ECONOMIC GROWTH & DEVELOPMENT ASSIGNMENT REPORT

ECON F244

GROUP 15

Determinants of Economic Growth in a Sample of Countries

Under the supervision of Prof Sudatta Bannerjee

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCES, PILANI HYDERABAD CAMPUS

Group Members

Name of Student	**ID Number**
Aarav Haran	2022B3A70880H
Aditya Pentyala	2022B3A70522H
Vaishnu Kanna	2022B3A71608H
Anant Gupta	2022B3A71282H
Samarth Tabakade	2022B3A71292H
Sashank Sivakumar	2022B3AA1303H
Anubhav Saha Roy	2022B3AA0877H

Table of Contents

Introduction

Economic growth is a multifaceted concept, and assessing it requires considering various factors and indicators. While there are numerous metrics available, the change in gross domestic product (GDP) over time is widely regarded as one of the most comprehensive measures of economic growth. GDP represents the total monetary value of all goods and services produced within a country's borders within a specific time frame, typically measured quarterly or annually.

The significance of GDP lies in its ability to capture the overall health and performance of an economy. It encompasses the output of all sectors, including agriculture, manufacturing, services, and government spending, providing a broad overview of economic activity. Changes in GDP reflect shifts in production, consumption, investment, and government expenditure, offering insights into the economy's trajectory.

Numerous factors can exert direct or indirect influence on the Gross Domestic Product (GDP) of an economy. In this report, we endeavor to analyze specific variables that could potentially impact the economic growth of a group of countries.

All the code used for the project, in Python and R, can be found at the following Github repository: https://github.com/adityapentyala/egd-assignment/tree/master

Sample

Countries

Our sample comprises countries from diverse regions globally, interconnected through membership in the United Nations Security Council (UNSC).
The United Nations Security Council (UNSC) is one of the principal organs of the United Nations. Its primary responsibility is maintaining international peace and security. Comprising 15 member states, including five permanent members (China, France, Russia, the United Kingdom, and the United States) and ten non-permanent members elected for two-year terms. The UNSC plays a crucial role in addressing global security challenges.

REASONS -We selected this group of countries primarily because it represents one of the most influential and powerful coalitions of nations. Additionally, it provided us with insights into a broad spectrum of countries, encompassing developed, developing, and underdeveloped nations. Another rationale for selecting these countries is their data availability for the specific variables we have chosen.
The list of countries we've chosen can be grouped into three categories:

- Developed Countries:
    - ➢ **USA -** The United States is a global economic powerhouse known for its technological innovation, diverse industries, and influential financial markets.
    - ➢ **UK -** The United Kingdom boasts a sophisticated financial sector, a strong services industry, and a long history of international trade and investment.
    - ➢ **France**- France is renowned for its rich cultural heritage, advanced manufacturing sector, and significant contributions to global luxury goods and tourism.
    - ➢ **Switzerland**- Switzerland is famous for its stable economy, banking sector expertise, and reputation as a global hub for finance, pharmaceuticals, and precision engineering.
    - ➢ **Norway**- Norway's economy is driven by its abundant natural resources, particularly oil and gas, coupled with a strong focus on sustainability, innovation, and social welfare.

- ➢ **Russia**- Russia possesses vast reserves of natural resources, including energy and minerals, and maintains a diversified economy with strengths in aerospace, defense, and technology sectors.
- ➢ **South Korea-** South Korea is a dynamic economy known for its leading technology companies, robust manufacturing sector, and global competitiveness in industries such as electronics, automotive, and shipbuilding.

- Developing countries
    - ➢ **India**- India is a rapidly growing economy with strengths in IT services, manufacturing, and agriculture, fueled by a young workforce and a burgeoning consumer market.
    - ➢ **China-** China is an economic powerhouse, leading in manufacturing, exports, and technological innovation, with a focus on infrastructure development and global trade.
    - ➢ **Albania-** Albania is transitioning to a market-based economy, with growing sectors in tourism, energy, and agriculture, supported by ongoing reforms and integration efforts.
    - ➢ **Brazil**- Brazil is a major emerging market, known for its diverse economy including agriculture, mining, manufacturing, and services, along with significant natural resource wealth.
    - ➢ **Mexico-** Mexico is a key player in global manufacturing, particularly in the automotive, electronics, and aerospace industries, with a growing services sector and proximity to major markets like the United States.

- Underdeveloped Countries
    - ➢ **Ghana-** Ghana faces economic challenges but has potential in agriculture, mining, and oil production, with ongoing efforts to improve infrastructure and attract investment.
    - ➢ **Nigeria-** Nigeria struggles with infrastructure deficits and economic diversification but possesses significant oil reserves, a growing telecommunications sector, and a young, entrepreneurial population.

The permanent members of the UNSC feature prominently on the list of developed nations. These include the USA, UK, France, Russia, and China which are still classified as a developing nation. The countries have large populations and possess readily available data.

The other non-permanent members include Norway, on the developed countries list. India, Brazil, Mexico and Albania on the developing list and Ghana on the under-developed list. These are growing economies with an abundance of resources and labor.

We have also considered Switzerland due to its political and military neutrality and South Korea. To round off the list of under-developed countries, we have taken Nigeria as it has a huge population and thus offers valuable insights into the variables considered.

Variables

We used 6 variables in our equation: Per Capita GDP, Total Labour Force, Gross Domestic Savings, Net Trade in Goods and Services, Consumer Price Index (CPI), and Foreign Direct Investment.

Per Capita GDP

Per Capita GDP captures the average economic output (GDP) per person. It offers insight into the average production and consumption of goods and services within a country, providing a gauge of the nation's overall growth. We took a log of the values, since the GDP growth is exponential, so the difference in log values provides a better insight into the growth.

Total Labor Force

We used Total Labor Force to represent Human Capital. The total labor force refers to the number of people who are either employed or actively seeking employment; it is a key indicator used to measure the size and composition of the workforce. Changes in the total labor force can affect employment and unemployment rates, which, in turn, influence economic growth. An additional reason we used was the trend of the Labour Force indicating the level of workers. If the trend was a drastic positive growth, it indicated that the workers were initially unskilled and that the workforce was rapidly transitioning into a skilled one and being more productive. This implies that investments have been made into education and healthcare, allowing them to qualitatively improve their productivity. Further, it was also observed that the labor force becomes extremely significant when we account for autocorrelation. Since linear regression cant account for autocorrelation, it appears to be less significant. However, the ARIMAX models can capture the autocorrelation and show its significance, so the issue is with the model and methods used, not with the variable itself.

Gross Domestic Savings

We used Gross Domestic Savings Variable to represent Savings. Gross domestic savings measures the total amount of savings generated within a country over a specific period, typically a year. It represents the difference between a country's total disposable income and its total consumption expenditure. When households, businesses, and the government save more than they spend on consumption, the excess savings can be channeled into productive investments, increased investment contributes to higher levels of production, leading to higher GDP growth rates. We used the same variable across two metrics, using both current USD as well as percentage of GDP, and compared the two results.

Net Trade in Goods and Services

Net Trade in Goods and Services is a key component of the balance of payments (BOP) and refers to the difference between a country's exports and imports of goods and services. It is also an important indicator of a country's international economic competitiveness. Net Trade is a component of GDP calculation. Net Trade affects aggregate demand, influences production levels and employment which further influence GDP of a country. Changes in net trade can influence exchange rates, which in turn, affect GDP growth.

We have taken the variable with current USD as well as percentage of GDP, and compared the results to see which metric gives us more conclusive results.

Consumer Price Index (CPI)

Consumer Price Index (CPI) is a measure that tracks the changes in the price level of a basket of consumer goods and services over time. It is one of the most commonly used indicators for measuring inflation. Changes in the CPI reflect changes in general price level. If prices increase it can lead to a decrease in consumption, which can dampen overall economic activity and GDP growth. Central banks often use CPI data as a key indicator for monetary policy decisions which further impacts GDP growth. Inflation, as indicated by CPI affects real wages of laborers which impacts unemployment rates and thus GDP growth rates.

Foreign Direct Investment (FDI)

Foreign Direct Investment (FDI) refers to the investment made by a company or individual in another country, in the form of either establishing business operations or acquiring business assets. FDI plays a significant role in global economic development. FDI often involves significant capital inflows in countries like India, which can be used to finance physical infrastructure**,** production and transportation which leads to GDP growth. FDI often brings advanced technology, which leads to increased efficiency and productivity, thereby enhancing GDP growth. FDI projects typically create jobs, which leads to higher household incomes and savings contributing to GDP growth. We have measured it in current USD, to get a clearer idea of the volume of investment in these countries, to accurately gauge the potential these investors see in the country they are investing in. We also used the variable as a percentage of GDP to see the variation in results, to get a clearer understanding.

Methodology

Data was pulled from worldbank.org and subsequent cleaning, preprocessing and regression were done in Python. The data from the excel was cleaned to include only our timeframe and moved into new excels, from which dataframes were read. Then, every 3rd year was taken into the final regression dataframe, with the growth rate in GDP calculated in 3 year intervals instead of 5 to account for the lack of data.

We used regression models to identify the relation between the various variables taken into consideration in our study. The primary equation we used is:

Growth rate of GDP = α + β1 (ln of initial PCGDP) +β2 (Gross Domestic Savings (current USD)) +β3 (Total labor force) +β4 (Trade in Goods and Services (current USD)) + β5 (Consumer Price Index) + β6 (Foreign Direct Investment (current USD))

The equation used with dummy variables to account for the level of development is:

Growth rate of GDP = α + β1 (ln of initial PCGDP) +β2 (Gross Domestic Savings (%age of GDP) +β3 (Total labor force) +β4 (Trade in Goods and Services (%age of GDP)) + β5 (Consumer Price Index) + β6 (Foreign Direct Investment (%age of GDP)) + δ1 (dummy for underdeveloped country) + δ2 (dummy for developing country)

We used a total of 7 models to interpret the data to varying degrees and effects, to gain an in-depth understanding of the variables and their interactions with each other. The first two models follow the standard OLS method. The first model uses absolute values for the variables wherever applicable, while the second model uses percentage of GDP as the unit for the variables wherever applicable, along with dummy variables to represent the underdeveloped and developing countries. This was done to get an in-depth understanding of how a change in the unit affects the model's results, and the possible effect on the model a country has by being developed, developing or underdeveloped.

Though outside the scope of this report, we included 2 models to account for the heteroskedasticity of the data, and to increase significance of the results we obtained from the models, and another one to observe the interactions the dummy variables have with other variables.

Again outside the scope of this report, we further ran ARIMAX models with current USD values as well as percentage of GDP values for the variables, to capture the autocorrelation that might be present in the model. We did this since OLS and ordinary linear regression cannot capture autocorrelation, but ARIMAX models can do this.

All these models will be explained in further depth in the next section, as and when needed.

Regression Analysis

In this section, we attempt to fit regression models to the datasets. Along the way, we have encountered problems with the data, and have tried to fit models that fix the problems to the best of our abilities.

