Exercício - Aula 4

Lucas Dutra

2024-08-29

Replication Exercise

- Microeconometrics 2024/II
- Prof. Carlos Charris
- Prof. Soraya Roman
- Topic: Panel Data
- Reference: Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." American Economic Journal: Applied Economics, 3 (2): 1–33.
- DOI: 10.1257/app.3.2.1
- URL: https://www.aeaweb.org/articles?id=10.1257/app.3.2.1
- INPUT FILES: Before running the codes, download the data folder and set your computer directory to this folder. The do-file uses the following datasets: final_HS_data.dta

DESCRIPTION: This do-file replicates the main results of Lavy et al (2001)'s paper, which includes the Tables 1-5 of the paper. For the most part of the replication exercise we will focus on the high-school sample. Because several cohorts are available, this sample allows us to test more complete models, including school trends when needed.

II.C. Evidence on the Validity of the Identification Strategy

Some ilustrations about the randomness of the within variation of female proportion. First, we create a table containing the year, school id and female proportion:

```
df_collapsed <- dados %>%
  group_by(semelmos, year) %>%
  summarize(mfemale = sum(mfemale, na.rm = TRUE))
```

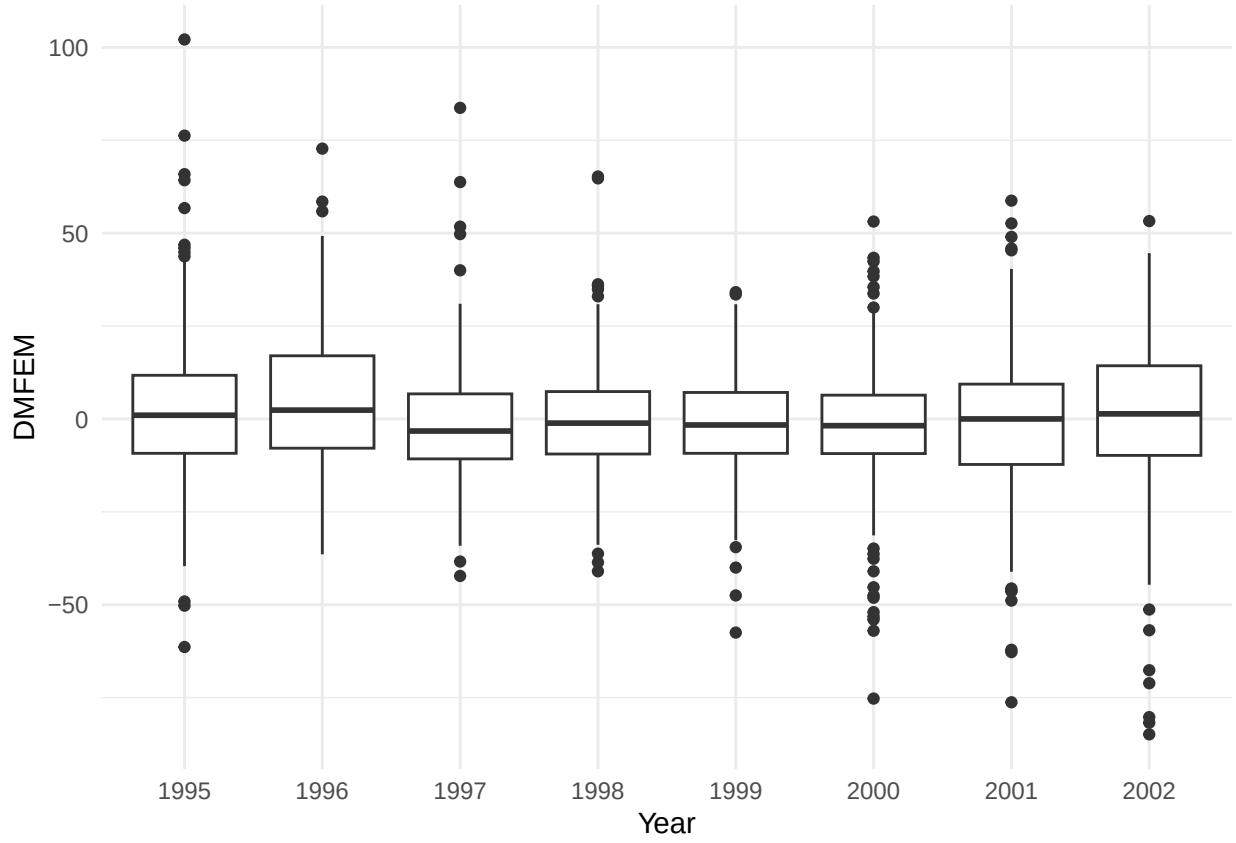
```
## `summarise()` has grouped output by 'semelmos'. You can override using the
## `.groups` argument.
```

Then, we estimate the within variation. Variable dmfem contains the difference between any year female proportion and the average female proportion by school:

```
df_collapsed <- df_collapsed %>%
  group_by(semelmos) %>%
  mutate(mfem_mean = mean(mfemale, na.rm = TRUE),  # Média de mfemale por semelmos
         dmfem = mfemale - mfem_mean)              # Subtrair a média da variável mfemale
```

This graph shows that the distribution of the female proportion deviations is similar across years:

```r
ggplot(df_collapsed, aes(x = factor(year), y = dmfem)) +
  geom_boxplot() +
  labs(x = "Year", y = "DMFEM") +
  theme_minimal()
```



Note that there are no visible difference (except for outliers) between two consecutive years. We can test if that is the case. We use non-parametric tests because of the asymmetries and outliers of the distributions.

Average differences of female proportion deviations across years:

```r
df_summary <- df_collapsed %>%
  group_by(year) %>%
  summarize(mean_dmfem = mean(dmfem, na.rm = TRUE))
```

Non-parametric tests:

```r
# Função para executar os testes para pares de anos consecutivos
test_comparisons <- function(df, start_year, end_year) {
  for (i in start_year:end_year) {
    # Filtra os dados para os dois anos consecutivos
    data_subset <- df_collapsed %>% filter(year %in% c(i, i + 1))

    # Teste de Wilcoxon (rank-sum)
    wilcox_test <- wilcox.test(dmfem ~ year, data = data_subset)
    cat("Wilcoxon test for years", i, "and", i + 1, "\n")
    print(wilcox_test)

    # Teste de Kolmogorov-Smirnov
    ks_test <- ks.test(data_subset$dmfem[data_subset$year == i],
```

```
                      data_subset$dmfem[data_subset$year == i + 1])
    cat("Kolmogorov-Smirnov test for years", i, "and", i + 1, "\n")
    print(ks_test)
  }
}

# Executa a função para os anos de 1995 a 2001
test_comparisons(df, 1995, 2001)
```

```
## Wilcoxon test for years 1995 and 1996
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 22454, p-value = 0.2499
## alternative hypothesis: true location shift is not equal to 0
##
## Kolmogorov-Smirnov test for years 1995 and 1996
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.10167, p-value = 0.2077
## alternative hypothesis: two-sided
##
## Wilcoxon test for years 1996 and 1997
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 29449, p-value = 0.001047
## alternative hypothesis: true location shift is not equal to 0
##
## Kolmogorov-Smirnov test for years 1996 and 1997
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.18328, p-value = 0.001099
## alternative hypothesis: two-sided
##
## Wilcoxon test for years 1997 and 1998
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 24747, p-value = 0.3369
## alternative hypothesis: true location shift is not equal to 0
##
## Kolmogorov-Smirnov test for years 1997 and 1998
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
```

```
## D = 0.071237, p-value = 0.608
## alternative hypothesis: two-sided
##
## Wilcoxon test for years 1998 and 1999
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 27354, p-value = 0.9896
## alternative hypothesis: true location shift is not equal to 0
##
## Kolmogorov-Smirnov test for years 1998 and 1999
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.044276, p-value = 0.9759
## alternative hypothesis: two-sided
##
## Wilcoxon test for years 1999 and 2000
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 31262, p-value = 0.3702
## alternative hypothesis: true location shift is not equal to 0
##
## Kolmogorov-Smirnov test for years 1999 and 2000
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.11207, p-value = 0.09305
## alternative hypothesis: two-sided
##
## Wilcoxon test for years 2000 and 2001
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 31497, p-value = 0.5939
## alternative hypothesis: true location shift is not equal to 0

## Warning in ks.test.default(data_subset$dmfem[data_subset$year == i],
## data_subset$dmfem[data_subset$year == : p-value will be approximate in the
## presence of ties

## Kolmogorov-Smirnov test for years 2000 and 2001
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.10355, p-value = 0.1306
## alternative hypothesis: two-sided
##
```

```
## Wilcoxon test for years 2001 and 2002
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  dmfem by year
## W = 30079, p-value = 0.05865
## alternative hypothesis: true location shift is not equal to 0
```

```
## Warning in ks.test.default(data_subset$dmfem[data_subset$year == i],
## data_subset$dmfem[data_subset$year == : p-value will be approximate in the
## presence of ties
```

```
## Kolmogorov-Smirnov test for years 2001 and 2002
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  data_subset$dmfem[data_subset$year == i] and data_subset$dmfem[data_subset$year == i + 1]
## D = 0.095969, p-value = 0.1857
## alternative hypothesis: two-sided
```

Results: show that for some of the adjacent cohorts in fact we find no systematic difference in the distribution of female proportion deviations, indicating that differences between these years happened almost as random events. However, in some years we find significant differences. In the estimation procedure, the authors include linear trends by school to try to control for long-term changes in school's performance that could be correlated to changes in female proportions. As the variation we are identifying relies on unexpected changes in female proportions through time, no much more can be done. If we adjust the time trends more flexibly (let's say allowing for a different slope in each year and school), we will end up absorbing the variation in the female proportion we are using to identify the effect of interest.

