UNIT 1 INTRODUCTION TO ECONOMETRICS *

Structure

1.0 OBJECTIVES

After going through this unit, you will be able to

- explain the significance of econometrics in the field of economics;
- distinguish between econometrics, mathematical economics and economic statistics;
- describe the steps to be followed in an econometric study; and
- distinguish between association and causation.

1.1 INTRODUCTION

Econometrics connects the real world to the existing economic theories. Econometrics is based on the development of statistical methods for testing economic relationships and various economic theories. Econometrics helps us in two ways so far as relationship among variables is concerned: (i) explaining the past relationship among the variables, and (ii) forecasting the value of one variable on the basis of other variables.

Econometrics is an interface between economics, mathematics and statistics. It is mainly concerned with the empirical estimation of economic theories. In a broad sense we can say that it is a branch of social science that combines the tools of mathematics and statistical inferences, and these tools are applied to analyse economic phenomena. Econometrics uses regression technique which establishes an association or relationship between various variables. You should note that such relationships do not imply causation. (i.e., cause and effect relationship). The notion of causation has to originate from some theory of economics.

*Dr. Pooja Sharma, Assistant Professor, Daulat Ram College, University of Delhi

1.2 MEANING OF ECONOMETRICS

As mentioned earlier, econometrics deals with 'economic measurement'. It can be defined as a stream of social science which uses techniques of mathematics, statistical inference and economic theory applied to analyze any economic phenomenon. It deals with applications of mathematical statistics to economic data. The objective is to provide empirical support to the economic models constructed with the help of mathematical relationship and therefore obtain numerical results. Thus econometrics makes use of economic theory, mathematical economics, and economic statistics.

Econometrics hence becomes a platform for interaction of economic theory, (microeconomics or macroeconomics) using sophisticated mathematical tools in the form of mathematical equations and economic statistics, that is, data. Economic statistics is developed by collection, processing and presentation of data.

The central concern of mathematical economics is to express economic theory in mathematical forms or equations. These equations are finally are expressed in the form of models. You should note that mathematical economics does not evaluate the measurability or empirical verification of theory.

Economic statistics is primarily concerned with collection, processing and presentation of economic data in the form of charts, diagrams and tables. These data could be on microeconomic variables pertaining to households and firms or it could pertain to macroeconomic variables such as GDP, employment, prices, etc. Data for econometric models could be primary data or secondary data. An economic statistician usually limits himself/ herself to tabulation and processing of data.

Econometrics is mainly interested in empirical verification of economic theories. An econometrician would build models and test economic theories. In mathematical economics the relationship is deterministic. For example,

$$Y_i = a + bX_i \qquad \ldots(1.1)$$

In (1.1) above, Y is the explained variable

X is the explanatory variable

a and b are parameters.

The nature of relationship in econometrics, on the other hand, is stochastic. We add a stochastic error variable u_i to equation (1.1). For example,

$$Y_i = a + bX_i + u_i \qquad \ldots(1.2)$$

We will discuss further in Unit 4 on stochastic relationship among variables. In econometrics we generally require special methods due to the unique nature of economic data since such data are not generated under controlled experiments. The aim of econometrics is to bridge the gap between economic theory and actual measurement simply using the technique of statistical inference.

Thus, you should note three prominent features of econometrics. First, econometrics deals with quantitative analysis of economic relationships. Second, it is based on economic theory and logic. Third, it requires appropriate estimation methods to draw inferences. Thus, if the relationship is not expressed in quantitative terms we cannot apply econometric tools. Further, the variables are related according to some theory or logic; otherwise it will be similar to spurious correlation that you studied in statistics.

1.3 ECONOMICS AND ECONOMETIRCS

In economic theory the statements could be qualitative in nature. On the other hand, as discussed above, econometrics is a composition of mathematical economics, economic statistics and mathematical statistics. Let us take an example. The law of demand states that *ceteris paribus* (i.e., other things remaining the same) a rise in price of a commodity is expected to decrease the quantity demanded of that commodity. Therefore, economic theory predicts a negative or inverse relationship between price and quantity demanded of a commodity.

The law of demand does not provide any numerical measure of the strength of relationship between the two variables namely, price and quantity demanded of the commodity. It fails to answer the question that by how much the quantity will go up or down as a result of a certain change in price of commodity.

Econometrics provides empirical content to most economic theories. The real application of economics in the applied world includes forecasting various crucial economic variables such as sales, interest rates, money supply, price elasticity, etc.

The role of an economist is of great significance for an economy when it comes to understand how the variables would behave over a period of time or how these variables are connected to each other. An economist may be required to assess the impact of a proposed price increase on quantity demanded. For example, the impact of increase in price of electricity can be estimated by an econometrician and the electricity board may increase in price accordingly.

Check Your Progress 1

1) Bring out the differences between econometrics, mathematical economics and statistics.

.....

.....

.....

.....

2) Bring out the prominent features of econometrics.

.....

.....

.....

.....

1.4 METHODOLOGY OF ECONOMETRICS

In econometrics we generally come across several types of economic issues. These issues could be from any branch of economics such as microeconomics, macroeconomics, public economics, international trade, etc. These also could be from any of the sectors of the economy such as agriculture, industry and services. The problem at hand could be different. However, there certain common steps to be followed in an econometric study. These steps are as follows:

1. Construction of a statement of theory or hypothesis
2. Specification of mathematical model of the theory
3. Specification of statistical or econometric model
4. Obtaining requisite data
5. Estimation of the parameters of econometric model
6. Testing of hypothesis
7. Forecasting or prediction
8. Interpretation of results

These eight steps need to be elaborated further. Let us consider an example so that we can comprehend the issues. As you know from introductory macroeconomics, consumption expenditure depends upon income of households. Let us see how an econometric study can be carried out on the above relationship.

Step 1:Construction of a Statement of Theory or Hypothesis

The relationship between consumption and income is complex in nature. There are several factors that that influence consumption expenditure of a household

such as size of family, education level, health status of family members, place of stay (rural/urban), etc. In a simple model, however, the Keynesian consumption function establishes the relationship between consumption expenditure and household income. There are two concepts used by Keynes: average propensity to consume (APC), and marginal propensity to consume (MPC). According to Keynes the APC has a tendency to decline as income level increases. We can take the above statement as a hypothesis. Recall that hypothesis is based on certain theory or logic.

Step 2: Specification of Mathematical Model of the Theory

The consumption function takes the following form:

$$C_i = C_0 + cY_i \quad ...(1.3)$$

The variables C and Y represent consumption expenditure and income respectively. Note that C_0 is autonomous consumption, which is the bare minimum needed for survival. Even if income of a household is zero, consumption will be C_0. We note that for APC to decline, the parameters of equation (1.3) should fulfil the following two conditions: $C_0 > 0$ and $0 < c < 1$. These two conditions will help us in formulation of hypothesis in mathematical form.

Step 3: Specification of Statistical or Econometric Model

The consumption income relationship specified in equation (1.3) is exact in nature. If we plot the graph for equation (1.4) we will obtain a straight line. As mentioned earlier, the nature of relationship in econometrics is *stochastic*. Let us consider two households with the same level of income. Their consumption expenditure would be different due to certain factors other than income (such as health status of family members). In order to incorporate such factors we include another variable, u_i, in our model. The variable u_i has to meet certain conditions (to be discussed in Unit 4). Thus the econometric specification of the consumption function would be as follows:

$$C_i = C_0 + cY_i + u_i \quad ...(1.4)$$

Step 4: Obtaining Requisite Data

Data can be obtained from primary sources or secondary sources. You should refer to Unit 1 of the course BECC 107: Statistical methods for Economics for details on primary data and secondary data. In that Unit we have discussed the procedure of conducting sample survey and the important sources of secondary data.

For estimation of our econometric model given at equation (1.4) we need data on two variables, viz., income (Y) and consumption expenditure (C). As you know, income and expenditure are flow variables. Thus we have to specify a time period for these variables. For convenience from measurement point of view, we can take monthly income and monthly expenditure. Second, we have to define

what constitutes a household – who all are members of a household and who all are not included in the household. Third, we have to decide on the nature of data we collect.

As you know, four types of data are available. (i) time series, (ii) cross- sectional, (iii) pooled-data, and (iii) panel data.

(i) Time Series

Time series data are collected on a variable regularly over a period of time. There are some variables on which data is available on a daily basis (e.g., SENSEX and NIFTY). In the case of some other variables, it is available on monthly basis (e.g., consumer price index), on a quarterly basis (e.g., GDP) or on an annual basis (e.g., fiscal deficit).

(ii) Cross-Sectional Data

Cross-sectional data refers to data on several variables at a point of time. For example through a sample survey we can collect household data on expenditure, income, saving, debt, etc. Remember that time series data focuses on the same variable over a period of time while cross-sectional data focuses on several variables at the same point of time. Census data is an example of cross-sectional data.

(iii) Pooled Data

In the pooled data we have elements of both the time series and cross-sectional data. It is a time series of cross-sections. The observations in each cross section may not refer to the same unit. Let us consider an example. The census data in India is collected decennially. The number of households in each census however differs. Such data can be pooled to analyse the shifts in population characteristics over time. You can think of several other examples of pooled data. Examples could be employment and unemployment surveys, workforce participation rates, human development index, etc.

(iv) Panel Data

It is a special type of pooled data. Here observations are taken on the same sample units at multiple points of time. Suppose we want to analyse the variability of returns across shares in the stock market. We can take a sample of 50 public limited companies and observe their net asset value (NAV) daily for the month of August 2021. Thus we get 31 cross sections (since the month August has 31 days) of 50 firms. This constitutes a panel data. We call it a 'balanced panel' if all observations (for time period 1 to t; and for sample units 1 to n) are available. We call it an 'unbalanced panel' if some observations are missing.

Step 5: Estimation of the Parameters of the Econometric Model

We have discussed about sampling procedure, statistical estimation and testing of hypothesis in Block 4 of BECC 107. You need a thorough understanding of those concepts. Remember that in econometric estimation, the number of equations is more than the number of parameters. In order to estimate such models we need certain estimation methods. As you will come to know in subsequent Units of this course, there are quite a few estimation methods. You have been introduced to the least squares method in Unit 5 of the course BECC 107: Statistical Methods for Economics. There are certain econometric software available for estimation purpose. You will learn about econometric software in the course BECE 142: Applied Econometrics.

Step 6: Testing of Hypothesis

Once you obtain the estimates of the parameters, there is a need for test of the hypothesis. As you know, in a sampling distribution of an estimator, the estimate varies across sample. The estimate that you have obtained could be a matter of chance, and the parameter may be quite different from the estimate obtained. We need to confirm whether the difference between the parameter and the estimate really exists or it is a matter of sampling fluctuation.

For the consumption function (1.4), we should apply one sided t-test for testing of the condition $C_0 > 0$. For the marginal propensity to consume we should apply two-sided t-test $H_0 : c = 0$. For testing both the parameters together we should apply F-test.

There is a need to check for the correct specification of the model. Two issues are important here: (i) how many explanatory variables should be there in the regression model, and (ii) what is the functional form of the model.

The consumption function (see equation (1.4)) is a case of two-variable regress model. There is one explained variable and one explanatory variable in the model. If we include more number of explanatory variables (such as education, type of residential area, etc.) it becomes a multiple linear regression model. The functional form again could be linear or non-linear.

Step 7: Forecasting or Prediction

The estimated model can be used for forecasting or prediction. We have the actual value of the dependent variable. On the basis of the estimated regression model, we obtain the predicted value of the dependent variable. The discrepancy between the two is the prediction error. This prediction error is required to be as small as possible.

