

# Introduction to Econometrics: Recitation 2

Seung-hun Lee\*

September 22nd, 2022

## 1 What is econometrics trying to achieve?

Econometrics is a field in economics that tries to answer real life questions, from finding how additional schooling affects our wage in the real market to how certain policies affect certain outcomes (e.g. Did a particular welfare program improve health?). Ultimately, it is about making a **quantitative** statement about two or more random events. We can make simple correlational statements using econometrics (e.g. wages and schooling is positively correlated). However, we want to go one extra mile and identify *causal relationships* - as in how much does our wage rise with one additional year of schooling? In order to achieve this goal, we collect data from a suitably defined population (more on this later) and use various methods to estimate a parameter that implies correlational/causal relationship.

To fully understand what econometrics is trying to achieve, we need to ask ourselves these three questions: What is the difference between correlational and causal relation? What is the suitably defined population? What are the methods that we need to use in econometrics?

### 1.1 Correlational vs causal relations

Suppose you have two random variables - call them  $X$  and  $Y$ . You want to identify if  $X$  *causes*  $Y$ , beyond being simply correlated. A **correlation** between  $X$  and  $Y$  is a statistical measure that describes how the two variables move together. It captures *any* type of statistical dependence that moves the two variables together. It can actually be causal in some cases

---

\*Contact me at [sl4436@columbia.edu](mailto:sl4436@columbia.edu) if you spot any errors or have suggestions on improving this note.

but in most cases, it is not. For instance, it may be the case that  $X$  and  $Y$  are correlated because  $X$  actually causes  $Y$  (think about more schooling and higher wages). However,  $Y$  can cause  $X$ , which makes them correlated but not  $X$  causing  $Y$ . It may be the case that  $X$  and  $Y$  are jointly moving because there is another factor  $Z$  that affects both, making the connection  $X$  and  $Y$  indirect.

For us to be able to claim **causal** relationship, it must be that cleanly changing variables  $X$  leads to changes in  $Y$ . More formally, if we are able to observe that exogenous changes in  $X$  affects  $Y$ , the two variables are in a causal relationship in that  $X$  causes  $Y$ . This is actually much more difficult to achieve than it seems. Changes in  $X$  may be a combination of many things - changes from  $X$  alone and changes from other factors that may indirectly affect  $X$ . Because real life events can rarely be controlled, we need to be able to tell the two apart. This is where randomized control trials and natural experiments can come in handy. Both methods, if correctly designed and the conditions are met, can help us tell whether changes in  $X$  affects  $Y$ , and by how much. Methods we use in econometrics are handy in that we can express concisely the relationship between  $X$  and  $Y$  variables in a single equation.

## 1.2 Suitably defined population

In econometrics, we need to be clear about the population that we are interested in. When we say we are interested in the relationship between schooling and wages, whose effects are we interested in? Are we interested in this for the entire US population, high school graduates, or college graduates? This matters because it determines what sampling methods we need to use in order to ensure that our observations in the data are comparable (or at least get close to it). In terms of sampling methods, we can use complete randomization, stratified randomization, or cluster randomization depending on our population characteristics.

This also raises another issue: We are almost surely never going to get the data from the entire population. Gathering data from the entire population is logistically (and maybe even ethically) difficult. The estimate we are obtaining through any econometric exercise is thus a *sample analogue* of the actual value we are trying to get, and we will do various diagnostic testing to see if they can be reasonably close to the true value.

## 1.3 What methods to use?

In econometrics, we will use many estimation methods to obtain the sample analogue of the parameter of interest. The most commonly used methods are:

- Ordinary least squares (OLS): Used in cross-sectional analysis where we observe multiple observations in one time period. Suitable for randomized control trials or in any case where the treatment assignment is as good as random.
- Panel estimation: If data has multiple individuals and multiple time periods.
- Instrumental variable methods: When we have proxy variables relevant to variable of interest and is reasonably exogenous
- Difference-in-differences: When we study 'before & after' events with multiple entities
- Regression discontinuity: In treatment with a cutoff determining treatment assignment
- Time series: When we observe one entity over multiple periods

Also, depending on the type of variables we use in our exercise, we have the following methods

- Univariate regression: One variable (besides an overall constant) controlled for
- Multivariate regression: Multiple variables (besides an overall constant) controlled for
- Nonlinear regression: Binary dependent variables
- Big data methods

## 2 Randomized control trials

In **randomized control trials**, we randomly categorize some individuals under treatment group and controlled group and run various tests as if this is a laboratory experiment. While this is difficult to pull off due to logistical and ethical concerns, many economists have successfully ran field experiments to answer interesting questions in economics (e.g. Does health interventions, such as deworming treatment, increase educational attainment for children in Kenya?<sup>1</sup>).

