Econometrics 2 CA1a

Ties van der Veen

1-2-2020

Rebecca Costello, Ties van der Veen, Matei van der Meer

I. Preliminaries

```r
#Clearing the workspace
rm(list = ls())
```

```r
#Installing needed packages
options(repos="https://cran.rstudio.com")
install.packages("interflex")
```

```
## package 'interflex' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpa2JwrQ\downloaded_packages
```

```r
install.packages("plm")
```

```
## package 'plm' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpa2JwrQ\downloaded_packages
```

```r
install.packages("margins")
```

```
## package 'margins' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\Rtmpa2JwrQ\downloaded_packages
```

```r
#Activating libraries
library(foreign)
library(tidyverse)
library(ggdag)
library(dplyr)
library(tinytex)
library(jtools)
library(huxtable)
library(summarytools)
library(ggstance)
library(pwr)
library(knitr)
```

```r
library(lemon)
library(AER)
library(lubridate)
library(ggplot2)
library(interflex)
library(plm)
library(margins)
```

```r
#Improving layout
knit_print.data.frame <- lemon_print
```

```r
#Setting summarytools to R Markdown
st_options(plain.ascii = FALSE, style = "rmarkdown")
st_css()
```

```
## <style type="text/css">
##  img {   background-color: transparent;   border: 0; }  .st-table td, .st-table th {   padding: 8px;
```

II. Treatment heterogeneity in Allcott (2011)

(a)

The boomerang effect mentioned in the paper describes the effect that people who use less energy than the average household and are made aware of this fact might lead increase their energy usage.

(b)

Conditional cooperation can be described as people changing their contribution depending on their information about others' contributions. Thus, if people are informed that they are below/above average energy usage, they might adapt their own energy consumption. This means that low/high energy users (pre-treatment) react differently, leading to a different outcome for them after the treatment. This is by definition heterogeneity of the treatment effect.

(c)

Allcott looks at the Conditional Average Treatment Effects, or CATEs. These are taken for "households in different percentiles of their experiment's distribution of baseline usage". This shows that the more energy a household used before the treatment, the more it saved after the treatment.

(d)

The energy savings that can be reached by implementing the treatment described in this research are on par with what a price increase of 5.2% could save. This is both more effective than a carbon cap and trade system (2.5% savings in 2020, 20% in 2030), and is less controversial. Overall, implementing this treatment seems like a better alternative.

III. Introduction to the computer assignment

```
theUrl_ca1a_ectrics2 <- "https://surfdrive.surf.nl/files/index.php/s/Jh0fYEzZK1JVZAi/download"
waste <- read.dta (file = theUrl_ca1a_ectrics2)
```

IV. Potential outcomes

(a)

Y(0,i) = no change in amount of waste seperated is recorded Y(1,i) = a change in amount of waste seperated is recorded

(b)

1. Repercussions for failing to seperate waste could move people to be more careful when disposing of an item.
2. The second treatment period could show a habit-forming effect. This could happen because of a variety of reasons: people find it easier to seperate waste, are more aware of their impact on the environment, etc.

V. Descriptive statistics

(a)

```
waste_sample <- waste[waste$calendar_week==34, ]
waste_sample %>%
  summarise(mean = mean(waste_sample$residual_weight, na.rm=TRUE), n = n())
```

mean	n
7.89	55

The mean of the main outcome variable, residual weight, is roughly 7.89 in week 34. This was calculated over all routes, as can be seen by the n-value (which indicates the number of routes) being 55. This implies that the average kilos of residual waste per household in this region was about 8 kg.

