Exploring potential of using ESS for Comparative Analysis of Rate of Returns to Education

Suhas Parandekar, Ekaterina Melianova and ArtÃ«m Volgin

October 17, 2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Together with RStudio, you need pandoc and Miktex programs installed on your MS-Windows machine. In practice, I need to change the codechink background and the only way I could determine to do this was to edit the intermediate .tex file, though it is also possible to force rmarkdown to follow a different template for latex, in which case you would need to follow the instructions in 3.3.7.4 at https://bookdown.org/yihui/rmarkdown/pdf-document.html

Getting the rawdata

The data was downloaded from the ESS website with the help of R package essurvey - you need an easily obtainable userid in the form of an e-mail for that purpose and login one time, then with R you only need to specify the login userid e-mail. As of October 19, 2019, latest available data was Round 8.

```
load("C:/Country/Russia/Data/SEABYTE/ESS/downloads/ess_all_rounds.rdata")

# Start with round 8 of ESS data

round8 <- as.data.frame(all_rounds[8])
```

Preliminaries

I successively filter for (i) Country; (ii) Main income is labor income; (iii) worked for pay last week;(iv) household income is not missing; (v) number of household members is at least 1; (vi) education level is clearly identifiued; and (vii) age is between 25 and 64 years of age (inclusive).

I then define the variables to use in a Mincerian regression - starting with income from deciles using ESS documentation.

First case is of Austria

The ESS data is harmonized across all countries in the sample.

```
round8_AT <- round8 %>% filter(cntry=="AT") %>% filter(hincsrca==1) %>%
  filter(pdwrk==1) %>%
  filter(!is.na(hinctnta)) %>%  filter(hhmmb>=1) %>% filter(edulvlb!=5555) %>%
  filter(agea %in% (25:64)) %>%
  select(brncntr,hhmmb,gndr,agea,edulvlb,eduyrs,hinctnta,dweight,pspwght,pweight)
```

```r
# Succesively select for main income source is labor income,
# at least 1 member of household, worked for pay last week,
# non-missing household income,education level identifiable,
# age between 25 and 64.

# Define variables of interest

# Mid-point income # from EESA Income definitions Appendix A2 page 3 for Austria
# atry_h atrributed labor income household
# income is nominal euros per year
round8_AT$atry_h[round8_AT$hinctnta==1] <- 14800/2
round8_AT$atry_h[round8_AT$hinctnta==2] <- (14800+19800)/2
round8_AT$atry_h[round8_AT$hinctnta==3] <- (19800+24200)/2
round8_AT$atry_h[round8_AT$hinctnta==4] <- (24200+29400)/2
round8_AT$atry_h[round8_AT$hinctnta==5] <- (29400+35000)/2

round8_AT$atry_h[round8_AT$hinctnta==6] <- (35000+41000)/2
round8_AT$atry_h[round8_AT$hinctnta==7] <- (41000+48000)/2
round8_AT$atry_h[round8_AT$hinctnta==8] <- (48000+56500)/2
round8_AT$atry_h[round8_AT$hinctnta==9] <- (56500+70700)/2
round8_AT$atry_h[round8_AT$hinctnta==10] <- 70700*2

# atry_l attributed labor income individual
round8_AT$atry_i <- round8_AT$atry_h/round8_AT$hhmmb
round8_AT$atry_lni <- log(round8_AT$atry_i)

# Gender
round8_AT$FEMALE[round8_AT$gndr==1] <-0
round8_AT$FEMALE[round8_AT$gndr==2] <-1

# Migrant
round8_AT$MIGRANT[round8_AT$brncntr==1] <-0
round8_AT$MIGRANT[round8_AT$brncntr==2] <-1

# Education
# Define some functions and vectors for later use

`%notin%` <- Negate(`%in%`)
voc <- c(421,422,423,520)
ter <- c(412,413,510,610,620,710,720,800)

round8_AT$edu_SEC[round8_AT$edulvlb < 412] <- 1
round8_AT$edu_SEC[round8_AT$edulvlb >= 412] <- 0

round8_AT$edu_VOC[round8_AT$edulvlb %in% voc] <- 1
round8_AT$edu_VOC[round8_AT$edulvlb %notin% voc] <- 0

round8_AT$edu_TER[round8_AT$edulvlb %in% ter] <- 1
round8_AT$edu_TER[round8_AT$edulvlb %notin% ter] <- 0

# Attributed experience
round8_AT$atr_exp=round8_AT$agea -(6+round8_AT$eduyrs)
```

```
blix <- papeR::summarize(round8_AT)
knitr::kable(blix, type = "numeric",format="pandoc",
             caption="Means of ESSA data Round 8 Austria")
```

