Assignment1

Jung Jae Kim

Questions

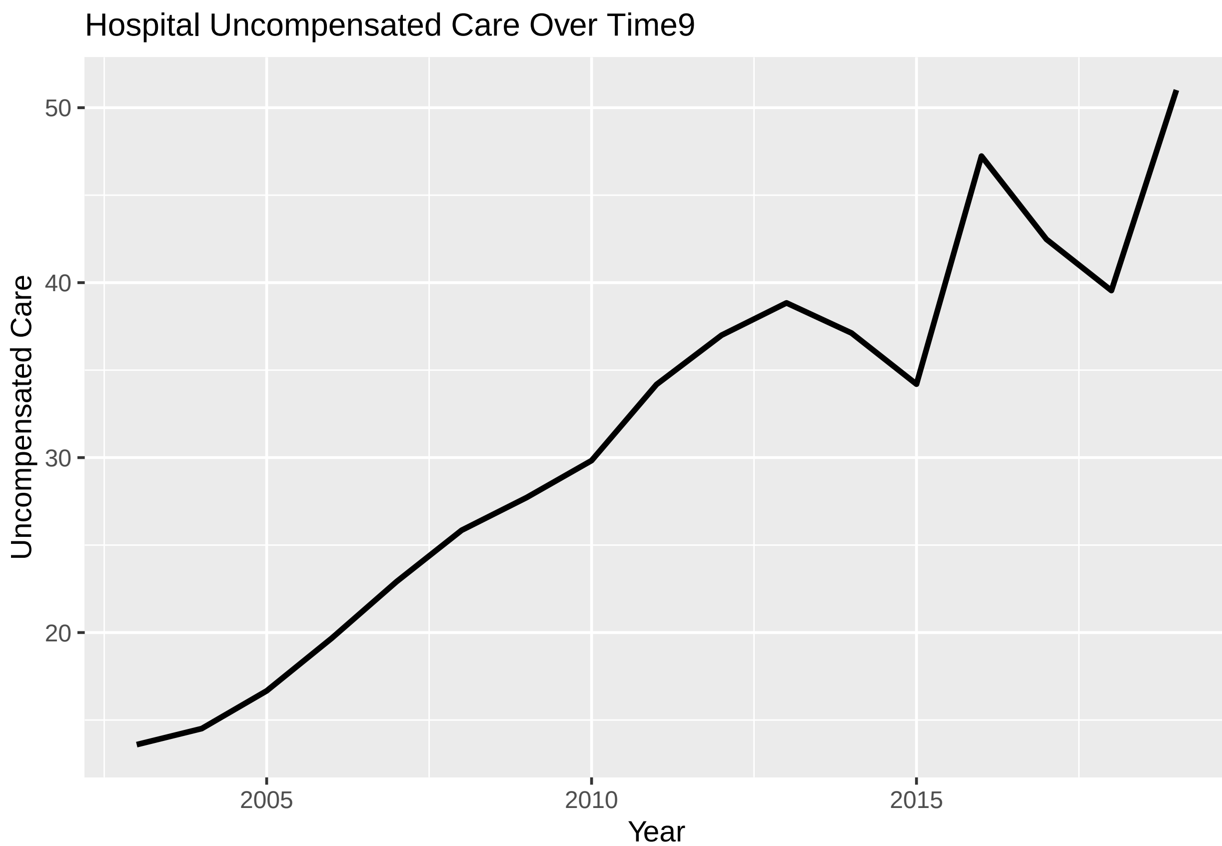
1. Provide and discuss a table of simple summary statistics showing the mean, standard deviation, min, and max of hospital total revenues and uncompensated care over time.

