Recession Resistant Portfolio Picking Strategies

Daniel Muniz
Alejandro Suarez

Abstract

This project seeks to explore portfolio building techniques that will do well in bear markets. We look at quarterly and daily data that point to a stock being resistant to recessions and we sample portfolios to compare between resistant and non-resistant stocks. We find that some treatments give promising portfolios during both bull and bear markets. These results warrant more investigation into the relationship between these naive metrics and portfolio performance.

Project Overview

Most financial investors have three main goals: minimize losses, grow a portfolio of assets, and maximize income through dividends and returns [1]. These three goals are often in competition with each other, where investors try to balance risk and return to better achieve one over the others. By looking at daily and quarterly data for a company, an investor can determine how the stock might contribute to their portfolio's goals. This project looks to explore ways to select stocks for a portfolio that minimize losses, while still maintaining the potential of growth in line with the market.

Problem Statement

The focus of our analysis will be on the Great Recession, which started in Quarter 3 of 2008 and lasted until Quarter 2 of 2009. For the portfolios, we will consider stocks that were "resistant" to the recession in comparison to their peers. The project will look at various selection criteria that emphasize the demonstrated resilience of stocks to an extreme bear market.

Since our selection criteria does not focus on maximizing diversity or returns in the general case, it's expected that the average "recession resistant" portfolio will not outperform the market during a bull market. However, the goal is to create portfolios that will provide a positive return during both bull and bear markets. We created our proof of concept by creating portfolios made up of S&P 500 stocks.

Metrics

A variety of metrics were analyzed before the development of our hypotheses [1] [2] [3]. From the daily data we queried and constructed the following metrics.

Daily Returns:
- (Close Price- Open Price)/Open Price

Askhi:
- The highest share price that a *seller* was wishing to sell for

Bidlo:
- The lowest share price that a *buyer* was willing to buy for

Price:
- The fair price of a stock, the agreed upon price between a buyer and a seller

Spread:
- (Askhi-Bidlo), it is a measure of discordance between buyer and seller of stocks. When the market is volatile, spreads may increase

Price/Askhi:
- The proportion of the Askhi price a seller was willing to sell for. Thought of in a different way, (1-P/A) is the *discount* a seller was willing to deduct from their desired sale price.

Bottom Spread/Spread:
- In other words (Price-Bidlo)/Spread
- The proportion of the total spread that was in favor of the buyer of a stock. It was thought that perhaps this could be used as a leading indicator of future returns

Shares Outstanding:
- The total number of stock shares publically available to trade on any given day

Trade Volume:
- The total number of shares *traded* on any given day. This number is typically much higher than the volume

Trade GDP/Market cap:
- (Volume*Price)/(Shares Outstanding*Price) = Volume/Sout
- The proportion of a company's market cap that was traded on a given day

Expected Return:
- The average daily, monthly, quarterly return of a stock price

Risk:
- A measure of volatility, it is the standard deviation of the returns of a stock

Risk Free Rate:
- The expected earning of the market with zero risk. For daily and monthly returns a common benchmark is the US 3 Month Treasury Bill [1]

Excess Return:

- Expected Return - Risk Free Rate

Sharpe Ratio:
- Excess Return/Risk
- It is a ratio that weighs the portfolio's excess returns against the portfolio's risk

Earnings per Share:
- Earings(net profit)/Shares Outstanding
- This metric is reported on a quarterly basis

Dividend per Share:
- Dividend Payout/Share
- The amount paid out to shareholders per unit of share

Return on Equity:
- Net Income / Average Shareholder Equity

Price per Book ratio:
- Price/Book Value per Share
- The ratio of the stock price to the balance sheet value recorded per share

Price per Earnings:
- Price / EPS
- The ratio of the stock price to the earnings incurred per stock share

Below we consider a more in-depth look at key quarterly metrics to this project. It's important to remember that these metrics are considered imperfect, each with limitations on their scope; therefore, testing different metrics will test different properties of stocks. A goal of our analysis will be to see if a metric that's resilient to change in an extreme bear market translates to high-performing stocks for a resilient portfolio.

The Dividend-adjusted Price to Earnings and Growth (PEG) takes the ratio of the price of the stock to a company's earnings, dividends, and projected earnings growth . It tells the investor how valuable a company's shares are. Theoretically, a ratio of less than 1 means that a stock is undervalued by the market, and greater than 1 is overvalued. Growth is projected when looking at future data, and can vary greatly based on the methodology used [4]. But when looking at historical data, it is simple and accurate to calculate growth from the earnings data already available.

Earnings per Share (EPS) is the ratio of quarterly earnings to the number of shares outstanding. It's a measure of how well the company is performing relative to its presence in the market, and it tells an investor how much benefit they're receiving per share. Generally, the higher the ratio, the better.

Return on Equity (ROE) is found by dividing a company's net income by shareholders' equity. Shareholder's equity is calculated by finding the difference between a company's assets and its debt, so ROE can be thought of as the return on net assets. Judging an ROE value to be good or bad depends on context (company direction, industry, market trends), but in general, the higher the ratio, the better.

Sharpe Ratio

Created in 1966 by William Sharpe [5], the Sharpe Ratio is the primary metric used to judge our portfolios. It is a ratio that weighs the portfolio's excess returns against the portfolio's risk. Essentially, it is a measure of how much return is gained per point of risk. It tells the investor how efficiently the portfolio is trading increases in risk for gains in returns.

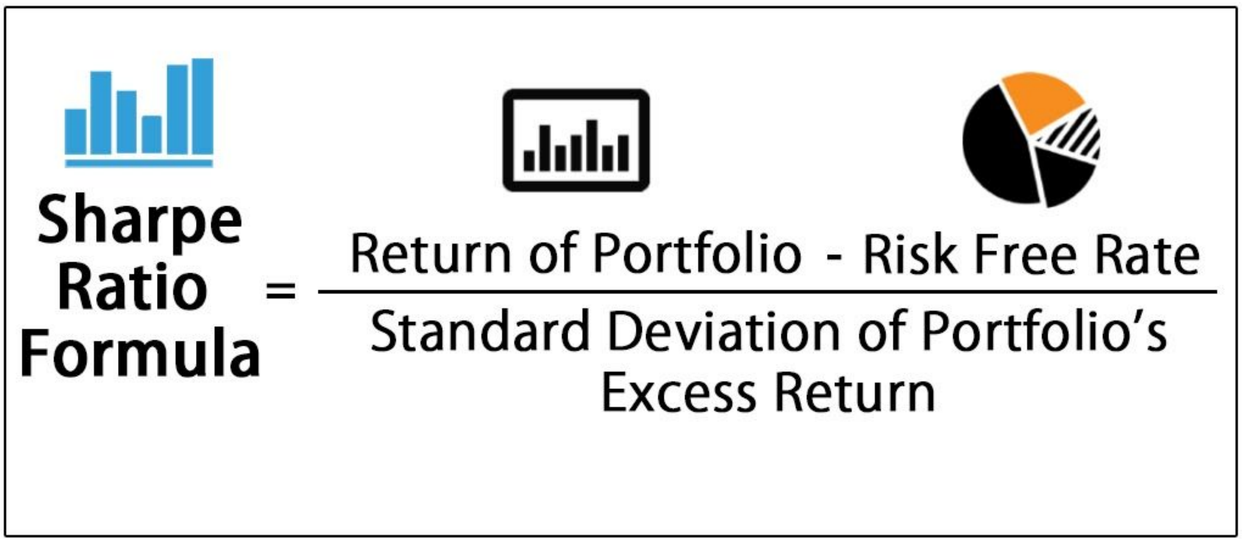

Figure 1: The Sharpe Ratio (Source: wallstreetmojo.com)

Almost every portfolio should seek to maximize this ratio. The Sharpe ratio can be considered the sample statistic of the Signal to Noise Ratio [6]. Theoretically, maximizing the Sharpe Ratio should get you between a high and max Signal to Noise ratio, given certain assumptions. We address these assumptions later in the report.

One thing to note about the Sharpe Ratio is that, in theory, it is primarily a measure of historical returns [6]; oftentimes when researchers and financial analysts look at future theoretical returns they seek to maximize some other function (such as utility). In practice, however, many investors evaluate performance of securities using Sharpe Ratio, and many firms tout their securities' high Sharpe Ratio. Since our training and testing data sets are historical, the Sharpe Ratio works well.

It should also be noted that oftentimes the Sharpe Ratio is used as a linear factor model for prediction [6], and other times modelled as a statistic from a sampling of returns and standard deviations [7] [8]. Our project does NOT use it for either of these purposes; instead, it is primarily used as a way to evaluate portfolios, and to determine/estimate the ideal weights for each portfolio.

Data Preprocessing

Daily data for S&P 500 stocks was acquired from two sources.

1. UPenn Wharton Research Data Services [9]
2. Alphavantage API [10]

Daily Source 1:

Daily metrics for 386 stocks from 2007-2019 such as Price, Return, Askhi, Bidlo, Volume, Shares Outstanding. Data for 2019 became available to the Finance Department in the first week of March 2020. Individual tables for each of the column metrics were created.

	date	symbol	name	bidlo	askhi	price	vol	ret	sout	open
0	2019-01-02	ORCL	ORACLE CORP	44.45	45.340	45.22	14320441.0	0.00155	3588919.0	44.48
1	2019-01-03	ORCL	ORACLE CORP	44.41	45.500	44.78	19868713.0	-0.00973	3588919.0	44.75
2	2019-01-04	ORCL	ORACLE CORP	45.25	46.950	46.71	20983953.0	0.0431	3588919.0	45.37
3	2019-01-07	ORCL	ORACLE CORP	46.47	48.105	47.45	17967907.0	0.015842	3588919.0	46.93
4	2019-01-08	ORCL	ORACLE CORP	47.20	48.360	47.88	16255688.0	0.009062	3588919.0	47.93

Figure 2 : UPenn Raw Data

date	2019-01-02	2019-01-03	2019-01-04	2019-01-07	2019-01-08	2019-01-09	2019-01-10	2019-01-11	2019-01-14	2019-01-15	...	2019-12-17	2019-12-18	20
ORCL_ORACLE CORP	0.001550	-0.009730	0.043100	0.015842	0.009062	-0.002089	0.000837	0.009829	-0.002278	0.008095	...	-0.021300	0.010409	-0.(
MSFT_MICROSOFT CORP	-0.004430	-0.036788	0.046509	0.001275	0.007251	0.014300	-0.006426	-0.007722	-0.007296	0.029005	...	-0.005401	-0.002069	0.(
TROW_T ROWE PRICE GROUP INC	-0.005199	-0.032992	0.035694	-0.001413	0.005117	0.003575	-0.003886	-0.022104	0.010416	0.017107	...	-0.001785	-0.005040	0.(
HON_HONEYWELL INTERNATIONAL INC	-0.002195	-0.013350	0.034981	0.005571	-0.002068	0.004664	0.014957	-0.002831	0.001820	0.001163	...	0.001701	-0.011658	0.(
ADM_ARCHER DANIELS MIDLAND CO	-0.000488	-0.004396	0.023056	0.005035	0.018607	0.008665	0.006965	-0.004842	-0.007183	0.014702	...	0.005002	-0.019476	0.(

Figure 3: Stock Returns

Spelling, numerical, and formatting errors were cleaned up. Once the data was cleaned it was uploaded to an Amazon cloud database. This data is the primary source for the exploration, and the source used for Experiment 1.

<u>Daily Source 2:</u>

Unfortunately the UPenn data could only provide daily stock data until 12/31/2019, so Alphavantage was also queried. Data from this source was more limited, and we could only acquire Price, Volume and Returns data.

	date	AAPL_open	AAPL_high	AAPL_low	AAPL_close	AAPL_volume	AAPL_return
0	2020-03-23	228.080	228.4997	213.68	215.25	31202470.0	-0.056252
1	2020-03-20	247.180	251.8300	228.00	229.24	100423346.0	-0.072579
2	2020-03-19	247.385	252.8400	242.61	244.78	67964255.0	-0.010530
3	2020-03-18	239.770	250.0000	237.12	246.67	75058406.0	0.028778
4	2020-03-17	247.510	257.6100	238.40	252.86	81013965.0	0.021615

Figure 4: Alphavantage Raw Data

Data from each of the 421 stocks was concatenated into a Price and Returns table. This data was used for Experiments 2 and 3.

