Working Paper 1: Draft

Suhas D. Parandekar, Ekaterina Melianova, Artëm Volgin

October 30, 2019

```
> library(foreign)
> library(plyr); library(dplyr)
> library(gmodels)
> library(lmtest)
> library(sqldf)
> library(XLConnectJars)
> library(questionr)
> library(labelled)
> library(tidyr)
> library(magrittr)
> library(ggplot2)
> library(data.table)
> library(pbapply)
> library(knitr)
> library(pander)
> library(gridExtra)
> library(rio)
> library(stargazer)
> library(xtable)

> # Working directory
> wd <- paste0(normalizePath(Sys.getenv("USERPROFILE"), winslash = "/"), "/Desktop")
> setwd(wd)
> # Connecting with SQLite
> db <- dbConnect(SQLite(), dbname=paste0(wd, "/rlms.db"))
> # Modified cbind
> cbind.all <- function (...) {
+   nm <- list(...)
+   nm <- lapply(nm, as.matrix)
+   n <- max(sapply(nm, nrow))
+   do.call(cbind, lapply(nm, function(x) rbind(x, matrix(, n -
+                                                           nrow(x), ncol(x)))))
+ }
> # A function for variable selection
> selectFromSQL <- function(column_names=NULL, column_blocks=NULL,
+                           wave_number=NULL, dbname = "rlms.db"){
+   "
+   1. column_names - select specific column/s
```

```
+
+   2. column_blocks - select specific block/s of columns. Available blocks:
+   Bank services
+   Children
+   Daily activities
+   Education
+   Elections
+   Employment
+   Employment/finance (retrospective)
+   Family
+   Finance
+   For women only
+   Health assessment
+   Identification variables
+   Inequity issues
+   Insurance
+   Interviewer's remarks
+   IT skills
+   Law
+   Living conditions
+   Maternal capital
+   Medical care
+   Migration
+   Military service
+   Nationality issues
+   Other
+   Pension
+   Personality assessment
+   Politics
+   Religion
+   Safety/crimes
+   Shopping
+   Socio-demographics
+   Sorces of news
+   State services
+   Transition period
+   Traveling
+   Trust
+
+   3. wave_number - select specific wave/s in RLMS
+
+   4. db_name - name of database in SQLite, by default rlms.db
+
+   "
+
+   # Add blocks to columns
+   if (is.null(column_blocks) == FALSE){
+     # Load table with blocks
```

```
+     blocks_df <- sqldf('SELECT * from rlms_blocks', dbname = dbname)
+     # Get columns from the selected blocks
+     columns_from_blocks <- c()
+     for (block in column_blocks){
+       columns_from_blocks <- c(columns_from_blocks, blocks_df[blocks_df$column_blo
+     }
+     column_names <- c(column_names, columns_from_blocks)
+   }
+
+   # Add ids to the columns
+   column_names <- unique(c("ID_W", "IDIND", "REDID_I", "ID_I", "ID_H", column_name
+   # Condition on a wave number
+   if (is.null(wave_number) == FALSE){
+     if(length(wave_number) == 1){
+       wave_condition <- paste('WHERE ID_W =', wave_number)
+     } else {
+       wave_condition <- paste('WHERE', paste(paste0('ID_W=', wave_number), collaps
+     }
+   } else {
+     wave_condition <- ""
+   }
+
+   # In case if the number of columns is 63 or larger (a default limitation of SQLi
+   if (length(column_names) >= 63){
+     # Create a list with column chunks
+     column_splits <- split(column_names, ceiling(seq_along(column_names)/63))
+     # Add the columns by parts
+     result_df <- data.frame()
+     for (column_split in column_splits){
+       if ('ID_W' %in% column_split == FALSE){
+         column_split <- c('ID_W', column_split)
+       }
+
+       command_line <- paste(c("SELECT", paste(column_split, collapse=', '),
+                               "FROM", paste(column_split, collapse=' NATURAL JOIN
+                               wave_condition), collapse = ' ')
+       cat('--- SQL command:', command_line, sep="\n")
+       temp_df <- data.frame(sqldf(command_line, dbname = dbname))
+       result_df <- data.frame(cbind.all(result_df, temp_df))
+     }
+     # Remove duplicates of the ID_W column
+     result_df <- result_df[, -grep("ID_W.", colnames(result_df))]
+
+   } else {
+     command_line <- paste(c("SELECT", paste(column_names, collapse=', '),
+                             "FROM", paste(column_names, collapse=' NATURAL JOIN ')
+                             wave_condition), collapse = ' ')
+     cat('--- SQL command:', command_line, sep="\n")
```

```
+     result_df <- sqldf(command_line, dbname = dbname)
+   }
+
+   return(result_df)
+ }
> ##############################################################################
>
> # Selecting the variables of interest
> df_ <- selectFromSQL(c("AGE", "J13_2", "J10", "J40", "EDUC", "J1",
+                        "J5A", "J5B", "H7_2", "H5",
+                        "J23", "I2", "I4", "YEAR", "J40", "J35_2Y", "J35_2M",
+                        "total_exper", "exper_main_", "exper_add_",
+                        "J5A_", "J5B_", "J35_2Y_", "J35_2M_"))

--- SQL command:
SELECT ID_W, IDIND, REDID_I, ID_I, ID_H, AGE, J13_2, J10, J40, EDUC, J1, J5A, J5B, H7_

> # Fixing system and user-defined missings in the RLMS database
>
> # Defining functions for a proper treatment of missing values
> SysMisFix <- function(df){
+   "SysMisFix changes chategorical NA to missing values"
+   temp <- df
+   for (i in colnames(df)){
+     temp[,i] <- mapvalues(df[,i], "NA", NA, warn_missing = F)
+   }
+   return(temp)
+ }
> UserMisFix <- function(df, na_range = 99999997:99999999){
+   "UserMisFix labels user-defined missings as missing value "
+   for (i in colnames(df)){
+     if (is.character(df[,i]) == T){
+       na_values(df[,i]) <- as.character(na_range)
+       }
+     else if (is.factor(df[,i]) == T){
+       na_values(df[,i]) <- NULL
+       }
+     else if (!i %in% c("ID_W", "IDIND","YEAR","REDID_I","ID_I","ID_H")){
+       na_values(df[,i]) <- na_range
+     }
+   }
+   return(df)
+ }
> df_ <- SysMisFix(df_) # determining system missings
> df_ <- UserMisFix(df_) # labelling user-defined missings
> # A function for calculating descriptive statistics: a slightly extended version o
> Freq <- function(var){
+   result <- freq(var, levels = "values", total = T)
+   result <- rbind(result,
```

```
+                    UserNA = apply(result[as.character(99999997:99999999),],2,sum),
+                    TotalNA = apply(result[c(99999997:99999999, "NA"),],2,sum, na.rm
+   return(result)
+ }
> ############################################################################################
>
> # Filtering age
> #Freq(df_$AGE)
> df <- df_[df_$AGE >= 25 & df_$AGE < 65,]
> # Filtering employed
> df <- df[df$J1 >= 1 & df$J1 < 5,]
> #Freq(df$J1)
>
> # Education
>
> # 4 categories:
> # 0 - lower than secondary
> # 1 - secondary
> # 2 - specialized / vocational
> # 3 - higher and above
>
> # Freq(df$EDUC)
> df$EDUC <- as.numeric(df$EDUC)
> df$edu_4 <- car::recode(df$EDUC, "0:7=0; 8:9=1; 10:11=2; 12=1; 13=2;
+                              14=1; 15:17=2; 18=2; 19:20=3; 21:23=3")
> # Freq(df$edu_4)
> df <- UserMisFix(df) # fixing missings after creating a new variable
> # Filtering 3 education levels
> df <- df[df$edu_4>0,]
> # Freq(df$edu_4)
>
> # Education as factor
> df$edu_4 <- factor(df$edu_4, levels=c(1,2,3),
+                               labels=c("Secondary",
+                                        "Vocational",
+                                        "Higher"))
> # Wage
>
> # Select max wage if there is an additional job
> # Question J13_2 is missing in 1994, 1995, 1996 - for those years let us use
> # J10 as an approximation (the amount of money earned within the previous 30 days)
> df$wage <- ifelse(df$YEAR < 1998, as.numeric(apply(df[,c("J10", "J40")], 1,
+                                                      max, na.rm=T)),
+                  as.numeric(apply(df[,c("J13_2", "J40")], 1, max, na.rm=T)))
> # There are a few of those whose additional wage is greater than the main one
> length(df[which(df$J40 > df$J13_2), "IDIND"]) # for 1998 - 2018

[1] 1954
```

```
> length(df[which(df$J40 > df$J10), "IDIND"]) # for 1994 - 1996

[1] 2099

> df <- UserMisFix(df) # fixing missings after creating a new variable
> # tail(Freq(df$wage), n = 7L) # ~ 20k NAs
>
> # Socio-demographics
>
> # non-Russian
> # Freq(df$I4)
> df$non_russ[df$I4 == 1] <- 0
> df$non_russ[df$I4 > 1] <- 1
> df$non_russ[is.na(df$I4)] <- 1
> Freq(df$non_russ)

             n     %  val%
0       118660  86.9  86.9
1        17842  13.1  13.1
Total   136502 100.0 100.0
UserNA      NA    NA    NA
TotalNA      0   0.0   0.0

> # Gender
> # Freq(df$H5)
> df$female[df$H5==2] <- 1
> df$female[df$H5==1] <- 0
> # Freq(df$female)
>
> # Generating a final dataset for the analysis
> names(df)[which(colnames(df) == "total_exper")] <- "exper"

> # Tagging instances (tag1) and unique respondents (tag2)
>
> # If a month of the start of work is missing but a year is not, let us use 1 (Janu
> df[is.na(df$J5B)&is.na(df$J5A)==F, "J5B"] <- 1 # for a main work
> df[is.na(df$J35_2M)&is.na(df$J35_2Y)==F, "J35_2M"] <- 1 # for an additional work
> df$tag1 <- ifelse(df$exper_main_ > df$exper_add_ &
+                     is.na(df$exper_main_) == F &
+                     is.na(df$exper_add_) == F &
+                     df$J5A == df$J5A_ &
+                     df$J5B == df$J5B_, 1,
+                   ifelse(df$exper_main_ < df$exper_add_ &
+                            is.na(df$exper_main_) == F &
+                            is.na(df$exper_add_) == F &
+                            df$J35_2Y == df$J35_2Y_ &
+                            df$J35_2M == df$J35_2M_, 1,
+                          ifelse(is.na(df$exper_main_) &
+                                   is.na(df$exper_add_) == F &
```

```
+                                 df$J35_2Y == df$J35_2Y_ &
+                                 df$J35_2M == df$J35_2M_, 1,
+                               ifelse(is.na(df$exper_add_) &
+                                        is.na(df$exper_main_) == F &
+                                        df$J5A == df$J5A_ &
+                                        df$J5B == df$J5B_, 1,
+                                      ifelse(is.na(df$exper_main_) == T
+                                                 & is.na(df$exper_add_) == T, -1, 0))
> table(df$tag1) # 51850 instances with inconsistencies, 77903 - without

   -1     0     1
 6749 51850 77903

> # sum(table(df$tag1))
>
> resp <- df %>%
+   group_by(IDIND) %>%
+   summarise(tag2 = ifelse(any(tag1 == 0), 0, 1))
> table(resp$tag2) # 13670 respondents with inconsistencies, 12674 - without

    0     1
13670 12674

> ################################################################################
>
> # Another method of calculating experience:
> # experience as a time between graduation and a current work
>
> # Working directory
> wd <- paste0(normalizePath(Sys.getenv("USERPROFILE"), winslash = "/"), "/Desktop")
> setwd(wd)
> # Vars of interest
> grad_dates <- c("J72_1E", "J72_1E2", "J72_1E3", "J72_1E4", "J72_1E5",
+                 "J72_2E", "J72_2E2",
+                 "J72_3E", "J72_3E2",
+                 "J72_4E", "J72_4E2",
+                 "J72_5E", "J72_5E2", "J72_5E3", "J72_5E4",
+                 "J72_6E", "J72_6E2")
> df_2 <- selectFromSQL(c(grad_dates,
+                              "J71", "J72_1", "J72_2", "J72_3", "J72_4", "J72_5", "
+                              select(-c("ID_H", "REDID_I", "ID_I"))

--- SQL command:
SELECT ID_W, IDIND, REDID_I, ID_I, ID_H, J72_1E, J72_1E2, J72_1E3, J72_1E4, J72_1E5, J

> df_2 <- SysMisFix(df_2) # determining system missings
> df_2 <- UserMisFix(df_2) # labelling user-defined missings
> # Merging with our subsample
> df_2 <- df %>%
```

```
+   left_join(df_2, by = c("IDIND", "ID_W"))
> # Format
> df_2[, grad_dates] <- sapply(df_2[, grad_dates], as.numeric)
> # Denoting user NA
> for (i in grad_dates){
+   df_2[,i] <- ifelse(df_2[,i] == 99999997|
+                         df_2[,i] == 99999998|
+                         df_2[,i] == 99999999, NA, df_2[,i])
+ }
> # Selecting the FIRST graduation among several (e.g., a person has 2 degrees)
> # BY levels of education
> df_2$J72_1E_min <- apply(df_2[, c("J72_1E", "J72_1E2",
+                                  "J72_1E3", "J72_1E4",
+                                  "J72_1E5")], 1, min, na.rm=T)
> df_2$J72_2E_min <- apply(df_2[, c("J72_2E", "J72_2E2")], 1, min, na.rm=T)
> df_2$J72_3E_min <- apply(df_2[, c("J72_3E", "J72_3E2")], 1, min, na.rm=T)
> df_2$J72_4E_min <- apply(df_2[, c("J72_4E", "J72_4E2")], 1, min, na.rm=T)
> df_2$J72_5E_min <- apply(df_2[, c("J72_5E", "J72_5E2",
+                                  "J72_5E3", "J72_5E4")], 1, min, na.rm=
> df_2$J72_6E_min <- apply(df_2[, c("J72_6E", "J72_6E2")], 1, min, na.rm=T)
> # A vector with the names of the selected graduation years
> grady_min <- c("J72_1E_min", "J72_2E_min", "J72_3E_min",
+                "J72_4E_min", "J72_5E_min", "J72_6E_min")
> # Replacing infinity which indicates the absence of a certain degree
> for (i in grady_min){
+   df_2[,i] <- ifelse(is.infinite(df_2[,i]), NA, df_2[,i])
+ }
> # Graduation age for each education level
> df_2$grad_age_1E <- df_2$J72_1E_min - df_2$H6
> df_2$grad_age_2E <- df_2$J72_2E_min - df_2$H6
> df_2$grad_age_3E <- df_2$J72_3E_min - df_2$H6
> df_2$grad_age_4E <- df_2$J72_4E_min - df_2$H6
> df_2$grad_age_5E <- df_2$J72_5E_min - df_2$H6
> df_2$grad_age_6E <- df_2$J72_6E_min - df_2$H6
> # A vector with the names of graduation ages
> grad_ages <- c("grad_age_1E", "grad_age_2E", "grad_age_3E",
+                "grad_age_4E", "grad_age_5E", "grad_age_6E")
> # Limiting the age of getting education by 25 years
> # This is necessary due to people recieving additional
> # degrees across their life-course: we are most interested
> # in the highest level recieved within this interval
> for (i in grad_ages){
+   df_2[,i] <- ifelse(df_2[,i] > 25, NA, df_2[,i])
+ }
> # summary(df_2[, grad_ages])
>
> # Defining this age for those with any level after secondary
> df_2$grad_age <- apply(df_2[,grad_ages], 1, max, na.rm=T)
```

```
> # Defining this age for those with only secondary level
> df_2$grad_age <- ifelse(is.na(df_2$J72_1E_min)&
+                           is.na(df_2$J72_2E_min)&
+                           is.na(df_2$J72_3E_min)&
+                           is.na(df_2$J72_4E_min)&
+                           is.na(df_2$J72_5E_min)&
+                           is.na(df_2$J72_6E_min), 17, df_2$grad_age)
> # Approximating this age for those who recieved their degree after 25 year
> df_2$grad_age <- ifelse(is.infinite(df_2$grad_age), 20, df_2$grad_age)
> # summary(df_2$grad_age)
> # Freq(df_2$grady)
>
> # Final graduation year
> df_2$grady <- df_2$H6 + df_2$grad_age
> # Final experience based on the time period after graduation
> df_2$exper2 <- df_2$YEAR - df_2$grady
> # summary(df_2$exper2
>
> ################################################################################
>
> # Note: df_ini is a df from mincer2a.R file with the initial experience
> # variable (without inconsistency fixing)
> setwd("C:/Country/Russia/Data/SEABYTE/RLMS/edreru/prelims")
> df_ini <- readRDS("df_ini.rds")
> # A table with instancies
>
> # Merging experience computed on the basis of the initial variables and
> # experience as a time between graduation and a current work
> df_t <- df %>%
+   left_join(df_ini[, c("IDIND", "ID_W", "exper_ini")], by = c("IDIND", "ID_W")) %>
+   left_join(df_2[, c("IDIND", "ID_W", "exper2")], by = c("IDIND", "ID_W"))
> df_t <- df_t[-which(is.na(df_t$exper)),]
> # Formatting
> df_t$exper <- as.numeric(df_t$exper)
> df_t$exper_ini <- as.numeric(df_t$exper_ini)
> # rlms_tbl0: means are aggregated by unique IDIND
> rlms_tbl0 <- df_t  %>%
+   group_by(IDIND) %>%
+   summarise(N = n(),
+             exp_mean_ = mean(exper, na.rm = T),
+             exp_mean_ini_ = mean(exper_ini, na.rm = T),
+             exp_mean_2_ = mean(exper2, na.rm = T))
> # rlms_tbl: means are aggregated by the number of instances
> rlms_tbl <- rlms_tbl0 %>%
+   group_by(N) %>%
+   summarise(N_resp = n(),
+             exp_mean = mean(exp_mean_, na.rm = T),
+             exp_mean_ini = mean(exp_mean_ini_, na.rm = T),
```

```
+              exp_mean_2 = mean(exp_mean_2_, na.rm = T))
> # Checking the distribution
> rlms_tbl$mult <- rlms_tbl$N*rlms_tbl$N_resp
> sum(rlms_tbl$mult) # ok
```

```
[1] 136428
```

```
> # A list to which a t-test will be applied
> list_splt_by_N <- split(rlms_tbl0, f = rlms_tbl0$N)
> # Conducting a t-test
> # Comparing between our EXP and the INITIAL one
> t.test.pval_1 <- c()
> for (i in 1:length(list_splt_by_N)){
+   base <- list_splt_by_N[[i]][['exp_mean_']]
+   against <- list_splt_by_N[[i]][['exp_mean_ini_']]
+   test <- t.test(base, against)
+   t.test.pval_1[i] <- round(test$p.value,3)
+ }
> # Comparing between our EXP and the one based on the difference
> # between graduation and a current work
> t.test.pval_2 <- c()
> for (i in 1:length(list_splt_by_N)){
+   base <- list_splt_by_N[[i]][['exp_mean_']]
+   against <- list_splt_by_N[[i]][['exp_mean_2']]
+   test <- t.test(base, against)
+   t.test.pval_2[i] <- round(test$p.value,3)
+ }
>
> # cbind.data.frame(rlms_tbl$N, t.test.pval)
```

RLMS Results

Data

To estimate returns to education in Russia we employed the Russian Longitudinal Monitoring Survey (RLMS) - the longest panel survey of individuals and households in Eastern Europe and Asia (Kozyreva and Sabirianova Peter 2015) and the only representative Russian survey with a sizable panel component allowing for a dynamic analysis. The data are notable for their reliability, diversity, and applicability to a variety of research questions. The RLMS embraces information on people's income and expenditure structure, their material well-being, educational and occupational behavior, health state and nutrition, migration, etc. High standards of the survey content accepted worldwide enable to conduct a comprehensive cross-country comparison. RLMS sampling procedures have been thoroughly and extensively described elsewhere (Kozyreva and Sabirianova Peter 2015). The present research uses all 23 waves that are currently available (1994 - 2018). A sub-sample selected for the empirical investigation includes working individuals aged 25-64 who are out of school and have positive labor market experience and income.

Methods

Our empirical analysis pertains to the examination of a slightly modified basic specification of a mincer-type wage equation (Mincer 1974). We present results for the general working population of the Russian Federation aged 25-64 as well as by gender and nationality. The specification of focus is as follows:

$$Log(Wage) = b_0 + b_1 \cdot Educ + b_2 \cdot Exp + b_3 \cdot Exp^2 + b_4 \cdot Gender + b_5 \cdot Nationality + \epsilon$$

where $Log(Wage)$ is a logarithm of wage, $Educ$ stands for the highest attained level of education, Exp and Exp^2 reflect the years of working experience and its quadratic term respectively, $Gender$ is a dummy variable for gender, $Nationality$ represents a person's nationality, b_0 is an intercept, $b_1 ... b_n$ are the respective slope estimates, ϵ refers to a normally distributed error term.

Measures

Dependent variable

For the dependent variable we used the logarithm of an average monthly wage within the past year on a person's primary job (J13.2 variable in the RLMS dataset). If a person had an additional job, the maximum wage value among the two (J13.2 and J40) was selected for the analysis. In the waves from 1994 to 1996 the above mentioned question was absent; for those waves we exploited a variable about the average amount of money earned by a respondent within the past 30 days (J10) as a reasonable approximation.

Independent variables

We distinguished 3 categories for the **education** variable (EDUC): (1) secondary, (2) vocational, and (3) higher. Incomplete levels were incorporated into the respective upper categories (e.g., incomplete higher - into higher). We are interested in exploring returns to vocational and higher education. Estimations of premiums to primary and secondary schooling levels are technically unreachable to us since the amounts of adults without primary education and with only primary level are minuscule in the general population. We also refused from a strategy of measuring education metrically in years due to a tangible variability of returns at each year of education: extrapolating yearly effects to any time period of schooling might be crude and erroneous.

To create a variable displaying a person's **experience** we leveraged four questions on a year and a month of both primary and additional job start (J5A, J5B, J35.2Y, and J25.2M). Based on these variables and the information on an interview date, we generated a labor market experience variable for each unique respondent in the sample by summing his or her experience time registered across RLMS waves. If an employed individual had missing values on a date of his or her start of work, we imputed those missings by the first non-missing record from previous waves. If only month of a job start lacked a valid response, we roughly assumed a respondent had started his or her work in January. In cases of the absence of a valid answer with regard to those questions in both primary and additional job in all RLMS waves a person have been surveyed, such cases ($< 1\%$) were dropped from the analysis. List-wise deletion strategy was also applied to the observations with "negative" experience ($< 1\%$) when according to one's responses a job started allegedly "after" the interview occurred.

In addition, we elaborated a routine for detecting and fixing inconsistent responses to the questions about experience. Such pitfalls, impeding the correct computation of the experience variable, originate from the following two circumstances:

1. individuals, "covering" a job starting date throughout one or several successive waves and "disclosing" it later (e.g., a respondent in 2005 mentioned he or she had started to work in 2000, however, in 2001 a different response was recorded);

2. individuals, naming an earlier job starting date compared to the mentioned preceding dates (e.g., a respondent in 2005 said he had started to work in 2000, but in 2001 he or she replied 1995 had been in fact the beginning year of his or her working activity).

In both situations the priority was given to earlier responses since the closeness to the actual job starting date can potentially downgrade the probability of making a mistake by a person. Incoherence correction involved 52% of unique respondents or 40% of instances in the pooled RLMS database (1994 - 2018). However, it should be noted that such an approach to the labor market experience measurement is not devoid of shortcomings. Tracking people's experience based solely on information present in the dataset implies we restrict people's career paths to what has been captured exclusively by the survey. Hence we assume everyone reports on their first job start when they are interviewed for the first time, which is not always the case. Therefore, we cannot be sure on the preciseness of calculations concerning experience variable for individuals surveyed little amount of times, albeit are certain about the experience data for people, participated in the survey many times. As a counter-argument we can contend that such systematic underestimation affects everyone equally, hence leaving a relative aspect intact.

There is another method of computing a person's labor market experience frequent in past research. This method equates one's experience with a time distance between his or her graduation from the last institution and a current work. Such an approach has some disadvantages as well. First, it ignores the existence of unemployment which is of particular point of concern for people lacking higher education; that is why the method disproportionately overestimates people's labor market activity. Second, it does not account for the fact that people can continuously receive their education across their life-course, thus the method neglects a qualification upsurge in older ages. Third, the approach erases the variability in experience between people of the same age.

Table 1 shows the results of averaging all the three ways of generating labor market experience by "cohorts" - groups of respondents aggregated on the basis of incidence number in the RLMS survey - and comparing the obtained means by a t-test. This table demonstrates that the two corresponding versions of experience computed with (EXP 1) and without (EXP 2) inconsistency correction gradually increase with the rise of the instance number, which is expected. Moreover, they are rather close in absolute values, nevertheless, entail a statistical difference (for the bulk of the cohorts t-test p-values are lower than 0.05). Therefore, perhaps, these variables cannot be leveraged interchangeably without introducing bias into the analysis. Importantly, both EXP 1 and EXP 2 take into consideration the fact that people can change their job, thus previous working experience is added to a current value. Besides, the table poses evidence that experience based on a graduation year (EXP 3) overestimates one's labor market activity since is greater than the corrected EXP 1 even for the most frequent RLMS participants (p-values of the respective t-tests are lower than 0.05). Overall, this indicates the necessity of harmonizing the data in the described manner.

```
> # The resulting table
> rlms_tbl_res <- cbind.data.frame(rlms_tbl, t.test.pval_1, t.test.pval_2)
> # Naming
> colnames(rlms_tbl_res) <- c("Cohort", "N resp", "EXP 1",
+                              "EXP 2", "EXP 3", "N inst",
+                              "t-tests 1: pv", "t-tests 2: pv")
> # Rounding
> rlms_tbl_res[, c("EXP 1", "EXP 2", "EXP 3")] <-
+   round(rlms_tbl_res[, c("EXP 1", "EXP 2", "EXP 3")], 2)
>
```

Table 1: Comparison between Initial and Corrected Versions of the Experience Variable

```
> print(xtable(rlms_tbl_res),
+   floating=FALSE,
+   include.rownames=F)
```

Cohort	N resp	EXP 1	EXP 2	EXP 3	N inst	t-tests 1: pv	t-tests 2: pv
1	6667	5.91	6.25	18.91	6667	0.01	0.00
2	3463	6.53	6.97	19.26	6926	0.02	0.00
3	2793	7.60	8.10	20.41	8379	0.03	0.00
4	2556	8.41	8.90	20.35	10224	0.04	0.00
5	1729	7.99	8.82	20.35	8645	0.00	0.00
6	1532	8.56	9.43	20.02	9192	0.00	0.00
7	1273	9.41	10.36	20.45	8911	0.00	0.00
8	1177	10.14	10.88	21.13	9416	0.02	0.00
9	910	10.66	11.75	21.32	8190	0.00	0.00
10	597	10.58	12.08	21.56	5970	0.00	0.00
11	541	10.52	12.04	20.85	5951	0.00	0.00
12	495	11.50	12.93	22.42	5940	0.00	0.00
13	508	12.27	13.95	22.46	6604	0.00	0.00
14	325	12.63	14.10	22.47	4550	0.01	0.00
15	339	13.60	15.28	22.95	5085	0.01	0.00
16	245	12.59	15.28	22.07	3920	0.00	0.00
17	235	13.77	15.99	22.91	3995	0.00	0.00
18	216	14.47	15.72	22.97	3888	0.06	0.00
19	212	14.52	16.34	23.67	4028	0.00	0.00
20	153	14.97	16.84	23.06	3060	0.00	0.00
21	96	16.34	17.13	25.34	2016	0.33	0.00
22	97	17.05	18.92	25.06	2134	0.04	0.00
23	119	19.63	20.88	25.87	2737	0.16	0.00

Finally, two socio-demographic variables were incorporated into the analysis, namely gender and nationality. We introduced gender in the form of a dummy variable with "1" standing for females, "0" - for males. Likewise, nationality reflected if a person did not identify him/herself as Russian.

