Homework 4

Economics 7103

Due Sunday, February 21 by 11:59 pm

You have imaginary data on the monthly yields for Pacific fish trawling companies (*fishbycatch.csv*). An environmental nonprofit targeted these firms and implemented a program designed to reduce bycatch. As part of the program, the nonprofit contacted firm managers and provided information about best practices to reduce bycatch. The program was implemented in two phases. In January 2018, the nonprofit contacted half of the firms. The next year in January 2019, the nonprofit contacted the remaining firms.

   You are interested in whether the program worked or not and decide to use this panel data to empirically estimate the effect of the program. You realize that you have a treatment and control group in pre- and post-treatment periods due to the program's rollout, so you think a difference-in-differences design is a good approach. You have the following data:

Variable	Description
firm	Firm identification number
shrimp*	Pounds of shrimp in month *
salmon*	Pounds of salmon in month *
bycatch*	Pounds of bycatch in month *
firmsize	Size of fishing fleet
treated	=1 if firm received information treatment in January 2018

Table 1: Variable descriptions for homework 3.

   Note that to convert these panel data from wide form to long form, you can use the Pandas `wide_to_long()` function.

1. Visually inspect the bycatch by month before and after treatment for treated and control groups by creating a line plot for months in 2017 and 2018. Does it appear that there are parallel trends before treatment? (Hint: I found the Pandas function `groupby()` useful.)

2. Estimate the treatment effect of the program on bycatch using the sample analog of the population difference-in-differences for treatment and control groups in December 2017 and January 2018. The population difference-in-differences is:

$$DID = \{E[Y_{igt}|g(i) = treat, t = Post] - E[Y_{igt}|g(i) = treat, t = Pre]\} \tag{1}$$
$$- \{E[Y_{igt}|g(i) = control, t = Post] - E[Y_{igt}|g(i) = control, t = Pre]\}. \tag{2}$$

   Simply report the estimate without a standard error. What is the intuition of the estimator?

3. Estimate the treatment effect of the program on bycatch using a regression-based two-period difference-in-differences estimator with estimating equation:

$$bycatch_{i,t} = \alpha + \lambda_{t=2017} + \gamma g(i) + \delta treat_{i,t} + \varepsilon_{i,t}, \tag{3}$$

   where $\lambda_{t=2017}$ is a separate intercept for the pre-period (December 2017), $g(i)$ is an indicator that firm i is in the treatment group, and $treat_{i,t}$ is an indicator variable equal to one when a firm is treated. Your estimating sample should include the observations in December 2017 and January 2018 only. Report the results in a table with standard errors or confidence intervals calculated using clustered standard errors at the firm level. How does this result compare to your previous calculation?