ECONOMETRICS PROJECT SEMESTER- I

ESTIMATING SHARE OF AGRICULTURE, FORESTRY AND FISHING IN THE TOTAL GDP OF DIFFERENT COUNTRIES (AS A % OF GDP) USING CHOICE VARIABLES

*by **ANKAN DUTTA(12)***

MA ECONOMICS , BATCH: 2024-26
ROLL NO - 12

Project: https://colab.research.google.com/drive/1qv2OVhN-RQ1N1sqPhZRiEmF3Egkpkt1D?usp=sharing

PROBLEM STATEMENT:

This project is part of the Semester course wherein I have tried to determine the factors that can best explain the variations in the share of Agriculture, forestry, and fishing in a percentage of GDP, seen across countries.

In this pursuit, I have incorporated the use of OLS regression theory to determine a suitable model. The procedures followed have been chosen to look for discrepancies in our data that may affect the model conclusions.

The data used in this project consists of observations related to different aspects of a countries development collected across 57 countries, for the year 2018.

DATASET

METHODOLOGY

(*SOURCE: WORLD BANK , TIME FRAME- 2018)*

My main objective was to see how factors such as Employment in agriculture, Gross fixed capital formation, and Rural population of different countries can explain the variations of Government spending on Agriculture, forestry, and fishing(as a percentage of GDP) across different countries. In the given data I have taken all the major economic factors that are both direct and indirect indicators and which could have effect on Government spending on Agriculture, forestry, and fishing.

Dependent Variable

Agriculture, forestry, and fishing(as a percentage of GDP)- This variable directly measures the share of agriculture in the total GDP

Independent Variables :

- Employment in agriculture (as a percentage of total employment) (modeled from ILO estimate) - Reflects the labour dependency of the agricultural sector.

- Gross fixed capital formation (as a percentage of GDP)- Measures investment in infrastructure and machinery, critical for agricultural productivity.
- Rural population (as a percentage of total population) - Reflects the proportion of the population dependent on agriculture.

Here, Employment in agriculture have been considered as direct indicators of Government spending on Agriculture, forestry, and fishing.

Gross fixed capital formation and Rural population have also been included as they indirectly affect agricultural growth in a country. Where, Rural population is predicted to have a positive relationship with share of Agriculture, forestry, and fishing, on GDP

DESCRIPTIVE STATISTICS:

The OLS regression was done in Python, where the results are as shown:

```
                                        OLS Regression Results
===============================================================================================
Dep. Variable:    Agriculture, forestry, and fishing, value added (% of GDP)   R-squared:                  0.853
Model:                                                                  OLS   Adj. R-squared:             0.845
Method:                                                       Least Squares   F-statistic:                102.5
Date:                                                      Sat, 28 Dec 2024   Prob (F-statistic):      4.73e-22
Time:                                                              05:47:36   Log-Likelihood:           -129.63
No. Observations:                                                        57   AIC:                        267.3
Df Residuals:                                                            53   BIC:                        275.4
Df Model:                                                                 3
Covariance Type:                                                  nonrobust
=====================================================================================================================
                                                             coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------------------
const                                                      0.0518      1.599      0.032      0.974      -3.156       3.260
Employment in agriculture (% of total employment) (modeled ILO estimate)   0.3216      0.028     11.429      0.000       0.265       0.378
Gross fixed capital formation (% of GDP)                   0.0171      0.077      0.223      0.824      -0.137       0.171
Rural population (% of total population)                   0.0134      0.028      0.483      0.631      -0.042       0.069
==============================================================================
Omnibus:                       21.123   Durbin-Watson:                   1.857
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               66.463
Skew:                          -0.869   Prob(JB):                     3.70e-15
Kurtosis:                       7.996   Cond. No.                         225.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Key Metrics

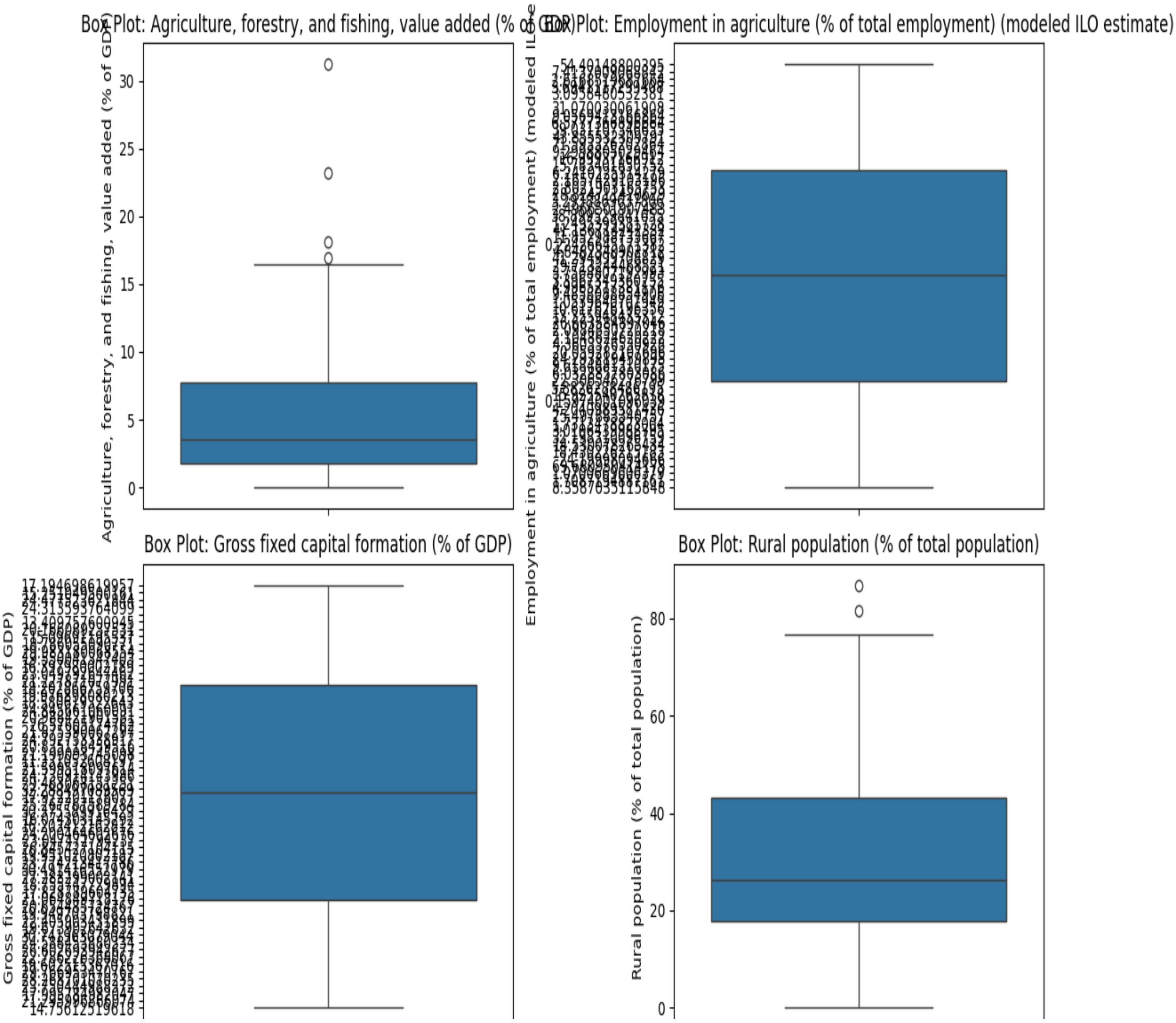
Metric	Value	Interpretation
R-squared	0.853	85.3% of the variation in the dependent variable is explained by the independent variables.
Adjusted R-squared	0.845	Adjusted for the number of predictors; the model explains most of the variability.
F-statistic	102.5	The overall model is highly significant (p-value <0.001).
