Social distancing, partisanship and SIPOs

Alec MacMillen

6/5/2020

1. Topic and motivation

Building on the work of Allcott et al (2020), I'm interested in exploring whether partisan differences in social distancing behavior persist even as states and counties start to transition out of shelter-in-place orders. The answer to this question has significant policy implications: voluntary social distancing even in absence of a formal SIPO could have positive public health benefits in continuing to hamper the spread of COVID-19 without a vaccine, while also dampening any potential economic impact of reopening society. I'll take Allcott et al's general framework for parsing partisan differences in mobility patterns and relate it directly to the start and end dates of SIPOs, to investigate whether the divergent trends they identify hold even on the back-end of shutdown policies.

2. Read in built data

Data gathering, cleaning and merging process stored at this repo: https://github.com/alecmacmillen/covid-research (data cleaning scripts in 'src' subfolder). Essentially, this built data has Safegraph social distancing metrics aggregated to the county-week level (devices leaving home, median time spent in and out of dwelling location, and the share of devices measures described in the footnote to Allcott Appendix A8); Trump vote share as the main independent variable, along with controls for demographics, public health (most significantly COVID-19 cases and deaths), weather, and policy responses (SIPOs).

```
data <- read_csv("../../../data/processed/allcott/social-distancing/allcott_social_distancing_build.csv
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   state = col_character(),
##   county = col_character(),
##   county_fips = col_character(),
##   week_start_date = col_date(format = "")
## )
```

```
## See spec(...) for full column specifications.
```

```
data %>% select(1:13) %>% head(10)
```

```
## # A tibble: 10 x 13
##    state county county_fips week_start_date device_count completely_home~
##    <chr> <chr>  <chr>       <date>                 <dbl>            <dbl>
##  1 alab~ autau~ 01001       2020-01-27             36156             6550
##  2 alab~ autau~ 01001       2020-02-03             39908             7742
##  3 alab~ autau~ 01001       2020-02-10             39326             7158
##  4 alab~ autau~ 01001       2020-02-17             38603             7160
##  5 alab~ autau~ 01001       2020-02-24             39264             7902
##  6 alab~ autau~ 01001       2020-03-02             41638             7229
```

```
##  7 alab~ autau~ 01001       2020-03-09            37727            6559
##  8 alab~ autau~ 01001       2020-03-16            36005            9968
##  9 alab~ autau~ 01001       2020-03-23            35287            9642
## 10 alab~ autau~ 01001       2020-03-30            37918           12498
## # ... with 7 more variables: devices_leaving_home <dbl>,
## #   log_devices_leaving_home <dbl>, median_home_dwell_time <dbl>,
## #   log_median_home_dwell_time <dbl>, median_non_home_dwell_time <dbl>,
## #   log_median_non_home_dwell_time <dbl>, share_of_devices <dbl>
```

```r
data %>% select(14:26) %>% head(10)
```

```
## # A tibble: 10 x 13
##    share_of_candid~ share_adj_attri~ share_asian share_black share_white
##               <dbl>            <dbl>       <dbl>       <dbl>       <dbl>
##  1            0.893            0.642     0.00804       0.185       0.766
##  2            0.884            0.674     0.00804       0.185       0.766
##  3            0.892            0.672     0.00804       0.185       0.766
##  4            0.893            0.662     0.00804       0.185       0.766
##  5            0.882            0.663     0.00804       0.185       0.766
##  6            0.892            0.692     0.00804       0.185       0.766
##  7            0.901            0.655     0.00804       0.185       0.766
##  8            0.855            0.596     0.00804       0.185       0.766
##  9            0.868            0.599     0.00804       0.185       0.766
## 10            0.815            0.600     0.00804       0.185       0.766
## # ... with 8 more variables: share_bachelors <dbl>, share_undergrad <dbl>,
## #   share_hhinc_100k <dbl>, share_poverty <dbl>, share_msta <dbl>,
## #   share_services <dbl>, share_sales <dbl>, share_resources <dbl>
```

```r
data %>% select(27:39) %>% head(10)
```

```
## # A tibble: 10 x 13
##    share_public_tr~ mean_precip mean_tmin mean_tmax sip_days
##               <dbl>       <dbl>     <dbl>     <dbl>    <dbl>
##  1         0.000783       0.696      31.1      60.0        0
##  2         0.000783       9.52       44.7      66.4        0
##  3         0.000783      13.2        44.6      65.9        0
##  4         0.000783      11.6        41.2      59.3        0
##  5         0.000783       2.64       35.9      61.4        0
##  6         0.000783      19.5        43.9      64.7        0
##  7         0.000783       0.877      54.5      76.4        0
##  8         0.000783       1.79       56.6      78.0        0
##  9         0.000783       0.332      59.2      82.3        0
## 10         0.000783       1.23       47.0      75.2        2
## # ... with 8 more variables: weekly_sipo_binary <dbl>,
## #   share_over_65 <dbl>, log_pop_density <dbl>, log_cases <dbl>,
## #   log_deaths <dbl>, trump_vote_share <dbl>, county_republican <dbl>,
## #   republican_vtsh_ntile <dbl>
```

