It

is therefore very important to find out in concrete applications whether this

assumption is fulfilled.

(a)

(a)



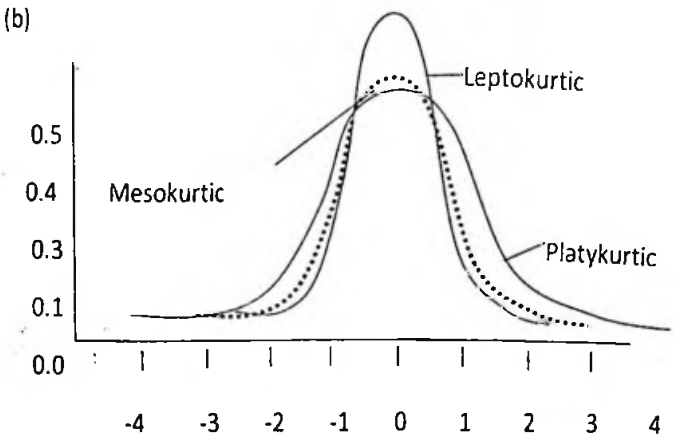


FIGURE 15

(a) Skewness; (b) kurtosis.

Some Important Theoretical Probability Distributions

In the text extensive use is made of the following probability distributions.

Normal Distribution

The best known of all the theoretical probability distributions is the normal distribution, whose bell-shaped picture is familiar to anyone with a modicum of statistical knowledge.

A (continuous) random variable X is said to be normally distributed if its

PDF has the following form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\quad\right) \qquad -\infty < X < \infty$$

where μ and σ^2, known as the parameters of the distribution, are, respectively, the mean and the variance of the distribution. The properties of this distribution are as follows:

1. It is symmetrical around its mean value.

2. Approximately 68 percent of the area under the normal curve lies between the values of $\mu \pm \sigma$, about 95 percent of the area lies between $\mu \pm 2\sigma$, and about 99.7 percent of the area lies between $\mu \pm 3\sigma$, as shown in Figure A.4.

3. The normal distribution depends on the two parameters μ and σ^2, so once these are specified, one can find the probability that X will lie within a certain interval by using the PDF of the normal distribution. But this task can be lightened considerably by referring to Table D.1 of Appendix D. To use this table, we convert the given normally distributed variable X with mean μ and σ^2 into a standardized normal variable Z by the following transformation:

$$Z = \frac{X - \mu}{\sigma}$$

An important property of any standardized variable is that its mean value is zero and its variance is unity. Thus Z has zero mean and unit variance. Substituting z into the normal PDF given previously, we obtain

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} Z^2\right) \quad f(Z) = ½ \exp(-1/2Z^2)$$

X and Y is such that X and Y is such that

which is the PDF of the standardized normal variable. The probabilities given in Appendix D, Table D.I, are based on this standardized normal variable.

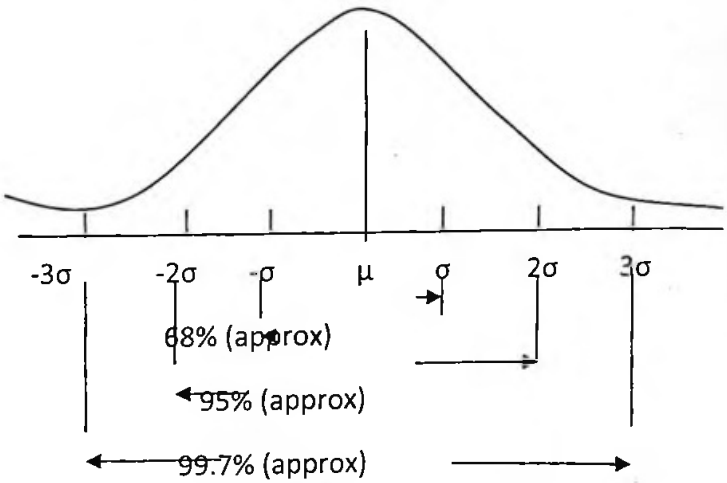
By convention, we denote a normally distributed variable as

$$X \sim N(\mu, \sigma^2)$$

where ~ means "distributed as," N stands for the normal distribution, and the quantities in the parentheses are the two parameters of the normal distribution, namely, the mean and the variance. Following this convention,

$X \sim N(0, 1)$

means X is a normally distributed variable with zero mean and unit variance. In other words, it is a standardized normal variable Z.



EXAMPLE 15.22

Assume that $X \sim N(8, 4)$. What is the probability that X will assume a value between $X_1 = 4$ and $X_2 = 12$? To compute the required probability, we compute the Z values as

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{4 - 8}{2} = -2$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12 - 8}{2} = +2$$

Now from Table D.1 we observe that $Pr(0 \leq Z \leq 2) = 0.4772$. Then, by symmetry, we have $Pr(-2 \leq Z \leq 0) = 0.4772$. Therefore, the required probability is $0.4772 + 0.4772 = 0.9544$.
(See Figure A.4.)

EXAMPLE 15.23

What is the probability that in the preceding example X exceeds 12?

The probability that X exceeds 12 is the same as that Z exceeds 2. From Table D.1 it is obvious that this probability is (0.5 - 0.4772) or 0.0228.

4. Let $X_1 \sim N(\mu_1, \sigma^2_1)$ and $X_2 \sim N(\mu_2, \sigma^2_2)$ and assume that they are independent. Now consider the linear combination

$$Y = aX_1 + bX_2$$

where a and b are constants. Then it can be shown that

$$Y \sim N[(a\mu_1 + b\mu_2), (a^2\sigma^2_1 + b^2a^2_2)]$$

This result, which states that a linear combination of normally distributed variables is itself normally distributed, can be easily

generalized to a linear combination of more than two normally distributed variables.

5. Central limit theorem. Let $X_1, X_2, \ldots, X_n$ denote n independent random variables, all of which have the same PDF with mean = μ and variance = σ^2. Let $X = \sum Xi/n$ (i.e., the sample mean). Then as n increases indefinitely, (i.e., n $\rightarrow\infty$).

$X \sim N$

$n \rightarrow\infty$)

That is, X approaches the normal distribution with mean μ and variance σ_2/n. Notice that this result holds true regardless of the form of the PDF As a result, it follows that

$$z = \frac{X - \mu}{\sigma / \sqrt{n}} = \frac{\sqrt{n}(X - \mu)}{\sigma} \sim N(0, 1)$$

That is, Z is a standardized normal variable.

6. The third and fourth moments of the normal distribution around the mean value are as follows:

Third moment: $E(X - \mu)^3 = 0$

Fourth moment: $E(X - \mu)_4 = 3\sigma^4$

Note: All odd-powered moments about the mean value of a normally distributed variable are zero.

7. As a result, and following the measures of skewness and kurtosis discussed earlier, for a normal PDF skewness = 0 and kurtosis = 3; that is, a normal distribution is symmetric and

mesokurtic. Therefore, a simple test of normality is to find out whether the computed values of skewness and kurtosis depart from the norms of 0 and 3. This is in fact the logic underlying the Jarque-Bera (JB) test of normality discussed in the text:

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

where S stands for skewness and K for kurtosis. Under the null hypothesis of normality, JB is distributed as a chi-square statistic with 2 df.

8. The mean and the variance of a normally distributed random variable are independent in that one is not a function of the other.

The X^2 (Chi-Square) Distribution

Let Z_1, Z_2, ...Z_k be independent standardized normal variables (i.e., normal variables with zero mean and unit variance).

$$Z = \sum_{i=1}^{k} Z_i^2$$

Then the quantity is said to possess the x^2 distribution with k degrees of freedom (df), where the term df means the number of independent quantities in the previous sum. A chi-square-distributed variable is denoted by xl, where the subscript k indicates the df. Geometrically, the chi-square distribution appears in Figure A.5.

Properties of the x^2 distribution are as follows:

1. As Figure A.5 shows, the X^2 distribution is a skewed distribution, the degree of the skewness depending on the df. For

comparatively few df, the distribution is highly skewed to the right; but as the number of df increases, the distribution becomes increasingly symmetrical. As a matter of fact, for df in excess of 100, the variable

$\sqrt{x^2} - (2k - 1)$

can be treated as a standardized normal variable, where k is the df.

