CMSE 202 Final Report

Macroeconomic Factors and Market Performance

Below is a list of group contributors' names and GitHub usernames:

Elizabeth Allard (allarde2), Sarah Bejleri (Sarahbejleri), Shaurya Arora (arorash2), Joseph Boberg(bobergjo), Ryleigh Bergmann (subbeeftobeans), Harshita Boddu (boddusri-edu)

GitHub Repository: https://github.com/allarde2/CMSE-202-Macroeconomic-Factors.git

RESEARCH QUESTION:

How strongly do macroeconomic factors explain or predict changes in major stock indices over time?

For our project, our goal was to investigate the relationship between macroeconomic indicators and the performance of the three major U.S. stock indices – S&P 500, Dow Jones, and NASDAQ. The scope of our project was to highlight which indicators are the strongest and hold the highest predictive power for stock indices over time. We emphasized the conceptual portion of each model we used to fully understand the statistical and predictive strength of macroeconomic variables including CPI, Real Interest Rate, Unemployment Rate, GDP Growth, M2 Money Supply, and the Federal Funds Rate.

METHODS:

We constructed three models to answer our question:

1. **Multiple Linear Regression**: examine the linear relationships and statistical significance of each variable's ability to predict the stock market index.
2. **Lasso Regression:** an extension of Multiple Linear Regression, to understand variable importance using regularization and variable selection.
3. **Time-Series with Lasso Regression:** incorporate first differencing to capture the predictive power of the previous quarter on the current quarter.

For all three models, each was implemented across all three indices using Python in Jupyter Notebooks. We sourced our data from **Yfinance** and **The Federal Reserve**, ranging quarterly from 1992 to 2023.

RESULTS:

1. For **Multiple Linear Regression**, our models for all three indexes had high R-squared values. For S&P 500 R-squared was 0.935, Dow Jones R-squared was 0.958, and NASDAQ R-squared was 0.886. In all three models, the top two predictors in terms of magnitude were Unemployment and M2 Money Supply. However, a caveat of Multiple Linear Regression is collinearity, which leads us into Lasso.
2. For **Lasso Regression,** it helped control collinearity as it sends unimportant variables to zero at a lower lambda value. We learned that M2 Money Supply, Unemployment Rate, and GDP Growth remained key drivers. However, the predictive accuracy decreased due to regularization penalties, but it was highly effective for feature selection.
3. For **Time-Series Lasso Regression**, all the RMSE values for all indices improved tremendously. GDP, Unemployment, and M2 Money Supply remained the three indicators with the highest predictive power. However, it was learned that the previous quarter's stock index is not a significant driver of that index for S&P 500 and Dow Jones but it was for NASDAQ specifically.

DISCUSSION:

Our findings confirm that macroeconomic factors, labor market (Unemployment Rate) and Growth (GDP), and Monetary Policy (M2 Money Supply) are strong predictors of stock market performance. Economically, these relationships make sense in terms of strength and direction, as unemployment negatively affects the economy, while GDP and M2 positively affect the economy. However, all our models failed to accurately predict the stock market index, particularly during 2020, as macroeconomic variables cannot explain pandemic fears and government regulation. Time-series modeling allowed us to see that adding first-differencing

terms did help predictive accuracy, but ultimately had a difficult time predicting sporadic changes, which are common in time-series data. Overall, the models were helpful for understanding which variables had the highest predictive power, but our models lacked true accuracy.

CONCLUSION:

In conclusion, macroeconomic indicators are a somewhat powerful tool for explaining and predicting S&P 500, Dow Jones, and NASDAQ index trends. Multiple Linear Regression and traditional Lasso regression provided good explanatory power of what variable is considered the most important for driving these indices. However, time-series allowed us better forecasting accuracy as for all three indexes it reported the smallest RMSE values. Ultimately, no model was perfect in terms of explanatory power and prediction accuracy, but they provided valuable insights that M2 Money Supply, Unemployment, and GDP Growth have the most pull-on stock market indices over time.

FUTURE EXPLORATION:

Based on the findings of our project, several interesting questions arose that could be useful to consider for future exploration. A particularly intriguing next step would be to investigate the non-linear relationships between macroeconomic indicators and stock indices using more advanced machine learning models such as Random Forests, Gradient Boosting, or Neural Networks, which may be better fit for capturing complex interactions missed by linear regression methods. It would also be crucial, if possible, to incorporate sentiment analysis from financial news or social media outlets to enhance the models' abilities to account for investor and consumer behavior. This would also be pivotal for taking account market reactions that are not

reflected in traditional economic variables such as during volatile periods like the COVID-19 pandemic and the Great Recession. Finally, another idea to consider for future exploration would be to expand on the dataset to include international macroeconomic variables and/or sector-specific indices to assess whether the relationships and trends we found hold true across global markets or different sectors of the economy.