Econometrics

Mrinal Kanti
Akshat Shankar
Mansi Goel

Econometrics

- Econometrics is the process of empirical determination of economic laws….
- Statement of theory or hypothesis.
- Specification of the mathematical model of the theory
- Specification of the statistical, or econometric, model
- Obtaining the data
- Estimation of the parameters of the econometric model
- Hypothesis testing
- Forecasting or prediction
- Using the model for control or policy purposes.

Regression Analysis

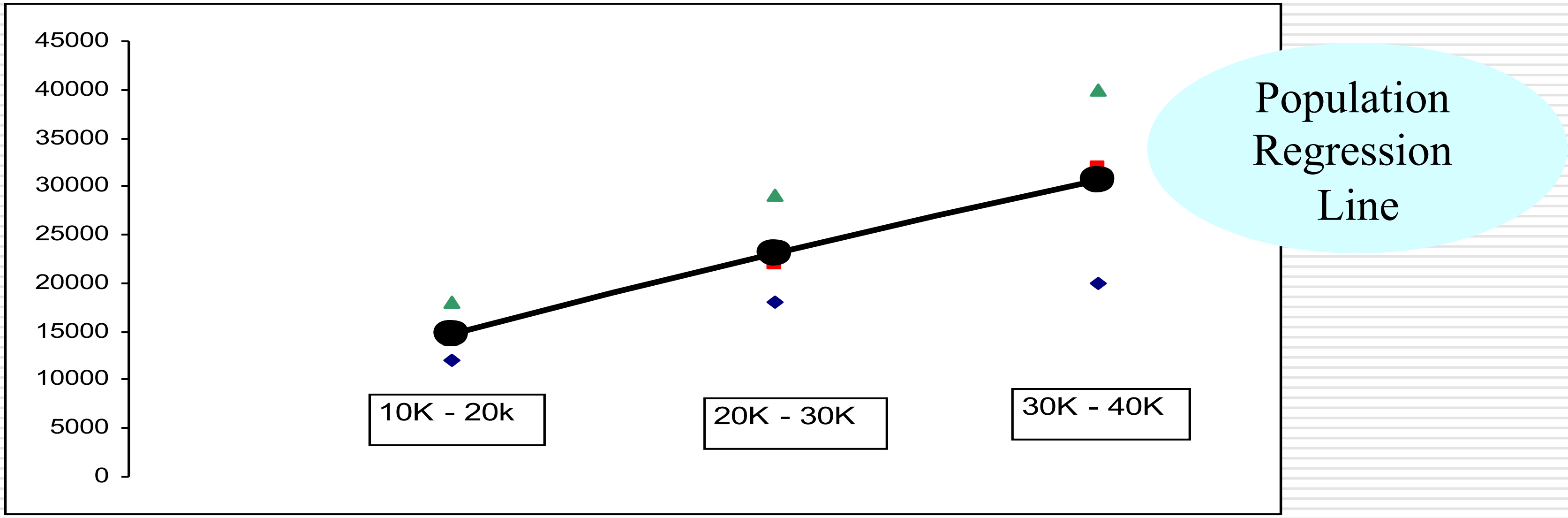
- …study of the dependence of one variable, the *dependent variable*, on one or more other variables, the *explanatory variables*,…
- Important: Dependence does *not* imply causation
- Regression vs. Correlation: Regression is concerned with finding an average value of the dependent variable on the basis of a sample of the explanatory variables

Types of Data

- Cross Sectional: Comparing different people at the same point of time. E.g. the height of all the participants in this class
- Time Series: Comparing values of the same entity across time: e.g. GDP of India from 1992- 2010
- Panel Data: Comparing a set of individuals across time. E.g. the height of all participants across time.

Bivariate Regression

	Income		
	10K - 20K	20K - 30K	30 K- 40 K
Expenditure	12000	18000	20000
	14000	22000	32000
	18000	29000	40000
Conditional Average	**14667**	**23000**	**30667**
Unconditional Average	**22778**		



On Average, the expenditure increases with income

The Population Regression Function

- The PRL is the curve connecting the means of subpopulations of Y given values of X
- Therefore, the conditional mean of Y, E(Y/X) is a function of X

