Econometrics Project Report

on

# Effect of various indicators of a nation on its economic growth(measured using GDP)

By
Suraj Kumar Sahoo (142202016)
Yugandhar Rajendra Dhande (142202023)

Guided by
**Dr.Amrita Roy**

May 2023

INDIAN INSTITUTE
OF TECHNOLOGY
PALAKKAD

# Contents

# 1 Introduction

The purpose of this project is to examine the relationship between several key economic indicators and Gross Domestic Product (GDP), which is an indicator of growth of an economy. Specifically, we will explore the dependencies of GDP on trade, unemployment, Foreign Direct Investment (FDI), inflation, savings, manufacturing value added (MVA), its recognition as developed or developing and its corruption perception index (CPI). We have expanded to such a large number of variables in order to explain our data well.

We will try to provide a brief understanding of all these indicators and thus, first discuss about the factors and theoretically how can they affect the GDP. GDP as is widely accepted, is a measure of the economic performance of a country and thus, is used as an indicator of its overall economic health. It represents the total value of all goods and services produced within a country's borders over a specific period.

# 2 About the data that we are planning to use

The source of data we used is from World Bank Database, United Nations Database and Transparency International Database. The data on variables including GDP, FDI, Inflation, Unemployment, Trade, Savings, MVA were obtained from World Bank and the remaining two 'developed or not' and "CPI" were obtained from UN Database and TI Database respectively. A brief idea of all these variables is provided in the table 2.

## 2.1 Pre-processing

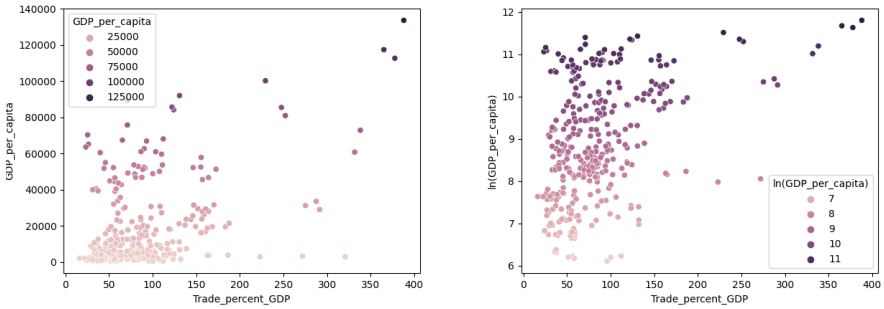
We followed a set of steps to make sure that our data is cleaned and is pre-processed to make our regression models more efficient and generalised.

- Cleaning all the Null values: Not every data point that we took from the sources mentioned were complete. There were instances (countries) whose information on some of the parameters was missing.
- Choosing the relevant data points manually: Since we took raw data from the World Bank, it contained data not just on particular countries but a set of countries forming a region of interest. For instance, there were entries such as, Middle East, Sub-Saharan Region, World etc., whose constituents were already taken into account individually. Adding those data again will only put negative impact on the regression in the form of multi collinearity.
- Scaling: We used Min-Max Scaling Technique to bring every variable into the same scale. This will help in decreasing the chance of model being biased towards one feature as compared to the other.

  At the end of all these steps we were left with a set of 119 countries whose data were up to date and had no missing values. Therefore, we took data points on a range of 3 years considering the years 2019, 2020 and 2021.

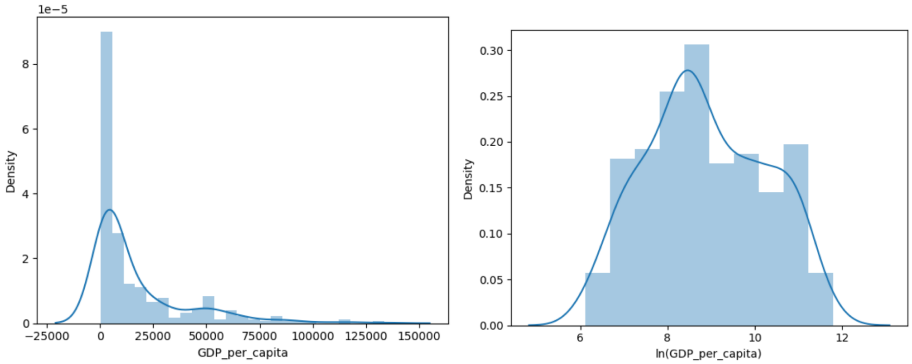
## 2.2 Why did we use log(GDP) instead of GDP?

The graph of log(GDP) with trade, will give a more distributed and clear trend, as compared to that with GDP. Thus by using log(GDP) as a measure; the model can fit more accurately.




(a) GDP vs Trade (b) ln(GDP) vs Trade

Figure 1: Relationship of GDP with Trade




(a) Density Distribution of GDP (b) Density Distribution of ln(GDP)

Figure 2: Comparison of both type of representation of GDP

## 2.3 Descriptive Statistics

The following section contains the basic descriptive statistical analysis of the data. Look at the table 1 for further information.

| Parameters | count | mean | median | std | min | max |
|---|---|---|---|---|---|---|
| FDI | 338 | 3.777 | 2.191 | 18.002 | -117.420 | 203.645 |
| ln(GDP_per_capita) | 338 | 8.879 | 8.712 | 1.369 | 6.109 | 11.802 |
| Inflation | 338 | 7.670 | 2.757 | 37.083 | -19.341 | 604.945 |
| Trade | 338 | 86.618 | 75.045 | 58.172 | 16.352 | 388.120 |
| Unemployment | 338 | 7.174 | 5.618 | 4.897 | 0.1 | 28.048 |
| MVA | 338 | 12.553 | 12.053 | 5.667 | 0.754 | 34.615 |
| savings | 338 | 22.846 | 22.428 | 9.391 | -5.623 | 53.553 |
| developed | 338 | 0.251 | 0.000 | 0.434 | 0.000 | 1.000 |
| cpi | 338 | 84.912 | 85 | 49.624 | 1 | 174 |

Table 1: Descriptive statistics of our data

# 3 Literature Study

This section contains a brief overview of all the variables we are choosing for our regression analysis of the economic growth of a nation.

By examining the relationship between these key economic indicators and GDP, this project aims to provide insights into the factors that drive economic growth, which can help one to understand what policymakers must take into account to make an informed decision.

## 3.1 Foreign Direct Investment (FDI)

FDI is the investment made by foreign companies or individuals in domestic companies or assets. When foreign companies invest in domestic companies or assets, they bring in new capital, technology and management practices. Which can lead to increased productivity and economic growth.

A higher GDP in general should attract foreign investments. But the relationship may not be that straightforward.
The FDI may as well have a negative impact on the GDP such as

- competing with the local market
- partitioning the profits made
- social impact such as labor exploitation

which can impact in general the population of the country and thus the GDP on a long run. Therefore, the relationship is complex and FDI cannot alone predict the change of GDP.

The graph 3 clearly suggests that there can be a negative as well as positive impact of FDI on GDP. A higher value of FDI may not necessarily imply a higher GDP. A study done on the relationship between FDI and economic growth in [2], shows that most of the countries exerts positive relationship between FDI

Table 2: Variable Description

| **Variables (Independent(I) and Dependent(D))** | **What does this variable describe?** | **Unit** |
|---|---|---|
| GDP (D) | Log of the Gross Domestic Product as a measure of the economic growth of a country in a particular year | USD per capita |
| Foreign Direct Investment (I) | Investment made by foreign companies or individuals in domestic companies or assets | Measured as the percent of GDP |
| Inflation (I) | increase in the general price level of goods and services in an economy over time | consumer prices (annual percentage increase) * |
| Unemployment (I) | percentage of total labor force that are unemployed | percentage |
| Developed (or not) (I) | Acceptance of a country as developed or not (as considered by the UN) | 0 or 1 if it is developing or developed |
| Trade (I) | The sum of all the imports and exports done annually | Percent of Gross Domestic product of the nation |
| Savings (I) | Gross savings are calculated as gross national income less total consumption, plus net transfers. | Percent of GDP of the nation |
| Manufacturing Value Added (I) | Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs | Percent of GDP of the nation |
| CPI (Corruption Perceptions Index) (I) | annual ranking of countries by their perceived levels of corruption | Rank (higher rank implies highly corrupt) |

* : The annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.

and economic growth while some countries have negative refer [8] and neutral effects as well.