(b)

```
min(waste$TimeElapsed, na.rm=TRUE)
```

```
## [1] 7
```

```
max(waste$TimeElapsed, na.rm=TRUE)
```

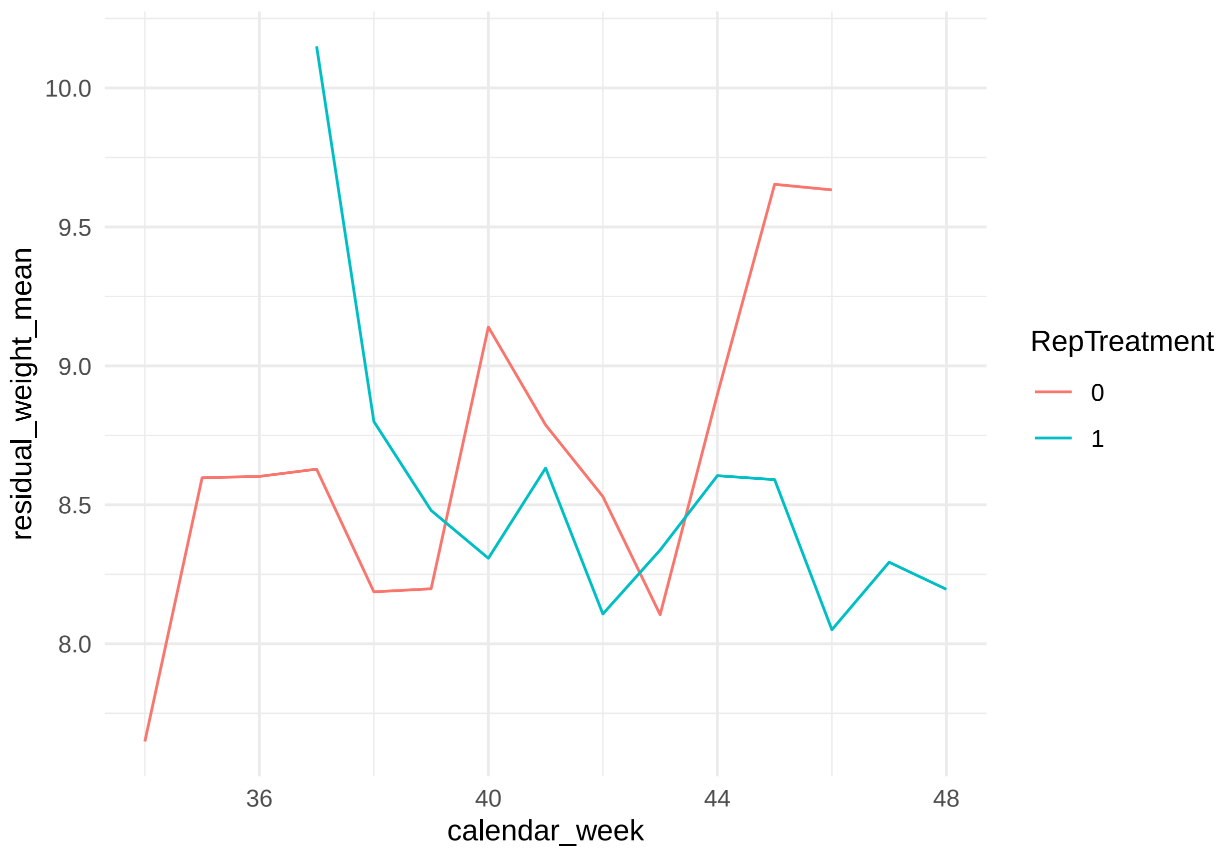
```
## [1] 35
```

This shows that the lowest amount of weeks between treatments was 7 weeks, and the highest was 35 weeks.

(c)

```r
#Creating mean values by treatment/control and calender week

groupmeans <- waste %>%
  filter(TimeElapsed>25) %>%
  group_by(RepTreatment, calendar_week) %>%
  summarise(residual_weight_mean = mean(residual_weight))

#Telling R that RepTreatment is a factor variable

groupmeans$RepTreatment <- as.factor(groupmeans$RepTreatment)

#Creating the line plot

ggplot(groupmeans, aes(x = calendar_week, y = residual_weight_mean, color=RepTreatment)) +
  stat_summary(geom = 'line') +
  theme_minimal()
```

```
## No summary function supplied, defaulting to `mean_se()
```



This graph seems to show overall lower means for garbage for the groups that have had the treatment administered than those who have not. Though this graph is not fully decisive so it seems wrong to draw conclusions this early.

