

Sample questions:

Question 1

Download the paper “Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization” and the data and programs that are available with the paper.

- a) Give a short description (one page) of the empirical strategy applied in the paper, with an emphasis on what happens before Section IV. In particular, explain how the approach is related to a difference-in-differences approach, and in particular, how Figure 3 is used to validate the approach.
- b) In Tables 2 and 3 standard errors are adjusted for 91 meso-region clusters. Explain what this means and show how standard errors would have been without using this clustering.
- c) In Tables 2 and 3, results from a two-stage least squares (TSLS) estimation procedure are reported, along with various OLS results. In this TSLS procedure, a control variable is instrumented because endogeneity in the control variable may contaminate other results. Explain the difference between instrumenting a control variable and using the instrument directly in place of the control variable. Illustrate this replicating the TSLS results from Tables 2 and 3 and comparing with the alternative suggested approach.
- d) In the analyses reported in Tables 2 and 3, results are not very powerful because the authors only use data for a single year. Redo Tables 2 and 3 using data for all the years 1996-2003 and 2004-2010 for the two analyses. How do you handle autocorrelation?
- e) Based on the analysis in d), apply quantile regression to see if the mean regressions reported in Tables 2 and 3 hide distributional effects. Specifically, do quantile regression for the three quartiles and compare these with mean regression results. Interpret the results.
- f) Describe and apply alternatives to quantile regression - such as unconditional quantile regression - in the case with control variables. What is the difference in interpretation between standard and unconditional quantile regression?

Question 2

- a) Generate data for a regression discontinuity analysis. Specifically, generate a running variable taking on values on the unit interval. Assign treatment based on whether the running variable is above 0.6. Let expected outcomes be third order polynomials, with different coefficients at either side of the threshold. Introduce noise in observed outcomes. Present standard RDD plots.
- b) Generate such data sets repeatedly. Choose parameter values for the expected

outcomes and demonstrate how local linear models can outperform global linear models and global quadratic models. Is there a trade-off between bias and variance.

Question 3

a) Construct a data set with three groups: always-takers, never-takers and compliers. Generate potential outcomes as treated and untreated for the two groups. Let the expected potential outcomes differ between groups. Randomize a binary instrument that affects treatment and assign treatment and outcomes (based on expected potential outcomes and noise). Present summary statistics from the data set.

b) Demonstrate that two stage least squares, indirect least squares, the indirect estimation method in Imbens and Rubin (1997) and the procedure for estimating LATE based on “Abadie’s kappa” presented in “Mostly harmless econometrics” give identical results.