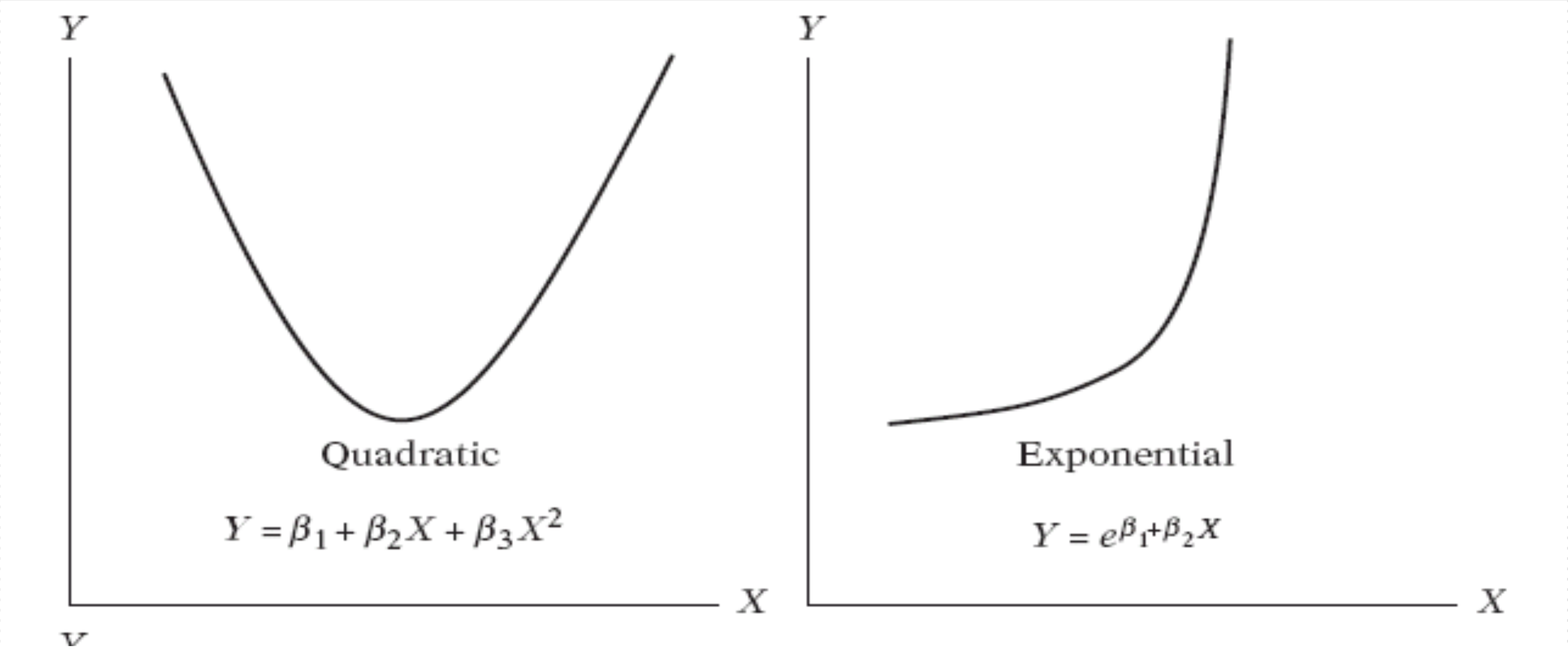
$$E(Y \mid X_i) = f(X_i)$$

- What kind of Function?
  - We are primarily concerned with Linear functions: $E(Y \mid X_i) = \beta_1 + \beta_2 X_i$
  - Linearity here is in parameters, so

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$$

is still linear

Examples of Linearity in Parameters



The Error Term

- An individual's actual value is clustered around the expected value
- Deviation of an individual Y_i around its expected value

$$u_i = Y_i - E(Y \mid X_i) \Rightarrow Y_i = E(Y \mid X_i) + u_i$$

- u_i is the stochastic error or disturbance term
- So any family's income can be expressed in two parts, the deterministic and the stochastic
- So we get,

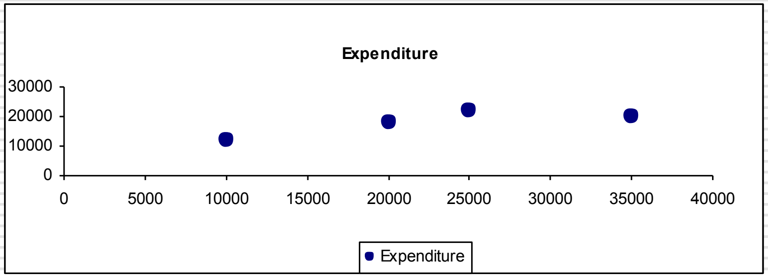
$$\begin{aligned} Y_i &= E(Y \mid X_i) + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned}$$