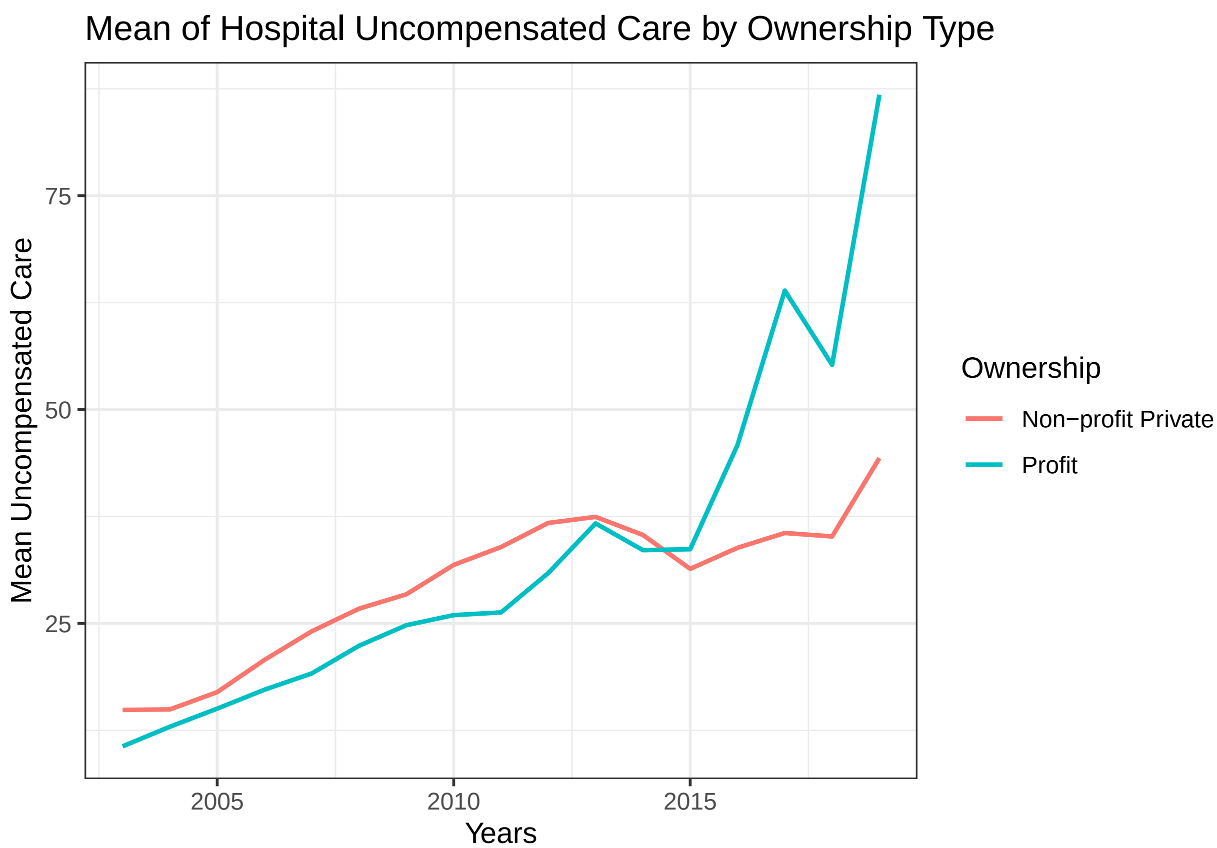
year	tot_rev_mean	tot_rev_sd	tot_rev_min	tot_rev_max	uncomp_care_mean	uncomp_care_sd	uncomp_care_min	uncomp_care_max
2003	183.7516	328.5340	-1.757898	4722.759	13.59386	32.34838	-0.128490	777.9874
2004	203.0854	368.4423	0.154394	5525.731	14.51141	32.17411	0.000001	684.0098
2005	222.4706	408.0067	0.000001	6398.554	16.66756	31.97184	0.000001	427.7000
2006	245.6002	451.4228	-0.104189	7784.095	19.67377	34.99964	0.000365	378.7916
2007	267.6509	493.6330	0.063650	8577.046	22.91398	44.85904	0.000001	736.3861
2008	292.2377	540.8699	0.000004	9293.788	25.84698	50.04529	0.003036	992.9241
2009	319.6343	595.6167	0.119236	9846.465	27.71623	48.80182	0.000010	583.9753
2010	342.5813	640.1367	0.306861	10185.416	29.83092	75.57166	0.000001	2793.9230
2011	368.1752	688.7359	-27.582223	10572.291	34.18110	77.86419	-17.191333	2057.8779
2012	390.8624	741.1380	-11.799711	11865.320	36.99496	89.15973	-1.238052	1881.0833
2013	413.2360	798.1295	0.094880	12751.708	38.84212	82.76625	-0.118659	1812.4911
2014	444.8683	860.3326	0.006624	13376.352	37.11968	92.36853	0.010376	1989.8858
2015	485.4881	930.6878	0.009368	14143.533	34.19745	91.00367	-0.529475	2037.4303
2016	525.9238	1017.6282	0.084952	15618.749	47.23335	431.83240	-0.036365	20404.4477
2017	566.1149	1121.9254	0.124513	16863.431	42.47821	108.49377	-0.027988	2746.8772
2018	613.1987	1241.4294	0.282914	18677.245	39.54799	104.75228	0.009091	2596.8667
2019	665.3059	1372.2255	0.000003	22000.932	51.01056	128.21166	-97.789186	2639.1503

