Introductory Econometrics

Multiple Linear Regression Model (I)

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

Basic Settings

- The multivariate linear regression model can be expressed as follows

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u.$$

- Affine population regression function is

$$\mathrm{E}\left(Y \mid X_1, X_2, \cdots X_k\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k.$$

- β_j is referred to as the **partial regression coefficient**, designating the extent to which $\mathrm{E}\left(Y\right)$ changes as X_j changes by one unit.

Basic Settings for Multiple Linear Regression Model

- For specific sample $\{(X_{i1}, X_{i2}, \ldots, X_{ik}, Y_i) : i = 1, 2, \cdots, n\}$,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i,$$

or in **matrix form**

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}_{n\times(k+1)}$$

Basic Settings

and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} \qquad \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{n\times 1}$$

Basic Settings

- Sample regression function for multiple linear regression model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k.$$

- For specific sample $\{(X_{i1}, X_{i2}, \ldots, X_{ik}, Y_i) : i = 1, 2, \cdots, n\}$, we have equivalent representation of sample regression function

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik} \\ Y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik} + e_i \end{aligned}$$

where e_i is referred to as the residual.

Basic Settings

- Sample regression function in matrix form

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$
$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$$

where

$$\hat{\boldsymbol{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \quad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Basic Assumptions

Assumption 1: Model is correctly specified, that is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}.$$

Assumption 2: Nonsingularity assumption. The rank of $\boldsymbol{X}'\boldsymbol{X}$ is $k+1$ with probability 1 and $\boldsymbol{X}'\boldsymbol{X}/n$ converges in probability to an invertible matrix, i.e. $P\lim \boldsymbol{X}'\boldsymbol{X}/n = \boldsymbol{Q}$ and $\boldsymbol{Q}$ is invertible.

Basic Assumptions

Assumption 3: Strict exogeneity.

$$\mathrm{E}\left(u_i \mid X_1, X_2, \cdots, X_k\right) = 0 \quad i = 1, 2, \cdots, n,$$

or in matrix form

$$\mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{X}\right) = \mathbf{0}.$$

Besides, **Assumption 3** implicitly suggests that $\forall i, j$

$$\mathrm{E}\left(u_i \mid X_{ij}\right) = 0,$$

or by denoting the i-th row of $\boldsymbol{X}$, we have

$$\mathrm{E}\left(\boldsymbol{X}_i' u_i\right) = \mathbf{0}.$$

Basic Assumptions

Assumption 4: Spherical error variance.

$$\operatorname{Var}\left(u_i \mid X_1, X_2, \cdots, X_k\right)=\sigma^2 \quad i=1,2, \cdots, n.$$

$$\operatorname{Cov}\left(u_i, u_j \mid X_1, X_2, \cdots, X_k\right)=0, i \neq j, \quad i, j=1,2, \cdots, n$$

Basic Assumptions

and in matrix form

$$\operatorname{Var}(\boldsymbol{u} \mid \boldsymbol{X}) = \operatorname{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \operatorname{E}\left(\begin{matrix} u_1^2 & \cdots & u_1 u_n \\ \vdots & & \vdots \\ u_n u_1 & \cdots & u_n^2 \end{matrix} \middle| \boldsymbol{X}\right)$$

$$= \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ denotes indentity matrix of n dimensions.

Basic Assumptions

Assumption 5: Normality assumption

$$u_i \mid X_1, X_2, \cdots, X_k \sim N\left(0, \sigma^2\right).$$

OLS Estimation

- Given the sample $\{(X_{i1}, X_{i2}, \cdots, X_{ik}, Y_i) : i = 1, 2, \cdots, n\}$, the target of OLS estimation finding $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ such that the sample regression function takes affine functional form

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_k X_{ik}$$

and

$$Q = \sum_{i=1}^{n} e_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2$$

is minimized.

OLS Estimation

- By taking partial derivatives of Q with respect to $\hat{\beta}_j$ $(j = 0, 1, \cdots, k)$ and setting the corresponding partial derivatives equal to 0, we obtain the system equations as follows

$$\begin{cases} \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \beta_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right) = 0 \\ \sum X_{i1} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right) = 0 \\ \sum X_{i2} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right) = 0 \\ \qquad\qquad\vdots \\ \sum X_{ik} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right) = 0 \end{cases}$$