- The disturbance term is a proxy for all the variables that are excluded from the model but which affect Y

The Sample Regression Function

- Since most of the times the data for the entire population is not available, we have to use sample information to estimate the PRF

Expenditur	Income
18000	20000
22000	25000
12000	10000
20000	35000



- Can write the Sample Regression Function as $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

In summary, our problem is to estimate the PRF

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

on the basis of the SRF

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + u_i$$

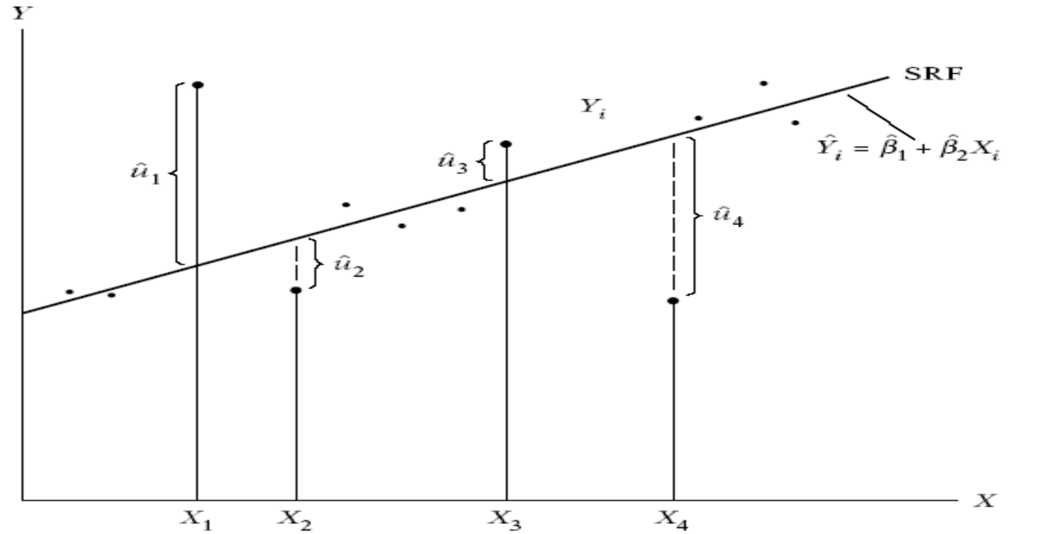
Estimation

- Need to estimate the SRF
- Such that the errors get minimized
- Can try to minimize the sum of the errors

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

$$\Sigma \hat{u}_i = \Sigma (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)$$



But giving equal weight to all the errors can be misleading… can get very low sum of errors as they could get cancelled out

Ordinary Least Squares

- Weight each error by itself
- Minimize Error Sum of Squares

$$\Sigma \hat{u}_i^2 = \Sigma (Y_i - \hat{Y}_i)^2$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + u_i$$

$$\Sigma \hat{u}_i^2 = \Sigma (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

- Next question:
  - Different beta's give us different set of errors
  - Is there an algorithm which will always give us the estimates with the lowest $\Sigma \hat{u}_i^2$
  - The OLS procedure give us estimates of the beta's with several desirable properties.
  - Derivation to be done on board.

$$\beta_2 = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{(\Sigma X_i - \bar{X})^2} \qquad \beta_1 = \bar{Y} - \beta_2 \bar{X}$$

Imp. Numerical Properties of the Estimators

$$\beta_1 = \bar{Y} - \beta_2 \bar{X}$$
$$\bar{Y} = \beta_1 + \beta_2 \bar{X}$$

The SRF line passes through the means of Y and X

$$E(\hat{Y}) = E(Y)$$

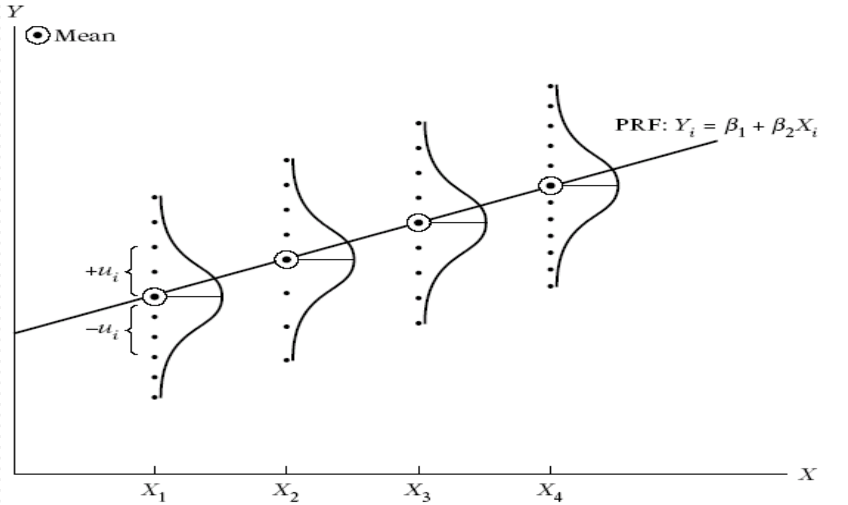
The means of the actual and estimated Y’s are equal

