Solow-Swan Model

Econometrics - Project 1

Alessandro Campi(3195247), Leonardo Tonelli(3216378),
Tommaso Ravasio(3192281), Edoardo Putignano(3195342)

Data cleaning

We obtained the data from the World Bank, we selected the following variables: GDP per capita (current US$)[1], Gross savings (% of GDP)[2], Population growth (annual %)[3] and Research and development expenditure (% of GDP)[4].

- **We removed from the dataset the countries for which oil production is the dominant industry** following the decision taken by Mankiw et Al[5] (see the list of countries removed in the R code); we did this as the Solow model should not be taken into account for countries that have the most of their GDP coming from the extraction of existing resources and not adding value to them.

- **We removed all the countries with a population lower than one million** (criteria and list defined in the R script) as they present more variable and unstable data.

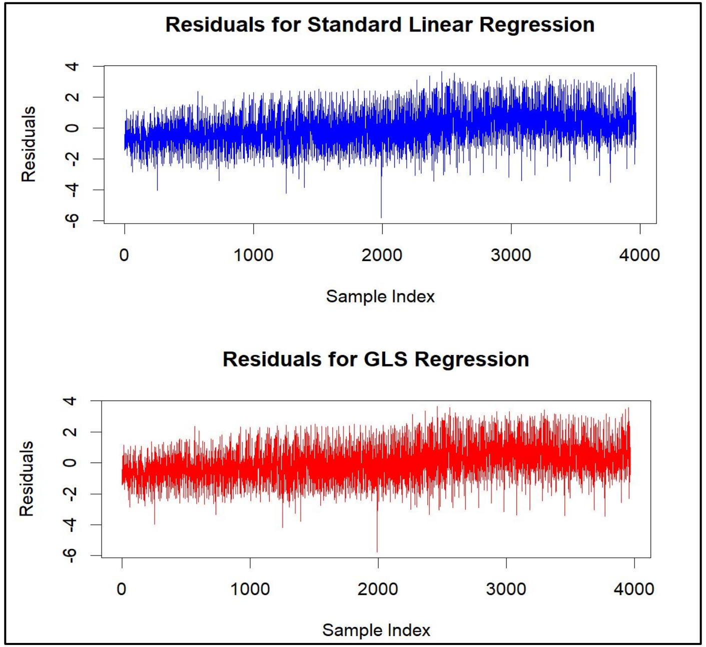
- **We removed all the observations without a specific name in the "country name" features** (list in R script), this is because there were some observations relating to generic geographical region (eg. "european union") which are obviously correlated to other countries and would have caused multicollinearity problems.

- **We removed all the observations with NA values** and all the observations that would produce a NA values when applying a logarithmic transformation, hence, we removed all negative saving rates and population growths. Even though it is not uncommon for developed countries to have a negative population growth, this did not significantly affect the number of data points available in our dataset. We find it unlikely that this change would cause problems of under-representation of developed countries in the regressions.

OLS shortcomings

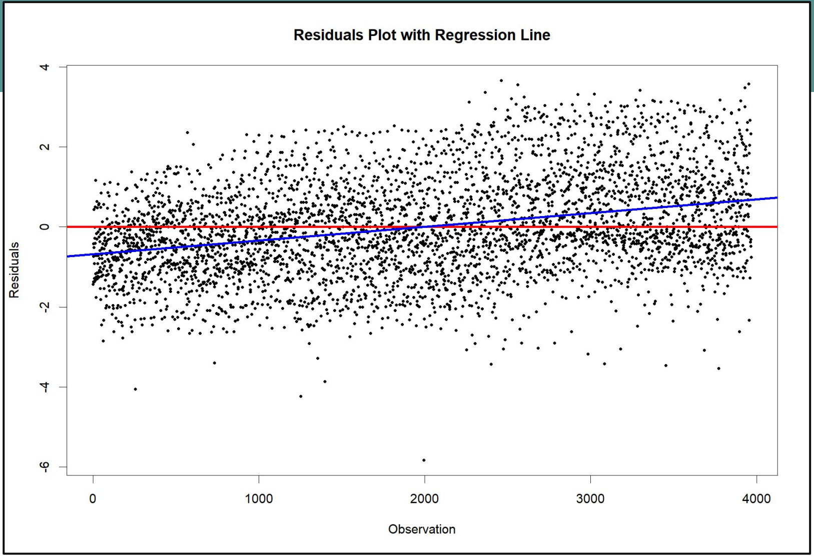
```
> bgtest(model_OLS)

	Breusch-Godfrey test for serial correlation of order
	up to 1

data:  model_OLS
LM test = 14.822, df = 1, p-value = 0.0001182
```