VI. Reporting and interpreting treatment heterogeneity

(a)

```
#Estimating ATE while ignoring serial correlation between observations within routes

ate <- lm(residual_weight ~ RepTreatment + factor(route) + factor(calendar_week), data=waste)
summ(ate)
```

```
## MODEL INFO:
## Observations: 825
## Dependent Variable: residual_weight
## Type: OLS linear regression
##
## MODEL FIT:
## F(69,755) = 16.85, p = 0.00
## R² = 0.61
## Adj. R² = 0.57
##
## Standard errors: OLS
## ------------------------------------------------------------
##                                   Est.   S.E.   t val.      p
## ----------------------------- ------- ------ -------- ------
## (Intercept)                       7.44   0.27    27.42   0.00
## RepTreatment                     -0.22   0.12    -1.85   0.06
## factor(route)103                 -2.24   0.34    -6.51   0.00
## factor(route)104                  2.20   0.34     6.40   0.00
## factor(route)105                 -0.06   0.35    -0.18   0.86
## factor(route)106                  1.83   0.34     5.33   0.00
## factor(route)107                 -0.38   0.34    -1.11   0.27
## factor(route)108                  1.55   0.34     4.50   0.00
## factor(route)109                 -1.68   0.34    -4.89   0.00
## factor(route)110                  0.15   0.34     0.45   0.65
## factor(route)111                  1.70   0.34     4.97   0.00
## factor(route)113                  0.98   0.34     2.85   0.00
## factor(route)202                  0.21   0.34     0.62   0.54
## factor(route)203                  1.92   0.34     5.61   0.00
## factor(route)204                  0.52   0.34     1.51   0.13
## factor(route)205                  0.87   0.34     2.52   0.01
## factor(route)206                 -0.91   0.34    -2.65   0.01
## factor(route)207                  2.03   0.34     5.90   0.00
## factor(route)208                 -1.02   0.34    -2.97   0.00
## factor(route)209                  0.90   0.35     2.61   0.01
## factor(route)210                  2.19   0.34     6.36   0.00
## factor(route)211                 -0.50   0.34    -1.47   0.14
## factor(route)213                  0.75   0.34     2.18   0.03
## factor(route)302                  1.83   0.34     5.34   0.00
## factor(route)303                  1.70   0.34     4.96   0.00
## factor(route)304                  1.38   0.34     4.02   0.00
## factor(route)305                  0.80   0.35     2.32   0.02
## factor(route)306                  1.96   0.34     5.70   0.00
## factor(route)307                  1.57   0.34     4.57   0.00
## factor(route)308                  1.99   0.34     5.82   0.00
```

```
## factor(route)310                   1.66   0.34    4.84   0.00
## factor(route)311                   0.05   0.34    0.15   0.88
## factor(route)312                   0.71   0.34    2.07   0.04
## factor(route)313                  -0.66   0.34   -1.91   0.06
## factor(route)402                  -0.77   0.34   -2.26   0.02
## factor(route)403                   0.65   0.34    1.90   0.06
## factor(route)404                  -1.16   0.34   -3.37   0.00
## factor(route)405                  -0.12   0.34   -0.34   0.73
## factor(route)406                  -0.63   0.34   -1.85   0.07
## factor(route)407                   1.24   0.34    3.61   0.00
## factor(route)408                   0.05   0.35    0.13   0.90
## factor(route)409                   0.65   0.34    1.90   0.06
## factor(route)411                  -0.02   0.34   -0.07   0.94
## factor(route)412                  -1.14   0.34   -3.32   0.00
## factor(route)413                   0.25   0.34    0.74   0.46
## factor(route)502                  -1.43   0.34   -4.15   0.00
## factor(route)503                   0.15   0.34    0.45   0.65
## factor(route)504                   0.27   0.35    0.78   0.43
## factor(route)505                  -0.35   0.34   -1.03   0.30
## factor(route)506                   0.37   0.34    1.09   0.27
## factor(route)507                   0.83   0.34    2.43   0.02
## factor(route)508                   0.07   0.34    0.21   0.83
## factor(route)509                  -0.10   0.34   -0.30   0.77
## factor(route)510                   0.76   0.34    2.21   0.03
## factor(route)511                   0.23   0.34    0.68   0.50
## factor(route)512                   0.60   0.34    1.76   0.08
## factor(calendar_week)35            0.48   0.18    2.71   0.01
## factor(calendar_week)36            1.18   0.18    6.60   0.00
## factor(calendar_week)37            1.09   0.18    6.03   0.00
## factor(calendar_week)38            0.81   0.18    4.48   0.00
## factor(calendar_week)39            0.82   0.18    4.44   0.00
## factor(calendar_week)40            1.33   0.18    7.21   0.00
## factor(calendar_week)41            1.14   0.19    6.00   0.00
## factor(calendar_week)42            0.68   0.19    3.59   0.00
## factor(calendar_week)43            0.93   0.20    4.70   0.00
## factor(calendar_week)44            1.23   0.20    6.21   0.00
## factor(calendar_week)45            1.63   0.21    7.78   0.00
## factor(calendar_week)46            1.18   0.21    5.67   0.00
## factor(calendar_week)47            0.94   0.21    4.39   0.00
## factor(calendar_week)48            0.77   0.21    3.59   0.00
## ------------------------------------------------------------
```

