Development Economics Assingment NYU Spring 2023

2023-02-06

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#install.packages("pwt10")

library("pwt10")
library(reshape2)
library(effectsize)
library(estimatr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("ggpubr")
```

```
## Loading required package: ggplot2
```

```
library(modelsummary)
library(magrittr)
library(readxl)
library(tinytex)
library(tibble)
library(tidyverse)  # ggplot(), %>%, mutate(), and friends
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## v purrr   0.3.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x purrr::set_names() masks magrittr::set_names()
```

```r
library(broom)  # Convert models to data frames
library(rdrobust)  # For robust nonparametric regression discontinuity
library(rddensity)  # For nonparametric regression discontinuity density tests
```

```r
#load the data from the package
data("pwt10.0")
```

1. What is the growth rate of gdp per worker? Show the derivation. Remember that, from the notation in Tuesday's class, the growth rate of gdp per worker can be written as ^y.

We are going to have Y = AK$^{\text{(alpha)(hL)}}$(1-alpha)

Then we have to translate it to nominal rates, and take the growth rates with logs:

g(y) = g(a) + (1-alpha)g(h) + (alpha)g(k)

(L is no longer here because we divided everything by L in order to get the terms in per capital)

(where g(y) = ˆy = y_hat and so on for the rest of the variables)

Where alpha is assumed to be 1/3

The parts of the dataset that we are going to take are: year, country, real gdp, capital, services, employment, and the human capital index.

```r
pwt = select(pwt10.0, 'year','country','rgdpna', 'rnna', 'emp','pop', 'hc')

head(pwt,5)
```

```
##          year country rgdpna rnna emp pop hc
## ABW-1950 1950   Aruba     NA   NA  NA  NA NA
## ABW-1951 1951   Aruba     NA   NA  NA  NA NA
## ABW-1952 1952   Aruba     NA   NA  NA  NA NA
## ABW-1953 1953   Aruba     NA   NA  NA  NA NA
## ABW-1954 1954   Aruba     NA   NA  NA  NA NA
```

2. At first glance head(pwt,5) might seem a bit concerning. Explain why it looks concerning, and in fact why it's totally reasonable what's going on.

We're not going to want to go back to 1950. Let's do something a little more recent, like 1980. We can do this using filter()

pwt_short = filter(pwt,year>1975)

It looks concerning because a lot of data is missing for several years and several countries (we only have year and name of the country for the most part). This is because our observations are starting in 1950, and that might be to early for the worldwide standardization of these numbers being published. That is why we will need to filter the data a little bit so that we have data points with information on it.

```r
pwt_short = filter(pwt,year>1975)

head(pwt_short,5)
```

```
##          year country   rgdpna     rnna emp      pop hc
## ABW-1976 1976   Aruba 442.7044 1218.427  NA 0.060586 NA
## ABW-1977 1977   Aruba 482.9395 1336.983  NA 0.060366 NA
## ABW-1978 1978   Aruba 526.8314 1466.154  NA 0.060103 NA
## ABW-1979 1979   Aruba 574.7123 1606.880  NA 0.059980 NA
## ABW-1980 1980   Aruba 626.9450 1760.181  NA 0.060096 NA
```

3. Is head(pwt short,5) less concerning? So let's instead only take the ones with no missing variables, using the very convenient drop na function

Now we have a little more information, but we are still missing some data from employment and human capital, which are even more recent. In order to have them all, we might want to either pick on countries that were pioneers using these statistics (dropping nans), or filter year a little bit higher (up to 1995 probably). We are going to proceed with the first method, dropping nans.

```r
pwt_short = drop_na(filter(pwt,year>1975))

head(pwt_short,5)
```

```
##          year country   rgdpna     rnna      emp      pop       hc
## AGO-1976 1976  Angola 57783.69 406898.4 4.405441 7.279509 1.030838
## AGO-1977 1977  Angola 58245.95 423176.3 4.537667 7.533735 1.033385
## AGO-1978 1978  Angola 55158.92 437047.6 4.670081 7.790707 1.035939
## AGO-1979 1979  Angola 55324.40 450429.5 4.807217 8.058067 1.038498
## AGO-1980 1980  Angola 56652.18 462300.9 4.952234 8.341289 1.041064
```

Now, we can see that we do not have Aruba as the first one because they never implemented HC information throughout the years, so now we only have countries that are compliant with all the fields.

4. Write out the code that reports all of the countries (once) Now you get to pick a country. Whatever country you want!

Now lets display all the different countries we have for our analysis

```r
head(distinct(pwt_short, country),5)
```

```
##                       country
## AGO-1976               Angola
## ALB-1976              Albania
## ARE-1976 United Arab Emirates
## ARG-1976            Argentina
## ARM-1990              Armenia
```

```r
print('Just printed the first 5 so its not a long list')
```

```
## [1] "Just printed the first 5 so its not a long list"
```

5. Which country did you pick? How can you make a new dataframe called "my country" that only contains the values for your country? Use our friend % > %, and go back to the optional assignment if you are having trouble remembering R logic for strings.

Now we are going to calculate the compound annual growth rate for your country (using whatever the start year and end year is for your country). Some useful commands.

First: I would create a dataframe with only two years in it (where 19xx and 20xx are the first and last years)

my_country_short <- my_country %>% filter(year==19xx | year == 20xx)

What does | stands for?

Now we want to add some variables with the growth rates. I'll show you how to do this for year, noting that we don't actually care about the growth rate of years (since they grow at a rate of one a year) but it is useful for showing the code. I will use the functions lag (which is exactly what you might think it means) and mutate (which takes a data frame and appends it)

stands for OR as a logical operator

And I am picking my country of origin: Venezuela (Republica Bolivaria de Venezuela).

And the years I am picking the growth from are 1976 to 2001

```
my_country = drop_na(filter(pwt_short, country == "Venezuela (Bolivarian Republic of)"))

my_country_short <- my_country %>% filter(year==1976 | year == 2001)

my_country_short
```

```
##          year                            country    rgdpna      rnna    emp
## VEN-1976 1976 Venezuela (Bolivarian Republic of)  9420.365  94488.21 3.7032
## VEN-2001 2001 Venezuela (Bolivarian Republic of) 13799.385 185277.69 9.4027
##               pop       hc
## VEN-1976 13.57221 1.595433
## VEN-2001 24.64647 2.189336
```

lag creates a lagged version of a time series, shifting the time base forward by a given number of observations. Lag does exactly the opposite, shifting the time base backwards by the given number of observations.

```
my_country_short <- my_country_short %>% mutate(dif_years = year - lag(year),
growth_years = ((year/ lag(year))^(1/dif_years))-1,growth_rgdpna = ((rgdpna/ lag(rgdpna))^(1/dif_years)

my_country_short
```

```
##          year                            country    rgdpna      rnna    emp
## VEN-1976 1976 Venezuela (Bolivarian Republic of)  9420.365  94488.21 3.7032
## VEN-2001 2001 Venezuela (Bolivarian Republic of) 13799.385 185277.69 9.4027
##               pop       hc dif_years growth_years growth_rgdpna growth_rnna
## VEN-1976 13.57221 1.595433        NA           NA            NA          NA
## VEN-2001 24.64647 2.189336        25 0.0005030247    0.01538719  0.02730126
##          growth_emp growth_hc
## VEN-1976         NA        NA
## VEN-2001 0.03797529 0.0127386
```

6. What are the compound annual growth rates for Y, K, and L? Now manipulate the data a bit more

The annual growth rate for Y is 1.538719%

The annual growth rate for K is 2.730126%

The annual growth rate for L is 3.797529%

7. What are the compound annual growth rates for y, k, and h? (the lower case letters mean in per capita terms dividing by L)

Lets create my__country__short__per__capital

```
my_country_short_per_capital <- my_country_short %>% mutate(gdp_per_capital = rgdpna/emp, capital_per_ca


my_country_short_per_capital
```

```
##          year                             country    rgdpna      rnna    emp
## VEN-1976 1976 Venezuela (Bolivarian Republic of)  9420.365  94488.21 3.7032
## VEN-2001 2001 Venezuela (Bolivarian Republic of) 13799.385 185277.69 9.4027
##               pop       hc dif_years growth_years growth_rgdpna growth_rnna
## VEN-1976 13.57221 1.595433        NA           NA            NA          NA
## VEN-2001 24.64647 2.189336        25 0.0005030247    0.01538719  0.02730126
##          growth_emp growth_hc gdp_per_capital capital_per_capital
## VEN-1976         NA        NA        2543.845            25515.29
## VEN-2001 0.03797529 0.0127386        1467.598            19704.73
##          growth_gdp_per_capital growth_capital_per_capital
## VEN-1976                     NA                         NA
## VEN-2001             -0.0217617                -0.01028352
```

The annual growth rate for y is -2.17617%

The annual growth rate for k is -1.028352%

The annual growth rate for h remains the same at 1.27386% (we calculated on last part)

So even though K and Y are growing positively, they are not doing so per capital (by using employment as population rates). This is mostly because of a big jump in population as well, going from 3.7 million employed, to 9.4 in 25 years.

We did it above using emp as population, but lets see how these changes if we use the actual population numbers (which weren't asked for in the doc assignment). But just for further analysis

```
my_country_short_per_capital_pop <- my_country_short %>% mutate(gdp_per_capital = rgdpna/pop, capital_pe


my_country_short_per_capital_pop
```

```
##          year                             country    rgdpna      rnna    emp
## VEN-1976 1976 Venezuela (Bolivarian Republic of)  9420.365  94488.21 3.7032
## VEN-2001 2001 Venezuela (Bolivarian Republic of) 13799.385 185277.69 9.4027
##               pop       hc dif_years growth_years growth_rgdpna growth_rnna
## VEN-1976 13.57221 1.595433        NA           NA            NA          NA
## VEN-2001 24.64647 2.189336        25 0.0005030247    0.01538719  0.02730126
##          growth_emp growth_hc gdp_per_capital capital_per_capital
## VEN-1976         NA        NA        694.0923            6961.889
```

```
## VEN-2001 0.03797529 0.0127386        559.8929           7517.412
##          growth_gdp_per_capital growth_capital_per_capital
## VEN-1976                     NA                         NA
## VEN-2001           -0.008557552                0.003075562
```

then, in this case the growth rate of y is -0.8557552% while growth rate of k is 0.3075562%, so not that much change. So we can say that population grew a lot (1.8 times bigger), but employment grew to an even greater rate (2.56 times greater than 25 years ago). That is why in our last analysis, this growth rate was way less. Which is an interesting characteristic of developing economies.

8. Given your answer to (1), what must TFP growth in your country have been? You could do this problem in pencil and paper, but try to do it in R too (it is very fast at math like "2+2")

Total Factor Productivity term is At in Cobb Douglas.

So the formula for the growth rate of At that we get from the derivation in part 1 is:

g(y) = g(a) + (1-alpha)g(h) + (alpha)g(k)

g(a) = g(y) - (1-alpha)g(h) - (alpha)g(k)

The TFP growth in my country shall be:

g(a) = g(y) - (1-1/3)g(h) - (1/3)g(k) (using emp as denominator for per capital)

g(a) = g(y) - (1-1/3)g(h) - (1/3)g(k) (using pop as denominator for per capital)

```r
#create TFP variable calculation and then add it to the dataset

growth_TFP <- my_country_short_per_capital[2, "growth_gdp_per_capital"]- (2/3)*my_country_short_per_capi

print(paste0(growth_TFP*100, ' percent'))
```

```
## [1] "-1.99705795497545 percent"
```

and then using pop

```r
#create TFP variable calculation and then add it to the dataset

growth_TFP <- my_country_short_per_capital_pop[2, "growth_gdp_per_capital"]- (2/3)*my_country_short_per_

print(paste0(growth_TFP*100, ' percent'))
```

```
## [1] "-2.01255107294912 percent"
```

Now we're going to go back to the all of the years in my country Create new variables for the growth rates of y, k, h, and TFP, calling them y growth,k growth,h growth, and TFP growth.

then, we will apply this formula ton the growth rate of a in the dataset

