Grace Saunders, Ravza Aykan, Mariam Seshan

Research Question

This study seeks to explore which measures of economic health are most useful in predicting the receipt of foreign aid. Specifically, we examine the relationship between various economic indicators, such as GDP growth, GNI per capita growth, and poverty headcount ratios, and the amount of Net Official Development Assistance (ODA) received by countries.

Data Collection and Observations

The dataset consists of observations from multiple countries over several years. Each observation represents a specific country-year combination, where each country's economic indicators (e.g., GDP growth, GNI per capita) and corresponding foreign aid receipt (Net ODA as a percentage of GNI) are recorded. The key variables in this study include Net Official Development Assistance (ODA) received as a percentage of Gross National Income (GNI), GDP growth (annual %), GNI per capita growth (annual %), Poverty headcount ratios at different poverty thresholds, and Control variables such as corruption, political stability, population size, and various other regional factors. Each row in the dataset represents one observation, capturing the economic status of a country at a specific point in time and its associated foreign aid receipt.

Approach: Supervised Learning & Regression

We are conducting supervised learning because our dataset consists of labeled input and output variables. This setup enables us to train a model using historical data to predict future outcomes. In particular, we aim to predict foreign aid receipt (both the likelihood and amount) based on several macroeconomic indicators such as GDP growth, GNI per capita growth, and poverty headcount ratios. Because our dependent variables—the likelihood of receiving foreign aid and the amount of aid received—are continuous rather than categorical, we will use regression models, which will allow us to quantify the relationship between macroeconomic health indicators and foreign aid, providing insights into which factors most significantly affect aid distribution.

We will primarily use Multiple Linear Regression (MLR) to examine the relationship between macroeconomic health indicators and the receipt of foreign aid. This model is suitable because we have multiple predictors (economic indicators, control variables), and MLR allows us to estimate the effect of each predictor while controlling for the effects of confounding variables (e.g., corruption, political stability).

The model will be in the following form:

$ODA = \beta_0 + \beta_1*GDP + \beta_2*GNI + \beta_3*poverty + \ldots + \beta_n*controls + \varepsilon$

Where:

- ODA (the amount of foreign aid received) is the dependent variable
- GDP (GDP growth), GNI (GNI per capita), poverty (poverty headcount), and controls (other control variables) are independent variables
- ε represents the error term

Additionally, we may use Time Series Regression Models to account for temporal dependencies (e.g., trends, seasonality) in the data. This will help account for delays in the impact of foreign aid on economic growth.

Data preprocessing will follow a structured process. First, we will inspect the dataset for missing values and use mean or median imputation when appropriate to fill in missing values. Rows with missing data may be excluded if imputation is not feasible. We will also scale and transform variables where necessary, as some of the economic variables (e.g., GDP growth, GNI per capita) may be on different scales, and some categorical variables may need to be converted into binary variables for use in the regression model. We will also check for multicollinearity between variables using Variance Inflation Factor (VIF). If any pair of predictors shows high correlation (VIF > 10), we will consider removing the redundant variables or applying Principal Component Analysis (PCA) to reduce dimensionality.

Once the data is preprocessed, we will fit the regression model using the aforementioned model to estimate the relationship between the independent variables and foreign aid receipt.

To evaluate the performance of the regression model, we will perform cross-validation. The data will be split into training and test sets. We will also evaluate the model by its R-squared value (which measures how well the independent variables explain the variation in the dependent variable), and the mean absolute and root mean squared error (which are used to assess the prediction accuracy of the model).

Interpretations and Results

In examining the relationships between economic health indicators and foreign aid receipt, we will focus on the magnitude and significance of the coefficients in our model, which will indicate the strength of the relationship between the economic indicator and the amount of foreign aid received, as well as whether that relationship is statistically significant. To present the results clearly, we will include a table of regression coefficients displaying the estimated coefficients for each independent variable, along with their standard errors and p-values. We will also compare R^2, mean absolute error, and root mean squared error values for the model, which will allow us to assess its performance. Scatter plots of key variables may also help visualize relationships and model fit.

Anticipated Challenges and Limitations

It is possible that foreign aid is allocated in response to economic conditions rather than causing economic changes. We plan to address this by including lagged variables to capture the delayed effects of economic growth on aid. Also, if foreign aid influences economic outcomes, it could be difficult to correctly determine the direction of causality, and so we may apply causal inference techniques or explore the issue using different model specifications. Finally, while we include key controls, there may be other unobserved variables influencing foreign aid, such as geopolitical events or international relations. We will attempt to minimize these influences by incorporating relevant control variables and time-fixed effects.