

# Financial Econometrics – Cross Section & Panel Data

Swiss Finance Institute PhD Program, Fall 2024 – Andreas Fuster, EPFL

## Problem Set 2 – “Solutions”

2. Replication of regressions for personal income growth.
  - a. See columns (1) and (2) in the table below. Results are close to the original ones from the paper, although not exactly identical.
  - b. The result is robust in the sense that the coefficients are still strongly statistically significant, but the magnitude reduces a bit – see columns (3) and (4). A justification for dropping the year of the reform is that it is not known (at least to us) when in the year the deregulation happened, so the treatment status of such a state-year is ambiguous. It is of course also unlikely that there is an immediate effect of deregulation on income growth.
  - c. I would argue that clustering at the state level is necessary, in order to account for within-state serial correlation of errors. This is what is implemented in columns (5)-(8) below – makes little difference to standard errors. One could argue that there should be two-way clustering by state and year, but then one has to deal with the fact that there are only 22 years in the sample, and potentially use wild bootstrap. There’s not a strong case for two-way clustering here though in my opinion.
  - d. See columns (7) and (8). Coefficient estimates are reduced a bit, but the standard errors also fall (since we have more observations) and so the coefficient estimates are strongly statistically significant.
  - e. So far, the results look quite robust to changes in the specification.

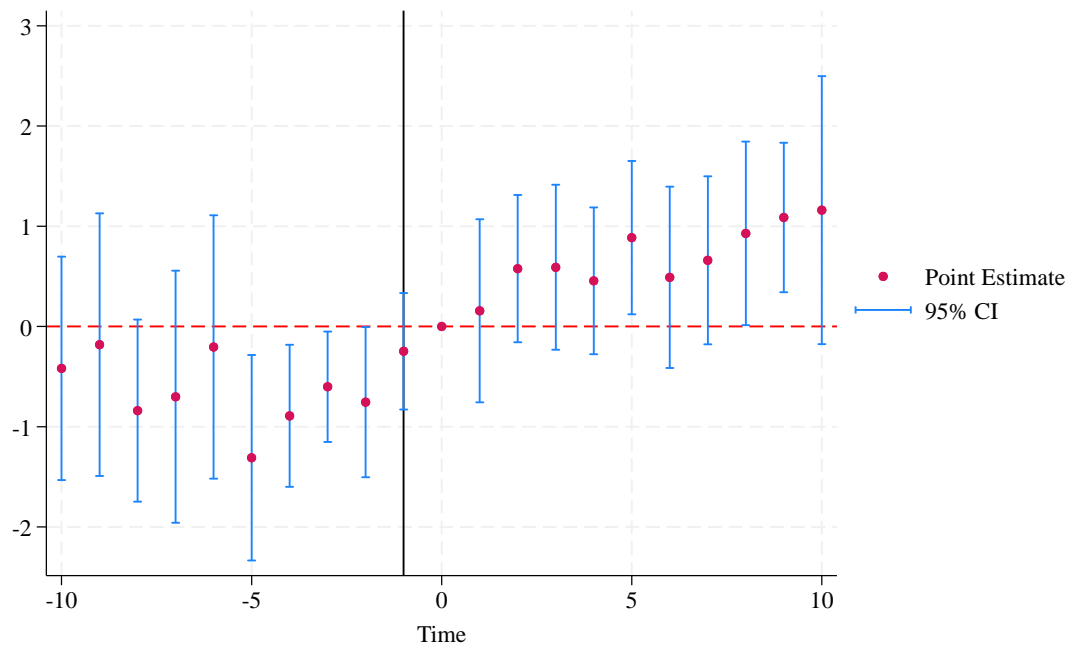
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2a Basic	2a Regional	2b Basic	2b Regional	2c Basic	2c Regional	2d Basic	2d Regional
After deregulation	0.940***	0.505**	0.816***	0.379*	0.940***	0.505*	0.761***	0.441**
	(0.251)	(0.232)	(0.229)	(0.211)	(0.289)	(0.290)	(0.210)	(0.192)
Constant	1.246***	1.481***	1.282***	1.523***	1.246***	1.481***	1.218***	1.453***
	(0.133)	(0.131)	(0.129)	(0.126)	(0.140)	(0.138)	(0.128)	(0.116)
Nr. obs.	1015	974	1050	1008	1015	974	1363	1308
Adj. R2	0.477	0.612	0.478	0.612	0.477	0.611	0.474	0.605
Mean (dep. var.)	1.70	1.72	1.69	1.71	1.70	1.72	1.68	1.72
SD (dep. var.)	3.50	3.46	3.46	3.42	3.50	3.46	3.13	3.09

