Recitation Two - Regression
Econometrics - Fall 2018

Nathaniel Mark

October 16, 2018

1 Topic One: Intuition of regression

1.1 Functional Form Assumption/ TRUE MODEL

In economics, a question we often want to answer is *what is the effect of variable A on variable B?* For example, *what is the effect of getting older on your expected income*? One way to answer this is to assume that variable B is a linear function of A. Using our example, we assume that **in the true world** the relationship takes the form:

$$Income_i = \beta_0 + \beta_1 Age_i + \epsilon_i$$

This says that the income of person i can be split into two pieces: 1) a linear function of age; 2) a part that is unexplained by age. β_1 is the increase in expected income by increasing age by one year. β_0 is the expected income at age 0. ϵ_i is everything else that effects income, by definition (just rearranging, $\epsilon_i = Income_i - (\beta_0 + \beta_1 Age_i)$. We call this the *error* or the *unexplained* part of income.

Note that we are assuming that in the **true world** there is a linear line that describes the relationship between income and age. This is called our *functional form* assumption. This is not necessarily correct, of course. Generally, we think of linear models as approximations of the true world. But, we have to assume some kind of model to estimate it and linear is convenient. Later in the class we talk about more realistic models for different kinds of data.

1.2 Estimation

We have now assumed what the true world looks like by writing down our model. Now, how do we use our data to learn about the true world? We use our data to estimate the parameters in our model. That is, if we can estimate what the value of β_0 and β_1 are, we can describe what the relationship between income and age is.

What estimator do we use to estimate β_0 and β_1? We want to find the values of β_0 and β_1 that best fits the data. To do so, we use a definition of "best fit": in *ordinary least squares* (aka regression), *the definition of the best fitting line is the line that minimizes the sum of squared errors.* That is, it is the line that minimizes the sum of squared distances between $Income_i$ and $\beta_0 + \beta_1 Age_i$. Therefore, our estimator of β_0 and β_1 are defined as the possible values that minimize that sum of squared distances:

$$min_{\beta_0,\beta_1} \sum (Income_i - (\beta_0 + \beta_1 Age_i))^2$$

After some algebra, we find that the solution to this minimization process is:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

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

2 Properties of OLS estimators

What properties do these estimators have? When?

2.1 Very Important Theorem

IF:
The following three assumptions hold **in the true world**:

- Assumption 1: $E[\epsilon_i|X_i] = 0$
  That is, X_i cannot explain ϵ_i as all. Imprecisely, ϵ_i and X_i are uncorrelated.
  This assumption implies a few important things:

1. $E[\epsilon_i|X_i] = 0 \Rightarrow Cov(X_i, \epsilon_i) = 0$
2. $E[\epsilon_i|X_i] = 0 \Rightarrow Corr(X_i, \epsilon_i) = 0$
3. and if iid, $E[\epsilon_i|X_i] = 0 \Rightarrow Cov(g(X_1, X_2, ...), \epsilon_i) = 0$ for any function g().

- Assumption 2: (X_i, Y_i) are i.i.d.
  That is, observation i is chosen randomly from the population. For example, we cannot draw random families and keep all people as observations, as siblings would be related.

- Assumption 3: No outliers ($E[X_i^4] < \infty$ and $E[Y_i^4] < \infty$).

THEN:

1. $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased.

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

and

$$[\hat{\beta}_0] = \beta_0$$

That is, on average, our estimator will be correct (think: if we run our experiment 999,999,999 times and take the average of our estimate, that average will be correct).

2. $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent.

$$\hat{\beta}_1 \to \beta_1$$

That is, as N goes to infinity, our estimate gets closer and closer to the true value. More precisely, as N goes to infinity, our estimate gets arbitrarily close to the true value.

3. $\hat{\beta}_0$ and $\hat{\beta}_1$ are asymptotically normal*.

$$\hat{\beta}_1 \to^d N(\beta_1, SE(\hat{\beta}_1))$$

That is, as N goes to infinity, the distribution of possible values that the estimator $\hat{\beta}_1$ could take on converges to a normal distribution.

In short, under the OLS assumptions, our estimators are good. If our OLS assumptions are not true, then our estimators are not good! This is the biggest problem in econometrics: trying to find situations where the OLS assumptions are valid.