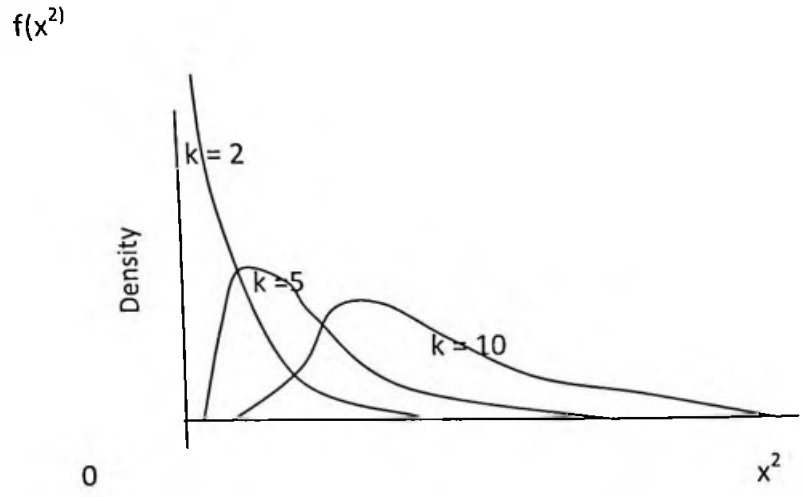
2. The mean of the chi-square distribution is k, and its variance is 2k, where k is the df.

3. If Z_1 and Z_2 are two independent chi-square variables with k_1 and k_2 df, then the sum 21 + 22 is also a chi-square variable with df = $k_1 + k_2$.

EXAMPLE 15.24

What is the probability of obtaining a x^2 value of 40 or greater, given the df of 20?

As Table D.4 shows, the probability of obtaining a x^2 value of 39.9968 or greater (20 df) is 0.005. Therefore, the probability of obtaining a X^2 value of 40 or greater is less than 0.005, a rather small probability.



Density function of the x^2 variable

Student's t Distribution

If Z, is a standardized normal variable [that is, $Z_1 \sim N(0, 1)$] and another variable Z_2 follows the chi-square distribution with k df and is distributed independently of Z_1, then the variable defined as

$$t = \frac{z_1}{\sqrt{z_2 / k}}$$

$$= \frac{z_1 \sqrt{k}}{\sqrt{z_2}}$$

follows Student's t distribution with k df. A t-distributed variable is often designated as t_k, where the subscript k denotes the df. Geometrically, the t distribution is shown in Figure A.6.

Properties of the Student's t distribution are as follows:

1. As Figure A.6 show, the **t distribution, like the normal distribution,** is symmetrical, but it is flatter than the normal distribution. But as the df increase, the t distribution approximates the normal distribution.

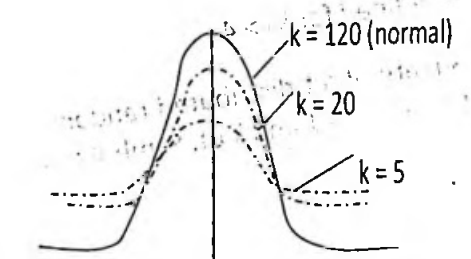
2. The mean of the t distribution is zero, and its variance is k/(k - 2).

3. The t distribution is tabulated in Table D.2.

EXAMPLE 15.25

Given df = 13, what is the probability of obtaining a t value (a) of about 3 or greater, (b) of about -3 or smaller, and (c) of /t/ of about 3 or greater, where /t/ means the absolute value (i.e., disregarding the sign) of t?

From Table D.2, the answers are (a) about 0.005, (b) about 0,005 because of the symmetry of the t distribution, and (c) about 0.01 = 2(0,005),



Student's t distribution for selected degrees of freedom,

The F Distribution

If Z_1 and Z_2 are independently distributed chi-square variables with k_1 and k_2 df, respectively, the variable

$$F = \frac{\quad}{\quad}_1$$

follows (Fisher's) F distribution with k_1 and k_2 df. An F-distributed variable is denoted by $F_{k1,k2}$, where the subscripts indicate the df associated with the two Z variables, k_1 being called the numerator df and k_2 the denominator df

Geometrically, the F distribution is shown in Figure A.7.

The F distribution has the following properties:

1. Like the chi-square distribution, the F distribution is skewed to the right. But it can be shown that as k_1 and k_2 become large, the F distribution approaches the normal distribution.

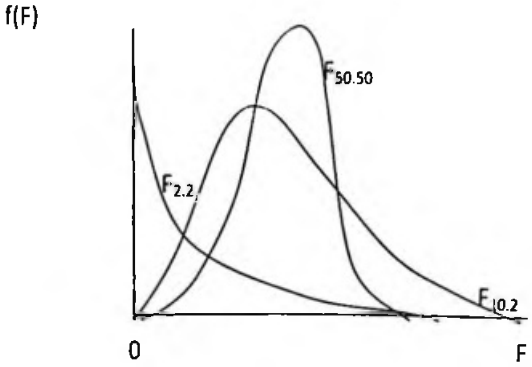
2. The mean value of an F-distributed variable is $k_2/(k_2 - 2)$, which is defined for $k_2 > 2$, and its variance is

$$\frac{2k^2_2(k_1 + k_2 - 2)}{k_1 (k_2 - 2)^2(k_2 - 4)}$$

which is defined for $k_2 > 4$.

3. The square of a t-distributed random variable with k df has an F distribution with 1 and k df. Symbolically,

$$t^2_k = F_{1,k}$$



F distribution for various degrees of freedom,

EXAMPLE 15.26

Given $k_1 = 10$ and $k_2 = 8$, what is the probability of obtaining an F value (a) of 3.4 or greater and (b) of 5.8 or greater?

As Table D.3 shows, these probabilities are (a) approximately 0.05 and (b) approximately 0.01.

4. If the denominator df, k_2, is fairly large, the following relationship holds between the F and the chi-square distributions:

$$K_1 F \sim x^2_{k1}$$

That is, for fairly large denominator df, the numerator df times the F value is approximately the same as a chi-square value with numerator df.

EXAMPLE15. 27

Let $k_1 = 20$ and $k_2 = 120$. The 5 percent critical F value for these df is 1.48. Therefore, k_1 F = (20)(1.48) = 29.6. From the chi-square

distribution for 20 df, the 5 percent critical chi-square value is about 31.41.

In passing, note that since for large df the t, chi-square, and F distributions approach the normal distribution, these three distributions are known as the distributions related to the normal distribution.

The Bernoulli Binomial Distribution

A random variable X is said to follow a distribution named after Bernoulli (a Swiss mathematician) if its probability density (or mass) function (PDF) is:

$$P(X = 0) = 1 - p$$

$$P(X = 1) = p$$

Where p, $0 \leq p \leq 1$, is the probability that some event is a "success," such as the probability of obtaining a head in a toss of a coin. For such a variable,

$$E(X) = [1 \times p(X = 1) + 0 \times p(X = 0)] = p$$

$$\text{var}(X) = pq$$

where $q = (1 - p)$, that is, the probability of a "failure."

Binomial Distribution

The binomial distribution is the generalization of the Bernoulli distribution. Let n denote the number of independent trials, each of which results in a "success" with probability p and a "failure" with a probability $q = (1 - p)$. If

X represents the number of successes in the n trials, then X is said to follow the binomial distribution whose PDF is:

$f(X) = \frac{n!}{x!(n-x)!} P^x(1-P)^{n-x}$

Where x represents the number of successes in n trials and where

where n!, read as n factorial, means n(n - 1)(n - 2)...1.

The binomial is a two-parameter distribution, n and p. For this distribution,

$E(X) = np$

$var(X) = np(1 - p) = npq$

For example. if you toss a coin 100 times and want to find out the probability of obtaining 60 heads, you put P = 0.5, n = 100 and x = 60 in the above formula. Computer routines exist to evaluate such probabilities.

You can see how the binomial distribution is a generalization of the Bernoulli distribution.

The Poisson distribution

A random X variable is said to have the Poisson distribution *if* its PDF is

$$f(X) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots, \lambda > 0$$

The Poisson distribution depends on a single parameter, λ. A distinguishing feature of the Poisson distribution is that its variance is equal to its expected value, which is λ. That is.

$E(X) = var(X) = \lambda$

The Poisson model, as we saw in the chapter on nonlinear regression models, is used to model rare or infrequent phenomena, such as the number of phone calls received in a span of say, 5 minutes or the number of speeding tickets received in a span of an hour, or the number of patents received by a firm say, in a year.

Statistical Inference: Estimation

In Section A.6 we considered several theoretical probability distributions.

Very often we know or are willing to assume that a random variable X follows a particular probability distribution but do not know the value(s) of the parameter(s) of the distribution. For example, if X follows the normal distribution, we may want to know the value of its two parameters, namely, the mean and the variance. To estimate the unknowns, the usual procedure is to assume that we have a random sample of size n from the known probability distribution and use the sample data to estimate the unknown parameters[5]. This is known as the problem of estimation. In this section, we take a closer look at this problem. The problem of estimation can be broken down into two categories: point estimation and interval estimation.

