Homework 4

Economics 7103

Due Monday, February 13th 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.

1 Python

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:

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

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

3. Estimate the treatment effect using the following regression specifications and report all coefficients, standard errors (or confidence intervals), and observations in a single table.

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

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

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.

(b) Suppose you would like to use the full monthly sample to improve on what you did in the previous question. Using the full monthly sample, estimate the treatment effect of the program on bycatch using a regression-based difference-in-differences estimator using the regression:

$$bycatch_{i,t} = c_i + \lambda_t + \gamma g(i) + \delta treat_{i,t} + \varepsilon_{i,t}. \tag{4}$$

where λ_t are indicator variables for each time period. Report and interpret the results using the same cluster-robust standard errors. How did your results change?

(c) Suppose now that you want to control for firm size and other covariates that change over time such as pounds of shrimp and salmon harvested. Estimate the difference-in-differences regression with added controls:

$$bycatch_{i,t} = c_i + \lambda_t + \gamma g(i) + \delta treat_{i,t} + \beta X_{i,t} + \epsilon_{i,t} \tag{5}$$

where $X_{i,t}$ includes firm size, pounds of shrimp harvested by firm i in month t, and pounds of salmon harvested by firm i in month t. Report and interpret the results using the same cluster-robust standard errors. How do your results change from question 1?

(d) Report the results from (a), (b), and (c) in a table with standard errors or confidence intervals calculated using clustered standard errors at the firm level. Omit the estimates of the coefficients on the month and firm indicators in your table. How do these results compare to your previous calculation?

2 Stata

Note that to convert these panel data from wide form to long form, you can use the Stata `reshape` function.

1. You now would like to allow and control for firm-specific fixed-effects. In particular, you would like to allow for an unobserved effect c_i that varies at the firm level but not over time:

$$bycatch_{i,t} = c_i + \lambda_t + \delta treat_{i,t} + \beta X_{i,t} + u_{i,t}. \tag{6}$$

(a) Generate indicator variables for each firm. Include these indicator variables in your OLS regression to control for fixed effects directly and estimate equation (6).

(b) Perform the "within-transformation" on all of the dependent and independent variables by demeaning each variable (i.e. instead of estimating $y_{i,t} = \beta x_{i,t} + \xi_{i,t}$, estimate $y_{i,t} - \bar{y}_i = \beta(x_{i,t} - \bar{x}_i) + e_{i,t}$) and estimate (6).

(c) Display the results of your estimates from (a) and (b) in the same table, reporting the same clustered standard errors or confidence intervals as previously. Omit the estimates of the coefficients on the month and firm indicators in your table. How do the results from (b) compare to (a)? How do these estimates compare to the previous estimates of the treatment effect and how does the interpretation change? (Note for the future that standard errors from (a) are typically "wrong", but do not worry about that for this homework. In addition, (a) is computationally costly when the panel size is large—in general you should use the within transformation to control for fixed effects.)