Team 14 Final Report

July 24, 2022

Contents

1 Team Members

- Aaditya Vinnakota (Aaditya316)
- Bao Vo (baovo022293)
- Frank Corrao (fcorrao)
- Michael Obin (Obin_michael)
- Nicholas Coyle (nscoyle)

2 Project Overview

2.1 Goals

The goal of our analysis is to identify stock attributes and performance trends which are indicative of resilience during economic downturns. To do so, we leverage data for securities in the S&P 500 during several 'down markets' - identifying the stocks which outperform the market and analyzing the traits they have in common (e.g. industry, company size, etc).

2.2 Approach

The metric we use to identify outperformance is Jensens Alpha. We run a standard Jensen's Alpha regression:

$Return_{Stock} - Return_{rf} = \alpha_{stock} + B_1[Return_{sp500} - Return_{rf}]$

For each period, we identify the top 30% of positive alpha stocks and code them as outperformers. This output beceomes the target of a logistic regression to identify features predective of that good performance:

$Outperform_{stock} = B_0 + B_1 * [LTMReturns] + B_2 * [LTMVolatility] + B_3 * [IsFinance?] + B_4 * [IsIndustrials?] + B_4 * [IsRealEstate?] + B_5 * [UnemploymentRate] + B_6 * [InterestRate]$

Variables that survive backwards-selection and are statistically significant will be deemed features that predict outperformance.

2.3 Initial Hypothesis / Literature Review

Through this research, we have formed several hypotheses about which types of companies might outperform in recession. We have identified two papers, (Woszczyk 2019) and (Ozkan 2009), which explore investment strategy during economic downturns. Researchers suggest that 'vice stocks' tend to outperform as folks resort to 'bad habits' (e.g. drinking, smoking, gambling) during times of unemployment Sources 2 (Ozkan 2009). They also find that healthcare companies are resilient, as demand can be relatively inelastic for medical care (Woszczyk 2019).

3 Overview of Data

Most of the data for this analysis was from sources that were relatively clean. The key componets we needed for the analysis:

- For identifying outperformers
  - Data to identify recessionary periods
  - Stock and index returns
  - Index components
- For identiying attributes that predict outperformance
  - Stock specific attributes (industry, common factor exposures, etc)
  - Macroeconomic factors

3.1 Identifying recessions

We observice market performance during several recession periods in the United States as idenfitied by the Federal Reserve GDP based recession indicator.

Due to limitations in access to historical data, we evalute performance to the three most recent recessions:

- January 2001 - October 2001
- December 2007 - June 2009
- January 2020 - September 2020

3.2 Stock returns

Corporate action adjusted stock and index prices were pulled from yahoo via the tidyquant R package. From this data, monthly returns were calculated. S&P500 Index components were idenfied from publicly available sources.

3.3 Stock specific attributes and Macroeconomic factors

Stock sectors were pulled from Yahoo via tidyquant. Macroeconomic factors such as interest rate and unemployment rate were pulled from Federal Reserve Bank of St. Louis (FRED), also via tidyquant.

