HCMG 901 Problem Set 3

Andrés Rovira

March 24, 2023

1 Stevenson and Wolfers (2006) Replication

1.1 Skim through SW2006 to orient yourself on the setting, empirical design, treated states, untreated states, etc. This will help when you try to replicate their results. What is their estimating equation? What is the key identification assumption? Do you have any concerns about this assumption? What evidence would you ideally provide to alleviate the concern you mentioned above? (Don't worry about whether you can do this with the data provided here)

Their estimating equation is the following two-way fixed effects specification:

$$Suicide\ rate_{s,t} = \sum_k \beta_k Unilateral^k_{s,t} + \gamma_s + \phi_t + Controls_{s,t} + \varepsilon_{s,t}$$

where $Unilateral^k_{s,t}$ is equal to 1 if state s adopted unilateral divorce k years ago from year t, and 0 otherwise.

The key identification assumption is parallel trends: in the absence of treatment, the treated units would have progressed on the same trend as the comparison units. However, I am concerned that, in this staggered differences-in-differences setting, already-treated and later-treated units are effectively used as comparison units for other treated units. As we know from class, this means that if treatment effects are heterogeneous or dynamic, the TWFE estimator will be biased. Ideally, I would like to see evidence from a "stacked" research design, where each treatment unit is compared to a clean group of never-treated comparison units, and then all of these data sets are stacked together and the treatment effects are estimated by one regression.

1.2 Remarkably, SW2006 does not have a descriptive statistics table. Prepare a descriptive statistics table that you would normally provide in such a paper. It should describe the sample size, means, and SD of suicide rates for men and women (key outcome) and the policy indicator (i.e., the key explanatory variable).

See Table 1.

Table 1: Descriptive Statistics

	Post Policy Change	Suicide Rate Men	Suicide Rate Women
Mean	0.53	18.48	5.06
SD	0.50	8.33	2.88
N	2,301	2,301	2,301

1.3 Replicate the analysis presented in Table 1. The data file does not contain any control variables, so just estimate a specification without controls, equivalent to columns 1f and 1m. Interpret the coefficients.

See Table 2. The interpretation of the coefficients is as follows: 9-10 years after the policy change, female suicide rates were down 10.4% (significantly different from 0) and male suicide rates were down 2.0% (not significantly different from 0). Overall, we can interpret the coefficients as showing that female suicides gradually dropped by a large amount over time, while male suicides remained fairly stable. Due to different data files, I am not able to replicate Table I exactly, but the results are very close, and the story remains the same.

1.4 Replicate Figure 5 in GB2021. Also produce the equivalent version of this figure for male suicides. Discuss how this event study is helpful in the context of the identification assumption.

See Figures 1 and 2.

Event studies of this type are useful for providing evidence on the absence of pre-trends, as long as the pre-treatment coefficients are close to 0 and flat. This is the case for the women, but perhaps not so for men, since their pre-treatment estimates are fairly negative already. The absence of pre-trends suggests that the parallel trends assumption could hold, though it cannot prove it definitively by any means.

The event studies also illustrate how the treatment effects are dynamic, varying over time. The effects particularly appear to grow over time for women. This justifies some concern about the usage of the TWFE estimator in this staggered DD design, since we know that bias can arise when treatment effects are dynamic.

Table 2: Replication of Table I

	(1) Female	(2) Male
Year of Change	1.1	-0.1
	(3.8)	(2.3)
1-2 Years Later	-1.7	2.4
	(3.8)	(1.7)
3-4 Years Later	-1.9	0.8
	(3.1)	(1.7)
5-6 Years Later	-3.3	1.3
	(2.9)	(1.7)
7-8 Years Later	-9.0	0.5
	(2.9)	(2.1)
9-10 Years Later	-10.4	-2.0
	(3.0)	(2.0)
11-12 Years Later	-10.7	-0.9
	(3.3)	(2.3)
13-14 Years Later	-12.2	-1.1
	(3.2)	(2.4)
15-16 Years Later	-13.1	-0.0
	(3.7)	(2.3)
17-18 Years Later	-16.5	-1.6
	(3.7)	(2.3)
19+ Years Later	-18.9	-3.7
	(3.2)	(2.2)
F Statistic	39.17	117.1
(p value)	0.000	0.000