II.C.2: Columns 5 to 7 of Table 1 (Balancing Tests for the Proportion of Female Students)

Re-scaling year variable to use as trend later:

```r
# Reescalar a variável 'year'
df_new <- dados %>%
  mutate(year = year - 1994)

# Variáveis para os testes de balanceamento
balvars <- c("educ_av", "educ_em", "m_ahim", "ole", "israel",
             "asiafr", "etyopia", "euram", "heveram", "enrollment")

# Regressões para cada variável
results <- lapply(balvars, function(v) {
  list(
    # OLS com efeitos fixos para ano
    ols = feols(as.formula(paste(v, "~ mfemale | year")),
                data = df_new,
                cluster = ~ semelmos),

    # Efeitos fixos para escola
    fixed_effects_school = feols(as.formula(paste(v, "~ mfemale | year + semelmos")),
                                 data = df_new, cluster = ~ semelmos),

    # Efeitos fixos para escola + tendências lineares
    fixed_effects_trends = feols(as.formula(paste(v, "~ mfemale | year + semelmos + semelmos:year")),
```

```
                                  data = df_new, cluster = ~ semelmos)
  )
})

# Nomear os resultados para cada variável
names(results) <- balvars

# Exibir resultados
results
```

```
## $educ_av
## $educ_av$ols
## OLS estimation, Dep. Var.: educ_av
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.606362   0.647609 0.936309  0.34997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 3.19571     Adj. R2: 0.003695
##                 Within R2: 5.034e-4
##
## $educ_av$fixed_effects_school
## OLS estimation, Dep. Var.: educ_av
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.517212   0.445086 1.16205  0.24626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.86988     Adj. R2: 0.195977
##                 Within R2: 5.742e-5
##
## $educ_av$fixed_effects_trends
## OLS estimation, Dep. Var.: educ_av
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.241719   0.371717 0.650278  0.51608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.83821     Adj. R2: 0.211578
##                 Within R2: 4.524e-6
##
##
## $educ_em
## $educ_em$ols
## OLS estimation, Dep. Var.: educ_em
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
```

```
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.539452    0.59712 0.903424  0.36712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.99746     Adj. R2: 0.006683
##                 Within R2: 4.529e-4
##
## $educ_em$fixed_effects_school
## OLS estimation, Dep. Var.: educ_em
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.372473   0.411613 0.90491  0.36634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.7012     Adj. R2: 0.192804
##                Within R2: 3.361e-5
##
## $educ_em$fixed_effects_trends
## OLS estimation, Dep. Var.: educ_em
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.382298   0.440272 0.868323  0.38601
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.6666     Adj. R2: 0.211305
##                Within R2: 1.282e-5
##
##
## $m_ahim
## $m_ahim$ols
## OLS estimation, Dep. Var.: m_ahim
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.356289   0.220041 1.61919   0.1066
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.24637     Adj. R2: 0.006433
##                 Within R2: 0.001142
##
## $m_ahim$fixed_effects_school
## OLS estimation, Dep. Var.: m_ahim
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.275105   0.293538 0.937207  0.34951
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## RMSE: 1.15312     Adj. R2: 0.148982
##                 Within R2: 1.006e-4
##
## $m_ahim$fixed_effects_trends
## OLS estimation, Dep. Var.: m_ahim
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
## mfemale 0.781154   0.298563 2.61638 0.0093988 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.11683     Adj. R2: 0.199629
##                 Within R2: 3.051e-4
##
##
## $ole
## $ole$ols
## OLS estimation, Dep. Var.: ole
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value  Pr(>|t|)
## mfemale -0.114509   0.035849 -3.19416 0.0015726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.246086     Adj. R2: 0.010383
##                  Within R2: 0.00302
##
## $ole$fixed_effects_school
## OLS estimation, Dep. Var.: ole
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value Pr(>|t|)
## mfemale -0.022898    0.03433 -0.666995  0.50536
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.239049     Adj. R2: 0.065564
##                  Within R2: 1.622e-5
##
## $ole$fixed_effects_trends
## OLS estimation, Dep. Var.: ole
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value Pr(>|t|)
## mfemale -0.013556   0.030811 -0.439979  0.66031
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.237244     Adj. R2: 0.077224
##                  Within R2: 2.037e-6
##
##
```

```
## $israel
## $israel$ols
## OLS estimation, Dep. Var.: israel
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.092122   0.046808 1.96806 0.050107 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.491376     Adj. R2: 0.006912
##                  Within R2: 4.915e-4
##
## $israel$fixed_effects_school
## OLS estimation, Dep. Var.: israel
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value Pr(>|t|)
## mfemale -0.041851   0.040756 -1.02687  0.30542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.477876     Adj. R2: 0.060118
##                  Within R2: 1.356e-5
##
## $israel$fixed_effects_trends
## OLS estimation, Dep. Var.: israel
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value Pr(>|t|)
## mfemale -0.040716   0.036482 -1.11604  0.26542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.476569     Adj. R2: 0.062817
##                  Within R2: 4.553e-6
##
##
## $asiafr
## $asiafr$ols
## OLS estimation, Dep. Var.: asiafr
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.027313   0.036851 0.741174  0.45925
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.437487     Adj. R2: 0.003782
##                  Within R2: 5.452e-5
##
## $asiafr$fixed_effects_school
## OLS estimation, Dep. Var.: asiafr
## Observations: 404,929
```

```
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.027521   0.025161 1.09376  0.27506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.425526     Adj. R2: 0.056896
##                  Within R2: 7.395e-6
##
## $asiafr$fixed_effects_trends
## OLS estimation, Dep. Var.: asiafr
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.022313   0.025422 0.877717   0.3809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.424659     Adj. R2: 0.058291
##                  Within R2: 1.722e-6
##
##
## $etyopia
## $etyopia$ols
## OLS estimation, Dep. Var.: etyopia
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value Pr(>|t|)
## mfemale -0.002483   0.002743 -0.905023  0.36628
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.057899     Adj. R2: 0.001437
##                  Within R2: 2.572e-5
##
## $etyopia$fixed_effects_school
## OLS estimation, Dep. Var.: etyopia
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value Pr(>|t|)
## mfemale -0.004676   0.005752 -0.812897  0.41701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.057496     Adj. R2: 0.014641
##                  Within R2: 1.169e-5
##
## $etyopia$fixed_effects_trends
## OLS estimation, Dep. Var.: etyopia
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value Pr(>|t|)
## mfemale -0.010303   0.005765 -1.78719 0.075053 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.057232     Adj. R2: 0.021109
##                  Within R2: 2.021e-5
##
##
## $euram
## $euram$ols
## OLS estimation, Dep. Var.: euram
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
## mfemale 0.082472   0.026547 3.10671 0.0020983 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.355289     Adj. R2: 0.001549
##                  Within R2: 7.532e-4
##
## $euram$fixed_effects_school
## OLS estimation, Dep. Var.: euram
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value Pr(>|t|)
## mfemale -0.013788   0.015291 -0.901655  0.36806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.348492     Adj. R2: 0.038759
##                  Within R2: 2.767e-6
##
## $euram$fixed_effects_trends
## OLS estimation, Dep. Var.: euram
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.001085   0.019964 0.054351   0.9567
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.347929     Adj. R2: 0.03937
##                  Within R2: 6.067e-9
##
##
## $heveram
## $heveram$ols
## OLS estimation, Dep. Var.: heveram
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value  Pr(>|t|)
## mfemale -0.199424   0.068653 -2.90483 0.0039858 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## RMSE: 0.375587     Adj. R2: 0.007555
##                  Within R2: 0.003929
##
## $heveram$fixed_effects_school
## OLS estimation, Dep. Var.: heveram
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.032794   0.054545 0.601228   0.5482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.357385     Adj. R2: 0.100833
##                  Within R2: 1.489e-5
##
## $heveram$fixed_effects_trends
## OLS estimation, Dep. Var.: heveram
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale 0.027621   0.037776 0.731166  0.46533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.355313     Adj. R2: 0.108915
##                  Within R2: 3.769e-6
##
##
## $enrollment
## $enrollment$ols
## OLS estimation, Dep. Var.: enrollment
## Observations: 404,929
## Fixed-effects: year: 8
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale   24.449    47.5095 0.514614  0.60725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 130.5     Adj. R2: 0.00447
##               Within R2: 4.911e-4
##
## $enrollment$fixed_effects_school
## OLS estimation, Dep. Var.: enrollment
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error   t value Pr(>|t|)
## mfemale -6.41751    28.0517 -0.228775  0.81922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 34.4     Adj. R2: 0.930838
##              Within R2: 6.162e-5
##
## $enrollment$fixed_effects_trends
```

```
## OLS estimation, Dep. Var.: enrollment
## Observations: 404,929
## Fixed-effects: year: 8,  semelmos: 265,  year:semelmos: 1,052
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error  t value Pr(>|t|)
## mfemale -74.6706    24.4591 -3.05288 0.002498 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 18.9     Adj. R2: 0.978942
##              Within R2: 0.0096
```

Notice that as expected, once we control for year and school fixed effects, the explanatory power of the female proportion on individual's socioeconomic characteristics is null in almost all cases. Once more, this shows the exogeneity (randomness) of the changes in female proportion.

II.C.3: Columns 5 to 6 of Table 2 (The Effect of the Proportion Female on Student's School Mobility)

```
# Variáveis de controle para os modelos de regressão
cvars <- c("educ_av", "av0", "av99", "av88", "avmiss", "educ_em", "em0",
           "em99", "em88", "emmiss", "m_ahim", "ah0", "ah88", "ah99",
           "ahmiss", "ole", "asiafr", "etyopia", "heveram", "euram",
           "m_educ_av", "m_educ_em", "m_m_ahim", "m_ole", "m_asiafr",
           "m_euram", "m_heveram", "m_etyopia",
           "enrollment", "enrollment2")

# Média para homens
mean_left_school_men <- dados %>%
  filter(female == 0) %>%
  summarize(mean_left_school = mean(left_school, na.rm = TRUE))

# Média para mulheres
mean_left_school_women <- dados %>%
  filter(female == 1) %>%
  summarize(mean_left_school = mean(left_school, na.rm = TRUE))

# Modelos de regressão para prever o efeito da proporção feminina
# sobre a evasão escolar no ano seguinte
# Para mulheres
regression_women <- feols(as.formula(paste("left_school ~ mfemale +",
                                           paste(cvars, collapse = " + "),
                                           "| year + semelmos + semelmos:year")),
                          data = dados %>% filter(female == 1),
                          cluster = ~ semelmos)
```

```
## The variables 'mfemale', 'ahmiss', 'm_educ_av', 'm_educ_em', 'm_m_ahim', 'm_ole' and 6 others have be
```

```
# Para homens
regression_men <- feols(as.formula(paste("left_school ~ mfemale +",
                                         paste(cvars, collapse = " + "),
                                         "| year + semelmos + semelmos:year")),
                        data = dados %>% filter(female == 0),
                        cluster = ~ semelmos)
```

```
## The variables 'mfemale', 'ahmiss', 'm_educ_av', 'm_educ_em', 'm_m_ahim', 'm_ole' and 6 others have be
```

```r
# Resultados
list(
  mean_left_school_men = mean_left_school_men,
  mean_left_school_women = mean_left_school_women,
  regression_women = regression_women,
  regression_men = regression_men
)
```

```
## $mean_left_school_men
## # A tibble: 1 x 1
##   mean_left_school
##              <dbl>
## 1           0.0975
##
## $mean_left_school_women
## # A tibble: 1 x 1
##   mean_left_school
##              <dbl>
## 1           0.0660
##
## $regression_women
## OLS estimation, Dep. Var.: left_school
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,896
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av -0.001277   0.000304 -4.197002 3.7031e-05 ***
## av0      0.474245   0.100334  4.726640 3.7232e-06 ***
## av99     0.005353   0.004938  1.084144 2.7929e-01
## av88     0.041357   0.025895  1.597111 1.1144e-01
## avmiss   0.005282   0.020005  0.264023 7.9197e-01
## educ_em -0.001539   0.000380 -4.052011 6.6900e-05 ***
## em0     -0.162455   0.094027 -1.727752 8.5207e-02 .
## em99     0.026092   0.010123  2.577660 1.0492e-02 *
## ... 11 coefficients remaining (display them with summary() or use argument n)
## ... 12 variables were removed because of collinearity (mfemale, ahmiss and 10 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.239771     Adj. R2: 0.056752
##                  Within R2: 0.029382
##
## $regression_men
## OLS estimation, Dep. Var.: left_school
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,892
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av -0.002403   0.000394 -6.103128 3.7099e-09 ***
## av0      0.432135   0.086485  4.996634 1.0653e-06 ***
## av99     0.015735   0.008283  1.899742 5.8561e-02 .
## av88    -0.002784   0.023257 -0.119700 9.0481e-01
## avmiss  -0.044426   0.011976 -3.709509 2.5335e-04 ***
## educ_em -0.001741   0.000459 -3.795095 1.8323e-04 ***
## em0     -0.072696   0.087568 -0.830171 4.0720e-01
```

```
## em99     0.036283   0.013808  2.627658 9.1026e-03 **
## ... 11 coefficients remaining (display them with summary() or use argument n)
## ... 12 variables were removed because of collinearity (mfemale, ahmiss and 10 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.285148     Adj. R2: 0.065507
##                  Within R2: 0.028068
```

Results show evidence that sorting across schools may not be a source of endogeneity (bias) because the mobility is low and the female proportion has no explanatory power on this variable once we control for fixed effects.