## 3.2 Inflation

Inflation is a sustained increase in the general price level of goods and services in an economy over time. There are several ways where GDP and inflation can influence each other. Consider a situation of high inflation where the purchasing power of the consumer is less, resulting in demand for goods to be less and thus

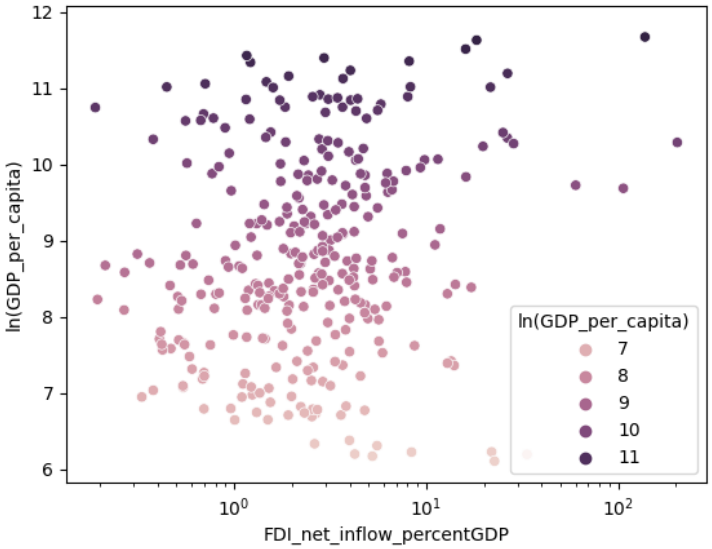



Figure 3: ln(GDP) vs FDI(on the scale of logarithm)

a declining economic growth. On the other hand, GDP growth can lead to inflation, i.e. when the economy is growing rapidly and demand for the goods and services is high the businesses may increase their prices in response to the increased demand. Other factors such as changes in productivity, supply chain disruptions and changes in government policy can also affect both inflation and GDP. Hence understanding their relation one on one can be a complex problem, without actually considering other factors that can influence both differently.

We shall look for the trend when we include other factors of the economy and then conclude how inflation is affecting the economy. The study done on inflation's effect on economic growth in [3] suggests there is significant negative correlation between inflation and income growth for rather long periods. This correlation survives even when considering trade, unemployment for econometric model and remains close to zero when our econometric model contains all other explainatory variables mentioned in table [2].

### 3.3 Unemployment

Unemployment measures the percentage of the workforce that is without work but actively seeking employment and is closely related to GDP. Growing GDP implies that economy is producing more goods and services,which in turn leads to higher demand for labor and results in employment opportunities. A shrinking GDP implies less production and services and will require less labor force hence unemployment rises. However the relationship may not be as simple as

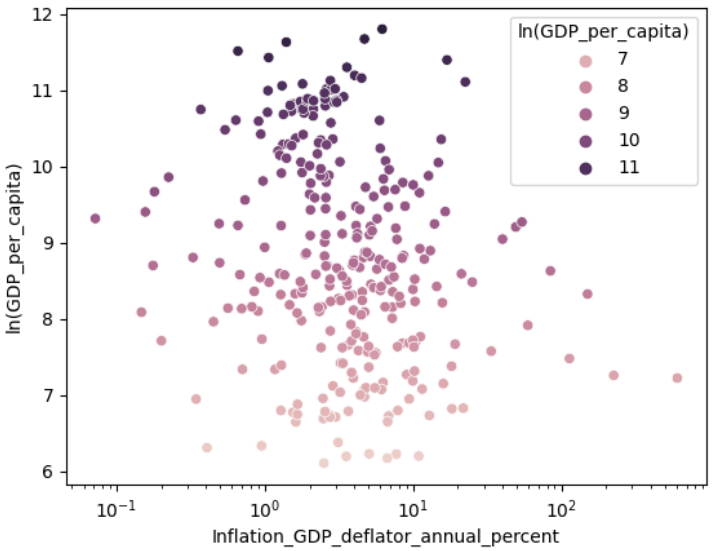



Figure 4: ln(GDP) vs Inflation

it may look. For instance, considering a nation with better infrastructure and technology may require lesser work force for the same amount of work. Also, in general unemployment may be high for countries with not-so-good government policies and immigration policies.

The graph illustrates the relationship of Unemployment with GDP.

Although a decreasing trend can be seen, there are several countries with less unemployment as well as less GDP. This can be a case, where workers are employed in low-skilled and low-paying jobs and GDP of the nation is relatively less. For example, Cambodia has a low GDP per capita as well as low unemployment owing to the fact that their workers are employed in low-skilled and low-paying jobs in agriculture and the garment industry.

The study done in [4] shows that economic growth impacts negatively on unemployment rate in Jordan for longer run. Which actually goes hand in hand with Okun's law that claimed an inverse relationship between the economic growth and unemployment.Our study also follows their trend.

### 3.4 Developed or not

A nation being declared developed or not is dependent on several factors, including GDP of the nation, extent of industrialisation and infrastructure development, Human Development Index (HDI) scores(which combines indicators of health, education, and income to provide a composite measure of human development.), poverty, political stability and democratic governance. There is no

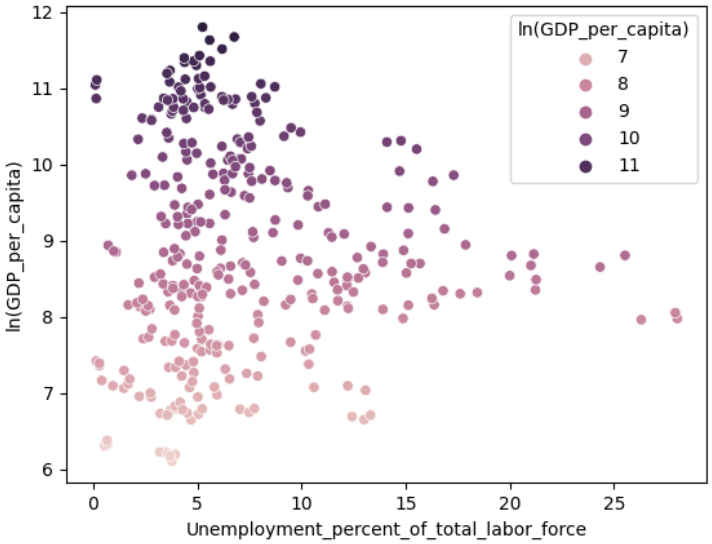



Figure 5: ln(GDP) vs Unemployment

actual official process to do so by the UN, but they look at the above mentioned indicators before deciding whether country is developed or not.

For instance, check the above plot for visualizing how the GDP of a nation varies among the developed countries and the developing countries. Clearly there are several countries that lie on a higher side of the GDP per capita but are still not considered as developed countries.

Above remark is explained in the paper [5],which analyzes time lags of nine development indicators in developing countries compared to 21 developed countries, showing progress in GDP, education, life expectancy, infant survival, telephone usage, and CO2 emissions. Employment outside agriculture has mixed results, while passenger car usage lags behind. The developing world has reduced the distance to the average benchmark country by a quarter in the past century.

### 3.5 Trade

Consider this study, given in [1] on the effect of trade liberalization on GDP. The results of the study suggest that international trade freedom has a positive and significant impact on GDP per capita. This implies that countries with more open trade policies tend to experience higher economic growth than those with more restrictive trade policies. Countries that have opened their economies in recent years (including India, Vietnam, and Uganda) have experienced faster growth and more poverty reduction.