Standard errors in parentheses

Figure 1: Replication of Figure 5 in GB2021 for women

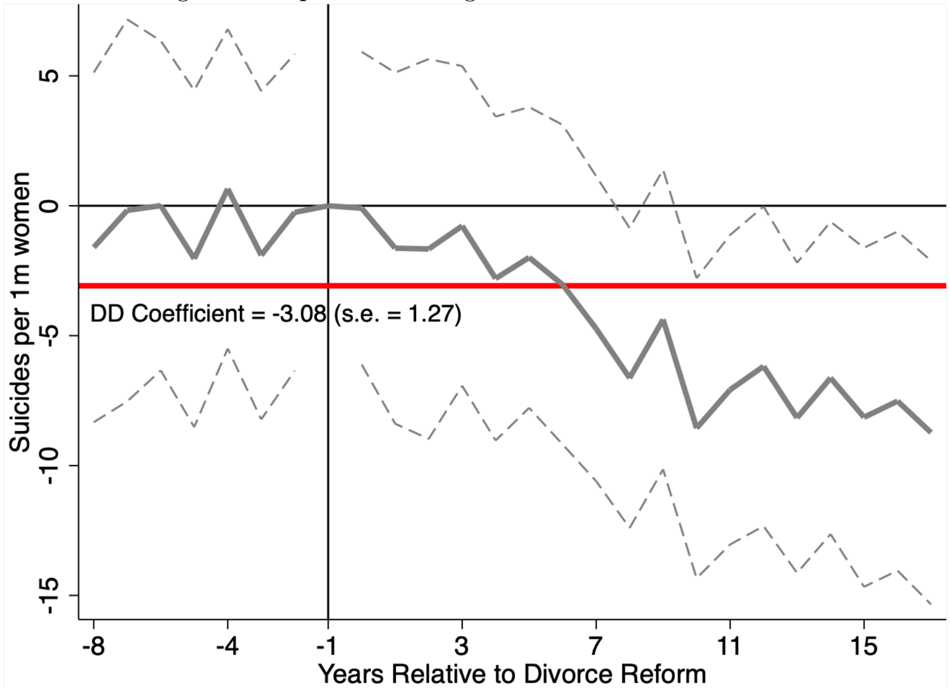
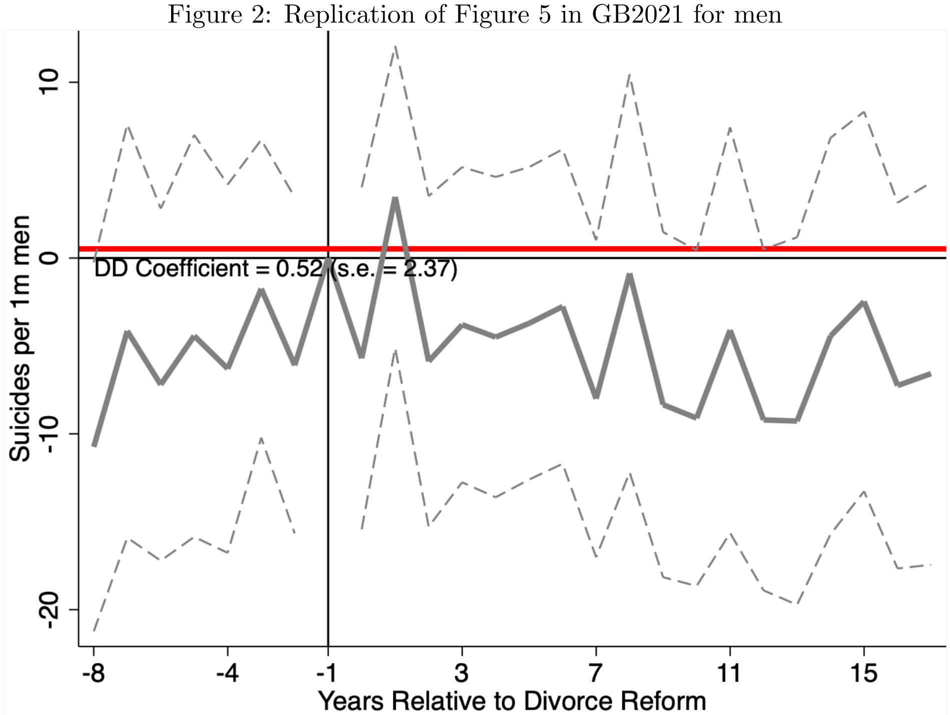


Figure 2: Replication of Figure 5 in GB2021 for men



1.5 Discuss the different types of 2x2 DD effects that are averaged together in a generalized D-D model. Following the notation in CH2020 that we discussed in class, why do TWFE estimates get biased in the case of staggered designs? What is the intuition behind the negative weights? Which groups or periods are more likely to get negative weights?

A generalized DD model averages together the following types of 2x2 DD effects:

1. Groups treated late vs groups treated early
2. Groups treated early vs groups treated late
3. Groups always treated vs "timing" groups (treated at some point in sample)
4. Groups never treated vs timing groups

TWFE estimates can be biased in staggered designs because when treatment effects are dynamic or heterogeneous, effects for some of these groups may receive negative weights. So, even if all of the 2x2 effects are positive, negative weights could potentially make the TWFE estimate negative!

The weights are defined as:

$$w_{fe,g,t} = \frac{\epsilon_{fe,g,t}}{\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \epsilon_{fe,g,t}}$$

Where $\epsilon_{fe,g,t}$ is the residual of observations in cell (g,t) from a regression of treatment indicator $D_{g,t}$ on a constant, group FE, and time FE. Clearly the weights can be negative since $\epsilon_{fe,g,t}$ can easily be negative. Intuitively, when treatment effects are dynamic or heterogeneous, mean outcomes across groups may follow different trends across periods either because more groups become treated or because of changes in treatment effects for groups treated earlier, and this is where the negative weights can arise. Groups treated for more periods and periods where a large fraction of groups are treated are more likely to get negative weights.

1.6 Replicate Figure 6 in GB2021 using the bacondecomp command that GB has produced (freely available for download). Estimate the model without controls and use the ddetail option. Briefly discuss the weights assigned to the different categories of estimators. Comment on the 2x2 estimate when we compare later treated states to early treated states as controls. What role is it playing in the overall estimate?

See Figure 3 and Table 3. Notice that the weights are positive for all 2x2 comparison categories, with the highest weight on the "Always treated vs timing" comparisons, and the lowest weight on "Early vs Late" comparisons.

The 2x2 estimate comparing later treated states to early treated states as controls is 3.51, which is notably the only positive estimate of the comparison categories. With a large weight of 0.25, this estimate is clearly pushing the overall estimate upward toward 0, which is likely a source of bias.