OLS Estimation

By rearranging the system of equations, we can obtain the matrix representation

$$\underbrace{\begin{pmatrix} n & \sum X_{i1} & \cdots & \sum X_{ik} \\ \sum X_{i1} & \sum X_{i1}^2 & \cdots & \sum X_{i1}X_{ik} \\ \vdots & \vdots & & \vdots \\ \sum X_{ik} & \sum X_{ik}X_{i1} & \cdots & \sum X_{ik}^2 \end{pmatrix}}_{\boldsymbol{X}'\boldsymbol{X}} \underbrace{\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}_{\hat{\boldsymbol{\beta}}}$$

$$= \underbrace{\begin{pmatrix} n & \sum X_{i1} & \cdots & \sum X_{ik} \\ \sum X_{i1} & \sum X_{i1}^2 & \cdots & \sum X_{i1}X_{ik} \\ \vdots & \vdots & & \vdots \\ \sum X_{ik} & \sum X_{ik}X_{i1} & \cdots & \sum X_{ik}^2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}}_{\boldsymbol{X}'\boldsymbol{Y}}$$

OLS Estimation

Given **Assumption 2**, which suggests that that $\boldsymbol{X}'\boldsymbol{X}$ is invertible, we obtain the OLS estimator expressed in matrix form:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y}.$$

Geometric Interpretation

- The system of equations above can be expressed in a compact form $\boldsymbol{X}'\boldsymbol{e} = \mathbf{0}$ since

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\underbrace{\left(\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}\right)}_{\boldsymbol{Y}}$$

- Using the property $\boldsymbol{X}'\boldsymbol{e} = \mathbf{0}$, we can show that

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1\overline{X}_1 + \hat{\beta}_2\overline{X}_2 + \cdots + \hat{\beta}_k\overline{X}_k.$$

Geometric Interpretation

- Projection Matrices.

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$$
$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} \equiv P\boldsymbol{Y}$$
$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - P\mathbf{Y} = (\boldsymbol{I}_n - P)\boldsymbol{Y} \equiv M\boldsymbol{Y}$$

- P and M are symmetric and idempotent.

$$P^2 = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' = P$$
$$M^2 = (\boldsymbol{I}_n - P)(\boldsymbol{I}_n - P) = \boldsymbol{I}_n - P - P + P^2 = \boldsymbol{I}_n - P = M$$

- For idempotent matrix P and M,

$$\operatorname{tr}(P) = \operatorname{rank}(P) \quad \operatorname{tr}(M) = \operatorname{rank}(M).$$

Geometric Interpretation

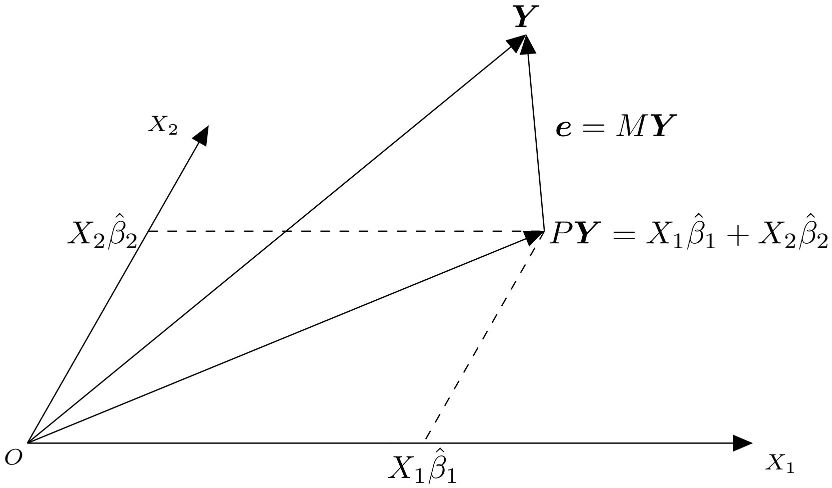
- $P\boldsymbol{X}=\boldsymbol{X}, \quad M\boldsymbol{X}=\mathbf{0}$, and $PM=\mathbf{0}$.
- Trace of P and M,

$$\begin{aligned} \operatorname{tr}(P) &= \operatorname{tr}\left[\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right]=\operatorname{tr}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right] \\ &= \operatorname{tr}\left(\boldsymbol{I}_k\right)=k+1 \\ \operatorname{tr}(M) &= \operatorname{tr}\left(\boldsymbol{I}_n-P\right)=\operatorname{tr}\left(\boldsymbol{I}_n\right)-\operatorname{tr}(P)=n-k-1 \end{aligned}$$

- Matrix representation of residuals $\boldsymbol{e}$ and dependent variables $\boldsymbol{Y}$.

