Supervised Learning for Long/Short Portfolio Bias

Mason Laney[1], Kevin Le[1], and Pierre-Nicolas Perrin[1]

[1]Department of Computer Science, The University of North Carolina at Chapel Hill

May 6, 2022

1 Motivation

Capital preservation is the foremost purpose of a traditional hedge fund, followed by absolute positive returns. To achieve this mandate, a fund can take the approach of a long-short (LS) strategy. Portfolio managers that run an LS approach are able to choose what risks their portfolio's positions are exposed to, whether those risks are market, sector, stock, or a combination. Standard LS portfolios have half their exposure in each direction. However, to promote the objective of absolute positive returns, a portfolio may take on net long or short exposure. Various implementations of the LS strategy decide this net exposure based on macroeconomic trends. This report analyzes the potential for supervised learning methods on quantitative macroeconomic data as a tool to help asset managers trading on a 20-60 day time frame decide their portfolio bias.

2 Background Work

Domain education will provide the reasoning behind initial available features. A proprietary trading process[1] taught by highly regarded traders from the Institute of Trading and Portfolio Management (ITPM) suggest the following macroeconomic leading indicators; 2-year US Government Treasury Yield, 10-year US Government Treasury Yield, 10-year TIPS (Inflation Indexed US Treasury) Yield, Inflation (10-year US Gov Treasury Yield - 10-year TIPS Yield), 10-year - 2-year US Treasury Yield Curve, Junk Corporate Debt Yield, Junk Corporate Debt Yield - 10-year US Government Treasury Yield Spread, VIX Volatility Index, ISM Manufacturing Index, M2 Money Supply, University of Michigan Consumer Sentiment Index (UMCSI) Composite, UMSCI Future Expectations, UMCSI Current Conditions, and New Residential Construction Permits. These indicators have the benefit of being interpretable which is significant for fund managers when discussing with investors. For this analysis, the percent change of these leading indicators will be used as the basis of the features.

Surveying the available data set before applying models is standard in quality data science practices. In Figure 1 below, we plot our target variable against each feature to examine potential correlation.