3 Proofs of properties

3.1 Unbiasedness

The first proof we went over in class is the unbiasedness of β_1 (slightly different than Prof. Bai's):

Step Zero (Not part of proof, just a reminder): Write out the "true" model: in our case

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

Step One: Write out the definition of your estimator:

$$\hat{\beta_1} = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Step Two: Plug in the true model::

$$\hat{\beta_1} = \frac{\frac{1}{n}\sum(X_i - \bar{X})(\beta_0 + \beta_1 X_i + \epsilon_i - \beta_0 - \beta_1 \bar{X} - \bar{\epsilon})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Where $\bar{\epsilon} = 0$ by construction, so

$$\hat{\beta_1} = \frac{\frac{1}{n}\sum(X_i - \bar{X})(\beta_0 + \beta_1 X_i + \epsilon_i - \beta_0 - \beta_1 \bar{X})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Step Three: Simplify:

$$\hat{\beta_1} = \frac{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})\beta_1}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})} + \frac{\frac{1}{n}\sum(X_i - \bar{X})\epsilon_i}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

$$\hat{\beta_1} = \beta_1 + \frac{\frac{1}{n}\sum(X_i - \bar{X})\epsilon_i}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})} \bigstar$$

Step Four: Take Expectations of both sides and simplify:

$$E[\hat{\beta}_1] = E[\beta_1] + E[\frac{\frac{1}{n}\sum(X_i - \bar{X})\epsilon_i}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}]$$

$$E[\hat{\beta}_1] = \beta_1 + E[E[\frac{\frac{1}{n}\sum(X_i - \bar{X})\epsilon_i}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})} |X_1, X_2, ...X_n]$$

by the Law of Total Expectation ($E[E[Z|W]] = E[Z]$)

$$E[\hat{\beta}_1] = \beta_1 + E[\frac{\frac{1}{n}\sum(X_i - \bar{X})E[\epsilon_i|X_1, X_2, ...X_n]}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}]$$

since, conditional on $X_1, ..., X_n$, X_i is just a constant – it can be pulled out of the expectation.

$$E[\hat{\beta}_1] = \beta_1 + E[\frac{\frac{1}{n}\sum(X_i - \bar{X})0}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}]$$

by the OLS assumption 1.

$$E[\hat{\beta}_1] = \beta_1 + 0 = \beta_1$$

Follow these same steps for the β_0 unbiasedness proof and you will be fine!

3.2 Consistency

The second proof we went over in class is the unbiasedness of β_1 (again, slightly different than Prof. Bai's):

Step Zero through three are almost the same as the unbiasedness proof, so you can start with ★ if you like (and add a $\bar{\epsilon} = 0$).

Step Zero (Not part of proof, just a reminder): Write out the "true" model: in our case

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

Step One: Write out the definition of your estimator:

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Step Two: Plug in the true model::

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum(X_i - \bar{X})(\beta_0 + \beta_1 X_i + \epsilon_i - \beta_0 - \beta_1\bar{X} - \bar{\epsilon})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

Step Three: Simplify:

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})\beta_1}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})} + \frac{\frac{1}{n}\sum(X_i - \bar{X})(\epsilon_i - \bar{\epsilon})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum(X_i - \bar{X})(\epsilon_i - \bar{\epsilon})}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})} \bigstar$$

Step Four: Apply Law of large numbers to all terms:

$$\beta_1 \to \beta_1$$

since β_1 is a constant.

$$\frac{1}{n}\sum(X_i - \bar{X})(\epsilon_i - \bar{\epsilon}) \to Cov(X_i, \epsilon_i)$$

since $\frac{1}{n}\sum(X_i - \bar{X})(\epsilon_i - \bar{\epsilon})$ is simply the sample estimate of $E[(X_i - E(X_i))(\epsilon_i - E(\epsilon_i))]$. i.e. the LLN applies.

$$\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X}) \to Cov(X_i, X_i)$$

since $\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})$ is simply the sample estimate of $E[(X_i - E(X_i))(X_i - E(X_i))]$. i.e. the LLN applies.
Therefore,

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum(X_i - \bar{X})\epsilon_i}{\frac{1}{n}\sum(X_i - \bar{X})(X_i - \bar{X})}$$
$$\to \beta_1 + \frac{Cov(X_i, \epsilon_i)}{Cov(X_i, X_i)}$$
$$\to \beta_1 + \frac{0}{Cov(X_i, X_i)}$$

by OLS assumption 1. Therefore,

$$\hat{\beta}_1 \to \beta_1$$