

Pol Sci 630: Problem Set 12: 2SLS, RDD, ggplot2

Prepared by: Anh Le (anh.le@duke.edu)

Due Date: Wed, Nov 23, 2016 (Beginning of Class)

1 2SLS

Insert your comments on the assignment that you are grading above the solution in bold and red text. For example write: "GRADER COMMENT: everything is correct! - 8/8 Points" Also briefly point out which, if any, problems were not solved correctly and what the mistake was. See below for more examples.

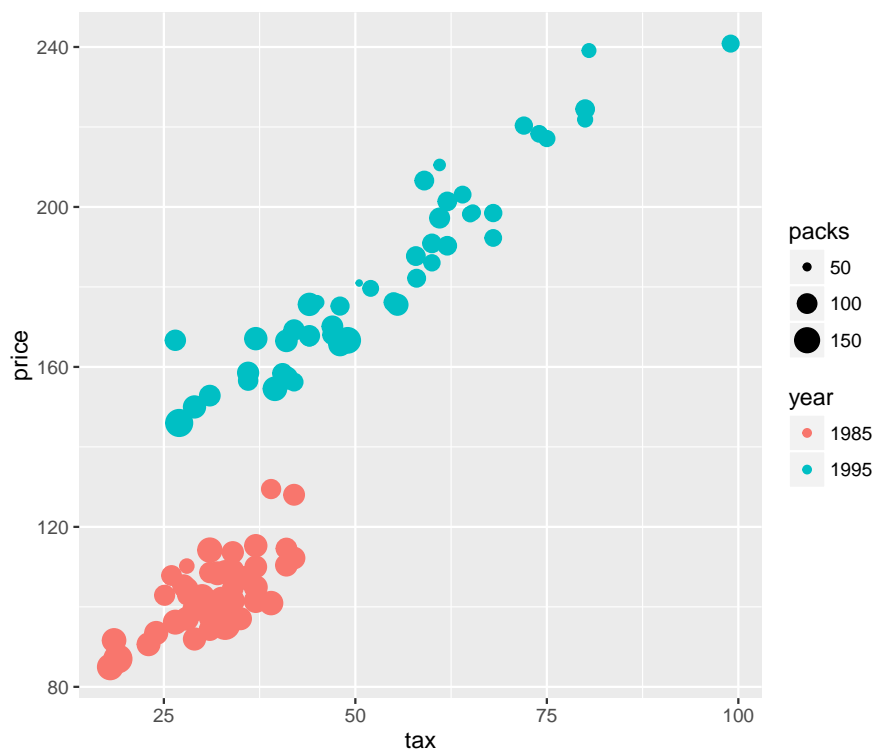
1.1 Load dataset CigarettesSW from package AER

```
library(AER)
data("CigarettesSW")
```

1.2 Plot the following using ggplot2

What can we say about the relationship between tax, price, and packs? Importantly, could sales tax be a valid instrument here? Explain your reasoning.

Note: This is a good way to show the relationship between 3 variables with a 2D plot.



1.3 Divide variable income by 1000 (for interpretability)

1.4 Run 2SLS

Run 2SLS with `ivreg`. Outcome: packs. Exogenous var: income. Endogenous var: price, whose instrument is tax. Interpret the coefficient of `income` and `price`.

Note: Different from the model during lab, this model has an exogenous independent variable that doesn't need to be instrumented for. See `'help(ivreg)'` > Details, which explains how to deal with this.

1.5 2SLS diagnostics: use F-test to check for weak instrument

1.6 2SLS by hand

Run the 2SLS by hand, i.e. not using `ivreg`, but run 2 stages of `lm`. Do you get the same estimate from `ivreg`?

1.7 Weak instrument test by hand

The weak instrument test aims to test whether the instrument is an important predictor of the endogenous variables, even after controlling for other variables.

We do it as follows:

- Run the standard 1st stage regression of endogenous var ~ instrument + exogenous vars
- Run a “modified” 1st stage regression of endogenous var ~ exogenous vars
- Use `waldtest(model1, model2)` to compare the two models (to see if the model with the instrument fits better). The null hypothesis is that the instrument has a statistically significant impact

The rule of thumb is that the F-statistic should be > 10

Implement the weak instrument test as described above and show that it gets the same F-statistic as given by `ivreg`.

2 Regression Discontinuity Design

Find the replication data here <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/900LQ7>

```
# Load the data like this
# load("replication_data.RData")
```

Variables that you’ll use: `elecpratio` (% of the population as the electorate), `treat` (whether got audited or not), `electorate.perpop07` (% registration in 07), `electorate.perpop08` (% registration in 08)

Read (by which I mean Ctrl + F) through the paper to figure out which bandwidth and the cutoff points the authors used. Read `help(rdrobust)` to see how to specify our own bandwidth and cutoff points.

In this exercise we’ll replicate their main results in Table 2 (p. 447)

2.1 Sharp RDD

Use `rdrobust` to estimate the RDD effect of `elecpratio` > 0.8 for Change in registration (%) ($\hat{\tau}_A$ in Table 2), using 1) the author’s bandwidth, and 2) the bandwidth chosen by `rdrobust` itself.

2.2 Fuzzy RDD

The design of this paper is a Fuzzy RDD because when `elecpratio` > 0.8 , a district may be audited but not necessarily.

`rdrobust` has an argument `fuzzy` to specify which observation is actually treated. Use it to get a Fuzzy RDD estimate for Change in registration (%) ($\hat{\tau}_R$ in Table 2)

2.3 Density test (graphical)

Plot the histogram of the number of observations on both sides of the cut-off to see if there's any difference