AIC	267.3	Used to compare models; lower values indicate better fit.
BIC	275.4	Penalizes for the number of predictors; lower values suggest better simplicity.
Durbin-Watson	1.857	No significant autocorrelation in residuals (close to the ideal value of 2).

Coefficients and Significance

Variable	Coefficient	Std. Error	t-statistic	P-value	95% CI (Lower, Upper)	Interpretation
Intercept	0.0518	1.599	0.032	0.974	(-3.156, 3.260)	The intercept is not significant, meaning no standalone effect when predictors are zero.
Employment in agriculture (% of total employment)	0.3216	0.028	11.429	<0.001	(0.265, 0.378)	Strong, positive, and significant relationship. A 1% increase in employment is associated with a 0.32% increase in agriculture's GDP share.
Gross fixed capital formation (% of GDP)	0.0171	0.077	0.223	0.824	(-0.137, 0.171)	Not statistically significant. Minimal or no influence on agriculture's GDP share.
Rural population (% of total population)	0.0134	0.028	0.483	0.631	(-0.042, 0.069)	Not statistically significant. Little evidence of a meaningful effect.

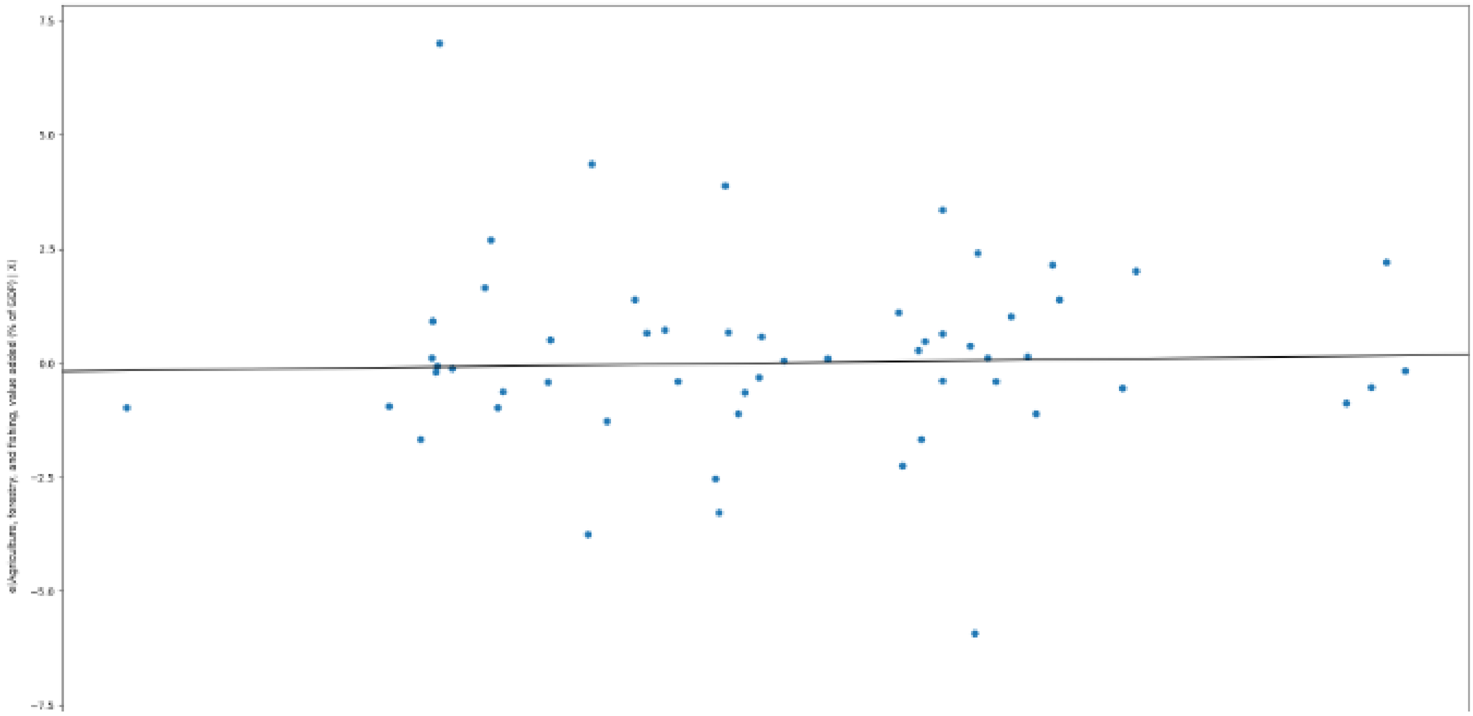
The summary statistics suggests that Employment in agriculture , 2018, Gross fixed capital formation, 2018, and Rural population ,2018, all these have positive coefficients which suggests that all of these explanatory variables have positive impact on the level of share of Agriculture, forestry, and fishing in a percentage of GDP, seen across countries. The overall fit of the model is quite moderately good as indicated by high values of R2 and adjusted R2, which are 0.853 and 0.845 respectively.

