AEM 4510/ECON 3865 Problem Set 4

Spring 2021

Due May 6

This problem set puts your new regression (`lm` or `fixest/feols`) skills to use. If you are unsure of how to do certain pieces of the problem set (e.g. take means of the data), Google and StackExchange are the best resources.

NIMBY Taxation Background

This exercise uses data from the application found in “State Taxes and Interstate Hazardous Waste Shipments,” by Arik Levinson (American Economic Review, 89(3)). Levinson studies the shipment of toxic waste between U.S. states. In an attempt to keep waste out, state governments often implement what are commonly known as “NIMBY” (Not In My BackYard) taxes on waste dumping. These taxes may differ depending upon where the waste comes from (i.e., a state typically charges a lower tax for dumping waste produced within its own borders than it charges for waste imported from another state).

Levinson’s question is whether these taxes work in deterring waste flows. This is an important question in environmental economics because, in addition to functioning as NIMBY taxes, these taxes also behave like Pigouvian taxes – interstate waste flows are a form of emissions whereby polluters in the origin state impose externalities (e.g., risk of leakage into groundwater) on the residents of the state. If waste flows are not responsive to these taxes, we may worry about the ability of Pigouvian policy to address these externalities.

The problem in determining whether these taxes are effective arises because we (as the economists) do not see everything that makes a state attractive or unattractive as a potential destination for waste. Unobserved state attributes that make some states more appropriate destinations for waste may be correlated with the tax that the state’s legislature sets, i.e. there might be omitted variable bias. To be more precise, state legislatures in states that are good destinations for waste may set higher taxes to deter waste flows. In the end, those states might still receive a higher volume of waste (despite their higher taxes), but a lower volume than they would have received in the absence of those taxes. By failing to adequately control for destination state attributes, we may mistakenly arrive at the result that NIMBY taxes encourage more waste flows.

The purpose of this exercise is for you to demonstrate this result using actual data. The data are available in an Excel spreadsheet on Canvas named “levinson_data.xls”. You should begin by downloading these data or using the csv that is uploaded on RStudio Cloud. You should have 16128 observations in the dataframe.

```r
library(tidyverse)
library(broom)
library(fixest)
library(readxl)

ps_dataset <- readxl::read_xls("data/levinson_data.xls")
ps_dataset
```

```
## # A tibble: 16,128 x 79
##    `orig id` `dest id` shipments  dist    dist2  oinc   oarea   opop odens  ocap
##        <dbl>     <dbl>     <dbl> <dbl>    <dbl> <dbl>   <dbl>  <dbl> <dbl> <dbl>
##  1         1         1     16.1     0        0   23.6 0.0508   4.04   79.6 1.10
##  2         2         1     10.5   358.  127812.  21.1 0.0521   2.35   45.1 0.835
##  3         3         1      0    1418. 2011561   27.5 0.114    3.66   32.2 2.26
##  4         4         1      0    1945. 3784079   35.8 0.156   29.8   191.  2.26
##  5         5         1      5.53 1129. 1275081   30.1 0.104    3.29   31.8 2.26
##  6         6         1      8.63  978.  956302.  41.7 0.00485  3.29  678.  0.439
##  7         7         1      9.24  761.  578994.  34.9 0.00196  0.666 341.  0
##  8         8         1     13.7   361.  130444.  27.5 0.0540  12.9   240.  0.347
##  9         9         1     15.0   181.   32916.  29.0 0.0579   6.48  112.  0.400
## 10        10         1     13.3   733.  537517   26.2 0.0559   2.78   49.7 0.122
## # … with 16,118 more rows, and 69 more variables: o65 <dbl>, ocoll <dbl>,
## #   ogen <dbl>, dinc <dbl>, darea <dbl>, dpop <dbl>, ddens <dbl>, dage65 <dbl>,
## #   dcollege <dbl>, dcapmil <dbl>, dwest <dbl>, dne <dbl>, dsouth <dbl>,
## #   yr90 <dbl>, yr91 <dbl>, yr92 <dbl>, yr93 <dbl>, yr94 <dbl>, yr95 <dbl>,
## #   samest <dbl>, tax <dbl>, d1 <dbl>, d2 <dbl>, d3 <dbl>, d4 <dbl>, d5 <dbl>,
## #   d6 <dbl>, d7 <dbl>, d8 <dbl>, d9 <dbl>, d10 <dbl>, d11 <dbl>, d12 <dbl>,
## #   d13 <dbl>, d14 <dbl>, d15 <dbl>, d16 <dbl>, d17 <dbl>, d18 <dbl>,
## #   d19 <dbl>, d20 <dbl>, d21 <dbl>, d22 <dbl>, d23 <dbl>, d24 <dbl>,
## #   d25 <dbl>, d26 <dbl>, d27 <dbl>, d28 <dbl>, d29 <dbl>, d30 <dbl>,
## #   d31 <dbl>, d32 <dbl>, d33 <dbl>, d34 <dbl>, d35 <dbl>, d36 <dbl>,
## #   d37 <dbl>, d38 <dbl>, d39 <dbl>, d40 <dbl>, d41 <dbl>, d42 <dbl>,
## #   d43 <dbl>, d44 <dbl>, d45 <dbl>, d46 <dbl>, d47 <dbl>, d48 <dbl>
```