We initially fit an ordinary OLS model with no dummies and receive the results. The problems seen were that the R-square value was extremely poor, and hence tried to fit another OLS model to the data, after having transformed metrics in current USD to %ages of GDP and adding non-interacting dummy variables for the level of development. We also add a third model that includes interactions between the exogenous and dummy variables in R to help maximize the R-square value.

Another problem we notice is the insignificance of coefficients with non-robust standard errors. We hypothesize that this is due to heteroskedastic nature of the data, and perform some graphical analysis to conclude that the data can be divided into independent clusters, each with their own similar covariances. We fit modified versions of the first two models with cluster-robust standard errors using a clustering algorithm and note the results which showcase higher levels of significance.

Finally, the Durbin-Watson statistic implies high degrees of autocorrelation among the variables, as is the case with most time series data. Hence, although outside the scope of this report, we perform basic autocorrelation time-series analysis and fit a dedicated time-series model, ARIMA, with exogenous regressors, to the same data passed to models 1 & 2 and note how the levels of significance drastically improve. Additionally, this model truly captures the true importance of total labor force as a development-invariant regressor, despite the challenges of large coefficients it poses.

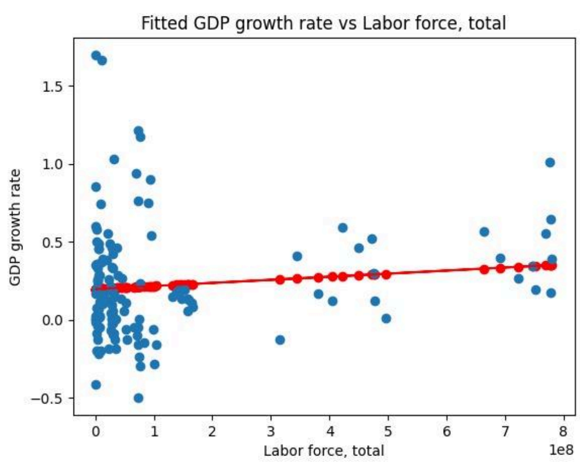
In total, we have implemented 7 different models to explain the growth rate of GDP with respect to our exogenous variables. Summaries of each of the models are given in the following pages, as well as visualizations of their fits.

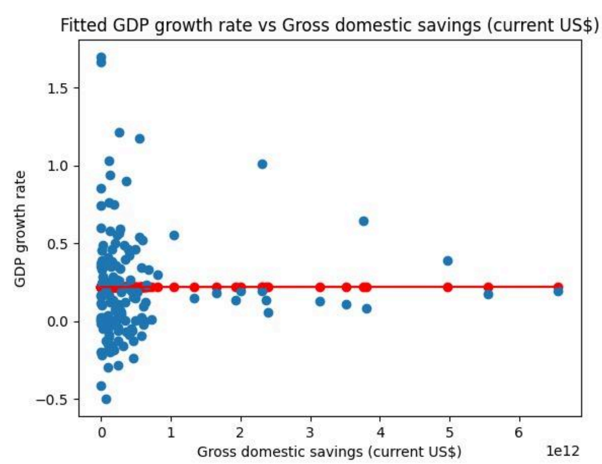
Link to Python notebook in which Models 1, 2, 4-7 are fit and analyzed: combined_regression.ipynb
Link to R source file for Model 3: interaction_regression.r

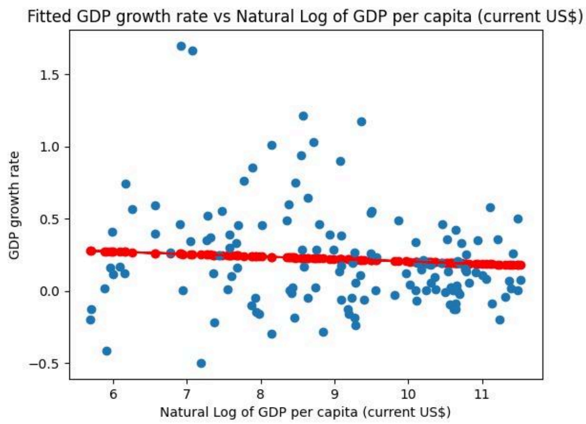
Model 1: OLS

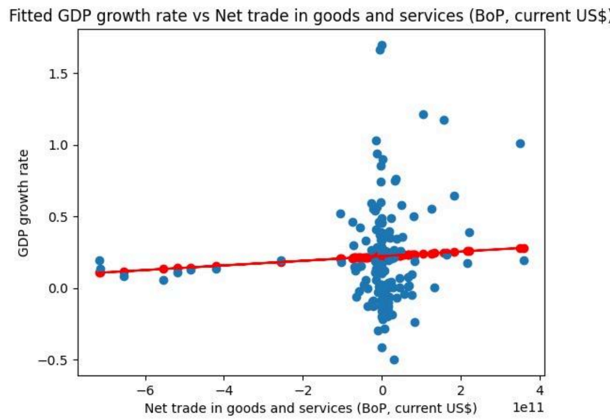
```
                            OLS Regression Results
==============================================================================
Dep. Variable:        gdp growth rate   R-squared:                       0.069
Model:                            OLS   Adj. R-squared:                  0.027
Method:                 Least Squares   F-statistic:                     1.634
Date:                Thu, 25 Apr 2024   Prob (F-statistic):              0.142
Time:                        10:33:12   Log-Likelihood:                -47.512
No. Observations:                 140   AIC:                             109.0
Df Residuals:                     133   BIC:                             129.6
Df Model:                           6
Covariance Type:            nonrobust
========================================================================================================================
                                                    coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------------
const                                             0.4444      0.216      2.053      0.042       0.016       0.873
Natural Log of GDP per capita (current US$)      -0.0171      0.024     -0.698      0.486      -0.066       0.031
Gross domestic savings (current US$)           8.092e-17   4.23e-14      0.002      0.998   -8.36e-14    8.38e-14
Labor force, total                              1.98e-10   2.31e-10      0.858      0.392   -2.58e-10    6.54e-10
Net trade in goods and services (BoP, current US$)  1.62e-13   2.25e-13      0.721      0.472   -2.83e-13    6.07e-13
Consumer price index (2010 = 100)                -0.0010      0.001     -1.386      0.168      -0.002       0.000
Foreign direct investment, net (BoP, current US$)  -4.427e-13   6.79e-13     -0.652      0.515   -1.79e-12       9e-13
==============================================================================
Omnibus:                       44.822   Durbin-Watson:                   1.760
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              115.071
Skew:                           1.275   Prob(JB):                     1.03e-25
```

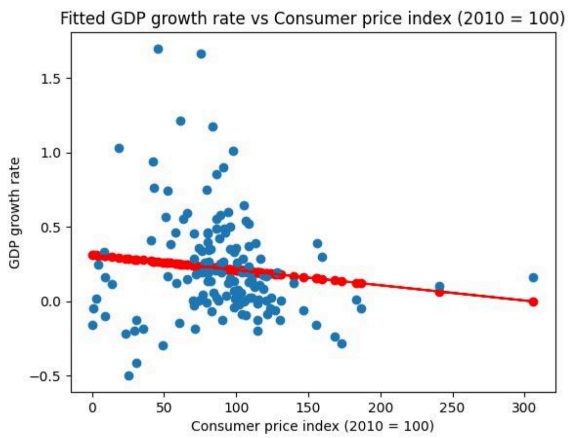
We notice that the R-squared value is 0.069, which is very low. This means that the model does not explain much of the variation in the dependent variable. Also, the p-value of the F-statistic is 0.142 which means that we cannot reject the null hypothesis. None of the coefficients, except for the constant term, are statistically significant at the 5% level. The value of the Durbin-Watson statistic is 1.760 which is close to 2, indicating that there is no strong evidence of autocorrelation in this model. Therefore, we can conclude that this model is not statistically significant.

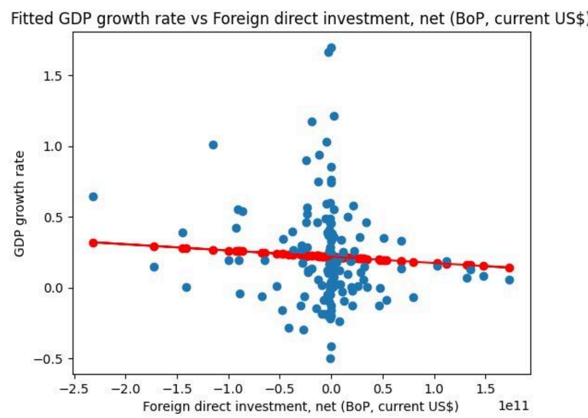





Fitted GDP growth rate vs Natural Log of GDP per capita (current US$)
GDP growth rate
Natural Log of GDP per capita (current US$)


Fitted GDP growth rate vs Net trade in goods and services (BoP, current US$)
GDP growth rate
Net trade in goods and services (BoP, current US$)
1e11


Fitted GDP growth rate vs Consumer price index (2010 = 100)
GDP growth rate
Consumer price index (2010 = 100)


Fitted GDP growth rate vs Foreign direct investment, net (BoP, current US$)
GDP growth rate
Foreign direct investment, net (BoP, current US$)
1e11

Problem 1: Low R^2 Score

The low R^2 score implies that our model does not generalize well enough to the data - that is, it cannot explain much of the variance in the model. A score of 0.069 implies that it can only explain 6.9% of the total variance in the data - a poor model.

To fix this error, we can use other variables - for example, changing the metrics from current USD to %age of GDP, and adding dummy regressors for the level of development of datapoints. This solution is implemented in Model 2, which showed a slight increase in R^2 score to 10.5%.

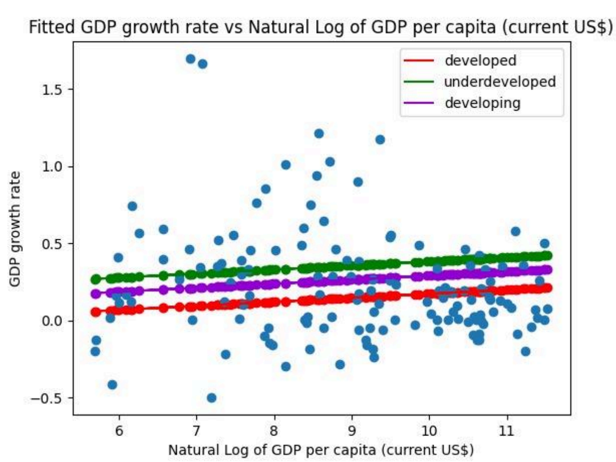
Additionally, to further improve R^2 score, we have implemented another model, Model 3, that includes interaction terms between the dummy variables and the exogenous variables. This improved the R^2 score of Models 1 & 2 by over 350% and 240% respectively, at 24.93%, although it is still a fairly poor R^2 score.

