Untitled

Sam Harper

29/10/2021

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:

```
d <- read_dta(here("data-clean", "hh_female_impact_rpt.dta")) %>%
    mutate(district = as_factor(hh1, levels = "labels"),
         time = as.numeric(delwave) + 1,
         sba_birth = as.numeric(sba_birth))

# restrict to variables for analysis
d_ind <- d %>%
  select(district, sba_birth, txdel, time) %>%
  mutate(dist_id = as.numeric(district)) %>%
  drop_na() %>%
  group_by(district, dist_id) %>%
  mutate(group = min(if_else(txdel==1, time, 5)),
         g2 = if_else(group==2, 1, 0),
         g3 = if_else(group==3, 1, 0),
         g4 = if_else(group==4, 1, 0))
```

Next let's aggregate up to the district level

```
d_sba <- d_ind %>%
  group_by(district, dist_id, time) %>%
  summarise(tsba = sum(sba_birth),
         tpop = n(),
         psba = tsba / tpop, # % SBA
         txdel = mean(txdel),
         group = mean(group),
         g2 = mean(g2),
         g3 = mean(g3),
         g4 = mean(g4))
```

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

```
glimpse(d_sba)
```

```
## Rows: 40
## Columns: 11
## Groups: district, dist_id [8]
## $ district <fct> Kaliua DC, Kaliua DC, Kaliua DC, Kaliua DC, Kaliua DC, Nzega ~
```

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

```
## $ dist_id  <dbl> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5~
## $ time     <dbl> 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1~
## $ tsba     <dbl> 409, 49, 63, 9, 94, 465, 60, 43, 9, 88, 165, 13, 12, 4, 21, 3~
## $ tpop     <int> 705, 73, 77, 10, 114, 634, 76, 57, 10, 101, 182, 13, 14, 5, 2~
## $ psba     <dbl> 0.5801418, 0.6712329, 0.8181818, 0.9000000, 0.8245614, 0.7334~
## $ txdel    <dbl> 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0~
## $ group    <dbl> 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4~
## $ g2       <dbl> 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ g3       <dbl> 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ g4       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1~
```

Since the CS DiD 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:

```
# create group-time aggregate data
d_agg <- d_sba %>% group_by(group, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop, # % SBA
            txdel = mean(txdel),
            g2 = mean(g2),
            g3 = mean(g3),
            g4 = mean(g4))
```

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):

```
gt <- d_agg %>% select(group, time, psba) %>%
  pivot_wider(names_from = time, names_prefix = "time",
              values_from = psba)

gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling()
```

Group Time ATTS

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

g22	time1	time2	Long diff	ATT_2_2
0	0.745	0.814	0.069	**NA**
1	0.631	0.711	0.080	**0.011**

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(2:3, color = "black", background = "red") %>%
  row_spec(2:4, background  = "lightgray") %>%
  column_spec(4:6, background = "white") %>%
  column_spec(1, background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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

```
d_agg %>% filter(time < 3) %>%
  mutate(g22 = if_else(group==2,1,0)) %>%
  group_by(g22, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g22, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time2` - `time1`) %>%
  group_by() %>%
  mutate(ATT_2_2 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

```
## `summarise()` has grouped output by 'g22'. You can override using the `.groups` argument.
```

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:

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

g23	time1	time3	Long diff	ATT_2_3
0	0.803	0.725	-0.078	**NA**
1	0.631	0.814	0.182	**0.26**

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(2,4), color = "black", background = "red") %>%
  row_spec(2, background = "white") %>%
  row_spec(3:4, background  = "lightgray") %>%
  column_spec(c(3,5,6), background = "white") %>%
  column_spec(1, background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

And we get:

```
d_agg %>% filter((time==1 | time==3) & group!=3) %>%
  mutate(g23 = if_else(group==2,1,0)) %>%
  group_by(g23, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g23, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time3` - `time1`) %>%
  group_by() %>%
  mutate(ATT_2_3 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

```
## `summarise()` has grouped output by 'g23'. You can override using the `.groups` argument.
```

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:

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(2,5), color = "black", background = "red") %>%
  row_spec(2:3, background = "white") %>%
```

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

g24	time1	time4	Long diff	ATT_2_4
0	0.854	0.800	-0.054	**NA**
1	0.631	0.846	0.215	**0.269**

```
  row_spec(4, background  = "lightgray") %>%
  column_spec(6, background = "white") %>%
  column_spec(c(1,3,4), background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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:

```
d_agg %>% filter((time==1 | time==4) & group!=3 & group!=4) %>%
  mutate(g24 = if_else(group==2,1,0)) %>%
  group_by(g24, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g24, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time4` - `time1`) %>%
  group_by() %>%
  mutate(ATT_2_4 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

```
## `summarise()` has grouped output by 'g24'. You can override using the `.groups` argument.
```

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:

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
```

