Code

[Christian Schluter's] MAG1 Applied Econometrics: Project 4 based on Acemoglou and Angrist (2000, NBER).

(https://christianschluter.github.io)

This project is based on

- Acemoglou, D. and J. Angrist (2000), ``How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws", *NBER Macroeconomics Annual*, 15, 9-59.

Questions:

- Q1. Explain the research design of the paper. In particular, explain why estimation by OLS is likely to be invalid, explain the nature of the instruments, and discuss whether the proposed instrumental variable strategy is credible.
- Q2 (Replication Tasks). Using the data file `AA_small.dta` (and either `stata` or R):
  - Replicate first the OLS benchmark, reported in Acemoglou and Angrist (2000, `AA2000`) in Table 2 column 3 (.075).
  - Replicate the summary table presented below (in section 2).
  - Subsample analysis: Replicate AA2000 Tables 4 (cols 5-8) and 6 (as per in section 2 below).
  - Briefly explain and comment on the main story of these Tables.
- Q3 (Data Wrangling). Extract from ipums USA (https://usa.ipums.org/usa/) the relevant variables from the 1980 census, compare your data extract to the 1980 subsample from `AA_small.dta`, as well as the OLS benchmark result. Hint: Consult `data wrangling exercises 2 (ipums)` of your homework.

NB: The full sample `AA_small.dta` is large (we have 722,343 persons), and a single regressions will take a few minutes if the sets of controls include a large number of interacted terms (i.e. `sob:yob`).

1 Preliminaries and code snippets

The following code snippets (in `R`) could help you in your replications.

Hide

```
library(tidyverse)
library(haven)
library(sandwich)   # for: vcovCL
library("AER")      # for: ivreg
library("lmtest")   # for: coeftest
library(stargazer)

child_labor <- read_dta(paste(path,"AA_small.dta", sep=""))
child_labor %>%
     mutate(year = factor(year),
            yob_fct = factor(yob),
            sob = factor(sob)) -> child_labor
```

```
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------
age             byte    %10.0g                age at census
avgEduc         double  %10.0g                average grade completed in SOR at cen
sus
ca              byte    %9.0g                 compulsory attendence requirment in S
OB at age 14
ca8             byte    %9.0g                 ca <= 8
ca9             byte    %9.0g                 ca = 9
ca10            byte    %9.0g                 ca = 10
ca11            byte    %9.0g                 ca = 11
cl              byte    %9.0g                 child labor threshold in SOB at age 1
4
cl6             byte    %9.0g                 cl <= 6
cl7             byte    %9.0g                 cl = 7
cl8             byte    %9.0g                 cl = 8
cl9             byte    %9.0g                 cl >= 9
indEduc         float   %9.0g                 highest grade completed
lnwkwage        double  %10.0g                log weekly wages
qob             byte    %10.0g                quarter of birth
sample5080      byte    %9.0g                 census years 1950-1980
sample5090      byte    %9.0g                 census years 1950-1990
sample6080      byte    %9.0g                 census years 1960-1980
sample6090      byte    %9.0g                 census years 1960-1990
sob             byte    %10.0g                state of birth
sor             byte    %10.0g                state of residence at census
weight          double  %10.0g                sample line weight
year            int     %10.0g                census year
yob             int     %9.0g                 year of birth
comp8           byte    %9.0g
```

Using `ipums`, AA2000 have generated an extract of white men in their forties, drawn from each of the U.S. census samples available every decade from 1950 to 1990 (a total of 5 census). Then, a manually generated data set coding the child labour laws has been merged by state. This extra variable is `cl`, and AA2000 use the indicators (`cl7`, `cl8`, `cl9`) as instruments.

2 Replication tasks

Question 2.2. Replicate (and interpret) the following table:

	Dependent variable			
	Years of schooling		Log weekly wages	
	(1)	(2)	(3)	(4)
A. First-stage and reduced-form estimates				
Child labor law req. 7 years	.166	−.024	.010	−.013
	(.067)	(.048)	(.011)	(.011)
Child labor law req. 8 years	.191	.024	.013	.005
	(.062)	(.051)	(.010)	(.010)
Child labor law req. 9 years or more	.400	.016	.046	.008
	(.098)	(.053)	(.017)	(.014)
B. Second-stage estimates				
Years of education			.124	.399
			(.036)	(.360)
State of birth dummies × linear year of birth trends	No	Yes	No	Yes

Panel A.col(1) is essentially column 7 of AA2000's Table 4. Panel B.col(3) is essentially column 6 of their Table 6 (lower part). All specifications use census year (`year`), year of birth dummies (`yob` coded as factors), and state of birth (`sob`). Columns 2 and 4 include, in addition, the full set of interactions between `sob` and `yob` in order to capture state-specific trends. All SEs are clustered at the state-of-birth (`sob`) level.

