3. Multiple Regression Analysis

The general linear regression with k explanatory variables is just an extension of the simple regression as follows

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i. \tag{1}$$

Because

$$\frac{\partial y_i}{\partial x_{ij}} = \beta_j \tag{2}$$

$j = 1, \ldots, k$, coefficient β_j indicates the marginal effect of variable x_j, and indicates the amount y is expected to change as x_j changes by one unit and other variables are kept constant (ceteris paribus).

The multiple regression opens up several additional options to enrich analysis and make modeling more realistic compared to the simple regression.

Example 3.1: Consider the hourly wage example. Enhance the model as

$$\log(w) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \tag{3}$$

where w = average hourly earnings, x_1 = years of education (educ), x_2 = years of labor market experience (exper), and x_3 = years with the current employer (tenure).

Dependent Variable: LOG(WAGE)
Method: Least Squares
Date: 08/21/12 Time: 09:16
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.284360	0.104190	2.729230	0.0066
EDUC	0.092029	0.007330	12.55525	0.0000
EXPER	0.004121	0.001723	2.391437	0.0171
TENURE	0.022067	0.003094	7.133070	0.0000

R-squared	0.316013	Mean dependent var	1.623268
Adjusted R-squared	0.312082	S.D. dependent var	0.531538
S.E. of regression	0.440862	Akaike info criterion	1.207406
Sum squared resid	101.4556	Schwarz criterion	1.239842
Log likelihood	-313.5478	Hannan-Quinn criter.	1.220106
F-statistic	80.39092	Durbin-Watson stat	1.768805
Prob(F-statistic)	0.000000		

For example the coefficient 0.092 means that, holding exper and tenure fixed, another year of education is predicted to increase wage by approximately 9.2%. Staying another year at the same firm (educ fixed, Δexper=Δtenure=1) is expected to result in a salary increase by approximately $0.4\% + 2.2\% = 2.6\%$.

Example 3.2: Consider the consumption function $C = f(Y)$, where Y is income. Suppose the assumption is that as incomes grow the marginal propensity to consume decreases.

In simple regression we could try to fit a level-log model or log-log model.

One possibility also could be

$$\beta_1 = \beta_{1l} + \beta_{1q} Y,$$

where according to our hypothesis $\beta_{1q} < 0$. Thus the consumption function becomes

$$\begin{aligned} C &= \beta_0 + (\beta_{1l} + \beta_{1q} Y)Y + u \\ &= \beta_0 + \beta_{1l} Y + \beta_{1q} Y^2 + u \end{aligned}$$

This is a multiple regression model with $x_1 = Y$ and $x_2 = Y^2$.

This simple example demonstrates that we can meaningfully enrich simple regression analysis (even though we have essentially only two variables, C and Y) and at the same time get a meaningful interpretation to the above polynomial model.

The response of C to a one unit change in Y is now

$$\frac{\partial C}{\partial Y} = \beta_{1l} + 2\beta_{1q} Y.$$

Estimation

In order to estimate the model we replace the classical assumption 3 as

3. None of the independent variables is constant, and no observation vector of any independent variable can be written as a linear combination of the observation vectors of any other independent variables.

The estimation method again is the OLS, which produces estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ by minimizing

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 \tag{4}$$

with respect to the parameters.

Again the first order solution is to set the $(k+1)$ partial derivatives equal to zero.

The solution is straightforward although the explicit form of the estimators become complicated.

Matrix form

Using matrix algebra simplifies considerably the notations in multiple regression.

Denote the observation vector on y as

$$\mathbf{y} = (y_1, \ldots, y_n)', \tag{5}$$

where the prime denotes transposition.

In the same manner denote the data matrix on x-variables enhanced with ones in the first column as an $n \times (k+1)$ matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \tag{6}$$

where $k < n$ (the number of observations, n, is larger than the number of x-variables, k).

