Appendix 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)
library(sandwich)

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"
```

```r
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

```r
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

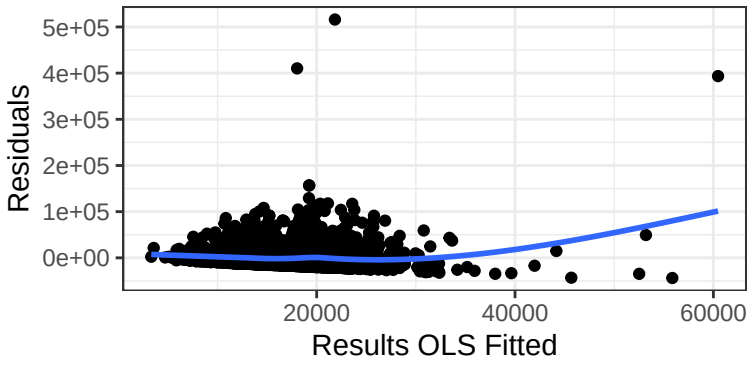
```
mdlEarn <- earn ~ cChildren + dPeriod + cChildren:dPeriod
rsltOLSEarn <- lm(mdlEarn, data=dfDiD.sub)

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

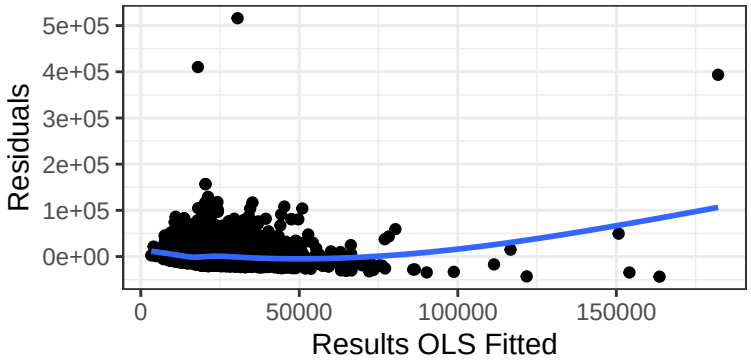
```
# adding urate, unearn and children as control variables
# Earn
mdl.control.earn <- earn ~ cChildren + dPeriod + cChildren:dPeriod +
  urate + unearn + children
rsltOLS.control.earn <- lm(mdl.control.earn, data=dfDiD.sub)

# 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

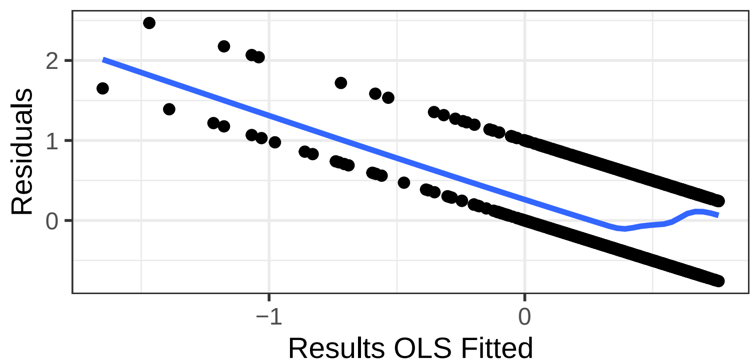
```
#Test for heteroskedasticity
rsltOLS.control.earn2 <- lm(mdl.control.earn, data=dfDiD.sub)
rsltOLS.control.finc2 <- lm(mdl.control.finc, data=dfDiD.sub)
rsltOLS.control.work2 <- lm(mdl.control.work, data=dfDiD)

# EARN
ggplot(data = data.frame(fit = fitted(rsltOLS.control.earn2),
  rsid = residuals(rsltOLS.control.earn2)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")
```




```
lmtest::bptest(rsltOLS.control.earn2)
# p < 0.01, heteroskedastiscity is detected.

#FINC
ggplot(data = data.frame(fit = fitted(rsltOLS.control.finc2),
  rsid = residuals(rsltOLS.control.finc2)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")
```




```
lmtest::bptest(rsltOLS.control.finc2)
# p < 0.01, heteroskedastiscity is detected.

#WORK
ggplot(data = data.frame(fit = fitted(rsltOLS.control.work2),
  rsid = residuals(rsltOLS.control.work2)),
  aes(fit, rsid)) +
  geom_point() +
  stat_smooth(se = F) +
  theme_bw() +
  labs(x = "Results OLS Fitted") +
  labs(y = "Residuals")
```




```r
lmtest::bptest(rsltOLS.control.work2)
# p < 0.01, heteroskedastiscity is detected
```

```r
#Standard  errors
seBasicEarn  <- sqrt(diag(vcov(rsltOLS.control.earn2)))
seWhiteEarn  <- sqrt(diag(vcovHC(rsltOLS.control.earn2, type="HC0")))
seClusterEarn <- sqrt(diag(vcov(rsltOLS.control.earn2, cluster="state")))
stargazer(rsltOLS.control.earn2, rsltOLS.control.earn2, rsltOLS.control.earn2,
          se=list(seBasicEarn, seWhiteEarn, seClusterEarn), type="text")
```