- We ran an **OLS** regression, and tested the assumptions with the following tests: Durbin Watson, BG, BP, Reset.
- All tests **rejected** their respective null hypothesis.
- Plotting the residuals we observed that the problem was in the <u>serial correlation of our observations</u>, due to the time-series nature of our data.
- The theory studied suggests us an **FGLS estimator** in case of serial correlation, being careful of its drawbacks (*biasedness, non-linearity, non efficiency in finite samples*).
- However, our dataset is ordered by year with multiple countries: then we **cannot** assume first order linear correlation between subsequent observations (between different countries and same year).
- As expected, using the FGLS estimator of first order serial correlation has no effect in the trend of residuals (in red).
- Therefore we will analyze the OLS estimators, **acknowledging its limitations**: not BLUE anymore because of serial correlation of residuals



OLS results

- Both variables are **statistically significant** (p-values close to 0)
- F-test is rejected
- Coefficients can be interpreted as **elasticities** since we are taking the logs of both dependent and independent variables
- *$\Delta_{\%}$ gdppc = 0.84 * $\Delta_{\%}$sav_rate, ceteris paribus.*
  *$\Delta_{\%}$ gdppc = -0.87 * $\Delta_{\%}$pop_growth, ceteris paribus*
- The Solow model predicts that a higher savings rate boosts GDP per capita, whereas a higher population growth rate diminishes GDP per capita
- OLS results yield parameters **consistent with economic theory**, at least in the direction of the associations.

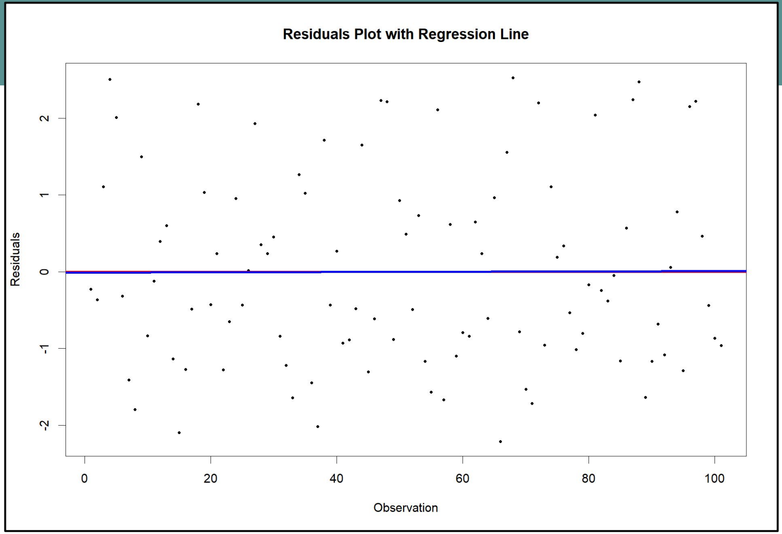



```
Call:
lm(formula = log(gdp_pc) ~ log(sav_rate) + log(pop_growth), data = data_ng)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.8259 -0.8522 -0.1368  0.8013  3.6574 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      5.62280    0.10858   51.79   <2e-16 ***
log(sav_rate)    0.83894    0.03582   23.42   <2e-16 ***
log(pop_growth) -0.86622    0.02040  -42.46   <2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.242 on 3964 degrees of freedom
  (4104 observations deleted due to missingness)
Multiple R-squared:  0.4172,	Adjusted R-squared:  0.417 
F-statistic:  1419 on 2 and 3964 DF,  p-value: < 2.2e-16
```