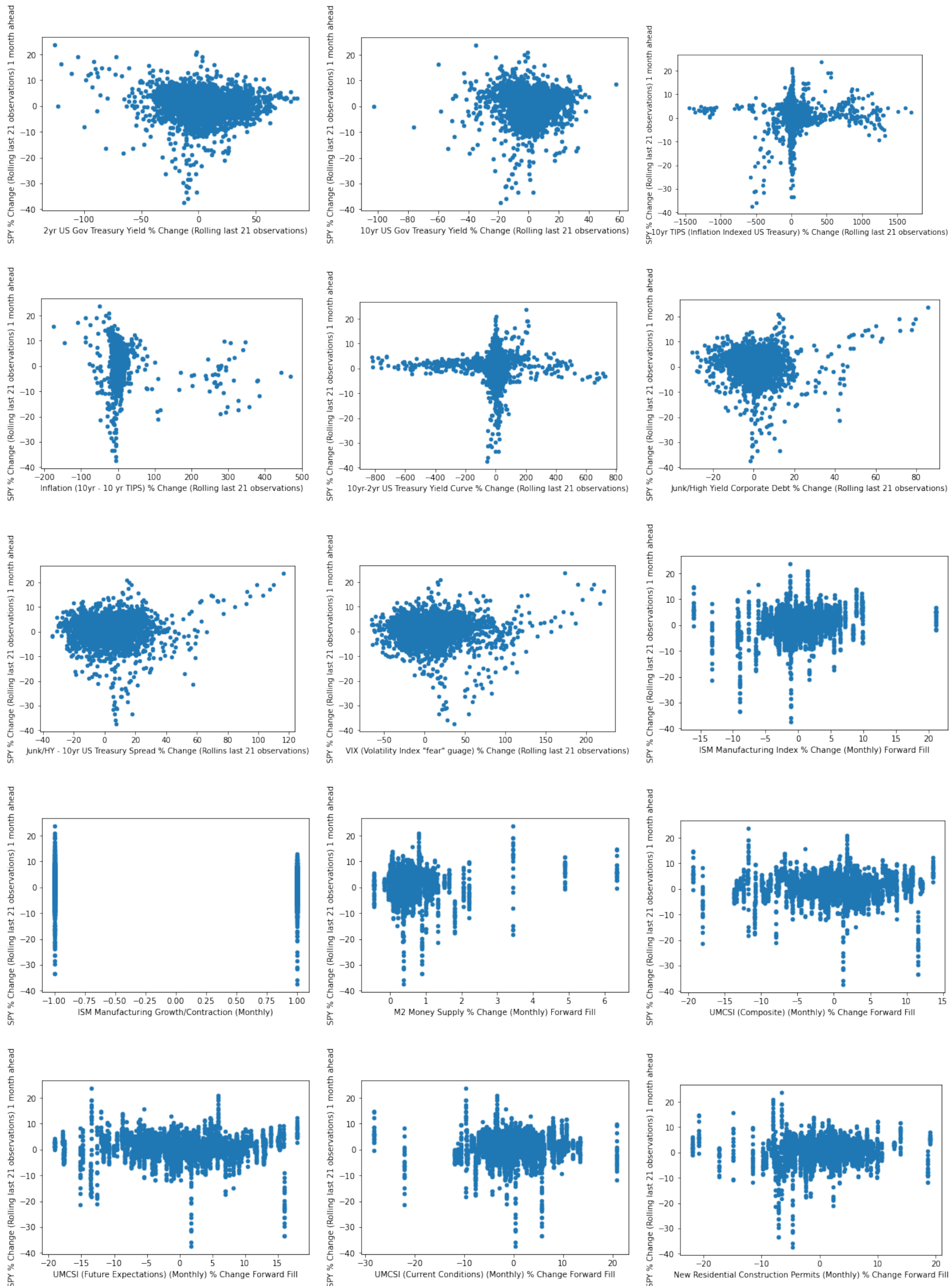

Figure 1: SPY % Change (last 21 observations) vs. Leading Indicator % Change (last 21 observation/monthly forward fill).

3 Data Processing

Raw values for the leading indicators were extracted from official sources [2-7], however, additional data processing was required to properly prepare the data for the models. For a portfolio manager trading on a 20-60 day time frame, predicting the amount of change in the market in one month's time is crucial in deciding net exposure. We decided to use the SPY ETF as our proxy for "the market" and the last 21 observations to represent the time period of one month (the stock market is closed on weekends and some holidays).

Our features are based on the percent change of the leading indicators and not the raw values. There are some leading indicators that are reported on a daily basis while others on a monthly basis. For each indicator that is reported on a daily basis, our corresponding feature is calculated as the total percent change of the last 21 observations. For indicators that are reported on a monthly basis, the corresponding feature records that singular monthly percent change for all samples that are observed during the reported monthly period (forward filled observations until the next reported value). Since we are trying to predict the next 21 observation period (next month) with the latest 21 observation period (latest month), for each sample, our SPY ETF target variable is one month (21 observations) ahead of the features in the data set.

4 Methods

We are comparing the following baseline models in our study: Linear Regression (LR), Ridge Regression (RR), Random Forest (RF), K-Nearest Neighbors (KNN), Support-Vector Machine (SVM), and Regression-Adapted Multilayer Perceptron (MLP). We decided to use Explained Variance, Max Error, and Mean Absolute Percentage Error as our metrics to compare the aforementioned models. While we understand that the data might not follow a linear relationship with respect to some features, we wanted to include the linear regression models in our baseline testing in order to contrast their results with those of the SVM and non-linear models and to extract insight on the causal relationships between the features and the SPY performance.

In addition to these baselines, we wanted to test the performance of Ridge Regression on a subset of our data. We hand-selected three features (Junk Corporate Debt Yield, Junk Corporate Debt Yield - 10-year US Government Treasury Yield Spread, and VIX Volatility Index) which appeared to have linear relationships with the target variable. The linear-seeming nature of these features meant that we could more confidently apply linear regression techniques to the data. We tested Ridge Regression on each of the three individual features versus the target in addition to the three features as a group versus the target. For these models, we decided to use the additional evaluation metric of R2 score, which we felt was appropriate now that the task was explicitly a linear regression one.

We implemented all of our models using scikit-learn. We used a random 80-20 train-test split to divide our dataset, and we calculated evaluation metrics using the metrics module from sklearn.

