**Regression Discontinuity Design (RDD) Approach**

# Regression Discontinuity Design (Sharp RD)

Use the data in schautonomy1.dta for the following questions. This questions is based on the study of ”The performance and competitive effects of school autonomy”, D. Clark (2009), *Journal of Political Economy* Vol 117, No 4.

The basic idea of the paper can be described very simply. Traditionally schools in the UK have been funded and managed by the Local Education Authorities (in London, this would be a borough e.g. Camden, Westminster) with rather little autonomy given to each school. But the 1988 Education Act allowed schools to opt out of LEA control and become funded by central not local government with much more autonomy

- this was called ‘grant-maintained’ (GM). Schools could become GM if the majority of parents chose that option in a ballot. In other words, if 51% of parents voted for GM status that school would become a GM-school, while if only 49% voted for it, it would remain under LEA control. This is the basis of the regression discontinuity design.

The study can be thought of as contributing more generally to the debate about how public institutions like schools or hospitals should be run: should they be given a budget and left to spend it how they want or should they be more tightly controlled. In the case of GM schools, becoming GM resulted not just in more autonomy, but also more resources were justified as the school now has to deal with some issues that were previously handled by the LEA. So the change to GM resulted in both more autonomy and possibly more resources.

The School Performance Tables (league tables) were first published in 1992 and contain the primary measure of student achievement, the fraction of grade 11 students who pass five or more GCSE examinations (the author calls this the school pass rate).

*scorei* = *β*0 + *β*1*GMi* + *ui* (1)

where *scorei* is the results of student *i* in a standardized math test and *GMi* is a dummy variable which takes on a value of one if the student is enrolled in a GM school and zero otherwise.

* 1. Describe and summarize the data.
  2. Estimate (1) by linear regression (using robust standard errors). Plot the results using linear fit and quadratic fit.

After 1988, a large number of schools held ballots about conversion to GM status. While many ballots were successful, there were also a large number of ballots where the majority of parents opposed to convert the school into a GM school. Suppose your dataset also include **the percentage of votes** that were cast in favour of conversion to a GM school.

* 1. Do a scatter-plot of the change in pass rate on the vote in favour of GM status.
  2. How would you exploit that additional information for an alternative estimation strategy to measure the causal effect of GM status on student achievement? Provide the appropriate regression specification using robust standard errors.
  3. What are the assumptions that need to hold for your approach to estimate the causal effect in D.
  4. Plot the results in D using quadratic fit.
  5. Instead of using *dpass* as the outcome variable, repeat your analyses in D using *passrate*2 as the outcome variable. Does the estimator of “win” change sign?
  6. Someone critical of the results suggests using *passrate*0 as the dependent variable. Suppose that the coefficient of the win variable had a *significant* negative sign. What does this imply?