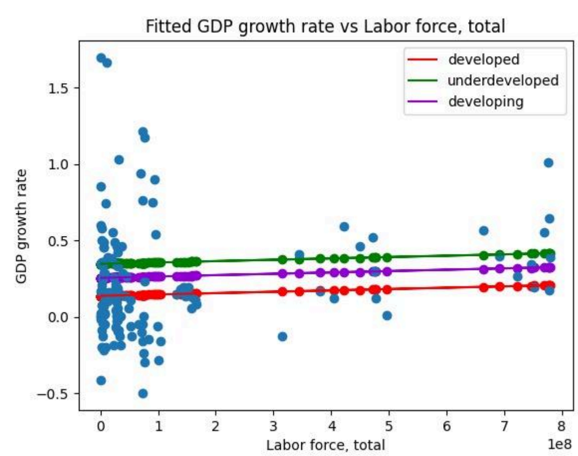
The score can be further improved by including more variables that are also relevant to the model, as well as fitting it to a larger and more homogenous dataset with more data points.

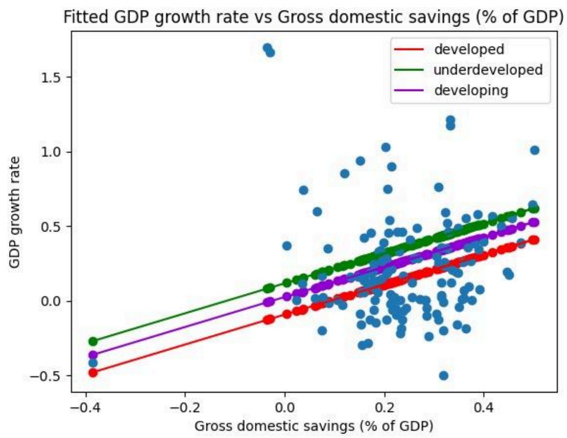
Model 2: OLS with dummy vars, metrics: %ages of GDP

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        gdp growth rate   R-squared:                       0.105
Model:                            OLS   Adj. R-squared:                  0.051
Method:                 Least Squares   F-statistic:                     1.931
Date:                Thu, 25 Apr 2024   Prob (F-statistic):             0.0605
Time:                        10:33:13   Log-Likelihood:                -44.690
No. Observations:                 140   AIC:                             107.4
Df Residuals:                     131   BIC:                             133.9
Df Model:                           8
Covariance Type:            nonrobust
==================================================================================================
                                             coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
const                                     -0.2285      0.361     -0.633      0.528      -0.942      0.485
Natural Log of GDP per capita (current US$)   0.0263      0.037      0.719      0.473      -0.046      0.099
Gross domestic savings (% of GDP)          1.0055      0.524      1.918      0.057      -0.031      2.042
Labor force, total                      8.865e-11   2.34e-10      0.378      0.706   -3.75e-10   5.52e-10
Net trade in goods and services (% of GDP)   -1.1513      0.629     -1.831      0.069      -2.395      0.093
Consumer price index (2010 = 100)         -0.0015      0.001     -1.840      0.068      -0.003      0.000
Foreign direct investment, net (% of GDP)   -0.3765      1.018     -0.370      0.712      -2.391      1.638
developing_dummy                           0.1178      0.105      1.118      0.266      -0.091      0.326
underdeveloped_dummy                       0.2097      0.135      1.555      0.122      -0.057      0.476
==============================================================================
Omnibus:                       59.521   Durbin-Watson:                   1.694
```

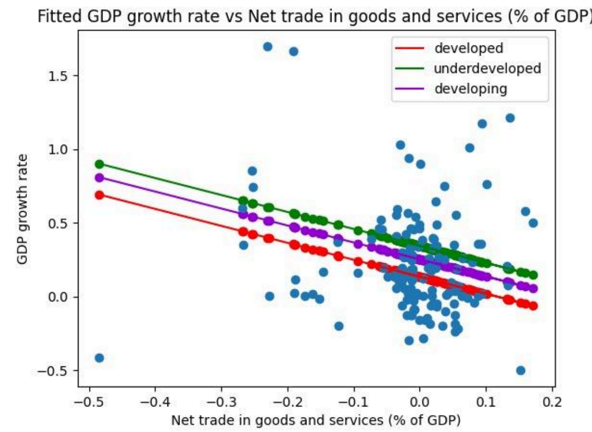
We notice that the R-squared value is 0.105, which indicates that the model explains only 10.5% of the variation in the dependent variable. Also, the p-value of the F-statistic is 0.0605 which means that we cannot reject the null hypothesis. None of the coefficients are statistically significant at the 5% level. Gross domestic savings(% of GDP), Net trade(% of GDP) and Consumer price index are significant at 90% confidence interval this time. The Durbin-Watson statistic of 1.694 suggests that there is no significant autocorrelation.



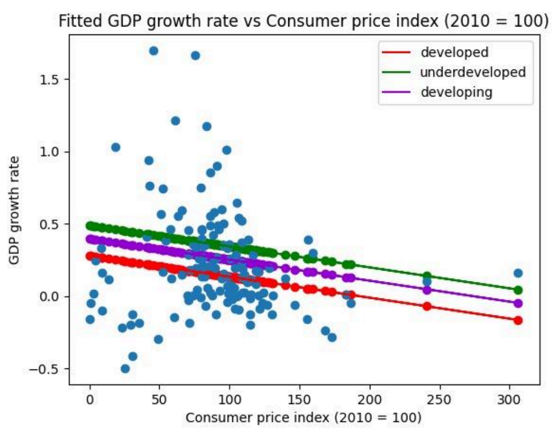




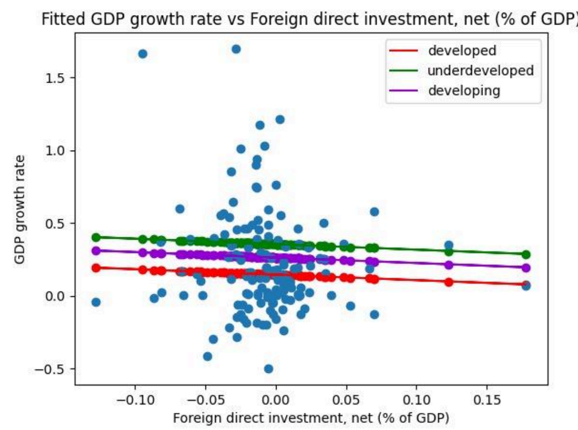
Fitted GDP growth rate vs Gross domestic savings (% of GDP)
developed
underdeveloped
developing
GDP growth rate
Gross domestic savings (% of GDP)



Fitted GDP growth rate vs Net trade in goods and services (% of GDP)
developed
underdeveloped
developing
GDP growth rate
Net trade in goods and services (% of GDP)



Fitted GDP growth rate vs Consumer price index (2010 = 100)
developed
underdeveloped
developing
GDP growth rate
Consumer price index (2010 = 100)



Fitted GDP growth rate vs Foreign direct investment, net (% of GDP)
developed
underdeveloped
developing
GDP growth rate
Foreign direct investment, net (% of GDP)

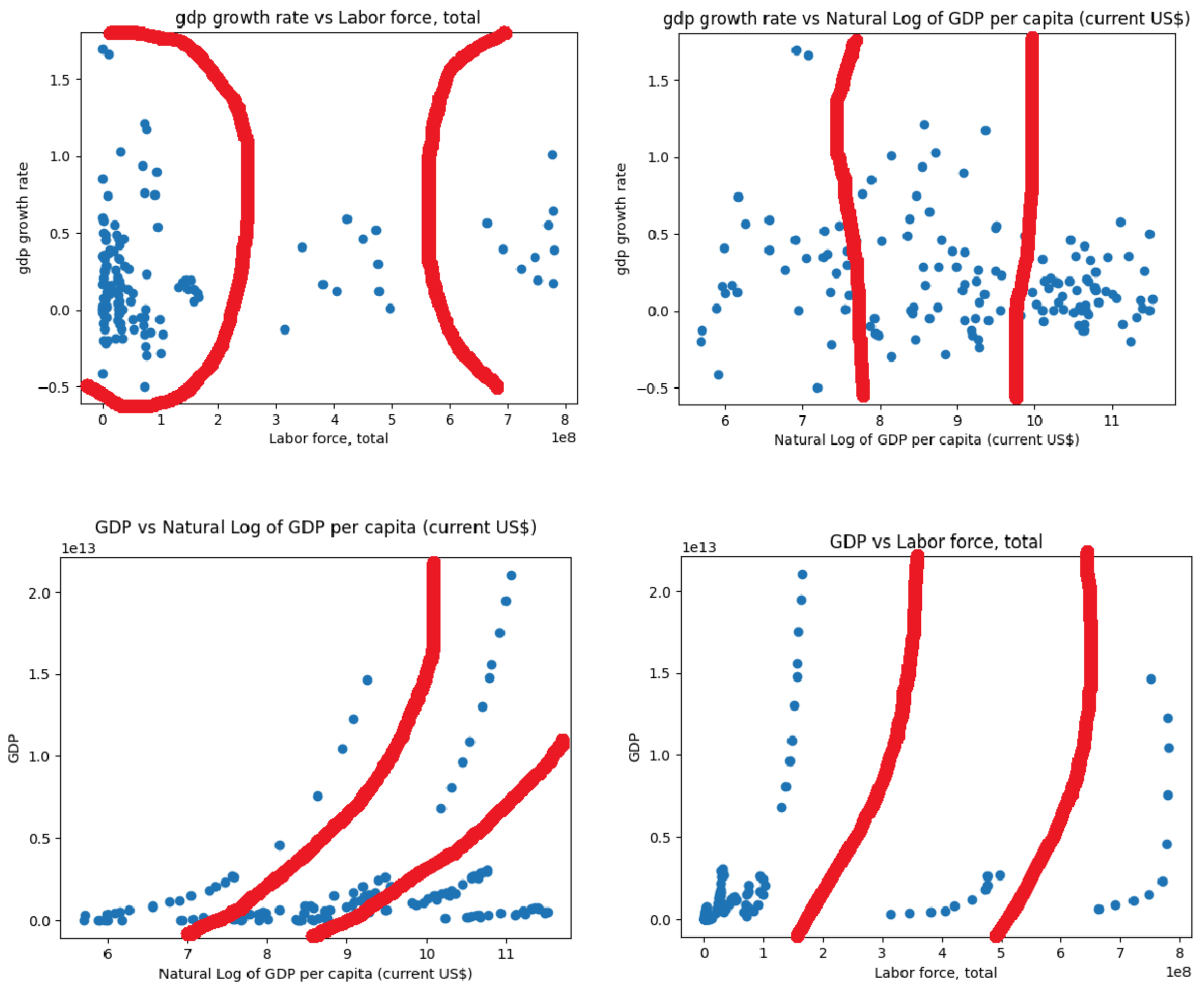
Model 3: OLS, metrics: %ages of GDP with exog:dummy interactions

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.63637 -0.15495 -0.04515  0.13583  1.48002 

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)  
(Intercept)                         -4.136e-01  5.850e-01  -0.707   0.4809  
lnPCGDP                              2.642e-02  5.503e-02   0.480   0.6320  
grossDomSavings                      1.715e+00  6.713e-01   2.555   0.0119 *
totalLaborForce                      2.331e-09  2.707e-09   0.861   0.3909  
NetTradePercent                     -1.732e+00  7.472e-01  -2.318   0.0221 *
CPI                                 -1.792e-03  1.667e-03  -1.075   0.2847  
developingDummy                     -3.508e-01  9.979e-01  -0.352   0.7258  
underdevelopedDummy                 -1.144e+00  1.645e+00  -0.695   0.4882  
lnPCGDP:totalLaborForce             -1.458e-10  2.421e-10  -0.602   0.5480  
lnPCGDP:developingDummy              9.996e-02  1.294e-01   0.772   0.4415  
lnPCGDP:underdevelopedDummy          2.840e-01  1.893e-01   1.500   0.1363  
grossDomSavings:developingDummy     -3.254e-01  1.043e+00  -0.312   0.7556  
grossDomSavings:underdevelopedDummy -4.276e+00  1.988e+00  -2.150   0.0335 *
totalLaborForce:developingDummy     -1.088e-09  1.097e-09  -0.992   0.3233  
totalLaborForce:underdevelopedDummy -4.068e-08  9.633e-08  -0.422   0.6736  
CPI:developingDummy                 -1.852e-03  2.240e-03  -0.827   0.4100  
CPI:underdevelopedDummy              1.806e-03  3.670e-03   0.492   0.6235  
developingDummy:FDIPercent           2.599e+00  3.142e+00   0.827   0.4097  
underdevelopedDummy:FDIPercent      -5.456e+00  4.716e+00  -1.157   0.2496  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3281 on 121 degrees of freedom
Multiple R-squared:  0.2493,	Adjusted R-squared:  0.1376 
F-statistic: 2.233 on 18 and 121 DF,  p-value: 0.005385
```