In 2016, the max value of uncompensated care is $20404 mil, which is 10 times bigger than max values in other periods. This might be reporting error, so I delete this observation for future analysis.

2

. Create a figure showing the mean hospital uncompensated care from 2003 to 2019. Show this trend separately by hospital ownership type (private not for profit and private for profit).




Mean of Hospital Uncompensated Care by Ownership Type
Mean Uncompensated Care
75
50
25
2005
2010
2015
Years
Ownership
Non−profit Private
Profit

3. Using a simple DD identification strategy, estimate the effect of Medicaid expansion on hospital uncompensated care using a traditional two-way fixed effects (TWFE) estimation:

$$y_{it} = \alpha_i + \gamma_t + \delta D_{it} + \varepsilon_{it}, \tag{1}$$

where $D_{it} = 1(E_i \leq t)$ in Equation @ref(eq:dd) is an indicator set to 1 when a hospital is in a state that expanded as of year t or earlier, γ_t denotes time fixed effects, α_i denotes hospital fixed effects, and y_{it} denotes the hospital's amount of uncompensated care in year t. Present four estimates from this estimation in a table: one based on the full sample (regardless of treatment timing); one when limiting to the 2014 treatment group (with never treated as the control group); one when limiting to the 2015 treatment group (with never treated as the control group); and one when limiting to the 2016 treatment group (with never treated as the control group). Briefly explain any differences.

	Model 1	Model 2	Model 3	Model 4
treatment	-31.387***	-38.359***	-38.089***	-39.504***
	(2.815)	(3.857)	(5.075)	(4.329)
Num.Obs.	36514	27789	16300	14822
R2	0.684	0.696	0.666	0.681
R2 Adj.	0.640	0.654	0.620	0.636
AIC	388040.1	298681.8	183824.2	167062.2
BIC	425940.4	326367.3	198983.4	181015.3
RMSE	43.49	46.25	60.28	59.91
Std.Errors	by: provider	by: provider	by: provider	by: provider

Note: ^^ + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Note: ^^ (1)-(4) representes full, 2014 group, 2015 group and 2016 group respectevely

4. Estimate an "event study" version of the specification in part 3:

$$y_{it} = \alpha_i + \gamma_t + \sum_{\tau<-1} D_{it}^{\tau}\delta_\tau + \sum_{\tau>=0} D_{it}^{\tau}\delta_\tau + \varepsilon_{it}, \tag{2}$$

where $D_{it}^{\tau} = 1(t - E_i = \tau)$ in Equation @ref(eq:event) is essentially an interaction between the treatment dummy and a relative time dummy. In this notation and context, τ denotes years relative to Medicaid expansion, so that $\tau = -1$ denotes the year before a state expanded Medicaid, $\tau = 0$ denotes the year of expansion, etc. Estimate with two different samples: one based on the full sample and one based only on those that expanded in 2014 (with never treated as the control group).

