The Model

The Solow-Swan Model analyzes changes in the level of output in an economy as a result of changes in the population growth rate and the savings rate.

$$\ln gdppc_t = \alpha + \beta \ln s_t - \gamma \ln n_t$$

With positive parameters, a higher savings rate (s_t) boosts capital accumulation, increasing GDP per capita (*gdppct*), while a higher population growth rate (n_t) dilutes capital per worker, reducing GDP per capita. To test whether the data support the Solow-Swan model's predictions, we sourced economic indicators from the World Bank, selecting variables that align with the model's framework:

- **GDP per capita** (current US$) – Measures economic output per person, adjusted for inflation. This serves as the dependent variable in our regression.
- **Gross domestic savings** (% of GDP) – Represents the savings rate, calculated as GDP minus total consumption, indicating the proportion of income saved by individuals and governments in a country.
- **Population growth** (annual %) – Captures the exponential rate of population increase, considering all residents, regardless of legal status or citizenship.

The dataset was originally in wide format, with years as separate columns. To enable regression analysis, we converted it into long format (country-year observations) and merged the variables into a unified dataset.

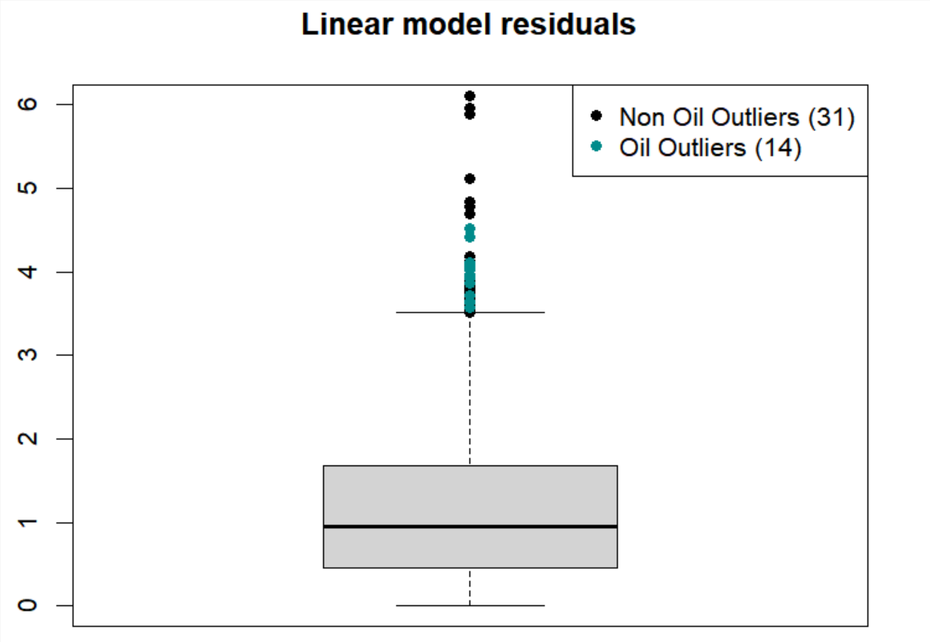
Data Cleaning

The resulting dataset is panel data with yearly observations for each country going from 1960 to 2023. However the model excludes country-specific and time-specific effects, therefore we treat our dataset as a pooled cross-sectional dataset, in which each country-year is treated as an independent observation.

Since the dataset included regional aggregates like 'Europe,' we excluded all non-country entries to avoid potential bias from repeated data, which could lead to over-representation and distort the analysis.

Traditional models like Solow's are not well-suited to explaining growth in oil-producing nations, whose economies rely heavily on resource extraction rather than value-added production. Moreover, the concentration of wealth in the hands of this minority further distorts GDP per capita figures, making their growth dynamics inconsistent with standard models [1].

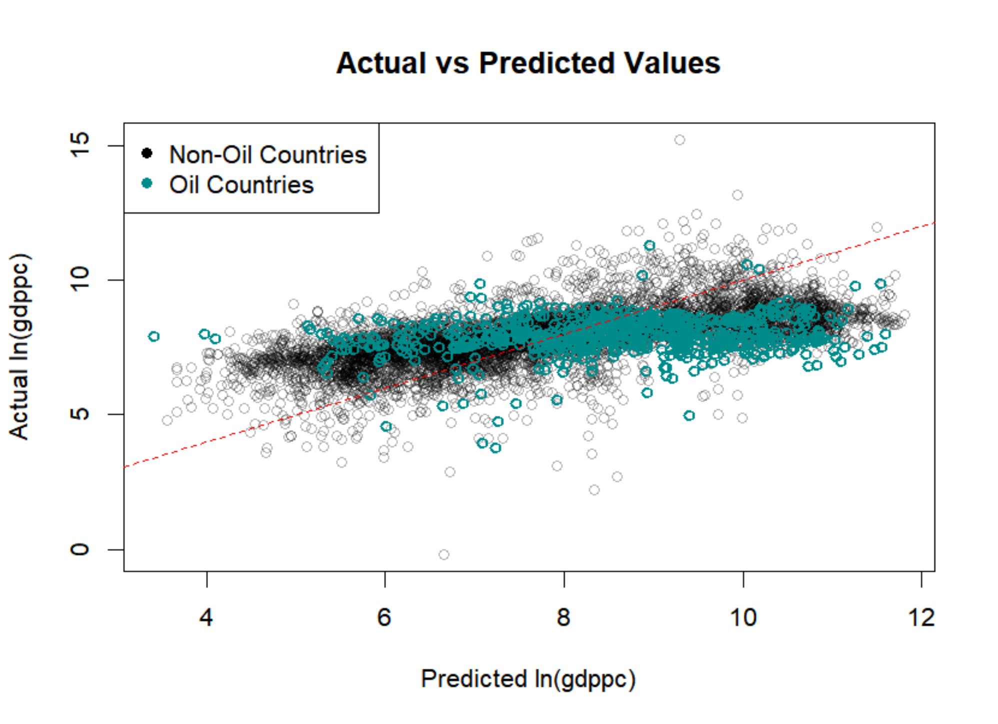
Therefore, after running an initial regression on the cleaned dataset, we conducted a residual analysis, which revealed, as expected, that oil-producing nations exhibited significant outlier behavior. In particular, as showed in the boxplot, a third of the detected outliers were oil producers, despite these countries accounting for only about 10% of all our observations (extensive calculations in the R script). To ensure a more representative analysis, we performed a second regression excluding major oil-producing nations [2].



[1] Mankiw, N.G., Romer, D. and Weil, D.N., 1992. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, *107*(2), pp.407–437.

[2] Angola, Azerbaijan, Bahrain, Brunei, Algeria, Gabon, Iran, Iraq, Kazakhstan, Kuwait, Libya, Nigeria, Oman, Qatar, Saudi Arabia, South Sudan, United Arab Emirates, Venezuela, Yemen.

Empirical Results

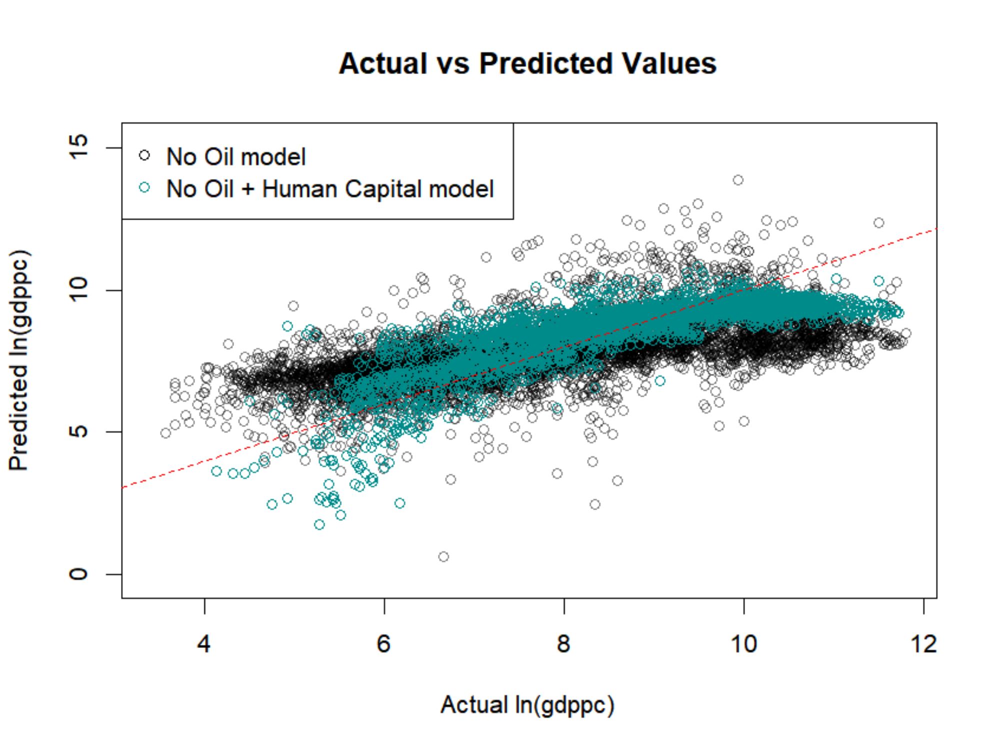


As showed in Table 1 (In slide 5), excluding oil producers reduces the intercept from 6.154 to 5.359, reflecting a lower baseline log GDP per capita in the absence of oil-rich countries, which are often wealthier due to resource extraction rather than capital accumulation. The savings coefficient decreases slightly from 0.861 to 0.745, while the population growth coefficient becomes more negative from -0.745 to -0.863 confirming that oil economies distort the relationship between these regressors and GDP per capita. However the coefficients are consistent with the Solow-Swan model hypothsis that higher savings and lower population growth lead to higher GDP per capita. Moreover, both models have coefficients that are statistically significant (p-value < 2.2e−16), but excluding oil-producing countries improves the model fit, as evidenced by the increase in R^2 from 0.36 to 0.40, as the graph shows, the model fits better for non-oil countries.

Consistent with the model, we applied a log transformation and excluded observations with negative or zero values for growth and saving rates, introducing a selection bias. Countries excluded due to negative savings had an exp(mean log GDP per capita) of 1675.24 , while those excluded for negative population growth averaged 4025.02 . This has two key implications: in poorer countries, the exclusion of negative savings values leads to an overestimation of their average savings rate, which in turn inflates their predicted GDP per capita. Conversely, in wealthier countries, the removal of negative population growth values results in an overestimated average population growth rate, which causes their predicted GDP per capita to be underestimated. This result is also shown in the graph above.

Model Extension - Human Capital

The relatively low explanatory power of the previous models, which is reasonable given the simplicity of the Solow framework, suggests the omission of key GDP determinants and underscores the need for an augmented model that incorporates factors such as human capital.
In our extended model we added as a regressor the **log-transformed secondary education,** which has a significant impact on productivity and long-term growth. To capture this effect, we introduce as human capital proxy the net enrollment rate data for secondary education from the World Bank, which accurately reflects the proportion of children enrolled at the appropriate age in secondary education, which equippes students with advanced skills essential for higher education and workforce integration.



Unlike gross enrollment ratios, which can overestimate educational attainment by including overage and underage students due to late entry, grade repetition, or variations in education systems, net enrollment rates provide a more precise measure of system coverage and efficiency.
Moreover, since individuals who attend secondary education have almost certainly attended primary school, the independent effect of primary education becomes negligible, making it redundant in the regression.
As the graph clearly shows, introducing a proxy for human capital leads to an overall better estimation, supported by an adjusted R^2 growing from about 0.40 to 0.64.

Extended Results

As before, the log transformation excludes poorer countries with low education and negative savings, as well as advanced economies with high education but declining populations. This selection bias potentially distorts the education-economic growth relationship by underrepresenting both extremes.

Nonetheless, including human capital, R^2 increased to 0.638, indicating that education plays a crucial role in explaining the variation in GDP per capita.
In all the regressions, the coefficients are all statistically signficiant, and being a «log-log» model, they represents the elasticities. For example the estimated coefficient for ln(educ_sec) is 1.63, suggesting that, ceteris paribus, a 1% increase in net enrollment rate data for secondary education is associated with an approximate 1.63% increase in GDP per capita. This underscores the importance of investing in education for long-term economic growth.

Table 1:

	Dependent variable:		
		ln_gdppc	
	Model All	Model No Oil	Model H
	(1)	(2)	(3)
ln_s	0.861***	0.745***	0.321***
	(0.021)	(0.022)	(0.030)
ln_n	−0.745***	−0.864***	−0.284***
	(0.017)	(0.018)	(0.022)
ln_educ_sec			1.626***
			(0.037)
Constant	6.154***	5.359***	8.616***
	(0.086)	(0.094)	(0.114)
Observations	6,914	6,199	2,931
R^2	0.358	0.396	0.638
Adjusted R^2	0.358	0.396	0.638
Residual Std. Error	1.402 (df = 6911)	1.374 (df = 6196)	0.917 (df = 2927)
F Statistic	1,928.565*** (df = 2; 6911)	2,032.340*** (df = 2; 6196)	1,721.278*** (df = 3; 2927)

Note: *p<0.1; **p<0.05; ***p<0.01

Moreover, due to high correlation between human capital and the other regressors, once human capital is added to the model, it captures part of the effect on GDP per capita that was previously attributed to saving rates and population growth. This re-allocation of explanatory power, often referred to as correcting for omitted variable bias, leads to a reduction in the absolute values of the coefficients for these variables.