Figure 3: Replication of Figure 6 in GB2021

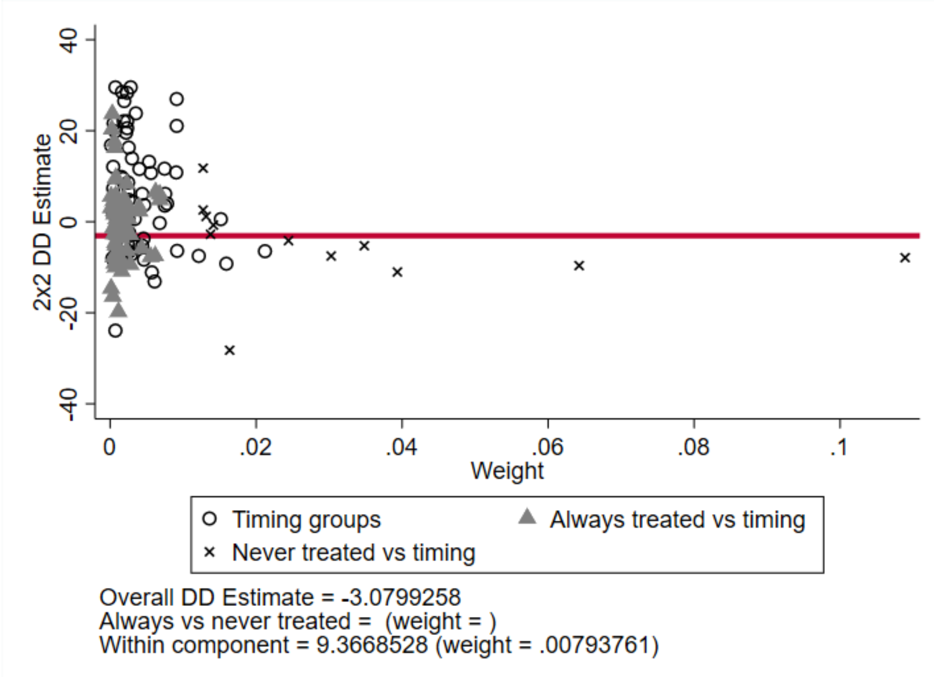


Table 3: DD Estimates and Weights by Category

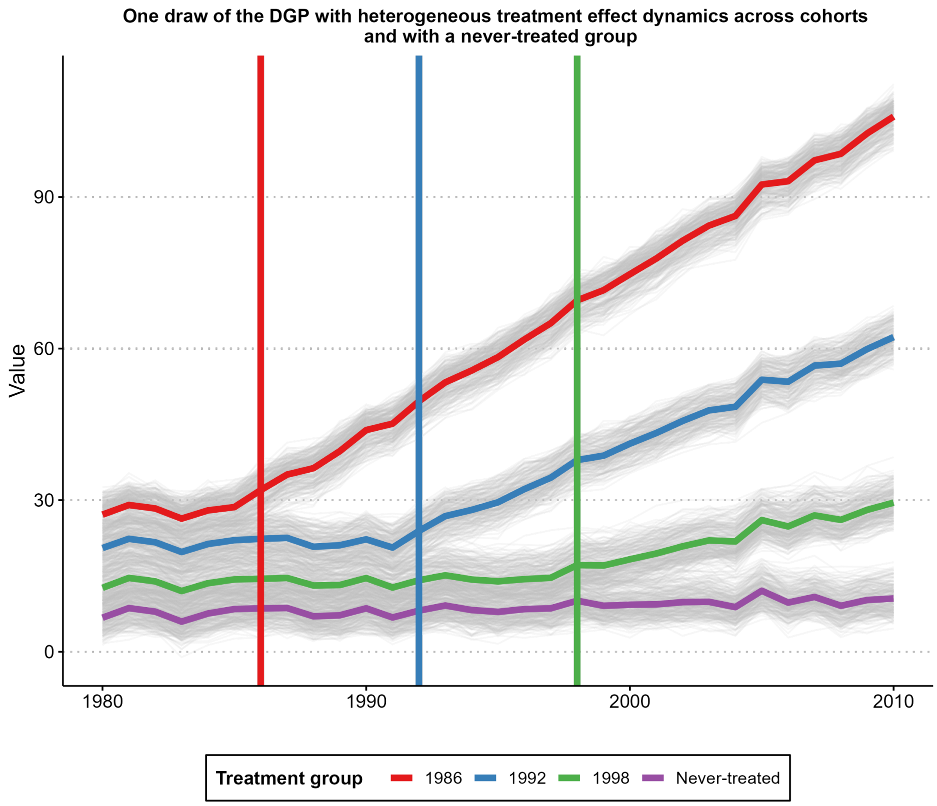
Group	DD Estimate	Weight
Late vs Early	3.51	0.26
Early vs Late	–0.19	0.11
Always treated vs timing	–7.04	0.38
Never treated vs timing	–5.33	0.24

2 Sant'anna Simulation Replication

2.1 Simulate the data and plot the raw trends in values by group over time. Check that it looks like his figure.

See Figure 4, which indeed looks like his figure.

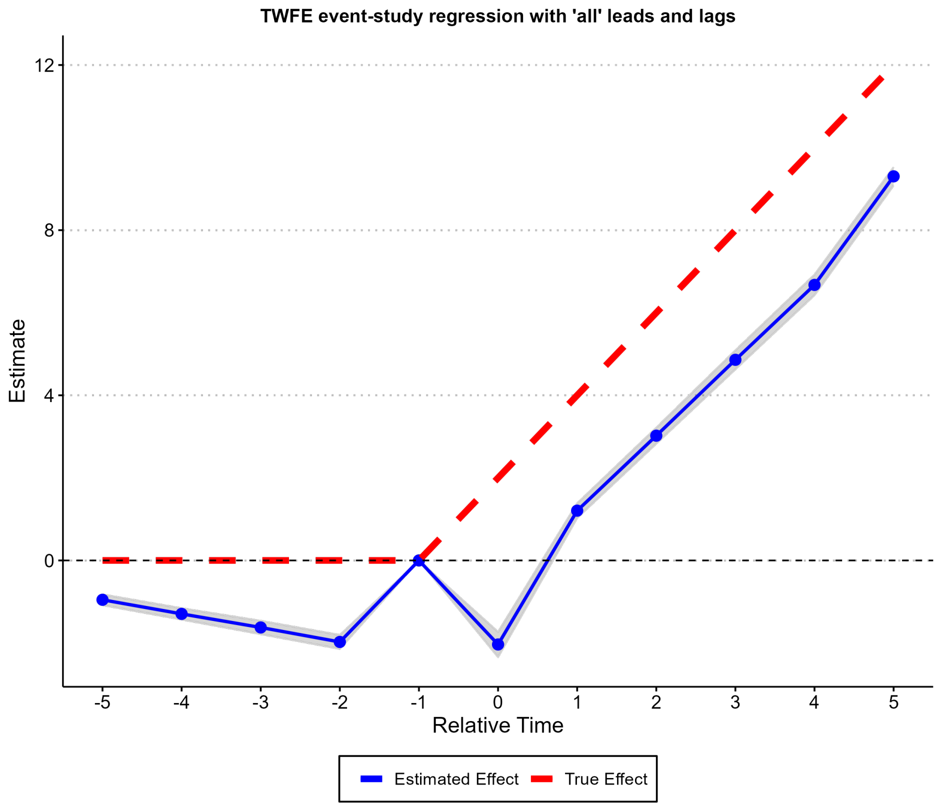
Figure 4:



2.2 Estimate the TWFE event study with all leads and lags (i.e., no binning). Does it approximately recover the true treatment effects in event time?

See Figure 5. The TWFE estimator is clearly biased downards, so it does not approximately recover the true treatment effects.

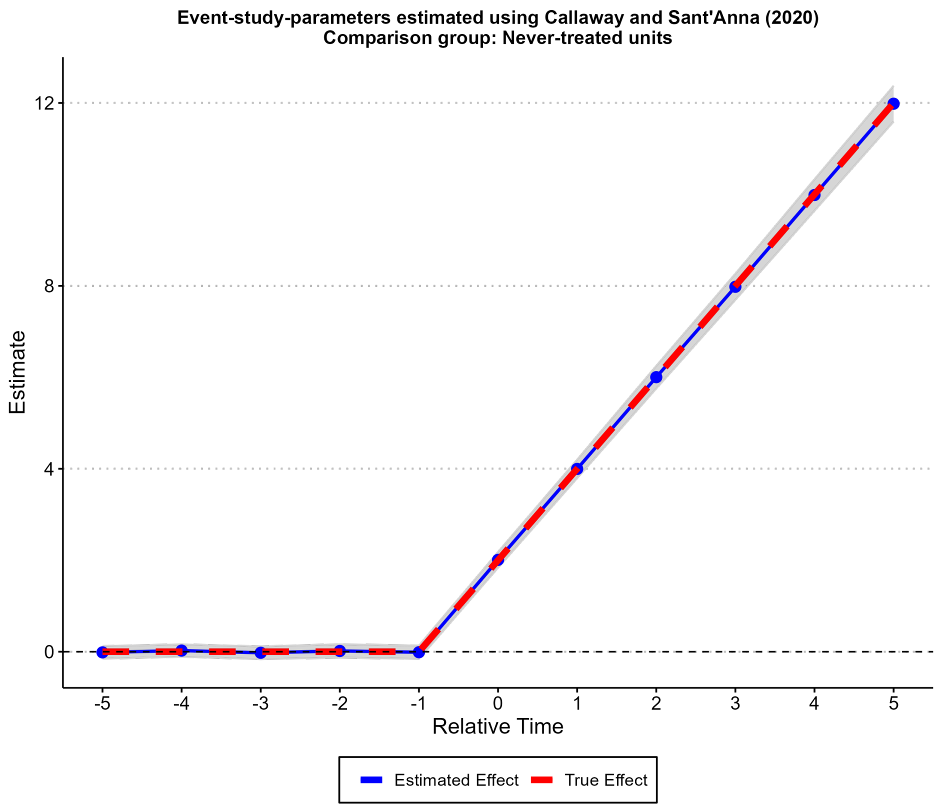
Figure 5:



2.3 Now use the Callaway Sant'anna estimator to produce the event study. Does this match the true effect?

See Figure 6. Indeed the estimates now match the true effects.

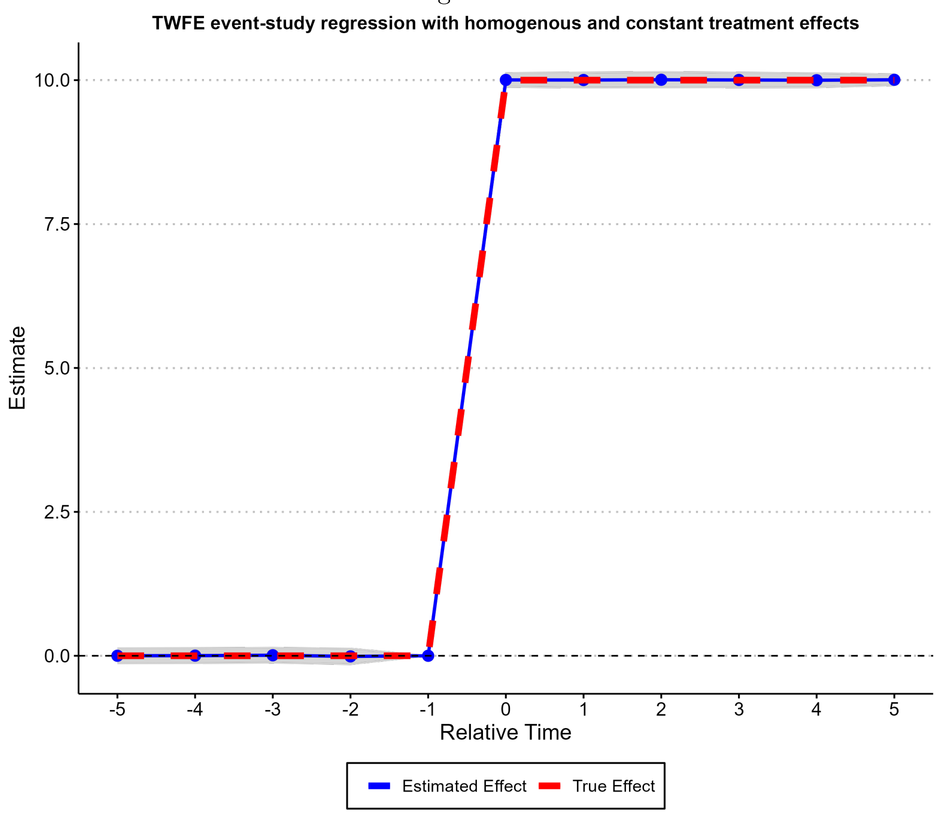
Figure 6:



