Facts About Growth:

ECON807 Macroeconomic Theory & Policy, PS2 Q2

Daniel Sánchez

1/24/23

This document goes through my code and output done to answer question 2 from problem set 2 (ECON807 Macroeconomic Theory & Policy).

Preliminaries

Below, I load the libraries which I will need for doing the assignment.

```
# Load libraries

library(tidyverse)
library(patchwork)
library(WDI)
library(modelsummary)
library(sandwich)
library(fixest)
```

Question 2 (a): Physical Capital Accumulation & Growth

Loading & Cleaning the Data

To answer this question, I will use data from the World Bank's World Development Indicators, which is a lot of data at the country level since 1960. I will extract the data from the World Bank's API by using the marvelous package **WDI** from Vincent Arel-Bundock. I load the data below:

```
# Create my vector of indicators that I want to use

indicators <- c(
  'inv_gdp'= 'NE.GDI.FTOT.ZS', #  Gross fixed capital formation (% of GDP)
  'gdp_pc' = 'NY.GDP.PCAP.KD',  # GDP per capita, PPP (constant 2017 international $)
  'gdp_g'= 'NY.GDP.PCAP.KD.ZG', # GDP per capita growth (annual %)
  'enrolment' = 'SE.PRM.ENRR' # Gross primary school enrolment
  )

# Now get my data

wdi_raw <-
  WDI(indicator = indicators,
      extra = T)
```

I will use gross fixed capital formation as a % of GDP to proxy Investment/GDP, as according to the database's documentation, this variable "includes land improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and the like, including schools, offices, hospitals, private residential dwellings, and commercial and industrial buildings" (2022a). I believe this variable most accurately represents the rate at which a country invests in physical capital.

I must clean the data to be able to use it. I drop all observations which are not actually countries.

```
# Create the dataframe object that I will use and filter out aggregates

df <-
  wdi_raw %>%
  filter(region != 'Aggregates')
```

I will use a sample of countries which have data for the first available year at the World Bank WDIs, which is 1960. To get the number of countries which have available data for all variables then, I will create a table with that information beginning 1961 (as GDP growth would only be available one period after the first one recorded at the WDIs).

```
countries_1961 <-
  wdi_raw %>%
  filter(year == 1961,
         !is.na(gdp_g),
         !is.na(gdp_pc),
         !is.na(inv_gdp))
```

There are 42, which means there can be about 2562 observations in total which are used for regressions. I will do an inner join of the 1960 GDP per capita for all countries which have it available.

```
df <-
  df %>%
  inner_join(
    wdi_raw %>%
      filter(year == 1960, !is.na(gdp_pc)) %>%
      select(iso3c, gdp_pc) %>%
      rename(gdp0 = 'gdp_pc'),
    by = 'iso3c'
  )
```

Regression Analysis

Run the regressions, which follow the model below:

$$\ln Y_{it}^{pc} = \beta_0 + \beta_1(\text{Investment}_{it}/Y_{it}) + \mathbf{x}_{it}'\gamma + u_{it}$$

where $\mathbf{x}_{it}$ is a vector of controls which may or may not include year and country fixed effects, and γ is the associated set of parameters. u_{it} is the unobserved error and we take Y_{it} to be the GDP (total or per capita,, pc) for country i in year t.

```
# Simple model without controlling for initial GDP level

reg1 <-
  lm(gdp_g ~ inv_gdp,
     data = df)

# Control for initial GDP level

reg2 <-
  lm(gdp_g ~ inv_gdp + gdp0,
     data = df)

# Control for initial GDP level with log

reg2l <-
  lm(gdp_g ~ inv_gdp + log(gdp0),
     data = df)
```

```
# Country fixed effects (without the value of )

reg3<-
  feols(gdp_g ~ inv_gdp | iso3c,
        cluster = ~iso3c,
        data = df)

# Year fixed effects

reg4<-
  feols(gdp_g ~ inv_gdp + log(gdp0) | year,
        cluster = ~year,
        data = df)

# Country & Year fixed effects

reg5<-
  feols(gdp_g ~ inv_gdp | year + iso3c,
        cluster = ~ (year+iso3c),
        data = df)
```

I display the models in the table below. Heteroskedasticity robust errors under the HC3 method are reported in some cases, whereas in FE models I report cluster-robust standard errors.

```
# Create a list of models for the table

models<-
  list(
    reg1,
    reg2,
    reg2l,
    reg3,
    reg4,
    reg5
  )

# Do a named vector for my coefficient names

names<-
```

```r
  c('inv_gdp' = 'Investment/GDP',
    'gdp0' = '1960 GDP per capita',
    'log(gdp0)' = 'Ln 1960 GDP per capita',
    '(Intercept)' = 'Intercept'
  )

# Do the table

modelsummary(models,
             vcov = list('robust','robust', 'robust', ~iso3c, ~year, ~iso3c + year),
             stars = c('*' = .1, '**' = 0.05, '***' = .01),
             coef_map = names,
             gof_omit = c('AIC|BIC|Log|RMSE'),
             output = 'markdown')
```