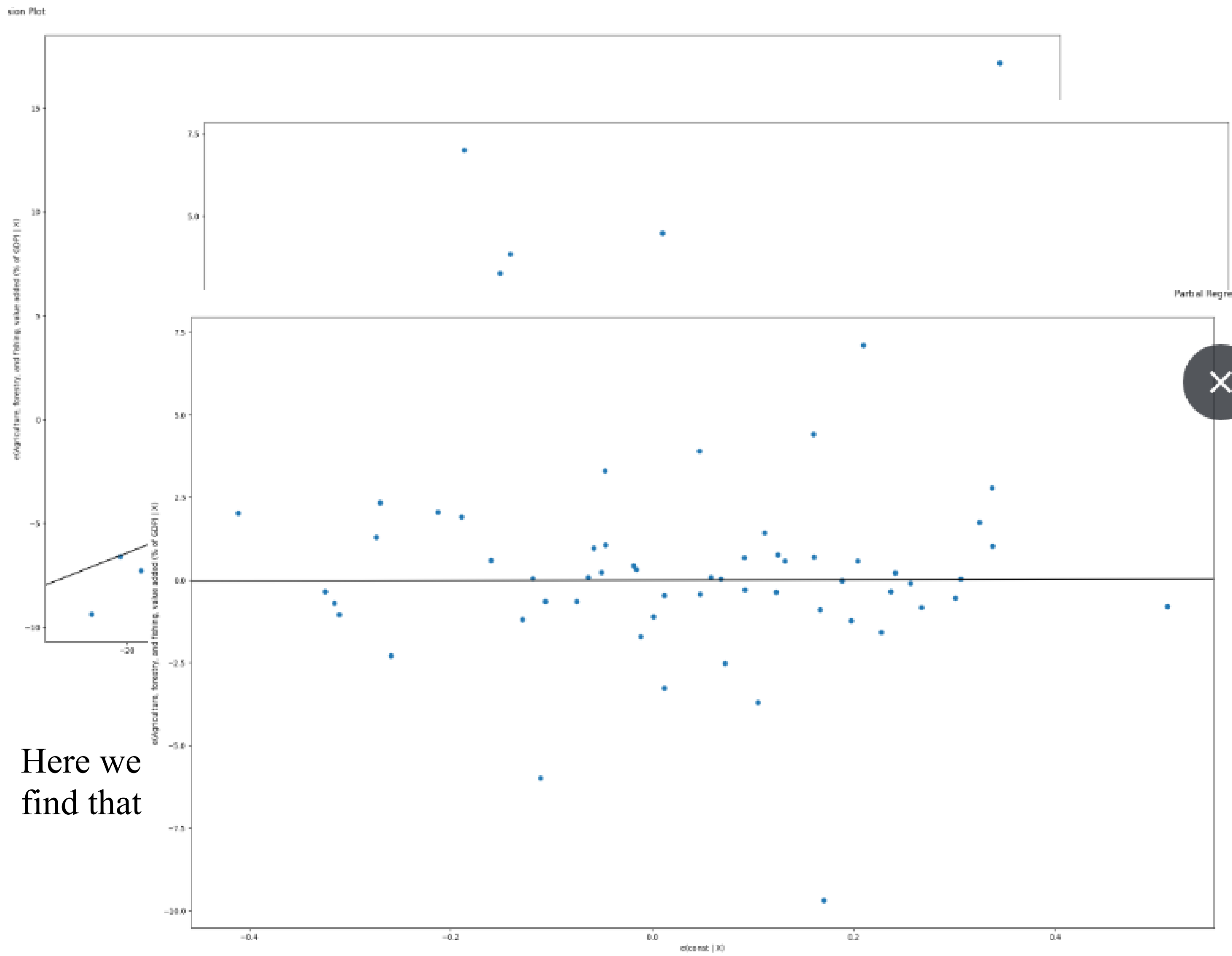
BOXPLOTS FOR RELEVANT VARIABLES



TESTING FOR LINEARITY OF A MODEL

We try to confirm visually if the dependent and independent variables have a linear relationship via partial regressor plots.





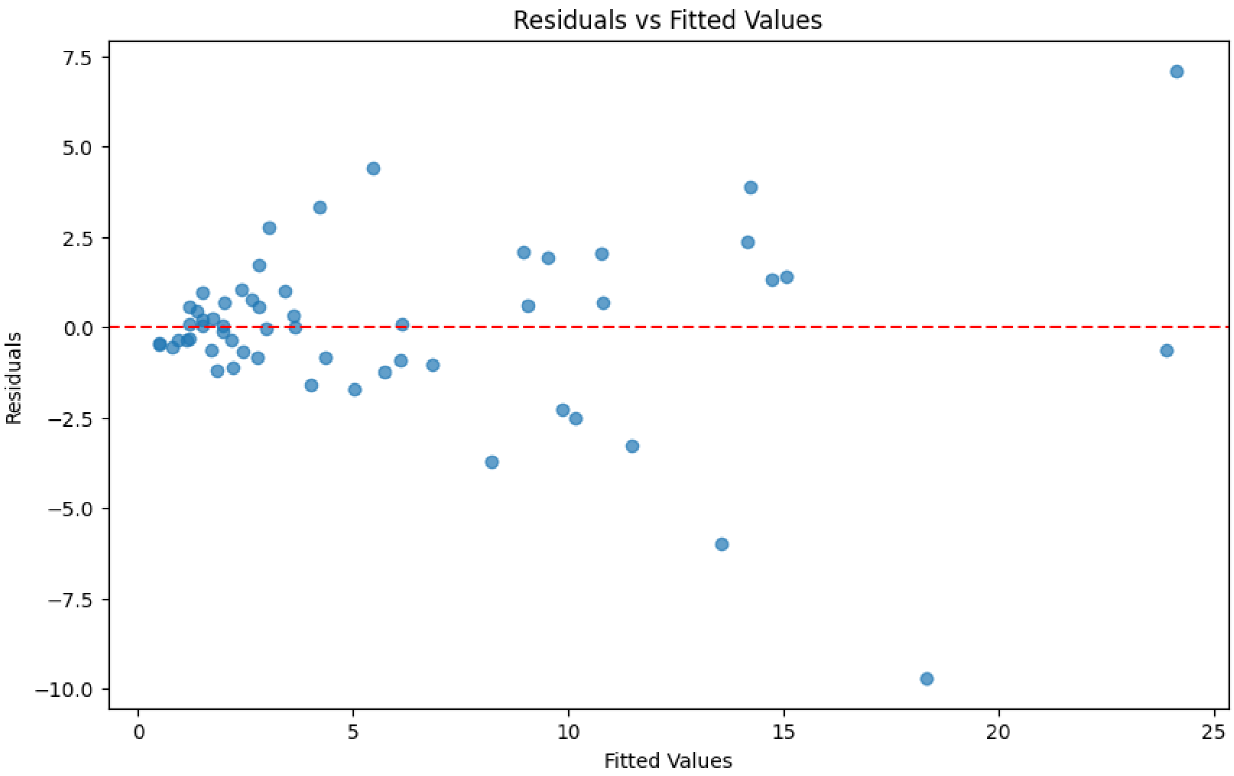
Here we find that

the partial regressor plots are approximately linear with the proximity being closer to Employment in agriculture , which suggests strong positive relation between Employment in agriculture with the dependent variable.

Analysis of Heteroskedasticity Tests

Heteroskedasticity refers to non-constant variance of residuals, which can affect the reliability of the regression model's coefficient estimates and standard errors.

We first proceed with visual inspection of the data to determine the presence of heteroscedasticity by plotting the residuals against the fitted values. We see there may be some presence of heteroscedasticity. We try confirm our suspicions by running the The **Goldfield-Quandt Test** for residuals which returns a p-value higher than 0.05. Thus, we may conclude our data does not have presence of heteroscedasticity.

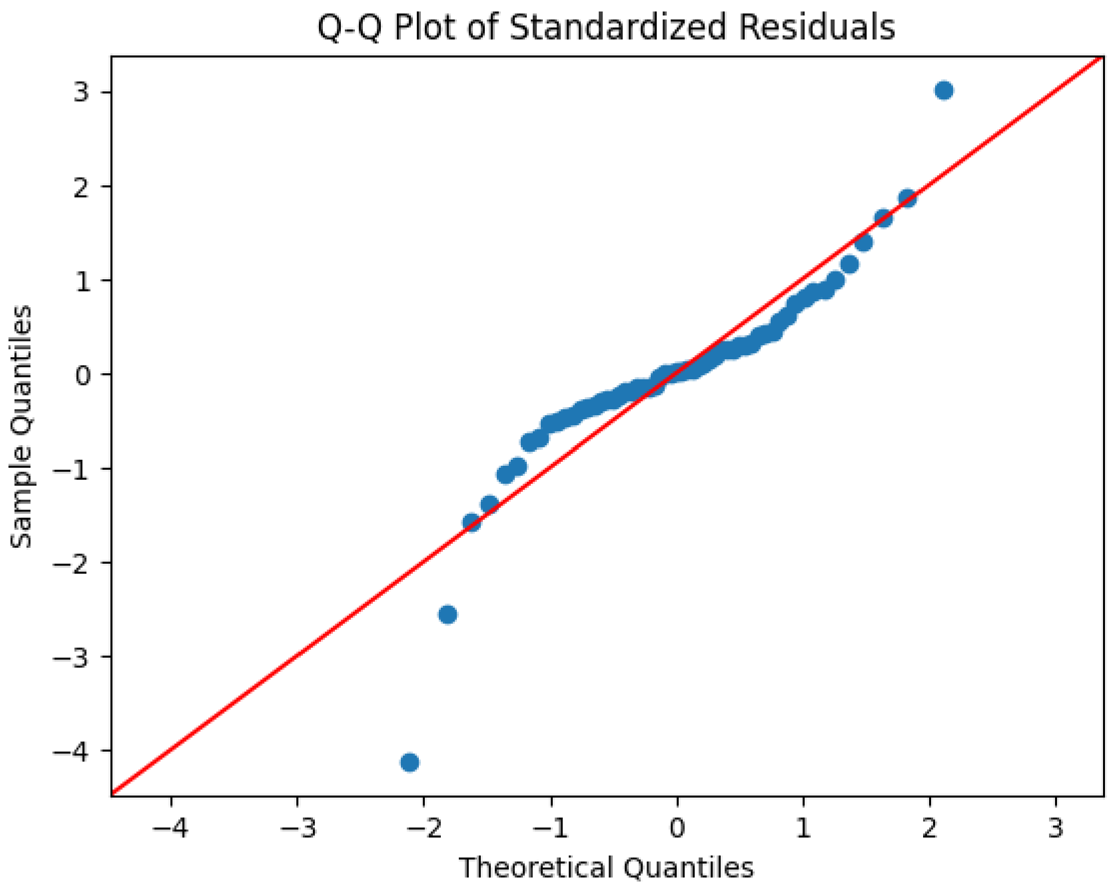