	A	AAL	AAP	AAPL	ABC	ABMD	ABT	ACN	ADBE
2020-03-20	66.46	10.38	75.03	229.24	80.61	130.51	68.00	149.94	295.34
2020-03-19	65.76	10.29	85.97	244.78	86.00	140.51	74.50	158.50	307.51
2020-03-18	68.87	11.65	88.74	246.67	84.97	146.55	79.26	151.15	294.61
2020-03-17	70.44	15.58	100.25	252.86	89.50	145.25	79.49	157.71	311.81
2020-03-16	64.13	15.92	95.25	242.21	80.50	142.44	73.66	152.54	286.03
...	...	...	...	...	...	...	...	...	...
2007-01-08	33.97	57.93	35.14	85.47	46.45	14.11	50.08	37.21	40.45
2007-01-05	34.09	58.29	35.02	85.05	45.73	14.14	49.90	36.68	40.62
2007-01-04	34.41	58.84	35.81	85.66	46.29	14.35	49.90	37.20	40.82
2007-01-03	34.30	56.30	35.58	83.80	46.11	14.05	48.97	36.38	39.92
2006-12-29	34.85	53.85	35.56	84.84	44.96	14.10	48.71	36.93	41.12

Figure 5: Alphavantage Returns

The quarterly earnings reports were scraped from Stockpup [11]. The earnings reports contained data on market shares, balance sheets, income statements, and cash flow statements, as well as some derived metrics and ratios. We pulled out shares split adjusted, EPS basic, Dividend per share, price, ROE, P/B, P/E, and long term debt to equity ratio, and organized them into their own dataframes based on company and quarter end.

A fundamental issue when dealing with this data was the date for each quarter's report. Most companies had at least one irregularity in the timing of their historical reporting; a few released their reports at seemingly arbitrary dates relative to the quarter ends, and a few even had multiple reports in a single quarter. To deal with this, two weeks' leeway was given around the quarter end date, and every other report was assigned to the quarter that it ended in.

Excess returns are calculated by subtracting the total return from the risk free rate. For monthly returns, the 3 month daily treasury yield [12] was deannualized to one month.

Data Exploration

In order to use the Sharpe ratio effectively, certain assumptions about the returns data need to be verified. We'll use statistical inference on returns of stocks judged to either violate or not violate these assumptions. With large samples (n=100), we can use the z test for proportion on a binomial distribution [13]. In doing so, we can judge whether the true percentage of 'bad' stocks is under a benchmark threshold.

We want to check for autocorrelation, heteroskedasticity, and normality of returns. If it can be inferred that the proportion of stocks that violate these assumptions is <10%, we will treat the assumptions as valid for the purposes of our analysis. They will validate the usage of the Sharpe ratio, and guarantee a certain level of efficacy [6].

For each of the tests, the null hypothesis assumes that of the 386 stocks in our sample, 10% or greater violate the assumption. The alternative states that the true proportion is less than 10%. We test at the 95% significance level. Each test will have its own sample of 100 stocks.

We'll judge relevant visual plots for our hypothesis tests. The reason we prefer this over well-defined tests is that we're more concerned with significant departures from the assumptions. Significance tests are generally too sensitive, and will throw out cases that don't hinder our analysis.

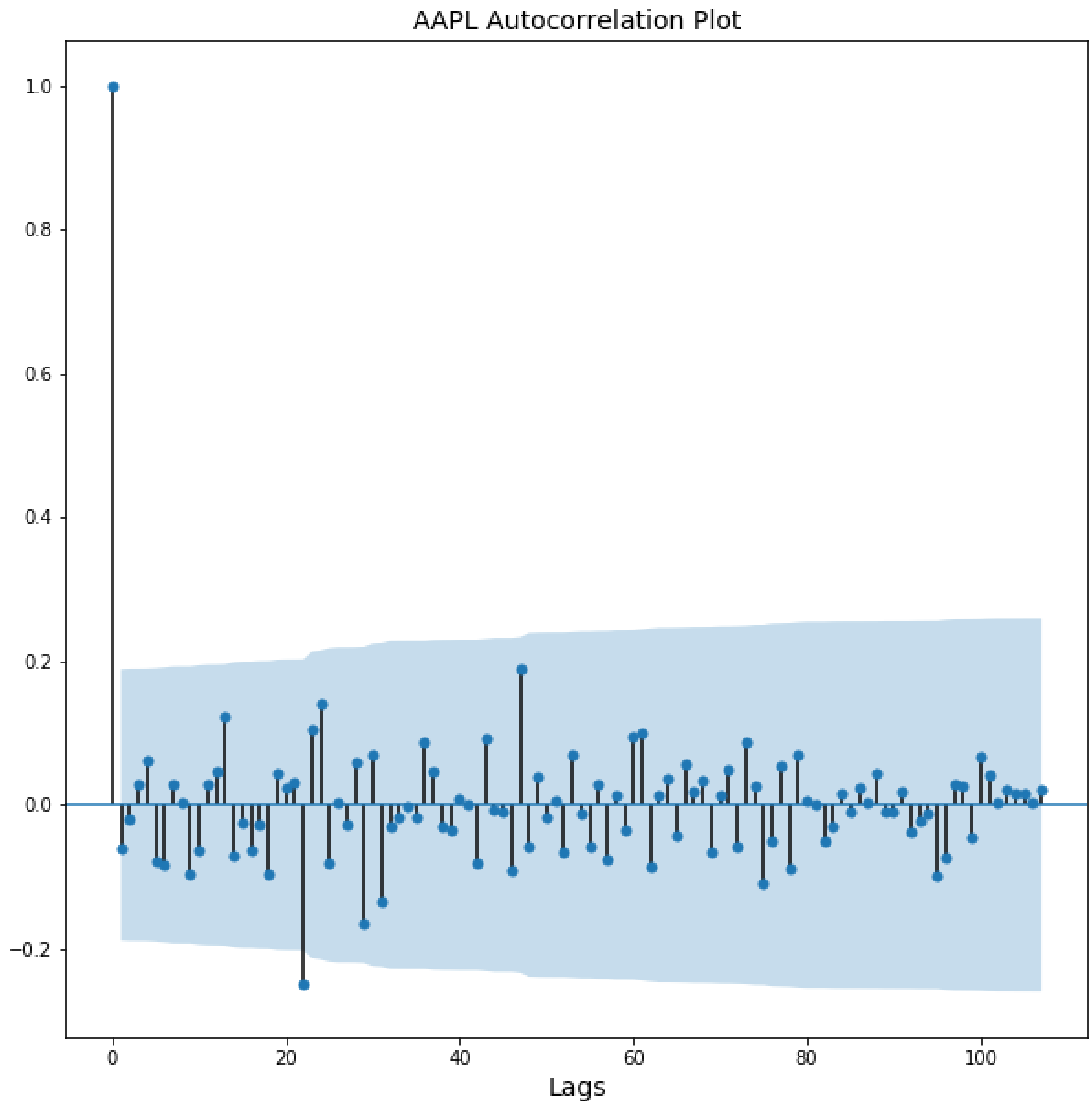

Figure 6: Example of Autocorrelation Plots on Returns

"Autocorrelation"

Figure 6 shows an example of the autocorrelation plots. The light blue band indicates the 95% confidence interval for not significant correlation. We judged a stock to have significant correlation if more than 4 lags showed significant correlation. Only one stock out of 100 showed significant correlation, with $z = -9.05$ and $p \cong 0$, so we reject the null and may assume minimal autocorrelation of returns.

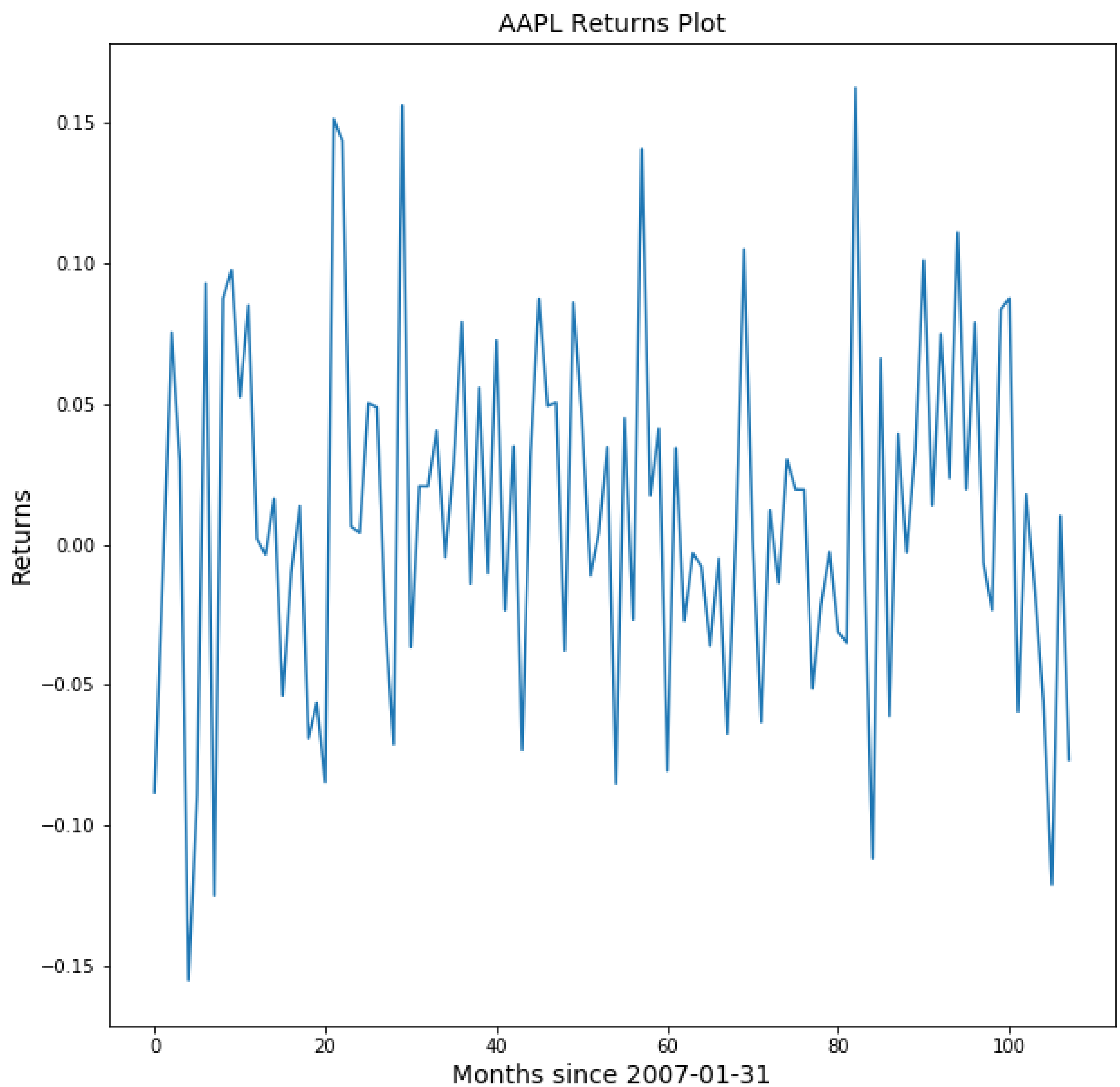

Figure 7: Sample of Heteroskedasticity Plots on Returns

"Heteroskedasticity"

Figure 7 shows a sample of the time series plots used to judge heteroskedasticity. We judged a stock to have heteroskedastic returns if its plot had a cone or hourglass shape. Three stocks were judged as heteroskedactic, with z = -4.1 and p = 0.00002, so we reject the null hypothesis and assume minimal heteroskedastic returns.

"Normality of Returns"

Normality of returns is the hardest assumption to meet, and the least relevant. It is mostly important for determining the distribution of Sharpe ratios taken from a single, well-defined portfolio, whereas we're using the Sharpe ratio as a way to weigh and judge multiple portfolios. In addition, a fat-tailed, mean-displaced t-distribution is generally a more accurate gauge for returns distributions.

For our purposes, we only care if returns show a significant departure from normality. By lowering the p-value, we "loosen the threshold", and can examine the extent of departure. Using the Shapiro-Wilk test, for $p = 0.001$, 48 stocks rejected the normality hypothesis. Gradually lowering the value also gradually lowered the number of rejections.

It's difficult to determine the conclusion that can be drawn from this. The Shapiro-Wilk test can oftentimes reject trivial departures, especially when the sample size is large ($n = 108$) [18]. But even so, the results aren't promising, and some stocks seem to wildly diverge from a Gaussian distribution. Future inquiry is recommended to determine the viability of the assumption, and potential effects on the Sharpe Ratio.