The model has far more terms due to the dummy:exogenous interaction terms that try to capture the variation of changes in variables according to level of development. In this case, the **R-squared value is 0.2493** which is far better than the previous models. It still suggests that a substantial portion of the variance is not explained by the model. The p-value of F-statistic here is 0.0054 which indicates that we can reject the null hypothesis that the model is insignificant both at the 5% and 1% significance levels. The coefficients of Gross domestic savings(% of GDP), Net trade(% of GDP) and Gross domestic savings(underdeveloped dummy) are statistically significant at 5% significance level.

Problem 2: Heteroskedasticity

Heteroskedasticity is said to occur when different datapoints have different variances - that is, the co-variances are not independent and identically distributed, violating one of the assumptions of OLS models. This leads to erroneous calculations of significance. Looking at some of the previous plots, we notice that there exist some irregularities:

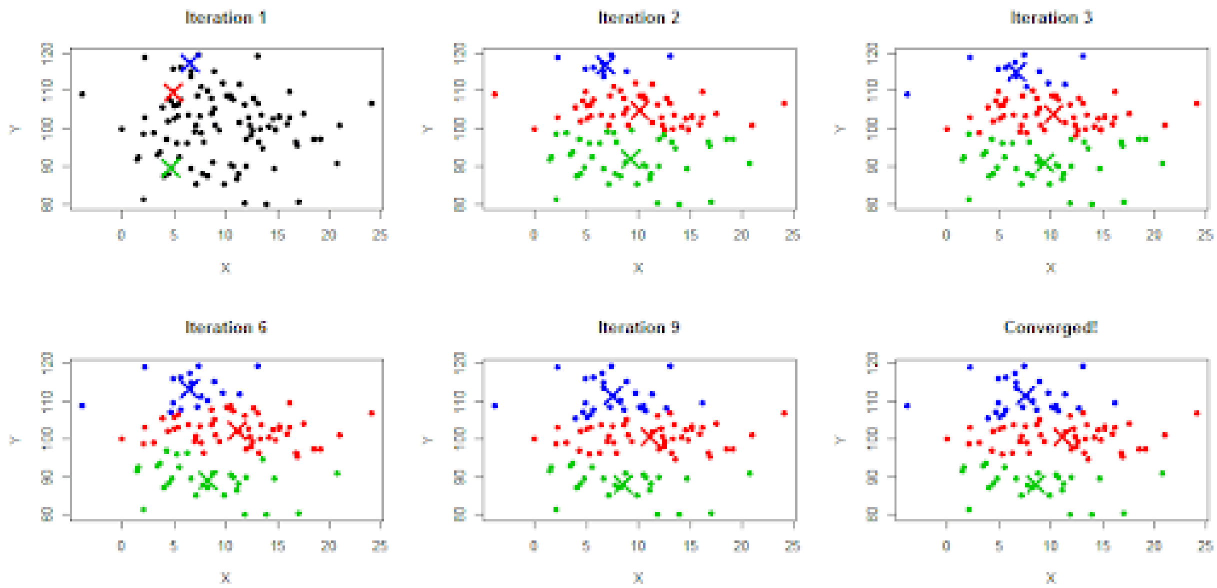


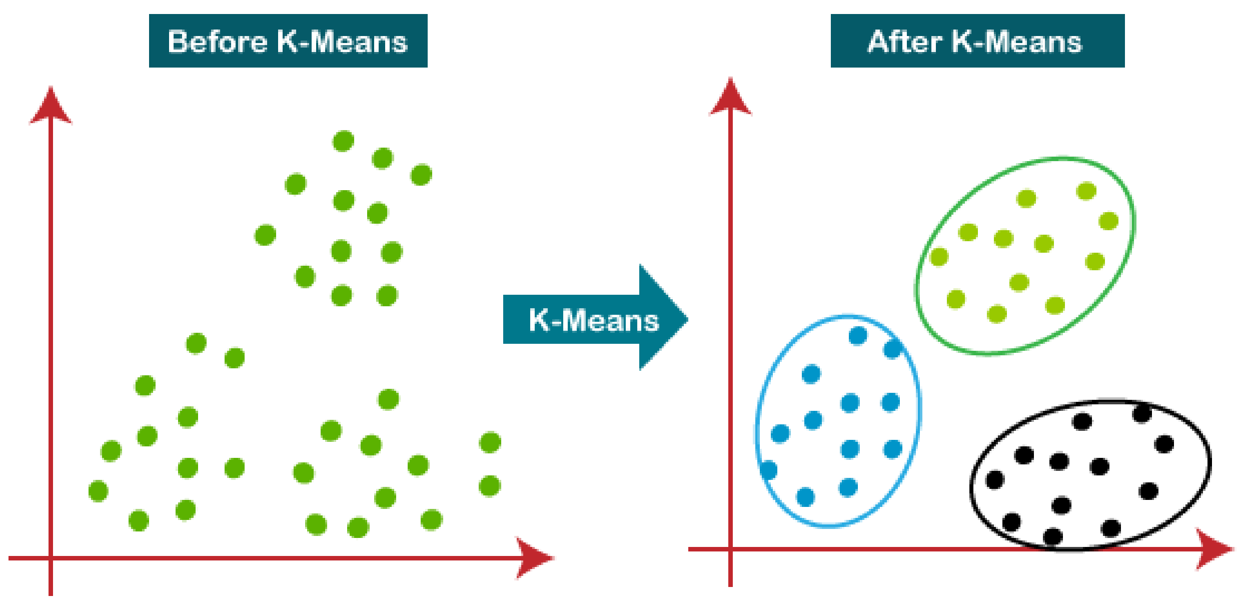
We can see that the datapoints can be divided into 3 regions, each of which has a conditional variance that differs from that of the other two. So, this allows us to hypothesize that developing, underdeveloped and developed nations all show different covariances with respect to each other, thus forming covariance groups, or covariance "clusters". This cluster-based covariance allows us to differentiate between the three groups and accommodate for the differences in covariances by calculating cluster-based-covariance robust standard errors.

Introduction to K-Means clustering

K-means clustering is an unsupervised ML algorithm that allows one to cluster together similar data. The goal is to partition the data into K distinct clusters, each having their own centroids, or mean value, hence

giving the name K-Means. Here, we use a K-Means model to find 3 distinct clusters by setting K as 3 and passing it the dataset. We then obtain the labels - that is, the numbers that signify which cluster a datapoint corresponds to - and pass them to the OLS model, specifying the covariance condition to be "cluster" instead of "non-robust", as in models 1 & 2.
Given below are visualizations of how the algorithm works.





Model 4: OLS with cluster-robust standard errors

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        gdp growth rate   R-squared:                       0.069
Model:                            OLS   Adj. R-squared:                  0.027
Method:                 Least Squares   F-statistic:                     2081.
Date:                Thu, 25 Apr 2024   Prob (F-statistic):           0.000480
Time:                        10:33:15   Log-Likelihood:                -47.512
No. Observations:                 140   AIC:                             109.0
Df Residuals:                     133   BIC:                             129.6
Df Model:                           6
Covariance Type:              cluster
========================================================================================================================
                                                         coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------------
const                                                  0.4444      0.076      5.874      0.000       0.296       0.593
Natural Log of GDP per capita (current US$)           -0.0171      0.009     -1.888      0.059      -0.035       0.001
Gross domestic savings (current US$)                8.092e-17   2.92e-14      0.003      0.998   -5.72e-14    5.74e-14
Labor force, total                                   1.98e-10   1.38e-10      1.438      0.150   -7.19e-11    4.68e-10
Net trade in goods and services (BoP, current US$)   1.62e-13   1.58e-13      1.024      0.306   -1.48e-13    4.72e-13
Consumer price index (2010 = 100)                     -0.0010   6.36e-05    -16.171      0.000      -0.001      -0.001
Foreign direct investment, net (BoP, current US$) -4.427e-13   2.34e-13     -1.889      0.059   -9.02e-13    1.66e-14
==============================================================================
Omnibus:                       44.822   Durbin-Watson:                   1.760
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              115.071
Skew:                           1.275   Prob(JB):                     1.03e-25
```

We notice that the coefficients for the model are the same as those of Model 1. However, the **F-value is now significant at a 99.99%** significance level. Additionally, the **significance of all coefficients have increased -** the const coefficient and CPI coefficient are significant at a 99.99% level, the coefficients of natural log of PCGDP and FDI are almost at a 95% level, and labor force at an 85% level. However, the gross domestic savings continues to be insignificant.

Since the coefficients are the same, the plots/fitted line will also remain the same as those of Model 1.

Model 5: OLS with dummy vars, metrics: %ages of GDP, cluster-robust standard errors

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        gdp growth rate   R-squared:                       0.105
Model:                            OLS   Adj. R-squared:                  0.051
Method:                 Least Squares   F-statistic:                     14.21
Date:                Thu, 25 Apr 2024   Prob (F-statistic):             0.0657
Time:                        10:33:16   Log-Likelihood:                -44.690
No. Observations:                 140   AIC:                             107.4
Df Residuals:                     131   BIC:                             133.9
Df Model:                           8
Covariance Type:              cluster
=================================================================================================
                                                 coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
const                                         -0.2285      0.176     -1.295      0.195      -0.574       0.117
Natural Log of GDP per capita (current US$)    0.0263      0.014      1.889      0.059      -0.001       0.054
Gross domestic savings (% of GDP)              1.0055      0.260      3.864      0.000       0.495       1.515
Labor force, total                          8.865e-11   4.39e-11      2.020      0.043    2.62e-12    1.75e-10
Net trade in goods and services (% of GDP)    -1.1513      0.271     -4.254      0.000      -1.682      -0.621
Consumer price index (2010 = 100)             -0.0015      0.000     -9.381      0.000      -0.002      -0.001
Foreign direct investment, net (% of GDP)     -0.3765      0.380     -0.992      0.321      -1.120       0.367
developing_dummy                               0.1178      0.011     10.870      0.000       0.097       0.139
underdeveloped_dummy                           0.2097      0.031      6.753      0.000       0.149       0.271
==============================================================================
Omnibus:                       59.521   Durbin-Watson:                   1.694
...
Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
```

We see that the model has a p(F-value) of 0.0657, nearly significant at the 95% level, around the same as that of Model 2, 0.0605. The R2 remains the same. Again, as in Model 3, the significance of coefficients is highly improved, with savings, net trade, CPI and the dummy coefficients being significant at the 99.99% level, and the labor force at a 95% level. However, even though the const and FDI coefficients show improvement, they still remain statistically insignificant at even a 90% level.

