DiD Problem Set

Problem 1: Diff-in-diff I

Please read Section II of Chetty, R., A. Looney and K. Kroft (2009), Salience and Taxation: Theory and Evidence, American Economic Review, 99(4), pp. 1145-1177.

i) The authors conducted an experiment in which certain product categories were assigned tax-inclusive price tags. As can be seen in Table 3, demand for the treated products decreased by 1.3 units. What would you have to assume to interpret this pre-post difference as a causal effect?

ii) The difference-in-difference estimate in Table 3 is -2.14 units. What assumptions can you relax when relying on this estimate in contrast to the pre-post difference?

iii) How do the authors calculate the standard errors and why? What is the purpose of footnote 10?

iv) In Panel B of Table 3, the authors add a second control group and calculate a "DDD Estimate". What potential concerns are addressed by this empirical strategy? Which concerns still remain?

Problem 2: Diff-in-diff II

In this problem, we are going to work with the dataset sodasales.dta, which you can download from github. This is data from 5,000 stores in 50 states (100 stores per state). A couple of states have implemented a tax on soda (indicated by treat=1), while others have not (treat=0). You have data from one year before the tax (indicated by post=0) and one year after the tax introduction (post=1). The variable sales1 measures the sales of soda in each of the stores in 1,000 liters.

i) Calculate the average sales by period (pre/post) and states (treated/control) using the data set and present the results in a 2x2 table. Calculate the difference-in-difference estimate of the introduction of the tax on soft drink sales based on these means (i.e. not using regression analysis). Are the findings consistent with economic theory?

ii) Suppose you would like to estimate the impact of the tax on sales in a DiD regression model. Write down the regression equation (do not forget to define the variables). Estimate the model using Stata and interpret the regression output.

iii) What are the assumptions you need to make so that the difference-in-difference estimator identifies the causal impact of the introduction of the tax? What are potential threats to estimating the causal effect of the tax on sugar consumption?

iii) How do you calculate your standard errors in this setting and why? Do your standard errors differ if you calculate robust standard errors instead?

iv) The variable sales2 measures the sales of bottled water from these stores. Assuming no substitution between soda and water, write down the DDD regression equation that you could estimate. Estimate the DDD regression model using Stata and interpret each of the coefficients. (Hint: You have to reshape the data first.) How would this estimate be biased if consumers do in fact substitute in reaction to the tax?