Public Economics Assignment

Miia Janatuinen, Ossi Tahvonen

23/03/2020

PART A

1. Read the macro study by Fölster and Henrekson (2001)

A) Summarize the main econometric approach and the findings in the study.

The authors use combined cross-section time series regressions with 5-year periods. They exploit within country variation to increase the efficiency. They use controls for periods and countries to get rid of short term endogeneity problems (caused by business cycles) and to allow for country specific production functions. They also report regressions using 2SLS (i.e. instrumental variables) to control for business cycles. The problem of selection on the OECD countries is controlled by running regressions with additional data set of more countries. The researchers correct for heteroscedasticity between countries and ________????

The paper finds significant (HOW BIG, NOT FOR TAXES) negative effects between government expenditure and economic growth. These effects persist in 2SLS (WHAT WAS THE INSTRUMENT) regressions and additional country data. SAVINGS ROBUSTNESS

The robustness tests show that the effect of government expenditure is more robust than the effect of the taxes. Neither is robust by the strict definition of the extreme bounds analysis.

B) How does the empirical strategy in the paper differ from the micro studies discussed in the class? How credible it is in comparison to studies presented at the course?

The amount of possible endogeneity/omitted variable problems and level of possible extrapolation hugely differs from the micro level studies. In many micro-level studies, the identification strategy is based on some kind of randomized controllized trial -type of situation but this is obviously not possible on macro level. The results of the reported checks are encouraging, as they mostly show no problems, but it is possible that there still are persisting problems in making a causal intrepretation of the estimates (and this is not what the paper suggests, they even admit that "In general, it is hardly possible to solve all econometric problems.").

C) How would you comment on the impact of the size of the public sector on economic growth on the based on its results?

It seems that there is, if the econometric approach is considered viable, a negative impact between the size of the public sector and the economic growth. As noted before, this is questionable. The robustness tests show that the result can be extended to differing combinations of controlling variables. This lends credibility to the result. Of course, as the title of the paper suggests, the results are only from the data in rich countries, and the result can not be extrapolated to developing countries.

There is also need to note that the results of this research only concern the sample of relatively high-income countries. If one were to exrapolate the results to some lower income level countries, it could result in a problematic interpretation. The question can therefore be answered only on behalf the countries that represent the sample. In addition, the paper does not concern the structure of the public expenditure. The result does not therefore mean that all public spending is bad for economomic growth and some expenditure might be beneficial (for instance government actions in coronavirus situation).

2. Read the evaluation of the British Working Families' Tax Credit by Blundell et al. (2005).

A) Summarize the main econometric approach and the findings in the study.

The paper uses a difference-in-difference methodology to study the labour market effects of the Working Family Tax Credit (WFTC) reform that took place in 1999 in the United Kingdom. Families with children are the treatment group and similar families without children are the control group. The effects of the reform are identified from other underlying effects by subtracting the change in employment of the control group from the change in employment of the treatment group. The authors use a probit model which gives estimates of the effect the reforms had on labour market outcomes. The outcome variable is the employment status and the treatment variable is an interaction term between a post-reform indicator variable and a indicator variable on the presence of children. (NOLLAHYPOTEESI?) The number of children, the age of the youngest child, partner's employment status, general economic conditions, seasonal controls, age and education level are controlled for in order to address the possible average differences between groups. In addition, the phase-in and adjustment periods are dropped from the sample.

In addition to the WFTC, there were other contemporaneous reforms that affected the parents and the results give the combined effects of these reforms. The estimated results show a significant positive effect of 3.6 ppt on single mothers' employment, corresponding an estimated 60 000 new workers. The effect was larger with more and younger children and underestimated due to other reforms. For all mothers in couples, there was a statistically insignificant treatment effect of +0.4ppt. There was however a significant effect of 2,6 ppt for those mothers whose partners were not employed. For single fathers, there was a positive and significant effect of 4.6ppt. For the fathers in couples, the treatment effect was statistically significant -0.5ppt and reforms increased employment rates of men whose partners were not working by +0.5ppt and reduced employment rates of men whose partners were working by -1.0ppt (vika omin sanoin + oliko significant?) The robustness checks indicate the existence of additional underlying trends for couples that are not controlled for. Also, time and heterogenous effects are tested and suggest .

