Econometric Analysis

Econometric Analysis:

An Applied Approach to Business and Economics

By

Sharif Hossain

Cambridge Scholars Publishing

Econometric Analysis: An Applied Approach to Business and Economics

By Sharif Hossain

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-6337-5
ISBN (13): 978-1-5275-6337-7

TABLE OF CONTENTS

PREFACE

During the last four decades, the uses of econometric tools and techniques in different areas of human activities-irrespective of any discipline have been increasing rapidly for solving problems and appropriate decision-making and policy design. It is now generally accepted in all societies that the study of modern business, economics, socio-economics, finance, banking and management is incomplete without proper knowledge of the basic and advanced level of econometric tools and techniques. Given the increasing complexity and variety of problems, in the fields of business, economics, socio-economics, finance, banking, management, medical sciences, psychology, sociology and natural sciences, students and researchers without a fair knowledge of econometrics may not be able to cope, and hence, may remain unfamiliar with many aspects of business, economic, socio-economic, finance, banking, managerial and scientific problems. Thus, the book **"Econometrics Analysis: An Applied Approach to Business and Economics"** is written to study basic and advanced studies of econometric courses for the undergraduate and graduate levels of students of business, economics, socio-economics, finance, banking, management and other social sciences as well as for the researchers already engaged in these fields who desire an introduction to econometric methods and their applications for solving real-life problems and reform policies in many aspects. In writing the first edition of this book, I aim to offer students and researchers a balanced presentation of fundamental and advanced levels of econometric concepts and methods along with practical advice on their effective application to real-life problems. This book covers almost all aspects of econometric techniques giving priority to modelling, inferential statistics, estimation, and testing of statistical hypotheses based on cross-sectional, time-series and panel data. This book includes the topics of basic concepts and ideas of econometrics; simple and multiple linear and non-linear regression analyses; the problems of single equation models: heteroscedasticity, autocorrelation, and multicollinearity; selection of the best regression equation along with the diagnostic tests; single variate and multivariate time-series econometric models; limited dependent-variable models; and econometric analyses of panel data models. It is known to us that econometrics is frequently interested in modelling to describe how the dependent and independent variables are related to each other. The principal objective of econometrics is to quantify the relationships between dependent and independent variables based on numerical data using statistical techniques, and for interpretation, and the resulting outcomes play a significant role in appropriate decision-making and reform policies. In general, econometrics deals with the macroeconomic relationships to reform policies for the economic development of any society. Nowadays, econometric tools and techniques are widely applicable and they play an important role in almost every field of social, business, economic, and scientific research and for the empirical verification of the theoretical construct. The wide application of econometrics in the field of finance has given rise to a new discipline namely Financial Econometrics. A preliminary study of econometrics requires a clear idea of the basic concepts and ideas. Keeping that in mind, primarily, the fundamental concepts and ideas of econometrics are discussed along with different functions which are associated with regression modelling. This book has been organised as follows:

Introductory: some basic concepts and ideas of econometrics are introduced in Chapter 1. In Chapter 2, the simple linear and non-linear regression models along with how to fit them to a set of data points using different estimation methods namely: the ordinary least-squares (OLS) method, maximum likelihood method, and the method of moments are presented. In this chapter, the important properties of OLS estimators are discussed along with their proofs. Later on, it is shown how to judge whether a relationship between a dependent variable (y) and independent variable (x) is statistically significant, how to use the model to estimate the expected value of y, and how to forecast a future value of y for a given value of x. In this chapter, proper justifications for the techniques employed in regression analysis are also provided along with numerical problems.

In Chapter 3, the multiple linear and non-linear regression models are presented along with their estimation techniques. In this chapter, the important properties of the OLS and ML estimators are discussed with their proofs. In this chapter, it is shown how to judge whether a relationship between a dependent variable (y) and independent variables X's are statistically significant or not. In this chapter, the Wald test, the F-test, and the LM test are also discussed for testing joint null hypotheses and general linear hypotheses. This chapter discusses some fundamental theorems based on the OLS estimators and their proofs. The ANOVA tables are also presented in fitting the multiple regression equations in different cases in this chapter. All the techniques are described along with their applications to the numerical problems.

In chapters 4, 5, and 6, problems of a single equation model namely: heteroscedasticity, autocorrelation and multicollinearity are presented along with the consequences and their proofs. Different tests for detecting the problems of heteroscedasticity, autocorrelation and multicollinearity and different estimation techniques of single-equation models with these problems are discussed. All the techniques are presented along with their applications to numerical problems.

In Chapter 7, different criteria for selecting the best regression equation along with the diagnostic tests are presented. In this chapter, all the techniques are discussed along with their applications to numerical problems.

In Chapter 8, different time-series econometrics models such as AR, MA, ARMA, ARIMA, ARCH, GARCH, EGARCH, MGARCH and GARCH-in Mean models are presented along with their important properties and estimation techniques. In this chapter, all the techniques are presented along with their applications to numerical problems. We know that the usual techniques of regression analyses can result in highly misleading conclusions when the time-series variables contain stochastic trends (Stock and Watson (1988), Nelson and Kang (1981), Granger and Newbold (1974)). In particular, if the dependent variable and at least one independent variable contain a stochastic trend, and if they are not co-integrated, the regression results are spurious (Phillips (1986), Granger and Newbold (1974)). Therefore, to identify the correct specification of the model, an investigation of the presence of stochastic trends in the time-series variables is very important. Therefore, in Chapter 9, the most popular and widely applicable tests to investigate whether time series data contain a stochastic trend or the unit root problem namely: Dickey and Fuller (DF), Phillips-Perron (PP), the Augmented Dickey-Fuller (ADF), and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) tests are discussed with their derivations. The DF, ADF and PP tests are discussed in four different cases. In this chapter, first, the meaning of unit root is explained with the numerical problem and later on, convergence rates of the OLS estimators of unit root processes are discussed. Brownian motion along with the functional central limit theorem are also discussed in this chapter. The chapter also discusses different unit root tests for different cases along with their applications to numerical problems. We know that the multivariate time-series models are very important relative to univariate time-series models, because they may improve forecasts, as a result, appropriate decisions can be taken and appropriate policies can be formulated. Therefore, in Chapter 10, multivariate time-series models such as dynamic, VAR, VEC and ARDL models are presented. In this chapter, the Engle and Granger test procedure is presented to test for detecting the causal relationships between pairs of variables. Three different types of models are presented for detecting the direction of causality between two variables X and Y depending upon the order of integration and the presence or absence of the cointegration relationship. In this chapter, the Toda-Yamamoto approach is also discussed for detecting the direction of causality between two variables X and Y, when X and Y may be integrated with different orders or not integrated or in both. In this chapter, all the techniques are discussed with their applications to numerical problems.

In Chapter 11, the most important and widely applicable limited dependent variable regression models namely the linear probability model, logit model, probit model and Tobit model are presented for detecting the probability of an event which is associated with different independent variables along with their important properties and estimation techniques. In this chapter, the censored and truncated regression models are also presented along with their important properties and estimation techniques. In this chapter, the Poisson regression model is also presented along with the important properties and estimation techniques to deal with the count-dependent variable. In this chapter, all the techniques are discussed with their applications to numerical problems. Sometimes, we have to deal with the panel data for econometric analyses of economic relationships in advanced studies programs. Therefore, in Chapter 12, an econometric analysis for panel data models is also presented. The most popular and widely applicable panel-data regression model, fixed-effects model, and random-effects model along with their important properties and estimation techniques are presented. In this chapter, different panel unit-root tests and cointegration tests are also presented. In this chapter, all the techniques are discussed with their applications to numerical problems. If the students of the undergraduate and graduate levels study this book, they should be able to know that the econometric tools and techniques are very fundamental and important for an in-depth understanding of the subject matter and help deal with the real-life problems irrespective of any discipline; to formulate, estimate, test and interpret suitable models for empirical studies irrespective of any discipline; to develop the ability for using econometric argumentation in the exposition of economic, socio-economic and business problems; to apply different econometric tools and techniques for solving different kinds of business, economic, socio-economic, financial, banking, and managerial problems based on cross-sectional, time series and panel data and for appropriate decision and policy making; to specify a time-series model for explanation or forecasting with an eye on the problems of non-stationarity, spuriousness, and seasonality; to apply different univariate time-series models such as AR, MA, ARMA, ARIMA, ARCH, GARCH, EGARCH, GARCH-in Mean models to deal with the real-life financial problems; to apply different multivariate time-series econometric models such as VAR , VEC and SURE models to deal with our real-life problems and to formulate short-run as well as long-run policies; to apply different econometric tools and techniques for panel-data analyses; to get the ideas in many software-based applications such as EViews, RATS and STATA to deal with numerical problems irrespective of any discipline; to apply econometric tools and techniques to write their Master's and Ph.D. thesis; and to apply econometric methods for different research projects by researchers. Thus, it can be said that the book will be appreciated by the students of undergraduate and graduate levels, professionals, trainers, and researchers who are doing research in the fields of business, economics, socio-economics, finance, banking, management and other physical and natural sciences. It can also be said that this book is a substantial addition to the voluminous literature of cross-sectional, time-series and panel data regression models and formally trained econometricians and statisticians interested in understanding many applications of econometric tools and techniques to deal with real-life problems might find this book helpful. I would like to express my thanks to my student Bablu Nasir for his sincere cooperation. I would also like to acknowledge my gratefulness to Cambridge Scholars Publishing and its editors and staff for taking special care towards the publication of this book.

