An analysis of the factors driving long-term economic growth

Lars Eirik Aanestad 715120290030
Natalie Malins 715120290006
David Koenig 715120290029
Nicolas Fernandez 715120290028

Exchange students

Content

An analysis of the factors driving long-term economic growth

Introduction

The aim of this project is to investigate which factors drive long-term economic growth. Throughout history, nations, companies, and humans have seized the opportunity to invest in different kinds of capital. Whether the investments materialized through humans or equipment is subsidiary to the original idea that investing, contrary to consumption, is the key to progress. This idea was theorized by Robert Solow who argued that economic growth was a function of capital accumulation, capital depreciation and a country's productivity (Jones, 2014). As a nation's capital increases, so will the cost of depreciation. Because economic theory states that investments have diminishing marginal return and depreciation cost is a fixed rate of the capital's value, Solow's model of economic growth ends in equilibrium when the return on invested capital equals the depreciation of the accumulated capital. The model predicts that nations will experience a decreasing growth rate over time, and that poor nations will converge towards richer nations through investments.

In contrast to Solow who saw productivity as an exogenous factor, Romer argued that productivity was a result of existing knowledge and allocation of labor between production and research within a nation. As Romer introduced knowledge as a non-rivalry asset, he simultaneously explained how investing in research can yield increasing marginal returns which ultimately make it possible for countries to appreciate prolonged economic growth (Jones, 2014).

Based on the theories briefly described above, we will investigate how initial return on capital explains return on capital at a later stage. Then we will expand our statistical model to include other variables as potential indicators for productivity.

The constructs that are utilized in Solow and Romer's theories are difficult to operationalize in real life. Capital accumulation is in their theories described as the sum of last year's capital and this year's investment. Measuring the total investment in a nation is difficult, but the annual return on these national investments would be the GDP, which is easier to access. Consequently, a lower GDP will be associated with lower capital accumulation, as we use absolute numbers. However, this measurement is not perfect as it does not show the return relative to the invested capital, nor does it allow us to measure growth rates. Nevertheless, as

GDP of previous years indicates available investment capital, we expect the GDP of 1960 to partially explain the growth in GDP from 1960 to 1996.

In Romer's theory productivity is described as the ratio of total labor conducting research and how efficient knowledge is developed (Jones, 2014). Again, this phenomenon is difficult to measure directly. An approximation would be looking at the education of the population, assuming that more people having access to the accumulated knowledge will influence economic growth.

A hinder to productivity would be the presence of illness. To investigate the effect of this, we can use the prevalence of Malaria, an illness that was thought to have great influence on the population numbers during the 1960s. Countries that experience greater outbreaks of Malaria are expected to suffer lower economic growth.

On the other side, exogenous shocks can promote sudden increases in investments: some countries receive gifts from heaven materialized as natural resources or beneficial changes in global demand. In economic theory, a sudden increase in income is associated with increased savings and investments. Thus, countries that find oil or other natural resources is expected to experience a burst in economic growth.

The dataset

Our data consist of observations from 182 countries on 13 variables. The dataset was obtained from the Norwegian School of Economics. Since our original dataset is written in Norwegian we renamed the variables. Our variables are as follows:

1. country – name of country
2. countrycode – 3 letter abbreviation of country name
3. lgdppc_1960 – the logarithm of the inflation adjusted GDP per capita in 1960
4. lgdppc_1996 - the logarithm of the inflation adjusted GDP per capita in 1996
5. lgddpc_2014 - the logarithm of the inflation adjusted GDP per capita in 2014
6. education_p_1960 – proportion of population in 1960 with primary education
7. education_p_1996 - proportion of population in 1996 with primary education
8. education_h_1960 - proportion of population in 1960 with higher education
9. education_h_1996 - proportion of population in 1996 with higher education
10. mines_1960 - proportion of GDP based of mining in 1960
11. oil_1960 – dummy-variable, 1 if country has oil production in 1960, 0 if not.
12. malaria_1960 – prevalence of malaria in 1960 on a scale from 0 to 1.

13. capitalism_1960 – A scale measuring capitalism. Higher value, more capitalism.

Analyzing which factors affect long-term economic growth requires data on economic growth. We therefore start of by creating columns for economic growth in the form of an arithmetic mean represented by the variables Growth_1960_1996 and Growth_1996_2014.

	lgdppc_1960	lgdppc_1996	lgdppc_2014	education_p_1960	education_h_1960	mines_1960	oil_1960	malaria_1960	capitalism_1960
count	111.000000	180.000000	182.000000	113.000000	114.000000	123.000000	126.000000	108.000000	126.000000
mean	8.078464	8.839345	9.295894	0.707522	0.031754	0.057724	0.095238	0.358718	3.373016
std	0.990250	1.244994	1.174845	0.311372	0.046435	0.091730	0.294715	0.429967	1.511207
min	5.783145	5.144375	6.345705	0.050000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.304172	7.946936	8.477570	0.460000	0.000000	0.000000	0.000000	0.000000	3.000000
50%	7.940775	8.883772	9.438845	0.830000	0.010000	0.020000	0.000000	0.046085	3.000000
75%	8.868837	9.764272	10.101928	1.000000	0.050000	0.080000	0.000000	0.867500	5.000000
max	10.260106	11.666056	12.003310	1.000000	0.320000	0.530000	1.000000	1.000000	5.000000

	Growth_1960_1996	Growth_1996_2014	education_p_1996
count	111.000000	180.000000	71.000000
mean	0.018822	0.024824	0.844991
std	0.018532	0.021610	0.176707
min	-0.030220	-0.025662	0.261370
25%	0.006020	0.011995	0.783642
50%	0.020702	0.022211	0.919949
75%	0.029158	0.034146	0.961339
max	0.069224	0.138050	0.999403

Figure 1 and 2 showing descriptive statistics of the dataset

Descriptive statistics of the dataset enables us to give an overview of our data. The dataset is far from complete and only holds 111 observations on growth. Many less developed countries are therefore left out, which can ultimately create a biased conclusion. In addition, many of the variables are quite skewed which could complicate the interpretation of the regressions and reduce the validity of the results.

Analyzing the effect of initial GDP on growth

$$\ln Growt_{1960,1996} = \beta_0 + \beta_1 lnGDPpc_{1960}$$

In the footsteps of Solow we start off by looking at capital accumulation's effect on growth. We do so by conducting a linear regression with Growth_1960_1996 as dependent variable and lgdppc_1960 as independent variable. The regression is coded utilizing the package "statsmodels".

OLS Regression Results

Dep. Variable:	Growth_1960_1996	R-squared:	0.022
Model:	OLS	Adj. R-squared:	0.013
Method:	Least Squares	F-statistic:	2.431
Date:	Tue, 12 Jun 2018	Prob (F-statistic):	0.122
Time:	13:37:12	Log-Likelihood:	286.92
No. Observations:	111	AIC:	-569.8
Df Residuals:	109	BIC:	-564.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0035	0.014	-0.243	0.808	-0.032	0.025
lgdppc_1960	0.0028	0.002	1.559	0.122	-0.001	0.006

Omnibus:	7.534	Durbin-Watson:	2.053
Prob(Omnibus):	0.023	Jarque-Bera (JB):	7.278
Skew:	0.517	Prob(JB):	0.0263
Kurtosis:	3.709	Cond. No.	68.2

Figure 3: Regression results model 1

The interpretation of the regression is that 1% increase in GDP will increase the growth rate with 0.0028. Results show a R2-value of 2,2% and a p-value for GDP of 12,2%, suggesting that initial GDP in 1960 is not by itself able to explain the variance in growth between countries. Empirically we have seen that some nations experience lasting economic growth while others suffer little growth or even stagnation. Return on capital at a previous point in time does not explain future growth in return on capital alone, and the simple version of Solow's theorem does not hold empirically.

Robustness Analysis 1

Linear regression is based on five assumptions: 1) there is a linear relationship between the dependent and independent variable, 2) the residuals are normally distributed, 3) no or little mulitcollinearity, 4) no autocorrelation, and 5) homoscedasticity.

