

ECON 244

APPLIED EXERCISE #3

Due 4 pm on November 21, In Ken Chay's mailbox

This exercise examines the following research question: What is the impact of Chile's 900 Schools Program (a.k.a., P900) on student achievement? For background on the program and the data source refer to the paper, "The Central Role of Noise in the Evaluation of Interventions that Use Test Scores to Rank Schools," by Kenneth Chay, Patrick McEwan, and Miguel Urquiola, *American Economic Review*, September 2005, 1237-1258. The paper can be downloaded at the following site:

<http://www.ingentaconnect.com/content/aea/aer/2005/00000095/00000004/art00015>.

Please include a concise summary of your empirical results when appropriate. We will analyze the following data set:

Data Source: p900_area2.dta

This STATA data extract is the same school-level data used in the referenced paper, but restricted to schools in regions 6, 7, and 8 in Chile.

Data Notes:

1. Link to data via command: `ln -s /class/e244/e244/data/p900_area2.dta .`

2. There are 658 school level observations.

3. The key variables are:

dx90, dx92 = 1988-1990 (1990 minus 1988) and 1988-1992 (1992 minus 1988) school-level change in average test scores (average of math and reading scores).

p90 = indicator equal to one if the school received the P900 program resources and equal to zero, otherwise.

x88 = average test score in the school in 1988.

ivrul2_area2 = indicator equal to one if school had an average 1988 score (x88) that made it eligible to receive the P900 program and equal to zero, otherwise.

rule2_area2 = 1988 average score minus the P900 eligibility cutoff in the region that the school is in.

4. The other relevant variables are:

ncas88, ncas90, ncas92 = the number of test takers in the school in 1988, 1990, and 1992.

ses90, ses92 = average "socioeconomic status" of students in the school in 1990 and 1992 – higher number implies higher SES (unavailable for 1988).

dses = 1990-1992 change in average SES.

rgn88 = region of Chile that school is in (6, 7, 8).

Research Question: Did the P900 program raise school achievement in 1990 and 1992?

The outcomes of interest are the 1988-1990 and 1988-1992 test score gains at the school-level. We want to estimate the "causal" effect of P900 on test score changes. Read the referenced paper for background and details on the P900 program, and sources of bias in conventional "difference-in-differences" evaluations of the program. In this exercise, we will use the regression discontinuity design implicit in the administration of the program to address potential issues of bias.

- a. Use a difference-in-differences regression model to estimate the impact of P900 on 1988-1990 and 1988-1992 test score gains – i.e., regress dx_{90} and dx_{92} on p_{90} . What do the estimates imply about the size of the effect of the P900 program?
- b. Describe how the estimates in (a) may be biased by noise and mean reversion. In particular, write down a simple error-components model that allows for a permanent school component, a transitory school component, and an individual component that are independent of each other. Use this model to derive the mean reversion bias function for 1988-1990 average test score gains – i.e., the correlation between x_{88} and dx_{90} that would prevail even in the absence of the P900 program. How is the bias dependent on the relative importance of the various error-component variances and the number of test takers in the school in 1988?
- c. In principle, only schools who scored below a region-specific threshold in 1988 were eligible to receive the P900 program in 1990 through 1992 (i.e., $ivrule2_area2 = 1$). Thus, the indicator for receiving P900 should be a discrete function of a school's average score in 1988. Describe how one could use this discontinuity in treatment assignment to derive an estimate of the program effects while controlling for the mean reversion bias. Under what conditions will these estimates be valid?
- d. In practice, the P900 program was assigned according to this rule imperfectly. However, over 97 percent of schools in regions 6-8 were “correctly” assigned implying a slippage of less than 3 percent. Using a nonparametric smoother (“ksm” command if using STATA 7; “lowess” if STATA 8) plot the “smoothed” relationship between the probability of receiving P900 (p_{90}) and the school's score in 1988 relative to the regional cutoff ($rule2_area2$), along with the actual data points. For this exercise, use the “running-mean” smoother and a bandwidth of 0.05. Describe your findings.
- e. Suppose that the P900 eligibility indicator ($ivrule2_area2$) is a potential instrumental variable for P900 program status. Under what assumptions will it be a valid instrument? Write down the first-stage and reduced-form equations for P900 status and 1988-1990 and 1988-1992 test score gains as a function of the eligibility instrument. How does two-stage least squares use these equations? Do you still need to control for potential mean reversion bias in the instrumental variables estimates?
- f. Using a nonparametric smoother, plot the smoothed relationship between dx_{90} and $rule2_area2$, separately for eligible and ineligible schools. Use a bandwidth of 0.3 for the eligible schools and 0.1 for the ineligible schools (and the running mean smoother). Describe what you find and what it implies about the impact of the P900 program on 1988-1990 test score gains. Now do the same analysis for dx_{92} and interpret your findings.
- g. Estimate the “first-stage” relationship between $ivrule2_area2$ and p_{90} and the “reduced-form” relationship between $ivrule2_area2$ and dx_{90} and dx_{92} , using the following three specifications: 1) no control variables; 2) controlling for a cubic in the 1988 score; and 3) controlling for a cubic in the 1988 score, a quadratic in the number of test takers in 1988, and the interaction of the 1988 score with 1988 school size. Interpret your findings. Now estimate the effects of P900 on 1988-1990 and 1988-1992 test score gains using two-stage least squares and the eligibility indicator as an instrument for the three specifications. Interpret the results.
- h. Now provide a concise synthesis/summary of your results similar in spirit to the one you wrote for Applied Exercise #2. Discuss the evidence you found on the importance of mean reversion bias in conventional difference-in-differences estimates of the program effects, and the “credibility” of the regression discontinuity design in removing this potential bias.