ARE 213 PS 2b

S. Sung, H. Husain, T. Woolley, A. Watt

2021-11-23

Contents

Problem 1

We first estimate an event study specification.

Part (a)

First determine the minimum and maximum event time values that you can estimate in this data set. Code up a separate event time indicator for each possible value of event time in the data set. Estimate an event study regression using all the event time indicators. What happens?

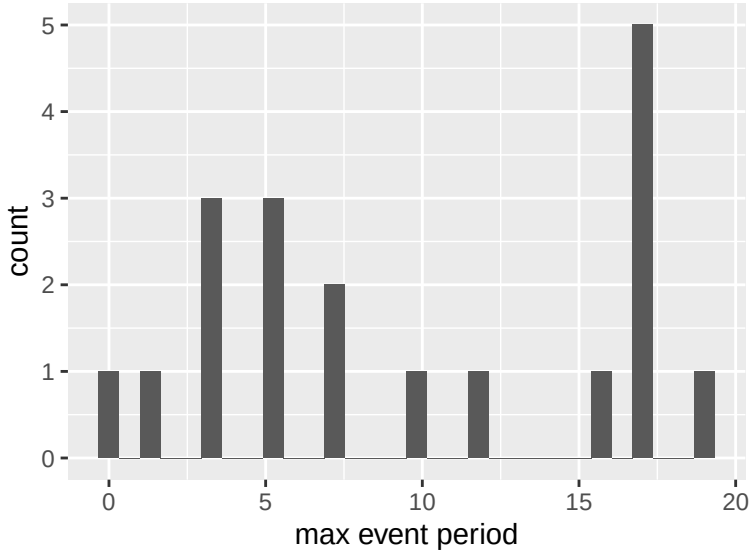
Table 1 lists the minimum and maximum event times that exist in the data for states that enacted a primary seat belt law in our study period (1981-2003).

Table 1: Maximum & Minimum Event Time Values

Max j	Min j
19	-22

Notice from Figure 1 that we have a wide range of maximum event times across our panel of states. One state even has a maximum event time of 0 – meaning they only added primary seat belt laws in the last year of our panel (2003). This means we will have an unbalanced panel if we run a regression on all possible event time dummies. In fact, we will still have an unbalance panel if we create max and min event time bins to aggregate early and late periods (an indicator for a event times greater than 5 and an indicator for all event times less than -5), we will still have an unbalanced panel.

Figure 1: Treated States, maximum event time histogram



To estimate the event study treatment effects, corresponding regression equation is:

$$Y_{st} = \alpha + \sum_{j=min\_t}^{max\_t} \tau_j D_{jst} + \gamma_s + \delta_t + \varepsilon_{st} + u_{st}$$

Note that we are estimating the regression with state and year fixed effects. In practice, we would want to omit a specific event time indicator so all our treatment effects are measured with respect to that event

time. If we keep all our indicators, then R will implicitly choose which to event time indicator to omit for us because the event time dummies, along with state and year fixed effects, are colinear.

We can see column (1) of Table 2 that the indicator for event time +19 was omitted for the regression. However, for interpretability, we'd rather have the treatment effects relative to a period closer to the year of initial treatment.

Part (b)

Estimate another event study regression using all the event time indicators save one that you choose to omit. Generate a plot of the event study coefficients.

We have chosen to omit the event time -1 from the regression so the other event time inidicator coefficients can be interpreted as relative to the year immediately before the passage of the primary seat belt law. Column (2) of Table 2 shows that the -1 event time period was omitted, and we we see that all of the treatment effects occuring before event time -1 are not significantly different from zero, whereas all the event time coefficient estimates for after event time -1 are negative and all are signficantly less than zero starting with event period 5.