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

g32	time1	time2	Long diff	ATT_3_2
0	0.803	0.85	0.046	**NA**
1	0.678	0.77	0.091	**0.045**

```
  column_spec(2:3, color = "black", background = "red") %>%
  row_spec(3:4, background  = "lightgray") %>%
  row_spec(1, background = "white") %>%
  column_spec(4:6, background = "white") %>%
  column_spec(1, background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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

```
d_agg %>% filter(time < 3 & group != 2) %>%
  mutate(g32 = if_else(group==3,1,0)) %>%
  group_by(g32, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g32, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time2` - `time1`) %>%
  group_by() %>%
  mutate(ATT_3_2 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

```
## `summarise()` has grouped output by 'g32'. You can override using the `.groups` argument.
```

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

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(3:4, color = "black", background = "red") %>%
  row_spec(3:4, background  = "lightgray") %>%
  row_spec(1, background = "white") %>%
  column_spec(c(1,2,5,6), background = "white") %>%
  column_spec(1, background = "white") %>%
  kable_minimal() %>%
  footnote(general = "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

g33	time2	time3	Long diff	ATT_3_3
0	0.85	0.725	-0.124	**NA**
1	0.77	0.664	-0.106	**0.018**

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

```
d_agg %>% filter(time >1 & time < 4 & group != 2) %>%
  mutate(g33 = if_else(group==3,1,0)) %>%
  group_by(g33, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g33, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time3` - `time2`) %>%
  group_by() %>%
  mutate(ATT_3_3 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

```
## `summarise()` has grouped output by 'g33'. You can override using the `.groups` argument.
```

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:

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(3,5), color = "black", background = "red") %>%
  row_spec(c(1,3), background = "white") %>%
  row_spec(4, background  = "lightgray") %>%
  column_spec(c(1,2,4,6), background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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

```
d_agg %>% filter((time == 2 | time == 4) & (group == 3 | group == 5)) %>%
  mutate(g34 = if_else(group==3,1,0)) %>%
  group_by(g34, time) %>%
```

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

g34	time2	time4	Long diff	ATT_3_4
0	0.939	0.8	-0.139	**NA**
1	0.770	0.8	0.030	**0.17**

```
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g34, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time4` - `time2`) %>%
  group_by() %>%
  mutate(ATT_3_4 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

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:

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(2,3), color = "black", background = "red") %>%
  row_spec(1, background = "white") %>%
  row_spec(c(2,4), background = "lightgray") %>%
  column_spec(c(1,4,5,6), background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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):

```
d_agg %>% filter(time < 3 & group >= 3) %>%
  mutate(g42 = if_else(group==4,1,0)) %>%
  group_by(g42, time) %>%
```

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

g42	time1	time2	Long diff	ATT_4_2
0	0.727	0.818	0.091	**NA**
1	0.778	0.807	0.029	**-0.062**

```
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g42, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time2` - `time1`) %>%
  group_by() %>%
  mutate(ATT_4_2 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

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):

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(3,4), color = "black", background = "red") %>%
  row_spec(1:2, background = "white") %>%
  row_spec(4, background = "lightgray") %>%
  column_spec(c(1,2,5,6), background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

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

g43	time2	time3	Long diff	ATT_4_3
0	0.939	0.854	-0.086	**NA**
1	0.807	0.673	-0.134	**-0.048**

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

```
d_agg %>% filter((time ==2 | time == 3) & group >= 4) %>%
  mutate(g43 = if_else(group==4,1,0)) %>%
  group_by(g43, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g43, time, psba) %>%
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time3` - `time2`) %>%
  group_by() %>%
  mutate(ATT_4_3 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

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:

```
gt %>%
  kbl(digits=3, escape = FALSE) %>%
  kable_styling() %>%
  column_spec(c(4,5), color = "black", background = "red") %>%
  row_spec(1:2, background = "white") %>%
  row_spec(4, background = "lightgray") %>%
  column_spec(c(1,2,3,6), background = "white") %>%
  kable_minimal() %>%
  footnote(general = "Red = treated, Gray = untreated")
```

And the resulting ATT(4,3) is:

```
d_agg %>% filter((time ==3 | time == 4) & group >= 4) %>%
  mutate(g44 = if_else(group==4,1,0)) %>%
  group_by(g44, time) %>%
  summarise(tsba = sum(tsba),
            tpop = sum(tpop),
            psba = tsba / tpop) %>%
  select(g44, time, psba) %>%
```

g44	time3	time4	Long diff	ATT_4_4
0	0.854	0.80	-0.054	**NA**
1	0.673	0.75	0.077	**0.13**

```
  pivot_wider(names_from = time, values_from = psba,
              names_prefix = "time") %>%
  mutate(`Long diff` = `time4` - `time3`) %>%
  group_by() %>%
  mutate(ATT_4_4 = `Long diff` - lag(`Long diff`, default = NA)) %>%
  kable(digits=3) %>%
  column_spec(5, bold=TRUE) %>%
  kable_minimal()
```

Weighted ATTs