Point Estimation

To fix the ideas, let X be a random variable with PDF $f(x; \theta)$, where θ is the parameter of the distribution (for simplicity of discussion only, we are assuming that there is only one unknown parameter; our discussion can be readily generalized). Assume that we know the functional form - that is, we know the

theoretical PDF, such as the t distribution-but do not know the value of θ. Therefore, we draw a random sample of size n from this known PDF and then develop a function of the sample values such that

$$\hat{\theta} = f(x_1, x_2, \ldots, x_n)$$

provides us an estimate of the true θ . $\hat{\theta}$ is known as a **statistic**, or an **estimator**, and a particular numerical value taken by the estimator is known as an estimate. Note that e can be treated as a random variable because it is a function of the sample data. $\hat{\theta}$ provides us with a rule, or formula, that tells us how we may estimate the true θ. Thus, if we let

$$\hat{\theta} = \frac{1}{n}(x_1 + X_2 + \ldots + x_n) = X$$

where X is the sample mean, then X is an estimator of the true mean value, say, μ. If in a specific case X = 50, this provides an estimate of μ. The estimator $\hat{\theta}$ obtained previously is known as a point estimator because it provides only a single (point) estimate of .

Interval Estimation

Instead of obtaining only a single estimate of e, suppose we obtain two estimates of e by constructing two estimators $\hat{\theta}_1$ (x_1, x_2. . . . ,x_n) and $\hat{\theta}_2$ (x_1, x_2. . . . ,x_n), and say with some confidence (i.e., probability) that the interval between $\hat{\theta}_1$ and $\hat{\theta}_2$ includes the true θ. Thus, in interval estimation, in contrast with point estimation, we provide a range of possible values within which the true $\hat{\theta}$ may lie.

Let X_1 X_2.....X_n be n random variables with joint PDF f (X_1. X_2..... X_n). If we can write (X_1.X_2.. .,xn,) = f (X_1) f (X_2) f (x_n)

where f (x) is the common PDF of each X, then x_1. $x_{2,....}x_n$ are said to constitute a random sample of size n from a population with PDF f(x_n).

The key concept underlying interval estimation is the notion of the sampling, or probability distribution, of an estimator. For example, it can be shown that if a variable X is normally distributed, then the sample mean X is also normally distributed with mean = μ (the true mean) and variance = σ^2/n, where n is the sample size. In other words, the sampling, or probability, distribution of the estimator X is X ~ N(μ, σ^2/n). As a result, if we construct the interval

$$X \pm 2\frac{\sigma}{\sqrt{n}}$$

and say that the probability is approximately 0.95. or 95 percent, that intervals like it will include the true μ we are in fact constructing an interval estimator for μ. Note that the interval given previously is random since it is based on X, which will vary from sample to sample.

More generally. in interval estimation we construct two estimators θ_1 and θ_2, both functions of the sample X values, such that

$$\Pr(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha \qquad 0 < \alpha < 1$$

That is, we can state that the probability is 1 - α that the interval from θ_1, to

$\hat{\theta}_2$ contains the true θ. This interval is known as a confidence interval of size 1 - α for θ, 1 - α being known as the confidence coefficient. If a = 0.05, then 1 - α = 0.95, meaning that if we construct a confidence interval with a confidence coefficient of 0.95, then in repeated such constructions resulting from repeated sampling we shall be right in 95 out of 100 cases if we maintain that the interval contains the true θ. When the confidence coefficient is 0.95, we often say that we have a 95% confidence interval. In general, if the confidence coefficient is 1 - α, we say that we have a 100(1-α)% confidence interval. Note that α is known as the level of significance, or the probability of committing a Type 1 error. This topic is discussed in Section A.8.

EXAMPLE 15.28

Suppose that the distribution of height of men in a population is normally distributed with mean = μ inches and a = 2.5 inches. A sample of 100 men drawn randomly from this population had an average height of 67 inches.

Establish a 95% confidence interval tor the mean height (= μ) in the population as a whole.

As noted. $X \sim N(\mu, \sigma^2/n)$, which in this case becomes $X \sim N(\mu, 2.5^2/100)$. From Table D.1 one can see that

$$X - 1.96\left(\frac{\cdot}{\cdot}\right) \leq \mu \leq X + 196$$

covers 95% of the area under the normal curve. Therefore, the preceding interval provides a 95% confidence interval for μ.

Plugging in the given values of X, σ and n, we obtain the 95% confidence interval as

$$66.51 \leq \mu \leq 67.49$$

In repeated such measurements, Intervals thus established will include the true μ with 95 percent confidence. A technical point may be noted here. Although we can say that the probability that the random interval $[X \pm 1.96(\sigma/\sqrt{n})]$ includes μ is 95 percent, we cannot say that the probability is 95 percent that the particular interval (66.51, 67.49) includes μ. Once this interval is fixed, the probability that it will include μ is either 0 or 1. What we can say is that if we construct 100 such intervals, 95 out of the 100 intervals will include the true μ; we cannot guarantee that one particular interval will necessarily include μ.

Methods of Estimation

Broadly speaking, there are three methods of parameter estimation: (1) least squares (LS), (2) maximum likelihood, and (3) method of moments (MOM) and its extension, the generalized method of moments (GMM). We-have devoted considerable time to illustrate the LS method. In Chapter 4 we introduced the ML method in the regression context. But the method is of much broader application.

The key idea behind the ML is the likelihood function. To illustrate this, suppose the random variable X has PDF $f(X,\theta)$ which depends on a single parameter θ. We know the PDF (e.g., Bernoulli or binomial) but do not know the parameter value. Suppose we obtain a random sample of nX values. The joint PDF of these n values is:

$$g(X_1, X_2, \ldots X_n; \theta)$$

Because it is a random sample, we can write the preceding joint PDF as a product of the individual PDF as

$$g(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta)$$

The joint PDF has a dual interpretation. If θ is known, we interpret it as joint probability of observing the given sample values. On the other hand, we can treat it as a function of θ for a given values of $x_1, x_2, \ldots, x_n$. On the latter interpretation, we call the joint PDF the likelihood function (LF) and write it as

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta)$$

Observe the role reversal of θ in the joint probability density function and the likelihood function.

The ML estimator of θ is that value of θ that maximizes the (sample) likelihood function, L. For mathematical convenience, we often take the log of the likelihood, called the log-likelihood function (log L). Following the calculus rules of maximization, we differentiate the log-likelihood function with respect to the unknown and equate the resulting derivative to zero. The resulting value of the estimator is called the maximum likelihood estimator. One can apply the second-order condition of maximization to assure that the value we have obtained is in fact the maximum value.

In case there is more than one unknown parameter, we differentiate the

log-likelihood function with respect to each unknown, set the resulting expressions to zero, and solve them simultaneously to obtain the values of the unknown parameters. We have already

shown this for the multiple regression model (see the appendix to Chapter 4).

EXAMPLE 15.29

Assume that the random variable X follows the Poisson distribution with the mean value of λ. Suppose x_1. x_2, . .x_n are independent Poisson random variables each with mean λ: Suppose we want to find out the ML estimator of λ , The ·lihood function here is:

$$L(x_1, x_2,\ldots,x_n;\lambda)= \frac{\sigma^{\lambda}\lambda^{x1}}{x_1!}\ \frac{\sigma^{\lambda}\lambda^{x2}}{x_2!}\ \ldots\ \frac{\sigma^{\lambda}\lambda^{xn}}{x_n!}$$

$$\frac{\sigma^{n\lambda}\lambda^{\Sigma x1}}{x_1!x_2!\ x_n!}$$

This is a rather unwieldy-expression, but if we take its log, it becomes

$$\log (x_1, x_2,\ldots,x_n; \lambda) -n\lambda + \Sigma x_i, \log\lambda -\log c$$

where $\log c = \Pi x!$. Differentiating the preceding expression with respect to λ, we obtain $(-n+ (\Sigma xi)/ \lambda)$. By setting this last expression to zero, we obtain $\lambda_{mi} = (\Sigma_{xi})/n= X$, which is the ML estimator of the unknown λ.

The Method of Moments. We have given a glimpse of MOM in exercise 3.4 in the so-called analogy principle in which the sample moments try to duplicate the properties of their population counterparts. The GMM, which is a generalization of

MOM, is now becoming more popular, but not at the introductory level. Hence we will not pursue it here.

The desirable statistical properties fall into two categories: small-sample, or finite-sample, properties and large-sample, or asymptotic, properties.

Underlying both these sets of properties is the notion that an estimator has a sampling, or probability, distribution.

