PhD Programme: Development Economics (micro)
University of Gothenburg
November 2009
Måns Söderbom
mans.soderbom@economics.gu.se

Computer Exercise I

Program Evaluation: Progresa[1]

In this exercise we will be using two rounds of survey data in order to see how the Progresa program can be evaluated. The first one is from October 1997 and the second from November 1999. Key variables have been combined in the Stata file **progresa.dta**. The data provided include children between 8--16 years of age in October 1997.

We will focus on the direct effects of conditional cash transfers for education: poor, rural households with children enrolled in grades 3--9 were eligible to receive educational grants. Payment is conditional on enrollment with unexcused absences on fewer than 15% of school days. Incentives are stronger for higher years of schooling, and at secondary level higher for girls than for boys.

Implementation:
Progresa was implemented in phases: First phase in August 1997. 140,544 households in 3,369 localities. Subsequently expanded through year 2000, with 2.6m families in 72,345 localities, all states. Our data include 24,000 households in 506 localities, spanning 7 states. Of these, 320 were randomly assigned to treatment and 186 to control (these were included in Phase II, and began receiving benefits in December 2000). Eligibility is therefore determined in two steps. First, one must be in a locality selected for treatment. Second, one must qualify as sufficiently poor to receive the subsidy. The calculation of poverty status was based on the baseline data (October 1997). On this basis, a cutoff rule for eligibility was defined for each region (regions are given by the variable *entidad*).

Questions:

a) Use a crosstab with Stata's -tab- command to document the share of households meeting the poverty criterion (pobre=1) in control and treatment (tcomm=1) communities. Why are some apparently non-poor households treated?

b) Investigate whether treatment status can be predicted by variables measured in the baseline year (1997). What are you learning from this?

c) Investigate whether there is any significant difference in the enrolment rates across treated and non-treated communities i) in 1997; ii) in 1999. Think about how you should compute the standard errors. Discuss the findings.

d) Produce a graph of enrolment rates in 1999 for each age group, and for boys and girls separately, distinguishing between the treatment group and the control group. What can you say about the impact of the program based on the results?

[1]This exercise draws on work by Andrew Zeitlin, University of Oxford.

e) Compute difference estimates, and difference-in-differences estimates, of the impact of the program, using a regression model that allows for different impacts for boys and girls. Test the hypothesis that the impact does not differ depending on gender. Discuss the results.

f) Compute difference estimates, and difference-in-differences estimates, of the impact of the program, using a regression model that allows for different impacts for boys and girls *and* across different age groups. Summarize your findings.

g) We've been using a linear probability modelling framework so far. Is this problematic?

h) Investigate whether there is any evidence that the program impact differs depending on initial per capita income.

i) Are there any other control variables in the dataset that you may want to use in your empirical analysis? Experiment (!) a little.

j) Why might the program impact on the enrolment rates of *non-poor* households? Investigate.

k) Regression discontinuity design. Progresa implemented a region-specific cut-off for program eligibility. This is given by variable **rcut**. Create a variable that takes a value of one if the observation has a score within 50 points of the cutoff. Then estimate the difference in differences on this subsample. How do the results compare with those obtained above? What's the logic behind adopting this particular identification strategy? How do the results change if you vary the size of the "neighborhood" in which you estimate the impact?