Randomized control trial serves as a benchmark for good program evaluation, where we assess the effectiveness of a policy in achieving its intended goals. As such, the topics we cover in this course will be largely based on randomized control trials and how to analyze/critique them.

---

<sup>1</sup>Edward Miguel, and Michael Kremer (2004) "Worms: Identifying impacts on education and health in the presence of treatment externalities", *Econometrica* 72(1), 159-217

## 2.1 Potential outcomes framework

Let  $Y_i$  be the observed outcome for individual  $i \in \{1, \dots, N\}$ . For each individual  $i$ , she can either be treated or be in the control group (not both). This will be captured by a variable  $W_i$  which equals 1 if  $i$  is treated, 0 if otherwise. Since each  $N$  individuals can be assigned to the treatment or not, we can define  $\mathbf{W}$ , an  $N$ -tuple vector of treatment assignment for all individuals. While it is realistic to think that individual treatment could be affected by how many others have taken the treatment, we will assume this away for the entire course until otherwise specified. This is known as **stable unit treatment value assumption (SUTVA)**. Mathematically, we can write  $Y_i(\mathbf{W}) = Y_i(w)$

A **potential outcome**, written as  $Y_i(w)$  where  $w = 1$  if individual  $i$  is in the treatment group and  $w = 0$  if not, is the outcome for the treated ( $Y_i(1)$ ) and the untreated individual ( $Y_i(0)$ ). Since an individual cannot be treated and untreated at the same time, we can only observe at most one of  $Y_i(1)$  or  $Y_i(0)$  for a given individual. This is known as a **fundamental problem of missing data** and raises problems in estimating the effect of a treatment.

We can write the relation between the observed outcome  $Y_i$  and the pair of potential outcomes  $Y_i(1)$  and  $Y_i(0)$  as follows.

$$\begin{aligned} Y_i &= Y_i(1)W_i + Y_i(0)(1 - W_i) \\ &= Y_i(0) + W_i(Y_i(1) - Y_i(0)) \end{aligned}$$

In here, we know  $W_i$  and  $Y_i$  for everyone regardless of treatment assignment. The problem is that we cannot see  $Y_i(0)$  for the treated group and  $Y_i(1)$  for the untreated group.

## 2.2 Treatment effects

What we are interested in is whether the treatment  $W_i$  affects the outcome of interest  $Y_i$ , or gains from the treatment that can be written as

$$Y_i(1) - Y_i(0)$$

However, because of the fundamental problem of missing data, we cannot capture the exact gain at all. Thus, the closest we can get to is the **average treatment effect (ATE)** and the

**average treatment effect on the treated (ATT)** which are defined as

$$ATE = E[Y_i(1) - Y_i(0)]$$

$$ATT = E[Y_i(1) - Y_i(0)|W_i = 1]$$

Furthermore, to characterize these concepts with something we can estimate, we need to set some assumptions. The strongest assumption possible is the **randomized assignment**, which states that  $Y_i(1)$  and  $Y_i(0)$  are independent of  $W_i$ . Or

$$E[Y_i(1)] = E[Y_i(1)|W_i = 1] = E[Y_i(1)|W_i = 0]$$

and similarly for  $E[Y_i(0)]$ . With this assumption and the definition of potential outcomes framework, we can write

$$\begin{aligned} E[Y_i|W_i] &= E[Y_i(1)W_i + Y_i(0)(1 - W_i)|W_i] \\ &= E[Y_i(1)|W_i]W_i + E[Y_i(0)|W_i](1 - W_i) \\ &= E[Y_i(1)]W_i + E[Y_i(0)](1 - W_i) \end{aligned}$$

In this framework, we can see that by letting  $W_i = 1$ , we get  $E[Y_i(1)] = E[Y_i|W_i = 1]$ . If  $W_i = 0$ , then  $E[Y_i(0)] = E[Y_i|W_i = 0]$ . This implies that under random assignment, we can identify the average treatment effect as

$$ATE = E[Y_i(1)] - E[Y_i(0)] = E[Y_i|W_i = 1] - E[Y_i|W_i = 0]$$

or by using the difference in the subsample means.