```r
my_country_new <- drop_na(my_country %>% mutate(y = rgdpna/emp, k = rnna/emp,
y_growth = ((y/ lag(y))^(1/1))-1, k_growth = ((k/ lag(k))^(1/1))-1, h_growth = ((hc/ lag(hc))^(1/1))-1,


head(my_country_new, 10)
```

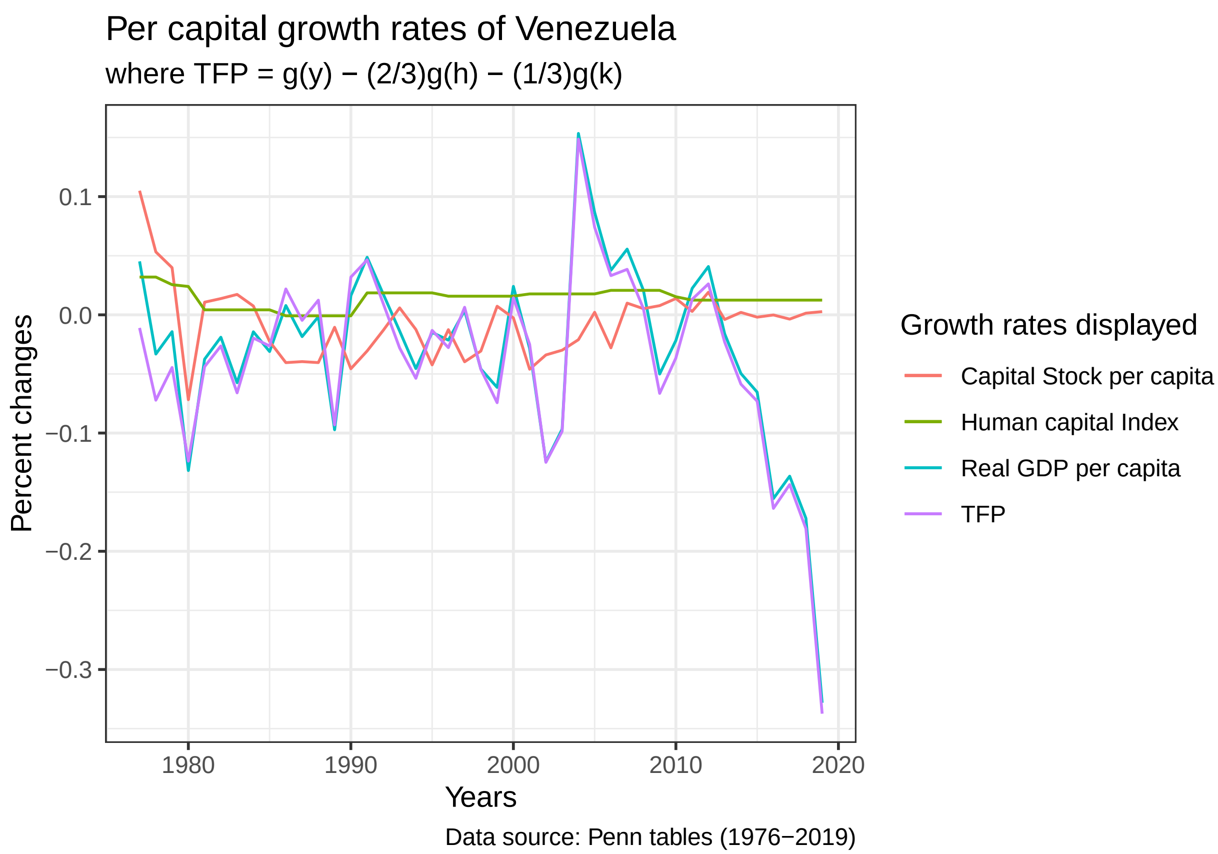
```
##          year                          country   rgdpna     rnna      emp
## VEN-1977 1977 Venezuela (Bolivarian Republic of) 10053.615 106606.1 3.781100
## VEN-1978 1978 Venezuela (Bolivarian Republic of) 10268.598 118616.5 3.994500
## VEN-1979 1979 Venezuela (Bolivarian Republic of) 10405.735 126798.9 4.106200
## VEN-1980 1980 Venezuela (Bolivarian Republic of) 10198.819 132854.2 4.635100
## VEN-1981 1981 Venezuela (Bolivarian Republic of) 10168.030 139087.8 4.801200
## VEN-1982 1982 Venezuela (Bolivarian Republic of) 10237.405 144701.5 4.927300
## VEN-1983 1983 Venezuela (Bolivarian Republic of)  9662.371 147398.2 4.934000
## VEN-1984 1984 Venezuela (Bolivarian Republic of)  9531.553 148619.8 4.938000
## VEN-1985 1985 Venezuela (Bolivarian Republic of)  9549.978 150217.3 5.106100
## VEN-1986 1986 Venezuela (Bolivarian Republic of) 10171.715 152327.3 5.395999
##               pop       hc        y        k    y_growth     k_growth
## VEN-1977 13.96438 1.646407 2658.913 28194.45  0.045233942  0.105002422
## VEN-1978 14.36473 1.699009 2570.684 29694.96 -0.033182103  0.053219786
## VEN-1979 14.77127 1.742319 2534.152 30879.86 -0.014211021  0.039902679
## VEN-1980 15.18261 1.784114 2200.345 28662.64 -0.131723402 -0.071801703
## VEN-1981 15.59789 1.791624 2117.810 28969.37 -0.037509969  0.010701574
## VEN-1982 16.01757 1.799166 2077.691 29367.30 -0.018943878  0.013736050
## VEN-1983 16.44313 1.806740 1958.324 29873.98 -0.057451572  0.017253188
## VEN-1984 16.87670 1.814345 1930.246 30097.17 -0.014338065  0.007471127
## VEN-1985 17.31952 1.821983 1870.308 29419.18 -0.031051989 -0.022526554
## VEN-1986 17.77200 1.820451 1885.047 28229.67  0.007880991 -0.040433381
##               h_growth  TFP_growth
## VEN-1977  0.0319500386 -0.01106689
## VEN-1978  0.0319499734 -0.07222201
## VEN-1979  0.0254910024 -0.04450592
## VEN-1980  0.0239882252 -0.12378165
## VEN-1981  0.0042095432 -0.04388352
## VEN-1982  0.0042095295 -0.02632891
## VEN-1983  0.0042095083 -0.06600897
## VEN-1984  0.0042095454 -0.01963480
## VEN-1985  0.0042095737 -0.02634952
## VEN-1986 -0.0008409503  0.02191942
```

