CHAPTER ONE: INTRODUCTION

1.1 What is econometrics?

Literally interpreted, *econometrics* means “economic measurement.” Although measurement is an important part of econometrics, the scope of econometrics is much broader. It is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy. The most common application of econometrics is the forecasting of such important macroeconomic variables as interest rates, inflation rates, and gross domestic product. While forecasts of economic indicators are highly visible and are often widely published, econometric methods can also be used in economic areas that have nothing to do with macroeconomic forecasting. For example, it is possible to study the effects of political campaign expenditures on voting outcomes.

Econometrics is a combination of *economic theory*, *mathematical economics* and *statistics*, but it is completely distinct from each one of these three branches of science. *Economic theory* makes statements or hypotheses that are mostly qualitative in nature. For instance, microeconomic theory states that, other things remaining the same, a reduction in the price of a commodity is expected to increase the quantity demanded of that commodity. But the theory itself does not provide any numerical measure of the relationship between the two, that is, it does not tell by how much the quantity will go up or down as a result of a certain change in the price of the commodity. It is the job of econometrician to provide such numerical statements.

The main concern of *mathematical economics* is to express economic theory in *mathematical form* (equations) without regard to measurability or empirical verification of the theory. Both economic theory and mathematical economics state the same relationships. Economic theory uses verbal exposition but mathematical economics employs mathematical symbolism. Neither of them allows for random elements which might affect the relationship and make it stochastic. Furthermore, they do not provide numerical values for the coefficients of the relationships. Although econometrics presupposes the expression of economic relationships in mathematical form, like mathematical economics it does not assume that economic relationships are exact (deterministic).

Economic statistics is mainly concerned with collecting, processing, and presenting economic data in the form of charts and tables. It is mainly a descriptive aspect of economics. It does not provide explanations of the development of the various variables and it does not provide measurement of the parameters of economic relationships.

In a nutshell, *econometrics is the application of statistical and mathematical techniques to the analysis of economic data with a purpose of verifying or refuting economic theories*. In this respect, econometrics is distinguished from mathematical economics, which consists of the application of mathematics only, and the theories derived need not necessarily have an empirical content.

1.2 Models, Economic models and Econometric models

A model is a simplified representation of a real world process. For instance, saying the quantity demanded of oranges depends on the prices of oranges is a simplified representation because there are a lot of other variables that may determine the demand for oranges (including income of consumers, change in the price of its substitutes/complements, etc.). In fact, there is no end to this stream of other variables.

In practice, we include in our model all the variables we think are relevant for our purpose and dump the rest of the variables in a basket called ‘disturbance.’ This brings us to the distinction between an economic model and an econometric model.

An economic model is a set of assumptions that approximately describes the behavior of an economy (or a sector of an economy). On the other hand, an *econometric model* consists of:

1. A set of behavioral equation derived from the economic model, these equations involve some observed variables and some “disturbances”.
2. A statement of whether there are errors of observation in the observed variables.

3. A specification of the probability distribution of the "disturbances" (and errors of measurement).

For instance, if we consider the simplest demand model where quantity demanded (q) depends on price (p), then the econometric model usually contains:

i) **Behavioral equation**

   $q = \beta_0 + \beta_1 p + u$ where u is a disturbance term.

ii) **Specification of the probability distribution of u**

   $E(u/p)=0$ and that the values of u for the different observations are independently and normally distributed with mean zero and variance σ^2. Based on these specifications, one proceeds to test empirically the law of demand or the hypothesis that $\beta_1 < 0$.

1.3 Goals of Econometrics

The three main goals of econometrics include:

A. Analysis: - Testing Economic Theory

Economists formulated the basic principles of the functioning of the economic system using verbal exposition and applying a deductive procedure. Economic theories thus developed in an abstract level were not tested against economic reality. Econometrics aims primarily at the verification of economic theories.

B. Policy-Making

In many cases, we apply the various econometric techniques in order to obtain reliable estimates of the individual coefficients of the economic relationships from which we may evaluate elasticities or other parameters of economic theory (multipliers, technical coefficients of production, marginal costs, marginal revenues, etc.) The knowledge of the numerical value of these coefficients is very important for the decisions of firms as well as for the formulation of the economic policy of the government. It helps to compare the effects of alternative policy decisions.

C. Forecasting

In formulating policy decisions, it is essential to be able to forecast the value of the economic magnitudes. Such forecasts will enable the policy-maker to judge whether it is necessary to take any measures in order to influence the relevant economic variables.

1.4 Methodology of Econometric Research

In any econometric research we may distinguish four basic stages:

A. Specification of the model

The first and the most important step the econometrician has to take in attempting the study of any relationship between variables, is to express this relationship in mathematical form, that is, to specify the model with which the economic phenomenon will be explored empirically. This is called the specification of the model or formulation of the maintained hypothesis. It involves the determination of:

i) The dependent and explanatory variables which will be included in the model. The econometrician should be able to make a list of the variables that might influence the dependent variable based on:

- General economic theories
- Previous studies in any particular field and
- Information about individual condition in a particular case, and the actual behavior of the economic agents may indicate the general factors that affect the dependent variable.

ii) The a priori theoretical expectations about the sign and the size of the parameters of the function. These priori definitions will be the theoretical criteria on the basis of which the results of the estimation of the model will be evaluated.

Example: Consider a simple Keynesian consumption function

$C = \beta_0 + \beta_1 Y + U$ Where: C = Consumption function Y = level of income.

