Econometrics 1 (ECON 4003)
Suggested Solutions - Tutorial 5

Max Schroder

October 31, 2019

Abstract

This guide is supposed to be complementary to the official solutions supplied by the lecturer. All errors are my own.

Question 1

This question is an example of an application of a *log-log model.* Economists like to use this transformation because it has a really nice interpretation. However, using a log transformation of a variable is not without its problems: whenever $x \leq 0$ then $\log(x)$ is not defined, so be careful. In the case of distance it seems like an innocuous transformation, but this is not going to be true in all cases.

a)

Because this can be confusing sometimes (especially to myself), we will derive, how to show that in the *log-log model, "a 1% increase in the explanatory variable is associated with a 1% increase in the dependent variable"*. Let's suppose we have the following regression model (ignoring the intercept for convenience):

$$\ln(y) = \beta \ln(x) \tag{1}$$

Now increase x by 1% and take the difference between the two predicted values:

$$\triangle \ln(y) = \beta \ln(x + \frac{1}{100}x) - \beta \ln(x) \tag{2}$$

$$\triangle \ln(y) = \beta(\ln(x + \frac{1}{100}x) - \ln(x)) \tag{3}$$

using the rule of logarithms:

$$\triangle \ln(y) = \beta \ln(\frac{x + \frac{1}{100}x}{x}) \tag{4}$$

$$\triangle \ln(y) = \beta \ln(1 + \frac{1}{100}) = \frac{\beta}{100} \tag{5}$$

using the approximation that $log(1+x) \approx x$ for small $x.Now$ the last thing we have to realize is that $\triangle \ln(y) \approx \frac{\triangle y}{y}$ which gives us:

$$100\frac{\triangle y}{y} = \beta \tag{6}$$

where the LHS is just the definition of a percentage change in y.

b)

Here we need to show that $E(u|X) \neq 0$ which violates one of the assumptions that makes the OLS estimator BLUE. I think the easiest way is to remember that $E(u|X) = 0$ implies $E(uX) = 0$ and then think of some reasons why the residual and the explanatory variable may be correlated. At the end of the day, determining whether the error term is uncorrelated with the regressors is more of an art than a science, and much of your job will be to justify, your particular assumptions.[1]

c)

Remember the residual/error term is essentially "all the things that are affecting our dependent variable, which we have not explicitly modelled". I always find it a good exercise to think: "If two units of observation have the same value of X. Do they have different values of Y? And if yes, what reasons might there be for this difference?" This will help you to think of a good interpretation for the error term, and might help you spot issues with your OLS assumptions.

Question 2

Random assignment is so popular, since if you do it correctly, it should theoretically guarantee $E(u|X) = 0$. However, randomization is not always easy to achieve. There might be issues with unconscious selection on part of the experimenter or the subjects, attrition (i.e. people (selectively) dropping out of the study over time), spillovers (people in the treatment group affecting those in the control) and many other issues.[2]

a)

It is important to remind yourself, that u_i is effectively measured in the same units that Y_i is. So if we think of the outcome being measured in # of questions answered correctly, then the residual will have a similar interpretation. This might help you to think about an interpretation for u_i.

[1] Always beware of the endogegeity taliban!

[2] Have a lookat Angrist's 1990 Vietnam Draft Lottery paper, which is quite possibly one of the most illustrative examples of these issues.

b)

If randomization is done correctly, then X is independent of u and the assumption is satisfied. The success of a randomization procedure relies in many ways on the process of "averaging out" other factors and might therefore be sensitive to sample size. For example, if you only had a handful of participants in your experiment, then there might very well be systematic unobserved differences in the two trial groups, even though randomization should have taken care of this in theory. In these settings, researchers sometimes resort to "stratified randomization", which is the process of sorting participants into groups according to some observable characteristic and then randomizing within these groups. For example, if you had two participants in your study that had an IQ of 200, you might want to assign them actively to separate groups.

Question 3

This question is really not very different to many others that we have solved before so I don't mean to dwell on it for too long. The main thing that I believe would be useful is to think about the coefficients representing not only numbers but also units. For example in this example the dependent variable is measured in lb's. Hence the intercept and the error term are denoted in terms of weight - rather than simply being arbitrary numbers. Similarly, the coefficient for $\hat{\beta}_1$ measures $\frac{lb}{inch}$. So for an individual that has a height of 70 inches we get:

$$\hat{\beta}_0 lb + \hat{\beta}_1 \frac{lb}{inch} * 70inch = \hat{\beta}_0 lb + \hat{\beta}_1 70lb \tag{7}$$

As you can see, explicitly writing down the units might help you with transforming them in question d).

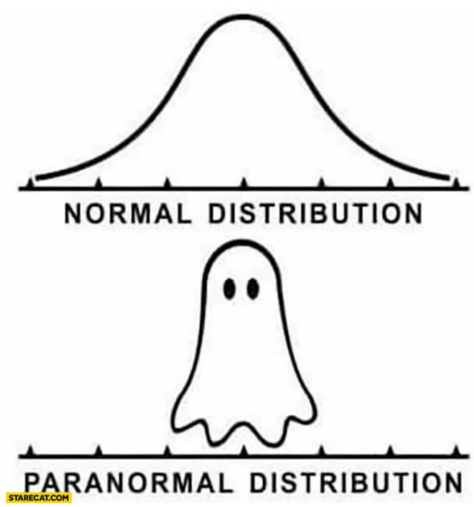

Figure 1: Source: starecat.com