ECO 344 - Applied Econometrics I
Chapter 3

Muhammad Salman Khalid

School of Economics & Social Sciences

September 17, 2026



Outline

1. Multiple Regression Analysis
2. Mechanics of Multiple Regression Model
3. Comparing Simple & Multiple Regression
4. The Expected Value of OLS Estimators
5. The Variance of OLS Estimators
6. Gauss Markov Theorem
7. Examples

Multiple Regression Analysis

Introduction

- It is difficult to establish ceteris paribus effect when we use the simple regression model.

Introduction

- It is difficult to establish ceteris paribus effect when we use the simple regression model.
- Remember SLR IV;

$$E[\mu|x] = E[\mu] = 0$$

Can we establish this using only one explanatory variable?

Introduction

- It is difficult to establish ceteris paribus effect when we use the simple regression model.
- Remember SLR IV;

$$E[\mu|x] = E[\mu] = 0$$

  Can we establish this using only one explanatory variable?
- One solution is to use multiple regression model to explicitly control for variables that are correlated with both x and y.

Introduction

- It is difficult to establish ceteris paribus effect when we use the simple regression model.
- Remember SLR IV;

$$E[\mu|x] = E[\mu] = 0$$

  Can we establish this using only one explanatory variable?
- One solution is to use multiple regression model to explicitly control for variables that are correlated with both x and y.
- Recall our wage and education example;

$$wage_i = \beta_0 + \beta_1 \times education_i + \mu$$

  We have talked about; How $E[\mu|education] \neq E[\mu]$

Introduction

- It is difficult to establish ceteris paribus effect when we use the simple regression model.
- Remember SLR IV;

$$E[\mu|x] = E[\mu] = 0$$

  Can we establish this using only one explanatory variable?
- One solution is to use multiple regression model to explicitly control for variables that are correlated with both x and y.
- Recall our wage and education example;

$$wage_i = \beta_0 + \beta_1 \times education_i + \mu$$

  We have talked about; How $E[\mu|education] \neq E[\mu]$
- Now we can take out the ability from the error term and explicitly control for it.
  But How do we measure ability? - measurement problem

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

- Does this satisfy our zero conditional mean?

$$E[\mu|education, SAT] = E[\mu]$$

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

- Does this satisfy our zero conditional mean?

$$E[\mu|education, SAT] = E[\mu]$$

  - It is important to note here that there is no explicit condition that two independent variable cannot be correlated with each other.

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

- Does this satisfy our zero conditional mean?

$$E[\mu|education, SAT] = E[\mu]$$

  - It is important to note here that there is no explicit condition that two independent variable cannot be correlated with each other.
  - However, the extent of this correlation can cause some problems. - (will discuss this later)

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

- Does this satisfy our zero conditional mean?

$$E[\mu|education, SAT] = E[\mu]$$

  - It is important to note here that there is no explicit condition that two independent variable cannot be correlated with each other.
  - However, the extent of this correlation can cause some problems. - (will discuss this later)
  - But there can be other factors that can still simultaneously affect education or ability and wage.

Introduction

- Now, let's assume that we can capture the ability using SAT Scores. - (Fair Assumption?)
- To fulfill zero conditional mean, we can rewrite our regression model as;

$$wage_i = \beta_0 + \beta_1 education_i + \beta_2 SAT + \mu$$

- Does this satisfy our zero conditional mean?

$$E[\mu|education, SAT] = E[\mu]$$

  - It is important to note here that there is no explicit condition that two independent variable cannot be correlated with each other.
  - However, the extent of this correlation can cause some problems. - (will discuss this later)
  - But there can be other factors that can still simultaneously affect education or ability and wage.
  - What to do?

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- However, the interpretation of our OLS estimators will be changed a bit.

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- However, the interpretation of our OLS estimators will be changed a bit.
- The intercept parameter β_0 will be the average value of y when all the independent variables are zero.

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- However, the interpretation of our OLS estimators will be changed a bit.
- The intercept parameter β_0 will be the average value of y when all the independent variables are zero.
- The slope parameter - One unit change in x_i will lead to β_i unit change in y keeping all other factors (including other independent variables) constant.

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- However, the interpretation of our OLS estimators will be changed a bit.
- The intercept parameter β_0 will be the average value of y when all the independent variables are zero.
- The slope parameter - One unit change in x_i will lead to β_i unit change in y keeping all other factors (including other independent variables) constant.
- When will be $\Delta y = \beta_i \Delta x$ in the following model?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

Introduction

- Essentially the solution is to include all such variable in our model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- However, the interpretation of our OLS estimators will be changed a bit.
- The intercept parameter β_0 will be the average value of y when all the independent variables are zero.
- The slope parameter - One unit change in x_i will lead to β_i unit change in y keeping all other factors (including other independent variables) constant.
- When will be $\Delta y = \beta_i \Delta x$ in the following model?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- When we treat all other x as constant ($\partial x_k = 0 for k \neq j$), only then

$$\partial y = \beta_j \partial x_j$$

Mechanics of Multiple Regression Model

Derivation of OLS for Multiple Regression Model

- Let we have k independent variable such that $x_1, x_2, x_3, ..., x_k$. Our multiple regression model would be;

Derivation of OLS for Multiple Regression Model

- Let we have k independent variable such that $x_1, x_2, x_3, ..., x_k$. Our multiple regression model would be;
- $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + \mu$$

  We have $k + 1$ estimators to estimate.