Standard errors in parentheses (heterosked.-robust in columns 1 to 4; clustered by state in columns 5 to 8)

\* p<0.1, \*\* p<0.05, \*\*\* p<0.01

3. For the event study graph, my intuition is that it is “cleanest” to drop the states that had already deregulated at the beginning of the period (since we don’t know when exactly the deregulation happened, we can’t assign “years relative to treatment” to them) and also the states that did not deregulate by the end (since they may have deregulated right after the end of the sample period). But this actually doesn’t seem to matter all that much for the results. Also note that for such a plot, we would not want to remove the year of the deregulation from the sample – but we could make that one the “omitted category” (baseline year), though could alternatively also use the previous year (-1). Following the advice in Baker et al, I keep all yearly coefficients for the estimation, without caps/bins, but I only plot from -10 to +10. This is done using the “eventdd”

package in Stata, though I also did it manually and got a similar chart. The resulting chart looks as follows:



Based on this chart, we would likely be worried about pretrends from year -5 to year 0.

I also experimented with the “xtevent” package of Freyaldenhoven et al., which we discussed in the lecture, but got a chart that looked somewhat different – presumably due to the binning of end points. I trust the one above more, given the discussion in Baker et al.

4. The worry is that, if treatment effects are heterogeneous or evolve dynamically with the length of treatment (which seems quite plausible), the standard TWFE estimator may lead to biased estimates of the average treatment effect. In particular, given that the early-treated and always-treated states effectively provide the control group for the late-treated states (along with the few never-treated states) this could in extreme cases even flip the sign of the coefficient.

The Bacon-Goodman decomposition in my case gave the following output (with N =1050, i.e. I kept the year of the deregulation in the sample):

```
.      bacondecomp incg d
Computing decomposition across 19 timing groups
including an always-treated group and a never-treated group
```

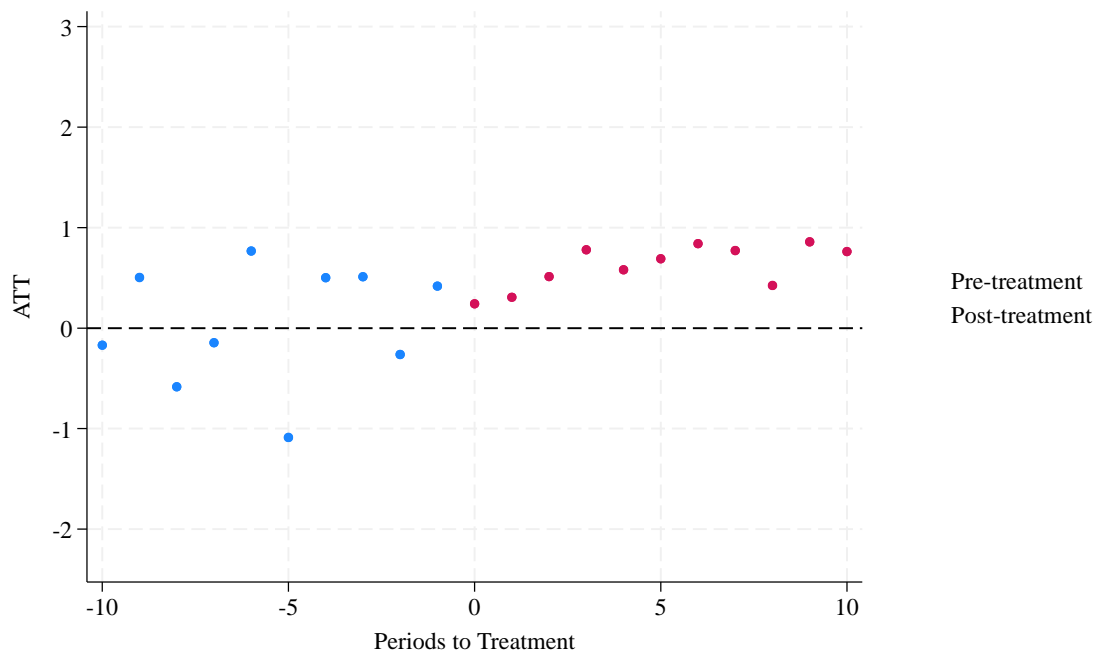
incg	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
d	.816199	.2821743	2.89	0.004	.2631476	1.36925

