Notes 12: OLS Theorems, Confidence Intervals

ECO 231W - Undergraduate Econometrics

Prof. Carolina Caetano

1 Does the OLS guess close?

We continue our study of the magnitude of the mistakes we will make if we use the OLS to estimate the coefficients of the linear model. Recall from last class:

Theorem 2. *If assumptions 1, 2 and 3 hold, then*

$$Var[\hat{\beta}_j] = \frac{\sigma^2/n}{\widehat{Var}(x_j)(1-R_j^2)}, \qquad \text{for all } j = 0, 1, \ldots, k.$$

where $\widehat{Var}(x_j) = \frac{1}{n}\sum_{i=1}^{n}(x_{ji} - \bar{x}_j)^2$*, and* R_j^2 *is the* R^2 *from a regression of* x_j *on all the other regressors.*

Both assumption 1 and 2 had two requirements, and assumption 3 had one. What were they?

1. ____
2. ____
3. ____
4. ____
5. ____

The theorem tells us that the OLS estimator has a variance with the formula above. However, it does not really tell us whether the OLS guesses close or not, because although we know the formula of the variance we don't really know its value. Why? __________

____

Since the variance of $\hat{\beta}_1$ is based on a quantity we don't know, we have to estimate it. Here is the estimator:

In other words, average the squared residuals of the OLS regression. Observe that you are not dividing them by n, you are dividing them by $n-k-1$. The reason why we divide by $n-k-1$ is a bit tricky to explain, but I will try. The sum $\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{1i}-\cdots-\hat{\beta}_k x_{ki})^2$ seems to have n terms. However, remember that the OLS estimates were obtained with $k+1$ equations (I suggest you go back to the class where we saw the multivariate regression method to review that). This means that the $\hat{\beta}$'s are such that the data has to satisfy $k+1$ equations. This is equivalent to saying that $k+1$ terms of this sum are already determined by the remaining terms. They are fixed, there is nothing random about them. There are only $n-k-1$ terms that are actual variables, that are free to vary. We even have a name for this, we say that our sum has $n-k-1$

____,

exactly because the remaining k terms are not free at all. So, since there are actually only $n-k-1$ terms in this sum, we should divide them by $n-k-1$ to average.

Little mathematical note for those interested: if you want to understand this point in depth, experiment opening up the sum in the univariate case, $\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{1i})^2$, and remember that the OLS coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy the equations: $\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{1i})=0$, and $\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{1i})x_{1i}=0$. Hint: you don't need to solve for $\hat{\beta}_0$ and $\hat{\beta}_1$ at all.

All right then, we have an estimator of the unknown part of the variance. Our variance estimator is

As always, we want to deal with a measure that has the right unit. So, we would like to estimate the standard deviation of $\hat{\beta}_1$. It turns out that the estimator of the standard deviation of $\hat{\beta}_1$ has a special name. We call it:

____

The formula is:

Now we know exactly how precise our estimator is. For example, suppose that we want to know the effect of class attendance on the final grade. The model is

$$grade = \beta_0 + \beta_1 class + \beta_2 OH + \beta_3 sections + u,$$

where $\mathbb{E}[u|class, OH, sections] = 0$, so assumption 1 is satisfied. Suppose that assumptions 2 and 3 are also satisfied. We run an OLS regression, and $\hat{\beta}_1 = 2$. We are tempted to say that each additional class increases the final grade in 2 points. However, $\hat{\beta}_1$ is not β_1, it is just our guess. We know that in average we will be right, because of theorem 1 (that says that the OLS is unbiased). However, how precise is our guess?

- Suppose that you calculated the standard error, and $SE(\hat{\beta}_1 = 2) = 0.1$. In that case, will you be confident about your guess? ______________________________
- Suppose that you calculated the standard error, and $SE(\hat{\beta}_1 = 2) = 1$. In that case, will you be confident about your guess? ______________________________

In the coming classes, we will formalize the idea of "confidence." We will study how to decide if we are confident or not, what is the threshold.

Before we finish this topic, there is still one question to answer, and it refers to the OLS estimator in comparison to other possible estimators of the coefficients of the linear model. By the way, although we are not going to study other estimators in this course, there are plenty of others to go by. Here are some other estimators you could look up: GLS (Generalized Least Squares), WLS (Weighted Least Squares), 2SLS (Two Stage-Least Squares), MM (Method of Moments), GMM (Generalized Method of Moments), MLE (Maximum Likelihood Estimator), LAD (Least Absolute Deviations estimator). Even more advanced, you could search online: kernel estimation, local polynomial estimation, Fourier series estimation, power series estimation, spline regression, penalized spline regression, wavelet regression.

2 Is OLS the best?

That depends on the assumptions, and on what you define as "best." We will compare the OLS to all the estimators that also guess right. Of course, in this case the best is the one that guesses the closest. Here is the definition:

- **B.L.U.E. (Best Linear Unbiased Estimator):** ______________________

  ____

  ____

So, is the OLS estimator BLUE? (That's how we write it. I know it doesn't make sense. It should be "Is the OLS estimator the B.L.U.E.?") The famous Gauss-Markov Theorem answers this question.

