final_report_notebook

July 23, 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).

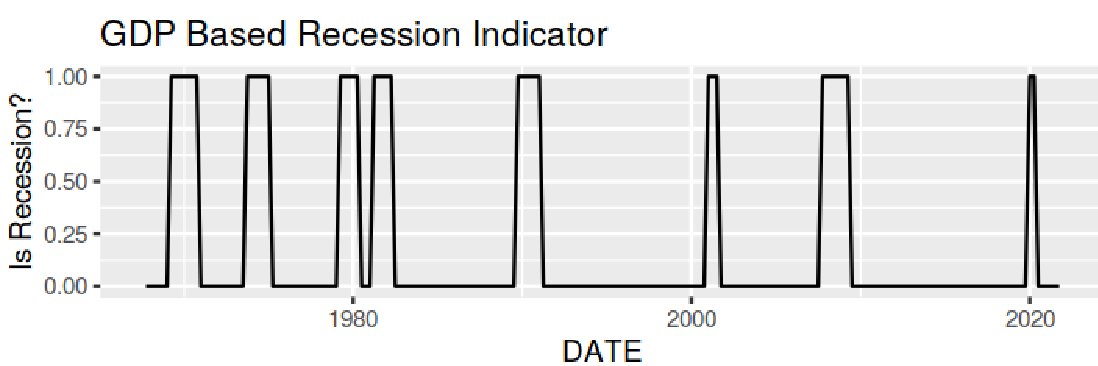
3 Overview of Data

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 identify 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 hard, we have a target to train a classifer on. We chose a logistic regression for its relative 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 + fin + inds + real_est + tech + tele, family = binomial(link = "logit"),
    data = combine_converted)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.8381  -0.3601  -0.0893   0.0580   2.6502

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   14.30926    2.77160   5.163 2.43e-07 ***
LMT_std        0.29190    0.02585  11.293  < 2e-16 ***
mean           1.79170    0.12388  14.463  < 2e-16 ***
interest_mean -2.17822    0.33619  -6.479 9.22e-11 ***
unemp_mean    -2.20710    0.32494  -6.792 1.10e-11 ***
fin            0.77164    0.33571   2.299   0.0215 *
inds           0.44253    0.31827   1.390   0.1644
real_est       0.88327    0.39513   2.235   0.0254 *
tech           0.33668    0.34981   0.962   0.3358
tele           1.34270    1.43504   0.936   0.3495
---
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:  581.14  on 1150  degrees of freedom
AIC: 601.14

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 + inds + real_est, family = binomial(link = "logit"),
    data = combine_converted)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.8332  -0.3675  -0.0883   0.0569   2.6245

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   14.35772    2.76455   5.194 2.06e-07 ***
LMT_std        0.29222    0.02563  11.400  < 2e-16 ***
mean           1.78874    0.12341  14.494  < 2e-16 ***
interest_mean -2.17711    0.33501  -6.499 8.11e-11 ***
unemp_mean    -2.20535    0.32434  -6.800 1.05e-11 ***
fin            0.70444    0.32953   2.138   0.0325 *
inds           0.37719    0.31205   1.209   0.2268
real_est       0.81622    0.38980   2.094   0.0363 *
---
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:  582.79  on 1152  degrees of freedom
AIC: 598.79

Number of Fisher Scoring iterations: 7
```

4.3 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.

4.4 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

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.

6 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)

7 Code

```
```