Group-Time ATTs

Sam Harper

2021-11-02

Standard Two-Way Fixed Effects

First, let's read in the Stata dataset, and create a smaller version with just the variables of interest for looking at skilled birth attendance:

Using this data, let's now run the standard TWFE analysis that is the workhorse of most DD applications with differential treatment timing.

```
## NOTE: 2,327 observations removed because of NA values (LHS: 2,327, Fixed-effects: 2,312).
```

```
## OLS estimation, Dep. Var.: sba_birth
## Observations: 5,555
## Fixed-effects: district: 8,  time: 5
## Standard-errors: Clustered (district)
##       Estimate Std. Error t value Pr(>|t|)
## txdel  0.05976   0.037325 1.60105   0.1534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.425631     Adj. R2: 0.057514
##                  Within R2: 5.313e-4
```

Specifying this using an OLS model with standard errors clustered at the district level, we get a DD estimate that we would interpret as showing that the intervention increased the probability of skilled birth attendance by 6 percentage points (95% CI -0.7 to 12.7).

Because the treatment is assigned at the cluster level and we aren't adjusting for any covariates, we can also analyze this at the cluster level. Next let's aggregate up to the district level. Here is a look at the dataset:

```
## `summarise()` has grouped output by 'district', 'dist_id'. You can override using the `.groups` argu
```

You can see that we are aggregating up to the district level, and how we have the total population (`tpop`) and the total number of births with a skilled attendant (`tsba`), as well as the proportion by district.

district	sba_birth	txdel	time	dist_id	group	g2	g3	g4
Kaliua DC	0	0	1	1	2	1	0	0
Kaliua DC	1	1	5	1	2	1	0	0
Kaliua DC	0	1	5	1	2	1	0	0
Kaliua DC	0	0	1	1	2	1	0	0
Kaliua DC	1	1	5	1	2	1	0	0
Kaliua DC	1	1	5	1	2	1	0	0

district	$\text{dist}_i d$	time	tsba	tpop	psba	txdel	group
Kaliua DC	1	1	409	705	0.580	0	2
Kaliua DC	1	2	49	73	0.671	1	2
Kaliua DC	1	3	63	77	0.818	1	2
Kaliua DC	1	4	9	10	0.900	1	2
Kaliua DC	1	5	94	114	0.825	1	2
Nzega DC	2	1	465	634	0.733	0	3
Nzega DC	2	2	60	76	0.789	0	3
Nzega DC	2	3	43	57	0.754	1	3
Nzega DC	2	4	9	10	0.900	1	3
Nzega DC	2	5	88	101	0.871	1	3
Nzega TC	3	1	165	182	0.907	0	5
Nzega TC	3	2	13	13	1.000	0	5
Nzega TC	3	3	12	14	0.857	0	5
Nzega TC	3	4	4	5	0.800	0	5
Nzega TC	3	5	21	24	0.875	1	5
Sikonge DC	4	1	310	374	0.829	0	5
Sikonge DC	4	2	49	53	0.925	0	5
Sikonge DC	4	3	23	27	0.852	0	5
Sikonge DC	4	4	4	5	0.800	0	5
Sikonge DC	4	5	44	52	0.846	1	5
Tabora MC	5	1	540	614	0.879	0	4
Tabora MC	5	2	53	59	0.898	0	4
Tabora MC	5	3	39	44	0.886	0	4
Tabora MC	5	4	6	6	1.000	1	4
Tabora MC	5	5	82	86	0.953	1	4
Urambo DC	6	1	158	193	0.819	0	2
Urambo DC	6	2	15	17	0.882	1	2
Urambo DC	6	3	20	25	0.800	1	2
Urambo DC	6	4	2	3	0.667	1	2
Urambo DC	6	5	25	30	0.833	1	2
Uyui DC	7	1	338	514	0.658	0	4
Uyui DC	7	2	60	81	0.741	0	4
Uyui DC	7	3	29	57	0.509	0	4
Uyui DC	7	4	6	10	0.600	1	4
Uyui DC	7	5	63	81	0.778	1	4
Igunga DC	8	1	522	821	0.636	0	3
Igunga DC	8	2	67	89	0.753	0	3
Igunga DC	8	3	40	68	0.588	1	3
Igunga DC	8	4	15	20	0.750	1	3
Igunga DC	8	5	94	131	0.718	1	3