CHAPTER ONE

INTRODUCTION: BASIC CONCEPTS AND IDEAS

1.1 Introduction

Econometrics is frequently interested in illustrating the behaviours of the relationships between dependent and independent economic variables. Thus, it can be said that the main concern of econometrics is to describe how the dependent and independent variables are related to each other by modelling. The principal objective of econometrics is to quantify the relationships between dependent and independent variables based on numerical data, using statistical tools and techniques to ease interpretation. The resulting outcomes play an important role in appropriate decision-making and policy formulation. In general, econometrics deals with macroeconomic relationships, but nowadays, econometric tools and techniques are also widely used in almost every field of social science, business, finance and economic research and for empirical verification of the theoretical construct. The wide application of econometrics in the field of finance has given rise to a new discipline called Financial Econometrics. Studying econometrics will introduce econometric tools and techniques which are fundamental and important for an in-depth understanding, and help to deal with business, economics, socio-economic, financial and scientific problems; to formulate, estimate, test and interpret suitable models for empirical study of economic, socio-economic, finance and business phenomenon; to develop the ability to use the econometric argumentation in exposition of economics, socio-economics, finance and business; to apply different econometric tools and techniques for solving different kinds of economic, socio-economic, financial and business problems and for appropriate decision making and policy formulations; to specify a time-series model for explanation or for forecasting with an eye on the problems of non-stationarity, spurious regression, and seasonality; to apply different econometric tools and techniques for panel-data analyses; to apply different software packages including EViews, GAUSS, LIMDEP, Python, R, RATS, SPSS, STATA, etc. for solving different problems irrespective of any discipline; to apply econometric methods to write Master's and Ph.D. thesis, and to apply econometric methods for research projects. Therefore, to study econometrics primarily, we have to get a clear idea about some basic concepts and ideas. In this chapter, some preliminary concepts and ideas which are associated with the study of econometrics are discussed. Also, different functions which are associated with regression modelling are discussed here.

1.2 Meaning of Econometrics

Econometrics is defined as the mathematical, empirical or numerical measurement of economic relationships in which the dependent variable is a function of deterministic components plus a random error term. In general, econometrics is defined as the branch of economics which deals with mathematical, empirical or numerical measurement of economic relationships. The main objective of econometrics is to deal with empirical measurement of economic theories and their verification.

Ex. 1-1: Suppose that an advertising agency might be interested in modelling the relationship between a firm's sales revenue (Y) and advertising costs (X) to know the impact of advertising costs on sales revenue. The relationship between Y and X can be written as

$$Y = f(X, \beta_0, \beta_1) + \varepsilon \tag{1.1}$$

where $f(X, \beta_0, \beta_1)$ is a deterministic factor and ε is a random error term. The mathematical or empirical measurement of relationship (1.1) is called econometrics.

1.3 Meaning of Theoretical Model

A theoretical model is defined as the functional relationship between economic variables in which one is the dependent variable and the remaining is/are independent variable(s). Theoretical models mainly deal with the development of appropriate methods for mathematical, empirical or numerical measurement of economic relationships.

Ex. 1-2: The relationship between profits (Y) and sales revenue (X) of a firm can be expressed by the following theoretical model:

$$Y = f(X) \quad (1.2)$$

If Y linearly depends on X, then the theoretical model between Y and X is given by

$$Y = \beta_0 + \beta_1 X \quad (1.3)$$

where β_0 and β_1 are the unknown and unsigned parameters connected with X and Y. β_1 is the rate of change of Y with respect to X and β_0 is the value of Y at the origin.

Types of Theoretical Models

There are two different types of theoretical models

(i) The economic model or Deterministic model.

(ii) The statistical model or Econometric model or Probabilistic model.

Economic Model or Deterministic Model

The economic or deterministic model is defined as the mathematical relationship between economic variables in which one is the dependent variable and the remaining is/are independent variable(s) which involves questions concerning the signs and magnitudes of unknown parameters. An economic or deterministic model hypothesised an exact relationship between economic variables. Thus, we can say that, in the economic model, the dependent variable Y is a function of the systematic components or deterministic components. The mathematical form of an economic model is given by

$$Y = f(X_1, X_2, \ldots, X_k, \beta_0, \beta_1, \ldots, \beta_k) \quad (1.4)$$

If the variable Y linearly depends on X_1, X_2,…….and X_k, then the economic model of the above relationship can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (1.5)$$

where Y is the dependent variable and X_1, X_2,…….and X_k, are the independent variables. β_1, β_2, …, and β_k are the unknown and unsigned parameters connected with the independent variables X_1, X_2,…….and X_k, and the dependent variable Y. These unknown and unobservable parameters are directly or indirectly related to elasticities and multipliers that play significant roles in economic decisions and actions. The economic model helps us to identify the relevant economic variables and economic parameters and gives us a basis for making economic conclusions and decisions.

Ex. 1-3: The economic model of the quantity demanded of a product can be written as

$$Q_d = f(P, P_s, P_c, Y, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4) \quad (1.6)$$

where Q_d is the quantity demanded of the product; P is the per unit price for the product; P_s is the per unit price for commodities that are substituted; P_c is the per unit price of items that are complements; and Y is the income level of consumers. β_1, β_2, β_3 and β_4 are the unknown and unsigned parameters connected with the price and income variables P, P_s, P_c, Y and the quantity variable Q_d. These unknown and unobservable parameters are directly or indirectly related to elasticities of the quantity demanded of the product with respect to price and income that will play significant roles in economic decisions and actions about demand function. If Q_d linearly depends on P, P_s, P_c, and Y, relationship (1.6) can be expressed as

$$Q_d = \beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 P_c + \beta_4 Y \quad (1.7)$$

β_1 indicates the rate of change of Q_d with respect to P given that all other independent variables are constant; β_2 indicates the rate of change of Q_d with respect to P_s given that all other independent variables are constant; β_3 indicates the rate of change of Q_d with respect to P_c given that all other independent variables are constant; β_4 indicates the rate of change of Q_d with respect to Y given that all other independent variables are constant; and β_0 is the value of Q_d when all the independent variables are zero.

Statistical or Econometric or Probabilistic Model

The statistical, econometric or probabilistic model is a mathematical relationship of economic variables in which the dependent variable Y can be expressed as a function of the systematic components and the nonsystematic random error term. The mathematical form of the statistical, econometric or probabilistic model is given by

$$Y = f(X_1, X_2, ..., X_k, \beta_0, \beta_1, \beta_2, ..., \beta_k) + \varepsilon \quad (1.8)$$

where $f(X_1, X_2, ..., X_k, \beta_0, \beta_1, ..., \beta_k)$ is the systematic component and ε is the nonsystematic component that we know is present but cannot be observed. This is called the random error term. If the variable Y linearly depends on X_1, X_2,…….and X_k, the econometric model of the above relationship can be expressed as

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_k + \varepsilon_i \quad (1.9)$$

where β_j (j=1, 2,……k) indicates the average impact of per unit change in X_j (j=1, 2,…,k) on Y given that all other independent variables are constant; and β_0 is the regression constant which indicates the average value of y when all X's are zero. A statistical or econometric model specifies a sampling process by which the sample data are generated and identifies the unknown parameters β_0, β_1,......, and β_k that are included in the model that describes the underlying probability distribution.

Ex. 1-4: The econometric model between the household consumption expenditure (C) and the level of income (Y) can be expressed as

$$C = f(Y, \beta_0, \beta_1) + \varepsilon \quad (1.10)$$

If C linearly depends on Y, the relationship can be written as

$$C_i = \beta_0 + \beta_1 Y_i + \varepsilon_i \quad (1.11)$$

where the term C_i indicates the consumption expenditure of the ith household; and Y_i indicates the income level of the ith household; ε_i is the random error term corresponding to the ith set of observations (i=1, 2,..….,N). β_0, and β_1 are the unknown and unsigned parameters connected with the income level (Y) and the consumption expenditure(C). β_1 is called the regression coefficient which indicates the average impact of per unit change in Y on C and β_0 is called the regression constant which indicates the average value of C when Y is zero.

1.4 Ideas behind Econometric Modeling

An economic model is a hypothetical or a theoretical construct which represents quantitative or logical relationships between two or more than two economic variables in which one is the dependent variable and the remaining is/are independent variable(s) describing the economic behaviours of an economic agent or of an economy or a sector of the economy. Economists are frequently using economic models to make several simplified statements or assumptions and then to see how different scenarios might play out. These types of theoretical constructs make statements or hypotheses that are mostly qualitative in nature. For example, microeconomic theory states that an increase in the price of a commodity is expected to increase the quantity supplied of that commodity given that other things are constant. Thus, the theoretical model illustrates a positive linear or non-linear relationship between the price and quantity supplied of the commodity. However, this theoretical construct does not provide any numerical measure of the relationship between the two variables. It also does not tell us how much the quantity supplied of the commodity will increase due to per unit change in the price of the commodity while all other independent variables are constant. However, the econometric model will provide us with numerical estimates. The econometric model gives the numerical, empirical or mathematical measurements of economic relationships. In addition, the main concern of mathematical economics is to express an economic theory in the mathematical/equational form without regard to empirical verification of the economic theory but econometric techniques are the outstanding methods for empirical verification of economic theorems. The economic models are confronted with the observational data. Using econometric techniques, we try to explore whether the theory can adequately explain the actual economic units. The main concern of economic statistics is to collect numerical data, record them, and represent them in tabular forms or graphs. Economists and statisticians attempt to describe their changes over time and also to detect relationships between economic variables. Mainly, economic statistics is a descriptive aspect of economics but it does not provide explanations of the development of economic variables and it does not provide any numerical measurement of economic relationships. However, econometric models will provide the mathematical or empirical measurement of economic relationships. The uses of econometric models for empirical or mathematical measurements or empirical verification of economic relationships have mainly two critics. First, sometimes econometric models may not be correctly specified for the empirical or mathematical measurement of economic relationships. For example, the use of a linear regression model for mathematical measurement of the relationship between marginal costs and the output level of an industrial sector is not correct. Another criticism is that an econometric model may not be realistically compatible with the type of data to be used. An econometric model based on the data of the individual economic agent or commodity may not be suitable for use with data aggregated over individuals or commodities. Problems may arise if the data are aggregated over a period of time longer than to which the economic model is applied. Despite these drawbacks, econometric models play an important role as the guide for economic policies.

Economic Theory, Mathematical Economics and Econometrics

An economic theory deals with the laws and principles which govern the functioning of an economy and its various parts. The main concern of an economic theory is to express statements or hypotheses about relationships of economic variables that are mostly qualitative in nature. The economic theory does not allow to inclusion of any random factor that might affect the relationship of economic variables, and thus, prevent the inclusion of a stochastic variable. In addition, the economic theory does not provide any mathematical or numerical measurement of economic relationships. For example, in the microeconomic theory, the law of demand states that the quantity demanded of a commodity and the price of that commodity are inversely related keeping that all other factors are constant. Thus, the economic theory implies a negative or inverse relationship between the price and the quantity demanded of the commodity but this theory does not tell us how a per unit change in the price of the commodity affects the quantity demanded of that product. The main concern of mathematical economics is to express the relationship of economic variables in mathematical form or an equational form without the mathematical measurement or empirical verification of the theory. The use of mathematical formulas does not allow the economist or researchers to include any random factor which might affect the relationship but allows economists or researchers to make assumptions regarding economic theories. It can be said that mathematical economics deals with the economic behaviours of deterministic models. Thus, there are no significant differences between economic theory and mathematical economics, as both of them state the same economic relationship in which the former is an expression in the linguistic form. However, mathematical economics uses mathematical or equational forms. Both express different economic relationships in an exact form but not in a probabilistic model. Neither economic theory nor mathematical economics provides the numerical measurement of economic relationships. However, econometrics differs significantly from economic theory and mathematical economics. The main concern of econometrics is to express the economic relationships in a mathematical form, but it does not assume the relationships to be exact as the model used is probabilistic in nature. Thus, econometric techniques are developed to take into account the effect of random disturbances which make differences from the exact behavioural patterns suggested by economic theory and mathematical economics. Furthermore, econometric methods provide us with the mathematical/empirical or numerical measurement of economic relationships. For example, if we apply the econometric methods to estimate the demand function of the type: $q_t = \alpha + \beta p_t + \varepsilon_t$, we find the numerical values of the coefficients of α and β respectively. Thus, it can be said that econometrics provides us with precise estimates of the coefficients or elasticities of economic relationships which play an important role for the policy-makers in policy reforms.

Division of Econometrics

Econometrics may be divided into two categories:

(i) Theoretical econometrics

(ii) Applied econometrics

Theoretical Econometrics

Theoretical econometrics is the branch of econometrics which deals with the development of appropriate methods for the mathematical or empirical or numerical measurement of economic relationships. For mathematical, empirical or numerical measurement of economic relationships, different techniques such as the OLS method, ML method, WLS, GLS, GMM, Unit Root Tests, Cointegration Analysis, Granger Causality Tests, etc. are developed.

Applied Econometrics

Applied econometrics is the branch of econometrics which deals with the application of econometric methods for the mathematical or empirical or numerical measurement of economic relationships in which one is the dependent variable and the remaining is/are independent variable(s) plus a random error term to specific branches of economic theory.

Ex. 1-5: Assume that the relationship between return on assets (RET) and the risk (BETA), is linear, so the econometric model is given by

$$RET_i = \beta_0 + \beta_1 BETA_i + \varepsilon_i \qquad (1.12)$$

where RET_i is the return of the ith company, $BETA_i$ is the beta coefficient from CAPM of the stock's price index of the ith company; and ε_i is the random error term corresponding to the ith set of observations (i =1, 2,…..,N). β_0 is the regression constant which indicates the average value of RET when BETA is zero and β_1 is the regression coefficient which indicates the average rate of change of RET with respect to BETA. If we apply the OLS method for the empirical measurement of the relationship in (1.12) based on the observed data of RET and BETA, it is called applied econometrics.

Scope and Limitation of Econometrics

Econometrics is the branch of economics which deals with the application of statistical methods and mathematical statistics to economic data for the mathematical measurement of economic relationships. It highly focuses on giving experimental content for finding out economic relationships. The main objective of econometrics is to compute the relationships between economic variables through statistical techniques. Therefore, the study of econometrics is very important for solving economic problems analytically and thereby providing effective and efficient solutions. Thus, econometrics is immensely important in the study of complex mathematical and statistical models which help in a detailed study of economic relationships for a specific branch of economics. In this section, the scopes of econometrics are discussed below:

Scope of Econometrics

(i) Econometric tools and techniques are very important for a solid understanding and dealing with business, economic, socio-economic, financial and scientific problems.

Ex. 1-6: If we are interested in examining empirically the risk-return relationship of Dhaka Stock Exchange (DSE), we can deal with this problem using an econometric method called GARCH-in-mean (GARCH(p, q) -M) model.

(ii) Econometric methods are widely applicable for appropriate decision-making and policy formulation by government agencies, businessmen, economists, statisticians, researchers and other policymakers irrespective of any discipline.

Ex. 1-7: Suppose the Bangladesh Govt. wants to devaluate its currency to correct the balance of payments. To devaluate BDT currency, the Govt. have to estimate the price elasticities of imports and exports demand functions of commodities. If exports and imports are of inelastic nature then devaluation will only ruin its economy. On the other hand, if the price elasticities are of elastic nature, then the devaluation plays a positive role in the economy. These price elasticities are to be estimated with the help of the demand functions of imports and exports commodities. Econometric tools and techniques may be applied to calculate the price elasticities of the demand for exports and imports of commodities. Then based on the calculated values, policymakers will make a decision, on whether the BDT currency will be devaluated or not.

(iii) Econometric techniques are the outstanding methods for empirical verification of economic theorems. The mathematical formulations in economic theory are called models which are confronted with observational data. Based on the econometric techniques, we try to explore whether the theory can adequately explain the actual economic units or not.

Ex. 1-8: The Keynesian consumption function is given by

$$C = f(Y) \quad (1.13)$$

where C is the aggregate consumption expenditure, and Y is the level of income.

If the variable C linearly depends on Y, then the Keynesian consumption function can be expressed by the following econometric model

$$C_t = \beta_0 + \beta_1 Y_t + \varepsilon_t \quad (1.14)$$

where β_0 is the average aggregate consumption expenditure at the zero income level, β_1 is the average rate of change of C with respect to X which is called the marginal propensity to consume (MPC) lying between 0 and 1. Using econometric technique(s), we can estimate β_0 and β_1, and verify how closely such estimates confirm to economic theory. These estimated MPC are used to calculate the size of the multiplier and they can be used for predicting changes in income due to changes in investment. Here, $\frac{dC}{dY} = \beta_1$ is the MPC and $\frac{1}{1-MPC}$ is the multiplier effect. If the estimated value of β_1 lies between 0 and 1, it can be said that the above relationship (1.14) satisfies the Keynes' law. Thus, econometrics helps us to investigate whether theory appears to be consistent with observed data, whether the functional relationship is stable over time, and whether changes over periods.

(iv) To establish new relationships between economic variables and to prove old theorems, econometricians frequently use econometric tools and techniques.

(v) Econometric methods are the most popular and widely applicable for forecasting the demand functions of commodities.