Derivation of OLS for Multiple Regression Model

- Let we have k independent variable such that $x_1, x_2, x_3, ..., x_k$. Our multiple regression model would be;
- $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + \mu$$

  We have $k + 1$ estimators to estimate.
- we will use the same method to estimate our OLS estimators;

$$Minimize : \sum_{i=1}^{n} \hat{Error}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik})$$

As $\hat{Error} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \hat{\beta}_3 x_3 - ..... - \hat{\beta}_k x_k$

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$
- F.O.C (First Order Conditions);

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$
- F.O.C (First Order Conditions);
- $$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$
- F.O.C (First Order Conditions);
- $$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- $$\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$
- F.O.C (First Order Conditions);
- $$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- $$\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- $$\sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$

Derivation of OLS for Multiple Regression Model

- Taking the derivatives w.r.t to each $\hat{\beta}$ and setting it equal to zero to minimize $\hat{\mu^2}$
- F.O.C (First Order Conditions);
- $$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- $$\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- $$\sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- .

- .

Derivation of OLS for Multiple Regression Model

- $$\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots\ldots\ldots - \hat{\beta}_k x_{ik}) = 0$$

Derivation of OLS for Multiple Regression Model

- $$\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- It is difficult to solve this system of equation manually with hand - However, the idea is similar as simple regression model.

Derivation of OLS for Multiple Regression Model

- $$\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- It is difficult to solve this system of equation manually with hand - However, the idea is similar as simple regression model.
- At the end we will get our beta estimates from solving these equations, only if we can uniquely find the solution of the above system.

Derivation of OLS for Multiple Regression Model

- $$\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - ..... - \hat{\beta}_k x_{ik}) = 0$$
- It is difficult to solve this system of equation manually with hand - However, the idea is similar as simple regression model.
- At the end we will get our beta estimates from solving these equations, only if we can uniquely find the solution of the above system.
- As long as there is no perfect collinearity among the regressors (M.L.R - III), we do find the unique solution of this system.

Changing more than one Variable

- we can find the effect of change in more than one variable in x on y using multiple regression model.

Changing more than one Variable

- we can find the effect of change in more than one variable in x on y using multiple regression model.
- Again;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

Changing more than one Variable

- we can find the effect of change in more than one variable in x on y using multiple regression model.
- Again;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- Taking Differential on Both Sides;

$$\partial y = \beta_1 \partial x_1 + \beta_2 \partial x_2 + ... + \beta_k \partial x_k$$

Changing more than one Variable

- we can find the effect of change in more than one variable in x on y using multiple regression model.
- Again;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- Taking Differential on Both Sides;

$$\partial y = \beta_1 \partial x_1 + \beta_2 \partial x_2 + ... + \beta_k \partial x_k$$

- This time only those $\partial \beta_k x_k$ will be zero which we do not want to change;

Changing more than one Variable

- we can find the effect of change in more than one variable in x on y using multiple regression model.
- Again;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

- Taking Differential on Both Sides;

$$\partial y = \beta_1 \partial x_1 + \beta_2 \partial x_2 + ... + \beta_k \partial x_k$$

- This time only those $\partial \beta_k x_k$ will be zero which we do not want to change;
- So if we want to see the effect of x_1 and x_2 on y then;

$$\Delta y = \beta_1 \Delta x_1 + \beta_2 \Delta x_2$$

OLS Predicted & Residual Values

- Similar to simple regression model, the predicted values will be;

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \ldots\ldots\ldots + \hat{\beta}_k x_{ik}$$

where $\hat{y}$ = Predicted Values

OLS Predicted & Residual Values

- Similar to simple regression model, the predicted values will be;

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + ... + \hat{\beta}_k x_{ik}$$

  where $\hat{y}$ = Predicted Values

- Similar to simple regression model, the residual values or error will be;

$$\hat{\mu}_i = y_i - \hat{y}_i$$

OLS Predicted & Residual Values

- Similar to simple regression model, the predicted values will be;

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \ldots\ldots\ldots + \hat{\beta}_k x_{ik}$$

  where $\hat{y}$ = Predicted Values

- Similar to simple regression model, the residual values or error will be;

$$\hat{\mu}_i = y_i - \hat{y}_i$$

- OR

$$\hat{\mu}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \hat{\beta}_3 x_{i3} - \ldots\ldots\ldots - \hat{\beta}_k x_{ik}$$

  where $\hat{\mu}$ = Residual Values or Error

Algebraic Properties of OLS in MLR

1. The sum of all the error terms is zero.

$$\sum_{i=1}^{n} \hat{\mu}_i = 0$$

Algebraic Properties of OLS in MLR

1. The sum of all the error terms is zero.

$$\sum_{i=1}^{n} \hat{\mu}_i = 0$$

2. The average of original values of y will be equals to the predicted values of y.

$$\bar{y} = \bar{\hat{y}}$$

Algebraic Properties of OLS in MLR

1. The sum of all the error terms is zero.

$$\sum_{i=1}^{n} \hat{\mu}_i = 0$$

2. The average of original values of y will be equals to the predicted values of y.

$$\bar{y} = \bar{\hat{y}}$$

3. The point $(\bar{y}, \bar{x_1}, \bar{x_2}, \bar{x_3}.....\bar{x_k})$ always lies on the regression line so ;

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x_1} + \hat{\beta}_2 \bar{x_2} + ... + \hat{\beta}_k \bar{x_k}$$