Then we can present the whole set of regression equations for the sample

$$
(7)\qquad
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{1k} + u_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \cdots + \beta_k x_{2k} + u_2 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_k x_{nk} + u_n
\end{aligned}
$$

in the matrix form as

(8)

$$
\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}
=
\begin{pmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1k} \\
1 & x_{21} & x_{22} & \cdots & x_{2k} \\
\vdots & \vdots & \vdots & & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{nk}
\end{pmatrix}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}
+
\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}
$$

or shortly

$$
(9)\qquad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u},
$$

where

$$
\mathbf{b} = (\beta_0, \beta_1, \ldots, \beta_k)'
$$

is the parameter vector and

$$
\mathbf{u} = (u_1, u_2, \ldots, u_n)'
$$

is the error vector.

The *normal equations* for the first order conditions of the minimization of (4) in matrix form are simply

$$\mathbf{X}'\mathbf{X}\widehat{\mathbf{b}} = \mathbf{X}'\mathbf{y} \tag{10}$$

which gives the explicit solution for the OLS estimator of $\mathbf{b}$ as

$$\widehat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{11}$$

where $\widehat{\mathbf{b}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k)'$ and existence of $(\mathbf{X}'\mathbf{X})^{-1}$ is granted by assumption 3.

The fitted model is

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_k x_{ik}, \tag{12}$$

$i = 1, \ldots, n$.

Remark 3.1:

Single and multiple regression do in general not produce the same parameter estimates on the same independent variables. For example, if you fit the simple regression $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, where $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are the OLS estimators, and fit a multiple regression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ then it turns out that $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$, where δ_1 is the slope coefficient from regressing x_2 on x_1. This implies that $\tilde{\beta}_1 \neq \hat{\beta}_1$ unless $\hat{\beta}_2 = 0$, or x_1 and x_2 are uncorrelated.

Goodness-of-Fit

Again in the same manner as with the simple regression we have

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{13}$$

or

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}, \tag{14}$$

where

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_k x_{ik}. \tag{15}$$

R-square:

R-square denotes again the sample variation in the fitted $\widehat{y}_i$'s as a proportion of the sample variation in the original y_i's:

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}. \tag{16}$$

Again as in the case of the simple regression, R^2 can be shown to be the squared correlation coefficient between the actual y_i and fitted $\widehat{y}_i$. This correlation is called the *multiple correlation*

$$R = \frac{\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum(y_i - \bar{y})^2}\sqrt{\sum(\hat{y}_i - \bar{\hat{y}})^2}}. \tag{17}$$

Remark 3.2: $\bar{\hat{y}} = \bar{y}$.

Remark 3.3: R^2 never decreases when an explanatory variable is added to the model.

Adjusted R-square:

$$(18)\quad \bar{R}^2 = 1 - \frac{s_u^2}{s_y^2} = 1 - \frac{n-1}{n-k-1}(1-R^2),$$

where

(19)

$$s_u^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2$$

is an estimator of the error variance $\sigma_u^2 = \mathbb{V}\mathrm{ar}[u_i]$. The square root of (19) is the so called *standard error of the regression*.

3.3 Expected values of the OLS estimators

Given observation tuples $(x_{i1}, x_{i1}, \ldots, x_{ik}, y_i)$, $i = 1, \ldots, n$ the classical assumptions read now:

Assumptions (classical assumptions):

1. $y = \beta_0 + \sum_{i=1}^{k} \beta_i x + u$ in the population.
2. $\{(x_{i1}, x_{i1}, \ldots, x_{ik}, y_i)\}$ is a random sample of the model above, implying uncorrelated residuals: $\mathbb{C}\text{ov}(u_i, u_j) = 0$ for all $i \neq j$.
3. All independent variables including the vector of constants are linearly independent, implying that $(\mathbf{X}'\mathbf{X})^{-1}$ exists.
4. $\mathbb{E}[u|x_1, \ldots, x_k] = 0$ implying $\mathbb{E}[u] = 0$ and $\mathbb{C}\text{ov}(u, x_1) = \ldots = \mathbb{C}\text{ov}(u, x_k) = 0$.
5. $\mathbb{V}\text{ar}[u|x_1, \ldots, x_k] = \sigma^2$ implying $\mathbb{V}\text{ar}[u] = \sigma^2$.