Table 2: Event Study Regressions

	Log(Fatality per Population)	
	Event Study a	Event Study b
	(1)	(2)
‘-22_ET‘	0.1385 (0.1578)	−0.1545 (0.1137)
‘-21_ET‘	0.4172*** (0.1372)	0.1242 (0.0828)
‘-20_ET‘	0.3697*** (0.1371)	0.0767 (0.0826)
‘-19_ET‘	0.3091** (0.1231)	0.0162 (0.0568)
‘-18_ET‘	0.2851** (0.1231)	−0.0079 (0.0567)
‘-17_ET‘	0.3460*** (0.1191)	0.0531 (0.0475)
‘-16_ET‘	0.3543*** (0.1191)	0.0613 (0.0477)
‘-15_ET‘	0.3526*** (0.1176)	0.0597 (0.0436)
‘-14_ET‘	0.3113*** (0.1175)	0.0184 (0.0440)
‘-13_ET‘	0.3412*** (0.1175)	0.0482 (0.0435)
‘-12_ET‘	0.3161*** (0.1169)	0.0231 (0.0425)
‘-11_ET‘	0.3235*** (0.1168)	0.0305 (0.0423)
‘-10_ET‘	0.3105*** (0.1163)	0.0176 (0.0411)
‘-9_ET‘	0.2946** (0.1162)	0.0017 (0.0412)
‘-8_ET‘	0.3096*** (0.1162)	0.0166 (0.0408)
‘-7_ET‘	0.3313*** (0.1161)	0.0383 (0.0411)
‘-6_ET‘	0.3385*** (0.1156)	0.0455 (0.0397)
‘-5_ET‘	0.3155*** (0.1144)	0.0225 (0.0368)
‘-4_ET‘	0.3268*** (0.1143)	0.0338 (0.0369)
‘-3_ET‘	0.3225*** (0.1138)	0.0295 (0.0359)
‘-2_ET‘	0.2981*** (0.1137)	0.0051 (0.0363)
‘-1_ET‘	0.2929** (0.1137)	
‘0_ET‘	0.2623** (0.1135)	−0.0306 (0.0363)
‘1_ET‘	0.2484** (0.1136)	−0.0445 (0.0365)
‘2_ET‘	0.2464** (0.1139)	−0.0466 (0.0375)
‘3_ET‘	0.2466** (0.1134)	−0.0463 (0.0375)
‘4_ET‘	0.2278** (0.1144)	−0.0651 (0.0397)
‘5_ET‘	0.2127* (0.1138)	−0.0803** (0.0401)
‘6_ET‘	0.1965* (0.1154)	−0.0965** (0.0432)
‘7_ET‘	0.2173* (0.1149)	−0.0756* (0.0438)
‘8_ET‘	0.1612 (0.1165)	−0.1317*** (0.0471)
‘9_ET‘	0.1790 (0.1165)	−0.1140** (0.0471)
‘10_ET‘	0.1772 (0.1162)	−0.1157** (0.0474)
‘11_ET‘	0.1594 (0.1174)	−0.1336*** (0.0494)
‘12_ET‘	0.1443 (0.1170)	−0.1487*** (0.0500)
‘13_ET‘	0.1522 (0.1185)	−0.1408*** (0.0526)
‘14_ET‘	0.1574 (0.1185)	−0.1355** (0.0533)
‘15_ET‘	0.1504 (0.1184)	−0.1425*** (0.0531)
‘16_ET‘	0.0973 (0.1181)	−0.1956*** (0.0533)
‘17_ET‘	0.0842 (0.1183)	−0.2087*** (0.0576)
‘18_ET‘	0.0144 (0.1520)	−0.2785** (0.1135)
‘19_ET‘		−0.2929** (0.1137)
Constant	−1.4815*** (0.1158)	−1.1885*** (0.0387)
Chose dummy to omit	No	Yes
Observations	1,104	1,104
R^2	0.9111	0.9111
Adjusted R^2	0.9013	0.9013
Note:		*p<0.1; **p<0.05; ***p<0.01

Part (c)

Create minimum and maximum event time indicators that correspond to bins of event time < -5 and event time > 5 respectively. Appropriately specify and estimate an event study regression using these min and max event time indicators. Generate a plot of the event study coefficients. Explain which specification you prefer, this one or the one in part (b).

Table 3: Event Study Regression with Threshold Indicators

	Log(Fatality per Population)
	Event Study c
below_ET	0.0261 (0.0275)
'-5_ET'	0.0276 (0.0355)
'-4_ET'	0.0372 (0.0356)
'-3_ET'	0.0300 (0.0347)
'-2_ET'	0.0073 (0.0350)
'0_ET'	−0.0329 (0.0350)
'1_ET'	−0.0506 (0.0352)
'2_ET'	−0.0498 (0.0362)
'3_ET'	−0.0467 (0.0361)
'4_ET'	−0.0671* (0.0381)
'5_ET'	−0.0806** (0.0384)
above_ET	−0.1232*** (0.0285)
Constant	−1.1828*** (0.0373)
Chose dummy to omit	Yes
Agg. Threshold Indicators	Yes
Observations	1,127
R^2	0.9090
Adjusted R^2	0.9019

Note: *p<0.1; **p<0.05; ***p<0.01
above_ and below_ variables are aggregate indicators for all event times that are below -5 and above 5, respectively.

Part (d)

What happens to your estimates from part (b) if you exclude the "pure control" states from your sample? What about if you exclude the pure controls in part (c)?

Part (e)

Overall, does the event study regression make you more confident or less confident that seat belt laws reduce fatalities (relative to the fixed effects results that you estimated on the last problem set)? Briefly explain.

Part (f*)

Building off the event study regression from part (c), estimate the interaction weighted event study estimator from Sun and Abraham (2020). As a reminder, the interacted event study regression takes the standard event time indicators (without any binning) and interacts each one with a cohort indicator (a cohort refers to a group of states that share the same date on which they were first treated). You then form the estimate for event time coefficient τ_j by averaging the estimates of the cohort-specific τ_j using the weights described in Sun and Abraham (2020).

