

How the U.S. Economy Has Changed Over Time

1. Introduction

1.1 Motivation for Analyzing U.S. Economic Growth and Government Borrowing

The U.S. economy is one of the biggest in the world, and it has changed a lot over time. Sometimes it grows quickly, and other times it faces big problems like the 2008 financial crisis or the COVID-19 pandemic. People often wonder what causes these changes and whether they're connected to how much money the government borrows. When the government borrows money, it can use that money to invest in things like building roads, schools, or healthcare. However, as the government borrows more, there are concerns about whether this borrowing could eventually cause problems, like higher interest payments that may slow down the economy. In this report, we're going to look at how the U.S. economy has changed over the years, and try to figure out if there's a link between the country's economic growth and the amount of money the government borrows.

1.2 Main Research Question

The central question this project seeks to answer is:

"Is the growth of the U.S. economy related to the increase in government borrowing over time?"

2. Used Data

For this project, we have identified two primary datasets: U.S. Economic Indicators and U.S. GDP vs. Debt. These datasets will help us analyze the relationship between the growth of the U.S. economy and the increase in government borrowing over time.

2.1 U.S. Economic Indicators (1974–2024)

The U.S. Economic Indicators Dataset gives detailed information about the U.S. economy from 1974 to 2024. It includes important numbers like the total value of goods and services produced in the country (GDP), the cost of living over time (CPI), and the unemployment rate (how many people don't have jobs). In this project, we use this data to calculate and look at these numbers for each year and then group them into decades to see long-term trends. After cleaning and organizing the data, we focus on the total value of goods and services, the unemployment rate, and the cost of living over time. This helps us understand how the economy has changed and how it can be linked to other factors, like government borrowing.

2.2 U.S. Public Debt vs. GDP (1947–2020)

The U.S. Public Debt vs GDP Dataset provides detailed information on the U.S. government's debt and its economic output (GDP) from 1947 to 2020. It includes data on the total public debt (the amount the government owes) and the GDP (the total value of all goods and services produced in the country) for each year. In this project, we use this dataset to understand how much the U.S. government has borrowed over time compared to the size of the economy. After organizing and cleaning the data, we focus on the relationship between the total public debt and GDP each year, which helps us see if the government's borrowing has been growing faster than the economy itself. This analysis helps us understand the impact of increasing government borrowing on the U.S. economy over time.

3. Analyses

3.1 Data Collection and Preparation

The data collection and preparation process begins by downloading two key datasets from Kaggle using the Kaggle API: one containing U.S. economic indicators and the other containing U.S. public debt vs. GDP data. These datasets are extracted and preprocessed to ensure they are in the correct format for analysis. Preprocessing involves cleaning the data, handling missing values, and ensuring consistent column names across all datasets.

To prepare the data for merging, the "DATE" column in each dataset is converted to a proper datetime format. For the economic indicators dataset, which includes monthly data on CPI, GDP, and unemployment, the data is aggregated into quarterly values by calculating the average for each quarter. Similarly, the public debt and GDP dataset, originally recorded at quarterly intervals, is processed to ensure consistency with the other datasets.

The next step is to merge the datasets. This is done by joining them on the "Quarter" column, which creates a unified dataset containing the key economic indicators: GDP, public debt, CPI, and unemployment rate, all at quarterly intervals.

Following the merging process, normalization is applied to standardize the data. Normalization ensures that all economic indicators—such as CPI, GDP, debt, and unemployment—are scaled to a range between 0 and 1. This makes it easier to compare them directly, even though they represent different units of measurement (e.g., debt in millions vs. unemployment rate as a percentage). The normalization is done by subtracting the minimum value from each data point and then dividing by the range (maximum value mi...

Finally, the processed and normalized data is saved as a SQLite file, which is then ready for further analysis. The entire pipeline—from downloading to merging and normalizing the data—ensures that the information is clean, consistent, and ready for meaningful analysis of trends and relationships within the U.S. economy over time.