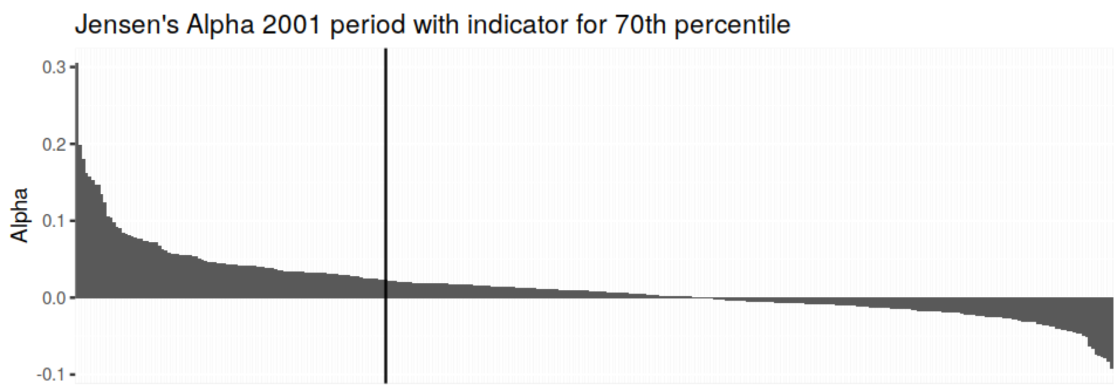
4 Overview of Modeling

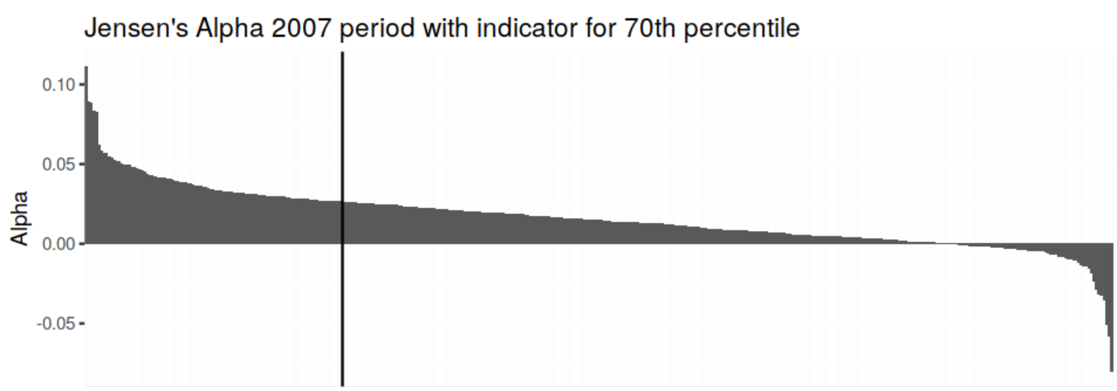
4.1 Identifying outperformers

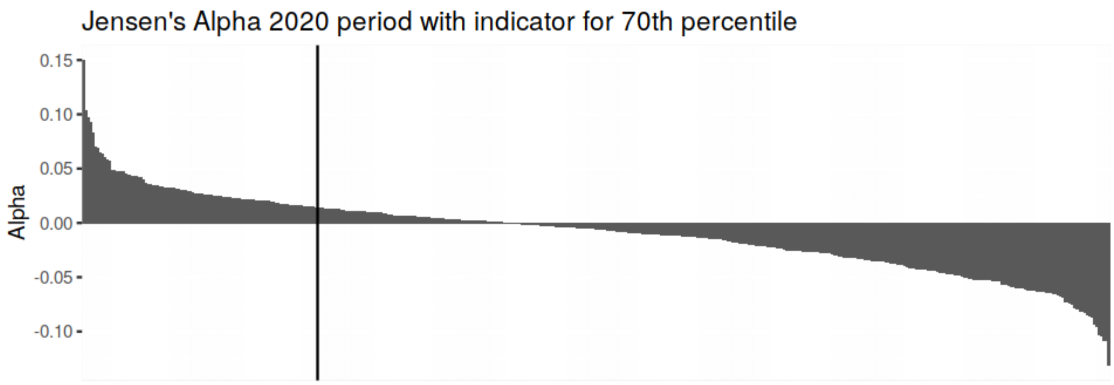
Identifying companies that do well in absolute return space during the peak to through phase of a business cycle is difficult. It may be the case that the vast majority of stocks are losers, simply due to exposure to market beta. Instead, we focus on identifying those companies that outperform the market, which could easily form the basis of a market neutral strategy. The metric we use to identify outperformance is Jensens Alpha. We run a standard Jensen's Alpha regression:

$$Return_{Stock} - Return_{rf} = \alpha_{stock} + B_1[Return_{sp500} - Return_{rf}]$$

We run this regression for S&P500 stocks for each recession period available to us, with the S&P500 return in placeof the market return. Fitted alpha values are mapped to class labels {outpeformer/Not outperformer} by finding the cutoff for the top 30% percentile alpha in each period, marking everything above that cutoff that is positive as an Outperformer. A threshold above the median was selected through exploratory analysis showing that this cutoff gave consistently positive returns of $> 2\%$, which should be enough to cover implentation costs and buffer against regression standard errror. With these labels now in place, we can now use our outperformers in a classifier.