Algebraic Properties of OLS in MLR

1. The sum of all the error terms is zero.

$$\sum_{i=1}^{n} \hat{\mu}_i = 0$$

2. The average of original values of y will be equals to the predicted values of y.

$$\bar{y} = \bar{\hat{y}}$$

3. The point $(\bar{y}, \bar{x_1}, \bar{x_2}, \bar{x_3}.....\bar{x_k})$ always lies on the regression line so ;

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x_1} + \hat{\beta}_2 \bar{x_2} + ... + \hat{\beta}_k \bar{x_k}$$

4. The Sample covariance between each independent variable and residuals are zero. Similarly, the covariance between predicted values and residuals is zero.

$$\sum_{i=1}^{n} x_{ij} \hat{\mu}_i = 0 \quad \text{for each } j$$

Regression Through Origin

In case of multiple linear regression, when we force β_0 to be zero, we estimate the following model;

$$y = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + ... + \tilde{\beta}_k x_k$$

However,

$$E[\mu] \neq 0 \quad \textit{Biased Estimators}$$

R-Squared

Similar to Simple Regression;

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

R-Squared

Similar to Simple Regression;

- $$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
- $$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

R-Squared

Similar to Simple Regression;

- $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- $$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- $$SSR = \sum_{i=1}^{n}\hat{\mu}_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

R-Squared

Similar to Simple Regression;

- $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$
- $$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$
- $$SSR = \sum_{i=1}^{n}\hat{\mu}_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$
- $$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

R-Squared

- There is a slight difference in the interpretation.

R-Squared

- There is a slight difference in the interpretation.
- R^2 is the proportion of variation in y that can be explained by the variation in **all independent variables**.

R-Squared

- There is a slight difference in the interpretation.
- R^2 is the proportion of variation in y that can be explained by the variation in **all independent variables**.
- The R-Squared generally increase when we add more and more variable (Regardless of its relevance).

R-Squared

- There is a slight difference in the interpretation.
- R^2 is the proportion of variation in y that can be explained by the variation in **all independent variables**.
- The R-Squared generally increase when we add more and more variable (Regardless of its relevance).
- We can also use adjusted R-squared to address this issue. (Will discuss in later chapters)

R-Squared

- There is a slight difference in the interpretation.
- R^2 is the proportion of variation in y that can be explained by the variation in **all independent variables**.
- The R-Squared generally increase when we add more and more variable (Regardless of its relevance).
- We can also use adjusted R-squared to address this issue. (Will discuss in later chapters)
- If we add a variable in the multiple regression model that has missing observation for 25 percent of the time. What would happen to R-Squared?

Comparing Simple & Multiple Regression

Another Perspective on Partial Effect

- Let's assume; we are interested in finding the effect of education on wage. So, we run a simple regression model;

$$wage = \eta_0 + \eta_1 education + \mu_1$$

Another Perspective on Partial Effect

- Let's assume; we are interested in finding the effect of education on wage. So, we run a simple regression model;

$$wage = \eta_0 + \eta_1 education + \mu_1$$

- However; $E[\mu_1|ability] \neq E[\mu_1]$ - Ability in the error that affect both education & wage. - (Education is Endogenous Variable)

Another Perspective on Partial Effect

- Let's assume; we are interested in finding the effect of education on wage. So, we run a simple regression model;

$$wage = \eta_0 + \eta_1 education + \mu_1$$

- However; $E[\mu_1|ability] \neq E[\mu_1]$ - Ability in the error that affect both education & wage. - (Education is Endogenous Variable)
- Let's run the following model first;

$$Education = \gamma_0 + \gamma_1 ability + \mu_2$$

Another Perspective on Partial Effect

- Let's assume; we are interested in finding the effect of education on wage. So, we run a simple regression model;

$$wage = \eta_0 + \eta_1 education + \mu_1$$

- However; $E[\mu_1|ability] \neq E[\mu_1]$ - Ability in the error that affect both education & wage. - (Education is Endogenous Variable)
- Let's run the following model first;

$$Education = \gamma_0 + \gamma_1 ability + \mu_2$$

- μ_2 is the residual from this regression, the part of education left unexplained by (uncorrelated with) ability. Therefore if we run the following model;

$$wage = \delta_0 + \delta_1 \mu_2 + \mu_3$$

Another Perspective on Partial Effect

- Let's assume; we are interested in finding the effect of education on wage. So, we run a simple regression model;

$$wage = \eta_0 + \eta_1 education + \mu_1$$

- However; $E[\mu_1|ability] \neq E[\mu_1]$ - Ability in the error that affect both education & wage. - (Education is Endogenous Variable)
- Let's run the following model first;

$$Education = \gamma_0 + \gamma_1 ability + \mu_2$$

- μ_2 is the residual from this regression, the part of education left unexplained by (uncorrelated with) ability. Therefore if we run the following model;

$$wage = \delta_0 + \delta_1 \mu_2 + \mu_3$$

- This δ_1 will be the same as β_1 from the multiple regression model given below; - (**Frisch - Waugh Theorem**)

$$wage = \beta_0 + \beta_1 education + \beta_2 ability + \mu_4$$

Comparison of Simple & Multiple Regression

- Turns out, there is a simple relationship between η_1 from simple regression model and β_1 from multiple regression model.

