The Drivers of GDP Across Countries

Authors:

Zaid Alotel (468411) & Abdurehman Yimer (448665)

6/6/2024

Contents

Introduction

The process of economic growth in developing countries is a complex-wide-ranging phenomenon. The Solow-Swan model provided by Robert Solow and Trevor Swan introduced the elementary model of sources of growth in income per capita back in 1956. The main idea of this model explains the difference between two primary sources of economic growth – capital accumulation and technological improvements. Capital accumulation is known as an endogenous process – it is elicited by investments inside the economy. Propelled by technological improvements of the capital, easy access to credit, and cheap investments, the economy creates a fertile environment for capital accumulation. Every increase in income or a productive lift in the country signals the development of physical capital and techniques that raised its efficiency in the previous period. Technological improvements are identified as exogenous to economic activities. They may occur due to the advancement and innovations not connected to economic processes. The approach of duality is used to spotlight the significance of both capital investment and technological progress for the development of the economy.

However, even with the essential theoretical clues given by the Solow-Swan model, the real-world implementation in the scenario of developing nations appears to be challenging. Developing economies are characterized by several limitations, such as restricted access to capital, uneven levels of technology utilization, and diverse advancement rates in human capital. Therefore, a more profound empirical analysis in required to determine what genuinely leads to GDP growth in developing countries. Knowing the most critical growth factors aids policy makers in developing a unique approach suited specifically for developing countries’ conditions and demands

This research aims to test the following hypotheses concerning the drivers of GDP growth in developing countries:

Technological Progress: Improvements in technology, measured as changes in total factor productivity (TFP), significantly contribute to GDP growth.
Capital Accumulation: Gross capital formation, which includes investments in physical capital, has a positive impact on GDP growth.
Human Capital Formation: The accumulation of human capital, measured by the average years of schooling and returns to education, positively affects GDP growth.

Share of Gross Capital Formation: The share of gross capital formation in GDP (csh_i) is a significant determinant of GDP growth.

This study is intended to identify and to quantify, through empirical study methodologies, the main causes for real per capita GDP movements in developing countries. By testing these hypotheses, the study hopes to provide insight into that there are contributing factors--for example technological progress, the accumulation of capital and human capital formation--all cumulatively enhance economic growth. Knowledge of this kind is crucial for policy makers who are seeking well-conceived economic strategies and policies that will promote sustainable development and growth--in other words instances where vested interests can be inadvertently sabotaging the country's long-term interests. In the end, this article hopes its analysis will yield insights into helpful economics: how to bring about a better life for people and simultaneously provide the conditions under which that happens.

Literature Review

Various factors determine economic growth in developing countries. This issue is addressed in the empirical research "Drivers of Economic Growth in Developing Countries" by Dao and the theoretical research "GDP and Tax Revenues-Causality Relationship in Developing Countries: Evidence from Palestine" by Iriqat and Anabtawi. Dao investigates the role of technological progress, gross capital formation, and human capital in per capita GDP growth and other factors in 38 developing economies from 1995 to 2010. As for Iriqat and Anabtawi, they measure the cyclical behaviour in Palestine of the two main variables in the aggregate national economy – tax revenues and GDP in Palestine. They base their theory on the Cochrane-Orcutt algorithm. Data do not reject the null hypothesis in the structural equation model. Therefore, it may not work correctly.

On the contrary, Iriqat and Anabtawi investigated the relationship between the GDP components and tax revenues in Palestine during the period from 1999 to 2014 and found that tax revenues do not cause the GDP, government spending, consumption, investment, and balance of trade. While these components are crucial for the fiscal policy, they are not the means to achieve economic growth. The authors refer to increasing taxes and investment facilitation and collectability, as well as streamlining government spending to be the relevant policy recommendations. Both studies prove the importance of technological progress, capital accumulation, and human capital for economic growth. At the same time, they underline the complexity of the relationships and the necessity of certain fiscal measures on the side of the developing countries.

The Data

The data section will explain the chosen variables from the dataset we are studying, the total number of variables was 52, but 14 was chosen for our study which is shown below:

The Variables:

Continuous Variables

Continuous variables are numerical variables that can take on an infinite number of values within a given range. For this dataset, the following columns are continuous:

1. pop: Population (in millions)
2. emp: Number of persons engaged (in millions)
3. avh: Average annual hours worked by persons engaged
4. hc: Human capital index, based on years of schooling and returns to education
5. xr: Exchange rate, national currency/USD (market+estimated)
6. ccon: Real consumption of households and government, at current PPPs (in mil. 2017US$)
7. ctfp: TFP level at current PPPs (USA=1)
8. rtfpna: TFP at constant national prices (2017=1)
9. labsh: Share of labour compensation in GDP at current national prices
10. pl_con: Price level of CCON (PPP/XR), price level of USA GDPo in 2017=1
11. pl_gdpo: Price level of CGDPo (PPP/XR), price level of USA GDPo in 2017=1

