# Module 1: Concept, Classical Methodology of Econometrics and Types of Data

## Econometrics

Econometrics is concerned with the systematic study of economic phenomena or theory using observed data. It means economic measurement. It is concerned with empirical determination of economic theory.

## Three important ingredients

1. Economic Theory
2. Economic Data
3. Statistical Methods

Econometrics mainly concerned for verifying or refuting economic theories or laws such as law of demand, consumption function. These laws are testable with economic data.

Econometrics also provides quantitative estimates of price or income elasticities of demand, estimates of returns to scale in production, the velocity of money.



## Methodology of Econometrics

Traditional econometric methodology proceeds along the following lines:

## 1. Statement of Theory or Hypothesis

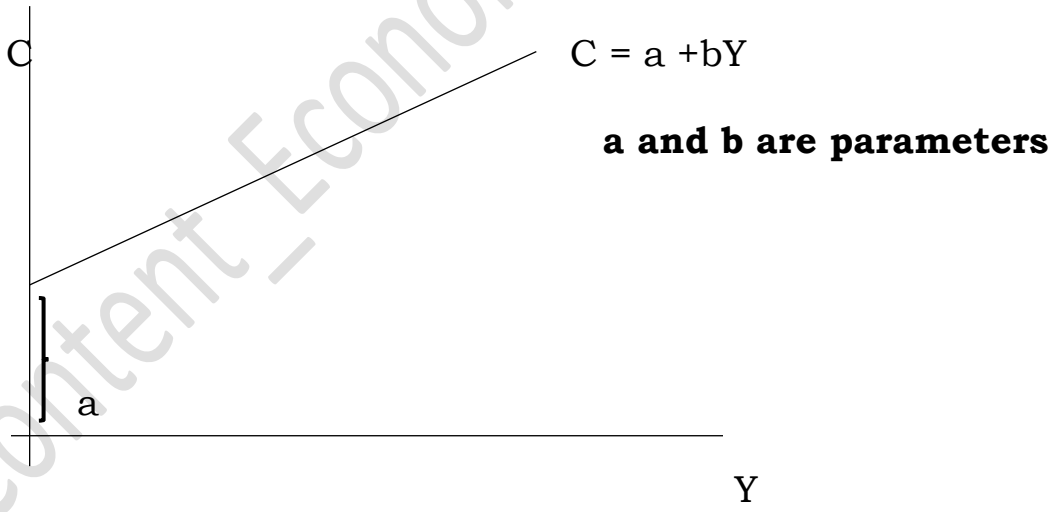
We have to state the theory to be empirically tested. In Keynesian consumption function, the hypothesis is, "Consumption is a function of income".

MPC < 1

## 2. Specification of the mathematical model of the theory

Mathematical form of Keynesian consumption function is

C = a +bY, 0 < b < 1, b: Marginal Propensity to Consume (MPC) = $\frac{dC}{dY}$



### 3. Specification of the Econometric Model

Mathematical relationship between consumption and income is deterministic or exact relationship.

**C = a + bY : Exact Relationship**

If Y = Rs. 1000/-, a = 50, b = 0.5

∴ C = 50 + 0.5*1000 = 550

In addition to income, other variables such as size of the family, ages of the members of the family, religion, wealth, geographic region, education also affect consumption.

Therefore, econometric model of consumption relationship is

**C = a + bY + u : Inexact or random relationship**

u : Disturbance or error term. It is random variable.

### 4. Obtaining Data

To estimate the econometric model, we need data. Estimation basically means to obtain the numerical values of the parameters a and b in the consumption income relationship.

### 5. Estimation of Econometric Model

Our next task is to estimate parameters of the consumption function. The estimated consumption function is

$$\widehat{C} = 50 + 0.5Y$$

The hat on C indicates that it is an estimate.