Comparison of Simple & Multiple Regression

- Turns out, there is a simple relationship between η_1 from simple regression model and β_1 from multiple regression model.
- $$\eta_1 = \beta_1 + \beta_2\gamma_1$$

  When would the estimates from both regression model would be same?

Comparison of Simple & Multiple Regression

- Turns out, there is a simple relationship between η_1 from simple regression model and β_1 from multiple regression model.
- $$\eta_1 = \beta_1 + \beta_2\gamma_1$$

  When would the estimates from both regression model would be same?

  1. When $\beta_2 = 0$ - There is no relationship between x_2 and y.

Comparison of Simple & Multiple Regression

- Turns out, there is a simple relationship between η_1 from simple regression model and β_1 from multiple regression model.
- $$\eta_1 = \beta_1 + \beta_2\gamma_1$$

  When would the estimates from both regression model would be same?
  1. When $\beta_2 = 0$ - There is no relationship between x_2 and y.
  2. When $\gamma_1 = 0$ - There is no relationship between x_1 and x_2.

Comparison of Simple & Multiple Regression

- Turns out, there is a simple relationship between η_1 from simple regression model and β_1 from multiple regression model.
- $$\eta_1 = \beta_1 + \beta_2\gamma_1$$

  When would the estimates from both regression model would be same?
  1. When $\beta_2 = 0$ - There is no relationship between x_2 and y.
  2. When $\gamma_1 = 0$ - There is no relationship between x_1 and x_2.
- Back to $E[\mu|x] = E[\mu] = 0$ - There should not be any variable in error term that can affect y and also correlated with any x.

The Expected Value of OLS Estimators

Un-biasedness of OLS Estimators

Under Assumptions MLR-I to MLR-IV; $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, ...\hat{\beta}_k$ will be unbiased estimators.

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

Un-biasedness of OLS Estimators

Under Assumptions MLR-I to MLR-IV; $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, ...\hat{\beta}_k$ will be unbiased estimators.

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

Un-biasedness of OLS Estimators

Under Assumptions MLR-I to MLR-IV; $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, ...\hat{\beta}_k$ will be unbiased estimators.

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

3. There is no Perfect Co-linearity - M.L.R - III

   In the sample and population, none of the independent variables are constant and there are no exact linear relationships among the independent variables.

Un-biasedness of OLS Estimators

Under Assumptions MLR-I to MLR-IV; $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, ...\hat{\beta}_k$ will be unbiased estimators.

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

3. There is no Perfect Co-linearity - M.L.R - III

   In the sample and population, none of the independent variables are constant and there are no exact linear relationships among the independent variables.

4. Zero Conditional Mean in Population - M.L.R - IV

$$E[\mu | x_1, x_2, ..., x_k] = 0$$

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

But instead we run the following model ;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

where x_3 is totally irrelevant variable.

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

But instead we run the following model ;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

where x_3 is totally irrelevant variable.

- Will the regression coefficient will still be unbiased?

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

But instead we run the following model ;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

where x_3 is totally irrelevant variable.

- Will the regression coefficient will still be unbiased?
- Let's look at MLR-I through MLR-IV ;

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

But instead we run the following model ;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

where x_3 is totally irrelevant variable.

- Will the regression coefficient will still be unbiased?
- Let's look at MLR-I through MLR-IV ;
- $\beta_3 = 0$ in the population, though $\hat{\beta}_3$ will be close to zero in any given sample, and;

$$E[\mu|x_1, x_2, x_3] = E[\mu|x_1, x_2] = 0$$

Including Irrelevant Variable in Regression Model

- Inclusion of an irrelevant variable - Over-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

But instead we run the following model ;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

where x_3 is totally irrelevant variable.

- Will the regression coefficient will still be unbiased?
- Let's look at MLR-I through MLR-IV ;
- $\beta_3 = 0$ in the population, though $\hat{\beta}_3$ will be close to zero in any given sample, and;

$$E[\mu|x_1, x_2, x_3] = E[\mu|x_1, x_2] = 0$$

- Therefore, the regression coefficient will still be unbiased (Variance might be larger).

Omitting a relevant variable

- Omitting a relevant variable from the model - Under-specifying the model.

