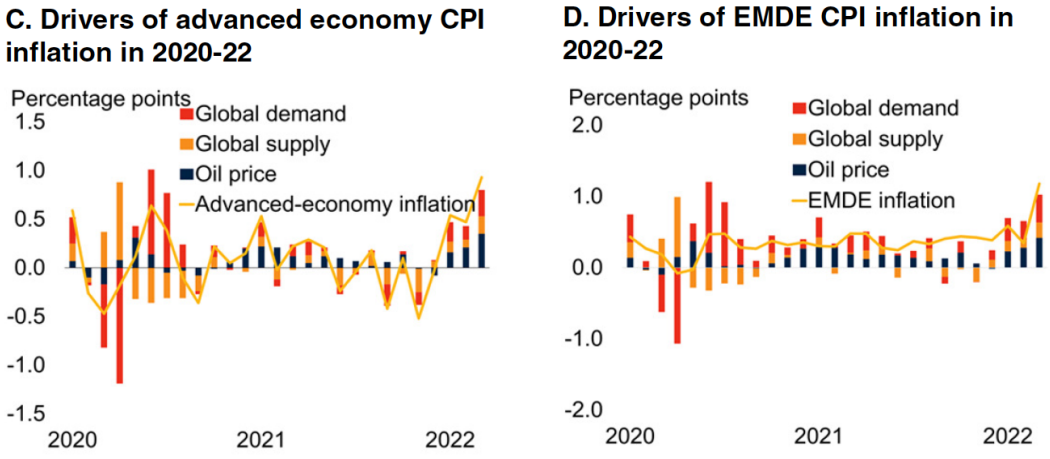
Group Project: The Effect of Macroeconomic Indicators on Stock Market Performance
Team 011: Thomas Byrne, Liangqu Chen, Jari Oinas, Anuj Shelat, David Chang

Project Progress Report

1. Overview

In this project, we are analyzing the relationship between various macroeconomic factors during 1967 – 2022 and the performance of the S&P 500 stock market index in the United States. Our primary research question is stated as follows: “What is the relationship between the U.S. stock market and macroeconomic indicators for the United States?”

Since 2020, global stock markets have been characterized by high volatility due to market disturbance caused by the COVID-19 pandemic (2020-Present) and the Russo-Ukrainian War (2022-Present). The current macroeconomic environment has been characterized by persistent, high inflation driven by increases in oil, food prices, and other supply/demand shocks.



Sources: Ha, Kose, and Ohnsorge (2021a); U.S. Bureau of Labor Statistics; World Bank.
Note: CPI = consumer price index; EMDEs = emerging market and developing economies.
A. Share of tradable goods and services in different inflation measures in the United States. PPI =producer price index.
B.-D. Contributions to month-on-month inflation in headline CPI for 83 countries, of which 31 are advanced economies and 52 are EMDEs, based on FAVAR models over the period of 2001M1-2022M3. Unexplained residual is omitted from the graph.

Subsequent fiscal policy has tightened at an unprecedented rate, driving consumer sentiment to a record-low. As a result, the S&P 500 market index has had its worst first half of a year since 1970.

As our project proposal stated:

> Our hypothesis is that a strong economy (as indicated by strong or healthy macroeconomic indicators) causes strong market performance. The underlying assumptions behind our hypothesis are the mainstream economic theories of price equilibrium, rational expectations, and the efficient market hypothesis. Given that the analysis will be done with data over a long period of time based on quarterly returns of major equity indices, the research group assumes that individual stocks will not have an effect and the market will be efficient. We assume that investors are rational and make rational investment decisions, and that stock market prices will move towards a price equilibrium that is defined by economic fundamentals such as supply and demand. In efficient markets, macroeconomic factors should be the only predictors of the value of market indices. Therefore, the regression analysis done by the team should result in a strong correlation and that we will be able to reject the null hypothesis.

In this project, we will build and fit a linear regression model taking the form of:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{21} + \ldots + \beta_{pj}X_{ip} + \varepsilon_{ij}$$

using S&P 500 stock market return data and examine the model to see if there is a linear relationship between the stock market and macroeconomic predictors in the United States.

