# Lab #12 - The Colonial Origins of Comparative Development

*Econ 224*

*October 18th, 2018*

## The Colonial Origins of Comparative Development

Today we'll work with a dataset from the 2001 paper "The Colonial Origins of Comparative Development" by Acemoglu, Johnson, and Robinson. To avoid repeatedly typing out all three names, I'll refer the paper and authors as AJR in the rest of this document. You can download this paper from the website of the *American Economic Review* or from Piazza, where I have posted a copy. The dataset is called `ajr.dta` and is available from the course website: http://ditraglia.com/econ224/ajr.dta. Note that this is file is in STATA format, so you'll need to convert to a format the R can understand before proceeding. Here is a description of the variables we'll use in this lab:

| Name | Description |
|---|---|
| longname | Full name of country, e.g. Canada |
| shortnam | Abbreviated country name, e.g. CAN |
| logmort0 | Natural log of early European settler mortality |
| risk | Avg. protection against expropriation risk 1985-1995 (0 to 10) |
| loggdp | Natural log of 1995 GDP/capita at purchasing power parity |
| latitude | Absolute value of latitude (scaled between 0 and 1) |
| meantemp | 1987 mean annual temperature in degrees Celsius |
| rainmin | Minimum monthly rainfall |
| malaria | % of Popn. living where falciporum malaria is endemic in 1994 |

The key variables in this analysis are `loggdp`, which is the *outcome* variable ($y$), `risk` which is the regressor of interest ($x$), and `logmort0`, which AJR propose as an *instrumental variable* ($z$) for `risk`. Both `loggdp` and `logmort0` are fairly self-explanatory, but `risk` is a bit strange. The *larger* the value of `risk`, the *more* protection a country has against expropriation. In other words, large values of `risk` indicate *better* institutions, as described in the first paragraph of AJR.

Note that for simplicity we will *not* consider heterogeneous treatment effects in this lab.

## Exercises

1. Read the abstract, introduction and conclusion of AJR and answer the following:
   (a) What is the key question that AJR try to answer?
   (b) Give an overview of AJR's key theory.
   (c) For $z$ to be a valid instrument, it must satisfy two assumptions. First it must be *relevant*: correlated with $x$. Second, it must be *excluded*: it must only be related to $y$ through its effect on $x$. (Exclusion is also called *validity* or *exogeneity* and in mathematical notation means that $z$ is uncorrelated with the error term in the structural equation). What do these assumptions mean in the context of AJR? Can either of them be checked using the available data?
2. OLS Regression:
   (a) Regress `loggdp` on `risk` and store the result in an object called `ols`.

(b) Display the results of part (a) in a cleanly formatted regression table, using appropriate R packages.
(c) Discuss your results from (b) in light of your readings from AJR. Can we interpret the results of `ols` causally? Why or why not?

3. IV Regression:

(a) Estimate the first-stage regression of `risk` on `logmort0` and store your results in an object called `first_stage`. Display and discuss your findings.
(b) Estimate the reduced-form regression of `loggdp` on `logmort0` and store your results in an object called `reduced_form`. Display and discuss your findings.
(c) Use the `ivreg` function from `AER` to carry out an IV regression of `loggdp` on `risk` using `logmort0` as an instrument for `risk` and store your results in an object called `iv`.
(d) Display your results from `iv`. How do they compare to the results of `ols`? Discuss in light of your answer to 2(c) above.
(e) Verify that you get the same estimate as in part (d) by running IV "by hand" using `first_stage` and `reduced_form`.

4. This question asks you consider a potential criticism of AJR. The critique depends on two claims. Claim #1: a country's current disease environment, e.g. the prevalence of malaria, is an important determinant of GDP/capita. Claim #2: a country's disease environment today depends on its disease environment in the past, which in turn affected early European settler mortality.

(a) Explain how claims #1 and #2 taken together would call into question the IV results from Question 3 above.
(b) Suppose that we consider re-running our IV analysis from Question 3 including `malaria` as an additional regressor. Explain why this might address the concerns you raised in the preceding part.
(c) Repeat Question 2 including `malaria` as an additional regressor.
(d) Repeat Question 3 part (a) adding `malaria` to the first-stage regression.
(e) Repeat Question 3 parts (c) and (d) including `malaria` in the IV regression. Treat `malaria` as *exogenous*. This means we will not need an instrument for this variable: instead it serves as its *own* instrument. See "Details" in the help file for `ivreg` to see how to specify this.
(f) In light of your results from this question, what do you make of the criticism of AJR based on a country's disease environment?

5. This question asks you to consider another potential criticism of AJR promoted by Jeffrey Sachs who stresses "geographical" explanations of economic development.

(a) Repeat part (e) from Question 4 but add `latitude`, `rainmin`, and `meantemp` as additional control regressors in addition to `malaria`. Each of these variables will serve as its own instrument. Continue to instrument `risk` using `logmort0`.
(b) Discuss your results. What do you make of AJR's view vis-a-vis Sachs's critique?