```
Goldfeld-Quandt Test:
F-statistic: 0.362769450955493, p-value: 0.9933115429371941
Observation: Reject H0 (homoskedasticity) if p-value < 0.05.
```

Analysis of Normality

From visual inspection of the quantile-quantile plot of the fitted residuals against the normal distribution we find there may be a possibility that the normality assumption is violated. The fitted residuals do not perfectly coincide with the normality line.

However, on conducting the Shapiro-Wilks test the p-value is quite low and is below our 5% probability of rejection. So, we conclude that the **distributional assumption of normality still holds.**




```
Shapiro-Wilk Test:
Statistic: 0.8872657558202263, p-value: 7.008803438607143e-05
Observation: Reject H0 (normality) if p-value < 0.05.
```

Influence Statistics and their interpretation

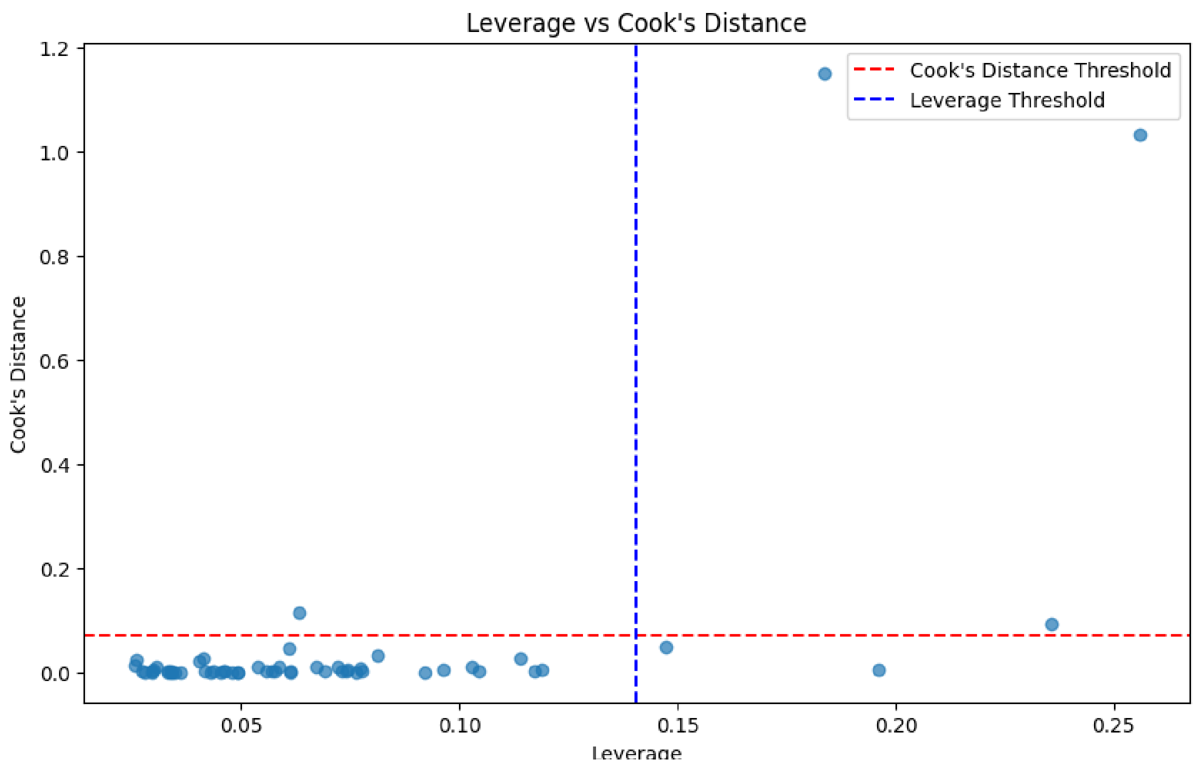
A scatter diagram of leverage, residuals, and Cook's distance can help identify influential data points in a regression model:

- Cook's distance

A measure of how much the model's estimates change when a specific observation is removed. A Cook's distance value above 1.0 indicates a highly influential point. Values above 0.5 indicate some influence, and values below 0.5 indicate no influence.

- Leverage

A measure of how far a predictor value is from its average. High leverage observations have predictor values that are far from their averages which greatly influence the fitted model.



Tests of Hypotheses: Theoretical Hypotheses and Testing for Relation among parameters. Model Comparison under Null with Unrestricted Model using ANOVA

The ANOVA test suggests that in the model the additional predictors do not improve model performance.