4.2 Identifying predictors of outperformance

With our coded outperformers over multiple recessionary periods in hand, we have a target to train a classifer on. We chose a logistic regression for its relatively good performance and interpertability.

$Outperform_{stock} = B_0 + B_1 * [LTMReturns] + B_2 * [LTMVolatility] + B_3 * [IsFinance?] + B_4 * [IsIndustrials?] + B_4 * [IsRealEstate?] + B_5 * [UnemploymentRate] + B_6 * [InterestRate]$

There is a long list of possible factors that could predict market returns, but our imagination is limted by our data as we could not secure access to a professional quality source of equities data. Some of the features we were able to consider are:

- Trailing 12 month returns
- Trailing 12 month volatility
- Company sector
- Average unemployment rate in the period
- Interest rates in the period

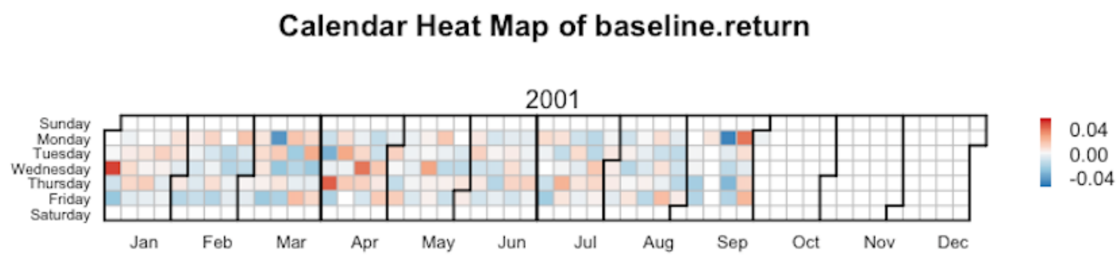
```
Call:
glm(formula = performance ~ LMT_std + mean + interest_mean +
    unemp_mean + con_dis + fin + inds + real_est + tech + tele,
    family = binomial(link = "logit"), data = combine_converted)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.8365  -0.3506  -0.0852   0.0579   2.5415

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    13.8584     2.7900   4.967 6.79e-07 ***
LMT_std         0.2873     0.0261  11.008  < 2e-16 ***
mean            1.7857     0.1242  14.379  < 2e-16 ***
interest_mean  -2.1478     0.3371  -6.372 1.87e-10 ***
unemp_mean     -2.1846     0.3260  -6.701 2.08e-11 ***
con_dis         0.5661     0.2920   1.939  0.05256 .
fin             1.0735     0.3725   2.882  0.00395 **
inds            0.7418     0.3558   2.085  0.03705 *
real_est        1.1853     0.4272   2.775  0.00552 **
```

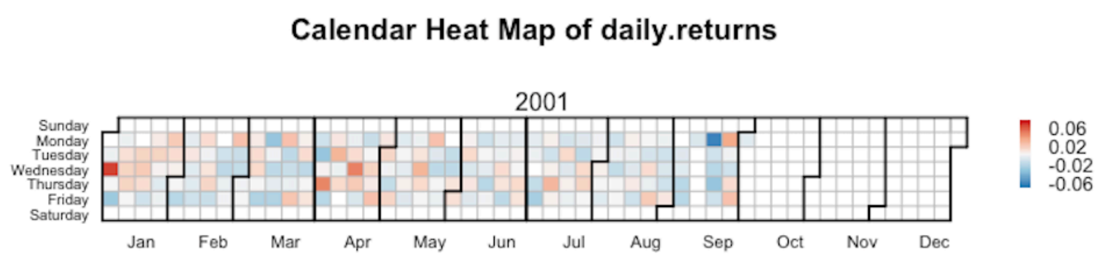
```
tech            0.6436     0.3857   1.669  0.09521 .
tele            1.6074     1.4452   1.112  0.26606
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1412.08  on 1159  degrees of freedom
Residual deviance:  577.36  on 1149  degrees of freedom
AIC: 599.36

Number of Fisher Scoring iterations: 7
```

