The Multiple Regression Model in Matrix Form

Hüseyin Taştan

1 Classical Regression Model

The regression model with k parameters can be written for each observation i:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \ldots + \beta_k x_{ik} + u_i, \quad i = 1, \ldots, n$$

where y_i is the dependent variable for observation i, and x_{ij}, $j = 2, 3 \ldots, k$, is ith observed value for jth independent variable. u_i is the random error term. We can also write a system of n equations with k unknowns:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Let us define the following vectors and matrices:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & \dots & x_{1k} \\ 1 & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Then we can write the model as below:

$$\underbrace{\boldsymbol{y}}_{n\times 1} = \underbrace{\boldsymbol{X}}_{n\times k} \underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\boldsymbol{u}}_{n\times 1}$$

Another way of writing this model is based on summoning the ith observed values for independent variables in a $k \times 1$ vector.

$$\boldsymbol{x}_i = \begin{bmatrix} 1 & x_{i2} & x_{i3} & \dots & x_{ik} \end{bmatrix}^\top$$

The regression model in this notational form:

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n$$

Assumptions of Classical Regression Model:

1. The model is linear in parameters: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$
2. rank($\boldsymbol{X}$) = k, (No perfect collinearity, i.e. The columns of X are independent of each other.)
3. $\mathrm{E}\left[\boldsymbol{u}|\boldsymbol{X}\right] = \mathbf{0}_{n\times 1}$, (Zero conditional mean)
4. $\mathrm{V}\left[\boldsymbol{u}|\boldsymbol{X}\right] = \mathrm{E}\left(\boldsymbol{u}\boldsymbol{u}^\top\right) = \sigma^2 \boldsymbol{I}_n$, (Homoscedasticity and no serial correlation)
5. $\boldsymbol{u}|\boldsymbol{X} \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_n\right)$, (the random error is distributed as multivariate normal.)

2 OLS Estimator

Sample Regression Function (SRF):

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}$$

where $\hat{\boldsymbol{\beta}}$ is a $k \times 1$ vector of OLS estimators , $\hat{\boldsymbol{u}}$ is a $n \times 1$ vector of residuals. The OLS method chooses $\hat{\boldsymbol{\beta}}$ vector by minimizing the sum of squared residuals,(SSR):

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{b}} SSR(\boldsymbol{b})$$

The optimization problem is based on finding the $\boldsymbol{b}$ vector by minimizing SSR. The sum of squared residuals in the above models can be stated with different notations:

$$SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}$$

or

$$SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} \right)^2$$

With the first notation, the optimization problem is:

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}}$$

or with the second notation:

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} \right)^2$$

Let us write explicitly the sum of squared residuals to obtain the first order conditions (FOCs) using the first notation.

$$\begin{aligned} SSR(\hat{\boldsymbol{\beta}}) &= \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}} \\ &= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} \end{aligned}$$

While finding the first derivative of SSR with respect to $\hat{\boldsymbol{\beta}}$, note that the second term in the above expression is a linear combination and the third term is in a quadratic form. Generally $\boldsymbol{z}$ is a $k \times 1$ vector, $\boldsymbol{A}$ is a $k \times n$ matrix and $\boldsymbol{B}$ is a $k \times k$ matrix:

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial \boldsymbol{z}} = \boldsymbol{A},$$

and

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial \boldsymbol{z}} = 2\boldsymbol{B}\boldsymbol{z}$$

For example $\boldsymbol{z} = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$ and

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Then

$$\begin{aligned} \boldsymbol{z}^\top \boldsymbol{A} &= \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} z_2 & z_1 + 2z_2 & -2z_1 \end{bmatrix} \end{aligned}$$

when we take the first derivatives with respect to z_1 and z_2

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_1} = \begin{bmatrix} 0 & 1 & -2 \end{bmatrix}$$

and

$$\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_2} = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix}$$

These derivatives can be summoned or arranged in a column vector:

$$\begin{aligned}
\frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial \boldsymbol{z}} = &= \begin{bmatrix} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_1} \\ \frac{\partial(\boldsymbol{z}^\top \boldsymbol{A})}{\partial z_2} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & -2 \\ 1 & 2 & 0 \end{bmatrix} \\
&= \boldsymbol{A}
\end{aligned}$$

