Econometrics: Lecture 7

Hang Miao

Rutgers University

January 21, 2021

Overview

1. Linear Regression
   - Linear Model
   - OLS estimator
   - Sum of Squares, R^2 and Adjusted $\bar{R}^2$
2. Evaluation for the OLS Estimator
   - Unbiasedness
   - Consistency
   - Efficiency: Gauss-Markov Theorem

Linear Regression

Population Space

Population Space

Suppose we have two population space (X, Y), where X is $(p+1)$-dimensional space while Y is $1-$dimensional space. Elements in space X is denoted as $\boldsymbol{x}' = (1, x_1, x_2, \cdots, x_p)$. Elements in space Y is denoted as y.

Linear Relationship between Two Space

Suppose there is a linear relationship between space (X,Y) such that $y = \boldsymbol{x}'\boldsymbol{\beta} + u$. Where $\boldsymbol{\beta}$ is also a $p-$dimensional column vector denoted as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \cdots, \beta_p)'$. It is the parameter of interest that characterize the linear relationship. u is a $1-$dimensional random variable with $\mathrm{E}[u] = 0$. It characterize the uncertain, stochastic part of the relationship An equivalent expression for this linear relationship could be derived by writing the matrix product of $\boldsymbol{x}'\boldsymbol{\beta}$ explicitly out:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + u$$

Sample Space

Sample Space

Suppose we have n observational samples $(\boldsymbol{x}_i', y_i), i = 1, 2, \cdots, n$ from population space, where $\boldsymbol{x}_i'$ is a $p-$dimensional row vector denoted as $\boldsymbol{x}_i' = (1, x_{i1}, x_{i2}, \cdots, x_{ip})$ while y_i is $1-$dimensional scaler. If we stack up all the sample $\boldsymbol{x}_i'$ and y_i, we have the sample matrix $\boldsymbol{X}$ and sample vector $\boldsymbol{y}$ denoted as follow:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Estimate the Linear Relationship

Same as using sample mean to estimate population mean, we want to construct an estimator by using the observational sample data to estimate the underlying population linear relationship, i.e β

Estimate the Linear Relationship

Ordinary Least Square Estimator

One way to estimate the underlying population linear relationship is to use OLS estimator defined as follow

$$
\begin{aligned}
\hat{\beta}_{ols} &= \text{argmin}_{\beta} \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2 \\
&= \text{argmin}_{\beta} (\boldsymbol{y} - \boldsymbol{X}\beta)' (\boldsymbol{y} - \boldsymbol{X}\beta) \\
&= \text{argmin}_{\beta} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\beta)^2 \\
&= \text{argmin}_{\beta_0, \beta_1, \beta_2, \cdots \beta_p} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_p x_{ip})^2 \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y}
\end{aligned}
$$

The last one is the explicit matrix form for the OLS estimator.

OLS Estimator as the Minimizer of the Mean Square Loss

If $\hat{\boldsymbol{\beta}}_{ols}$ minimize the objective loss function $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$, it must satisfy the first order condition:

$$
\begin{aligned}
0 &= \left.\frac{\partial \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{ols}} = \left.\frac{\partial \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{ols}} \\
&= -2\sum_{i=1}^{n} \boldsymbol{x}_i\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} {}_{ols} = -2\sum_{i=1}^{n} \boldsymbol{x}_i\left(y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}_{ols}}\right) \\
&= -2\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{pmatrix} y_1 - \boldsymbol{x}_1'\hat{\boldsymbol{\beta}}_{ols} \\ y_2 - \boldsymbol{x}_2'\hat{\boldsymbol{\beta}}_{ols} \\ \vdots \\ y_n - \boldsymbol{x}_n'\hat{\boldsymbol{\beta}}_{ols} \end{pmatrix} \\
&= -2\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{ols}\right)
\end{aligned}
$$

OLS Estimator as the Minimizer of the Mean Square Loss

From the First Order Condition we could derive the closed form OLS estimator.