2. What Progress has been Made?

Project progress as of current date has been significant. The following tasks were completed by Team 011:

- We completed a preliminary academic literature review of the relationship between the stock market and macroeconomic predictors.
- We selected and compiled data from various sources (FRED, yfinance APIs) and completed requisite data transformations.
- We have begun exploratory data analysis.
- We have begun the process of fitting the predictors to the linear regression model and checking the model's adherence to the assumptions of the linear regression model (especially with regards to multicollinearity).

3. Literature Review

The link between stock market behavior and macroeconomic variables, have been illustrated by different theories in the academic literature, including the Efficient Market Hypothesis, the Dividend Discount Model and the Arbitrage Price Theory (Alshogeathri 2011).

The Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH), is a hypothesis that states that share prices reflect all information and consistent alpha generation is impossible, regardless of the investment strategies used. According to the EMH, stocks always trade at their fair value on exchanges, making it impossible for investors to purchase undervalued stocks or sell stocks for inflated prices. The only way for an investor can obtain higher returns is by purchasing riskier investments (Investopedia).

The Efficient Market Hypothesis can be distinguished into three forms: weak, semi-strong and strong market efficiency (Alshogeathri 2011):

- The weak form: current stock prices incorporate all relevant past information
- The semi-strong form: current stock prices fully reflect all available public information such as the past price of the stock, how the company is performing, expectations regarding macroeconomic factors as well as public information about GDP, money supply and interest rate etc.
- The strong form: in addition to past and public information, stock prices also reflect private information about the specific company

Dividend Discount Model

The Dividend Discount Model (DDM) is a quantitative method used for predicting the price of a company's stock based on the theory that its present-day price is worth the sum of all of its future dividend payments when discounted back to their present value. Fair value of a stock is calculated irrespective of the prevailing market conditions (Investopedia).

DDM estimates the value of a common share at time t, using the relationship:

$$E_t[P_t] = \sum_{i=t+1}^{\infty} E_t[D_i] / (1 + r_t)^{i-t}$$

Where $E_t[P_t]$ is the expected intrinsic value or price that we would expect to pay for the share in the year t based on the information we have at the time. D_i is the nominal annual dividends we expect to be paid at time i, and r_t is the discount rate investors demand at time t (Foerster and Sapp 2005).

In further model development, macroeconomic factors are considered in dividend growth estimates. The growth rate of GNP is frequently argued to be the maximum sustainable growth rate for a firm's dividends. Other macroeconomic factors also may influence the firm's abilities to pay dividends. Booth (1998) forecasts the dividend growth using the model:

$$Divagro_t = \alpha_0 + \alpha_1 Yield + \alpha_2 Inflation_t + \alpha_3 GNPGrowth_t + \epsilon_t$$

where $Divagro_t$ is the dividend growth rate at time t, r_t is the yield on a long-term government bond to represent the opportunity costs for the firm and investors. $Inflation_t$ is the overall rate of inflation based on the year-over-year changes in the consumers price index (CPI) to capture changes in overall level of risk. $GNPGrowth_t$ is the year-over-year growth in GNP and is the residual error term.

The Arbitrage Pricing Theory

Arbitrage Pricing Theory (APT) is a multi-factor asset pricing model based on the idea that an asset's returns can be predicted using the linear relationship between the asset's expected return and a number of macroeconomic variables that capture systematic risk (Investopedia / Ross 1976).

$$R_{it} = r_i^f + \beta_i X_t + \epsilon_t$$

where R_{it} is the return of stock i at time t,R_i^f is the risk free interest rate or the expected return at time t, X_t is a vector of the predetermined macroeconomic factors or the systematic risks and β_t is the measure of the stocks sensitivity to each of these macroeconomic factors. ϵ_t is the error term representing unsystematic risk (Alshogeathri 2011).

The Arbitrage Price Theory does however not specify which or how many macroeconomic factors to include in the modelling of the stock return. Hence, a large number of different macroeconomic factors such as the interest rate, money supply, inflation and exchange rates have been included in a large body of empirical studies based on reasonable theory (Alshogeathri 2011).

4. Data Preparation and Wrangling

For this project, macroeconomic data and market return data has been pulled from Federal Reserve Economic Data (FRED) and yfinance using their respective APIs. Macroeconomic and finance data is from Jan 1967 until Mar 2022. The data points and data sources are shown in Appendix 1.

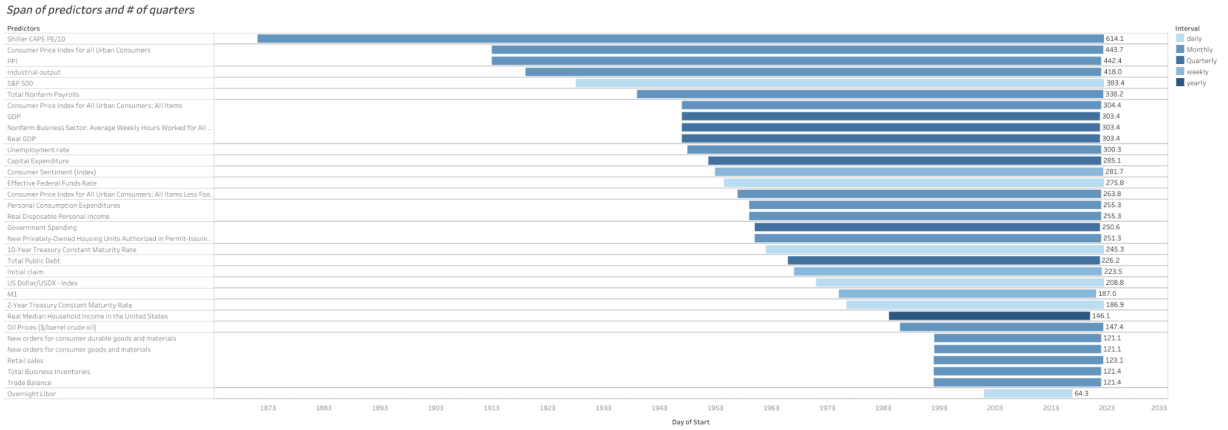


Figure 1: Time Span of Predictors and Number of Quarters of Data available for study

A key consideration was selecting a time frame we should study and which predictors we should include in our data set for our model. The data set we initially compiled included 29 variables (28 independent variable, 1 dependent variable for the S&P 500 market returns) but the availability of data is different for each variable, with some variables having data going all the way back to 1873 (through present date) and others only as far back as 1993 (through present date). Ultimately, we decided on the current time frame due to a desire to capture a wide range of different macroeconomic environments (such as the inflationary period during the 1970's, dotcom tech bubble in the late 1990's, etc) in our analysis of the stock market.

An issue that has risen in our data wrangling is on what frequency our analysis should be. The time frequency of our data ranges from daily to quarterly, depending on the variables. Our team has decided that we will fit the model to both quarterly and monthly time frames, because based on initial data exploration it appears that we achieve a better fit, when using longer time frames. It is particularly interesting that the results vary, and we are interested to analyze this more in detail.

Correlation between the different macroeconomic variables and SP500 is shown in Figure 2:

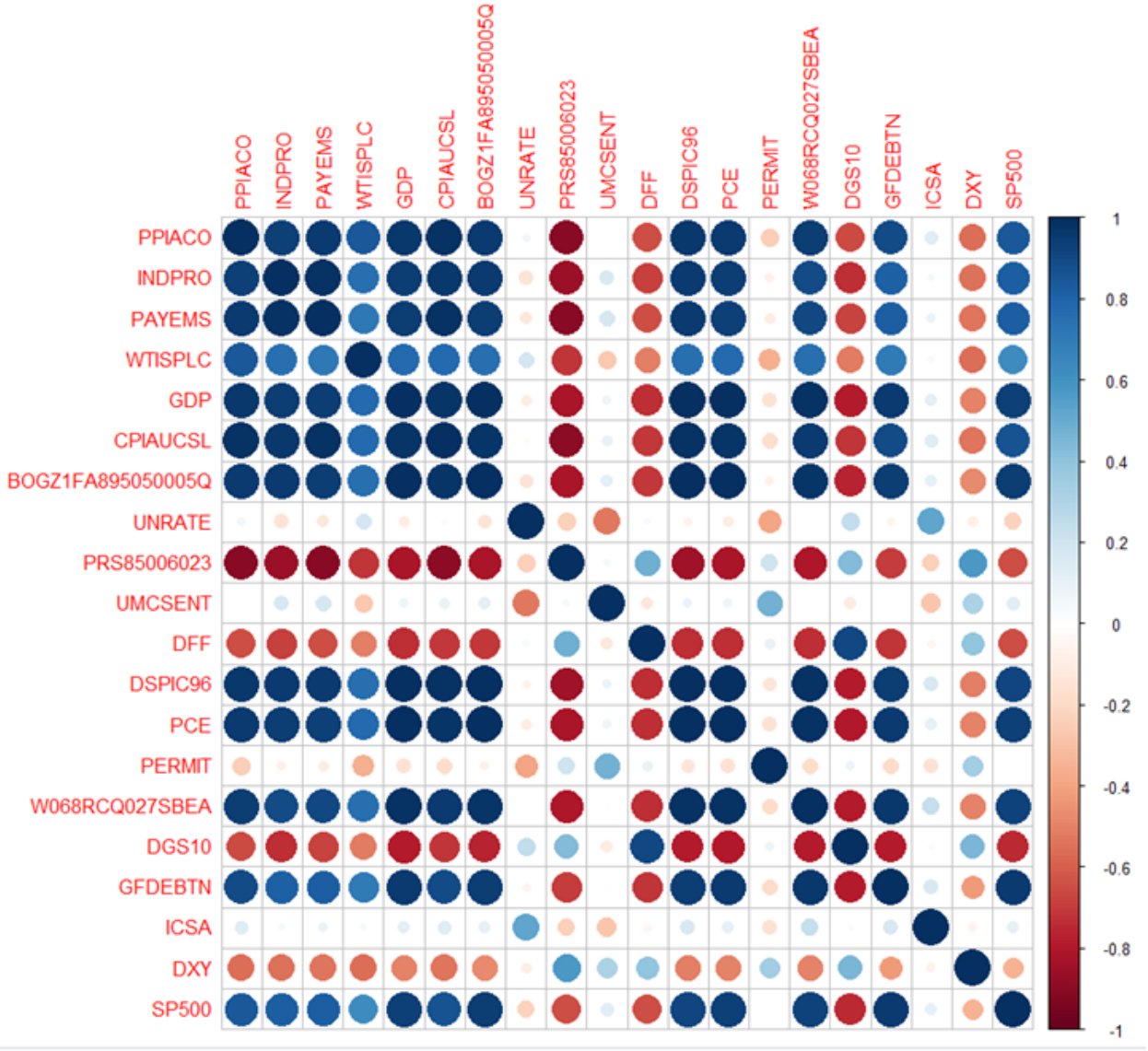


Figure 2: Correlation between macroeconomic variables and SP500 (Quarterly)

As expected based on the above reviewed financial theories, there is:

- Positive correlation between SP500 and:
    - Producer price index (PPIACO)
    - Industrial output (INDPRO)
    - Total number of employees (PAYEMS)
    - Oil price (WTISPLC)
    - Gross Domestic Product (GDP)
    - Consumer price index (CPIAUCSL)
    - Capital expenditure BOGZ..)
    - Real disposable personal income (DSPIC96)
    - Personal consumption expenditures
    - Government spending (W068RCQ027SBEA)

    - Total public debt (GFDEBTN)
- Negative correlation between SP500 and:
    - Avg. hours worked for all employed (PRS85006023)
    - Federal Funds Effective Rate (DFF)
    - 10-year treasury constant maturity rate (DGS10)
- Unemployment rate (UNRATE), initial employment claims (ICSA), new privately-owned housing units permits (PERMIT), US Dollar index (DXY) had weaker correlations and we will remove them from further analysis.
- A bit surprisingly consumer sentiment (UMCSENT) didn't have a big correlation on SP500 during the investigated time period, but this is probably due to the longer time frame and overtime the effects balance out. We'll still keep this indicator for analysis on SP500 performance during shorter time frames.

Correlation between remaining macroeconomic variables and SP500 is shown in Figure 3:

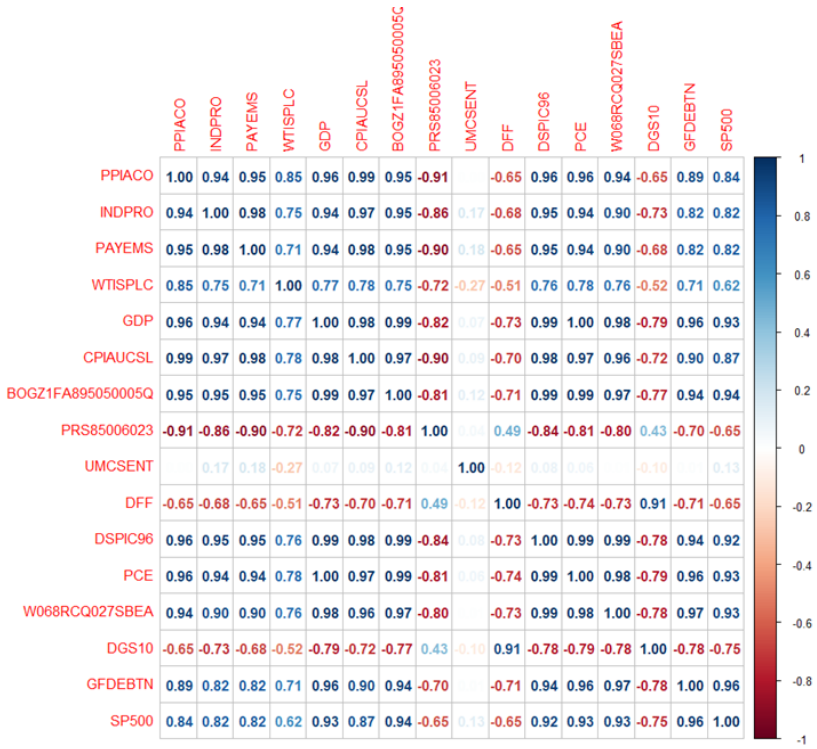


Figure 3: Correlation between macroeconomic variables and SP500 (Quarterly)

As we work on improving the linear regression model's fit, we have been reviewing each predictor and checking for model adherence to the assumptions of linear regression: multivariate normality, no or little multicollinearity, no auto-correlation, amd homoscedasticity. Of particular concern is multicollinearity; given that all of the independent variables are measures or indicators of the macroeconomic environment, we would expect at least some of the variables to be correlated with one another. We will discuss our assessment in the below section.

5. Data Exploration & Analysis

Trend for SP500 and macroeconomic variables is shown in Figure 4.