Omitting a relevant variable

- Omitting a relevant variable from the model - Under-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

But we run the following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

Omitting a relevant variable

- Omitting a relevant variable from the model - Under-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

But we run the following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

- Let's look at MLR-I through MLR-IV ;

Omitting a relevant variable

- Omitting a relevant variable from the model - Under-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

But we run the following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

- Let's look at MLR-I through MLR-IV ;
-

$$E[\mu|x_1, x_2] \neq E[\mu|x_1, x_2, x_3] = 0$$

Omitting a relevant variable

- Omitting a relevant variable from the model - Under-specifying the model.
- Let we have following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \mu$$

But we run the following model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

- Let's look at MLR-I through MLR-IV ;
-

$$E[\mu|x_1, x_2] \neq E[\mu|x_1, x_2, x_3] = 0$$

- Therefore, the regression coefficient will be Biased (Omitted Variable Bias).

Direction of Bias

- we know;

$$\hat{\eta}_1 = \hat{\beta}_1 + \hat{\beta}_2\gamma_1$$

Direction of Bias

- we know;

$$\hat{\eta_1} = \hat{\beta}_1 + \hat{\beta}_2\gamma_1$$

-

$$E[\hat{\eta_1}] = E[\hat{\beta}_1 + \hat{\beta}_2\gamma_1]$$

Direction of Bias

- we know;

$$\hat{\eta_1} = \hat{\beta}_1 + \hat{\beta}_2\gamma_1$$

- $$E[\hat{\eta_1}] = E[\hat{\beta}_1 + \hat{\beta}_2\gamma_1]$$

- $$E[\hat{\eta_1}] = \beta_1 + \beta_2\gamma_1$$

Direction of Bias

- we know;

$$\hat{\eta_1} = \hat{\beta}_1 + \hat{\beta}_2\gamma_1$$

- $$E[\hat{\eta_1}] = E[\hat{\beta}_1 + \hat{\beta}_2\gamma_1]$$

- $$E[\hat{\eta_1}] = \beta_1 + \beta_2\gamma_1$$

- $$Bias[\hat{\eta_1}] = E[\hat{\eta_1}] - \beta_1 = \beta_2\gamma_1$$

Direction of Bias

- we know;

$$\hat{\eta}_1 = \hat{\beta}_1 + \hat{\beta}_2\gamma_1$$

- $$E[\hat{\eta}_1] = E[\hat{\beta}_1 + \hat{\beta}_2\gamma_1]$$

- $$E[\hat{\eta}_1] = \beta_1 + \beta_2\gamma_1$$

- $$Bias[\hat{\eta}_1] = E[\hat{\eta}_1] - \beta_1 = \beta_2\gamma_1$$

- Direction of Bias

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ When x_2 Is Omitted in Estimating Equation (3.40)

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

The Variance of OLS Estimators

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

- If the assumption M.L.R - V holds, we call the errors to be homoscedastic.

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

- If the assumption M.L.R - V holds, we call the errors to be homoscedastic.
- However, if the assumption M.L.R - V does not hold, we call the errors to be heteroscedastic.

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

- If the assumption M.L.R - V holds, we call the errors to be homoscedastic.
- However, if the assumption M.L.R - V does not hold, we call the errors to be heteroscedastic.
- In STATA, we visualize the error with independent variable to see the pattern in variance.

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

- If the assumption M.L.R - V holds, we call the errors to be homoscedastic.
- However, if the assumption M.L.R - V does not hold, we call the errors to be heteroscedastic.
- In STATA, we visualize the error with independent variable to see the pattern in variance.
- OR we can also use some statistical test to determine homoscedasticity/heteroscedasticity. (Will be covered in later chapters).

Homoscadasticity

- The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu|(x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

- If the assumption M.L.R - V holds, we call the errors to be homoscedastic.
- However, if the assumption M.L.R - V does not hold, we call the errors to be heteroscedastic.
- In STATA, we visualize the error with independent variable to see the pattern in variance.
- OR we can also use some statistical test to determine homoscedasticity/heteroscedasticity. (Will be covered in later chapters).
- The assumptions M.L.R - I through M.L.R - V are collectively known as **Gauss Markov Assumptions**.

The Variance of Error Term

Variance of the OLS Slope Estimators

- Under Assumptions M.L.R - I through M.L.R - V, the variance of the slope estimators would be;

$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

where σ^2 is the variance of error.

$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$

R_j^2 is the R-Squared from the regression of x_j on all the remaining independent variables.

The Variance of Error Term

Variance of the OLS Slope Estimators

- Under Assumptions M.L.R - I through M.L.R - V, the variance of the slope estimators would be;

$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

where σ^2 is the variance of error.

$$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

R_j^2 is the R-Squared from the regression of x_j on all the remaining independent variables.

- Therefore, there are three components of the variance of β_j estimators.

The Variance of Error Term

1. The Variance of Error Term

- We can estimate the variance of the residual term using the following equation;

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} \hat{\mu^2}}{(n-k-1)} = \frac{SSR}{(n-k-1)}$$

where n-k-1 is the degree of freedom.

The Variance of Error Term

1. The Variance of Error Term

- We can estimate the variance of the residual term using the following equation;

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} \hat{\mu^2}}{(n-k-1)} = \frac{SSR}{(n-k-1)}$$

  where n-k-1 is the degree of freedom.

- The degree of freedom is the degree to which your observations are more than the technically required values.