$$
\begin{aligned}
& -2\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{ols}\right) = 0 \\
\Longrightarrow \quad & \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_{ols} = \boldsymbol{X}'\boldsymbol{y} \\
\Longrightarrow \quad & \hat{\boldsymbol{\beta}}_{ols} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}
\end{aligned}
$$

Predicted Response (Fitted Value) $\hat{y}$

After we derived the explicit formula for $\hat{\boldsymbol{\beta}}_{ols}$, given $\boldsymbol{x}$, we could forecast the $\mathrm{E}[y|\boldsymbol{x}] = \boldsymbol{x}'\boldsymbol{\beta} + \mathrm{E}[u|\boldsymbol{x}] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \mathrm{E}[u|\boldsymbol{x}]$ by the predicted y denoted as $\hat{y} = \boldsymbol{x}'\hat{\boldsymbol{\beta}}_{ols} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$. Under the assumption $\mathrm{E}[u|\boldsymbol{x}] = 0$ and $\mathrm{Var}[u|\boldsymbol{x}] = \sigma^2$, we will see later that $\hat{\boldsymbol{\beta}}_{ols} \xrightarrow{p} \boldsymbol{\beta}$. As a result, $\hat{y} \xrightarrow{p} \mathrm{E}[y|\boldsymbol{x}]$

OLS Residual $\hat{u}_i$

The residual $\hat{u}_i$ is the the difference between observed y_i and predicted $\hat{y}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{ols}$ denoted as $\hat{u}_i = y_i - \hat{y}_i$. If we stack up all the n $\hat{u}_i$ together, we have the OLS Residual Vector $\hat{\boldsymbol{u}}$ defined as follow:

$$\hat{\boldsymbol{u}} = \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{pmatrix} = \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{pmatrix} = \begin{pmatrix} y_1 - \boldsymbol{x}_1'\hat{\boldsymbol{\beta}}_{ols} \\ y_2 - \boldsymbol{x}_2'\hat{\boldsymbol{\beta}}_{ols} \\ \vdots \\ y_n - \boldsymbol{x}_n'\hat{\boldsymbol{\beta}}_{ols} \end{pmatrix} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{ols}$$

where $\boldsymbol{X}$, $\boldsymbol{y}$ is defined as same as those from previous slides.

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \ \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ \hat{\boldsymbol{\beta}}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}$$

Properties of FOC

After we defined the OLS residual $\hat{u}$, the FOC of the Mean square loss $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{ols}) = 0$ could be represented by $\boldsymbol{X}'\hat{\boldsymbol{u}} = 0$ which is:

$$\boldsymbol{X}'\hat{\boldsymbol{u}} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \hat{u}_i \\ \sum_{i=1}^n x_{i1}\hat{u}_i \\ \sum_{i=1}^n x_{i2}\hat{u}_i \\ \vdots \\ \sum_{i=1}^n x_{ip}\hat{u}_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

- We have the same result as in the simple regression. In simple regression, there are two equations $\sum_{i=1}^n \hat{u}_i = 0$ and $\sum_{i=1}^n x_i\hat{u}_i$ for two estimator $\hat{\beta}_0$ and $\hat{\beta}_1$.
- For multivariate regression, we have $p+1$ equations $\sum_{i=1}^n \hat{u}_i = 0$, $\sum_{i=1}^n x_{i1}\hat{u}_i = 0$, $\sum_{i=1}^n x_{i2}\hat{u}_i = 0$, $\cdots$, $\sum_{i=1}^n x_{ip}\hat{u}_i = 0$ and $p+1$ estimator $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$

Sum of Squares

Definition (Total Sum of Squares (TSS or SST))

$$\mathrm{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Definition (Explained Sum of Squares (ESS or SSE))

$$\text{ESS} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Definition (Sum of Squared Residuals (SSR or RSS))

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

More Defination

Property of Sum of Squares for OLS

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

Definition (Coefficient of determination: R^2)

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

Definition (Reduced Chi-Squared: s^2)

$$s^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-p-1} = \frac{\text{RSS}}{n-p-1}$$

