Econ 140 - Spring 2016
Section 12

GSI: Fenella Carpena

April 28, 2016

1 Experiments and Quasi-Experiments

Exercise 1.0. Consider the STAR Experiment discussed in lecture where students were randomly assigned to one of three groups: small class size, regular class size, or regular class size with a teacher's aide. Using data collected from the experiment, we estimate the regression:

$$Y_i = \beta_0 + \beta_1 SmallClass_i + \beta_2 RegAide_i + u_i$$

where Y_i is test score for student i, $SmallClass_i = 1$ if student i is in a small class and 0 otherwise, and $RegAide_i = 1$ if student i is in a regular class with an aide and 0 otherwise. How would you expect including $TeacherExp$ (teacher's years of experience) as an additional variable would alter the coefficient on $\overline{SmallClass}$: would it increase, decrease, or stay the same? Explain.

Exercise 1.1. (Stock & Watson, Review the Concepts 13.1) A researcher studying the effects of a new fertilizer on crop yields plans to carry out an experiment in which different amounts of fertilizer are applied to 100 different 1-acre parcels of land. There will be four treatment levels. Treatment level 1 has no fertilizer, treatment level 2 is 50% of the manufacturer's recommended amount of fertilizer, treatment level 3 is 100%, and treatment level 4 is 150%. The researcher plans to apply treatment 1 to the first 25 parcels of land, treatment level 2 to the second 25 parcels, and so forth. Can you suggest a better way to assign treatment levels? Why is your proposal better than the researcher's method?

Exercise 1.2. (Stock & Watson, Review the Concepts 13.2) A clinical trial is carried out for new cholesterol-lowering drug. The drug is given to 500 patients, and a placebo is given to another 500 patients, using random assignment of the patients.

(a) How would you estimate the treatment effect of the drug?

(b) Suppose that you had data on the weight, age and gender of each patient. Could you use these data to improve your estimate? Explain.

Exercise 1.3. (Stock & Watson, Review the Concepts 13.3) Researchers studying the STAR data report anecdotal evidence that school principals were pressured by some parents to place their children in the small classes.

(a) Suppose that some principals succumbed to their children in the small classes. How would such transfers compromise the internal validity of the study?

(b) Suppose that you had data on the original random assignment of each student before the principal's intervention. How could you use this information to restore the internal validity of the study?

Exercise 1.4. (Adapted from Stock & Watson, Exercise 13.4) Going back to the Card and Krueger (1994) example, consider the difference-in-difference regression:

$$emp_{it} = \beta_0 + \beta_1 NJ_i + \beta_2 POST_t + \beta_3 NJ_i * POST_t + u_{it}$$

(a) In terms of coefficients β_0, β_1, β_2, β_3, what is the expected number of employees in:

(i) A New Jersey restaurant in 1991?

(ii) A New Jersey restaurant in 1993?

(iii) A Pennsylvania restaurant in 1991?

(iv) A Pennsylvania restaurant in 1993?

(b) In terms of the coefficients β_0, β_1, β_2, β_3, what is the average causal effect of the minimum wage on employment?

(c) Explain why Card and Kruger used the difference-in-difference estimator of the causal effect instead of the "New Jersey after − New Jersey before" or the "1993 New Jersey − 1993 Pennsylvania" differences estimator.

Exercise 1.5. (Stock and Watson, Exercise 13.3) Suppose that, in a randomized controlled experiment of the effect of an SAT preparatory course on SAT scores, the following results are reported:

	Treatment Group	Control Group
Average SAT Score ($\overline{X}$)	1241	1201
Standard deviation of SAT score (S_X)	93.2	97.1
Number of women	55	45
Number of men	45	55

(a) Estimate the average treatment effect on test scores.

(b) Is there evidence of non-random assignment? Explain.

2 Final Exam Review

Final Exam Spring 2014, Question 3. You are hired by the Government of Ghana to study the impact of income on the level of education. Using data on rural villages, you estimate the following population regression using OLS:

$$Educ_i = \beta_0 + \beta_1 Income_i + \beta_2 Pop_i + \beta_3 School_i + \beta_4 Age_i + u_i$$

where $Educ_i$ is average years of formal education in the village, $Income_i$ is average annual income per capita in the village, Pop_i is the number of village residents, $School_i$ is the number of schools in the village, and Age_i is the average age of the village population.

(a) (5 points) Explain what econometric problem is likely to arise that leads to biased and inconsistent estimates as a result of including $Income$ as a regressor in the education regression as is done above.

You learn from Ghana's Minister of Agriculture that the country's citizens derive the bulk of their income from agriculture. As a result, you cleverly infer that average annual rainfall ($Rainfall$) may be a good instrument for income.

(b) (5 points) You recall from your econometrics course that an instrument can be used in a procedure called Two Stage Least Squares that is designed to solve this econometric problem. Describe carefully the first of the two stages and why TSLS will generate a consistent estimate of β_1.

You want to check the Minister's suggestion that rainfall has an impact on incomes in Ghana. You have information on average annual incomes in 1996 and 1997 for two regions: the "coastal region," which had the same precipitation level in both years, and the "hill region," which experienced a 30% increase in rainfall. Comparing 1996 and 1997, income in the coastal region fell from 124 to 104, while income in the hill region fell from 98 to 96. You also recall from your econometrics course that this situation might represent a "natural" or "quasi" experiment, allowing you to estimate the "treatment effect" of rainfall.

(c) (8 points) Perform a difference in differences analysis of the effect of rainfall on average income. Summarize the analysis in a table.

(d) (6 points) Describe a multivariate regression that when estimated using OLS will generate exactly the same estimate of the effect of rainfall on income as was generated by the analysis in part (c).

(e) (6 points) Describe in detail one threat to the internal validity of the OLS estimates when treating these data as a quasi-experiment, and how it would bias the coefficient estimate.

Final Exam Spring 2011, Question 5. In 1980, due to a temporary easing of Cuban emigration rules, there was a huge influx of Cuban immigrants into the state of Florida. As a result of this so-called "Mariel boatlift," the low-skilled labor force of Miami increased by 7%. David Card compared the average hourly wages in Miami and comparison cities (Atlanta, Houston, Los Angeles, and Tampa-St. Petersburg). The average hourly wages expressed in logarithms are given in the below table:

		Cities	
		Miami	Comparison
Year	1979	1.85	1.93
	1981	1.85	1.91

(a) (10 points) Calculate the percentage change in average hourly wages in the treatment group and in the control group, and uses those changes to the differences-in-differences ("DiD") estimate. Is the sign of the DID estimate what would be predicted by economic theory? Explain.

(b) (10 points) Give an example of a relevant variable that is omitted from the DiD estimation, and predict the likely bias it would cause.

(c) (12 points) To accommodate other determinants of metropolitan wage rates, you suggest including a measure of the size of the metropolitan manufacturing sector M_i since it might reflect ability to absorb low-skilled workers. Write down a linear regression that generates a DiD estimate while incorporating this control variable. Why would you believe that this regression approach would change your estimate of the effect of the Mariel boatlift from (a)?