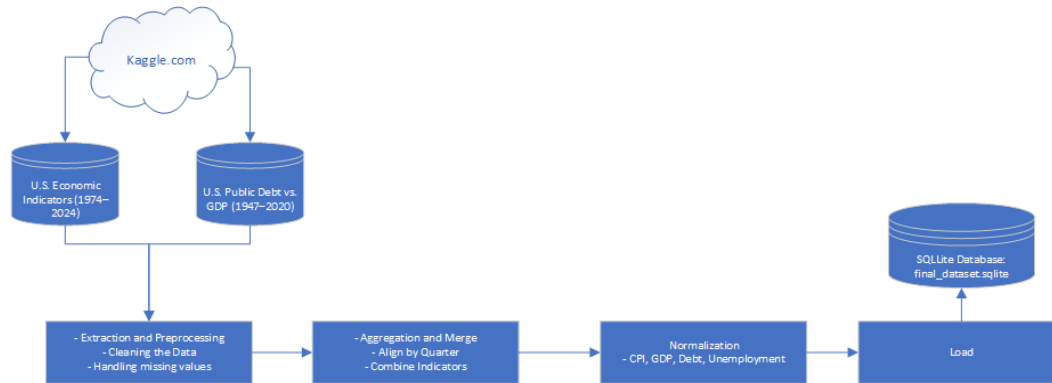
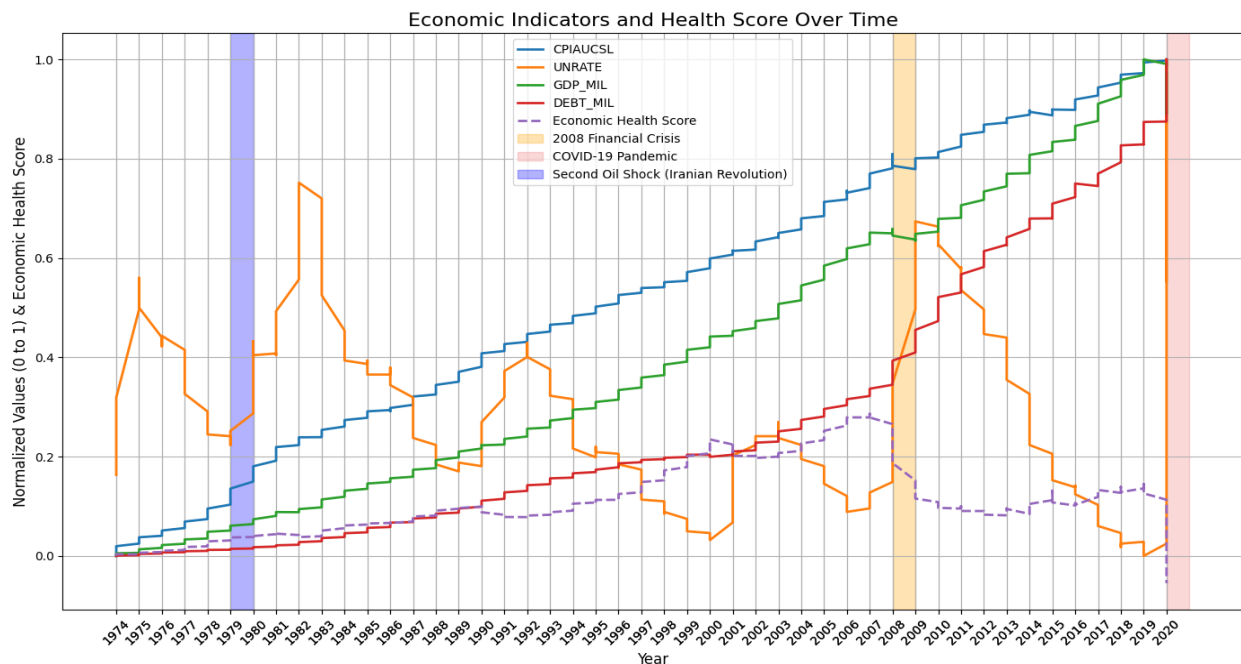


Figure 1: Improved Data Pipeline Structure

3.2 Analysis

Based on the final dataset containing quarterly GDP, government debt, unemployment rate, and CPI, we can create visualizations to analyze the U.S. economy over time. Figure 2 displays the normalized values of these economic indicators as line plots, allowing for easy comparison of their trends. Additionally, the Economic Health Score, which reflects the relationship between GDP, debt, and unemployment, is plotted as a dashed line. Shaded regions in the plot highlight significant economic events such as the Iranian Revolution in 1979 and the 2008 financial crisis and the COVID-19 pandemic, providing context to the observed trends. These visualizations offer insights into how the U.S. economy has evolved and the potential impact of key events.



In Figure 2, it is evident that the U.S. GDP has experienced a significant increase over the years, reflecting consistent economic growth. Similarly, government debt has also shown a sharp upward trend, especially during major economic events such as the Iranian Revolution in 1979, the 2008 financial crisis, and the COVID-19 pandemic.

Unemployment rates, on the other hand, display fluctuations with spikes during periods of economic downturns. CPI (inflation) demonstrates a steady upward trend over time, indicating a gradual increase in the cost of living.

The Economic Health Score, calculated as the difference between GDP and public debt divided by the unemployment rate plus one, provides a comprehensive measure of the economy's overall performance, capturing its response to major events and underlying trends. Additionally, the Economic Health Score shows variability, with notable declines during crises, highlighting the adverse impacts of these events on overall economic health.

3.3 Problems

The "U.S. Public Debt vs GDP" dataset only contains data up to 2020, while the "U.S. Economic Indicators" dataset extends to 2024. To address this discrepancy, the analysis focuses on the overlapping period of 1974 to 2020 to ensure consistency across datasets. Additionally, since some of the datasets were recorded at monthly intervals, the data was aggregated to quarterly values to align with other datasets and facilitate meaningful comparisons.

4. Conclusions

The growth of the U.S. economy appears to be closely related to the increase in government borrowing over time, particularly during economic crises such as the Iranian Revolution, the 2008 financial crisis, and the COVID-19 pandemic. Government borrowing has been a critical mechanism for responding to economic shocks, supporting recovery, and fostering growth. However, this dependency on borrowing may pose risks to economic stability in the long run, necessitating careful management of public debt during periods of economic stability.