EC 320: Introduction to Econometrics
Lecture Notes

David M. Hall
University of Oregon

Fall 2025

1 Session Overview

- Date:
- Topic:
  - Chapters (Dougherty):
- Goals/Objectives:
  1. One
  2. Two
  3. Three

1.1 Before Class:

2 During Class:

2.1 Welcome

Today's Goals:

- Discuss types of estimators.
- Understand the assumptions of the simple linear regression model.
- Use the assumptions of OLS to show that OLS is BLUE.

2.1.1 Warm Up:

Example 2.1. Write the following equations:

1. Assumed simple linear model.
2. Regression model.
3. Residual.
4. $\hat{\beta}_0$
5. $\hat{\beta}_1$
6. The mean of $X$and Y

2.2 Content Section #1:

Time: Thus far, I have shown you how to obtain the estimates of the OLS parameters. I have said nothing about why we use OLS, or what you may considered caring about in regards to an "estimator." We will in particular discuss two important properties of an estimator: unbiasedness and efficiency. The goal of today is to show, under certain assumptions, that OLS is unbiased.

In the texbook, Dougherty makes a note that these assumptions are all in the context of cross-sectional data. Time does some weird things to data, so there may be different and/or added assumptions based on the estimator used when dealing with *time series* or **panel** data. A cross-section of data can be thought of as a single snapshot in time. For example, you might collect data on the number of parking spots in all of the parking lots on campus on October 26th of this year. We might be able to tell something about the relationship between number of students and parking spots, for example.

A time series would track the number of parking spots over time. Did the total number of spots grow from 1998 through 2025? Panel data would show you the same, but may also distinguish between units of observation by tracking the number of parking spots in each parking lot across the campus.

Definition 2.1 (A.1 Linearity)**.** The model is linear in parameters and correctly specified. In the simple model, this is written as:

$$Y = \beta_0 + \beta_1 X + u \tag{1}$$

Linear in parameters simply means that Xenters linearly, or that there is a linear relationship between X and Y (if one exists). You have most certainly encountered other types of relationships; think of the "function families" you probably discussed sometime in middle (?) school: quadratic, exponential, etc.

If you are excited about non-linear estimators, just note that there are also an abundance of them that you can search up!

Definition 2.2 (Variation Exists)**.** X, the independent variable, must have some variation

Intuitively, we an often think about a regression as determining how the variation in Xdetermines the variation in Y. So if there is no variation in Xthere cannot be any in Y.

Mathematically, if Xhas no variation, then $X_i = \bar{X}\ \forall i$. Take a look at what you wrote for the warm up for $\hat{\beta}_1$. Notice that if $X_i - \bar{X} = 0 \forall i$ then $(X_i - \bar{X})^2 = 0 \forall i$ and we are left with a denominator of 0. $\hat{\beta}_1$ would be undefined and therefore also cannot solve for $\hat{\beta}_0$.

Definition 2.3 (Assumption 3: Disturbance has expectation of 0)**.**

$$E(u_i) = 0 \forall i$$

As Dougherty puts it, this is highly reasonable when we include an intercept in the equation. The intercept represents any constant tendency in Yunaccounted for. If u has a non-zero expectation, we can just subtract that and move it to the intercept.

Visually, we can show that when we allow the disturbance term to be non-zero in expectation, it might not fit the data very closely.

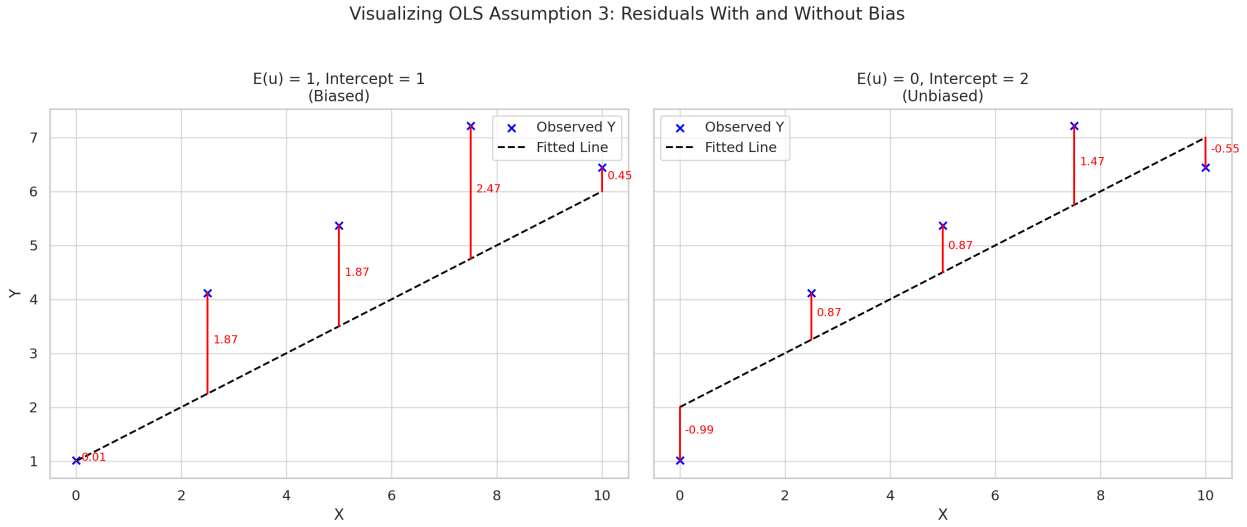

Figure 1: Code for this can be found in the appendix.

