Module 4

Eugénie de Jong

22/04/2021

In the next assignment we want to replicate some plots from the paper "Female Socialization: How Daughters Affect Their Legislator Fathers' Voting on Women's Issues" (Washington, 2008). The paper explores whether having a daughter makes politicians more sensitive to women's rights issues and how this is reflected in their voting behavior. The main identifying assumption is that after controlling for the number of children, the gender composition is random. This might be violated if families that have a preference for girls keep having children until they have a girl. In this assignment we will prepare a dataset that allows us to test whether families engage in such a "female child stopping rule".

Setup

- Load the libraries "Rio" and "tidyverse"
- Change the path of the working directory to your working directory.

```
# Set working directory
setwd("/Users/eusje96/Documents/1. Educatie/1. Stockholm School of Economics/1. Year 1/2. Semester 2/Te

library("rio")
library("tidyverse")
library("stargazer")
library("magrittr")
```

- import the data sets *basic.dta* and *genold108.dta*

```
# Import the data sets
basic <- import("basic.dta")
genold <- import("genold108.dta")
```

- create a subset of the 108th congress from the *basic* dataset

```
Basic_108 <- basic %>% filter(congress==108)
```

- join this subset with the *genold* dataset

```
Joined_Data <- left_join(genold, Basic_108, by = c("name","district", "statenam"))
```

Data preparation

- check table 1 in the appendix of the paper and decide which variables are necessary for the analysis (check the footnote for control variables)
- drop all other variables.

```
Joined_Data <- Joined_Data %>% select(name, ngirls, genold, totchi, party, rgroup,
                                      region, age, female, district, statenam, srvlng,
                                      age, female, white)
```

- Recode *genold* such that gender is a factor variable and missing values are coded as NAs.

```
# Recode "genold" such that gender is a factor variable
Joined_Data$genold <- as.factor(Joined_Data$genold)

# Checking whether recoded correctly
is.factor(Joined_Data$genold)
```

```
## [1] TRUE
```

```
# Coding missing values as NAs.
Joined_Data$genold %<>% na_if("") %>% as_factor()
```

- Recode *party* as a factor with 3 levels (D, R, I)

```
# Recode party as a factor variable
Joined_Data$party <- as.factor(Joined_Data$party)

# Checking whether recoded correctly
is.factor(Joined_Data$party)
```

```
## [1] TRUE
```

```
# Generate 3 levels (D, R, I)
 print(Joined_Data$party)
```

```
##   [1] 2 2 2 2 1 2 1 2 2 2 2 1 2 2 1 2 1 1 2 1 1 2 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1
##  [38] 1 2 1 2 2 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2 1 1
##  [75] 1 2 2 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 1 2 2 1 2 2
## [112] 2 1 1 1 1 2 2 2 2 2 2 2 1 1 1 2 2 1 1 1 1 1 2 1 2 1 2 2 1 2 2 2 2 1 2 2 1
## [149] 2 2 2 2 2 1 2 1 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 1 2 1 1 1 2 1 1 1 1
## [186] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 2 2 2 2 2 2 1 1 1 1 2 2 2 1 1 2 1 1 2
## [223] 1 2 1 1 2 1 1 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 1 1 2 2 1
## [260] 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 2 2 2 2 2 1 2 2 1 1 2 1 2 2 1 2
## [297] 2 2 2 1 1 1 2 2 2 2 2 1 2 2 1 1 1 2 1 2 2 2 1 2 2 1 2 2 2 1 2 1 1 1 1 1 2
## [334] 2 2 2 2 2 2 2 1 1 1 1 2 2 1 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 1 1 1 2 1 1 1 1 2
## [371] 1 2 2 2 2 2 1 1 1 2 2 2 1 1 1 1 2 1 2 2 2 1 1 2 1 1 1 1 2 2 2 1 2 3 2 2 1 2
## [408] 3 2 1 1 1 2 2 1 1 1 2 2 1 1 2 1 1 2 1 2 1 1 1 2 2 1 2 2
## Levels: 1 2 3
```

```
 Joined_Data$party <- recode(Joined_Data$party, "1" = "D", "2" = "R", "3" = "I")
```

- Recode *rgroup* and *region* as factors.

```
# Recode *rgroup* and *region* as factors
 Joined_Data$rgroup <- as.factor(Joined_Data$rgroup)
 Joined_Data$region <- as.factor(Joined_Data$region)

# Check whether recoding was successful
 is.factor(Joined_Data$rgroup)
```

```
## [1] TRUE
```

```
 is.factor(Joined_Data$region)
```

```
## [1] TRUE
```