III.A. Effects on High School Students' Achievements

Columns 1 to 4 of Table 3 (Estimates of the Effect of Proportion Female on Scholastic Outcomes in High School):

```
# Lista de variáveis dependentes
depvars <- c("wmean", "zakaibag", "units", "madmug", "university")

# Especificação 1 e 2 para mulheres
results_women <- lapply(depvars, function(depvar) {
  list(
    # Especificação 1: Efeitos fixos de escola + tendências lineares + covariáveis
    spec1 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 1),
            cluster = ~ semelmos),
      error = function(e) e
    ),

    # Especificação 2: Efeitos fixos de escola + tendências lineares
    spec2 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale | year + semelmos")),
            data = dados %>% filter(female == 1),
            cluster = ~ semelmos),
      error = function(e) e
    )
  )
})
```

```
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
```

```
# Especificação 1 e 2 para homens
results_men <- lapply(depvars, function(depvar) {
  list(
    # Especificação 1: Efeitos fixos de escola + tendências lineares + covariáveis
    spec1 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 0),
```

```
            cluster = ~ semelmos),
      error = function(e) e
    ),

    # Especificação 2: Efeitos fixos de escola + tendências lineares
    spec2 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale | year + semelmos")),
            data = dados %>% filter(female == 0),
            cluster = ~ semelmos),
      error = function(e) e
    )
  )
})
```

```
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
## The variable 'ahmiss' has been removed because of collinearity (see $collin.var).
```

```
# Exibir resultados
list(
  results_women = results_women,
  results_men = results_men
)
```

```
## $results_women
## $results_women[[1]]
## $results_women[[1]]$spec1
## OLS estimation, Dep. Var.: wmean
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error  t value   Pr(>|t|)
## mfemale   9.322583   2.745674  3.39537 7.9124e-04 ***
## educ_av   0.633568   0.034314 18.46406  < 2.2e-16 ***
## av0     -31.325197   6.090544 -5.14325 5.2815e-07 ***
## av99     -2.718889   0.518447 -5.24429 3.2279e-07 ***
## av88     -5.710665   2.066720 -2.76315 6.1288e-03 **
## avmiss    1.725482   1.673592  1.03101 3.0349e-01
## educ_em   0.756620   0.044725 16.91711  < 2.2e-16 ***
## em0      13.528779   6.636533  2.03853 4.2497e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 21.8     Adj. R2: 0.181756
##              Within R2: 0.060815
##
## $results_women[[1]]$spec2
## OLS estimation, Dep. Var.: wmean
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
```

```
## mfemale  9.95435    3.04003 3.27443 0.0012007 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 22.5     Adj. R2: 0.129181
##              Within R2: 3.277e-4
##
##
## $results_women[[2]]
## $results_women[[2]]$spec1
## OLS estimation, Dep. Var.: zakaibag
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value   Pr(>|t|)
## mfemale  0.203096   0.058029  3.49991 5.4647e-04 ***
## educ_av  0.014185   0.000713 19.89815  < 2.2e-16 ***
## av0     -0.233706   0.050392 -4.63779 5.5549e-06 ***
## av99    -0.047345   0.008659 -5.46785 1.0588e-07 ***
## av88    -0.086235   0.028577 -3.01761 2.7977e-03 **
## avmiss   0.063771   0.029344  2.17319 3.0659e-02 *
## educ_em  0.018008   0.001015 17.75030  < 2.2e-16 ***
## em0      0.156417   0.061120  2.55916 1.1053e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.438263     Adj. R2: 0.184163
##                  Within R2: 0.042471
##
## $results_women[[2]]$spec2
## OLS estimation, Dep. Var.: zakaibag
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
## mfemale 0.211748   0.062688 3.37783 0.0008412 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.447794     Adj. R2: 0.148415
##                  Within R2: 3.727e-4
##
##
## $results_women[[3]]
## $results_women[[3]]$spec1
## OLS estimation, Dep. Var.: units
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value   Pr(>|t|)
## mfemale  2.836818   1.313280  2.16010 3.1668e-02 *
## educ_av  0.292872   0.015392 19.02790  < 2.2e-16 ***
## av0     -8.742312   1.788164 -4.88899 1.7656e-06 ***
## av99    -1.293042   0.205593 -6.28934 1.3223e-09 ***
## av88    -2.619527   0.650670 -4.02589 7.4291e-05 ***
```

```
## avmiss   1.117469   0.686150  1.62861 1.0459e-01
## educ_em  0.378914   0.021630 17.51762  < 2.2e-16 ***
## em0      4.607329   2.000940  2.30258 2.2083e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.47384     Adj. R2: 0.19581
##                 Within R2: 0.060097
##
## $results_women[[3]]$spec2
## OLS estimation, Dep. Var.: units
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale  3.11131     1.3794 2.25555 0.024921 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.73962     Adj. R2: 0.144692
##                 Within R2: 2.113e-4
##
##
## $results_women[[4]]
## $results_women[[4]]$spec1
## OLS estimation, Dep. Var.: madmug
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value   Pr(>|t|)
## mfemale  0.160171   0.077677  2.06202 4.0188e-02 *
## educ_av  0.038861   0.001866 20.82912  < 2.2e-16 ***
## av0     -0.171823   0.072567 -2.36779 1.8619e-02 *
## av99    -0.074638   0.014390 -5.18677 4.2760e-07 ***
## av88    -0.079571   0.042715 -1.86284 6.3600e-02 .
## avmiss   0.130296   0.080697  1.61462 1.0759e-01
## educ_em  0.037468   0.001864 20.09797  < 2.2e-16 ***
## em0      0.144666   0.078728  1.83755 6.7257e-02 .
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.814391     Adj. R2: 0.167708
##                  Within R2: 0.059863
##
## $results_women[[4]]$spec2
## OLS estimation, Dep. Var.: madmug
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.177576   0.079769 2.22613 0.026853 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## RMSE: 0.839887     Adj. R2: 0.114903
##                  Within R2: 7.453e-5
##
##
## $results_women[[5]]
## $results_women[[5]]$spec1
## OLS estimation, Dep. Var.: university
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value   Pr(>|t|)
## mfemale  0.195181   0.056918  3.42916 7.0270e-04 ***
## educ_av  0.016680   0.000791 21.07854  < 2.2e-16 ***
## av0     -0.170718   0.050077 -3.40914 7.5399e-04 ***
## av99    -0.049377   0.007936 -6.22155 1.9298e-09 ***
## av88    -0.081536   0.023518 -3.46692 6.1477e-04 ***
## avmiss   0.080152   0.031510  2.54371 1.1541e-02 *
## educ_em  0.020262   0.001074 18.87190  < 2.2e-16 ***
## em0      0.146624   0.057845  2.53475 1.1833e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.435252     Adj. R2: 0.23045
##                  Within R2: 0.05757
##
## $results_women[[5]]$spec2
## OLS estimation, Dep. Var.: university
## Observations: 205,891
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value   Pr(>|t|)
## mfemale 0.203905   0.061078 3.33846 0.00096425 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.448272     Adj. R2: 0.183837
##                  Within R2: 3.449e-4
##
##
##
## $results_men
## $results_men[[1]]
## $results_men[[1]]$spec1
## OLS estimation, Dep. Var.: wmean
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error   t value   Pr(>|t|)
## mfemale   9.945394   3.518905  2.826275 5.0708e-03 **
## educ_av   0.819462   0.044569 18.386211  < 2.2e-16 ***
## av0     -25.002759   5.273618 -4.741101 3.4866e-06 ***
## av99     -3.886962   0.715447 -5.432917 1.2631e-07 ***
## av88     -2.529313   2.253397 -1.122445 2.6270e-01
## avmiss    5.120185   2.094466  2.444625 1.5157e-02 *
```

```
## educ_em   0.744917   0.052156 14.282383  < 2.2e-16 ***
## em0       3.131293   5.045105  0.620660 5.3536e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 25.0     Adj. R2: 0.180288
##              Within R2: 0.052141
##
## $results_men[[1]]$spec2
## OLS estimation, Dep. Var.: wmean
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
## mfemale  11.1191    3.73427 2.97758 0.0031764 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 25.7     Adj. R2: 0.135612
##              Within R2: 3.355e-4
##
##
## $results_men[[2]]
## $results_men[[2]]$spec1
## OLS estimation, Dep. Var.: zakaibag
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## mfemale  0.133462   0.072981  1.828740 6.8571e-02 .
## educ_av  0.016622   0.000828 20.066705  < 2.2e-16 ***
## av0     -0.157000   0.054998 -2.854653 4.6516e-03 **
## av99    -0.049957   0.009949 -5.021283 9.4777e-07 ***
## av88    -0.027577   0.034719 -0.794282 4.2775e-01
## avmiss   0.107796   0.076499  1.409121 1.5998e-01
## educ_em  0.015941   0.001009 15.806198  < 2.2e-16 ***
## em0      0.020563   0.049757  0.413271 6.7974e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.448243     Adj. R2: 0.193404
##                  Within R2: 0.040257
##
## $results_men[[2]]$spec2
## OLS estimation, Dep. Var.: zakaibag
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.158657   0.088547 1.79178 0.074318 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.457498     Adj. R2: 0.159874
```

```
##                  Within R2: 2.152e-4
##
##
## $results_men[[3]]
## $results_men[[3]]$spec1
## OLS estimation, Dep. Var.: units
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## mfemale  3.549615   1.848444  1.920326 5.5898e-02 .
## educ_av  0.409986   0.019871 20.632302  < 2.2e-16 ***
## av0     -7.118728   1.636620 -4.349651 1.9521e-05 ***
## av99    -1.511811   0.272795 -5.541929 7.2625e-08 ***
## av88    -0.999887   0.824726 -1.212386 2.2645e-01
## avmiss   3.150052   1.174965  2.680975 7.8048e-03 **
## educ_em  0.386983   0.025105 15.414832  < 2.2e-16 ***
## em0      0.917554   1.636552  0.560663 5.7550e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 9.93693     Adj. R2: 0.205133
##                 Within R2: 0.058044
##
## $results_men[[3]]$spec2
## OLS estimation, Dep. Var.: units
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale  4.61323    2.28506 2.01886 0.044516 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.2     Adj. R2: 0.156582
##              Within R2: 3.634e-4
##
##
## $results_men[[4]]
## $results_men[[4]]$spec1
## OLS estimation, Dep. Var.: madmug
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## mfemale  0.331023   0.113682  2.911842 3.9013e-03 **
## educ_av  0.040711   0.001938 21.003717  < 2.2e-16 ***
## av0     -0.185827   0.094439 -1.967690 5.0154e-02 .
## av99    -0.088566   0.014538 -6.091912 3.9449e-09 ***
## av88    -0.090604   0.038892 -2.329591 2.0585e-02 *
## avmiss   0.368217   0.114825  3.206780 1.5083e-03 **
## educ_em  0.034524   0.002037 16.949942  < 2.2e-16 ***
## em0      0.033410   0.082833  0.403344 6.8702e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
```

```
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.814478     Adj. R2: 0.201688
##                  Within R2: 0.05955
##
## $results_men[[4]]$spec2
## OLS estimation, Dep. Var.: madmug
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value  Pr(>|t|)
## mfemale 0.388539   0.141527 2.74534 0.0064617 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.839708     Adj. R2: 0.151587
##                  Within R2: 3.831e-4
##
##
## $results_men[[5]]
## $results_men[[5]]$spec1
## OLS estimation, Dep. Var.: university
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## mfemale  0.139738   0.077190  1.810307 7.1389e-02 .
## educ_av  0.018392   0.000867 21.216157  < 2.2e-16 ***
## av0     -0.116618   0.050621 -2.303750 2.2017e-02 *
## av99    -0.045913   0.009648 -4.758967 3.2142e-06 ***
## av88    -0.037270   0.027437 -1.358409 1.7550e-01
## avmiss   0.130413   0.082498  1.580809 1.1512e-01
## educ_em  0.017487   0.001082 16.164754  < 2.2e-16 ***
## em0      0.022151   0.045938  0.482191 6.3007e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ... 1 variable was removed because of collinearity (ahmiss)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.43791     Adj. R2: 0.229502
##                 Within R2: 0.053472
##
## $results_men[[5]]$spec2
## OLS estimation, Dep. Var.: university
## Observations: 199,038
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##         Estimate Std. Error t value Pr(>|t|)
## mfemale 0.173691   0.092302 1.88178  0.06097 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.450049     Adj. R2: 0.18631
##                  Within R2: 2.665e-4
```