```
=========================================================================
                                          Dependent variable:
                                -----------------------------------------
                                                 earn
                                     (1)           (2)           (3)
-------------------------------------------------------------------------
cChildren                       -4,734.562*** -4,734.562*** -4,734.562***
                                  (945.280)     (835.140)     (945.280)

dPeriod                           -567.581      -567.581      -567.581
                                  (696.500)     (799.615)     (696.500)

urate                            429.112***    429.112***    429.112***
                                  (161.063)     (146.978)     (161.063)

unearn                           339.835***     339.835*     339.835***
                                  (37.979)      (192.934)     (37.979)

children                        -1,505.236*** -1,505.236*** -1,505.236***
                                  (301.752)     (264.835)     (301.752)

cChildren:dPeriod                 1,490.156     1,490.156     1,490.156
                                  (945.477)     (916.133)     (945.477)

Constant                        16,363.220*** 16,363.220*** 16,363.220***
                                 (1,285.565)   (1,383.430)   (1,285.565)

-------------------------------------------------------------------------
Observations                        7,052         7,052         7,052
```

```
R2                                  0.043         0.043         0.043
Adjusted R2                         0.042         0.042         0.042
Residual Std. Error (df = 7045)  19,061.410    19,061.410    19,061.410
F Statistic (df = 6; 7045)        52.984***     52.984***     52.984***
=========================================================================
Note:                                         *p<0.1; **p<0.05; ***p<0.01
```

```r
#No impact on the significance of the DiD effect, still insignificant
#Standard error of seWhite seems smaller than basic and clustered

seBasicFinc  <- sqrt(diag(vcov(rsltOLS.control.finc2)))
seWhiteFinc  <- sqrt(diag(vcovHC(rsltOLS.control.finc2, type="HC0")))
seClusterFinc <- sqrt(diag(vcov(rsltOLS.control.finc2, cluster="state")))
stargazer(rsltOLS.control.finc2, rsltOLS.control.finc2, rsltOLS.control.finc2,
          se=list(seBasicFinc, seWhiteFinc, seClusterFinc), type="text")
```

```
=========================================================================
                                           Dependent variable:
                                 ----------------------------------------
                                                  finc
                                      (1)           (2)           (3)
-------------------------------------------------------------------------
cChildren                        -4,734.562*** -4,734.562*** -4,734.562***
                                   (945.280)     (835.140)     (945.280)

dPeriod                             -567.581      -567.581      -567.581
                                   (696.500)     (799.615)     (696.500)

urate                              429.112***    429.112***    429.112***
                                   (161.063)     (146.978)     (161.063)

unearn                           1,339.835***  1,339.835***  1,339.835***
                                    (37.979)     (192.934)      (37.979)

children                         -1,505.236*** -1,505.236*** -1,505.236***
                                   (301.752)     (264.835)     (301.752)

cChildren:dPeriod                  1,490.156     1,490.156     1,490.156
                                   (945.477)     (916.133)     (945.477)

Constant                         16,363.220*** 16,363.220*** 16,363.220***
                                 (1,285.565)   (1,383.430)   (1,285.565)

-------------------------------------------------------------------------
Observations                        7,052         7,052         7,052
R2                                  0.172         0.172         0.172
Adjusted R2                         0.171         0.171         0.171
Residual Std. Error (df = 7045)  19,061.410    19,061.410    19,061.410
F Statistic (df = 6; 7045)       244.155***    244.155***    244.155***
=========================================================================
Note:                                         *p<0.1; **p<0.05; ***p<0.01
```

```r
#No impact on the significance of the DiD effect, still insignificant
#Standard error of seWhite seems smaller than basic and clustered

seBasicWork  <- sqrt(diag(vcov(rsltOLS.control.work2)))
seWhiteWork  <- sqrt(diag(vcovHC(rsltOLS.control.work2, type="HC0")))
seClusterWork <- sqrt(diag(vcov(rsltOLS.control.work2, cluster="state")))
stargazer(rsltOLS.control.work2, rsltOLS.control.work2, rsltOLS.control.work2,
          se=list(seBasicWork, seWhiteWork, seClusterWork), type="text")
```

```
=================================================================
                                      Dependent variable:
                                  -------------------------------
                                               work
                                     (1)        (2)        (3)
-----------------------------------------------------------------
cChildren                          -0.017     -0.017     -0.017
                                   (0.017)    (0.016)    (0.017)

dPeriod                            -0.031**   -0.031**   -0.031**
                                   (0.013)    (0.013)    (0.013)

urate                             -0.021***  -0.021***  -0.021***
                                   (0.003)    (0.003)    (0.003)

unearn                            -0.017***  -0.017***  -0.017***
                                   (0.001)    (0.001)    (0.001)

children                          -0.052***  -0.052***  -0.052***
                                   (0.004)    (0.004)    (0.004)

cChildren:dPeriod                  0.035**    0.035**    0.035**
                                   (0.017)    (0.017)    (0.017)

Constant                           0.817***   0.817***   0.817***
                                   (0.024)    (0.024)    (0.024)

-----------------------------------------------------------------
Observations                        13,746     13,746     13,746
R2                                  0.090      0.090      0.090
Adjusted R2                         0.089      0.089      0.089
Residual Std. Error (df = 13739)    0.477      0.477      0.477
F Statistic (df = 6; 13739)       225.866*** 225.866*** 225.866***
=================================================================
Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

```r
#No impact on the significance of the DiD effect, all three significant (p<0.05).
#Standard error for all three remains the same
```