$$E(u_i) = 0$$

Mean of the disturbances is zero

Classical Linear Regression Model-Assumptions

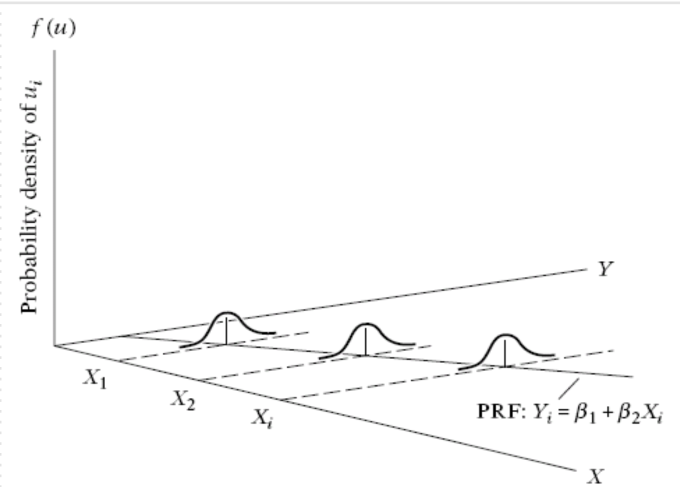
1. Model is Linear in Parameters $Y_i = \beta_1 + \beta_2 X_i + u_i$
2. X values are fixed, non-stochastic
3. Conditional mean of u_i is 0 $E(u_i | X_i) = 0$

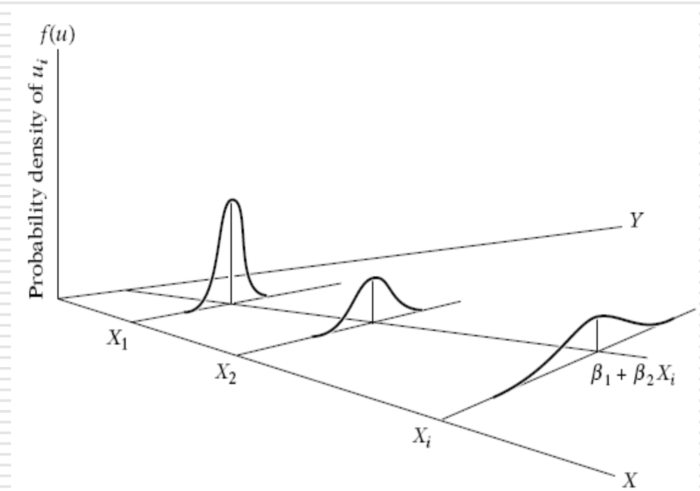


Assumptions contd.

4. Homoscedasticity – equal variance of u_i- Conditional variance of u_i are same

$$\begin{aligned}\text{var}(u_i \mid X_i) &= E[u_i - E(u_i \mid X_i)]^2 \\ &= E\left(u_i^2 \mid X_i\right) \text{ because of Assumption 3} \\ &= \sigma^2\end{aligned}$$





Assumptions contd.

5. No Autocorrelation between disturbances- given any two X's, the correlation between the corresponding u's is zero.

$$\text{cov}(u_i, u_j \mid X_i, X_j) = E\{[u_i - E(u_i)] \mid X_i\}\{[u_j - E(u_j)] \mid X_j\}$$

$$= E(u_i \mid X_i)(u_j \mid X_j)$$

$$= 0$$

6. Zero covariance between u_i and X_i
   $E(u_i X_i) = 0$
7. Number of observations must be greater than the number of variables in the model
8. Variance of X must be a positive number- $\text{var}(X) = \frac{\sum (X_i - \bar{X})^2}{n-1}$
9. Regression model is correctly specified
10. No perfect Multicollinearity- no perfect relationship among the X variables