Now let's take the derivative of $\boldsymbol{z}^\top \boldsymbol{B} \boldsymbol{z}$ term with respect to the elements of $\boldsymbol{z}$ vector.

$$\begin{aligned}
\boldsymbol{z}^\top \boldsymbol{B} \boldsymbol{z} &= \begin{bmatrix} z_1 & z_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\
&= 2z_1^2 + 2z_1 z_2 + 2z_2^2
\end{aligned}$$

where the first derivatives are:

$$\begin{aligned}
\frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial \boldsymbol{z}} &= \begin{bmatrix} \frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial z_1} \\ \frac{\partial(\boldsymbol{z}^\top \boldsymbol{B}\boldsymbol{z})}{\partial z_2} \end{bmatrix} \\
&= \begin{bmatrix} 4z_1 + 2z_2 \\ 2z_1 + 4z_2 \end{bmatrix} \\
&= 2\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 2\boldsymbol{B}\boldsymbol{z}
\end{aligned}$$

Let's come back to SSR:

$$SSR(\hat{\boldsymbol{\beta}}) = \boldsymbol{y}^\top \boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

where $\hat{\boldsymbol{\beta}} = \boldsymbol{z}$, $\boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{A}$ and $\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{B}$. Then the first order conditions of OLS problem can be written:

$$\frac{\partial SSR(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{0}_k$$

giving the normal equations:

$$\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{y}$$

The second assumption of the classical regression model implies the following rank condition:

$$rank(\boldsymbol{X}) = rank(\boldsymbol{X}^\top \boldsymbol{X}) = k$$

Therefore $\boldsymbol{X}^\top \boldsymbol{X}$ is invertible. When we multiply the both sides of the normal equations by $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$, we can find the vector of OLS estimators:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

Remember that the optimization problem with the second notation can be written as:

$$\min_{\hat{\boldsymbol{\beta}}} SSR(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} \right)^2$$

Here the first order conditions are:

$$\frac{\partial SSR(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2 \sum_{i=1}^{n} \boldsymbol{x}_i \left(y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} \right) = \mathbf{0}_k$$

giving the respective normal equations:

$$\left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n} \boldsymbol{x}_i y_i$$

Hence there are two ways to write the vector of the OLS estimator:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \\ &\equiv (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}\end{aligned}$$

For any i observation, the elements of $\boldsymbol{x}_i \boldsymbol{x}_i^\top$ $(k \times k)$ matrix are as follows:

$$\boldsymbol{x}_i \boldsymbol{x}_i^\top = \begin{bmatrix} 1 \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix} \begin{bmatrix} 1 & x_{i2} & \dots & x_{ik} \end{bmatrix} = \begin{bmatrix} 1 & x_{i2} & x_{i3} & \dots & x_{ik} \\ x_{i2} & x_{i2}^2 & x_{i2}x_{i3} & \dots & x_{i2}x_{ik} \\ x_{i3} & x_{i3}x_{i2} & x_{i3}^2 & \dots & x_{i3}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ik} & x_{ik}x_{i2} & x_{ik}x_{i3} & \dots & x_{ik}^2 \end{bmatrix}$$

Now we have a matrix sequence with n elements. The sum of these matrices:

$$\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top = \begin{bmatrix} n & \sum x_{i2} & \sum x_{i3} & \dots & \sum x_{ik} \\ \sum x_{i2} & \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \dots & \sum x_{i2}x_{ik} \\ \sum x_{i3} & \sum x_{i3}x_{i2} & \sum x_{i3}^2 & \dots & \sum n x_{i3}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ik} & \sum x_{ik}x_{i2} & \sum x_{ik}x_{i3} & \dots & \sum x_{ik}^2 \end{bmatrix} = \boldsymbol{X}^\top \boldsymbol{X}$$

The resulting matrix is a square, symmetric, positive definite matrix. Along with the same motivation:

$$
\begin{aligned}
\sum_{i=1}^{n} \boldsymbol{x}_i y_i &= \sum_{i=1}^{n} \begin{bmatrix} 1 \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix} y_i \\
&= \begin{bmatrix} \sum y_i \\ \sum x_{i2} y_i \\ \vdots \\ \sum x_{ik} y_i \end{bmatrix} \\
&= \boldsymbol{X}^\top \boldsymbol{y}
\end{aligned}
$$