$$\begin{aligned} \boldsymbol{e} &= M\boldsymbol{Y}=M\left(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}\right)=M\boldsymbol{u} \\ \boldsymbol{Y} &= (P+M)\boldsymbol{Y}=\hat{\boldsymbol{Y}}+\boldsymbol{e} \end{aligned}$$

Geometric Interpretation



Geometric Interpretation

- Matrix representation of sum of squared residuals.

$$\begin{aligned}\sum e_i^2 &= \boldsymbol{e}'\boldsymbol{e} \\ &= (M\boldsymbol{u})'(M\boldsymbol{u}) \\ &= \boldsymbol{u}'M\boldsymbol{u} \\ &= \boldsymbol{Y}'M\boldsymbol{Y}\end{aligned}$$

- Unbiased estimation of σ^2:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1}.$$

Moment Estimation

- Recall that based on **Assumption 3** we have $\mathrm{E}\left(\boldsymbol{X}_i' u_i\right) = \mathbf{0}$, which serve as the **population moment condition**. By using the **sample moment condition** as the proxy for population moment condition

$$\frac{1}{n}\sum \boldsymbol{X}_i'\left(Y_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{\mathrm{MM}}\right) = \mathbf{0}.$$

- The sample moment condition can be expressed in matrix form as follows

$$\frac{1}{n}\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{MM}}) = \mathbf{0}.$$

  Hence, the moment estimator is equivalent to the OLS estimator.

Maximum Likelihood Estimation

- Given the normality assumption in **Assumption 5** and the $i.i.d.$ assumption in **Assumption 4**, we can write down the **likelihood function** in terms of parameters to estimated,

$$
\begin{aligned}
L\left(\boldsymbol{\beta}, \sigma^2\right) &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma[Y_i-(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik})]^2} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})}
\end{aligned}
$$

Maximum Likelihood Estimation

- By taking log of the likelihood function, we obtain the log likelihood function as follows

$$
\begin{aligned}
L^* &= \ln L \\
&= -n \ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})
\end{aligned}
$$

- By taking partial derivatives of L^* with respect to σ^2 and $\boldsymbol{\beta}$ and setting the partial derivatives equal to 0, we obtain

$$
\begin{aligned}
\frac{\partial L^*}{\partial \hat{\boldsymbol{\beta}}_{\mathrm{ML}}} &= \frac{1}{\hat{\sigma}^2_{\mathrm{ML}}}\boldsymbol{X}'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}\right) = 0 \\
\frac{\partial L^*}{\partial \hat{\sigma}^2_{\mathrm{ML}}} &= \frac{n\pi}{2\pi\hat{\sigma}^2_{\mathrm{ML}}} + \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}\right)}{2\hat{\sigma}^4_{\mathrm{ML}}} = 0
\end{aligned}
$$

Maximum Likelihood Estimation

and we finally solves as follows

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = (\boldsymbol{X}'\boldsymbol{X})\,\boldsymbol{X}'\boldsymbol{Y}$$

$$\hat{\sigma}^2_{\text{ML}} = \frac{\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}}\right)\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}}\right)}{n} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n}$$

- $\hat{\boldsymbol{\beta}}_{\text{ML}}$ is equivalent to the OLS estimator but $\hat{\sigma}^2_{\text{ML}}$ is biased.

Goodness of Fit

- Recall the previous definition

$$
\begin{aligned}
\text{TSS} &\equiv \sum \left(Y_i - \bar{Y}\right)^2 = \sum y_i^2 \text{ Total Sum of Squares} \\
\text{ESS} &\equiv \sum \left(\hat{Y}_i - \bar{Y}\right)^2 = \sum \hat{y}_i^2 \text{ Explained Sum of Squares} \\
\text{RSS} &\equiv \sum \left(Y_i - \hat{Y}_i\right)^2 = \sum e_i^2 \text{ Residual Sum of Squares}
\end{aligned}
$$

- It can be shown that

$$
\text{TSS} = \text{ESS} + \text{RSS}\,.
$$

Goodness of Fit

- Based on the decomposition above

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}.$$

- With an intercept included in the regression model, it can be shown that $0 \leqslant R^2 \leqslant 1$.
- R^2 never decreases when additional regressors are included.
- A better measure of goodness-of-fit is given by the adjusted coefficient of determination,

$$\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)} = 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right).$$

Summary

- Basic settings and assumptions for multiple linear regression model.
- OLS estimation for multiple linear regression model.
- Geometric interpretation for OLS.
- Properties of OLS estimator.
- Goodness of fit and interpretation using matrix language.