# Overall report

Huy Le Quang

6/17/2020

## 1. Introduction

The European Value Study (EVS) covers a time span from 1981 to 2017 with a core questionnaire repeated over time on adult population (at least 18 years old at the time of survey) in 33 countries in Europe. The study aims at understanding human values in various topics, namely, life, family, work, region, politics and society. The EVS data is managed by GESIS Leibniz Institute for Social Sciences in Germany, and can be obtained free of charge.

The data used in this report is taken from EVS 2017. In 2017, researchers used a mixed mode data collection. The sample size is approximately 1,200 in each country. This report focuses on analyzing two key questions: - v72: When a mother works for pay, the children suffer. - v80: when jobs are scarce, employers should give priority to national people over immigrants.

In particular, we want to look at if people agree or disagree with these two statements, and how it changes with ages. We also present a simple regression model to further investigate the relationship between sex, age, education and the degree that people agree with the statements.

## 2. Descriptive statistics

This section presents the descriptive statistics of the variables in our study. First, I show the summary table for categorical variables, and then for a continuous variable. After cleaning, all missing cases are omitted, therefore, we have a balanced dataset.

There are total 52,823 observations, with the average age of around 49.5. The proportion of each country in the database is quite similar with roughly 3, except for Denmark and Switzerland with 6 each. The percentage of female participants is slightly higher than males with 55. Most of participants have medium education (46).

For the first question of interest, 43 participants disagree that children would suffer from a working mother. But for the second question, a higher proportion of people agree that when jobs are scarce, employers should give priority to national people over immigrants (28 agree and 34 agree strongly).

```
# set  working directory

setwd ("G:/My Drive/IPSDS/Modern Workflow in Data Science/Assignments/Assignment-2/")

## Load necessary packages

library(tidyverse)
```

```
## -- Attaching packages ---------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.1     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(foreign)
library(pander)
library(qwraps2)
library(texreg)
```

```
## Version:  1.37.5
## Date:     2020-06-17
## Author:   Philip Leifeld (University of Essex)
##
## Consider submitting praise using the praise or praise_interactive functions.
## Please cite the JSS article in your publications -- see citation("texreg").
```

```
##
## Attaching package: 'texreg'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
# load data

load("data_final.RData")

# make smaller data for descriptive:

explore_data <- data_final %>%
    select(-v72, -v80, -v225, -v243_r, -education)

# get the type of variables

type_var <- unlist(map(explore_data, class))

# make frequency table

freq_tab <- function(x) {
    tab <- cbind(Count = table(x, useNA = "always"),
                 Prop = round(prop.table(table(x, useNA = "always")),2))
    tab <- as.data.frame(tab) %>%
        tbl_df() %>%
        mutate(Cat = row.names(tab)) %>%
        select(Cat, Count, Prop)
}

# select only factor variables and make table

props <- map(explore_data[, type_var == "factor"],
             freq_tab)


# get the variable names and put all proportions together
```

```
vars <- unlist(map(props, nrow))
props_tab <- reduce(props, rbind)
props_tab <- props_tab %>%
    mutate(Variable = rep(names(vars), vars))

pander(props_tab,
       caption = "Descriptive statistics of factor variables. Data: EVS 2017")
```