3. Exploratory data analysis to replicate Allcott et al's findings

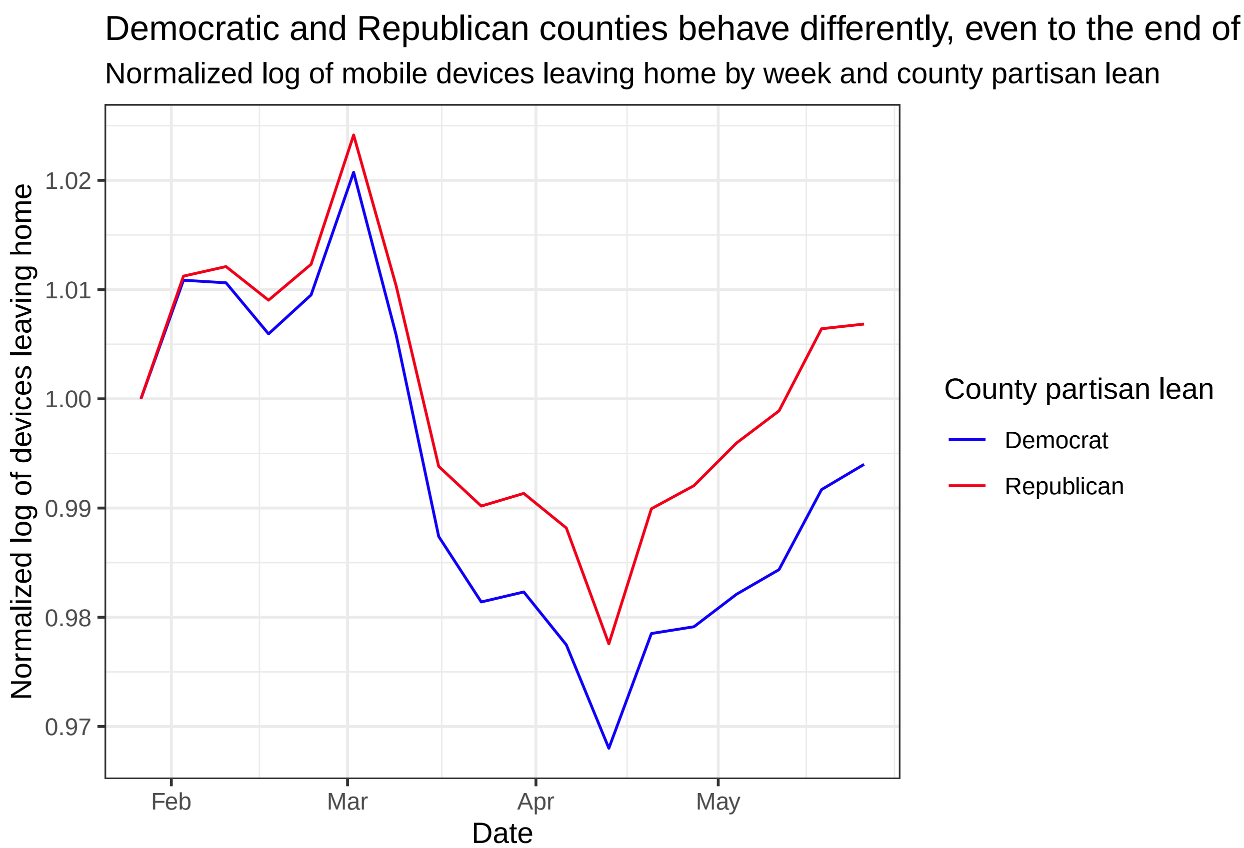
For a quick visual reference validating the data build process, let's look at raw by-county partisanship differences in Safegraph social distancing measures. **Note: Allcott et al use POI visits as their main**

specification of the dependent variable. I'm still running the data build code for that data, since it's huge: 50+ GB. This is one of the major additions I'll make to this analysis for next week, Friday 6/12 (see "Next Steps" section at the end for a fuller summary).

Essentially, we want confirmation that partisan differences in social distancing behavior holds in this replicated dataset. I'll use Safegraph's social distancing measures (Allcott's alternative specification that they report in Appendix A8) as a proxy for now. The following code examines two variables: log devices leaving home (log of raw count of devices that leave their calculated geographic home area) and log median time spent at home, both for the county-week level. I set the measures for the week of 1/27/20 as the baseline values and calculate future weeks relative to that baseline.

3A. Log of mobile devices leaving home

```
graphs1 <- data %>%
  group_by(county_fips) %>%

  # Calculate baselines for dependent vars
  mutate(baseline_ldlh = ifelse(week_start_date == as.Date("2020-01-27", "%Y-%m-%d"),
                                log_devices_leaving_home, NA),
         baseline_lmhdt = ifelse(week_start_date == as.Date("2020-01-27", "%Y-%m-%d"),
                                 log_median_home_dwell_time, NA)) %>%
  fill(baseline_ldlh, baseline_lmhdt) %>%

  # Calculate relative measure for future weeks
  mutate(normalized_ldlh = log_devices_leaving_home / baseline_ldlh,
         normalized_lmhdt = log_median_home_dwell_time / baseline_lmhdt) %>%
  ungroup() %>%
  group_by(week_start_date, county_republican) %>%

  # Calculate mean of relative measures by county partisanship
  summarize(mean_ldlh = mean(normalized_ldlh, na.rm=TRUE),
            mean_lmhdt = mean(normalized_lmhdt, na.rm=TRUE)) %>%
  mutate(county_republican = as.factor(county_republican))

# Custom color palette for viz
party.colors <- c("#1000f7", "#f20019")

# Plot weekly log of devices leaving home by week and county partisan lean
ggplot(graphs1, aes(x=week_start_date, y=mean_ldlh, color=county_republican)) +
  geom_line() +
  scale_color_manual(values=party.colors, labels=c("Democrat", "Republican")) +
  theme_bw() +
  guides(color=guide_legend(title="County partisan lean")) +
  labs(x="Date",
       y="Normalized log of devices leaving home",
       title="Democratic and Republican counties behave differently, even to the end of May",
       subtitle="Normalized log of mobile devices leaving home by week and county partisan lean")
```

