Predict the likely hood of a recession in the US

Aishwarya Hastak[1], Manali Shelar[1]

Abstract

Recession can occur due to a variety of reasons, but at the base, it's a simple supply-demand imbalance. When the demand for goods peaks and then starts to fall, we are left with an extra supply of goods and services for which there is not enough demand or consumption. This leads to less production and downsizing of the workforce in companies. Due to the rise in unemployment, individuals can't purchase as much as before.

Building a system to predict the possibility of a recession is intended to help detect and foresee economic dangers so that preventive steps can be taken to lessen their effects. It can offer early warning signs and support investors in making de- feasible choices regarding how to react to changing economic conditions.

Keywords

Recession — Yield Curve — Probit Model

[1] Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA

Contents

1. Problem and Data Description

When a recession occurs, it negatively affects businesses, country economies, and individuals. It would be helpful if we could predict how likely it is for a recession to occur in the current or upcoming years.

Recession can have a lot of effects on all of us including, but not limited to, a drop in a country's economy, a drop in company shares and stock prices, a rise in unemployment, inflation, and difficulty in finding jobs, higher interest rates, housing issues. If we were able to predict an impending recession, both businesses and individuals could take steps to prepare for it; such as proper budgeting, clearing off high amounts of debts, focusing on savings, invest capital in the right sectors. These are just some of the many things that could be done to cushion the blow of a recession.

We have used 6 recession indicators: unemployment rate [1], consumer price index (inflation rate) [2], GDP growth/change rate [3], and stock index percent change [4], 10-year-2-month treasury bond rate, and federal funds rate for the USA. We have also collected data for quarters from the years 1872 to 2022 of past recession/expansion periods in the US. We will analyze how these indicators have affected the occurrence of recession in the past.

We have sourced all our data from Federal Reserve Economic Data, which is a public site for economic data. We have used quarterly data since it is more appropriate for analyzing the GDP rate.

## 2. Data Preprocessing & Exploratory Data Analysis

2.1 Handling Missing Values

We have some missing data values for the indicators of the Stock index percent change and Unemployment Percent Change. We will fill those rows with the mean of all values.

2.2 Exploratory Data Analysis
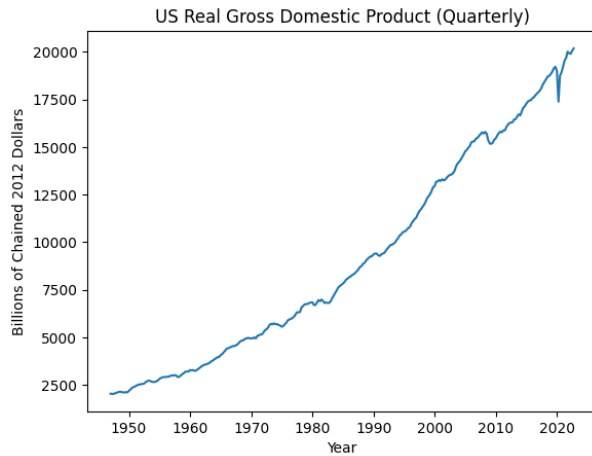
The plot of the US Quarterly Real GDP(in billions of dollars) shows an upward trend indicating a gradual increase in GDP from the year 1950 to 2023. We can also see a sharp decline in the year 2020 which was also a year of recession. Similarly, there are smaller declines during several other years, such as 2008, 1990, and 1981, that were also hit by recession.
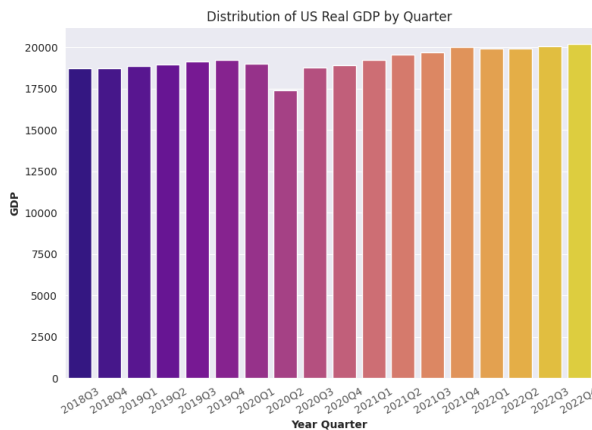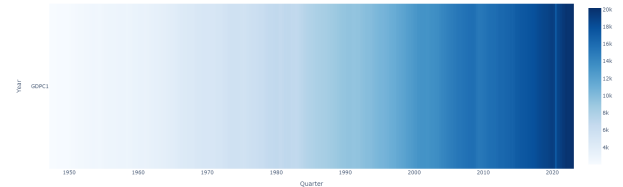

