Microeconometrics

Problem set 2

Note. Upload a copy of your solutions on Moodle. The solutions should include a .pdf document with the answers to each one of the questions **including the screenshot of command used to generate the output and the screenshot of results or the output** generated from the statistical software (feel free to use STATA or R, or other softwares). Please check the submission deadline on moodle, late problem sets will not be accepted.

1. (50%) **Regression Discontinuity.** De Kadt and Rosenzweig investigate partisan ties and their effects on national resource allocation in Ghana. The specific research question of interest is whether constituencies in Ghana that elect MPs who are from the same party as the President (the "ruling party") receive more electrification over the next four years. Using electoral data from the 1996 parliamentary elections and nightlights data the authors use a sharp regression discontintuity (RD) design to investigate the effect of the treatment (an MP from the ruling party winning the 1996 election) on the outcome (change in nightlights over the next four years). The forcing variable the authors use is voteshare of the ruling party MP candidate. The unit of analysis is the constituency. In this problem you will similarly conduct a sharp RD analysis using the dataset **votes_Ghana**. The dataset contains 152 observations each corresponding to a constituency in Ghana, with the following variables:

   - *constit*: name of consistuency
   - *voteshare*: voteshare (% votes obtained) for the ruling party MP candidate
   - *treatment*: treatment status indicator (1 if the ruling party MP won the 1996 election in that constituency, 0 if the ruling party candidate lost)
   - *changeNL_1996_2000*: change in nightlights between 1996 and 2000
   - *mean_1995*: mean level of nightlights in 1995

   (a) Look at the data to see if sharp RD makes sense for our dataset. To do so, plot treament (y- axis) as a function of the forcing variable (x-axis), where the forcing variable is the margin of victory/loss for the ruling party MP candidate. Does it seem appropriate to use a sharp regression discontinuity design in this case? [*HINT:* to do the plot you can use the STATA command `twoway scatter`. For the rest of the exercise remember that some commands requires the definition of the cut-off value or for simplicity they require that the forcing variable has a discontinuity at zero.]

   (b) Estimate the local average treatment effect at the threshold. What are assumptions required for this estimation strategy? Interpret your resulting estimate.

   (c) Conduct the same analysis as in part (b) except that you should now use a *linear* model with different slopes for treated and control units [*HINT*: add the interaction between the forcing variable and the treatment variable]

   (d) Conduct the same analysis as in part (b) except that you should now use a *quadratic* model with different model coefficients for the treated and control groups.

   (e) Use the `rd` command in STATA (or similar in other softwares) to estimate the Local Average Treatment Effect at the threshold using a local linear regression with a triangular kernel. Report your estimate with the s.e. and discuss your result in light of points (b), (c) and (d) [*HINT:* use the stata command `rd` with the default options, report the coefficient and the s.e. using a bandwidth of 100 percent]

(f) How do the estimates of the LATE at the threshold differ based on your results from parts (b) to (e)? In other words, how robust are the results to different specifications of the regression? What other types of robustness checks might be appropriate?

(g) Finally conduct a placebo test using nightlights measured in 1995 ($mean_1995$) as the outcome in a sharp RD analysis for the 1996 election. Use local linear regression as you did in part (e). What does this placebo test say about the relationship between the 1996 election of ruling party MPs and nightlights measured in 1995?

2. (50%) **Panel data models.** Consider a subset of the data used by Vella and Verbeek (1998) "*Whose Wages do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men*", **wagepan**, to estimate the effects of unions on workers' wages. The dataset is comprised of 545 men (identified by variable *nr*) who worked in every year from 1980 through 1987 (identified by variable *year*) in the United States. Consider the following model:

$$lwage_{it} = a_i + \theta_t D_t + \beta x_{it} + \pi z_i + u_{it}$$

where i denotes the worker and t the year. The vector x_{it} is comprised of *exper* (labor market experience) and its square *expersq*, *married* equals 1 if the individual is married, and *union* equals 1 if the worker is unionized. The vector z_i includes the variables *black* equals 1 for blacks, *hisp* equals 1 for hispanic workers, and *educ* denotes the number of years of education.

(a) Study the distribution of the variable *lwage*. Produce an histogram of the overall distribution and estimate the non-parametric density of the variable. Then compare the non-parametric density for black, hispanics and the remaining group ($black == 0$ and $hisp == 0$).

(b) Explain which effects parameters θ_t and a_i are likely to capture.

(c) If unions are successful in their wage negotiations with employers, what should be the sign of β_{union}?

(d) Estimate the equation by pooled OLS. Do you find any evidence for a union effect? Are the assumptions required for these estimates to be consistent plausible? If not, what would be the asymptotic bias you would expect in the union estimate?

(e) Considering the time-varying variables, estimate the equation using the within (FE) estimators.[1] What are the necessary assumptions for consistency of the estimator? Can we estimate the returns to education? Why? What about race effects and experience?

(f) Compare results from the pooled OLS and within (FE) estimator. What do you learn?

[1] In most softwares you need to define what is the panel. In STATA, for instance, you would run the following code: *xtset nr year*. This tells the software that i = nr and t = year.