Table 1: Descriptive statistics of factor variables. Data: EVS 2017

| Cat | Count | Prop | Variable |
|---|---|---|---|
| Albania | 1417 | 0.03 | country |
| Azerbaijan | 1726 | 0.03 | country |
| Austria | 1567 | 0.03 | country |
| Armenia | 1480 | 0.03 | country |
| Bosnia and Herzegovina | 1650 | 0.03 | country |
| Bulgaria | 1470 | 0.03 | country |
| Belarus | 1495 | 0.03 | country |
| Croatia | 1436 | 0.03 | country |
| Czechia | 1650 | 0.03 | country |
| Denmark | 3222 | 0.06 | country |
| Estonia | 1237 | 0.02 | country |
| Finland | 1127 | 0.02 | country |
| France | 1815 | 0.03 | country |
| Georgia | 2130 | 0.04 | country |
| Germany | 2067 | 0.04 | country |
| Greece | 0 | 0 | country |
| Hungary | 1466 | 0.03 | country |
| Iceland | 1558 | 0.03 | country |
| Italy | 2168 | 0.04 | country |
| Lithuania | 1360 | 0.03 | country |
| Montenegro | 953 | 0.02 | country |
| Netherlands | 2292 | 0.04 | country |
| Norway | 1093 | 0.02 | country |
| Poland | 1280 | 0.02 | country |
| Romania | 1449 | 0.03 | country |
| Russia | 1711 | 0.03 | country |
| Serbia | 1419 | 0.03 | country |
| Slovakia | 1396 | 0.03 | country |
| Slovenia | 1045 | 0.02 | country |
| Spain | 1174 | 0.02 | country |
| Sweden | 1144 | 0.02 | country |
| Switzerland | 3069 | 0.06 | country |
| North Macedonia | 1036 | 0.02 | country |
| Great Britain | 1721 | 0.03 | country |
| NA | 0 | 0 | country |
| Female | 29212 | 0.55 | male |
| Male | 23611 | 0.45 | male |
| NA | 0 | 0 | male |
| agree strongly | 5150 | 0.1 | working_mom_fct |
| agree | 14808 | 0.28 | working_mom_fct |
| disagree | 22615 | 0.43 | working_mom_fct |
| disagree strongly | 10250 | 0.19 | working_mom_fct |

| Cat | Count | Prop | Variable |
|---|---|---|---|
| NA | 0 | 0 | working_mom_fct |
| agree strongly | 18205 | 0.34 | immigrant_job_fct |
| agree | 14787 | 0.28 | immigrant_job_fct |
| neither agree nor disagree | 8102 | 0.15 | immigrant_job_fct |
| disagree | 7608 | 0.14 | immigrant_job_fct |
| disagree strongly | 4121 | 0.08 | immigrant_job_fct |
| NA | 0 | 0 | immigrant_job_fct |
| lower | 10496 | 0.2 | education_fct |
| medium | 24406 | 0.46 | education_fct |
| higher | 17921 | 0.34 | education_fct |
| NA | 0 | 0 | education_fct |

```
# make table for continuous variables

cont_var <- list("Age" =
                     list("min" = ~ min(explore_data$age),
                          "median" = ~ median(explore_data$age),
                          "max" = ~ max(explore_data$age),
                          "mean (sd)" = ~ mean_sd(explore_data$age)))

knitr::kable(summary_table(explore_data, cont_var),
             row.names = TRUE,
             caption = "Descriptive statistics of Age. Data: EVS 2017")
```

Table 2: Descriptive statistics of Age. Data: EVS 2017

| | explore_data (N = 52,823) |
|---|---|
| min | 18 |
| median | 50 |
| max | 82 |
| mean (sd) | 49.48 ± 17.62 |

## 3. Graphs

In this section, we will explore how people's attitude towards working mothers and jobs for immigrants changes with age.

Figure 1: Age and the attitude towards working mothers

```
# Working mom and age

explore_data %>%
    ggplot(aes(x = age,
               y = working_mom))+
    stat_summary(fun.y = mean,
                 geom = "line",
                 color = "red")+
    ylim(1,4)+
    theme_bw()+
    labs(x = "Age",
         y = "Attitude towards working mothers",
```

```
        caption = "Data: EVS 2017")
```