	Individual	Aggregate
txdel	0.060	0.060
	(0.037)	(0.040)
Num.Obs.	5555	40
Std.Errors	by: district	by: district
FE: district	X	X
FE: time	X	X

Now, we can also fit the TWFE model to this aggregate data, controlling for district fixed effects and survey wave.

```
## OLS estimation, Dep. Var.: psba
## Observations: 40
## Fixed-effects: district: 8,  time: 5
## Standard-errors: Clustered (district)
##       Estimate Std. Error t value Pr(>|t|)
## txdel  0.05976   0.039958 1.49557  0.17842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.466034     Adj. R2: 0.826702
##                  Within R2: 0.058009
```

The estimates from these two models are identical (SEs are very minimally different), which is just to say that it is fine to work with the aggregate data to estimate the impact of the intervention (absent control for covariates):

Callaway/Sant'Anna Approach

Okay so we have the traditional DD estimate of around a 6 percentage point increase in SBA. Now let's look at the Callaway-Sant'Anna DD estimate. Since the CS DD approach functions by aggregating different kinds of group-time DDs, we can also aggregate up to the group level, since we have 2 groups being treated at each wave post-baseline. Let's define a new variable `group` that indicates the time at which each group was *first* treated. So for Kaliua DC and Urambo DC that were the first groups to be treated post-baseline (what we will call time=2), they are assigned a value of `group=2`, and so on for the other groups. Meanwhile, we are still aggregating up the number of births with SBA and the total population of each group.

Now, just to get a look at what we are comparing when we run the group-time DDs, let's reshape this data to wide (just to see it):

Okay, now let's estimate the impact using the CS DD method:

```
# Use not-yet-treated as comparison group
atts_cs <- did::att_gt(yname = "psba", # name of the LHS variable
                       tname = "time", # name of the time variable
                       idname = "dist_id", # name of the id variable
                       gname = "group", # name of the first treatment period
                       data = d_sba, # name of the data
                       xformla = NULL,
                       weightsname = "tpop",
                       est_method = "reg", # estimation method.
                       control_group = "notyettreated", # set the control group
                       bstrap = TRUE, # if TRUE compute boostrapped SE
```

group	time	tsba	tpop	psba	txdel
2	1	567	898	0.631	0
2	2	64	90	0.711	1
2	3	83	102	0.814	1
2	4	11	13	0.846	1
2	5	119	144	0.826	1
3	1	987	1455	0.678	0
3	2	127	165	0.770	0
3	3	83	125	0.664	1
3	4	24	30	0.800	1
3	5	182	232	0.784	1
4	1	878	1128	0.778	0
4	2	113	140	0.807	0
4	3	68	101	0.673	0
4	4	12	16	0.750	1
4	5	145	167	0.868	1
5	1	475	556	0.854	0
5	2	62	66	0.939	0
5	3	35	41	0.854	0
5	4	8	10	0.800	0
5	5	65	76	0.855	1

group	psba_t1	psba_t2	psba_t3	psba_t4	psba_t5	tpop_t1	tpop_t2	tpop_t3	tpop_t4	tpop_t5
2	0.631	0.711	0.814	0.846	0.826	898	90	102	13	144
3	0.678	0.770	0.664	0.800	0.784	1455	165	125	30	232
4	0.778	0.807	0.673	0.750	0.868	1128	140	101	16	167
5	0.854	0.939	0.854	0.800	0.855	556	66	41	10	76

term	group	time	estimate	std.error	conf.low	conf.high
ATT(2,2)	2	2	0.011	0.090	-0.203	0.226
ATT(2,3)	2	3	0.260	0.143	-0.078	0.599
ATT(2,4)	2	4	0.269	0.072	0.098	0.440
ATT(3,2)	3	2	0.045	0.086	-0.158	0.249
ATT(3,3)	3	3	0.018	0.142	-0.319	0.356
ATT(3,4)	3	4	0.170	0.062	0.022	0.318
ATT(4,2)	4	2	-0.062	0.117	-0.339	0.214
ATT(4,3)	4	3	-0.048	0.149	-0.403	0.306
ATT(4,4)	4	4	0.130	0.275	-0.522	0.783

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

```
                         biters = 1000, # number of boostrap interations
                         print_details = FALSE, # if TRUE, print detailed results
                         panel = FALSE) # panel or repeated cross-sectional
summary(atts_cs)
```