### 6. Hypothesis Testing

Keynes expected the MPC to be positive but less than one. In our example we found the MPC to be about 0.5. But before we accept this finding as confirmation of Keynesian consumption theory, we must enquire whether this estimate is sufficiently below unity to convince us that this is not a chance occurrence or peculiarity of the particular data we have used. In other words, is 0.5 is statistically less than one?

Such confirmation or refutation of economic theories on the basis of sample data can be done by statistical inference (hypothesis testing).

### 7. Forecasting or Prediction

We may use the model of consumption function to predict the values of the dependent variable C on the basis of known or expected values of the explanatory variable Y.

Suppose that we want to predict the consumption expenditure (for example of Durgapur City) for 2018. The per capita income for 2018 was Rs. 9000/-. Then we obtain

$$\widehat{C} = 50 + 0.5 * 9000 = 50 + 4500 = 4550$$

Predicted value of consumption expenditure is Rs. 4550/-.

Now the actual value of the consumption expenditure in 2018 is Rs. 4400/-. The estimated model thus over predicted the actual expenditure by about Rs. 150/-. This is the forecast error.

One can predict the future course of income, consumption expenditure and employment following a change in the government's fiscal policy.

## 8. Using the Model for Control or Policy Purposes

We have already estimated the consumption function given in the above equation. Now we suppose that government believes that consumer expenditure of Rs. 5000/- will keep the unemployment rate at its current level of 4.2% (in 2019). What level of income will guarantee the target amount of consumption expenditure?

We get $5000 = 50 + 0.5*Y$

Or, $Y = 9900$

Thus Rs. 9900/- per capita income will guarantee the unemployment rate of 4.2%.

Thus, an estimated model can be used for control or policy purposes. By appropriate fiscal and monetary policy mix, the government can manipulate the control variable Y to produce the desired level of the target variable C.

# Types of Data

## 1. Cross Sectional Data

Cross section data are data on one or more variables collected at the same point in time. It consists of a sample of individuals, households, firms, cities, states, countries, taken at a given point in time.

Cross sectional data are widely used in economics and other social sciences. In economics and, the analysis of cross-sectional data is closely aligned with the applied econometrics fields such as labour economics, state & local public finance, industrial organization, urban economics, demography and health economics. Data on individuals, households, firms, cities at a given point in time are important for testing microeconomic hypotheses and evaluating economic policies.

A cross sectional data on wages and other individual characteristics

| Observation | Wage | Education | Experience |
|---|---|---|---|
| 1 | 3100 | 11 | 2 |
| 2 | 3500 | 12 | 22 |
| 3 | 3000 | 16 | 18 |
| 4 | 3200 | 15 | 20 |
| 5 | 3400 | 14 | 15 |

## 2. Time Series Data

Time series data are data on a variable or several variables collected at regular intervals of time, such as daily (e.g. stock prices, weather reports), weekly (e.g. money supply figures) monthly (unemployment rate), quarterly (e.g. GDP), annually (e.g. Budget), quinquennially i.e. every five years (e.g. the census of manufacturers) or decennially (e.g. census of population).