The bar graph of the US Quarterly Real GDP(in billions of dollars) for only the last 5 years shows a clear decline during the second quarter of the year 2020, when a recession occurred.
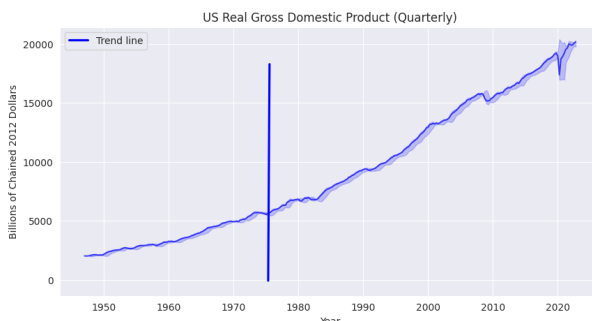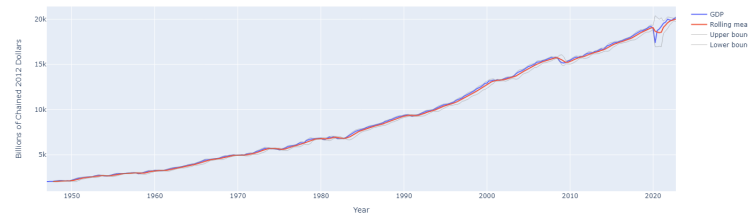
The heatmap of the quarterly GDP data by year again shows a gradual increase in the GDP over time. The line plot shows that the rolling mean of the GDP is within the upper and lower bounds of the confidence interval.
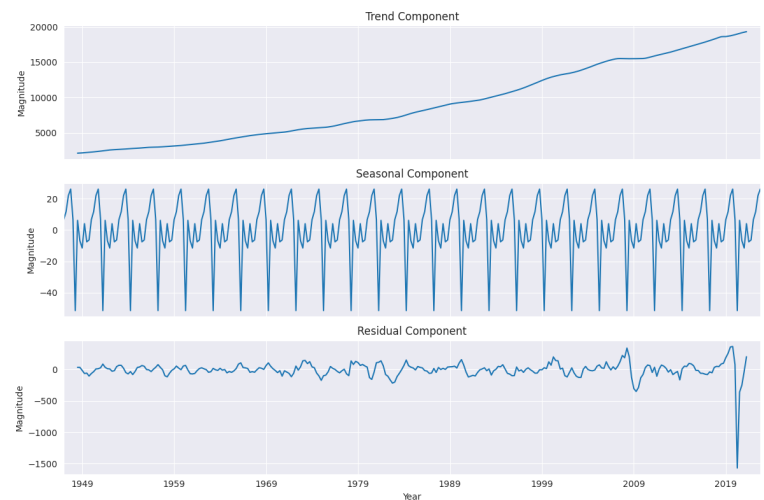
The seasonal decomposition plot ran on an additive model shows an upward trend in the growth of GDP. There is also a linear seasonality in the data. The residual components lie around zero which is a good sign that shows that the data is centered around the mean.
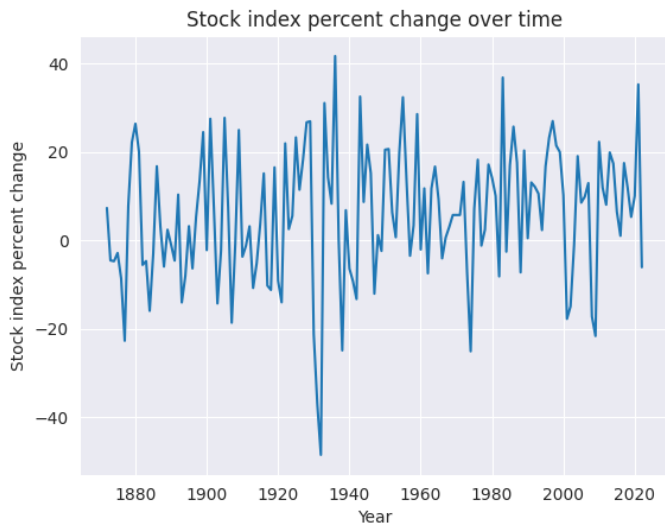
The line chart of the stock index percent change over time shows a sharp upward and downward edge with no apparent pattern. However, the most significant sharp declines align with the years of recession. The top 5 lowest percent changes in stock prices are in the years 1929, 1937, 1887, 1973, and 2008 which are also years of the recession period. Hence, stock prices and the occurrence of a recession show a clear correlation.
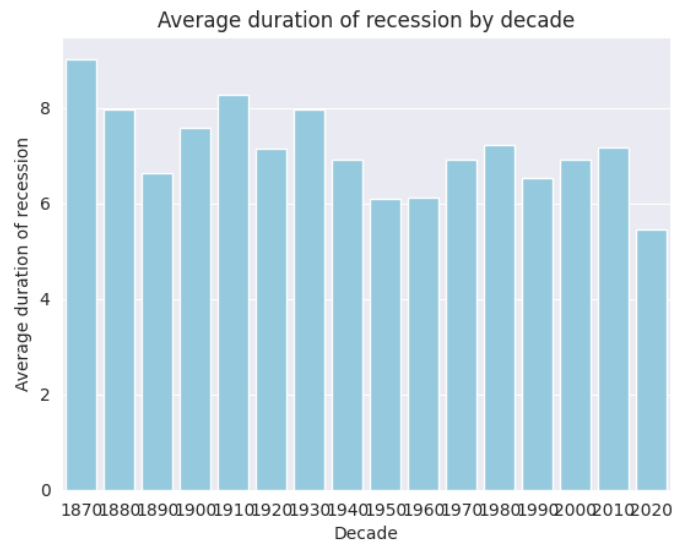
The bar chart of the average duration of recession by decade shows a period of the recession lasting for at least 5 to 6 months for all decades, which means each decade sees at least half a year of the recession period. The highest duration of a recession was during the decade 1870 and the lowest was in the last decade of 2020.
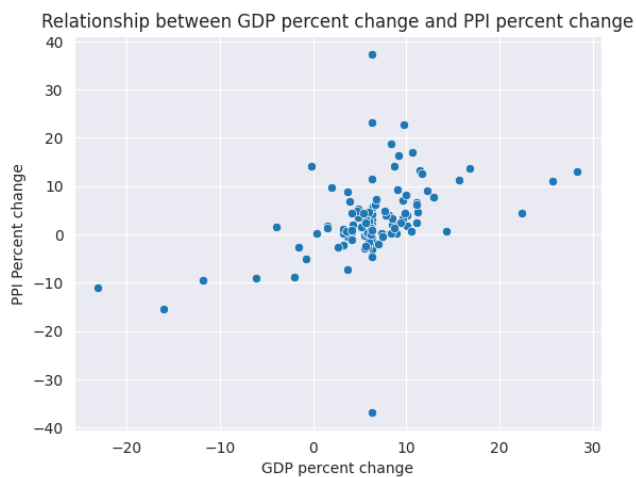