Small-Sample Properties

Unbiasedness. An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if the expected value of $\hat{\theta}$ is equal to the true θ; that is,

$E(\hat{\theta}) = \theta$

or

$E(\hat{\theta}) - \theta = 0$

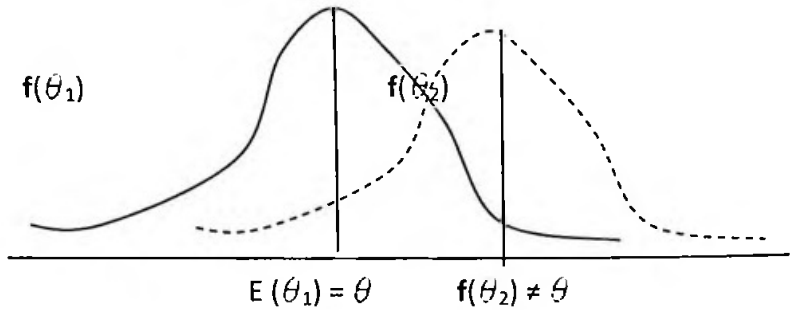
If this equality does not hold, then the estimator is said to be biased, and the bias is calculated as

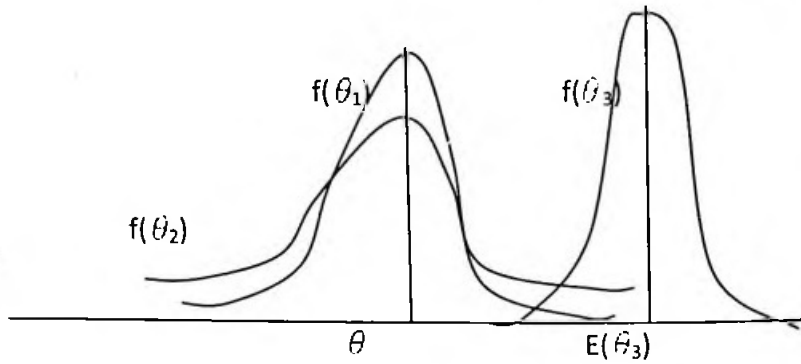
$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$

Of course, if $E(\hat{\theta}) = \theta$ that is, $\hat{\theta}$ is an unbiased estimator-the bias is zero.

Geometrically, the situation is as depicted in Figure A.8. In passing, note that unbiasedness is a property of repeated sampling, not of any given sample: keeping the sample size fixed, we draw several samples, each time obtaining an estimate of the unknown parameter. The average value of these estimates is

expected to be equal to the true value if the estimator is to be unbiased.





Minimum Variance. $\hat{\theta}_1$ is said to be a minimum-variance estimator of θ if the variance of $\hat{\theta}_1$ is smaller than or at most equal to the variance of $\hat{\theta}_2$ which is any other estimator of θ.

Geometrically, we have Figure A.9, which shows three estimators of θ, namely $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$, and their probability distributions. As shown, the variance of $\hat{\theta}_3$ is smaller than that of either $\hat{\theta}_1$ or $\hat{\theta}_2$. Hence, assuming only the three possible estimators, in this case $\hat{\theta}_3$ is a minimum-variance estimator. But note that $\hat{\theta}_3$ is a biased estimator (why?).

Best Unbiased, or Efficient, Estimator. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators of θ, and the variance of $\hat{\theta}_1$ is smaller than or at most equal to the variance of $\hat{\theta}_2$, then $\hat{\theta}_1$ is a minimum-variance unbiased, or best unbiased, or efficient, estimator. Thus, in Figure A.9, of the two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, $\hat{\theta}_1$ is best unbiased, or efficient.

Linearity. An estimator $\hat{\theta}$ is said to be a linear estimator of θ if it is a linear function of the sample observations. Thus, the sample mean defined as

$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{n}(x_i + x_2 + \ldots + x_n)$ is a linear estimator because it is a linear function of the X values.

Best Linear Unbiased Estimator (BLUE). If $\hat{\theta}$ is linear, is unbiased, and has minimum variance in the class of all linear unbiased estimators of θ, then it is called a best linear unbiased estimator, or BLUE for short.

Minimum Mean-Square-Error (MSE) Estimator. The MSE of an estimator $\hat{\theta}$ is defined as

$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$

This is in contrast with the variance of $\hat{\theta}$, which is defined as

$$\text{var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

The difference between the two is that var $(\hat{\theta})$ measures the dispersion of the distribution of $\hat{\theta}$ around its mean or expected value, whereas MSE$(\hat{\theta})$ measures dispersion around the true value of the parameter. The relationship between the two is as follows:

$$MSE(\theta) = E(\hat{\theta} - \theta)^2$$

$$= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2$$

$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta)$$

$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 \text{ since the last term is zero}^6$$

$$= \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2$$

= variance of $\hat{\theta}$ plus square bias

Of course, if the bias is zero, MSE$(\hat{\theta})$ = var$(\hat{\theta})$.

The minimum MSE criterion consists in choosing an estimator whose MSE is the least in a competing set of estimators. But notice that even if such an estimator is found, there is a tradeoff involved-to obtain minimum variance you may have to accept some bias. Geometrically, the situation is as shown in Figure A.10. In this figure, $\hat{\theta}_2$ is slightly biased, but its variance is smaller

than that of the unbiased estimator θ_1. In practice, however, the minimum MSE criterion is used when the best unbiased criterion is incapable of producing estimators with smaller variances.

The last term can be written as $2[E\theta)]^2 - [E\ \theta_1)]^2 - \theta E(\theta) + E(\theta)) = 0$. Also note that $E[\ E(\theta - \theta)^2 = [E(\theta) - \theta\}2 =$. since the expected value of a constant is simply the constant itself.

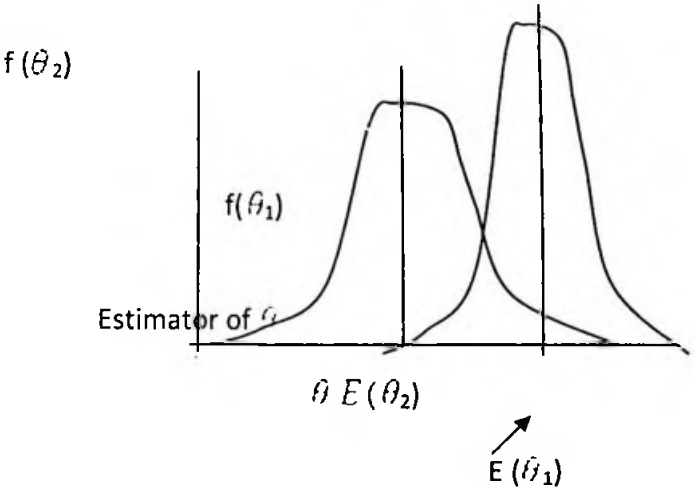


FIGURE A.1 0 Tradeoff between bias and variance.

Large-Sample Properties

It often happens that an estimator does not satisfy one or more of the desirable statistical properties in small samples. But as the sample size increases indefinitely, the estimator possesses several desirable statistical properties. These properties are known as the large-sample, or asymptotic, properties.

Asymptotic Unbiasedness. An estimator e is said to be an asymptotically unbiased estimator of θ if

$$\lim_{n \to \infty} E(\hat{\theta}_n) = \theta$$

where $\hat{\theta}_n$ means that the estimator is based on a sample size of n and where lim means limit and $n \to \infty$ means that n increases indefinitely. In words, θ is an asymptotically unbiased estimator of θ if its expected, or mean, value approaches the true value as the sample size gets larger and larger. As an example, consider the following measure of the sample variance of a random variable X:

$$S^2 = \frac{\Sigma(X_i - X)^2}{n}$$

It can be shown that

$$E(S^2) = \sigma^2\left(1 - \frac{1}{n}\right)$$

where σ^2 is the true variance. It is obvious that in a small sample S^2 is biased, but as n increases indefinitely, $E(S^2)$ approaches true σ^2; hence it is asymptotically unbiased.

Consistency. $\hat{\theta}$ is said to be a consistent estimator if it approaches the true value 0 as the sample size gets larger and larger. Figure A.11 illustrates this property.

In this figure we have the distribution of $\hat{\theta}$ based on sample sizes of 25, 50, 80 and 100. As the figure shows, $\hat{\theta}$ based on n = 25 is

biased since its sampling distribution is not centered on the true θ. But as n increases, the distribution of $\hat{\theta}$ not only tends to be more closely centered on θ (i.e., $\hat{\theta}$ becomes less biased) but its variance also becomes smaller. If in the limit (i.e., when n increases indefinitely) the distribution of $\hat{\theta}$ collapses to the single point θ, that is, if the distribution of $\hat{\theta}$ has zero spread, or variance, we say that $\hat{\theta}$ is a consistent estimator of θ.

More formally, an estimator $\hat{\theta}$ is said to be a consistent estimator of θ if the probability that the absolute value of the difference between $\hat{\theta}$ and θ is less

than δ (an arbitrarily small positive quantity) approaches unity. Symbolically,

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \delta) = 1 \quad \delta > 0$$

where P stands for probability. This is often expressed as

$$\operatorname{plim}_{n \to \infty} \hat{\theta} = \theta$$

where plim means probability limit.

Note that the properties of unbiasedness and consistency are conceptually very much different. The property of unbiasedness can hold for any sample size, whereas consistency is strictly a large-sample property.

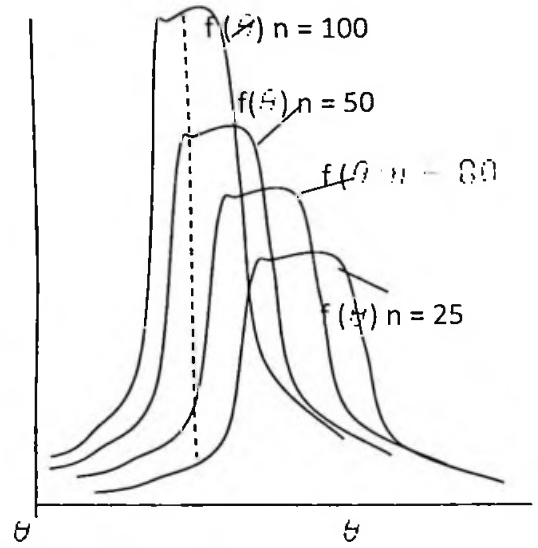


FIGURE A.11

The distribution of $\hat{\theta}$ as sample size increases.

A sufficient condition for consistency is that the bias and variance both tend to zero as the sample size increases indefinitely.[7] Alternatively, a sufficient condition for consistency is that the MSE(B) tends to zero as n increases indefinitely. (For MSE($\hat{\theta}$), see the discussion presented previously.

EXAMPLE 15.30

Let X_1 X_2...X_n be a random sample from a distribution with mean μ and variance σ^2. Show that the sample mean X is a consistent estimator of μ.

From elementary statistics it is known that E(X) = μ. And var(X) = σ^2/n. Since E(X) = μ regardless of the sample size, it is unbiased. Moreover, as n increases indefinitely, var (X) tends toward zero. Hence, X is a consistent estimator of μ.

The following rules about probability limits are noteworthy.

1. Invariance (Slutsky property). If $\hat{\theta}$ is a consistent estimator of θ and if

h(θ) is any continuous function of θ, then

plim$_{n\to\infty}$ h($\hat{\theta}$) = h(θ)

What this means is that if $\hat{\theta}$ is a consistent estimator of θ, then $1/\hat{\theta}$ is also a consistent estimator of $1/\theta$ and that log ($\hat{\theta}$) is also a consistent estimator of log (θ). Note that this property does not hold true of the expectation operator; that is, if $\hat{\theta}$ is an unbiased estimator of θ [that is, E($\hat{\theta}$) = θ], it is not true that $1/\theta$ is an unbiased estimator of $1/\hat{\theta}$; that is, E($1/\hat{\theta}$) ≠ 1/E($\hat{\theta}$) ≠ $1/\theta$.

2. If b is a constant, then

plim$_{n\to\infty}$ b = b

That is, the probability limit of a constant is the same constant.

3. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent estimators, then

plim ($\hat{\theta}_1 + \hat{\theta}_2$) = plim $\hat{\theta}_1$ + plim $\hat{\theta}_2$

plim($\hat{\theta}_1 \hat{\theta}_2$) = plim $\hat{\theta}_1$ plim $\hat{\theta}_2$

plim ($\frac{\hat{\theta}_1}{\hat{\theta}_2}$) $\frac{\text{plim } \hat{\theta}_1}{\text{plim } \hat{\theta}_2}$

The last two properties, in general, do not hold true of the expectation operator E. Thus, $E(\theta_1/\theta_2) \neq (\theta_1)/E(\theta_1)$. Similarly, $E(\theta_{1/}\theta_2) \neq E(\theta_1])E(\theta_2)$. If, however. θ_1 and θ_2 are independently distributed. $E(\theta_1\ \theta_2) = E(\theta_1\ \theta_2)$, as noted previously.

Asymptotic Efficiency. Let θ be an estimator of θ. The variance of the asymptotic distribution of θ is called the asymptotic variance of θ. If θ is consistent and its asymptotic variance is smaller than the asymptotic variance of all other consistent estimators of θ, θ is called asymptotically efficient.

Asymptotic Normality. An estimator θ is said to be asymptotically normally distributed if its sampling distribution tends to approach the normal distribution as the sample size n increases indefinitely. For example, statistical theory shows that if $X_1, X_2,..., X_n$" are independent normally distributed variables with the same mean μ. and the same variance σ^2, the sample mean X is also normally distributed with mean μ. and variance σ^2 /n in small as well as large samples. But if the X_i are independent with mean μ and variance σ^2 but are not necessarily from the normal distribution, then the sample mean X is asymptotically normally distributed with mean μ and variance σ^2/n; that is, as the sample size n increases indefinitely, the sample mean tends to be normally distributed with mean μ and variance σ^2/n. That is in fact the central limit theorem discussed previously.

Statistical Inference: Hypothesis Testing

Estimation and hypothesis testing constitute the twin branches of classical statistical inference. Having examined the problem of

estimation, we briefly look at the problem of testing statistical hypotheses.

The problem of hypothesis testing may be stated as follows. Assume that we have an rv X with a known PDF f(x; θ), where e is the parameter of the distribution. Having obtained a random sample of size n, we obtain the point estimator $\hat{\theta}$. Since the true e is rarely known, we raise the question: Is the estimator $\hat{\theta}$ "compatible" with some hypothesized value of θ, say, $\theta = \theta$, where $\hat{\theta}$ is a specific numerical value of θ? In other words, could our sample have come from the PDF f(x; θ) = θ^*? In the language of hypothesis testing $\theta = \theta^*$ is called the null (or maintained) hypothesis and is generally denoted by Ho. The null hypothesis is tested against an alternative hypothesis, denoted by Hi, which, for example, may state that $\theta = \theta^*$. (Note: In some textbooks, Ho and Hi are designated by Hi and H_2, respectively.)

The null hypothesis and the alternative hypothesis can be simple or composite. A hypothesis is called simple if it specifies the value(s) of the parameter(s) of the distribution; otherwise it is called a composite hypothesis.

Thus, if $X \sim N(\mu\ \sigma^2)$ and we state that

Ho: $\mu = 15$ and $\sigma = 2$

It is a simple hypothesis, whereas

Ho: $\mu = 15$ and $\sigma > 2$

is a composite hypothesis because here the value of a is not specified. To test the null hypothesis (i.e., to test its validity), we use the sample information to obtain what is known as the test

statistic. Very often this test statistic turns out to be the point estimator of the unknown parameter. Then we try to find out the sampling, or probability, distribution of the test statistic and use the confidence interval or test of significance approach to test the null hypothesis. The mechanics are illustrated below.

To fix the ideas, let us revert to Example 23, which was concerned with the height (X) of men in a population. We are told that

$$Xi \sim N(\mu, \sigma^2) = N(u, 2.5^2)$$

$X = 67$ $n = 100$

Let us assume that

Ho: $\mu = \mu = 69$

Hi: $\mu \neq 69$

The question is: Could the sample with $X = 67$, the test statistic, have come from the population with the mean value of 69? Intuitively, we may not reject the null hypothesis if X is "sufficiently close" to μ'; otherwise we may reject it in favor of the alternative hypothesis. But how do we decide that X is "sufficiently close" to μ^*? We can adopt two approaches, (1) confidence interval and (2) test of significance, both leading to identical conclusions in any specific application.

The Confidence Interval Approach

Since $Xi \sim N(\mu, \sigma^2)$, we know that the test statistic X is distributed as

$$X \sim N(\mu, \sigma^2/n)$$

Since we know the probability distribution of X, why not establish, say, a

100(1 - α) confidence interval for based on X and see whether this confidence interval includes $\mu = \mu^*$? If it does, we may not reject the null hypothesis; if it does not, we may reject the null hypothesis. Thus, if a = 0.05, we will have a 95% confidence interval and if this confidence interval includes μ', we may not reject the null hypothesis - 95 out of 100 intervals thus established are likely to include μ^*.

The actual mechanics are as follows: since $X \sim N(\mu, \sigma^2/n)$, it follows that

$$Zi = \frac{X-\mu}{\sigma/\sqrt{n}} \qquad N(0, 1)$$

that is, a standard normal variable. Then from the normal distribution table we know that

$$Pr(-1.96 \leq Z_i \leq \leq 1.96) = 0.95$$

That is,

$$Pr\left(-1.96 \leq \frac{X-\mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

which, on rearrangement, gives

$$Pr\left(X - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq X + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

This is a 95% confidence interval for μ. Once this interval has been established, the test of the null hypothesis is simple. All that we have to do is to see whether $\mu = \mu$ lies in this interval. If it does, we may not reject the null hypothesis; if it does not, we may reject it.