Notice that there are no systematic differences between specification 1 and 2, which indicates results

are robust to the introduction of individual and cohort characteristics. Sign and significance of effects is as expected. Having more peers is beneficial for girls and prejudicial for boys. Authors indicate that the size of the effects are modest when compared to alternative educational policies.

III.B. Falsification tests

Columns 5 to 8 of Table 3 (Estimates of the Effect of Proportion Female on Scholastic Outcomes in High School)

IDEA: We intend to capture the effect of an exogenous change in female proportion on scholastic outcomes. The authors want to prove that this effect is not coming from a spurious correlation (the observed effect comes from another unobservable variable correlated with the changes in female proportion). If this was the case, we would observe that the female proportion of the adjacent cohorts (which are different groups of individuals) can also significantly affect the scholastic outcomes of the current cohort. The authors show that these variables are statistically non-significant for the majority of the regression models.

```
# Lista de variáveis dependentes
depvars <- c("wmean", "zakaibag", "units", "madmug", "university")

# Especificação de covariáveis
cvars <- c("educ_av", "av0", "av99", "av88", "avmiss",
           "educ_em", "em0", "em99", "em88", "emmiss",
           "m_ahim", "ah0", "ah88", "ah99", "ahmiss",
           "ole", "asiafr", "etyopia", "heveram", "euram",
           "m_educ_av", "m_educ_em", "m_m_ahim", "m_ole",
           "m_asiafr", "m_euram", "m_heveram", "m_etyopia",
           "enrollment", "enrollment2")

# Falsificação para mulheres
results_women <- lapply(depvars, function(depvar) {
  list(
    # Falsificação 1: Proporção feminina no t-1
    falsification1 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale_past +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 1),
            cluster = ~ semelmos),
      error = function(e) e
    ),

    # Falsificação 2: Proporção feminina no t+1
    falsification2 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale_fut +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 1),
            cluster = ~ semelmos),
      error = function(e) e
    )
  )
})
```

```
## NOTE: 3,644 observations removed because of NA values (RHS: 3,644).

## NOTE: 525 observations removed because of NA values (RHS: 525).

## NOTE: 3,644 observations removed because of NA values (RHS: 3,644).
```

```
## NOTE: 525 observations removed because of NA values (RHS: 525).
## NOTE: 3,644 observations removed because of NA values (RHS: 3,644).
## NOTE: 525 observations removed because of NA values (RHS: 525).
## NOTE: 3,644 observations removed because of NA values (RHS: 3,644).
## NOTE: 525 observations removed because of NA values (RHS: 525).
## NOTE: 3,644 observations removed because of NA values (RHS: 3,644).
## NOTE: 525 observations removed because of NA values (RHS: 525).
```

```r
# Falsificação para homens
results_men <- lapply(depvars, function(depvar) {
  list(
    # Falsificação 1: Proporção feminina no t-1
    falsification1 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale_past +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 0),
            cluster = ~ semelmos),
      error = function(e) e
    ),

    # Falsificação 2: Proporção feminina no t+1
    falsification2 = tryCatch(
      feols(as.formula(paste(depvar, "~ mfemale_fut +",
                             paste(cvars, collapse = " + "), "| year + semelmos")),
            data = dados %>% filter(female == 0),
            cluster = ~ semelmos),
      error = function(e) e
    )
  )
})
```

```
## NOTE: 3,537 observations removed because of NA values (RHS: 3,537).
## NOTE: 799 observations removed because of NA values (RHS: 799).
## NOTE: 3,537 observations removed because of NA values (RHS: 3,537).
## NOTE: 799 observations removed because of NA values (RHS: 799).
## NOTE: 3,537 observations removed because of NA values (RHS: 3,537).
## NOTE: 799 observations removed because of NA values (RHS: 799).
## NOTE: 3,537 observations removed because of NA values (RHS: 3,537).
## NOTE: 799 observations removed because of NA values (RHS: 799).
## NOTE: 3,537 observations removed because of NA values (RHS: 3,537).
## NOTE: 799 observations removed because of NA values (RHS: 799).
```

```r
# Exibir resultados
list(
  results_women = results_women,
  results_men = results_men
)
```

```
## $results_women
## $results_women[[1]]
## $results_women[[1]]$falsification1
## OLS estimation, Dep. Var.: wmean
## Observations: 202,247
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past   7.012158   2.436511  2.877951 4.3326e-03 **
## educ_av        0.638811   0.034531 18.499359  < 2.2e-16 ***
## av0          -31.341560   6.135180 -5.108499 6.2604e-07 ***
## av99          -2.652116   0.519556 -5.104582 6.3795e-07 ***
## av88          -5.514106   2.072273 -2.660898 8.2744e-03 **
## avmiss         1.266344   1.681599  0.753059 4.5209e-01
## educ_em        0.753451   0.044757 16.834435  < 2.2e-16 ***
## em0           13.498476   6.681510  2.020273 4.4373e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 21.8     Adj. R2: 0.180463
##              Within R2: 0.060842
##
## $results_women[[1]]$falsification2
## OLS estimation, Dep. Var.: wmean
## Observations: 205,366
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut    1.733466   2.778445  0.623898 5.3324e-01
## educ_av        0.634282   0.034361 18.459176  < 2.2e-16 ***
## av0          -31.476426   6.128639 -5.135957 5.4711e-07 ***
## av99          -2.719263   0.520063 -5.228717 3.4841e-07 ***
## av88          -5.611278   2.078981 -2.699052 7.4042e-03 **
## avmiss         1.721065   1.680215  1.024312 3.0663e-01
## educ_em        0.755881   0.044802 16.871734  < 2.2e-16 ***
## em0           13.557172   6.658770  2.035987 4.2754e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 21.8     Adj. R2: 0.180808
##              Within R2: 0.060401
##
##
## $results_women[[2]]
## $results_women[[2]]$falsification1
## OLS estimation, Dep. Var.: zakaibag
## Observations: 202,247
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error  t value   Pr(>|t|)
## mfemale_past   0.155763   0.056514  2.75617 6.2589e-03 **
## educ_av        0.014172   0.000720 19.67557  < 2.2e-16 ***
## av0           -0.233426   0.050537 -4.61896 6.0518e-06 ***
## av99          -0.046714   0.008783 -5.31842 2.2451e-07 ***
```

```
## av88         -0.081621   0.028705 -2.84348 4.8141e-03 **
## avmiss        0.056512   0.030193  1.87167 6.2367e-02 .
## educ_em       0.018015   0.001022 17.62603  < 2.2e-16 ***
## em0           0.155553   0.061874  2.51401 1.2537e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.438303     Adj. R2: 0.183218
##                  Within R2: 0.042416
##
## $results_women[[2]]$falsification2
## OLS estimation, Dep. Var.: zakaibag
## Observations: 205,366
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error  t value   Pr(>|t|)
## mfemale_fut  0.111869   0.057841  1.93409 5.4174e-02 .
## educ_av      0.014183   0.000713 19.89398  < 2.2e-16 ***
## av0         -0.237835   0.049944 -4.76208 3.1689e-06 ***
## av99        -0.047442   0.008714 -5.44463 1.1906e-07 ***
## av88        -0.084287   0.028716 -2.93519 3.6281e-03 **
## avmiss       0.063367   0.029474  2.14995 3.2471e-02 *
## educ_em      0.018030   0.001017 17.73716  < 2.2e-16 ***
## em0          0.158393   0.061371  2.58089 1.0397e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.438385     Adj. R2: 0.183532
##                  Within R2: 0.042281
##
##
## $results_women[[3]]
## $results_women[[3]]$falsification1
## OLS estimation, Dep. Var.: units
## Observations: 202,247
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error  t value   Pr(>|t|)
## mfemale_past  1.965641   1.230173  1.59786 1.1128e-01
## educ_av       0.293849   0.015479 18.98341  < 2.2e-16 ***
## av0          -8.743905   1.800394 -4.85666 2.0556e-06 ***
## av99         -1.293710   0.207483 -6.23527 1.7967e-09 ***
## av88         -2.543435   0.651440 -3.90433 1.2025e-04 ***
## avmiss        0.977103   0.693320  1.40931 1.5993e-01
## educ_em       0.378497   0.021662 17.47256  < 2.2e-16 ***
## em0           4.600378   2.014478  2.28366 2.3192e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.48131     Adj. R2: 0.194419
##                  Within R2: 0.060044
##
## $results_women[[3]]$falsification2
## OLS estimation, Dep. Var.: units
```

```
## Observations: 205,366
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut  0.943167   1.154822  0.816721 4.1483e-01
## educ_av      0.293091   0.015411 19.018069  < 2.2e-16 ***
## av0         -8.795513   1.797845 -4.892254 1.7390e-06 ***
## av99        -1.293474   0.206404 -6.266717 1.5006e-09 ***
## av88        -2.588505   0.654971 -3.952089 9.9608e-05 ***
## avmiss       1.114424   0.688687  1.618186 1.0682e-01
## educ_em      0.378952   0.021656 17.498645  < 2.2e-16 ***
## em0          4.626197   2.007860  2.304043 2.2000e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.47676     Adj. R2: 0.194847
##                 Within R2: 0.059874
##
##
## $results_women[[4]]
## $results_women[[4]]$falsification1
## OLS estimation, Dep. Var.: madmug
## Observations: 202,247
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past  0.075896   0.090562  0.838058 4.0276e-01
## educ_av       0.038895   0.001885 20.636241  < 2.2e-16 ***
## av0          -0.172765   0.071452 -2.417916 1.6292e-02 *
## av99         -0.075252   0.014533 -5.178018 4.4731e-07 ***
## av88         -0.075293   0.042807 -1.758883 7.9765e-02 .
## avmiss        0.123413   0.081581  1.512761 1.3155e-01
## educ_em       0.037481   0.001878 19.959576  < 2.2e-16 ***
## em0           0.144865   0.078567  1.843833 6.6337e-02 .
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.815232     Adj. R2: 0.168072
##                  Within R2: 0.059821
##
## $results_women[[4]]$falsification2
## OLS estimation, Dep. Var.: madmug
## Observations: 205,366
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut  0.050824   0.068223  0.744977 4.5695e-01
