Assignment 4 - Econometrics II

Floris Holstege, Stanislav Avdeev

```
## [1] 0.2666667
```

Problem 1

We use the allocation to the two judges as a instrument. This instrument is relevant, since one judge (Jones) will sentence more than the other (Smith). But it is also exogenous, since the defendants are randomly allocated to a judge.

I) In order to assess the effect of a prison sentence of future arrests, we use the Wald-estimator, defined as follows:

$$\beta_{wald} = \frac{P(Y = 1|Z = 1) - P(Y = 1|Z = 0)}{P(D = 1|Z = 1) = P(D = 1|Z = 0)}$$

Where $P(Y = 1)$ is the probability of an arrest in the subsequent 3 years, Z is our instrumental variable, with $Z = 1$ being judged by Jones, and $Z = 0$ being judged by Smith. $P(D = 1)$ is the probability of an prison sentence.

To estimate the probabilities, we simply use the % of each event occuring in our sample. The Wald estimator then becomes:

$$\beta_{wald} = \frac{0.46 - 0.38}{0.7 - 0.4} = \frac{0.08}{0.3} = 0.26\frac{2}{3}$$

II) The result implies that if one receives a prison sentence, one is $26\frac{2}{3}$% more likely to be arrested in the subsequent 3 years. This effect holds for all compliers.

Since we are working with an instrument, the definition of compliers and always takers changes. Instead of related to the prison treatment, it is not related to the instrument that we use to estimate the actual effect. That is because non-compliance is now only relevant insofar as it effects the observed outcome from this instrument. The definition here is therefore:

$$Z(1) = 1, Z(0) = 0$$

In this case, those are people who under Jones went to prison, but would not under smith, and those that under Smith would go to prison, but not under Jones. $(0.7 \cdot 0.6) + (0.4 \cdot 0.3) = 0.54$.

III) The definition of always takers is in this case

$$Z(1) = Z(0) = 1$$

More specifically, an always taker is someone who is sent to prison by both judges. This fraction equals $0.7 \cdot 0.4 = 0.28$.

```
get_size_givenMDE <- function(MDE, t_Alpha, t_Power, p, Sigma2){

  # get the MDE
  size <- (((t_Alpha - t_Power)/MDE)^2) * sigma2/(p*(1-p))
  size <- round(size, 0)

  return(size)
}


MDE = 0.1
t_Alpha = 1.96
t_Power = -0.524
p = 0.5
sigma2 <- p* (1- p)

size <- get_size_givenMDE(0.1, t_Alpha, t_Power, p, sigma2)
size
```

```
## [1] 617
```

```
perc_nonComply <- 0.2

new_size = (1/(1-perc_nonComply)^2)*size
new_size
```

```
## [1] 964.0625
```

Problem 2

$$n = (\frac{(1.96 + 0.524)}{MDE})^2 * \frac{\sigma^2}{p \cdot (1-p)}$$

I) 617
II) 965

```
# get data on flu shots, divide in treatment and control
dfFlu <- import("FluData.dta")
dfFlu_treatment <- dfFlu[dfFlu$TreatGroup == 1,]
dfFlu_control <- dfFlu[dfFlu$TreatGroup == 0,]

# percentage that got the flu in the treatment group (got flu shot)
p_flu <- 0.8
sigma2_flu <- p_flu*(1-p_flu)

t_Alpha_flu = 1.96
t_Power_flu = -0.84

# size of the experiment, given p and sigma2
size_flu <- get_size_givenMDE(0.05, t_Alpha_flu, t_Power_flu, p_flu, sigma2_flu)
size_flu
```

```
## [1] 4900
```

```r
# get df with group that actually got the treatment
dfFlu_actualTreatment <- dfFlu_treatment[dfFlu_treatment$Treatment == 1,]
perc_comply_flu <- nrow(dfFlu_actualTreatment)/nrow(dfFlu_treatment)

# get new size given the rate that did not get treatment
new_size_flu = (1/(perc_comply_flu)^2)*size_flu
new_size_flu
```

```
## [1] 10982.51
```

```r
type_group <- ifelse(dfFlu$TreatGroup ==1 & dfFlu$Treatment == 0, "Untreated treatment",ifelse(dfFlu$Tre
dfFlu$type_group <- type_group

dfSummary <- dfFlu %>%
  group_by(type_group) %>%
  summarise(perc_boy = sum(GenderChild)/n(),
            mean_age_mother = mean(AgeMother),
            mean_edu_mother = mean(EducationMother),
            perc_married = sum(Married)/n(),
            perc_nationality = sum(Nationality)/n(),
            mean_HHincome = mean(Hhincome),
            perc_group = n()/nrow(dfFlu)
            )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
xtable(dfSummary)
```

```
## % latex table generated in R 3.6.3 by xtable 1.8-4 package
## % Fri Jan 29 12:26:37 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{rlrrrrrrr}
##   \hline
##  & type\_group & perc\_boy & mean\_age\_mother & mean\_edu\_mother & perc\_married & perc\_nationali
##   \hline
## 1 & control & 0.51 & 26.09 & 12.34 & 0.96 & 0.28 & 2269.88 & 0.20 \\
##   2 & Treated treatment & 0.50 & 26.59 & 12.52 & 0.98 & 0.24 & 2373.87 & 0.54 \\
##   3 & Untreated treatment & 0.50 & 24.88 & 11.83 & 0.94 & 0.34 & 2110.71 & 0.27 \\
##    \hline
## \end{tabular}
## \end{table}
```

```r
# define the two models
model_ols_simple <- lm(Flu ~ Treatment, data=dfFlu_treatment)
model_ols_extensive <- lm(Flu ~ Treatment + GenderChild + AgeMother + EducationMother + Married + Natio

# check the output
summary(model_ols_simple)
```

```
##
## Call:
## lm(formula = Flu ~ Treatment, data = dfFlu_treatment)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.6749 -0.4009 -0.4009  0.5990  0.5990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.674925   0.008346   80.87   <2e-16 ***
## Treatment   -0.273976   0.010212  -26.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.483 on 10087 degrees of freedom
## Multiple R-squared:  0.06661,    Adjusted R-squared:  0.06652
## F-statistic: 719.8 on 1 and 10087 DF,  p-value: < 2.2e-16
```

```
summary(model_ols_extensive)
```

```
##
## Call:
## lm(formula = Flu ~ Treatment + GenderChild + AgeMother + EducationMother +
##     Married + Nationality + Hhincome, data = dfFlu_treatment)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.99840 -0.40130 -0.04587  0.41294  1.10535
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.185e+00  4.456e-02  49.025   <2e-16 ***
## Treatment       -1.630e-01  9.947e-03 -16.387   <2e-16 ***
## GenderChild      1.592e-02  8.962e-03   1.777   0.0757 .
## AgeMother       -4.751e-02  1.799e-03 -26.411   <2e-16 ***
## EducationMother -2.994e-02  3.118e-03  -9.602   <2e-16 ***
## Married         -2.477e-02  2.467e-02  -1.004   0.3155
## Nationality      9.160e-02  1.014e-02   9.035   <2e-16 ***
## Hhincome         5.065e-06  4.688e-06   1.080   0.2800
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.45 on 10081 degrees of freedom
## Multiple R-squared:  0.1903, Adjusted R-squared:  0.1897
## F-statistic: 338.4 on 7 and 10081 DF,  p-value: < 2.2e-16
```

```
calc_robust_se <- function(model){

  cov_model <- vcovHC(model, type = "HC1")
  robust_se <- sqrt(diag(cov_model))
  return(robust_se)
}


# Adjust standard errors
robust_se_simple    <- calc_robust_se(model_ols_simple)
robust_se_extensive <- calc_robust_se(model_ols_extensive)

# create stargazer output
stargazer(model_ols_simple, model_ols_extensive, se = list(robust_se_simple, robust_se_extensive))
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 29, 2021 - 12:26:37
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{\textit{Dependent variable:}} \\