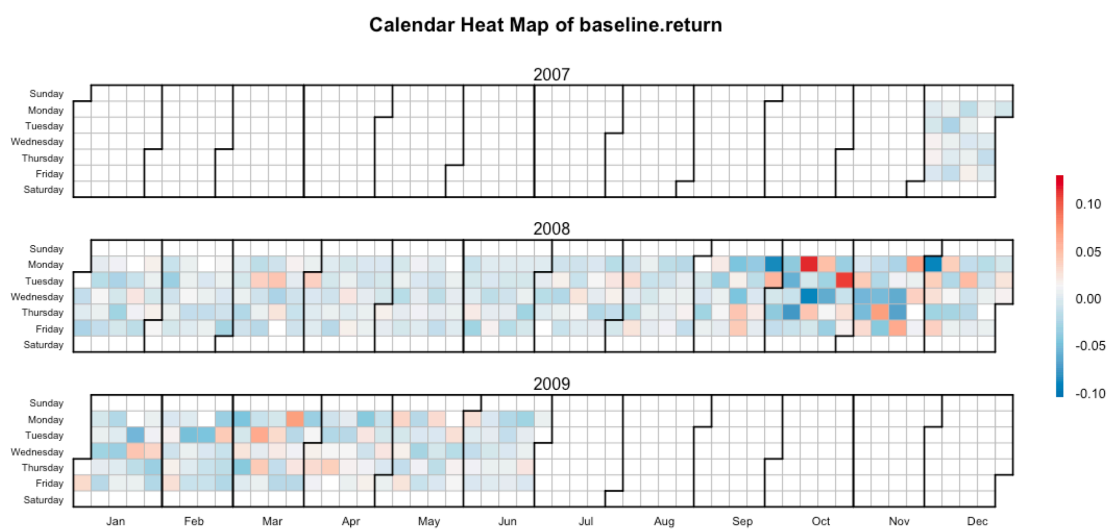
We perform backwards-selection to remove some insignicant variables to get a final regression output

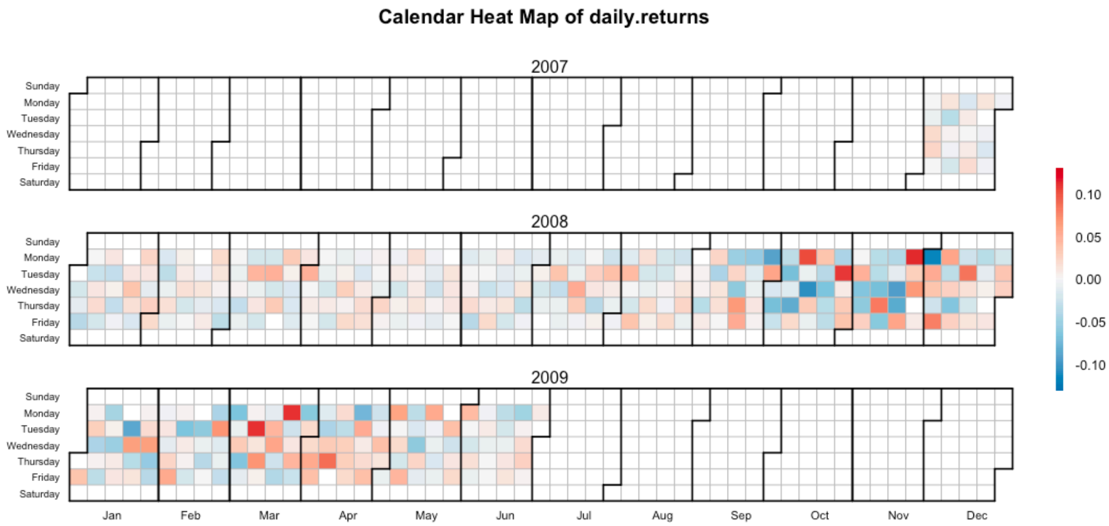
```
Call:
glm(formula = performance ~ LMT_std + mean + interest_mean +
    unemp_mean + fin + real_est, family = binomial(link = "logit"),
    data = combine_converted)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.8360  -0.3618  -0.0892   0.0545   2.6010

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   14.23872    2.75610   5.166 2.39e-07 ***
LMT_std        0.29344    0.02563  11.451  < 2e-16 ***
mean           1.79070    0.12362  14.485  < 2e-16 ***
interest_mean -2.15924    0.33401  -6.465 1.02e-10 ***
unemp_mean    -2.18552    0.32307  -6.765 1.33e-11 ***
fin            0.64191    0.32525   1.974   0.0484 *
real_est       0.75411    0.38624   1.952   0.0509 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1412.08  on 1159  degrees of freedom
Residual deviance:  584.25  on 1153  degrees of freedom
AIC: 598.25

Number of Fisher Scoring iterations: 7
```

