Problem Set #4

Prof. Matt Reimer*

Econ A390: Methods for Public Policy Evaluation

Please answer all the questions below within this Microsoft Word document. For all parts, your answer should consist of a concise explanation, along with any STATA commands and results. See these instructions for importing STATA output and graphs into Microsoft Word documents. You are encouraged to work as a group; however, there may be no more than three students per group. Everyone is responsible for submitting one typed file for grading on Blackboard under Assignments. Please indicate the members of your group at the top of the file. You might find it helpful to read through the assignment, then read your class notes and textbook, and only then try and do the questions. *Attempting this assignment without consulting your notes and the text may be hazardous to your health and your grade!!*

Motivation: Causal effect of class size on educational attainment

Class size is a variable at the heart of policy debates on school quality and the allocation of school resources in many countries; however, causal effects of class size on pupil achievement have proved very difficult to measure. This is because there is a selection process associated with the students (and parents) that attend smaller or larger classes. For instance, if smaller class sizes are viewed as beneficial to educational outcomes, then more educated or wealthier families may take extra steps to ensure that their children are in smaller class sizes. In this sense, higher educational achievement due to smaller class sizes may be conflated with socioeconomic status. On the other hand, smaller classes may be designed to help students who need extra one-on-one attention who are at risk of lower educational achievement. In this sense, lower educational achievement from smaller class sizes may be conflated with remedial training.

In this problem set, we will attempt to estimate the causal effect of class size on test scores using regression discontinuity and instrumental variables techniques. The data we will use (described below) comes from a national testing program in Israeli elementary schools, all of which have a unique way of determining classroom size. Since 1969, the maximum number of students in an Israeli classroom is 40, a rule referred to as "Maimonides' rule", named after the great twelfth century Rabbinic scholar, Maimonides, who used the Babylonian Talmud to derive a rule suggesting there should be no more than 40 students in a classroom. Specifically, according to Maimonides' rule, class size increases one-for-one with enrollment until 40 pupils are enrolled, but when 41 students are enrolled, students must be divided between two classes. Thus, there will be a sharp drop in class size, to an average of 20.5 pupils. Similarly, when 80 pupils are enrolled, the average class size will again be 40, but when 81 pupils are enrolled, students are divided be-

*Telephone: 1-907-786-5430, e-mail: *mreimer2@alaska.edu*.

tween three classes and the average class size drops to 27. All schools in Israel must abide by Maimonides' rule; however, class sizes vary for other reasons as well. For example, some schools might cap classrooms at a number lower than 40 or may not split class size evenly across two classrooms once total enrollment is above 40.

You will make use of enrollment, class size, and test score data for Israeli elementary schools that have less then 80 students enrolled in grades four or five. In June of 1991, all Israeli fourth and fifth graders were given achievement tests designed to measure mathematics and reading skills. Composite scores were derived from the test results so that all scores lie between 1 and 100. The unit of observation in our data is a class. For each class, our dataset consists of average reading (`avgread`) and math (`avgmath`) scores from the exams, as well as class size (`classize`), and additional school-level characteristics, such as total enrollment (`enrollment`), and the percentage of students that are from "disadvantaged backgrounds" (`pct_dis`). The variable `m_rule` is the class size that would prevail in a classroom if schools used Maimonides' rule perfectly. For this problem set, you will need to download and install the user-written STATA command `binscatter`. In the STATA prompt, type `ssc install binscatter`. You should also make use of the commands in the example dofiles on Blackboard.

```
. describe

Contains data from econ390/Spring2015/Stata/pset4.dta
  obs:         2,290
 vars:            11                          14 Apr 2015 15:09
 size:        66,410
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
schlcode        long   %12.0g                 school code
enrollment      int    %8.0g                  Number of students enrolled
c_boys          int    %8.0g                  Number of boys enrolled
c_girls         int    %8.0g                  Number of girls enrolled
grade           byte   %8.0g                  class grade
classize        byte   %8.0g                  Number of students in class
avgmath         float  %9.0g                  Average composite math score in class
avgread         float  %9.0g                  Average composite reading score in class
pct_dis         byte   %8.0g                  Percent of students from
                                                ´disadvantaged backgrounds´
m_rule          float  %9.0g                  Maimonides´ rule for class size
D               float  %9.0g                  Treatment: =1 if enrollment > 40;
                                                0 otherwise
-------------------------------------------------------------------------------
Sorted by:  grade
```

Problem 1:

In this problem, we will exploit the nature of Maimonides' rule to estimate the causal effect of class size on test scores using an instrumental variables approach.

A. Suppose that your friend has not taken Econ 390 Methods for Public Policy Evaluation. In ignorance, your friend suggests that you regress test scores on class size to infer what the causal effect of class size on educational attainment is.

i) Estimate separate regressions for grades four and five, with average reading score (`avgread`) as your outcome variable and class size (`classize`) as your treatment variable. What does your estimated model suggest about the relationship between class size and test scores? How does this change when you include the control variables `pct_dis` and `enrollment`? Explain.

ii) Explain why your friend's suggestion in part (i) may not be a very good idea for inferring the causal effect of class size on test scores.

B. One possible way to address your concern about your friend's suggestion is to use an instrumental variable (IV) for class size.

i) What is an IV? How does an IV approach differ from the approach suggested by your friend?

ii) What assumptions must be met by an IV? Discuss whether or not Maimonides' rule (`m_rule`) is likely to meet these assumptions as an IV for class size.

iii) Empirically investigate whether `m_rule` satisfies the assumptions required for an IV. Things to investigate include the first-stage regression, the reduced-form regression, and a scatter plot of `classize` and `m_rule`. Is it possible to be certain that `m_rule` meets the requirements of an IV? Why or why not?

iv) Estimate the effect of class size (`classize`) on average reading test scores (`avgread`) by IV using the `ivreg` command. Do this separately for grades four and five, with and without the control variables `pct_dis` and `enrollment`. How does your answer differ from Part A? Explain.

Problem 2:

In this problem, we will exploit the nature of Maimonides' rule to conduct a regression discontinuity design (RDD) to estimate the causal effect of class size on test scores.

A. Suppose you have another friend who decided not to take Econ 390 Methods for Public Policy Evaluation (what a shame....).

i) Explain to your friend the main idea of a RDD. In particular, how does Maimonides' rule fit into a RDD for estimating the causal effect of class size on test scores?

ii) What are the three main assumptions underlying a RDD that we talked about in class? Describe how these assumptions relate to our specific application in this problem set.

B. Let's examine the data to see if our problem fits into a RDD.

i) Using the `binscatter` command, plot `classize` against `enrollment` separately for grades four and five. Is there evidence of Maimonides' rule? Explain.

ii) Using the `binscatter` command, plot `avgread` against `enrollment` separately for grades four and five. Is there any evidence that class size has an effect on test scores? Explain.

C. Estimate the causal effect of class size on test scores using a RDD.

i) Run the appropriate regression to obtain an RDD estimate of the causal effect of class size on average reading test score. Do this separately for grades four and five. *Hint: don't forget to center your data!*

ii) Discuss some possible robustness checks that explore whether the assumptions required for an RDD hold. Apply these robustness checks to your RDD model above. Do you find any evidence that make your results in part (i) questionable? Explain.

D. Compare and contrast your IV and RDD estimates of the effect of class size on test scores.

i) Why are your RDD estimates of class size effects on test scores so different from your IV estimates in Problem 1?

ii) Is there any reason you prefer the IV results to the RDD (or vice versa)?

iii) How do you feel about the internal and external validity of both sets of estimates? Explain.