Definition (Standard Error of the Regression: SER)

$$\text{SER} = \sqrt{s^2} = \sqrt{\frac{\text{RSS}}{n-p-1}}$$

Proof of TSS = ESS + RSS

$$
\begin{aligned}
\text{TSS} &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n} \left[(y_i - \hat{y}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \right] \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= \text{RSS} + \text{ESS} + 2\sum_{i=1}^{n} \hat{u}_i(\hat{y}_i - \bar{y})
\end{aligned}
$$

Proof of $\sum_{i=1}^n \hat{u}_i(\hat{y}_i - \bar{y}) = 0$

$$
\begin{aligned}
\sum_{i=1}^n \hat{u}_i(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n \hat{u}_i\hat{y}_i - \sum_{i=1}^n \hat{u}_i\bar{y} \\
&= \sum_{i=1}^n \hat{u}_i\hat{y}_i - \bar{y}\sum_{i=1}^n \hat{u}_i \\
&= \sum_{i=1}^n \hat{u}_i(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip}) - \bar{y}\sum_{i=1}^n \hat{u}_i \\
&= \hat{\beta}_0 \sum_{i=1}^n \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^n \hat{u}_i x_{i1} + \hat{\beta}_2 \sum_{i=1}^n \hat{u}_i x_{i2} + \cdots \\
&\quad + \hat{\beta}_p \sum_{i=1}^n \hat{u}_i x_{ip} - \bar{y}\sum_{i=1}^n \hat{u}_i = 0
\end{aligned}
$$

The last equation is valid is due to the FOC condition, we have $\sum_{i=1}^n \hat{u}_i = 0$, $\sum_{i=1}^n x_{i1}\hat{u}_i = 0$, $\sum_{i=1}^n x_{i2}\hat{u}_i = 0$, $\cdots$, $\sum_{i=1}^n x_{ip}\hat{u}_i = 0$

Adjusted $\bar{R}^2$

- In multiple regression, the R^2 will be always be increasing if we add more regressors even if these regressors are irrelevant to the response variable y.
- This is because the $\hat{\beta}_{ols}$ is estimated by minimizing SSR. If we add more regressor thus increasing the dimension $\hat{\beta}_{ols}$, the new SSR must be reduced unless the coefficient for the newly added regressor is 0.
- The original choice is feasible while we choose the new estimator, thus the new estimator must have lower loss.
- In response to this issue, the original R^2 is modified. The adjusted $\bar{R}^2$ is defined as follow:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}\frac{\text{RSS}}{\text{TSS}}$$

- $\frac{n-1}{n-p-1}$ is always greater than 1. More regressor would result a higher p, a even larger $\frac{n-1}{n-p-1}$, thus a lower $\bar{R}^2$

Evaluation for the OLS Estimator

Summary of Assumptions so Far

Assumptions For Getting the OLS Estimator

1. Underlying population space (X, Y) has a linear relationship $y = \boldsymbol{x}'\boldsymbol{\beta} + u = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + u$ with $\mathrm{E}[u] = 0$
2. Rank of the matrix $\boldsymbol{X}$ is $p+1$ so that $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exist. This is also called No Multicollinearity

With the above two assumptions, we could get the OLS estimator $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Two questions needs to be answered:

- how well dose this estimated linear relationship $\hat{\boldsymbol{\beta}}_{ols}$ close to the true linear relationship $\boldsymbol{\beta}$?
- dose the linear relationship imply causal effect?

Unbiasedness

In order to answer the first question, we have to propose one more relatively strong assumption:

Assumptions For Getting Unbiased and Consistent OLS estimator

3. Strict Exogeneity Condition: $\mathrm{E}[\boldsymbol{u}|\boldsymbol{X}] = 0$

Under this assumption, the estimated $\hat{\boldsymbol{\beta}}_{ols}$ is unbiased, meaning the $\mathrm{E}[\hat{\boldsymbol{\beta}}_{ols}|\boldsymbol{X}] = \boldsymbol{\beta}$.