A scatter plot of GDP towards growth indicates that there is some linear relationship between the variables, which is that richer countries have a lower (and more similar) growth rate than poorer countries. A plot of the sample residuals versus the theoretical residuals, a Q-Q plot in combination with a large sample (111 obs.) indicate that the assumption of normality holds. Multicollinearity, which

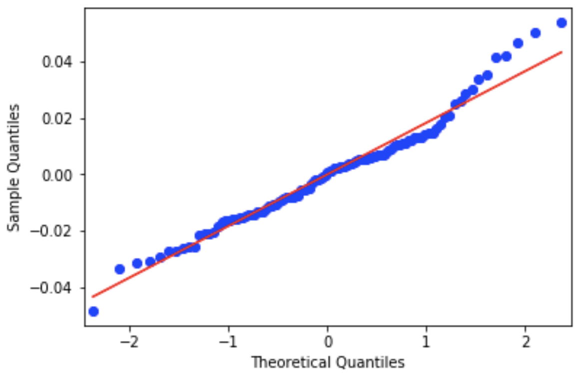

Figure 4: Q-Q plot

is a measure of correlation between the independent variables, will not be a problem while only having one independent variable. Autocorrelation will also not be a problem in our analysis as the fluctuations in the growth have been smoothed out and presented as an average for the combined years.

A Breusch-Pagan test indicates that we have problems with heteroskedasticity in our model. Heteroskedasticity means that the residuals do not have constant variance. This will influence the standard error and ultimately the precision of our model. There is also a concern that the growth that we are trying to explain is ultimately explaining the initial GDP. Overall it seems to be a poor model explaining the drivers of economic growth.

Productivity as a supplement to the Solow-model

Initial GDP as the only independent variable gave unsatisfactory results.

Productivity, the efficiency of turning input into output, has the potential to explain why some countries apparently are able to refine their starting capital more than other countries. We add education variables as proxy for competence to our original model to get a more nuanced picture of the drivers of long-term growth.

$$\ln Growth_{1960,1996} = \beta_0 + \beta_1 lnGDPpC_{1960} + \beta_2 education_p + \beta_3 education_h$$

We create a new data frame including data on GDP, growth and higher and lower education. This is in order to keep as many observations as possible through to the regression state. A regression with the original data frame would exclude all those countries that have missing values, also for those variables that are not tested in our regression. We will therefore make a new data frame for every regression.

Dep. Variable:	Growth_1960_1996	**R-squared:**	0.417
Model:	OLS	**Adj. R-squared:**	0.399
Method:	Least Squares	**F-statistic:**	22.88
Date:	Sun, 10 Jun 2018	**Prob (F-statistic):**	2.93e-11
Time:	22:08:53	**Log-Likelihood:**	285.21
No. Observations:	100	**AIC:**	-562.4
Df Residuals:	96	**BIC:**	-552.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0514	0.015	3.453	0.001	0.022	0.081
lgdppc_1960	-0.0086	0.002	-4.042	0.000	-0.013	-0.004
education_p_1960	0.0502	0.007	7.670	0.000	0.037	0.063
education_h_1960	0.0410	0.039	1.041	0.301	-0.037	0.119

Omnibus:	13.155	**Durbin-Watson:**	2.131
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	15.404
Skew:	0.726	**Prob(JB):**	0.000452
Kurtosis:	4.260	**Cond. No.**	234.

Figure 5: Regression result model 2

The regression returns some interesting results. GDP is now significant with a p-value of 0,000. Also primary education has turned out significant (p-value of 0,000), while higher education is far from significant at a 5% significance level. With the inclusion of education we also see that initial GDP has a negative effect on the expected growth rate. This supports the assumption of diminishing returns on investments, and Solow and Romer's idea that poor countries will converge towards rich ones. Education also helps to explain more of the variance in the data with the R2 climbing up to 41,7%. Even though the regression supports our assumption that both GDP and education are factors that can explain long-term economic growth one should be aware that the distribution for primary education is skewed. There is for example a group of countries that is remarkably little educated (8 countries with primary education less than 20%). Consequently, there may be other variables that could explain both the growth and the level of primary education in a country, variables that are not included in this dataset.

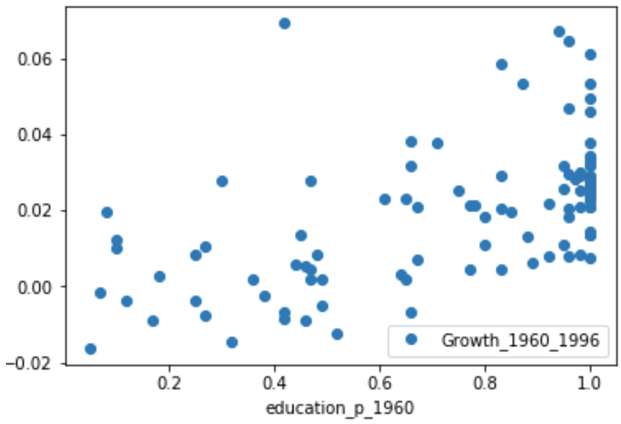

Figure 6: Primary education vs. Growth

Robustness analysis 2

Scatter plots of education versus growth back the strong linear relationship between primary education and growth. Our residuals seem to be slightly non-normal, but this is considered a minor issue as our sample is of some size (100). There are reasons to suspect that the proportion of the population having completed primary education would affect the proportion of the population moving towards higher education, and therefore causing problems with multicollinearity. However, variables affecting each other are quite normal and it does not seem to be an extreme trend visible when plotting the two against each other. Another issue with the regression is that USA has a proportion of their population with higher education of over 30%. This seems unreasonable. Excluding the US from the dataset when running the model show that the USA has a noticeable influence on the coefficients.

While a Durbin-Watson test implies that our model does not have a problem with auto-correlation, a Breusch-Pagan test once again implies that we have a problem with heteroskedasticity. The Breusch-Pagan test is conducted with the help of 'statsmodels.stats'-package.

Including symptoms of luck and failure

Our previous regression returned results that support a theory where education and capital accumulation is part of explaining differences in growth rates between countries. Even though this brings us closer to understanding which factors drive long-term economic growth, there is still a lot of unexplained variance in our model. We expand our model to include illness as a hinder to long-term growth and possible exongenous shocks that are thought to boost future GDP through a sudden increase in investments.

Regression including all our variables

$$\ln Growt_{1960,1996} = \beta_0 + \beta_1 lnGDPpC_{1960} + \beta_2 education_p + \beta_3 education_h \\ + \beta_4 oil_{dummy} + \beta_5 malaria_{1960} + \beta_6 capitalism_{1960} + \beta_7 mines_{1960}$$

To try to account for a wider specter of economic factors we run a regression including all of our current variables. 4 of the 7 variables turn out significant.

We remove the insignificant variables and run a regression with the remaining ones.

$$\ln Growt_{1960,1996} = \beta_0 + \beta_1 lnGDPpC_{1960} + \beta_2 education_p + \beta_5 malaria_{1960} + \beta_7 mines_{1960}$$

Dep. Variable:	Growth_1960_1996	**R-squared:**	0.492
Model:	OLS	**Adj. R-squared:**	0.469
Method:	Least Squares	**F-statistic:**	21.56
Date:	Mon, 11 Jun 2018	**Prob (F-statistic):**	1.84e-12
Time:	00:40:56	**Log-Likelihood:**	276.10
No. Observations:	94	**AIC:**	-542.2
Df Residuals:	89	**BIC:**	-529.5
Df Model:	4		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0725	0.015	4.891	0.000	0.043	0.102
lgdppc_1960	-0.0097	0.002	-5.090	0.000	-0.014	-0.006
education_p_1960	0.0408	0.007	6.145	0.000	0.028	0.054
mines_1960	0.0493	0.019	2.586	0.011	0.011	0.087
malaria_1960	-0.0182	0.004	-4.075	0.000	-0.027	-0.009

Omnibus:	1.593	**Durbin-Watson:**	2.097
Prob(Omnibus):	0.451	**Jarque-Bera (JB):**	1.508
Skew:	0.304	**Prob(JB):**	0.471
Kurtosis:	2.874	**Cond. No.**	117.

Figure 7: Regression result

All of the variables turn out significant, mines at a 5% significance level while primary education, malaria, and initial GDP turn out significant at a 0.1% significance level. A plot of mines versus growth make us aware of Botswana. With over 50% of its GDP and with a growth rate of over 6% over a 36-year period we suspect that Botswana is a major influence in the significance of the mining variable.

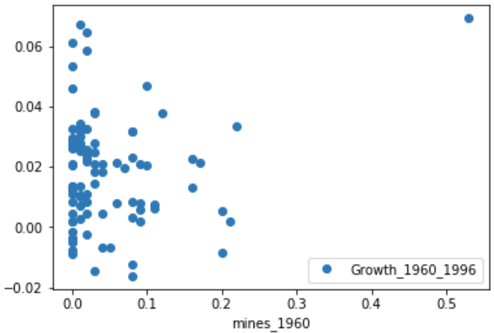

Figure 8: Mining vs. Growth - exposing Botswana

A new regression run without the presence of Botswana confirms this suspicion – mining is no longer a significant variable.

The oil-dummy turns out insignificant. Oil-producing countries would get their natural resources accounted for in their GDP and would maybe be expected to have a higher value of GDP in 1960 than their non-producing counterparts, and that effect of oil could be caught by the effect of GDP. A regression excluding GDP does however not turn the oil dummy significant.

Some theories like the 'Dutch Disease' illustrate how a sudden increase in capital can cause rapid changes in demand of public goods, inflation and subsequent painful restructuring of a nation's economy causing dramatic fluctuations in growth. There is also empiricism that shows how a sober approach of including natural resources into the economy can lay the foundation of prolonged growth, exemplified by Norway and their sovereign wealth fund. Even thought there are many scenarios where oil can affect a country's growth, our model is not able to observe this effect.

Our final and most important finding in this regression is the negative influence the prevalence of malaria apparently has had on a country's ability to grow. The significance of malaria immediately raises the question of what it is about countries placed in the habitat of the anopheles gambiae - more commonly 'malaria mosquito'. Is it the mosquito that denies nations to reach their full economic potential, or is it a symptom of a more serious underlying issue? To acknowledge the fact that it might not be the mosquito that is the sole denier of economic growth in the equatorial latitudes we create a dummy variable that has the potential to differentiate between the states that handle the malaria mosquito better and worse. By putting a threshold at 85% prevalence of malaria for the dummy-variable (value of the variable is 1 if value of malaria >0,85, 0 if not) we hope to catch an underlying effect of poor governance or some other unfortunate state(war, famine, AIDS make people more susceptible to malaria e.g.).

The regression with the dummy-variable has an adjusted R2 of 49,3% with all 3 variables significant at a 0,1% level. Other regressions also showed promising results with adjusted R2 above 50% when Botswana was excluded. We decide to not exclude outliers, as there may be circumstances and variables outside of this dataset that can shed light on their behavior.

Dep. Variable:	Growth_1960_1996	**R-squared:**	0.508
Model:	OLS	**Adj. R-squared:**	0.493
Method:	Least Squares	**F-statistic:**	33.03
Date:	Thu, 14 Jun 2018	**Prob (F-statistic):**	9.36e-15
Time:	23:39:16	**Log-Likelihood:**	293.69
No. Observations:	100	**AIC:**	-579.4
Df Residuals:	96	**BIC:**	-569.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0652	0.013	5.054	0.000	0.040	0.091
education_p_1960	0.0436	0.006	7.065	0.000	0.031	0.056
dummy_malaria_1960	-0.0159	0.004	-4.363	0.000	-0.023	-0.009
lgdppc_1960	-0.0091	0.002	-5.093	0.000	-0.013	-0.006

Omnibus:	6.472	**Durbin-Watson:**	2.183
Prob(Omnibus):	0.039	**Jarque-Bera (JB):**	5.935
Skew:	0.508	**Prob(JB):**	0.0514
Kurtosis:	3.628	**Cond. No.**	83.8

Figure 9: Regression result for model with malaria as a dummy

Robustness Analysis 3

Analyses of the assumptions for multiple regression for the regressions with the 3 and 4 variables are almost identical. The Q-Q plot looks ok, and accounting for the sample size of 100 observations we conclude that the residuals are satisfactory. The Durbin-Watson test for auto-correlation signals no auto-correlation while the Breusch-Pagan test for heteroskedasticity concludes that the residuals do not have constant variance, which ultimately reduces the precision of the model.