Turning to our example, we have already established a 95% confidence interval for μ, which is

$$66.51 \leq \mu \leq 67.49$$

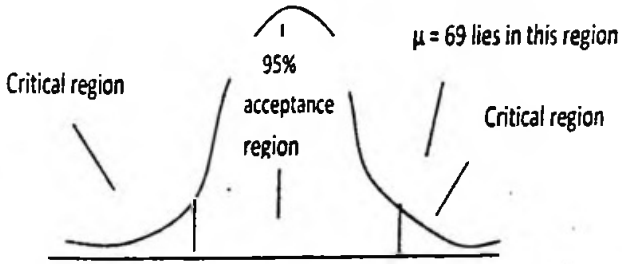
This interval obviously does not include $\mu = 69$. Therefore, we can reject the null hypothesis that the true μ is 69 with a 95% confidence coefficient.

Geometrically, the situation is as depicted in Figure A.I2.

In the language of hypothesis testing, the confidence interval that we have established is called the acceptance region and the area(s) outside the acceptance region is (are) called the critical region(s), or region(s) of rejection of the null hypothesis. The lower and upper limits of the acceptance region (which demarcate it from the rejection regions) are called the critical values. In this language of hypothesis testing, if the hypothesized value falls inside the acceptance region, one may not reject the null hypothesis; otherwise one may reject it.

It is important to note that in deciding to reject or not reject Ho, we are likely to commit two types of errors: (1) we may reject Ho when it is, in fact true; this is called a type 1 error (thus, in the preceding example X = 67 could have come from the population with a mean value of 69), or (2) we may not reject Ho when it is, in fact, false; this is called a type II error.

Therefore, a hypothesis test does not establish the value of true μ. It merely provides a means of deciding whether we may act as if $\mu = \mu$.



95% Confidence Interval for μ

Type 1 and Type II Errors. Schematically, we have

State of nature / Decision	Ho is true	Ho is false
Reject	Type I error	No error
Do not reject	No error	Type II error

Ideally, we would like to minimize both type I and type II errors. But unfortunately, for any given sample size, it is not possible to minimize both the errors simultaneously. The classical approach to this problem, embodied in the work of Neyman and Pearson, is to assume that a type I error is likely to be more serious in practice than a type II error. Therefore, one should try to keep the probability of committing a type I error at a fairly low level, such as 0.01 or 0.05, and then try to minimize the probability of having a type II error as much as possible.

In the literature, the probability of a type I error is designated as α and is called the level of significance, and the probability of a type II error is designated as β. The probability of not committing a type II error is called the power of the test. Put differently, the power of a test is its ability to reject a false null hypothesis. The classical approach to hypothesis testing is to fix α at levels such as 0.01 (or 1 percent) or 0.05 (5 percent) and then try to maximize the power of the test; that is to minimize β.

It is important that the reader understands the concept of the power of a test, which is best explained with an example.[8]

Let $X \sim N(\mu, 100)$; that is, X is normally distributed with mean μ and variance 100. Assume that $\alpha = 0.05$. Suppose we have a sample of 25 observations, which gives a sample mean value of X. Suppose further we entertain the hypothesis Ho: $\mu = 50$. Since X is normally distributed, we know that the sample mean is also normalcy distributed as: $X \sim N(\mu, 100/25)$. Hence under the stated null hypothesis that $\mu = 50$, the 95% confidence interval] for X is $(\mu \cdot 1.96\underline{(\sqrt{100/25})} = \mu - 3.92$, that is, (46.08 to 53.92). Therefore, the critical region consists of all values of X less than 46.08 or greater than 53.92. That is, we will reject the null hypothesis that the true mean is 50 if a sample mean value is found below 46.08 or greater than 53.92.

But what is the probability that X will lie in the preceding critical region(s) if the true μ has a value different from 50? Suppose there are three alternative hypotheses: $\mu = 48$. $\mu = 52$, and $\mu = 56$. If any of these alternatives is true, it will be the actual mean of the distribution of X. The standard error is unchanged for the three alternatives since σ^2 is still assumed to be 100.

The shaded areas in Figure A.13 show the probabilities that X will fall in the critical region if each of the alternative hypotheses is true. As you can check, these probabilities are 0.17 (for $\mu = 48$). 0.05 (for $\mu = 50$), 0.17 (for

$\mu = 52$) and 0.85 (for $\mu = 56$). As you can see from this figure, whenever the true value of μ differs substantially from the hypothesis under consideration (which here is $\mu = 50$), the probability of rejecting the hypothesis is high but when the true value is not very different from the value given under the null hypothesis, the probability of rejection is small. Intuitively, this should make sense if the null and alternative hypotheses are very closely bunched.

This can be seen further if you consider Figure A.14, which is called the power function graph, and the curve shown there is called the power curve.

The reader will by now realize that the confidence coefficient ($1 - \alpha$) discussed earlier is simply one minus the probability of committing a type

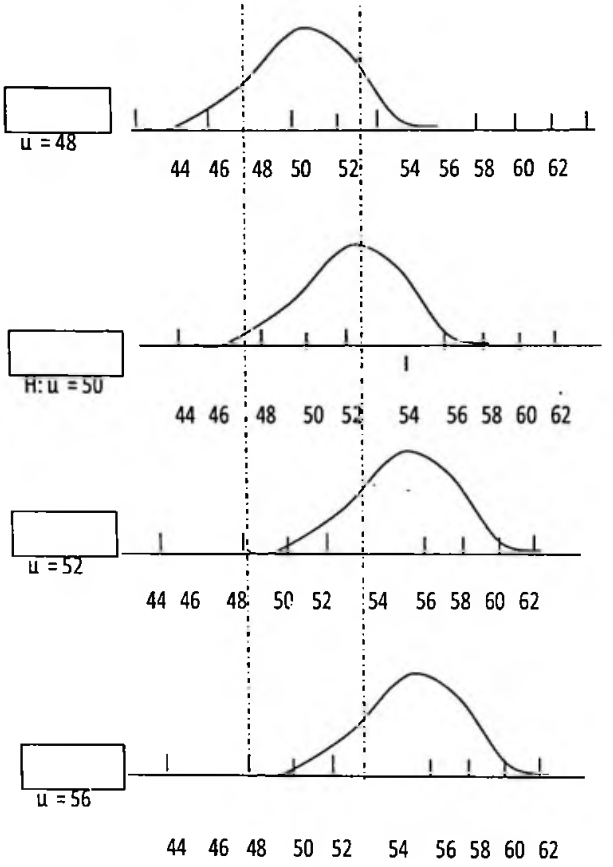


Figure 13: Distribution of X when N = 25, σ = 10, and μ = 48, 50,52, 0r 56. Under H: μ = 50, the critical region with α = 0.005 is X < 46.1 and X > 53.9. The shaded area indicates the probability that X will fall into the critical region. This probability is:

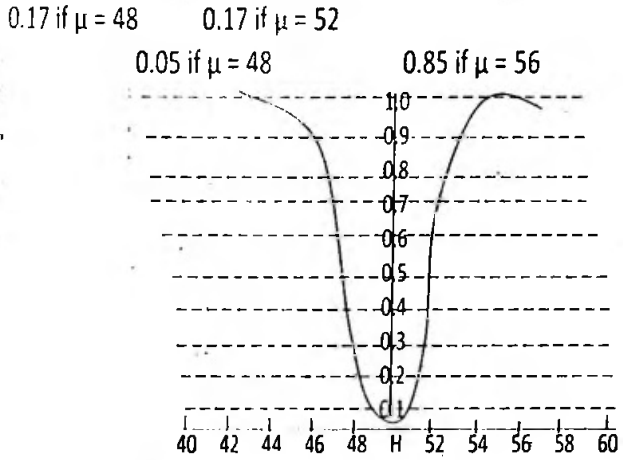


Figure a 14: power function confidence coefficient means that we are prepared to accept at the most a 5% probability of committing a type 1 error – we do not want to reject the true hypothesis by more than 5 out of 100 times.

The p value or Exact Level of Significance. Instead of preselecting α at arbitrary levels, such as 1,5, or 10 percent, one can obtain the p (probability) value, or exact level of significance of a test statistic. The p value is defined as the lowest significance level at which the null hypothesis can be rejected.

Suppose that in an application involving 20 df we obtain a t value of 3.552. now the p value, or the exact probability, of obtaining a t value of 3.552 or greater can be seen from Table .2 as 0.001 (one-tailed) or 0.002 (two-tailed). We can say that the observe t value of 3.552 is statistically significant at the 0.001 0r 0.002 level, depending on whether we are using a one- tail or two-tail test.

Several statistical packages now routine print out the p value of the estimate test statistics. Therefore, the reader is advised to give the p value wherever possible.

The test significance of Approach

Recall that

$$Z\,i = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

In any given application, X and n are known (or can be estimated), but the true μ and σ are not known. But if σ is specified and we assume (under Ho) that $\mu = \mu$, a specific numerical value, then Zi can be directly computer and we can easily look at the normal distribution table to find the probability of obtaining the computed Z value. If this probability is small, yes, less than 5% percent or 1 percent, we can reject the null hypothesis – if the hypothesis were true, the chances of obtaining the particular Z value should be very high. This is the general idea behind the test of significance approach to hypothesis testing. The key idea here is the test statistic (here the Z statistic) and its probability distribution under the assumed value $\mu = \mu$. Appropriately, in the present case, the test is known as the Z test, since we use the Z (standardized normal) value.