The scatter plot of GDP percent change vs. PPI percent change shows a proportional relationship, where if GDP percent change increases then so does the PPi percent change. This makes sense because the if the production prices are lower then it means that the consumers will pay lesser at the retail level. This could indicate an economic slowdown which could lower the GDP. If the PPI(Producer Price Index) is high then it indicates a possibility of inflation.
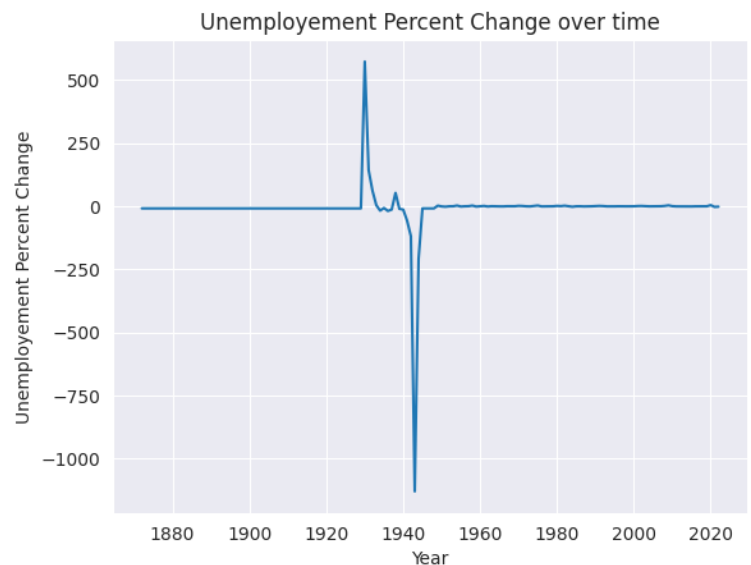
The line chart of the unemployment percent change over time does not show to have any effect on the duration of a recession. Hence, we can assume that unemployment is not a strong indicator for predicting the likelihood of a recession.

