ECO383 lecture - October 27, 2011

Supplementary notes.

Given we covered a couple of somewhat complicated things today, I would like to provide a quick sketch that may help you see how the pieces fit together.

Our starting point was the formula we derived for b^{OLS}, namely

$$b^{OLS} = \beta + \frac{\sum_i (C_i - \bar{C}) u_i}{\sum_i (C_i - \bar{C})^2}. \qquad (1)$$

This comes from the earlier formula

$$b^{OLS} = \frac{\sum_i (C_i - \bar{C}) T_i}{\sum_i (C_i - \bar{C})^2}, \qquad (2)$$

then substituting for T_i using

$$T_i = \alpha + \beta C_i + u_i, \qquad (3)$$

which is our 'true' model. [Hint: you should definitely make sure you can derive (1) from (2) and (3).]

Equation (1) is important. It makes clear that b^{OLS} depends on a (weighted) sum of random variables $\{u_1, u_2, u_3, \dots, u_N\}$. These impart randomness to our estimator, implying that b^{OLS} is itself a random variable.

Given that b^{OLS}, our slope estimator, is a random variable, we want to know at least three things:

i) what is its expected value, $E(b^{OLS})$?

ii) what is its variance, $Var(b^{OLS})$?

iii) how is the random variable distributed?

We should approach these questions in that order, and be very clear what minimum assumptions are needed at each stage.

i) $E(b^{OLS})$.

If we assume

$$E(u_i) = 0, \text{ for all } i \quad [\text{Assumption 1}]$$

then we can show $E(b^{OLS}) = \beta$.

This is the important property of <u>unbiasedness</u>.

ii/ $Var(b^{OLS})$

We saw that demonstrating that

$$Var(b^{OLS}) = \frac{\sigma^2}{\sum_i (C_i - \bar{C})^2} \qquad (4)$$

is more involved. We started by noting that

$$Var(b^{OLS}) = E\left[\left(b^{OLS} - E(b^{OLS})\right)^2\right]$$

$$= E\left[\left(\beta + \frac{\sum_i (C_i - \bar{C})u_i}{\sum_i (C_i - \bar{C})^2} - \beta\right)^2\right]$$

$$= E\left[\left(\frac{\sum_i (C_i - \bar{C})u_i}{\sum_i (C_i - \bar{C})^2}\right)^2\right]$$

Now, we can expand the bracket, and this will yield $N \times N$ terms in total. (If N were 2, then we would generate 4 terms, as we saw.) Of these:

- there are N terms of the form

$$(C_i - \bar{C})^2 u_i^2$$

- and there are $(N-1) \times N$ terms of the form

$$(c_i - \bar{c})u_i \times (c_j - \bar{c})u_j, \text{ where } i \neq j.$$

We can make a lot of progress by assuming

$$Var(u_i) = E(u_i^2) = \sigma^2, \text{ for all } i.$$

[We showed in class the equivalence between $Var(u_i)$ and $E(u_i^2)$ if $E(u_i) = 0$, which we assume according to Assumption 1.]

Let us label this as Assumption 2. Also,

$$Cov(u_i, u_j) = E(u_i u_j) = 0, \text{ for all } i \neq j.$$

We will call this Assumption 3.

By Assumption 2, each of the N terms of the form

$$(c_i - \bar{c})^2 u_i^2$$

has an expected value of

$$(c_i - \bar{c})^2 \sigma^2,$$

giving a total sum of

$$\sigma^2 \sum_i (c_i - \bar{c})^2.$$

By Assumption 3, each of the $(N-1) \times N$ terms, which we can write as

$$(c_i - \bar{c})(c_j - \bar{c}) u_i u_j \quad , \text{ for } i \neq j$$

has an expectation $= 0$. So all these drop out.

This allows us to obtain, with a bit more manipulation, the expression in (4) above, namely

$$\mathrm{Var}\left(b^{OLS}\right) = \frac{\sigma^2}{\sum (c_i - \bar{c})} \, .$$

[Please make sure you are clear how the relevant assumptions come into play in the above derivations. If not, please let me know.]

iii/ Distribution of b^{OLS}

Note that we have yet to say anything about the <u>distribution</u> of b^{OLS}, even though we have been able to characterize some of its properties.

Yet in order to carry out statistical inference (hypothesis testing), we need to commit to the error term following a certain distribution. And given that b^{OLS} is a weighted sum of the error terms, it will follow a certain distribution also, as a consequence.

The typical assumption made is as follows:

$$u_i \sim N(0, \sigma^2), \text{ for all } i.$$

We will label this Assumption 4.

It turns out that one very convenient property of the Normal distribution (which we will acknowledge and use) is that the sum of normal random variables is itself normally distributed.

In this instance, given Assumption 4, it turns out that

$$b^{OLS} \sim N\left(\beta, \frac{\sigma^2}{\sum_i (C_i - \bar{C})^2}\right).$$

Once we have this, we are then in a position to do some hypothesis testing (to follow...)