Table 1: Means of ESSA data Round 8 Austria

	N	Mean	SD	Min	Q1	Median	Q3	Max
brncntr	695	1.11	0.32	1.00	1.00	1.00	1.00	2.00
hhmmb	695	2.37	1.24	1.00	1.00	2.00	3.00	7.00
gndr	695	1.54	0.50	1.00	1.00	2.00	2.00	2.00
agea	695	42.82	10.51	25.00	34.00	44.00	52.00	64.00
edulvlb	695	391.96	144.19	113.00	322.00	322.00	423.00	800.00
eduyrs	695	13.38	2.80	4.00	12.00	12.00	15.00	25.00
hinctnta	695	5.58	2.27	1.00	4.00	6.00	7.00	10.00
dweight	695	1.03	0.48	0.37	0.62	0.99	1.23	3.36
pspwght	695	0.99	0.65	0.18	0.54	0.79	1.19	4.00
pweight	695	0.37	0.00	0.37	0.37	0.37	0.37	0.37
atry_h	695	40276.91	25267.84	7400.00	26800.00	38000.00	44500.00	141400.00
atry_i	695	19511.60	12787.62	3700.00	12666.67	17300.00	22250.00	141400.00
atry_lni	695	9.74	0.51	8.22	9.45	9.76	10.01	11.86
FEMALE	695	0.54	0.50	0.00	0.00	1.00	1.00	1.00
MIGRANT	695	0.11	0.32	0.00	0.00	0.00	0.00	1.00
edu_SEC	695	0.69	0.46	0.00	0.00	1.00	1.00	1.00
edu_VOC	695	0.14	0.35	0.00	0.00	0.00	0.00	1.00
edu_TER	695	0.17	0.38	0.00	0.00	0.00	0.00	1.00
atr_exp	695	23.44	11.17	0.00	13.50	24.00	32.50	50.00

Now, run the regression.

Table 2: Returns to Education by level: Austria

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.86	0.0826	119.4	0
edu_VOC	0.1537	0.05997	2.562	0.01061
edu_TER	0.2552	0.04744	5.379	1.029e-07
atr_exp	-0.02592	0.007354	-3.525	0.000452
I(atr_exp^2)	0.0006019	0.0001589	3.788	0.0001655
FEMALE	-0.0507	0.0391	-1.297	0.1952
MIGRANT	-0.165	0.05618	-2.937	0.003424

Table 3: Returns to Education by year: Austria

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.413	0.1311	71.82	3.557e-322
eduyrs	0.03677	0.006306	5.831	8.468e-09
atr_exp	-0.02563	0.007324	-3.499	0.0004966
I(atr_exp^2)	0.0006144	0.0001582	3.884	0.0001126
FEMALE	-0.02454	0.03861	-0.6357	0.5252
MIGRANT	-0.1742	0.05595	-3.113	0.00193

Case of Russian Federation

```
round8_RU <- round8 %>% filter(cntry=="RU") %>% filter(hincsrca==1) %>%
  filter(pdwrk==1) %>%  filter(!is.na(hinctnta)) %>%  filter(hhmmb>=1) %>%
  filter(edulvlb!=5555) %>%   filter(agea %in% (25:64)) %>%
select(brncntr,hhmmb,gndr,agea,edulvlb,eduyrs,hinctnta,dweight,pspwght,pweight)

# Define variables of interest

# Mid-point income # from EESA Income definitions Appendix A2 page 24 for Russia
#atry_h atrributed labor income household
# Income is nominal rubles per month
round8_RU$atry_h[round8_RU$hinctnta==1] <- 12000/2
round8_RU$atry_h[round8_RU$hinctnta==2] <- (12000+15000)/2
round8_RU$atry_h[round8_RU$hinctnta==3] <- (15000+18000)/2
round8_RU$atry_h[round8_RU$hinctnta==4] <- (18000+21000)/2
round8_RU$atry_h[round8_RU$hinctnta==5] <- (21000+25000)/2

round8_RU$atry_h[round8_RU$hinctnta==6] <- (25000+30000)/2
round8_RU$atry_h[round8_RU$hinctnta==7] <- (30000+40000)/2
round8_RU$atry_h[round8_RU$hinctnta==8] <- (40000+60000)/2
round8_RU$atry_h[round8_RU$hinctnta==9] <- (60000+80000)/2 # typo in text
round8_RU$atry_h[round8_RU$hinctnta==10] <- 80000*2

# atry_l attributed labor income individual
round8_RU$atry_i <- round8_RU$atry_h/round8_RU$hhmmb
round8_RU$atry_lni <- log(round8_RU$atry_i)

# Gender
round8_RU$FEMALE[round8_RU$gndr==1] <-0
round8_RU$FEMALE[round8_RU$gndr==2] <-1

# Migrant
round8_RU$MIGRANT[round8_RU$brncntr==1] <-0
round8_RU$MIGRANT[round8_RU$brncntr==2] <-1

# Education
# Define some functions and vectors for later use

`%notin%` <- Negate(`%in%`)
voc <- c(421,422,423,520)
ter <- c(412,413,510,610,620,710,720,800)

round8_RU$edu_SEC[round8_RU$edulvlb < 412] <- 1
round8_RU$edu_SEC[round8_RU$edulvlb >= 412] <- 0

round8_RU$edu_VOC[round8_RU$edulvlb %in% voc] <- 1
round8_RU$edu_VOC[round8_RU$edulvlb %notin% voc] <- 0

round8_RU$edu_TER[round8_RU$edulvlb %in% ter] <- 1
round8_RU$edu_TER[round8_RU$edulvlb %notin% ter] <- 0

# Attributed experience
```

```
round8_RU$atr_exp=round8_RU$agea -(6+round8_RU$eduyrs)
```

```
blix <- papeR::summarize(round8_RU)
knitr::kable(blix, type = "numeric",format="pandoc",
             caption="Means of ESSA data Round 8 Russia")
```

