

## 5 Difference-in-Difference

### 1. Identification in the DiD model with constant effects.

Assume the usual DiD setup (two groups, two periods) and the following simple regression model:

$$y_{it} = \delta_0 + \delta_1 \mathbf{1}(T_t = 2) + \delta_2 \mathbf{1}(D_i = 1) + \delta_3 \mathbf{1}(T_t = 2) \mathbf{1}(D_i = 1) + u_{it}, \quad (1)$$

and write the regression error as

$$u_{it} \equiv \lambda_t + \theta_i + \epsilon_{it} \quad (2)$$

- (a) Show that  $\delta_3$  is the DiD estimator, and that this simple estimator eliminates both  $\lambda_t$  and differences in  $\theta_i$  across groups.
- (b) State the assumptions about the time-varying component of the error term,  $\epsilon_{it}$ , needed for identification of  $\delta_3$ .

### 2. Replication of Card and Krueger (1994)

In this exercise we ask you to replicate some of the results in Card and Krueger (1994). On April 1, 1992 New Jersey raised its minimum wage from \$4.25 to \$5.05. The minimum in Pennsylvania remained at the federal level of \$4.25. Download the dataset called `did.dta` from the OLAT. It contains data collected by David Card and Alan Krueger on fast food restaurants in New Jersey (NJ) and eastern Pennsylvania (PA) during March and November/December of 1992. Use this data to analyze the impact of the minimum wage increase in New Jersey on employment in the fast food industry.

Throughout, variable names with a trailing “2” refer to the second (Nov./Dec.) wave of the data, the same names without any number refer to the corresponding variable from the March wave. We constructed a few additional variables not on the original dataset: “fte” and “fte2” are full time equivalent employment, it is the sum of the number of full time employees and one half the number of part time employees, excluding managers; “dfte” refers to the change in full time equivalent employment between the second and first interview (fte2 - fte); “dw” refers to the change in the starting wage between the second and first interview; “gap” is the proportionate increase in the March starting wage which is necessary to comply with the new minimum wage, or

$gap = (5.05 - wage\_st) / wage\_st$  if in NJ and  $wage\_st < 5.05$   
 $gap = 0$  if in NJ and  $wage\_st \geq 5.05$   
 $gap = 0$  if in PA

and “sample” is dummy variable which is 1 if both wage and employment data are available in both the first and second interview wave, and 0 otherwise. Conduct the following analysis for the part of the dataset with `sample` equal to 1. If you don’t specify this, Stata will make calculations with the full set of available observations for each variables, so you may not be comparing the same set of restaurants between March and November, or you may compare wages and employment for different restaurants.

- (a) Calculate the average employment separately for restaurants in NJ and in PA, both for each interview wave. Use full time equivalent employment. Calculate the difference in average employment between the second and first interviews. Now calculate the

difference between NJ and PA of the time differences just obtained. Provide a short interpretation of the difference-in-differences estimate of the employment effect. Under what conditions does this provide a valid estimate of the employment effect of the minimum wage increase? Interpret your finding.

- (b) An alternative to comparing NJ and PA restaurants is comparing restaurants within NJ who have high and low wages before the minimum wage increase. Restrict your sample to restaurants in NJ and calculate the employment and wage change for those restaurants paying starting wages of \$5.00 and more before and the minimum wage increase and those paying less than \$5.00. What is the relative impact of the minimum wage on starting wages and employment within NJ?
- (c) The difference-in-difference estimates in (a) to (b) can also be obtained by regressing the change of employment (or the change in the wage) on a constant and a dummy variable for the state you are in (called “state” in the dataset). Repeat the exercises in (a) to (b) using regressions. Regression also allows you to control for other factors. Repeat the regressions, entering a dummy variable for whether the restaurant is company owned “co.owned”, as compared to franchised) and three dummy variables for three of the four chains in the dataset (Burger King, KFC, Roy Rogers, and Wendy’s; you will have to construct the dummies from the variable “chain”). Do your results change?
- (d) Remember that the variable gap is set to zero for all restaurants in PA because they weren’t affected by a minimum wage increase. Create a new variable just like gap for restaurants in PA, but pretend that there was a minimum wage increase to \$5.05 in PA over this period. Now run a regression of changes in employment and wages just for PA using this new variable. How do your results differ from those for NJ? Why is this last regression a check on how well our methodology is doing in uncovering effects of the minimum wage increase?

### 3. Synthetic Control Method: The Effect of Reforms

In this exercise we ask you to use the synthetic control method to estimate the effect of political reforms on GDP growth and economic mobility across generations.

