Problem Set 1

(1) Read Section 11.1 of Johnston and DiNardo (1996), available at bspace. Following their lead write a program in Stata (or some other language) that draws 100 observations from the following iid data generating process:

$$Y_i = 1 + 0.5X_i + \varepsilon_i$$

where X_i is distributed uniform on [0,1] and ε_i is distributed standard normal and is independent of X_i. Estimate the slope and intercept and report them along with their standard errors in a table (those reporting t-ratios in parentheses rather than standard errors will be punished). I recommend the use of `outreg` and Excel, or `outreg` and LaTeX, but you are free to produce the table the way you see fit.

(2) Create 10,000 such data sets. For each data set, estimate the coefficients and standard errors for the slope and intercept. For each data set, test 3 null hypotheses: (1) that the slope is equal to 0.5, (2) that the intercept is equal to 1, and (3) that the slope is equal to 0.5 and jointly that the intercept is equal to 1. Report the fraction of the 10,000 data sets for which you reject these true null hypotheses.

(3) For the same 10,000 data sets, test 3 different null hypotheses: (1) that the slope is equal to 0.6, (2) that the intercept is equal to 1.1, and (3) that the slope is equal to 0.6 and jointly that the intercept is equal to 1.1. Report the fraction of the 10,000 data sets for which you reject these false null hypotheses. Discuss the difference between the results in questions (2) and (3).

(4) Create another data generating process corresponding to the measurement error model discussed in class. (A) Show that the classical linear regression model assumptions are not met. (B) Generate data from the measurement error model discussed in class and replicate the analysis from above. Discuss.