Correlation matrix
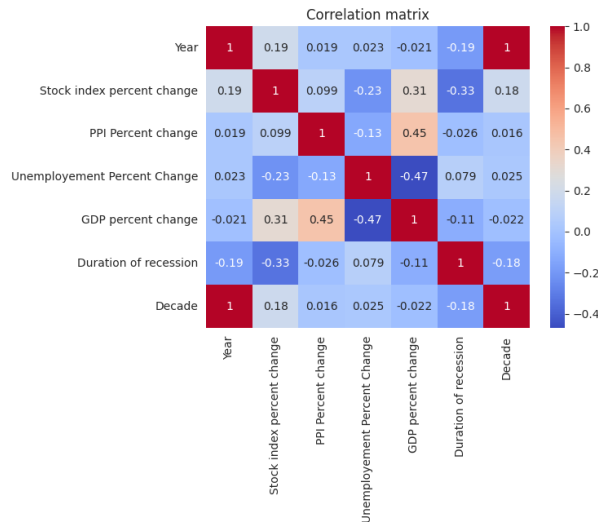
The heat map of the correlation matrix for all the indicators shows a strong correlation between GDP percent change and PPI percent change, which we also concluded from the scatter plot above. It also shows a correlation between Stock index percent change and GDP percent change.

Plot of 10Y3M Treasury Rates over time with occurences of Recession
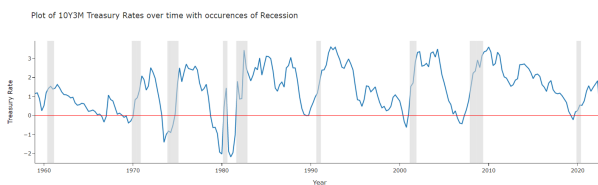
The plot of the 10Y3M treasury bond rate with respect to the recession period is known as Yield Curve. This indicator has been used in a lot of State of the Art models as the only predictor for recession, which is telling of how important this indicator is. This plot shows us that treasury bond is able to predict the occurrence of recession up to 3 quarters earlier.