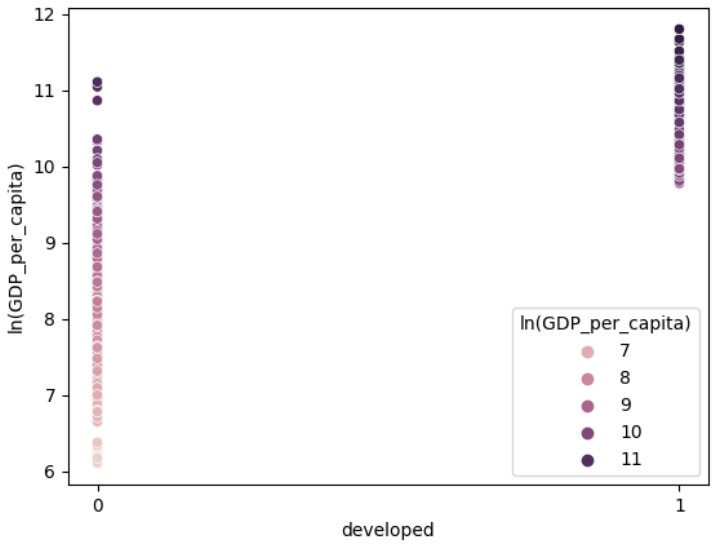



Figure 6: ln(GDP) vs Developed/Not Developed

We can see a similar positive impact when we look at the data that we collected from a set of 125 countries for the years 2019, 2020 and 2021.

Notice that there are countries with trade more than their GDP, which can be viewed as that their trade being more prominent than their internal production. Also, countries that have one highly profitable sector and they export that commodity will have more than average amount of trade as compared to other countries. Countries that also rely on other sectors that they do not trade with (including services and finance) have high GDP with very less trade. They come towards the top left part of the graph. Examples include Switzerland, Luxembourg etc. There can be several countries whose GDP might be very high and produce a lot more and thus export a lot more. Which increases the trade percentage for example USA, China and Germany etc. Countries like Argentina and Brazil have a very low GDP and they interact a lot less with other countries for trade. Hence we can find a lot of outliers that do not fit well with the long accepted trend of positive impact of trade on GDP.

### 3.6 Savings

Gross savings are calculated as Gross National Income(GNI) less total consumption, plus net transfers. When (GNI - total consumption*) is high, it indicates that there is a large surplus of income available for saving and investing. This can contribute to economic growth if that income is used for productive investments or if it is used to fund increased consumption in the future. The net

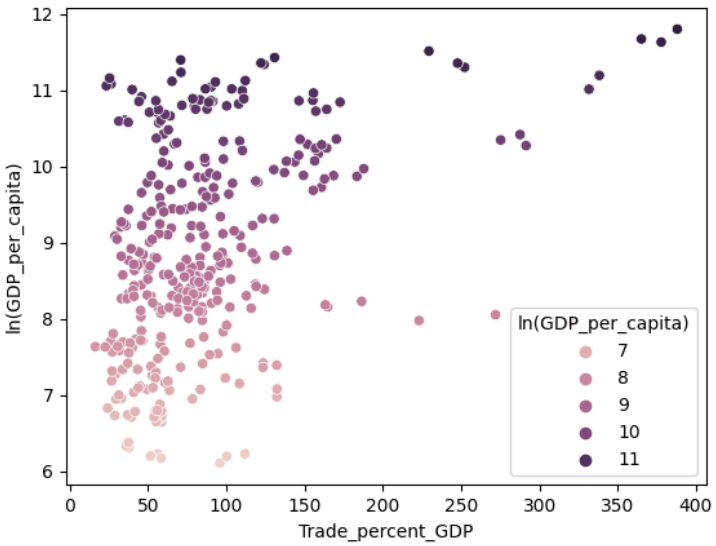



Figure 7: ln(GDP) vs Trade

transfers* funds have the similar effect. A lesser savings indicates that there is a lack of capital available for investment, which can constrain economic growth.

The research in [6] found that increased savings from commercial banks, as well as remittances and loans, have a positive impact on Kosovo's economic growth and investment. Increasing domestic savings in the banking system can reduce unemployment, promote technological development and improve people's well-being by boosting GDP.

**Definitions:**

- *Total consumption* is the total amount of spending on goods and services within an economy.
- *Net transfers* refer to the difference between the amount of money received from other countries (in the form of aid, remittances, etc.) and the amount of money sent out of the country.

## 3.7 Manufacturing Value Added (MVA)

MVA refers to the net contribution of the manufacturing sector to an economy's GDP. The value of goods produced by the manufacturing sector, including both intermediate and final goods, is included in the calculation of GDP. A study conducted in [7] makes the same conclusion that manufacturing is positively

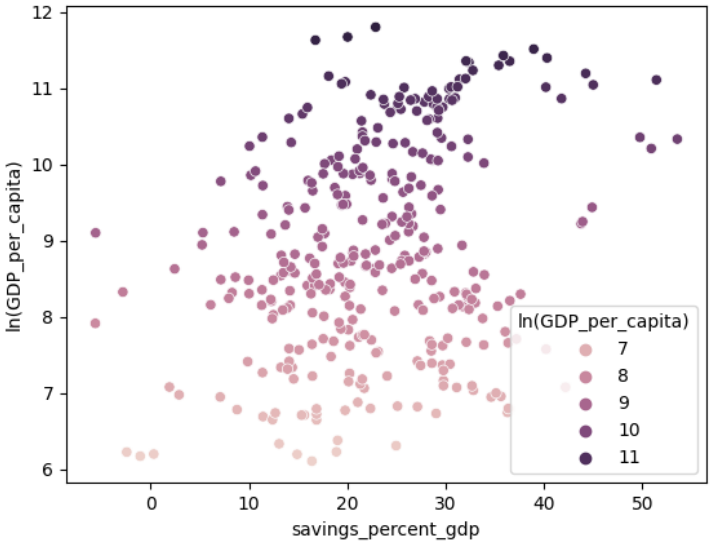



Figure 8: ln(GDP) vs Savings

related to economic growth and this effect is more pronounced for the poorer countries.

The graph clearly suggests that there are other countries that rely less on manufacturing yet their GDP is high enough. This can be a result of them relying on other sectors of the economy including service sector, which includes tourism, finance, and transportation. For instance UAE, Switzerland, Norway and so on rely more on the service sector than manufacturing.

## 3.8 Corruption Perception Index(CPI)

CPI gives an annual ranking of countries by their perceived levels of corruption. The CPI is produced by Transparency International, an international non-governmental organization that works to combat corruption around the world. The CPI ranks countries on a scale of 0 (highly corrupt) to 100 (very clean) based on a composite score that takes into account a range of factors including bribery, embezzlement, nepotism and other forms of corruption.

Countries with low levels of corruption tend to have a more conducive business environment along with greater transparency and accountability in government decision-making. This can encourage both domestic and foreign investment which leads to higher levels of economic growth and development, which can positively impact their GDP. Similarly, a higher level of corruption can adversely affect the economy.

The analysis in paper [10] suggests, how corruption perceptions affect eco-

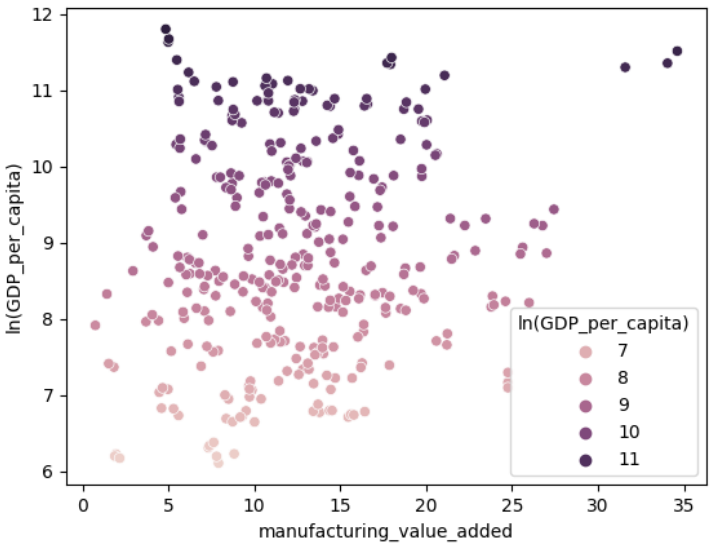



Figure 9: ln(GDP) vs Manufacturing Value Added

nomic development in a country and using empirical data with statistical methods how we can determine the impact of corruption on economic growth. The paper [9] investigates the relationship between corruption and economic factors such as GDP, exports, and FDI, using a large dataset and econometric models to explore the complex interactions between corruption and economic development. Both papers contribute to a better understanding of the role of corruption in shaping economic outcomes.

# 4 Diagnostic Test

## 4.1 Assumptions of OLS

We shall start with how well we are in line with the OLS regression assumptions.

Note that we checked these assumptions on our most complex model that included all the variables described earlier.

### 4.1.1 Assumption 1

***The models are linear in parameter***

We are considering linear models only i.e., the parameters appear in the power of 1 and each term is either a constant or the product of a parameter and an independent variable.

For instance we have considered this model (with all the independent variables):

$$log(GDP) = \beta_0 + \beta_1 * \text{Trade} + \beta_2 * \text{Unemployment} + \beta_3 * \text{Inflation} + \beta_4 * \text{FDI} + \beta_5 * \text{MVA} + \beta_5 * \text{Savings} + \beta_6 * \text{Developed} + \beta_7 * \text{CPI}$$

Therefore, we meet the first assumption of the model.

#### 4.1.2 Assumption 2

***There is variation in the samples i.e. the values of $x_i$ are not fixed and there is a standard deviation in the dependent variable.***

The data is a random sample drawn from the population, since we are considering more than 100 countries with data taken for a period of 3 years annually. Therefore we can considering this assumption valid in our case. And looking at the graphs we can easily infer that we are in line with this assumption and that $\sum_{i=0}^{N}(x_i - \bar{x})^2 \neq 0$.

#### 4.1.3 Assumption 3

***The expected value of the error term is zero and the error terms follow a normal distribution***.

We obtained the mean of the residuals calculated by regressing over all the independent variables as *-7.898819938044593e-18 (≈0)*. Therefore, we are abiding by this assumption.

Also our error terms follow a normal distribution, whose mean is clearly around zero.

#### 4.1.4 Assumption 4

***Homoscedasticity: The variance across all the possible observations is constant.***

For this we may look at the graphs that we plotted with every feature. Evidently there are few variables that may not completely adhere to this assumption because of the spread in data points. Therefore we performed **the Goldfeld Quandt Test** (has been describe in Supplementary part of the report), which yielded the following f-stat values depicted in table 3.

In GQT, we are looking at the difference in variation in two groups of data, taken from two ends of the independent variables. Considering the significance level at 0.05 and degrees of freedom at 337 and 337 respectively, the f- statistics critical value stands at 1.19624615.

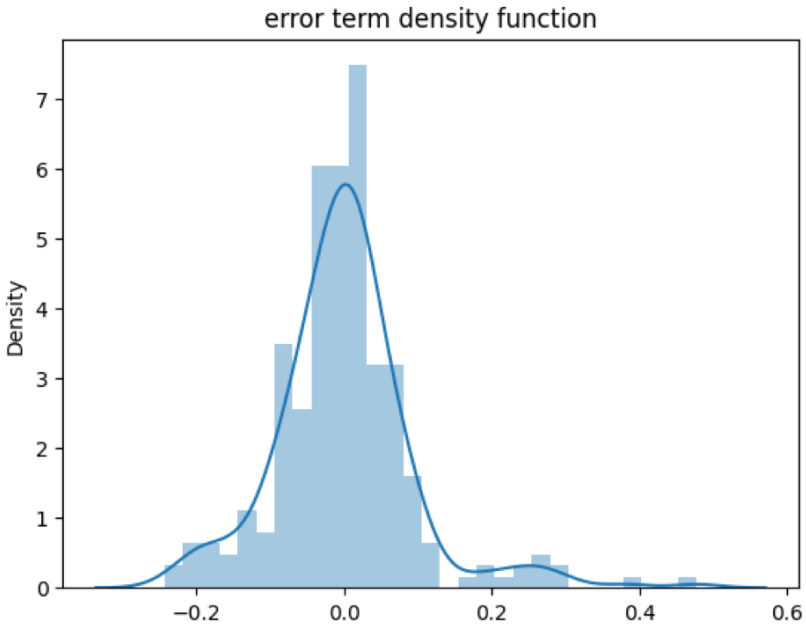



Figure 10: Error terms following a Normal Distribution i.e., $e_i \sim N(0, \sigma^2)$

| | |
|---|---|
| FDI-net-inflow-percentGDP | 1.1321745719866247 |
| Inflation-GDP-deflator-annual-percent | 1.1391530457844363 |
| Trade-percent-GDP | 1.0880750478343302 |
| Unemployment-percent-of-total-labor-force | 1.1304259326848742 |
| manufacturing-value-added | 1.1184710057743383 |
| savings-percent-gdp | 1.051213022971721 |
| cpi-ranking | 1.1491797194504925 |

Table 3: GoldFeld Quandt Test values for each independent factor

(no of observations = 338, split for Goldfeld Quandt Test = 0.4 )

Since our f-statistics value is less than or equal to this critical value, we fail to reject the null hypothesis of homoscedasticity and conclude that there is no significant evidence of heteroscedasticity in this data.

A similar f-stat for a multi regression model is 0.6475070455384426 and is less than the critical value of 1.19883902 (considering the significance level at 0.05 and degrees of freedom at 330 and 330 respectively).

#### 4.1.5 Assumption 5

***There is no correlation between the error terms of two observations i.e.,*** $Covar(u_i, u_j) = 0, \forall i, j \in \{1, \ldots, N\}, i \neq j$

We are initially claiming this to be true owing to the fact that the data we collected is sampled randomly across the population and hence the errors would not be as correlated. We may not be able to achieve the absolute zero but we can at least claim this to be true till some extent.

For instance we have approximately considered following number of countries from each continent:
Africa:
Asia: 28
Europe: 48
North America: 3
South America: 10
Australia : 1

A similar inference can also be drawn from the plot given below in figure 11 that shows the relationship between the residuals(error terms) and the predicted values or the dependent variable values. If the errors were correlated, there would be some pattern with the change in values of the dependent variable. Since the values of errors are abruptly above or below zero, we cannot but deny the fact that the errors are in anyway interrelated.

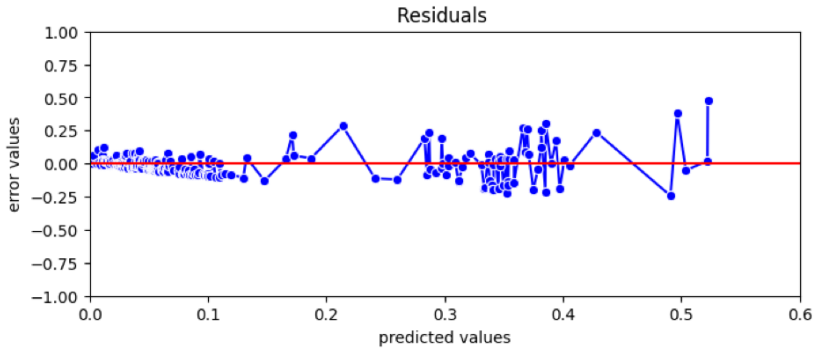



Figure 11: Plot depicting the relationship of the dependent variable with the residuals, i.e., $(y_i - \hat{y_i})$

If there were any patterns in the error values, then we may not be able to say that the error values were uncorrelated.

So, considering all the assumptions that we discussed earlier, according to the Gauss Markov Theorem, we may be able to get the

Best Linear Unbiased Estimators (BLUE),

that have the least variance and are unbiased.

However, running these regressions may not yield the best results that perfectly explains the relationship of GDP with every variable but we will try to find the best approximation of it and try to explain the results we observe.

## 4.2 Multi-collinearity

Since we will also deal with multi variable regression, we need to take care of another problem that can cause a bad inference of the multi regression, the problem of multi-collinearity.

### 4.2.1 Assumption

***There is no perfect linear relationship among the explanatory variables.***

To check if our data follows this assumption we shall look at the covariance matrix for initial conclusions. See figure 12.

The matrix clearly shows some dependencies but they may or may not be significant. Looking closely at the matrix, we can however see that there are variables that are more correlated than others. These observations, such as that of saving and percent of labour force unemployed or manufacturing value added, do confer with our earlier explanation of how savings can lead to economic growth that in turn results in higher investment and hence more manufacturing value is added, hence employment opportunities. Similarly one can also see being developed or not influences trade, savings and CPI ranking.

Such type of inferences can be drawn with our earlier knowledge. But the main question remains that, will such correlation hurt the multiple regression or not.

For getting a better answer to this, we calculated the Variance Inflation Factor. This is the factor that comes up in the calculation of variance of the parameter we are considering (see equation 4.2.2). Since for a BLUE estimator we require the least variance, a larger VIF factor can increase the variance, in turn giving a less efficient parameter estimation.

### 4.2.2 Variance Inflation Factor(VIF)

A variance inflation factor (VIF) is a measure of the amount of multicollinearity in regression analysis.Variance inflation factors allow a quick measure of how much a variable is contributing to the standard error in the regression. When significant multicollinearity issues exist, the variance inflation factor will be very large for the variables involved.The usually followed rule of thumb is that if VIF is greater than 5 then there exists a multicolinearity.
The formula for VIF is:

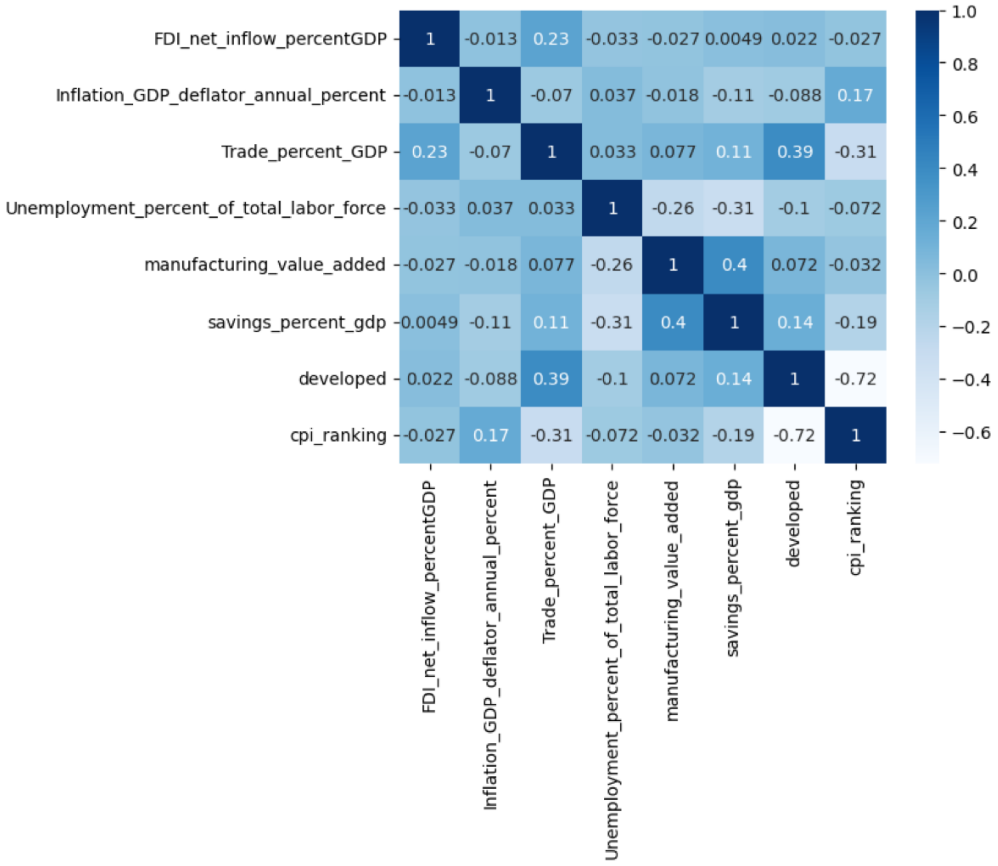



Figure 12: The Correlation Matrix, considering all the independent variable

$$VIF_i = 1/(1 - R_i^2)$$

where:

$R_i^2$ = Unadjusted coefficient of determination for regressing the $i^{th}$ independent variable on the remaining ones

$$Var(\beta_i) = \frac{\sigma_u^2}{\sum_{j=1}^{N}(X_{i,j} - \bar{X}_i)^2} * VIF_i \tag{1}$$

We have calculated the VIF factor of all the independent variable possible in our case and observations have been recorded in the table 4.

| | |
|---|---|
| FDI-netinflow-percentGDP | 1.0661876124712002 |
| Inflation-GDP-deflator-annual-percent | 1.0410776409613947 |
| Trade-percent-GDP | 1.275818729197952 |
| Unemployment-percent-of-total-labor-force | 1.2394716697229557 |
| manufacturing-value-added | 1.2267672819145403 |
| savings-percent-gdp | 1.3368297433199912 |
| developed | 2.4047391793044013 |
| cpi-ranking | 2.3358995191087044 |

Table 4: Variance Inflation Factor of the independent variable in consideration

# 5 Choice of models, results and their explanation

## 5.1 Model 1

***log(GDP) vs. Trade***

We considered the following econometric model for single variable regression:

$$log(GDP) = \beta_0 + \beta_1 * Trade + u$$

And we estimated the following:

$$log(GDP) = 0.3706 + 0.6142 * Trade$$

**Explanation pertaining to the observations:**

Looking at the regression output, we can clearly see that trade has a positive relationship with the economic growth of the country, which was established in the previous section by the literature study. The complete OLS summary can be seen in the figure 13 in the supplementary section which includes all the regressions done on the dataset.

We will now consider a multiple regression model, with parameters that in theory should affect the GDP positively based on a very clear cause-effect relationship that we established earlier.

## 5.2 Model 2

***ln(GDP) vs. Foreign Direct Investment, Trade, Gross Savings***

Econometric/True model:

$$log(GDP) = \beta_0 + \beta_1 * Trade + \beta_2 * FDI + \beta_3 * Savings + u$$

Predicted Model:

$$log(GDP) = 0.2799 + 0.5915 * Trade - 0.1994 * FDI + 0.3537 * Savings$$

According to the predictions, we can clearly see the positive effect of trade and savings on the GDP of a nation. However the effect of FDI is not as was expected. This negative effect can be described by the data we had in our hand. Look at the figure 3, in which the data is scattered enough so as to actually be considered to have only a positive effect on GDP.

**Explanation pertaining to the observations:**

This type of observation can be the reason for the inclusion of a significant number of countries that have high GDP even though they receive a very less foreign direct investment, examples include Switzerland, Japan, Norway etc. Also, considering the factors by which the increase in FDI can have negative impact on the GDP, we can say that the observation although unlikely is actually possible. This was established in the literature survey section.

One other reason why this result seems insignificant is that we did not take into account the time it might take for a country to get affected by the FDI done now. Since the FDI of a nation is not always constant and rising (depends on the source country and sink country's financial, political and social condition). This has also been affirmatively concluded by a study in [8].

This study (done on sub-Saharan African countries) has showed that in the long run foreign direct investment has a favorable and significant effect but, it is statistically insignificant in the shorter run.

## 5.3 Model 3

***ln(GDP) vs. (Unemployment, Inflation)***

Earlier we considered only those variables that were supposed to affect GDP positively; here we considered only the negatively impacting variables namely Unemployment and Inflation etc.

Econometric model:

$$log(GDP) = \beta_0 + \beta_1 * Unemployment + \beta_2 * Inflation + u$$

Predicted Model:

$$log(GDP) = 0.5134 - 0.0199 * Unemployment - 0.5012 * Inflation$$

**Explanation pertaining to the observations:**

The observations gives us the idea of the negative impact clearly. However, we can see from the results table [5] and hypothesis testing table [6] that the

standard error is too high for both the factors increasing their p-value. Therefore, we may not be able to say that the values obtained are not just by a chance. This is mostly due to the fact that we neglected so many other factors knowingly to see the nature of the relationship.

## 5.4 Model 4

***log(GDP) vs. all the features***

In this model we regressed with of all the features in hand. This, as was theoretically claimed earlier in theory, should be able to give us explain our dependent variable perfectly. The predicted model has been given below. The summary of the regression is provided in figure 16.