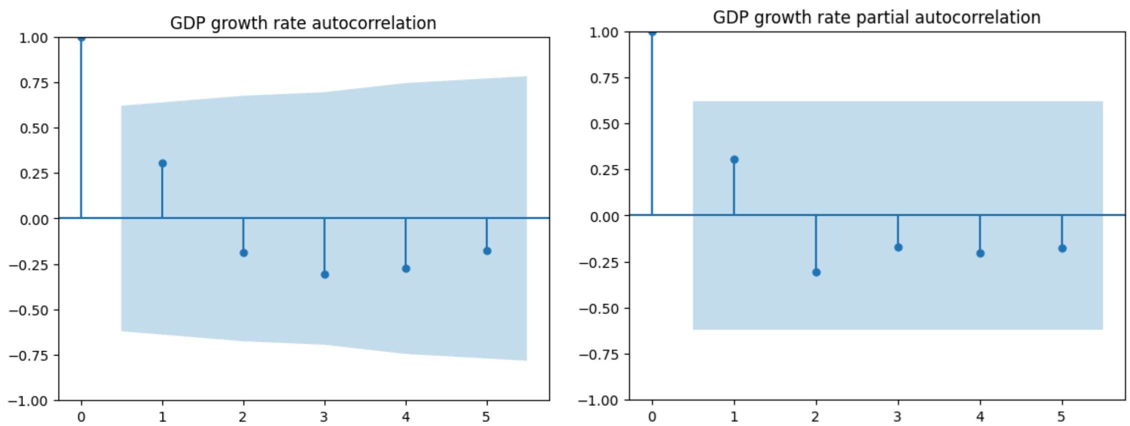
As with Model 3 analogous to Model 1, the plots/fitted line will be the same as those of Model 2.

Problem 3: Autocorrelation

Also known as serial correlation, this represents a degree of correlation between a datapoint and the time-lagged version of itself. That is, the datapoint at time *t* might be correlated with its own value at *t-1, t-2, ...* etc. The problem of autocorrelation might cause errors in calculating significance of coefficients. We try to solve this problem by fitting a new time-series model with regressors, ARIMAX.

Visualizing auto-correlation between GDP growth rates, averaged across nations per year

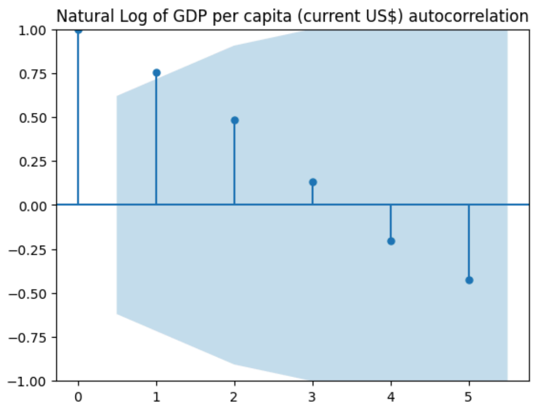
Here, we plot Auto-Correlation Function (corresponding to autoregression) and Partial Auto-Correlation Function (corresponding to moving average) of the GDP growth rates averaged by country, for each year.

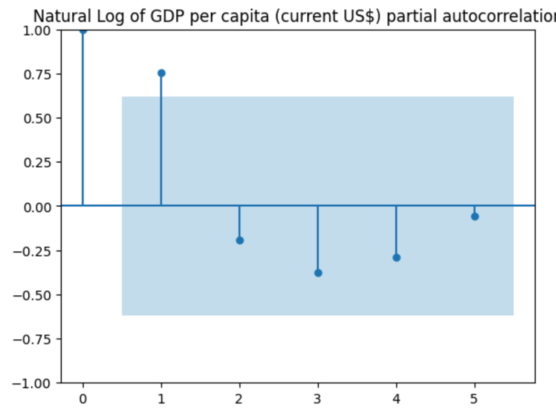


The x-axis represents the number of lagged periods to take into account (for us, 1 lagged period = 3 years). We see from the plots that gdp growth rate is autocorrelated, both absolutely and partially, by 1 lag period (3 years). Hence, growth rate at time *t* is correlated with its value at *t-1*.

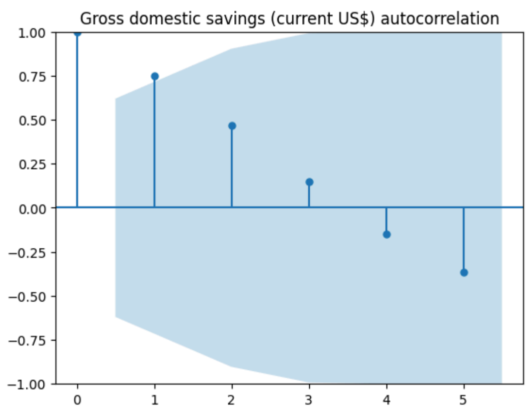
Visualizing auto-correlation between exogenous variables, averaged across nations per year

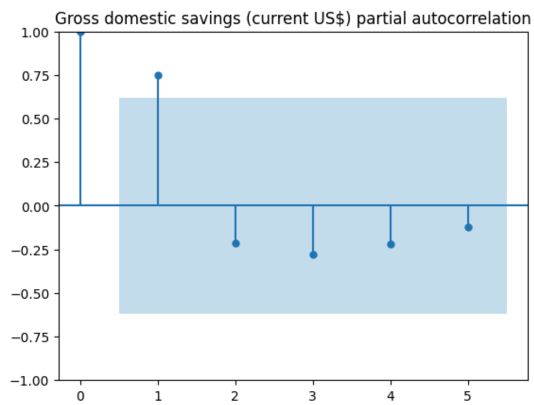
1. Natural log of per-capita GDP:

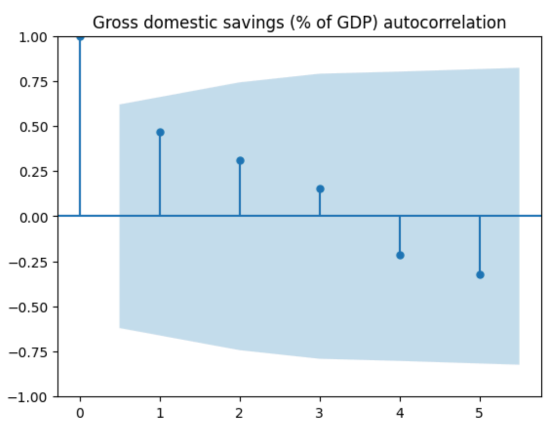


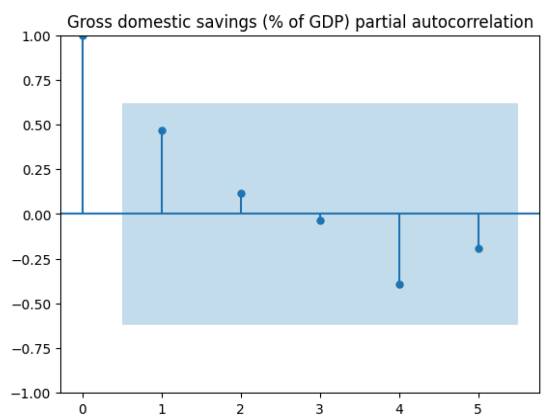


2. Gross domestic savings (current USD & %age of GDP):

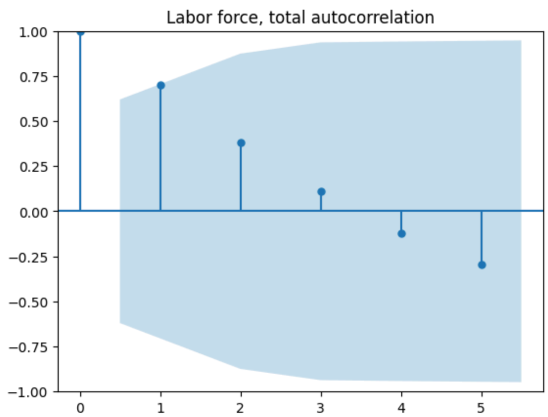


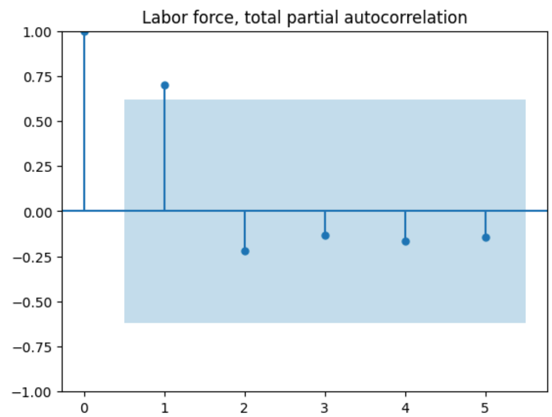




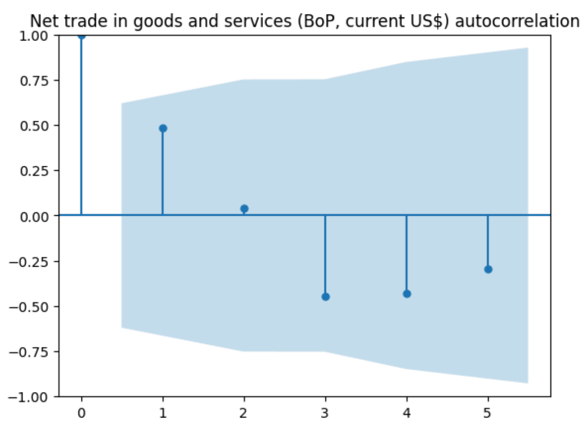


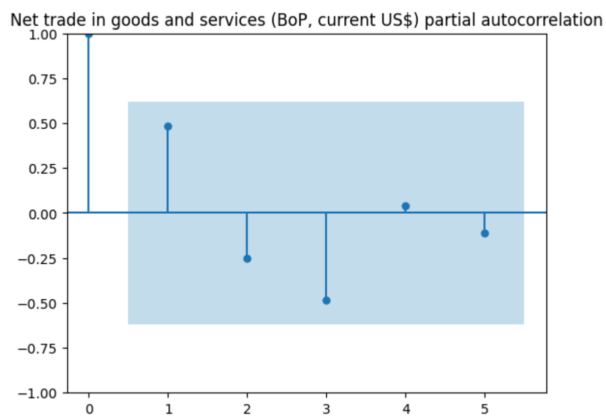
3. Labor force (Total):

