FINAL CAPSTONE SUBMISSION. Please see readme called: CAPSTONE FINAL README. Relevant data files are: fred.csv, fred\_sns.csv, fred\_clsf.csv. The jupyter notebook has been pre-run and saved with the charts and data tables already outputted so that the reader need to recreate the process to rerun the code.

**PROBLEM STATEMENT**

Is there a consistent actionable relationship between macroeconomic data of US to predict real GDP, mortgage rates and other market indicators? If yes, how might macroeconomic data influence asset allocation decisions for portfolio management?

**RELEVANCE OF THE PROJECT**

Markets, market indicators and the economy are crucial to everyone’s well being. Understanding and predicting the path of economy and global market indicators can be extremely valuable in making better life choices as well as offer an opportunity to systematically profit (and avoid losses) by making accurate prediction of market indicators. This question bridges a gap in my understanding of what's happening in the economy and what's happening in the market. It connects the two in logical ways that can help me in making more educated and informed portfolio and trading decisions.

**DATA USED FOR ANALYSIS**

* St. Louis Fed (FRED) - [Federal Reserve Economic Data | FRED | St. Louis Fed (stlouisfed.org)Links to an external site.](https://fred.stlouisfed.org/)
* Yahoo Finance: [Yahoo Finance - Stock Market Live, Quotes, Business & Finance NewsLinks to an external site.](https://finance.yahoo.com/)
* Nasdaq Data Link: [Retail Trading Activity Tracker: Keep track of retail sentiment (nasdaq.com)Links to an external site.](https://data.nasdaq.com/institutional-investors)

**METHODS APPLIED IN ANALYSIS**

* Visualization of data using various types of plots to gain an intuitive understanding of the relationships
* ARIMA models to evaluate stationarity and cointegration among macroeconomic data vectors.
* Regression: various linear, non-linear regression methods to further assess macroeconomic data cross-relationships
* Classification methods: logistic regression, KNN, SVM, decision trees to classify data into relevant categories to forecast recessions.
* Advanced methods: Use of methods such as random forest to identify the relationships with high degree of confidence and predict recessions.

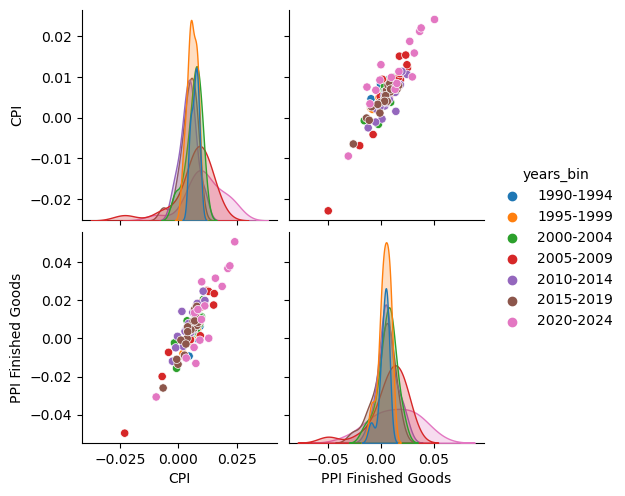
**EXPECTED RESULTS**

* Quantitative description of relationships amongst macroeconomic variables and between macroeconomic variables and key market indicators.
* Quantitative description of relationships between above mentioned macroeconomic variables and economic regimes (recessions and periods of growth).
* Regression and classification analyses to better understand leading indicator drivers of GDP and mortgage rates.
* Actionable recommendations for portfolio allocation and recommendation on trading decisions based on this analysis.

**RESULTS**

There is clearly a quantifiable relationship between various macroeconomic indicators:

* Difference between CPI and PPI % change quarter over quarter tend to show stationarity – I have witnessed it practically in markets where – if CPI data is high, eventually the PPI data catches up to it and vice-versa. As a result, when CPI data is released or PPI data is released, having an educated sense of which way the data may lean can guide positioning of a portfolio. Specifically, if PPI is strong, it would make sense to have low exposure to the market and it may be followed by a selloff in equity markets and a rise in interest rates.



Inflation has tended to go up when oil prices go up. History is the best evidence: during the Yom-Kippur war to Iranian revolution period, oil prices went up a few 100% and as a result, US economy faced one of the largest inflationary periods of all its history. This period was accompanied by initially failed attempts at containing inflation by raising rates and later, by high rates that led to deep recession. As a result, the level of US10Y govt. bond vs. Fed Funds effective rate (used by US Federal Reserve to implement its monetary policy) can be instructive in terms of effectiveness of a rate regime at containing a recession. Predicting recessions is part of the objective of this project.

A group of graphs showing different sizes of data

Description automatically generated with medium confidence

While personal consumption rises as inflation rises, a lot is attributable to nominal increase in prices but eventually might result in a recession.

A group of graphs showing different colored lines

Description automatically generated

Industrial Production and Capacity utilization, similarly tend to move in line and are indicative of a growing or slowing economy

A group of graphs showing different colored lines

Description automatically generated

Similarly LMV sales and retail sales have a positively correlated relationship

A graph of different colored dots

Description automatically generated with medium confidence

A comparable relationship may be observed between non-investment grade corporated debt yield (BBB yield) and mortgage borrowing rate.

A graph of different colored lines

Description automatically generated with medium confidence

This relationship is also corroborated by slicing the scatter plot by bins of yield curve steepness measured as US10Y yield – Fed Funds effective rate changes.

