Economics 202   Professor Zabel
Econometrics   Spring 2016

Homework 5
Due Thursday March 17, 1:30 PM
Full Credit will not be Given Unless You Show Your Work

Part I - Monte Carlo Simulation

A   Create a "do" file (myreg.do) that contains the following code

```
program myreg, eclass
drop _all
set obs 10
generate x = rnormal()
gen u = chi2den(2,uniform())
generate y = 1+ 3*x + u
regress y x
end
```

then run this do file and on the Stata command line type

simulate _b _se, seed(xxxxx) reps(1000): myreg

where you replace xxxxx with an integer $< 2^{31}$-1 (or 2,147,483,647).

This will generate 1000 OLS estimates of β_1 (called _b_x) from the population
model

   y = 1 + 3*x + u

That is, β_0 = 1 and β_1 = 3. The number of observations is first set to 10, a
"small" sample.

Summarize these estimates using the "sum" command along with the option "detail". Also graph the estimates using the "hist" command and compare to the normal distribution using the option "normal."

One can test that the sampling distribution in normal using a number of different tests. First, one can test for skewness (0 under the null) and kurtosis (3 under the null) using the command "sktest: and for overall normality using the command "mvtest"

1. Write down the mean, standard deviation, skewness, and kurtosis for the sampling distribution for $\hat{\beta}_1$.

2. Include the histogram for each sampling distribution.

3. Report the results for the normality tests.

Run this simulation again with the number of observations set to 100 (keep the number of repetitions at 1000). Again, include the above information in 1-3 in your homework.

4. How do the results from the two simulations compare?

Part II - Computer Problem

Using the CPS dataset that you generated for Homework 3, run a regression of $\ln(wsal)$ on age, age^2, hs, col, white, metro, married, separated, divorced, female, female*age, nkids, nper, south, midwest, and west. (call this Model 1). Include the results in a Table (Table 1) using outreg.

1. What is the percent difference in salary for a woman who is 30 years old and a woman who is 40 years old? Is this difference statistically significant? Is this difference economically significant? Use the F-test to test for statistical significance. Include the restricted model in Table 1. Make sure to write out the computation for the F statistic.

2. Let Model 2 be the same as Model 1 but the specify the dependent variable as wsal (versus lnwsal). Compare Model 1 and Model 2 using R^2's (and pseudo-R^2's). Then compare then using hypothesis tests. Which Model do you prefer? Why?

Part III - Problems from Text: Chapter 15: 2, 8 (skip vi)

For problem 2, use the data file fertil2.txt (1,293 obs). The variables are in order:

children	number of living children
age	age in years
electric	=1 if has electricity, 0 otherwise
radio	=1 if has radio, 0 otherwise
tv	=1 if has tv, 0 otherwise
bicycle	=1 if has bicycle, 0 otherwise
educ	years of education, 0 otherwise
frsthalf	=1 if mnthborn <= 6, 0 otherwise

For problem 8, use the data file 401ksubs.txt (9,275 obs). The variables are:

e401k	=1 if eligble for 401(k)
inc	annual income, $1000s
age	age
nettfa	net total fin. assets, $1000
p401k	=1 if participate in 401(k)
pira	=1 if have IRA

Part IV – Read Sections 3 and 4, and 5 (pages 29 – top of page 37) of "The Causal Effect of Studying on Academic Performance," by Ralph and Todd R. Stinebrickner and answer the following questions.

In Section 3, the authors investigate the direct impact of RGAME on GPA by regressing GPA on RGAME and X.

1. What variables are included in X?

2. For the results in column (1) of Table 2, is the impact of RGAME on GPA economically significant?

3. What are the different avenues through which having a roommate with a video game can affect one's GPA? Which one(s) appear to be able to explain the difference in GPA between students whose roommates do and do not have video games? How do the authors show this to be the case?

4. Write down the main model used to estimate the impact of studying on student performance.

5. What is the OLS estimate of this impact? Is it economically significant?

6. Give three reasons why the OLS estimator is likely to be biased. Instrumental variables (IV) estimation is an estimation approach that does not suffer from the biases that affect the OLS estimator. For IV to be viable, a valid "instrument" is needed.

7. What are the necessary properties of a valid instrument for the model in this article? Explain why RGAME satisfies these properties.

8. A simple version of the IV estimator is called the "Wald Estimator". Write down the formula for the Wald estimator that uses RGAME as the valid instrument and show that it takes a value of 0.358 for males and 0.274 for females.