We fine-tuned the hyperparameters for the models by hand. We found that $\alpha = 1.0$ was optimal for Ridge Regression, while KNN was most effective when $k = 3$. Random Forest worked best with 50 trees and 30% of the features being used. For the MLP regressor, we found that $\alpha = 0.01$ led to the consistently best results.

5 Results

None of our baseline models performed particularly well. Random Forest was ahead of the pack by a significant margin, as seen in Table 1; however even this model had a value of 95% on the Mean Absolute Percentage Error metric.

Model	MAPE	EV	Max
Linear Regression	1.4594	0.1060	31.306
Ridge Regression	1.4594	0.1060	31.306
Random Forest	**0.9501**	**0.8432**	**10.619**
KNN (k = 3)	1.3249	0.6149	21.665
SVM	1.6690	0.1575	28.012
MLP ($\alpha = 0.01$)	1.5223	0.5285	22.248

Table 1: Evaluation metrics of different baseline models. MAPE: Mean Absolute Percent Error; EV: Explained Variance; Max = Max Error. Low scores are good for MAPE and Max, while high scores are good for EV. The Random Forest model (which uses 50 trees and 30% of features; fine-tuned by hand) achieves best performance across all three metrics.

K-Nearest Neighbors was the next-best performing model, with a MAPE score of 132%, while the Multi Layer Perceptron Regressor came in third with MAPE score of 152%. The Linear Regression, Ridge Regression, and Support Vector Machine models performed especially poorly at this task, with all of them having Explained Variance scores of well under 0.5, meaning the target is only lightly correlated with the variables according to the model.

The results from performing linear regression on a subset of features are shown in Table 2.

Model	MAPE	EV	Max	R2 Score
RR (Feature 5)	1.3030	0.0007	32.290	-0.0021
RR (Feature 6)	1.3029	0.0035	32.324	0.0008
RR (Feature 7)	1.3035	-0.0002	32.287	-0.0030
RR (All three)	1.3099	0.0046	32.406	0.0019

Table 2: Performance of Ridge Regularized Linear Regression using three individual features that appear to be linearly related to the target variable, as well as performance of Ridge Regression using the three features as a group. The model with all three features performs similarly to the models that only use a single feature. (Feature 5: Junk Corporate Debt; Feature 6: Junk Corporate Debt - 10yr US Treasury Spread; Feature 7: Volatility Index)

The three individual models performed comparably to the model which used all three features. All had MAPE values of around 130%, extremely low Explained Variance scores, and Max error scores of over 32. All models had R2 scores very close to 0, indicating that there is essentially no linear relationship between the variables. Thus, we can determine that this data is not suited for a linear regression approach.

6 Discussion

The results were not satisfying enough to suggest using our leading indicators to predict market change in the proceeding month. Given our best model had a MAPE over .9, the potential error is too much for a portfolio manager to confidently base their potfolio's net exposure on. We believe that one or more of these are possible reasons behind our poor performance: (1) relatively low amount of data (while we have around 4000 data points for training, more data might help extract the sometimes vague relationships between market indicators and SPY price change); (2) the non-sequential models we used are not optimal for this task - Seq2Seq models (thus implementing position-encoding for the data time series) and other structured prediction models might help prediction performance; (3) our current features don't account for external factors.

References

[1] Institute of Trading and Portfolio Management (ITPM), 2022. https://www.itpm.com/

[2] Federal Reserve Board Data Release. Board Of Governors of the Federal Reserve System, 2022. https://www.federalreserve.gov/datadownload/

[3] ICE BofA US High Yield Index Effective Yield. Federal Reserve Bank of St. Louis, 2022. https://fred.stlouisfed.org/series/BAMLH0A0HYM2EY

[4] Money Stock Measures. Board Of Governors of the Federal Reserve System, 2022. https://www.federalreserve.gov/releases/h6/current/default.htm

[5] Survey of Consumers. University of Michigan, 2022. http://www.sca.isr.umich.edu/

[6] ISM Report On Business. Institute for Supply Management, 2022. https://www.ismworld.org/supply-management-news-and-reports/reports/ism-report-on-business/

[7] New Residential Construction. United States Census Bureau, 2022. https://www.census.gov/construction/nrc/index.html