Economics 312 Spring 2020

Project #2 Assignment Due: 11:59pm, Mon., February 10

Partner assignments

Shisham Adhikari	George Zhao
Blaise Albis-Burdige	Matt Wan
Isabelle Caldwell	Roy Shannon
Anne Cao	Amrita Sawhney
Aditya Gadkari	Lauren Rabe
Catherine Gong	Grisha Post
Lirui Jiao	Evian Oosthuizen
Jonathan Li	Max Nobel
Charlie Lyu	Paul Nguyen
Peter Mulgrew	

This project involves repeating the Monte Carlo simulation that was performed in class, then modifying it to allow for alternative data-generating processes. Each student team is to submit one paper that reflects your joint work on the project. Your joint work should follow the guidelines at https://www.reed.edu/economics/parker/312/asgns/guide.html. Because the number of students in the class is odd, one student will work alone each week. I will expect that the solo student might end up asking me more questions through the process.

1. **Regression with normal errors.** Repeat the in-class Monte Carlo simulation using 10,000 replications, saving the results for both the slope coefficient (b = _b[x]) and the standard error of the slope coefficient (se = _se[x]) in a Stata data set. The data file and do file used in class can be downloaded from the links on the assignment Web page. Then analyze the following questions about the distribution of the resulting estimates:
   a. Show the summary statistics for your OLS estimates of the coefficient and the standard error. Does the coefficient estimator seem unbiased? How did you come to that conclusion?
   b. Theory says that the variance of the sampling distribution of the OLS slope estimator is given by Wooldridge's equation [2.57]. In our case, we are controlling σ^2, but in order to apply this formula to our sample, you must calculate $SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$, the sum of squared x deviations. It is possible to calculate this mechanically, but the easiest way is by working backwards from the sample standard deviation of x, $s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, which is reported by the summarize command. After computing $\sum_{i=1}^{n}(x_i - \bar{x})^2$, compute the *theoretical*

standard deviation of the OLS slope estimator's sampling distribution using equation [2.57]. Compare the computed repeated-sample standard deviation of your OLS slope estimates from the Monte Carlo simulation with the theoretical standard deviation you calculated from equation [2.57]. Are they close or is there a sizable discrepancy between the actual dispersion of your estimator (in repeated sampling) vs. the theoretical value it should have?

c. The OLS *standard error* is an estimator for this standard deviation of $\hat{\beta}_1$. Examine the mean of the distribution of your 10,000 estimated OLS standard errors. Is the OLS standard error a good estimator of the standard deviation of the coefficient estimator in your simulation?

d. Plot a histogram of your OLS coefficient estimates. Theory says that they should follow the normal distribution. Does this seem plausible?

e. Use the Stata Help menu to find a suitable test for whether a random variable follows the normal distribution. Look up in the pdf or printed Stata manuals the procedure used for your chosen test. Describe the test both intuitively (what's the basic idea) and computationally (what's the formula), and justify why it is an appropriate test in this context. Based on the results of the test, are you comfortable concluding that the OLS estimator follows a normal distribution? [*Students frequently "find" estimators or tests in Stata that are not in the textbook. This is OK, but, as in this problem, you need to use the Stata documentation to determine how it works and to justify its use.*]

2. **Regression with uniform errors.** We now re-run the simulation with an error term that follows a uniform distribution rather than the normal distribution. According to theory, the sampling distribution of the OLS estimator in this case is *asymptotically* normal, so it should converge to a normal distribution as *n* gets large. Your *n* (157) is pretty large, but not extremely large, so it is an interesting question whether asymptotic normality is valid for this sample.

   a. The two parameters of the uniform distribution are *a* and *b*, the minimum and maximum values that the variable can have. Any value between *a* and *b* is equally likely. We want to make the mean and variance of our uniform distribution the same as the normal distribution in the previous part as possible, so we want a mean of zero and standard deviation of 0.3 (variance of 0.09). If *u* follows a uniform distribution between *a* and *b*, then its mean is $E(u) = (a+b)/2$ and its variance is $\text{var}(u) = (b-a)^2/12$. Use these formulas to calculate the values of *a* and *b* that make the mean zero and the variance 0.09. (Suggestion to check your work: Once you think you have the formula right, try creating a variable and use the `summarize` command see if the mean and standard deviation are close to zero and 0.3, respectively.)

b. Run a Monte Carlo simulation with 10,000 replications drawing the values of u from a uniform distribution with the a and b values that you calculated above. [You can use the Stata function `runiform` in place of the `rnormal`, but be sure to look at the Stata help file to see how it is used.] As before, save the values of the coefficient estimates and their standard errors.

c. Repeat the analysis from the previous part (standard deviation of coefficient estimates, average standard error vs. standard deviation, histogram, and normality tests) and interpret the results.

d. What do you conclude about the validity of the asymptotic properties of the OLS estimator for uniform errors and n = 157?

3. **Sample size and asymptotic normality of OLS estimators.** You examined whether the asymptotic property of normality applied with n = 157. What happens if the sample is much smaller?
    a. Apply the analysis of question 2 (using the uniform distribution) to samples of sizes 15 and compare your results. To get a smaller sample of x, take the first 15 observations. You can easily create a new dataset with smaller n with the command `keep if _n<=15` to get a dataset with just the first 15 observations. (It's probably a good idea to "save as" a different file name so as not to save over your earlier dataset.) Note that you will need to recompute the sum of squared deviations of x for the smaller sample.
    b. What do you conclude from this experiment about the size of sample at which asymptotic properties are appropriate when the error is uniformly distributed?