Econometric model:

$$log(GDP) = \beta_0 + \beta_1 * \text{Trade} + \beta_2 * \text{Unemployment} + \beta_3 * \text{Inflation} + \beta_4 * \text{FDI} + \beta_5 * \text{MVA} + \beta_5 * \text{Savings} + \beta_6 * \text{Developed} + \beta_7 * \text{CPI} + u$$

Predicted Model:

$$\begin{aligned} log(GDP) = 0.5631 + 0.1515 * \text{Trade} + 0.0069 * \text{Unemployment} + \\ 0.0679 * \text{Inflation} - 0.0055 * \text{FDI} + 0.0785 * \text{MVA} + \\ 0.1455 * \text{Savings} + 0.1452 * \text{Developed} - 0.4945 * \text{CPI} \end{aligned}$$

**Explanation pertaining to the observations:**

Given the nature of coefficient we can see there are a few variables that do not comply with their claimed nature. Those are-

- Unemployment: (Predicted Nature: Positive)

  This case can be attributed to the fact that we are here trying to get the causal inference on our model. If our regression was just done for prediction purposes we would not have cared about the sign and magnitude of the coefficient of the parameters in the predicted models.

  **Note:** The aim of the study now is to determine whether a particular independent variable really affects the dependent variable and to estimate the magnitude of that effect. This is causal analysis. In a **causal analysis**, the independent variables are regarded as the causes of the dependent variable. (*taken from the book "Multiple Regression, A Primer" by Paul D. Allison*)

  $\sim$ How is this **Note** related to our case?

  We are assuming that Unemployment is the cause of change in GDP, which although being true, doesn't necessarily imply that a change in GDP, first

hand, will not affect the unemployment. Along side this, we might also be facing issues due to the moderate level of multi collinearity that we encountered.

- Inflation: (Predicted Nature: Positive) This case is similar to that with Unemployment. We earlier established that inflation can negatively affect GDP. We did not consider the other way of cause and effect. It could be a possible reason along side other reasons such as multi collinearity.
- FDI: (Predited Nature: Negative) This case has been discussed earlier.

We can also infer about Manufacturing Value Added from the t-statistics, which is lower than the t-critical value(1.96). This implies that we may not be able to reject the null hypothesis ($\beta_{MVA} = 0$). A reason could be the increase of correlation with other factors.

### 5.5 Model 5

***ln(GDP) vs. Trade, Gross Savings, Developed, CPI***

Econometric model:

$$log(GDP) = \beta_0 + \beta_1 * \text{Trade} + \beta_2 * \text{Savings} + \beta_3 * \text{Developed} + \beta_4 * \text{CPI} + u$$

Predicted Model:

$$log(GDP) = 0.5751 + 0.1541 * \text{Trade} + 0.1742 * \text{Savings} + 0.1475 * \text{Developed} - 0.4880 * \text{CPI}$$

**Explanation pertaining to the observations:**

The coefficients of the variables have the same sign as expected. If we look at the table 5, we can clearly compare the values of coefficients that resulted from Model 4 and Model 5. The values of coefficients of Savings and CPI are a bit different and that can be attributed to the problem of **Omitted Variable Bias**, since looking at the covariance matrix done in Fig. 12, it is clear that the factors that we omitted in Model 5, had a good amount of covariance with factors Savings and CPI. The hypothesis testing inferred from the table (6) shows the statistical significance of the coefficients and about the goodness of fit.

Nonetheless, we can see that, we are still unable to confer everything from the data, since not all the variance of dependent variable is being explained by our independent variables. There is still scope of adding other variables.

## 6 Results Table

A cumulative table has been given, that comprise of all the estimated coefficients for that model and their standard error(in parentheses).

| Variables | Model1 | Model2 | Model3 | Model4 | Model5 |
|---|---|---|---|---|---|
| Foreign Direct Investment (I) | | -0.1994 (0.213)** | | -0.0055 (0.118)** | |
| Inflation (I) | | | -0.5012 (0.220)** | 0.0679 (0.110)** | |
| Unemployment (I) | | | -0.0199 (0.074)** | 0.0069 (0.041)** | |
| Developed (or not) (I) | | | | 0.1452 (0.023)* | 0.1475 (0.022)* |
| Trade (I) | 0.6142 (0.077)* | 0.5915 (0.077)* | | 0.1515 (0.046)** | 0.1541 (0.045)* |
| Savings (I) | | 0.3537 (0.074)* | | 0.1455 (0.047)** | 0.1742 (0.041)* |
| Manufacturing Value Added (I) | | | | 0.0785 (0.042)** | |
| CPI (Corruption Perceptions Index) (I) | | | | -0.4945 (0.034)* | -0.4880 (0.032)* |
| constant | 0.3706 (0.019)* | 0.2799 (0.087)** | 0.5134 (0.025)* | 0.5631 (0.059)* | 0.5751 (0.031)* |
| $R^2$ | 0.160 | 0.216 | 0.016 | 0.766 | 0.763 |

* : significance at 1%
**: significance at 5%

Table 5: Results(coefficients and standard errors) for each model

The table gives idea of the $R^2$ of every model, that tells one about the fraction of the dependent variable being explained by the independent variables we chose. We also later used the F-Statistics values that were calculated using this $R^2$ for further analysis of goodness of fit.

# 7 Hypothesis Testing Table

We have created this table that contains all the information required for the t-tests and f-tests required for inferences that goes with testing significance of the goodness of fit.

The table has the following columns:

- Parameters: The parameters in each model used.
- p-values for t-statistics or f-statistics : The probability of observing a certain result or a more extreme result. It is calculated as the area under

| Model | Parameters | p_value t-Stat & F-Stat | Null hypothesis | Alternative hypothesis | $\alpha$ |
|---|---|---|---|---|---|
| **Model 1** | $\beta_0$ (constant) | 0.000 | $\beta_0 = 0$ | $\beta_0 \neq 0$ | 0.01 |
| | $\beta_1$ (trade) | 0.000 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.01 |
| | $R^2$ (Model 1) | 0.000 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.01 |
| | | | | | |
| **Model 2** | $\beta_0$ (constant) | 0.001 | $\beta_0 = 0$ | $\beta_0 \neq 0$ | 0.01 |
| | $\beta_1$ (Trade) | 0.000 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.01 |
| | $\beta_2$ (FDI) | 0.351 | $\beta_2 = 0$ | $\beta_2 \neq 0$ | 0.05 |
| | $\beta_3$ (savings) | 0.000 | $\beta_3 = 0$ | $\beta_3 \neq 0$ | 0.01 |
| | $R^2$ | 0.000 | $\beta_i = 0, \forall i \in \{1, 2, 3\}$ | $\exists j \ni \beta_j \neq 0$ | 0.01 |
| | | | | | |
| **Model 3** | $\beta_0$ (constant) | 0.000 | $\beta_0 = 0$ | $\beta_0 \neq 0$ | 0.01 |
| | $\beta_1$ (Inflation) | 0.023 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.05 |
| | $\beta_2$ (Unemployment) | 0.790 | $\beta_2 = 0$ | $\beta_2 \neq 0$ | 0.05 |
| | $R^2$ | 0.0712 | $\beta_i = 0, \forall i \in \{1, 2\}$ | $\exists j \ni \beta_j \neq 0$ | 0.05 |
| | | | | | |
| **Model 4** | $\beta_0$ (constant) | 0.000 | $\beta_0 = 0$ | $\beta_0 \neq 0$ | 0.01 |
| | $\beta_1$ (Trade) | 0.001 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.01 |
| | $\beta_2$ (Unemployment) | 0.866 | $\beta_2 = 0$ | $\beta_2 \neq 0$ | 0.05 |
| | $\beta_3$ (Inflation) | 0.539 | $\beta_3 = 0$ | $\beta_3 \neq 0$ | 0.05 |
| | $\beta_4$ (FDI) | 0.963 | $\beta_4 = 0$ | $\beta_4 \neq 0$ | 0.05 |
| | $\beta_5$ (MVA) | 0.065 | $\beta_5 = 0$ | $\beta_5 \neq 0$ | 0.05 |
| | $\beta_6$ (savings) | 0.002 | $\beta_6 = 0$ | $\beta_6 \neq 0$ | 0.01 |
| | $\beta_7$ (developed) | 0.000 | $\beta_7 = 0$ | $\beta_7 \neq 0$ | 0.01 |
| | $\beta_8$ (cpi_ranking) | 0.000 | $\beta_8 = 0$ | $\beta_8 \neq 0$ | 0.01 |
| | $R^2$ | 0.000 | $\beta_i = 0, \forall i \in \{1, \ldots, 8\}$ | $\exists j \ni \beta_j \neq 0$ | 0.01 |
| | | | | | |
| **Model 5** | $\beta_0$ (constant) | 0.000 | $\beta_0 = 0$ | $\beta_0 \neq 0$ | 0.01 |
| | $\beta_1$ (Trade) | 0.001 | $\beta_1 = 0$ | $\beta_1 \neq 0$ | 0.01 |
| | $\beta_2$ (savings) | 0.000 | $\beta_2 = 0$ | $\beta_2 \neq 0$ | 0.01 |
| | $\beta_3$ (developed) | 0.000 | $\beta_3 = 0$ | $\beta_3 \neq 0$ | 0.01 |
| | $\beta_4$ (cpi_ranking) | 0.000 | $\beta_4 = 0$ | $\beta_4 \neq 0$ | 0.01 |
| | $R^2$ | 0.000 | $\beta_i = 0, \forall i \in \{1, \ldots, 4\}$ | $\exists j \ni \beta_j \neq 0$ | 0.01 |