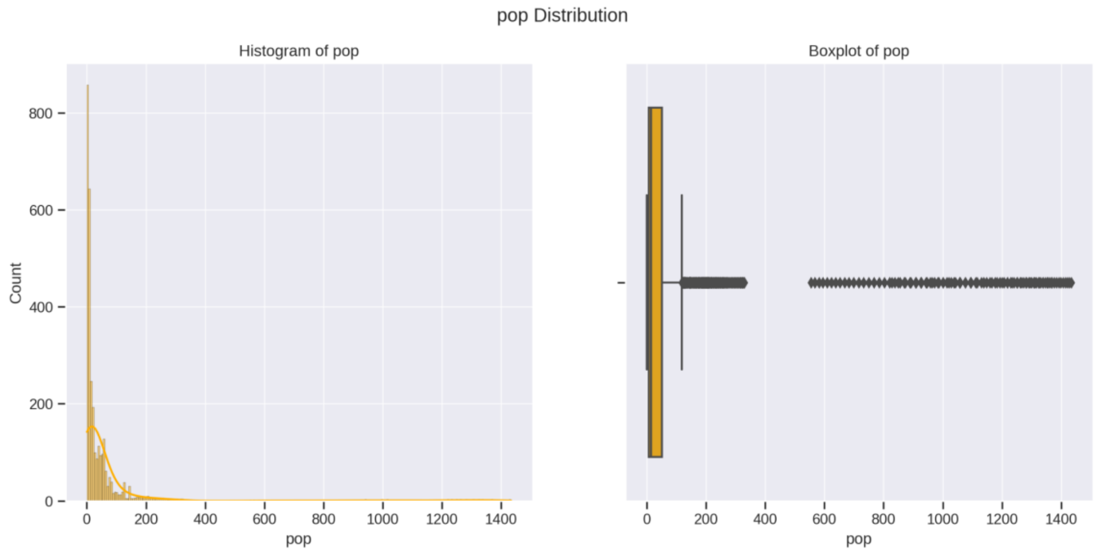
Discrete Variables

Discrete variables are numerical variables that have a countable number of values. For this dataset, the following columns are discrete:

1. csh_c: Share of household consumption at current PPPs
2. csh_i: Share of gross capital formation at current PPPs
3. csh_x: Share of merchandise exports at current PPPs

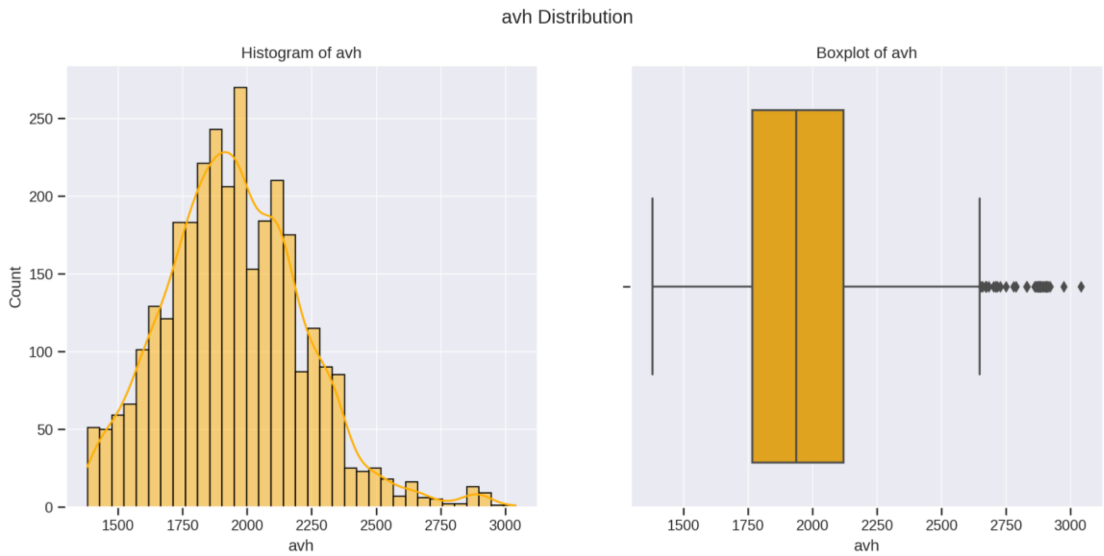
Data Visualization:

Population Analysis:



Histogram: The population data reveals a right-skewed distribution, which means most developing countries have populations below a certain threshold. However, there are a few countries with significantly larger populations, creating an imbalance in the overall data.

Boxplot: The boxplot clearly shows outliers with extremely high population values. This emphasizes the stark contrast in population sizes among developing countries, with a few countries having much higher populations than the rest.



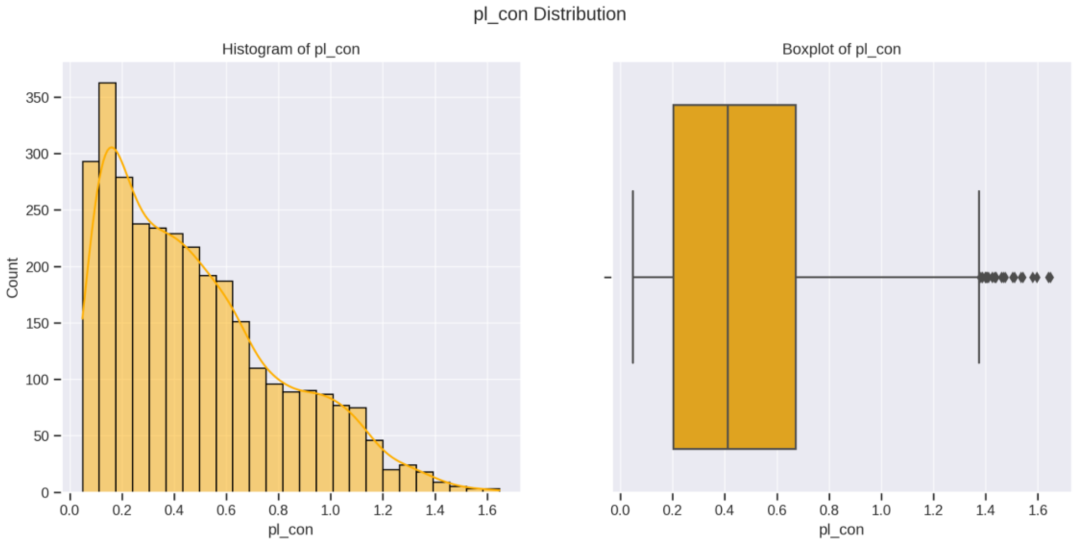
Average Annual Hours Worked (avh)

