5304 Econometrics Autumn 2023[1]

Problem set 5
Panel data and DiD

Part 1a: Basic and dynamic differences-in-differences

In this assignment, we are using an excerpt from the data set used in Acemoglu and Angrist (2001). The paper investigates the labor market effects for disabled workers from the Americans with Disabilities Act that came into effect in 1992. The authors use a DiD design where non-disabled workers act as a control group for disabled workers affected by the new law. We focus on the sample of males aged 21–39, available in *PS5_ AngristAcemoglu01.dta*.

1. First, you need to prepare the data set. Keep only workers with a business income equal to zero (*incbus*). Also, the variable *year* refers to the survey year, which asks questions about work related matters in the previous year. Generate a new year variable that captures the *working* year instead of the survey year. Also generate a dummy for being disabled. *Hint: check value labels to figure out how the variables are coded.*

2. It is good practice to begin by plotting the data. Show the time trends in weeks worked ($wkswork1$) for disabled and non-disabled workers and indicate with a vertical line when the ADA came into effect. Describe what you see and assess whether the parallel trends assumption seem to hold.

3. I will now guide you on how to write down the econometric specification of a so-called dynamic DiD model. Denote weeks worked of individual i at working year t as Y_{it}. Let D_i be a dummy for whether an individual is disabled or not. Let y_t be a dummy for year t (so that y_{1988} is a dummy that takes on the value 1 in year 1998, and 0 otherwise). Now, write out a specification of Y_{it} on being disabled, year dummies, and a set of interactions that allow the coefficients on the year dummies to vary depending on whether the individual is disabled or not.[2] Let year 1987 be the omitted category.

4. Estimate the dynamic DiD model that you wrote down. Use the function $coefplot$ (available in both R and Stata) to show the coefficients on the interactions between year and disability in a figure. Does there seem to be an effect of ADA on weeks worked for disabled workers? Does the parallel trends assumption seem to hold?[3]

5. Now add year×age, year×race, year×schooling and year×region dummies. Intuitively, what do these interactions control for? Output another coefficient plot. Do the controls change the results compared to the simple specification?

---

[1]TA: Petter Berg, petter.berg@phdstudent.hhs.se

[2]The coefficients on these interactions show the dynamic treatment effect, i.e. the difference between disabled and non-disabled people in each year. The hypothesis in the paper is that this difference increased after the introduction of ADA.

[3]We have a slightly different sample than Acemoglu and Angrist (2001), so don't expect the same results as in the paper. Also, they drop outliers in a quite particular way. See this as an exercise somewhat divorced from the results of the paper.

Part 1b: New developments in the DiD literature [OPTIONAL]

Recent years have seen a huge surge in the econometric literature on DiD.[4] In particular, the standard two-way fixed effects estimator that has been widely used for DiD designs where treatment is staggered has been shown to be very problematic when treatment effects are heterogeneous across groups and time. For this exercise, you are given a simulated panel data set ($PS5\_simulated.dta$) with three countries $\{1, 2, 3\}$ and 10 time periods from 2011 through 2020. Country 1 is never treated, 2 is treated in 2014 and country 3 is treated in 2017. You have access to the potential outcomes under no treatment, $Y_{it}(0)$, and under treatment, $Y_{it}(1)$, as well as the "actual, observed outcome" Y_{it} (which is just given by the potential outcome under whatever treatment status a country has in a given time, plus some random noise).

1. Plot the time series of $Y_{it}(0)$ and $Y_{it}(1)$ against the time variable. Does the parallel trends assumption hold? Describe the potential outcome paths under treatment for all three countries (even the one that is, in fact, not treated). Are the treatment effects dynamic, heterogeneous or both? Also, plot the "observed" outcome Y_{it} against time.

2. What is the true Average Treatment Effect on the Treated (ATT)? *Hint: the individual treatment effect for a given country/time period cell is simply the difference between the potential outcomes. The ATT is the average of these over all treated cells. Don't compute this by hand!*

3. Now we pretend that we only observe Y_{it} and not the underlying potential outcomes. Estimate the ATT, δ, by two-way fixed effects (TWFE):

$$Y_{it} = \lambda_t + \alpha_i + \delta Treatment_{it} + u_{it}$$

   where i denotes country and t denotes year. Is $\hat{\delta}^{TWFE}$ close to the true ATT?[5]

4. Exclude country 3 so that we have no heterogeneity (there is only one treated unit!). Compute the ATT and estimate by TWFE again. Do you get closer?

