CA3aVdVeen

Ties van der Veen

12-9-2019

Rebecca Costello - Ties van der Veen - Niels van Opstal

I. Introduction to the assignment

Packages

```
options(repos="https://cran.rstudio.com")
install.packages("jtools")
```

```
## package 'jtools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```
install.packages("huxtable")
```

```
## package 'huxtable' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```
install.packages("ggstance")
```

```
## package 'ggstance' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```
install.packages("summarytools")
```

```
## package 'summarytools' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```
install.packages("pwr")
```

```
## package 'pwr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```r
install.packages("knitr")
```

```
## package 'knitr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```r
install.packages("lemon")
```

```
## package 'lemon' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```r
install.packages("tinytex")
```

```
## package 'tinytex' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```r
install.packages("AER")
```

```
## package 'AER' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tiess\AppData\Local\Temp\RtmpqqKHD0\downloaded_packages
```

```r
library(foreign)
library(tidyverse)
library(ggdag)
library(dplyr)
library(tinytex)
library(jtools)
library(huxtable)
library(summarytools)
library(ggstance)
library(pwr)
library(knitr)
library(lemon)
library(AER)
```

```r
knit_print.data.frame <- lemon_print
```

```r
st_options(plain.ascii = FALSE, style = "rmarkdown")
st_css()
```

```
## <style type="text/css">
##  img {   background-color: transparent;   border: 0; }  .st-table td, .st-table th {   padding: 8px;
```

Data

```
theUrl_ca3a <- "https://surfdrive.surf.nl/files/index.php/s/aRyyi4GbwUjF9DF/download"
colonies <- read.dta (file = theUrl_ca3a)
```

```
colonies <- colonies[colonies$excolony==1, ]
colonies <- colonies[is.na(colonies$excolony)==FALSE, ]
```

II. Potential outcomes

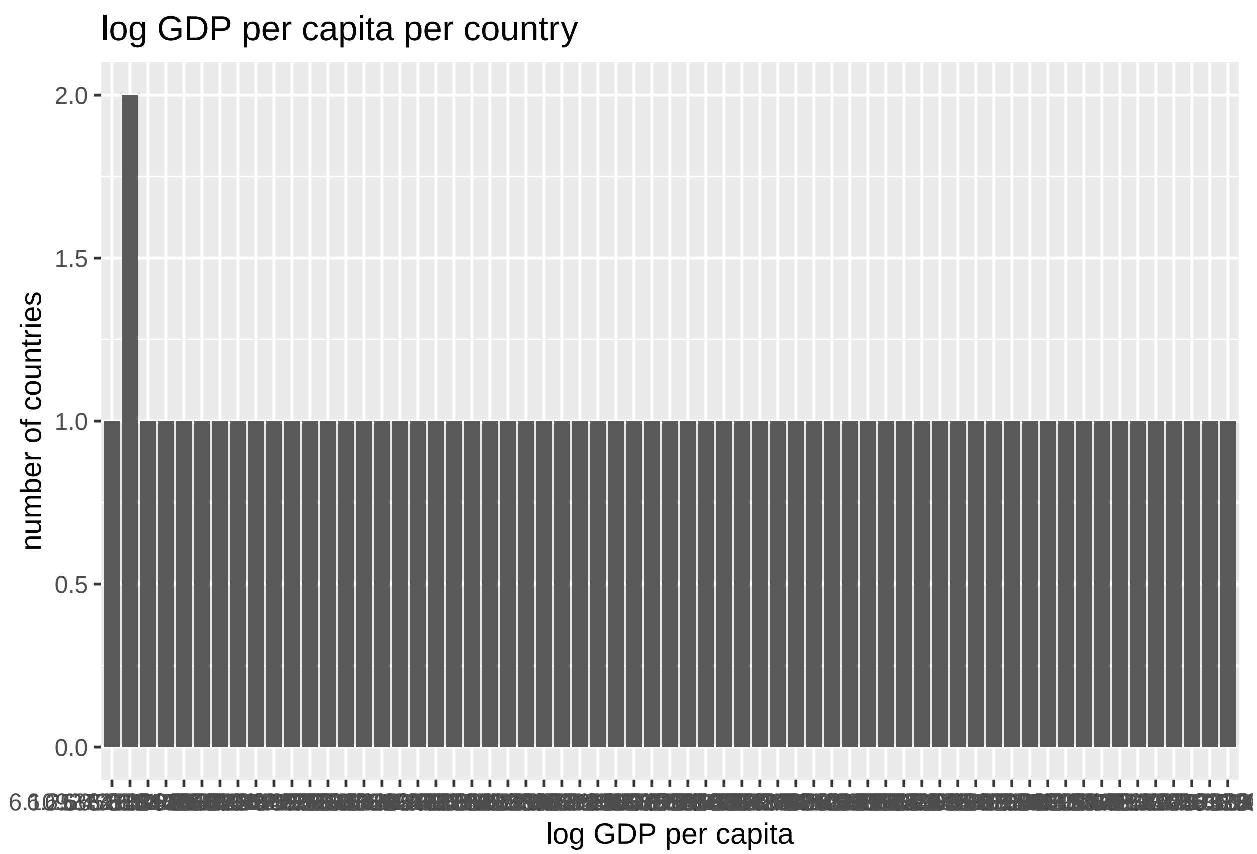
(a) Y(0,i) = low GDP per capita Y(1,i) = high GDP per capita