2.4 Now, let's change the data generating process so that the treatment effects are constant over time and across all groups. So, the outcome jumps on treatment but stays stable over time instead of increasing. Re-estimate the TWFE estimator on this new constant treatment effects data. This time does the TWFE recover the true effects?

See Figure 7. Yes, now the TWFE estimator does recover the true effects.

Figure 7:



Stata Code for Part 1

```
/*****************************************************************************
AUTHORS: Andres Rovira, Stephanie Grove
CREATED: 2023-03-17
PURPOSE: Solve 1st part of HCMG 901 Problem Set 3
*****************************************************************************/

global home "~"
if "$S_OS" == "Windows" global home "`:env USERPROFILE'"
global code "$home/Dropbox (Penn)/Classes/2_health_applied_metrics/ps3/code"
global tex  "$home/Dropbox (Penn)/Apps/Overleaf/hcmg901_ps3"
cd "$code"

// ssc install texsave
// ssc install outreg2
// ssc install bacondecomp
// ssc install estout

******************************************************************************
**# 1.2
******************************************************************************

use "../input/divorce_example.dta", clear
isid year stfips sex

gen post = (year >= _nfd)
replace post = 1 if nfd == "PRE"

keep year stfips sex asmrs post
reshape wide asmrs, i(year stfips post) j(sex)
isid year stfips

local desc_vars post asmrs1 asmrs2
local stats count sd mean
tempfile desc_table
foreach stat of local stats {
	preserve
		collapse (`stat') `desc_vars'
		gen statistic = "`stat'"
		if "`stat'" == "count" {
			save `desc_table', replace
		}
		else {
			append using `desc_table'
			save `desc_table', replace
		}
	restore
}
use `desc_table', clear
```

```
order statistic
replace statistic = "Mean" if statistic == "mean"
replace statistic = "SD" if statistic == "sd"
replace statistic = "N" if statistic == "count"
label var post   "Post Policy Change"
label var asmrs1 "Suicide Rate Men"
label var asmrs2 "Suicide Rate Women"
format post asmrs? %13.2fc
tostring *, replace force usedisplayformat
foreach var of varlist * {
        replace `var' = subinstr(`var', ".00", "", .)
}
texsave * using "$tex/tab_2.tex", varlabels frag replace location("H") ///
        title("Descriptive Statistics")


********************************************************************************
**# 1.3
********************************************************************************

use "../input/divorce_example.dta", clear
keep if year >= 1964

gen post = (year >= _nfd)
replace post = 1 if nfd == "PRE"

gen years_post = year - _nfd
gen cat_post = 100 if years_post < 0 | nfd == "NRS"
replace cat_post = years_post if years_post == 0
forvalues i = 1(2)17 {
        replace cat_post = `i' if years_post == `i' | years_post == `i' + 1
}
replace cat_post = 19 if years_post >= 19 | nfd == "PRE"

label define lbl_years_post 0 "Year of Change" 1 "1-2 Years Later"
label define lbl_sex        1 "Male" 2 "Female"
label values sex lbl_sex

eststo clear
forvalues i = 1/2 {
        cap drop asmrs_elas_`i'
        summarize asmrs if sex == `i' & post == 0, meanonly
        gen asmrs_elas_`i' = 100 * asmrs / r(mean)
        if `i' == 1 label var asmrs_elas_`i' "Male Suicide Elas"
        if `i' == 2 label var asmrs_elas_`i' "Female Suicide Elas"
        eststo: regress asmrs_elas_`i' ib100.cat_post i.stfips i.year if sex == `i', robust
}

esttab est2 est1 using "$tex/tab_3.tex", replace ///
        nostar keep(*.cat_post) b(%9.1fc) se(%9.1fc) noobs ///
        coeflab(0.cat_post   "Year of Change" ///
```

```
            1.cat_post   "1-2 Years Later" ///
                    3.cat_post   "3-4 Years Later" ///
                    5.cat_post   "5-6 Years Later" ///
                    7.cat_post   "7-8 Years Later" ///
                    9.cat_post   "9-10 Years Later" ///
                    11.cat_post  "11-12 Years Later" ///
                    13.cat_post  "13-14 Years Later" ///
                    15.cat_post  "15-16 Years Later" ///
                    17.cat_post  "17-18 Years Later" ///
                    19.cat_post  "19+ Years Later" ///
                    100.cat_post "preceding policy change") ///
        drop(100.cat_post) varwidth(25) mtitles("Female" "Male") ///
        title("Replication of Table I") ///
        stats(F p, fmt(a3 3) labels("F Statistic" "(p value)"))


********************************************************************************
**# 1.4
********************************************************************************

do "DD_figure5_SW_replication_edited.do" 2
do "DD_figure5_SW_replication_edited.do" 1


********************************************************************************
**# 1.6
********************************************************************************

use "../input/divorce_example.dta" if year >= 1964 & sex == 2, clear

gen post = (year >= _nfd)
replace post = 1 if nfd == "PRE"

replace asmrs = asmrs * 10
keep asmrs year stfips post
xtset stfips year

bacondecomp asmrs post, ddetail stub("gb_")
egen wgt = total(gb_S), by(gb_cgroup)
collapse (mean) mean_b = gb_B [pweight = gb_S], by(gb_cgroup wgt)
egen tot_wgt = total(wgt)
gen norm_wgt = wgt / tot_wgt
drop wgt tot_wgt

graph export "$tex/fig_6.png", replace as(png)

label var gb_cgroup "Group"
label var mean_b    "DD Estimate"
label var norm_wgt  "Weight"
format mean_b norm_wgt %13.2fc
tostring mean_b norm_wgt, replace force usedisplayformat
```

```
decode gb_cgroup, gen(group)
drop gb_cgroup
order group
texsave * using "$tex/tab_6.tex", varlabels frag replace location("H") ///
        title("DD Estimates and Weights by Category")
```

