Introduction to Econometrics

From micro: Demand curve is relationship b/w Quantity demanded & mkt price level.

$$Q_d = \alpha - \beta P \quad , \quad \alpha, \beta > 0$$ ---- equation/mathematical formulation of the demand curve.

=> Numerical estimates of parameters of the equation.

=> For forecasting purposes/mkt studies etc.

(i) For estimating an equation from economic model data needs to be collected.

Population: entire set under study.

Sample: a small portion of the population/subset of the population from which data is collected.

The sample data used for analysis is required to draw inferences about the population in general.

The sample should be a "true" representative of the population.

Eg: Suppose we want estimate the mkt demand curve for a particular good:

$$[Q = \alpha + \beta P] \quad [\text{expected}: \alpha > 0, \beta < 0]$$

dependent variable. independent variable

$$Q_{jio} = \alpha + \beta P_{jio} + \gamma \cdot P_{airtel} + \delta \cdot M + \cdots$$

deterministic factors.

In real world, the value of the dependent variable might not only be influenced by deterministic factors [variables obtained from economic theory] but also random factors which are completely uncorrelated with the given variables.

Economic model: $Q = \alpha + \beta P$ $[\alpha > 0, \beta < 0]$.

Econometric model: $Q = \alpha + \beta P + u$ → random disturbance term.

Data used for Econometric analysis:-

(i) Cross-sectional data.

Eg: estimate the dd fn: $Q = \alpha + \beta P$ for India.

data collected at the same point in time from different individuals on the same objective

a sample of size 'n' is chosen/surveyed and data is collected on (Q, P).

(ii) Time-series data:

Eg: the fluctuation of GDP overtime.

Collect data on the same variable from the same unit overtime.

(iii) Panel data/Longitudinal data:

Eg: To study how the preference pattern of individuals changes overtime for a particular good.

First fix the individuals in the sample and then collect data from the must fix the individuals in the sample and then collect data from those individuals overtime.

As different sampling units are used [cross-sectional aspect] and data collected overtime [time series aspect].

Setup: Define Y = Dependent Variable.

X = Independent Variable.

The relationship b/w X & Y in the population is given by:

True Model: $Y = \alpha + \beta X + u$.

where α, β = parameters of the model [to be estimated]

u = random disturbance term/ stochastic term.

To estimate the True Model, data is collected from the population. Let the sample size by 'n'.

Ind 1 : Y_1, X_1

Ind 2 : Y_2, X_2

⋮

Ind n : Y_n, X_n

$\Rightarrow$ Let a sample of size 'n' $(Y_i, X_i)_{i=1}^{n}$ be collected from the population.

Using the data we will now estimate the model:

True Model: $Y_i = \alpha + \beta X_i + u_i, \; i = 1, 2, \ldots, n$

Estimated Model: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$

where $\hat{\alpha}$ is the estimate of α [from the sample]

& $\hat{\beta}$ is the estimate of β [from the sample] estimate of α [from the sample]

& $\hat{\beta}$ is the estimate of β [from the sample]

$\boxed{\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i}$ $\Rightarrow$ obtained relationship b/w Y & X.

using this eqn, we will now derive inferences about the population.

Simple Linear Regression Model (SLRM)

A sample of size n $(Y_i, X_i)_{i=1}^{n}$ is collected from the population.

True Model: $Y_i = \alpha + \beta X_i + u_i$, $i = 1, 2, \ldots, n$ [unknown]

Estimated Model: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$, $i = 1, 2, \ldots, n$ [known].

$\hookrightarrow$ why write $\hat{Y}_i$ instead of Y_i? (HW)