Table 6: Hypothesis testing

the t-distribution in the range t-statistics and $\infty/-\infty$ (depending on one tailed or two-tailed). [same for F-test with F-distribution(only one tailed)]

- Null-Hypothesis: It is the hypothesis that the particular parameter/model is not significant.
- Alternate Hypothesis: The statement that contradicts the null hypothesis.
- significance level: The significance level of a parameter is the probability that the parameter value could have occurred by chance. Suppose we set the level is quite low, and the p-value stands lower than the significant

level. It implies that the probability of occurring by chance is quite small and thus, we say the parameter is significant.

Using this table we can infer about the parameters such as:

- In Model 1, the parameter 'trade' is significant in nature and p-value $< 0.01$, implies the parameter is statistically significant and thus it is unlikely that this value has occurred by chance.
- In Model 2, parameters trade and savings are significant to a level of 1%. We therefore reject the null hypothesis that the parameters $\beta_1, \beta_3$ are zeros. Similarly, we fail to reject the null hypothesis that $\beta_2 = 0$(for FDI), with the significance level of 5%. F-statistics suggests that the model is itself statistically significant at 1%, since the p-value $< 0.01$.
- A similar inference can be drawn for the other models 3 and 4. In Model 3, Inflation is statistically significant but unemployment is not(given a significance level of 5%). We can refer the previous table 5 to check that the standard error is high for this, which results in a larger t-critical value and hence the high p-value. Several reasons can affect the s.e.($\beta_2$) including lesser data points, lesser variation in the data points(MSD) or collinearity. The F-statistics suggests that the model is not significant. In Model 4, we dropped variables based on the hypothesis testing done for deciding their statistical significance.
- We can clearly see for Model 5 that the p-value is less than 0.01 for every variable thus we can say that they are significant to the model. Same is true for the model's significance.

# 8 Conclusion

We conducted a regression analysis on the various factors that can affect the GDP of a nation. The analysis showed us that finding the exact relationship is a difficult task. There are a range of factors that we did not include and could be a necessary factors in deriving the real GDP. We noticed that these factors, that we considered, may have a combined effect that lead to a moderate amount of multi collinearlity. We established that we could, by choice of some variables over others, get a better fit and do a decent inference on the observations that were statistically established and were significant. We were able to quantify the degree of influence that each variable has on GDP. Nevertheless, a huge scope of variables lie on the forefront to be used for a even better statistical analysis. Such kind of analysis of GDP is a powerful tool for policymakers, investors, and researchers to understand the drivers of economic growth and make informed decisions.

# References


[1] Kaltrina Bajraktari, Roberta Bajrami, Medain Hashani (2013), *"The Effect of Trade Openness and Economic Freedom on Economic Growth: the Case of Middle East and East Asian Countries"*, International Journal of Economics and Financial Issues 3(2):376-385

[2] Mohammad Amin Almfraji, Mahmoud Khalid Almsafir, (2014), *"Foreign Direct Investment and Economic Growth Literature Review from 1994 to 2012"*, Procedia - Social and Behavioral Sciences Volume 129, Pages 206-213

[3] Javier Andrés and Ignacio Hernando, (1997), "*Inflation and economic growth: some evidence for the OECD countries*", Monetary Policy and the Inflation Process, BANK FOR INTERNATIONAL SETTLEMENTS Monetary and Economic Department BASLE

[4] Hala Hjazeen, Mehdi Seraj and Huseyin Ozdeser, (2021), "*The nexus between the economic growth and unemployment in Jordan*", Future Business Journal

[5] Dominik Paprotny, (2021), "*Convergence Between Developed and Developing Countries:A Centennial Perspective*", Social indicators research, 2021 - Springer.

[6] Artur Ribaj and Fitim Mexhuani, (2021), "*The impact of savings on economic growth in a developing country (the case of Kosovo)*", Journal of Innovation and Entrepreneurship, 2021 - Springer

[7] Eman Attiah,(2019), "*The Role of Manufacturing and Service Sectors in Economic Growth: An Empirical Study of Developing Countries*", University of Piraeus. International Strategic Management Association

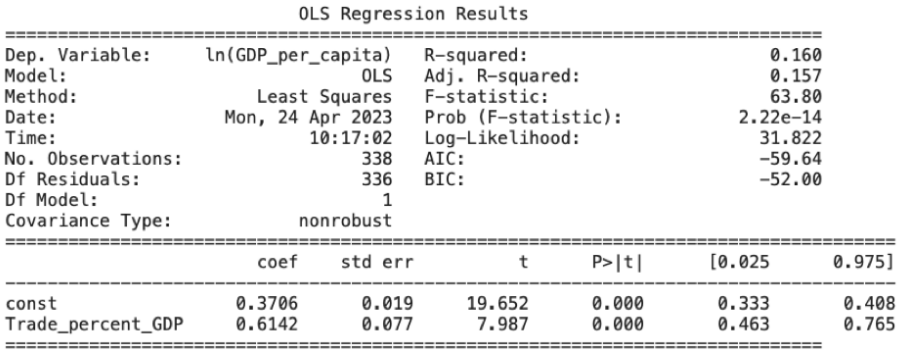
[8] Belesity Bekalu Ayenew, "*The effect of foreign direct investment on the economic growth of Sub-Saharan African countries: An empirical approach*", Cogent Economics  Finance, 2022 - Taylor  Francis

[9] Jia Sha, Plamen Ch. Ivanov, Boris Podobnik, H. Eugene Stanley, "*Quantitative relations between corruption and economic factors*", The European Physical Journal B volume 56, pages 157–166 (2007)

[10] Liviu-Stelian BEGU,Simona-Andreea APOSTU and Andreea-Oana ENACHE , "*Corruption perceptions index and economic development of the country*", Proceedings of the International Conference on Applied Statistics, Volume 1 (2019) ,Page: 115 - 122

# 9 Supplementary Content

## 9.1 The Regression Outputs

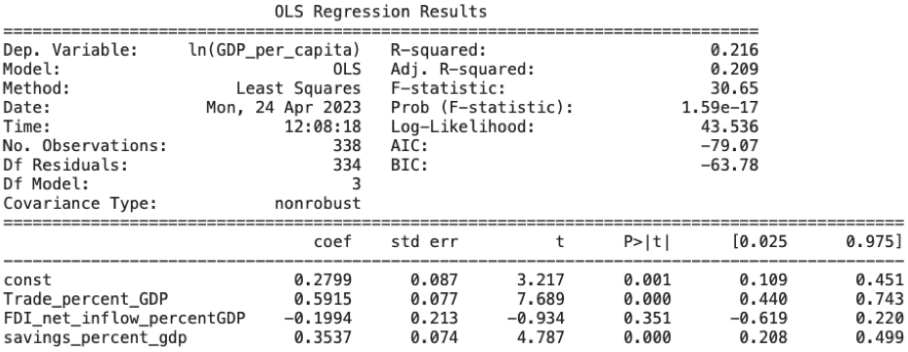


OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | ln(GDP_per_capita) | R-squared: | 0.160 |
| Model: | OLS | Adj. R-squared: | 0.157 |
| Method: | Least Squares | F-statistic: | 63.80 |
| Date: | Mon, 24 Apr 2023 | Prob (F-statistic): | 2.22e-14 |
| Time: | 10:17:02 | Log-Likelihood: | 31.822 |
| No. Observations: | 338 | AIC: | -59.64 |
| Df Residuals: | 336 | BIC: | -52.00 |
| Df Model: | 1 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.3706 | 0.019 | 19.652 | 0.000 | 0.333 | 0.408 |
| Trade_percent_GDP | 0.6142 | 0.077 | 7.987 | 0.000 | 0.463 | 0.765 |

