

## Problem Set IV - Difference-in-Difference

*Enrijeta Shino*

June 22, 2018

### 1 Problem 1: Panel Data

Take the data set "jtrain1" from <http://www.stata.com/texts/eacsap/> This file has data on firms and the amount of job training they get<sup>1</sup>

**1.1:** Use data from 1987 and 1988 only. Construct the difference-in-differences estimator in the following two ways:

**1.1a:** Construct the four means (control, treatment  $\times$  before, after) and calculate the differences<sup>2</sup>

$$\begin{aligned} DID &= \mathbb{E}[Y(i)|T = 1, G = 1] - \mathbb{E}[Y(i)|T = 0, G = 1] - \left\{ \mathbb{E}[Y(i)|T = 1, G = 0] - \mathbb{E}[Y(i)|T = 0, G = 0] \right\} \\ &= 35.978 - 7.672 - 9.671 + 8.889 = \mathbf{27.524} \end{aligned}$$

---

<sup>1</sup>See Holzer, H. J., Block, R. N., Cheatham, M., & Knott, J. H. (1993). Are training subsidies for firms effective? The Michigan experience. *ILR Review*, 46(4), 625-636.

<sup>2</sup>Your results should look similar to the Table 2 in the Holzer et al (1993) paper.

Table 1: Annual Hours of Training per Employee, by Year and Receipt of Grant. (Means and Standard Deviations in Parentheses)

Description	1987	1988	1989
Firms Receiving Grants in 1988	7.672 (19.5748)	35.978 (36.9560)	10.056 (19.4746)
Firms Receiving Grants in 1989	6.125 (11.3093)	9.316 (17.3605)	50.650 (36.2177)
Firms Not Receiving Grants	8.889 (17.4030)	9.671 (18.1773)	11.594 (22.3806)

Note: The sample sizes for the firms receiving grants in 1988 and firms receiving grants in 1989 are 35 and 28. We replicate Table 2 in Holzer et al. (1993). We are able to exactly match the first two rows of the table however the results are slightly different for the last row but with minor changes.

**1.1.b:** Run the regression

Table 2

	<i>Dependent variable:</i>	
	<b>log(hrsemp)</b>	<b>hrsemp</b>
	(1)	(2)
Grant	-0.518 (0.436)	-1.667 (4.157)
Year1988	-0.391 (0.303)	0.332 (3.046)
Grant×Year1988	<b>1.809***</b> (0.554)	<b>27.974***</b> (6.007)
Constant	2.178*** (0.227)	9.339*** (2.165)
Observations	152	256
R <sup>2</sup>	0.099	0.150
Adjusted R <sup>2</sup>	0.081	0.140
Residual Std. Error	1.536 (df = 148)	20.993 (df = 252)
F Statistic	5.431*** (df = 3; 148)	14.833*** (df = 3; 252)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

**1.2b:** Use the same data from 1987 and 1988. Run the fixed effect regression ( $\alpha_i$  is the unobserved heterogeneity)

Table 3

	<i>Dependent variable:</i>	
	log(hrsemp)	hrsemp
	(1)	(2)
Grant	−0.825* (0.462)	−3.015 (3.854)
Year1988	−0.391 (0.311)	0.333 (2.523)
Grant × Year1988	<b>1.814***</b> (0.537)	<b>27.971***</b> (7.787)
Fixed Effects (fcode)	0.0001** (0.00004)	0.0004 (0.0003)
Constant	−36.002** (14.916)	−177.176 (119.876)
Observations	152	256
R <sup>2</sup>	0.142	0.156
Adjusted R <sup>2</sup>	0.119	0.142
Residual Std. Error	1.505 (df = 147)	20.966 (df = 251)
F Statistic	6.083*** (df = 4; 147)	11.562*** (df = 4; 251)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

**1.3:** Do you get exactly the same answer between 1.1(a,b) and 1.2, why or why not?

No we do not get exactly the same answer on the interaction term as in part 1.1(a,b) when compared to 1.2, but the answers are very close. The estimated coefficient on the interaction which represents the DID estimate is lower when compared to both parts 1.1(a,b). The reason is that we did not consider the fact that firms may be different for reasons unobservable to the econometrician however we control for firms fixed effects under the assumptions that they do not change over time. (e.g. we cannot observe if some firms are more experienced and have better managerial skills than other firms.).