The Variance of Error Term

1. The Variance of Error Term

- We can estimate the variance of the residual term using the following equation;

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n} \hat{\mu^2}}{(n-k-1)} = \frac{SSR}{(n-k-1)}$$

  where n-k-1 is the degree of freedom.
- The degree of freedom is the degree to which your observations are more than the technically required values.
- As the $\hat{\mu}$ is from the sample the $\hat{\sigma^2}$ is also an estimator. It will be unbiased under the assumption M.L.R - I through M.L.R-V.

$$E[\hat{\sigma^2}] = \sigma^2$$

The Total Sample Variation in x_j

2. The Total Sample Variation in x_j

- $$SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

The Total Sample Variation in x_j

2. The Total Sample Variation in x_j

- $$SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$
- This is the variance of x_j.

The Total Sample Variation in x_j

2. The Total Sample Variation in x_j

- $$SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$
- This is the variance of x_j.
- As
$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

The Total Sample Variation in x_j

2. The Total Sample Variation in x_j

- $$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$
- This is the variance of x_j.
- As
$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$
  - The bigger the variance is, the less variance of the slope estimators. (Ideal)

    The smaller the variance is, the more variance of the slope estimators.

3. R_j^2 The Linear Relationship among Independent Variables

- R_j^2 is the R-Squared from the regression of x_j on the remaining independent variables.

3. R_j^2 The Linear Relationship among Independent Variables

- R_j^2 is the R-Squared from the regression of x_j on the remaining independent variables.
- This Term was not in the variance of slope estimators in the simple regression model because we had just one independent variable.

3. R_j^2 The Linear Relationship among Independent Variables

- R_j^2 is the R-Squared from the regression of x_j on the remaining independent variables.
- This Term was not in the variance of slope estimators in the simple regression model because we had just one independent variable.
- This is the extent to which independent variables are related to each other. (The Covariance between x_j and all other independent variables.

3. R_j^2 The Linear Relationship among Independent Variables

- R_j^2 is the R-Squared from the regression of x_j on the remaining independent variables.
- This Term was not in the variance of slope estimators in the simple regression model because we had just one independent variable.
- This is the extent to which independent variables are related to each other. (The Covariance between x_j and all other independent variables.
- As

$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

3. R_j^2 The Linear Relationship among Independent Variables

- R_j^2 is the R-Squared from the regression of x_j on the remaining independent variables.
- This Term was not in the variance of slope estimators in the simple regression model because we had just one independent variable.
- This is the extent to which independent variables are related to each other. (The Covariance between x_j and all other independent variables.
- As

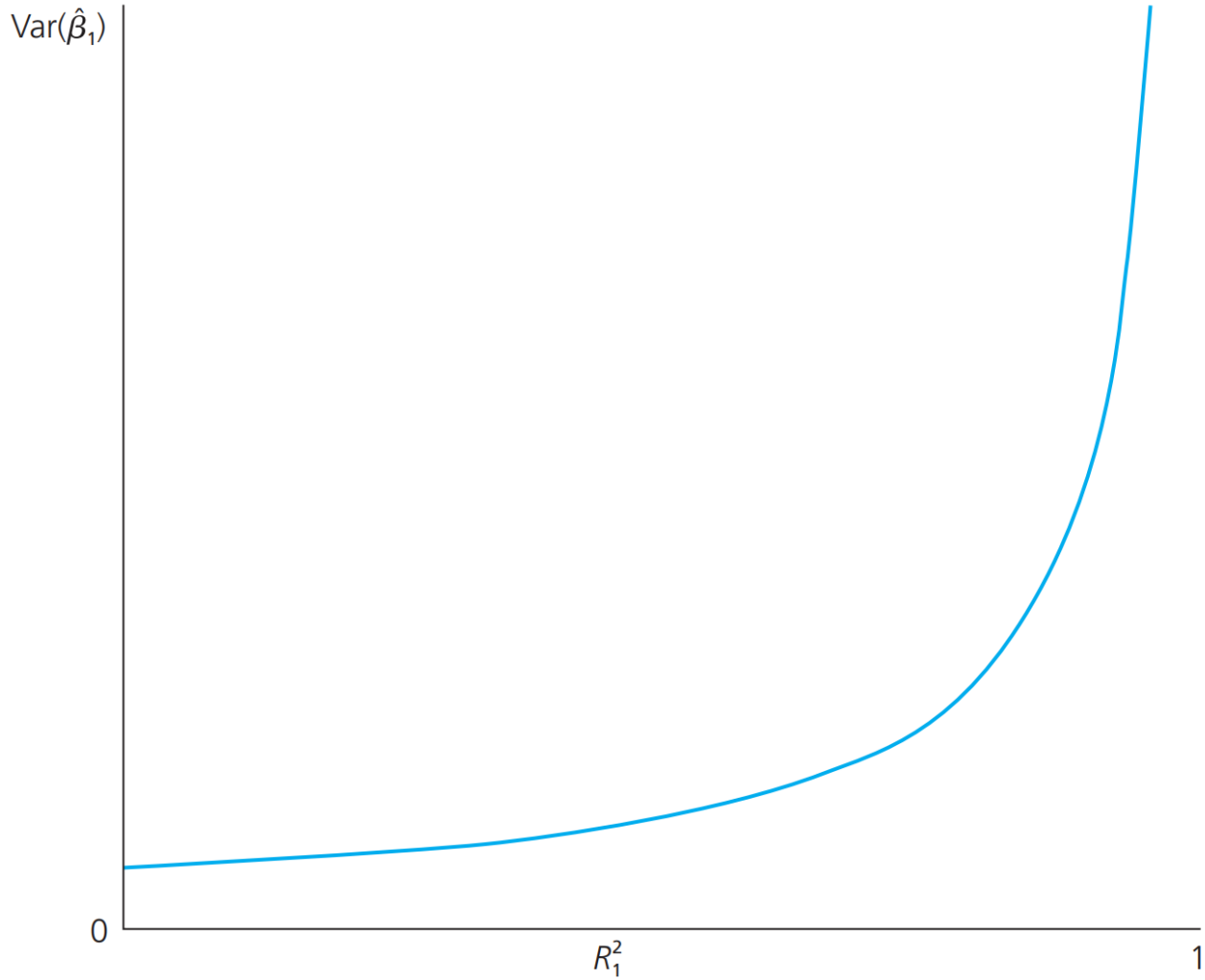
$$Var[\hat{\beta}_j] = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