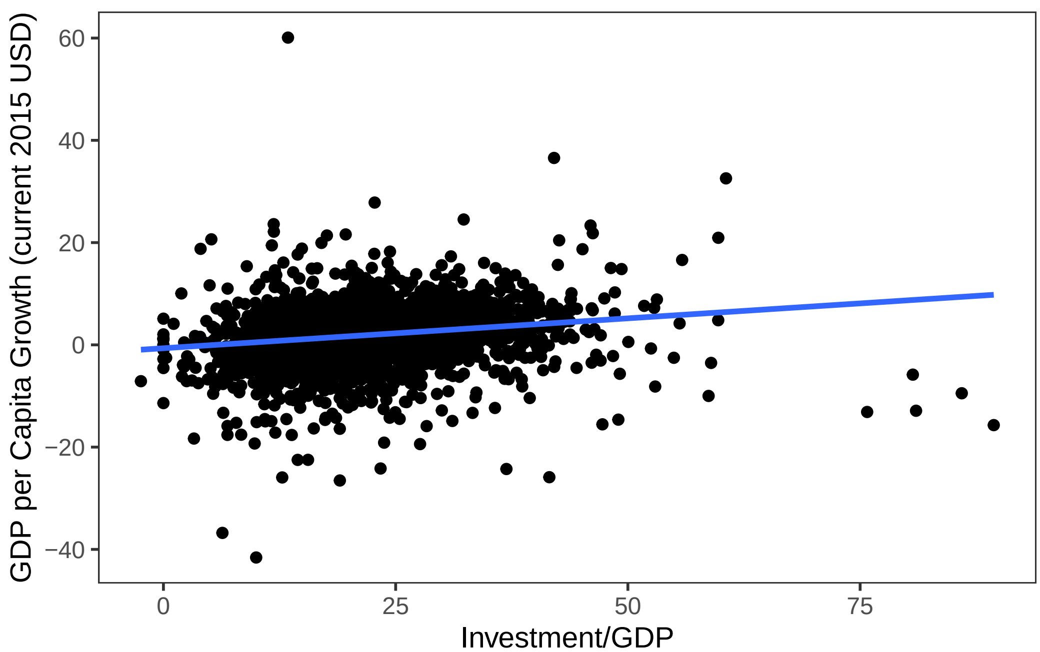
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Investment/GDP	0.117***	0.118***	0.120***	0.115***	0.132***	0.135***
	(0.017)	(0.017)	(0.017)	(0.035)	(0.017)	(0.037)
1960 GDP per capita		-0.00002**				
		(0.00001)				
Ln 1960 GDP per capita			-0.182***		-0.201***	
			(0.056)		(0.070)	
Intercept	-0.660*	-0.592	0.661			
	(0.369)	(0.367)	(0.544)			
Num.Obs.	4233	4233	4233	4233	4233	4233
R2	0.036	0.037	0.038	0.100	0.159	0.215
R2 Adj.	0.036	0.036	0.038	0.082	0.147	0.187
R2 Within				0.022	0.050	0.031
R2 Within Adj.				0.021	0.049	0.031
Std.Errors	HC3	HC3	HC3	by: iso3c	by: year	by: iso3c & year
FE: iso3c				X		X
FE: year					X	X

Note: ^^ * p < 0.1, ** p < 0.05, *** p < 0.01

This shows that investment/GDP is related to greater GDP per capita. An increase in one percentage point of investment/GDP is associated with a 0.11-0.14 increase in the growth rate. The most conservative effect is when not controlling for country and year fixed effects. Note that when I consider country fixed effects, there is no need to control for 1960 (initial) GDP per capita, as it is actually perfectly collinear with the fixed effects. All of the effects are statistically significant at the 99% level.

Below, I also plot the data for the level of GDP per capita (not growth) as well as the growth rate in scatterplotS to show the observed relationship.

```
ggplot(df, aes(inv_gdp, gdp_g)) +
  geom_point()+
  stat_smooth(method = 'lm', se = F)+
  theme_bw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = 'Investment/GDP',
       y = 'GDP per Capita Growth (current 2015 USD)')
```



Question 2(b): Including human capital

I now allow for the buildup of human capital by including the primary schooling rate (gross). According to the World Bank, this rate "is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown" (2022b).

```r
regh <-
  lm(gdp_g ~ enrolment,
     data = df)

# Control for initial GDP level with log

reg1h <-
  lm(gdp_g ~ inv_gdp + log(gdp0) + enrolment,
     data = df)

# Country fixed effects

reg2h<-
  feols(gdp_g ~ inv_gdp  + enrolment | iso3c,
        cluster = ~year,
        data = df)


# Year fixed effects

reg3h<-
  feols(gdp_g ~ inv_gdp + log(gdp0) + enrolment | year,
        cluster = ~year,
        data = df)

# Country & Year fixed effects

reg4h<-
  feols(gdp_g ~ inv_gdp + enrolment | year + iso3c,
        cluster = ~ (year+iso3c),
        data = df)
```