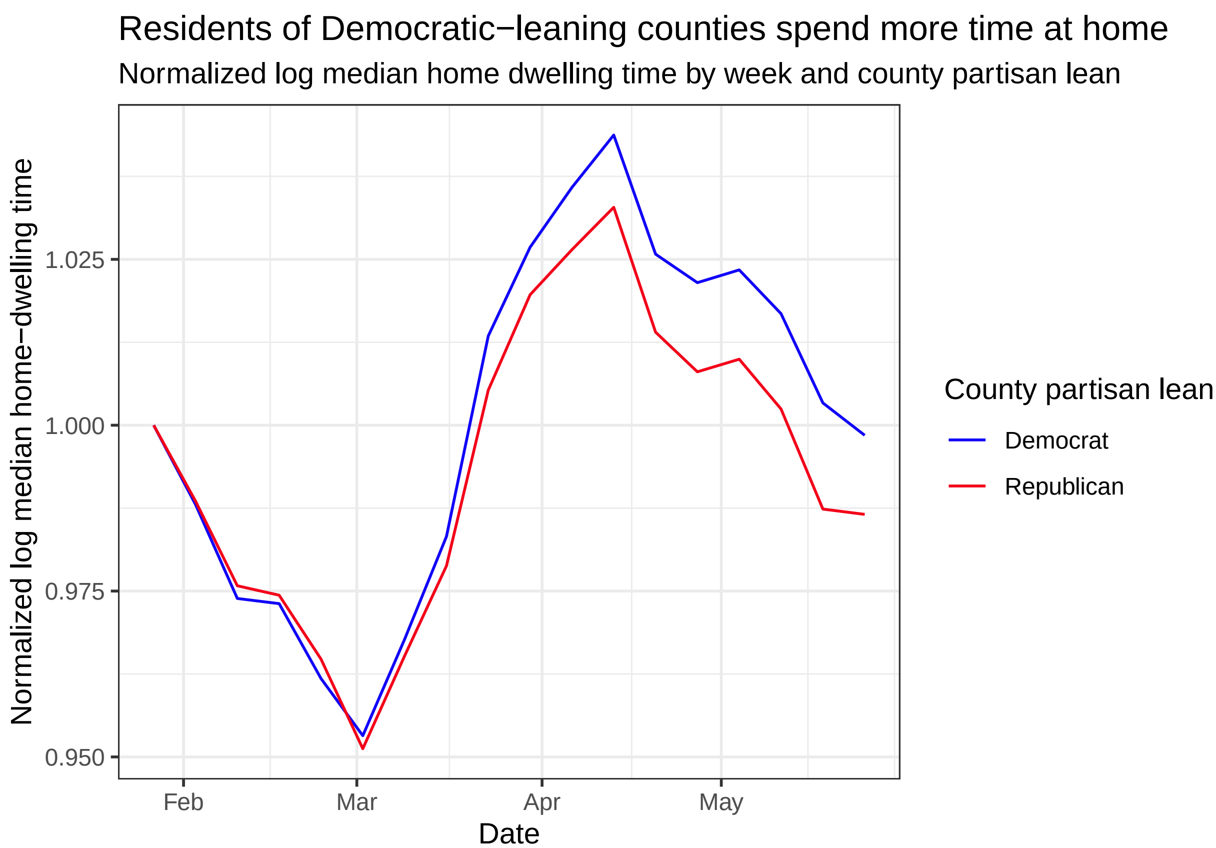


From this figure we see that the trends Allcott et al identified hold in this replication dataset as well: by mid-March, the number of devices leaving home areas in Democratic and Republican-leaning counties diverged sharply, and this difference persists even to the end of May.

As another sanity check, we can make a similar plot using log median time at home dwelling as the dependent variable.

3B. Log of median time spent at home

```
ggplot(graphs1, aes(x=week_start_date, y=mean_lmhdt, color=county_republican)) +
  geom_line() +
  scale_color_manual(values=party.colors, labels=c("Democrat", "Republican")) +
  theme_bw() +
  guides(color=guide_legend(title="County partisan lean")) +
  labs(x="Date",
       y="Normalized log median home-dwelling time",
       title="Residents of Democratic-leaning counties spend more time at home",
       subtitle="Normalized log median home dwelling time by week and county partisan lean")
```



Again, we see that Democratic counties spent *more* time at home starting by mid-March, and this disparity again persists into the end of May. Of course, these visual checks use absolute dates without reference to variation in SIPO orders (or other types of variation).

4. Regression model

4A. Data prep

To examine the impact of SIPOs, I create two new variables in the dataset: *weeks_since_start*, which measures the number of weeks since the beginning of a SIPO (with the week of the SIPO's enactment labeled 0), and *weeks_since_end*, which measures the number of weeks since the end of a SIPO. Counties where a SIPO was never in effect are dropped.

```
# First, identify county_fips where there ever was a SIPO
ever.sipo <- data %>%
  group_by(county_fips) %>%
  summarize(sipo_binary = max(weekly_sipo_binary, na.rm=TRUE)) %>%
  ungroup() %>%
  filter(sipo_binary > 0)

# Generate new dataframe with the two weeks_since variables
data.weeks.since <- data %>%
  # Drop counties that never had a SIPO
  filter(county_fips %in% ever.sipo$county_fips) %>%
  group_by(county_fips) %>%
```

```r
  # Generate weeks_since variables using the weekly_sipo_binary variable that signals whether there was
  # at any time in the county-week observation
  mutate(sip_start_week_relative = ifelse(weekly_sipo_binary == 1 & dplyr::lag(weekly_sipo_binary) == 0
         sip_end_week_relative = ifelse(weekly_sipo_binary == 1 & dplyr::lead(weekly_sipo_binary) == 0,
         weeknum = row_number(),
         sip_start_week_sub = ifelse(sip_start_week_relative == 0, weeknum, NA),
         sip_end_week_sub = ifelse(sip_end_week_relative == 0, weeknum, NA)) %>%
  fill(sip_start_week_sub, sip_end_week_sub) %>%
  fill(sip_start_week_sub, sip_end_week_sub, .direction="up") %>%
  ungroup() %>%
  mutate(weeks_since_start = as.factor(weeknum - sip_start_week_sub),
         weeks_since_end = as.factor(weeknum - sip_end_week_sub)) %>%

  # Clean up unnecessary vars
  select(-c(sip_start_week_relative, sip_end_week_relative, weeknum, sip_start_week_sub, sip_end_week_su
```