  - The bigger the R_j^2 is, the bigger variance of the slope estimators.

    The smaller the R_j^2 is, the smaller variance of the slope estimators. (Ideal)

3. R_j^2 The Linear Relationship among Independent Variables

FIGURE 3.1 $\text{Var}(\hat{\beta}_1)$ as a function of R_1^2.



Multicollinearity

- The multicollinearity is when the R_j^2 approaches to 1. The higher the R-Squared is, the higher multicollinearity will exist causing the variance of slope estimators to be high.

Multicollinearity

- The multicollinearity is when the R_j^2 approaches to 1. The higher the R-Squared is, the higher multicollinearity will exist causing the variance of slope estimators to be high.
- The extreme case of multicollinearity is perfect collinearity where $R_j^2 = 1$. In this case there exists an exact relationship between at least two variables violating M.L.R - III.

Multicollinearity

- The multicollinearity is when the R_j^2 approaches to 1. The higher the R-Squared is, the higher multicollinearity will exist causing the variance of slope estimators to be high.
- The extreme case of multicollinearity is perfect collinearity where $R_j^2 = 1$. In this case there exists an exact relationship between at least two variables violating M.L.R - III.
- For Example:

$$wage = \beta_0 + \beta_1 education + \beta_2 Male + \beta_3 Female + \mu$$

Multicollinearity

- The multicollinearity is when the R_j^2 approaches to 1. The higher the R-Squared is, the higher multicollinearity will exist causing the variance of slope estimators to be high.
- The extreme case of multicollinearity is perfect collinearity where $R_j^2 = 1$. In this case there exists an exact relationship between at least two variables violating M.L.R - III.
- For Example:

$$wage = \beta_0 + \beta_1 education + \beta_2 Male + \beta_3 Female + \mu$$

- Exact relationship between male and female;

$$Male + Female = 1$$

In this case the regression would not run.

Variance Inflation Factor - VIF

- Let's Define Variance Inflation Factor $\frac{1}{1-R_j^2}$ is the component of variance that depends on the covariance between independent variable.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} VIF$$

Variance Inflation Factor - VIF

- Let's Define Variance Inflation Factor $\frac{1}{1-R_j^2}$ is the component of variance that depends on the covariance between independent variable.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} VIF$$

- Ideally, we want to have the VIF to be as low as possible (Low R_j^2).

Variance Inflation Factor - VIF

- Let's Define Variance Inflation Factor $\frac{1}{1-R_j^2}$ is the component of variance that depends on the covariance between independent variable.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} VIF$$

- Ideally, we want to have the VIF to be as low as possible (Low R_j^2).
- However, it is not in our hand when we are trying to remove endogeneity.

Variance Inflation Factor - VIF

- Let's Define Variance Inflation Factor $\frac{1}{1-R_j^2}$ is the component of variance that depends on the covariance between independent variable.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j} VIF$$

- Ideally, we want to have the VIF to be as low as possible (Low R_j^2).
- However, it is not in our hand when we are trying to remove endogeneity.
- One trick is to only care about the our main variable while looking at the VIF.

Variances in Misspecified Models

- There can be tradeoff between Bias and Variance.

Variances in Misspecified Models

- There can be tradeoff between Bias and Variance.
- Let we have the following population model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

Variances in Misspecified Models

- There can be tradeoff between Bias and Variance.
- Let we have the following population model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

- So, our estimated model would be

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\mu}$$

Variances in Misspecified Models

- There can be tradeoff between Bias and Variance.
- Let we have the following population model;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

- So, our estimated model would be

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\mu}$$

- And our misspecified model would be;

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\mu}$$

Variances in Misspecified Models

- The Variance of Correctly specified model would be;

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1} VIF$$

Variances in Misspecified Models

- The Variance of Correctly specified model would be;

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1} VIF$$

- The Variance of Misspecified model would be;

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST_1}$$

Variances in Misspecified Models

- The Variance of Correctly specified model would be;

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1} VIF$$

- The Variance of Misspecified model would be;

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST_1}$$

- As $VIF = \frac{1}{1-R_j^2} > 1$ So,

$$Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$$

Variances in Misspecified Models

- The Variance of Correctly specified model would be;

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1} VIF$$

- The Variance of Misspecified model would be;

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST_1}$$

- As $VIF = \frac{1}{1-R_j^2} > 1$ So,

$$Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$$

- Therefore, there is a clear trade off between Bias and Variance.

