Homework 6

Gergel Anastasia

6/12/2021

Exercise 1

```r
library(haven)

data <- read_dta("/Users/herrhellana/Dropbox/_NYU studies/Quant I/home assignments/HW6/gendereq.dta")

library(sandwich)
library(stargazer)

m1 <- lm(data$femjobs ~ data$relig)
heterosked1 <- vcovHC(m1, type = 'HC1')

m2 <- lm(data$femjobs ~ data$relig + data$gdp_k)
heterosked2 <- vcovHC(m2, type = 'HC1')

m3 <- lm(data$femjobs ~ data$relig + data$gdp_k + data$univ)
heterosked3 <- vcovHC(m3, type = 'HC1')
# m4 <- lm(data$femjobs ~ data$relig + data$gdp_k + data$univ + data$region)


stargazer(m1, m2, m3, type = 'text',
          omit.stat=c('f', 'rsq'),
          dep.var.caption = 'Models',
          dep.var.labels.include = F,
          digits = 2,
          df = F,
          #dep.var.labels = c("% saying religion \"very important\"",
                             # "GDP per capita ($1,000s)",
                             # '% with university degree',
                             # 'Intercept'),
          covariate.labels = c("% saying religion \"very important\"",
                               "GDP per capita ($1,000s)",
                               '% with university degree',
                               'Intercept'),
          se = list(sqrt(diag(heterosked1)),sqrt(diag(heterosked2)),
                    sqrt(diag(heterosked3))))
```

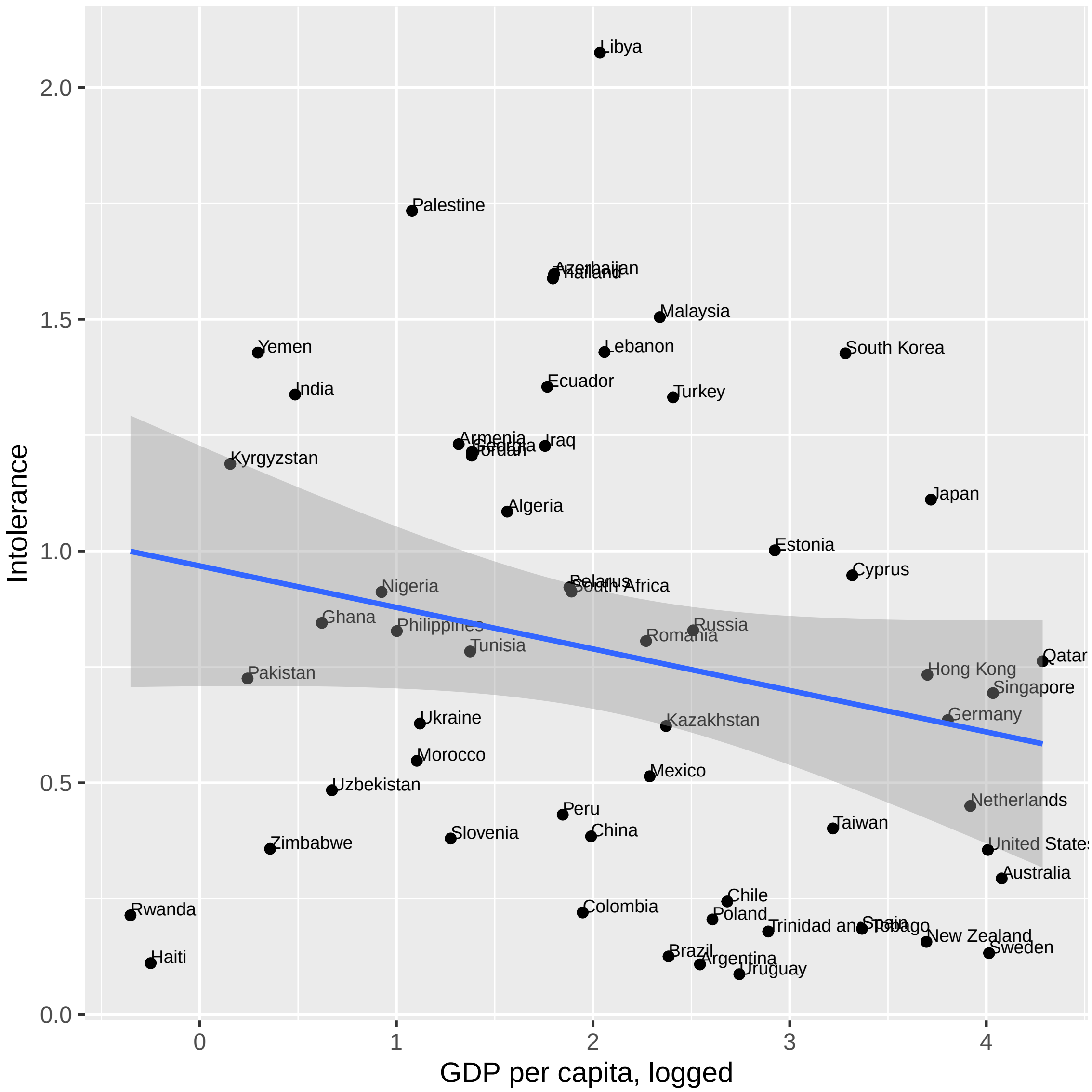
```
================================================================
                                             Models
                                   -----------------------------
                                      (1)       (2)       (3)
----------------------------------------------------------------
% saying religion "very important" -0.38***  -0.32***  -0.33***
                                    (0.07)    (0.08)    (0.07)
```

```
GDP per capita (1,000s)                          0.24*    0.34**
                                                (0.13)   (0.13)

% with university degree                                  -0.50**
                                                          (0.22)

Intercept                               60.76***  53.60***  60.83***
                                         (5.19)    (5.80)    (6.71)

-----------------------------------------------------------------
Observations                               60        60        60
Adjusted R2                               0.29      0.31      0.34
Residual Std. Error                       18.25     17.92     17.50
=================================================================
Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

Exercise 2

(a) Are rich countries generally more tolerant than poor ones?

```r
intolerance <- read_dta("/Users/herrhellana/Dropbox/_NYU studies/Quant I/home assignments/HW5/intoleran


library(ggplot2)
ggplot(data = intolerance, aes(x = log(gdpcap), y = n_sum,
                               label=countryname)) + geom_point() +
  geom_text(aes(label=countryname), size=2.5, hjust=0, vjust=0) +
  geom_smooth(method = 'lm') +
  labs(x = "GDP per capita, logged", y = "Intolerance")
```




```r
m_gdp <- lm(data=intolerance, n_sum ~ log(gdpcap))
heterosked_gdp <- vcovHC(m_gdp, type = 'HC1')


stargazer(m_gdp, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Intolerance',
          covariate.labels = c('GDP per capita', 'Intercept'),
          se = list(sqrt(diag(heterosked_gdp))))
```

```
===============================================
                             Outcome
                    ---------------------------
                           Intolerance
-----------------------------------------------
GDP per capita               -0.090*
                             (0.051)
```

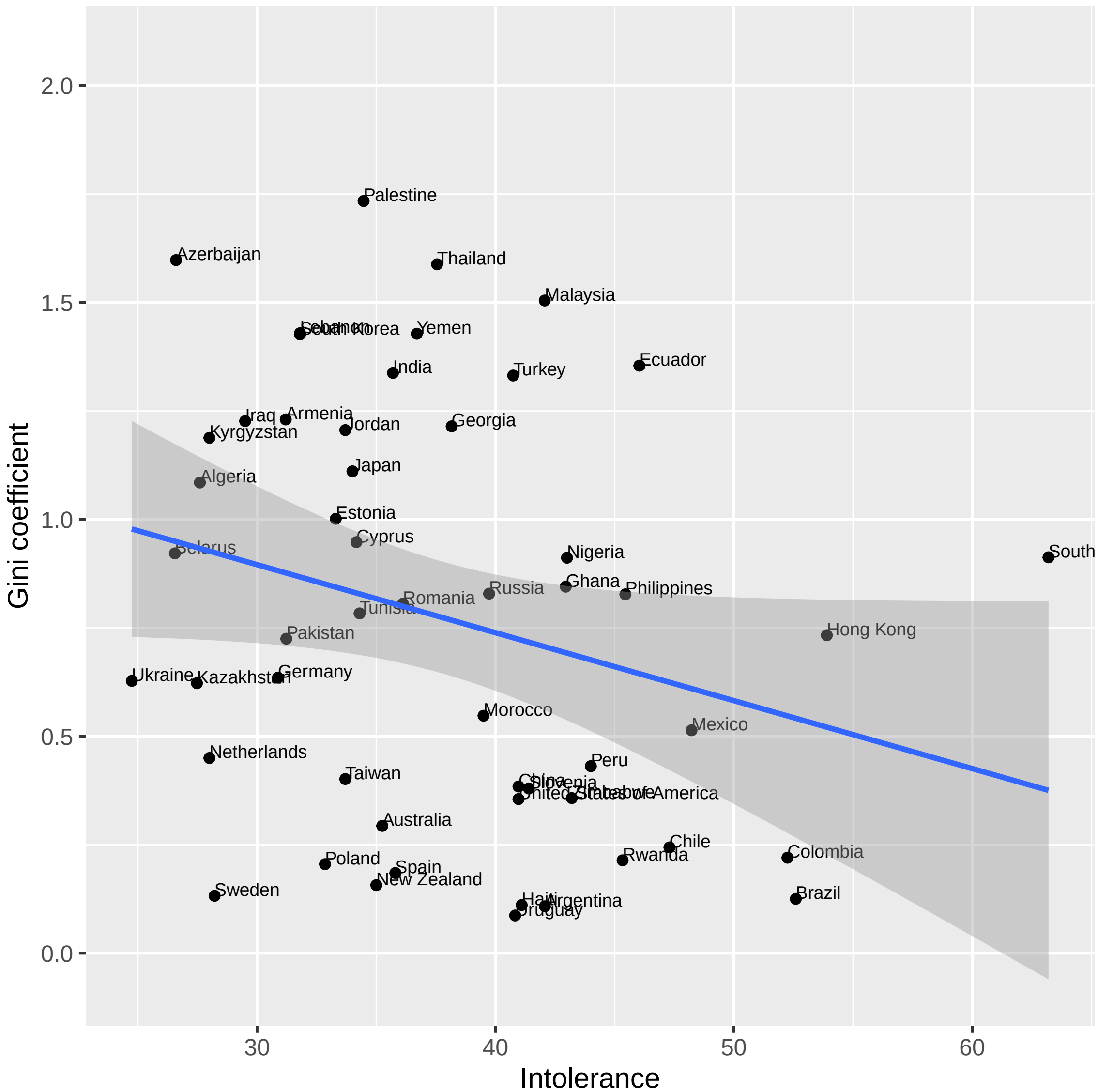
```
Intercept                        0.968***
                                  (0.136)

-----------------------------------------------
Observations                       58
R2                                0.048
Adjusted R2                       0.031
Residual Std. Error          0.489 (df = 56)
F Statistic                2.799* (df = 1; 56)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Rihcer countries are associated with lower rates of the intolerance index, this correlation is significant at the 90% confidence level but the effect of GDP per capita on intolerance is very small. The average effect of one-percent change in logged GDP per capita on intolerance is $-0.09 * 0.01 = -0.001$ (negative). The R^2 statistics also points out at the low explanatory power of the model. It is reasonable to assume that GDP per capita does not explain a lot of variation in the intolerance measure, and other factors play a more important role in predicting intolerance rates.

Some outliers: Libya, Palestine, Rwanda, Haiti, South Korea (a similar set to the one we found in the previous homework).

(b) Are unequal countries less tolerant than equal ones?

```
ggplot(data = intolerance, aes(x = gini, y = n_sum,
                               label=countryname)) + geom_point() +
  geom_text(aes(label=countryname), size=2.5, hjust=0, vjust=0) +
  geom_smooth(method = 'lm') +
  labs(x = "Intolerance", y = "Gini coefficient")
```




```r
m_gini <- lm(data=intolerance, n_sum ~ gini)
heterosked_gini <- vcovHC(m_gini, type = 'HC1')


stargazer(m_gini, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Intolerance',
          covariate.labels = c('Gini coefficient', 'Intercept'),
          se = list(sqrt(diag(heterosked_gini))))
```

```
===============================================
                          Outcome
                ---------------------------
                         Intolerance
-----------------------------------------------
Gini coefficient           -0.016**
                           (0.008)
```

```
Intercept                         1.366***
                                   (0.297)

------------------------------------------------
Observations                         53
R2                                  0.068
Adjusted R2                         0.049
Residual Std. Error           0.467 (df = 51)
F Statistic                 3.696* (df = 1; 51)
================================================
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

Higher levels of economic inequality are associated with lower rates of intolerance, this correlation is significant at the 95% confidence level. The average effect of one-unit change in the Gini coefficient on the intolerance rate is -0.016 (negative). However, the small R^2 statistics points out at the low explanatory power of the model. The Gini coefficient explains variation in intolerance better than GDP per capita but other impactful factors have to be accessed.

Some outliers: Palestine, Thailand, Malaysia, Ecuador, Azzerbaijan, Sweden (a similar set to the one we found in the previous homework).

(c) What is the relationship between inequality and intolerance, holding national wealth constant?

```r
m_econ1 <- lm(data=intolerance, n_sum ~ gini + log(gdpcap))
heterosked_econ1 <- vcovHC(m_econ1, type = 'HC1')


stargazer(m_econ1, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Intolerance',
          covariate.labels = c('Gini coefficient',
                               'logged GDPpc', 'Intercept'),
          se = list(sqrt(diag(heterosked_econ1))))
```

```
================================================
                                  Outcome
                          ----------------------------
                                Intolerance
------------------------------------------------
Gini coefficient                 -0.017**
                                  (0.008)

logged GDPpc                      -0.108*
                                  (0.056)

Intercept                        1.634***
                                  (0.310)

------------------------------------------------
Observations                         53
R2                                  0.138
Adjusted R2                         0.103
Residual Std. Error           0.454 (df = 50)
F Statistic                4.000** (df = 2; 50)
================================================
Note:                    *p<0.1; **p<0.05; ***p<0.01
```

The average effect of one-unit change in the Gini coefficient on the intolerance rate is -0.017 (negative), conditioning on other regressors set to zero. This relationship is significant at the 95% confidence level. The R^2 statistics point out at better explanatory power of this model in comparison to models estimated above.

(d) What is the relationship between inequality and intolerance, holding national wealth and economic growth constant?

```
m_econ2 <- lm(data=intolerance, n_sum ~ gini + log(gdpcap) + growth)
heterosked_econ2 <- vcovHC(m_econ2, type = 'HC1')


stargazer(m_econ2, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Intolerance',
          covariate.labels = c('Gini coefficient',
                               'logged GDPpc', 'Growth','Intercept'),
          se = list(sqrt(diag(heterosked_econ2))))
```

```
===============================================
                         Outcome
                  ---------------------------
                       Intolerance
-----------------------------------------------
Gini coefficient        -0.017**
                        (0.008)

logged GDPpc            -0.107*
                        (0.062)

Growth                   0.003
                        (0.035)

Intercept               1.627***
                        (0.332)

-----------------------------------------------
Observations              53
R2                       0.138
Adjusted R2              0.085
Residual Std. Error   0.458 (df = 49)
F Statistic         2.616* (df = 3; 49)
===============================================
Note:             *p<0.1; **p<0.05; ***p<0.01
```

The inequality coefficient value remains the same with additional regressors included in the model. The R^2 statistics is similar to the previous model's. The average effect of one-unit change in the Gini coefficient on the intolerance rate is -0.017 (negative), conditioning on other regressors set to zero. This relationship is significant at the 95% confidence level.

```
stargazer(m_gini, m_econ1, m_econ2, type = 'text',
          omit.stat=c('f', 'rsq'),
          dep.var.caption = 'Models',
          dep.var.labels.include = F,
          digits = 3,
          df = F,
          #dep.var.labels = c("% saying religion \"very important\"",
```

```
                           # "GDP per capita ($1,000s)",
                           # '% with university degree',
                           # 'Intercept'),
          covariate.labels = c("Gini coef",
                            "GDP per capita (logged)",
                            'Growth',
                            'Intercept'),
          se = list(sqrt(diag(heterosked_gini)),sqrt(diag(heterosked_econ1)),
                    sqrt(diag(heterosked_econ2))))
```

```
=====================================================
                                 Models
                        -----------------------------
                          (1)       (2)       (3)
-----------------------------------------------------
Gini coef               -0.016**  -0.017**  -0.017**
                         (0.008)   (0.008)   (0.008)

GDP per capita (logged)            -0.108*   -0.107*
                                   (0.056)   (0.062)

Growth                                        0.003
                                             (0.035)

Intercept               1.366***  1.634***  1.627***
                         (0.297)   (0.310)   (0.332)

-----------------------------------------------------
Observations               53        53        53
Adjusted R2              0.049     0.103     0.085
Residual Std. Error      0.467     0.454     0.458
=====================================================
Note:                     *p<0.1; **p<0.05; ***p<0.01
```

The models tell that higher rates of inequality are associated with lower rates of intolerance, the magnitude and sign of this relationship remain the same when including additional controls. So the effect of inequality on intolerance is robust.

Exercise 3

Do people with higher income vote more?

(a)

```
cps <- read_dta("/Users/herrhellana/Dropbox/_NYU studies/Quant I/home assignments/HW6/cpsnov2018abr.dta

library(dplyr)
cps <- cps %>% mutate(voted01 = case_when(voted == 2 ~ 1,
                                          voted == 1 ~ 0))

cps <- cps %>% mutate(faminc_new = case_when(faminc == 100 ~ 2500,
                                             faminc == 110 ~ 500,
```

```
                                  faminc == 111 ~ 250,
                                  faminc == 112 ~ 750,
                                  faminc == 120 ~ 1500,
                                  faminc == 121 ~ 1250,
                                  faminc == 122 ~ 1750,
                                  faminc == 130 ~ 2500,
                                  faminc == 131 ~ 2250,
                                  faminc == 132 ~ 2750,
                                  faminc == 140 ~ 3500,
                                  faminc == 141 ~ 3250,
                                  faminc == 142 ~ 3750,
                                  faminc == 150 ~ 4500,
                                  faminc == 200 ~ 6500,
                                  faminc == 210 ~ 6250,
                                  faminc == 220 ~ 5500,
                                  faminc == 230 ~ 7000,
                                  faminc == 231 ~ 6750,
                                  faminc == 232 ~ 6500,
                                  faminc == 233 ~ 7250,
                                  faminc == 234 ~ 7500,
                                  faminc == 300 ~ 8750,
                                  faminc == 310 ~ 7750,
                                  faminc == 320 ~ 8250,
                                  faminc == 330 ~ 8750,
                                  faminc == 340 ~ 8500,
                                  faminc == 350 ~ 9500,
                                  faminc == 400 ~ 12500,
                                  faminc == 410 ~ 10500,
                                  faminc == 420 ~ 11500,
                                  faminc == 430 ~ 11250,
                                  faminc == 440 ~ 11000,
                                  faminc == 450 ~ 12500,
                                  faminc == 460 ~ 13500,
                                  faminc == 470 ~ 13750,
                                  faminc == 480 ~ 13500,
                                  faminc == 490 ~ 14500,
                                  faminc == 500 ~ 17500,
                                  faminc == 510 ~ 15500,
                                  faminc == 520 ~ 16500,
                                  faminc == 530 ~ 17500,
                                  faminc == 540 ~ 16250,
                                  faminc == 550 ~ 18750,
                                  faminc == 560 ~ 19000,
                                  faminc == 600 ~ 22500,
                                  faminc == 700 ~ 37500,
                                  faminc == 710 ~ 27500,
                                  faminc == 720 ~ 32500,
                                  faminc == 730 ~ 37500,
                                  faminc == 740 ~ 45000,
                                  faminc == 800 ~ 75000,
                                  faminc == 810 ~ 62500,
                                  faminc == 820 ~ 55000,
                                  faminc == 830 ~ 67500,
```