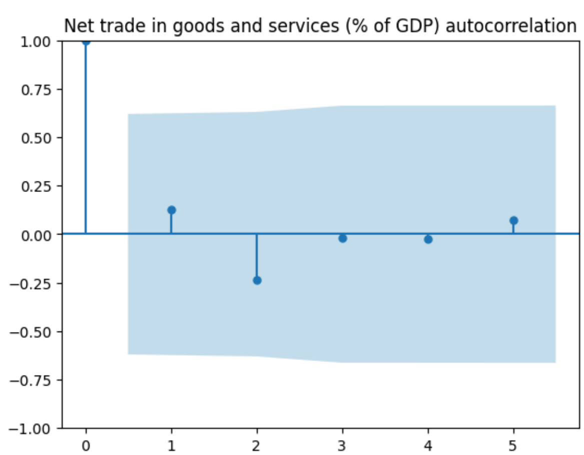


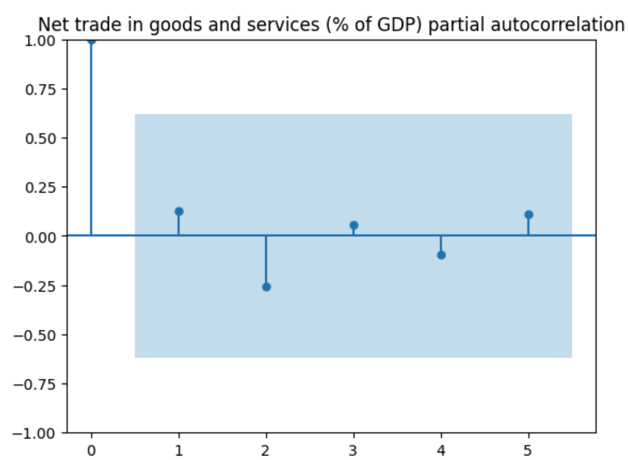


4. Net trade (current USD and %age of GDP):

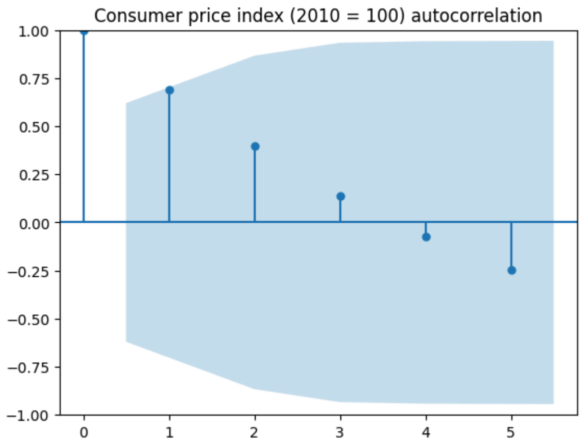


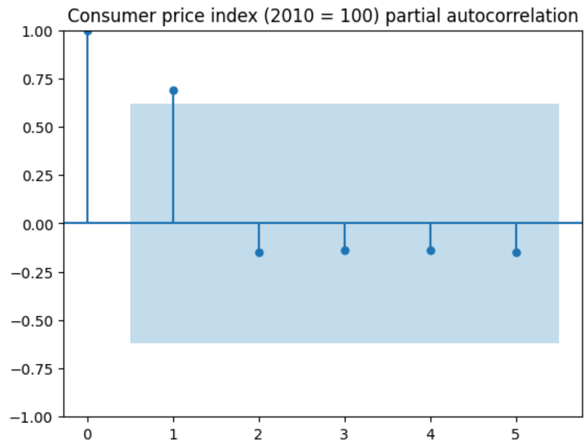




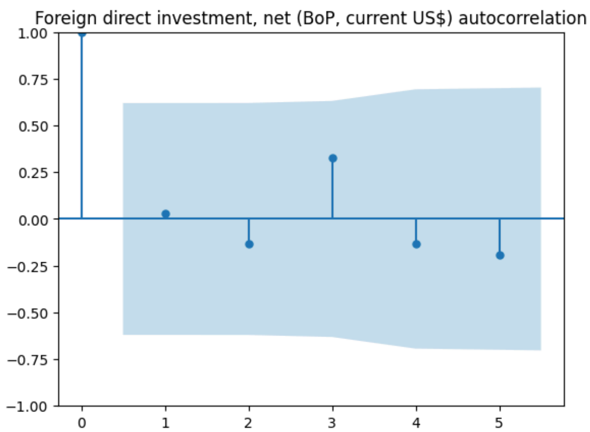


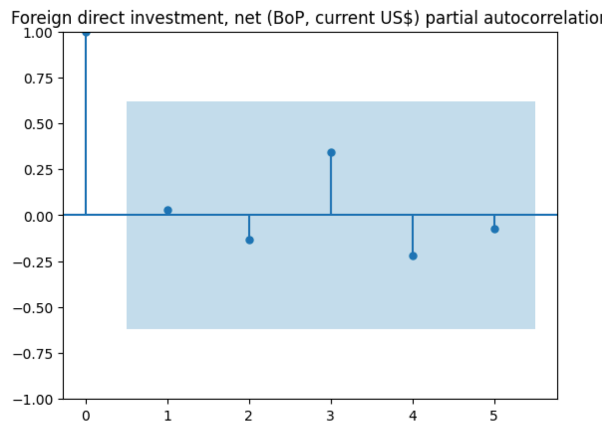
5. Consumer Price Index:

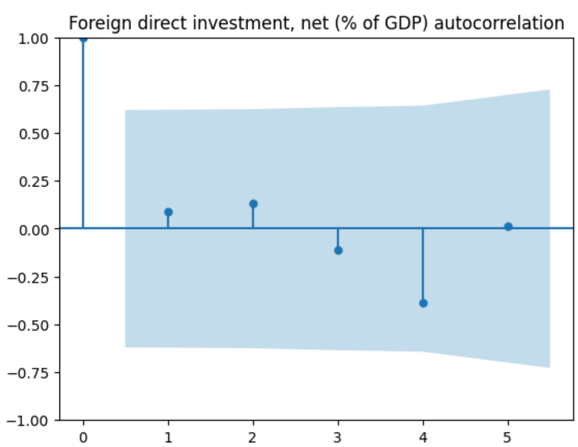


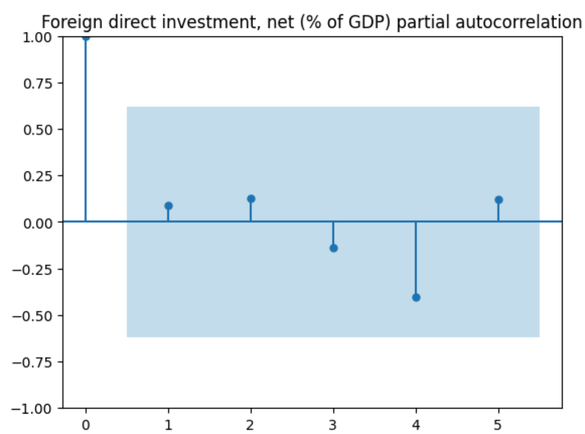


6. Net Foreign Direct Investment (current USD and %age of GDP)









Introduction to the ARIMAX model

ARIMAX stands for Auto Regressive Integrated Moving Average - Exogenous. ARIMA models are a series of models with 3 parameters - p, d & q that decide the Auto Regressive, Integrated & Moving Average parts of the model. The family is represented by the functional form ARIMA(p, d, q). The parameters are explained below:

1. p (AR term): It signifies the number of lagged periods the current period is correlated with. For instance, p=2 signifies that the datapoint at time *t* is dependent on that of times *t-1* and *t-2*.
2. d (I term): It signifies the differencing term - differences in timesteps. This is equivalent to transforming a variable *Y(t)* as *Z(t) = Y(t) - Y(t-1)*
3. q (MA term): It signifies size of MA window - the number of preceding periods for which we need to take the moving average of the model.

From the above autocorrelation plots, we see that a lag of 1 is significant for most variables for both AR and MA terms. Hence, we opt for an ARIMAX(1,0,1) model, keeping a simple 0-order integration term and single lag AR and MA terms. The X simply signifies that we would like to pass our exogenous variables to obtain coefficient estimates and autocorrelation-corrected robust standard errors.

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p}$$

AR equation

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \ldots + \omega_q \varepsilon_{t-q} + \varepsilon_t$$

MA equation

Combined ARIMA equation, *d*=0:

$$\begin{aligned} Y_t = {} & (\beta_1 + \beta_2) \\ & + (\Phi_1 Y_{t-1} + \ldots + \Phi_p Y_{t-p}) \\ & + (\omega_1 \varepsilon_{t-1} + \ldots + \omega_q \varepsilon_{t-q} + \varepsilon_t) \end{aligned}$$