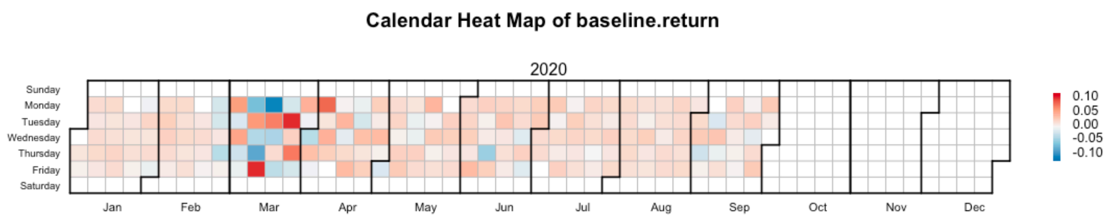
4.3 Outperformance Group Vs. Market During Each Period

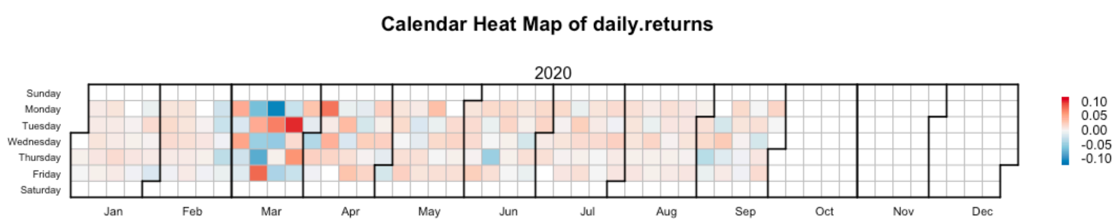












4.4 Findings

Six variables are significant at the 95% confidence interval:

- Trailing Company Returns
- Trailing Company Volatility
- Is Finance Sector
- Is Real Estate
- Average interest rate
- average unemployment rate

