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:

Review the sheet that shows how to create a data frame in R, and also what a for loop is. You'll need/want to bring a laptop with R capabilities to this class period. If you cannot, that is OK as I will share the code after.

2 During Class:

2.1 Welcome

- Explain the difference between our point estimate and the distribution we get.
- Show a Monte Carlo Experiment
- Prove OLS is efficient.

2.1.1 Warm Up:

Example 2.1. Exercise 2.3 in Dougherty (p. 125)

Suppose that the true model is $Y_i = \beta_1 + u_i$ and we are examining an estimator $\hat{\beta}_1 = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Show that this is unbiased.

Answer:

$$\hat{\beta}_1 = \bar{Y} = \frac{1}{n}\sum_i Y_i$$

$$\text{Substitute In } Y_i = \beta_1 + u_i$$

$$\hat{\beta}_1 = \frac{1}{n}\sum_i (\beta_1 + u_i)$$

$$\text{Since } \beta_1 \text{ is a constant, } \sum_i \beta_1 = n \cdot \beta_1 \text{ thus}$$

$$\hat{\beta}_1 = \beta_1 + \frac{1}{n}\sum_i u_i = \beta_1 + \bar{u}$$

$$\text{Taking expectations and noting that } E(u_i|X) = 0 \implies E(\bar{u}) = 0$$

$$E(\hat{\beta}_1) = \beta_1 \quad \square$$

2.2 Content Section #1:

Time:

So last time we discussed the assumptions of the OLS model that make possible the unbiasedness of our estimator. We also discussed some implications of assumption 6: that the disturbance term has a normal distribution.

Namely, we discussed the implication that our estimate has a **distribution** over it, and tehrefore we can do some fancy inference testing.

I want to show you what that looks line in practice. In particular, I want to hammer down the pouint that what you estimate in a regression is a **single point** within this distribution.

I am going to show you that we can simulate many draws from some fake data through a process known as Monte Carlo experiment. Oftentimes people use the words Monte Carlo experiment and simulating interchangably. Researchers use this simulation to test specific properties of their estimators. For instance, if you propose an estimator that you claim is distributed uniformly, you can simulate many different estimations using your esitmator to see if the ending distribution is uniform.

In our case, we will do the same but with the normal distribution. We will generate some data, estimate regressions and then see what the distribution of those estimates looks like.

I now move to R.

2.3 Activity #1:

Discuss with someone what would happen to the distribution if I decreased or increased the standard deviation of either x or u in the simulation. Knowing what you know about hypothesis testing, would you prefer a small or large standard deviation on the estimate?

If you pair with someone with a computer, show another chart of what the distribution of $\hat{\beta}_1$ looks like.

2.4 Content Section #2:

Time:

We will move on to discussing the "precision" of the regression coefficient. You will recall that our estimate of β has a distribution that is characterized by its mean and variance. We showed that OLS estimates the mean, what about the variance? In practice, we do not know the actual variance of the estimate but we can get an estimate.

I will note that you can see in the beginning of Section 2.5 of Dougherty that you can get an expression for the population variance. This hinges on knowing the true data generating process for Y and our assumptions. I will cover in class if there is time. The more important part is to just know that you can write down the population variance of both $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\sigma^2_{\hat{\beta}_1} = \sigma^2_u(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2}) \tag{1}$$

$$\sigma^2_{\hat{\beta}_0} = \frac{\sigma^2_u}{\sum_i (X_i - \bar{X})^2} \tag{2}$$

If Time Allows

We know the definition of the variance for some estimator is the difference between the estimate and the expectation of that estimator squared. For example, for some estimator β:

$$\sigma^2_\beta = E\{(\hat{\beta} - E(\hat{\beta}))^2\} \tag{3}$$

We know a few things that we will use:

- $\hat{\beta}$ is unbiased s.t. $E(\hat{\beta}) = \beta$
- $\hat{\beta}$ can be decomposed into a constant β and a random part s.t. $\hat{\beta} = \beta + \sum_i a_i u_i$.
- $a_i = \frac{X_i - \bar{X}}{\sum_i (X_i - \bar{X})^2}$
- Homoskedasticity: $E(u_i^2) = \sigma^2$
- iid Distribution $E(u_i u_j) = 0$

From the first equation, we can rewrite that in parentheses as $\sum_i a_i u_i$ so that we have:

$$\sigma^2_\beta = E\{(\sum_i a_i u_i)^2\} \tag{4}$$

Expanding this we get:

$$\sigma^2_\beta = \sum_i a_i^2 E(u_i^2) + 2 \sum_i \sum_j a_i a_j E(u_i u_j)$$

Using the two assumptions from above, since $E(u_i^2) = \sigma^2$ and $E(u_i u_j) = 0$ we get:

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

b.t.w. that last part we got to from the Lecture on 2.1-2.3 where I showed that $\sum_i a_i^2 = \frac{1}{\sum_i (X_i - \bar{X})^2}$

We are attempting to estimate the standard deviation of the estimator, which we know has a distribution that follows that depends on the standard deviation of u, our disturbance term and also observed values of X. So if we just get a good estimator for the disturbance term, we can get an estimator for the regression coefficient.

The proposed estimator is the following:

$$\hat{\sigma^2} = \frac{n}{n-2} MSD(\hat{u})$$
$$\text{where } MSD(\hat{u}) = \frac{1}{n}\sum_i (u_i - \hat{\bar{u}})^2 = \frac{1}{n}\sum_i \hat{u}_i^2$$

We use the MSD because it is a good measure of how much the residuals are scattered all around.

What we want to get is the estimate of the standard deviation of the regression coefficient, which we will call the **standard error**. I have already given you the population variance for β_0 and β_1 in our standard linear model (will do so again below), and now you have a nice estimator for σ_u^2 as well. Using these three equations, take some time to get two expressions: one for the standard error of $\hat{\beta}_0$ and one for $\hat{\beta}_1$. Hint: the standard deviation is, of course, the square root of the variance.

Equations:

1. $\sigma^2_{\hat{\beta}_1} = \sigma_u^2(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2})$

2. $\sigma^2_{\hat{\beta}_0} = \frac{\sigma_u^2}{\sum_i (X_i - \bar{X})^2}$

3. $\hat{\sigma_u^2} = \hat{\sigma^2} = \frac{n}{n-2} MSD(\hat{u})$
   where $MSD(\hat{u}) = \frac{1}{n}\sum_i (u_i - \hat{\bar{u}})^2 = \frac{1}{n}\sum_i \hat{u}_i^2$

You should get something along these lines:

$$s.e.(\hat{\beta}_1) = \sqrt{\hat{\sigma}_u^2(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (X_i - \bar{X})^2}}$$
$$s.e.(\hat{\beta}_0) = \sqrt{\frac{\hat{\sigma_u^2}}{\sum_i (X_i - \bar{X})^2}}$$

2.5 Activity #2 (Optional):

Time:

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

1. What two sources of variation contribute to the standard error of the estimator?

2. Why do we care about the standard errors?

3. On a scale of 1-10 how do you feel about your understanding of Monte Carlo Simulation?

3 Be Able To

- Write out a be able to - separate but related to the goals of this lecture.

4 Key Words

- Fill in Here