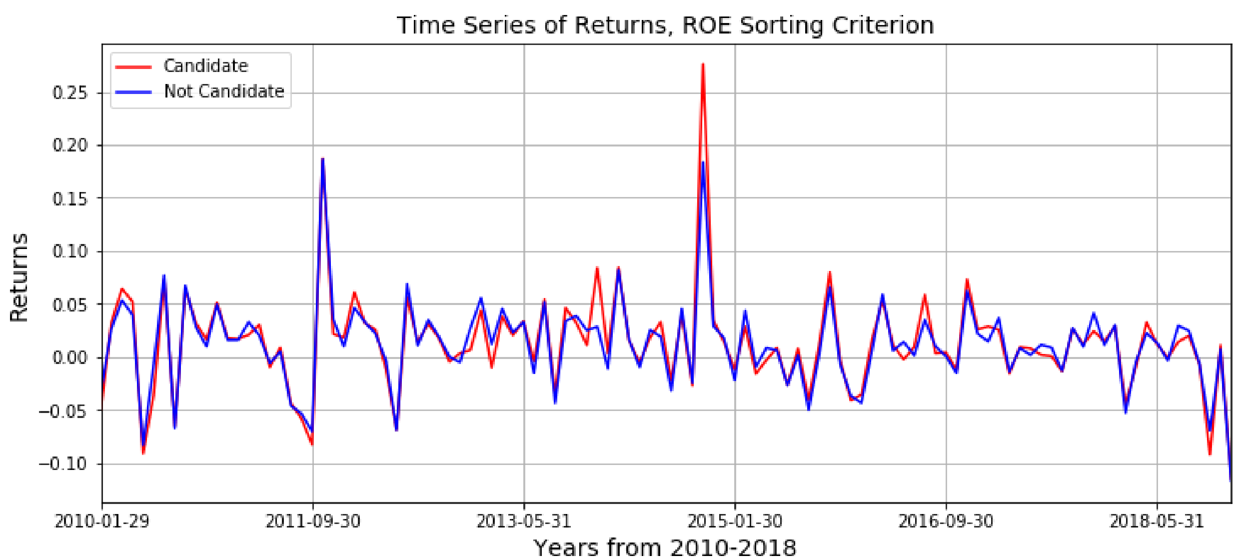

Figure 8: Comparison of Control and Alternative Returns for Recession Resistant Stocks (ROE based criteria)

We also explore the recession resistant returns and compare them to the control stocks. As can be seen in Figure 8, the candidate returns generally mirror the control returns, no matter the criteria used to sort the stocks. One thing to note is that on the highest month of returns in our data (the middle of 2014), only the ROE sorting criterion leads to that enormous spike in candidate returns.

Figure 9 is an example of the “recession resistant” candidate pool selection process, when looking at quarterly metrics. The selection process is equivalent to taking the L1 norm of each stock’s averages and picking the 20% smallest values. As can be seen, there are quite a few stocks that kept similar, or even better performance during the recession compared to the total time period. The average is computed from the unweighted quarterly data for each respective time period. The Average 07-18 EPS value includes the recession data, which occurred between 2008 and 2009.

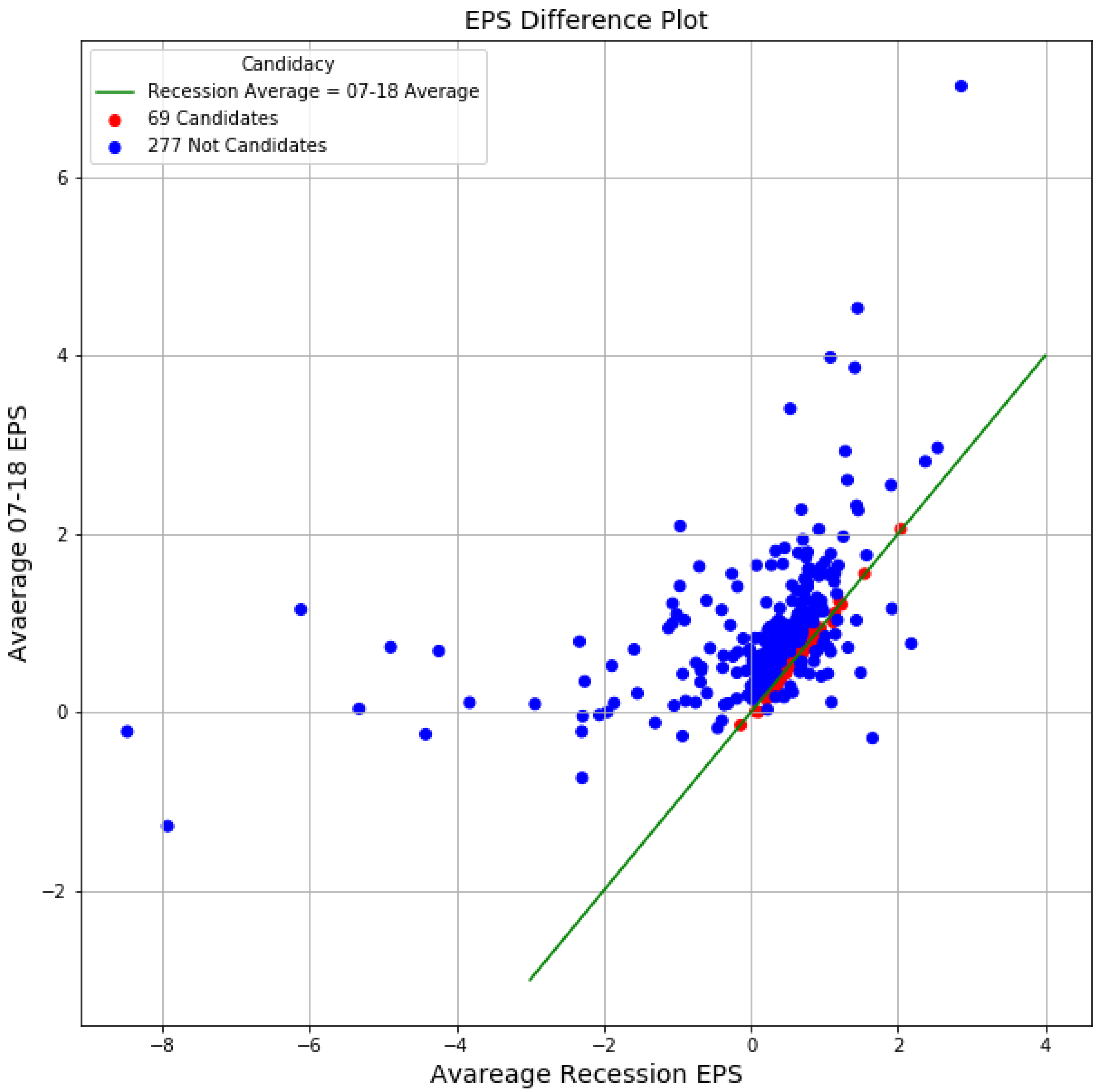

Figure 9: Average EPS Difference Plot

In order to focus on building a resistant portfolio, finding measures related to stock returns was also important. With respect to the daily data, this meant trying to find a leading indicator for stock returns, negative returns in particular.

For each of the daily metrics time series, histograms and pair plots were created for each stock. A visual aid was needed to get some intuition of the behavior of each stock from 2007-2018. We decided to create a portfolio of visualizations for six major metrics, namely: price, returns, spread, and volume/sharesout time series; scaled price (price/max-min) histogram, and a Price/Askhi ~ Returns pairplot (Figures 10, 11). 386 of these portfolios were made (from UPenn data). The rest can be found in the Appendix.

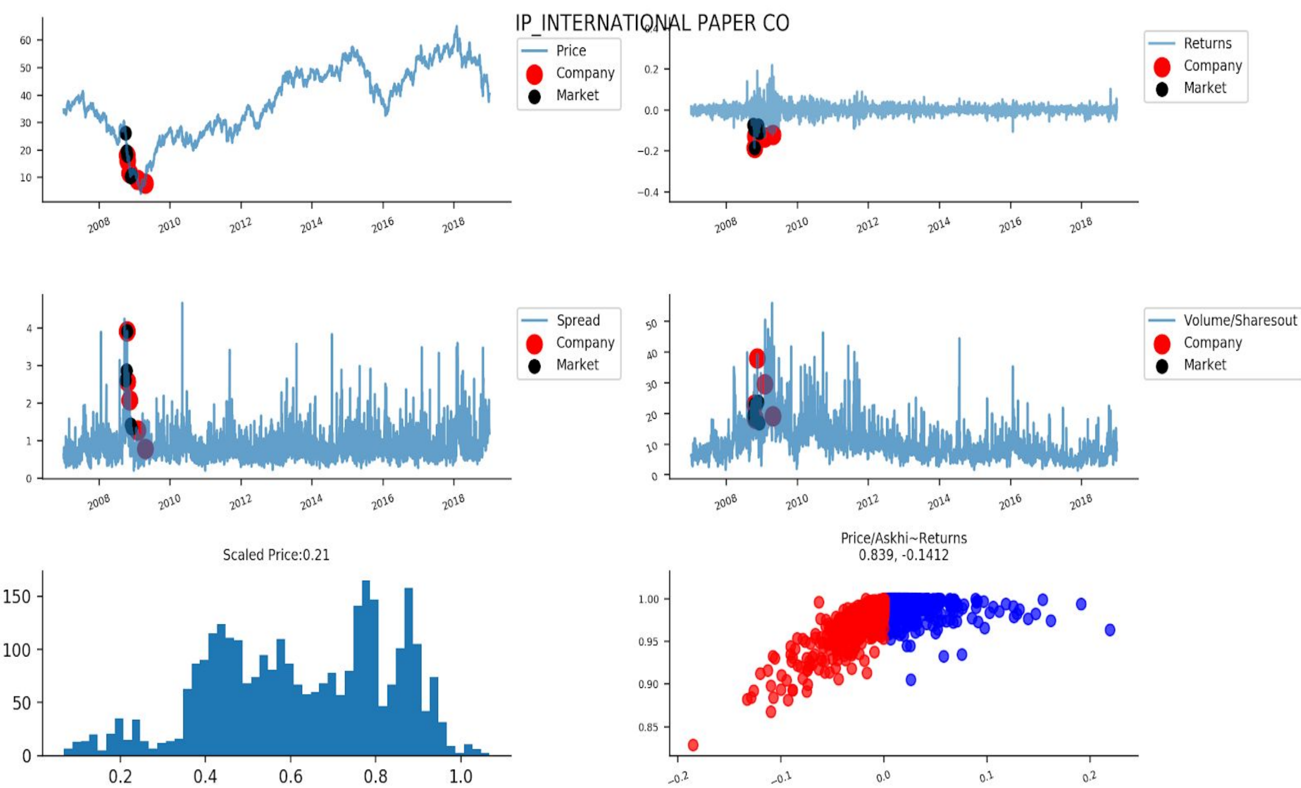

Figure 10: IP stock visualizations

For each of the timeseries we wanted to examine how the worst days (return-wise) of each stock compared to those of the market [14] [15] [16]. Red dots indicate the 5 worst return days for each of the companies while black dots indicate the 5 worst return dates for each of the companies.

We noted that the 5 worst return dates for each company did not necessarily coincide with the worst return dates of the market. Figures 10 and 11 demonstrate the differences between stocks that coincided with the overall market during the market's worst return dates (IP) and stocks that did not (NFLX).

The correlation plot on the bottom right of each portfolio relates the Price/Askhi to the stock returns. Notice that for negative returns days there was a higher correlation to the Price/Askhi metric than for positive return days. This prompted us to explore further and determine whether the Price/Askhi metric was thought to be useful as a leading indicator of a negative return date.

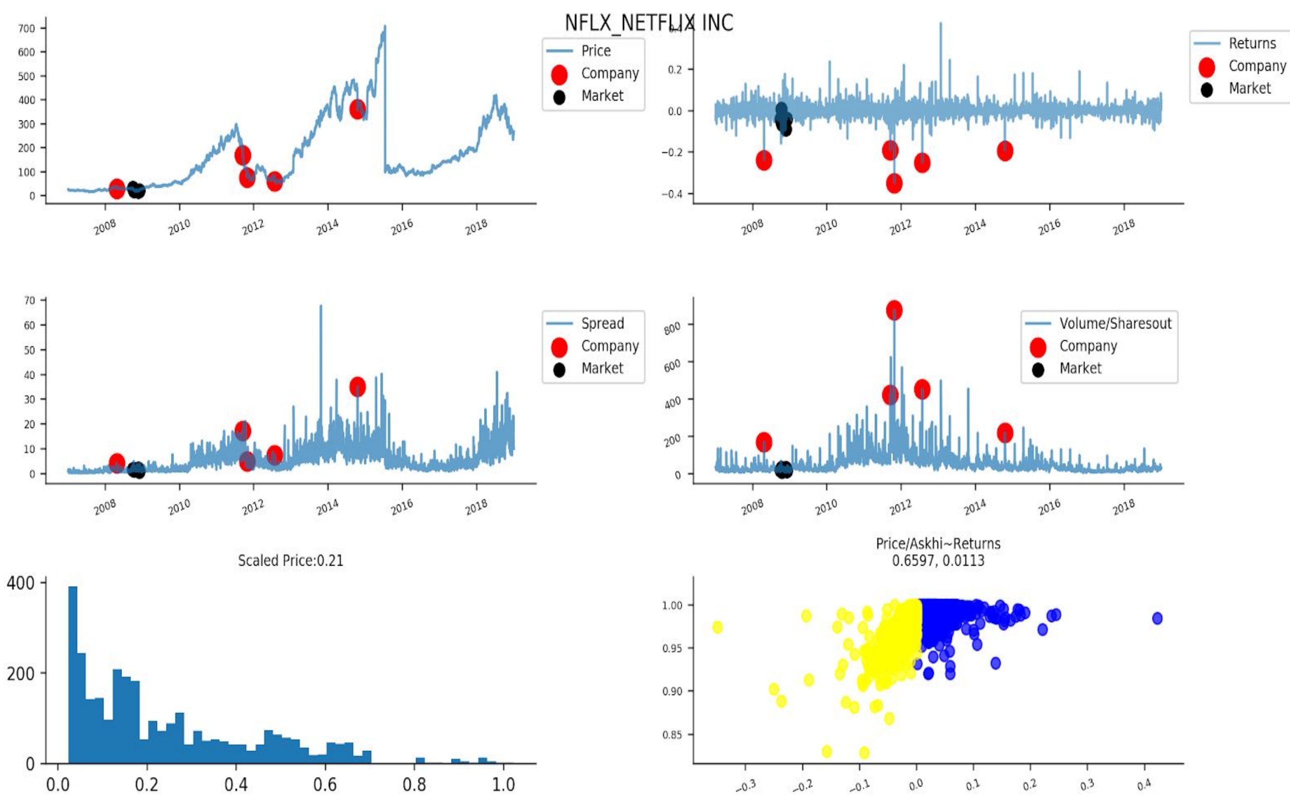

Figure 11: NFLX stock visualizations

In Figure 12 we have the distribution of the correlations between Price/Askhi and Returns 1 day later. The return threshold is the absolute value of the lower bound of the returns that were examined for correlation. For example, when the return threshold was set to 0.03, the histograms showcase the distribution of the returns with Price/Askhi when the returns were either less than -0.03 or greater than 0.03. As the return threshold grows, the distributions get noisier.