Plot of GDP change over time with occurences of Recession

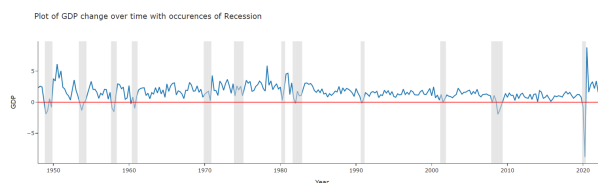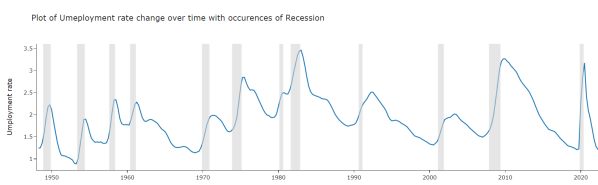The plot of the GDP rate with respect to the recession period shows us that the GDP rate is able to predict the occurrence of a recession up to 2 quarters earlier, where 2 consequent decrease in GDP usually indicates a likelihood of a recession.

Plot of Umeployment rate change over time with occurences of Recession

We took a three-month moving average of the unemployment rate and found that it is able to predict the occurrence of a recession up to 1 quarter earlier.

3. Algorithm and Methodology

3.1 Building a model for Recession Prediction

The world's economies have seen a roller-coaster ride in 2022. The conflict in Russia and Ukraine caused recessions in both those regions.

In India, we have already witnessed the value of currency fluctuating like never before, and the same is true for all the major currencies throughout the world. China is currently experiencing an economic crisis as a result of both covid and real estate concerns.

Additionally, the possibility that the United States is entering a recession is currently causing shockwaves throughout the entire world.

Because if the biggest economy in the world witnesses a recession its going to have a ripple effect on economies all across the world.

Building a system to predict the possibility of a recession is intended to help detect and foresee economic dangers so that preventive steps can be taken to lessen their effects. It can offer early warning signs and support investors in making defensible choices regarding how to react to changing economic conditions.

For several reasons, including policy-making, corporate planning, economic planning, and financial stability, predicting the risk of a recession is essential.

There are several intriguing attributes of predicting the recession probability using different economic indicators. First of all, since we are working with economic data(usually having the monthly/quarterly update frequency), we only have a few hundred data points to work with.

Secondly, it's time series data which usually is the case for projects involving economic data. Also, since the underlying model is the time series data, methods like K-fold cross-validation cannot be used while dividing the data into training and testing data sets because we have to make predictions based on the data available at the current time.

Lastly, it's a rare-event prediction since recessions are rare. The two main assumptions that we made while working on this project are that US recessions give early warning signs and that future recessions will be similar to past recessions.

Feature Selection: In these types of projects, while selecting the features, domain knowledge is important. Automated feature selection might result in over-fitting of the models.

So, after careful consideration, we decided to use 6 economic parameters viz. Unemployment Rate, Quarterly GDP rates,

Consumer Price Index(Inflation rates), Stocks index, 10Y-3M Treasury bond rates, and Federal Funds.

3.2 Indicators Used for Recession Prediction
Currently used indicators for recession prediction:

1. Sahm Rule Recession Indicator: The Sahm Rule identifies a recession based on the three-month average of the unemployment rate. A recession is judged to have begun when the average unemployment rate climbs by 0.5 percentage points or more above its previous year's low.

2. New York Fed Recession Indicator: Its drawbacks include only providing a 12-month projection and relying on a single variable (the spread between 10-year and 3-month Treasury rates).

3. Guggenheim model: This model uses a combination of economic and market indicators. From this model, we got the idea of predicting the recession for different time frames.

4. Rabobank Model: It is also solely based on one variable (the difference between the 10-year and 1-year Treasury rates) and covers only one time period (17 months).

5. Wells Fargo Economics Probability Recession Model: This model makes use of both economic and market data. However, they only forecast for the next six months.

6. The Stockton-Malki Recession Indicator: It was created by two economists from the Federal Reserve Bank of San Francisco. It forecasts the likelihood of a recession in the following 12 months by combining the Treasury yield curve and the unemployment rate.

3.3 Models Used for Recession Prediction
1. Probit Model: The probit model is a statistical technique for analyzing binary dependent variables. Using the same fundamental framework as the generic probit model, the probit model estimates the chance of a recession occurring as a function of these explanatory variables. One advantage of the probit model for recession prediction is that it allows for the introduction of uncertainty and ambiguity into the study. Rather than predicting whether or not a recession will occur, the probit model provides a measure of the possibility of a recession occurring, which can be useful for decision-making and risk management.

