Appendix

Eline van Groningen, Paola Priante, Valery Maasdamme, Yuhu Wang

9/25/2020

Difference-in-Difference Analysis: Female Labor Participation

```r
#Downloading the libraries, setting the working directory and importing the data set
library(tidyverse)
library(stargazer)
library(dagitty)
library(gridExtra)
library(tinytex)
library(ggplot2)
library(tidyr)
library(dplyr)
library(plyr)
library(reshape2)

dir <- "/Users/valeriemaasdamme/Documents/BAM_ASP_A2"
dirProg <- paste0(dir, "/programs/")
dirData <- paste0(dir, "/Data/")

dfDiD <- read.csv(file=paste0(dirData, "DiD_dataset.csv"))
```

Preparing and analyzing the dataset

```r
# no need to transform the dataset, already in the long format
str(dfDiD) # all variables are numeric or integer, no need to tranform

dfDiD$dPeriod = ifelse(dfDiD$year >= 1993, 1, 0) # dummy variable for period
dfDiD$cChildren = ifelse(dfDiD$children >= 1, 1, 0) # dummy for different groups

dfDiD.sub <- subset(dfDiD, work=="1") #creating a subset of employed women
```

1 Plotting the dependent variables

```r
#Earn
#6 years for both groups, total of 12 averages (average by year and children (0/1)
earn.agg = aggregate(dfDiD.sub$earn, list(dfDiD.sub$year, dfDiD.sub$cChildren == 1),
```

```
                        FUN = mean, na.rm = TRUE)
names(earn.agg) = c("Year","Children","Earn") #rename variables
#new variable with group name
earn.agg$Group[1:6] = "Women without children"
earn.agg$Group[7:12] = "Women with children"

Earn.plot <- qplot(Year, Earn, data=earn.agg, geom=c("point","line"),
  colour = Group,
  xlab="Year", ylab="Annual earnings") +
  geom_vline(xintercept = 1993) +
  theme_bw()
ggsave(file="Earn.pdf", width=7, height=4)


#Finc
finc.agg = aggregate(dfDiD.sub$finc, list(dfDiD.sub$year, dfDiD.sub$cChildren == 1),
                     FUN = mean, na.rm = TRUE)
names(finc.agg) = c("Year","Children","Finc")
finc.agg$Group[1:6] = "Women without children"
finc.agg$Group[7:12] = "Women with children"

Finc.plot <- qplot(Year, Finc, data=finc.agg, geom=c("point","line"),
  colour = Group,
  xlab="Year", ylab="Annual Family Income") +
  geom_vline(xintercept = 1993) +
  theme_bw()
ggsave(file="Finc.pdf", width=7, height=4)


#Work
work.agg = aggregate(dfDiD$work, list(dfDiD$year, dfDiD$cChildren == 1),
                     FUN = mean, na.rm = TRUE)
names(work.agg) = c("Year","Children","Work")

work.agg$Group[1:6] = "Women without children"
work.agg$Group[7:12] = "Women with children"

Work.plot <- qplot(Year, Work, data=work.agg, geom=c("point","line"),
  colour = Group,
  xlab="Year", ylab="Work")+
  geom_vline(xintercept = 1993) +
  theme_bw()
ggsave(file="Work.pdf", width=7, height=4)
```

2 Summary statistics of the dataset

```
stargazer(dfDiD, type = "text")
stargazer(dfDiD[, c("children", "finc", "earn", "age", "work", "unearn")], type = "text")
```

3 Difference-in-Difference

```r
# creating averages per group per period
avgEarn <- ddply (dfDiD.sub, .(dPeriod, cChildren), summarise,
                  avgEarn = mean(earn, na.rm=TRUE))

avgFinc <- ddply (dfDiD.sub, .(dPeriod, cChildren), summarise,
                  avgFinc = mean(finc, na.rm=TRUE))

avgWork <- ddply (dfDiD, .(dPeriod, cChildren), summarise,
                  avgWork = mean(work, na.rm=TRUE))

#Remodel the avg table from long to wide, add  row for the difference in averages
avgtable.Earn <- dcast (avgEarn, dPeriod ~ cChildren, value.var = "avgEarn")
avgtable.Earn <- rbind(avgtable.Earn, avgtable.Earn[2,]-avgtable.Earn[1,])
rownames(avgtable.Earn) <- c("Before", "After", "Difference") # renaming the rows
colnames(avgtable.Earn) <- c("dPeriod", "Women without children (0)",
                             "Women with children (1)") # renaming the columns
avgtable.Earn[3, "dPeriod"] <- NA

avgtable.Finc <- dcast (avgFinc, dPeriod ~ cChildren, value.var = "avgFinc")
avgtable.Finc <- rbind(avgtable.Finc, avgtable.Finc[2,]-avgtable.Finc[1,])
rownames(avgtable.Finc) <- c("Before", "After", "Difference")
colnames(avgtable.Finc) <- c("dPeriod", "Women without children (0)",
                             "Women with children (1)")
avgtable.Finc[3, "dPeriod"] <- NA

avgtable.Work <- dcast (avgWork, dPeriod ~ cChildren, value.var = "avgWork")
avgtable.Work <- rbind(avgtable.Work, avgtable.Work[2,]-avgtable.Work[1,])
rownames(avgtable.Work) <- c("Before", "After", "Difference")
colnames(avgtable.Work) <- c("dPeriod", "Women without children (0)",
                             "Women with children (1)")
avgtable.Work[3, "dPeriod"] <- NA

stargazer(avgtable.Earn, summary=FALSE, align = TRUE, type="text",
          title = "Average Annual Earnings")
stargazer(avgtable.Finc, summary=FALSE, align = TRUE, type="text",
          title = "Average Indicator Annual Family Income")
stargazer(avgtable.Work, summary=FALSE, align = TRUE, type="text",
          title = "Average Indicator Work Status")
```

4 Regression analysis

```r
mdlEarn <- earn ~ cChildren + dPeriod + cChildren:dPeriod
rsltOLSEarn <- lm(mdlEarn, data=dfDiD.sub)

mdlFinc <- finc ~ cChildren + dPeriod + cChildren:dPeriod
rsltOLSFinc <- lm(mdlFinc, data=dfDiD.sub)

mdlWork <- work ~ cChildren + dPeriod + cChildren:dPeriod
rsltOLSWork <- lm(mdlWork, data=dfDiD)
```

```
stargazer(rsltOLSEarn, rsltOLSFinc, rsltOLSWork,
          intercept.bottom = FALSE, align = TRUE, no.space=TRUE,
          type="text")
```

Control variables

```
# adding urate, unearn and children as control variables
# Earn
mdl.control.earn <- earn ~ cChildren + dPeriod + cChildren:dPeriod +
  urate + unearn + children
rsltOLS.control.earn <- lm(mdl.control.earn, data=dfDiD.sub)

# Finc
mdl.control.finc <- finc ~ cChildren + dPeriod + cChildren:dPeriod +
  urate + unearn + children
rsltOLS.control.finc <- lm(mdl.control.finc, data=dfDiD.sub)

# Work
mdl.control.work <- work ~ cChildren + dPeriod + cChildren:dPeriod +
  urate + unearn + children
rsltOLS.control.work <- lm(mdl.control.work, data=dfDiD)

stargazer(rsltOLS.control.earn, rsltOLS.control.finc,
          rsltOLS.control.work,
          intercept.bottom = FALSE,
          align = TRUE,
          no.space=TRUE, type="text")
```

Robust standard errors