Under these assumptions we can show that the estimators of the regression coefficients are unbiased. That is

Theorem 3.1: Given observations on the x-variables

(20) $$\mathbb{E}[\hat{\beta}_j] = \beta j$$

Using matrix notations, the proof of Theorem 3.1 is pretty straightforward. To do this write

$$\begin{aligned}
\widehat{\mathbf{b}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\mathbf{b}+\mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \mathbf{b}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.
\end{aligned} \tag{21}$$

Given $\mathbf{X}$, the expected value of $\widehat{\mathbf{b}}$ is

$$\begin{aligned}
\mathbb{E}[\widehat{\mathbf{b}}] &= \mathbf{b}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{u}] \\
&= \mathbf{b}
\end{aligned} \tag{22}$$

because by Assumption 4 $\mathbb{E}[\mathbf{u}]=\mathbf{0}$. Thus the OLS estimators of the regression coefficients are unbiased.

Remark 3.4: If $\mathbf{z}=(z_1,\ldots,z_k)'$ is a random vector, then $\mathbb{E}[\mathbf{z}]=(\mathbb{E}[z_1],\ldots,\mathbb{E}[z_n])'$. That is the expected value of a vector is a vector whose components are the individual expected values.

Irrelevant variables in a regression

Suppose the correct model is

$$y = \beta_0 + \beta_1 x_1 + u \tag{23}$$

but we estimate the model as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u. \tag{24}$$

Thus, $\beta_2 = 0$ in reality. The OLS estimation results yield

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 \tag{25}$$

By Theorem 3.1 $\mathbb{E}[\widehat{\beta}_j] = \beta_j$, thus in particular $\mathbb{E}[\widehat{\beta}_2] = \beta_2 = 0$, implying that inclusion of extra variables to a regression does not bias the results.

However, as will be seen later, they decrease accuracy of estimation by increasing variance of the OLS estimates.

Omitted variable bias

Suppose now as an example that the correct model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \tag{26}$$

but we misspecify the model as

$$y = \beta_0 + \beta_1 x_1 + v, \tag{27}$$

where the omitted variable is embedded into the residual term $v = \beta_2 x_2 + u$.

OLS estimator for β_1 for specification (27)is

$$\tilde{\beta}_1 = \frac{\sum(x_{i1} - \bar{x}_1)y_i}{\sum(x_{i1} - \bar{x}_1)^2}. \tag{28}$$

From Equation (2.37) we have

$$\tilde{\beta}_1 = \beta_1 + \sum_{i=1}^{n} a_i v_i, \tag{29}$$

where

$$a_i = \frac{(x_{i1} - \bar{x}_1)}{\sum(x_{i1} - \bar{x}_1)^2}. \tag{30}$$

Thus because $\mathbb{E}[v_i] = \mathbb{E}[\beta_2 x_{i2} + u_i] = \beta_2 x_{i2}$

$$
\begin{aligned}
\mathbb{E}[\tilde{\beta}_1] &= \beta_1 + \sum a_i \mathbb{E}[v_i] \\
&= \beta_1 + \sum a_i \beta_2 x_{i2} \\
&= \beta_1 + \beta_2 \underbrace{\frac{\sum (x_{i1} - \bar{x}_1) x_{i2}}{\sum (x_{i1} - \bar{x}_1)^2}}_{\tilde{\delta}_1}
\end{aligned}
\tag{31}
$$

i.e.,

$$
\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2 \tilde{\delta}_1, \tag{32}
$$

where $\tilde{\delta}_1$ is the slope coefficient of regressing x_2 upon x_1, implying that $\tilde{\beta}_1$ is biased for β_1 unless $x_1$and x_2 are uncorrelated (or $\beta_2 = 0$). This is called the omitted variable bias.