(b) If colonists came to settle, they brought in institutions that helped the country flourish. If they did the opposite and just extracted resources, these institutions were not established and the country had less of an opportunity to develop. These development differences explain current GDP per capita.

III. Descriptive statistics

```
ggplot(colonies, aes(x=as.factor(logpgp95)))+
  geom_histogram(stat='count', binwidth = "0.5")+
  labs(x='log GDP per capita', y='number of countries', title='log GDP per capita per country')
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```r
summary(is.na(colonies$logpgp95))
```

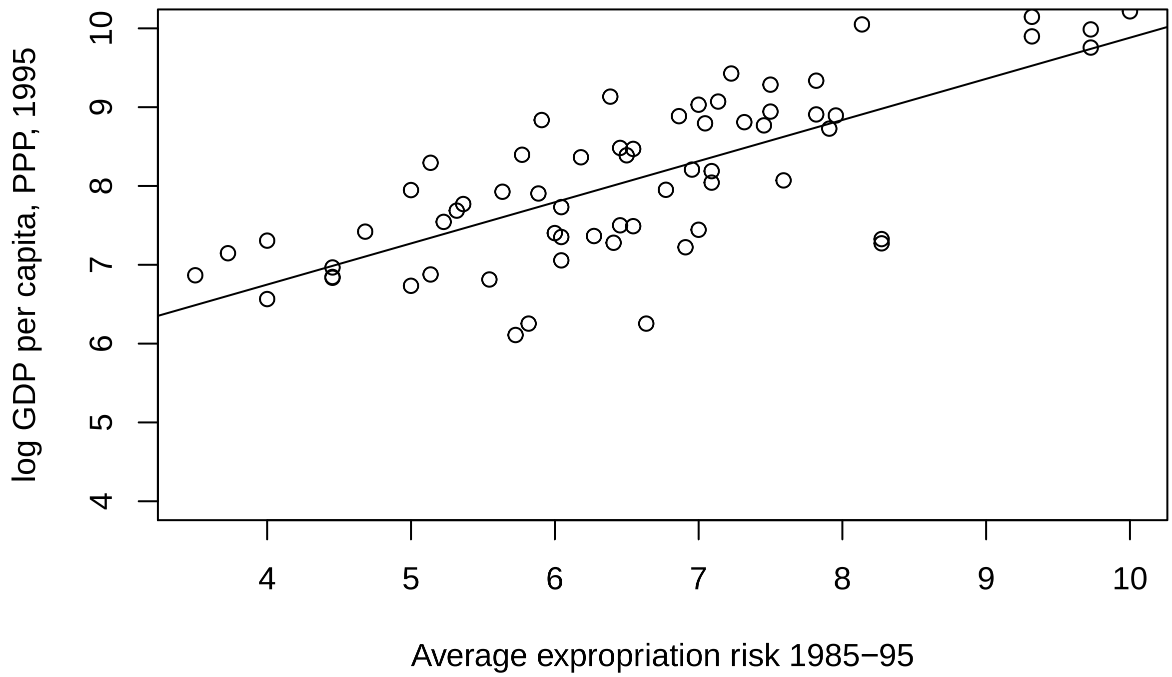
```
##    Mode   FALSE
## logical      64
```

No output, so no missing entries.

