Replication of *Income and Democracy* (2008)

Dora Simon and Julian Langer

6/9/2018

In this document, we try to replicate columns 1 and 2 of table 2 of "Income and Democracy" by Acemoglu et al (2008). The original paper can be found here.

It is always good to first load the libraries needed and not to have them all over the place.

Libraries

```
library(readxl)          # read excel files
library(tibble)          # cuter dataframes
library(dplyr)           # data manipulation
library(lfe)             # fixed effects models
library(stargazer)       # nice tables
library(lmtest)          # for coeftest function
library(multiwayvcov)    # (multiway) clustered standard errors
library(AER)             # instrumental variables
library(ivpack)          # robust standard errors for ivreg
#library(plm)
```

First, we read in the data and adjust it to our needs.

Loading the data for estimation

```
ajry_df = read_xls("./raw_data/ajry.xls",
                   sheet = 2) %>%
  arrange(code_numeric, year_numeric) %>%
  rename(log_gdp_pc = lrgdpch,
         freedom_house = fhpolrigaug)
```

Then, we generate lagged values by year and keep only the observations which belong to sample 1.

```
# generate lagged variables
ajry_df = ajry_df %>%
  group_by(code_numeric) %>%
  mutate(lag_log_gdp_pc = lag(log_gdp_pc, order_by = year_numeric),
         lag_freedom_house = lag(freedom_house, order_by = year_numeric),
         lag2_nsave = lag(nsave, 2, order_by = year_numeric),
         lag_worldincome = lag(worldincome, order_by = year_numeric)) %>%
  filter(sample == 1)

#saveRDS(ajry_df, "./tidy_data/tidy-data.rds")
```

For the regressions, we are going to first replicate columns 1 and 2 using the standard `lm` functions. Then, we are repeating this step for the newer `felm` functions.

Pooled OLS with Time Effects

```
#ajry_df<-readRDS("./tidy_data/tidy-data.rds")
# pooled ols with lm
pooled_est = lm(freedom_house ~ -1 + lag_freedom_house + lag_log_gdp_pc +
             factor(year_numeric), data = ajry_df)
# standard errors clustered by country
vcov_pooled_lm <- cluster.vcov(pooled_est, ajry_df$code_numeric)
pooled.lm.se   <- sqrt(diag(vcov_pooled_lm ))
```

```
stargazer(pooled_est, se=list(pooled.lm.se),  omit="year_numeric",
          covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy", type="latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:51:56

Table 1:

	Dependent variable:
	democracy
Democracy lag	0.706***
	(0.035)
Log GDP per capita lag	0.072***
	(0.010)
Observations	945
R^2	0.920
Note:	*p<0.1; **p<0.05; ***p<0.01

Fixed Effects with the `lm` function

```
# fe with lm
fe_est = lm(freedom_house ~ -1 + lag_freedom_house + lag_log_gdp_pc +
             factor(year_numeric) + factor(code_numeric), data = ajry_df)

# standard errors clustered by country
vcov_fe_lm <- cluster.vcov(fe_est, ajry_df$code_numeric)
fe.lm.se   <- sqrt(diag(vcov_fe_lm ))
```

```
stargazer(pooled_est, fe_est, se=list(pooled.lm.se, fe.lm.se),
          omit=c("year_numeric", "code_numeric"),
          covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy", type="latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:52:01

Table 2:

	Dependent variable:	
	democracy	
	(1)	(2)
Democracy lag	0.706***	0.379***
	(0.035)	(0.051)
Log GDP per capita lag	0.072***	0.010
	(0.010)	(0.035)
Observations	945	945
R^2	0.920	0.941
Note:	*p<0.1; **p<0.05; ***p<0.01	

Pooled OLS and FE with the `lfe` package

```
# pooled OLS
pool.felm <- felm(freedom_house ~ lag_freedom_house + lag_log_gdp_pc | year_numeric | 0 | code_numeric,
             data = ajry_df)
```

```
# pooled OLS
stargazer(pool.felm, covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy",type = 'latex')
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:52:01

Table 3:

	Dependent variable:
	democracy
Democracy lag	0.706***
	(0.035)
Log GDP per capita lag	0.072***
	(0.010)
Observations	945
R^2	0.725
Note:	*p<0.1; **p<0.05; ***p<0.01

```
fe.felm <- felm(freedom_house ~ lag_freedom_house + lag_log_gdp_pc | year_numeric + code_numeric | 0 |
               code_numeric, data = ajry_df)
```

```
stargazer(pool.felm, fe.felm, covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy",type = 'latex')
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:52:01

Table 4:

	Dependent variable:	
	democracy	
	(1)	(2)
Democracy lag	0.706***	0.379***
	(0.035)	(0.051)
Log GDP per capita lag	0.072***	0.010
	(0.010)	(0.035)
Observations	945	945
R^2	0.725	0.796
Note:	*p<0.1; **p<0.05; ***p<0.01	

Pooled OLS and FE with the `plm` package

If I load the `plm` package before the `lm` and `felm` estimations, it destroys the regression results. I have not figured out why, but probably, the `plm` overrides some functions that are used for the other estimations.

```
library(plm)
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'plm'
```

```
## The following object is masked from 'package:lfe':
##
##     sargan
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```
plm_dataframe<-pdata.frame(ajry_df, index=c("code_numeric", "year_numeric"))

pooled_plm <- plm(freedom_house ~ -1 + lag_freedom_house + lag_log_gdp_pc +
            factor(year_numeric), data = plm_dataframe, model="pooling")

vcov_pooled_plm <- vcovHC(pooled_plm, type="HC0", cluster="group")
pooled.plm.se   <- sqrt(diag(vcov_pooled_plm ))
```

```
stargazer(pooled_plm, se=list(pooled.lm.se, pooled.plm.se),  omit="year_numeric",
          covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy", type="latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:52:01

```
# fe with lm
fe_plm = plm(freedom_house ~ -1 + lag_freedom_house + lag_log_gdp_pc +
            factor(year_numeric) + factor(code_numeric), data = plm_dataframe, model="within")
```

Table 5:

	Dependent variable:
	democracy
Democracy lag	0.706***
	(0.035)
Log GDP per capita lag	0.072***
	(0.010)
Observations	945
R^2	0.725
Note:	*p<0.1; **p<0.05; ***p<0.01

```
# standard errors clustered by country
vcov_fe_lm <- cluster.vcov(fe_est, ajry_df$code_numeric)
fe.lm.se   <- sqrt(diag(vcov_fe_lm ))
```

```
stargazer(pooled_est, fe_est, se=list(pooled.lm.se, fe.lm.se),
          omit=c("year_numeric", "code_numeric"),
          covariate.labels = c("Democracy lag", "Log GDP per capita lag"),
          keep.stat = c("rsq", "n"), dep.var.labels   = "democracy", type="latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sa, Sep 08, 2018 - 20:52:02

Table 6:

	Dependent variable:	
	democracy	
	(1)	(2)
Democracy lag	0.706***	0.379***
	(0.035)	(0.051)
Log GDP per capita lag	0.072***	0.010
	(0.010)	(0.035)
Observations	945	945
R^2	0.920	0.941
Note:	*p<0.1; **p<0.05; ***p<0.01	