3. Linear Regression and OLS

ECON8011 Microeconometrics

Burcu Duzgun Oncel *

*Marmara University Department of Economics

Spring2023

Table of contents

Simple Regression Model

$$Y = \beta_0 + \beta_1 X + u$$

- u: disturbance or error term
  - Represents the factors other than X that affect Y
  - Treated as unobserved
- β_0 : intercept parameter
- β_1 : slope parameter

$$Y = \beta_0 + \beta_1 X + u$$

- Taking the expected value of the above equation conditional on X and using $E(u) = 0$ gives:

$$E(Y/X) = \beta_0 + \beta_1 X$$

- population regression function (PRF), $E(Y/X)$, is a linear function of X. The linearity means that a one-unit increase in X changes the expected value by the amount β_1.

$$Y = \underbrace{\beta_0 + \beta_1 X}_{systematic} + \underbrace{u}_{unsystematic}$$

- Now, suppose we choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to make the sum of squared residuals as small as possible.

$$\sum_{i=1}^{n} u_1^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

- OLS estimators for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\hat{\beta}_0 = \overline{Y} - \beta_1 \overline{X}$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

k-Variable Regression Model

- Set of n simultaneous equations:

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + ... + \beta_k X_{k1} + u_1$$

$$Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + ... + \beta_k X_{k2} + u_2$$

....

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + ... + \beta_k X_{kn} + u_n$$

- Systems of equations: The matrix representation of the general k variable model:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & . & . & X_{k1} \\ 1 & X_{22} & X_{32} & . & . & X_{k2} \\ 1 & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & X_{2n} & X_{3n} & . & . & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix}$$

- More generally, for the i^{th} observation:

  $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i$

- Usually $X_{1i} = 1$ (an intercept)

- Regressor vector $\mathbf{X_i}$ and parameter vector β are $Kx1$ column vectors.

$$\underset{Kx1}{\mathbf{X_i}} = \begin{bmatrix} X_{1i} \\ X_{2i} \\ . \\ . \\ X_{Ki} \end{bmatrix} \text{ and } \underset{Kx1}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_K \end{bmatrix}$$

$$\mathbf{X_i}'\beta = \begin{bmatrix} X_{1i} & X_{2i} & .... & X_{ki} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_n \end{bmatrix} = \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

- Note that all vectors are defined to be column vectors.

- For the i^{th} observation: $Y_i = \mathbf{X_j}'\beta + u_i$

- Now combine all N observations from sample $(Y_i, \mathbf{X_i}), i = 1, ..., N$.
- The linear regression model is:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} \mathbf{X_1}\beta \\ \mathbf{X_2}\beta \\ . \\ . \\ \mathbf{X_N}\beta \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix}$$

- This is : $\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$

$$\text{where } \underset{Nx1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} \qquad \underset{NxK}{\mathbf{X}} = \begin{bmatrix} X_1' \\ X_2' \\ . \\ . \\ X_n' \end{bmatrix} \qquad \underset{Nx1}{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix}$$

- The OLS estimator: $\hat{\beta} = (X'X)^{-1}X'Y$

CLRM Assumptions

- The regression model is linear in parameters.
- The values of the regressors, the X's are fixed in repeated sampling.
- For given X's the mean value of the disturbance u_i is zero.
- For given X's, the variance u_i is constant or homoscedastic.
- For given X's, there is no autocorrelation in the disturbances.
- If the X's are stochastic, the disturbance term and the (stochastic) X's are independent or at least uncorrelated.

- The number of observations must be greater than the number of regressors.
- There must be sufficient variablity in the values taken by the regressors.
- The regression model is correctly specified.
- There is no exact linear relationship (i.e., multicollinearity) in the regressors.
- The stochastic (disturbance) term u_i is normally distributed.

- The mean value of the disturbance u_i is zero.

$$E\begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{bmatrix} = E\begin{bmatrix} E(u_1) \\ E(u_2) \\ . \\ . \\ E(u_n) \end{bmatrix} = E\begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}$$

- Var-Cov matrix of the disturbances:

$$E(\mathbf{u}\mathbf{u}') = \begin{bmatrix} E(u_1^2) & E(u_1 u_2) & \dots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \dots & E(u_2 u_n) \\ . & . & \dots & . \\ E(u_n u_1) & E(u_n u_2) & \dots. & E(u_n^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots. & \sigma^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots. & 1 \end{bmatrix} = \sigma^2 \mathbf{I}$$

OLS Estimation

- The OLS estimator minimizes the sum of squared errors:

$$Q(\beta) = min \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (Y_i - X_i'\beta)^2$$

$$Q(\beta) = min \sum_{1}^{N} \hat{u}'\hat{u} = min \sum_{i=1}^{N} (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

where $\sum_{i=1}^{N} (Y - X\hat{\beta})'(Y - X\hat{\beta})$

$$= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

- Minimization of sum of squared errors lead:

$$Q(\beta) = \frac{\partial(\hat{u}'\hat{u})}{\partial\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

$= (X'X)\beta = X'Y \rightarrow$ **normal equation**

$$\hat{\beta_{OLS}} = (X'X)^{-1}X'Y$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_1 \end{bmatrix}$$

$$(\mathbf{X'X}) = \begin{bmatrix} 1 & 1 & 1 & .... & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \dots & \dots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & .... & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 & \\ Y_2 & \\ Y_3 & \\ \dots & \dots \\ Y_n & \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Example

- Example: $N = 4$ with (X, Y) equal to (1,1), (2,3), (2,4) and (3,4).
- Then **Y** is 4x1 and **X** is 4x2 with

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} \qquad X = \begin{bmatrix} X_1' \\ X_2' \\ X_3' \\ X_4' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

- Thus;

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 4 & 8 \\ 8 & 18 \end{bmatrix}^{-1} \begin{bmatrix} 12 \\ 27 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$$

- Intercept $\hat{\beta}_1 = 0$ and slope coefficient $\hat{\beta}_2 = 1.5$

Properties of OLS Estimators

- $\hat{\beta}_{OLS}$ is always estimable, provided $rank[X] = K$
- But properties $\hat{\beta}_{OLS}$ depend on the true model
  - called the data generating process (d.g.p.)
- Essential result:
  - If the d.g.p. is correctly specified and the error term u_i is uncorrelated with regressors X_i
  - Then
    1. $\hat{\beta}$ is consistent for β
    2. $\hat{\beta}$ is normally distributed in large samples (asymptotically)
    3. Variance of $\hat{\beta}$ varies with assumptions on error u_i

Unbiasedness of $\hat{\beta}$

- $\hat{\beta}$'s are linear, unbiased and consistent.
- To establish the unbiasedness of $\hat{\beta}$'s :

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon$$

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(\epsilon) = \beta \text{ since } E(\epsilon) = 0$$

Consistency of $\hat{\beta}$

- Consistency means that the probability limit (*plim*) of $\hat{\beta}$ equals β
  - That is : $lim_{N\to\infty} Pr(|\hat{\beta}-\beta|<\epsilon)=1$ for any $\epsilon>0$
- We have:

$$plim\hat{\beta} = plim\{\beta + (X'X)^{-1}X'u\}$$

$$= plim\beta + plim\{(\sum_i X_iX_i')^{-1}\sum_i X_iu_i\}$$

$$= plim\beta + plim(\tfrac{1}{N}\sum_i X_iX_i')^{-1} x plim\tfrac{1}{N}\sum_i X_iu_i$$

$$= \beta + plim(\tfrac{1}{N}\sum_i X_iX_i')^{-1} x0$$

$$= \beta$$

Variance of $\hat{\beta}$

$\hat{\beta} - \beta = (X'X)^{-1}X'u$ (from the proof of unbiasedness)

$$var - cov(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$= E\{[(X'X)^{-1}X'u][(X'X)^{-1}X'u]'\}$$

$$= E[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'\underbrace{E(uu')}_{\sigma^2 I}X(X'X)^{-1}$$

$$= \sigma^2 \underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}$$

$= var - cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$ where $\hat{\sigma^2} = \frac{\hat{u}'\hat{u}}{n-k}$

- Variance-covariance matrix of $\hat{\beta}$ can be written as:

$$var - cov(\hat{\beta}) = \begin{bmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) & \dots & cov(\hat{\beta}_1, \hat{\beta}_k) \\ cov(\hat{\beta}_2, \hat{\beta}_1) & var(\hat{\beta}_2) & \dots & cov(\hat{\beta}_2, \hat{\beta}_k) \\ \dots & \dots & \dots & \dots \\ cov(\hat{\beta}_k, \hat{\beta}_1) & cov(\hat{\beta}_k, \hat{\beta}_2) & \dots & var(\hat{\beta}_k) \end{bmatrix}$$

Efficiency of $\hat{\beta}$

- $\beta^* = [(X'X)^{-1}X' + C]Y$ where C is a matrix of constants.
- $\beta^* = [(X'X)^{-1}X' + C](X\beta + u)$
$$= \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'u + CX\beta + Cu$$
- $\beta^* = \beta + CX\beta + (X'X)^{-1}X'u + Cu$ for unbiasedness $CX = 0$
- $\beta^* - \beta = (X'X)^{-1}X'u + Cu$
- $E[(\beta^* - \beta)(\beta^* - \beta)'] = Var(\beta^*) =$
$E\{[(X'X)^{-1}X'u + Cu][(X'X)^{-1}X'u + Cu]'\}$

$$E[\underbrace{(X'X)^{-1}X'uu'X(X'X)^{-1}}_{\sigma^2(X'X)^{-1}} + \underbrace{(X'X)^{-1}X'uuC'}_{0} + \underbrace{Cuu'X(X'X)^{-1}}_{0 \text{ since } CX=0} + \underbrace{Cuu'C'}_{\sigma^2 CC'}]$$

$$Var(\beta^*) = \sigma^2(X'X)^{-1} + \sigma^2 CC'$$

then C must be 0 for $var(\hat{\beta})$ be minimum.

- The coefficient of determination R^2 is determined as $R^2 = \frac{ESS}{TSS}$
- In matrix:

$$R^2 = \frac{\hat{\beta}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}$$