Variable Names and Descriptions

Note the actual variable names in the dataframe do not have the subscripts, I include them here so you know whether they refer to the origin state, destination state, or year.

- Subscript i: origin state
- Subscript j: destination state
- Subscript t: year
- $shipments_{i,j,t}$: natural log of shipments from state i to state j in year t
- $dist_{i,j}$: distance (miles) from state i to state j
- $oinc_i$: origin state i income in 1989 (thousands of dollars)
- $opop_i$: origin state population in 1990 (millions)
- $oarea_i$: origin state area (millions of square miles)
- $odens_i$: origin state population density (persons per square mile)
- $o65_i$: origin state percent over age 65
- $ocoll_i$: origin state percent with a college degree
- $ocap_i$: origin state waste disposal capacity
- $ogen_i$: origin state natural log of waste generation
- $dinc_j$: destination state j income in 1989 (thousands of dollars)
- $dpop_j$: destination state population in 1990 (millions)
- $darea_j$: destination state area (millions of square miles)
- $ddens_j$: destination state population density (persons per square mile)
- $d65_j$: destination state percent over age 65
- $dcoll_j$: destination state percent with a college degree

- $dcap_j$: destination state waste disposal capacity
- $dwest_j$: equal to 1 if destination state is in the West region, equal to 0 otherwise
- dne_j: equal to 1 if destination state is in the Northeast region, equal to 0 otherwise
- $dsouth_j$: equal to 1 if destination state is in the South region, equal to 0 otherwise
- $y90_t$: equal to 1 if year of waste shipment is 1990, equal to 0 otherwise
- $y91_t$: equal to 1 if year of waste shipment is 1991, equal to 0 otherwise
- $y92_t$: equal to 1 if year of waste shipment is 1992, equal to 0 otherwise
- $y93_t$: equal to 1 if year of waste shipment is 1993, equal to 0 otherwise
- $y94_t$: equal to 1 if year of waste shipment is 1994, equal to 0 otherwise
- $y95_t$: equal to 1 if year of waste shipment is 1995, equal to 0 otherwise
- $tax_{i,j,t}$: disposal tax per ton faced by waste shipped from origin state i to destination state j in year t
- $samest_{i,j}$: equal to 1 if origin state i and destination state j are the same, equal to 0 otherwise
- $d1_j - d48_j$: equal to 1 if destination state j is the same state as the variable name (1, 2, …, 48), equal to 0 otherwise (only one of these will be equal to one for a given row of the dataframe)

Problems

1. Determine the mean tax charged on waste shipments in the data set.
2. Next, determine the mean tax charged on waste shipments that cross state boundaries and compare it to the mean tax charged on waste that is produced “in-state”. Which is bigger?
3. Regress the natural log of waste shipments on the tax $tax_{i,j,t}$; distance $dist_{i,j}$ and distance squared $dist2_{i,j}$; all the attributes of the origin state $oinc_i - ogen_i$; the dummy variable for whether they are the same state $samest_{i,j}$; and all the year dummy variables $y90_t - y95_t$.
   a. What is the coefficient on tax?
   b. What is the elasticity of waste shipment with respect to the tax (i.e. the percent change in shipments given a 1% increase in the tax)?
      - Recall regression formulas in R are:
        `lm(dependent variable ~ ind. var. 1 + ind. var. 2 + ..., data = dataframe)` or
        `feols(dependent variable ~ ind. var. 1 + ind. var. 2 + ..., data = dataframe)`
4. Repeat 3, but control for (i.e. add them as independent variables) all the destination state attributes that might capture attractiveness of the state to dumpers: $dinc_j - dsouth_j$ above.
5. Repeat 3, but instead include all destination dummy variables: $d1_j - d48_j$ above. These will control for *all* attributes of the destination states that are not changing over time, that may make them attractive to dumpers. When you include these dummy variables you no longer will be able to include destination-specific variables like $dinc_j$.
6. Compare and contrast your results for 3b, 4b, 5b. You may have different estimates of the elasticity across the different questions, potentially even with different signs. Explain why controlling for the different sets of variables results in different estimates for the effect of the tax on shipments. Feel free to use examples to explain.