5. There are many suggested estimators that address this. One is Callaway and Sant'Anna (2021) and can be run with the `csdid` package (or the function *att_gt* in the R package *did*). Estimate $\hat{\delta}^{CS}$ again with all three countries. Is it close to the true ATT? *Hint: use the option* `method(reg)` *(*`est_method = "reg"` *in R). To output the estimated ATT, also specify* `agg(simple)` *(*`aggte(model, type="simple")` *where* `model` *is the model created by* `att_gt` *in R).*

Part 2: Interpreting published results

In Kleven et al. (2019), the authors investigate the differential impact of having children on labor market outcomes for men and women (i.e. fathers and mothers) respectively. To do so, they run event-study regressions with the event of interest being defined as the year t in which the man/woman had its first child. They estimate the following baseline regression *separately* for both genders g:

$$Y^g_{ist} = \pi + \sum_{j \neq -1} \alpha^g_j \cdot \mathbb{1}[j = t] + \sum_k \beta^g_k \cdot \mathbb{1}[k = age_{is}] + \sum_y \gamma^g_y \cdot \mathbb{1}[y = s] + u^g_{ist}$$

[4] A good starting point is the set of slides by Scott Cunningham (the guy behind the Mixtape causal inference book) available here `https://www.dropbox.com/s/l2vmb8rujumsqjj/did_talk.pdf?dl=0`

[5] Okay, I won't leave you speculating on what close means – let's say within 0.5 of the true value.

where i refers to an individual, s refers to an actual year and t refers to the event-time, i.e. the number of years *until* or *after* the individual's first child-birth. Hence, the coefficients of interest are α_j^g for time periods $j \geq 0$, which measures the effect on labor market outcomes in time periods *after* child birth. The reference period is the year just before child birth, i.e. $t = -1$. The intercept is given by π.

1. The coefficients β_k^g are dummies for the age of an individual i at time s. Given that the authors are investigating the impact on labor market outcomes, such as earnings, for men and women of having their first child, why is it important to control for age?

2. Let's assume that childbirth is correlated with the business cycle in such a way that when the economy is doing poorly, more women decide to have their first child. Further, assume that the timing of the first child for men is uncorrelated with the business cycle. Would this be a problem for the specification given above? Explain why or why not.

Figure 1: The main result of the paper, for those who are interested!

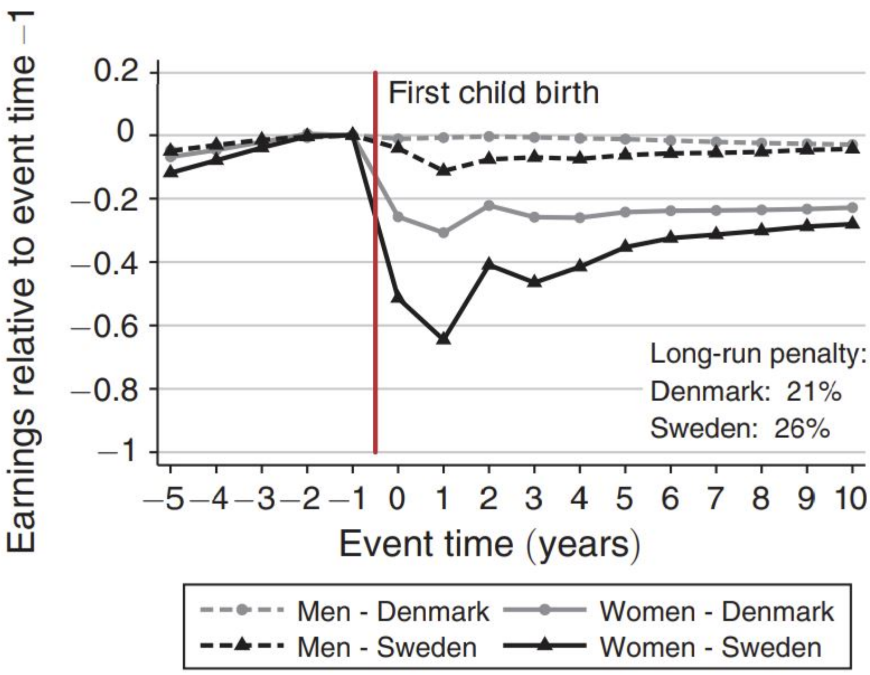

FIGURE 1. CHILD PENALTIES IN EARNINGS IN SCANDINAVIAN COUNTRIES

References

ACEMOGLU, D. AND J. ANGRIST (2001): "Consequences of Employment Protection: The Case of the Americans with Disability Act," *Journal of Political Economy*, 109.

CALLAWAY, B. AND P. H. SANT'ANNA (2021): "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 225, 200–230, themed Issue: Treatment Effect 1.

KLEVEN, H., C. LANDAIS, J. POSCH, A. STEINHAUER, AND J. ZWEIMULLER (2019): "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, 109, 122–26.