Ex. 1-9: Let there exist a linear demand function between the quantity demanded (Y) of a product and the time trend (t) of the type:

$$Y_t = \alpha+\beta t+\varepsilon_t \quad (1.15)$$

where Y_t represents the quantity demanded of a product at time t, ε_t is the random error term corresponding to the tth set of observations, α is constant which indicates the average value of Y at the point of origin, and β is the regression coefficient which indicates the average change of Y for 1 unit change in time t. Let $\hat{\alpha}$ and $\hat{\beta}$ be the least squares estimates of α and β. They can be expressed by $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{t}$, and $\hat{\beta} = \frac{\sum_{t=1}^{T} Y_t t - n\bar{t}\bar{y}}{\sum_{t=1}^{T} t^2 - n\bar{t}^2}$. Then we can forecast the demand of the product for time period (t+1) as given by $\hat{Y}_{t+1} = \hat{\alpha}+\hat{\beta}(t+1)$, and so on.

(vi) Econometric methods are widely applicable for forecasting business cycles.

(vii) Econometric methods are widely applicable for market research purposes, such as the analysis of a profit function, production function, cost function, supply function, distribution of wealth, etc.

(viii) Econometric techniques are widely applicable for solving different kinds of socio-economic problems such as poverty, income inequality, crimes, divorce etc.

(ix) Now-a-days, the study of econometrics is very important for hypothesis testing, estimation of the parameters, identifying proper functional forms of economic variables, and for measuring the impact of independent variable(s) on a dependent variable.

Limitation of Econometrics

Econometrics has its own limitations. Econometrics cannot be applied to answer all kinds of queries. Some of its limitations are given below:

(i) Econometric methods are widely applicable for quantitative analyses, but for qualitative analyses, econometric methods are not directly applicable or has limited application.

Ex. 1-10: If we are interested in knowing the impact of democracy or political crises on economic development, econometric methods fail to answer these questions directly, as such problems cannot be transformed into a mathematical model. However, econometrics can answer indirectly by using the dummy variable(s) in the regression equation.

(ii) The econometric techniques are based upon certain statistical assumptions which are not always true with economic data.

(iii) Econometric methods do not give priority to moral judgment, but for policy formulations and decision-making, moral judgments play an important role.

(iv) Econometrics methods are time consuming, very complex and not easy to understand.

(v) Econometric techniques should be applied by econometricians or skilled persons who know econometrics to deal with the problems irrespective of any discipline but these persons are not available in our country.

Steps Involved in an Econometric Analysis of Economic Models

The following steps are involved in an econometric analysis of an economic model:

Step 1: First, we have to formulate an economic model that is appropriate to give the answers to the questions. For example, the economic model for the demand analysis of a commodity can be written as

$$Q_d = f(P, P_s, P_c, Y, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4) \quad (1.16)$$

where all the terms of equation (1.16) have already been defined previously.

Step 2: Second, we have to formulate an econometric model from an economic model, which is appropriate for the empirical measurement of underlying assumptions associated with the problem under consideration. For example, the econometric model of the economic model given in equation (1.16) can be written as

$$Q_d = f(P, P_s, P_c, Y, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4) + \varepsilon \tag{1.17}$$

If the quantity demanded Q_d of the commodity linearly depends on P, P_s, P_c and Y, the econometric model can be written as

$$Q_{di} = \beta_0 + \beta_1 P_i + \beta_2 P_{si} + \beta_3 P_{ci} + \beta_4 Y + \varepsilon_i \tag{1.18}$$

Step 3: Third, we have to collect an appropriate data set that is properly defined and matches the concepts of the econometric model for empirical measurement. For example, for the empirical measurement of the economic relationship (1.17), we have to collect the data of the variables P, P_s, $P_{c,}$ Y and Q_d respectively.

Step 4: Fourth, we have to use a suitable software package such as Microsoft Excel, GAUSS, LIMDEP, RATS, SPSS, EViews, STATA, R, Python, TSP, etc. to estimate and test the model to be carried out with the observed data.

Step 5: In the final stage, we use the estimated model for prediction and policy purposes. For example, using the software RATS with the help of the observed data of the variables P, P_s, $P_{c,}$ Y and Q_d we can estimate the values of $\beta_0, \beta_1, \beta_2, \beta_2$, and β_4 respectively, we can predict the value of Q_d , and test whether the relationship is significant or not. Based on these estimated values, a decision can be made about the functional form and also a new policy can be formulated about the quantity demanded of that particular commodity.

Different steps that are involved in an econometric analysis of economic models are shown below graphically:

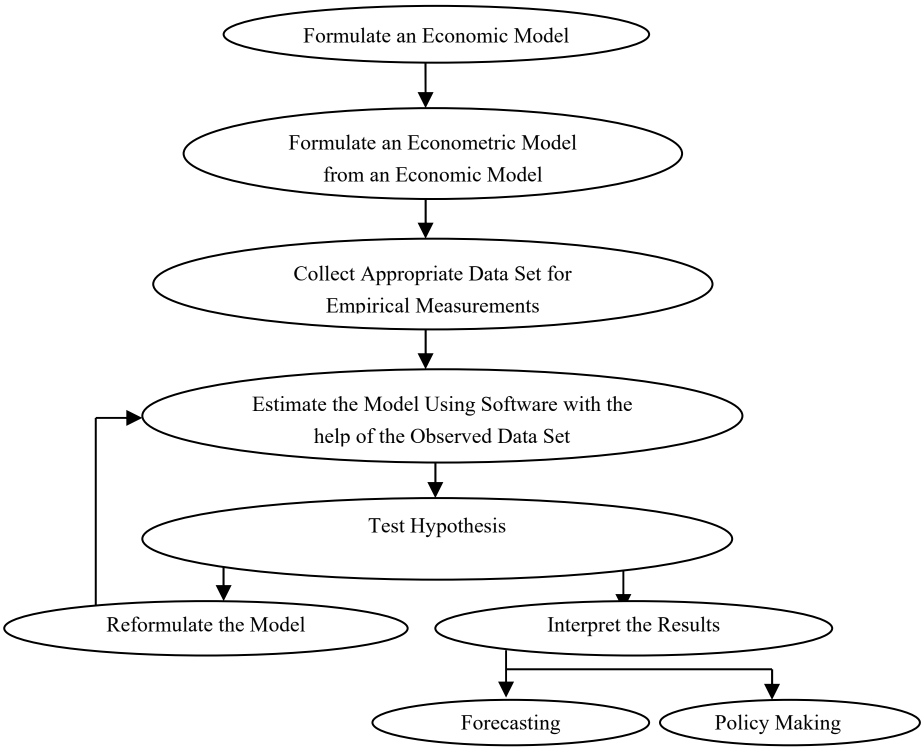

Fig. 1-1: Flowchart for the steps involved in an econometric analysis of economic models

1.5 Types of Relationships

A relationship refers to the functional form between two or more economic variables in which one is the dependent variable and the remaining is/are independent variable(s). For the time being, based on the functional form, the values of the dependent variable are determined for the given values of the independent variable(s) or explanatory variable(s). There are many different types of relationships of economic variables that we have to deal with. This section focuses on different types of relationships with corresponding examples.

Behavioral Relationship

The behavioral relationships tell us how the dependent variable will change its behavior on an average to respond to changes in the independent variable(s). In a behavioral relationship, the dependent variable is one that takes values depending on the values of independent variables.

Ex. 1-11: The relationship between the quantity demanded of a product and its price of the type

$$q = \alpha + \beta p \tag{1.19}$$

is called a behavioural relationship. Here, q is the quantity demanded, and p is the price per unit of the commodity. α and β are two parameters. This relationship describes that every additional unit of price (p) of the commodity generates β amount of change in quantity the demanded (q) of the commodity, and at the origin, the value of q will be α. Thus, it can be said that, in a behavioral relationship, the dependent variable will be detected based on the given values of the independent variable(s).

Technical Relationship

The functional relationship between a firm's physical production output and the factors associated with production inputs is called the technical relationship. Thus, the production function is purely a technical relationship which connects factor inputs and output. The production function of a firm depends on the technological development of a country. With every stage of technological development of a country, the production function of the firm undergoes the changes.

Ex. 1-12: The Cobb-Douglas production function of the type

$$P = A_0 K^{\alpha} L^{\beta} \tag{1.20}$$

is called a technical relationship. Here, P is the output, A_0 is a technical constant, K is the input capital, L is the input labour, α is the output elasticity with respect to capital, and β is the output elasticity with respect to labour. The Cobb-Douglas production function reflects how much output we can expect if we have a combination of labour and capital, and for the known values of A_0, α, and β.

Definitional Relationship

A relationship is said to be a definitional relationship from which we can obtain a meaningful definition by relating the variables.

Ex. 1-13: The productivity of a firm may be measured either on an aggregate basis or on an individual basis, which are called total and partial measures respectively. The total productivity measure is obtained from the following definitional relationship

Total productivity index/measure = Total output/ Total input

The value of a product can be obtained by multiplying the selling price of the product and the quantity sold. Thus, the relationship of the type $v = p \times q$, is called a definitional relationship.