Returning to our example, if $\mu = \mu = 69$, the Z statistic becomes

$$Z\,i = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$= -.2/0.25 = -8.$

If we look at the normal distribution table D.1, we see the probability of obtaining such a Z value is extremely small. (Note: The probability of a Z value exceeding 3 or -3 is about 0.001. therefore, the probability of Z exceeding 8 is even smaller). Therefore we can reject the null hypothesis that $\mu = 69$; given this value, our chance of obtaining X of 67 is extremely small. We therefore doubt our sample came from the population with a mean of 69. Diagrammatically, the situation is depicted in Figure a. 15.

In language of test of significance, when we say that a test (statistic) is significant, we generally mean that we can reject the null hypothesis. And the test statistic is regarded as significant if the probability of our obtaining it is equal to or less than α, the probability of obtaining committing a type I error. Thus, if α= 0.05, we know that the probability of obtaining a Z value of -1.96 or 1.96 is 5 percent (or 2.5 percent in each tail of the standardised normal distribution). I our illustrative example, Z was -8. Hence the probability of obtaining such a Z value is much smaller than 2.5 percent, well below our pre-specified probability of committing a type I error. That is why the compute value of Z = -8 is statistically significant; that is, we reject the null hypothesis.

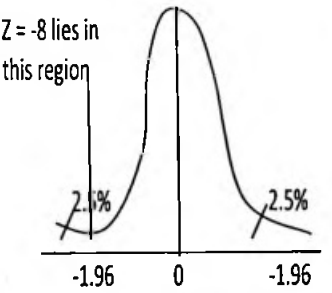


Figure A.15 The distribution of the Z statistic

Hypothesis that the true μ is 69. Of course, we reached the same conclusion using the confidence interval approach to hypothesis testing.

We now summarize the steps involved in testing a statistical hypothesis again:

Step 1. State the null hypothesis Ho and the alternative hypothesis H_1 (e.g., Ho: μ = 69 and H_1: μ = 69).

Step 2. Select the test statistic (e.g., X).

Step 3. Determine the probability distribution of the test statistic (e.g., $X \sim N(\mu, \sigma^2/n)$.

Step 4. Choose the level of significance (i.e., the probability of committing a type 1 error) α

Step 5. Using the probability distribution of the test statistic, establish a 100 (1 -α)% confidence interval. If the value of the parameter under the null hypothesis (e.g., μ = μ = 69) lies in this confidence region, the region of acceptance, do not reject the null hypothesis. But if it falls outside this interval (i.e., it falls into the region of rejection), you may reject the null hypothesis. Keep

in mind that in not rejecting or rejecting a null hypothesis you are taking a chance of being wrong a percent of the time.

REFERENCES

Arise Spanos (1968) *Probability theory and statistical inference: Econometrics modeling with observational data.*

Allen, R.G.O. (1966), *Mathematical Economics*, 2ns ed. Macmillan, London,.

Almon, S. (1965), *"The Distributed Lag between Capital Appropriations and Expenditures"*, Econometrica, Vol.30, pp. 178-96.

Amemiya, T. (1985), *Advanced Econometrics, Basil Blackwell, Oxford,.*

Anderson, R.L. (1952) and Bancroft, T.A., Statistical Theory in Research, New York: McGraw Hill.

Barnett, R. A. (1979), *Finite Mathematic*, 2nd edition; Califonia; Dellen publishing company.

Basmann, R.L., (1957), *"A Generalized Classical Method for Linear Estimation of Coefficients in Structural Equation, Econometric*, Vol.25, pp.77-83.

Damodar, Gujarati, (1970) *Basic econometrics* 2nd edition London.

Dominick Salvatore (1982) *Theory and problems of statistics and Economics;*

Damodar Gujarati N. (2005); *Basic Econometrics*; New delhi; Tata McGraw-Hill publishing company limited.

Desai, M. (1976) *Applied Econometrics*, London: Philip – Allan,.

Dhrymes, P.J. (1970) *Econometrics: Statistical Foundation and Application* New York: Harper and Row,.

Dhrymes, P.J. (1978) *Introductory Econometrics*, New York: Springer – Verlag,.

Durbin J. (1953*)*, "A note on Regression when there is Extraneous Information on one of the Coefficient", Journal of the American Statistical Association, Vol.48, pp. 799-808.

Durbin, J., and Watson, G.S., *"Testing for Serial Correlation in Least Squared Regression, Biometrica*, Vol.37, pp. 409-28.

Epstein, R.J., (1989) *A History of Econometrics*, Amsterdam: North Holland Publishers,.

Frank, C.R.J. (1971), *Statistics and Econometrics*, Holt, Rinehart and Winston.

Golberger Aethus S. (1998); *Introductory Econometrics*, Harvard University Press.

Gujarati Damdar N. (1999), New York; *Essentials of Econometrics* second edition McGraw-Hills.

Hill, Carter, William Griffiths and George Judge (2001); *Undergraduate Econometrics* New York, John Willis and sons.

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

Katz, David A (1982); *Econometric theory and Applications,* Prentice Hall, Engle-Wood Cliffs, NJ.

Klein, Lawrence R; (1962); *An Introduction to Enconometrics,* Prentice Hall, Engle-Wood Cliffs N.J.

Koop, Gary, (2000); *Analysis of Economic data*; New York; John Willis and Sons.Shamim's outline series New York: MaCGraw Hill-INC.

Haires, Brain (1978) *Introduction to Quantitative Economic;* London: George Allen & Unwin.

Johnson J. (1963) *Econometrics methods*, 2nd Edition, Aukland International Students edition: MaCGraw Hill-INC.

Johnson S. (1991) *Estimation and Hypothesis Testing of Intergration Vector: Econometrics* London.

Omotosho M.Y (2000) *Practical Econometrics* Ibadan.

Upender M. (1977) *Applied Econometrics* 1st Edition: London.

Panneerselvan R. (2000) *Research Methodology* New Delhi-110001 PHI Learning private limited.

Walters, A.A. (1968) *An Introduction Of Economics*; London, Macmillan.

Oyeniyi, T.A. (1997)) *Fundamental Principles of Economics;* Lagos; Cedar Group Nigeria Limited.

Farrar, D.E., and Glauber, R.R., "*Multicollinearity in Regression Analysis.* The problem Resisted, Review of Economics and Statistics. Vol. 49, pp. 92-107.

Freund, J.E., and Walpole, R.E.,(1987) *Mathematical Statistics*, 4th ed., New York: Prentice Hall International

Fox K.A., (1958) Econometric *Analysis for public policy*, Iowa State College Press, 1958

Goldbeger, A.S., (1964) Econometric Theory New York John Wisley & Sons,

Goldberger, A. S., A Course in Econometrics, Cambridge, Mass: Harward University Press, 1991.

Gujarati D., Basic Econometrics, 2nd edition, London: McGraw-Hill Publishing Company, 1978.

Haavelmo, Tryge, "The Role of the Econometrics in the Advancement of Economic Theory", *Economietrica*, Vol.26. No.3 pp.351-357.

Haavelmo, T., "The Probability Approach in Econometrics", ***Supplement to Econometrica***, Vol.12, pp. 1-118.

Hadley, G., (1965)Liner Algebra, Mass: Addition – Wesley,.

Harvey, A. C., The Econometric Analysis of Time Series, 2nd ed., Philip. Allan Hasel Hempstead.

Johnston, J., (1984) Econometric Methods, 3rd ed. New York: Mc Grill Hill,.

Kadiyala, K.R.(1968) "A Transformation Used to Circumvent the Problem of Autocorrelation", *Econometrica*, Vol.36, No.1,. pp. 93-96.

Kendall, M., and Ord, J. K., (1990) Time Series, 3rd edition, London: Editorial Arnold,.

P.A., A Guide to Econometrics, 2nd edition, Martin Roberston, Oxford, 1990

Kelejian, H.H., and Oats, W.E., Introduction to Econometrics – Principles and Applications, 3rd edition, New York; Harper & Row Publishers, 1989.

Klein, L.R., A Textbook of Econometrics, Row Peterson. Evanston, 111, 1953.

Klein, L.R., Whither Econometrics, Journal of the American Statistical Association, Vol.66, 1971 pp. 415-421.

Kmenta, J., Elements of Economics, 2nd edition; New York Macmillan, 1986.

Koopmans, T.C., Measurement Without Theory, Review of Economics and Statistics, Vol.29, 1974 pp.161-172.

Koutsoyiannis, A., Theory of Economics, 2nd edition, London: Macmillian Publishing Company, 1977.

Koyck, L.M., Distributed Lags and Investment Analysis, Amsterd: North Holland Publishing Company, 1954.