Figure 13 plots the distribution of the Correlations between Price/Askhi and Returns when the returns came 5 days later. You can see the similarity in distributions between 1 day and 5 day return thresholds for each of the corresponding return thresholds.

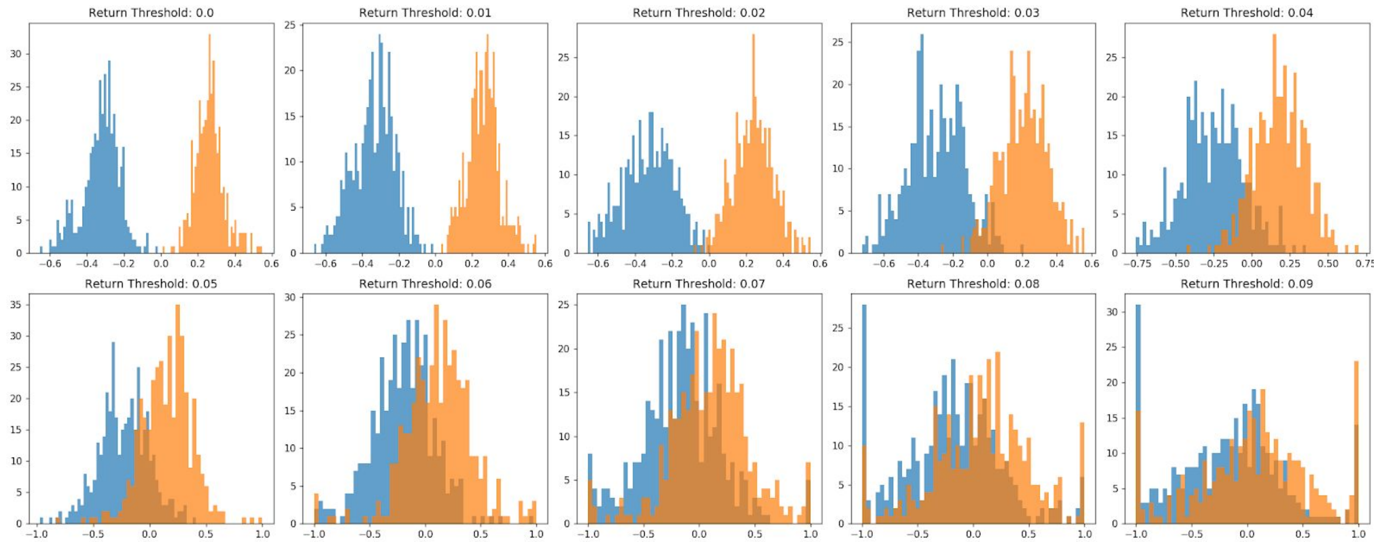

Figure 12: 1 Day Price/Askhi ~ Returns

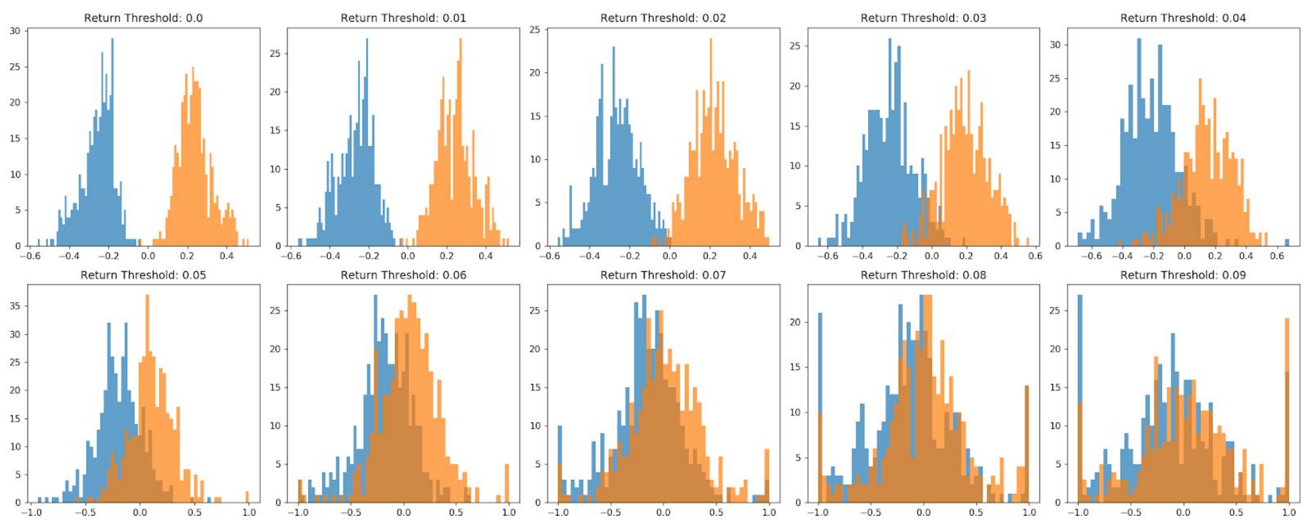

Figure 13: 5 Day Price/Askhi ~ Returns

Algorithms and Techniques, Implementation and Refinement

“Sharpe Ratio Algorithm”

The major algorithm used in our project is the one needed to find the maximum Sharpe ratio. Given a set of stocks in a portfolio, the algorithm takes in the expected excess return for each stock and calculates the weights for each stock that give the theoretically highest possible Sharpe Ratio. A high Sharpe ratio implies that a portfolio is balancing diversity with potential returns, a must for any reasonable investor.

The algorithm calculates the covariance matrix from a set of returns, and its inverse. Then the inverse can be matrix multiplied with the vector of excess returns. The resulting vector (commonly called the zeta) will be a scalar multiple of the vector of ideal weights, so to find these weights it is sufficient to normalize the zeta vector [17].

Care has to be taken that there is enough returns data for the chosen portfolio size. Even with 9 years of monthly returns data (108 observations), a 30 stock portfolio was not guaranteed to converge to a positive Sharpe ratio (a theoretical impossibility given at least one positive expected return). This is due to the covariance matrix, which needs many more observations per stock to be invertible.

What's important to note about this algorithm is that it finds the ideal weights by looking at historical data. For this project, we'll be training the weights on historical data and applying the weights to unknown testing data. This has many similarities to the training/testing split when training a prediction model. However, it needs to be noted that there is no way to avoid "overfitting" the weights on the training data; the algorithm only allows for ideal weight calculation, given the available data. This limitation can only be mitigated with more data.

This algorithm assumes that unlimited short-sales are available to the investor. A short-sale is when the investor borrows a stock to sell now, then buys back the stock at a later date to return to the lender. In reality, short-sales are difficult to pull off, with high margins that can affect the rate of return; we ignore these limitations when conducting our analyses, but note that they must be considered before any real world portfolios are created using the techniques mentioned here.

"Bootstrap Random Sampling"

To test whether the treatment of "recession resistant" stocks is significant, we'll randomly sample portfolios from a sampling of stocks: the Control 1 stocks consisting of "not recession resistant" stocks; Control 2 stocks consisting of all the stocks we have complete data for (346 stocks); and the Alternative stocks consisting of "recession resistant" stocks. The number of stocks in each subsample will vary based on the metric used and how much complete data we have on it, but the percent distribution will be 80/100/20 for A, B, and alternative, respectively.

"Statistical Inference"

To test the hypothesis that alternative portfolios will outperform the control, we'll implement a two sample statistical analysis on the difference of means. The experiment will be set up as a two comparison problem, comparing Control 1 with alternative, and the comparing Control 2 with alternative. Since this project does not seek a confidence level for the general collection of tests, there is no need to implement a multiple-comparison controlling procedure.

Benchmark

All of our hypotheses will have two benchmarks: Control 1 and Control 2. The first control has all the non-candidate stocks for our hypotheses, i.e. stocks not shown to be recession resistant. It is used as part of the standard two sample inference test, to test treatment versus no treatment. This is in line with traditional statistical testing.

The second control contains all the stocks in our dataset. It is useful as a representation of the S&P 500 Index, and the broader market. Since there is minimal bias in the sample of stocks, you can expect the sampled portfolios to mimic the performance of a complete market sampled portfolio.

For hypotheses 1-2, we test whether the alternative portfolios outperform the controls, using a one-sided two sample inference z-test on n = 1000 sampled portfolios per subsample, with a 95% confidence level. For hypotheses 3-6, we use a two sample inference z-test on n = 10000 sampled portfolios per subsample, with a 95% confidence level.

Experiment 1: Hypotheses 1 and 2

Quarterly metrics for the 386 stocks were used to split the sample between Control and Alternative

Control Portfolios 1: Random Sampling of 25 stocks from the S&P 500 that are not considered to be recession resistant

Control Portfolios 2: Random Sampling of 25 stocks from the S&P 500 stocks that have trading and earnings data from 1/1/2007 - 12/31/2018

Alternative Portfolio: Random Sampling of 25 stocks chosen from the top 20% pool of recession resistant stocks, based on the treatment: EPS, ROE, or PEG.

<u>Hypothesis 1:</u> The average Alternative Portfolio will outperform the average Control portfolios on 2019 data

<u>Hypothesis 2:</u> The average Alternative Portfolio will outperform the average Control

portfolios on The Great Recession data (Q3 2008-Q2 2009)

Methodology and Model Validation

For the three subsamples of stocks (Control 1, Control 2, Alternative), 1000 random samples of portfolios with 25 random stocks will be generated. Each portfolio will then undergo the Sharpe ratio algorithm to find the maximal Sharpe ratio weights on the 2010-2018 monthly returns data, and then those weights will be applied to two different sets of returns (corresponding to hypotheses 1 and 2): 2019 returns and Great Recession returns (Q3 2008 - Q2 2009).

Those generated portfolio/weight combinations will produce returns and risks for each subsample, from which we'll compute the sample of Sharpe ratios. We'll use the sample to make inferences on the mean Sharpe ratio of each subsample's possible portfolios. For each hypothesis, a one-sided, two-sample z test was run, with alpha = 0.05. Z-scores, p-values, and 95% confidence intervals are reported. Two separate tests were run: on the difference between Control 1 and Alternative, and on the difference between Control 2 and Alternative. See Appendix for related histograms and QQ-plots.

This type of inference does not require a normality assumption about the individual Sharpe ratios. The difference of the means, along with their variances, form the random variable

$$Z = \frac{\overline{A} - \overline{C} - (\mu_A - \mu_C)}{\sqrt{\sigma_A^2/n_A + \sigma_C^2/n_C}}$$

where $\overline{A}$ and $\overline{C}$ are the sample means for control and alternative portfolio populations. Z is a r.v. with mean = 0 and variance = 1, and if the sample sizes are large, Z is approximately normal by the Central Limit Theorem [13]. For large samples, the population variance can be replaced by the sample variances.

The samples were engineered to be drawn using a p-RNG technique in Python; as such, we can assume that the random variables are independent and identically distributed.