Findings

Figure 1 displays rates of returns to higher and vocational education in Russia in 1994-2018. The results suggest that on average wage premiums to university schooling in Russia are roughly 3-5 times greater than to vocational schooling depending on the year under focus. Overall, there is a moderate curved growth in both return types, achieving their peak in the early 2000s (83% for higher education and 26% for vocational education compared to the average earnings of workers with the secondary level), which is followed by a downward pattern (see Figure 1). This goes in line with the previous meta-analytic research (Lukyanova 2010).

```
> df_mincer <- df[, c("IDIND", "YEAR", "edu_4", "wage", "exper", "non_russ", "female
> df_mincer$exper <- as.numeric(df_mincer$exper)
> # summary(df_mincer)
> # df[which(df$wage == 0), "ID_I"]
>
> # Filtering the missings left
> df_mincer <- df_mincer %>%
+   filter(!is.na(wage) & !is.na(exper) & wage > 0)
> # Empty list where the regression output will be written
> lm_mincer_all <- vector("list", length(unique(df_mincer$YEAR)))
> lm_mincer_f = lm_mincer_m = lm_mincer_rus = lm_mincer_nrus = lm_mincer_all
> vec_year <- unique(df_mincer$YEAR)
> # Looping over each year
> # all
> for(i in seq(length(vec_year))){
+   lm_mincer_all[[i]] <- lm(log(wage) ~ edu_4 + exper + I(exper^2) + non_russ + fem
+                           data = df_mincer[df_mincer$YEAR == vec_year[i],])
+ }
> names(lm_mincer_all) <- vec_year
> # by gender
> for(i in seq(length(vec_year))){
+    lm_mincer_f[[i]] <- lm(log(wage) ~ edu_4 + exper + I(exper^2) + non_russ,
+                           data = df_mincer[df_mincer$YEAR == vec_year[i] &
+                                                df_mincer$female == 1,])
+    lm_mincer_m[[i]] <- lm(log(wage) ~ edu_4 + exper + I(exper^2) + non_russ,
+                                   data = df_mincer[df_mincer$YEAR == vec_year[i] &
+                                                        df_mincer$female == 0,])
+ }
> # by nationality
> for(i in seq(length(vec_year))){
+   lm_mincer_rus[[i]] <- lm(log(wage) ~ edu_4 + exper + I(exper^2) + female,
+                             data = df_mincer[df_mincer$YEAR == vec_year[i] &
+                                                  df_mincer$non_russ == 0,])
+   lm_mincer_nrus[[i]] <- lm(log(wage) ~ edu_4 + exper + I(exper^2) + female,
+                             data = df_mincer[df_mincer$YEAR == vec_year[i] &
+                                                  df_mincer$non_russ == 1,])
+ }
> names(lm_mincer_f) <- vec_year
> names(lm_mincer_m) <- vec_year
> names(lm_mincer_rus) <- vec_year
> names(lm_mincer_nrus) <- vec_year
> # Summary if needed
> smry_all <- lapply(lm_mincer_all, summary)
> smry_f <- lapply(lm_mincer_f, summary)
> smry_m <- lapply(lm_mincer_m, summary)
> smry_rus <- lapply(lm_mincer_rus, summary)
> smry_nrus <- lapply(lm_mincer_nrus, summary)
> # Calculating returns by year for higher and vocational education
```

```
>
> RoREs <- as.data.frame((matrix(ncol = 21, nrow = length(vec_year))))
> colnames(RoREs) <-  c("YEAR", "returns_to_HE_all", "p_for_HE_all", "returns_to_VE_
+                       "returns_to_HE_f", "p_for_HE_f", "returns_to_VE_f", "p_for_V
+                       "returns_to_HE_m", "p_for_HE_m", "returns_to_VE_m", "p_for_V
+                       "returns_to_HE_rus", "p_for_HE_rus", "returns_to_VE_rus", "p
+                       "returns_to_HE_nrus", "p_for_HE_nrus", "returns_to_VE_nrus",
> # A function for percentages
> percent <- function(x, digits = 1, format = "f", ...) {
+   paste0(formatC(100 * x, format = format, digits = digits, ...), "%")
+ }
> # Obtaining the values
> for (i in seq(length(vec_year))){
+   RoREs[i,] <- c(vec_year[i], (percent(exp(smry_all[[i]]$coefficients[3,1]) - 1)),
+                  formatC(smry_all[[i]]$coefficients[3,4], digits = 2),
+                  (percent(exp(smry_all[[i]]$coefficients[2,1]) - 1)),
+                  formatC(smry_all[[i]]$coefficients[2,4], digits = 2),
+
+                  (percent(exp(smry_f[[i]]$coefficients[3,1]) - 1)),
+                  formatC(smry_f[[i]]$coefficients[3,4], digits = 2),
+                  (percent(exp(smry_f[[i]]$coefficients[2,1]) - 1)),
+                  formatC(smry_f[[i]]$coefficients[2,4], digits = 2),
+
+                  (percent(exp(smry_m[[i]]$coefficients[3,1]) - 1)),
+                  formatC(smry_m[[i]]$coefficients[3,4], digits = 2),
+                  (percent(exp(smry_m[[i]]$coefficients[2,1]) - 1)),
+                  formatC(smry_m[[i]]$coefficients[2,4], digits = 2),
+
+                  (percent(exp(smry_rus[[i]]$coefficients[3,1]) - 1)),
+                  formatC(smry_rus[[i]]$coefficients[3,4], digits = 2),
+                  (percent(exp(smry_rus[[i]]$coefficients[2,1]) - 1)),
+                  formatC(smry_rus[[i]]$coefficients[2,4], digits = 2),
+
+                  (percent(exp(smry_nrus[[i]]$coefficients[3,1]) - 1)),
+                  formatC(smry_nrus[[i]]$coefficients[3,4], digits = 2),
+                  (percent(exp(smry_nrus[[i]]$coefficients[2,1]) - 1)),
+                  formatC(smry_nrus[[i]]$coefficients[2,4], digits = 2))
+ }
> x_axis <- c(c(1994, 1996), seq(2000, 2018, 2))
> # Converting to data.table and melting in order to visualize
> RoREs <- as.data.table(RoREs)
> RoREs_1 <- melt(RoREs, measure=c("returns_to_HE_all", "returns_to_VE_all"))
> RoREs_1$value <- as.numeric(substr(RoREs_1$value, 1, nchar(RoREs_1$value)-1))
> # Plotting all
> p1 <- ggplot(RoREs_1, aes(YEAR, value, group = variable, color = variable)) +
+   geom_point(size = 3) +
+   geom_smooth(se = F) +
+   scale_y_continuous(limits = c(-10, 110), breaks = seq(-50, 110, 10)) +
```

```
+   theme(legend.title = element_blank(),
+         legend.position = "bottom",
+         panel.grid.minor = element_blank(),
+         axis.text.x = element_text(angle = 30, hjust = 1, size = 16),
+         axis.text.y = element_text(size = 16),
+         axis.title = element_text(size = 16),
+         legend.text = element_text(size = 16),
+         legend.key = element_rect(size = 16)) +
+   labs(color = "Education level") +
+   scale_color_manual(labels = c("Higher education", "Vocational education"),
+                      values = c("darkgreen", "red")) +
+   scale_x_discrete(breaks = x_axis) +
+   ylab("Rate of returns, %") +
+   xlab("Year")
> # Saving
> ggsave("p1.png")
```

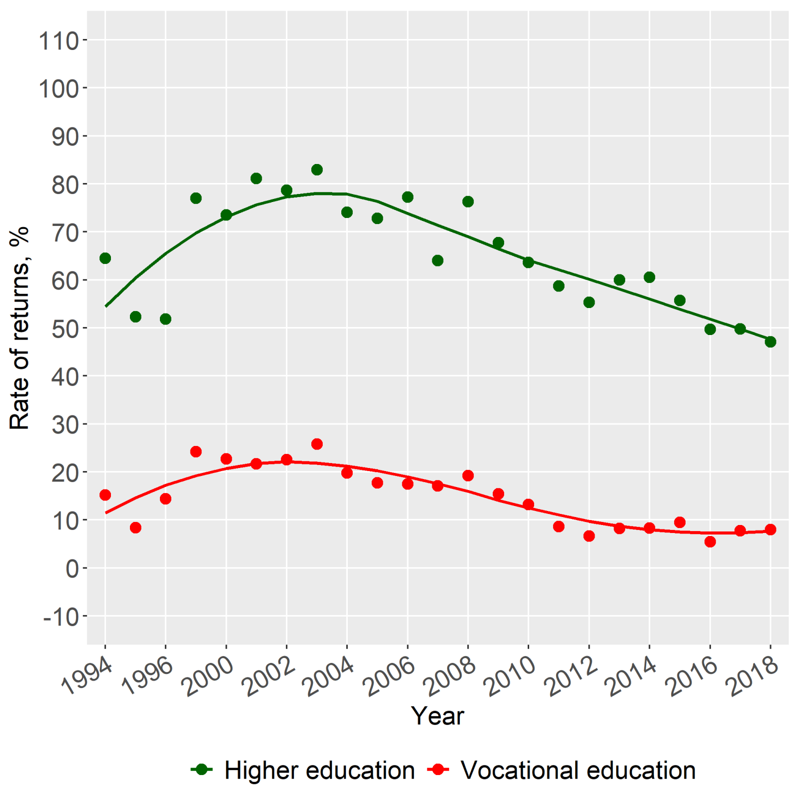
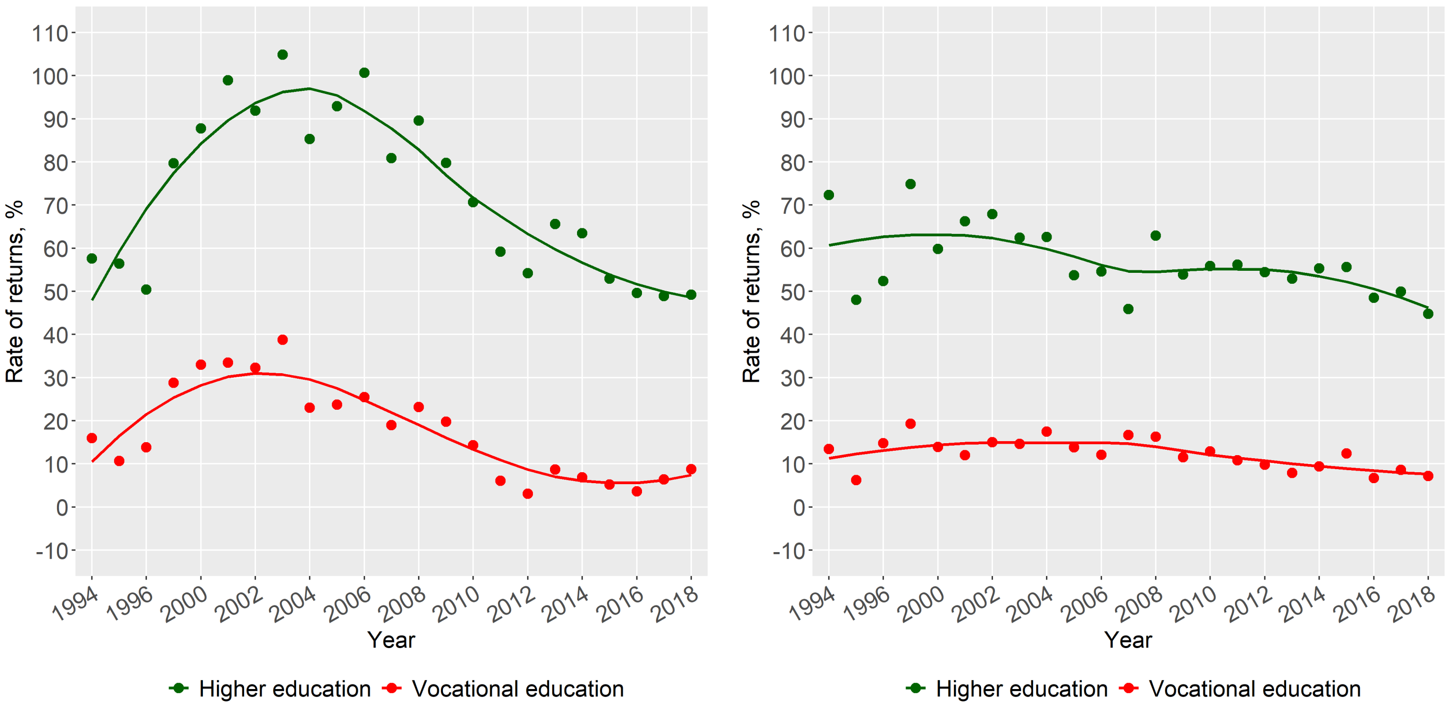