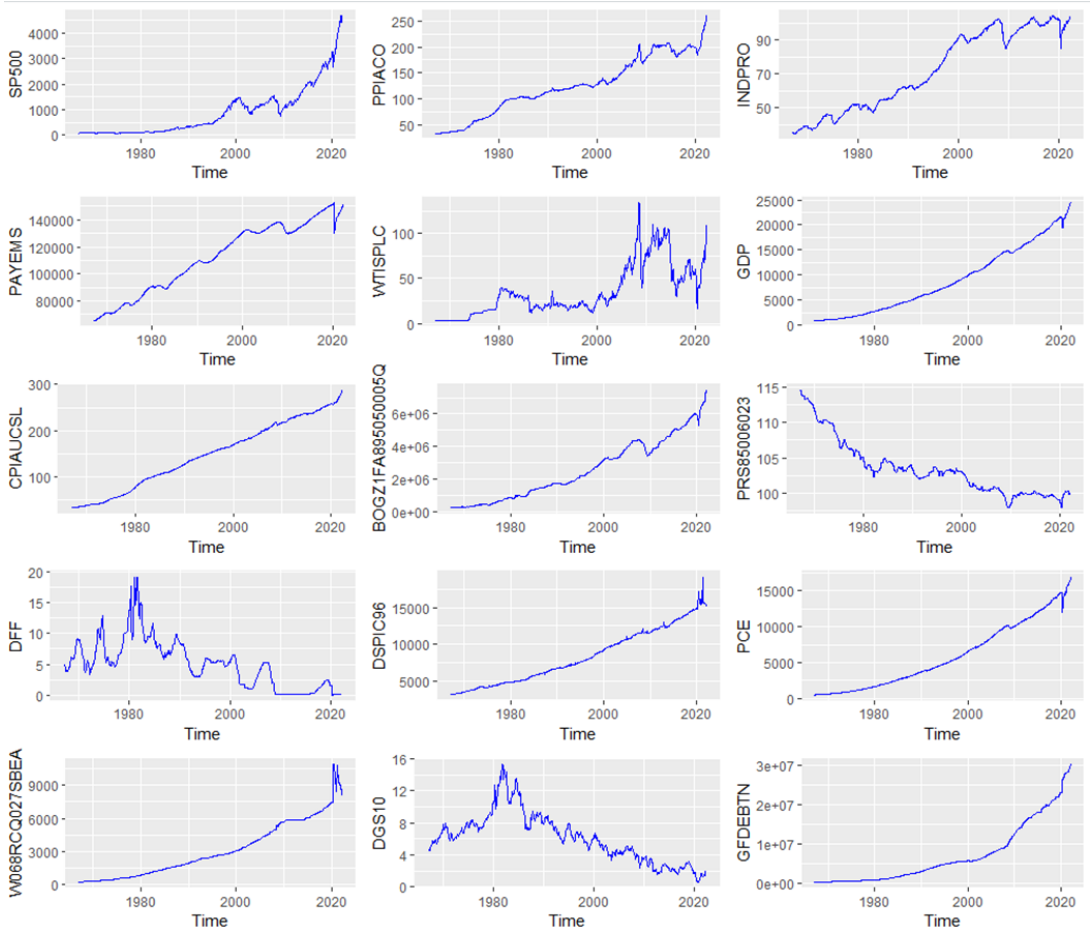


Figure 4: Time Series trends of SP500 and Macroeconomic variables

Our initial exploratory data analysis of the data (on a monthly time frame) had a focus especially on model fit for the assumptions of linear regression.

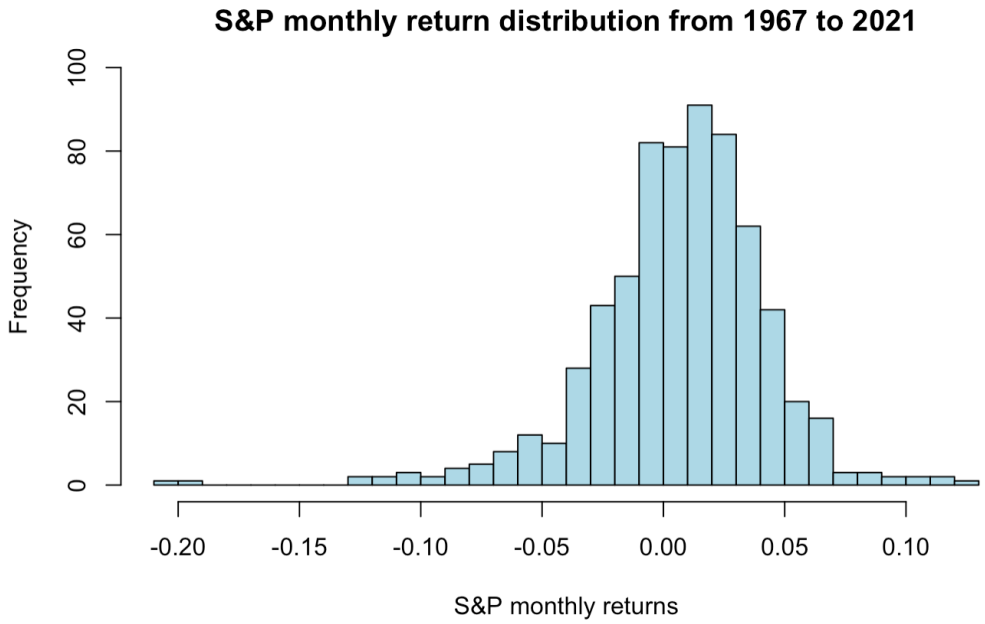


Figure 5: S&P monthly return distribution shows outlier or high leverage point data points on the left tail, with -20% monthly returns.