Results: Hypothesis 1 (2019 returns)

For each of the following treatments, the null and alternative hypotheses are the same.

$$H_0^1 : \mu_{alt} \leq \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} \leq \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Treatment: EPS

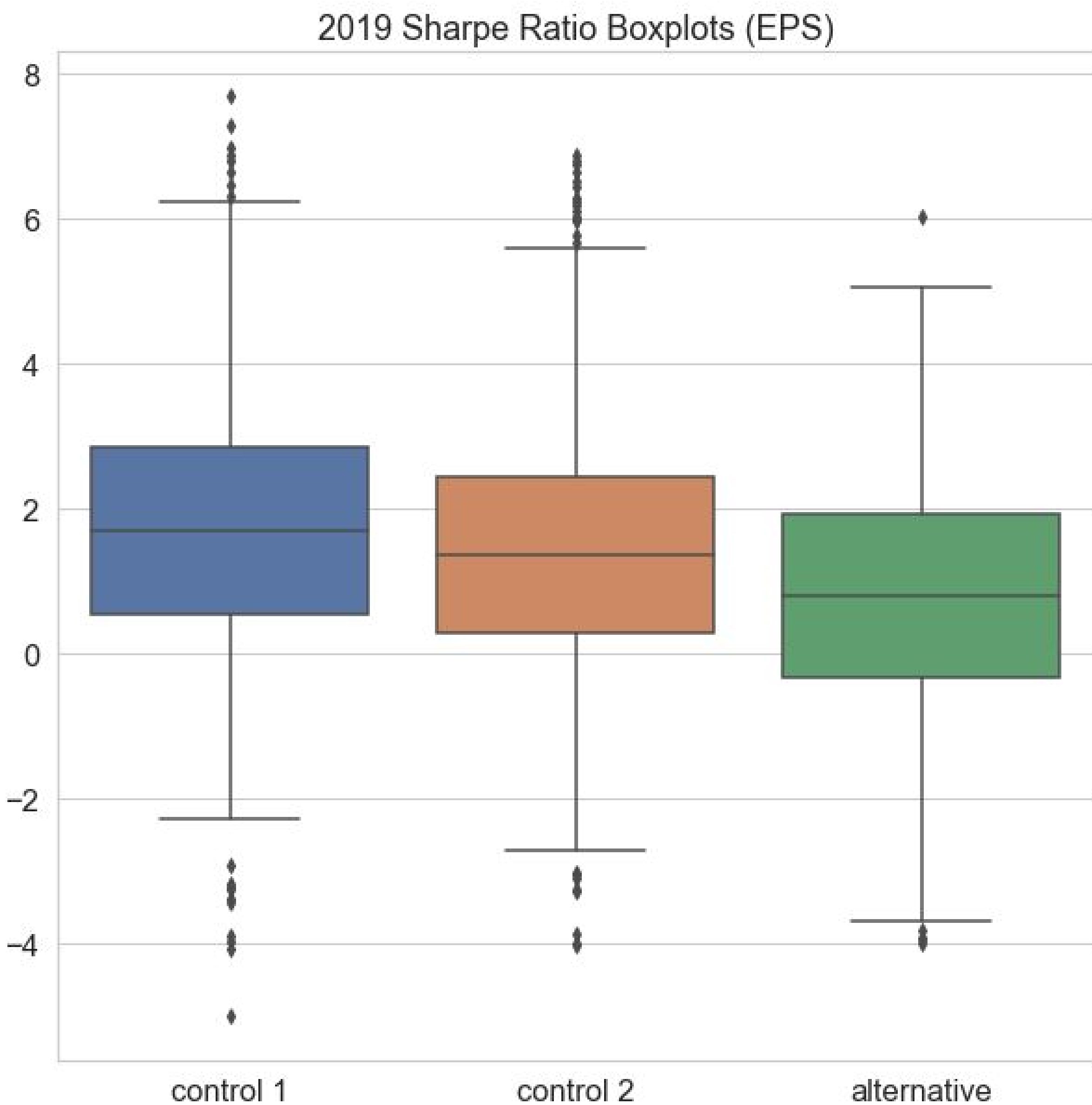

Figure 14: Hypothesis 1, Sharpe Ratio Boxplots, EPS Treatment

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	1.697805	1.777549	0.056211	(-4.982864, 7.687493)
Control 2	1.388431	1.740022	0.055024	(-4.011585, 6.888352)
Alternative	0.816962	1.654256	0.052312	(-3.984463, 6.025371)

Table 1: Hypothesis 1, Sharpe Ratio Statistics, EPS Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = -11.47$, $p \simeq 1$.
95% Confidence Interval: (-1.009714, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-1.009714, infinity).

Alternative~Control 2 Test:

2 Sample Z-test produced $z = -7.52$, $p \simeq 1$.
95% Confidence Interval: (-0.696713, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-0.696713, infinity).

Treatment: ROE

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	1.397289	1.793121	0.056703	(-4.955471, 8.921529)
Control 2	1.453059	1.831128	0.057905	(-4.246647, 8.396932)
Alternative	2.047810	1.392352	1.392352	(-1.959722, 7.051484)

Table 2: Hypothesis 1, Sharpe Ratio Statistics, ROE Treatment

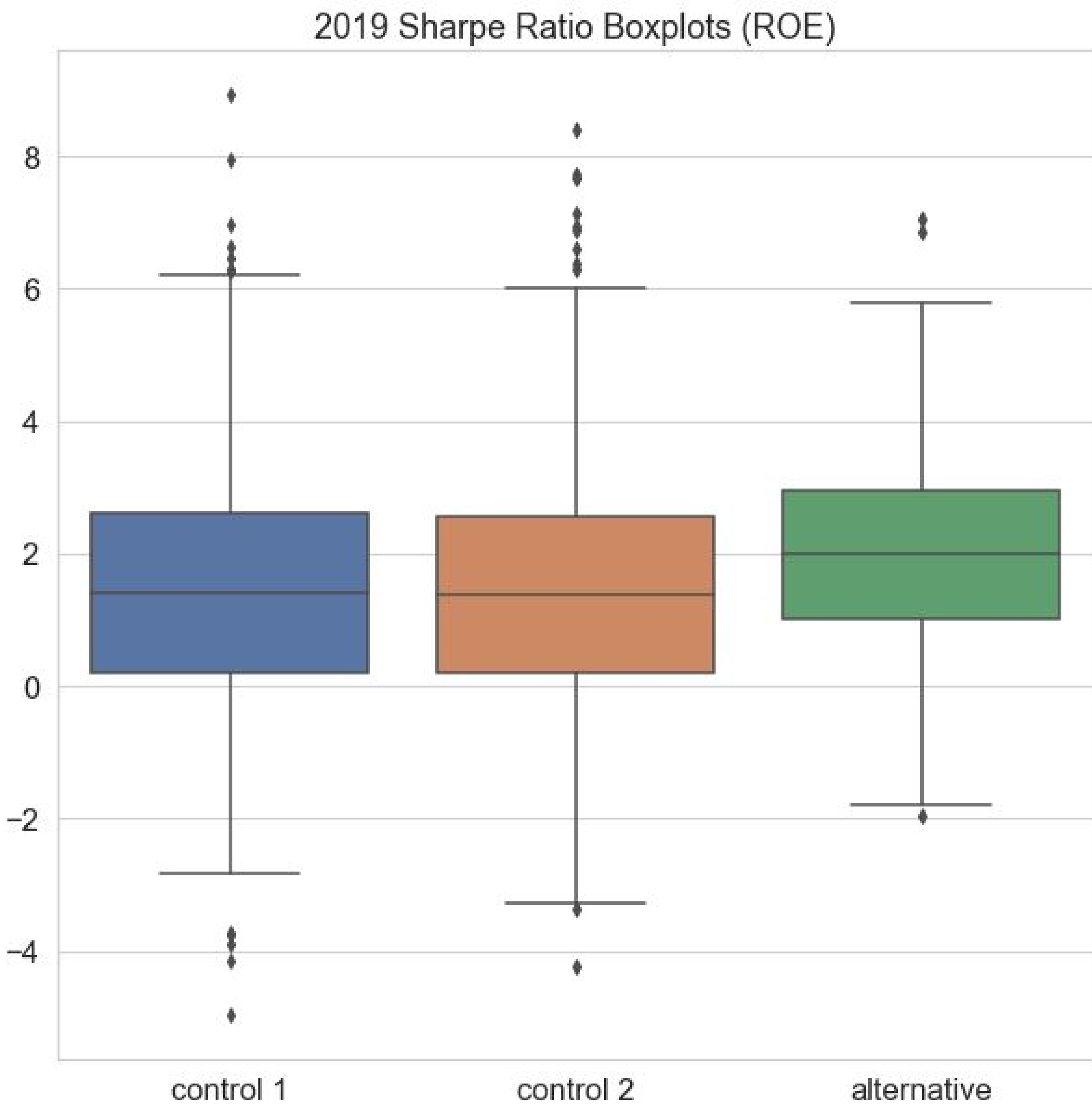

Figure 15: Hypothesis 1, Sharpe Ratio Boxplots, ROE Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = 9.06$, $p \simeq 0$.
95% Confidence Interval: (0.532210, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (0.532210, infinity).

Alternative~Control 2 Test:

2 Sample Z-test produced $z = 8.18$, $p \simeq 0$.
95% Confidence Interval: (0.474724, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (0.474724, infinity).

Treatment: PEG

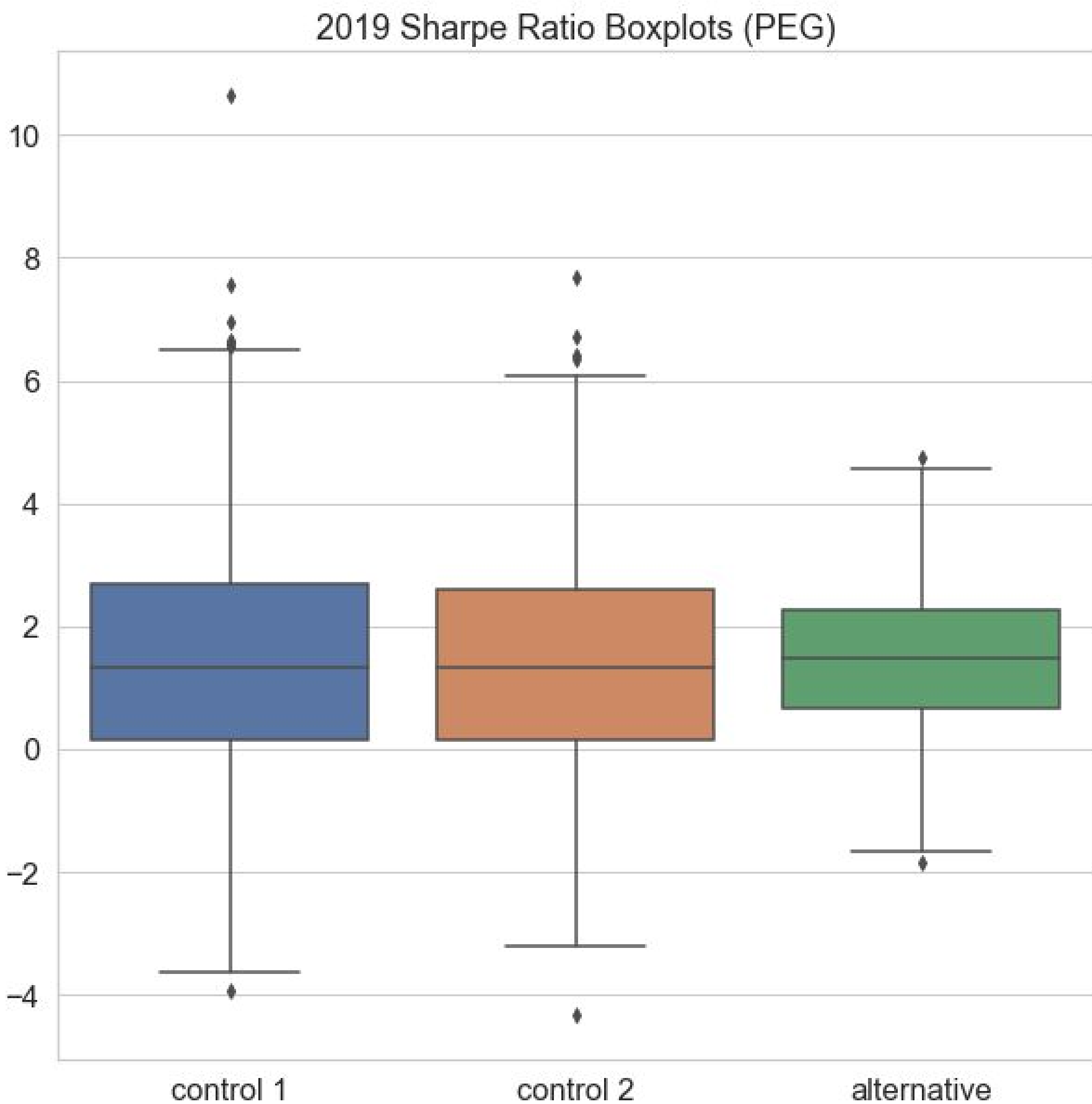

Figure 16: Hypothesis 1, Sharpe Ratio Boxplots, PEG Treatment

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	1.437675	1.924398	0.060855	(-3.920387, 10.628283)
Control 2	1.397051	1.819409	0.057348	(-4.311681, 7.674974)
Alternative	1.461365	1.148058	0.036305	(-1.848123, 4.742309)

Table 3: Hypothesis 1, Sharpe Ratio Statistics, PEG Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = 0.33$, $p = 0.37$.
95% Confidence Interval: (-0.092774, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-0.092774, infinity).