Stochastic Relationship

A functional relationship of economic variables is said to be a stochastic relationship if the dependent variable Y can be expressed as a function of the systematic components and the nonsystematic random error term. The mathematical form of the stochastic relationship is given by

$$Y = f(X_1, X_2, \ldots, X_k, \beta_0, \beta_1, \ldots, \beta_k) + \varepsilon \tag{1.21}$$

where $f(X_1, X_2, \ldots, X_k, \beta_0, \beta_1, \ldots, \beta_k)$ is the systematic component and ε is the nonsystematic component that we know is present but cannot be observed. This ε is called a random error term. If the variable Y linearly depends on $X_1, X_2, \ldots\ldots$, and X_k then the stochastic relationship can be expressed as

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon \tag{1.22}$$

All the terms of equation (1.22), have already been defined.

Ex. 1-14: The linear function of the returns (RET) of firms with some firm-specific factors of the type

$$RET_i = \beta_0 + \beta_1 S_i + \beta_2 PE_i + \beta_3 BETA_i + \varepsilon_i \tag{1.23}$$

is called a stochastic relationship. Here, RET_i is the annual return in the percentage of the ith firm, S_i is the size of the ith firm which is measured in terms of sales revenue, PE_i is the price to earnings (P/E) ratio of the ith firm, $BETA_i$ is the ith firm's CAPM beta coefficient, and ε_i is the random error term corresponding to the ith set of observation which is assumed to be independently, identically and normally distributed with zero mean and constant variance for all i, i.e., $\varepsilon_i \sim IIND(0, \sigma^2), \forall$ i. $\beta_0, \beta_1, \beta_2$, and β_3 are the parameters. This stochastic relationship describes that every additional unit of the jth independent variable (j =1, 2, 3) generates β_j amount of average change in returns (RET) and β_0 indicates the average value of RET when the values of S, PE and BETA are zero. In this relationship, RET is determined by S, PE and BETA.

Static Relationship

A functional relationship of economic variables is said to be a static relationship if the dependent variable Y can be expressed as a function only of the deterministic components or systematic components in one specified period with a constant increment. Thus, a static relationship is a non-probabilistic model. The mathematical form of the static relationship is given by

$$Y = f(X_1, X_2, ..., X_k, \beta_0, \beta_1, ..., \beta_k) \quad (1.24)$$

where $f(X_1, X_2, ..., X_k, \beta_0, \beta_1, ..., \beta_k)$ is the deterministic component. The static relationship does not allow to include the random factor in the equation.

Ex. 1-15: The linear consumption function of the type

$$C = \alpha + \beta Y \quad (1.25)$$

is called a static relationship. Here, C is the per capita consumption expenditure, Y is per capita income, α, and β are two parameters. The parameter β indicates the rate of change of C with respect to Y and α is the value of C at the origin. In this relationship, C is determined only by Y.

Dynamic Relationship

A functional relationship is said to be a dynamic relationship if the dependent variable Y at time t can be expressed as a function of its past values including current and lagged values of another independent variable X. If the dependent variable Y linearly depends on X, then the dynamic relationship is given by

$$Y_t = \beta_0 + \beta_1 X_t + \beta_1 X_{t-1} + \alpha_1 Y_{t-1} \quad (1.26)$$

Both behavioural and technical relationships can be expressed as static and dynamic.

Ex. 1-16: The linear consumption function of the type:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \alpha_1 C_{t-1} \quad (1.27)$$

is called a dynamic relationship. Here, C_t = Per capita consumption expenditure at time t, C_{t-1} = Per capita consumption expenditure at time (t-1), Y_t = Per capita income at time t, Y_{t-1} = Per capita income at time (t-1), $\beta_0, \beta_1, \beta_2$, and α_1 are the unknown and unsigned parameters which are connected with the variables Y and C.

Micro Relationship

A functional relationship between the dependent variable (Y) and an independent variable (X) is said to be a micro relationship if the functional relationship holds to study the economic behaviours of a single firm, single household, for a consumer and so on. The demand function, supply function, cost function, profit function, etc. for a single firm are the micro relationships.

Ex. 1-17: If we consider a production function to study the behaviours of a single firm of the type

$$q = CL^{\alpha}K^{\beta} \quad (1.28)$$

it is called a micro relationship. Here, q is the output level of an output variable Q of a firm, L is the quantity of the input variable labour of the firm, and K is the capital investment of the firm. Here, α is the output elasticity with respect to labour and β is the output elasticity with respect to capital of the firm. If we consider a consumption function to study the behaviours of a single consumer it is called a micro relationship.

Macro Relationship

A functional relationship between the dependent variable (Y) and an independent variable (X) is said to be a macro relationship, if the function holds to study the behaviours of a nation or aggregate of consumers or the aggregate of firms. The national consumption function, national savings function, national investment function, and the aggregate demand function are the macro relationships.

Ex. 1-18: The functional relationship between savings and gross national income of the type

$$S = \alpha+\beta Y \qquad (1.29)$$

is called a macro relationship. This is called a national saving function. Here, the dependent variable S is savings and Y is the gross national income. α and β are two parameters in which α is the constant which indicates the amount of savings at the origin, which will be negative, if the gross national income becomes zero at no savings. However, one has to borrow money from someone to survive. Thus, α is always negative. The parameter β is the rate of change of savings with respect to Y. β is less than or equal to one because savings cannot be greater than income. Thus, the ratio of the change in savings to the change in income is less than 1. It will be equal to 1 only when the consumption expenditure is zero. The scenario can be expressed by the following equation

$$\beta = \frac{\Delta S}{\Delta Y} = \frac{\Delta(Y-C)}{\Delta Y} = \frac{\Delta Y - \Delta C}{\Delta Y} = 1 - \frac{\Delta C}{\Delta Y} = 1 - b$$

where b is the marginal propensity to consume (MPC)

Thus, the saving function can be written as

$$S = \alpha+(1\text{-}b)Y \qquad (1.30)$$

Here, (1-b) is called the marginal propensity to savings (MPS). Based on the functional relationship, we must determine the impact of Y on S but the influence of other factors is assumed to be constant.

1.6 Random Error Term

In an economic model, the dependent variable can be expressed as a function of the deterministic or systematic components, which describes the exact relationship between economic variables. However, in practice, to deal with any problem irrespective of any discipline, we feel that all the observations do not fall exactly on a straight line or any other smoothing curve. We can expect that the observed values will be closer to the straight line, or a curve. For this reason, in regression models we introduce a stochastic disturbances term to measure the deviation of the actual value from the central value of the dependent variable. The introduced stochastic disturbance term is known as a random error term, because this term is introduced in the regression model to find the effects of all the factors which are not included in the equation by the investigators or researchers. These types of errors are errors of omitted variables, errors of misspecification of the mathematical form of the model, errors of the measurement of the dependent variable, etc.

Ex. 1-19: Suppose we want to study the relationship between profit (Y) and investment (X) of a number of listed companies at DSE at time t. Let us have n sample observations which are divided into m sub-groups on the basis of investment level. We have X_1, X_2,…., and X_m investment levels whose profits are Y_1, Y_2,…..,and Y_m. We cannot expect that all the listed companies within one sub-group having X_1 level of investment will have an identical profit Y_1. Some companies have profits more than Y_1 while some others will have less than Y_1. But all values will vary around a central figure of Y_1. Hence, the disturbance term is introduced in the regression model which will show the deviations of actual values from the central value. Here, the profit of different listed companies of subgroup one will be $Y_1+\varepsilon_1$, $Y_1+\varepsilon_2$, $Y_1+\varepsilon_3$, $Y_1+\varepsilon_4$, and so on. Error term may have positive or negative values and is drawn at random.

Reasons to Include the Random Error Term in a Model

We know that econometric models are probabilistic models or statistical models. Thus, econometric models give priority to probabilistic judgments and like all statistical judgments, they are subject to errors. These errors may arise for the following reasons:

(i) The error may arise due to the omission of relevant variables in the model. We know that each economic variable is affected by so many economic variables at the same time. If any one of the relevant variables is left out and cannot be included in the model, errors may arise. In such a case, an error is bound to occur in the model.

(ii) Errors may also arise due to the non-availability of data. The econometric study is frequently based upon the assumption that we have large samples of accurate data. But unfortunately, we will not find an example in which reliable and representative data are available. Thus, in the absence of accurate data, the sampling error arises in the model.

(iii) Errors may also arise due to the misspecification of the functional form between economic variables. Sometimes, we assume that the relationship is linear between variables, but it may be non-linear. In such a case, the forecast is bound to be incorrect. This type of error arises mainly due to the misunderstanding of the investigator.

1.7 Data Associated with Econometric Analysis

A variety of economic data sets are applied in an econometric analysis for the empirical measurement of economic models. The most important economic data sets which are widely applicable for econometric analysis of economic relationships are described below:

Cross-Sectional Data

The data which are observed at a particular point in time or at one point in time are called the cross-sectional data. For cross-sectional data the subscript i is used. For example, let $Y_1, Y_2,....,Y_i,....,Y_n$ is a sample of total infected people due to coronavirus of n countries in the year 2020. This is a cross-sectional data in which Y_i indicates the total infected people due to coronavirus of the country i (i =1, 2,……..,n).