Figure 1: Rates of Returns to Higher and Vocational Education in Russia, RLMS 1994-2018

Notably, gender-based trends in Russia have a different shape across time with regard to schooling premiums. Particularly, males' payoffs to higher education (varying from 45% to 76%) turn out to be on a slightly decreasing slope, whereas women' returns are described by an inversely U-shaped pattern, reaching their maximum of 104% in 2001. Within the last roughly 5 years wage premiums to higher education for women have stabilized around the level of men (50%). A similar comparative picture is observed with respect to vocational education, however, the described regularities are way less pronounced (see Figure 2): returns for males are almost flat within the time period under focus and the parabolic association for females is tangibly less concave. The overall outcome concerning payoffs to schooling isolated by gender has been confirmed in a similar fashion by past studies (e.g., Cheidvasser and Benítez-Silva 2007; Lukyanova 2010).

```
> x_axis <- c(c(1994, 1996), seq(2000, 2018, 2))
> # The same procedure for females
> RoREs_2a <- melt(RoREs, measure=c("returns_to_HE_f", "returns_to_VE_f"))
> RoREs_2a$value <- as.numeric(substr(RoREs_2a$value, 1, nchar(RoREs_2a$value)-1))
> # Plotting females
> p2a <- ggplot(RoREs_2a, aes(YEAR, value, group = variable, color = variable)) +
+   geom_smooth(se = F) +
+   geom_point(size = 3) +
+   scale_y_continuous(limits = c(-10, 110), breaks = seq(-50, 110, 10)) +
+   theme(legend.title = element_blank(),
+         legend.position = "bottom",
+         panel.grid.minor = element_blank(),
+         axis.text.x = element_text(angle = 30, hjust = 1, size = 16),
+         axis.text.y = element_text(size = 16),
+         axis.title = element_text(size = 16),
+         legend.text = element_text(size = 16),
+         legend.key = element_rect(size = 16)) +
+   scale_color_manual(labels = c("Higher education", "Vocational education"),
+                      values = c("darkgreen", "red")) +
+   scale_x_discrete(breaks = x_axis) +
+   ylab("Rate of returns, %") +
+   xlab("Year")
> # Saving
> ggsave("p2a.png")
> # The same procedure for males
> RoREs_2b <- melt(RoREs, measure=c("returns_to_HE_m", "returns_to_VE_m"))
> RoREs_2b$value <- as.numeric(substr(RoREs_2b$value, 1, nchar(RoREs_2b$value)-1))
> # Plotting males
> p2b <- ggplot(RoREs_2b, aes(YEAR, value, group = variable, color = variable)) +
+   geom_smooth(se = F) +
+   geom_point(size = 3) +
+   scale_y_continuous(limits = c(-10, 110), breaks = seq(-50, 110, 10)) +
+   theme(legend.title = element_blank(),
+         legend.position = "bottom",
+         panel.grid.minor = element_blank(),
+         axis.text.x = element_text(angle = 30, hjust = 1, size = 16),
+         axis.text.y = element_text(size = 16),
+         axis.title = element_text(size = 16),
+         legend.text = element_text(size = 16),
+         legend.key = element_rect(size = 16)) +
+   scale_color_manual(labels = c("Higher education", "Vocational education"),
+                      values = c("darkgreen", "red")) +
```

```
+   scale_x_discrete(breaks = x_axis) +
+   ylab("Rate of returns, %") +
+   xlab("Year")
> # Saving
> ggsave("p2b.png")
```



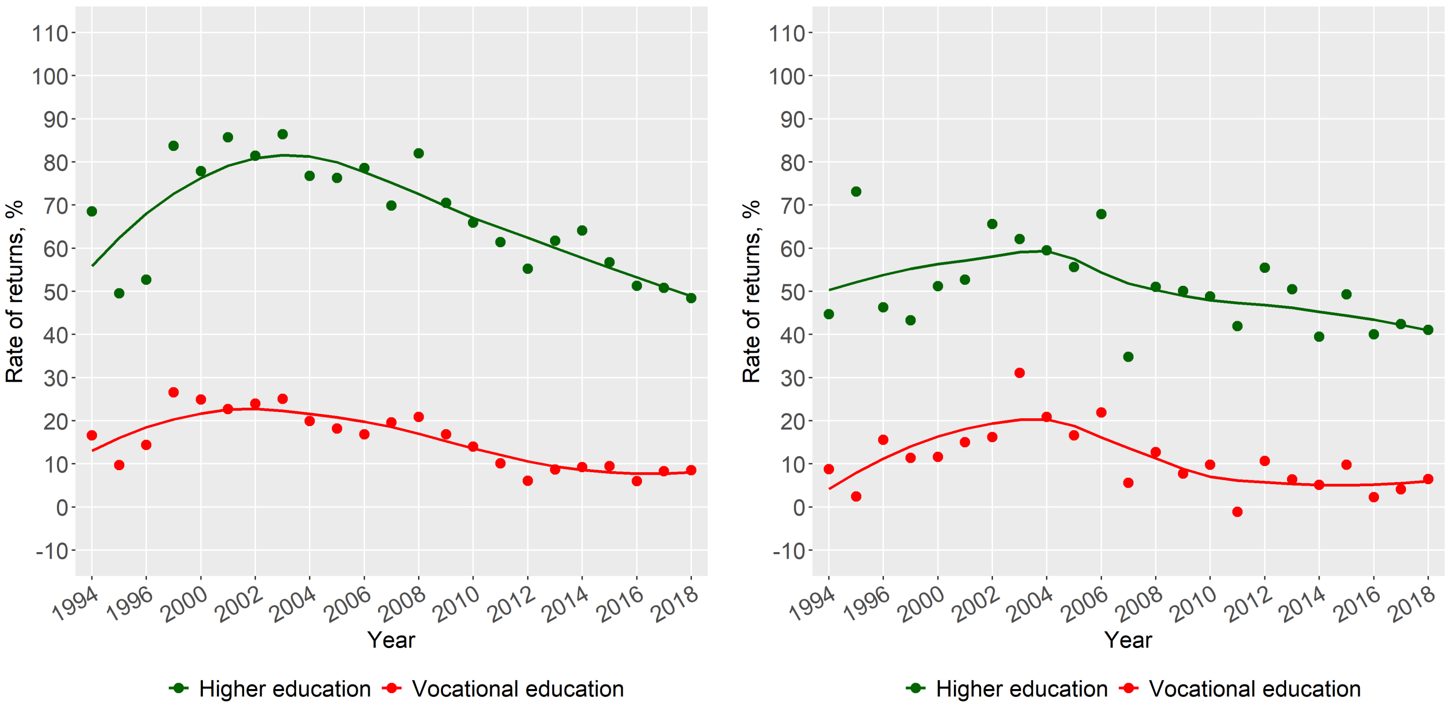
(a) Females **(b)** Males

Figure 2: Rates of Returns to Higher and Vocational Education in Russia, RLMS 1994-2018

For Russians payoffs to higher and vocational education are characterized by a pattern almost identical to the one uncovered for the whole population. As for non-Russians, the estimates of wage advantages regarding people with university education level compared to those with only secondary level are not statistically significant in the majority of time periods investigated (except for 2002-2006 and 2008). Nevertheless, the payoffs to vocational education for those who identified themselves as non-Russians are significant and the respective time trend is loosely distinguishable from the one registered for Russians. In other words, nationality seem to affect returns to higher education, but does not play a similar part with respect to vocational education.

```
> # The same procedure for Russians
> RoREs_3a <- melt(RoREs, measure=c("returns_to_HE_rus", "returns_to_VE_rus"))
> RoREs_3a$value <- as.numeric(substr(RoREs_3a$value, 1, nchar(RoREs_3a$value)-1))
> # Plotting Russians
> p3a <- ggplot(RoREs_3a, aes(YEAR, value, group = variable, color = variable)) +
+   geom_smooth(se = F) +
+   geom_point(size = 3) +
+   scale_y_continuous(limits = c(-10, 110), breaks = seq(-50, 110, 10)) +
+   theme(legend.title = element_blank(),
+         legend.position = "bottom",
+         panel.grid.minor = element_blank(),
+         axis.text.x = element_text(angle = 30, hjust = 1, size = 16),
+         axis.text.y = element_text(size = 16),
+         axis.title = element_text(size = 16),
+         legend.text = element_text(size = 16),
```

```
+         legend.key = element_rect(size = 16)) +
+   scale_color_manual(labels = c("Higher education", "Vocational education"),
+                      values = c("darkgreen", "red")) +
+   scale_x_discrete(breaks = x_axis) +
+   ylab("Rate of returns, %") +
+   xlab("Year")
> # Saving
> ggsave("p3a.png")
> # The same procedure for non-Russians
> RoREs_3b <- melt(RoREs, measure=c("returns_to_HE_nrus", "returns_to_VE_nrus"))
> RoREs_3b$value <- as.numeric(substr(RoREs_3b$value, 1, nchar(RoREs_3b$value)-1))
> # Plotting non-Russians
> p3b <- ggplot(RoREs_3b, aes(YEAR, value, group = variable, color = variable)) +
+   geom_smooth(se = F) +
+   geom_point(size = 3) +
+   scale_y_continuous(limits = c(-10, 110), breaks = seq(-50, 110, 10)) +
+   theme(legend.title = element_blank(),
+         legend.position = "bottom",
+         panel.grid.minor = element_blank(),
+         axis.text.x = element_text(angle = 30, hjust = 1, size = 16),
+         axis.text.y = element_text(size = 16),
+         axis.title = element_text(size = 16),
+         legend.text = element_text(size = 16),
+         legend.key = element_rect(size = 16)) +
+   scale_color_manual(labels = c("Higher education", "Vocational education"),
+                      values = c("darkgreen", "red")) +
+   scale_x_discrete(breaks = x_axis) +
+   ylab("Rate of returns, %") +
+   xlab("Year")
> # Saving
> ggsave("p3b.png")
```



(a) Russians **(b)** Non-Russians

Figure 3: Rates of Returns to Higher and Vocational Education in Russia, RLMS 1994-2018

References

Kozyreva, Polina and Klara Sabirianova Peter (2015). "081. Economic Change in Russia: Twenty Years of the Russian Longitudinal Monitoring Survey". In: *Economics of Transition* 23.2, pp. 293–298.