Table 4: Means of ESSA data Round 8 Russia

	N	Missing	Mean	SD	Min	Q1	Median	Q3	Max
brncntr	870	2	1.05	0.21	1.00	1.00	1.00	1.00	2.00
hhmmb	872	0	2.56	1.18	1.00	2.00	2.00	3.00	8.00
gndr	872	0	1.52	0.50	1.00	1.00	2.00	2.00	2.00
agea	872	0	40.69	10.28	25.00	32.00	40.00	49.00	64.00
edulvlb	872	0	543.60	164.56	113.00	520.00	520.00	720.00	800.00
eduyrs	872	0	13.67	2.39	7.00	12.00	14.00	15.00	21.00
hinctnta	872	0	6.34	2.32	1.00	5.00	7.00	8.00	10.00
dweight	872	0	1.04	0.46	0.34	0.54	0.98	1.38	3.00
pspwght	872	0	0.91	1.05	0.10	0.32	0.48	0.90	4.00
pweight	872	0	4.99	0.00	4.99	4.99	4.99	4.99	4.99
atry_h	872	0	43091.17	35574.81	6000.00	23000.00	35000.00	50000.00	160000.00
atry_i	872	0	18732.74	15508.90	1500.00	9166.67	14000.00	23000.00	160000.00
atry_lni	872	0	9.60	0.67	7.31	9.12	9.55	10.04	11.98
FEMALE	872	0	0.52	0.50	0.00	0.00	1.00	1.00	1.00
MIGRANT	870	2	0.05	0.21	0.00	0.00	0.00	0.00	1.00
edu_SEC	872	0	0.24	0.42	0.00	0.00	0.00	0.00	1.00
edu_VOC	872	0	0.38	0.48	0.00	0.00	0.00	1.00	1.00
edu_TER	872	0	0.39	0.49	0.00	0.00	0.00	1.00	1.00
atr_exp	872	0	21.02	11.03	1.00	12.00	20.00	30.00	50.00

Table 5: Returns to Education by level: Russia

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.676	0.1065	90.85	0
edu_VOC	0.01137	0.06127	0.1856	0.8528
edu_TER	0.289	0.06354	4.548	6.19e-06
atr_exp	-0.01506	0.00881	-1.709	0.08775
I(atr_exp^2)	0.0002405	0.000175	1.375	0.1696
FEMALE	-0.05401	0.04534	-1.191	0.2339
MIGRANT	-0.1087	0.08962	-1.213	0.2255

Table 6: Returns to Education by year: Russia

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.255	0.1918	48.27	1.686e-247
eduyrs	0.04108	0.01111	3.697	0.0002322
atr_exp	-0.01909	0.008699	-2.195	0.02845
I(atr_exp^2)	0.0003303	0.0001733	1.906	0.05703
FEMALE	-0.05299	0.04564	-1.161	0.246
MIGRANT	-0.09445	0.09015	-1.048	0.2951

Just some fun with R markdown in knitr.

There is a lot going on here between R, latex and pandoc, but the intermediate tex file is always a recourse in case things become iffy trying to run latex commands from inside the .rmd files.

```
writeLines("help")
```

```
## help
```

Red text with a gray background.

Red text.

Black text.

Black text with a red background.

Black text with a darker red background. Trying this for a code chunk will show the begin and end latex commands which I don't know yet how to supress, if at all it is possible in the current set-up.