Team 32 Project Progress Report

Hank Corrion, Josh Santiago, Ben Metzger

10/26/25 | CPTS 475, Section 1 | Dr. Assefaw Gebremedhin

Redefinition of Project Goals and Scope

Previously, our project goal was to create a composite economic health indicator to predict market sentiment and direction. To expand the scope and depth of our project, we have refined our question to **test how macroeconomic indicators predict market direction vs. sentiment-based indicators**. To test our question, we will be using three different statistical models: **Long Short-Term Memory Neural Network, Random Forest Regressor with Bootstrapping and Boosting, and a Logistic Regressor Model**. We will predict market values using our models and compare the average accuracy metrics of each model using macroeconomic and sentiment-based indicators.

Macroeconomic Indicators

Macroeconomic indicators are official statistics that summarize recent economic conditions. These indicators are produced by agencies like the BEA and BLS and are published via the St. Louis Fed's FRED database. Since many of these indicators are published months after the time series that they represent, they are typically lagging or coincident indicators. Some Macroeconomic Indicators we will use in this project are *GDP, CPI, unemployment rate, initial jobless claims, and federal funds rate.*

Sentiment-based Indicators

Sentiment-based indicators reflect investors' perceptions of current economic conditions. Because they respond rapidly to news and shifts in investments, they are more indicative of short-term market movements. Given their volatility, these series are noisy and benefit from light preprocessing, and because many are correlated, **PCA can** reduce dimensionality. Since these indicators are based on current sentiments of the market, they are categorized as **coincident or leading indicators**. Some Sentiment-Based Indicators we will use in this project are the *VIX*, *S&P trading volume/direction, put/call ratio, 10y-3m treasury yields, and the fear and greed index.*

Data

Regarding the data necessary for this project, we chose to collect economic and financial data from 1970 through the most recent available dates to capture a sufficiently long and relevant history that represents today's market dynamics for our models. This timeframe is also convenient because many FRED and Yahoo Finance series reliably extend back to 1970. In

Python, we imported the raw CSVs from FRED (fredapi) and Yahoo Finance (yfinance) for all the listed indicators above with the exception of put/call ratio. Since FRED and Yahoo Finance both don't provide a put/call ratio time series we can import, we plan to pull daily SPX or SPY option volume data from the CBOE (Chicago Board Options Exchange) and compute the daily ratio ourselves. Currently, all of the CSVs have been consolidated into a single pandas dataframe organized by date, and we're working on cleaning and preprocessing it (date alignment, standardizing time periods, handling missing data, handling extreme outliers, etc) to prepare it for modeling.

Project Progress

Our team met up on two different occasions to discuss refining our project question and researching the subject thoroughly to find different datasets to work with. Since redefining our project question and expanding our scope, we have made decent progress towards data collection and methodology. We have updated our research focus which compares the power of macroeconomic and sentiment-based indicators when it comes to market prediction.

When expanding our datasets to use, beyond the initial list of variables, we have sourced additional datasets to include outlets. We hope to continue progressing by beginning to develop our code and models and get closer to answering our questions.