Mincer, Jacob A. (1974). "082. Schooling, Experience, and Earnings". In: *NBER Books*.

Appendix

```
> for (i in 1:length(vec_year)){
+   stargazer(lm_mincer_all[i],
+             lm_mincer_m[i],
+             lm_mincer_f[i],
+             lm_mincer_rus[i],
+             lm_mincer_nrus[i],
+             type = "latex",
+             column.labels = c("Total Sample",
+                               "Males",
+                               "Females",
+                               "Russians",
+                               "Non-Russians"),
+             covariate.labels = c("Vocational education",
+                                  "Higher education",
+                                  "Experience",
+                                  "Experience squared",
+                                  "Non-Russian",
+                                  "Female"),
+             title = paste0("Results of Mincer Analysis, RLMS ",
+                            as.character(vec_year[i])),
+             dep.var.caption = "",
+             dep.var.labels.include = F,
+             df = F)
+   cat("\n\\newpage\n")
+ }
```

Table 2: Results of Mincer Analysis, RLMS 1994

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.141***	0.126**	0.149***	0.153***	0.085
	(0.042)	(0.064)	(0.055)	(0.046)	(0.102)
Higher education	0.498***	0.544***	0.455***	0.522***	0.370***
	(0.047)	(0.072)	(0.062)	(0.051)	(0.121)
Experience	0.007	0.008	0.004	0.001	0.036***
	(0.005)	(0.008)	(0.007)	(0.006)	(0.013)
Experience squared	−0.0002	−0.0004	0.0001	0.00000	−0.001**
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0004)
Non-Russian	−0.004	−0.056	0.042		
	(0.045)	(0.069)	(0.059)		
Female	−0.499***			−0.513***	−0.438***
	(0.033)			(0.036)	(0.085)
Constant	12.104***	12.132***	11.588***	12.128***	11.979***
	(0.046)	(0.065)	(0.059)	(0.050)	(0.110)
Observations	3,041	1,394	1,647	2,564	477
R^2	0.103	0.049	0.041	0.109	0.083
Adjusted R^2	0.101	0.045	0.038	0.108	0.073
Residual Std. Error	0.906	0.957	0.858	0.905	0.906
F Statistic	57.786***	14.271***	14.057***	62.886***	8.534***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Results of Mincer Analysis, RLMS 1995

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.081*	0.060	0.101*	0.092*	0.023
	(0.044)	(0.066)	(0.060)	(0.049)	(0.109)
Higher education	0.421***	0.392***	0.447***	0.402***	0.549***
	(0.048)	(0.072)	(0.066)	(0.053)	(0.127)
Experience	0.002	−0.002	0.006	0.002	0.004
	(0.006)	(0.009)	(0.008)	(0.006)	(0.015)
Experience squared	−0.0001	−0.00004	−0.0001	−0.0001	−0.00005
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.001)
Non-Russian	−0.041	−0.061	−0.021		
	(0.047)	(0.070)	(0.064)		
Female	−0.431***			−0.437***	−0.386***
	(0.035)			(0.038)	(0.090)
Constant	12.882***	12.936***	12.400***	12.893***	12.790***
	(0.049)	(0.067)	(0.064)	(0.052)	(0.118)
Observations	2,690	1,235	1,455	2,262	428
R^2	0.085	0.032	0.040	0.082	0.105
Adjusted R^2	0.083	0.028	0.037	0.080	0.094
Residual Std. Error	0.897	0.923	0.875	0.891	0.925
F Statistic	41.523***	8.181***	12.128***	40.396***	9.873***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: Results of Mincer Analysis, RLMS 1996

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.134^{**}	0.138^{*}	0.129^{*}	0.134^{**}	0.145
	(0.052)	(0.080)	(0.069)	(0.058)	(0.118)
Higher education	0.417^{***}	0.421^{***}	0.408^{***}	0.423^{***}	0.381^{***}
	(0.056)	(0.086)	(0.074)	(0.062)	(0.133)
Experience	0.008	0.008	0.007	0.007	0.019
	(0.006)	(0.010)	(0.008)	(0.007)	(0.018)
Experience squared	-0.0004^{*}	-0.001^{*}	-0.0002	-0.0003	-0.001
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.001)
Non-Russian	0.039	0.043	0.034		
	(0.055)	(0.083)	(0.074)		
Female	-0.477^{***}			-0.475^{***}	-0.499^{***}
	(0.039)			(0.043)	(0.100)
Constant	13.208^{***}	13.230^{***}	12.713^{***}	13.205^{***}	13.245^{***}
	(0.059)	(0.084)	(0.075)	(0.064)	(0.130)
Observations	2,281	1,033	1,248	1,941	340
R^2	0.087	0.036	0.029	0.084	0.104
Adjusted R^2	0.084	0.031	0.025	0.082	0.091
Residual Std. Error	0.929	0.976	0.888	0.934	0.906
F Statistic	35.992^{***}	7.709^{***}	7.501^{***}	35.671^{***}	7.774^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Results of Mincer Analysis, RLMS 1998

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.217***	0.176***	0.253***	0.236***	0.108
	(0.037)	(0.054)	(0.050)	(0.040)	(0.093)
Higher education	0.571***	0.559***	0.586***	0.608***	0.360***
	(0.041)	(0.061)	(0.054)	(0.044)	(0.106)
Experience	0.015***	0.010	0.020***	0.014***	0.022*
	(0.005)	(0.007)	(0.006)	(0.005)	(0.011)
Experience squared	−0.0004***	−0.0004*	−0.0004**	−0.0004**	−0.001*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Non-Russian	−0.032	−0.073	−0.0004		
	(0.038)	(0.059)	(0.049)		
Female	−0.470***			−0.483***	−0.406***
	(0.028)			(0.030)	(0.071)
Constant	6.364***	6.447***	5.817***	6.358***	6.373***
	(0.043)	(0.061)	(0.057)	(0.046)	(0.107)
Observations	3,101	1,434	1,667	2,614	487
R^2	0.133	0.064	0.079	0.141	0.095
Adjusted R^2	0.131	0.061	0.076	0.140	0.086
Residual Std. Error	0.768	0.806	0.733	0.767	0.770
F Statistic	79.074***	19.512***	28.374***	85.755***	10.119***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Results of Mincer Analysis, RLMS 2000

	Total Sample (1)	Males (2)	Females (3)	Russians (4)	Non-Russians (5)
Vocational education	0.205***	0.131**	0.285***	0.222***	0.110
	(0.039)	(0.056)	(0.053)	(0.042)	(0.101)
Higher education	0.551***	0.469***	0.630***	0.576***	0.413***
	(0.043)	(0.064)	(0.057)	(0.046)	(0.113)
Experience	0.009*	0.001	0.016**	0.008	0.014
	(0.005)	(0.008)	(0.007)	(0.006)	(0.016)
Experience squared	−0.0002	−0.0002	−0.0002	−0.0002	−0.0004
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.001)
Non-Russian	−0.001	−0.016	0.002		
	(0.042)	(0.065)	(0.055)		
Female	−0.534***			−0.536***	−0.511***
	(0.030)			(0.032)	(0.081)
Constant	7.071***	7.207***	6.398***	7.059***	7.120***
	(0.046)	(0.065)	(0.062)	(0.049)	(0.123)
Observations	3,213	1,475	1,738	2,765	448
R^2	0.128	0.042	0.078	0.133	0.104
Adjusted R^2	0.127	0.038	0.075	0.131	0.094
Residual Std. Error	0.830	0.861	0.799	0.829	0.835
F Statistic	78.748***	12.777***	29.233***	84.557***	10.307***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Results of Mincer Analysis, RLMS 2001

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.196***	0.113**	0.289***	0.204***	0.140
	(0.036)	(0.052)	(0.050)	(0.039)	(0.101)
Higher education	0.594***	0.508***	0.688***	0.619***	0.423***
	(0.039)	(0.059)	(0.053)	(0.042)	(0.110)
Experience	0.0002	−0.0004	−0.0002	−0.002	0.017
	(0.005)	(0.007)	(0.006)	(0.005)	(0.015)
Experience squared	−0.00001	−0.0001	0.0001	0.00004	−0.0004
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0005)
Non-Russian	−0.027	−0.103*	0.025		
	(0.040)	(0.063)	(0.051)		
Female	−0.463***			−0.478***	−0.361***
	(0.027)			(0.029)	(0.081)
Constant	7.491***	7.591***	6.921***	7.501***	7.383***
	(0.041)	(0.058)	(0.057)	(0.044)	(0.116)
Observations	3,604	1,673	1,931	3,128	476
R^2	0.125	0.056	0.091	0.136	0.066
Adjusted R^2	0.124	0.053	0.089	0.135	0.056
Residual Std. Error	0.813	0.853	0.774	0.805	0.858
F Statistic	85.935***	19.738***	38.626***	98.701***	6.661***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 8: Results of Mincer Analysis, RLMS 2002

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.203***	0.140***	0.280***	0.215***	0.150*
	(0.033)	(0.047)	(0.046)	(0.036)	(0.082)
Higher education	0.581***	0.518***	0.652***	0.595***	0.504***
	(0.035)	(0.052)	(0.049)	(0.039)	(0.092)
Experience	0.008*	0.001	0.013**	0.006	0.019
	(0.004)	(0.007)	(0.006)	(0.005)	(0.013)
Experience squared	−0.0001	−0.0001	−0.0002	−0.0001	−0.0004
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0004)
Non-Russian	0.069*	0.066	0.062		
	(0.035)	(0.055)	(0.046)		
Female	−0.444***			−0.441***	−0.459***
	(0.025)			(0.026)	(0.068)
Constant	7.755***	7.863***	7.193***	7.753***	7.815***
	(0.038)	(0.053)	(0.052)	(0.041)	(0.097)
Observations	3,803	1,748	2,055	3,286	517
R^2	0.133	0.062	0.098	0.134	0.125
Adjusted R^2	0.131	0.059	0.096	0.133	0.116
Residual Std. Error	0.748	0.774	0.722	0.748	0.749
F Statistic	96.883***	22.835***	44.584***	101.827***	14.558***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Results of Mincer Analysis, RLMS 2003

