Economics 202 Professor Zabel
Econometrics Spring 2016

Homework 6
Due Thursday March 31 @ 1:30 PM
Full Credit will not be Given Unless You Show Your Work
Use Outreg to Display Regression Results

Part I – This problem uses data on 4th and 8th grade math test scores for 269 school districts in Massachusetts for 1999-2006. Note that because some school districts only include elementary schools, there are only 220 school districts with an 8th grade. Starting in 1999, Massachusetts implemented the Massachusetts Comprehensive Assessment System (MCAS) at state-wide standardized test. Because the tests are not necessarily comparable from year to year, the test scores are standardized on an annual basis. Variable definitions are given below. The Stata dataset mass_testscore_data is on trunk.

Variable	**Definition**
std_gr4m	Standardized 4th grade math test score
std_gr8m	Standardized 8th grade math test score
lnppcurexp	Natural log of current per pupil spending
pctlowinc	Percent of students from low income households
pctlep	Percent of students with limited English proficiency
pctsped	Percent of Special Education students
pct_black	Percent of African-American students
pct_hispanic	Percent of Hispanic students
totstu	Total district student enrollment
district_number	School district code

1. Include a table of summary statistics for the above variables. Call this Table 1.

Note that *pctlep* is missing for 107 observations. In order not to lose these observations when including this variable in the regressions, we can make the following fix.

- Generate a dummy variable that is 1 for the missing observations and 0 otherwise. Call this variable *pctlep_flag*.
- Replace the missing values for *pctlep* with "0"s.
- Include both *pctlep* and *pctlep_flag* in the regression.

The coefficient on *pctlep_flag* is the mean difference in the dependent variable between the group that is missing values for *pctlep* and the group that is not (conditional on other explanatory variables). If this coefficient is not significantly different from zero, we say that the observations for *pctlep* are "missing at random."

2. Run OLS regressions of 4^{th} and 8^{th} grade test scores on the other variables in the above table of definitions (other than district_number) and include year dummies in the models. Include the results in Table 2 (but exclude the year dummies). Interpret the coefficient estimates for *lnppcurexp, pctlowinc*, and *totstu* for the regression with 4^{th} grade test scores as the dependent variable.
3. Test that the year dummies are jointly zero for each model. What is the result? Is this surprising? Why or why not?
4. Any there any differences between 4^{th} and 8^{th} grade regressions you ran in part 2? In particular, which regression parameters (other than those for the year dummies) in the 4^{th} and 8^{th} grade regressions are significantly different and which ones are not (at the 1% significance level)? To determine this you need to run these two regressions as a system so that you can use the test command to compare coefficients across the two equations. The Stata command to do this is "sureg"
5. Estimate the same two models using (district level) fixed effects. Interpret the coefficient estimates for *lnppcurexp*, *pctlowinc*, and *totstu* for the regression with 4^{th} grade test scores as the dependent variable. Explain why the results have changed.

Part II - Problems from the Text Book: Chapter 8: C6, Chapter 14: C7

For problem C6 from Chapter 8, download the data set crime1.txt from Trunk. There are 2,725 individual observations. The variables are (in order):

narr86	number of times arrested, 1986
pcnv	proportion of prior convictions
avgsen	average sentence length, mos.
tottime	time in prison since 18 (mos.)
ptime86	months in prison during 1986
qemp86	number of quarters employed, 1986

For problem C7 from Chapter 14, download the data set murder.txt from Trunk. There are 153 state-level observations (includes DC). The variables are (in order):

id	State identifier
state	State abbreviation (two letters)
year	87, 90, or 93
mrdrte	murders per 100,000 population
exec	total executions, past 3 years
unem	annual unemployment rate

Note that because state is a character variable, it is necessary to precede “state” with the string “str2” in the infile statement.

Part III - Read "Are Restaurants Really Supersizing America", by Michael L. Anderson and David A. Matsa, Sections III - VI. Answer the following questions

1. The authors use the Two Sample Two-Stage Least Squares estimator (TS2SLS). Explain how this estimator works - what are the first and second-stages of this estimator? (See page 4 of the class notes on Instrumental Variables).

2. What are the IV results - does distance to a restaurant have a statistically significant impact on obesity? Does it have an economically significant impact on obesity? Explain.

3. What possible alternative explanations for the lack of a relationship between the availability of restaurants and obesity do the authors raise? What evidence do they provide that contradicts these explanations?

4. What reasons do the authors provide for why restaurants don't affect obesity?

Part IV – Empirical Project: Hand in updated project information **separately** with

1. Title and (your) name
2. Section 1: Introduction
3. Section 2: Literature Review of 1 paper
4. Section 3: Include information on data set(s)