Question 2.3 Subsamples analysis. Consider now the subsamples used in AA2000 Tables 4 (cols 5-8) and 6. The first stage regression results for the subsamples are:

Without State-of-Residence Controls

1960–1980 (5)	*1950–1980* (6)	*1950–1990* (7)	*1980* (8)
0.105	0.115	0.175	0.062
(0.077)	(0.051)	(0.043)	(0.041)
0.120	0.119	0.202	0.143
(0.093)	(0.075)	(0.059)	(0.034)
0.269	0.225	0.410	0.182
(0.098)	(0.084)	(0.059)	(0.041)

Turning to Table 6 (cols 1-3) using the Quarter-of-Birth instrument (`qob:yob_fct`), and then the Child Labour Laws instruments (cols 4-6):

	QOB Instruments			SOB-CL Instruments		
	1960–1980 (1)	1980 (2)	1960–1970 (3)	1960–1980 (4)	1950–1980 (5)	1950–199((6)
Including state-of-residence main effects	0.073 (0.012)	0.090 (0.016)	0.063 (0.017)	0.076 (0.034)	0.103 (0.038)	0.113 (0.018)
No state-of-residence main effects	0.073 (0.012)	0.088 (0.016)	0.063 (0.017)	0.080 (0.064)	0.112 (0.060)	0.126 (0.027)
N	609,852	376,479	233,373	609,852	626,511	729,695

You might not get exactly the same results. For instance, I have obtained:

```
## IV estimation: QoB
## ===================================
##
##                (1)     (2)     (3)
## -----------------------------------
## indEduc       0.072   0.089   0.065
##              (0.014) (0.016) (0.019)
## -----------------------------------
## Observations 603,993 372,597 231,396
## R2            0.459   0.106   0.283
## ===================================
```

3 Data wrangling: Using ipums 1980

For census year 1980, generate the main data from scratch using ipums (https://www.ipums.org/) ([Census] Integrated Public Use Micro Samples) using R or `stata` . Recall that you need to create an extract of white US-born men in their forties with positive earnings.

If you want to use R, recall the following code snippet (from `data wrangling exercises 2 (ipums)` of your homework):

Hide

```
library("ipumsr")
ddi   <- read_ipums_ddi(paste(path.ipums,"usa_1980_extract_bpl.xml",sep=""))
dat80 <- read_ipums_micro(ddi, data_file =
                          paste(path.ipums,"usa_1980_extract_bpl.dat",sep=""))
```

```
## Use of data from IPUMS USA is subject to conditions including that users should
## cite the data appropriately. Use command `ipums_conditions()` for more details.
```

Hide

```
dat80
```

```
## # A tibble: 11,343,120 x 21
##     YEAR       SAMPLE SERIAL  HHWT CLUSTER STRATA      GQ PERNUM PERWT     SEX
##    <int>    <int+lbl>  <dbl> <dbl>   <dbl>  <dbl> <int+l>  <dbl> <dbl> <int+l>
##  1  1980 198001 [198…      1    20 1.98e12     22 1 [Hou…      1    20 2 [Fem…
##  2  1980 198001 [198…      1    20 1.98e12     22 1 [Hou…      2    20 1 [Mal…
##  3  1980 198001 [198…      1    20 1.98e12     22 1 [Hou…      3    20 1 [Mal…
##  4  1980 198001 [198…      1    20 1.98e12     22 1 [Hou…      4    20 1 [Mal…
##  5  1980 198001 [198…      1    20 1.98e12     22 1 [Hou…      5    20 1 [Mal…
##  6  1980 198001 [198…      2    20 1.98e12     24 1 [Hou…      1    20 2 [Fem…
##  7  1980 198001 [198…      2    20 1.98e12     24 1 [Hou…      2    20 1 [Mal…
##  8  1980 198001 [198…      3    20 1.98e12     28 1 [Hou…      1    20 1 [Mal…
##  9  1980 198001 [198…      3    20 1.98e12     28 1 [Hou…      2    20 2 [Fem…
## 10  1980 198001 [198…      3    20 1.98e12     28 1 [Hou…      3    20 2 [Fem…
## # … with 11,343,110 more rows, and 11 more variables: AGE <int+lbl>,
## #   BIRTHQTR <int+lbl>, BIRTHYR <dbl+lbl>, RACE <int+lbl>, RACED <int+lbl>,
## #   BPL <int+lbl>, BPLD <int+lbl>, EDUC <int+lbl>, EDUCD <int+lbl>,
## #   WKSWORK1 <dbl+lbl>, INCWAGE <dbl+lbl>
```

As a benchmark, replicate the sample characteristics (AA2000 Table 1, col 1980) and the OLS result. Note that education is slightly differently coded, the replication will not be exact.