```r
                                              faminc == 840 ~ 112500,
                                              faminc == 841 ~ 87500,
                                              faminc == 842 ~ 125000,
                                              faminc == 843 ~ 225000))


m_income <- lm(data=cps, voted01 ~ faminc_new)
heterosked_income <- vcovHC(m_income, type = 'HC1')

# Model I
stargazer(m_income, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Voted',
          covariate.labels = c('Income','Intercept'),
          se = list(sqrt(diag(heterosked_income))))
```

```
===============================================
                             Outcome
                    ---------------------------
                              Voted
-----------------------------------------------
Income                     0.00000***
                            (0.00000)

Intercept                   0.550***
                             (0.030)

-----------------------------------------------
Observations                   702
R2                            0.027
Adjusted R2                   0.026
Residual Std. Error      0.471 (df = 700)
F Statistic          19.393*** (df = 1; 700)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

The relationship between x and y is statistically significant at the 99% confidence level, the average effect of income on the voting variable is extremely small (approximates zero) since we operate with a non-transformed (not logged) income variable.

The relationship between x and y cannot be linear because of the diminishing marginal return: each additional increase in income makes less difference in the economic status of a person. Political and social behavior is less contrasted among the richest groups but the variety in behavior is more pronounced when we compare groups with lower wages.

Why education is a potential confounder: People with higher education levels are more likely to get higher salaries (each additional degree or educational effort signals about an employee's higher ability). More educated people are more likely to be politically active as they understand how and why to vote, or where to get information about candidates to make a more conscious voting decision.

Other potential confounders: - `age` can be a potential confounder – underage respondents are not eligible to vote, young people are more politically active than old people. The older a person get, the more working experience they get, hence the higher wage they can receive. However, this trend is generally linear until an old-age-threshold (when an employer does not want to hire an old person being skeptical of their working capabilities). - The citienship status, `citizen`, is also a confounder. Non-citizens cannot vote and they frequently get lower wages as the American market is economically protectionist.

- `nativity` is another potential confounder. Immigrants and/or their children have political behavior different from americans (they may care less about the U.S. politics or may not express the U.S. patriotic sentiment; or, otherwise, can voice their grievances more fiercely as a discriminated group). The same problem with the lower wages for immigrants and non-citizens persists.

These variables confound the relationship between income and voting behavior, making it spurious, and, if not included in the model, the variables lead to the omitted variable bias.

(b)

```
m_income_log <- lm(data=cps, voted01 ~ log(faminc_new))
heterosked_income_log <- vcovHC(m_income_log, type = 'HC1')

# Model II
stargazer(m_income_log, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Voted',
          covariate.labels = c('Income (log)','Intercept'),
          se = list(sqrt(diag(heterosked_income_log))))
```

```
===============================================
                          Outcome
                    ---------------------------
                           Voted
-----------------------------------------------
Income (log)              0.096***
                          (0.019)

Intercept                 -0.409*
                          (0.213)

-----------------------------------------------
Observations                702
R2                         0.036
Adjusted R2                0.035
Residual Std. Error   0.469 (df = 700)
F Statistic         26.194*** (df = 1; 700)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

A 1% change in logged income is associated with a positive change of $0.01 \cdot 0.096 = 0.00096$ in the voted variable. This relationship is significant at the 99% confidence level.

Models differ from each other in the magnitude of the coefficient. The R^2 statistics grew a bit after the log transformation – Model II fits the data better.

(c) I am recoding the `educ` variable as an ordinal discrete variable `educ_new` with six categories: 0 – no schooling, 1 – no more than high school, 2 – some undergrad (no more than 2 years), 3 – some additional schooling (Associate's degrees), 4 – MA and professional degrees, 5 – doctorate degree.