### 3 Ordinary least squares

So at this point you may be asking “Why go over randomized control trials and what does this have to do with econometrics?” The answer to that is that if we have an experimental data at hand, econometrics gives us the method that allows us to analyze key properties in a simple manner.

Suppose that the **population linear regression model** (also known as data generating process in some books) is

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where  $Y_i$  is the dependent variable,  $X_i$  is an independent variable, and  $u_i$  is the error term. We want to estimate the values of  $\beta_0$  and  $\beta_1$ . However, we do not know what the true values of the population parameters -  $\beta_0$  and  $\beta_1$  in this case- are. An alternative way to approach the problem is to use the **sample linear regression model** (or just model)

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$$

where  $\hat{u}_i$  is a residual (not the error, crucial difference!), and the fitted (or predicted) value of  $Y_i$  given some value of  $X_i$  is  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ . The ideal estimator minimizes the squared sum of residuals. Mathematically, this can be obtained by solving the following minimization problem and the first order conditions

$$\begin{aligned} \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \\ [\hat{\beta}_0] : -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \\ [\hat{\beta}_1] : -2 \sum_{i=1}^n X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \end{aligned}$$

The resulting **least squares estimators** are

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

### Proof for deriving $\hat{\beta}_0, \hat{\beta}_1$

From  $[\hat{\beta}_0] : -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$  Note that

$$\begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 &\implies \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \\ \implies \sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i = 0 &\implies \sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i \\ \implies \sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i \end{aligned}$$

Using the fact that  $\sum_{i=1}^n Y_i = n\bar{Y} \iff \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}$ , I get

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

As for  $[\hat{\beta}_1]$  :  $-2 \sum_{i=1}^n X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$  divide both sides by  $-2$  and rearrange

$$\begin{aligned} \sum_{i=1}^n X_i Y_i &= \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 \implies \sum_{i=1}^n X_i Y_i = (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 \\ \implies \hat{\beta}_1 \left( \sum_{i=1}^n X_i^2 - \bar{X} \sum_{i=1}^n X_i \right) &= \sum_{i=1}^n X_i Y_i - \bar{Y} \sum_{i=1}^n X_i \implies \hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i - \bar{Y} \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - n \bar{X}^2} \end{aligned}$$

Note that

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n X_i Y_i - \bar{X} \sum_{i=1}^n Y_i - \bar{Y} \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X} \bar{Y} \\ &= \sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y} - \bar{Y} \sum_{i=1}^n X_i + n \bar{X} \bar{Y} = \sum_{i=1}^n X_i Y_i - \bar{Y} \sum_{i=1}^n X_i \end{aligned}$$

and similarly,  $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n \bar{X}^2$ . So  $\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$

### 3.1 Main Assumptions of the OLS Estimators

For the ordinary least squares to have desired properties of unbiasedness, consistency, efficiency, and asymptotic normality, the following assumptions must be made

**Assumption 3.1.** Here are the assumptions for the classical linear regression model

**A0 Linearity:** The regression is assumed to be linear in parameters.

$$\text{Okay: } Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

$$\text{Not: } Y_i = \beta_0 + \beta_1 X_i + \beta_2^2 X_i + u_i,$$

**A1**  $E(u_i | X_i) = 0$ : This means that conditional on letting  $X_i$  take a certain value, we are not making any systematic error in the linear regression. This is required for the OLS to be unbiased. (Alternately, you can say  $\text{cov}(X_i, u_i) = 0$ , an exogeneity condition.)

**A2 i.i.d. (random sampling):**  $(X_i, Y_i)$  is assumed to be from independent, identical distribution

**A3 No Outliers:** Outlier has no impact on the regression results.  $(E(X_i^4), E(Y_i^4) < \infty)$

**A4 Homoskedasticity:**  $\text{var}(u_i) = \sigma_2$ , or variance of  $u_i$  does not depend on  $X_i$ . If this condition is broken, there exists a heteroskedasticity

**A5 No Autocorrelation (Serial Correlation):** For  $i \neq j$ ,  $\text{cov}(u_i, u_j) = 0$ . In other words, error at the previous period does not have any impact on the current period. This is usually broken in time series settings, where the error in the previous period carries over to the next period.