```
ANOVA Results (Model Comparison):
   df_resid         ssr  df_diff    ss_diff        F    Pr(>F)
0      55.0  317.909759      0.0        NaN      NaN       NaN
1      53.0  315.271709      2.0   2.638051  0.22174  0.801863

Theoretical Hypotheses Testing (Full Model):
                                                 OLS Regression Results
========================================================================================================================
Dep. Variable:     Agriculture_forestry_and_fishing_value_added_percent_of_GDP   R-squared:                       0.853
Model:                                                                     OLS   Adj. R-squared:                  0.845
Method:                                                          Least Squares   F-statistic:                     102.5
Date:                                                         Sat, 28 Dec 2024   Prob (F-statistic):           4.73e-22
Time:                                                                 07:13:12   Log-Likelihood:                -129.63
No. Observations:                                                           57   AIC:                             267.3
Df Residuals:                                                               53   BIC:                             275.4
Df Model:                                                                    3
Covariance Type:                                                     nonrobust
=======================================================================================================================================================
                                                                         coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------------------------------------------------
Intercept                                                              0.0518      1.599      0.032      0.974      -3.156       3.260
Employment_in_agriculture_percent_of_total_employment_modeled_ILO_estimate     0.3216      0.028     11.429      0.000       0.265       0.378
Gross_fixed_capital_formation_percent_of_GDP                           0.0171      0.077      0.223      0.824      -0.137       0.171
Rural_population_percent_of_total_population                           0.0134      0.028      0.483      0.631      -0.042       0.069
==============================================================================
Omnibus:                       21.123   Durbin-Watson:                   1.857
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               66.463
Skew:                          -0.869   Prob(JB):                     3.70e-15
Kurtosis:                       7.996   Cond. No.                         225.
------------------------------------------------------------------------------
```

we find AIC and BIC have quite high values, indicating our models are good fits and can explain the variations in the data.

ANOVA Results (Model Comparison)

Metric	Value	Interpretation
Residual Degrees of Freedom (Restricted)	55.0	Residual degrees of freedom in the null model (without additional predictors).
Residual Sum of Squares (SSR, Restricted)	317.91	Total unexplained variance in the restricted model.
Residual Degrees of Freedom (Full)	53.0	Residual degrees of freedom in the full model.

Metric	Value	Interpretation
Residual Sum of Squares (SSR, Full)	315.27	Total unexplained variance in the full model.
Difference in Residuals (SSR Difference)	2.64	Reduction in unexplained variance by adding predictors in the full model.
F-statistic	0.22	Test statistic comparing restricted and full models.
p-value	0.80	Indicates whether the additional predictors improve the model significantly.

Conclusion:

- The **p-value (0.80)** is much greater than 0.05, meaning we **fail to reject the null hypothesis** (H0H_0H0).
- The additional predictors (**Gross fixed capital formation** and **Rural population**) **do not significantly improve** the model compared to the restricted model with only **Employment in agriculture**.

Checking for Multicollinearity: Condition Indices, Condition Number, VIFs and IVIFs for full model

```
Condition Indices:
[1.         3.49017329 5.93550365]

Condition Number: 5.935503649731208

Variance Inflation Factors (VIFs) and IVIFs:
                                         Variable       VIF      IVIF
0  Employment in agriculture (% of total employme...  3.991867  0.250509
1          Gross fixed capital formation (% of GDP)  4.507501  0.221852
2          Rural population (% of total population)  9.016731  0.110905
```

Condition Indices and Condition Number

Metric	Value	Interpretation
Condition Indices	[1.0, 3.49, 5.94]	All indices are below 30, indicating no severe multicollinearity.
Condition Number	5.94	The condition number is far below the critical threshold of 30, confirming no severe multicollinearity.

Observation: The predictors do not exhibit problematic multicollinearity based on condition indices or the condition number.

Variance Inflation Factors (VIFs)

Variable	VIF	IVIF	Interpretation
Employment in agriculture (% of total employment)	3.99	0.25	Moderate multicollinearity; acceptable as VIF is below 10.
Gross fixed capital formation (% of GDP)	4.51	0.22	Moderate multicollinearity; acceptable as VIF is below 10.
Rural population (% of total population)	9.02	0.11	High multicollinearity; approaching the threshold of concern (VIF = 10).

CONCLUSION:-

The model we have chosen is quite adequate in explaining the variations in Gross National Income seen across countries. We can conclude that the variables Employment in agriculture , Gross fixed capital formation and Rural population (as a percentage of total population), considering the year 2018 , as adequate estimators of the share of agriculture , forestry and fishing in the total GDP of different countries.

Project: https://colab.research.google.com/drive/1qv2OVhN-RQ1N1sqPhZRiEmF3Egkpkt1D?usp=sharing

THANK YOU