## educ_av      0.038893   0.001870 20.800786  < 2.2e-16 ***
## av0         -0.174490   0.071443 -2.442358 1.5251e-02 *
## av99        -0.074613   0.014427 -5.171764 4.5997e-07 ***
## av88        -0.077936   0.042879 -1.817572 7.0267e-02 .
## avmiss       0.130084   0.080777  1.610403 1.0851e-01
## educ_em      0.037442   0.001868 20.047274  < 2.2e-16 ***
## em0          0.145815   0.078581  1.855607 6.4628e-02 .
```

```
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.814723     Adj. R2: 0.167564
##                  Within R2: 0.05981
##
##
## $results_women[[5]]
## $results_women[[5]]$falsification1
## OLS estimation, Dep. Var.: university
## Observations: 202,247
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error  t value   Pr(>|t|)
## mfemale_past  0.136485   0.054336  2.51187 1.2611e-02 *
## educ_av       0.016677   0.000800 20.84803  < 2.2e-16 ***
## av0          -0.170691   0.049760 -3.43026 7.0038e-04 ***
## av99         -0.048729   0.008035 -6.06471 4.5972e-09 ***
## av88         -0.077277   0.023657 -3.26659 1.2337e-03 **
## avmiss        0.074450   0.032270  2.30711 2.1829e-02 *
## educ_em       0.020239   0.001081 18.72188  < 2.2e-16 ***
## em0           0.145813   0.058263  2.50267 1.2936e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.435242     Adj. R2: 0.23007
##                  Within R2: 0.057382
##
## $results_women[[5]]$falsification2
## OLS estimation, Dep. Var.: university
## Observations: 205,366
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error  t value   Pr(>|t|)
## mfemale_fut  0.119011   0.051617  2.30564 2.1909e-02 *
## educ_av      0.016659   0.000791 21.06021  < 2.2e-16 ***
## av0         -0.174784   0.049166 -3.55498 4.4807e-04 ***
## av99        -0.049434   0.007989 -6.18762 2.3293e-09 ***
## av88        -0.079863   0.023688 -3.37145 8.6010e-04 ***
## avmiss       0.079754   0.031649  2.51993 1.2331e-02 *
## educ_em      0.020287   0.001076 18.85563  < 2.2e-16 ***
## em0          0.148587   0.057931  2.56487 1.0877e-02 *
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.435348     Adj. R2: 0.230035
##                  Within R2: 0.057395
##
##
##
## $results_men
## $results_men[[1]]
## $results_men[[1]]$falsification1
## OLS estimation, Dep. Var.: wmean
```

```
## Observations: 195,501
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past   3.849387   3.336930  1.153572 2.4973e-01
## educ_av        0.819477   0.044685 18.338811  < 2.2e-16 ***
## av0          -25.144944   5.270134 -4.771216 3.0450e-06 ***
## av99          -3.775279   0.719741 -5.245328 3.2198e-07 ***
## av88          -2.342293   2.257277 -1.037663 3.0038e-01
## avmiss         5.130043   2.100241  2.442598 1.5243e-02 *
## educ_em        0.747590   0.052740 14.174934  < 2.2e-16 ***
## em0            3.022502   5.048634  0.598677 5.4991e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 25.0     Adj. R2: 0.177055
##              Within R2: 0.0521
##
## $results_men[[1]]$falsification2
## OLS estimation, Dep. Var.: wmean
## Observations: 198,239
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut    3.780991   3.531351  1.070693 2.8529e-01
## educ_av        0.821360   0.044510 18.453382  < 2.2e-16 ***
## av0          -25.067682   5.277479 -4.749935 3.3492e-06 ***
## av99          -3.921326   0.705919 -5.554920 6.7953e-08 ***
## av88          -2.496057   2.250744 -1.108992 2.6845e-01
## avmiss         5.159929   2.099657  2.457511 1.4636e-02 *
## educ_em        0.745294   0.052339 14.239867  < 2.2e-16 ***
## em0            3.100662   5.038662  0.615374 5.3884e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 25.0     Adj. R2: 0.179955
##              Within R2: 0.052052
##
##
## $results_men[[2]]
## $results_men[[2]]$falsification1
## OLS estimation, Dep. Var.: zakaibag
## Observations: 195,501
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##                Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past   0.104365   0.069332  1.505291 1.3345e-01
## educ_av        0.016531   0.000833 19.836033  < 2.2e-16 ***
## av0           -0.158881   0.055235 -2.876455 4.3526e-03 **
## av99          -0.048257   0.010079 -4.788101 2.8187e-06 ***
## av88          -0.026809   0.034923 -0.767649 4.4339e-01
## avmiss         0.107495   0.076443  1.406207 1.6085e-01
## educ_em        0.016007   0.001018 15.717848  < 2.2e-16 ***
## em0            0.021823   0.050003  0.436441 6.6288e-01
```

```
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.448554     Adj. R2: 0.191983
##                  Within R2: 0.040213
##
## $results_men[[2]]$falsification2
## OLS estimation, Dep. Var.: zakaibag
## Observations: 198,239
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut  0.065616   0.068831  0.953288 3.4132e-01
## educ_av      0.016651   0.000830 20.053237  < 2.2e-16 ***
## av0         -0.159117   0.054360 -2.927100 3.7207e-03 **
## av99        -0.050816   0.009750 -5.212035 3.7803e-07 ***
## av88        -0.027063   0.034871 -0.776065 4.3841e-01
## avmiss       0.108204   0.076552  1.413477 1.5870e-01
## educ_em      0.015960   0.001011 15.791062  < 2.2e-16 ***
## em0          0.020887   0.049600  0.421104 6.7402e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.448394     Adj. R2: 0.192826
##                  Within R2: 0.040257
##
##
## $results_men[[3]]
## $results_men[[3]]$falsification1
## OLS estimation, Dep. Var.: units
## Observations: 195,501
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past  1.376410   1.741871  0.790191 4.3013e-01
## educ_av       0.409679   0.019910 20.576408  < 2.2e-16 ***
## av0          -7.153918   1.635633 -4.373791 1.7636e-05 ***
## av99         -1.496057   0.275964 -5.421210 1.3437e-07 ***
## av88         -0.944144   0.826305 -1.142610 2.5424e-01
## avmiss        3.150471   1.177300  2.676012 7.9199e-03 **
## educ_em       0.388537   0.025366 15.317457  < 2.2e-16 ***
## em0           0.938458   1.639883  0.572271 5.6763e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 9.95026     Adj. R2: 0.20261
##                 Within R2: 0.057965
##
## $results_men[[3]]$falsification2
## OLS estimation, Dep. Var.: units
## Observations: 198,239
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
```

```
## mfemale_fut  2.174342   1.677919  1.295857 1.9616e-01
## educ_av      0.411219   0.019910 20.654008  < 2.2e-16 ***
## av0         -7.162965   1.635623 -4.379349 1.7199e-05 ***
## av99        -1.528438   0.266510 -5.735012 2.6719e-08 ***
## av88        -0.964895   0.833833 -1.157181 2.4825e-01
## avmiss       3.160418   1.173797  2.692475 7.5478e-03 **
## educ_em      0.387108   0.025166 15.382366  < 2.2e-16 ***
## em0          0.919726   1.633119  0.563171 5.7380e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 9.93989     Adj. R2: 0.203918
##                 Within R2: 0.057958
##
##
## $results_men[[4]]
## $results_men[[4]]$falsification1
## OLS estimation, Dep. Var.: madmug
## Observations: 195,501
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past  0.101936   0.108591  0.938721 3.4874e-01
## educ_av       0.040920   0.001941 21.077604  < 2.2e-16 ***
## av0          -0.188957   0.094121 -2.007592 4.5713e-02 *
## av99         -0.086247   0.014712 -5.862273 1.3689e-08 ***
## av88         -0.088766   0.038778 -2.289067 2.2871e-02 *
## avmiss        0.368813   0.115150  3.202899 1.5287e-03 **
## educ_em       0.034520   0.002063 16.734949  < 2.2e-16 ***
## em0           0.037163   0.083102  0.447201 6.5510e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.81589     Adj. R2: 0.201565
##                 Within R2: 0.059577
##
## $results_men[[4]]$falsification2
## OLS estimation, Dep. Var.: madmug
## Observations: 198,239
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut  0.123269   0.107142  1.150521 2.5097e-01
## educ_av      0.040779   0.001940 21.021220  < 2.2e-16 ***
## av0         -0.190580   0.093411 -2.040234 4.2326e-02 *
## av99        -0.088658   0.014730 -6.018786 5.8772e-09 ***
## av88        -0.087841   0.039572 -2.219768 2.7288e-02 *
## avmiss       0.369406   0.115082  3.209934 1.4925e-03 **
## educ_em      0.034562   0.002045 16.899726  < 2.2e-16 ***
## em0          0.033328   0.083018  0.401454 6.8841e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.814991     Adj. R2: 0.201484
```

```
##                   Within R2: 0.059378
##
##
## $results_men[[5]]
## $results_men[[5]]$falsification1
## OLS estimation, Dep. Var.: university
## Observations: 195,501
## Fixed-effects: year: 8,  semelmos: 263
## Standard-errors: Clustered (semelmos)
##               Estimate Std. Error   t value   Pr(>|t|)
## mfemale_past  0.088866   0.069719  1.274629 2.0357e-01
## educ_av       0.018334   0.000865 21.192612  < 2.2e-16 ***
## av0          -0.118447   0.050770 -2.333032 2.0404e-02 *
## av99         -0.043678   0.009738 -4.485341 1.0902e-05 ***
## av88         -0.036465   0.027692 -1.316784 1.8906e-01
## avmiss        0.130188   0.082538  1.577323 1.1593e-01
## educ_em       0.017525   0.001094 16.016952  < 2.2e-16 ***
## em0           0.023140   0.046208  0.500775 6.1695e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.438238     Adj. R2: 0.228591
##                   Within R2: 0.053488
##
## $results_men[[5]]$falsification2
## OLS estimation, Dep. Var.: university
## Observations: 198,239
## Fixed-effects: year: 8,  semelmos: 264
## Standard-errors: Clustered (semelmos)
##              Estimate Std. Error   t value   Pr(>|t|)
## mfemale_fut  0.068063   0.071635  0.950136 3.4292e-01
## educ_av      0.018416   0.000870 21.163750  < 2.2e-16 ***
## av0         -0.118636   0.050054 -2.370145 1.8503e-02 *
## av99        -0.046437   0.009471 -4.902880 1.6550e-06 ***
## av88        -0.035099   0.027616 -1.270977 2.0486e-01
## avmiss       0.130783   0.082569  1.583937 1.1441e-01
## educ_em      0.017542   0.001083 16.193372  < 2.2e-16 ***
## em0          0.022405   0.045843  0.488728 6.2544e-01
## ... 22 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.437994     Adj. R2: 0.229228
##                   Within R2: 0.053459
```

Note: Results are qualitatively similar, but effect size differ with respect to the paper's effects. Differences seem to come from different sample selection processes. For more information, see SAS code.