Histogram:

The histogram shows that most countries have a similar number of average annual hours worked, forming a bell-shaped curve around the average.

Boxplot:

The boxplot indicates that while there is some variation in the average annual hours worked across different countries, there are very few countries that have significantly higher or lower values than the rest.

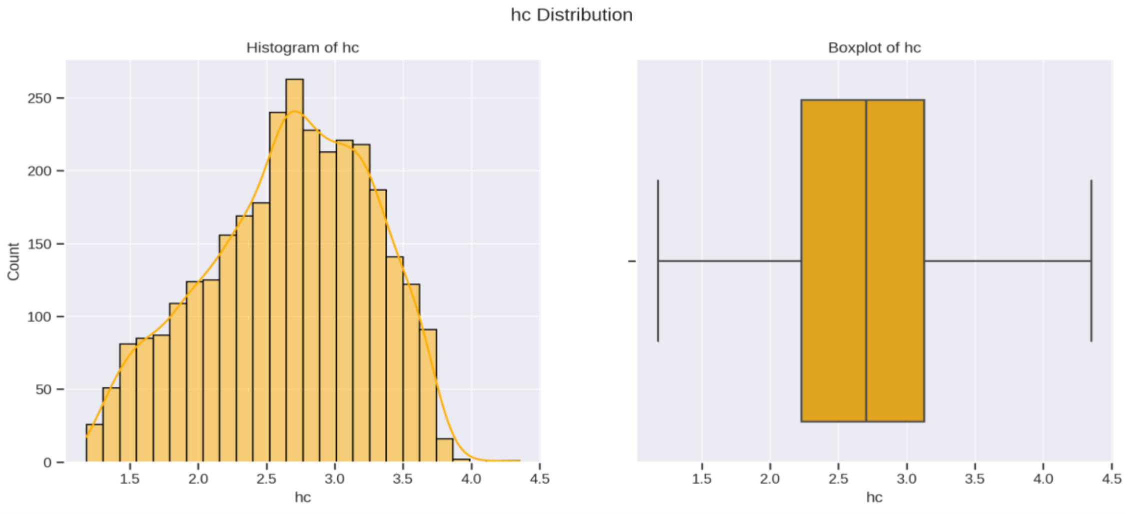


Price Level of CCON (pl_con):

Histogram: The distribution shows some variability with a right-skewed pattern.

Boxplot: Some outliers are present, reflecting countries with higher price levels.

Interpretation: Price levels impact purchasing power and cost of living. Higher price levels can reduce real income and consumption, affecting GDP growth.



Human Capital Index (hc)

Histogram:

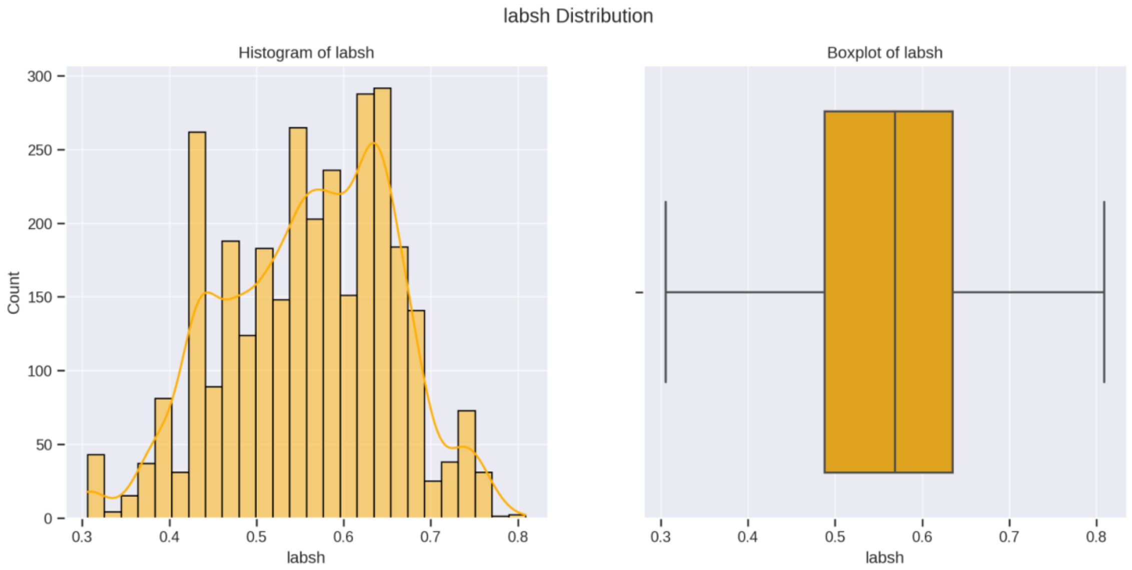
The histogram illustrates that the human capital index data is spread out in a bell-shaped curve, suggesting that the levels of education and skills vary widely among developing countries.

Boxplot:

The boxplot reveals a fairly even distribution of human capital indexes, with only a few countries having values that are much higher or lower than the rest.

Interpretation:

Human capital is a crucial factor in driving economic growth. Countries with higher human capital indexes generally have more educated and skilled workforces, which boosts productivity and leads to greater GDP growth.



Share of Labor Compensation (labsh):

Histogram:

The histogram shows that the share of labour compensation is fairly normally distributed, meaning most countries fall around the average.

Boxplot:

The box plot indicates some variation in the share of labour compensation across different countries, but there are only a few countries with values that are much higher or lower than the rest.

Interpretation:

The share of labour compensation in GDP represents the portion of income going to workers. A higher share can increase consumer spending, which boosts the economy, but it might also reduce business profits and investment, so it'll be studied later.

Methodology

1. Constructing the General Model

Given the high dimensionality of our dataset with 55 features, we exercised our economic intuition to identify and filter relevant variables pertaining to our research question.

Addressing Multicollinearity: Acknowledging the presence of multicollinearity among variables, we employed a strategic approach, leveraging both intuition and correlation matrices to select features most pertinent to our research inquiry. This meticulous process allowed us to develop our General Model effectively."

2. Poolability Test or Breusch-Pagan LM test for Individual effects

The poolability test assesses whether the panel data model should include entity (or individual) fixed effects, time fixed effects, or both, or if a pooled ordinary least squares (OLS) model without fixed effects is appropriate. This test helps determine the appropriate model specification by examining whether the individual effects (entity and/or time) are statistically significant.

```
        F test for individual effects

data:  rgdpo ~ pop + emp + avh + hc + xr + ccon + ctfp + rtfpna + labsh +  ...
F = 10.249, df1 = 63, df2 = 3057, p-value < 2.2e-16
alternative hypothesis: significant effects
```

Result:

Based on a p-value approaching zero, we reject the null hypothesis, which states "There are no individual-specific effects (fixed effects) present in the data." This indicates the presence of significant individual-specific effects. Consequently, employing a simple OLS model is inadequate, as the data exhibits non-poolable characteristics. This necessitates the use of panel data models to properly account for individual heterogeneity and ensure accurate estimation of model parameters.

3. Hausman Test:

The Hausman test, utilised in panel data analysis to choose between fixed effects (FE) and random effects (RE) models, assesses the consistency of estimators. By comparing their efficiency and bias, the test examines whether the individual-specific effects in the FE model correlate with the regressors, a condition that would contravene the assumptions of the RE model.

```
        Hausman Test

data:  rgdpo ~ pop + emp + avh + hc + xr + ccon + ctfp + rtfpna + labsh +  ...
chisq = 443.49, df = 14, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Result:

The Hausman test yields a test statistic of 443.49 with 14 degrees of freedom, resulting in a p-value significantly less than 0.05. This indicates strong evidence to reject the null hypothesis that says "There is no systematic difference between the estimators of the fixed effects and random effects models," in Favor of the fixed effect model. Consequently, we infer that the random effect model is inconsistent. Given this result, we conclude that the fixed effect model is superior, as it is consistent, while the random effect model is likely to be biased for this dataset. Therefore, we opt for the fixed effect model to ensure robust estimation of model parameters.

4. Breusch-Godfrey/Wooldridge test for Serial Correlation

The Breusch-Godfrey/Wooldridge test for serial correlation detects autocorrelation in regression models. It examines whether residuals from a regression display correlation across time or observations. This test is essential for identifying violations of the independence assumption among residuals, which can bias parameter estimates and compromise the reliability of regression analysis.

```
        Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  rgdpo ~ pop + emp + avh + hc + xr + ccon + ctfp + rtfpna + labsh +  ...
chisq = 2733.8, df = 12, p-value < 2.2e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

Result:

The Breusch-Godfrey/Wooldridge test for serial correlation in panel models yields a chi-squared test statistic of 2733.8 with 12 degrees of freedom, resulting in a p-value significantly less than 0.05. This indicates strong evidence to reject the null hypothesis that says "there is no serial correlation present in the model's errors." Therefore, there is suggestive evidence of serial correlation in the errors, suggesting potential violations of the model's assumptions. Consequently, measures need to be taken to address this issue in order to ensure the reliability of the regression analysis.

5. Breusch-Pegan Test for Heteroskedasticity

The Breusch-Pagan test for heteroskedasticity is used to detect whether the variance of the errors in a regression model is constant across all levels of the independent variables, a condition known as homoscedasticity.

By examining the relationship between the squared residuals from the regression model and the independent variables, the test assesses whether there is a systematic pattern in the variance of the errors. If there is evidence of significant heteroskedasticity, it indicates that the assumptions of the regression model are violated, potentially leading to biased estimates and unreliable inferences.

```
        studentized Breusch-Pagan test

data:  fixed_model
BP = 1746.4, df = 14, p-value < 2.2e-16
```

Result:

The result of the Studentized Breusch-Pagan test indicates that there is strong evidence of heteroscedasticity in the model's errors. The test statistic (BP) is 1746.4 with 14 degrees of freedom, and the p-value is extremely small (less than 2.2e-16). This very low p-value suggests that we reject the null hypothesis that assumes homoscedasticity, meaning that the variance of the errors is constant across observations.

Since we reject the null hypothesis that says "there is homoscedasticity in the model's errors, meaning that the variance of the errors is constant across observations", we conclude that there is evidence of heteroscedasticity in the model's errors. This violation of the assumption of constant error variance indicates that the model may be biased and inconsistent. Consequently, it suggests that the results of the model might not be reliable or trustworthy.

Given this evidence of heteroscedasticity, it's important to take measures to address this issue in the modeling process. Possible measures could include using robust standard errors, transforming the data, or employing alternative modeling techniques that are robust to heteroscedasticity. Failure to address heteroscedasticity could lead to biased estimates and unreliable inferences from the model.

6. Robust Standard Error Model

The robust standard error model provides reliable estimates of parameter variability in regression analysis, especially when traditional assumptions like homoscedasticity and independence are violated. It adjusts for issues such as heteroscedasticity and autocorrelation, ensuring the validity of statistical inferences. Researchers commonly employ this model when analyzing data prone to these complexities, enhancing the reliability and robustness of regression analyses.

In panel data models, where observations are collected over both cross-sectional units and time periods, the presence of heteroscedasticity and autocorrelation can pose significant challenges to accurate estimation and inference. Robust standard errors offer a valuable solution to mitigate these challenges.

Since the initial model has the problem with heteroscedasticity and autocorrelation, we decided to use the Robust standard error model to address those problems. When confronted with heteroscedasticity and autocorrelation in a panel data model, employing robust standard errors becomes particularly crucial. By accounting for these issues, robust standard errors provide estimates of parameter variability that are less influenced by violations of traditional assumptions. This ensures the reliability of statistical inferences, even in the presence of heteroscedasticity and autocorrelation across both cross-sectional units and time periods in the panel data.

7. General to Specific Procedure

The General-to-Specific (GETS) procedure, a widely-used approach in econometrics, offers a systematic method for variable selection in regression analysis. Rather than starting with a pre-specified model, this method begins with a general model that includes all potentially relevant variables. Through a series of iterative steps, it systematically refines the model by sequentially testing the significance of individual variables and excluding those that are found to be statistically insignificant. This process continues until a final model is reached that contains only statistically significant variables.

GETS is valued for its ability to efficiently identify the most relevant variables while controlling for overfitting and mitigating the risk of including spurious or irrelevant variables. By iteratively testing and refining the model based on statistical criteria, the GETS procedure helps to construct parsimonious yet robust regression models that accurately capture the underlying relationships in the data. Additionally, GETS reduces the risk of Type I error, ensuring that significant variables are not falsely removed from the model, thus preserving their importance in explaining the variation in the dependent variable.

Our Model has only two insignificant Variables after filtering the variables based on the intuition and Going through each feature one by one to remove highly correlated variables. So first we check the joint significant of the insignificant variables.

```
t test of coefficients:

          Estimate  Std. Error t value  Pr(>|t|)
pop     -2.5989e+03  7.5658e+02 -3.4351 0.0006003 ***
emp      1.0565e+04  1.9623e+03  5.3840 7.835e-08 ***
avh      4.2538e+02  7.8510e+01  5.4181 6.490e-08 ***
hc      -2.4662e+04  1.7631e+04 -1.3987 0.1619955
xr      -2.1857e+01  5.8339e+00 -3.7465 0.0001827 ***
ccon     1.2984e+00  2.9735e-02 43.6663 < 2.2e-16 ***
ctfp     1.3658e+05  2.5442e+04  5.3683 8.541e-08 ***
rtfpna   1.5915e+05  2.7997e+04  5.6845 1.436e-08 ***
labsh    2.7308e+04  7.5809e+04  0.3602 0.7187012
pl_con   5.0283e+05  7.6533e+04  6.5701 5.887e-11 ***
pl_gdpo -4.0532e+05  6.9891e+04 -5.7993 7.341e-09 ***
csh_c   -4.5805e+05  7.8027e+04 -5.8703 4.816e-09 ***
csh_i    4.0892e+05  7.0987e+04  5.7605 9.221e-09 ***
csh_x    4.8418e+04  1.8182e+04  2.6630 0.0077863 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Step 1:

Test the Joint Significant of the insignificant Variables.

Null Hypothesis (H0):

The null hypothesis posits that the coefficients of both HC and labsh are simultaneously equal to zero. (hc=0 and labsh=0)

Alternative Hypothesis (H1):

The alternative hypothesis suggests that at least one of the coefficients of HC and labsh is not zero.

```
Linear hypothesis test

Hypothesis:
hc = 0
labsh = 0

Model 1: restricted model
Model 2: Robust_std_err_model

Note: Coefficient covariance matrix supplied.

  Res.Df Df  Chisq Pr(>Chisq)
1   3059
2   3057  2 2.1419     0.3427
```

Since the p-value of the linear hypothesis test is greater than the alpha value (0.05), we fail to reject the null hypothesis, which states that the coefficients of HC and labsh are zero simultaneously. This means these variables are jointly insignificant. Therefore, we can remove them a once from the model without committing a Type I error.

Comparing the Models

I. Stargazer

Comparison of Different Models

	Dependent variable:			
	Dependent Variable: GDP			
	panel linear		coefficient test	
	(1)	(2)	(3)	(4)
pop	-3,365.201***	-2,598.897***	-2,598.897***	-2,638.609***
	(187.872)	(311.096)	(756.575)	(756.575)
emp	9,025.047***	10,564.960***	10,564.960***	10,565.130***
	(382.565)	(566.739)	(1,962.276)	(1,962.276)
avh	258.106***	425.380***	425.380***	447.399***
	(34.750)	(57.410)	(78.510)	(78.510)
hc	38,937.780**	-24,661.610	-24,661.610	
	(16,661.620)	(23,962.840)	(17,631.430)	
ccon	1.298***	1.298***	1.298***	1.298***
	(0.005)	(0.007)	(0.030)	(0.030)
ctfp	125,063.300***	136,579.000**	136,579.000***	143,087.200***
	(40,303.010)	(59,805.720)	(25,441.560)	(25,441.560)
rtfpna	126,603.600***	159,150.700***	159,150.700***	161,156.400***
	(30,859.310)	(50,958.910)	(27,997.460)	(27,997.460)
labsh	-119,266.900	27,308.480	27,308.480	
	(77,252.560)	(118,262.500)	(75,808.800)	

xr	-3.552	-21.857***	-21.857***	-23.430***
	(6.226)	(7.269)	(5.834)	(5.834)
pl_con	381,569.100***	502,833.100***	502,833.100***	521,033.500***
	(109,987.400)	(123,656.400)	(76,533.260)	(76,533.260)
pl_gdpo	-319,621.100***	-405,318.600***	-405,318.600***	-437,977.000***
	(115,922.200)	(129,964.100)	(69,891.360)	(69,891.360)
csh_c	-346,026.800***	-458,047.100***	-458,047.100***	-428,511.200***
	(80,758.600)	(96,932.700)	(78,027.490)	(78,027.490)
csh_i	560,948.800***	408,920.900***	408,920.900***	422,144.200***
	(82,396.370)	(92,767.240)	(70,987.350)	(70,987.350)
csh_x	-43,963.240*	48,418.460	48,418.460***	42,178.030**
	(26,139.270)	(38,943.890)	(18,182.190)	(18,182.190)
Constant	-703,948.700***			
	(133,796.900)			
Observations	3,135	3,135		
R2	0.973	0.962		
Adjusted R2	0.973	0.961		
F Statistic	114,929.700***	5,498.738*** (df = 14; 3057)		
Note:				*p<0.1; **p<0.05; ***p<0.01

II. AIC Criteria

The Akaike Information Criterion (AIC) is a metric used for model selection that balances goodness of fit with model complexity. It quantifies how well a model fits the data while penalizing for the number of parameters. AIC values are calculated from the likelihood function and the number of parameters, with lower values indicating better model fit. It is used to choose the most appropriate model among alternatives.

Apart from the Stargazer we used Akaike Information Criterion to Compare between the models

```
> print(AIC_Random)
[1] 87009.9
> print(AIC_Fixed)
[1] 86610.5
> print(AIC_Final)
[1] 86607.7
> print(AIC_Robust)
[1] 86607.7
```

Interpretation of AIC Scores for Model Comparison:

The Akaike Information Criterion (AIC) serves as a valuable tool for comparing different models and assessing their relative goodness of fit. Upon analysis of the AIC scores:

1. The Fixed Effect model exhibits a lower AIC value compared to the Random Effect model. This disparity suggests that the Fixed Effect model is more preferable over the Random Effect model, indicating a better fit to the data.

2. Furthermore, the final model, which includes only the significant variables, demonstrates a lower AIC than the General Model. This reduction in AIC implies that the final model is superior to the more complex General Model, suggesting a more parsimonious representation of the data without sacrificing explanatory power.
3. Notably, the Robust Model exhibits a similar AIC to the Reduced Model. However, it still maintains a lower AIC than both the General and Random models. This consistency in lower AIC values indicates that the Robust Model offers improved model fit compared to the broader General and Random models.

Overall, the comparison of AIC scores highlights the superiority of the Fixed Effect model over the Random Effect model, the preference for the Reduced Model over the General Model, and the favourable performance of the Robust Model relative to both the General and Random models.

Hypothesis Testing

1.Population Size and GDP

We aim to investigate the hypothesis that countries with the highest population numbers exhibit the lowest Gross Domestic Product (GDP) levels, utilizing an economic model as the framework for our analysis.

Formulating the Hypothesis:

- **Null Hypothesis (H0)**: Population size (**pop**) has no effect on GDP (**rgdpo**).
- **Alternative Hypothesis (H1)**: Population size (**pop**) has an effect on GDP (**rgdpo**).

```
              Estimate   Std. Error t value  Pr(>|t|)
pop        -2.6386e+03  7.5658e+02 -3.4876 0.0004943 ***
```

Given that the p-value for **pop** is 0.0004943, which is significantly less than 0.05, we strongly reject the null hypothesis that says "population size has insignificant to the model". This indicates that population size has a significant effect on GDP.

Interpretation of the Parameter:

The coefficient for population size is negative, indicating an inverse relationship between population size and GDP. Which indicates that countries with high population size have the lowest GDP per capita like sub-Saharan countries and China.

Specifically, an increase in the population by 1 million people is associated with a decrease in the GDP of the country by 2.6 billion dollars.

2. Variation of GDP across Different Economy

We seek to examine Growth Fact 1 proposed by the Swan-Solow model, which posits that there exists substantial variation in GDP per capita across different economies.