4B. Dependent variable: log_devices_leaving_home

Now we can start with a simple regression model. Here we examine the difference in *log_devices_leaving_home* by county partisanship indexing time by weeks since the SIPO *start*.

```r
# Run regression
ldlh.weeks.since.start <- plm(log_devices_leaving_home ~ trump_vote_share + weeks_since_start,
                              data = data.weeks.since,
                              index = c("county_fips", "weeks_since_start"),
                              model = "within")
```

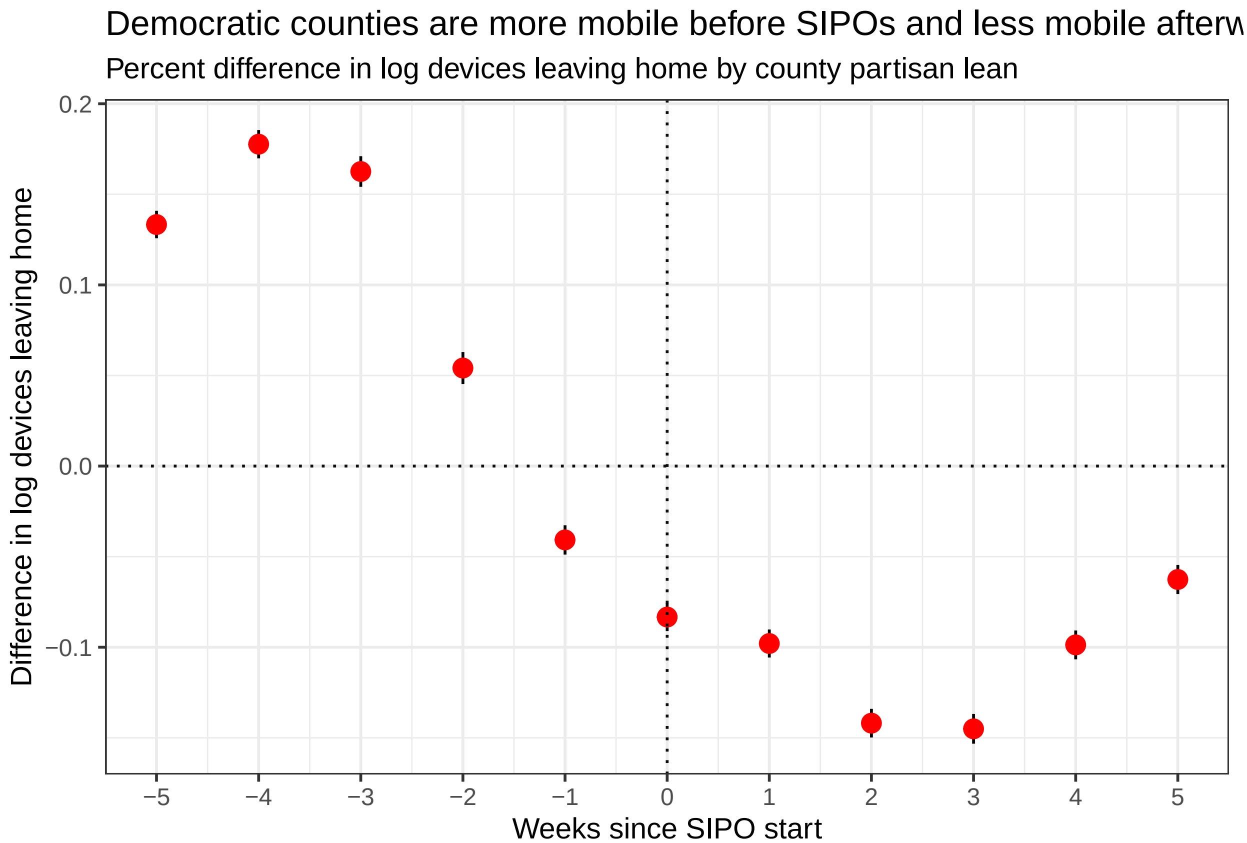
```
## at least one couple (id-time) has NA in at least one index dimensionin resulting pdata.frame
##  to find out which, use e.g.table(index(your_pdataframe), useNA = "ifany")
```

```r
# Extract coefficients
ldlh.weeks.since.start.coef <- tidy(coeftest(
  ldlh.weeks.since.start, vcov=vcovHC(ldlh.weeks.since.start, type="sss", cluster="group"))) %>%
  mutate(significance = ' ',
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.95), '*', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.975), '**', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.995), '***', significance),
         week = row_number() - 10)

# Print coefficient estimates
ldlh.weeks.since.start.coef %>%
  select(week, estimate, std.error, significance) %>%
  knitr::kable(.)
```

week	estimate	std.error	significance
-9	0.0197253	0.0038662	***
-8	0.0922171	0.0038223	***
-7	0.1269768	0.0037606	***
-6	0.1276466	0.0037564	***
-5	0.1333269	0.0037749	***
-4	0.1776781	0.0039030	***