Problem 2

We now apply the synthetic control methods from Abadie et al (2010).

Part (a)

We created an aggregate "treatment" state (state number 99 or "TU") which combines the (population weighted) data from the first 4 states to have a primary seatbelt law (CT, IA, NM, TX). Please use this state as the "treatment" state in the synthetic control analysis.

—— a.i

Compare the average pre-period log traffic fatalities per capita of the TU site to that of the average of all the "control" states. Next, graph the pre-period log traffic fatalities by year for the pre-period for both the TU and the average of the control group. Interpret.

—— a.ii

Compare the dependent variable between the TU site and each control state for the year before the treatment. Which control state best matches the TU? Now compare this state's covariates with the TU covariates. Do they appear similar? What might this imply for in terms of using this state as the counterfactual state?

Part (b)

Apply the synthetic control method using the available covariates and pre-treatment outcomes to construct a synthetic control group.

—— **b.i**

Discuss the synthetic control method including its benefits and potential drawbacks.

—— **b.ii**

Use the software package provided by Abadie et al to apply the synthetic control method. (You are free to use either Stata, Matlab, or R but answers will be provided in Stata and R only). Please be sure to state precisely what the command is doing and how you determined your preferred specification.

Part (c)

Graphical interpretation and treatment significance.

—— c.i

Generate graphs plotting the gap between the TU and the synthetic control group under both your preferred specification and a few other specifications you tried.

—— c.ii

Compare the graph plotting the gap between the TU and the synthetic control group under your preferred specification with the graphs plotting the gap between each control state and its "placebo" treatment. Do you conclude that the treatment was significant? Why or why not?

—— c.iii

Create a graph of the post-treatment/pre-treatment prediction ratios of the Mean Squared Prediction Errors (MSPE) for the actual and "plecebo" treatment gaps in (ii). [See Abadie et al. for an example]. Do you conclude that the treatment was significant? Why or why not?

Part (d)

How do your synthetic control results compare to your fixed effects results from Question (3) in the last problem set? Interpret any differences.