Single year results

- To test our assumptions on the serial correlation of the errors between years we decided to take into account data from a **single year** picked at random
- In this case we **can't reject** the assumptions of **homoscedasticity** and **no serial correlation** of the errors by Breusch-Godfrey and Breusch-Pagan tests
- Sometimes the RESET test would reject the assumption of linearity of parameters, which would suggest that some **model respecification** would be in order
- The direction of coefficients is once again consistent with economic theory. However, the betas present a **higher standard error**, which makes them less statistically significant. This is unsurprising due to the limited availability of data points

```
> bgtest(model_single)

        Breusch-Godfrey test for serial correlation of order
        up to 1

data:  model_single
LM test = 0.16582, df = 1, p-value = 0.6839
```




```
> summary(sy_solow)

Call:
lm(formula = log(gdp_pc) ~ log(sav_rate) + log(pop_growth), data = single_year)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2124 -0.9591 -0.3176  0.9290  2.5259

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       6.7365     0.7401   9.103 1.07e-14 ***
log(sav_rate)     0.6148     0.2422   2.539   0.0127 *
log(pop_growth)  -0.6468     0.1251  -5.172 1.23e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.273 on 98 degrees of freedom
  (40 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.2842,    Adjusted R-squared:  0.2696
F-statistic: 19.45 on 2 and 98 DF,  p-value: 7.682e-08
```

Research and development expenditure

We enhanced the Solow model by including '**R&D expenditure**' as a percentage of GDP, recognizing its role in driving technological progress and long-term economic growth. R&D investment fosters innovation and productivity, thus it is likely correlated **positively with GDP**.

Developed nations with lower population growths typically allocate more to R&D, while developing ones with higher population growth invest less, this suggest a potential **negative correlation between rnd_exp and pop_growth**. Savings often serve as a primary investment capital source, we expect a **positive correlation between sav_rate positively and rnd_exp.**

Our economic model:

$$ln(gdp\ pc) = a + b\ ln(s) - c\ ln(n) + d\ ln(rnd\ exp)$$

```
Call:
lm(formula = log(gdp_pc) ~ log(sav_rate) + log(pop_growth) +
    log(rnd_exp), data = data_gls_2)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9335 -0.5629  0.1494  0.6566  2.3750

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.57434    0.21229  40.389  < 2e-16 ***
log(sav_rate)    0.25164    0.06653   3.782 0.000162 ***
log(pop_growth) -0.23530    0.02639  -8.915  < 2e-16 ***
log(rnd_exp)     0.82340    0.02357  34.941  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9595 on 1360 degrees of freedom
Multiple R-squared:  0.6019,	Adjusted R-squared:  0.601
F-statistic: 685.4 on 3 and 1360 DF,  p-value: < 2.2e-16
```

```
> cor(log(data_gls_2$pop_growth), log(data_gls_2$rnd_exp))
[1] -0.3683664
> cor(log(data_gls_2$sav_rate), log(data_gls_2$rnd_exp))
[1] 0.317079
```

We run a OLS regression on the new dataset. The results we obtained were **in line with economic theory.**

- $\Delta_{\%}$ *gdppc* = ***0.2516 * $\Delta_{\%}$sav_rate***, *ceteris paribus (it went down from 0.84)*
- $\Delta_{\%}$ *gdppc* = ***-0.2353 * $\Delta_{\%}$pop_growth***, *ceteris paribus (it went up from -0.87)*
- $\Delta_{\%}$ *gdppc* = ***0.8234 * $\Delta_{\%}$rnd_exp***, *ceteris paribus*

The introduction of this new variable, lead to a drop in the coefficient of the “sav_rate” and an increase in the coefficient of the “pop_growth”, this because the new variable reduced an upward bias resulting from the positive correlation between the sav_rate and the rnd_exp and it reduced a downward bias resulting from the negative correlation between the pop_growth and the rnd_exp.

Bibliography

[1] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?view=chart (GDP per Capita)

[2] https://data.worldbank.org/indicator/NY.GNS.ICTR.ZS?view=chart (Saving rate)

[3] https://data.worldbank.org/indicator/SP.POP.GROW?view=chart (Population Growth)

[4] https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS?view=chart (R&D Expenditure)

[5] Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth.

[6] Baxter, Marianne & Crucini, Mario. (1993). Explaining Saving-Investment Correlations. American Economic Review. 83. 416-36.

[7] Marcellino, M. (2007). Applied Econometrics: An Introduction. Princeton University Press.