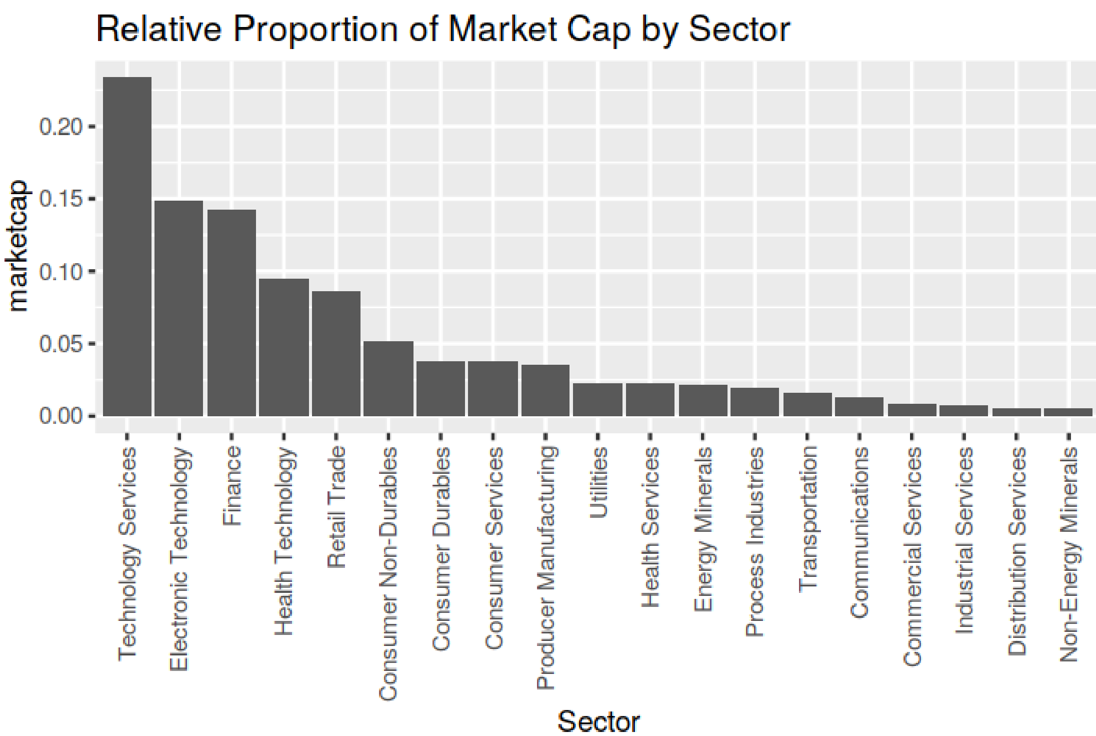
Strong performance in the year leading up to the recessionary period was indicative of outperformance, as is stock volatility. The finance an real estate sectors show outperformance during recessions. As unemployment and interest rates rise, peformance decreases.

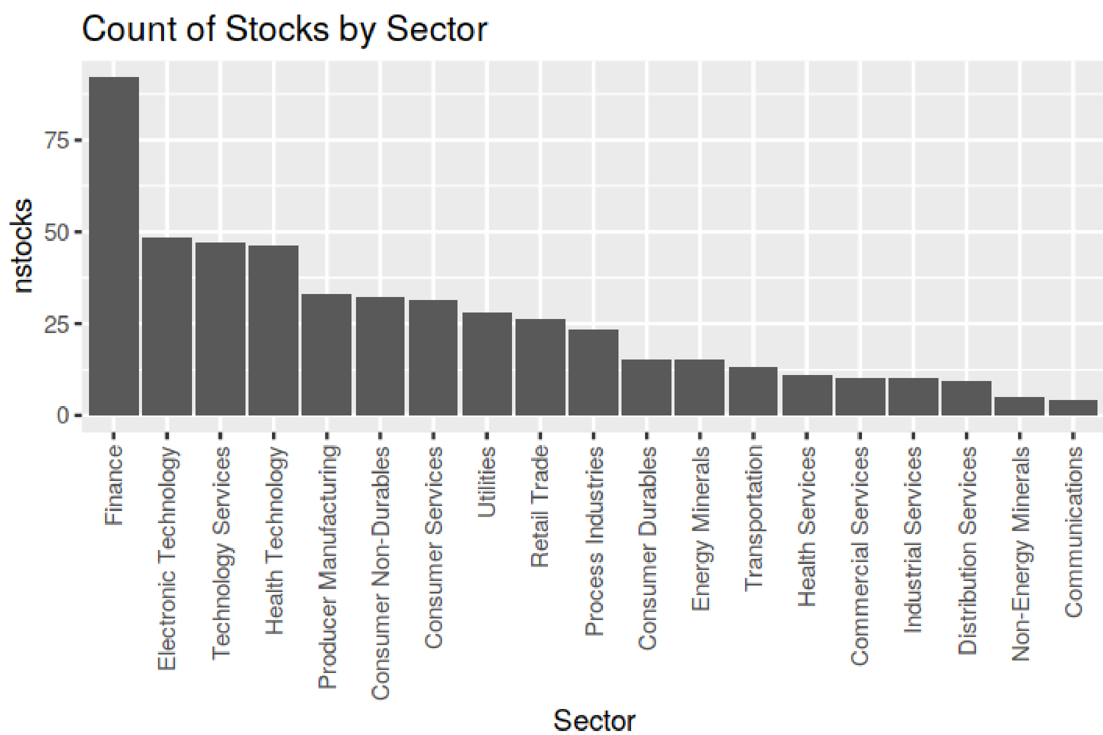
4.5 Challenges

There are several key challenges that must be overcome to make these results more robust and realistic:

1. Limited Historical Data: as our access to historical S&P 500 data was limited, we could only evaluate three recession periods. Further, within those recession periods, there were some individual stocks for which data was unavailable (e.g. for the 2001 recession period, we only had data for 340 stocks in the S&P 500). We did our best to mitigate this by using code to track the changes in the S&P 500 listed in wikipedia
2. Limited Operational Performance Data: operational performance data (e.g. market cap, operating income, profit margin) may have been useful in predicting resilience during downturns. However, access to this data required significant investment
3. Non-stationarity: long periods of time between recessions means the patterns we find in our training data may simply no longer exist due to naturally changing market and macroeconomic dynamics

An additional caveat is that by using a market cap weighted index like the S&P 500, some bias in industry composition might introduced.





As can be seen, there is a large disparity in sector coverage within the index. Industries such as real estate, materials, & telecommunications consist of a dramatically smaller subset of the securities used. Consequently, for industries such as telecommunications, it is reasonable to assume that there is a degree of overfitting for stocks in that industry, and others with a relatively small composition. Although we could have arbitrarily added additional stocks to have a more balanced weighting between industries, this would have resulted in selection bias.

Another natural challenge was due to the fact we are dealing with non-stationarity in our financial time series data. In financial data, this is the case when there are significantly different regimes within the data that may affect how our response variable changes. bAlthough there are many ways the market dynamics changed drastically over the periods that we examined, the most noticeable difference that caught our attention was the difference in interest rates between the three recession periods. As shown in the table below, the interest rates were markedly different during the three periods.

```
library(IRdisplay)
display_png(file="../../../Recession-Proof_Portfolio/Visualizations/Picture1.png",
    height=1000, width=500)
```

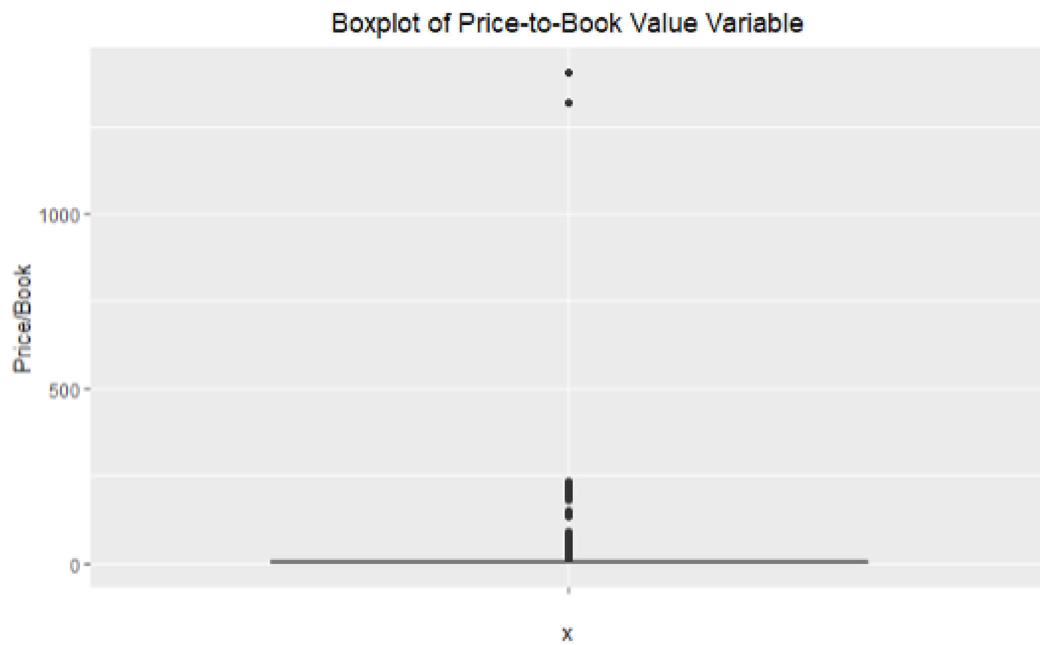
Recession Period	Average Interest Rate During Period
January 2001 - October 2001	4.27%
December 2007 - June 2009	1.50%
January 2020 - September 2020	0.47%