Alternative~Control 2 Test:

2 Sample Z-test produced $z = 0.94$, $p = 0.17$.
95% Confidence Interval: (-0.046517, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-0.046517, infinity).

Results: Hypothesis 2 (Great Recession returns)

Similar to hypothesis 1, the null and alternative hypotheses are the same for each treatment.

$$H_0^1 : \mu_{alt} \leq \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} \leq \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Treatment: EPS

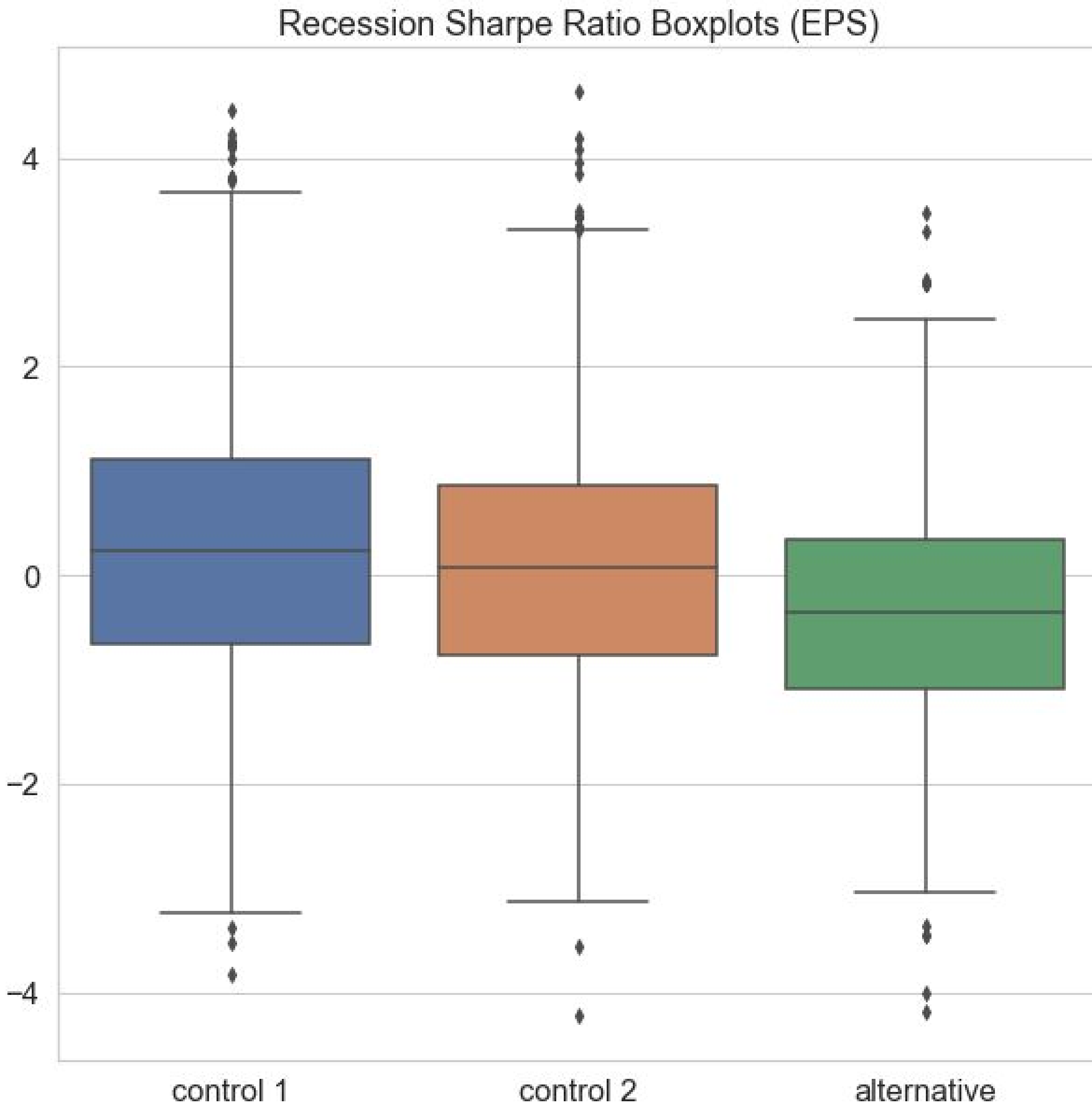

Figure 17: Hypothesis 2, Sharpe Ratio Boxplots, EPS Treatment

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	0.255074	1.305625	0.041287	(-3.820197, 4.451172)
Control 2	0.102449	1.262305	0.039918	(-4.212520, 4.630670)
Alternative	-0.367781	1.086242	0.034350	(-4.190690, 3.472917)

Table 4: Hypothesis 2, Sharpe Ratio Statistics, EPS Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = -11.59$, $p \simeq 1$.
95% Confidence Interval: (-0.710668, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-0.710668, infinity).

Alternative~Control 2 Test:

2 Sample Z-test produced $z = -8.93$, $p \simeq 1$.
95% Confidence Interval: (-0.554440, infinity)

Conclusion: Failure to reject null hypothesis. There is not sufficient evidence to conclude the alternative is higher than the control. We are 95% confident the true difference of means lies within (-0.554440, infinity).

Treatment: ROE

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	-0.307397	1.205478	0.038121	(-4.033574, 3.735260)
Control 2	0.129449	1.170542	0.037016	(-3.276263, 5.031241)
Alternative	1.305175	0.967089	0.030582	(-2.522042, 4.645807)

Table 5: Hypothesis 2, Sharpe Ratio Statistics, ROE Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = 33.00$, $p \simeq 0$.
95% Confidence Interval: (1.529901, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (1.529901, infinity).

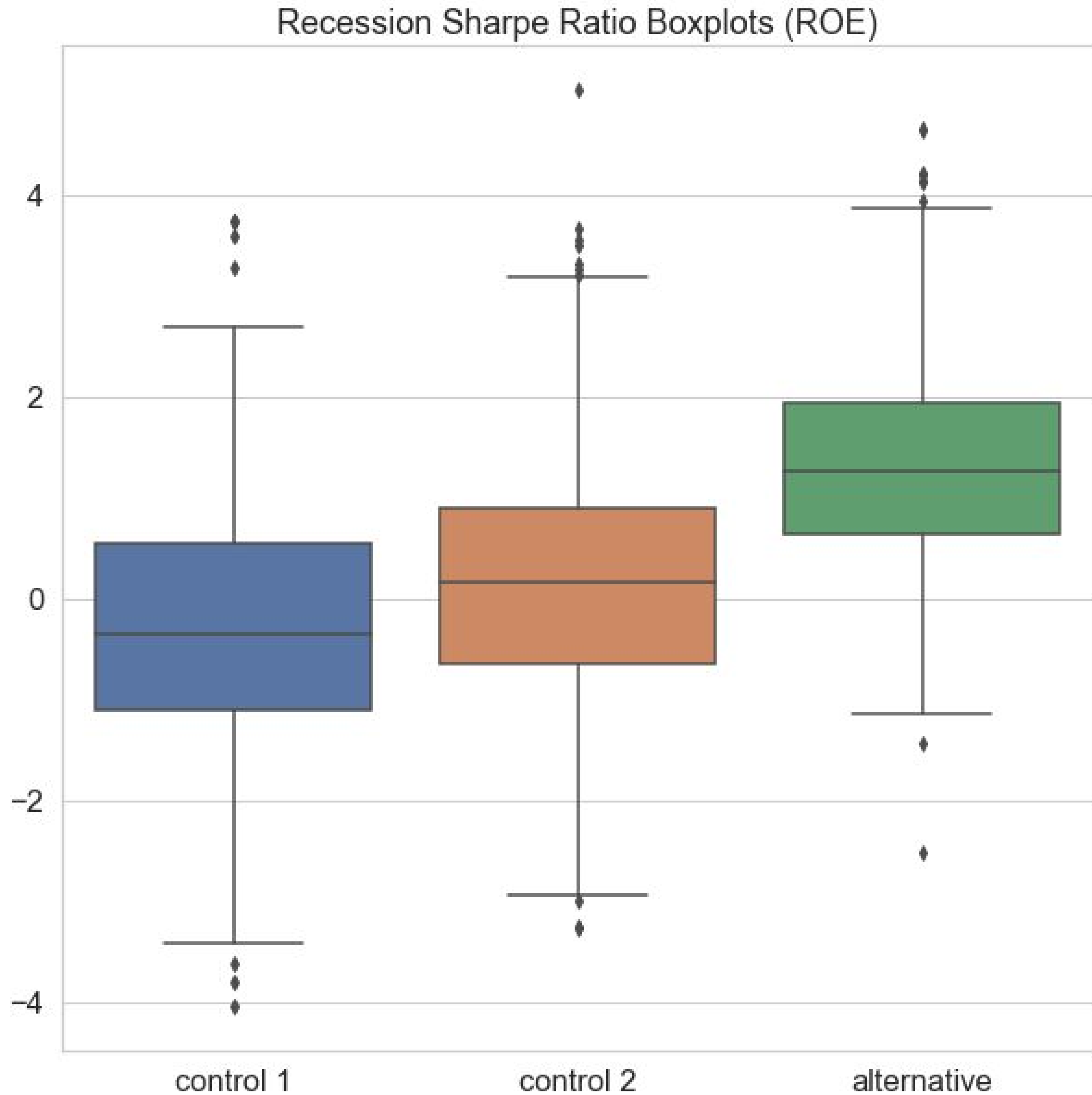

Figure 18: Hypothesis 2, Sharpe Ratio Boxplots, ROE Treatment

Alternative~Control 2 Test:

2 Sample Z-test produced $z = 24.49$, $p \simeq 0$.
95% Confidence Interval: (1.094825, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (1.094825, infinity).

Treatment: PEG

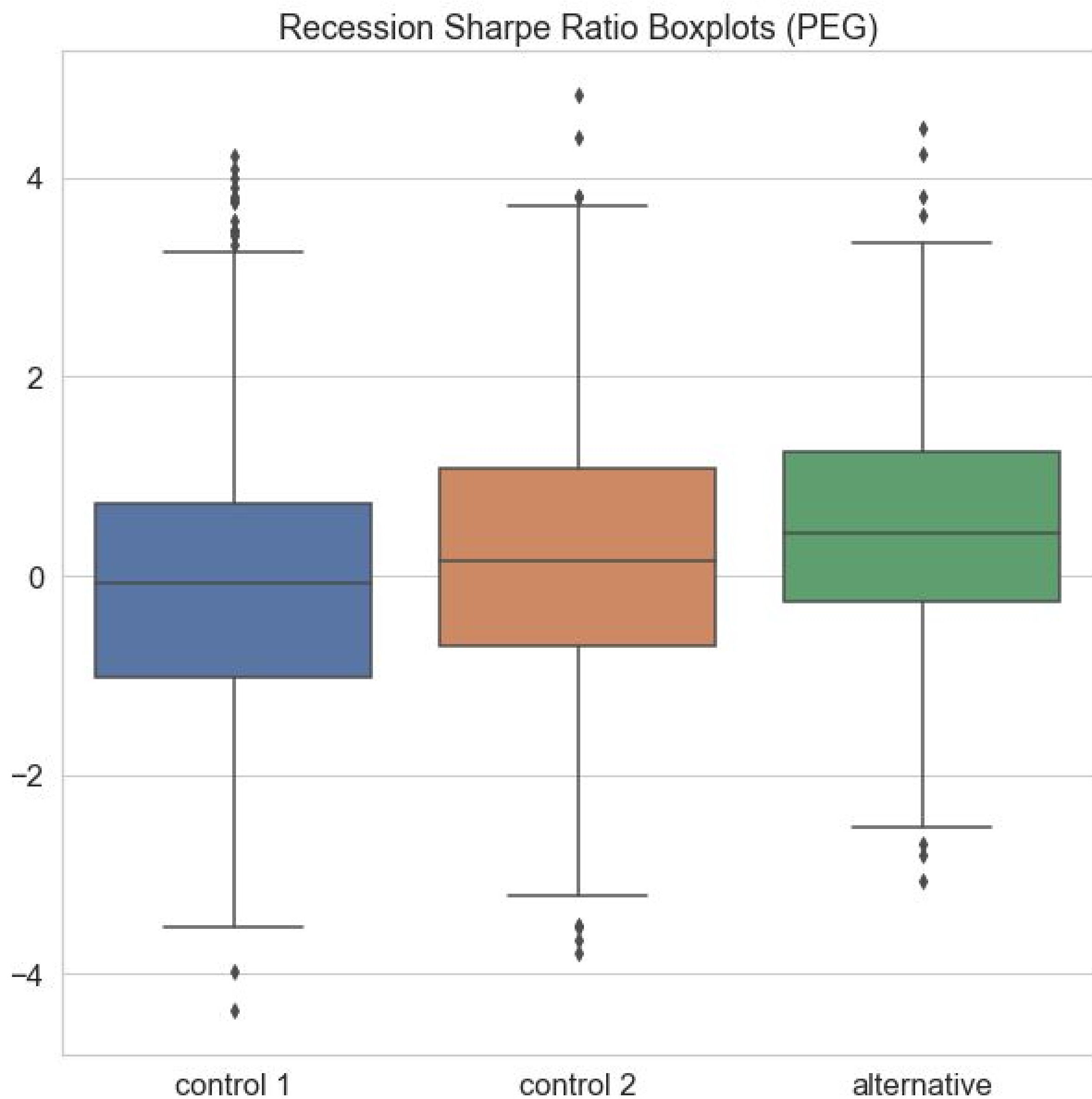

Figure 19: Hypothesis 2, Sharpe Ratio Boxplots, PEG Treatment

Pool	Mean	Standard Dev.	Standard Error of Mean	Range
Control 1	-0.072585	1.332292	0.042131	(-4.355690, 4.210346)
Control 2	0.189825	1.296360	0.040995	(-3.794867, 4.822705)
Alternative	0.486088	1.153768	0.036485	(-3.058994, 4.495340)

Table 6: Hypothesis 2, Sharpe Ratio Statistics, PEG Treatment

Alternative~Control 1 Test:

2 Sample Z-test produced $z = 10.02$, $p \simeq 0$.
95% Confidence Interval: (0.467000, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (0.467000, infinity).