Variances in Misspecified Models

- The Variance of Correctly specified model would be;

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1} VIF$$

- The Variance of Misspecified model would be;

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST_1}$$

- As $VIF = \frac{1}{1-R_j^2} > 1$ So,

$$Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$$

- Therefore, there is a clear trade off between Bias and Variance.
- Generally, we care more about variance in forecasting and care more about bias while finding the effect of x on y.

Gauss Markov Theorem

Gauss Markov Assumptions

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

Gauss Markov Assumptions

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

Gauss Markov Assumptions

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

3. There is no Perfect Co-linearity - M.L.R - III

   In the sample and population, none of the independent variables are constant and there are no exact linear relationships among the independent variables.

Gauss Markov Assumptions

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

3. There is no Perfect Co-linearity - M.L.R - III

   In the sample and population, none of the independent variables are constant and there are no exact linear relationships among the independent variables.

4. Zero Conditional Mean in Population - M.L.R - IV

$$E[\mu | x_1, x_2, ..., x_k] = 0$$

Gauss Markov Assumptions

1. The population model is linear in parameters - M.L.R - I

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \mu$$

2. We have a random sample of n observations. - M.L.R - II

$$(x_{i1}, x_{i2}, ..., x_{ij}, y) : i = 1, 2, 3, ..., n$$

3. There is no Perfect Co-linearity - M.L.R - III

   In the sample and population, none of the independent variables are constant and there are no exact linear relationships among the independent variables.

4. Zero Conditional Mean in Population - M.L.R - IV

$$E[\mu | x_1, x_2, ..., x_k] = 0$$

5. The error μ has the same variance given any value/combination of the explanatory values. - M.L.R - V

$$Var(\mu | (x_1, x_2, x_3, ..., x_k)) = \sigma^2$$

Gauss Markov Theorem

- Under Gauss Markov Assumptions (M.L.R - I through M.L.R - V), the OLS estimators are considered to be **Best Linear Unbiased Estimators (BLUE)**.

Gauss Markov Theorem

- Under Gauss Markov Assumptions (M.L.R - I through M.L.R - V), the OLS estimators are considered to be **Best Linear Unbiased Estimators (BLUE)**.
- When the standard set of assumptions holds, we need not look for alternative unbiased estimators (Minimizing other forms of errors): none will be better than OLS.

Examples

Example 1:

1 Using the data in GPA2 on 4,137 college students, the following equation was estimated by OLS:

$$\widehat{colgpa} = 1.392 - .0135\ hsperc + .00148\ sat$$
$$n = 4{,}137,\ R^2 = .273,$$

where *colgpa* is measured on a four-point scale, *hsperc* is the percentile in the high school graduating class (defined so that, for example, *hsperc* = 5 means the *top* 5% of the class), and *sat* is the combined math and verbal scores on the student achievement test.

(i) Why does it make sense for the coefficient on *hsperc* to be negative?

(ii) What is the predicted college GPA when *hsperc* = 20 and *sat* = 1,050?

(iii) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A's SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?

(iv) Holding *hsperc* fixed, what difference in SAT scores leads to a predicted *colgpa* difference of .50, or one-half of a grade point? Comment on your answer.

Example 2:

7 Which of the following can cause OLS estimators to be biased?

(i) Heteroskedasticity.

(ii) Omitting an important variable.

(iii) A sample correlation coefficient of .95 between two independent variables both included in the model.

Example 3:

15 The following estimated equations use the data in MLB1, which contains information on major league baseball salaries. The dependent variable, *lsalary*, is the log of salary. The two explanatory variables are years in the major leagues (*years*) and runs batted in per year (*rbisyr*):

$$\widehat{lsalary} = \underset{(.098)}{12.373} + \underset{(.0132)}{.1770}\ years$$

$$n = 353,\ \text{SSR} = 326.196,\ \text{SER} = .964,\ R^2 = .337$$

$$\widehat{lsalary} = \underset{(.084)}{11.861} + \underset{(.0118)}{.0904}\ years + \underset{(.0020)}{.0302}\ rbisyr$$

$$n = 353,\ \text{SSR} = 198.475,\ \text{SER} = .753,\ R^2 = .597$$

(i) How many degrees of freedom are in each regression? Why is the SER smaller in the second regression than the first?

(ii) The sample correlation coefficient between *years* and *rbisyr* is about 0.487. Does this make sense? What is the variance inflation factor (there is only one) for the slope coefficients in the multiple regression? Would you say there is little, moderate, or strong collinearity between *years* and *rbisyr*?

(iii) How come the standard error for the coefficient on *years* in the multiple regression is lower than its counterpart in the simple regression?