### 3.2 Sampling distribution of OLS

Note that the OLS estimate that we are getting is a random variable - the estimate we get is different depending on which sample we work with. This is why we can discuss the distributional properties - mean and variance, in particular - of the OLS.

- $\hat{\beta}_1$ : Recall that we can write

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Now, replace  $Y_i$  and  $\bar{Y}$  with

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad \bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u},$$

which allows us to write

$$(Y_i - \bar{Y}) = (\beta_1(X_i - \bar{X}) + (u_i - \bar{u}))$$

and get

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

We are now ready to discuss the distributional properties

- $E[\hat{\beta}_1]$ : It can be written as

$$\begin{aligned} E[\hat{\beta}_1] &= E \left[ \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \\ &= \beta_1 + E \left[ \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \end{aligned}$$

Note that the  $\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})$  can be written to something simpler. This is



equal to

$$\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n X_i u_i - \bar{u} \sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n u_i + n\bar{X}\bar{u}$$

Since  $\bar{X}$  is a sample mean of  $X$ ,  $\sum_{i=1}^n X_i = n\bar{X}$ . Thus, we get

$$\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^n X_i u_i - \bar{X} \sum_{i=1}^n u_i = \sum_{i=1}^n (X_i - \bar{X})u_i$$

The assumption that conditional mean is zero and  $(X_i, u_i)$  are uncorrelated means that the term on the left hand side is zero. Therefore, IF THE CLASSICAL ASSUMPTIONS ARE VALID,  $E[\hat{\beta}_1] = \beta_1$ .

- $var[\hat{\beta}_1]$ : We use the definition of the variances and the fact that the expected value of  $\hat{\beta}_1$  is unbiased (at least for now) to get

$$\begin{aligned} var(\hat{\beta}_1) &= E \left[ (\hat{\beta}_1 - E[\hat{\beta}_1])^2 \right] \\ &= E \left[ (\hat{\beta}_1 - \beta_1)^2 \right] \\ &= E \left[ \left( \frac{\sum_{i=1}^n (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^2 \right] \\ &= E \left[ \left( \frac{(X_1 - \bar{X})(u_1 - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2} + \dots + \frac{(X_n - \bar{X})(u_n - \bar{u})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)^2 \right] \end{aligned}$$

At the moment, we are assuming homoskedasticity and no autocorrelation. Since  $X_i$  is from the data<sup>2</sup> and  $u_i$  is a random error term, we can take all the  $X_i$  terms in and keep the  $u_i$  terms in the expectation to get (i.i.d assumption is also useful here)

$$\begin{aligned} var(\hat{\beta}_1) &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 E[(u_i - \bar{u})^2]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_u^2}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} (\because E[(u_i - \bar{u})^2] = var(u_i)) \\ &= \sigma_u^2 \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} = \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{aligned}$$

Note that to decrease the variance in the estimates, the variance of the error should

---

<sup>2</sup>Here, I am taking a slightly different angle from class. In class, we take  $X_i$  as purely random variable. In that version, you get results that looks like the ones from class. Key takeaways, however, remain the same

be small relative to the variation in the  $X_i$ . Moreover, as the number of observations increase, the variance decreases through increase in the denominator.

At the end of the day, we can say the following about the distribution of our  $\hat{\beta}_1$  estimator and use this to test our hypothesis

$$\hat{\beta}_1 \sim N \left( \beta_1, \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

- $\hat{\beta}_0$ : The formula for  $\hat{\beta}_0$  is  $\bar{Y} - \hat{\beta}_1 \bar{X}$ . By changing  $\bar{Y}$ , we can get

$$\begin{aligned} \hat{\beta}_0 &= (\beta_0 + \beta_1 \bar{X} + \bar{u}) - \hat{\beta}_1 \bar{X} \\ &= \beta_0 + (\beta_1 - \hat{\beta}_1) \bar{X} + \bar{u} \end{aligned}$$

Then we can say the following about the sampling distribution

- $E[\hat{\beta}_0]$ : We can write

$$E[\hat{\beta}_0] = \beta_0 + E[(\beta_1 - \hat{\beta}_1) \bar{X}] + E[\bar{u}] = \beta_0$$

since  $\hat{\beta}_1$  is unbiased and conditional expectation of  $u_i$  is zero. Thus, under our current assumptions,  $\hat{\beta}_0$  is unbiased.