## Solutions

```r
#--------------------------- Load Data etc.
library(haven)
library(AER)
library(stargazer)
ajr <- read_dta('http://ditraglia.com/econ224/ajr.dta')

#--------------------------- Custom Stargazer Output
mystar <- function(...) {
  stargazer(..., type = 'latex', header = FALSE, digits = 2, omit.stat = c('f', 'ser', 'adj.rsq'))
}
```

```r
#--------------------------- Question 2: OLS Regression
ols <- lm(loggdp ~ risk, ajr)
mystar(ols, title = 'OLS Results')
```

Table 2: OLS Results

| | *Dependent variable:* |
|---|---|
| | loggdp |
| risk | 0.51*** |
| | (0.06) |
| Constant | 4.73*** |
| | (0.41) |
| Observations | 62 |
| $R^2$ | 0.52 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

```r
#--------------------------- Question 3: IV Regression
# First-stage
first_stage <- lm(risk ~ logmort0, ajr)
mystar(first_stage, title = 'First Stage Regression')
```

Table 3: First Stage Regression

| | *Dependent variable:* |
|---|---|
| | risk |
| logmort0 | −0.62*** |
| | (0.13) |
| Constant | 9.39*** |
| | (0.63) |
| Observations | 62 |
| $R^2$ | 0.27 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

```r
# Reduced Form
reduced_form <- lm(loggdp ~ logmort0, ajr)
mystar(reduced_form, title = 'Reduced Form Regression')
```

```r
# IV Regression
iv <- ivreg(loggdp ~ risk | logmort0, data = ajr)
mystar(iv, title = 'IV Results')
```

Table 4: Reduced Form Regression

| | *Dependent variable:* |
|---|---|
| | loggdp |
| logmort0 | $-0.56^{***}$ |
| | (0.08) |
| Constant | $10.63^{***}$ |
| | (0.38) |
| Observations | 62 |
| $R^2$ | 0.46 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 5: IV Results

| | *Dependent variable:* |
|---|---|
| | loggdp |
| risk | $0.91^{***}$ |
| | (0.16) |
| Constant | $2.13^{**}$ |
| | (1.01) |
| Observations | 62 |
| $R^2$ | 0.19 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

```r
# IV by hand
coef(reduced_form)[2] / coef(first_stage)[2]
```

```
 logmort0
0.9052929
```

```r
#-------------------------- Question 4: Malaria Critique
ols2 <- lm(loggdp ~ risk + malaria, ajr)
mystar(ols, ols2, title = 'OLS Results')
```

Table 6: OLS Results

| | *Dependent variable:* | |
|---|---|---|
| | loggdp | |
| | (1) | (2) |
| risk | 0.51*** | 0.34*** |
| | (0.06) | (0.06) |
| malaria | | −1.15*** |
| | | (0.18) |
| Constant | 4.73*** | 6.30*** |
| | (0.41) | (0.41) |
| Observations | 62 | 62 |
| $R^2$ | 0.52 | 0.71 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

```r
first_stage2 <- lm(risk ~ logmort0 + malaria, ajr)
mystar(first_stage, first_stage2, title = 'First Stage Regression')
```

```r
iv2 <- ivreg(loggdp ~ risk + malaria | logmort0 + malaria, data = ajr)
mystar(iv, iv2, title = 'IV Results')
```

```r
#-------------------------- Question 5: Sachs Critique
iv3 <- ivreg(loggdp ~ risk + malaria + latitude + rainmin + meantemp |
               logmort0 + malaria + latitude + rainmin + meantemp,
             data = ajr)
mystar(iv, iv2, iv3, title = 'IV Results')
```

Table 7: First Stage Regression

| | *Dependent variable:* | |
|---|---|---|
| | risk | |
| | (1) | (2) |
| logmort0 | $-0.62^{***}$ | $-0.44^{**}$ |
| | (0.13) | (0.19) |
| malaria | | $-0.70$ |
| | | (0.52) |
| Constant | $9.39^{***}$ | $8.83^{***}$ |
| | (0.63) | (0.75) |
| Observations | 62 | 62 |
| $R^2$ | 0.27 | 0.29 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

Table 8: IV Results

| | *Dependent variable:* | |
|---|---|---|
| | loggdp | |
| | (1) | (2) |
| risk | $0.91^{***}$ | $0.59^{**}$ |
| | (0.16) | (0.22) |
| malaria | | $-0.75^{*}$ |
| | | (0.40) |
| Constant | $2.13^{**}$ | $4.50^{***}$ |
| | (1.01) | (1.59) |
| Observations | 62 | 62 |
| $R^2$ | 0.19 | 0.61 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

Table 9: IV Results

| | *Dependent variable:* | | |
|---|---|---|---|
| | loggdp | | |
| | (1) | (2) | (3) |
| risk | $0.91^{***}$ | $0.59^{**}$ | 0.68 |
| | (0.16) | (0.22) | (1.30) |
| malaria | | $-0.75^{*}$ | $-0.75$ |
| | | (0.40) | (0.82) |
| latitude | | | $-0.44$ |
| | | | (2.51) |
| rainmin | | | $-0.001$ |
| | | | (0.02) |
| meantemp | | | 0.01 |
| | | | (0.09) |
| Constant | $2.13^{**}$ | $4.50^{***}$ | 3.83 |
| | (1.01) | (1.59) | (9.66) |
| Observations | 62 | 62 | 62 |
| $R^2$ | 0.19 | 0.61 | 0.54 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | |