Introduction

This project aims to predict the economic performance of countries using key financial and macroeconomic indicators. The original goal was to classify economic collapse as a binary outcome, but we shifted to a more nuanced approach using a continuous response: the **economic_index**, a 0–100 score that reflects a country's economic strength.

We explore the structure of the data—focusing on distribution, outliers, and relationships between variables—before building and evaluating linear regression models.

Data Analysis

The dataset includes 5,950 country-year observations from 1990 to 2023. The response variable, economic_index, reflects overall economic health. The six predictor variables are:

- **gdp_growth**: annual percentage change in GDP

- **gov_debt**: government debt as a percentage of GDP

- **inflation_rate**: annual rate of inflation

- **interest_rates**: average interest rate for the country

- **black_market**: an index of black-market activity

- **cpi**: consumer price index

There are no missing values, and the data is ready for modeling.

The economic_index ranges from 14.26 to 95.29, with a mean of 54.88 and a standard deviation of 16.05, indicating a moderately wide spread in global economic performance. gdp_growth ranges from –64% to 150%, reflecting both severe contractions and rapid expansions. gov_debt shows extreme variation, from under 1% to 677% of GDP. inflation_rate and cpi include very large outliers, which likely reflect hyperinflation events. interest_rates span from –97.69% to 139.96%, possibly including data errors or highly irregular monetary policies. black_market is more stable, with a typical range between 7 and 233.

Histograms show that economic_index is approximately bell-shaped, but variables like gov_debt, inflation_rate, and cpi are heavily skewed. Boxplots confirm the presence of outliers, especially in inflation-related variables. We chose not to remove these extreme values because they reflect valid economic realities in some countries.
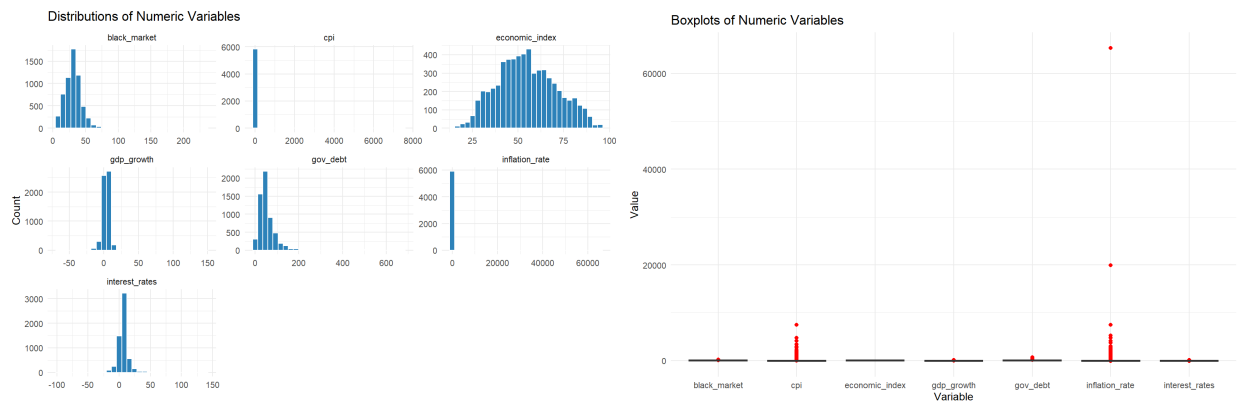
The correlation matrix shows a moderate negative relationship between economic_index and black_market (−0.57), while most other predictors are weakly correlated. No strong multicollinearity is evident, making this a suitable dataset for multiple linear regression.
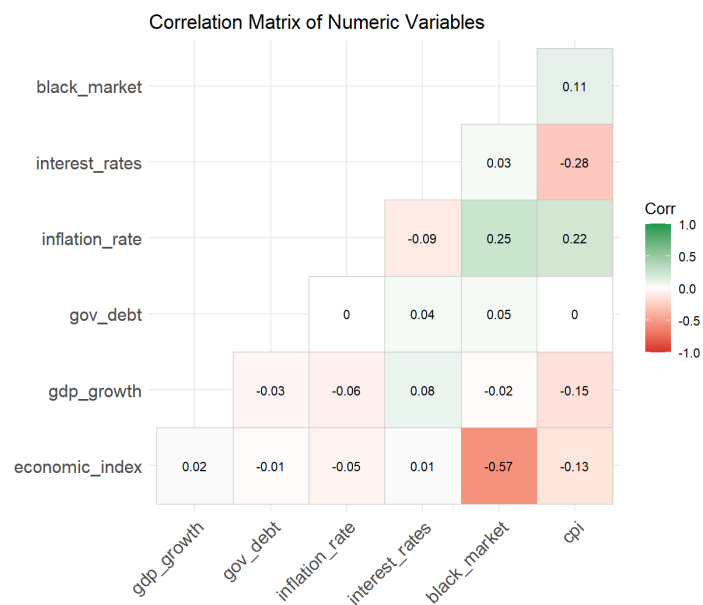
Full Model

