Economics 312 — Spring 2013

Project #1 Assignment — Due: Midnight, Monday, February 11

Partner assignments

Jon Chai	Michael Zhao
Logan Donoughe	Andrew Watson
Orphelia Ellogne	Sri Shanmugam
Kevin Gallagher	Ted Sand
Allison Giffin	A.J. Roetker
Johannes Harkins	Torrey Payne
Tristan Hechtel	Maha Pasha
Weiqi Hu	Jacob Melnick
John Iselin	Cody Melcher
Martha Janicki	Andrea Lim
Maya Jarrad	Tess Lallemant
Jason Jin	Kathleen Kruzich
Marina Kaminsky	Tom Koskores
Bryan Kim	Shruti Korada

This project involves repeating the Monte Carlo simulation that was performed in class on February 4, then modifying it to allow for alternative data-generating processes.

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?
    b. Theory says that the variance of the OLS slope estimator is given by HGL's equation (2.15). In order to apply this formula to our sample, you must calculate $\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2$, the sum of squared x deviations. You can calculate this sum from the sample standard deviation of x, $s_x=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}$, which is reported by the `summarize` command. After computing the sum, compute the theoretical standard deviation of the OLS slope estimator using (2.15). Compare the standard deviation of your OLS slope estimates from the Monte Carlo simulation with the theoretical standard deviation from equation (2.15). Are they close?
    c. The OLS standard error attempts to measure this standard deviation. 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, 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. This is OK, but, as in this problem, you need to use the Stata documentation to determine how it works and to be able to justify its use.*]

2. **Regression with uniform errors.** We now re-run the simulation with an error term that follows a uniform distribution. According to theory, the distribution of the OLS estimator is asymptotically normal, so that 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. HGL discuss the uniform distribution in Section B.3.4 of Appendix B. The two parameters of the uniform distribution are a and b, the minimum and maximum values that the variable can have. We want to make our uniform distribution as much like the normal distribution of the previous part as possible, so we want a mean of zero and standard deviation of 0.2 (variance of 0.04). Use the formulas for the mean and variance of the uniform distribution on page 679 to calculate the values of a and b that make the mean zero and the variance 0.04. (Once you think you have the formula right, try creating a variable with the appropriate command and check to see if the mean and standard deviation are close to the values you expect.)

   b. Run a Monte Carlo simulation with 10,000 replications drawing the values of e 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. **Regression when the error is correlated with the regressor.** We shall see later in the course that the most critical assumption of OLS estimation is that the error term is independent of the regressor. To see this, we now consider a case in which the error term e depends on the regressor x according to $e_i = \varepsilon_i + 0.1x_i$, where ε is a normally distributed error term that is independent of x.
    a. We want the distribution of e to have a mean of zero and a standard deviation of 0.2, just like the ones in the previous problems. Given the mean and variance of x (calculated above from the results of the summarize command) and that x and ε have zero covariance, what must the mean and standard deviation of ε be if e is to have the desired moments?
    b. As a test of your procedure, generate a normal variable ε with the calculated mean and standard deviation, and add it to $0.1x$ to get a series for e. Use the summarize command to verify that e has mean near zero and standard deviation near 0.2.
    c. Once you are confident that your procedure works, incorporate it into a do-file and run a Monte Carlo simulation with 10,000 samples, each having an e series generated by your procedure.
    d. Examine the properties of your estimated coefficients. Do they seem unbiased?