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{ols} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
&= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
\mathrm{E}[\hat{\boldsymbol{\beta}}_{ols}|\boldsymbol{X}] &= \mathrm{E}[\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{\beta} + \mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}|\boldsymbol{X}] \\
&= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}[\boldsymbol{u}|\boldsymbol{X}] \\
&= \boldsymbol{\beta}
\end{aligned}$$

Consistency

Suppose we have large number of observation n, from strong law of large number, we have

$$\frac{\boldsymbol{X}'\boldsymbol{X}}{n} \xrightarrow{p} \mathrm{E}[\boldsymbol{x}\boldsymbol{x}'], \qquad \frac{\boldsymbol{X}'\boldsymbol{u}}{n} \xrightarrow{p} \mathrm{E}[\boldsymbol{x}u]$$

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{ols} &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
&= \boldsymbol{\beta} + \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{u}}{n}\right) \\
&\xrightarrow{p} \boldsymbol{\beta} + \mathrm{E}[\boldsymbol{x}\boldsymbol{x}']^{-1}\mathrm{E}[\boldsymbol{x}u]
\end{aligned}$$

Under the Strict Exogeneity Condition: $\mathrm{E}[u|\boldsymbol{x}'] = 0$,

$$\mathrm{E}[\boldsymbol{x}u] = \mathbf{0}$$

the estimated $\hat{\boldsymbol{\beta}}_{ols}$ is consistent: $\hat{\boldsymbol{\beta}}_{ols} \xrightarrow{p} \boldsymbol{\beta}$.

$\boldsymbol{X}'\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{X} &= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} & \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \boldsymbol{x}_3' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'
\end{aligned}
$$

Matrix Expression for $\boldsymbol{x}_i \boldsymbol{x}_i'$

$$
\begin{aligned}
\boldsymbol{x}_i \boldsymbol{x}_i' &= \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} \begin{pmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \end{pmatrix} \\
&= \begin{bmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} & \dots & x_{i1}x_{ip} \\ x_{i2} & x_{i2}x_{i1} & x_{i2}^2 & \dots & x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ip} & x_{ip}x_{i1} & x_{ip}x_{i2} & \dots & x_{ip}^2 \end{bmatrix}
\end{aligned}
$$

Matrix Expression for $X'X$

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{X} &= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \\
&= \sum_{i=1}^{n} \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} & \cdots & x_{i1}x_{ip} \\ x_{i2} & x_{i2}x_{i1} & x_{i2}^2 & \cdots & x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ip} & x_{ip}x_{i1} & x_{ip}x_{i2} & \cdots & x_{ip}^2 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ip} \\ \sum_{i=1}^n x_{i2} & \sum_{i=1}^n x_{i2}x_{i1} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{ip}x_{i1} & \sum_{i=1}^n x_{ip}x_{i2} & \cdots & \sum_{i=1}^n x_{ip}^2 \end{bmatrix}
\end{aligned}
$$

$\frac{\boldsymbol{X}'\boldsymbol{X}}{n} \xrightarrow{p} \mathrm{E}[\boldsymbol{x}\boldsymbol{x}']$

$$
\begin{aligned}
\frac{\boldsymbol{X}'\boldsymbol{X}}{n} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \\
&= \begin{bmatrix}
\frac{\sum_{i=1}^n 1}{n} & \frac{\sum_{i=1}^n x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{ip}}{n} \\
\frac{\sum_{i=1}^n x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i1}^2}{n} & \frac{\sum_{i=1}^n x_{i1}x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{i1}x_{ip}}{n} \\
\frac{\sum_{i=1}^n x_{i2}}{n} & \frac{\sum_{i=1}^n x_{i2}x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i2}^2}{n} & \cdots & \frac{\sum_{i=1}^n x_{i2}x_{ip}}{n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\sum_{i=1}^n x_{ip}}{n} & \frac{\sum_{i=1}^n x_{ip}x_{i1}}{n} & \frac{\sum_{i=1}^n x_{ip}x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{ip}^2}{n}
\end{bmatrix} \\
&\xrightarrow{p} \begin{bmatrix}
1 & \mathrm{E}[x_1] & \mathrm{E}[x_2] & \cdots & \mathrm{E}[x_p] \\
\mathrm{E}[x_1] & \mathrm{E}[x_1^2] & \mathrm{E}[x_1x_2] & \cdots & \mathrm{E}[x_1x_p] \\
\mathrm{E}[x_2] & \mathrm{E}[x_1^2] & \mathrm{E}[x_2^2] & \cdots & \mathrm{E}[x_2x_p] \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathrm{E}[x_p] & \mathrm{E}[x_1x_p] & \mathrm{E}[x_2x_p] & \cdots & \mathrm{E}[x_p^2]
\end{bmatrix} \equiv \mathrm{E}[\boldsymbol{x}\boldsymbol{x}']
\end{aligned}
$$