III.C. Heterogeneous Treatment Effects

Table 4 (Heterogeneous Effects by Parental Education and Immigration Status). Parent's average education (if the education of one parent is missing, use the other's):

```
# Calculando a média de educação dos pais
dados <- dados %>%
  mutate(
```

```r
    edtot = case_when(
      av0 != 1 & av88 != 1 & av99 != 1 & avmiss != 1 & em0 != 1 & em88 != 1 & em99 != 1 & emmiss != 1 ~
      is.na(rowMeans(select(., educ_av, educ_em))) & av0 != 1 & av88 != 1 & av99 != 1 & avmiss != 1 ~ ed
      is.na(rowMeans(select(., educ_av, educ_em))) & em0 != 1 & em88 != 1 & em99 != 1 & emmiss != 1 ~ ed
      TRUE ~ NA_real_
    )
  )

# Lista de variáveis dependentes
depvars <- c("wmean", "zakaibag", "units", "madmug", "university")

# Análise para mulheres
results_women <- lapply(depvars, function(depvar) {
  list(
    # Resultados por educação dos pais: baixa educação (coluna 2)
    low_education = feols(as.formula(paste(depvar, "~ mfemale +",
                                           paste(cvars, collapse = " + "),
                                           "| year + semelmos + semelmos:year")),
                          data = dados %>% filter(female == 1 & edtot <= 12),
                          cluster = ~ semelmos),

    # Resultados por educação dos pais: alta educação (coluna 3)
    high_education = feols(as.formula(paste(depvar, "~ mfemale +",
                                            paste(cvars, collapse = " + "),
                                            "| year + semelmos + semelmos:year")),
                           data = dados %>% filter(female == 1 & edtot > 12 & !is.na(edtot)),
                           cluster = ~ semelmos),

    # Resultados para novos imigrantes (coluna 4)
    new_immigrants = feols(as.formula(paste(depvar, "~ mfemale +",
                                            paste(cvars, collapse = " + "),
                                            "| year + semelmos + semelmos:year")),
                           data = dados %>% filter(female == 1 & ole == 1),
                           cluster = ~ semelmos)
  )
})
```

```
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca

## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee

## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca

## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee

## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca

## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee

## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca

## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee

## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
```

```
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
```

```
# Análise para homens
results_men <- lapply(depvars, function(depvar) {
  list(
    # Resultados por educação dos pais: baixa educação (coluna 6)
    low_education = feols(as.formula(paste(depvar, "~ mfemale +",
                                           paste(cvars, collapse = " + "),
                                           "| year + semelmos + semelmos:year")),
                          data = dados %>% filter(female == 0 & edtot <= 12),
                          cluster = ~ semelmos),

    # Resultados por educação dos pais: alta educação (coluna 7)
    high_education = feols(as.formula(paste(depvar, "~ mfemale +",
                                            paste(cvars, collapse = " + "),
                                            "| year + semelmos + semelmos:year")),
                           data = dados %>% filter(female == 0 & edtot > 12 & !is.na(edtot)),
                           cluster = ~ semelmos),

    # Resultados para novos imigrantes (coluna 8)
    new_immigrants = feols(as.formula(paste(depvar, "~ mfemale +",
                                            paste(cvars, collapse = " + "),
                                            "| year + semelmos + semelmos:year")),
                           data = dados %>% filter(female == 0 & ole == 1),
                           cluster = ~ semelmos)
  )
})
```

```
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'av0', 'av99', 'av88', 'avmiss', 'em0' and 14 others have been removed beca
## The variables 'mfemale', 'ahmiss', 'ole', 'm_educ_av', 'm_educ_em', 'm_m_ahim' and 7 others have bee
```

```
# Exibir resultados
list(
  results_women = results_women,
  results_men = results_men
)
```

```
## $results_women
## $results_women[[1]]
## $results_women[[1]]$low_education
## OLS estimation, Dep. Var.: wmean
## Observations: 113,282
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,890
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.416638   0.067105   6.208797 2.0714e-09 ***
## educ_em   0.515148   0.081773   6.299703 1.2477e-09 ***
## m_ahim   -0.869993   0.100487  -8.657751 4.9200e-16 ***
## ah0     -54.988693   3.001381 -18.321129  < 2.2e-16 ***
## ah88      7.848675   2.426643   3.234375 1.3750e-03 **
## ah99     -0.540424   0.461481  -1.171064 2.4263e-01
## ole     -12.013428   0.864109 -13.902679  < 2.2e-16 ***
## asiafr    0.202986   0.173316   1.171190 2.4258e-01
## etyopia  -0.220257   1.509206  -0.145943 8.8408e-01
## heveram  -0.836491   0.404037  -2.070333 3.9397e-02 *
## euram     1.293019   0.267508   4.833579 2.2826e-06 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 22.3     Adj. R2: 0.145224
##              Within R2: 0.019493
##
## $results_women[[1]]$high_education
## OLS estimation, Dep. Var.: wmean
## Observations: 80,759
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,868
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.523245   0.032949  15.880346  < 2.2e-16 ***
## educ_em   0.627462   0.034153  18.371961  < 2.2e-16 ***
## m_ahim    0.030516   0.092568   0.329664 7.4192e-01
## ah0     -65.171080   5.040239 -12.930156  < 2.2e-16 ***
## ah88     -3.607767   4.405454  -0.818932 4.1357e-01
## ah99     -1.710458   0.489101  -3.497145 5.5191e-04 ***
## ole     -10.222344   1.133646  -9.017230  < 2.2e-16 ***
## asiafr   -1.329662   0.243739  -5.455267 1.1283e-07 ***
## etyopia  -8.976555   4.663532  -1.924840 5.5328e-02 .
## heveram  -3.251587   0.341703  -9.515825  < 2.2e-16 ***
## euram    -1.030060   0.179552  -5.736830 2.6465e-08 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 18.7     Adj. R2: 0.155324
##              Within R2: 0.040722
##
## $results_women[[1]]$new_immigrants
## OLS estimation, Dep. Var.: wmean
## Observations: 13,729
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,525
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error   t value   Pr(>|t|)
```

```
## educ_av   0.616088   0.148222  4.156527 4.4362e-05 ***
## av0     -25.662631   9.562411 -2.683699 7.7656e-03 **
## av99     -3.351544   1.509974 -2.219604 2.7340e-02 *
## av88     -5.708993   3.701788 -1.542226 1.2428e-01
## avmiss   13.025446   9.999018  1.302673 1.9388e-01
## educ_em   0.931511   0.156963  5.934572 9.7391e-09 ***
## em0       3.838061   8.164186  0.470109 6.3869e-01
## em99    -12.151306   2.640012 -4.602746 6.6258e-06 ***
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 29.4     Adj. R2: 0.250454
##              Within R2: 0.059101
##
##
## $results_women[[2]]
## $results_women[[2]]$low_education
## OLS estimation, Dep. Var.: zakaibag
## Observations: 113,282
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,890
## Standard-errors: Clustered (semelmos)
##            Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.00826189   0.001181   6.997267 2.1634e-11 ***
## educ_em  0.01252217   0.001671   7.495006 1.0134e-12 ***
## m_ahim  -0.01589671   0.001734  -9.169785  < 2.2e-16 ***
## ah0     -0.46159011   0.036973 -12.484479  < 2.2e-16 ***
## ah88     0.11561136   0.101721   1.136557 2.5676e-01
## ah99     0.00518107   0.007357   0.704202 4.8193e-01
## ole     -0.09905877   0.013869  -7.142397 8.9850e-12 ***
## asiafr  -0.00000332   0.004106  -0.000808 9.9936e-01
## etyopia -0.02718379   0.025658  -1.059472 2.9036e-01
## heveram  0.02168032   0.008017   2.704435 7.2886e-03 **
## euram    0.02505041   0.005876   4.262909 2.8149e-05 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.453802     Adj. R2: 0.158827
##                  Within R2: 0.009274
##
## $results_women[[2]]$high_education
## OLS estimation, Dep. Var.: zakaibag
## Observations: 80,759
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,868
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.010303   0.000775  13.289122  < 2.2e-16 ***
## educ_em  0.012562   0.000818  15.364087  < 2.2e-16 ***
## m_ahim  -0.003973   0.001656  -2.398705 1.7150e-02 *
## ah0     -0.734815   0.038422 -19.124944  < 2.2e-16 ***
## ah88    -0.062847   0.083849  -0.749521 4.5421e-01
## ah99    -0.014500   0.007867  -1.843054 6.6446e-02 .
## ole     -0.117633   0.012559  -9.366636  < 2.2e-16 ***
## asiafr  -0.021942   0.005416  -4.051068 6.7154e-05 ***
```

```
## etyopia -0.180064   0.088340  -2.038307 4.2520e-02 *
## heveram -0.034367   0.005812  -5.913324 1.0381e-08 ***
## euram   -0.017585   0.003902  -4.506514 9.9250e-06 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.388515     Adj. R2: 0.121493
##                  Within R2: 0.018421
##
## $results_women[[2]]$new_immigrants
## OLS estimation, Dep. Var.: zakaibag
## Observations: 13,729
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,525
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.009738   0.001875  5.194712 4.2439e-07 ***
## av0     -0.222690   0.090619 -2.457421 1.4671e-02 *
## av99    -0.043777   0.022528 -1.943203 5.3110e-02 .
## av88    -0.073348   0.049833 -1.471874 1.4231e-01
## avmiss   0.152706   0.113791  1.341986 1.8081e-01
## educ_em  0.019020   0.002260  8.417772 2.9611e-15 ***
## em0      0.071426   0.079462  0.898871 3.6958e-01
## em99    -0.109290   0.034659 -3.153264 1.8110e-03 **
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $co
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.419004     Adj. R2: 0.190546
##                  Within R2: 0.048715
##
##
## $results_women[[3]]
## $results_women[[3]]$low_education
## OLS estimation, Dep. Var.: units
## Observations: 113,282
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,890
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.215437   0.030261   7.119352 1.0338e-11 ***
## educ_em   0.314495   0.038734   8.119445 1.8203e-14 ***
## m_ahim   -0.409075   0.043868  -9.325166  < 2.2e-16 ***
## ah0     -17.128951   0.784879 -21.823689  < 2.2e-16 ***
## ah88      4.330327   0.803817   5.387204 1.5892e-07 ***
## ah99      0.021629   0.165711   0.130522 8.9625e-01
## ole      -3.847670   0.342089 -11.247581  < 2.2e-16 ***
## asiafr    0.050742   0.071886   0.705868 4.8089e-01
## etyopia  -0.537984   0.626710  -0.858427 3.9144e-01
## heveram   0.585140   0.163925   3.569553 4.2496e-04 ***
## euram     0.630967   0.121955   5.173748 4.5556e-07 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.74869     Adj. R2: 0.174555
##                  Within R2: 0.019552
```