```
        F test for individual effects

data:  rgdpo ~ pop + emp + avh + hc + xr + ccon + ctfp + rtfpna + labsh +  ...
F = 10.249, df1 = 63, df2 = 3057, p-value < 2.2e-16
alternative hypothesis: significant effects
```

From the poolability (pF) test, we observe that there is a significant individual effect across countries. Since the p-value is significantly less than the value of alpha (0.05), we strongly reject the null hypothesis, which states, "There are no individual-specific effects (fixed effects) present in the data." This indicates the presence of significant individual-specific effects or variations in GDP per capita across countries, supporting the hypothesis that "There is enormous variation in GDP per capita across economies.

3. Human capital and GDP

Our third hypothesis, rooted in the Solow-Swan economic model, proposes that countries with a higher number of people engaged in the economy, also known as human capital (emp), tend to exhibit higher GDP levels. This hypothesis seeks to investigate whether variations in workforce participation, measured by emp, contribute to differences in GDP across countries. The null hypothesis suggests that emp has no effect on GDP, while the alternative hypothesis proposes that a larger and more skilled labour force leads to higher economic output.

Formulating the Hypothesis:

- **Null Hypothesis (H0):** There is no significant relationship between the number of persons engaged in economic activities (emp) and Gross Domestic Product (GDP) across countries.
- **Alternative Hypothesis (H1):** A larger and more skilled workforce, as indicated by a higher number of persons engaged (emp), corresponds to higher GDP figures across countries.

```
Linear hypothesis test

Hypothesis:
emp = 0

Model 1: restricted model
Model 2: FINAL_ROBUST_STD.ERR_MODEL

Note: Coefficient covariance matrix supplied.

  Res.Df Df Chisq Pr(>Chisq)
1   3060
2   3059  1 28.71  8.406e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

After conducting the linear hypothesis test, the obtained p-value is substantially lower than the chosen significance level of 0.05. Consequently, we confidently reject the null hypothesis, which posits that the coefficient of the variable is zero. This rejection indicates that the coefficient of the variable differs significantly from zero, underscoring its importance and significance within the model

Interpretation of the Parameter

In interpreting the parameter, with a coefficient of 10565, we observe that a one-million increase in the number of persons engaged results in a corresponding GDP increase of 10.565 billion dollars

4. Technology Level and GDP

This hypothesis aims to test a fundamental tenet of the Solow-Swan model regarding the role of technology in shaping Gross Domestic Product (GDP) across economies. The null hypothesis challenges the notion proposed by the Solow-Swan model, suggesting that technology level or growth is not the leading factor influencing GDP. Conversely, the alternative hypothesis aligns with the Solow-Swan model, asserting that technology level or growth is indeed the primary determinant of GDP, indicative of a significant positive correlation between technological advancement and economic output.

Formulating the Hypothesis:

- **Null Hypothesis (H0)**: Contrary to the Solow-Swan model, technology level or growth is not the primary determinant of Gross Domestic Product (GDP) across economies.
- **Alternative Hypothesis (H1)**: In line with the Solow-Swan model, technology level or growth serves as the primary driver of GDP across economies, suggesting a significant positive relationship between these variables.

```
ctfp     1.4309e+05  2.5442e+04  5.6242 2.032e-08 ***
```

From the Final Model The P-value of Technology variable is very close to zero, so we reject the Null Hypothesis that says "the coefficient of the variable is equal to zero, implying that the variable has no effect on the response variable (Gross Domestic Product or GDP)". So, Technology is statistically Significant to the model.

Therefore, technology level is primary determinant of Gross Domestic Product (GDP) across economies.

Interpretation of the parameter

An additional Unit increase in Total Factor Productivity (TFP) Increases the GDP by 143 billion Dollar. It is worth to remember that TFP level is set to 1 for the United States, it means that the productivity level in the United States is used as the reference point or benchmark. Other countries' TFP levels are then expressed relative to the United States.

5. Investment share and GDP

This hypothesis aims to investigate the association between the share of gross capital formation (investment share) and Gross Domestic Product (GDP) across countries. The null hypothesis says that there is no significant relationship between investment share and GDP, implying that investment share does not influence GDP levels. Conversely, the alternative hypothesis suggests that countries with a higher investment share tend to have higher GDP, indicating a positive correlation between investment levels and economic output.

Formulating the Hypothesis:

- **Null Hypothesis (H0):** There is no significant relationship between the share of gross capital formation (investment share) and Gross Domestic Product (GDP) across countries.
- **Alternative Hypothesis (H1):** Countries with a higher share of gross capital formation (investment share) exhibit higher Gross Domestic Product (GDP), suggesting a positive relationship between these variables.

```
csh_i    4.2214e+05  7.0987e+04  5.9468 3.044e-09 ***
```

Since the P-value is close to zero we reject the null hypothesis that says "the coefficient of Investment share is Zero" suggesting that this variable is statistically significant in determining the response variable.

So, there is a significant relationship between the share of gross capital formation (investment share) and Gross Domestic Product (GDP) across countries.

Interpretation of the parameter:

Increasing the investment share by one unit leads to a GDP increase of $422 billion.

Economic Intuition:

The countries which have high investment share/rate will save more of the capital produced and tends to accumulate more capital which results in high GDP.