Example 2.1. *Suppose that there is only an intercept in the model with no independent variable:*

$$y_i = \beta_1 + u_i, \quad i = 1, \ldots, n$$

In this case $\boldsymbol{X}$ *matrix is a* $n \times 1$ *vector in which all elements are one. Let's call this vector as* $\boldsymbol{\imath}$*:*

$$\boldsymbol{\imath} = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}^\top = \boldsymbol{X}$$

The OLS estimator of β_1 *can be written as:*

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} = (\boldsymbol{\imath}^\top \boldsymbol{\imath})^{-1} \boldsymbol{\imath}^\top \boldsymbol{y}$$

$$\boldsymbol{\imath}^\top \boldsymbol{\imath} = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = n$$

and

$$\boldsymbol{\imath}^\top \boldsymbol{y} = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} y_i$$

then the OLS estimator is

$$\beta_1 = n^{-1} \sum_{i=1}^{n} y_i \equiv \bar{y}$$

Example 2.2. *Let's consider regression model with a binary (dummy) variable and an intercept:*

$$y_i = \delta_0 + \delta_1 D_i + u_i, \quad i = 1, \dots, n$$

For the sake of simplicity, suppose that the dependent variable consists of just 5 observations and the dummy variable is defined as follows:

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad D_i = \begin{cases} 1, & \textit{if} \;\; y_i \leq 3 \\ 0, & \textit{otherwise} \end{cases} \Leftrightarrow \boldsymbol{D} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

In this case, $\boldsymbol{X}$*:*

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\delta}_0 & \hat{\delta}_1 \end{bmatrix}^\top$ *Now let us find the OLS estimator*

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 5/6 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} \sum y_i \\ \sum y_i D_i \end{bmatrix} = \begin{bmatrix} 15 \\ 6 \end{bmatrix}$$

the vector of OLS estimators:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 5/6 \end{bmatrix} \begin{bmatrix} 15 \\ 6 \end{bmatrix} = \begin{bmatrix} 4.5 \\ -2.5 \end{bmatrix}$$

The estimated regression is

$$\hat{y}_i = 4.5 - 2.5 D_i$$

It is obvious that the intercept term is the mean of the base category (y is greater than 3) in this simple example $((4+5)/2 = 4.5)$*. The mean of the* *other category is 2. And the estimated coefficient of the dummy variable D is the difference of the means of these two categories (-2.5). In our example, we can also find the fitted (estimated) values for the dependent variable and the vector of residuals.*

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4.5 \\ -2.5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 4.5 \\ 4.5 \end{bmatrix},$$

$$\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \\ 4.5 \\ 4.5 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ -0.5 \\ 0.5 \end{bmatrix}$$

Another way of including the dummy variable in the model is to add separately the dummies for each category by excluding the intercept term. Consider the below model in matrix notation form:

$$y_i = \gamma_0 D_{i1} + \delta_0 D_{i2} + u_i, \quad i = 1, \dots, n$$

Now $\boldsymbol{X}$ and $\boldsymbol{X}^\top \boldsymbol{X}$:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} \sum y_i D_{i1} \\ \sum y_i D_{i2} \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

Then the vector of OLS estimators is

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\delta}_0 \end{bmatrix} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 4.5 \end{bmatrix}$$

The estimated regression function is

$$\hat{y}_i = 2D_{i1} + 4.5D_{i2}$$

If we added an intercept term in this model, we would be in a dummy variable trap . In this case, the model and corresponding data matrix are

$$y_i = \beta_0 + \gamma_0 D_{i1} + \delta_0 D_{i2} + u_i, \quad i = 1, \ldots, n$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Clearly seen that the sum of the second and third column of $\boldsymbol{X}$ *matrix is the first column of this matrix. The classical assumption about Rank condition is violated:* $rank(\boldsymbol{X}) < 3$. *Another way of detecting this violation is to calculate cross product matrix:*

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad \mid \boldsymbol{X}^\top \boldsymbol{X} \mid = 0$$

The sum of the second and third column of this matrix is its first column and the sum of the second and third row of this matrix is its first row. The model can not be estimated in this form, because this matrix is singular and its determinant is 0. The OLS estimator is undefined. So one of the columns in this matrix is redundant, it should be excluded.

Example 2.3. *Simple regression model: intercept term + one independent variable:*

$$y_i = \beta_1 + \beta_2 x_i + u_i, \quad i = 1, \ldots, n$$

In this model, the cross product matrix and its inverse are as below:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix},$$

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \frac{1}{n \sum x_i^2 - \left(\sum x_i\right)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix}$$

The OLS estimator:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\
&= \frac{1}{n\sum x_i^2 - \left(\sum x_i\right)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\sum x_i^2 \sum y_i - \sum x_i \sum y_i x_i}{n\sum x_i^2 - (\sum x_i)^2} \\ \dfrac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \end{bmatrix}
\end{aligned}
$$

or

$$\hat{\beta}_1 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum y_i x_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\hat{\beta}_2 = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

The numerator and denominator of the slope parameter can be stated in a *simpler fashion. Remember that*

$$\begin{aligned}
\sum (y_i - \bar{y})(x_i - \bar{x}) &= \sum (y_i - \bar{y}) x_i \\
&= \sum y_i x_i - \bar{y} \sum x_i \\
&= \sum y_i x_i - \frac{1}{n} \sum y_i \sum x_i,
\end{aligned}$$

$$\begin{aligned}
\sum (x_i - \bar{x})^2 &= \sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2 \\
&= \sum x_i^2 - n\bar{x}^2 \\
&= \frac{n \sum x_i^2 - (\sum x_i)^2}{n}
\end{aligned}$$

Based on this, the simplified expression is

$$\begin{aligned}
\frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} &= \frac{\sum y_i x_i - \frac{1}{n} \sum y_i \sum x_i}{\frac{n \sum x_i^2 - (\sum x_i)^2}{n}} \\
&= \frac{n \sum y_i x_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \\
&= \hat{\beta}_2
\end{aligned}$$

3 Unbiasedness of the OLS estimator

Under the three classical assumptions, the OLS estimator $\hat{\boldsymbol{\beta}}$ is unbiased of $\boldsymbol{\beta}$. The proof is very simple. In the formula of the OLS estimator, writing the population regression function instead of the dependent variable and then taking the expectation conditional on $\boldsymbol{X}$ would be enough to prove the unbiasedness of the OLS estimator.

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u} \\
&= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] &= \boldsymbol{\beta} + \mathrm{E}\left[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{u}|\boldsymbol{X}\right] \\
&= \boldsymbol{\beta} + \boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \underbrace{\mathrm{E}\left[\boldsymbol{u}|\boldsymbol{X}\right]}_{=\mathbf{0}} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

The OLS estimator is determined as the linear combination of random error terms about the unknown true value. On average, the random vector of estimator is equal to the true parameter vector.

4 Variance-Covariance Matrix of the OLS Estimator

The fourth classical assumption is about the covariance matrix of the random error term:

$$\mathrm{V}(\boldsymbol{u}|\boldsymbol{X}) = \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n$$

This assumption implies that the error terms are uncorrelated and homoscedastic. Let us take a closer look at the covariance matrix. Note that all expectations are conditional on $\boldsymbol{X}$:

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) &= \mathrm{E}\left\{ \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \right\} \\
&= \mathrm{E}\left\{ \begin{bmatrix} u_1^2 & u_1u_2 & \dots & u_1u_n \\ u_2u_1 & u_2^2 & \dots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_1 & u_nu_2 & \dots & u_n^2 \end{bmatrix} \right\} \\
&= \begin{bmatrix} \mathrm{E}(u_1^2) & \mathrm{E}(u_1u_2) & \dots & \mathrm{E}(u_1u_n) \\ \mathrm{E}(u_2u_1) & \mathrm{E}(u_2^2) & \dots & \mathrm{E}(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(u_nu_1) & \mathrm{E}(u_nu_2) & \dots & \mathrm{E}(u_n^2) \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \\
&= \sigma^2 \boldsymbol{I}_n
\end{aligned}
$$

Under this assumption, let us derive the variance-covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$:

$$
\begin{aligned}
\mathrm{V}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})\right)^\top\right] \\
&= \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^\top\right] \\
&= \mathrm{E}\left[\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\right)\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\right)^\top\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{u}\boldsymbol{u}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\right] \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\underbrace{\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top)}_{\sigma^2\boldsymbol{I}_n}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}^\top\boldsymbol{I}_n\boldsymbol{X}}_{\boldsymbol{X}^\top\boldsymbol{X}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2\underbrace{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}}_{\boldsymbol{I}_k}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}
\end{aligned}
$$

In this formula the error variance σ^2 is unknown, hence it should be estimated by using the given data. An unbiased estimator of σ^2 can be found as follows:

$$\begin{aligned}
s^2 &= \frac{1}{n-k} SSR \\
&= \frac{1}{n-k} \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}} \\
&= \frac{1}{n-k} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})
\end{aligned}$$

It can be shown that $\mathrm{E}(s^2) = \sigma^2$. Therefore the variance-covariance matrix of the OLS estimator is written below.

$$\widehat{\mathrm{V}(\hat{\boldsymbol{\beta}})} = s^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

This matrix is a dimension of $k \times k$, symmetric, square and positive definite. The diagonal elements shows the variances of the OLS estimators, and the off-diagonal elements shows the covariances.

GAUSS-MARKOV THEOREM: *Under the first four classical assumptions, the OLS estimators are the best (minimum variance) linear unbiased estimators among all the competing linear unbiased estimators.* (**Best Linear Unbiased Estimator - BLUE**).

If any of the first four assumptions fails, then the Gauss-Markov theorem no longer holds. For example, if heteroscedasticity exists, (the violation of the forth assumption) the variance-covariance matrix that we have just derived will not be valid and the OLS estimators are not BLUE any more. For the Gauss-Markov theorem to be valid, the random error terms does not need to follow the normal distribution (the fifth classical assumption).The classical assumption of the normally distributed random error terms is required only to derive the sampling distributions of the OLS estimators and to conduct t and F tests in small (finite) samples.

Example 4.1. *Now let us give a numerical example by using a simple regression model. Our* ***data set*** *is as follows:*

$$
\boldsymbol{y} = \begin{bmatrix} 3 \\ 8 \\ 18 \\ 3 \\ 2 \\ 6 \\ 6 \\ 11 \\ 6 \\ 6 \\ 1 \\ 16 \\ 10 \\ 20 \\ 12 \\ 10 \\ 18 \\ 10 \\ 5 \\ 15 \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 8 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 6 \\ 1 & 3 \\ 1 & 2 \\ 1 & 0 \\ 1 & 7 \\ 1 & 4 \\ 1 & 9 \\ 1 & 5 \\ 1 & 4 \\ 1 & 8 \\ 1 & 5 \\ 1 & 2 \\ 1 & 7 \end{bmatrix}
$$

To find the OLS estimator, firstly calculate the required quantities:

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 20 & 80 \\ 80 & 468 \end{bmatrix}, \quad (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} 186 \\ 1042 \end{bmatrix}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\ &= \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix} \begin{bmatrix} 186 \\ 1042 \end{bmatrix} \\ \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \begin{bmatrix} 1.2459 \\ 2.0135 \end{bmatrix} \end{aligned}$$

The sample regression function:

$$\hat{y} = 1.2459 + 2.0135x$$

$$SSR = \hat{\boldsymbol{u}}^\top \hat{\boldsymbol{u}} = 20.173, \quad s^2 = \frac{1}{18} SSR = 1.1207$$

The variance-covariance matrix:

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = s^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = 1.1207 \begin{bmatrix} 0.1581 & -0.0270 \\ -0.0270 & 0.0068 \end{bmatrix} = \begin{bmatrix} 0.1772 & -0.0303 \\ -0.0303 & 0.0076 \end{bmatrix}$$

Then,

$$\mathrm{V}(\hat{\beta}_1) = 0.1772, \quad \mathrm{V}(\hat{\beta}_2) = 0.0076, \quad Cov(\hat{\beta}_1, \hat{\beta}_2) = -0.0303$$

Standard Errors:

$$se(\hat{\beta}_1) = \sqrt{0.1772} = 0.4209$$

$$se(\hat{\beta}_2) = \sqrt{0.0076} = 0.087$$

t-ratios:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{1.2459}{0.4209} = 2.9599$$

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{2.0135}{0.087} = 23.1386$$