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.229***	0.136***	0.328***	0.224***	0.271***
	(0.033)	(0.046)	(0.046)	(0.035)	(0.083)
Higher education	0.604***	0.485***	0.717***	0.623***	0.483***
	(0.035)	(0.052)	(0.048)	(0.038)	(0.090)
Experience	0.008**	0.008	0.009*	0.009**	0.006
	(0.004)	(0.007)	(0.006)	(0.005)	(0.012)
Experience squared	−0.0002	−0.0004*	−0.0001	−0.0002	−0.0002
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0004)
Non-Russian	0.066*	0.049	0.075		
	(0.035)	(0.053)	(0.047)		
Female	−0.491***			−0.495***	−0.468***
	(0.024)			(0.026)	(0.066)
Constant	7.970***	8.086***	7.359***	7.963***	8.078***
	(0.038)	(0.054)	(0.052)	(0.041)	(0.099)
Observations	3,857	1,765	2,092	3,335	522
R^2	0.151	0.060	0.110	0.154	0.137
Adjusted R^2	0.150	0.057	0.108	0.153	0.129
Residual Std. Error	0.747	0.761	0.731	0.749	0.731
F Statistic	114.015***	22.345***	51.464***	121.088***	16.378***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Results of Mincer Analysis, RLMS 2004

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.181***	0.161***	0.207***	0.181***	0.189**
	(0.031)	(0.044)	(0.044)	(0.034)	(0.080)
Higher education	0.554***	0.486***	0.617***	0.570***	0.467***
	(0.033)	(0.049)	(0.046)	(0.036)	(0.086)
Experience	0.0001	−0.006	0.005	0.001	−0.009
	(0.004)	(0.006)	(0.005)	(0.004)	(0.011)
Experience squared	0.00003	0.00002	0.00001	−0.00002	0.0004
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0003)
Non-Russian	0.039	0.037	0.036		
	(0.033)	(0.050)	(0.044)		
Female	−0.497***			−0.497***	−0.505***
	(0.023)			(0.025)	(0.063)
Constant	8.294***	8.380***	7.708***	8.284***	8.393***
	(0.036)	(0.050)	(0.049)	(0.039)	(0.089)
Observations	3,967	1,824	2,143	3,439	528
R^2	0.157	0.061	0.101	0.159	0.151
Adjusted R^2	0.156	0.058	0.099	0.158	0.143
Residual Std. Error	0.711	0.730	0.692	0.712	0.707
F Statistic	123.291***	23.549***	48.085***	129.807***	18.629***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 11: Results of Mincer Analysis, RLMS 2005

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.163***	0.129***	0.212***	0.167***	0.154*
	(0.031)	(0.043)	(0.044)	(0.034)	(0.079)
Higher education	0.547***	0.430***	0.657***	0.567***	0.442***
	(0.033)	(0.048)	(0.046)	(0.036)	(0.085)
Experience	−0.003	−0.007	0.001	−0.002	−0.008
	(0.004)	(0.006)	(0.005)	(0.004)	(0.011)
Experience squared	0.0001	0.00002	0.0001	0.0001	0.0002
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0003)
Non-Russian	0.063*	0.016	0.109**		
	(0.033)	(0.049)	(0.044)		
Female	−0.503***			−0.517***	−0.427***
	(0.023)			(0.025)	(0.064)
Constant	8.532***	8.644***	7.906***	8.524***	8.633***
	(0.036)	(0.050)	(0.051)	(0.039)	(0.092)
Observations	3,913	1,801	2,112	3,367	546
R^2	0.162	0.053	0.120	0.169	0.124
Adjusted R^2	0.160	0.050	0.118	0.167	0.116
Residual Std. Error	0.704	0.723	0.684	0.700	0.733
F Statistic	125.554***	20.059***	57.218***	136.313***	15.263***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Results of Mincer Analysis, RLMS 2006

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.161***	0.114***	0.227***	0.155***	0.198***
	(0.027)	(0.038)	(0.039)	(0.029)	(0.069)
Higher education	0.572***	0.436***	0.696***	0.580***	0.518***
	(0.029)	(0.043)	(0.041)	(0.032)	(0.075)
Experience	−0.003	0.002	−0.006	−0.002	−0.005
	(0.003)	(0.005)	(0.004)	(0.004)	(0.009)
Experience squared	0.00003	−0.0002	0.0002	0.00002	0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)
Non-Russian	0.092***	0.066	0.117***		
	(0.029)	(0.044)	(0.039)		
Female	−0.456***			−0.463***	−0.411***
	(0.020)			(0.022)	(0.054)
Constant	8.733***	8.791***	8.200***	8.735***	8.819***
	(0.031)	(0.042)	(0.043)	(0.033)	(0.078)
Observations	4,804	2,172	2,632	4,179	625
R^2	0.163	0.059	0.133	0.164	0.151
Adjusted R^2	0.162	0.057	0.131	0.163	0.144
Residual Std. Error	0.681	0.694	0.667	0.684	0.659
F Statistic	156.012***	27.246***	80.635***	163.550***	22.059***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Results of Mincer Analysis, RLMS 2007

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.158***	0.155***	0.174***	0.179***	0.054
	(0.026)	(0.035)	(0.037)	(0.028)	(0.066)
Higher education	0.495***	0.378***	0.593***	0.530***	0.298***
	(0.028)	(0.039)	(0.039)	(0.030)	(0.072)
Experience	−0.001	0.002	−0.003	−0.003	0.011
	(0.003)	(0.005)	(0.004)	(0.003)	(0.009)
Experience squared	−0.00005	−0.0002*	0.0001	0.00000	−0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)
Non-Russian	0.045	−0.019	0.110***		
	(0.028)	(0.041)	(0.038)		
Female	−0.429***			−0.447***	−0.318***
	(0.019)			(0.020)	(0.053)
Constant	8.955***	9.000***	8.476***	8.953***	8.996***
	(0.029)	(0.040)	(0.040)	(0.031)	(0.076)
Observations	4,726	2,153	2,573	4,136	590
R^2	0.152	0.050	0.113	0.161	0.097
Adjusted R^2	0.150	0.048	0.111	0.160	0.089
Residual Std. Error	0.640	0.641	0.636	0.641	0.631
F Statistic	140.495***	22.629***	65.411***	158.888***	12.539***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 14: Results of Mincer Analysis, RLMS 2008

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.175***	0.151***	0.209***	0.189***	0.120*
	(0.027)	(0.037)	(0.040)	(0.030)	(0.069)
Higher education	0.567***	0.488***	0.640***	0.599***	0.412***
	(0.029)	(0.041)	(0.041)	(0.032)	(0.072)
Experience	−0.0002	−0.001	0.0001	−0.001	0.002
	(0.003)	(0.005)	(0.005)	(0.004)	(0.009)
Experience squared	−0.0001	−0.0002	−0.00000	−0.0001	−0.0002
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)
Non-Russian	0.043	−0.031	0.116***		
	(0.028)	(0.040)	(0.039)		
Female	−0.453***			−0.476***	−0.322***
	(0.020)			(0.022)	(0.053)
Constant	9.176***	9.249***	8.650***	9.173***	9.235***
	(0.032)	(0.043)	(0.044)	(0.034)	(0.081)
Observations	4,827	2,170	2,657	4,140	687
R^2	0.162	0.079	0.114	0.172	0.107
Adjusted R^2	0.161	0.077	0.112	0.171	0.101
Residual Std. Error	0.683	0.680	0.683	0.681	0.689
F Statistic	155.203***	37.314***	68.048***	172.104***	16.395***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 15: Results of Mincer Analysis, RLMS 2009

	Total Sample (1)	Males (2)	Females (3)	Russians (4)	Non-Russians (5)
Vocational education	0.143^{***}	0.109^{***}	0.181^{***}	0.155^{***}	0.075
	(0.027)	(0.037)	(0.040)	(0.029)	(0.071)
Higher education	0.517^{***}	0.431^{***}	0.587^{***}	0.534^{***}	0.406^{***}
	(0.028)	(0.040)	(0.040)	(0.030)	(0.076)
Experience	0.012^{***}	0.008	0.016^{***}	0.011^{***}	0.021^{**}
	(0.003)	(0.005)	(0.004)	(0.003)	(0.010)
Experience squared	-0.0004^{***}	-0.0003^{**}	-0.0004^{***}	-0.0003^{***}	-0.001^{**}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)
Non-Russian	0.054^{*}	0.036	0.066		
	(0.030)	(0.044)	(0.041)		
Female	-0.436^{***}			-0.439^{***}	-0.409^{***}
	(0.019)			(0.020)	(0.055)
Constant	9.174^{***}	9.263^{***}	8.659^{***}	9.172^{***}	9.234^{***}
	(0.031)	(0.042)	(0.043)	(0.033)	(0.082)
Observations	4,803	2,146	2,657	4,267	536
R^2	0.159	0.069	0.112	0.161	0.147
Adjusted R^2	0.158	0.067	0.110	0.160	0.139
Residual Std. Error	0.650	0.640	0.657	0.654	0.621
F Statistic	151.376^{***}	31.959^{***}	66.940^{***}	163.424^{***}	18.259^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

Table 16: Results of Mincer Analysis, RLMS 2010

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.124^{***}	0.121^{***}	0.133^{***}	0.131^{***}	0.094
	(0.021)	(0.030)	(0.031)	(0.023)	(0.062)
Higher education	0.492^{***}	0.444^{***}	0.535^{***}	0.506^{***}	0.398^{***}
	(0.023)	(0.032)	(0.032)	(0.024)	(0.066)
Experience	0.007^{***}	0.009^{**}	0.004	0.005^{*}	0.019^{**}
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)
Experience squared	-0.0002^{***}	-0.0004^{***}	-0.0001	-0.0002^{***}	-0.001^{**}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.083^{***}	0.072^{**}	0.095^{***}		
	(0.024)	(0.036)	(0.033)		
Female	-0.405^{***}			-0.408^{***}	-0.392^{***}
	(0.015)			(0.016)	(0.049)
Constant	9.315^{***}	9.333^{***}	8.889^{***}	9.318^{***}	9.361^{***}
	(0.023)	(0.032)	(0.033)	(0.025)	(0.067)
Observations	7,325	3,318	4,007	6,532	793
R^2	0.149	0.071	0.104	0.153	0.124
Adjusted R^2	0.149	0.069	0.103	0.152	0.119
Residual Std. Error	0.645	0.659	0.633	0.641	0.675
F Statistic	214.077^{***}	50.335^{***}	93.184^{***}	234.981^{***}	22.323^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 17: Results of Mincer Analysis, RLMS 2011