Ex. 1-20: The number of deaths due to coronavirus pandemic in the year 2020, the GDP of Asian countries for the year 2023, the number of car accidents recorded in different big cities in the year 2023, etc. will constitute cross-section data.

Time Series Data

The data observed over a period of time are called time-series data. In other words, an arrangement of statistical data in chronological order, i.e., in accordance with the occurrence of time is known as time series data. Time series data may be defined as a set of figures observed over a period of time. Mathematically, a time series is defined as the functional relationship such as

$$Y_t = f(t), \text{ where } t = t_1, t_2,.....,t_n. \tag{1.31}$$

where Y_t is the value of the variable Y at time t.

Ex. 1-21: The GDP, growth rate of GDP, export values, import values etc. of Bangladesh for a specified period of time will constitute time series data.

Panel Data or Longitudinal Data

The combination of cross-sectional and time-series data is called panel data. For panel data, the subscripts i and t are used. The subscript i is used for cross-sectional and t is used for time-series data. For example $\{Y_{it}, i=1, 2,...,n; t=1, 2,...,T\}$, a set of values of the variable Y, will constitute a panel data. Here, Y_{it} is the value of the variable Y corresponding to the ith (i=1, 2,…,n) individual and tth (t=1, 2,…….,T) time.

Ex. 1-22: The GDP of SAARC countries over a period of time; FDI of Asian, European, OECD countries over a period of time; per capita income of the people of different districts of Bangladesh over a period of time; agricultural production of different districts of Bangladesh over a period of time; profits of different firms over a period of time will constitute panel data.

Experimental Data

The data which are collected by experiments, i.e., by natural science are called experimental data. Sometimes, the investigators collect data to find the effect of some factors on a given phenomenon provided that the effects of some other factors are constants, then, these types of data are called experimental data.

Ex. 1-23: Suppose, to find the impact of consumption of ice-creams on weight gain, a researcher collected data keeping that the eating, smoking, and drinking habits of the people are constant will constitute experimental data.

1.8 Variables

In this section, the variables which are the most popular and commonly used in econometric analyses are discussed.

Variable

A variable is defined as a characteristic which can take on different values at different times, places, or situations within its domain. The domain of a variable X is a set of all permissible values that X can take in a given context.

Ex. 1-24: Household income, household expenditure, national income, national expenditure, prices, interest rates, wages, profits, industrial output, agricultural output etc. are all variables, whose values are obtained from published external data sources. For the income variable of a family, if the lowest value is BDT 0 and the highest value is BDT 50,000, the domain of the income variable is 0 to 50,000.

In general, variables can be classified into two broad categories namely: (i) quantitative variables and (ii) qualitative variables

Quantitative Variable

Quantitative variables are those variables whose values can be measured and expressed on a numerical scale.

Ex. 1-25: GDP, GNP, export values, import values of Bangladesh are the quantitative variables.

Types of Quantitative Variable

Quantitative variables can be classified into two broad categories namely: (i) discrete variables and (ii) continuous variables.

Discrete Variable

A quantitative variable X is said to be discrete if it takes only a particular finite or countably infinite value to its domain.

Ex. 1-26: The number of persons in a family, the number of crimes that are happening every day in Dhaka city, the number of deaths due to road accidents, and the number of deaths due to coronavirus every day in Dhaka City are all discrete variables.

Continuous Variable

A quantitative variable X is said to be a continuous variable if it takes value within a specific interval or range. The variable X is said to be continuous if it takes values 1 to 5 in a straight line. That is, $1 \leq X \leq 5$.

Ex. 1-27: Height and weight of a person, household income and expenditure, distance between two places etc. are all continuous variables.

Qualitative Variable

Qualitative variables are those variables which cannot be measured and expressed numerically but can be classified into several groups or categories. The characteristics which are used to classify an object or an individual into different categories are called attributes.

Ex. 1-28: Sex of persons, education level of a university, and outcomes of a coin tossing problem are qualitative variables. The sex of persons has two categories namely: male or female. The education level of a university can be classified into three categories namely: good, medium or bad, the outcomes of a coin tossing problem gives two categories namely: head or tail. These categories are sometimes called attributes.

Categorical Variable

A categorical variable is one that has two or more categories. There are two types of categorical variables: nominal and ordinal. A categorical variable is said to be a nominal variable if it has no intrinsic ordering to its categories but an ordinal variable has a clear ordering to its categories.

Ex. 1-29: Gender is a categorical variable having two categories namely: (i) male and (ii) female with no intrinsic ordering to the categories. The categorical variable temperature is a variable with three orderly categories namely: (i) low, (ii) medium and (iii) high. A frequency table is a way of counting how often each category of the variable in question occurs. It may be enhanced by the addition of percentages that fall into each category.

Dichotomous Variable

The categorical variables with only two categories or levels are called the dichotomous variables.

Ex. 1-30: The outcomes of a coin tossing problem-head or tail, classification of the employees of a public university according to gender: male or female, classification of all the politicians of Bangladesh according to income class: rich or poor, classification of the politicians according to political party namely: Democrat or Republican, classification of the students according to pass or fail. The number of death due to coronavirus are classified according to age: under age 65 or 65 and over are all dichotomous variables

Binary Variable

Binary variables are those variables which take only two values 1 or zero. 1 is for success and 0 for failure.

Ex. 1-31: Let us define a variable Y such that

$$Y = \begin{cases} 1, \text{ if a family owns a new car} \\ 0, \text{ otherwise} \end{cases}$$

then y is called a binary variable.

Dependent and Independent Variables

A variable is said to be a dependent variable if its values depend on changes in other variables in a functional relationship of two or more variables. In general, the dependent variable is denoted by Y. A variable is said to be an independent variable if its values do not depend on changes in other variables in a functional relationship. This is in contrast to the definition of the dependent variable. An independent variable is denoted by X.

Ex. 1-32: If we consider the demand function of the type $q = a+bp$, then the variable quantity demanded q is called the dependent variable and the variable per unit price p of the product is called the independent variable. a and b are two parameters in which b is the rate of change of q with respect to p and a is the value of q at the origin.

Control variable

A variable that is held constant in order to assess or clarify the relationship between two or more variables is called a control variable (sometimes called a controlled variable). A control variable can strongly influence the results in an experiment.

Ex. 1-33: Let us consider an example to understand the meaning ofa control variable. Suppose, we want to perform a regression analysis and to investigate how the amount of sunlight received affects the growth of a plant. Here, the amount of sunlight received is the independent variable and growth of the plant is the dependent variable. As the independent variable changes, we can see the corresponding changes in the dependent variable i.e., the growth of the plant. A control variable is another factor in a regression analysis. In our example of a plant growth, control variables could be water and fertilizers supplied. If control variables are not kept constant, they could ruin the regression analysis. We may conclude that plants grow optimally at 6 hours of light a day. However, if our plant is receiving different levels of fertilizers and water supply, our experiment becomes invalid. Hence, we need to identify the variables that may affect the outcome of our analysis and should take actions to control them.

Intervening Variable/Mediating Variable

An intervening variable is a hypothetical variable used to explain causal links between other variables. Intervening variables cannot be observed in an experiment, that is why they are called the hypothetical variable. In psychology, the intervening variable is sometimes called a mediator variable. In statistics, an intervening variable is usually considered to be a sub-type of a mediating variable. However, the lines between the two terms are somewhat fuzzy, and they are often used interchangeably.

Ex. 1-34: We know there is an association between being poor and having a shorter life span. Just because someone is poor it does not mean that he/she will face an early death. So, other hypothetical variables are used to explain the phenomenon. These intervening variables could include lack of access to healthcare or poor nutrition, environmental pollution etc.

Exercises

1-1: Explain the meaning of econometrics with an example.

1-2: Define a theoretical model with an example.

1-3: Discuss different types of theoretical models with an example of each.

1-4: Explain the ideas behind econometric modeling of economic relationships with an example.

1-5: Distinguish between economic theory, mathematical economics and econometrics with an example of each.

1-6: Distinguish between economic and econometric models with an example of each.

1-7: Distinguish between theoretical and applied econometrics.

1-8: Discuss the advantages of econometrics.

1-9: What are the limitations of econometrics?

1-10: Write different steps that are involved in an econometric analysis of economic models.

1-11. Why do we prefer an econometric model rather than an economic model? Discuss.

1-12: Distinguish between economic theory and econometrics.

1-13. Distinguish between mathematical economics and econometrics.

1-14: Define different types of relationships with an example of each.

1-15: Distinguish between static and dynamic relationships with an example of each.

1-16: Distinguish between static and stochastic relationships with an example of each.

1-17: Distinguish between micro and macro relationships with an example of each.

1-18: Define a random error term with an example. Discuss the reasons to include the random error term in a model.

1-19: Define data with an example. Discuss different types of data with an example of each.

1-20: Distinguish between cross-sectional and time-series data with an example of each.

1-21: Distinguish between time-series and panel data with an example of each.

1-22: Define a variable with an example. Discuss different types of variables with an example of each.

1-23: Distinguish between quantitative and qualitative variables with an example of each.