Theorem 3. ____

____

Let me just go over this in detail. This theorem states that if the population satisfies assumption 1 and 3, and the data satisfies assumption 2, then the OLS is, among all the linear estimators that are unbiased, the one with the smallest variance. You know what that means? It means that if the assumptions above are satisfied, then you should not have any doubt, you should pick the OLS! The question is always whether the assumptions are satisfied or not. For now, take a moment to savor the fact that at least we know exactly what to do in this case.

3 Confidence Intervals

We will assume that we still are in this nice clean world where assumptions 1, 2 and 3 are true. We know that we should use the OLS estimator. We calculated the $\hat{\beta}_1$ using our data, and we even calculated the standard error $SE(\hat{\beta}_1)$. Now, what do we do? Do we say that our $\hat{\beta}_1$ is the same as the true β_1? No, we cannot say that, ever. The fact is that to get the β_1 exactly right would be a miracle. That is the problem with just looking at $\hat{\beta}_1$. It is a number, and thus it is either equal to β_1 or not. That doesn't help us at all.

However, suppose that instead of concerning ourselves with a number, which can be right or wrong, we asked ourselves: I want to think of an interval, and I want to be 95% confident that my interval has got the true β_1 trapped inside. You see, we can never be 100% confident, after all, we don't have the whole population. All we are doing is based on the sample we got, and there is always a possibility that our sample is not very good. Sure, we are requiring that the sample be random, which is good, but we could be unlucky.

We could have gotten a sample that was not representative of the population at all, just by chance. So, we can never be 100% confident, but it turns out that we can be very confident anyway. How confident? As confident as you want. The problem is, the more confident you want to be, the bigger the interval. So, the interval depends on your confidence level.

Suppose you want to be 95% confident. What is the 95% confidence interval of β_1?

Observe the details of this definition. I didn't say "It is the interval that 95% of the times has the true β_1 inside." This would not make sense. An interval either has β_1 or not, there is nothing random about it. No, I will give you a formula that, if applied to random samples from the population with the same number of observations as ours, 95% of times the resulting interval would cover the true β_1. We gamble with a 95% chance that the sample we got will yield one of those intervals that covers the true β_1. However, once we calculated the interval in our sample, the resulting interval either has the β_1 or it doesn't, and we have no way of knowing it.

So, if you want to be 95% confident, I can give you an interval. What if you want to be 99% confident? I can give you an interval as well. Would that interval be bigger or smaller than the 95% one? ____________________

Sure, if you want to be more confident, you will have to take less risk, and therefore pick a bigger interval.

In order to give you the formula, let me use the notation usually adopted in other books. If you want to be $(1-\alpha)\cdot 100\%$ confident. Wait, stop there. What? Yeah, this is how they write it, there is a reason, you will see. So, for example, if you want to be 95% confident, then it's the same as having $\alpha = 0.05$. If you want to be 99% confident, then it's the same as having $\alpha = 0.01$. If you want to be 90% confident, then it's the same as having $\alpha = 0.1$. Do the math, you will see.

Anyway, if you want to be $(1-\alpha)\cdot 100\%$ confident, the interval formula for β_1 is

where $z_{1-\frac{\alpha}{2}}$ is the critical value of the normal distribution which corresponds to the probability $1-\frac{\alpha}{2}$.

That seems obscure, but it's not. Suppose you get a normal (some people say gaussian) random variable X, then you want the value $z_{1-\frac{\alpha}{2}}$, which satisfies

$$\mathbb{P}(X \leqslant z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}.$$

So, for example, if the confidence is 95%, then $\alpha = 0.05$, and thus $z_{0.95}$ is the 95th percentile of the normal distribution. The most common confidence values are 90%, 95%, and 99%, and the critical values are respectively $z_{0.95} = 1.64$, $z_{0.975} = 1.96$, and $z_{0.995} = 2.58$.

So, if we want to be 95% confident, the formula of our interval is:

- NOTICE: Most of the time, we like to use the confidence level of 95%, which corresponds to $z_{0.975} = 1.96$. However, you can always use 2 instead of 1.96. It is a little higher, but then again, it's not as if our model was exactly precise either, right? So, it doesn't harm to go just a tad above, and also simplify the math, right? Do you see the trick now? The 95% confidence interval is roughly 2 standard errors below and above $\hat{\beta}_1$. No bad for a rule of thumb.

So, suppose that we estimated the coefficients and got $\hat{\beta}_1 = 2$, and $SE(\hat{\beta}_1) = 0.1$, then the 95% confidence interval is

This is very precise. For example, if we want to decide whether to miss two classes, we can be quite sure that our grade will decrease between 3.6 and 4.4 points. This is quite informative. It means that we will probably go down one grade level, say from an A to an A-, probably no more and no less. However, suppose that $SE(\hat{\beta}_1) = 1$, then the 95% confidence interval is

This means that at the same level of certainty as in the other example, we can expect our grade to decrease between 0.08 and 7.92. This is not informative at all. We could basically not be affected at all, since 0.08 is almost nothing out of 100, or we could be in big trouble, since 7.92 points could be the difference between an A- and a C+!

So, the OLS is more precise than other estimators, but how precise is it? It depends on the situation. In the case of the problem above, $SE(\hat{\beta}_1) = 0.1$, the OLS is quite precise. If $SE(\hat{\beta}_1) = 1$, the OLS is not precise enough.