The synthetic control method (SCM), developed by [Abadie and Gardeazabal, 2003], is optimal when you have a single unit that is treated (in this case a country) and many other units that are not treated, all of which could be a control but none of which is a perfect one. This method creates an appropriately weighted average of the control units which best approximates the evolution of the outcome variable in the treated unit *before* treatment.

The zip file `SCM.zip` contains the following datasets:

- **IMFreforms.dta**: A panel dataset of countries constructed by the IMF, taken from [Giuliano et al., 2013]<sup>1</sup>, containing liberalization reform indicators in the following sectors:
  - (a) Agriculture
  - (b) Networks: Telecoms and Electricity Industries
  - (c) Trade: Tariff Rates
  - (d) External Capital Account Liberalization
  - (e) Current-account restrictions

<sup>1</sup>The full dataset is available here <https://www.aeaweb.org/articles?id=10.1257/mac.5.4.179>. See [Ostry et al., 2009] for more details on how it was constructed

## (f) Domestic Financial Sector Liberalization

Indicators are increasing in the degree of liberalization.

- `GDMIMay2018.csv`: The Global Database on Intergenerational Mobility ([GDMI](#)): a panel dataset of economic mobility across generations constructed by the World Bank.
- (a) Choose the treatment  $D$ : pick a country and a specific reform from the list provided here below. Succinctly describe what happened: what reform is it? when was it enacted? what is the background to this reform? (*Hint: you might need to do some online research to better understand this background*)
  - (b) Choose the control group: what set of countries can be included in the synthetic control to estimate the counterfactual? Describe the criteria used to make this choice, and justify each restriction in time or geography that you decide to make, if any.
  - (c) Merge all of the dataset and describe the pattern of missing values, with special focus on the countries that you decided to focus on. Construct any additional variable needed for the regression (e.g. GDP growth rates, population growth, etc.). Do you need to include any variables as controls? Which ones? Should you construct the weights of the synthetic control using the outcome variable only, or also adding some controls? Why? (*Hint: Look for relevant literature online*)
  - (d) Focus first on GDP growth. Perform the SCM estimation. The [Synth Package](#) provided on Jens Heinmüller web-page can be very useful. Describe your findings: How are the pre-trends? Did the algorithm work correctly? Which countries have a big/small weight? What are the estimated effects of this reform? Are the result statistically significant? Are the results economically significant (i.e. is the magnitude of the effect big enough to be relevant)?
  - (e) Now focus on intergenerational mobility. Since the outcome variable is present only every decade, how should you format your data? Perform again the SCM estimation. Do you need different controls? Describe your findings.

### List of Case Study (Reforms)

- Indonesia 1970 (capital account)
- UK 1979 (capital account)
- Ecuador 1975 (capital account)
- Japan 1979 (capital account)
- Ireland 1979 (capital account)
- Chile 1979 (capital account)
- Netherlands 1977 (financial sector)
- Germany 1973 (financial sector)
- Switzerland 1977 (financial sector)
- Chile 1974 (financial sector)
- Chile 1975 (agriculture)
- Chile 1978 (networks)
- Jordan 1965 (trade)

## Working with Synthetic Control Method in R: some tips

When using the Synth Package in R, it might happen to have an "unbalanced panel" error. One solution might be to re-write your code following the structure of this code here:

<https://github.com/ulfelder/coups-and-growth/blob/master/synth.THL.R#L63>

Something else that might work is to "clean" your dataset in the following way:

1. delete all the rows which have a "NA" in the variable for GDP per capita (`rgdpch`);
2. restrict a bit the time period you look at (making sure you have a reasonable number of pre- and post- observations);
3. make sure that for *each* country you have the same number of periods;
4. replace all the missing values in your variable of interest with the average of that variable of interest;
5. set up the standard code for synthetic control looking at the pdf on Jens Heinmüller web-page.

These two ideas are not bulletproof solutions, but we hope will help you completing the exercise.

## References

- [Abadie and Gardeazabal, 2003] Abadie, A. and Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1):113–132.
- [Giuliano et al., 2013] Giuliano, P., Mishra, P., and Spilimbergo, A. (2013). Democracy and Reforms: Evidence from a New Dataset. *American Economic Journal: Macroeconomics*, 5(4):179–204.
- [Ostry et al., 2009] Ostry, J. D., Prati, A., and Spilimbergo, A. (2009). Structural Reforms and Economic Performance in Advanced and Developing Countries. *IMF Occasional Paper Series*, 268(iv).