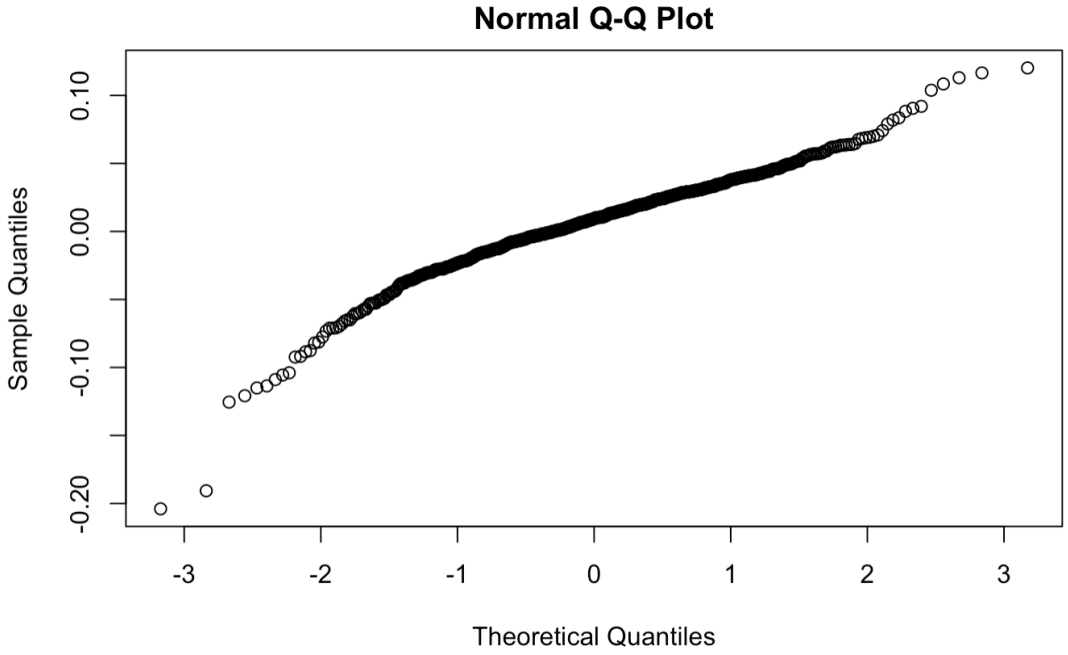


Figure 6: Q-Q Plot: data is mostly normally distributed, but there are potentially outliers.

We completed Grubbs' Test on both tails of the S&P 500 data, and found that 4 data points could be classified as outliers. In our most recent team discussion, it was

generally agreed that the outlier data points should be removed, though there was some debate due to said data being a result of historical real returns.

As mentioned in our project proposal, of particular concern, was the possibility of multicollinearity. We reviewed the data's variance inflation factor chart and checked for multicollinearity in a correlation plot of the predictors.

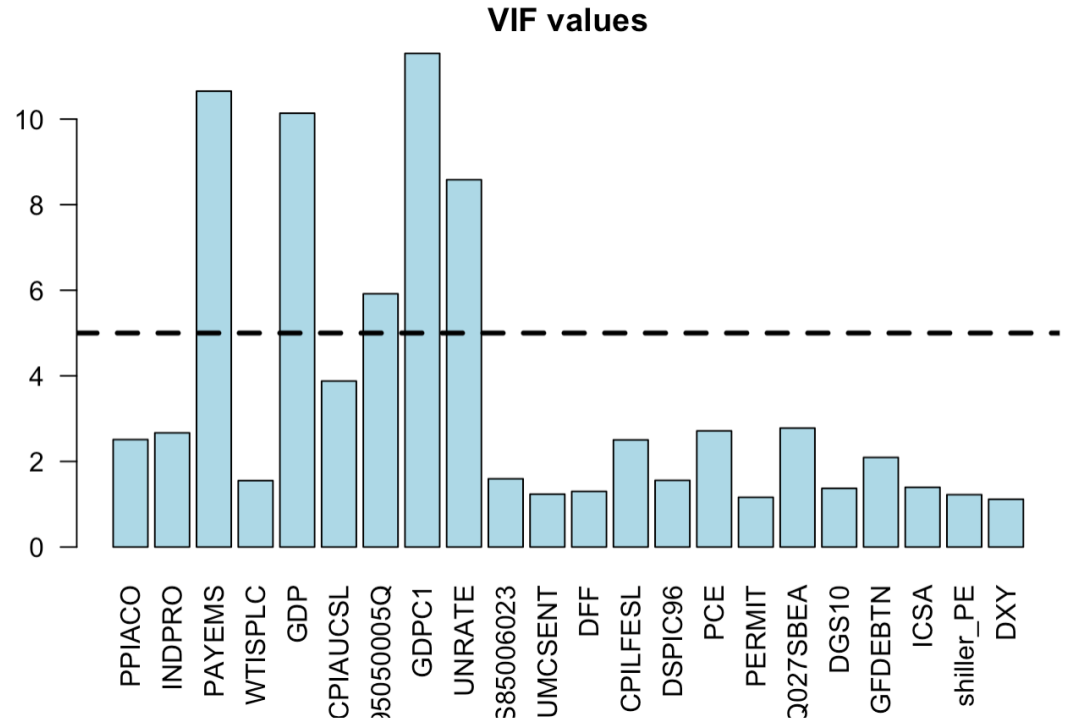


Figure 7: Problematic collinearity may be indicated where VIF > 5.