Bacon Decomposition

	Beta	TotalWeight
Timing_groups	1.156110989	.4656488568
Always_v_timing	.450889961	.4274809145
Never_v_timing	.7963901968	.1068702286

“Timing\_groups” are late vs. early and early vs. late – a previous version of the “bacondecomp” command had shown those separately (which I thought was helpful – not sure why they changed it). In the above, we see that 42.7% of total weight in the overall DiD estimate is actually assigned to Treated vs. Already Treated (“Always\_v\_timing”), which is potentially problematic to the extent that treatment effects evolve dynamically over time. However, we also see that the average DiD estimate for this category is actually lowest, so that this may bias the overall treatment effect estimate downward (i.e. not overstate the effect of treatment, which we would be most concerned about). The treatment estimate for the “Never\_v\_timing” group (i.e. using never-treated as control group) is close to the overall treatment effect.

5. Callaway-Sant’Anna: the ‘csdid’ package in Stata gets an estimated treatment effect of 0.44 with st. error of 0.42, so a p-value of 0.29. Thus, the effect is not statistically significant and quite a bit smaller than with the TWFE approach. Their method also allows to generate event study plots – I am showing one below. While none of the effects post-treatment are individually significant, they are all positive, and with their method there are no clear pretrends visible (although as noted in Roth (2024)<sup>1</sup>, the interpretation of these charts is not the same as for “usual” event study charts).



For the imputation estimator, I used the intuitive method proposed by [Gardner et al. \(2024\)](#), which essentially ‘imputes’ each unit’s counterfactual based on time fixed effects and unit fixed effects that are estimated only based on units while they are untreated. With that method (implemented in the [did2s](#) package) I get a treatment effect of 0.45 with st. error of 0.23 (clustered at state level), so a p-value of 0.051. Again, this is quite a bit smaller than the original, but close to statistically significant at the 5% level.

A recent paper by Zdrojewski and Butler (published in the *Critical Finance Review*, 2024) revisits JS in detail (also using additional data) and claims that their results are spurious:

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3871311](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3871311)

<sup>1</sup> <https://www.jonathandroth.com/assets/files/HetEventStudies.pdf>. Note, he recommends using the “long2” option to make event study plots that are closer to the usual ones, but that led to an error message for me.

We are using the same data as them, and the Callaway-Sant'Anna ATT estimate I get (0.44) is identical to theirs, although the confidence interval shown in their Figure 3 is slightly wider than what I get in Stata. Given that the imputation estimator yields a tighter estimate and may be more efficient in this application, the conclusions of Zdrojewski and Butler may be more negative than is warranted. Note that their conclusions are also partly based on a "falsification test" where they put per capita cigarette consumption as the dependent variable, and find a strongly significant effect using the TWFE approach. While I think there are indeed concerns here, I don't find the falsification approach very convincing, in the sense that one could often find some other dependent variable for which one would also obtain significant results "by chance" – that does not necessarily mean that the original results are "wrong."