	Model 1	Model 2
relative_year = -5 × treated_ever	15.731***	15.259***
	(3.200)	(3.899)
relative_year = -4 × treated_ever	6.176*	8.692*
	(2.410)	(3.798)
relative_year = -3 × treated_ever	4.754***	8.312**
	(1.375)	(2.794)
relative_year = -2 × treated_ever	1.227	1.986
	(0.951)	(2.227)
relative_year = 0 × treated_ever	-9.143***	-11.571***
	(1.143)	(2.380)
relative_year = 1 × treated_ever	-18.248***	-16.890***
	(1.614)	(2.225)
relative_year = 2 × treated_ever	-27.695***	-28.346***
	(2.000)	(2.730)
relative_year = 3 × treated_ever	-35.154***	-37.836***
	(2.592)	(3.274)
relative_year = 4 × treated_ever	-38.225***	-39.463***
	(3.236)	(3.872)
relative_year = 5 × treated_ever	-47.131***	-53.887***
	(4.332)	(5.218)
Num.Obs.	36514	27789
AIC	378771.8	291732.5
BIC	378865.4	291823.1
RMSE	43.27	46.05
Std.Errors	by: provider	by: provider
FE: provider	X	X
FE: year	X	X

Note: ^^ + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5. Sun and Abraham (SA) show that the δ_τ coefficients in Equation @ref(eq:event) can be written as a non-convex average of all other group-time specific average treatment effects. They propose an interaction weighted specification:

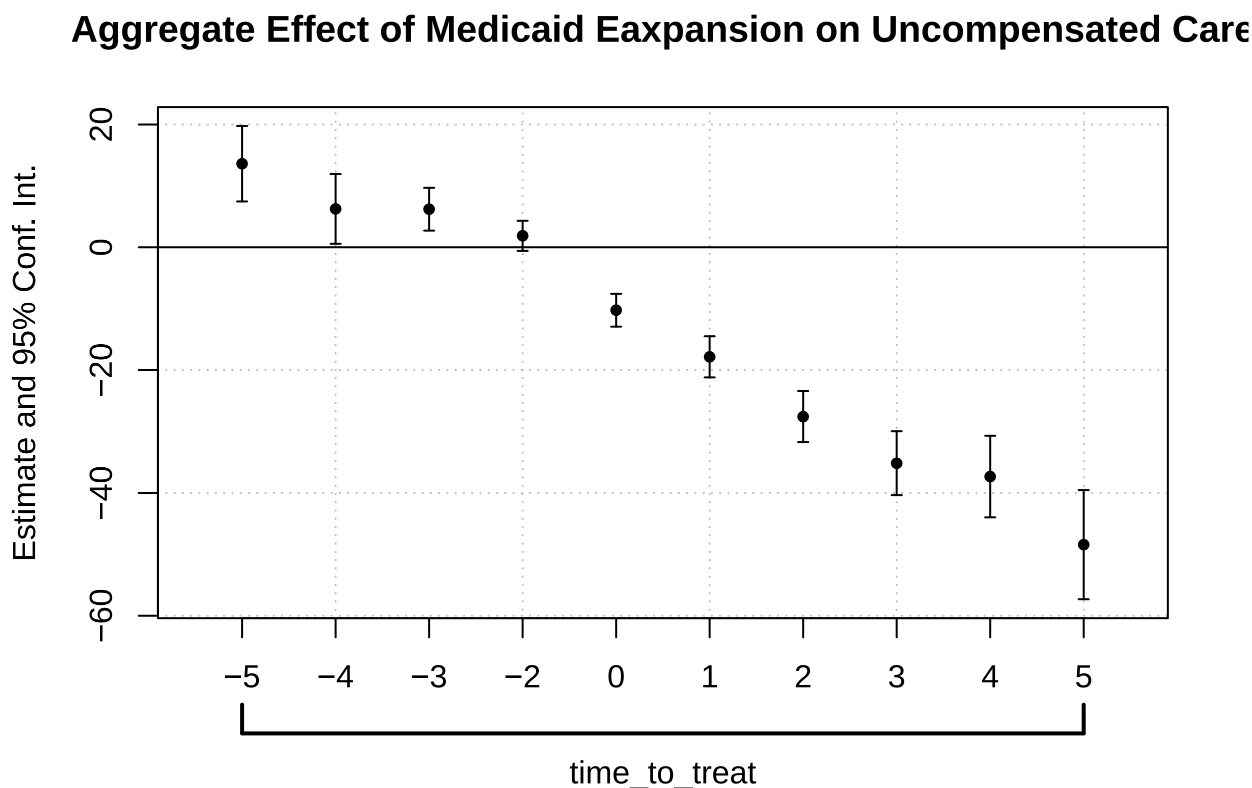
$$y_{it} = \alpha_i + \gamma_t + \sum_{e} \sum_{\tau \neq -1} \left(D_{it}^{\tau} \times 1(E_i = e)\right) \delta_{e,\tau} + \varepsilon_{it}. \tag{3}$$