week	estimate	std.error	significance
-3	0.1625966	0.0042257	***
-2	0.0541158	0.0044151	***
-1	-0.0407591	0.0040496	***
0	-0.0833423	0.0038659	***
1	-0.0979371	0.0038705	***
2	-0.1419081	0.0039507	***
3	-0.1450353	0.0041077	***
4	-0.0986998	0.0039800	***
5	-0.0625965	0.0040109	***
6	-0.0349922	0.0040218	***
7	0.0045771	0.0040653	
8	0.0384857	0.0039822	***
9	0.0519301	0.0041215	***
10	0.0373334	0.0051166	***

```r
# Plot estimates and confidence intervals
ldlh.weeks.since.start.coef %>%
  filter(week >= -5 & week <= 5) %>%
  ggplot(.) +
  geom_pointrange(aes(x=week, y=estimate, ymin=estimate-2*std.error, ymax=estimate+2*std.error)) +
  geom_point(aes(x=week, y=estimate), color="red", size=3) +
  scale_x_continuous(name="Weeks since SIPO start", breaks=c(-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)) +
  geom_vline(xintercept=0, linetype="dotted", alpha=0.95) +
  geom_hline(yintercept=0, linetype="dotted", alpha=0.95) +
  scale_y_continuous(name="Difference in log devices leaving home") +
  labs(title="Democratic counties are more mobile before SIPOs and less mobile afterwards",
       subtitle="Percent difference in log devices leaving home by county partisan lean")  +
  theme_bw()
```



County partisanship is measured on a 0-1 scale where 0 is most Democratic and 1 is most Republican, so my understanding is that these estimates suggest that residents of Democrat counties were more mobile than their Republican counterparts before SIPOs went into effect (perhaps anticipating that SIPOs would soon be implemented), but adjusted their behavior and became *less* mobile even before SIPOs went into place, a finding that has been reported elsewhere in the literature *(citations to come)*. These estimates also suggest that residents of Democratic counties under SIPOs adhered more closely to them than did residents of Republican counties. A few notes/questions:

- The estimates look to be very precise, which makes me wonder about how clustered standard errors should be specified.
- I also know I need to think about controlling for time trends, because it looks like there are pretty big differences between county types even long before SIPOs were put in place.
- **Any guidance on how to think about these two issues?**

Now I repeat the analysis for time since SIPO *end*.

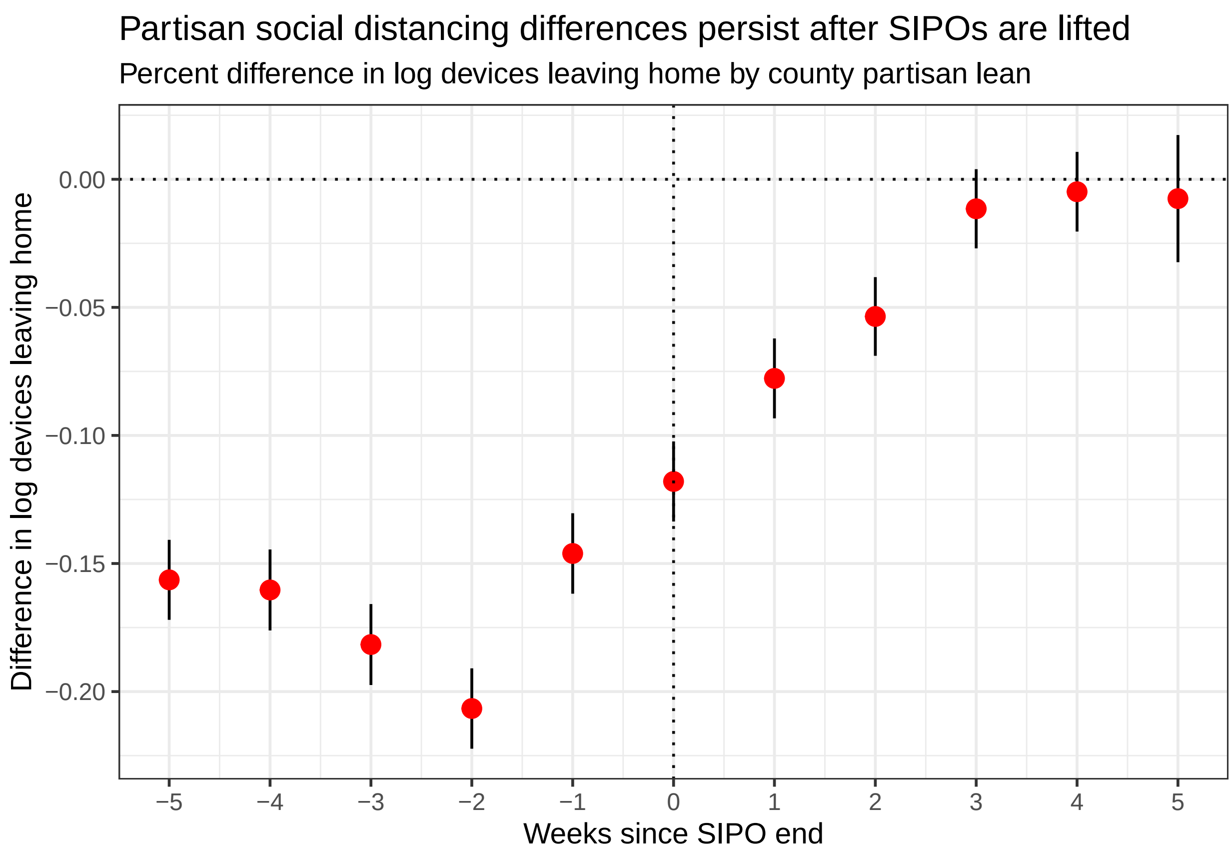
```
# Run regression
ldlh.weeks.since.end <- plm(log_devices_leaving_home ~ trump_vote_share + weeks_since_end,
                             data = data.weeks.since,
                             index = c("county_fips", "weeks_since_end"),
                             model = "within")
```

```
## at least one couple (id-time) has NA in at least one index dimensionin resulting pdata.frame
##  to find out which, use e.g.table(index(your_pdataframe), useNA = "ifany")
```

```r
# Extract coefficients
ldlh.weeks.since.end.coef <- tidy(coeftest(ldlh.weeks.since.end, vcov=vcovHC(ldlh.weeks.since.end, type=
  mutate(significance = ' ',
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.95), '*', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.975), '**', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.995), '***', significance),
         week = row_number() - 16)

# Print coefficient estimates
ldlh.weeks.since.end.coef %>%
  select(week, estimate, std.error, significance) %>%
  knitr::kable(.)
```