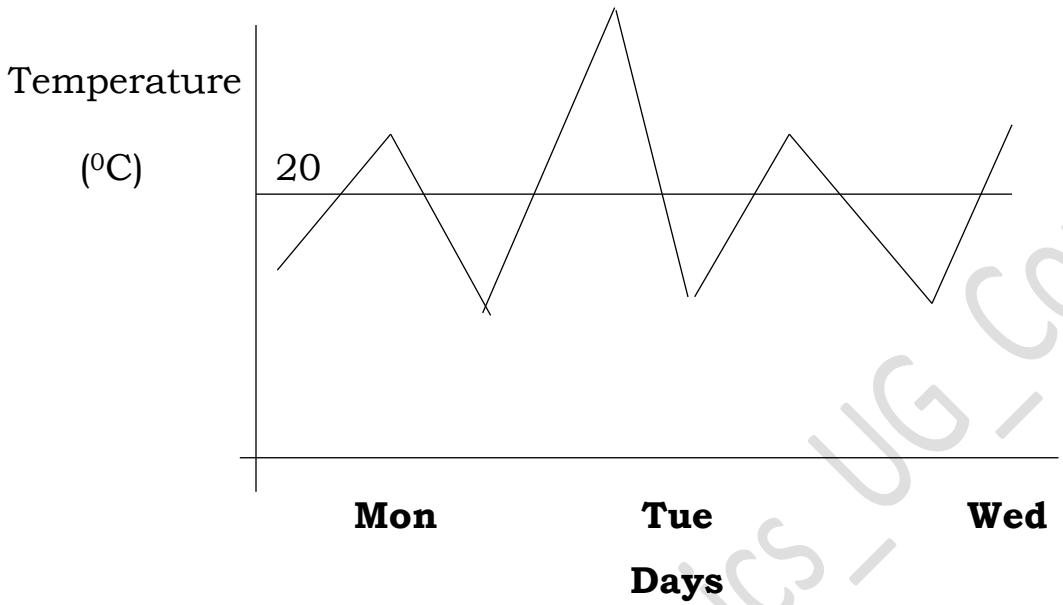
**Sample of Time Series Data**

| Observation | Year | Minimum Wage (Rs.) | Unemployment (%) | GDP (Mn Rs.) |
|---|---|---|---|---|
| 1 | 2016 | 200 | 2 | 900 |
| 2 | 2017 | 210 | 3 | 950 |
| 3 | 2018 | 220 | 4 | 970 |
| 4 | 2019 | 240 | 5 | 980 |
| 5 | 2020 | 260 | 6 | 1000 |

Time series data are heavily used in econometric studies. However, they present special problems for econometrician. Most empirical work based on time series data assumes that underlying time series is stationary.

A time series is stationary if its mean and variance do not vary systematically over time.

**Stationary Time Series Data**




### 3. Pooled Data

Some data sets have both cross-sectional and time series components. Such data sets are called pooled or combined data. For example, suppose that two cross sectional household surveys are taken in India in 20018 and 2019 respectively. In 2018, a 100 sample of households is surveyed for variables such as income, savings, family size and so on. In 2019, a new sample of 150 households is taken using the same survey questions. To increase our sample size, we can form a pooled data by combining the two years (2018 and 2019).

**Therefore, the data set can be represented as:**

| Observation | Year | Income (Rs) | Savings (Rs) |
|---|---|---|---|
| 1 | 2018 | 4000 | 1000 |
| 2 | 2018 | 5000 | 800 |
| 3 | 2018 | 3500 | 1200 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 100 | 2018 | 4500 | 800 |
| 101 | 2019 | 6000 | 1500 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 250 | 2019 | 7000 | 1700 |

Pooling cross sections from different years is often effective way of analysing the effects of new government policy. The idea is to collect data from years before and after a key policy change. Example – impact of reduction of income tax on savings.

### 4. Panel or Longitudinal Data

This is a special type of pooled data in which the same cross-sectional unit (say, a family or a firm) is surveyed over time. As an example, suppose same households (or people living in the same address) are

surveyed at periodic intervals in a census of households. We may collect information such as wage, education and job experience for a set of households fore the years 2015 and 2016.

**A Two-Year Panel Data on Households**

| Households | Year | Wage | Education (in Years) | Job Experience (In Years) |
|---|---|---|---|---|
| 1 | 2015 | 3100 | 10 | 5 |
| | 2016 | 3200 | 10 | 5 |
| 2 | 2015 | 3500 | 8 | 3 |
| | 2016 | 3700 | 8 | 3 |
| . | . | . | . | . |
| . | . | . | . | . |
| 200 | 2015 | 5400 | 9 | 10 |
| | 2016 | 5600 | 9 | 10 |

The key feature of panel data that distinguishes them from a pooled cross section is that the same cross-sectional units (individuals, firms or countries) are followed over a given time period.