We can show mathematically as well that the intercept should be picking up any constant disturbance term expectation:

Supposing $Y_i = \beta_0 + \beta_1 X_i + u_i$, assume that A3 is not true s.t. $E(u_i) = \mu \neq 0 \forall i \in \{1, ...n\}$. We can define some new residual, $v_i = u_i - \mu$ and substitute this in for our equation s.t.

$$Y_i = \beta_0 + \beta_1 X_i + v_i + \mu$$

then we can simply write this as a re-specified linear model:

$$\begin{aligned} Y_i &= \beta_0^* + \beta_1 X_i + v_i \\ \beta_0^* &= \beta_1 + \mu \\ \text{and } E(v_i) = E(u_i - \mu) &= E(u_i) - E(\mu) = \mu - \mu = 0 \end{aligned}$$

Definition 2.4 (Assumption 4: Homoskedastic)**.** We also assume that the disturbance term is **homoskedastic**, meaning that there can be variance in the residual but it must be constant (must not depend on anything).
$E(u_i^2) = E\{(u_i - \mu)^2\} = \sigma^2 \ \forall i$

This is sort of similar to assumption four in that, if there were something systematic about our disturbance term, we might believe that the model should be able to pick it up. For example, suppose you were taking a cross-sectional sample of data of a strawberry field. If we want to test the effect of a pesticide on the number of strawberries, we would want to know how much pesticide was sprayed onto the plant and how many berries are on each. Maybe the ground is uneven on the borders of the plot, though, so you know there are pools of water and that some of the strawberry plants get plenty of water but some on the edge do not. Our model of pesticide to strawberry yield might not do very well on the border; i.e. the variance of our estimator might be higher for border plants than plants in the center. Since this is not a constant variance across all units, this assumption would fail.

Definition 2.5 (Assumption 5: Disturbance is iid)**.** u_i is distributed independently of u_j for all $i \neq j$

There should also be no systematic association between any two observations. The strawberry plants are a good example of this as well: if plant number 3 and number 20 are planted too close together, then they may compete for units and one may end up killing the other. Or maybe there is a really sunny patch where all of the plants get much more son than the others. There would be a tendency for plants in that area to be systematically larger than others.

If this assumption is unmet then we have "autocorrelation." If you plan on working with macroeconomic data or any time series (or panel, but for some reason they/we do not seem to care about autocorrelation) then I recommend researching a bit more on the subject.

Definition 2.6 (Assumption 6: Normal)**.** The disturbance term is distributed normally.

For the most part this is a convenience assumption since we know so much about the normal distribution (we can use t and F tests!), but this is also due to the Central Limit Theorem (CLT).

First, why do we care about the normal and t tests and F tests? Since the estimate of our parameter will have a distribution, we would like have the ability to infer how much we can trust this estimate. So we will be able to do things like test whether or not we can reject a null hypothesis that the true parameter is 0.

The CLT states that *if a random variable is a result of the effects of a large number of other random variables, it will have an approximately normal distribution even if its components do not*. So as long as u is some combination of factors, even if those are non-normal, u will have a tendency to be normal.

2.3 Activity #1:

Discuss with a partner for 1.5 minutes each which assumption makes the most sense, and which makes the least sense. For the one that makes the least sense, why?

2.4 Content Section #2:

Time:

We will spend the second half of class showing that, under assumptions 1-6, OLS is unbiased.

Some important equations/formulas etc from what we've ccovered so far:

1. "True" Model: $Y_i = \beta_0 + \beta_1 X_i + u_i$
2. "Fitting": $\hat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_i$
3. $\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i = \bar{Y})}{\sum_i (X_i - \bar{X})^2}$

Remark 2.1. We make the point to note taht $\hat{\beta}_1$ has a *random* and a *non-random* component. Why? Any randomness enters throguh the disturbance term u, which only Y depends on. Therefore since $\hat{\beta}_1$ is dependent on Y which is dependent on u, $\hat{\beta}_1$ is dependent on u and thus has some randomness.

Let's separate it!

Much of what we are going to do next is simply substituting the definition of Y_i into the equation for the estimator. So let's start with $\hat{\beta}_1$ and sub!