A group of graphs showing different colored lines

Description automatically generated with medium confidence

Difference between Personal income and savings change quarter over quarter shows stationarity. In other words, as persona income rises, so do savings. Therefore, measuring the difference between the two can be helpful in understanding if consumers are building wealth, therefore, contributing to the growth in GDP. A decline in savings may be indicative of a slowing economy (to be tested in following segment)

A graph showing a line of blue dots

Description automatically generated with medium confidence

A graph showing a line of savings

Description automatically generated with medium confidence

Difference between Retail Sales and LMV sales quarter over quarter shows stationarity. In other words, growth in retail sales tends to be corroborated by growth in LMV sales and vice-versa. These may be helpful in predicting a growing or slowing economy.

A graph showing a line of savings

Description automatically generated with medium confidence

A graph showing a line of sales

Description automatically generated

Difference between wages and CPI change quarter over quarter shows stationarity. This is perhaps the most intuitive and one of the more important observations. As wages grow, consumers have more disposable income which leads to higher expenditures, higher money supply in economy and eventually, above a certain threshold, inflation.

A graph showing a line of blue dots

Description automatically generated

A graph with blue dots and numbers

Description automatically generated

In 2022-23 some of the most volatile days of stock markets were during the release of CPI and PPI. The two, consumer price inflation and producer price inflation tend to be positively correlated and the difference between their % changes QoQ exhibits stationarity. What this means is that when CPI goes up, PPI may eventually catch up and vice versa. This is very relevant because once we know the CPI data, we can have an educated sense of where PPI data might print and as a result, how may the market react to that data release.

A graph showing a line of dots

Description automatically generated

A graph of a function

Description automatically generated

The next step involved analysis of correlations across all data features and cointegration analysis (not shown here but available in the jupyter notebook).

Based on the analyses above, I developed regression models to predict real GDP change QoQ% and mortgage rate change QoQ%. I also developed a framework for classification models using logistic regression, k nearest neighbors, support vector machines and decision tree classifiers to classify data features and predict recessions. Finally, I used random forest analysis to optimize the process to predict recessions.

Theclassification exercise was performed by classifying all economic features into 3 ordinal categories: Low, Medium, High where:

Low(Lo): shows negative values (so negative inflation, negative wage growth, negative industrial production, negative retail sales for example)

Medium (Med): shows generally average levels in the economy that are necessary to achieve` the break even real growth rate of about 2% (so inflation level of 0-2% per year for example).

High (Hi): represents the high range of values where inflation is > 2% per year, wage growth, industrial production and so on are also higher than normally indicative of 0-2% real GDP growth per year.

Regression results:

Sequential feature selection model was employed first to establish a baseline.

Simiarly other models were employed such as ridge regression.

The models broadly identified Industrial production, retail sales and real private investment as the key drivers of QoQ change in real GDP.

These models were then used to predict mortgage rates in the economy.

Wages, industrial production, consumer credit and investment grade corporate bond yields were found to be the primary drivers of QoQ changes in mortgage rates.

Classification results:

These are perhaps the more powerful results due to the better categorization of features into 3 main categories.

The metrics of the performance of various models are shown below:

|  | **Model** | **Train Accuracy** | **Test Accuracy** | **Precision** | **Recall** | **F1\_Score** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | Logistic Regression | 0.951220 | 0.911111 | 0.797619 | 0.6875 | 0.725610 |
| 1 | K Nearest Neighbors | 1.000000 | 0.911111 | 0.772556 | 0.8625 | 0.807692 |
| 2 | Decision Tree | 0.939024 | 0.866667 | 0.687970 | 0.7500 | 0.711538 |
| 3 | Support Vector Machines | 0.951220 | 0.888889 | 0.703488 | 0.5875 | 0.612737 |

The results of the random forest analysis and its key metrics are shown below:

Optimal tree depth: 5

Training Accuracy: 0.9878048780487805

Test Accuracy: 0.9333333333333333

Precision: 0.6666666666666666

Recall: 0.8

F1 Score: 0.7272727272727272

Confusion Matrix:

[[38 2]

[ 1 4]]

**KEY TAKEAWAYS**

These all facilitate a better understanding of macroeconomic trends and knowing one data can help guide judgement about the other data release. In turn, one is better able to assess the impact on the market due to the new data release.

As such, one can use the release of data on these variables to guide one’s judgement of economy which in turn can guide one’s judgement of how the markets would behave.

**These models may be used to predict recessions, QoQ changes in mortgage rates and real GDP growth %. In turn, these data can help guide:**

**Home purchase decisions: one may want to delay home purchase decisions ahead of an impeding recession and into a sharply rising interest rate environment (such as the one experienced in 2022).**

**Investment allocation decisions:**

1. **one may prefer to delay deploying capital in mortgage bonds for example when mortgage interest rates are rising and real GDP is slowing down.**
2. **2. Similarly, one may was to delay investing in broad stock markets in a recessionary environment accompanied by rising inflation rates (comparable to the 2022 market environment).**
3. **One may consider repositioning the portfolio ahead of release of CPI/PPI data based on the prior release of the corresponding PPI/CPI data respectively.**
4. **One may want to develop this analysis further to find a leading economic indicator index based on the features identified here.**
5. **One may want to keep a close eye on rising wages: a wage price spiral is well documented as a major economic risk of any environment.**