week	estimate	std.error	significance
-15	0.0284073	0.0066917	***
-14	0.0300855	0.0072199	***
-13	-0.0073014	0.0074633	
-12	0.0480585	0.0073093	***
-11	0.0577294	0.0072640	***
-10	0.0572676	0.0074676	***
-9	0.0633585	0.0077818	***
-8	0.0631710	0.0080263	***
-7	-0.0341332	0.0079975	***
-6	-0.1210097	0.0077971	***
-5	-0.1563775	0.0078042	***
-4	-0.1603255	0.0079047	***
-3	-0.1816309	0.0079005	***
-2	-0.2066126	0.0078465	***
-1	-0.1460789	0.0078540	***
0	-0.1179687	0.0078167	***
1	-0.0777284	0.0077975	***
2	-0.0535571	0.0076779	***
3	-0.0115049	0.0077245	
4	-0.0048580	0.0077756	
5	-0.0075410	0.0124081	

```r
# Plot estimates and confidence intervals
ldlh.weeks.since.end.coef %>%
  filter(week >= -5) %>%
  ggplot(.) +
  geom_pointrange(aes(x=week, y=estimate, ymin=estimate-2*std.error, ymax=estimate+2*std.error)) +
  geom_point(aes(x=week, y=estimate), color="red", size=3) +
  scale_x_continuous(name="Weeks since SIPO end", breaks=c(-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)) +
  geom_vline(xintercept=0, linetype="dotted", alpha=0.95) +
  geom_hline(yintercept=0, linetype="dotted", alpha=0.95) +
  scale_y_continuous(name="Difference in log devices leaving home") +
  labs(title="Partisan social distancing differences persist after SIPOs are lifted",
       subtitle="Percent difference in log devices leaving home by county partisan lean") +
  theme_bw()
```



These results look pretty much like what I expected. In the weeks leading up to a SIPO end, Democratic counties social distance more than Republican counties (i.e., residents in Democratic counties adhere more closely to the SIPO). Even in the week that the SIPO is lifted (week 0) and the few weeks after, Democratic counties social distance more - it's not until 3-4 weeks after a SIPO is lifted that the partisan difference in distancing fully attenuates. This is a pretty intriguing finding and I'll be interested to pursue it further, with more fully fleshed-out model specifications, and potentially exploring this trend at the state level.

4C. Dependent variable: log median time spent at home

I'll run through the same model with a different dependent variable, log median time spent at home by devices per county-week, as an initial robustness check.

```
# Run regression
mhdt.weeks.since.start <- plm(log_median_home_dwell_time ~ trump_vote_share + weeks_since_start,
                              data = data.weeks.since,
                              index = c("county_fips", "weeks_since_start"),
                              model = "within")
```