The results are presented below:

```r
# Create a list of models for the table

models<-
  list(
    regh,
    reg1h,
    reg2h,
    reg3h,
```

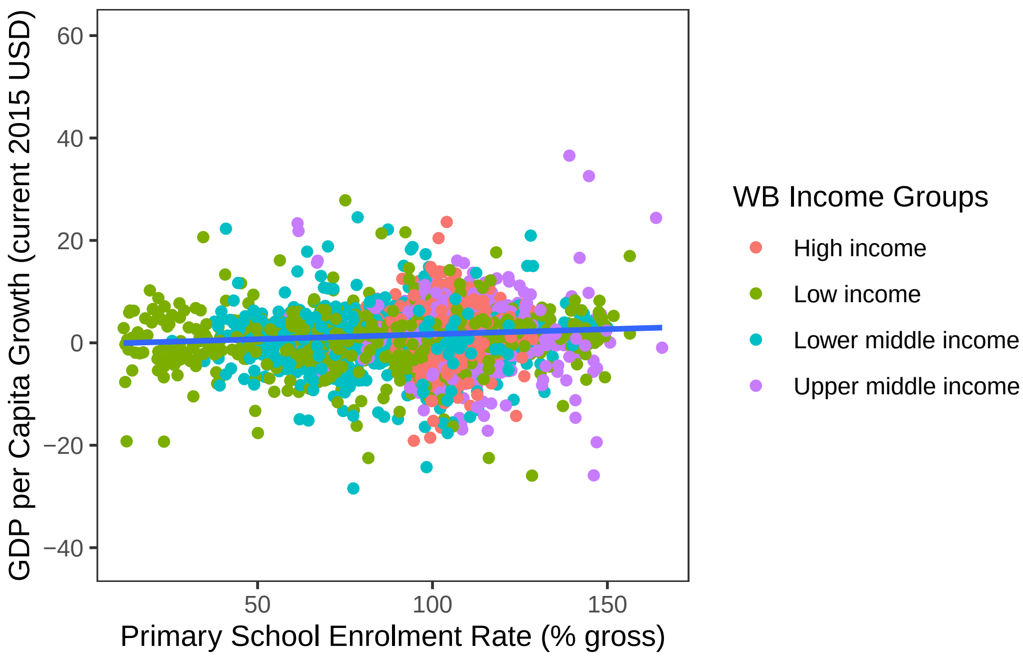
```
    reg4h
  )

# Do a named vector for my coefficient names

names<-
  c('inv_gdp' = 'Investment/GDP',
    'enrolment' = 'Primary school enrolment rate (% gross)',
    'log(gdp0)' = 'Ln 1960 GDP per capita',
    '(Intercept)' = 'Intercept'
  )

# Do the table

modelsummary(models,
             vcov = list('robust', 'robust',~iso3c, ~year, ~iso3c + year),
             stars = c('*' = .1, '**' = 0.05, '***' = .01),
             coef_map = names,
             gof_omit = c('AIC|BIC|Log|RMSE'),
             output = 'markdown')
```

	Model 1	Model 2	Model 3	Model 4	Model 5
Investment/GDP		0.118***	0.121**	0.118***	0.125***
		(0.021)	(0.047)	(0.021)	(0.046)
Primary school enrolment rate (% gross)	0.020***	0.009*	0.0005	0.013***	0.010
	(0.004)	(0.005)	(0.010)	(0.005)	(0.010)
Ln 1960 GDP per capita		-0.262***		-0.273***	
		(0.055)		(0.063)	
Intercept	-0.272	0.204			
	(0.432)	(0.660)			
Num.Obs.	3805	3306	3306	3306	3306
R2	0.008	0.046	0.112	0.156	0.220
R2 Adj.	0.008	0.045	0.089	0.142	0.187
R2 Within			0.024	0.053	0.030
R2 Within Adj.			0.024	0.052	0.030
Std.Errors	HC3	HC3	by: iso3c	by: year	by: iso3c & year
FE: iso3c			X		X
FE: year				X	X

Note: ^^ * p < 0.1, ** p < 0.05, *** p < 0.01

The results show that while a naive regression might relate higher schooling rates to higher growth rates, however, when controlling for the ratio of physical capital accumulation used in the previous question, as well as for country fixed effects, this effects seems to go away. I show this relationship graphically below, by using the World Bank's income groups.

```
ggplot(df, aes(enrolment, gdp_g)) +
  geom_point(aes(colour = income))+
  stat_smooth(method = 'lm', se = F)+
  theme_bw()+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())+
  labs(x = 'Primary School Enrolment Rate (% gross)',
       y = 'GDP per Capita Growth (current 2015 USD)',
       colour = 'WB Income Groups' )
```



The grouping above allows to illustrate what the regressions above are also illustrating: countries with high income are already well-educated, so must of the variation across countries in schooling can be explained through differences in income.

References

The World Bank. 2022a. "Gross Fixed Capital Formation (% of GDP)."

———. 2022b. "School Enrollment, Primary (% Gross)."