The direction of the omitted variable bias is as follows:

	$\mathbb{C}\text{orr}(x_1, x_2) > 0$	$\mathbb{C}\text{orr}(x_1, x_2) < 0$
$\beta_2 > 0$	positive bias	negative bias
$\beta_2 < 0$	negative bias	positive bias

3.4 The variance of OLS estimators

Write the regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i \tag{33}$$

$i = 1, \ldots, n$ in the matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}. \tag{34}$$

Then we can write the OLS estimators compactly as

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{35}$$

Under the classical assumption 1–5, and assuming $\mathbf{X}$ fixed we can show that the variance-covariance matrix $\mathbf{b}$ is

$$\mathbb{C}\text{ov}[\hat{\mathbf{b}}] = (\mathbf{X}'\mathbf{X})^{-1}\sigma_u^2. \tag{36}$$

Variances of the individual coefficients are obtained form the main diagonal of the matrix, and can be shown to be of the form

$$\mathbb{V}\text{ar}[\hat{\beta}_j] = \frac{\sigma_u^2}{(1 - R_j^2)\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}, \tag{37}$$

$j = 1, \ldots, k$, where R_j^2 is the R-square when regressing x_j on the other explanatory variables and the constant term.

Multicollinearity

In terms of linear algebra, we say that vectors

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$$

are linearly independent if

$$a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k = 0$$

holds only if

$$a_1 = \cdots = a_k = 0.$$

Otherwise $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent. In such a case some $a_\ell \neq 0$ and we can write

$$\mathbf{x}_\ell = c_1\mathbf{x}_1 + \cdots + c_{\ell-1}\mathbf{x}_{\ell-1} + c_{\ell+1}\mathbf{x}_{\ell+1} + \cdots + c_k\mathbf{x}_k,$$

where $c_j = -a_j/a_\ell$ that is $\mathbf{x}_\ell$ can be represented as a linear combination of the other variables.

In statistics the multiple correlation measures the degree of linear dependence. If the variables are perfectly linearly dependent. That is, if for example, x_j is a linear combination of other variables, the multiple correlation $R_j = 1$.

A perfect linear dependence is rare between random variables. However, particularly between macro economic variables dependencies are often high.

From the variance equation (37)

$$\mathbb{V}\text{ar}[\widehat{\beta}_j] = \frac{\sigma_u^2}{(1-R_j^2)\sum_{i=1}^n (x_{ij}-\bar{x}_j)^2},$$

we see that $\mathbb{V}\text{ar}[\widehat{\beta}_j] \to \infty$ as $R_j^2 \to 1$. That is, the more the explanatory variables are linearly dependent the larger the variance becomes. This implies that the coefficient estimates become increasingly instable.

High (but not perfect) correlation between two or more explanatory variables is called <u>multicollinearity</u>.

Symptoms of multicollinearity:

(1) High correlations between explanatory variables.

(2) R^2 is relatively high, but the coefficient estimates tend to be insignificant (see the section of hypothesis testing)

(3) Some of the coefficients are of wrong sign and some coefficients are at the same time unreasonably large.

(4) Coefficient estimates change much from one model alternative to another.

Example 3.3: Variable E_t denotes cost expenditures in a sample of Toyota Mark II cars at time point t, M_t denotes the milage and A_t age.

Consider model alternatives:

Model A: $E_t = \alpha_0 + \alpha_1 A_t + u_{1t}$

Model B: $E_t = \beta_0 + \beta_1 M_t + u_{2t}$

Model C: $E_t = \gamma_0 + \gamma_1 M_t + \gamma_2 A_t + u_{3t}$

Estimation results: (t-values in parentheses)

Variable	Malli A	Malli B	Malli C
Constant	-626.24	-796.07	7.29
	(-5.98)	(-5.91)	(0.06)
Age	7.35		27.58
	(22.16)		(9.58)
Miles		53.45	-151.15
		(18.27)	(-7.06)
df	55	55	54
$\bar{R}^2$	0.897	0.856	0.946
$\hat{\sigma}$	368.6	437.0	268.3

Findings:

Apriori, coefficients α_1, β_1, γ_1, and γ_2 should be positive. However, $\hat{\gamma}_2 = -151.15$ (!!?), but $\hat{\beta}_1 = 53.45$. Correlation $r_{M,E} = 0.996$!