Step 8: Interpretation of Results

There is a need for correct interpretation of the estimates. In later Units of this course we will discuss issues such as model specification and interpretation of the result. The estimated model can be used for policy recommendation also.

1.5 ASSOCIATION AND CAUSATION

As you know from 'BECC 107: Statistical Methods for Economics' correlation implies association between two variables. Technically we can find out the correlation coefficient between any two variables (say the number of students visiting IGNOU library and the number of road accidents in Delhi). In some cases we find the correlation coefficients to be high also. Such relationship between variables however leads to spurious correlation. If we take two such variables (where correlation coefficient is high) and carry out a regression analysis we will find the estimates to be statistically significant. Such regression lines are meaningless. Thus regression analysis deals with the association or dependence of one variable on the other. It does not imply 'causation' however. The notion of causation has to come from existing theories in economics. Therefore a statistical relationship can only be statistically strong or suggestive. Unless causality is established between the variables the purpose of testing the economic theory would not make any sense. Most of the economic theories test the hypothesis whether one variable has a causal effect on the other.

Thus logic or economic theory is very important in regression analysis. We should not run a regression without establishing the logic for the relationship between the variables. Let us look into the case of the law of demand. While analysing consumer demand, we need to understand the effect of changing price of the good on the quantity demanded holding the other factors such as income, price of other goods, tastes and preferences of individuals unchanged. However, if the other factors are not held fixed, then it would be impossible to know the causal effect of price change on quantity demanded.

Check Your Progress 2

1) Explain the steps you would follow in an econometric study.

.....

.....

.....

.....

.....

2) Assume that you have to carry out an econometric study on Keynesian consumption function. Write down the steps you would follow.

.....

.....

.....

.....

.....

3) What do you understand by cause and effect relationship? How is it different from association?

.....

.....

.....

.....

.....

1.6 LET US SUM UP

In this Unit we dealt with the significance of econometrics in the field of economics. Econometrics connects the real world with theory. It helps us to ascertain the validity of theory.

Behind every econometric model there should be certain logic. The relationship between variables should come from certain economic theory or logic. Mere estimation of a regression model may give up meaningless results.

In this Unit we described the steps of carrying out econometric analysis. There are eight steps that we should follow while conducting an econometric study.

1.7 ANSERS TO CHECK YOUR PORGRESS EXERCISES

Check Your Progress 1

1) In Section 1.2 we have shown that econometrics and interface between economics, statistics and mathematical economics. Elaborate on that.

2) There are three prominent features of econometrics. First, econometrics deals with quantitative analysis of economic relationships. Second, it is based on economic theory and logic. Third, it requires appropriate estimation methods to draw inferences.

Check Your Progress 2

1) You should explain the eight steps mentioned in Section 1.4.

2) You should follow the eight steps given in Section 1.4. Your answer may include the following:

   (i) Statement of the theory: $0< MPC<1$

   (ii) Mathematical specification of the model: $C = \beta_1 + \beta_2 Y$, $0< \beta_2 <1$

   (iii) Econometric specification the model: $C = \beta_1 + \beta_2 Y + u$

   (iv) Collection of Data: Secondary data from RBI Handbook of Statistics

   (v) Parameter Estimation: $\widehat{C_i} = -184.08 + 0.7164Y_i$

   (vi) Hypothesis Test: $\beta_1 > 0$ or $\beta_2 > 0$

   (vii) Prediction: what is the value of C, given the value of Y?

3) Regression analysis deals with the association or dependence of one variable on the other. It does not imply causation. The notion of causation has to come from outside statistics. It could be some existing theory in economics. Therefore a statistical relationship can only be statistically strong or suggestive. Most of the economic theories test the hypothesis whether one variable has a causal effect on the other. Regression *per se* is all about association between two or more variables; this association might be suggestive. Unless causality is established between the variables the purpose of testing the economic theory would not make any sense.