- $var[\hat{\beta}_0]$ : Using the definition of the variance, we can write

$$\begin{aligned} var(\hat{\beta}_0) &= E \left[ (\hat{\beta}_0 - E[\hat{\beta}_0])^2 \right] \\ &= E \left[ (\hat{\beta}_0 - \beta_0)^2 \right] \\ &= E \left[ ((\beta_1 - \hat{\beta}_1) \bar{X} + \bar{u})^2 \right] \\ &= \bar{X}^2 E \left[ (\beta_1 - \hat{\beta}_1)^2 \right] + 2\bar{X} E \left[ (\beta_1 - \hat{\beta}_1) \bar{u} \right] + E[\bar{u}^2] \end{aligned}$$

Under the assumption (A2), we can ignore the middle term as this is zero. The rest of the terms are  $\bar{X}^2 var(\hat{\beta}_1)$  and  $\frac{\sigma_u^2}{n}$ . the final result is

$$var(\hat{\beta}_0) = \frac{\sigma_u^2 \bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} + \frac{\sigma_u^2}{n} = \frac{\sigma_u^2}{n} \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

So we can write

$$\hat{\beta}_0 \sim N \left( \beta_0, \frac{\sigma_u^2}{n} \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

### 3.3 Measure of Fitness

These numbers tell us how informative the sample linear regression we used is in telling us about the population data. In other words, they tell us how closely does our sample regression capture the data. We discussed three types of measure

- **R<sup>2</sup>**: It is defined as a fraction of total variation which is explained by the model. Mathematically, this is

$$\begin{aligned} Y_i &= \underbrace{\hat{\beta}_0 + \hat{\beta}_1 X_i}_{\hat{Y}_i} + u_i, \quad \bar{Y} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 \bar{X}}_{\bar{\hat{Y}}} + \bar{u}, \\ \implies Y_i - \bar{Y} &= (\hat{Y}_i - \bar{\hat{Y}}) - (u_i - \bar{u}) \\ \implies \sum_{i=1}^n (Y_i - \bar{Y})^2 &= \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_{i=1}^n (u_i - \bar{u})^2 - 2 \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(u_i - \bar{u}) \end{aligned}$$

Note that  $\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})(u_i - \bar{u}) = \sum_{i=1}^n \hat{Y}_i u_i - \bar{\hat{Y}} \sum_{i=1}^n u_i - \bar{u} \sum_{i=1}^n \hat{Y}_i + n \bar{u} \bar{\hat{Y}}$ . Since  $\sum_{i=1}^n u_i = n \bar{u}$ ,  $\sum_{i=1}^n \hat{Y}_i = n \bar{\hat{Y}}$  and  $\sum_{i=1}^n \hat{Y}_i u_i = n \bar{u} \bar{\hat{Y}}$ , all the terms cancel each other out. So we are left with

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}_{ESS} + \underbrace{\sum_{i=1}^n (u_i - \bar{u})^2}_{RSS} \implies 1 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} + \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Thus, the  $R^2$  can be found as

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Intuitively, higher  $R^2$  implies that the model explains more of the total variance, which implies that the regression fits the data well.

- **SER**: Standard Error of Regression. It estimate the standard deviation of the error term

in  $Y_i$ , or mathematically

$$SER = \sqrt{\frac{1}{n-2} \sum_{i=1}^n u_i^2}$$

where  $u_i = y_i - \hat{y}_i$  and we use  $n - 2$  since there is loss of d.f. by two due to  $\hat{\beta}_0, \hat{\beta}_1$ . If SER turns out to be large, this implies that our model might be missing a key variable.

- **RMSE:** Root mean squared error. It is similar to SER in terms of how it looks,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{u}_i^2}$$

this is used to assess the accuracy of the predictions. Numerically, the difference between SER and RMSE is minimal and even approximate to identical figure in large sample.