Model 6: ARIMAX(1, 0, 1), metrics in current USD

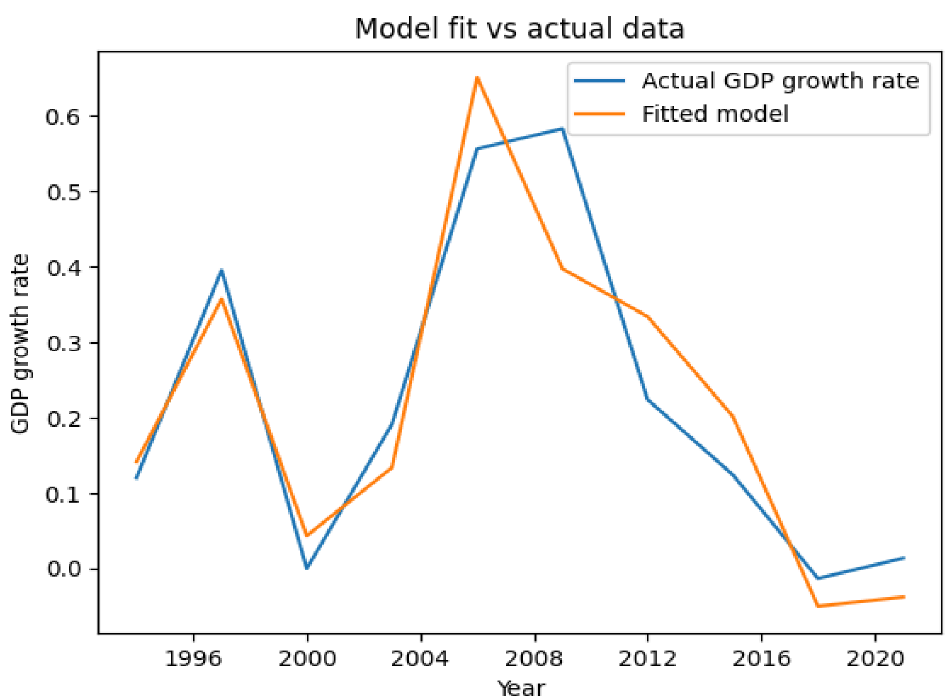
```
                               SARIMAX Results
==============================================================================
Dep. Variable:        gdp growth rate   No. Observations:                   10
Model:                 ARIMA(1, 0, 1)   Log Likelihood                  17.394
Date:                Thu, 25 Apr 2024   AIC                            -14.788
Time:                        10:33:20   BIC                            -11.762
Sample:                    12-31-1993   HQIC                           -18.108
                         - 12-31-2020
Covariance Type:                  opg
=======================================================================================================================
                                                         coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------------------
const                                                -13.3843   5.59e-18  -2.39e+18      0.000     -13.384     -13.384
Natural Log of GDP per capita (current US$)            2.3266   4.44e-17   5.25e+16      0.000       2.327       2.327
Gross domestic savings (current US$)               -6.054e-12   1.24e-13    -48.795      0.000    -6.3e-12   -5.81e-12
Labor force, total                                 -5.978e-08   5.18e-10   -115.327      0.000   -6.08e-08   -5.88e-08
Net trade in goods and services (BoP, current US$)  6.059e-12   1.77e-12      3.427      0.001    2.59e-12    9.52e-12
Consumer price index (2010 = 100)                      0.0323   1.28e-16   2.53e+14      0.000       0.032       0.032
Foreign direct investment, net (BoP, current US$)  -7.937e-12   6.82e-12     -1.164      0.245   -2.13e-11    5.43e-12
ar.L1                                                 -0.3918   9.12e-20    -4.3e+18      0.000      -0.392      -0.392
ma.L1                                                 -0.2215   1.39e-19    -1.6e+18      0.000      -0.221      -0.221
sigma2                                                 0.0021   8.01e-18   2.59e+14      0.000       0.002       0.002
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):                 2.56
Prob(Q):                              0.86   Prob(JB):                         0.28
```

We see that for the average values of variables over time, the coefficients and results are highly significant at the 99.99%, except for Net Trade (current USD), significant at the 99% confidence level and FDI (current USD) which is barely significant. Additionally, ar.L1 corresponds to the autoregressive lag value for the *t-1* period, and ma.L1 corresponds to the moving average lag value at *t-1*.

Model 7: ARIMAX(1, 0, 1), metrics as %age of GDP

```
                               SARIMAX Results
==============================================================================
Dep. Variable:        gdp growth rate   No. Observations:                   10
Model:                 ARIMA(1, 0, 1)   Log Likelihood                   8.939
Date:                Thu, 25 Apr 2024   AIC                              2.122
Time:                        10:33:23   BIC                              5.148
Sample:                    12-31-1993   HQIC                            -1.197
                         - 12-31-2020
Covariance Type:                  opg
==================================================================================================================
                                                  coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------
const                                          -0.2556   2.65e-06  -9.65e+04      0.000      -0.256      -0.256
Natural Log of GDP per capita (current US$)    -0.7436   2.75e-06   -2.7e+05      0.000      -0.744      -0.744
Gross domestic savings (% of GDP)              40.2278   5.46e-07   7.37e+07      0.000      40.228      40.228
Labor force, total                          -2.792e-08     1.5e-09    -18.559      0.000   -3.09e-08    -2.5e-08
Net trade in goods and services (% of GDP)    -25.8196   2.96e-07  -8.73e+07      0.000     -25.820     -25.820
Consumer price index (2010 = 100)               0.0063      0.002      3.728      0.000       0.003       0.010
Foreign direct investment, net (% of GDP)       6.7119   4.27e-07   1.57e+07      0.000       6.712       6.712
ar.L1                                           0.5066   6.25e-06   8.11e+04      0.000       0.507       0.507
ma.L1                                               -0   7.62e-06         -0      1.000   -1.49e-05    1.49e-05
sigma2                                          0.0038      0.001      3.684      0.000       0.002       0.006
===================================================================================
Ljung-Box (L1) (Q):                   4.20   Jarque-Bera (JB):                 0.58
Prob(Q):                              0.04   Prob(JB):                         0.75
```

We see that, when taken as %ages of GDP, all coefficients for the exogenous variables are significant at the 99.99% level of confidence. However, when taken as averages, we notice that the model has no moving average lag component for the *t-1* period, as seen above with a coefficient of 0 and no significance at any confidence level.



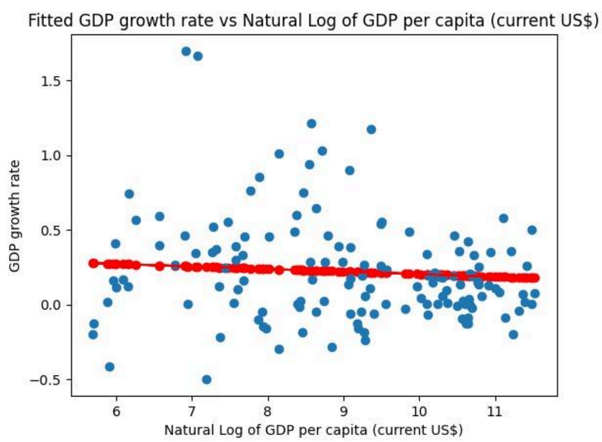
Summary Statistics of Models

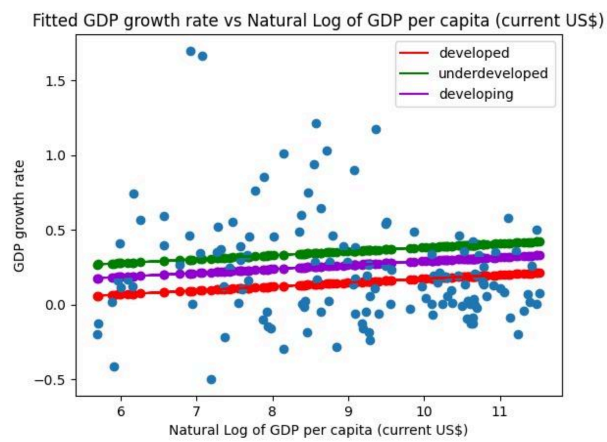
Model	R^2 (%)	P(F-val)	β_0	β_1	β_2	β_3	β_4	β_5	β_6	δ_1	δ_2
1	6.9	0.142	0.444	-0.017	-8.092 * e-17	-1.98* e-10	-1.62* e-13	-0.001	-4.427 * e-13	-	-
2	10.5	0.0605	-0.228	0.026	1.005	-8.865 * e-11	-1.151	-0.001	-0.376	0.117	0.209
3	24.9	0.00538	-4.136 *e-01	-2.642 *e-02	1.715 *e+00	-2.331 *e-09	-1.732 *e-00	-1.792 *e-01	-	-0.3508 *e-01	-1.44 4*e-0
4	6.9	0.0657	0.444	-0.017	-8.092 * e-17	-1.98* e-10	-1.62* e-13	-0.001	-4.427 * e-13	-	-
5	10.5	0.00048	-0.228	0.026	1.005	-8.865 * e-11	-1.151	-0.001	-0.376	0.117	0.209
6	-	-	-13.38	2.326	-6.1* e-12	-5.97* e-8	6.1*e-12	.323	-7.9* e-12	-	-
7	-	-	-.256	-.744	40.23	-2.8* e-8	-25.8	.0063	6.712	-	-

Interpretation of Results

In this section, we interpret the results of Models 1 & 4 and 2 & 5, since they are in the scope of this report, are fit on all relevant data, and possess the same coefficients as their non-robust counterparts. We will be interpreting these results both numerically and graphically, with Models 1&4 on the left and 2&5 on the right under their relevant exogenous variables.

Effect of ln(PCGDP) on GDP growth rate

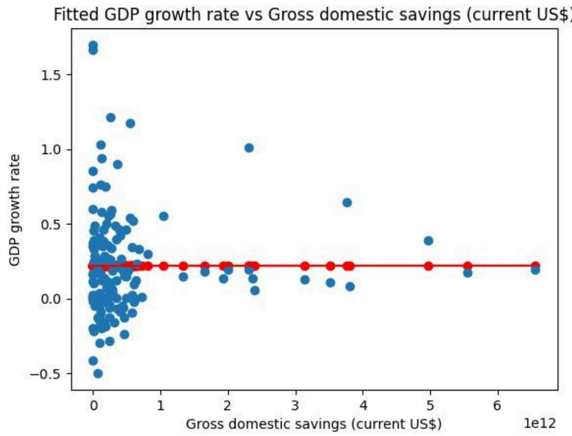


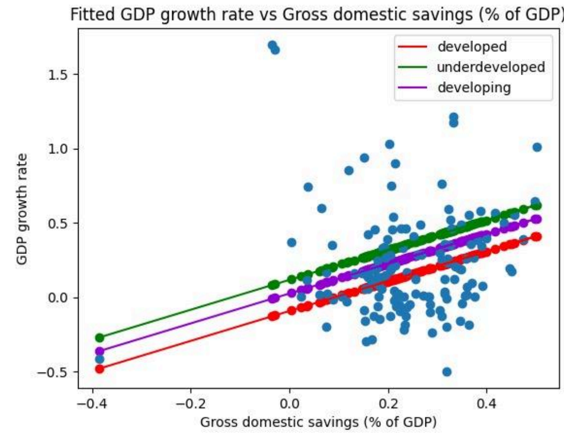


Models 1 & 2 appear to disagree on the effect of ln(PCGDP) on the GDP growth rate. The model without dummies regressed on terms with the current USD metric appears to show a negative trend as PCGDP increases, whereas the model with dummies regressed on variables with the %age of GDP metric show a positive correlation between ln(PCGDP) and GDP growth rate. However, we explain the result of Model 1 as erroneous, with the same proof of multicollinearity with gross savings in USD as explained during the presentation (the proof for the same can be found at the end of this Python notebook). Hence, we take the results of the second model, which is also more statistically significant, to show that the rate of increase of GDP growth rate with respect to PCGDP is 0.0263 divided by the PCGDP at the point.

