EC 425/525
Spring 2022

Professor Eric Zou
Department of Economics, University of Oregon

Empirical Project 2
The Effect of China's Huai River Policy on Air Pollution

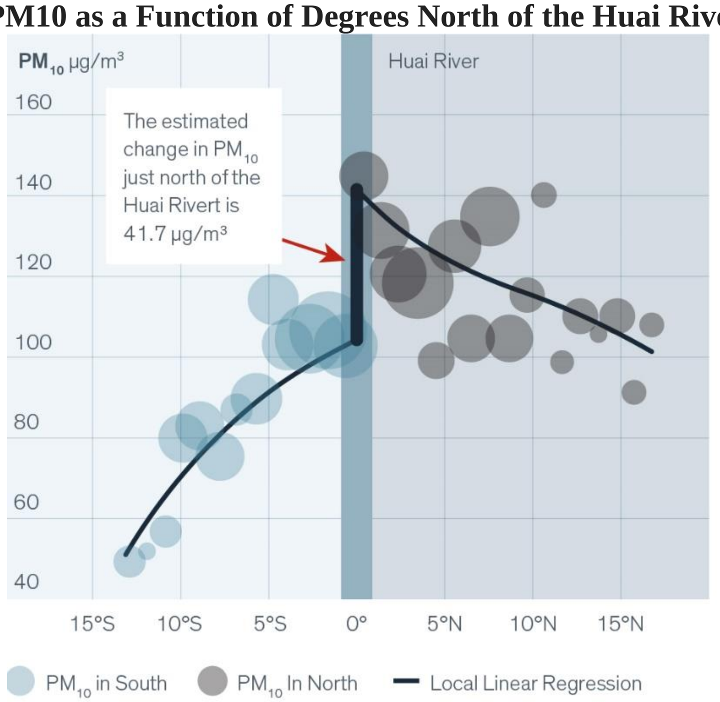
Due at midnight (Pacific Time) on Friday, May 20, 2022

In this empirical project, you will use a regression discontinuity design to estimate the causal effect of Huai River Policy on air pollution. To answer some of the questions, you will need to refer to the following paper:

Ebenstein, Avraham, Maoyong Fan, Michael Greenstone, Guojun He, and Maigeng Zhou. "New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy." Proceedings of the National Academy of Sciences 114, no. 39 (2017): 10384-10389. (https://www.pnas.org/content/114/39/10384; also available on Canvas: Files/reading/Ebenstein_et_al_2017.pdf)

The Stata data file **huairiver.dta** consists of geographic, weather, and air quality information for 161 cities in China. These data were originally used in Ebenstein et al. (2017). The life expectancy variables used in the original paper are confidential and not available. For this empirical project, we will focus on the impact of the Huai River Policy on PM10 air pollution. The graphs below were drawn using the Ebenstein et al. data.

Fig. 2 of Ebenstein et al. (2017)
PM10 as a Function of Degrees North of the Huai River



Note: This figure plots PM10 concentration as a function of cities' distance to the Huai River.

Instructions

Please submit your Empirical Project on Canvas. Your submission should be a single PDF file containing three parts:

1. A 4-6 page summary (double spaced and including references, graphs, and tables)
2. A do-file with your STATA code or an .R script file with your R code
3. Your STATA or R output

Specific questions to address in your summary

1. Explain why a simple comparison of air pollution in northern cities versus southern cities would not measure the causal effect of the Huai River Policy. Explain how did the Ebenstein et al. paper overcome this problem by using a regression discontinuity design.

2. Explain what is the outcome variable and what is the assignment variable in Fig.2 of the Ebenstein et al. paper?

3. What is a binned scatter plot? Explain how it is constructed.
   Hint: Binscatter plot is covered in week1's lecture.

4. Graphical regression discontinuity analysis.

   a. Draw a binned scatter plot to visualize how PM10 changes at the Huai River line. Display fitted lines (linear, or quadratic, or whatever functional form you see fit) based on what you see in the data.

   b. Draw binned scatter plots to test whether (i) temperature, (ii) precipitation, and (iii) wind speed changes at the Huai River line. Display fitted lines (linear, or quadratic, or whatever functional form you see fit) based on what you see in the data.

5. Regression analysis. Run the regressions that correspond to your three graphs in 4a and 4b to quantify the discontinuities that you see in the data. Report a 95% confidence interval for each of these estimates.

   Note: Part of this question is to get you to think about how to do bandwidth selection in a regression discontinuity design. Carefully read through how Ebenstein et al. choose bandwidth (e.g., page 4 of the paper) and the associated material in the SI Appendix. Try your best to replicate their bandwidth selection in your work.

6. Recall that any quasi experiment requires an identification assumption to make it as good as an experiment. What is the identification assumption for regression discontinuity design? Explain whether your graphs in 4b are consistent with that assumption.

7. Another type of validity test for regression discontinuity design is the manipulation test. Do we need to worry about manipulation in this study context? Explain why or why not. If you believe a manipulation test should be done, report such a test.

8. Consider the "placebo test" in Fig. 4 of the Ebenstein et al. paper.

    a. Explain the logic of the "placebo test" underlying Fig. 4. Why did the authors estimate regression discontinuity using *false* locations of the Huai River? What do the results of this test tell us?

    b. Replicate Fig. 4 of Ebenstein et al. (2017). Hint: To obtain cities' distance to a "placebo" Huai River that is 1-degree North of the true Huai River, simply add 1 to the "dist_huai" variable.

DATA DESCRIPTION, FILE: huairiver.dta

The data consist of *n* = 161 cities. For more details on the construction of the variables included in this data set, please see Ebenstein et al. (2017).

Variable Definitions in huairiver.dta

Variable	Label
dsp6	Unique identifier of city
wspd	Average wind speed (m/s)
temp	Average temperature (degree F)
prcp	Average precipitation (millimeter)
north_huai	1 if city to the north of the Huai River
dist_huai	Distance (in degrees) north of the Huai River. Negative distance = south of the river
pm10	Average PM10 level (ug/m3)

Note: This table describes the variables included in huairiver.dta.