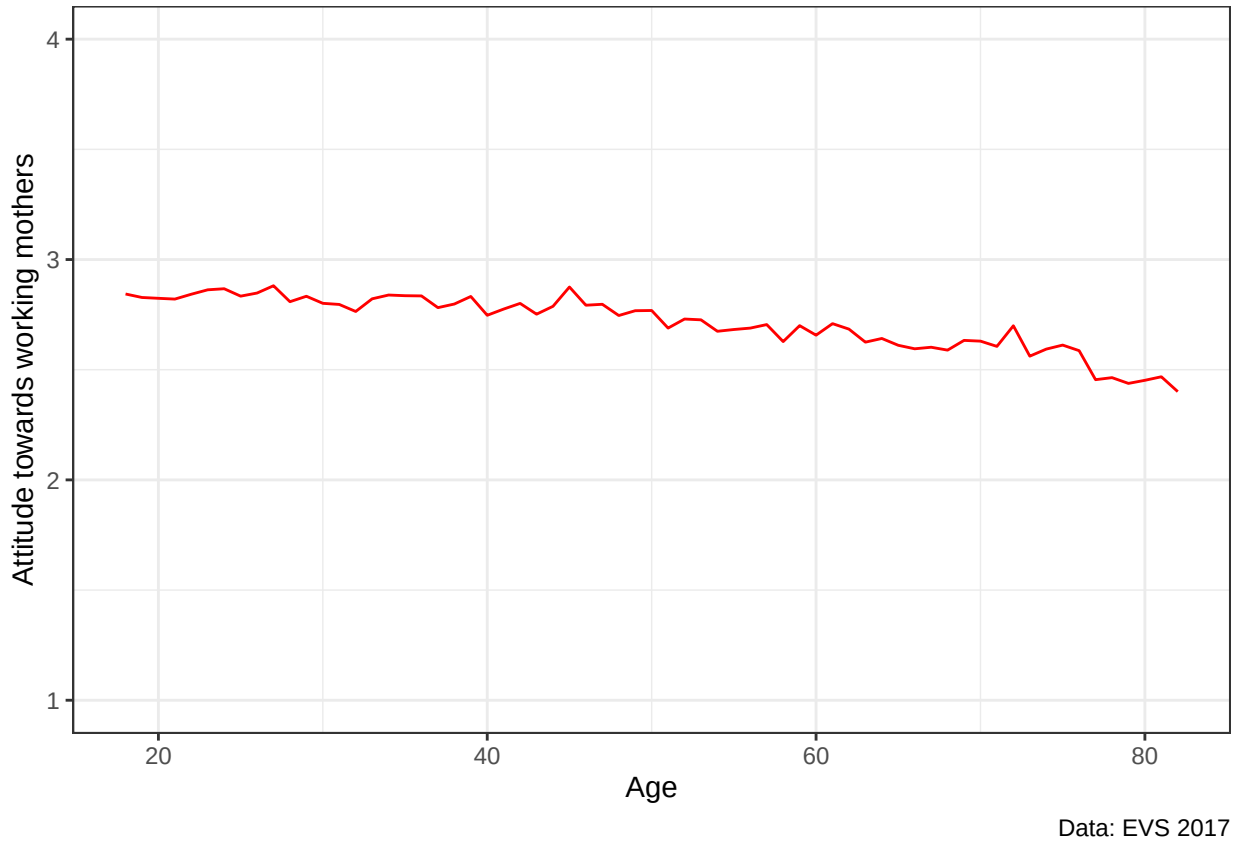
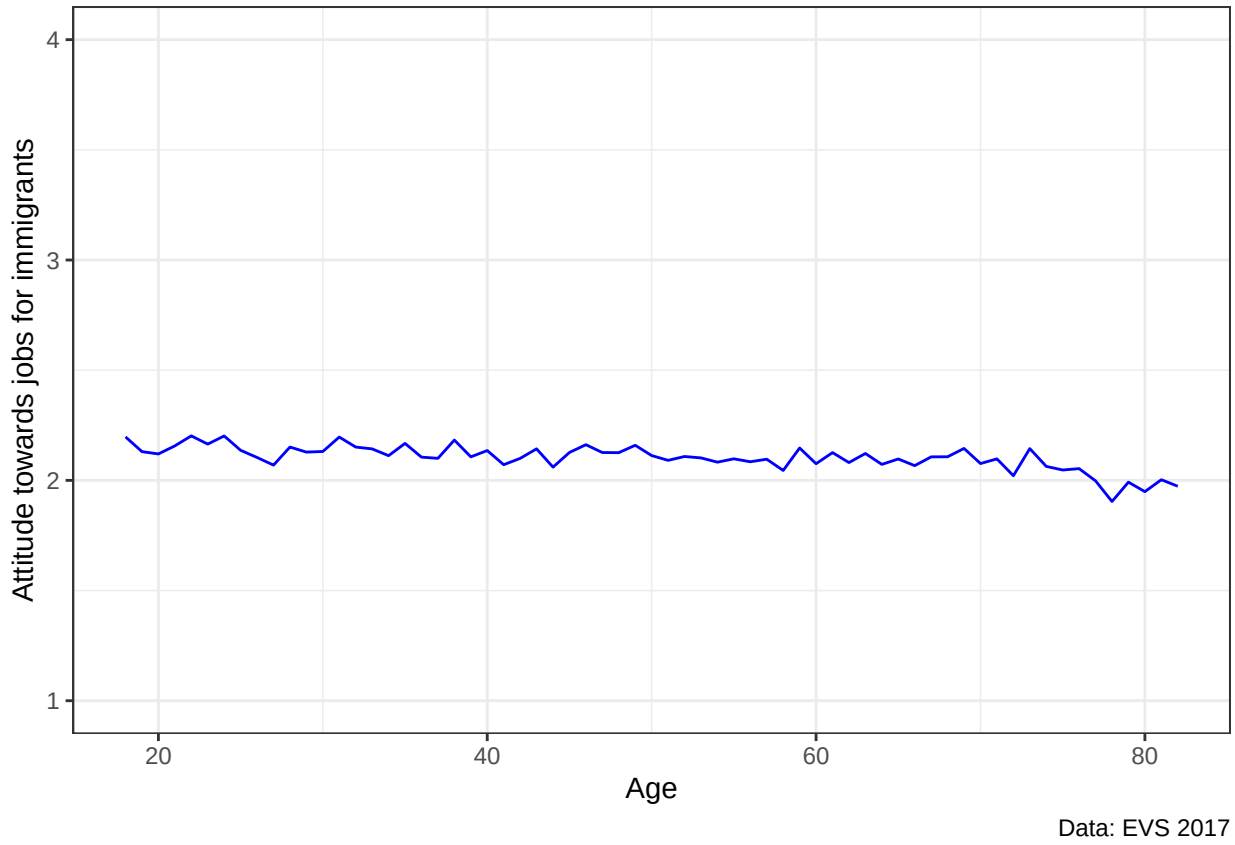



