Midterm Exam for Causal Inference in the Social Sciences

Fall 2007

11/14/2007
This exam will analyze the following data set: [http://are.berkeley.edu/~rocio/f2007_PS236/cross_section_wfl.csv](http://are.berkeley.edu/~rocio/f2007_PS236/cross_section_wfl.csv) The variables are described in the following file: [http://are.berkeley.edu/~rocio/f2007_PS236/codebook.water.txt](http://are.berkeley.edu/~rocio/f2007_PS236/codebook.water.txt).

This dataset has 435 observations, and was used in the article “Water for Life: The Impact of the Privatization of Water Services on Child Mortality”, by S. Galiani, P. Gertler, and E. Schargrodsky (2005, Journal of Political Economy, volume 113). This article is in the class syllabus.

The units of observations are municipalities in Argentina, and the treatment under study is the privatization of municipal water services. All 435 municipalities in this sample had public water services in the year 1990, but by the year 1999, 123 municipalities had privatized their water services. Of the 123 municipalities which privatized between 1990 and 1999, 83 municipalities did it between 1998 and 1999. The original panel data structure of the dataset has been simplified to a cross section: for each municipality, the dataset you will be working with has the covariates for each year between 1990 and 1999.

The treatment indicator has been defined as equal to one if the municipality privatized its water services sometime between 1991 and 1999, and equal to zero if a municipality whose water services were public in 1990 never privatized between 1991 and 1999.

The outcomes of interest are total child mortality and child mortality from infectious and parasitic diseases, i.e., waterborne diseases. Perinatal mortality is also of interest for the reasons discussed in Galiani, Gertler, and Schargrodsky (2005).

For the purpose of all questions on this exam except for the bonus question, when trying to find optimal balance, you can safely restrict yourself to the following baseline covariates and the nonlinear functions listed here: \( y_{1990}.tasatot; y_{1990}.tmaccid; y_{1990}.tmperin; y_{1990}.tmconge; y_{1990}.tmsangr; y_{1990}.tmcircu; y_{1990}.tmdiges; y_{1990}.tmgenit; y_{1990}.tmosteo; y_{1990}.tmrespi; y_{1990}.tmendoc; y_{1990}.tminfec; y_{1990}.unemp; y_{1990}.pbirpc; y_{1990}.pubspendpc; y_{1990}.ineq; y_{1990}.peron; y_{1990}.hognbi1; y_{1990}.hognbi2; y_{1990}.tasatot*y_{1990}.peron; y_{1990}.tminfec*y_{1990}.peron; and y_{1990}.unemp * y_{1990}.peron.
Use a matching method to find excellent balance for this dataset, and answer the following questions.

1. Estimate the causal effect for the treated of privatizing water services between 1991 and 1999 on child mortality in 1999, without dropping any observations. Are these effects significant? What is the most interesting summary statistic when comparing child mortality across the two groups? How informative are mean differences? What are the mean differences?

2. Estimate the same causal effect as in Question 1 this time dropping at most 5 observations. Does your balance improve with respect to the balance you found in Question 1? Are these effects significant? What are the mean differences?

3. Estimate the causal effect for the treated of privatizing water services between 1991 and 1995 on child mortality in 1996, using the matched dataset of Question 2. Note that since the baseline year has not changed, you need not create a different matched dataset. But you do need to exclude municipalities whose water services were privatized after 1996 from your matched dataset (and you have to make sure to exclude the corresponding matched controls also). Are the effects significant? What are the mean differences?

4. Estimate a post-matching parametric bias adjustment method on the matched dataset you obtained in Question 2 such as a regression model. Does this make a difference for the various outcomes of interest?

5. How do your results differ from those in Galiani, Gertler and Schargrodsky (2005)? Are your results and their results comparable?

6. In this study, does SUTVA limit what we can say? If so, how? If not, why not?

7. (BONUS QUESTION). Freed of the constraints in the previous questions, find the best matching method (and possibly post-matching adjustment model) to answer the
substantive question at hand. Note that in the previous questions we ignored the fact that some municipalities privatized water earlier than others. Does privatizing water services reduce child mortality? What should the policy maker learn?