1-24: Distinguish between discrete and continuous variables with an example of each.

1-25: Distinguish between dependent and independent variables with an example of each.

1-26: Distinguish between control and independent variables with an example of each.

1-27: Distinguish between categorical and dichotomous variables with an example of each.

1-28: Distinguish between categorical and binary variables with an example of each.

1.29: Identify the following functions:

(i): $q_d = \alpha+\beta p$; (ii): $q_s = \alpha+\beta p$; (iii): $GDP_t = GDP_0(1+r)^t$; (iv): $GDP_t = e^{\alpha+\beta t+\varepsilon_t}$; (v): $q_d = A_0 y^{\alpha}$; (vi): $q_s = A_0 y^{\alpha} e^{\varepsilon_t}$;

(vii): $GDP_t = A_0 L_t^{\alpha} K_t^{\beta}$; (ix): Quick Ratio $= \dfrac{\text{Current Assets} - \text{Inventories}}{\text{Current Liabilities}}$; (x): Total Debt Ratio $= \dfrac{\text{Total Debt}}{\text{Total Assets}}$;

(xi): $Y_t = \beta_0 + \beta_1 Y_{t-1}$; and (xii): $C_t = \alpha + \beta Y_t$.

CHAPTER TWO

SIMPLE REGRESSION MODELS

2.1 Introduction

Most of the research problems irrespective of any discipline are associated with the topic of modelling, i.e., they try to describe how the dependent and independent variables are related to each other. For example, businessmen might be interested in modelling the relationship between the amount of capital investment and returns. A producer might be interested in modelling the relationship between the level of output and production costs. An advertising agency might be interested in modelling the relationship between a firm's sales revenue and advertising costs. A government agency might be interested in modelling the relationship between economic growth and investment and also might be interested in relating the growth rate of GDP to the time periods, etc. An investment firm might be interested in modelling the relationship between risk and stock returns and also might be interested in relating the performance of the stock market to the time periods etc. A doctor might be interested in modelling the relationship between the weight of children and breastfeeding. Therefore, to deal with these types of problems, students and researchers should be able to know the simple linear and non-linear regression models. Thus, in this chapter, simple linear and non-linear regression models are discussed and show how to fit them to a set of data points using different estimation methods namely: the method of least squares, maximum likelihood method, and method of moments. Then, we will show how to judge whether a relationship between a dependent variable(y) and independent variable(x) is statistically significant, how to use the model to estimate the expected value of y, and how to forecast a future value of y for a given value of x. In this chapter, proper justifications for the techniques employed in a regression analysis are also provided. For the empirical measurement of the relationships, the software packages RATS, EViews and STATA are used.

2.2 Simple Linear Regression Models and Estimation Methods

In this section, the meaning of simple population regression equation, sample regression equation, and simple linear regression equations and the methods which are the most popular and widely applicable to estimate simple linear regression equations are discussed along with the assumptions associated with the models.

Population Regression Function or Population Regression

The functional relationship between dependent variable (Y) and independent variable (X) of the type

$$Y_i = E(Y|X_i)+\varepsilon_i \tag{2.1}$$

is called a (two-variable) population regression function (PRF) or population regression equation. Here, $E(Y|X_i)$ is called the conditional mean or conditional expectation of Y and read as the expected value of Y given that X takes the specific value X_i which is defined as a function of X_i. Mathematically it can be written as

$$E(Y|X_i) = f(X_i)\ , (i = 1, 2,\ldots\ldots,N) \tag{2.2}$$

Equation (2.2) tells us how the average response of Y varies due to a change in X. ε_i is the random error term. Thus, it can be said that, in a PRF, Y_i is equal to the conditional mean of Y plus the random error term ε_i. If Y linearly depends on X, then the PRF is defined as

$$Y_i = \beta_0+\beta_1X_i+\varepsilon_i \tag{2.3}$$

Here, $E(Y|X_i)$ is defined as

$$E(Y|X_i) = \beta_0+\beta_1X_i \tag{2.4}$$

where β_0 is called the intercept or constant term which indicates the value of $E(Y|X_i)$ at the origin, and β_1 is the called regression coefficient or slope coefficient which gives the marginal effect of X on Y. Equation (2.3) is also known as the linear population regression function (LPRF) or simply the linear population regression. In the literature, sometimes alternatively we call it a linear population regression model (LPRM) or a linear population regression equation (LPRE). In real-life problems, we can't deal with the entire population for examination. That is why we have to estimate/predict/forecast based on the sample information. Simply, regression is defined as the mathematical or empirical measurement of the average relationship of a dependent variable with one or more than one independent variable.

Ex. 2-1: Assume that the consumption expenditure (C) of households depends on income level Y. Then, the population regression equation between C and Y can be written as

$$C = f(Y) + \varepsilon \qquad (2.5)$$

where C is called the dependent variable/explained variable/regressand variable or response variable; Y is called the independent variable/explanatory variable, or regressor variable; and ε is called the random error term. If C linearly depends on Y, the average relationship of C can be expressed as

$$C_i = \beta_0 + \beta_1 Y_i + \varepsilon_i \qquad (2.6)$$

where β_0 is the regression constant, which indicates the average consumption expenditures of the households when the income level Y is zero, β_1 is the regression coefficient, which indicates the average change of consumption expenditures of the household for changing one unit of Y, and ε_i is the ith random error term.

The graph of this equation is given below:

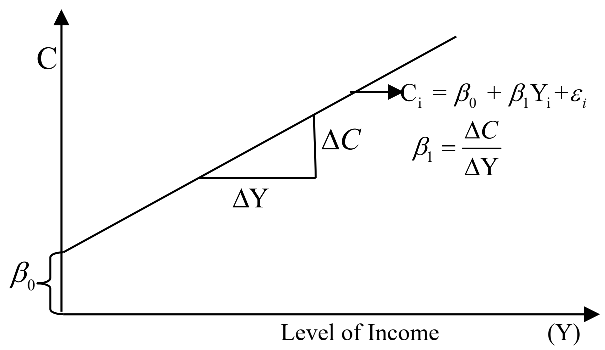

Fig. 2-1: Regression equation between consumption expenditure and level of income

A regression analysis attempts to estimate the nature of the mathematical/empirical relationship between economic variables and thereby provides a mechanism for the prediction and forecasting of the PRFs.

Simple Linear Regression Model

The linear relationship between two variables Y and X of the type

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (2.7)$$

is called a simple linear regression model. Here, Y is called the dependent variable/explained variable/regressand variable/response variable, X is called the independent variable or explanatory variable or regressor variable, ε is called the disturbance term or random error term, β_0 is called the regression constant which indicates the average value of Y when X is zero, and β_1 is called the regression coefficient which indicates the average impact of per unit change in X on Y. The variable Y is called the dependent variable because the value of Y will change due to the change in the value of X and the disturbance term ε. For time-series data Y_i and X_i represent the value of Y and X at the ith period while in cross-section data, it represents observation of the ith individual or object. Thus, the simple linear regression model indicates the average relationship of the dependent variable Y with one independent variable X plus the random error term ε. The stochastic nature of the regression model implies that, for each value of X, there is a sampling distribution of the values of Y. The graphical presentation of a simple linear regression model between two variables Y and X is given below:

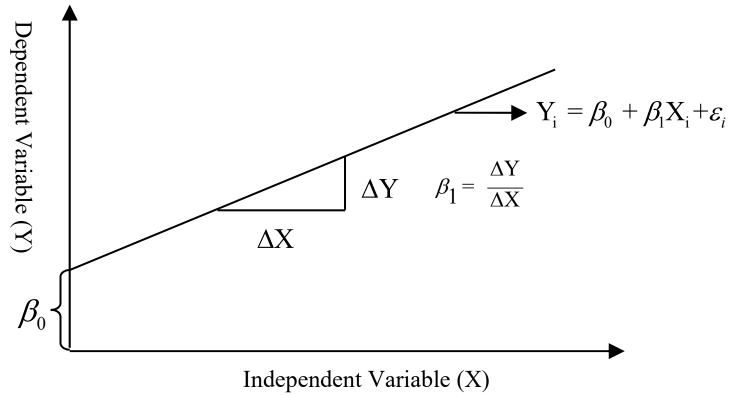

Fig. 2-2: Simple linear regression model between Y and X.

Ex. 2-2: The simple linear regression model between profit and investment of several industries is given by

$$prof_i = \beta_0 + \beta_1 inv_i + \varepsilon_i ; \ (i = 1, 2, \ldots\ldots\ldots, N) \quad (2.8)$$

where inv_i is the ith value of the independent variable investment, $prof_i$ is the corresponding ith value of the dependent variable profit, ε_i is the ith value of the random error term ε, β_0 is called the regression constant which indicates the average value of profit when investment level is zero, and β_1 is called the regression coefficient which indicates the average impact of per unit change in investment on profit.