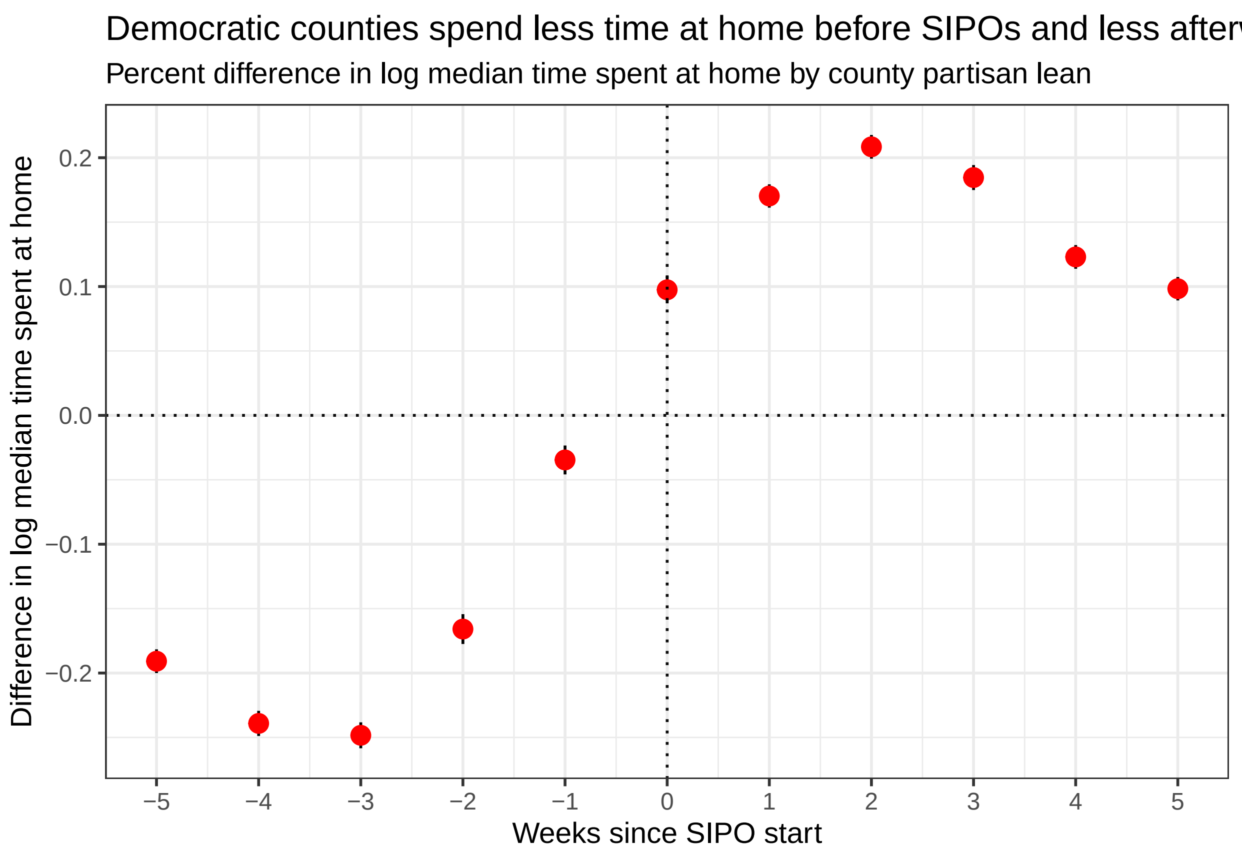
```
## at least one couple (id-time) has NA in at least one index dimensionin resulting pdata.frame
##  to find out which, use e.g.table(index(your_pdataframe), useNA = "ifany")
```

```
# Extract coefficients
mhdt.weeks.since.start.coef <- tidy(coeftest(
  mhdt.weeks.since.start, vcov=vcovHC(mhdt.weeks.since.start, type="sss", cluster="group"))) %>%
  mutate(significance = ' ',
```

```r
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.95), '*', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.975), '**', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.995), '***', significance),
         week = row_number() - 10)

# Print coefficient estimates
mhdt.weeks.since.start.coef %>%
  select(week, estimate, std.error, significance) %>%
  knitr::kable(.)
```

week	estimate	std.error	significance
-9	0.0127936	0.0045569	***
-8	-0.0200895	0.0046298	***
-7	-0.0937300	0.0045704	***
-6	-0.1417949	0.0045057	***
-5	-0.1908306	0.0046047	***
-4	-0.2391029	0.0048754	***
-3	-0.2483389	0.0050340	***
-2	-0.1658770	0.0057910	***
-1	-0.0346103	0.0056080	***
0	0.0974672	0.0052540	***
1	0.1702863	0.0045189	***
2	0.2084504	0.0045993	***
3	0.1845985	0.0048436	***
4	0.1229792	0.0045932	***
5	0.0982999	0.0045428	***
6	0.0716817	0.0047322	***
7	0.0190718	0.0048223	***
8	-0.0277602	0.0047053	***
9	-0.0537055	0.0052511	***
10	-0.0321176	0.0056857	***

```r
# Plot estimates and confidence intervals
mhdt.weeks.since.start.coef %>%
  filter(week >= -5 & week <= 5) %>%
  ggplot(.) +
  geom_pointrange(aes(x=week, y=estimate, ymin=estimate-2*std.error, ymax=estimate+2*std.error)) +
  geom_point(aes(x=week, y=estimate), color="red", size=3) +
  scale_x_continuous(name="Weeks since SIPO start", breaks=c(-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)) +
  geom_vline(xintercept=0, linetype="dotted", alpha=0.95) +
  geom_hline(yintercept=0, linetype="dotted", alpha=0.95) +
  scale_y_continuous(name="Difference in log median time spent at home") +
  labs(title="Democratic counties spend less time at home before SIPOs and less afterwards",
       subtitle="Percent difference in log median time spent at home by county partisan lean")  +
  theme_bw()
```



Roughly same findings and issues as the SIPO start analysis above - will need to think about clustering standard errors and controlling for time trends.

