Appendix Difference-in-Difference Analysis: Female Labor Participation

Cleaning data

```r
# Load libraries
library(tidyverse)
library(stargazer)
library(dagitty)
library(gridExtra)
library(tinytex)
library(ggplot2)
library(tidyr)
library(dplyr)
library(plyr)
library(reshape2)
library(sandwich)

# Set directories
dir <- "/Users/valeriemaasdamme/Desktop/BAM_ASP_A2"
dirProg <- paste0(dir, "/programs/")
dirData <- paste0(dir, "/data/")
dirResults <- paste0(dir, "results/")

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

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

1 Visual evidence

```r
## Variable == earn
# Compute group average for both women with and without children
earn.agg = aggregate(dfDiD$earn, list(dfDiD$year, dfDiD$cChildren == 1),
                     FUN = mean, na.rm = TRUE)

# Rename column names
names(earn.agg) = c("Year","Children","Earn")

# New variable with group name
earn.agg$Group[1:6] = "Women without children"
earn.agg$Group[7:12] = "Women with children"
```

```r
# Make and save plot
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(base_size = 15) +
  theme(axis.title = element_text(size = 19))
Earn.plot
```

```r
ggsave(file="Earn.pdf", width=7, height=4)
```

```r
# Variable == finc
# Compute group average for both women with and without children
finc.agg = aggregate(dfDiD$finc, list(dfDiD$year, dfDiD$cChildren == 1),
                     FUN = mean, na.rm = TRUE)
names(finc.agg) = c("Year","Children","Finc")

# New variable with group name
finc.agg$Group[1:6] = "Women without children"
finc.agg$Group[7:12] = "Women with children"

# Make and save plot
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(base_size = 15) +
  theme(axis.title = element_text(size = 19))
Finc.plot
```

```r
ggsave(file="Finc.pdf", width=7, height=4)
```

```r
# Variable == work
# Compute group average for both women with and without children
work.agg = aggregate(dfDiD$work, list(dfDiD$year, dfDiD$cChildren == 1),
                     FUN = mean, na.rm = TRUE)
names(work.agg) = c("Year","Children","Work")

# New variable with group name
work.agg$Group[1:6] = "Women without children"
work.agg$Group[7:12] = "Women with children"

# Make and save plot
Work.plot <- qplot(Year, Work, data=work.agg, geom=c("point", "line"),
                   colour = Group, xlab="Year", ylab="Work") +
  geom_line() +
  geom_vline(xintercept = 1993) +
  theme_bw(base_size = 15) +
  theme(axis.title = element_text(size = 19))
Work.plot
```

```r
ggsave(file="Work.pdf", width=7, height=4)
```

2 Summary statistics of the data

```r
# Convert categorical variables to vector
dfDiD$year <- as.factor(dfDiD$year)
dfDiD$state <- as.factor(dfDiD$state)
# dfDiD$nonwhite <- as.factor(dfDiD$nonwhite)
# dfDiD$dWork <- as.factor(dfDiD$work)
# dfDiD$cChildren = as.factor(dfDiD$cChildren)
# dfDiD$dPeriod <- as.factor(dfDiD$dPeriod)

# Check for converting categorical variables
str(dfDiD)

# Tabulate summary statistics
stargazer(dfDiD, title="Summary Statistics")

# Generate frequency tables categorical variables
year_freq <- as.data.frame(table(dfDiD$year))
state_freq <- as.data.frame(table(dfDiD$state))
nonwhite_freq <- as.data.frame(table(dfDiD$nonwhite))
work_freq <- as.data.frame(table(dfDiD$work))
dPeriod_freq <- as.data.frame(table(dfDiD$dPeriod))
cChildren_freq <- as.data.frame(table(dfDiD$cChildren))

# Tabulate frequency tables with stargazer
stargazer(year_freq, summary=FALSE, title="year")
stargazer(state_freq, summary=FALSE, title="state")
stargazer(nonwhite_freq, summary=FALSE, title="nonwhite")
stargazer(work_freq, summary=FALSE, title="work")
stargazer(dPeriod_freq, summary=FALSE, title="period")
stargazer(cChildren_freq, summary=FALSE, title="cChildren")
```

3 Difference in difference effect

```r
avgEarn <- ddply(dfDiD, .(dPeriod, cChildren), summarise,
                 avgEarn = mean(earn, na.rm=TRUE))

avgFinc <- ddply(dfDiD, .(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  additional row for the
# difference in averages and rename the rows
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")
colnames(avgtable.Earn) <- c("dPeriod", "Women without children (0)",
                             "Women with children (1)")
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

# Tabulate DiD tables with stargazer
stargazer(avgtable.Earn, summary=FALSE, align = TRUE,
          title = "Average Annual Earnings")
stargazer(avgtable.Finc, summary=FALSE, align = TRUE,
          title = "Average Indicator Annual Family Income")
stargazer(avgtable.Work, summary=FALSE, align = TRUE,
          title = "Average Indicator Work Status")
```

4 Regression analysis

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

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

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

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

4.1 Regression with control variables

```
# DiD regression earn with control variables
mdl.control.earn <- earn ~ cChildren + dPeriod + cChildren:dPeriod + urate +
  unearn + ed + nonwhite + children
rsltOLS.control.earn <- lm(mdl.control.earn, data=dfDiD)

# DiD regression earn with control variables
mdl.control.finc <- finc ~ cChildren + dPeriod + cChildren:dPeriod + urate +
  unearn + ed + nonwhite + children
rsltOLS.control.finc <- lm(mdl.control.finc, data=dfDiD)

# DiD regression earn with control variables
```

```
mdl.control.work <- work ~ cChildren + dPeriod + cChildren:dPeriod + urate +
  unearn + ed + nonwhite + 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")

stargazer(rsltOLS.control.earn, rsltOLS.control.finc, rsltOLS.control.work,
          intercept.bottom = FALSE, align = TRUE, no.space=TRUE,
        title="DiD regression for earn, finc and work with control variables")
```

4.2 Regression with robust standard errors

```
#Plots
ggplot(data = data.frame(fit = fitted(rsltOLS.control.earn),
  rsid = residuals(rsltOLS.control.earn)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")

ggplot(data = data.frame(fit = fitted(rsltOLS.control.finc),
  rsid = residuals(rsltOLS.control.finc)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")

ggplot(data = data.frame(fit = fitted(rsltOLS.control.work),
  rsid = residuals(rsltOLS.control.work)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")

# BP test for the three models
lmtest::bptest(rsltOLS.control.earn)
lmtest::bptest(rsltOLS.control.finc)
lmtest::bptest(rsltOLS.control.work)
```