## \cline{2-3}
## \\[-1.8ex] & \multicolumn{2}{c}{Flu} \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment & $-$0.274$^{***}$ & $-$0.163$^{***}$ \\
##   & (0.010) & (0.010) \\
##   & & \\
##  GenderChild &  & 0.016$^{*}$ \\
##   &  & (0.009) \\
##   & & \\
##  AgeMother &  & $-$0.048$^{***}$ \\
##   &  & (0.002) \\
##   & & \\
##  EducationMother &  & $-$0.030$^{***}$ \\
##   &  & (0.003) \\
##   & & \\
##  Married &  & $-$0.025 \\
##   &  & (0.022) \\
##   & & \\
##  Nationality &  & 0.092$^{***}$ \\
##   &  & (0.010) \\
##   & & \\
##  Hhincome &  & 0.00001 \\
##   &  & (0.00000) \\
##   & & \\
##  Constant & 0.675$^{***}$ & 2.185$^{***}$ \\
##   & (0.008) & (0.038) \\
##   & & \\
## \hline \\[-1.8ex]
## Observations & 10,089 & 10,089 \\
## R$^{2}$ & 0.067 & 0.190 \\
## Adjusted R$^{2}$ & 0.067 & 0.190 \\
## Residual Std. Error & 0.483 (df = 10087) & 0.450 (df = 10081) \\
## F Statistic & 719.847$^{***}$ (df = 1; 10087) & 338.396$^{***}$ (df = 7; 10081) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```r
# run both 2sls regressions - simple and extensive model
model_ols_simple_iv <- ivreg(Flu ~ Treatment | TreatGroup, data=dfFlu)
model_ols_extensive_iv <- ivreg(Flu ~ Treatment + GenderChild + AgeMother + EducationMother + Married + 
                                | TreatGroup + GenderChild + AgeMother + EducationMother + Married + Na
```

```r
summary(model_ols_simple_iv)
```

```
##
## Call:
## ivreg(formula = Flu ~ Treatment | TreatGroup, data = dfFlu)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6207 -0.4279  0.3793  0.3793  0.5721
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.620690   0.009705   63.95   <2e-16 ***
## Treatment   -0.192779   0.016227  -11.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4847 on 12581 degrees of freedom
## Multiple R-Squared: 0.05932, Adjusted R-squared: 0.05924
## Wald test: 141.1 on 1 and 12581 DF,  p-value: < 2.2e-16
```

```r
summary(model_ols_extensive_iv)
```

```
##
## Call:
## ivreg(formula = Flu ~ Treatment + GenderChild + AgeMother + EducationMother +
##     Married + Nationality + Hhincome | TreatGroup + GenderChild +
##     AgeMother + EducationMother + Married + Nationality + Hhincome,
##     data = dfFlu)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00902 -0.40838  0.03057  0.40263  1.09651
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.153e+00  4.011e-02  53.675   <2e-16 ***
## Treatment       -1.983e-01  1.509e-02 -13.143   <2e-16 ***
## GenderChild      1.343e-02  8.048e-03   1.668   0.0953 .
## AgeMother       -4.628e-02  1.633e-03 -28.344   <2e-16 ***
## EducationMother -2.732e-02  2.789e-03  -9.795   <2e-16 ***
## Married         -2.848e-02  2.182e-02  -1.305   0.1918
## Nationality      9.024e-02  9.134e-03   9.880   <2e-16 ***
## Hhincome         3.569e-06  4.225e-06   0.845   0.3982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4513 on 12575 degrees of freedom
## Multiple R-Squared: 0.1847,  Adjusted R-squared: 0.1842
## Wald test: 369.2 on 7 and 12575 DF,  p-value: < 2.2e-16
```

```r
# Adjust standard errors
robust_se_simple_iv    <-  calc_robust_se(model_ols_simple_iv)
robust_se_extensive_iv <-  calc_robust_se(model_ols_extensive_iv)
```

```r
# create stargazer output
stargazer(model_ols_simple_iv, model_ols_extensive_iv, se = list(robust_se_simple_iv, robust_se_extensi
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 29, 2021 - 12:26:45
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{\textit{Dependent variable:}} \\
## \cline{2-3}
## \\[-1.8ex] & \multicolumn{2}{c}{Flu} \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment & $-$0.193$^{***}$ & $-$0.198$^{***}$ \\
##   & (0.016) & (0.015) \\
##   & & \\
##  GenderChild &  & 0.013$^{*}$ \\
##   &  & (0.008) \\
##   & & \\
##  AgeMother &  & $-$0.046$^{***}$ \\
##   &  & (0.002) \\
##   & & \\
##  EducationMother &  & $-$0.027$^{***}$ \\
##   &  & (0.003) \\
##   & & \\
##  Married &  & $-$0.028 \\
##   &  & (0.019) \\
##   & & \\
##  Nationality &  & 0.090$^{***}$ \\
##   &  & (0.009) \\
##   & & \\
##  Hhincome &  & 0.00000 \\
##   &  & (0.00000) \\
##   & & \\
##  Constant & 0.621$^{***}$ & 2.153$^{***}$ \\
##   & (0.010) & (0.034) \\
##   & & \\
## \hline \\[-1.8ex]
## Observations & 12,583 & 12,583 \\
## R$^{2}$ & 0.059 & 0.185 \\
## Adjusted R$^{2}$ & 0.059 & 0.184 \\
## Residual Std. Error & 0.485 (df = 12581) & 0.451 (df = 12575) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```r
# run partial regression
## not affraid of irrelevant instrument, since 67% overlap...
```

```r
Partial_ols_FluShot <- lm(Treatment ~ TreatGroup, data=dfFlu)
summary(Partial_ols_FluShot)
```

```
##
## Call:
## lm(formula = Treatment ~ TreatGroup, data = dfFlu)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.668 -0.668  0.332  0.332  0.332
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.380e-14  8.445e-03    0.00        1
## TreatGroup  6.680e-01  9.431e-03   70.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4217 on 12581 degrees of freedom
## Multiple R-squared:  0.2851, Adjusted R-squared:  0.285
## F-statistic:  5016 on 1 and 12581 DF,  p-value: < 2.2e-16
```