B) Compare the approach to the US studies on the EITC discussed in the class.

Both studies study the extensive margin labour market responses due to a tax reform for families. In the US study, the control group are families without children and the treatment group are families with children. The only group that is studied in the US case are single mothers, while in the UK the study is extended to single fathers and parents in couples. Both studies also test different family sizes (katso tarkemmin miten blundellilla oli) and find heterogenous effects. Both studies use a difference-in-difference methodology with similar controls and treatment variables. In addition, Kleven studies multiple reforms.

C) How would you comment on the credibility of the research design? Are there significant threats to its identification strategy?

PART B

1. Read the article.
2. Open data and provide summary statistics similar to those in table 1. (2p.)

We note here that the data provided in the ".dta" format and the data in ".xls" format are different and provide differing results. We were able to obtain the correct regression results using ".dta" -file so we use it throughout the exercise. The main problem seems to be the totrob-variable, which gets values between 0 and 1 in the "dta" -file but is rounded up or down in the ".xls" -file. The picture draw in part 4 is also different with both data, and is more similar to the picture in the Assignment session slides.

```
#packages
library(MASS)
#tibble etc
library(tidyverse)
library(dplyr)
#tables
library(xtable)
```

```
#plots
library(ggplot2)
library(stargazer)
#statadata
library(foreign)
#robust SE:s
library(lmtest)
library(sandwich)
#fixed effects
library(lfe)

#data
data <- read.dta("MonthlyPanel.dta")

#new dummy that has 0 if no jewish institution in same or neighbouring block
data$jewin <- as.numeric(data$institu1|data$institu3)

#table
data2 <- as.tibble(data)
data2 <- data %>% dplyr::select(distanci, edpub, estserv, banco,totrob, jewin)
table <- cbind(apply(data2[data2$jewin==0,],2,mean),apply(data2[data2$jewin==0,],2,sd),
               apply(data2[data2$jewin==1,],2,mean),
               apply(data2[data2$jewin==1,],2,sd))[1:5,]
table <- cbind(table,table[,1]-table[,3])

colnames(table) <- c("Census tracts without Jewish institutions (A)",
                     "SD(A)", "Census tracts with Jewish institutions (B)", "SD(B)",
                     "Difference (C)=(A)-(B)")
table1 <- xtable(table, caption="DEMOGRAPHIC CHARACTERISTICS OF CONTROL AND TREATMENT AREAS")
```

	Census tracts without Jewish institutions (A)	SD(A)	Census tracts with Jewish institutions (B)	SD(B)
distanci	3.66	1.83	0.81	0.39
edpub	0.03	0.17	0.05	0.21
estserv	0.02	0.15	0.02	0.12
banco	0.08	0.27	0.08	0.26
totrob	0.08	0.23	0.08	0.23

Table 1: DEMOGRAPHIC CHARACTERISTICS OF CONTROL AND TREATMENT AREAS

VIKAT SD PUUTTUUUUUUU!!!!!!!!!!!

3. Replicate table 3. (3p.)

The authors use mostly Hubert-White standard errors. In table 5, they also examine the clustering in standard errors. The clustering seems to have no effect. As shown in the exercise session, we use clustered standard errors in regressions A-C and E. The specifications of functions in R differ between packages and are different from Stata, so the results we obtain are different but accurate up to 4 decimals.

In the data set, institu3 variable is 1 also when there is a protected institution in the same block (i.e. when institut1 = 1). Creating new institu3 variable that equals 1 only when institu1 is 0:

```
data$institu3_neww <- data$institu3-data$institu1

table(data$institu1,data$institu3)
```

```
##
```

```
##        0    1
##   0 7458 1771
##   1    0  407
```

```
table(data$institu1,data$institu3_neww)
```

```
##
##        0    1
##   0 7458 1771
##   1  407    0
```

Creating new indicator variables for two-block distance from the nearest protected institution (twoblock) and for time after the terrorist attack (postt):

```
data$twoblock <- ifelse(data$distanci==2,1,0)

data$postt <- ifelse(data$mes>7 & data$mes<70,1,0)
```

Column A: The difference-in-difference estimator is

$$Car\ Theft_{it} = \alpha_0\ Same\ Block\ Police_{it} + M_t + F_i + \varepsilon_{it}$$

where $Car\ Theft_{it}$ are the number of car thefts in block i for month i, $ Same ; Block ; Police_{it} $ is the interaction term between institu1 and postt (=the treatment variable), M_t are the month fixed effects, F_t are block fixed effects and ε_{it} is the error term. Since block fixed effects are used, we use clustered standard errors.

Creating the treatment variables and dropping months 72 and 73:

```
data$SameBlock <- data$institu1*data$postt
data$OneBlock <- data$institu3_neww*data$postt
data$TwoBlock <- data$twoblock*data$postt
data <- data[data$mes < 70,]
```

Fixed effect regression (mes and observ as the fixed effects, clustering wrt. observ):

```
m3a <- felm(data=data, totrob ~ SameBlock | mes + observ | 0 | observ)
summary(m3a)
```

```
##
## Call:
##    felm(formula = totrob ~ SameBlock | mes + observ | 0 | observ,      data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67021 -0.09455 -0.03133  0.01656  2.14999
##
## Coefficients:
##           Estimate Cluster s.e. t value Pr(>|t|)
## SameBlock -0.07753      0.02350    -3.3  0.00101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.23 on 6999 degrees of freedom
## Multiple R-squared(full model): 0.1983   Adjusted R-squared: 0.09707
## Multiple R-squared(proj model): 0.001277   Adjusted R-squared: -0.1249
## F-statistic(full model, *iid*):1.959 on 884 and 6999 DF, p-value: < 2.2e-16
## F-statistic(proj model): 10.89 on 1 and 875 DF, p-value: 0.001007
```

Column B: The difference-in-difference estimator is like in A, but now also includes a treatment variable for one-block distance (One-Block Police = institu3_neww*postt):

$$Car\ Theft_{it} = \alpha_0\ Same\ Block\ Police_{it} + \alpha_1 One\ Block\ Police_{it} + M_t + F_i + \varepsilon_{it}$$

```r
m3b <- felm(data=data, totrob ~ SameBlock + OneBlock | mes + observ | 0 | observ)
summary(m3b)
```

```
##
## Call:
##    felm(formula = totrob ~ SameBlock + OneBlock | mes + observ |      0 | observ, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67135 -0.09356 -0.03063  0.01568  2.14886
##
## Coefficients:
##           Estimate Cluster s.e. t value Pr(>|t|)
## SameBlock -0.08007      0.02360  -3.393 0.000723 ***
## OneBlock  -0.01326      0.01456  -0.911 0.362575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.23 on 6998 degrees of freedom
## Multiple R-squared(full model): 0.1984   Adjusted R-squared: 0.09707
## Multiple R-squared(proj model): 0.001414   Adjusted R-squared: -0.1249
## F-statistic(full model, *iid*):1.958 on 885 and 6998 DF, p-value: < 2.2e-16
## F-statistic(proj model): 5.927 on 2 and 875 DF, p-value: 0.002776
```

Column C: Like in B, but adding a treatment variable for Two-Block Police (=twoblock*postt):

$$Car\ Theft_{it} = \alpha_0\ Same\ Block\ Police_{it} + \alpha_1 One\ Block\ Police_{it} + \alpha_2 Two\ Block\ Police_{it} + M_t + F_i + \varepsilon_{it}$$

```r
m3c <- felm(data=data, totrob ~ SameBlock + OneBlock + TwoBlock | mes + observ | 0 | observ)
summary(m3c)
```

```
##
## Call:
##    felm(formula = totrob ~ SameBlock + OneBlock + TwoBlock | mes +      observ | 0 | observ, data = d
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.67070 -0.09394 -0.03023  0.01560  2.14854
##
## Coefficients:
##            Estimate Cluster s.e. t value Pr(>|t|)
## SameBlock -0.080802     0.023928  -3.377 0.000765 ***
## OneBlock  -0.013988     0.015078  -0.928 0.353818
## TwoBlock  -0.002185     0.012384  -0.176 0.860011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.23 on 6997 degrees of freedom
```

```
## Multiple R-squared(full model): 0.1984   Adjusted R-squared: 0.09694
## Multiple R-squared(proj model): 0.001419   Adjusted R-squared: -0.125
## F-statistic(full model, *iid*):1.955 on 886 and 6997 DF, p-value: < 2.2e-16
## F-statistic(proj model): 3.965 on 3 and 875 DF, p-value: 0.008015
```

Column D: The cross section regression uses data only from after the terrorist attack and doesn't include block-fixed effects:

$$Car\ Theft_{it} = \beta_0 + \alpha_0\ Same\ Block\ Police_{it} + \alpha_1 One\ Block\ Police_{it} + \alpha_2 Two\ Block\ Police_{it} + M_t + \varepsilon_{it}$$

Now $Same\ Block\ Police_{it}$, $One\ Block\ Police_{it}$ and $Two\ Block\ Police_{it}$ have not been interacted with the post variable, since data is only from when postt=1. We use robust standard errors.

```r
data_post <- data[data$mes>7,]
data_post$SameBlock <- data_post$institu1
data_post$OneBlock <- data_post$institu3_neww
data_post$TwoBlock <- data_post$twoblock
model3d <- lm(data = data_post, formula = totrob ~ SameBlock + OneBlock + TwoBlock +
                as.factor(mes))
summary(model3d)
```

```
##
## Call:
## lm(formula = totrob ~ SameBlock + OneBlock + TwoBlock + as.factor(mes),
##     data = data_post)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11425 -0.10854 -0.10112 -0.04153  2.38575
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.114247   0.009442  12.099  < 2e-16 ***
## SameBlock        -0.072719   0.019611  -3.708 0.000211 ***
## OneBlock         -0.011581   0.010526  -1.100 0.271306
## TwoBlock         -0.003429   0.009343  -0.367 0.713624
## as.factor(mes)9  -0.013128   0.012254  -1.071 0.284072
## as.factor(mes)10 -0.002283   0.012254  -0.186 0.852201
## as.factor(mes)11 -0.010845   0.012254  -0.885 0.376191
## as.factor(mes)12 -0.005708   0.012254  -0.466 0.641379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2564 on 4372 degrees of freedom
## Multiple R-squared:  0.003623,   Adjusted R-squared:  0.002028
## F-statistic: 2.271 on 7 and 4372 DF,  p-value: 0.02626
```

Robust standard errors:

```r
coeftest(model3d, vcov= vcovHC(model3d, "HC1"))
```

```
##
## t test of coefficients:
##
##                    Estimate Std. Error t value  Pr(>|t|)
## (Intercept)       0.1142467  0.0102112 11.1883 < 2.2e-16 ***
```

```
## SameBlock        -0.0727188  0.0113886 -6.3852  1.89e-10 ***
## OneBlock         -0.0115807  0.0109013 -1.0623    0.2881
## TwoBlock         -0.0034292  0.0092500 -0.3707    0.7109
## as.factor(mes)9  -0.0131279  0.0126751 -1.0357    0.3004
## as.factor(mes)10 -0.0022831  0.0127582 -0.1790    0.8580
## as.factor(mes)11 -0.0108447  0.0123883 -0.8754    0.3814
## as.factor(mes)12 -0.0057078  0.0127109 -0.4490    0.6534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Column E: The time-series regression only uses data from the Jewish blocks (up to two block distance):

$$Car\ Theft_{it} = \beta_0 + \alpha_0\ Same\ Block\ Police_{it} + \alpha_1 One\ Block\ Police_{it} + \alpha_2 Two\ Block\ Police_{it} + F_i + \varepsilon_{it}$$

Also the different lengths of the months are taken into account in the new totrob2 variable:

```
data$totrobc <- ifelse(data$mes==7, data$totrob*(30/17), data$totrob)
data$totrob2 <- ifelse(data$mes==5|data$mes==8|data$mes==10|data$mes==12,
                        data$totrob*(30/31), data$totrobc)
```

Regression using clustered standard errors:

```
data_close <- data[data$distanci<3,]
m3e <- felm(data = data_close, totrob2 ~ SameBlock+OneBlock+TwoBlock | observ | 0 | observ)
summary(m3e)
```

```
##
## Call:
##    felm(formula = totrob2 ~ SameBlock + OneBlock + TwoBlock | observ |      0 | observ, data = data_c
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74844 -0.08883 -0.02776  0.00002  1.77780
##
## Coefficients:
##             Estimate Cluster s.e. t value Pr(>|t|)
## SameBlock -5.843e-02    2.342e-02  -2.495    0.013 *
## OneBlock  -4.538e-05    1.464e-02  -0.003    0.998
## TwoBlock   1.701e-02    1.074e-02   1.585    0.114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2396 on 3389 degrees of freedom
## Multiple R-squared(full model): 0.1891   Adjusted R-squared: 0.08719
## Multiple R-squared(proj model): 0.002186   Adjusted R-squared: -0.1232
## F-statistic(full model, *iid*):1.855 on 426 and 3389 DF, p-value: < 2.2e-16
## F-statistic(proj model): 2.912 on 3 and 423 DF, p-value: 0.03422
```

Collecting the results:

```
stargazer(m3a,m3b, m3c, model3d, m3e, column.separate=c(3,1,1),
          column.labels=c("Difference-in-difference","Cross section"," Time series"),
          dep.var.labels.include = FALSE, dep.var.caption = "", model.names = FALSE,
          summary.stat = c("sd"),  title = "THE EFFECT OF POLICE PRESENCE ON CAR THEFT",
          keep = c("SameBlock", "OneBlock","TwoBlock"), label = "3",
          keep.stat = c("n", "rsq") )
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Mon, May 04, 2020 - 11:41:53
## \begin{table}[!htbp] \centering
##   \caption{THE EFFECT OF POLICE PRESENCE ON CAR THEFT}
##   \label{3}
## \begin{tabular}{@{\extracolsep{5pt}}lccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{3}{c}{Difference-in-difference} & Cross section &  Time series \\
## \\[-1.8ex] & (1) & (2) & (3) & (4) & (5)\\
## \hline \\[-1.8ex]
##  SameBlock & $-$0.078$^{***}$ & $-$0.080$^{***}$ & $-$0.081$^{***}$ & $-$0.073$^{***}$ & $-$0.058$^{
##   & (0.023) & (0.024) & (0.024) & (0.020) & (0.023) \\
##   & & & & & \\
##  OneBlock &  & $-$0.013 & $-$0.014 & $-$0.012 & $-$0.00005 \\
##   &  & (0.015) & (0.015) & (0.011) & (0.015) \\
##   & & & & & \\
##  TwoBlock &  &  & $-$0.002 & $-$0.003 & 0.017 \\
##   &  &  & (0.012) & (0.009) & (0.011) \\
##   & & & & & \\
## \hline \\[-1.8ex]
## Observations & 7,884 & 7,884 & 7,884 & 4,380 & 3,816 \\
## R$^{2}$ & 0.198 & 0.198 & 0.198 & 0.004 & 0.189 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{5}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

(fixing the robust standard errors for column D by hand)

Table 2: THE EFFECT OF POLICE PRESENCE ON CAR THEFT

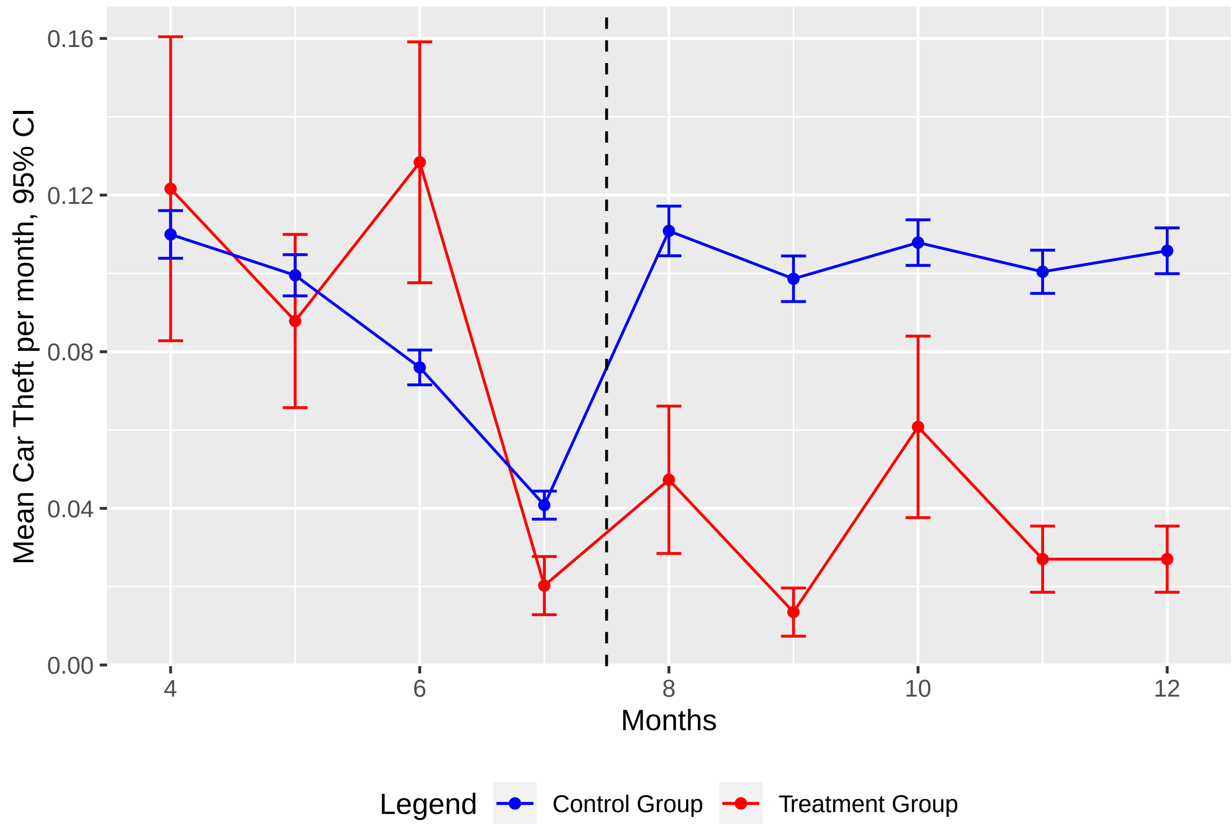
	Difference-in-difference			Cross section	Time series
	(1)	(2)	(3)	(4)	(5)
SameBlock	-0.078^{***}	-0.080^{***}	-0.081^{***}	-0.073^{***}	-0.058^{**}
	(0.023)	(0.024)	(0.024)	(0.011)	(0.023)
OneBlock		-0.013	-0.014	-0.012	-0.00005
		(0.015)	(0.015)	(0.011)	(0.015)
TwoBlock			-0.002	-0.003	0.017
			(0.012)	(0.009)	(0.011)
Observations	7,884	7,884	7,884	4,380	3,816
R^2	0.198	0.198	0.198	0.004	0.189
Note:					$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4. Plot a figure of choice describing the DID-effects. (2p.)

```
datamean_treat <- data[data$institu1==1,] %>% as.tibble() %>%
  group_by(mes) %>%
```

```
  summarise_at(vars(totrob),
               list(name2 = mean, sdev2 = sd))

datamean_contr <- data[data$institu1==0,] %>% as.tibble() %>%
  group_by(mes) %>%
  summarise_at(vars(totrob),
               list(name2 = mean, sdev2 = sd))
datamean_treat

datamean <- cbind(as.data.frame(datamean_contr),as.data.frame(datamean_treat[,2:3]))[1:9,]
datamean <-  cbind(datamean, datamean_contr[1:9,2] - datamean_treat[1:9,2])
#the correct values for the standard errors are sd/sqrt(n)

n_contr <- nrow(data[data$institu1==0,])
n_treat <- nrow(data[data$institu1==1,])

colnames(datamean) <- c("mes","name2", "sdev2", "name", "sdev", "dif")


colors1 <- c("Treatment Group" = "red", "Control Group" = "blue")

p1 <- ggplot(data = datamean, aes(x=mes)) +
  geom_point(aes(x=mes, y = name, color = "Treatment Group")) + geom_line(aes(x=mes, y = name, color =
  geom_point(aes(x = mes, y = name2, colour = "Control Group")) +
  geom_line(data = datamean, aes(x = mes, y = name2, color = "Control Group")) +
  geom_errorbar(data = datamean, aes(ymin = name - 1.96*sdev/sqrt(n_treat), ymax = name + 1.96*sdev/sqr
  geom_errorbar(data = datamean, aes(ymin = name2 - 1.96*sdev2/sqrt(n_contr), ymax = name2 + 1.96*sdev2/
  geom_vline(xintercept = 7.5, linetype = 2) +
  xlab("Months") + ylab("Mean crime") +
  theme(legend.position="bottom") +
  labs(x = "Months",
       y = "Mean Car Theft per month, 95% CI",
       color = "Legend") +
  scale_color_manual(values = colors1)
```

```
p1
```