R Code for Part 2

```
################################################################################
# AUTHORS: Pedro H. C. Sant'Anna and Brantly Callaway (light edits by Andres Rovira)
# CREATED: 2023-03-19 copied from source: https://psantanna.com/posts/twfe
# PURPOSE: Solve 2nd part of HCMG 901 Problem Set 3
################################################################################


# Load libraries and set baseline parameters
library(tidyverse)
library(lfe)
library(fastDummies)
library(ggthemes)
library(did)
library(fs)
theme_set(theme_clean() + theme(plot.background = element_blank()))

setwd(path_expand("~/Dropbox (Penn)/Classes/2_health_applied_metrics/ps3/code"))
tex_path <- path_expand("~/Dropbox (Penn)/Apps/Overleaf/hcmg901_ps3")

iseed = 20201221
nrep <- 100
true_mu <- 1
set.seed(iseed)
colors <- c("True Effect" = "red", "Estimated Effect" = "blue")


################################################################################
# 2.1
################################################################################

## Generate data - treated cohorts consist of 250 obs each, with the treatment effect still = true_mu
make_data3 <- function(nobs = 1000,
                       nstates = 40) {

  # unit fixed effects (unobservd heterogeneity)
  unit <- tibble(
    unit = 1:nobs,
    # generate state
    state = sample(1:nstates, nobs, replace = TRUE),
    unit_fe = rnorm(nobs, state/5, 1),
    # generate instantaneous treatment effect
    #mu = rnorm(nobs, true_mu, 0.2)
    mu = true_mu
  )

  # year fixed effects (first part)
  year <- tibble(
    year = 1980:2010,
    year_fe = rnorm(length(year), 0, 1)
```

```
  )

  # Put the states into treatment groups
  treat_taus <- tibble(
    # sample the states randomly
    state = sample(1:nstates, nstates, replace = FALSE),
    # place the randomly sampled states into 1\{t \ge g \}G_g
    cohort_year = sort(rep(c(1986, 1992, 1998, 2004), 10))
  )

  # make main dataset
  # full interaction of unit X year
  expand_grid(unit = 1:nobs, year = 1980:2010) %>%
    left_join(., unit) %>%
    left_join(., year) %>%
    left_join(., treat_taus) %>%
    # make error term and get treatment indicators and treatment effects
    # Also get cohort specific trends (modify time FE)
    mutate(error = rnorm(nobs*31, 0, 1),
           treat = ifelse((year >= cohort_year)* (cohort_year != 2004), 1, 0),
           mu = ifelse(cohort_year==1992, 2, ifelse(cohort_year==1998, 1, 3)),
           tau = ifelse(treat == 1, mu, 0),
           year_fe = year_fe + 0.1*(year - cohort_year)
    ) %>%
    # calculate cumulative treatment effects
    group_by(unit) %>%
    mutate(tau_cum = cumsum(tau)) %>%
    ungroup() %>%
    # calculate the dep variable
    mutate(dep_var = (2010 - cohort_year) + unit_fe + year_fe + tau_cum + error) %>%
    # Relabel 2004 cohort as never-treated
    mutate(cohort_year = ifelse(cohort_year == 2004, Inf, cohort_year))

}
#----------------------------------------------------------------------------
# make data
data <- make_data3()

# plot
plot3 <- data %>%
  ggplot(aes(x = year, y = dep_var, group = unit)) +
  geom_line(alpha = 1/8, color = "grey") +
  geom_line(data = data %>%
              group_by(cohort_year, year) %>%
              summarize(dep_var = mean(dep_var)),
            aes(x = year, y = dep_var, group = factor(cohort_year),
                color = factor(cohort_year)),
            size = 2) +
  labs(x = "", y = "Value",  color = "Treatment group   ") +
  geom_vline(xintercept = 1986, color = '#E41A1C', size = 2) +
  geom_vline(xintercept = 1992, color = '#377EB8', size = 2) +
```

```
  geom_vline(xintercept = 1998, color = '#4DAF4A', size = 2) +
  #geom_vline(xintercept = 2004, color = '#984EA3', size = 2) +
  scale_color_brewer(palette = 'Set1') +
  theme(legend.position = 'bottom',
        #legend.title = element_blank(),
        axis.title = element_text(size = 14),
        axis.text = element_text(size = 12)) +
  scale_color_manual(labels = c("1986", "1992", "1998", "Never-treated"),
                     values = c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3")) +
  ggtitle("One draw of the DGP with heterogeneous treatment effect dynamics across cohorts \n and wit
  theme(plot.title = element_text(hjust = 0.5, size=12))

plot3
ggsave("fig_2-1.png", path = tex_path)




################################################################################
# 2.2
################################################################################

# function to run ES DID
run_ES_DiD_sat_never_het <- function(...) {

  # resimulate the data
  data <- make_data3()

  # make dummy columns
  data <- data %>%
    # make relative year indicator
    mutate(rel_year = year - cohort_year)

  # get the minimum relative year - we need this to reindex
  min_year <- min(data$rel_year * (data$rel_year != -Inf), na.rm = T)

  # reindex the relative years
  data <- data %>%
    mutate(rel_year2 = rel_year) %>%
    mutate(rel_year = rel_year - min_year) %>%
    dummy_cols(select_columns = "rel_year") %>%
    select(-("rel_year_-Inf"))