Re-estimate your event study using the SA specification in Equation (3). Show your results for $\hat{\delta}_{e,\tau}$ in a Table, focusing on states with $E_i = 2014$, $E_i = 2015$, and $E_i = 2016$.

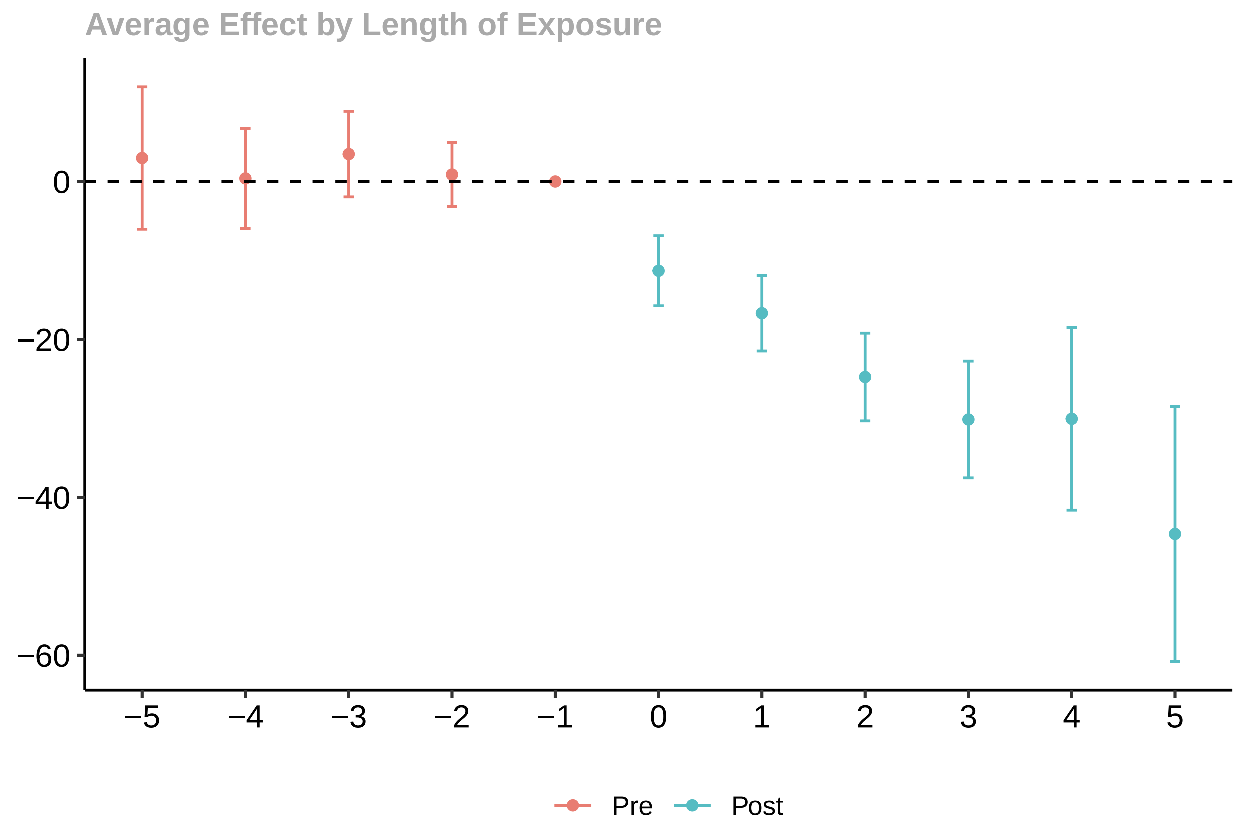
	E = 14	E = 15	E = 16
Dependent Var.:	uncomp_care	uncomp_care	uncomp_care
time_to_treat = -5	11.56*** (3.189)	17.34*** (4.464)	10.26. (5.767)
time_to_treat = -4	6.265. (3.364)	3.479 (4.017)	-2.906 (4.035)
time_to_treat = -3	6.855** (2.112)	5.394* (2.094)	-3.301 (3.566)
time_to_treat = -2	1.628 (1.656)	3.013 (2.360)	-2.386 (2.553)
time_to_treat = 0	-11.17*** (1.843)	-1.575 (2.580)	-11.51*** (2.589)
time_to_treat = 1	-16.81*** (1.926)	-15.99*** (3.337)	-34.81*** (4.593)
time_to_treat = 2	-27.07*** (2.331)	-27.15*** (3.776)	-36.59*** (4.829)
time_to_treat = 3	-34.79*** (2.745)	-31.35*** (4.743)	-51.83*** (5.550)
time_to_treat = 4	-36.54*** (3.350)	-41.95*** (5.794)	NA
time_to_treat = 5	-48.43*** (4.535)	NA	NA
Fixed-Effects:	————	————	————
provider	Yes	Yes	Yes
year	Yes	Yes	Yes
————————	————————	————————	————————
S.E.: Clustered	by: provider	by: provider	by: provider
Observations	36,514	36,514	36,514
R2	0.68704	0.68704	0.68704
Within R2	0.03264	0.03264	0.03264