```
##
## $results_women[[3]]$high_education
## OLS estimation, Dep. Var.: units
## Observations: 80,759
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,868
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.210053   0.013351  15.733642  < 2.2e-16 ***
## educ_em   0.250993   0.015040  16.687940  < 2.2e-16 ***
## m_ahim   -0.022724   0.034437  -0.659878 5.0991e-01
## ah0     -22.507872   0.909260 -24.754055  < 2.2e-16 ***
## ah88     -0.098479   1.509255  -0.065250 9.4802e-01
## ah99     -0.570397   0.177609  -3.211530 1.4846e-03 **
## ole      -3.462941   0.385067  -8.993083  < 2.2e-16 ***
## asiafr   -0.248761   0.094002  -2.646342 8.6272e-03 **
## etyopia  -3.468748   1.789260  -1.938649 5.3614e-02 .
## heveram  -0.245919   0.126863  -1.938459 5.3637e-02 .
## euram    -0.323471   0.074233  -4.357510 1.8879e-05 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 7.30116     Adj. R2: 0.131745
##                 Within R2: 0.027678
##
## $results_women[[3]]$new_immigrants
## OLS estimation, Dep. Var.: units
## Observations: 13,729
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,525
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.271534   0.046931  5.785776 2.1417e-08 ***
## av0     -7.721988   2.963103 -2.606048 9.7068e-03 **
## av99    -1.364257   0.565555 -2.412244 1.6573e-02 *
## av88    -2.162536   1.150878 -1.879031 6.1399e-02 .
## avmiss   3.423700   3.037172  1.127266 2.6071e-01
## educ_em  0.477899   0.054100  8.833590  < 2.2e-16 ***
## em0      1.602079   2.238364  0.715736 4.7482e-01
## em99    -3.516378   0.930625 -3.778512 1.9712e-04 ***
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.3     Adj. R2: 0.257551
##              Within R2: 0.065478
##
##
## $results_women[[4]]
## $results_women[[4]]$low_education
## OLS estimation, Dep. Var.: madmug
## Observations: 113,282
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,890
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.009531   0.002469  3.860950 1.4223e-04 ***
```

```
## educ_em  0.014229   0.002688  5.292612 2.5442e-07 ***
## m_ahim  -0.018055   0.002290 -7.884221 8.4607e-14 ***
## ah0     -0.308212   0.056657 -5.439991 1.2188e-07 ***
## ah88    -0.085627   0.108740 -0.787452 4.3173e-01
## ah99     0.023883   0.011341  2.106027 3.6150e-02 *
## ole     -0.121116   0.017650 -6.862010 4.8549e-11 ***
## asiafr  -0.005933   0.005854 -1.013371 3.1181e-01
## etyopia -0.038158   0.036001 -1.059913 2.9016e-01
## heveram  0.088352   0.011523  7.667258 3.4088e-13 ***
## euram    0.059673   0.009235  6.461559 4.9965e-10 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.696406     Adj. R2: 0.111169
##                  Within R2: 0.006471
##
## $results_women[[4]]$high_education
## OLS estimation, Dep. Var.: madmug
## Observations: 80,759
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,868
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.046255   0.001975 23.420745  < 2.2e-16 ***
## educ_em  0.041274   0.002073 19.914224  < 2.2e-16 ***
## m_ahim  -0.010595   0.003636 -2.913828 3.8773e-03 **
## ah0     -0.955410   0.123910 -7.710500 2.5869e-13 ***
## ah88     0.057316   0.143894  0.398319 6.9072e-01
## ah99    -0.007611   0.015241 -0.499393 6.1792e-01
## ole     -0.237071   0.025968 -9.129492  < 2.2e-16 ***
## asiafr  -0.030579   0.011607 -2.634517 8.9254e-03 **
## etyopia -0.380501   0.155208 -2.451553 1.4875e-02 *
## heveram  0.055768   0.014260  3.910918 1.1709e-04 ***
## euram   -0.014495   0.010751 -1.348295 1.7872e-01
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.91765     Adj. R2: 0.124522
##                 Within R2: 0.030654
##
## $results_women[[4]]$new_immigrants
## OLS estimation, Dep. Var.: madmug
## Observations: 13,729
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,525
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.020501   0.003212  6.383351 8.3146e-10 ***
## av0     -0.182210   0.086956 -2.095423 3.7136e-02 *
## av99    -0.028242   0.035506 -0.795429 4.2712e-01
## av88    -0.117190   0.079088 -1.481769 1.3966e-01
## avmiss   0.360283   0.189669  1.899533 5.8641e-02 .
## educ_em  0.026896   0.003301  8.148898 1.7447e-14 ***
## em0      0.181885   0.112445  1.617539 1.0702e-01
## em99    -0.103875   0.053491 -1.941932 5.3264e-02 .
## ... 10 coefficients remaining (display them with summary() or use argument n)
```

```
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.701673     Adj. R2: 0.193067
##                  Within R2: 0.035308
##
##
## $results_women[[5]]
## $results_women[[5]]$low_education
## OLS estimation, Dep. Var.: university
## Observations: 113,282
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,890
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.009141   0.001221   7.488065 1.0586e-12 ***
## educ_em  0.013012   0.001790   7.267947 4.1626e-12 ***
## m_ahim  -0.019170   0.001723 -11.124452  < 2.2e-16 ***
## ah0     -0.344732   0.043100  -7.998463 4.0248e-14 ***
## ah88     0.024555   0.147039   0.166999 8.6750e-01
## ah99     0.016058   0.007359   2.182110 2.9987e-02 *
## ole     -0.149333   0.011929 -12.518756  < 2.2e-16 ***
## asiafr  -0.000903   0.004471  -0.201918 8.4014e-01
## etyopia -0.047840   0.023717  -2.017069 4.4704e-02 *
## heveram  0.037416   0.007150   5.232955 3.4125e-07 ***
## euram    0.048623   0.005941   8.184439 1.1852e-14 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.441792     Adj. R2: 0.192086
##                  Within R2: 0.013983
##
## $results_women[[5]]$high_education
## OLS estimation, Dep. Var.: university
## Observations: 80,759
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,868
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.012672   0.000859  14.745878  < 2.2e-16 ***
## educ_em  0.014643   0.000865  16.921405  < 2.2e-16 ***
## m_ahim  -0.007976   0.001747  -4.566415 7.6281e-06 ***
## ah0     -0.650010   0.047540 -13.672843  < 2.2e-16 ***
## ah88    -0.031861   0.083552  -0.381329 7.0327e-01
## ah99    -0.001590   0.008568  -0.185632 8.5288e-01
## ole     -0.167811   0.013404 -12.519777  < 2.2e-16 ***
## asiafr  -0.021882   0.005441  -4.021889 7.5491e-05 ***
## etyopia -0.246324   0.079130  -3.112884 2.0572e-03 **
## heveram -0.033455   0.005990  -5.585346 5.8127e-08 ***
## euram   -0.009995   0.004024  -2.483977 1.3616e-02 *
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.39912     Adj. R2: 0.152371
##                 Within R2: 0.027026
##
```

```
## $results_women[[5]]$new_immigrants
## OLS estimation, Dep. Var.: university
## Observations: 13,729
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,525
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error  t value   Pr(>|t|)
## educ_av  0.012042   0.001769  6.80795 7.2562e-11 ***
## av0     -0.106252   0.062158 -1.70939 8.8616e-02 .
## av99    -0.049478   0.021707 -2.27933 2.3487e-02 *
## av88    -0.073597   0.035752 -2.05855 4.0570e-02 *
## avmiss   0.175195   0.142331  1.23090 2.1951e-01
## educ_em  0.018589   0.002304  8.06711 2.9736e-14 ***
## em0      0.082344   0.071656  1.14915 2.5159e-01
## em99    -0.074324   0.030887 -2.40635 1.6836e-02 *
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.402102     Adj. R2: 0.213766
##                  Within R2: 0.048079
##
##
##
## $results_men
## $results_men[[1]]
## $results_men[[1]]$low_education
## OLS estimation, Dep. Var.: wmean
## Observations: 105,251
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,887
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.530460   0.074479   7.122260 1.0157e-11 ***
## educ_em   0.423997   0.088012   4.817471 2.4585e-06 ***
## m_ahim   -0.985380   0.107169  -9.194642  < 2.2e-16 ***
## ah0     -46.051849   2.809998 -16.388568  < 2.2e-16 ***
## ah88     10.580514   3.842309   2.753686 6.3037e-03 **
## ah99     -0.853869   0.614091  -1.390461 1.6556e-01
## ole     -10.753417   1.212357  -8.869840  < 2.2e-16 ***
## asiafr   -0.022798   0.265447  -0.085886 9.3162e-01
## etyopia  -2.326832   2.094822  -1.110754 2.6769e-01
## heveram  -0.032151   0.453275  -0.070931 9.4351e-01
## euram     2.334478   0.340394   6.858167 4.9670e-11 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 26.0     Adj. R2: 0.138139
##              Within R2: 0.013587
##
## $results_men[[1]]$high_education
## OLS estimation, Dep. Var.: wmean
## Observations: 81,977
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,883
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
```

```
## educ_av   0.652527   0.036977  17.646650  < 2.2e-16 ***
## educ_em   0.563760   0.045760  12.319832  < 2.2e-16 ***
## m_ahim   -0.008653   0.117929  -0.073377 9.4156e-01
## ah0     -57.501593   3.853251 -14.922877  < 2.2e-16 ***
## ah88      2.743144   6.707522   0.408965 6.8290e-01
## ah99     -0.637906   0.542047  -1.176846 2.4032e-01
## ole     -11.243360   1.307647  -8.598160 7.3834e-16 ***
## asiafr   -1.670755   0.254804  -6.557010 2.8911e-10 ***
## etyopia -13.222405   4.108512  -3.218295 1.4513e-03 **
## heveram  -3.745012   0.317402 -11.798970  < 2.2e-16 ***
## euram    -1.266560   0.207274  -6.110572 3.5615e-09 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 21.5     Adj. R2: 0.15435
##              Within R2: 0.034658
##
## $results_men[[1]]$new_immigrants
## OLS estimation, Dep. Var.: wmean
## Observations: 12,787
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,503
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error   t value   Pr(>|t|)
## educ_av   0.623929   0.144999  4.302987 2.4164e-05 ***
## av0     -14.813199  10.183368 -1.454646 1.4702e-01
## av99     -4.589250   1.656753 -2.770027 6.0240e-03 **
## av88     -5.241424   3.418430 -1.533284 1.2647e-01
## avmiss   -3.673219  12.895182 -0.284852 7.7599e-01
## educ_em   0.427244   0.177535  2.406533 1.6828e-02 *
## em0      -6.387463   7.314083 -0.873310 3.8333e-01
## em99    -11.499199   2.543398 -4.521195 9.4842e-06 ***
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 29.3     Adj. R2: 0.257112
##              Within R2: 0.058938
##
##
## $results_men[[2]]
## $results_men[[2]]$low_education
## OLS estimation, Dep. Var.: zakaibag
## Observations: 105,251
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,887
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.008689   0.001283  6.774676 8.1368e-11 ***
## educ_em  0.007036   0.001609  4.373764 1.7615e-05 ***
## m_ahim  -0.015955   0.001719 -9.281097  < 2.2e-16 ***
## ah0     -0.286163   0.034800 -8.223035 9.1773e-15 ***
## ah88     0.149386   0.111181  1.343631 1.8023e-01
## ah99     0.002502   0.008300  0.301478 7.6329e-01
## ole     -0.069986   0.017117 -4.088809 5.7664e-05 ***
## asiafr  -0.002960   0.004991 -0.593195 5.5356e-01
```