```
cps <- cps %>% mutate(educ_new = case_when(educ <= 2 ~ 0, # no degree
                                           educ <= 73 ~ 1, # school only
                                           educ <= 90 ~ 2, # some undergrad
                                           educ <= 92 ~ 3, # additional
                                           educ <= 122 ~ 4, # bachelor
                                           educ <= 124 ~ 5, # MA and prof
                                           educ == 125 ~ 5)) # PhD
```

```r
m_income_educ <- lm(data=cps, voted01 ~ log(faminc_new) + educ_new)
heterosked_income_educ <- vcovHC(m_income_educ, type = 'HC1')

# Model III
stargazer(m_income_educ, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Voted',
          covariate.labels = c('Income (log)', 'Education', 'Intercept'),
          se = list(sqrt(diag(heterosked_income_educ))))
```

```
================================================
                             Outcome
                    ----------------------------
                              Voted
------------------------------------------------
Income (log)                0.063***
                             (0.021)

Education                   0.051***
                             (0.013)

Intercept                    -0.176
                             (0.218)

------------------------------------------------
Observations                   702
R2                            0.056
Adjusted R2                   0.054
Residual Std. Error      0.464 (df = 699)
F Statistic           20.872*** (df = 2; 699)
================================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

Education confounds the income-voting relationship because (a) it is statistically significantly correlated with the voting variable and (b) it changes the magnitude of the income coefficient (from 0.063 to 0.096). The R^2 statistics grew in comparison with Model II – this model fits the data better.

(d) Model III but with other confounders (nativity and age).

```r
# recode nativity
cps <- cps %>% mutate(nativity_new = case_when(nativity >= 4 ~ 0, # non-native
                                               nativity >= 2 ~ 1, # partly
                                               nativity == 1 ~ 2)) # native



m_final <- lm(data=cps, voted01 ~ log(faminc_new) +
                educ_new + nativity_new + age)

heterosked_final <- vcovHC(m_final, type = 'HC1')

# Model III
stargazer(m_final, type = 'text', dep.var.caption = 'Outcome',
          dep.var.labels = 'Voted',
          covariate.labels = c('Income (log)', 'Education',
```

```r
                                 'Nativity', 'Age', 'Intercept'),
          se = list(sqrt(diag(heterosked_final))))
```

```
===============================================
                         Outcome
                    ---------------------------
                          Voted
-----------------------------------------------
Income (log)              0.073***
                           (0.021)

Education                 0.049***
                           (0.013)

Nativity                  0.096***
                           (0.027)

Age                       0.006***
                           (0.001)

Intercept                 -0.749***
                           (0.229)

-----------------------------------------------
Observations                 702
R2                          0.128
Adjusted R2                 0.123
Residual Std. Error      0.447 (df = 697)
F Statistic           25.500*** (df = 4; 697)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

We see that both `age` and `nativity_new` are confounders: they have a statistically significant correlation with the voting variable and they change the magnitude of the income coefficient. The R^2 is the highest among all estimated models implying that Model IV has the highest explanatory power.

All models together:

```r
stargazer(m_income, m_income_log, m_income_educ, m_final, type = 'text',
          omit.stat=c('f', 'rsq'),
          dep.var.caption = 'Models',
          dep.var.labels.include = F,
          digits = 3,
          df = F,
          #dep.var.labels = c('Voted'),
          covariate.labels = c('Income', 'Income (log)',
                               'Education', 'Nativity',
                               'Age', 'Intercept'),
          se = list(sqrt(diag(heterosked_income)),
                    sqrt(diag(heterosked_income_log)),
                    sqrt(diag(heterosked_income_educ)),
                    sqrt(diag(heterosked_final))))
```

	Models			
	(1)	(2)	(3)	(4)
Income	0.00000***			
	(0.00000)			
Income (log)		0.096***	0.063***	0.073***
		(0.019)	(0.021)	(0.021)
Education			0.051***	0.049***
			(0.013)	(0.013)
Nativity				0.096***
				(0.027)
Age				0.006***
				(0.001)
Intercept	0.550***	-0.409*	-0.176	-0.749***
	(0.030)	(0.213)	(0.218)	(0.229)
Observations	702	702	702	702
Adjusted R2	0.026	0.035	0.054	0.123
Residual Std. Error	0.471	0.469	0.464	0.447
Note:		*p<0.1; **p<0.05; ***p<0.01		

Overall, income and voting are positively correlated (this relationship is statistically significant and robust when including other economic variables). However, we should access potential confounders and include it to the model to eliminate the bias of the effect of income on voting. This data showed that a set of such confounders comprise education level, nativity, and age of a respondent.