5 Future Direction

There are a number possible ways to continue this work, the most important being to acquire a longer history of data, and more stock specific data, like common equity factor exposures, to regress against. Another important future direction would be to include the use of non-linear classifiers, or non-linear feature transformation as some patterns may be complicated. For instance, though we found that trailing performance/momentum is predictive of outperformance during a recession, perhaps the biggest outperformers could perform poorly if they are part of a bubble that has formed. Lastly, should a recession be upon us, it will be important to use that to validate our results.

Although we used a logistic regression in the end to determine which types of stocks would outperform, we had different ideas initially. We wanted to replace this model with a k-means clustering model. The k-means model would still use the output from the Jensen's alpha regression. With those results, our plan was to then find categories that distinguished between the groups of underperformers and outperformers and then see if there were any interesting commonalities between the subgroups that the algorithm developed.

As we went deeper into our exploratory data analysis, we noticed that there were some fundamental characteristics of some of our features that we wanted to use that made it difficult to implement this method. When we examined the price to book value ratio variable, we noticed that there was a very severe degree of skewness in its distribution as shown in the boxplot below:



Further investigation of data transformation of by Y and X variables is likely warrented

6 Conclusion

Upon completion of this ambitious project, there were many important learning outcomes for us, and important practical applications of this exercise that we will be able to apply to our personal investment portfolios, and the portfolios of others that want to have an edge in turbulent markets. The main learning point that resonated with us through all of our team discussions was the fact that this undertaking was more of an art than a science. Although we had complex models at our disposal, the majority of the value add that we brought to this endeavor was from being able to discern between the strengths and weaknesses of the tools that were available and choose the best combination in-light of the problem that we were facing. Although there were some instances where imperfect information, or lack of information altogether lead us to situations that left us wanting to do more, we were mindful of the business problem with all of our decisions and were able to develop a model that all groups' members have expressed interest in using in their personal lives.

We can summarize our key findings in two parts. From a macroeconomic level, rising unemployment rates & rising interest rates are indicative of outperformance. From a firm specific level, strong performance and volatility in the period leading to a downturn in addition to being classified as a real estate or financials company are conducive to outperformance. These findings are powerful and have immediate relevance given the growing level of hearsay concerning a potential forthcoming economic downturn.

7 Sources

1. Woszczyk K, 2019, 'Do mutual funds invest in recession-proof industries prior to crisis?', MSc thesis, Erasmus University Rotterdam, Rotterdam (link)
2. Ozkan F C, Xiong Y, 2009, 'Wise Investing: Analysis of the recession-proof sin stocks', MBA thesis, Simon Fraser University, British Columbia (link)
3. Tidyquant, source of stock data (via Yahoo) and macroeconomic times series (via Fred) (link)
4. Kaggle, guidance for performing S&P 500 Analysis in R (link)
5. Schwab, 'Macro-economic factors of fundamental analysis' (link)
6. Stijn Claessens, M. Ayhan Kose, and Marco E. Terrones, 2008, 'What Happens During Recessions, Crunches and Busts?' (link)
7. U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/UNRATE, July 18, 2022 (link)
8. Board of Governors of the Federal Reserve System (US), Federal Funds Effective Rate [FEDFUNDS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/FEDFUNDS, July 18, 2022 (link)
9. Hamilton, James, Dates of U.S. recessions as inferred by GDP-based recession indicator [JHDUSRGDPBR], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/JHDUSRGDPBR, July 18, 2022 (link)