In this function, the coefficient β_1 is the marginal propensity to consume (MPC) and should be positive with a value less than unity ($0<\beta_1<1$). The constant intercept, β_0 of the function is expected to be positive.

iii) The mathematical form of the model (number of equations, linear or non-linear form of these equations, etc).

In general, the specification of the econometric model will be based on economic theory and on any available information relating to the phenomenon being studied.

B. Estimation of the Model

Having specified the econometric model, the next task of the econometrician is to obtain estimates (numerical values) of the parameters of the model from the data available. In the above Keynesian consumption function, if $\beta_1 = 0.8$, this value provides a numerical estimates of the marginal propensity to consume (MPC). It also supports Keynes' hypothesis that MPC is less than 1.

C. Evaluation of Estimates

After the estimation of the model, the econometrician must proceed with the evaluation of the results, that is, with the determination of the reliability of these results. The evaluation consists of deciding whether the estimates of the parameters are theoretically meaningful and statistically satisfactory. Various criteria may be used.

- *Economic a prior criteria*: – These are determined by the principles of economic theory and refer to the sign and the size of the parameters of economic relationships. In econometric jargon, we say that economic theory imposes restrictions on the signs and values of the parameters of economic relationships.
- *Statistical criteria*: – These are determined by statistical theory and aim at the evaluation of the statistical reliability of the estimates of the parameters of the model. The most widely used statistical criteria are the *correlation coefficient* and the *standard deviation (or the standard error)* of the

estimates. Note that the statistical criteria are secondary only to the a priori theoretical criteria. The estimates of the parameters should be rejected in general if they happen to have the wrong sign or size even though they pass the statistical criteria.

- *Econometric criteria*: – are determined by econometric theory. It aims at the investigation of whether the assumptions of the econometric method employed are satisfied or not in any particular case. When the assumptions of an econometric technique are not satisfied it is customary to re-specify the model.

D. Evaluation of the forecasting power of the estimated model

The final stage of any econometric research is concerned with the evaluation of the forecasting validity of the model. Estimates are useful because they help in decision-making. A model, after the estimation of its parameters, can be used in forecasting the values of economic variables. The econometrician must ascertain how good the forecasts are expected to be. In other words, he must test the forecasting power of the model.

It is conceivably possible that the model is economically meaningful and statistically and econometrically correct for the sample period for which the model has been estimated, yet it may not be suitable for forecasting due to, for example, rapid change in the structural parameters of the relationship in the real world.

Therefore, the final stage of any applied econometric research is the investigation of the stability of the estimates, their sensitivity to changes in the size of the sample. One way of establishing the forecasting power of a model is to use the estimates of the model for a period not included in the sample. The estimated value (forecast value) is compared with the actual (realized) magnitude of the relevant dependent variable. Usually, there will be a difference between the actual and the forecast value of the variable, which is tested with the aim of establishing whether it is (statistically) significant. If after conducting the relevant test of significance, we find that the difference between the realized value of the dependent variable and that estimated from the model is statistically significant, we conclude that the forecasting power of the model, its extra - sample performance, is poor.

Another way of establishing the stability of the estimates and the performance of the model outside the sample of the data, from which it has been estimated, is to re-estimate the function with an expanded sample, that is, a sample including additional observations. The original estimates will normally differ from the new estimates. The difference is tested for statistical significance with appropriate methods.

Reasons for a model's poor forecasting performance

a) The values of the explanatory variables used in the forecast may not be accurate
b) The estimates of the coefficients $(\beta's)$ may be poor, due to deficiencies of the sample data.
c) The estimates are 'good' for the period of the sample, but the structural background conditions of the model may have changed from the period that was used as the basis for the estimation of the model, and therefore the old estimates are not 'good' for forecasting. The whole model needs re-estimation before it can be used for prediction.

Example: Suppose that we estimate the demand function for a given commodity with a single equation model using time-series data for the period 1950 – 68 as follows

$$\hat{Q}_t = 100 + 5Y_t - 30P_t$$

This equation is then used for 'forecasting' the demand of the commodity in the year 1970, a period outside the sample data.

Given $Y_{1970} = 1000$ and $P_{1970} = 5$

$$\hat{Q}_t = 100 + 5(1000) - 30(5) = 4950 \text{ units.}$$

If the actual demand for this commodity in 1970 is 4500, there is a difference of 450 between the estimated from the model and the actual market demand for the product. The difference can be tested for significance by various methods. If it is found significant, we try to find out what are the sources of the error in the forecast, in order to improve the forecasting power of our model.

1.5 The structure of economic data

Economic data sets come in a variety of types. The most important data structures encountered in applied work are the following:

- **Cross-sectional data:** consists of a sample of individuals, households, firms, cities states, countries, or a variety of other units, taken at a given point in time. The population census conducted by CSA is a good example.
- **Time series data:** consists of observations on a variable or several variables over time. Examples of time series data include stock prices, money supply, consumer price index, gross domestic product, annual homicide rates, and automobile sales figures.
- **Pooled Cross Section data**: have both cross-sectional and time series features. A pooled cross section is analyzed much like a standard cross section, except that we often need to account for secular differences in the variables across the time.
- **Panel data (or longitudinal data):** This is a special type of pooled data in which the same cross-sectional units (say, individuals, firms or countries) are surveyed over time. It consists of a time series for each cross sectional member in the data set. Hence, the key feature of panel data that distinguishes it from a pooled cross section is the fact that the same cross-sectional units are followed over a given time period.