IV. Estimating the treatment effect using an instrumental variable

(a)

```r
plot(colonies$avexpr, colonies$logpgp95,
     xlab="Average expropriation risk 1985-95", ylab="log GDP per capita, PPP, 1995",
     ylim = c(4, 10))
abline(lm(colonies$logpgp95 ~ colonies$avexpr))
```



(b)

```r
reg3a1 <- lm(logpgp95 ~ avexpr, data=colonies)
summ(reg3a1, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 64
## Dependent Variable: logpgp95
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,62) = 72.82, p = 0.00
## R² = 0.54
## Adj. R² = 0.53
##
## Standard errors: OLS
## --------------------------------------------------------
##                      Est.   2.5%   97.5%   t val.      p
## ----------------- ------ ------ ------- -------- ------
## (Intercept)          4.66   3.84    5.48    11.41   0.00
## avexpr               0.52   0.40    0.64     8.53   0.00
## --------------------------------------------------------
```

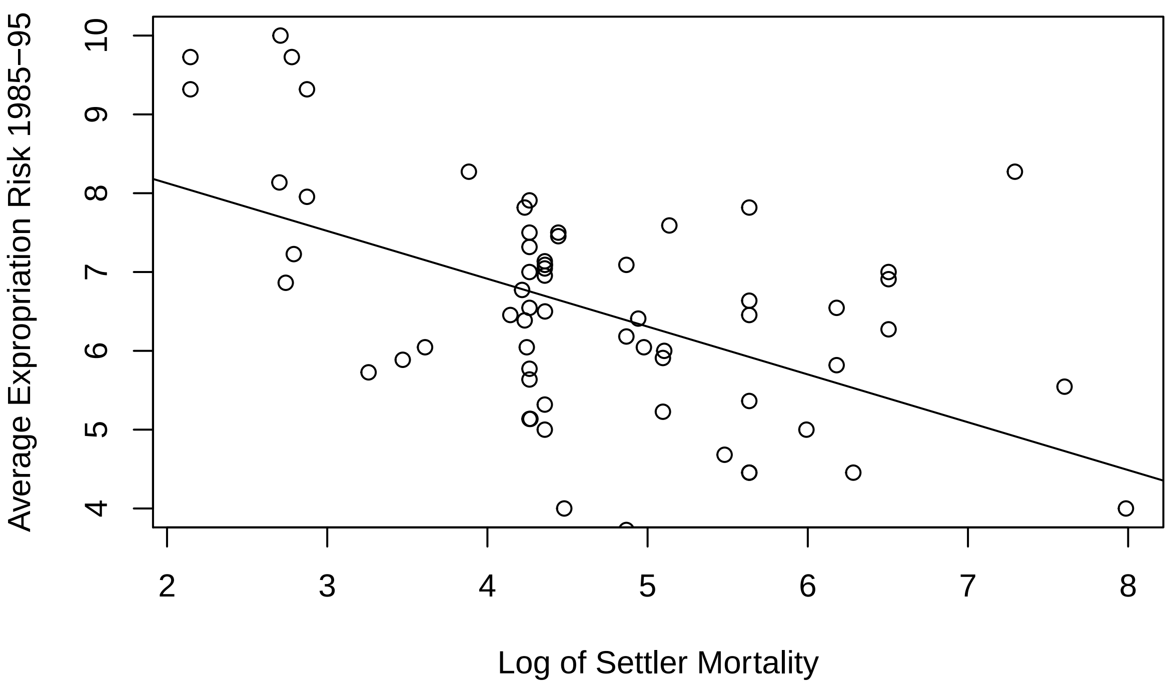
The regression results suggest a higher GDP per capita for countries with good institutions compared to the countries that don't have good institutions.

(c)

There are multiple confounding factors that influence this result. Seeing this as a causal relationship would be shortsighted, especially since there's such a gap between the assumed variable's origin (strategy of colonization) and "current' (1995) GDP per capita.

(d)

```r
plot(colonies$logem4, colonies$avexpr,
     xlab="Log of Settler Mortality", ylab="Average Expropriation Risk 1985-95",
     ylim = c(4, 10))
abline(lm(colonies$avexpr ~ colonies$logem4))
```



(e)

```r
reg3a2 <- lm(logpgp95 ~ logem4, data=colonies)
summ(reg3a2, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 64
## Dependent Variable: logpgp95
## Type: OLS linear regression
##
## MODEL FIT:
```

```
## F(1,62) = 56.60, p = 0.00
## R² = 0.48
## Adj. R² = 0.47
##
## Standard errors: OLS
## -----------------------------------------------------------
##                       Est.    2.5%   97.5%    t val.      p
## ----------------- -------- -------- -------- --------- ------
## (Intercept)          10.73   10.00   11.46     29.22   0.00
## logem4               -0.57   -0.73   -0.42     -7.52   0.00
## -----------------------------------------------------------
```

The instrumental variable is statistically significant (p=0.00) as can be seen in the regression output. Economically this variable is not that significant, since there is no causal relationship between mortality in the 1800s and current economic welfare. However, this does not mean that the variable is useless; far from it. It does its job as an instrumental variable.