As shown in the VIF chart, independent variables that might indicate problematic collinearity are: PAYEMS, GDP, BOGZ1FA895050005Q, GDPC1, UNRATE. As discussed in our progress report video presentation, some of these variables are different measures of the same macroeconomic phenomenon (e.g. GDP and GDPC1 (real GDP) are different indicators of gross domestic product). As such, it would make the most sense to include only one of the respective variables. We have proposed excluding or removing different combinations or sets of variables.

It is readily apparent that a key step in our model development will be variable selection. We plan to use either stepwise (using the Bayes Information Criterion) or LASSO regression to inform the model's variable selection decisions. We have met as a team on a weekly basis to discuss these above issues. We discussed next steps at our latest team meeting and plan to proceed accordingly.

6. Conclusions

In Team 011's first two weeks, we have compiled data for our model, completed data transformations, begun exploratory data analysis, and have started the process of fitting our linear regression model. We have addressed various issues that have arisen over the course of the first half of the team project. Initial model development appears to support the team's hypothesis that there is a strong relationship between macroeconomic predictors and the performance of the stock market.

In next steps, the team plans to complete model fitting over the next week, using variable selection and regularization methods to improve fit. We hope to complete a comprehensive academic literature review. If time permits, we will consider developing other models or examining particular time periods of stock market performance.

Appendix 1: Data Points

Predictors	start	end	Interval	API	quarters	FRED Code	Var_name
PPI	1/1/1913	5/1/2022	Monthly	FRED	444	PPIACO	PPIACO
Industrial output	1/1/1919	5/1/2022	Monthly	FRED	419	INDPRO	INDPRO
S&P 500	12/30/1927	6/22/2022	daily	CSV WSJ+kaggle	383		SP500
Total Nonfarm Payrolls	1/1/1939	5/1/2022	Monthly	FRED	338	PAYEMS	PAYEMS
Oil Prices ($/barrel crude oil)	1/1/1946	5/2/2022	Monthly	FRED	310	WTISPLC	WTISPLC
GDP	1/1/1946	1/1/2022	Quarterly	FRED	308	GDP	GDP
Consumer Price Index for All Urban Consumers: All Items	1/1/1947	5/1/2022	Monthly	FRED	306	CPIAUCSL	CPIAUCSL
Capital Expenditure	10/1/1946	1/1/2022	Quarterly	FRED	305	BOGZ1FA895050005Q	BOGZ1FA895050005Q
Unemployment rate	1/1/1948	5/1/2022	Monthly	FRED	302	UNRATE	UNRATE
Nonfarm Business Sector: Average Weekly Hours Worked for All Employed Persons	1/1/1948	1/1/2022	Quarterly	FRED	300	PRS85006023	PRS85006023
Consumer Sentiment (Index)	11/1/1952	5/1/2022	weekly	FRED	282	UMCSENT	UMCSENT
Effective Federal Funds Rate	7/1/1954	6/24/2022	daily	FRED	276	DFF	DFF
Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	1/1/1957	5/1/2022	Monthly	FRED	265	CPILFESL	CPILFESL
Real Disposable Personal Income	1/1/1959	4/1/2022	Monthly	FRED	257	DSPIC96	DSPIC96
Personal Consumption Expenditures	1/1/1959	4/1/2022	Monthly	FRED	257	PCE	PCE
New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units	1/1/1960	5/1/2022	Monthly	FRED	253	PERMIT	PERMIT
Government Spending	1/1/1960	1/1/2022	Quarterly	FRED	252	W068RCQ027SBEA	W068RCQ027SBEA
10-Year Treasury Constant Maturity Rate	1/2/1962	6/24/2022	daily	FRED	245	DGS10	DGS10
Total Public Debt	1/1/1966	1/1/2022	Quarterly	FRED	227	GFDEBTN	GFDEBTN
Initial claim	1/7/1967	6/18/2022	weekly	FRED	225	ICSA	ICSA
US Dollar/USDX - Index	1/31/1967	6/21/2022	daily	CSV yahoo finance	225		DXY