To quantify how the predictors affect a country's economic performance, we fit a multiple linear regression model with economic_index as the response and all six predictors included. The fitted model is:

$$\textbf{economic\_index} = 79.28 + 0.0097 \cdot \text{gdp\_growth} + 0.0057 \cdot \text{gov\_debt} + 0.0020 \cdot \text{inflation\_rate} + 0.0155 \cdot \text{interest\_rates} - 0.7993 \cdot \text{black\_market} - 0.0061 \cdot \text{cpi}$$

The coefficient for black_market is strongly negative and statistically significant, meaning higher black_market activity is associated with lower economic performance. On the other hand, gdp_growth has a very small coefficient and is not statistically significant. Overall, the model explains 34.5% of the variance in the economic index ($R^2 = 0.345$), and the F-test indicates the model is statistically significant ($p < 2.2e\text{-}16$).

Model Assumptions

The residual vs. fitted plot shows that residuals are centered around zero, but with slight curvature, suggesting mild non-linearity or heteroscedasticity. The Q-Q plot shows that residuals are approximately normally distributed, with some deviation at the tails. Overall, the assumptions of linear regression appear to be reasonably satisfied.
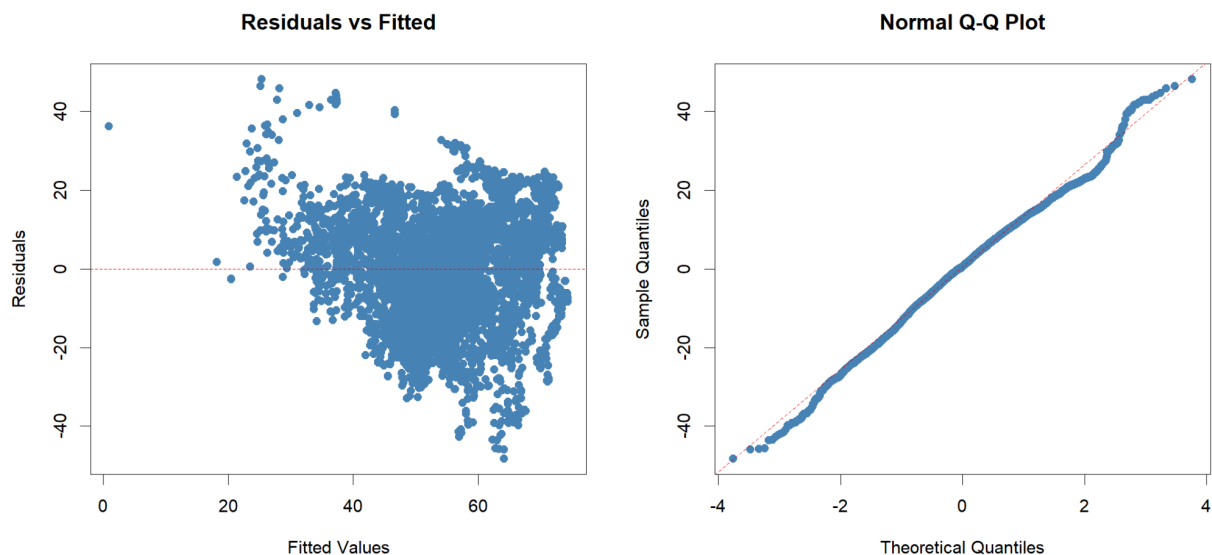
Reduced Model

economic_index $= 79.37 + 0.0058 \cdot \text{gov\_debt} + 0.002031 \cdot \text{inflation\_rate} - 0.7981 \cdot \text{black\_market}$
$- 0.00634 \cdot \text{cpi}$

F-Test: Full vs Reduced Model

To evaluate whether the non-significant variables gdp_growth and interest_rates add value to the model, we compared the full model to a reduced model without them.

Hypotheses
H_0: gdp_growth and interest_rates do not significantly improve the model.
H_1: At least one of them does.

Results
F-statistic $= 0.5891$
df $= (2, 5943)$
p-value $= 0.5549$

Conclusion
Because the p-value (0.5549) is greater than 0.05, we fail to reject the null hypothesis. This means that removing gdp_growth and interest_rates does not significantly reduce the model's explanatory power. Therefore, we prefer the reduced model for its simplicity.

Confidence Intervals for a Typical Observation

Using the average values of each predictor, we predicted the economic_index for a typical country-year:

- Full Model: 54.88 (95% CI: 54.55 to 55.21)

- Reduced Model: 54.88 (95% CI: 54.55 to 55.21)

Both models yield virtually identical predictions, confirming that the dropped predictors are not meaningful contributors to the outcome. The reduced model is simpler and just as effective.

Appendix

Authors

This report was completed by:

- **Viswa Sushanth Karuturi**

- **Naeem Almohtaseb**

GitHub Repository

All data and analysis scripts for this project are publicly available in the public GitHub repository:
https://github.com/sushanthvk02/financial-collapse-indicators

The repository contains:

- **/data/** — All cleaned and final datasets, including the merged predictor dataset used in modeling

- **/scripts/** — R scripts for data cleaning, exploratory data analysis, and linear model fitting