(1) The coefficient for RepTreatment is estimated at -0.22. This tells us that the treatment caused a lower average waste disposal for subjects that were administered the treatment.
(2) Since these values are in kilograms, the reduction is 0.22 or 220 grams. This is a rather small decrease. To see this result in percentages, we divide the estimate by the mean of the outcome variable calculated earlier:

```
-0.22/7.887636
```

```
## [1] -0.02789175
```

In percentages, this is a 0.03% decrease.

```r
#Estimating the ate without ignoring serial correlation

ate1 <- plm(residual_weight ~ RepTreatment, data=waste, effect="twoways",
            model = "within", index=c("route", "calendar_week"))
coeftest(ate1, vcov=vcovHC(ate1, cluster="group"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## RepTreatment -0.21965    0.10217 -2.1498  0.03189 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b)

```r
#Interaction between treatment variable and the time since last treatment

ate2 <- plm(residual_weight ~ RepTreatment*TimeElapsed, data=waste, effect = "twoways",
            model="within", index=c("route", "calendar_week"))
coeftest(ate2, vcov=vcovHC(ate2, cluster="group"))
```

```
##
## t test of coefficients:
##
##                           Estimate Std. Error t value Pr(>|t|)
## RepTreatment              0.311854   0.457120  0.6822   0.4953
## RepTreatment:TimeElapsed -0.023826   0.018592 -1.2816   0.2004
```

```r
#Verifying the ate being similar to the ate under (a)
print(0.31185+(mean(waste$TimeElapsed)*(-0.0238)))
```