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.082^{***}	0.102^{***}	0.059^{*}	0.097^{***}	-0.011
	(0.020)	(0.027)	(0.031)	(0.022)	(0.059)
Higher education	0.462^{***}	0.446^{***}	0.465^{***}	0.479^{***}	0.350^{***}
	(0.021)	(0.030)	(0.031)	(0.023)	(0.061)
Experience	0.005^{*}	0.003	0.006^{*}	0.004	0.007
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)
Experience squared	-0.0002^{***}	-0.0003^{**}	-0.0002^{**}	-0.0002^{***}	-0.0003
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.050^{**}	0.047	0.054		
	(0.024)	(0.034)	(0.033)		
Female	-0.443^{***}			-0.445^{***}	-0.425^{***}
	(0.015)			(0.016)	(0.046)
Constant	9.492^{***}	9.512^{***}	9.035^{***}	9.483^{***}	9.590^{***}
	(0.022)	(0.029)	(0.032)	(0.023)	(0.065)
Observations	7,166	3,270	3,896	6,415	751
R^2	0.174	0.088	0.100	0.176	0.158
Adjusted R^2	0.173	0.087	0.099	0.175	0.152
Residual Std. Error	0.617	0.608	0.624	0.616	0.629
F Statistic	250.624^{***}	63.004^{***}	86.557^{***}	272.876^{***}	27.904^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 18: Results of Mincer Analysis, RLMS 2012

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.064***	0.094***	0.031	0.059***	0.102*
	(0.021)	(0.027)	(0.031)	(0.022)	(0.057)
Higher education	0.440***	0.434***	0.433***	0.440***	0.441***
	(0.021)	(0.030)	(0.031)	(0.023)	(0.059)
Experience	0.008***	0.011***	0.006*	0.006**	0.021***
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)
Experience squared	−0.0003***	−0.0005***	−0.0002*	−0.0003***	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.041*	0.042	0.038		
	(0.023)	(0.033)	(0.033)		
Female	−0.462***			−0.461***	−0.470***
	(0.015)			(0.016)	(0.045)
Constant	9.643***	9.637***	9.195***	9.655***	9.596***
	(0.022)	(0.029)	(0.033)	(0.024)	(0.062)
Observations	7,427	3,366	4,061	6,603	824
R^2	0.171	0.088	0.090	0.169	0.183
Adjusted R^2	0.170	0.086	0.089	0.169	0.178
Residual Std. Error	0.635	0.621	0.645	0.635	0.633
F Statistic	254.601***	64.701***	80.544***	269.058***	36.546***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 19: Results of Mincer Analysis, RLMS 2013

	Total Sample (1)	Males (2)	Females (3)	Russians (4)	Non-Russians (5)
Vocational education	0.078***	0.076***	0.083***	0.083***	0.062
	(0.021)	(0.028)	(0.032)	(0.023)	(0.058)
Higher education	0.470***	0.424***	0.504***	0.480***	0.409***
	(0.022)	(0.030)	(0.032)	(0.024)	(0.061)
Experience	0.011***	0.010***	0.012***	0.008***	0.029***
	(0.002)	(0.004)	(0.003)	(0.003)	(0.008)
Experience squared	−0.0003***	−0.0005***	−0.0003***	−0.0003***	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	−0.045**	−0.043	−0.049		
	(0.023)	(0.032)	(0.032)		
Female	−0.432***			−0.431***	−0.436***
	(0.015)			(0.016)	(0.045)
Constant	9.684***	9.724***	9.211***	9.693***	9.559***
	(0.023)	(0.031)	(0.033)	(0.025)	(0.063)
Observations	7,324	3,359	3,965	6,440	884
R^2	0.165	0.082	0.111	0.167	0.154
Adjusted R^2	0.164	0.081	0.110	0.166	0.149
Residual Std. Error	0.629	0.625	0.631	0.625	0.659
F Statistic	240.156***	59.946***	99.247***	258.110***	31.873***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 20: Results of Mincer Analysis, RLMS 2014

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.080***	0.089***	0.067**	0.088***	0.050
	(0.023)	(0.031)	(0.034)	(0.024)	(0.065)
Higher education	0.473***	0.440***	0.492***	0.495***	0.333***
	(0.024)	(0.033)	(0.034)	(0.025)	(0.067)
Experience	0.008***	0.011**	0.007**	0.005*	0.037***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.009)
Experience squared	−0.0002***	−0.0004***	−0.0002	−0.0002*	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)
Non-Russian	−0.018	−0.040	0.002		
	(0.025)	(0.037)	(0.034)		
Female	−0.409***			−0.416***	−0.363***
	(0.016)			(0.017)	(0.051)
Constant	9.780***	9.793***	9.359***	9.794***	9.632***
	(0.025)	(0.035)	(0.036)	(0.027)	(0.073)
Observations	6,147	2,794	3,353	5,449	698
R^2	0.160	0.081	0.117	0.168	0.128
Adjusted R^2	0.160	0.080	0.116	0.167	0.121
Residual Std. Error	0.616	0.627	0.606	0.611	0.652
F Statistic	195.451***	49.417***	88.665***	220.189***	20.278***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 21: Results of Mincer Analysis, RLMS 2015

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.091***	0.117***	0.050	0.091***	0.093
	(0.023)	(0.029)	(0.036)	(0.024)	(0.063)
Higher education	0.443***	0.442***	0.425***	0.449***	0.401***
	(0.023)	(0.031)	(0.036)	(0.025)	(0.064)
Experience	0.010***	0.016***	0.007*	0.006**	0.036***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.008)
Experience squared	−0.0003***	−0.001***	−0.0001	−0.0002**	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	−0.039	−0.056*	−0.024		
	(0.024)	(0.033)	(0.034)		
Female	−0.422***			−0.426***	−0.391***
	(0.016)			(0.016)	(0.047)
Constant	9.818***	9.794***	9.426***	9.838***	9.627***
	(0.025)	(0.033)	(0.038)	(0.027)	(0.070)
Observations	6,230	2,844	3,386	5,515	715
R^2	0.163	0.093	0.093	0.165	0.167
Adjusted R^2	0.162	0.091	0.091	0.164	0.161
Residual Std. Error	0.599	0.587	0.607	0.597	0.613
F Statistic	202.247***	58.161***	69.116***	217.040***	28.468***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 22: Results of Mincer Analysis, RLMS 2016

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.052**	0.064**	0.036	0.058**	0.023
	(0.024)	(0.030)	(0.038)	(0.026)	(0.063)
Higher education	0.403***	0.396***	0.403***	0.414***	0.336***
	(0.024)	(0.031)	(0.037)	(0.026)	(0.064)
Experience	0.011***	0.017***	0.007*	0.009***	0.032***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.008)
Experience squared	−0.0003***	−0.001***	−0.0001	−0.0002***	−0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.013	0.016	0.011		
	(0.025)	(0.035)	(0.036)		
Female	−0.406***			−0.406***	−0.404***
	(0.016)			(0.017)	(0.046)
Constant	9.872***	9.849***	9.489***	9.881***	9.800***
	(0.027)	(0.034)	(0.040)	(0.028)	(0.074)
Observations	6,296	2,905	3,391	5,611	685
R^2	0.146	0.083	0.080	0.145	0.167
Adjusted R^2	0.145	0.081	0.079	0.144	0.161
Residual Std. Error	0.619	0.592	0.641	0.621	0.598
F Statistic	179.466***	52.260***	58.799***	190.040***	27.237***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 23: Results of Mincer Analysis, RLMS 2017

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.074***	0.083***	0.062	0.080***	0.040
	(0.024)	(0.029)	(0.040)	(0.026)	(0.061)
Higher education	0.404***	0.404***	0.398***	0.411***	0.354***
	(0.025)	(0.031)	(0.040)	(0.027)	(0.062)
Experience	0.013***	0.019***	0.009**	0.013***	0.020***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.008)
Experience squared	−0.0004***	−0.001***	−0.0002**	−0.0004***	−0.001**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.061**	0.058*	0.063		
	(0.026)	(0.035)	(0.038)		
Female	−0.442***			−0.443***	−0.432***
	(0.016)			(0.017)	(0.044)
Constant	9.920***	9.890***	9.506***	9.919***	9.973***
	(0.027)	(0.034)	(0.043)	(0.029)	(0.072)
Observations	6,355	2,945	3,410	5,715	640
R^2	0.153	0.089	0.066	0.149	0.201
Adjusted R^2	0.152	0.087	0.065	0.148	0.195
Residual Std. Error	0.628	0.578	0.668	0.636	0.550
F Statistic	191.331***	57.371***	48.265***	199.302***	31.864***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 24: Results of Mincer Analysis, RLMS 2018

	Total Sample	Males	Females	Russians	Non-Russians
	(1)	(2)	(3)	(4)	(5)
Vocational education	0.077***	0.070**	0.085**	0.082***	0.063
	(0.023)	(0.031)	(0.036)	(0.025)	(0.067)
Higher education	0.386***	0.370***	0.400***	0.394***	0.345***
	(0.024)	(0.032)	(0.036)	(0.025)	(0.069)
Experience	0.013***	0.015***	0.013***	0.012***	0.022***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.008)
Experience squared	−0.0004***	−0.0005***	−0.0003***	−0.0003***	−0.001**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Non-Russian	0.004	−0.019	0.025		
	(0.025)	(0.035)	(0.035)		
Female	−0.393***			−0.398***	−0.354***
	(0.015)			(0.016)	(0.048)
Constant	9.993***	10.007***	9.582***	10.000***	9.930***
	(0.027)	(0.036)	(0.039)	(0.028)	(0.079)
Observations	6,120	2,806	3,314	5,468	652
R^2	0.144	0.073	0.077	0.147	0.126
Adjusted R^2	0.143	0.072	0.075	0.146	0.119
Residual Std. Error	0.593	0.581	0.602	0.591	0.610
F Statistic	171.170***	44.408***	55.095***	187.836***	18.631***

Note: *p<0.1; **p<0.05; ***p<0.01