Apart from the above hypothesis, there are other Crucial Driving Factors of GDP. Those are: -

Avh (Average annual hours worked by persons engaged): refers to the average number of hours worked per year by individuals who are actively participating in the labour force.

Share of merchandise exports: refers to the proportion of a country's total exports that consists of tangible goods, also known as merchandise.

Share of household consumption: the Higher the GDP of the country, the higher will be the household consumption. Even if this is not the driver of Growth, it shows whether a certain country have high or low GDP.

Results

General Model Construction

In constructing the General Model, we carefully selected variables from a dataset comprising 55 features, leveraging economic intuition and correlation matrices to address multicollinearity. This process allowed us to develop a robust model to investigate the drivers of GDP in developing countries.

Poolability Test

The Poolability Test results, with a p-value approaching zero, indicated significant individual-specific effects, leading us to reject the null hypothesis of no fixed effects. This confirmed the necessity of using panel data models to account for individual heterogeneity.

Hausman Test

The Hausman Test yielded a test statistic of 443.49 with 14 degrees of freedom and a p-value significantly less than 0.05. This led us to reject the null hypothesis and conclude that the fixed effect model is more appropriate than the random effect model for our dataset.

Breusch-Godfrey/Wooldridge Test for Serial Correlation

The Breusch-Godfrey/Wooldridge test for serial correlation produced a chi-squared statistic of 2733.8 with 12 degrees of freedom and a p-value significantly less than 0.05. This indicated the presence of serial correlation in the model's errors, necessitating corrective measures.

Breusch-Pagan Test for Heteroskedasticity

The Breusch-Pagan test for heteroskedasticity resulted in a test statistic of 1746.4 with 14 degrees of freedom and a p-value less than 2.2e-16. This strong evidence of heteroskedasticity suggested that the variance of errors was not constant, requiring adjustments to ensure reliable model estimates.

Robust Standard Error Model

Given the issues of heteroskedasticity and autocorrelation, we employed the Robust Standard Error Model. This approach adjusted for these problems, providing reliable estimates of parameter variability and ensuring the validity of statistical inferences.

General to Specific Procedure

Using the General-to-Specific (GETS) procedure, we systematically refined our model by testing and excluding statistically insignificant variables. This process confirmed the insignificance of human capital (hc) and the share of labour compensation in GDP (labsh), allowing us to remove them without risk of Type I error.

Hypothesis Testing

Population Size and GDP: The hypothesis that population size affects GDP was confirmed, with a p-value of 0.0004943. The negative coefficient indicated an inverse relationship, suggesting that higher population sizes are associated with lower GDP per capita.

Variation of GDP across Different Economies: The significant p-value from the Poolability Test supported the hypothesis of substantial variation in GDP per capita across different economies.

Human Capital and GDP: The p-value for the variable representing the number of persons engaged in the economy (emp) was significantly low, confirming a positive relationship between human capital and GDP, with a coefficient indicating that a one-million increase in persons engaged corresponds to a GDP increase of 10.565 billion dollars.

Technology Level and GDP: The p-value for the technology variable was close to zero, confirming that technological level is a primary determinant of GDP, with a unit increase in Total Factor Productivity (TFP) increasing GDP by 143 billion dollars.

Investment Share and GDP: The p-value for the investment share variable was close to zero, indicating a significant relationship between investment share and GDP, with a one-unit increase in investment share leading to a GDP increase of 422 billion dollars.

Other significant drivers of GDP included average annual hours worked (avh), share of merchandise exports, and share of household consumption, which further validated the robustness of our model.

Conclusion

In this study the drivers of GDP in developing countries using panel data approach is thoroughly investigated. We have used a variety of powerful statistical tests and methodologies and we observed some important factors that affect or changed the economic growth. The technological progress, capital accumulation, human capital, the size of population and share of investment are found to be the determinant factors for GDP.

Outcomes: The results are consistent with the theoretical framework of the Solow-Swan model as they demonstrate that these variables are fundamental in exhibiting economic growth.

Robust standard error methods were required for accurate parameter estimates because the baseline model had problems with heteroskedasticity and serial correlation. Our model was further improved with the General-to-Specific procedure by removing insignificant variables, making our final model generally more robust and explanatory.

In conclusion, our study offers empirical evidence of the critical part played by technological progressiveness, capital accumulation and human capital in determining GDP growth in developing economies. Those insights provided valuable information for policymakers hoping to promote economic growth through the support of technology, education and physical capital. Correcting the pertinent heteroskedasticity and serial correlation issues is absolutely essential in order to make future economic models more reliable, leading to stronger policy implications.

References

1. Iriqat, R. A., & Anabtawi, A. N. (2016). GDP and Tax Revenues-Causality Relationship in Developing Countries: Evidence from Palestine. International Journal of Economics and Finance, 8(4), 54-62. Retrieved from ARAB AMERICAN UNIVERSITY

2. Landau, L. (2014). Economic Growth and Tax Revenues. International Advances in Economic Research, 20(2), 135-147. Retrieved from Taylor & Francis Online

3. Dritsakis, N. (1996). The Causal Relationship Between Government Revenue and Expenditure in Greece. National Bureau of Economic Research. Retrieved from NBER

4. Cheng, B. S., & Lai, T. W. (1997). Government Expenditures and Economic Growth in South Korea: A VAR Approach. Journal of Economic Development, 22(1), 11-24. Retrieved from Eastern Illinois University