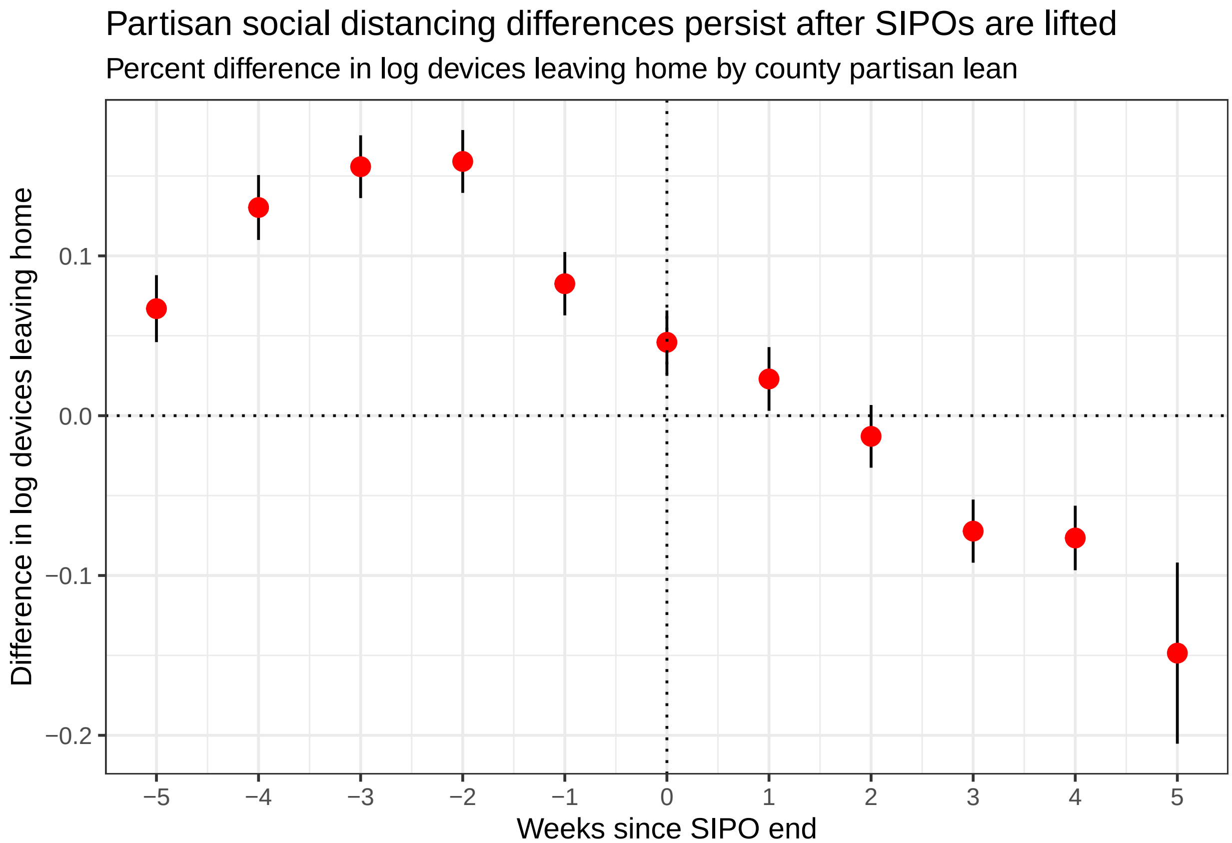
```
# Run regression
mhdt.weeks.since.end <- plm(log_median_home_dwell_time ~ trump_vote_share + weeks_since_end,
                         data = data.weeks.since,
                         index = c("county_fips", "weeks_since_end"),
                         model = "within")
```

```
## at least one couple (id-time) has NA in at least one index dimensionin resulting pdata.frame
##  to find out which, use e.g.table(index(your_pdataframe), useNA = "ifany")
```

```
# Extract coefficients
mhdt.weeks.since.end.coef <- tidy(coeftest(
  mhdt.weeks.since.end, vcov=vcovHC(mhdt.weeks.since.end, type="sss", cluster="group"))) %>%
  mutate(significance = ' ',
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.95), '*', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.975), '**', significance),
         significance = ifelse(abs(estimate/std.error) >= qnorm(0.995), '***', significance),
         week = row_number() - 16)

# Print coefficient estimates
mhdt.weeks.since.end.coef %>%
  select(week, estimate, std.error, significance) %>%
  knitr::kable(.)
```

week	estimate	std.error	significance
-15	-0.0387034	0.0099094	***
-14	-0.0804196	0.0092238	***
-13	-0.0723132	0.0094309	***
-12	-0.1189145	0.0097034	***
-11	-0.1809589	0.0098415	***
-10	-0.2022650	0.0099007	***
-9	-0.2359995	0.0100300	***
-8	-0.2219012	0.0104355	***
-7	-0.1547491	0.0104582	***
-6	-0.0611158	0.0108584	***
-5	0.0669749	0.0104693	***
-4	0.1302666	0.0101456	***
-3	0.1557959	0.0098204	***
-2	0.1590782	0.0098278	***
-1	0.0825787	0.0099040	***
0	0.0459046	0.0099806	***
1	0.0229775	0.0099664	**
2	-0.0129421	0.0098076	
3	-0.0722451	0.0098791	***
4	-0.0765291	0.0100951	***
5	-0.1485794	0.0283449	***

```r
# Plot estimates and confidence intervals
mhdt.weeks.since.end.coef %>%
  filter(week >= -5) %>%
  ggplot(.) +
  geom_pointrange(aes(x=week, y=estimate, ymin=estimate-2*std.error, ymax=estimate+2*std.error)) +
  geom_point(aes(x=week, y=estimate), color="red", size=3) +
  scale_x_continuous(name="Weeks since SIPO end", breaks=c(-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5)) +
  geom_vline(xintercept=0, linetype="dotted", alpha=0.95) +
  geom_hline(yintercept=0, linetype="dotted", alpha=0.95) +
  scale_y_continuous(name="Difference in log devices leaving home") +
  labs(title="Partisan social distancing differences persist after SIPOs are lifted",
       subtitle="Percent difference in log devices leaving home by county partisan lean") +
  theme_bw()
```



Again, the data are telling a fairly consistent story - Democratic counties spend more time at home pre-SIPO end and relatively less afterwards. It's also heartening in a way that these analyses face some of the same issues surrounding clustered SEs and time trends, because it suggests that once these issues are addressed there will be some clearer conclusions.

5. Next steps and questions

The questions I need help with are:

1. **How to control for time trends in the event study analysis?**
2. **How to think about clustering standard errors, and whether it's an issue that some of my estimates have extremely high precision?**
3. Allcott et al report panel results with *just* county-time fixed effects, and *also* state-time fixed effects. **How can I think about composing these two types of fixed effects together? Is there any sample code I could take a look at to address this?**

The steps I'll take for next Friday are:

- Finish POI data build and run models with POI visits as dependent variable, including additional analysis by NAICS code
- Add controls for policies, public health, weather and demographics
- Incorporate feedback, especially as pertains to the questions above
- Organize paper more traditionally (abstract, introduction, lit review, data & methods, results, conclusion, appendices)