Sample Regression Function (SRF)

In a population regression function (PRF), we do not know the values of β_0, β_1 and ε_i but we can have some estimates based on sample information. So, the sample counterpart of equation (2.4) can be written as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \text{ (Deterministic form)} \quad (2.9)$$

where $\hat{Y}_i$ is read as Y_i-hat or Y_i-cap, $\hat{Y}_i$ is the estimator of $E(Y_i|X_i)$, $\hat{\beta}_0$ is the estimator of β_0, and $\hat{\beta}_1$ is the estimator of β_1.

We want to estimate the population regression function (2.3) based on the following regression function

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i \quad (2.10)$$

Thus, $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i$ (stochastic form) is called a sample regression function (SRF). When we are dealing with the sample data, the individual values of Y are defined as

$$Y_i = \hat{Y}_i + e_i \quad (2.11)$$

where e_i is the estimate of ε_i which is called the residual. Thus, it can be said that the individual value of Y is equal to the sum of the estimate of the conditional mean of Y and the residual (e_i). Equation (2.11) can be written as

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \quad (2.12)$$

The SRF along with PRF are shown below graphically:

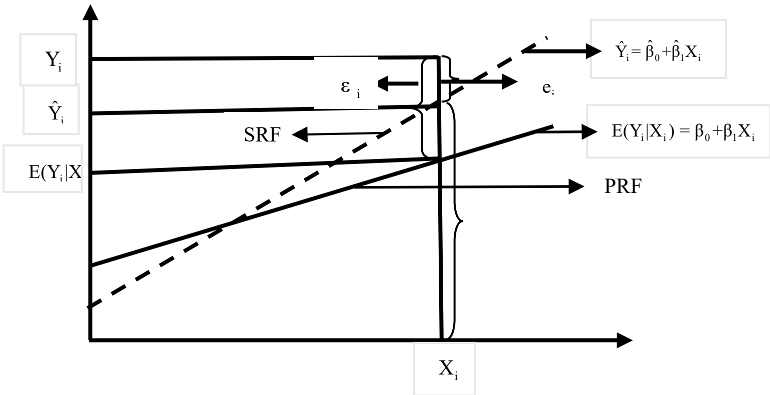

Fig. 2-3: Sample and population regression equations

Ex. 2-3: Economic theory tells us that the quantity demanded (q) of a commodity will be a function of its price (p), i.e., q = f(p). However, the economic theory does not tell us whether the function will be linear or non-linear. If we assume a linear relationship between q and p, we can write

$$q_i = \beta_0 + \beta_1 p_i \quad (2.13)$$

β_1 is always expected to be negative. From the samples, the price elasticity of demand is given by

$$e_p = \hat{\beta}_1 \times \frac{\overline{p}}{\overline{q}} \quad (2.14)$$

where $\hat{\beta}_1$ is the regression estimate of β_1. The relationship between q and p that is produced here will not be exact because q is influenced by many other factors like consumers' income, price of other related commodities, time period, age of consumers, etc. To count their effects, we use an unknown disturbances term ε_i. Let e_i be the residual. Then the SRF is given by

$$q_i = \hat{\beta}_0 + \hat{\beta}_1 p_i + e_i \quad (2.15)$$

where $\hat{\beta}_0$ is the estimate of β_0 and $\hat{\beta}_1$ is the regression estimate of β_1. Equation (2.15) is called a sample regression function (SRF). The predicted line of equation (2.15) is shown for the given data in Table 2-1 .

Table 2-1: Random sample from the population of per unit price and consumption of a commodity

Price Per Unit (p, in TK)	Consumption (q, in kg)
30	112
48	55
25	120
55	40
40	75
20	125
35	98
50	45
36	90
45	65

The actual values with the predicted values are shown below graphically:

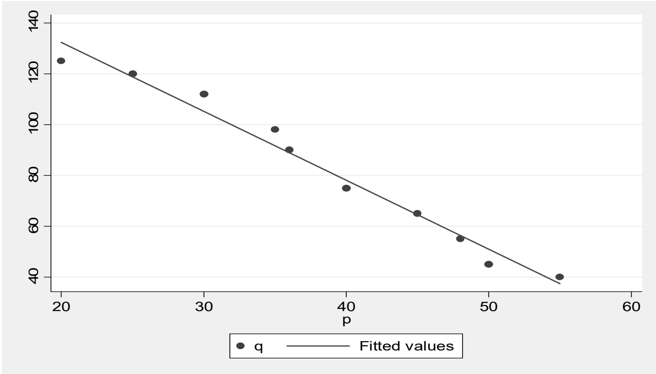

Fig. 2-4: Actual versus predicted values

Statistical Assumptions of Simple Linear Regression Models

Let us consider the simple linear regression model between Y and X of the type

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; \ (i = 1, 2, \ldots\ldots, N) \quad (2.16)$$

The application of econometric techniques for the mathematical or empirical measurement of the simple linear regression model is based on some basic assumptions. These assumptions are given below:

Normality: The first assumption is that the disturbance term ε_i is normally distributed.

Zero Mean: The mean value of the error term ε_i is zero, i.e., $E(\varepsilon_i) = 0, \ \forall \ i$ (i=1, 2,….,N). This means that, for each value of X, some values of ε_i will be positive and some will be negative, but their summation will be zero.

Homoscedasticity: Every disturbance has the same variance around its zero mean whose value is unknown, i.e., $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2, \ \forall \ i.$ It means that, for all values of X's, ε_i shows the same dispersion around the zero mean. This implies that the variance of the disturbance term does not depend on the value of X. For small as well as for large values of X, the sampling distribution of ε_i for all i will be identical. Thus, we can say $\varepsilon_i \sim N(0, \sigma^2)$ for all i.

Non-autoregression: This assumption requires that various disturbance terms are not related to each other i.e. $Cov(\varepsilon_i, \varepsilon_j) = 0,$ for $i \neq j,$ and $i, j = 1, 2, \ldots\ldots\ldots, N$. The covariance of ε_i and ε_j is equal to zero means the value of the disturbance term in one period does not depend on the value of that in the previous period.

Non-stochastic: The independent or regressor variable X is a non-random variable whose values are fixed in repeated samples.

Stochastic: The dependent or response variable Y is a random variable or stochastic variable. So, it has a sampling distribution. Y_i is also normally distributed with mean $\beta_0 + \beta_1 Y_i$ and constant variance σ^2. Thus, we can write $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$.

Independence: X_i and ε_j are independent for all i and j. This assumption automatically follows that, if X is considered as a non-random variable, then $E(X_i, \varepsilon_j) = 0, \ \forall$ i and j.

Methods for Estimating Simple Linear Regression Models or Parameters

The following three methods are the most popular and widely applicable for estimating simple linear regression models or parameters of simple linear regression models:

(1) Method of moments

(2) Ordinary least squares (OLS) method

(3) Maximum likelihood (ML) method.

Method of Moments

The simple linear regression model between two variables Y and X is given by

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i;\ (i = 1, 2, \ldots, N) \tag{2.17}$$

where Y is called the dependent variable/explained variable/regressand variable/response variable, X is called the independent variable/explanatory variable/regressor variable, ε is called the disturbance term or random error term, β_0 is called the regression constant which indicates the average value of Y when X is zero, β_1 is the regression coefficient which indicates the average change in Y for per unit change in X, Y_i is the ith observation of the dependent variable Y, X_i is the corresponding ith observation of the independent variable X, and ε_i is the random error term corresponding to the ith set of observations. Since the value of Y_i depends on ε_i, which is unknown to us, we have to take the expected value of Y_i for a given value of X_i. The expected value of Y_i for a given value of X_i is given by

$$E(Y_i|X_i = x_i) = \beta_0 + \beta_1 x_i + E(\varepsilon_i)$$

$$y_i = \beta_0 + \beta_1 x_i + E(\varepsilon_i) \tag{2.18}$$

Assumptions:

$E(\varepsilon_i) = 0,\ \forall$ i, $Var(\varepsilon_i) = \sigma^2,\ \forall$ i, $Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$, $\varepsilon_i \sim NIID(0, \sigma^2)$, and $E(X_i, \varepsilon_j) = 0,\ \forall$ i and j. The regressor variable X is a non-random variable whose values are fixed in repeated samples. Thus, $Cov(X_i, \varepsilon_j) = 0,\ \forall$ i and j. .

Let y_i be the observed value of the dependent variable y corresponding to the ith set of observations in the sample, x_i be the corresponding ith value of the independent variable X, and e_i be the estimated value of ε_i (residual). Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the estimators of β_0 and β_1 respectively. Thus, for sample observations, the regression equation is given by

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i;\ \ (i = 1, 2, \ldots, n)$$

$$e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \tag{2.19}$$

For the sample observations, we can write

$$\sum_{i=1}^{n} e_i = 0 \Rightarrow \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2.20}$$

$$\sum_{i=1}^{n} e_i x_i = 0 \Rightarrow \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \tag{2.21}$$

From equations (2.20), it can be written as

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i \tag{2.22}$$

and from equation (2.21), we have

$$\sum_{i=1}^{n} y_i x_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 \tag{2.23}$$

From equation (2.23), we have

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 \tag{2.24}$$

Putting the value of $\hat{\beta}_0$ in equation (2.23), we have

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{2.25}$$