Remedies:

In the collinearity problem the question is the there is not enough information to reliably identify each variables contribution as an explanatory variable in the model. Thus in order to alleviate the problem:

(1) Use non-sample information if available to impose restrictions between coefficients.

(2) Increase the sample size if possible.

(3) Drop the most collinear (on the base of R_j^2) variables.

(4) If a linear combination (usually a sum) of the most collinear variables is meaningful, replace the collinear variables by the linear combination.

Variances of misspecified models

Consider again as in (26) the regression model

(38) $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Suppose the following models are estimated by OLS

(39) $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

and

(40) $$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1.$$

Then by (37)

(41) $$\mathbb{V}\text{ar}[\hat{\beta}_1] = \frac{\sigma_u^2}{(1 - r_{12}^2)\sum(x_{i1} - \bar{x}_1)^2}$$

and in analogy to (2.40)

(42)
$$\mathbb{V}\text{ar}[\tilde{\beta}_1] = \frac{\mathbb{V}\text{ar}[\beta_2 x_2 + u]}{\sum(x_{i1} - \bar{x}_1)^2} = \frac{\sigma_u^2}{\sum(x_{i1} - \bar{x}_1)^2},$$

where r_{12} is the sample correlation between x_1 and x_2.

Thus $\mathbb{V}\text{ar}[\tilde{\beta}_1] \leq \mathbb{V}\text{ar}[\hat{\beta}_1]$, and the inequality is strict if $r_{12} \neq 0$.

In summary (assuming $r_{12} \neq 0$):

(1) If $\beta_2 \neq 0$, then $\tilde{\beta}_1$ is biased, $\hat{\beta}_1$ is unbiased, and $\mathbb{V}\text{ar}[\tilde{\beta}_1] < \mathbb{V}\text{ar}[\hat{\beta}_1]$

(2) If $\beta_2 = 0$, then both $\tilde{\beta}_1$ and $\hat{\beta}_1$ are unbiased, but $\mathbb{V}\text{ar}[\tilde{\beta}_1] < \mathbb{V}\text{ar}[\hat{\beta}_1]$

Estimating error variance σ_u^2

An unbiased estimator of the error variance $\mathbb{V}\mathrm{ar}[u] = \sigma_u^2$ is

$$\hat{\sigma}_u^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} \hat{u}_i^2, \tag{43}$$

where

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k u_{ik}. \tag{44}$$

The term $n-k-1$ in (42) is the degrees of freedom (*df*).

It can be shown that

$$\mathbb{E}[\hat{\sigma}_u^2] = \sigma_u^2, \tag{45}$$

i.e., $\hat{\sigma}_u^2$ is unbiased estimator of σ_u^2.

$\hat{\sigma}_u$ is called the *standard error of the regression*.

Standard errors of $\widehat{\beta}_k$

Standard deviation of $\widehat{\beta}_j$ is the square root of (37), i.e.

$$\sigma_{\widehat{\beta}_j} = \sqrt{\mathbb{V}\text{ar}[\beta_j]} = \frac{\sigma_u}{\sqrt{(1 - R_j^2)\sum(x_{ij} - \bar{x}_j)^2}} \tag{46}$$

Substituting σ_u by its estimate $\hat{\sigma}_u = \sqrt{\hat{\sigma}_u^2}$ gives the <u>standard error of $\widehat{\beta}_j$</u>

$$\text{se}(\widehat{\beta}_j) = \frac{\hat{\sigma}_u}{\sqrt{(1 - R_j^2)\sum(x_{ij} - \bar{x}_j)^2}}. \tag{47}$$

3.5 The Gauss-Markov Theorem

Theorem 3.2: Under the classical assumptions 1–5 $\widehat{\beta}_0, \widehat{\beta}_1, \ldots \widehat{\beta}_k$ are best linear unbiased estimators (BLUEs) of $\beta_0, \beta_1, \ldots \beta_k$, respectively.

BLUE:

Best: The variance of the OLS estimator is smallest among all linear unbiased estimators of β_j

Linear: $\widehat{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$.

Unbiased: $\mathbb{E}[\widehat{\beta}_j] = \beta_j$.