5. Discuss the results. (3p.)

a. What do the estimation results in table 3 tell?

There are statistically significant effects for the Police on the same block across all specifications.

Instead, the coefficients for the other police presence variables are not statistically significantly different from zero in any regressions.

b. What are the potential concerns in the research set up?

THINGS TO ADDRESS

-data

-regression E

Similar conlusions can be reached with different regressions, ie. time series and cross-sectional data.

Using differences in differences estimation, the most important threat to estimation is violation if the parallel trends assumption. This is required for the control group to provide a credible counterfactual to treatment. There is no statistical test for this assumption. It can be assessed visually from graphs like Picture 1. The picture however only shows the means and not the controls, namely block and month dummies, used in regressions. The trends seem to be somewhat parellel before the treatment. Table 4 of the paper also presents estimates for different time periods to check for this assumption, and finds that there are effects even if the treatment is started later.

The authors also explore the possibility that there were other simultaneous treatments in addition to the increase in police presence, such as parking restrictions. They present evidence that this alone can not explain the change in car thefts. They also examine the chance that people (drivers or criminals) adjusted their behaviour because of fear and this caused the reduction in thefts instead of the police presence. They find no evidence of short effects like this, which still leaves the possibility of the effect being longer than the data.

c. What do we learn from the paper? Discuss the internal and external validity of the study.

The internal validity is addressed in the previous question and somewhat in the other parts of the exercise. We find it to be adequate.

On the external validity, some questions can be expressed. The authors examine effects of car price, time of day and day of the week. There are still effects, which lends credibility to idea that the results could be extrapolated. However, the neighborhoods considered were high-middle income areas and represents a small part of the city, with most thefts concentrating on low-income areas.

Important question regarding the validity is the chance that crime simply switched neighborhoods. There are no significant effects on the neighboring blocks, where the crime could have moved, as shown in table 3. Other than this, the data does not allow us to learn about the displacement of crime.

References