```
## [1] -0.2217027
```

This is similar to the -0.22 ate we found in (a).

(c)

At the value of 1 (?)

It does not seem logical to report this CATE since it is insignificant and has a very small effect. (d)

RepTreatment:TimeElapsed is the value for RepTreatment multiplied by the number of weeks it took for the second round of treatment to begin for that route. This does not seem to be very informative since it is a simple multiplication without any other balances.

The statistical significance of this term is 0.2004, meaning that in 20% of repeated tests we would get different results of this estimator. This is not a generally accepted significance level, but could be sizable enough for policy effects.

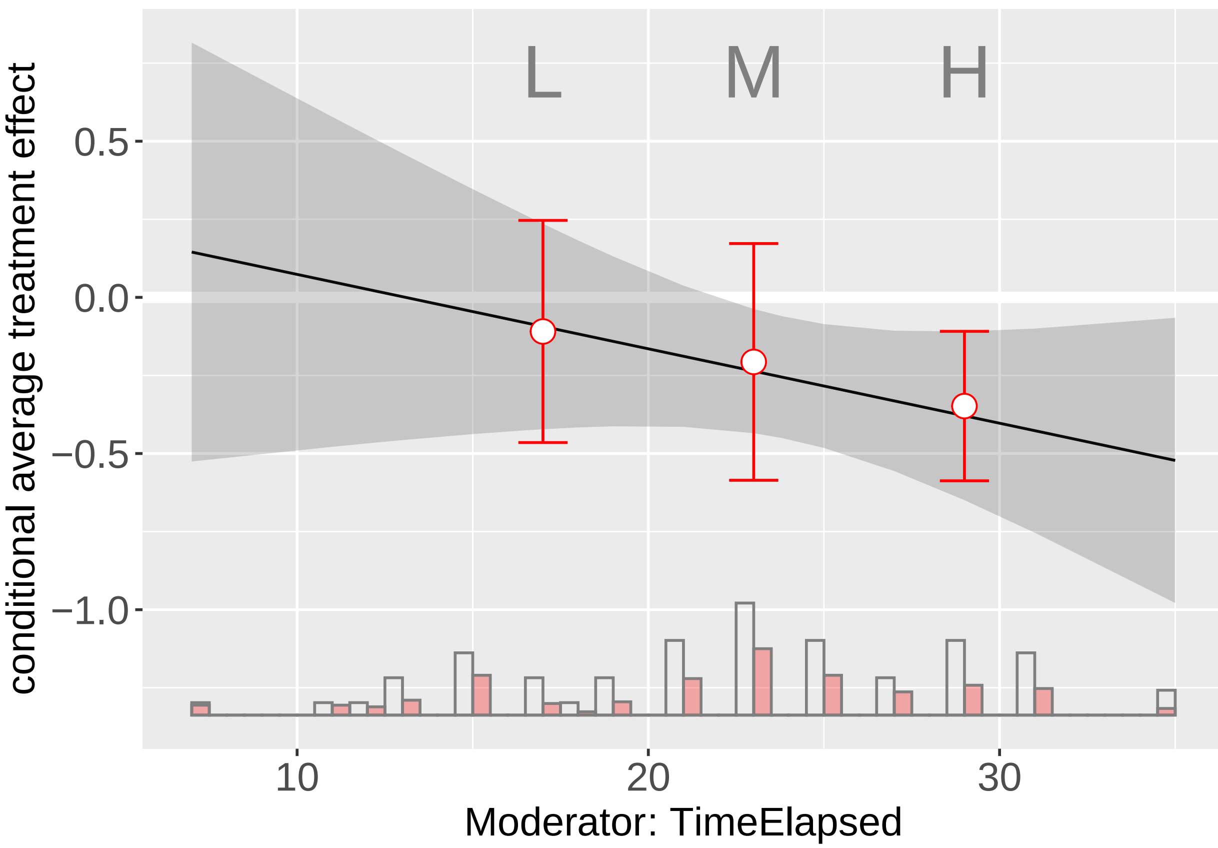
(e)

```r
#Reloading the data

theUrl_ca1a_ectrics2 <- "https://surfdrive.surf.nl/files/index.php/s/Jh0fYEzZK1JVZAi/download"
waste <- read.dta (file = theUrl_ca1a_ectrics2)
```

```r
#Creating the linear marginal effect plot

out <- inter.binning(Y = "residual_weight", D = "RepTreatment", X = "TimeElapsed",
                     Z = NULL, FE = c("route", "calendar_week"), data = waste,
                     vartype = "cluster", cl="route", main=NULL,
                     wald=TRUE, ylab="conditional average treatment effect")
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite

## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite

## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```r
out$graph
```



This plot suggests that the CATE decreases as time elapsed increases. The marginal effects line up because it is forced to take a line-shape by the function.

(f)

The linear approximation seems decent enough, though the bins do show a rather big range which could be an issue.

(g)

No, just correlational.

(h)

The graph tells you taht the CATE is higher for a lower time elapsed, so it would be better to have a short time between the first treatment and the repeated treatment.