A summary of the estimates, SEs, and 95% CIs (based on 1000 bootstrapped replications) is below:

One thing that is immediately different from the TWFE method is that we get an entire suite of ATTs for each group and time period, which allows us to see:

- heterogeneity across both groups treated at different times (e.g., the ATT in the first year of treatment is much larger for the group first treated at time=2 (ATT(2,2) = 0.01) relative to the estimated effect in the first year of treatment for the group first treated at time 4, i.e., ATT(4,4) = 0.13, though both estimates are imprecise); and
- as well as different times for each treatment group, e.g.., the impact at time 3 for the group first treated at time 2 is 0.26 relative to 0.01 for the first year of treatment.

Let's take a quick detour to understand where these estimates are coming from.

Group Time ATTs

The core of the CS estimator is the group-time ATT. They define groups based on when they were first treated and

Group 2

The basic idea of the group-time ATTs is to estimate a series of ATTs for each group G that is treated at time T. So if we wanted to estimate the ATT at time=2 for the group that is first treated at time=2, we calculate the 'long' difference, i.e., post minus pre for the treated group ($G = 2$) and the difference for all groups *not already treated* for the same period. Thus, the groups we are comparing to estimate ATT(2,2) are:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

If we take the population-weighted average SBA proportions for these two groups, we get

Now, we can ask about how the effect in Group 2 changes with time after the intervention. The ATT(2,3) asks about the estimated treatment effect at time 3 ($t = 3$) for the group that was first treated at time 2 ($g = 2$). To get this estimate, we now create a similar 2x2 table but are using time 3 as the 'post' estimate. But note here that, since group 3 ($G = 3$) is treated at time 3, we don't want to include it as a part of our control group, we that means we restrict our control comparison to only those groups that are **not** treated by time 3. So we are comparing:

And we get:

We can of course extend our view of how the treatment effect evolves for Group 2 by calculating the effect of being treated at time 4 for the group that was first treated at time 2, i.e., ATT(2,4). This is comparing:

Note that the control group here also changes, since at $t = 4$ group 4 has now been treated, so we exclude them from the control group for this comparison. And the estimate is:

Group 3

Now, what about the groups that are first treated at time 3, i.e., ($t = 3$)? Since we actually have more than one pre-period for this group, we can also see whether there is some evidence of non-parallel trends by looking at, for example, the ATT(3,2), which is the effect of being treated at **time=2** for the group that is first treated at time 3. In essence, we are comparing the pre-intervention 'long difference' between $t = 1$ and $t = 2$ for the group eventually treated at time 3 with the same long difference among the controls:

The estimate of the ATT(3,2) is:

Now for the treatment effect at time 3 for the groups first treated at time 3 we are comparing:

The estimate of the ATT(3,3) is:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

The groups for comparison for the estimated treatment effect at time 4 for the groups first treated at time 3 now also need to exclude group 4, since it has received the treatment at time 4, but note also that the 'pre-intervention' period is still ($t = 2$) since group 3 was actually treated at time 3. So, we care comparing:

The estimate of the ATT(3,4) is:

Group 4

Since we have untreated groups remaining at time 5, the last group we can estimate a 'clean' ATT for is Group 4. In this case, we also now have 2 pre-intervention periods we can use to assess the parallel trends assumption. For the time 2 periods before treatment starts for Group 4, we can compare the change in SBA between time 1 and time 2 for our treated group, i.e., the group first treated at time 4, and a control group. In this case the control group will only include Group 3 and Group 5, since Group 2 is treated at ($t = 2$) and we have to exclude it. So our comparison groups are:

Again, taking the population-weighted estimate for the control group and the estimates for Group 4, we can calculate the ATT at time 2 for the group first treated at time 4, i.e., ATT(4,2):

As you might expect, we now make the same progression for the addition ATTs for the the group first treated at time 4. First, for the next ATT for the pre-intervention period we want the treatment effect at time 3 for the group that is first treated at time 4. That means comparing only Groups 4 and 5 (since group 3 is treated at time 3, it has to be excluded from the control population):

And the resulting ATT(4,3) is:

Finally, the last ATT we can estimate is the effect of being treated at time 4 for the group that is first treated at time 4. Again, we can only compare groups 4 and 5 here, since every other group has already been treated. So we are comparing:

And the resulting ATT(4,3) is:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

Weighted ATTs