Module 4 - Instructions

Oliver Engist

3/25/2020

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".

I encourage you to take a look at the paper, as we will come back to it later in the course.

Setup

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

```
setwd("~/R projects/Module_4")
library(rio)
library(tidyverse)
library(knitr)
```

- import the data sets *basic.dta* and *genold108.dta*
- create a subset of the 108th congress from the *basic* dataset
- join this subset with the *genold* dataset

```
library(readstata13)
basic <- read.dta13("basic.dta")
genold108 <- read.dta13("genold108.dta")

library("magrittr")
basic108 <- basic %>% filter(congress == 108)

genold108 <- left_join(genold108, basic108, by = "name")
```

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.

```
keeps <- c("ngirls", "totchi", "genold",
           "white", "female", "party", "age",
           "srvlng", "rgroup", "region", "name")
genold108 <- genold108[keeps]
```

- Recode *genold* such that gender is a factor variable and missing values are coded as NAs.
- Recode *party* as a factor with 3 levels (D, R, I)
- Recode *rgroup* and *region* as factors.
- generate variables for age squared and service length squared
- create an additional variable of the number of children as factor variable

```
genold108$genold <- as.factor(na_if(genold108$genold,""))
genold108$party <- as.factor(genold108$party)
genold108$rgroup <- as.factor(genold108$rgroup)
genold108$region <- as.factor(genold108$region)
genold108$age2 <- genold108$age^2
genold108$srvlng2 <- genold108$srvlng^2
genold108$children <- as.factor(genold108$totchi)
genold108$white <- as.factor(genold108$white)
genold108$female <- as.factor(genold108$female)
```

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. 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)

```
m.nchi <- lm(as.numeric(totchi) ~ genold + white + female +
                party + age + age2 + srvlng +
                srvlng2 + rgroup + region, data = genold108)
beta.nchi <- summary(m.nchi)$coef[2,1:2]

m.ngirls <- lm(as.numeric(ngirls) ~ genold + white + female +
                  party + age + age2 + srvlng +
                  srvlng2 + rgroup + region + as.factor(totchi), data = genold108)
beta.ngirls <- summary(m.ngirls)$coef[2,1:2]
```

- Run the same regression separately for Democrats (including Bernie) and Republicans. Save the coefficient and standard error of *genold*

```
#Finding the coefficients for democrats:
genold108.dem <- genold108 %>% filter(genold108$party==1 | name == "SANDERS, BERNARD")
m.nchi.dem <- lm(as.numeric(totchi) ~ genold + white + female +
```

[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.

```
                    age + age2 + srvlng + srvlng2 +
                    rgroup + region, data = genold108.dem)
beta.nchi.dem <- summary(m.nchi.dem)$coef[2,1:2]

m.ngirls.dem <- lm(ngirls ~ genold + white + female +
                     age + age2 + srvlng + srvlng2 +
                     rgroup + region + as.factor(totchi), data = genold108.dem)
beta.ngirls.dem <- summary(m.ngirls.dem)$coef[2,1:2]

#Finding the coefficients for republicans:
genold108.rep <- genold108 %>% filter(genold108$party==2)
m.nchi.rep <- lm(as.numeric(totchi) ~ genold + white + female +
                    age + age2 + srvlng + srvlng2 +
                    rgroup + region, data = genold108.rep)
beta.nchi.rep <- summary(m.nchi.rep)$coef[2,1:2]

m.ngirls.rep <- lm(ngirls ~ genold + white +
                     female + age + age2 + srvlng + srvlng2 +
                     rgroup + region + as.factor(totchi), data = genold108.rep)
beta.ngirls.rep <- summary(m.ngirls.rep)$coef[2,1:2]
```

- Collect all the *genold* coefficients from the six regressions, including their standard errors and arrange them in a table as in the paper.
- print the table

```
library(sjPlot)
tab_model(m.ngirls, m.nchi, m.ngirls.dem, m.nchi.dem, m.ngirls.rep, m.nchi.rep,
          dv.labels = c("Daughters congress", "Total children congress",
                        "Daughters democrats", "Total children democrats",
                        "Daughters republicans", "Total children republicans"),
          p.style = "a",
          show.ci = FALSE,
          show.p = FALSE,
          collapse.se = TRUE,
          show.intercept = FALSE,
          terms = c("genold [G]"),
          pred.labels = c("First child female"),
          title = "APPENDIX TABLE 1: EVIDENCE ON LEGISLATOR
          CHILD GENDER MIX SELECTION, 108TH CONGRESS",
          file = "Appendix1.doc"
          )
```

APPENDIX TABLE 1: EVIDENCE ON LEGISLATOR CHILD GENDER MIX SELECTION, 108TH CONGRESS

Daughters congress

Total children congress

Daughters democrats

Total children democrats

Daughters republicans

Total children republicans

Predictors

Estimates

Estimates

Estimates

Estimates

Estimates

Estimates

First child female

1.36 ***(0.08)

-0.08 (0.15)

1.39 ***(0.11)

0.09 (0.18)

1.23 ***(0.11)

-0.28 (0.23)

Observations

227

227

104

104

122

122

R2 / R2 adjusted

0.785 / 0.754

0.145 / 0.058

0.824 / 0.773

0.187 / 0.003

0.816 / 0.768

0.213 / 0.076

- p<0.05 ** p<0.01 *** p<0.001

```
#For some reason tab_model() looks nice only in the html format, but not in pdf.
```