Figure 13: Model 1

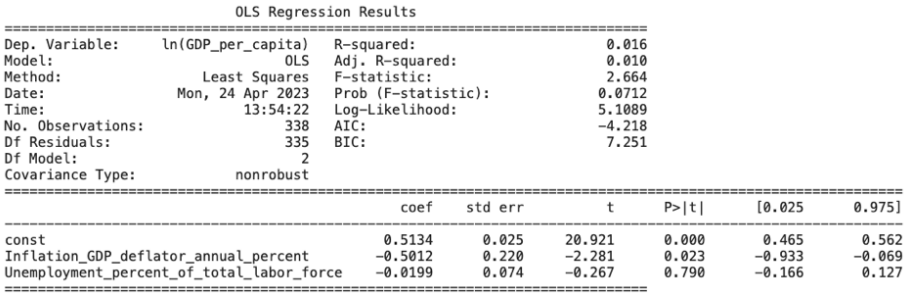


OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | ln(GDP_per_capita) | R-squared: | 0.216 |
| Model: | OLS | Adj. R-squared: | 0.209 |
| Method: | Least Squares | F-statistic: | 30.65 |
| Date: | Mon, 24 Apr 2023 | Prob (F-statistic): | 1.59e-17 |
| Time: | 12:08:18 | Log-Likelihood: | 43.536 |
| No. Observations: | 338 | AIC: | -79.07 |
| Df Residuals: | 334 | BIC: | -63.78 |
| Df Model: | 3 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.2799 | 0.087 | 3.217 | 0.001 | 0.109 | 0.451 |
| Trade_percent_GDP | 0.5915 | 0.077 | 7.689 | 0.000 | 0.440 | 0.743 |
| FDI_net_inflow_percentGDP | -0.1994 | 0.213 | -0.934 | 0.351 | -0.619 | 0.220 |
| savings_percent_gdp | 0.3537 | 0.074 | 4.787 | 0.000 | 0.208 | 0.499 |

Figure 14: Model 2



OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | ln(GDP_per_capita) | R-squared: | 0.016 |
| Model: | OLS | Adj. R-squared: | 0.010 |
| Method: | Least Squares | F-statistic: | 2.664 |
| Date: | Mon, 24 Apr 2023 | Prob (F-statistic): | 0.0712 |
| Time: | 13:54:22 | Log-Likelihood: | 5.1089 |
| No. Observations: | 338 | AIC: | -4.218 |
| Df Residuals: | 335 | BIC: | 7.251 |
| Df Model: | 2 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.5134 | 0.025 | 20.921 | 0.000 | 0.465 | 0.562 |
| Inflation_GDP_deflator_annual_percent | -0.5012 | 0.220 | -2.281 | 0.023 | -0.933 | -0.069 |
| Unemployment_percent_of_total_labor_force | -0.0199 | 0.074 | -0.267 | 0.790 | -0.166 | 0.127 |

Figure 15: Model 3

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | ln(GDP_per_capita) | R-squared: | 0.766 |
| Model: | OLS | Adj. R-squared: | 0.760 |
| Method: | Least Squares | F-statistic: | 134.6 |
| Date: | Mon, 24 Apr 2023 | Prob (F-statistic): | 6.33e-99 |
| Time: | 18:04:32 | Log-Likelihood: | 247.85 |
| No. Observations: | 338 | AIC: | -477.7 |
| Df Residuals: | 329 | BIC: | -443.3 |
| Df Model: | 8 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.5631 | 0.059 | 9.474 | 0.000 | 0.446 | 0.680 |
| Trade_percent_GDP | 0.1515 | 0.046 | 3.270 | 0.001 | 0.060 | 0.243 |
| Unemployment_percent_of_total_labor_force | 0.0069 | 0.041 | 0.169 | 0.866 | -0.073 | 0.087 |
| Inflation_GDP_deflator_annual_percent | 0.0679 | 0.110 | 0.616 | 0.539 | -0.149 | 0.285 |
| FDI_net_inflow_percentGDP | -0.0055 | 0.118 | -0.046 | 0.963 | -0.238 | 0.227 |
| manufacturing_value_added | 0.0785 | 0.042 | 1.850 | 0.065 | -0.005 | 0.162 |
| savings_percent_gdp | 0.1455 | 0.047 | 3.112 | 0.002 | 0.054 | 0.237 |
| developed | 0.1452 | 0.023 | 6.341 | 0.000 | 0.100 | 0.190 |
| cpi_ranking | -0.4945 | 0.034 | -14.652 | 0.000 | -0.561 | -0.428 |

| | | | |
|---|---|---|---|
| Omnibus: | 8.140 | Durbin-Watson: | 2.077 |
| Prob(Omnibus): | 0.017 | Jarque-Bera (JB): | 9.371 |
| Skew: | -0.256 | Prob(JB): | 0.00923 |
| Kurtosis: | 3.634 | Cond. No. | 27.6 |

Figure 16: Model 4

OLS Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | ln(GDP_per_capita) | R-squared: | 0.763 |
| Model: | OLS | Adj. R-squared: | 0.760 |
| Method: | Least Squares | F-statistic: | 268.2 |
| Date: | Mon, 24 Apr 2023 | Prob (F-statistic): | 9.01e-103 |
| Time: | 18:05:03 | Log-Likelihood: | 245.86 |
| No. Observations: | 338 | AIC: | -481.7 |
| Df Residuals: | 333 | BIC: | -462.6 |
| Df Model: | 4 | | |
| Covariance Type: | nonrobust | | |

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.5751 | 0.031 | 18.683 | 0.000 | 0.515 | 0.636 |
| Trade_percent_GDP | 0.1541 | 0.045 | 3.450 | 0.001 | 0.066 | 0.242 |
| savings_percent_gdp | 0.1742 | 0.041 | 4.226 | 0.000 | 0.093 | 0.255 |
| developed | 0.1475 | 0.022 | 6.676 | 0.000 | 0.104 | 0.191 |
| cpi_ranking | -0.4880 | 0.032 | -15.145 | 0.000 | -0.551 | -0.425 |

| | | | |
|---|---|---|---|
| Omnibus: | 12.419 | Durbin-Watson: | 2.080 |
| Prob(Omnibus): | 0.002 | Jarque-Bera (JB): | 14.421 |
| Skew: | -0.364 | Prob(JB): | 0.000739 |
| Kurtosis: | 3.703 | Cond. No. | 9.94 |

Figure 17: Model 5

## 9.2 Goldfeld Quandt Test

The Goldfeld-Quandt test is a statistical test used to determine if the variance of the error terms in a regression model is heteroscedastic, meaning that the vari-

ance of the errors is not constant across the range of the independent variable. This test is particularly useful when working with time-series data or when the sample size is small.

This test involves dividing the data into two parts(we did in the ratio 0.4), by sorting the independent variable ($x_i$) in ascending order, and then estimating two separate regression models for each part.

It involves finding the similarity between the two regression's RSS values and performing the F-test to determine if there is evidence of heteroscedasticity. The F-critical value can be calculated by the following formula:

$$GQ = (RSS_1/(n_1 - k))/(RSS_2/(n_2 - k))$$

where $RSS_1$ and $RSS_2$ are the residual sum of squares for the first and second subsamples, respectively, $n_1$ and $n_2$ are the sample sizes for the first and second subsamples, and $k$ is the number of parameters estimated in the regression model.
$(n_i - k)$ : degrees of freedom.
Null and Alternate Hypothesis of Goldfeld-Quandt Test

- Null Hypothesis: Heteroscedasticity is not present.
- Alternate Hypothesis: Heteroscedasticity is present.