(f)

```r
reg3a3 <- lm(avexpr ~ logem4, data=colonies)
summ(reg3a3, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 64
## Dependent Variable: avexpr
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,62) = 22.95, p = 0.00
## R² = 0.27
## Adj. R² = 0.26
##
## Standard errors: OLS
## -----------------------------------------------------------
##                       Est.    2.5%   97.5%    t val.      p
## ----------------- -------- -------- -------- --------- ------
## (Intercept)           9.34    8.12   10.56     15.30   0.00
## logem4               -0.61   -0.86   -0.35     -4.79   0.00
## -----------------------------------------------------------
```

These results show that mortality rate has a sizable effect on settlement strategy, which is what was already predicted. Thus, with the significance reported, these results support the theory.

The F-test reported here gives F=22.95, which is big enough (considering the minimum of 10).

(g)

```r
avexpr_hat <- predict(reg3a3)
colonies$pred_avexpr=avexpr_hat
```

```r
reg3a4 <- lm(logpgp95 ~ pred_avexpr, data=colonies)
summ(reg3a4, confint=TRUE)
```

```
## MODEL INFO:
## Observations: 64
## Dependent Variable: logpgp95
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,62) = 56.60, p = 0.00
## R² = 0.48
## Adj. R² = 0.47
##
## Standard errors: OLS
## ------------------------------------------------------
##                     Est.   2.5%   97.5%   t val.      p
## ----------------- ------ ------ ------- -------- ------
## (Intercept)         1.91   0.26    3.56     2.32   0.02
## pred_avexpr         0.94   0.69    1.20     7.52   0.00
## ------------------------------------------------------
```

By taking the prediction of the colonization strategy (avexpr), it seems like almost all of the variation in log GDP per capita (logpgp95) is explained. From (b), which showed an estimate of 0.52, this is a big jump.

(h)

```r
reg3a5 <- ivreg(logpgp95 ~ avexpr | logem4, data=colonies)
summary(reg3a5, confint=TRUE, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = logpgp95 ~ avexpr | logem4, data = colonies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.44903 -0.56242  0.07311  0.69564  1.71752
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.9097     1.0267   1.860   0.0676 .
## avexpr        0.9443     0.1565   6.033  9.8e-08 ***
##
## Diagnostic tests:
##                  df1 df2 statistic  p-value
## Weak instruments   1  62     22.95 1.08e-05 ***
## Wu-Hausman         1  61     24.22 6.85e-06 ***
## Sargan             0  NA        NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9483 on 62 degrees of freedom
## Multiple R-Squared: 0.187,   Adjusted R-squared: 0.1739
## Wald test: 36.39 on 1 and 62 DF,  p-value: 9.799e-08
```

(i)

```r
reg3a6 <- ivreg(logpgp95 ~ avexpr + lat_abst | lat_abst + logem4, data=colonies)
summary(reg3a6, confint=TRUE, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = logpgp95 ~ avexpr + lat_abst | lat_abst + logem4,
##     data = colonies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5611 -0.6557  0.0732  0.7572  1.8803
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6918     1.2930   1.308    0.196
## avexpr        0.9957     0.2217   4.492 3.21e-05 ***
## lat_abst     -0.6472     1.3351  -0.485    0.630
##
## Diagnostic tests:
##                  df1 df2 statistic  p-value
## Weak instruments   1  61     13.09 0.000603 ***
## Wu-Hausman         1  60     18.75 5.75e-05 ***
## Sargan             0  NA        NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.005 on 61 degrees of freedom
## Multiple R-Squared: 0.1025,  Adjusted R-squared: 0.07305
## Wald test: 17.01 on 2 and 61 DF,  p-value: 1.351e-06
```

Including this covariate allows us to control for environmental situations that would give a certain country an advantage (or disadvantage) over the other, creating differences in GDP growth. The new estimation results show that including this covariate makes the colonization strategy (avexpr) able to pretty much explain the whole effect on GDP per capita.

(j)

The exclusion restriction means that the independent variables do not directly impact the dependent variables. The instrumental variable is used when this exclusion restriction cannot be guaranteed through randomization. In this case, the instrumental variable "mortality rate of settlers in the 1800s" is unlikely to affect other variables in the study, but can be used to check the independence of the colonization strategy variable. Thus it is likely the effect will hold.