Matrix Expression for $\boldsymbol{X}'\boldsymbol{u}$

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{u} &= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n u_i \\ \sum_{i=1}^n x_{i1} u_i \\ \sum_{i=1}^n x_{i2} u_i \\ \vdots \\ \sum_{i=1}^n x_{ip} u_i \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} u_i \\ x_{i1} u_i \\ x_{i2} u_i \\ \vdots \\ x_{ip} u_i \end{pmatrix} = \sum_{i=1}^n u_i \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} \\
&= \sum_{i=1}^n u_i \boldsymbol{x}_i
\end{aligned}$$

$$
\begin{aligned}
\frac{\boldsymbol{X}'\boldsymbol{u}}{n} &= \begin{pmatrix} \frac{\sum_{i=1}^{n} u_i}{n} \\ \frac{\sum_{i=1}^{n} x_{i1} u_i}{n} \\ \frac{\sum_{i=1}^{n} x_{i2} u_i}{n} \\ \vdots \\ \frac{\sum_{i=1}^{n} x_{ip} u_i}{n} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} u_i \boldsymbol{x}_i \\
&\xrightarrow{p} \begin{pmatrix} \mathrm{E}[u] \\ \mathrm{E}[x_1 u] \\ \mathrm{E}[x_2 u] \\ \vdots \\ \mathrm{E}[x_p u] \end{pmatrix} \equiv \mathrm{E}[u\boldsymbol{x}] \\
\mathrm{E}[u\boldsymbol{x}] &= \mathrm{E}_{\boldsymbol{x}}\Big[\mathrm{E}_u[u\boldsymbol{x}|\boldsymbol{x}]\Big] = \mathrm{E}_{\boldsymbol{x}}\Big[\boldsymbol{x}\mathrm{E}_u[u|\boldsymbol{x}]\Big] = \mathrm{E}_{\boldsymbol{x}}\Big[\boldsymbol{x}0\Big] = 0
\end{aligned}
$$

Gauss-Markov Theorem

Assumptions For Getting the BLUE Estimator

4. Spherical Error Variance: $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \sigma^2 \boldsymbol{I}_n$
   - Homoskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{X}] = \sigma^2 \quad \forall\, i = 1, 2, \cdots, n$
   - No Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

If the linear model satisfied the assumption 1 - 4, then the β_{ols} would be the best linear unbiased estimator (blue). It is the most efficient estimator in the sense that achieves the least variance.

Central Limit Theorem

Definition (Lindeberg–Levy CLT)

Suppose $\{X_n\}$ is a sequence of i.i.d. random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Or equivalently,

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

Multivariate Central Limit Theorem

Definition (Multivariate Central Limit Theorem)

Suppose $\{\boldsymbol{V}_n\}$ is a sequence of m-dimensional i.i.d. random variables with $\mathrm{E}[\boldsymbol{V}_i] = \boldsymbol{\mu}$ and $\mathrm{Var}[\boldsymbol{V}_i] = \boldsymbol{\Sigma}_{\boldsymbol{V}}$, where $\boldsymbol{\Sigma}_{\boldsymbol{V}}$ is the covariance matrix defined as $\mathrm{E}[(\boldsymbol{V}_i - \boldsymbol{\mu})(\boldsymbol{V}_i - \boldsymbol{\mu})']$. Then as $n \to \infty$,

$$\sqrt{n}(\overline{\boldsymbol{V}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{V}})$$