Our model's ability to predict growth from 1996 to 2014

Our regressions have given us a lot of insight into the factors that might influence a country's chance of appreciating prolonged economic growth. Even though we have looked at a handful of different variables we thought it would be interesting to evaluate how our model would predict growth in the 18-year period from '96 to '14.

Our dataset did not include any observations of the malaria situation in 1996, we were therefore forced to find other data covering this. World Health Organization (WHO) provided us with data for the total number of incidents for different countries. The data was incomplete, and was mostly covering less developed countries in the danger zone of malaria. To be able to have a minimum foundation to base the comparison on we added data on countries that were obvious cases in our dummy-test of malaria exceeding 85%. This method has its obvious flaws of biasedness but it was, as we saw it, our only way to be able to evaluate our model's ability to predict growth.

Furthermore we downloaded a dataset on countries' population in '96 in order to calculate a ratio of malaria incidents in a population. The original data on malaria prevalence cannot be properly compared to our new ratio. However, we believe that comparing their respective dummy-variables is a step toward diminishing the differences and thus is our best possible option. These datasets were merged using the 'panda'-package, opening up for numpy's array division to create a new column containing our calculated malaria ratio.

In order to predict future growth we ran two regressions with our preferred model with dummy-malaria, GDP, and primary education. One regression was based on a sample of the 100 countries that had observations for the given variables in 1960, while the other regression was based on the 43 countries that had observations on the variables in both 1960 and 1996. These two regressions returned different coefficients.

We then predicted the growth rate by multiplying the coefficients with each country's respective data. This was again solved by utilizing numpy's arrays package. The growth rate was compared with the actual growth and tested using Wilcoxon's Signed Rank test (statsmodels). The null hypothesis for the test is that the difference between the pairs follows a symmetric distribution around 0, while the alternative hypothesis is that the difference between the pair does not follow a symmetric distribution around 0. The null hypothesis was rejected in both cases, on a significance level of 1% for the model based on the smaller sample, and on a significance level of 0,1% for the model based on the larger sample. In both cases our models fail to give any good predictions for the future. This was expected, as there is a lot of unexplained variance and few of our variables that manage to explain it.

Robustness analysis 4

As our residuals have suffered from non-normality and a lack of constant variance we could not test the difference in means between the predicted growth and the actual growth using the student's t-test. We therefore had to go for a non-parametric alternative. Wilcoxon signed rank test seemed like a good option as it checks whether the samples come from the same distributions or not. The test makes clear that the distributions of our predicted growth and the actual growth are significantly different.

Conclusion

In this project we have analyzed which factors drive long-term economic growth. We have done this using statistical techniques. With the help of macroeconomics, especially the Solow-model and the extended Solow-Romer model we have tried to give an explanation to why some factors have turned out significant while others have not. We have seen that inital GDP is unable to explain the growth in GDP between 1960 and 1996. Furthermore a lot of the growth that we have appreciated the last years may be due to Romer's productivity variable, a variable that is represented in our model by primary education. Our last finding is that capital accumulation and productivity have the possibility of explaining some of the variation in growth rates between countries, but fail to explain the growth rate in those countries that suffer from other issues such as widespread diseases. Finally it is important to emphasize that our regressions do not prove anything, but are merely a tool to discover possible links between variables. The subject of econometrics is complex and we acknowledge that there are nuances of the subject that we lack a deeper understanding of, nonetheless our project has enabled us to gain valuable insight into the subject and especially how to practically implement statistics into Python.

Bibliography

Jones, Charles I. (2014). Macroeconomics (3. Edition). New York: W. W. Norton & Company, Inc.