- generate variables for age squared and service length squared

```
Joined_Data %<>% mutate(age_sq = age^2)
Joined_Data %<>% mutate(srvlng_sq = srvlng^2)
```

- create an additional variable of the number of children as factor variable

```
Joined_Data %<>% mutate(totchi_factor = totchi)
Joined_Data$totchi_factor <- as.factor(Joined_Data$totchi_factor)
```

Replicationg Table 1 from the Appendix

We haven't covered regressions in R yet. Use the function *lm()*. The function takes the regression model (formula) and the data as an input. The model is written as $y \sim x$, where x stands for any linear combination of regressors (e.g. $y \sim x_1 + x_2 + female$). Use the help file to understand the function.

- Run the regression $total.children = \beta_0 + \beta_1 gender.oldest + \gamma' X$ where γ stands for a vector of coefficients and X is a matrix that contains all columns that are control variables.[1]
- Save the main coefficient of interest (β_1)
- Run the same regression separately for Democrats and Republicans (assign the independent to one of the parties). Save the coefficient and standard error of *genold*
- Collect all the *genold* coefficients from the six regressions, including their standard errors and arrange them in a table as in the paper.

```
# Create vectors of control variables
control_totchi <- c("totchi", "genold", "party", "rgroup","region", "srvlng", "srvlng_sq",
                    "age", "age_sq", "female", "white")
control_ngirls <- c("ngirls", "totchi", "genold", "party", "rgroup","region", "srvlng",
                    "srvlng_sq", "age", "age_sq", "female", "white")

# Regression 1, dependent variable = total children, all data
reg1 <- lm(totchi ~ ., data = Joined_Data[, control_totchi])

# Regression 2, dependent variable = ngirls, all data
reg2 <- lm(ngirls ~ ., data = Joined_Data[, control_ngirls])

# Regression 3, dependent variable = total children, Democrat data
reg3 <- lm(totchi ~ .-party, data = subset(Joined_Data[, control_totchi], party= "D"))

# Regression 4, dependent variable = ngirls, Democrat data
reg4 <- lm(ngirls ~ .-party, data = subset(Joined_Data[, control_ngirls], party= "D"))

# Regression 5, dependent variable = total children, Republican data
reg5 <- lm(totchi ~ .-party, data = subset(Joined_Data[, control_totchi], party= "R"))

# Regression 6, dependent variable = ngirls, Republican data
reg6 <- lm(ngirls ~ .-party, data = subset(Joined_Data[, control_ngirls], party= "R"))

# Saving coefficient for genold
beta_1 <- c(summary(reg1)$coefficients[2,1], summary(reg2)$coefficients[2,1],
            summary(reg3)$coefficients[2,1], summary(reg4)$coefficients[2,1],
            summary(reg5)$coefficients[2,1], summary(reg6)$coefficients[2,1])
```

[1] This is just a short notation instead of writing the full model with all control variables $totchi = \beta_0 + \beta_1 genold + \gamma_1 age + \gamma_2 age^2 + \gamma_3 Democrat + ... + \epsilon$ which quickly gets out of hand for large models.

```r
# Saving genold sd:
sd <- c(summary(reg1)$coefficients[2,2], summary(reg2)$coefficients[2,2],
        summary(reg3)$coefficients[2,2], summary(reg4)$coefficients[2,2],
        summary(reg5)$coefficients[2,2],summary(reg2)$coefficients[2,2])

# Round variables
beta_1 %<>% round(2)
sd %<>% round(2)

# Creating variable N to be included in the table
N <- c(227,227,105,105,122,122)
N <- as.integer(N)

# Creating a table
table <- matrix(c(beta_1, sd, N), ncol=6,nrow=3, byrow=TRUE)

# Giving the columns and rows the right names
colnames(table) <- c("Full Congress, No. children", "Full Congress, No. daughters",
                     "Democrats, No. children", "Democrats, No. daughters",
                     "Republicans, No. children", "Republicans, No. daughters")
rownames(table) <- c("First child female", "Std. Error", "N")

# Print the table
print(table)
```

```
##                   Full Congress, No. children Full Congress, No. daughters
## First child female                       -0.08                         0.66
## Std. Error                                0.15                         0.04
## N                                       227.00                       227.00
##                   Democrats, No. children Democrats, No. daughters
## First child female                   -0.12                    0.65
## Std. Error                            0.15                    0.04
## N                                   105.00                  105.00
##                   Republicans, No. children Republicans, No. daughters
## First child female                     -0.12                       0.65
## Std. Error                              0.15                       0.04
## N                                     122.00                     122.00
```

- print the table