  # make regression formula
  indics <- paste("rel_year", (1:max(data$rel_year))[-(-1 - min_year)], sep = "_", collapse = " + ")
  keepvars <- paste("rel_year", c(-5:-2, 0:5) - min_year, sep = "_")
  formula <- as.formula(paste("dep_var ~", indics, "| unit + year | 0 | state"))

  # run mod
  mod <- felm(formula, data = data, exactDOF = TRUE)
```

```
  # grab the obs we need
  # grab the obs we need
  mod2 <- tibble(
    estimate = mod$coefficients,
    term1 = rownames(mod$coefficients)
  )

  es <-
    mod2 %>%
    filter(term1 %in% keepvars) %>%
    mutate(t = c(-5:-2, 0:5)) %>%
    select(t, estimate)
  es
}

data_sat_never_het <- map_dfr(1:nrep, run_ES_DiD_sat_never_het)

ES_plot_sat_never_het <- data_sat_never_het %>%
  group_by(t) %>%
  summarize(avg = mean(estimate),
            sd = sd(estimate),
            lower.ci = avg - 1.96*sd,
            upper.ci = avg + 1.96*sd) %>%
  bind_rows(tibble(t = -1, avg = 0, sd = 0, lower.ci = 0, upper.ci = 0)) %>%
  mutate(true_tau = ifelse(t >= 0, (t + 1)* 2, 0)) %>%
  ggplot(aes(x = t, y = avg)) +
  #geom_linerange(aes(ymin = lower.ci, ymax = upper.ci), color = 'darkgrey', size = 2) +
  geom_ribbon(aes(ymin = lower.ci, ymax = upper.ci), color = "lightgrey", alpha = 0.2) +
  geom_point(color = 'blue', size = 3) +
  geom_line(aes(color = 'Estimated Effect'), size = 1) +
  geom_line(aes(x = t, y = true_tau, color = 'True Effect'), linetype = "dashed", size = 2) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  scale_x_continuous(breaks = -5:5) +
  labs(x = "Relative Time", y = "Estimate") +
  theme(axis.title = element_text(size = 14),
        axis.text = element_text(size = 12))+
  ggtitle("TWFE event-study regression with 'all' leads and lags")+
  scale_color_manual(values = colors) +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        legend.position = "bottom",
        legend.title = element_blank())

ES_plot_sat_never_het
ggsave("fig_2-2.png", path = tex_path)


################################################################################
# 2.3
################################################################################
```

```
# function to run ES DID
run_CS_never_het <- function(...) {

  # resimulate the data
  data <- make_data3()
  data$cohort_year[data$cohort_year==Inf] <- 0

  mod <- did::att_gt(yname = "dep_var",
                     tname = "year",
                     idname = "unit",
                     gname = "cohort_year",
                     control_group= "nevertreated", # AR changed this from "never_treated"
                     bstrap = FALSE,
                     data = data,
                     print_details = FALSE)
  event_std <- did::aggte(mod, type = "dynamic")

  att.egt <- event_std$att.egt
  names(att.egt) <- event_std$egt

  # grab the obs we need
  broom::tidy(att.egt) %>%
    filter(names %in% -5:5) %>%
    mutate(t = -5:5, estimate = x) %>%
    select(t, estimate)
}

data_CS_never_het <- map_dfr(1:nrep, run_CS_never_het)

ES_plot_CS_never_het <- data_CS_never_het %>%
  group_by(t) %>%
  summarize(avg = mean(estimate),
            sd = sd(estimate),
            lower.ci = avg - 1.96*sd,
            upper.ci = avg + 1.96*sd) %>%
  mutate(true_tau = ifelse(t >= 0, (t + 1)* 2, 0)) %>%
  ggplot(aes(x = t, y = avg)) +
  #geom_linerange(aes(ymin = lower.ci, ymax = upper.ci), color = 'darkgrey', size = 2) +
  geom_ribbon(aes(ymin = lower.ci, ymax = upper.ci), color = "lightgrey", alpha = 0.2) +
  geom_point(color = 'blue', size = 3) +
  geom_line(aes(color = 'Estimated Effect'), size = 1) +
  geom_line(aes(x = t, y = true_tau, color = 'True Effect'), linetype = "dashed", size = 2) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  scale_x_continuous(breaks = -5:5) +
  labs(x = "Relative Time", y = "Estimate") +
  theme(axis.title = element_text(size = 14),
        axis.text = element_text(size = 12))+
  ggtitle("Event-study-parameters estimated using Callaway and Sant'Anna (2020)\nComparison group: Ne
  scale_color_manual(values = colors) +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        legend.position = "bottom",
```

```
        legend.title = element_blank())

ES_plot_CS_never_het
ggsave("fig_2-3.png", path = tex_path)



################################################################################
# 2.4
################################################################################


## Generate data - treated cohorts consist of 250 obs each, with the treatment effect still = true_mu
make_data4 <- function(nobs = 1000,
                       nstates = 40) {

  # unit fixed effects (unobservd heterogeneity)
  unit <- tibble(
    unit = 1:nobs,
    # generate state
    state = sample(1:nstates, nobs, replace = TRUE),
    unit_fe = rnorm(nobs, state/5, 1),
    # generate instantaneous treatment effect
    #mu = rnorm(nobs, true_mu, 0.2)
    mu = true_mu
  )

  # year fixed effects (first part)
  year <- tibble(
    year = 1980:2010,
    year_fe = rnorm(length(year), 0, 1)
  )

  # Put the states into treatment groups
  treat_taus <- tibble(
    # sample the states randomly