```r
ggplot(data = my_country_new, mapping = aes(year)) +
geom_line(aes(y = y_growth, color = "y_growth")) +
geom_line(aes(y = k_growth, color = "k_growth"))+
geom_line(aes(y = h_growth, color = "h_growth")) +
geom_line(aes(y = TFP_growth, color = "TFP_growth"))
```

9. Plot the relationship between year and all of the growth rates. I already gave you most of the code, but I want you to make two change to the default ggplot choices. First, axis and legend names should never ever be variable names (unless the variable name is plain english like "year"). Make the names reflect the data. Second, make it so that the graph is legible when printed (I would suggest changing two things: first I would pick a different theme, getting rid of the gray background, and then I would figure out how to make the scale colors gray)

```
#make some changes to the graph


ggplot(data = my_country_new, mapping = aes(year)) +
geom_line(aes(y = y_growth, color = "Real GDP per capita")) +
geom_line(aes(y = k_growth, color = "Capital Stock per capita"))+
geom_line(aes(y = h_growth, color = "Human capital Index")) +
geom_line(aes(y = TFP_growth, color = "TFP")) +xlab("Years") + ylab("Percent changes")+ theme_bw() +  g
```



"'

10. Briefly describe the growth experience of the country you have selected. Which factor was the most important driver of growth? Was growth relatively constant over the period, or did it vary substantially?

Growth variated substantially for the country I choose. It was volatile but still in good levels before 2010. After that, political turmoil together with many other factors conducted GDP down, going down even further year by year afterwards (we can see this by the growth rate being increasingly negative as time goes by, i.e negatively steeper). TFP follows real GDP per capita rather closely due to the constantness of capital stock and human capital index, which did not vary a lot through the years. We can also see that as real GDP started to go down in recent years, capital stock did not move (growth rate stagnated at 0). So TFP growth has been virtually in the same rate as real GDP per capita growth.

Then, one could say that judging by the first half of the graph, capital stock was indeed an important driver of real gdp growth rates (capital stock being a leading indicator, and gdp a lagging indicator of growth). However, this is no longer true in the second part of the graph. Many papers have discussed about this, and one of the clearest conclusions is that after 2010, Venezuela has not been a reliable source in administrating clear accounts of capital stock (which should be in a negative rate by now, since gdp has been declining exponentially after 2010).