$$\Delta GDP\ growth\ rate/\Delta PCGDP\ =\ 0.0263/PCGDP$$

Effect of Gross Domestic Savings on GDP growth rate

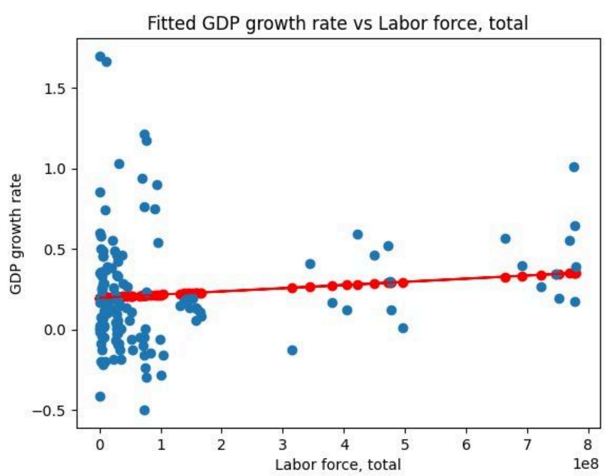


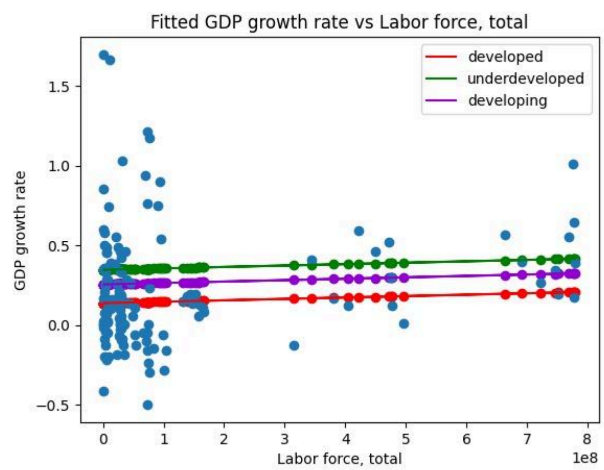


Again, the problem of multicollinearity as described above makes models 1 & 2 disagree on the effect of gross domestic savings on GDP growth rate. We take the results of the second model to be accurate, and postulate that with an increase of 1% in gross domestic savings relative to GDP, the GDP growth rate increases by a factor of 1.055.

$$\Delta GDP\ growth\ rate / \Delta Gross\ Domestic\ Savings\ (\%age\ of\ GDP)\ =\ 1.055$$

Effect of Total Labor Force on GDP growth rate

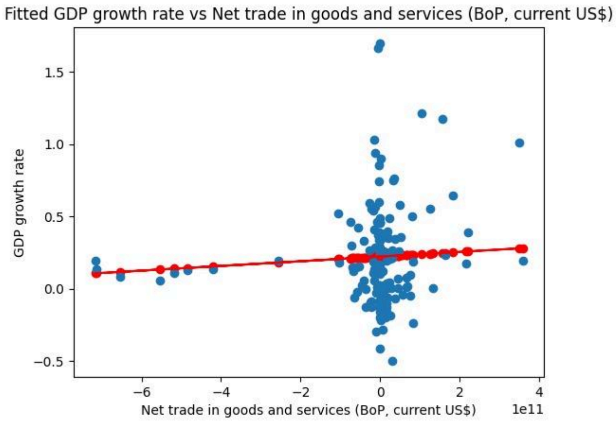


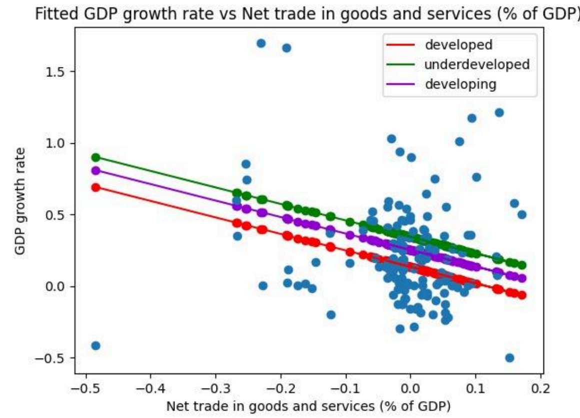


Both models agree on the fact that GDP growth rate has a slightly positive effect on the GDP growth rate of a country. They signify that an increase in the labor force by 100 million laborers increases GDP growth rate by a factor of 0.008865 to 0.0198.

$$\Delta GDP\ growth\ rate / \Delta Total\ Labour\ Force\ (in\ 100M)\ \approx\ 0.01\ to\ 0.02$$

Effect of Net Trade of Goods & Services on GDP growth rate

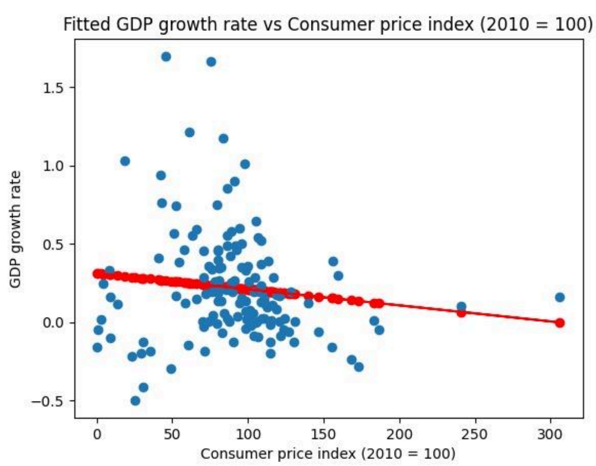


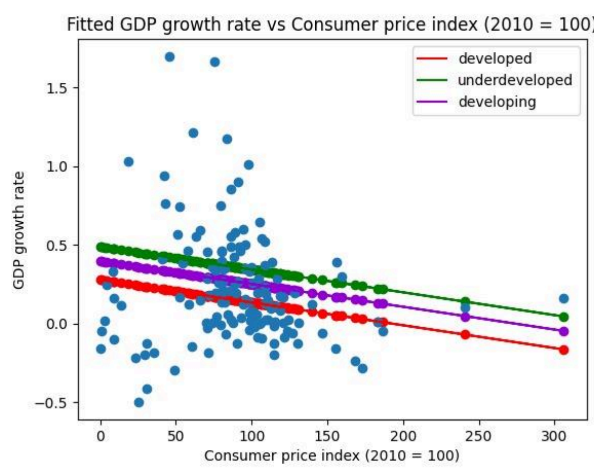


The models appear to tell us two different sides of the same story. The model fitted to net trade in current USD says that with an increase of $1Billion in net trade, the GDP growth rate increases by a factor of 0.000162. The model on the right tells us that with an increase in net trade of 1% relative to GDP, the growth rate decreases by a factor of 1.1513.

$$\Delta GDP\ growth\ rate/\Delta Net\ trade\ (1B\ USD) \ = \ 0.000162$$
$$\Delta GDP\ growth\ rate/\Delta Net\ trade\ (\%age\ of\ GDP) \ = \ 1.1513$$

Effect of Consumer Price Index on GDP growth rate

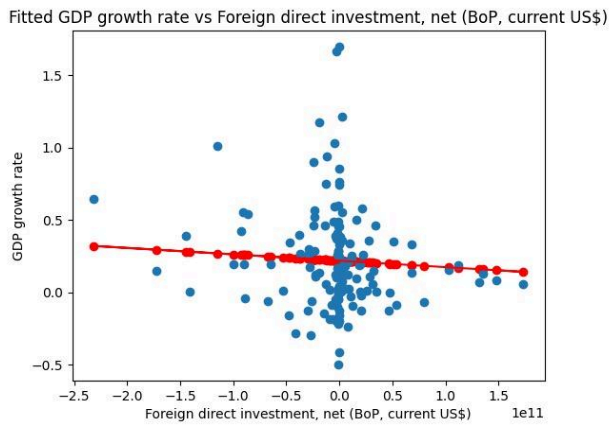


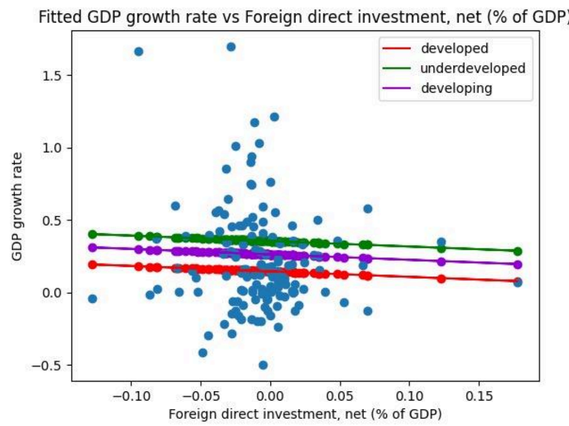


Both models strongly agree that GDP growth rate decreases with an increase in Consumer Price Index, by factors of 0.001 and 0.0015 respectively.

$$\Delta GDP\ growth\ rate/\Delta CPI \ = \ 0.001\ to\ 0.0015$$

Effect of Foreign Direct Investment on GDP growth rate





The model on the left tells us that with an increase of $1Billion in FDI, the GDP growth rate decreases by a factor of 0.00044. The model on the right shows us that with an FDI increase of 1% relative to GDP, the GDP growth rate declines by a factor of 0.3765.

$$\Delta GDP\ growth\, rate/\Delta FDI(\$1\ Billion)\ =\ 0.00044$$
$$\Delta GDP\ growth\, rate/\Delta FDI(\%age\ of\ GDP)\ =\ 0.3765$$

Effect of Level of Development on GDP growth rate

All the plots on the right above display the underdeveloped regression line intercept as higher than the developing regression line intercept, which is in turn higher than the developed dummy intercept. This agrees with empirical evidence that underdeveloped countries show higher rates of growth than developed countries in their early stages, given the same parameters as each other.

Conclusion

After conducting an exhaustive regression analysis, our report shows a detailed understanding of the relationship between GDP growth rates and our selected variables.

Our analysis indicates that as per capita GDP increases, so does the growth rate, with the degree of growth being dependent upon the level of per capita GDP at any given time.

It also shows an increase in growth rate with an increase in domestic savings, which is consistent with the theoretical growth models that have been discussed.

Another notable deduction from our analysis is that while a substantial increase in the labor force is required to effect a marginal change in GDP, the significance of this variable becomes extremely pronounced when considering its autocorrelation. Even though the choice of labor force as a variable was controversial, the final heteroskedastic and time-series models indicate that it is an invaluable indicator.

As for net trade, we used both metrics, current value in USD and % relative to GDP, both of them resulting in different marginal effects on GDP growth rate. Our analysis also revealed a negative relationship between CPI and growth rate, indicating that as CPI increases, the growth rate tends to decrease. Similar to net trade, we used both metrics for FDI and reached the same conclusion, an increase in the FDI had a negative effect on the growth rate.

By analyzing these variables, our analysis not only shows their individual contributions but also displays their interrelation and collective impact on economic performance. These insights give policymakers and stakeholders a comprehensive understanding needed for formulating targeted strategies aimed at reaching sustainable and inclusive economic growth. By making use of these findings, economists can make more informed decisions, ultimately steering the economy towards economic and social prosperity.