```
## etyopia -0.036692   0.028790 -1.274464 2.0362e-01
## heveram  0.017507   0.008310  2.106585 3.6101e-02 *
## euram    0.047648   0.006503  7.326798 2.8938e-12 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.445514     Adj. R2: 0.161489
##                  Within R2: 0.007426
##
## $results_men[[2]]$high_education
## OLS estimation, Dep. Var.: zakaibag
## Observations: 81,977
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,883
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.013251   0.000780 16.988383  < 2.2e-16 ***
## educ_em  0.012346   0.001007 12.260458  < 2.2e-16 ***
## m_ahim  -0.003297   0.002007 -1.642638 1.0165e-01
## ah0     -0.595168   0.061489 -9.679313  < 2.2e-16 ***
## ah88     0.038299   0.064289  0.595739 5.5186e-01
## ah99    -0.011256   0.008117 -1.386782 1.6668e-01
## ole     -0.116771   0.013812 -8.454506 1.9523e-15 ***
## asiafr  -0.036731   0.005441 -6.750362 9.3876e-11 ***
## etyopia -0.323427   0.075018 -4.311314 2.2965e-05 ***
## heveram -0.057124   0.006121 -9.331898  < 2.2e-16 ***
## euram   -0.032504   0.004160 -7.814181 1.3301e-13 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.426234     Adj. R2: 0.141361
##                  Within R2: 0.018967
##
## $results_men[[2]]$new_immigrants
## OLS estimation, Dep. Var.: zakaibag
## Observations: 12,787
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,503
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.010349   0.002035  5.085233 7.1887e-07 ***
## av0     -0.094149   0.069575 -1.353196 1.7721e-01
## av99    -0.048521   0.020937 -2.317487 2.1282e-02 *
## av88    -0.047658   0.034756 -1.371200 1.7154e-01
## avmiss   0.093996   0.157379  0.597258 5.5087e-01
## educ_em  0.007190   0.002407  2.987147 3.0948e-03 **
## em0     -0.048907   0.058082 -0.842040 4.0057e-01
## em99    -0.067606   0.027874 -2.425447 1.5996e-02 *
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $co
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.406157     Adj. R2: 0.202057
##                  Within R2: 0.029171
##
##
```

```
## $results_men[[3]]
## $results_men[[3]]$low_education
## OLS estimation, Dep. Var.: units
## Observations: 105,251
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,887
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error    t value   Pr(>|t|)
## educ_av   0.277226   0.033814   8.198521 1.0797e-14 ***
## educ_em   0.226194   0.043777   5.166920 4.7092e-07 ***
## m_ahim   -0.485717   0.046772 -10.384809  < 2.2e-16 ***
## ah0     -12.746412   0.997074 -12.783813  < 2.2e-16 ***
## ah88      4.907604   2.098050   2.339126 2.0078e-02 *
## ah99     -0.141217   0.221209  -0.638390 5.2378e-01
## ole      -3.039790   0.476897  -6.374105 8.2087e-10 ***
## asiafr   -0.088301   0.118002  -0.748307 4.5494e-01
## etyopia  -0.397165   0.838156  -0.473855 6.3600e-01
## heveram   1.006536   0.221246   4.549392 8.2229e-06 ***
## euram     1.305807   0.151538   8.617009 6.4948e-16 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.1     Adj. R2: 0.176865
##              Within R2: 0.014797
##
## $results_men[[3]]$high_education
## OLS estimation, Dep. Var.: units
## Observations: 81,977
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,883
## Standard-errors: Clustered (semelmos)
##           Estimate Std. Error   t value   Pr(>|t|)
## educ_av   0.292458   0.016460  17.76787  < 2.2e-16 ***
## educ_em   0.251097   0.020285  12.37875  < 2.2e-16 ***
## m_ahim   -0.052901   0.048105  -1.09969 2.7247e-01
## ah0     -19.914833   1.183155 -16.83197  < 2.2e-16 ***
## ah88      3.185175   2.294500   1.38818 1.6626e-01
## ah99     -0.312020   0.206025  -1.51448 1.3111e-01
## ole      -4.128305   0.419428  -9.84270  < 2.2e-16 ***
## asiafr   -0.684495   0.106148  -6.44848 5.3831e-10 ***
## etyopia  -6.312271   1.760367  -3.58577 4.0056e-04 ***
## heveram  -0.410282   0.144660  -2.83618 4.9207e-03 **
## euram    -0.492042   0.086073  -5.71660 2.9424e-08 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 8.90692     Adj. R2: 0.138606
##                 Within R2: 0.025569
##
## $results_men[[3]]$new_immigrants
## OLS estimation, Dep. Var.: units
## Observations: 12,787
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,503
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.305890   0.060839  5.027863 9.4436e-07 ***
```

```
## av0     -3.659320   3.171074 -1.153968 2.4961e-01
## av99    -1.601244   0.610827 -2.621438 9.2906e-03 **
## av88    -1.763425   1.362334 -1.294415 1.9671e-01
## avmiss   4.150319   5.553973  0.747270 4.5560e-01
## educ_em  0.201187   0.072628  2.770117 6.0224e-03 **
## em0     -2.117334   2.360476 -0.896995 3.7058e-01
## em99    -3.317848   0.908007 -3.653989 3.1436e-04 ***
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $co
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 11.1     Adj. R2: 0.261454
##              Within R2: 0.050041
##
##
## $results_men[[4]]
## $results_men[[4]]$low_education
## OLS estimation, Dep. Var.: madmug
## Observations: 105,251
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,887
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.008270   0.002395   3.453347 6.4512e-04 ***
## educ_em  0.007820   0.003158   2.476260 1.3906e-02 *
## m_ahim  -0.021785   0.002103 -10.359051  < 2.2e-16 ***
## ah0     -0.325286   0.040448  -8.042024 3.0270e-14 ***
## ah88     0.112400   0.126056   0.891671 3.7338e-01
## ah99     0.019492   0.014172   1.375435 1.7017e-01
## ole     -0.084351   0.021302  -3.959759 9.6637e-05 ***
## asiafr  -0.008108   0.007297  -1.111246 2.6748e-01
## etyopia -0.075915   0.037756  -2.010652 4.5383e-02 *
## heveram  0.097031   0.012233   7.931635 6.2205e-14 ***
## euram    0.072916   0.010280   7.093173 1.2120e-11 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.68864     Adj. R2: 0.138879
##                 Within R2: 0.005981
##
## $results_men[[4]]$high_education
## OLS estimation, Dep. Var.: madmug
## Observations: 81,977
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,883
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.046385   0.001932 24.013517  < 2.2e-16 ***
## educ_em  0.038104   0.002215 17.203896  < 2.2e-16 ***
## m_ahim  -0.006127   0.004076 -1.503265 1.3397e-01
## ah0     -1.049811   0.147170 -7.133329 9.4954e-12 ***
## ah88     0.100829   0.139670  0.721913 4.7099e-01
## ah99    -0.022481   0.015955 -1.409096 1.5999e-01
## ole     -0.222116   0.027925 -7.954109 5.3745e-14 ***
## asiafr  -0.076806   0.010805 -7.108490 1.1044e-11 ***
## etyopia -0.380168   0.132960 -2.859261 4.5866e-03 **
```

```
## heveram  0.017446   0.015145  1.151952 2.5039e-01
## euram   -0.049377   0.010763 -4.587421 6.9511e-06 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.910458     Adj. R2: 0.153624
##                  Within R2: 0.029641
##
## $results_men[[4]]$new_immigrants
## OLS estimation, Dep. Var.: madmug
## Observations: 12,787
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,503
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.020534   0.004036  5.087816 7.1006e-07 ***
## av0      0.025233   0.126775  0.199037 8.4239e-01
## av99    -0.068215   0.033452 -2.039187 4.2479e-02 *
## av88    -0.173509   0.045189 -3.839638 1.5605e-04 ***
## avmiss   0.224918   0.283938  0.792137 4.2903e-01
## educ_em  0.014758   0.004515  3.268730 1.2314e-03 **
## em0     -0.120188   0.098417 -1.221220 2.2315e-01
## em99    -0.060095   0.041693 -1.441371 1.5073e-01
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $co
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.693338     Adj. R2: 0.230968
##                  Within R2: 0.029922
##
##
## $results_men[[5]]
## $results_men[[5]]$low_education
## OLS estimation, Dep. Var.: university
## Observations: 105,251
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,887
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.009056   0.001271   7.12562 9.9514e-12 ***
## educ_em  0.007240   0.001699   4.26168 2.8294e-05 ***
## m_ahim  -0.019859   0.001557 -12.75330  < 2.2e-16 ***
## ah0     -0.224595   0.029106  -7.71637 2.4916e-13 ***
## ah88     0.127686   0.089135   1.43250 1.5319e-01
## ah99     0.020631   0.008013   2.57449 1.0586e-02 *
## ole     -0.115545   0.014155  -8.16259 1.3694e-14 ***
## asiafr  -0.007106   0.004843  -1.46724 1.4351e-01
## etyopia -0.063454   0.025549  -2.48366 1.3628e-02 *
## heveram  0.032725   0.008775   3.72955 2.3496e-04 ***
## euram    0.066784   0.006473  10.31695  < 2.2e-16 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.424151     Adj. R2: 0.186856
##                  Within R2: 0.012228
##
```

```
## $results_men[[5]]$high_education
## OLS estimation, Dep. Var.: university
## Observations: 81,977
## Fixed-effects: year: 8,  semelmos: 264,  year:semelmos: 1,883
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error    t value   Pr(>|t|)
## educ_av  0.015346   0.000853  17.991643  < 2.2e-16 ***
## educ_em  0.014002   0.001076  13.017852  < 2.2e-16 ***
## m_ahim  -0.006017   0.002119  -2.839828 4.8665e-03 **
## ah0     -0.544887   0.070893  -7.686093 3.0232e-13 ***
## ah88     0.009020   0.063060   0.143037 8.8637e-01
## ah99    -0.007265   0.007648  -0.949878 3.4305e-01
## ole     -0.161952   0.012873 -12.580612  < 2.2e-16 ***
## asiafr  -0.043526   0.005544  -7.850731 1.0507e-13 ***
## etyopia -0.354693   0.068817  -5.154157 5.0099e-07 ***
## heveram -0.050656   0.006499  -7.794248 1.5121e-13 ***
## euram   -0.025376   0.004226  -6.004325 6.3567e-09 ***
## ... 20 variables were removed because of collinearity (mfemale, av0 and 18 others [full set in $colli
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.429634     Adj. R2: 0.166189
##                  Within R2: 0.025688
##
## $results_men[[5]]$new_immigrants
## OLS estimation, Dep. Var.: university
## Observations: 12,787
## Fixed-effects: year: 8,  semelmos: 252,  year:semelmos: 1,503
## Standard-errors: Clustered (semelmos)
##          Estimate Std. Error   t value   Pr(>|t|)
## educ_av  0.012338   0.001963  6.284331 1.4462e-09 ***
## av0     -0.063792   0.062361 -1.022949 3.0732e-01
## av99    -0.039242   0.019475 -2.014935 4.4978e-02 *
## av88    -0.069283   0.051137 -1.354856 1.7668e-01
## avmiss   0.056695   0.152463  0.371859 7.1031e-01
## educ_em  0.006298   0.002391  2.634395 8.9527e-03 **
## em0     -0.022403   0.057008 -0.392981 6.9467e-01
## em99    -0.050250   0.024141 -2.081521 3.8400e-02 *
## ... 10 coefficients remaining (display them with summary() or use argument n)
## ... 13 variables were removed because of collinearity (mfemale, ahmiss and 11 others [full set in $co
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.385835     Adj. R2: 0.198758
##                  Within R2: 0.025975
```

Note: As shown in the paper, results are coming from children with low parent's average education and new immigrants