Alternative~Control 2 Test:

2 Sample Z-test produced $z = 5.40$, $p \simeq 0$.
95% Confidence Interval: (0.206303, infinity)

Conclusion: Reject null hypothesis. There is sufficient evidence to conclude the difference between alternative and control is greater than 0. We are 95% confident the true difference of means lies within (0.206303, infinity).

Experiment 2: Hypotheses 3-4

Hypotheses 3 and 4 check whether the Alternative Portfolio will yield a positive return in 2019, and will outperform Control Portfolio 1 and Control Portfolio 2 on each of the 5 worst S&P 500 return dates in 2019 and each of the 5 worst return dates of 2020.

Alternative Criteria:

- None of the 5 worst trading days (by daily return) from 2007-2018 coincided with *any* of the 5 worst trading days of the S&P 500 from 2007-2018
- Pool size: 32 stocks

Control 1 Criteria:

- The stock did not meet the Alternative criteria but still traded in the S&P 500
- Pool size: 421-32 = 379 stocks

Control 2 Criteria:

- The stock traded in the S&P 500
- Pool size: 421 stocks

Hypothesis 3: The Alternative Portfolio will yield a positive return in 2019

$$H_0 : \mu_{alt} = 0 \qquad H_A : \mu_{alt} > 0$$

Hypothesis 4a: The Alternative Portfolio will outperform Control 1 and The Alternative Portfolio will outperform Control 2 in each of the 5 worst S&P 500 trading days of 2019.

$$For\ each\ Date\ in\ the\ 5\ Worst\ Dates\ of\ 2019$$

$$H_0^1 : \mu_{alt} = \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} = \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Hypothesis 4b: Is the exact hypothesis as 4a, except for 2020

$$For\ each\ Date\ in\ the\ 5\ Worst\ Dates\ of\ 2020$$

$$H_0^1 : \mu_{alt} = \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} = \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Methodology

10,000 bootstraps were generated for the testing of each hypothesis. Portfolios of size 25 were generated for the Alternative, Control 1, and Control 2 stock subsamples. Then, one sided z-tests were performed on the portfolio returns to test the performances of the null hypotheses. An alpha of 0.05 was used to test each hypothesis.

Results: Hypothesis 3

X Bar Alt	Lower	Upper	P-Value	Hypothesis Result
0.259152	0.123751	0.394553	0	Reject Null

Table 7: Hypothesis 3 Results

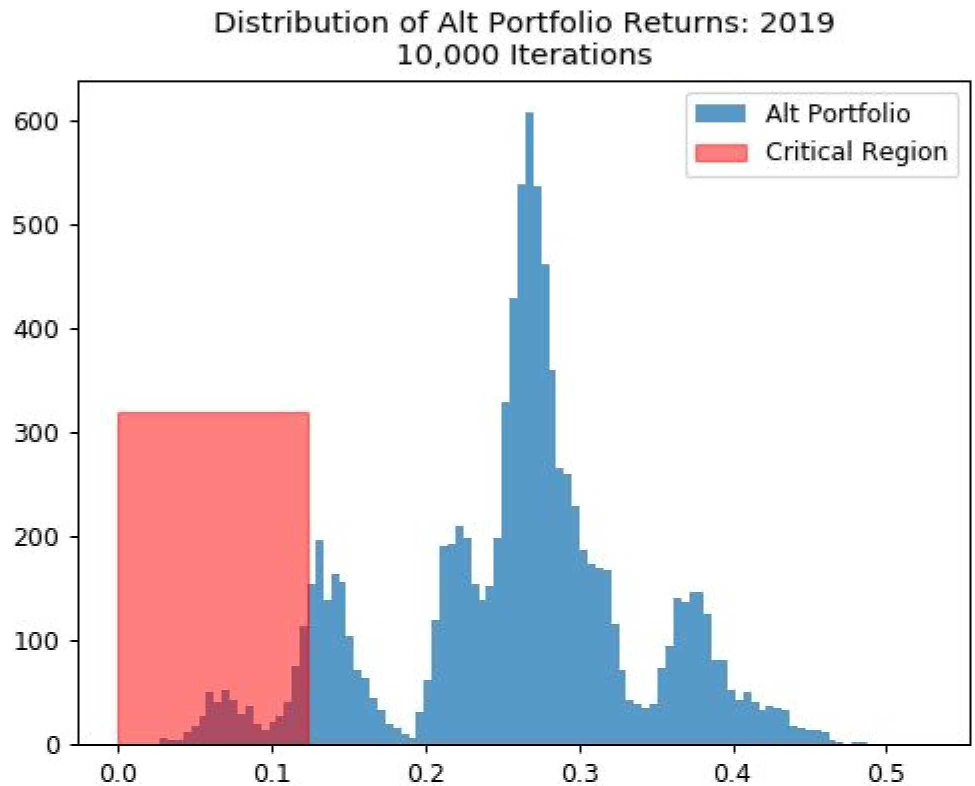

Figure 20: Distribution of Alt Portfolio Returns

According to the experiment results the Alternative Portfolio did generate a positive return in 2019.

Results: Hypothesis 4a

Check Date	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
1/3/2019	-0.0191	-0.0294	-0.0088	0.3390	Accept
3/22/2019	-0.0251	-0.0357	-0.0145	0.8097	Accept
8/5/2019	-0.0243	-0.0314	-0.0172	0.1473	Accept
8/14/2019	-0.0328	-0.0389	-0.0268	0.8789	Accept
8/23/2019	-0.0268	-0.0332	-0.0203	0.6287	Accept

Table 8: Hypothesis 4a, Control 1~Alternative Results

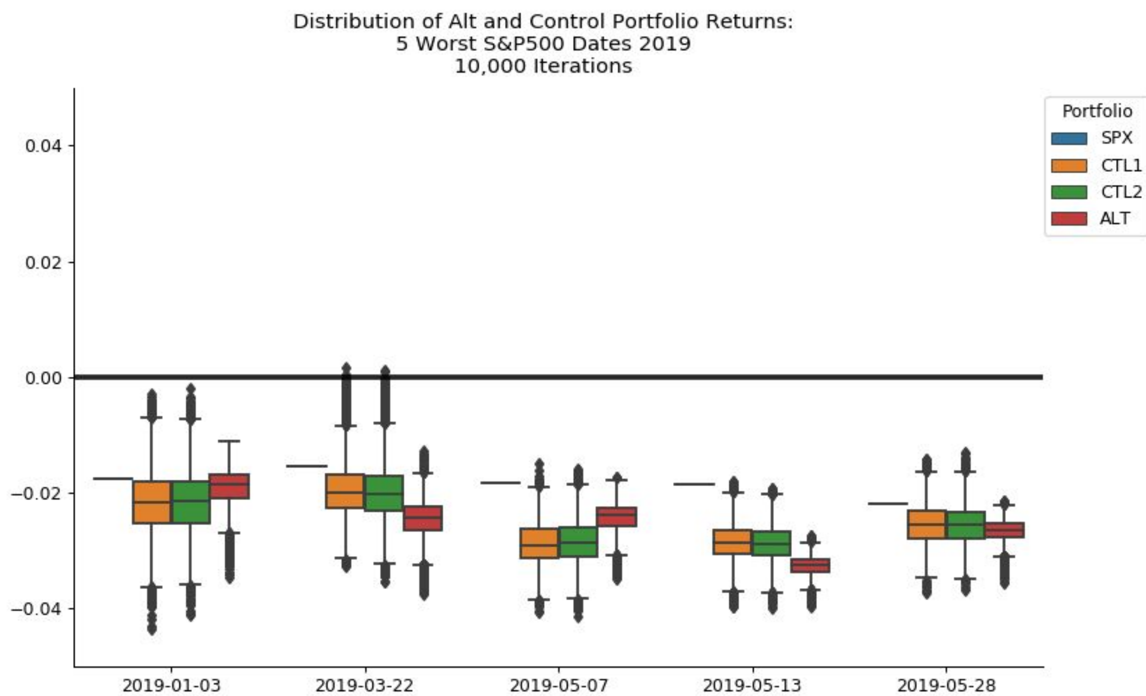

Figure 21: Hypothesis 4a, Returns Boxplots

Check Date	X Bar Alt	Lower 2	Upper 2	P-Value 2	Hypothesis Result (against Ctl2)
1/3/2019	-0.0191	-0.02927	-0.00897	0.343899	Accept
3/22/2019	-0.0251	-0.03589	-0.01433	0.793564	Accept
8/5/2019	-0.0243	-0.03145	-0.01713	0.166157	Accept
8/14/2019	-0.0328	-0.03886	-0.02677	0.862956	Accept
8/23/2019	-0.0268	-0.03315	-0.02039	0.622217	Accept

Table 9: Hypothesis 4a, Control 2~Alternative Result

Results: Hypothesis 4b

Check Date	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
2/25/2020	-0.0302	-0.0379	-0.0224	0.2907	Accept
3/9/2020	-0.0831	-0.1069	-0.0593	0.6725	Accept
3/12/2020	-0.1129	-0.1366	-0.0893	0.6778	Accept
3/16/2020	-0.1197	-0.1596	-0.0798	0.1886	Accept
3/20/2020	-0.0192	-0.0561	0.0176	0.0486	Reject

Table 10: Hypothesis 4b, Control 1~Alternative Results

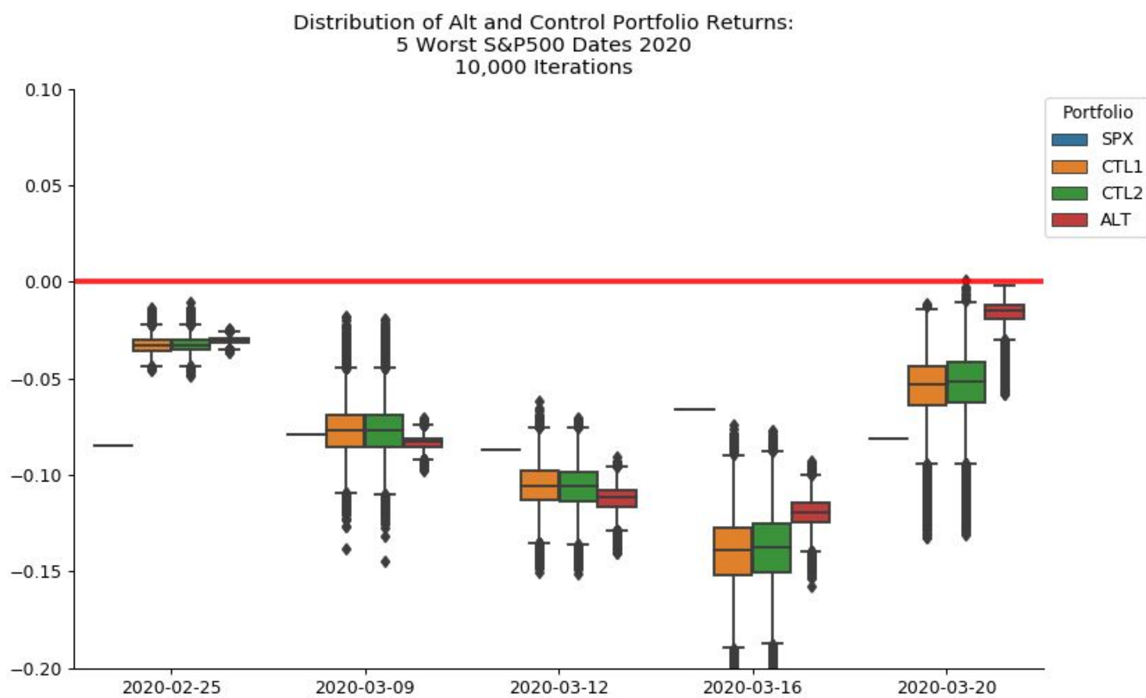

Figure 22: Hypothesis 4b, Returns Boxplots

Check Date	**X Bar Alt**	**Lower 2**	**Upper 2**	**P-Value 2**	**Hypothesis Result (against Ctl2)**
2/25/2020	-0.0302	-0.03793	-0.02239	0.308574	Accept
3/9/2020	-0.0831	-0.10718	-0.05901	0.669281	Accept
3/12/2020	-0.1129	-0.1365	-0.08939	0.674088	Accept
3/16/2020	-0.1197	-0.15923	-0.08018	0.206702	Accept
3/20/2020	-0.0192	-0.0569	0.018425	0.063688	Accept

Table 11: Hypothesis 4b, Control 2~Alternative Results

Experiment 3: Hypotheses 5-6

Hypotheses 5 and 6 concern whether the Alternative Portfolio will outperform Control Portfolio 1 and Control Portfolio 2 in 2019 and in 2020.