Kramer, W., Ploberger, W., and Alt, R., "Testing for Structural Change in Dynamic Models", Econometrical, Vo. 56, pp. 1355-1369.

Leamer, E.E., "Let's take the con out of Econometrics", American Economic Review, Vol.23, pp.31-43, 1983 reprinted in Granger, C.W.J., (ed) Modeling Economic Series: Readings in Econometric Methodology, Oxford University Press, Oxford.

Liu, T.C., "Underindentification, Structural Estimation and Forecasting", Econometrical, Vol.28, pp.835-865.

Maddala, G.S., (ed) "Model Selection", Journal of Econometrics. Vol.16, 1981.

Maddala, G.S., Introduction to Econometrics, London Macmillan, 1988.

Mills, T.C., Time Series Techniques for Economists, Cambridge: Cambridge University Press, 1990.

Mogan, M.S., The History of Economic Ideas, Cambridge: Cambridge University Press, 1990.

Nerlove, M., Distributed Lags and Demand Analysis, Agricultural handbook, 141, US Department of Agriculture, Washington, 1958.

Nerlove, M., Returns to Scale in Electricity Supply, in Christ, C., (ed), Measurement in Economics, Standford University Press, 1963.

Orcult, G.H., et al. Micro analysis of Socioeconomic Systems: A Simulation Study New York: Harper & Row, Publishers, 1981.

Pagan, A.R., "Three Econometric Methodologies: A critical Appraisal", Journal of Economic Surveys Vol.8, pp. 3-24, 1987, reprinted in Granger, C.W.J., (ed) Modeling Economic series: Reading in Econometric Methodology, Oxford: Oxford University Press, 1990.

Quandt, R.E., "On Certain Small Sample Properties of K-class Estimators",

Roos, C.F., "Survey of Econometric Forecasting Techniques", Econometrica, Vol.23, 1955 pp.363-395.

Schumpeter, J.A., "The Common Sence of Econometrics", Econometrica, Vol.1, No.1, pp.5-12, 1833.

Stewart, J., Understanding Economics, 2nd ed. Hutchinson Publishing Company, 1984.

Stone, J., and Richard, N., The Role of Measurement in Economics, Cambridge University Press, 1951.

Theil, Henri, Estimation and Simultaneous Correlation in Complete Equation System, (The Hague: Central Planning Bureau, 1953.

Theil, Henri, Economic Forecast and Policy 2nd ed., North – Holland Publishing Company, 1961.

Thomas, R.L., Introductory Economics: Theory and Application, New York. Longman Inc., 1985.

Walters, A.A., An Introduction to Econometrics, 2nd ed., Macmillan Publishing Company, 1970.

Zellner, A., Huang, D.S and Chan, L.C., "Further Analysis of the Short-run Consumption Functions with Emphasis on the Role of Liquid Assets'. Econometrica, Vol. 33, 1965 pp. 571-81.

Zelliner, A., and Geisel, M.k., Analysis of Distributed Lag models with Applications to Consumption Function Estimation, Econometrica, Volume 38., 1970 pp.865-88.

A

Absolute change. 69, 70.
Acceptance region.101,167.
Adjusted coefficient of
Determination.51,94.
Almon distributed lag
Models.107.
Alternative hypothesis.216.
Analysis of variance
Models.157,197..
ANOVA table. 204..
Augmented Dickey-Fuller
Test.138.
Augmented Engle-
Granger test.142
Autocorrelation .173.
Autocorrelation
Coefficient.180..
Autocorrelation function.1l74
Autoregressive scheme
Autoregression.122.
Auto regression models.122.
A priori expectation.13..

B

Best linear unbiased
Estimator, 33,61,283.
Best unbiased estimator
Bias.61,283.
Bilateral causality.155..
Bivariate regression.30..
BLUE:see Best linear 61.
unbiased
BLUE: see Best unbiased
Estimator
Binomial distribution.272.

C

Causality.155.
Causation. 18.
Central limit theorem. 52.
Chi-square distribution. 205.
Chi-square goodness of fit. 205.
Chi-square test. 211.
Chow test. 105..
Classical econometric. 33..
Classical linear regression
Classical normal linear
Regressions.30.
Coefficient
Coefficient of
Determination. 33..
Coefficient of expectation,
Conditional regression
Analysis. 30.
Confidence interval. 59.
Confidence limits
Consistency. 61, 283.
Consistent estimator.283,
Constant elasticity model. 82.
Constant growth model. 70.
Correlation analysis. 20.
Correlation coefficient. 22.
Covariance. 253.

D

Dependent variable. 8,18.
Deterministic trend. 69.
Detrended time series. 126.
Dichotomous dependent
Dichotomous response
Model. 159.
Dickey-Fuller test. 133.
Dickey-Fuller unit root
Test. 133.
Difference equation
Difference-stationary
Difference stationary
Stochastic. 1126.
Differential slope
coefficient
Distributed lag model. 107.
Distributed lag multiplier. 114.
Downward trend. 69.
Dummy regressor. 160.
Dummy variables. 159.
Dummy-variable trap,159.
Durbin's *h* statistic, 144, 173.
Durbin's *h* test, 173.
Durbin-Watson
Cointegration test,142.

A
Absolute change. 69, 70.
Acceptance region.101,167.
Adjusted coefficient of
Determination.51,94.
Almon distributed lag
Models.107.
Alternative hypothesis.216.
Analysis of variance
Models.157,197..
ANOVA table. 204..
Augmented Dickey-Fuller
Test.138.
Augmented Engle-
Granger test.142
Autocorrelation .173.
Autocorrelation
Coefficient.180..
Autocorrelation function.1i74
Autoregressive scheme
Autoregression.122.
Auto regression models.122.
A priori expectation.13..

B
Best linear unbiased
Estimator, 33,61,283.
Best unbiased estimator
Bias.61,283.
Bilateral causality.155..
Bivariate regression.30..
BLUE:see Best linear 61.
unbiased
BLUE: see Best unbiased
Estimator
Binomial distribution.272.

C
Causality.155.
Causation. 18.
Central limit theorem. 52.
Chi-square distribution. 205.
Chi-square goodness of fit. 205.
Chi-square test. 211.
Chow test. 105..
Classical econometric. 33..
Classical linear regression
Classical normal linear
Regressions.30.
Coefficient
Coefficient of
Determination. 33..
Coefficient of expectation,
Conditional regression
Analysis. 30.
Confidence interval. 59.
Confidence limits
Consistency. 61, 283.
Consistent estimator.283,
Constant elasticity model. 82.
Constant growth model. 70.
Correlation analysis. 20.
Correlation coefficient. 22.
Covariance. 253.

D
Dependent variable. 8,18.
Deterministic trend. 69.
Detrended time series. 126.
Dichotomous dependent
Dichotomous response
Model. 159.
Dickey-Fuller test. 133.
Dickey-Fuller unit root
Test. 133.
Difference equation
Difference-stationary
Difference stationary
Stochastic. 1126.
Differential slope
coefficient
Distributed lag model. 107.
Distributed lag multiplier. 114.
Downward trend. 69.
Dummy regressor. 160.
Dummy variables. 159.
Dummy-variable trap,159.
Durbin's *h* statistic, 144, 173.
Durbin's *h* test, 173.
Durbin-Watson
Cointegration test,142.

A
Absolute change. 69, 70.
Acceptance region.101,167.
Adjusted coefficient of
Determination.51,94.
Almon distributed lag
Models.107.
Alternative hypothesis.216.
Analysis of variance
Models.157,197..
ANOVA table. 204..
Augmented Dickey-Fuller
Test.138.
Augmented Engle-
Granger test.142
Autocorrelation .173.
Autocorrelation
Coefficient.180..
Autocorrelation function.1i74
Autoregressive scheme
Autoregression.122.
Auto regression models.122.
A priori expectation.13..

B
Best linear unbiased
Estimator, 33,61,283.
Best unbiased estimator
Bias.61,283.
Bilateral causality.155..
Bivariate regression.30..
BLUE:see Best linear 61.
unbiased
BLUE: see Best unbiased
Estimator
Binomial distribution.272.

C
Causality.155.
Causation. 18.
Central limit theorem. 52.
Chi-square distribution. 205.
Chi-square goodness of fit. 205.
Chi-square test. 211.
Chow test. 105..
Classical econometric. 33..
Classical linear regression
Classical normal linear
Regressions.30.
Coefficient
Coefficient of
Determination. 33..
Coefficient of expectation,
Conditional regression
Analysis. 30.
Confidence interval. 59.
Confidence limits
Consistency. 61, 283.
Consistent estimator.283,
Constant elasticity model. 82.
Constant growth model. 70.
Correlation analysis. 20.
Correlation coefficient. 22.
Covariance. 253.

D
Dependent variable. 8,18.
Deterministic trend. 69.
Detrended time series. 126.
Dichotomous dependent