Figure 2: Age and the attitude towards jobs for immigrants

```
# Jobs for immigrants and age

explore_data %>%
    ggplot(aes(x = age,
               y = immigrant_job))+
    stat_summary(fun.y = mean,
                 geom = "line",
                 color = "blue")+
    ylim(1,4)+
    theme_bw()+
    labs(x = "Age",
         y = "Attitude towards jobs for immigrants",
         caption = "Data: EVS 2017")
```




## 4. Empirical analysis

This section presents the results of two regression models. Model 1 investigates the relationship between the attitude towards working mothers and other independent variables: age, sex and education level. Model 2 investigates the relationship between the attitude towards jobs for immigrants and the same independent variables.

```
# working mom and age

model_1 <- lm(data = explore_data,
              formula = working_mom ~ age + I(age^2) + male + education_fct)

# Immigrant jobs and age

model_2 <- lm(data = explore_data,
              formula = immigrant_job ~ age + I(age^2) + male + education_fct)

# make table

texreg(list(model_1, model_2),
        table = F,
        use.packages = F,
       digits = 5,
       caption = "Regression table. Data: EVS 2017")
```

| | Model 1 | Model 2 |
|---|---|---|
| (Intercept) | 2.73608*** | 2.26034*** |
| | (0.02986) | (0.04379) |
| age | −0.00205 | 0.00029 |
| | (0.00126) | (0.00185) |
| age$^2$ | −0.00003* | −0.00004* |
| | (0.00001) | (0.00002) |
| maleMale | −0.07184*** | 0.02772* |
| | (0.00758) | (0.01111) |
| education_fctmedium | 0.11810*** | −0.02871 |
| | (0.01034) | (0.01516) |
| education_fcthigher | 0.40458*** | 0.46218*** |
| | (0.01096) | (0.01608) |
| $R^2$ | 0.04775 | 0.03513 |
| Adj. $R^2$ | 0.04766 | 0.03504 |
| Num. obs. | 52823 | 52823 |

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

From the above table, we can see that