6

. Present an event study graph based on the results in part 5. Hint: you can do this automatically in R with the fixest package (using the sunab syntax for interactions), or with eventstudyinteract in Stata. These packages help to avoid mistakes compared to doing the tables/figures manually and also help to get the standard errors correct.



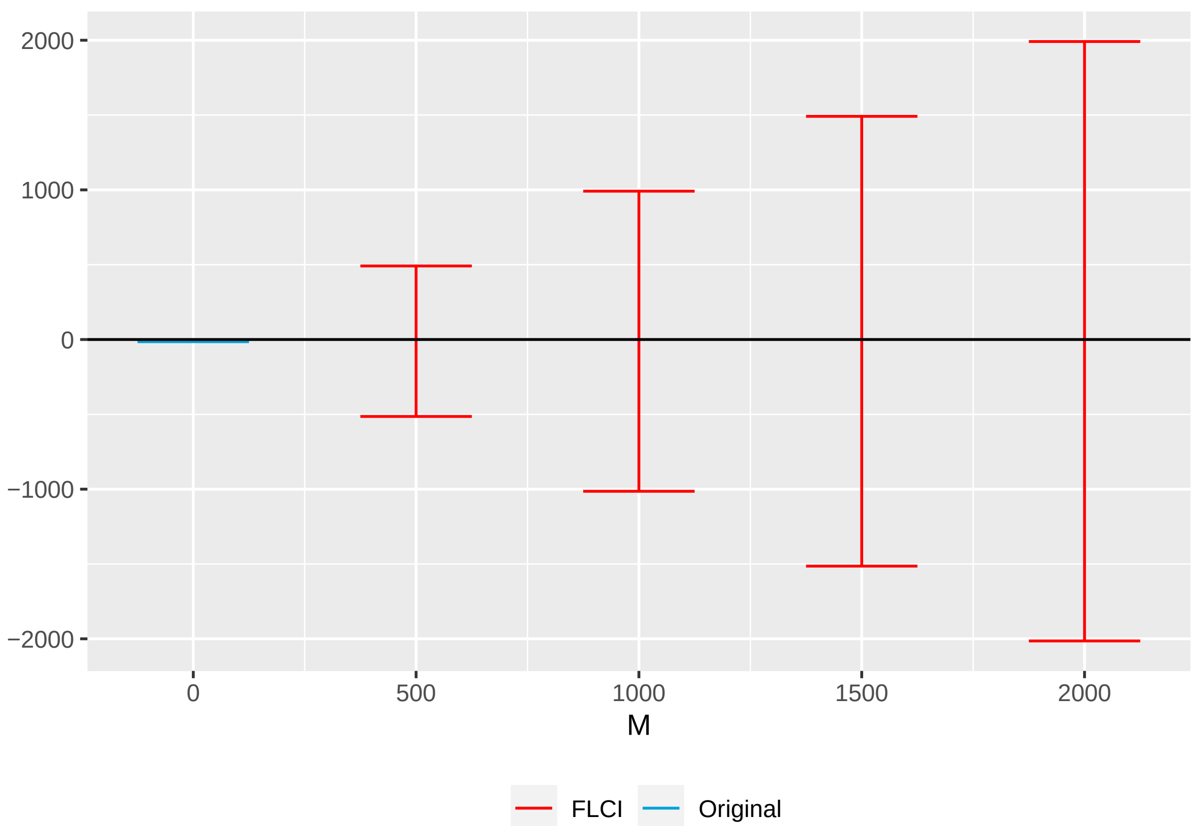
7. Callaway and Sant'Anna (CS) offer a non-parametric solution that effectively calculates a set of group-time specific differences, $ATT(g,t) = E[y_{it}(g) - y_{it}(\infty)|G_i = g]$, where g reflects treatment timing and t denotes time. They show that under the standard DD assumptions of parallel trends and no anticipation, $ATT(g,t) = E[y_{it} - y_{i,g-1}|G_i = g] - E[y_{it} - y_{i,g-1}|G_i = \infty]$, so that $\hat{ATT}(g,t)$ is directly estimable from sample analogs. CS also propose aggregations of $\hat{ATT}(g,t)$ to form an overall ATT or a time-specific ATT (e.g., ATTs for τ periods before/after treatment). With this framework in mind, provide an alternative event study using the CS estimator. Hint: check out the `did` package in `R` or the `csdid` package in `Stata`.