    state = sample(1:nstates, nstates, replace = FALSE),
    # place the randomly sampled states into 1\{t \ge g \}G_g
    cohort_year = sort(rep(c(1986, 1992, 1998, 2004), 10))
  )

  # make main dataset
  # full interaction of unit X year
  expand_grid(unit = 1:nobs, year = 1980:2010) %>%
    left_join(., unit) %>%
    left_join(., year) %>%
    left_join(., treat_taus) %>%
    # make error term and get treatment indicators and treatment effects
    # Also get cohort specific trends (modify time FE)
    mutate(error = rnorm(nobs*31, 0, 1),
           treat = ifelse((year >= cohort_year)* (cohort_year != 2004), 1, 0),
```

```
          mu = 10, #ifelse(cohort_year==1992, 2, ifelse(cohort_year==1998, 1, 3)),
          tau = ifelse(treat == 1, mu, 0),
          year_fe = year_fe + 0.1*(year - cohort_year)
    ) %>%
    # calculate cumulative treatment effects
    group_by(unit) %>%
    mutate(tau_cum = cummax(tau)) %>%
    ungroup() %>%
    # calculate the dep variable
    mutate(dep_var = (2010 - cohort_year) + unit_fe + year_fe + tau_cum + error) %>%
    # Relabel 2004 cohort as never-treated
    mutate(cohort_year = ifelse(cohort_year == 2004, Inf, cohort_year))

}
#----------------------------------------------------------------------------
# make data
data <- make_data4()

# plot
plot4 <- data %>%
  ggplot(aes(x = year, y = dep_var, group = unit)) +
  geom_line(alpha = 1/8, color = "grey") +
  geom_line(data = data %>%
              group_by(cohort_year, year) %>%
              summarize(dep_var = mean(dep_var)),
            aes(x = year, y = dep_var, group = factor(cohort_year),
                color = factor(cohort_year)),
            size = 2) +
  labs(x = "", y = "Value",  color = "Treatment group   ") +
  geom_vline(xintercept = 1986, color = '#E41A1C', size = 2) +
  geom_vline(xintercept = 1992, color = '#377EB8', size = 2) +
  geom_vline(xintercept = 1998, color = '#4DAF4A', size = 2) +
  #geom_vline(xintercept = 2004, color = '#984EA3', size = 2) +
  scale_color_brewer(palette = 'Set1') +
  theme(legend.position = 'bottom',
        #legend.title = element_blank(),
        axis.title = element_text(size = 14),
        axis.text = element_text(size = 12)) +
  scale_color_manual(labels = c("1986", "1992", "1998", "Never-treated"),
                     values = c("#E41A1C", "#377EB8", "#4DAF4A", "#984EA3")) +
  ggtitle("One draw of the DGP with heterogeneous treatment effect dynamics across cohorts \n and wit
  theme(plot.title = element_text(hjust = 0.5, size=12))

plot4



# function to run ES DID
run_ES_DiD_sat_never_het_4 <- function(...) {

  # resimulate the data
```

```
  data <- make_data4()

  # make dummy columns
  data <- data %>%
    # make relative year indicator
    mutate(rel_year = year - cohort_year)

  # get the minimum relative year - we need this to reindex
  min_year <- min(data$rel_year * (data$rel_year != -Inf), na.rm = T)

  # reindex the relative years
  data <- data %>%
    mutate(rel_year2 = rel_year) %>%
    mutate(rel_year = rel_year - min_year) %>%
    dummy_cols(select_columns = "rel_year") %>%
    select(-("rel_year_-Inf"))


  # make regression formula
  indics <- paste("rel_year", (1:max(data$rel_year))[-(-1 - min_year)], sep = "_", collapse = " + ")
  keepvars <- paste("rel_year", c(-5:-2, 0:5) - min_year, sep = "_")
  formula <- as.formula(paste("dep_var ~", indics, "| unit + year | 0 | state"))

  # run mod
  mod <- felm(formula, data = data, exactDOF = TRUE)

  # grab the obs we need
  # grab the obs we need
  mod2 <- tibble(
    estimate = mod$coefficients,
    term1 = rownames(mod$coefficients)
  )

  es <-
    mod2 %>%
    filter(term1 %in% keepvars) %>%
    mutate(t = c(-5:-2, 0:5)) %>%
    select(t, estimate)
  es
}

data_sat_never_het_4 <- map_dfr(1:nrep, run_ES_DiD_sat_never_het_4)

ES_plot_sat_never_het_4 <- data_sat_never_het_4 %>%
  group_by(t) %>%
  summarize(avg = mean(estimate),
            sd = sd(estimate),
            lower.ci = avg - 1.96*sd,
            upper.ci = avg + 1.96*sd) %>%
  bind_rows(tibble(t = -1, avg = 0, sd = 0, lower.ci = 0, upper.ci = 0)) %>%
  mutate(true_tau = ifelse(t >= 0, 10, 0)) %>%
```

```
  ggplot(aes(x = t, y = avg)) +
  #geom_linerange(aes(ymin = lower.ci, ymax = upper.ci), color = 'darkgrey', size = 2) +
  geom_ribbon(aes(ymin = lower.ci, ymax = upper.ci), color = "lightgrey", alpha = 0.2) +
  geom_point(color = 'blue', size = 3) +
  geom_line(aes(color = 'Estimated Effect'), size = 1) +
  geom_line(aes(x = t, y = true_tau, color = 'True Effect'), linetype = "dashed", size = 2) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  scale_x_continuous(breaks = -5:5) +
  labs(x = "Relative Time", y = "Estimate") +
  theme(axis.title = element_text(size = 14),
        axis.text = element_text(size = 12))+
  ggtitle("TWFE event-study regression with homogenous and constant treatment effects")+
  scale_color_manual(values = colors) +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        legend.position = "bottom",
        legend.title = element_blank())

ES_plot_sat_never_het_4
ggsave("fig_2-4.png", path = tex_path)
```