2. Elastic Net Model: To increase the accuracy and interpretability of the model, the elastic net model combines the L1 and L2 regularization approaches. As input variables, the elastic net model would use a collection of economic data such as GDP growth rate, unemployment rate, consumer confidence index, and so on. The model would next examine the historical links between these characteristics and the occurrence of recessions to determine the most important predictors. The L1 regularization method assists in identifying a subset of input variables that are most important for anticipating recessions. By reducing the coefficients towards each other, the L2 regularization method helps to lessen the impacts of multicollinearity among the input variables.

3. Support Vector Machine(SVM): SVMs are a form of machine learning technique that is used for classification and regression analysis. They can anticipate future recessions by analyzing previous economic data and discovering patterns and linkages.

4. XGBoosts: It is a gradient-boosting approach that creates a powerful prediction model by combining decision trees. Based on these indicators, the algorithm would generate a series of decision trees that are tuned to anticipate the occurrence of recessions.

4. Experiments and Results

4.1 Likelohood Prediction
We tested our data on four models and found the following results for the likelihood of the occurrence of a recession in the Current quarter:

Model	Likelyhood
Probit	0.08126163
ElasticNetCV	0.13968719177426533
XGBoost	0
SVM	0

Next quarter:

Model	Likelyhood
Probit	0.04208493
ElasticNetCV	0.05264960351830204
XGBoost	0
SVM	0

Next 2 quarters:

Model	Likelyhood
Probit	0.14938127
ElasticNetCV	0.2100915155207727
XGBoost	0
SVM	0

Next 4 quarters:

Model	Likelyhood
Probit	0.48919047
ElasticNetCV	0.33625819764
XGBoost	0
SVM	0

The probit model seems to give the most accurate prediction with some likelihood of a recession in the next 2 or the next 4 quarters.

4.2 Model Performance
Because this is a classification problem, metrics like accuracy, prediction/recall, and F1 score are acceptable. However, for probabilistic predictions, Log Loss Errors are the optimal error metric.

In binary classification issues, it is a common error metric. In

the case of binary classification, log-loss indicates how close the prediction probability is to the matching actual/true value (0 or 1).

The values of log loss error for our models were as follows:

Model Name	Current Quarter	Next Quarter	Next 2 Quarters	Next 4 Quarters
Probit	0.150	0.634	0.281	0.260
Elastic Net	0.189	0.543	0.614	0.280
XGBoost	0.560	0.715	0.893	0.915
SVM	0.657	0.698	0.923	1.150

5. Deployment and Maintenance

We have deployed our project [5] using the Flask framework and using an online tool called PythonAnywhere. Currently, our website is linked to a static database. We plan to connect our database through APIs which will be a part of our future work. As part of the maintenance, we will continuously update our database with current and relevant data to ensure that our system stays robust and effective.

6. Summary and Conclusions

Our data analysis established that treasury bond rate, GDP rate, and unemployment were crucial in predicting recession earlier. We found that the Probit model gives the best results with an average log loss error of 0.3312, which was the best amongst all models.

Our future work includes connecting the database through APIs, refining our choice of indicators, and including more economic and financial indicators. We also plan to consider current events, market sentiments, and news as a factor to predict a recession.

Acknowledgments

References

[1] Unemployment in the United States. (2012, August 17). https://fred.stlouisfed.org/series/M08I4AUSM175NNBR0

[2] Producer Price Index by Commodity: All Commodities. (2023, March 15). https://fred.stlouisfed.org/series/PPIACO0

[3] Gross Domestic Product. (2023, March 30). https://fred.stlouisfed.org/series/GDPA0

[4] Index of All Common Stock Prices, Cowles Commission and Standard and Poor's Corporation for the United States. (2012, August 15). https://fred.stlouisfed.org/series/M1125AUSM343NNBR0

[5] http://recessionprediction.pythonanywhere.com/