Department of Economics UN3412
Columbia University Fall 2021

Problem Set 5
Introduction to Econometrics
Seyhan Erden and Tamrat Gashaw
for all sections.

1. (33p) The data file ***rental.dta*** include rental prices and other variables for college towns in 1980 and in 1990. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is

$$\log(rent_{it}) = \beta_0 + \delta_0\, y90_t + \beta_0 \log(pop_{it}) + \beta_2 \log(avginc_{it}) + \beta_3\, pctstu_{it} + a_i + u_{it}$$

Variables needed are explained in below

Variables in RENTAL.dta

Variable	Definition
pop	City population
avginc	Average income
pctstu	Student population as a percentage of city population (during the school year)
y90	=1 for 1990, zero otherwise.

(a) (5p) Estimate the equation by pooled OLS and report the results in standard form. What do you make of the estimate on the 1990 dummy variable?

(b) (5p) Interpret the sample coefficient of pctstu

(c) (5p) Are the standard errors you report in part (a) valid? Explain.

(d) (6p) Now, difference the equation and estimate by OLS. Compare your estimate of β_3 with that of part (a). Does the relative size of the student population appear to affect rental prices?

(e) (6p) Obtain the heteroskedasticity-robust standard errors for the first-differenced equation in part(d)

(f) (6p) Estimate the model by fixed effects to verify that you get identical estimates and standard errors to those in part (d) (use “areg” and “xtreg” commands and report both results)

2. (34p) Use the state-level data on murder rates and executions in ***murder.dta*** for the following exercise.

(a) (4p) Consider the unobserved effects model

$$mrdrte_{it} = \eta_t + \beta_1 exec_{it} + \beta_2 unem_{it} + a_i + u_{it}$$

where η_t simply denotes different year intercepts and a_i is the unobserved state effect. If past executions of convicted murderers have a deterrent effect, what should be the sign of β_1? What sign do you think β_2 should have? Explain.

(b) (4p) Using just the years 1990 and 1993, estimate the equation from part (i) by pooled OLS. Ignore the serial correlation problem in the composite errors. Do you find any evidence for a deterrent effect?

(c) (4p) Now, using 1990 and 1993, estimate the equation by fixed effects. You may use first differencing since you are only using two years of data. Is there evidence of a deterrent effect? How strong?

(d) (5p) Compute the heteroskedasticity-robust standard error for the estimation in part (ii).

(e) (5p) Find the state that has the largest number for the execution variable in 1993. (The variable *exec* is total executions in 1991, 1992, and 1993.) How much bigger is this value than the next highest value?

(f) (6p) Estimate the equation using first differencing, dropping Texas from the analysis. Compute the usual and heteroskedasticity-robust standard errors. Now, what do you find? What is going on?

(g) (6p) Use all three years of data and estimate the model by fixed effects. Include Texas in the analysis. Discuss the size and statistical significance of the deterrent effect compared with only using 1990 and 1993.

3. (33p) The file ***pension.dta*** contains information on participant-directed pension plans for U.S. workers. Some of the observations are for couples within the same family, so this data set constitutes a small cluster sample (with cluster sizes of two).

(a) (5p) Ignoring the clustering by family, use OLS to estimate the model

$$\begin{aligned} pctstck = \beta_0 &+ \beta_1 choice + \beta_2 prftshr + \beta_3 female + \beta_4 age + \beta_5 educ \\ &+ \beta_6 finc25 + \beta_7 finc35 + \beta_8 finc50 + \beta_9 finc75 + \beta_{10} finc100 \\ &+ \beta_{11} finc101 + \beta_{12} wealth89 + \beta_{13} stckin89 + \beta_{14} irain89 + u \end{aligned}$$

where the variables are defined in the data set. The variable of most interest is *choice*, which is a dummy variable equal to one if the worker has a choice in how to allocate pension funds among different investments. What is the estimated effect of *choice*? Is it statistically significant?

(b) (5p) Are the income, wealth, stock holding, and IRA holding control variables important? Explain.

(c) (5p) Determine how many different families there are in the data set.
(d) (6p) Now, obtain the standard errors for OLS that are robust to cluster correlation within a family. Do they differ much from the usual OLS standard errors? Are you surprised?
(e) (6p) Estimate the equation by differencing across only the spouses within a family. Why do the explanatory variables asked about in part (ii) drop out in the first-differenced estimation?
(f) (6p) Are any of the remaining explanatory variables in part (v) significant? Are you surprised?

Following questions will not be graded, they are for you to practice and will be discussed at the recitation:

Question 1:

U.S. airlines were deregulated in 1975, allowing them to charge whatever prices they wished and to choose routes for their flights more freely than previously. One anticipated gain from deregulations was cost reduction, to be derived in part by allowing airlines to reduce excess capacity. Baltagi, Griffin and Vadali estimate that airlines did, indeed, reduce excess capacity following deregulations[1]. Their analysis combined data on variable costs and factor shares to efficiently estimate excess capacity for 23 airlines in the years 1971-1986. Data file **deregulate.dta** contain the following variables:

Variable	Description
airline	A number indicating the airline in the observation.
pf	The price of fuel
pl	The price of labor
pm	The price of materials
reg	=1 if the observation is from the regulated period =0 otherwise
stage	Average length of the airline's flights that year
vc	Variable cost (fuel+labor+materials)
y	An index of annual passenger miles flown by the airline
year	The year of the observation

(a) Regress the log of costs on the regulation dummy, year and the natural logs of three price variables and of *stage* (i) using OLS (ii) using firm-specific fixed effects without cluster (iii) with cluster
(b) What is the interpretation of regulation dummy's coefficient?
(c) What is the interpretation of year's coefficient?

[1] Badi H. Baltagi, James M. Griffin, and Sharada R. Vadali, "Excess Capacity: A Permanent Characteristic of U.S. Airlines," *Journal of Applied Econometrics* 13, no.5 (1998): 645-657

(d) Briefly explain why we can conclude that the estimated standard errors reported for OLS are probably incorrect as well as the ones in fixed effects regression without cluster errors?
(e) What does the fixed effects regression imply about the effect of deregulation on airlines' variable cost?
(f) How do you counter the objection that technical change would have reduced airline costs even without the deregulation?
(g) Add the squares of the logged regressors to the fixed effects regression in (a). What does this regression suggest about the conclusions in (e)?
(h) Are the added terms in regression (g), taken together, jointly statistically significant? Show the needed test results.
(i) Some have argued that deregulation enables airlines to better plan their flight. This could mean that more efficient flight lengths were chosen after deregulation. How does this affect the interpretations in (e) and (g), and how would you take this consideration into account?

Question 2: SW Exercise 10.1

Question 3: SW Exercise 10.5

Question 4: SW Empirical Exercise 10.2