Appendix A: R Code

```
rm(list=ls())
knitr::opts_chunk$set(echo = F)
# stargazer table type (html, latex, or text)
# Change to latex when outputting to PDF, html when outputting to html
table_type = "latex"

# install.packages("Synth")
library(tidyverse)
library(haven)
library(stargazer)
library(ggplot2)
library(tinytex)
library(Synth)
library(kableExtra)
# library(plm)
# library(lmtest)
# library(sandwich)
# library(gridExtra)
# library(grid)
# library(gtable)
library(fastDummies)
# library(EnvStats)
# Load data from PS2a with previous log variables
data = read_dta('traffic_safety2.dta') %>%
    mutate(fat_pc = fatalities/population,
           ln_fat_pc = log(fat_pc),
           ln_tvmt_pc = log(totalvmt/population),
           ln_precip = log(precip),
           ln_rspeed = log(rural_speed),
           ln_uspeed = log(urban_speed))

# Create list of event dates for states that passed primary laws in our study
event_dates = data %>%
    group_by(state) %>%
    mutate(event = primary - lag(primary), # event=1 ==> first year primary=1
           event_year = year) %>%
    filter(event == 1) %>%
    select(state, event_year)

# Add year of primary event and event time (t) to dataframe
data = data %>%
    left_join(event_dates, by='state') %>%
    mutate(j = ifelse(is.na(event_year), 99, year - event_year))
    # t = 99 ==> control state (doesn't pass primary during study period)

# Table of max and min event times
# Shouldn't these be our event study thresholds?
df_temp = data %>%
    filter(j < 99) %>%
    group_by(state) %>%
    summarize(min_j = min(j), max_j = max(j))
```

```r
max_j_inclusive = min(df_temp$max_j, na.rm = T)
min_j_inclusive = max(df_temp$min_j, na.rm = T)
max_j = max(filter(data, j<99)$j, na.rm = T)
min_j = min(filter(data, j<99)$j, na.rm = T)
data.frame(max_j = max_j, min_j = min_j) %>%
  kbl(caption = "Maximum \\& Minimum Event Time Values",
      col.names = c('Max j', 'Min j'),
      align = 'cc') %>%
  kable_styling(latex_options = "HOLD_position")
df_temp %>%
    arrange(max_j) %>%
    filter(!is.na(max_j), max_j < 99) %>%
    select(max_j) %>%
    ggplot(aes(x=max_j), data=.) +
    geom_histogram() +
    xlab("max event period")
# Function for adding dummies to a dataframe for all uniuqe values between given numbers
create_dummies = function(df, colname, min_value, max_value) {
    # Create dummies for each value of colname between min_value and max_value
    df1 = df
    for (val in min_value:max_value) {
        df1 = mutate(df1, "{colname}_{val}" := ifelse(eval(as.symbol(colname)) == val, 1, 0))
    }
    return(df1)
}

# Create order of dummies for dataframe (then used for regression table)
name_order1 = paste('j', min_j_inclusive:max_j_inclusive, sep='_')
# Create Dummies that make a balanced panel
# (only dummies for event times j that are shared across all states)
df_few_dummies = create_dummies(data,
                                colname = 'j',
                                min_value = min_j_inclusive,
                                max_value = max_j_inclusive) %>%
    relocate(all_of(name_order1)) %>%
    # Change "j_..." to "..._ET" because LaTeX doesn't like j_-3 type variable names
    rename_with(~ paste0(str_replace(., 'j_', ''), '_ET'), contains("j_"))

# Create order of dummies for dataframe (then used for regression table)
name_order2 = paste('j', min_j:max_j, sep='_')
# Create Dummies for all possible event times
# (results in unbalanced panel over event times j)
df_all_dummies = dummy_cols(data, select_columns = 'j') %>%
    select(-j_99) %>%
    filter(state != 99)  %>%
    relocate(all_of(name_order2)) %>%
    rename_with(~ paste0(str_replace(., 'j_', ''), '_ET'), contains("j_"))
reg_1a = df_all_dummies %>%
    mutate(state=factor(state), year=factor(year)) %>%
    select(ln_fat_pc, state, year, contains('_ET')) %>%
    lm(ln_fat_pc ~ ., data = .)

reg_1b = df_all_dummies %>%
```

```
    mutate(state=factor(state), year=factor(year)) %>%
    select(ln_fat_pc, state, year, contains('_ET'), -`-1_ET`) %>%
    lm(ln_fat_pc ~ ., data = .)

stargazer(reg_1a, reg_1b,
          title = "Event Study Regressions\\label{tab:event-study-dummy-trap}",
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE,
          # model.names = FALSE,
          column.labels = c("Event Study a", "Event Study b"),
          # keep = paste0('^j', (-5):5, '$'),
          # order = paste0('/^t_', (-5):5, '$/'),
          # perl = T,
          # covariate.labels = as.character((-5:5)),
          omit = c("state", "year"),
          add.lines=list(c('Chose dummy to omit', 'No', 'Yes')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE,
          omit.stat=c("f", "ser"),
          single.row = TRUE,
          digits = 4, type = table_type, header = FALSE)


# Function for adding dummies to a dataframe for all uniuqe values between given numbers
create_dummies_threshold = function(df, colname, min_value, max_value, suffix = NULL) {
    # Create indicator variables for each value of colname between min_value and max_value
    # then create indicator variables for all values of colname below min_value
    # and another for above max_value
    if (is.null(suffix)) {suffix = colname}
    df1 = df
    # add aggregate indicator for all values below min_value
    df1 = mutate(df1, "below_{suffix}" := ifelse(eval(as.symbol(colname)) < min_value, 1, 0))
    # add all indicators in between min and max_value
    for (val in min_value:max_value) {
        df1 = mutate(df1, "{val}_{suffix}" := ifelse(eval(as.symbol(colname)) == val, 1, 0))
    }
    # add aggregate indicator for all values above max_value
    df1 = mutate(df1, "above_{suffix}" := ifelse(eval(as.symbol(colname)) > max_value, 1, 0))
    return(df1)
}


# Create Dummies that make a balanced panel
# (only dummies for event times j that are shared across all states)
df_threshold_dummies = create_dummies_threshold(data,
                                                colname = 'j',
                                                min_value = -5,
                                                max_value = 5,
                                                suffix = 'ET')

reg_1c = df_threshold_dummies %>%
    mutate(state=factor(state), year=factor(year)) %>%
    select(ln_fat_pc, state, year, contains('_ET'), -`-1_ET`) %>%
    lm(ln_fat_pc ~ ., data = .)
```

```r
stargazer(reg_1c,
          title = "Event Study Regression with Threshold Indicators\\label{tab:event-study-thresholds}"
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE,
          # model.names = FALSE,
          column.labels = c("Event Study c"),
          # keep = paste0('^j', (-5):5, '$'),
          # order = paste0('/^t_', (-5):5, '$/'),
          # perl = T,
          # covariate.labels = as.character((-5:5)),
          omit = c("state", "year"),
          add.lines=list(c('Chose dummy to omit', 'Yes'), c('Agg. Threshold Indicators', 'Yes')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE,
          omit.stat=c("f", "ser"),
          single.row = TRUE,
          notes = c("above\\_ and below\\_ variables are aggregate indicators for",
                    "all event times that are below -5 and above 5, respectively."),
          digits = 4, type = table_type, header = FALSE)
```