For the numerator, note:

$$\begin{aligned}\sum_i (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_i (X_i - \bar{X})(\beta_0 + \beta_1 X_i + u_i - \beta_0 - \beta_1 \bar{X} - \bar{u}) \\ &= \sum_i (X_i - \bar{X})(\beta_1 [X_i - \bar{X}] + [u_i - \bar{u}] \\ &= \beta_1 \sum_i (X_i - \bar{X})^2 + \sum_i (X_i - \bar{X})(u_i - \bar{u})\end{aligned}$$

therefore, we can show that:

$$\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} = \frac{\beta_1 \sum_i (X_i - \bar{X})^2 + \sum_i (X_i - \bar{X})(u_i - \bar{u})}{\sum_i (X_i - \bar{X})^2}$$

$$= \underbrace{\beta_1}_{\text{non-random}} + \underbrace{\frac{\sum_i (X_i - \bar{X})(u_i - \bar{u})}{\sum_i (X_i - \bar{X})^2}}_{\text{random}}$$

It helps to rewrite the random part a bit. Since $\bar{u} = 0$, then the numerator can be written as $\sum_i (X_i - \bar{X}) u_i$ and we can write the whole thing as:

$$\hat{\beta}_1 = \beta_1 + \sum_i a_i u_i \tag{2}$$

where $a_i = \frac{(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}$

Just a quick note: you **cannot** in practice decompose the estimate like this because we do not actually know what u is. It is unobserved.

2.4.1 OLS Unbiasedness

I want to cover unbiasedness in this lecture and tackle efficiency next. What does it **mean** for an estimator to be unbiased, though?

Definition 2.7 (Unbiasedness)**.** An estimator is unbiased if the expected value of it is equal to the true value of the parameter being estimated.

So, in our case of $\hat{\beta}_1$, we would like to show that:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1$$

Starting with the OLS estimator, we know that:

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

Substitute in the true model $Y_i = \beta_0 + \beta_1 X_i + u_i$. Since both Y_i and $\bar{Y}$ are affected by the error term, we get:

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

Plugging this into the formula for $\hat{\beta}_1$:

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

Now take expectations:

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

Under the assumption that $\mathbb{E}[u_i|X_i] = 0$, it follows that:

- $\mathbb{E}[u_i] = 0$
- $\mathbb{E}[\bar{u}] = 0$
- $\text{Cov}(X_i, u_i) = 0$

Therefore, the numerator has expectation zero:

$$\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})\right] = 0$$

So we conclude:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1$$

□

In general, many unbiasedness proofs follow these steps:

- Start with the definition of the estimator.
- Substitute in the true model (which includes the error term).
- Take expectations, applying assumptions on the error.

2.5 Activity #2 :

Time:

Page 123 of Dougherty

Suppose that $Y_i = \beta_0 + \beta_1 X_i + u_i \ \forall i \in \{1, ..., n\}$. Suppose that our estimator uses the slope from the first to the last point:

$$\hat{\beta}_1^* = \frac{Y_n - Y_1}{X_n - X_1}$$

Show that this is an unbiased estimator of $\hat{\beta}_1$ following the same steps as the last estimator.

2.6 Summary Activity (5 Q's of Reflection):

1. What is the definition of unbiasedness in your own words?
2. Which assumption was the most confusing for you and why?
    - Follow up: how will you try to understand that assumption better?
3. Why do we care about our estimator having a "distribution"?

3 Be Able To

- Given an estimator and a proposed true model, be able to prove an estimator is (un)biased.
- Explain each of the six OLS assumptions in your own words.
- Decompose the OLS estimator into a random and a non-random portion.

4 Key Words

- Fill in Here

5 Appendix:

Code for Figure 1:

```
    # Load libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(patchwork)  # For side-by-side plots

set.seed(320)

# Generate 5 X values and base error terms
x <- seq(0, 10, length.out = 5)
u_base <- rnorm(5, mean = 0, sd = 1)

# Case 1: E(u) = 1, Intercept = 1
u1 <- u_base + 1
y1 <- 1 + 0.5 * x + u1
yhat1 <- 1 + 0.5 * x

# Case 2: E(u) = 0, Intercept = 2
```

```
u2 <- u_base
y2 <- 2 + 0.5 * x + u2
yhat2 <- 2 + 0.5 * x

# Create data frames
df1 <- data.frame(x = x, y = y1, yhat = yhat1, residual = y1 - yhat1, case = "E(u) = 1,
df2 <- data.frame(x = x, y = y2, yhat = yhat2, residual = y2 - yhat2, case = "E(u) = 0,

# Function to plot residuals
plot_residuals <- function(df, title) {
  ggplot(df, aes(x = x)) +
    geom_point(aes(y = y), color = "blue", size = 3) +
    geom_line(aes(y = yhat), linetype = "dashed") +
    geom_segment(aes(xend = x, y = yhat, yend = y), color = "red") +
    geom_text(aes(y = (y + yhat) / 2, label = sprintf("%.2f", residual)),
              color = "red", hjust = -0.2, size = 3) +
    labs(title = title, x = "X", y = "Y") +
    ylim(min(c(df1$y, df2$y)) - 1, max(c(df1$y, df2$y)) + 1) +
    theme_minimal()
}

# Combine plots
p1 <- plot_residuals(df1, unique(df1$case))
p2 <- plot_residuals(df2, unique(df2$case))

p1 + p2 + plot_annotation(title = "Visualizing OLS Assumption 3: Residuals With and With
```