8. Rambachan and Roth (RR) show that traditional tests of parallel pre-trends may be underpowered, and they provide an alternative estimator that essentially bounds the treatment effects by the size of an assumed violation in parallel trends. One such bound RR propose is to limit the post-treatment violation of parallel trends to be no worse than some multiple of the pre-treatment violation of parallel trends. Assuming linear trends, such a relative violation is reflected by

$$\Delta(\bar{M}) = \left\{ \delta : \forall t \geq 0, |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq \bar{M} \times \max_{s<0} |(\delta_{s+1} - \delta_s) - (\delta_s - \delta_{s-1})| \right\}.$$

The authors also propose a similar approach with what they call "smoothness restrictions," in which violations in trends changes no more than M between periods. The only difference is that one restriction is imposed relative to observed trends, and one restriction is imposed using specific values. Using the `HonestDiD` package in `R` or `Stata`, present a sensitivity plot of your CS ATT estimates using smoothness restrictions, with assumed violations of size $M \in \{500, 1000, 1500, 2000\}$. Check out the GitHub repo here for some help in combining the `HonestDiD` package with CS estimates. Note that you'll need to edit the function in that repo in order to use pre-specified smoothness restrictions. You can do that by simply adding `Mvec=Mvec` in the `createSensitivityResults` function for `type=smoothness`.



9. Discuss your findings and compare estimates from different estimators (e.g., are your results sensitive to different specifications or estimators? Are your results sensitive to violation of parallel trends assumptions?).

Answers: Simple DiD estimation result indicate that the treatment effect is significant and robust to different treatment timing. The treatment effect is still robust when estimating event study. Sun and Abraham (SA) and Callaway and Sant'Anna (CS) also shows stable treatment effects. Rambachan and Roth (RR) method indicates that the results are robust to parallel trend assumption.

10. Reflect on this assignment. What did you find most challenging? What did you find most surprising?

Answers: Understanding and cleaning datasets are most challenging. I was surprised that there are multiple methods to do DiD and sensitivity analysis. This assignmet is really helpful to understand DiD deeper.