Alternative Criteria:

- The stock had a higher return than the S&P 500 between 8/2008-12/2008 (ie. Return Stock - Return S&P 500 > return threshold
- 5 Return thresholds were set, namely: 0.0, 0.05, 0.10, 0.15, 0.20
- Pool sizes: 195, 157, 121, 94, and 57 stocks respectively

Control 1 Criteria:

- The stock did not meet the Alternative criteria but still traded in the S&P 500
- Pool sizes: 226, 264, 300, 327, 364 stocks respectively

Control 2 Criteria:

- The stock traded in the S&P 500
- This pool consisted of 421 stocks

Hypothesis 5a: The Alternative Portfolio will outperform Control 1 and The Alternative Portfolio will outperform Control 2 in 2019

$$For\ each\ Threshold\ in\ the\ Return\ Thresholds$$

$$H_0^1 : \mu_{alt} = \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} = \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Hypothesis 5b: the same as 5a, except that performance is being tested in 2020

Hypothesis 6: In Hypothesis 6, the performance of stocks was augmented by selecting the optimal set of weights that would optimize the sharpe ratio of each portfolio. Once the weighted returns were calculated, the Alternative Portfolio was compared to the two control Portfolios. Otherwise, the statements are the same as the Hypotheses 5.

Hypothesis 6a: The Alternative Portfolio will outperform Control 1 and The Alternative Portfolio will outperform Control 2 in 2019

$$For\,each\,Threshold\,in\,the\,Return\,Thresholds$$

$$H_0^1 : \mu_{alt} = \mu_{ctl1} \qquad H_A^1 : \mu_{alt} > \mu_{ctl1}$$

$$H_0^2 : \mu_{alt} = \mu_{ctl2} \qquad H_A^2 : \mu_{alt} > \mu_{ctl2}$$

Hypothesis 6b: the same as 6a, except for 2020

Methodology

10,000 bootstraps were generated for the testing of each hypothesis. I.e The average returns of portfolios of sample size 25 were generated for the Alternative, Control 1, and Control 2 portfolios. This was done 10,000 times. After the bootstrapping was completed, one sided z-tests were performed to test the performances of the Null Hypotheses. An alpha of 0.05 was used to test each hypothesis.

Results: Hypothesis 5a

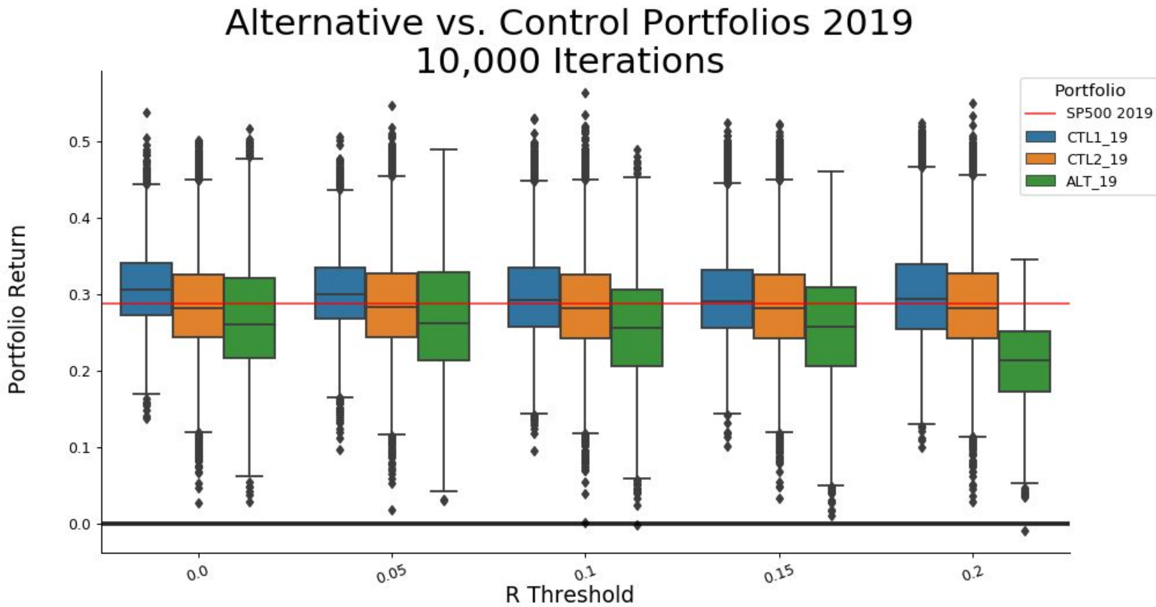

Figure 23: Hypothesis 5a, Stock Returns Threshold and Portfolio Returns

R - Threshold	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
0.0	0.2715	0.1163	0.426613	0.650469	Accept
0.05	0.2722	0.113749	0.430597	0.621812	Accept
0.10	0.2561	0.097353	0.414886	0.677422	Accept
0.15	0.2567	0.09905	0.41425	0.668998	Accept
0.20	0.2093	0.067387	0.351165	0.859842	Accept

Table 12: Hypothesis 5a, Control 1~Alternative Results

R - Threshold	X Bar Alt	Lower 2	Upper 2	P-Value 2	Hypothesis Result (against Ctl2)
0.0	0.2715	0.098538	0.444375	0.560774	Accept
0.05	0.2722	0.096414	0.447933	0.561629	Accept
0.10	0.2561	0.090635	0.421604	0.621661	Accept
0.15	0.2567	0.090687	0.422614	0.62134	Accept
0.20	0.2093	0.062263	0.356289	0.810406	Accept

Table 13: Hypothesis 5a, Control 2~Alternative Results

Results: Hypothesis 5b

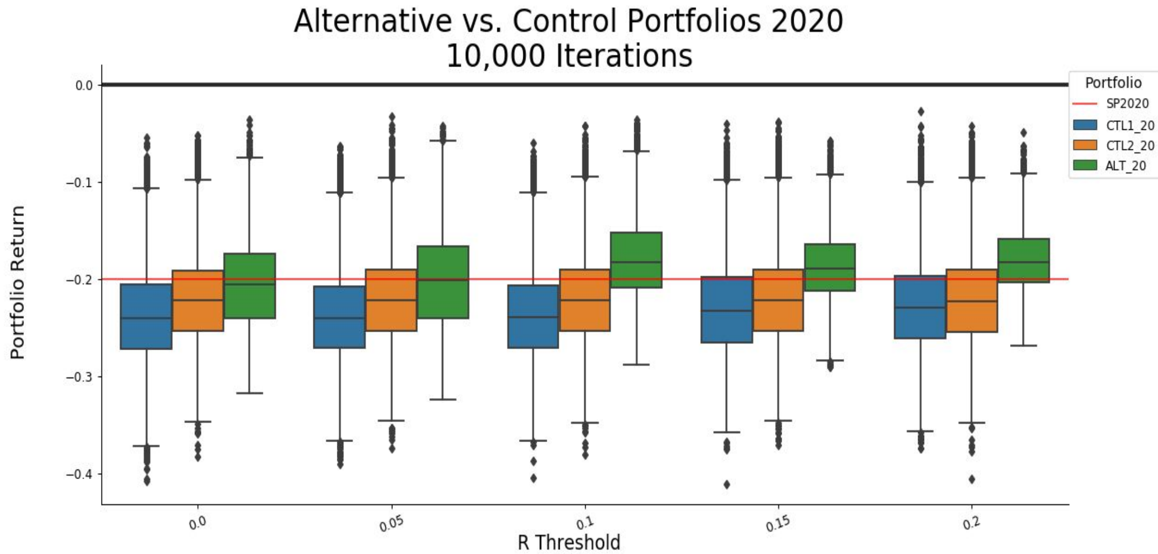

Figure 24: Hypothesis 5b, Stock Returns Threshold and Portfolio Returns Boxplots

R - Threshold	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
0.0	-0.20499	-0.3240	-0.0859	0.3394	Accept
0.05	-0.20125	-0.3200	-0.0825	0.3194	Accept
0.10	-0.17963	-0.2862	-0.0730	0.1948	Accept
0.15	-0.18738	-0.2901	-0.0847	0.2554	Accept
0.20	-0.1804	-0.2790	-0.0818	0.2260	Accept

Table 14: Hypothesis 5b, Control 1~Alternative Results

R - Threshold	X Bar Alt	Lower 2	Upper 2	P-Value 2	Hypothesis Result (against Ctl2)
0.0	-0.20499	-0.3162	-0.0938	0.4149	Accept
0.05	-0.20125	-0.3145	-0.0880	0.3970	Accept
0.10	-0.17963	-0.2849	-0.0744	0.2669	Accept
0.15	-0.18738	-0.2857	-0.0891	0.2989	Accept
0.20	-0.1804	-0.2774	-0.0834	0.2531	Accept

Table 15: Hypothesis 5b, Control 2~Alternative Results

Results: Hypothesis 6a

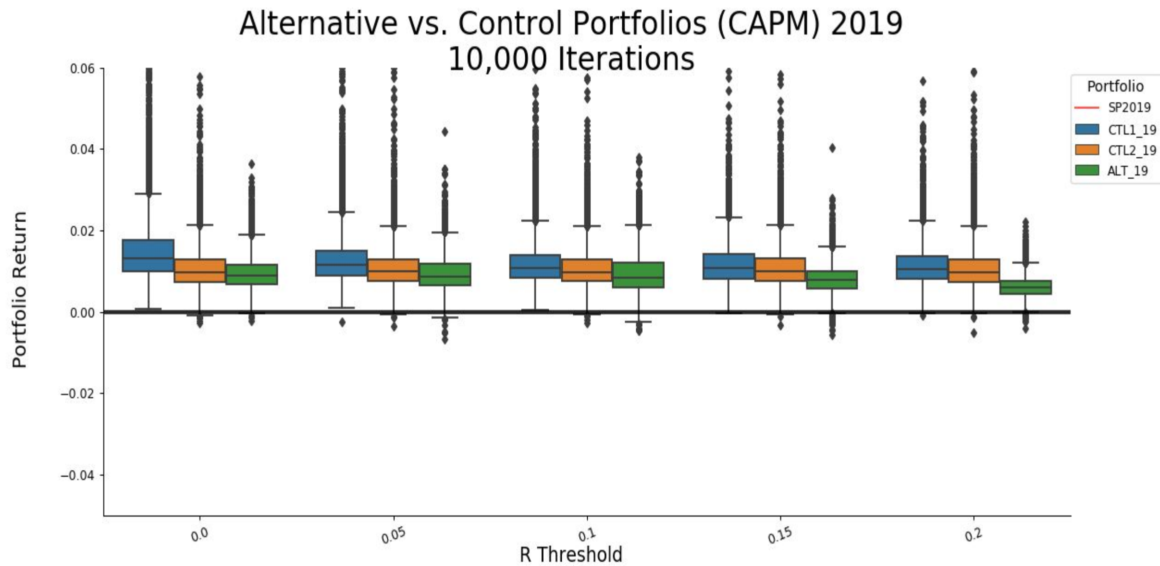

Figure 25: Hypothesis 6a, Stock Returns Threshold and Portfolio Returns Boxplots

R - Threshold	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
0.0	0.0095	-0.1037	0.1226	0.5285	Accept
0.05	0.0094	-0.0160	0.0349	0.5900	Accept
0.10	0.0093	-0.0045	0.0232	0.6190	Accept
0.15	0.0079	-0.0073	0.0232	0.6612	Accept
0.20	0.0061	-0.0038	0.0159	0.8155	Accept

Table 16: Hypothesis 6a, Control 1~Alternative Results

R - Threshold	X Bar Alt	Lower 2	Upper 2	P-Value 2	Hypothesis Result (against Ctl2)
0.0	0.0095	-0.1745	0.1935	0.4994	Accept
0.05	0.0094	-0.0073	0.0262	0.5445	Accept
0.10	0.0093	-0.0019	0.0206	0.5785	Accept
0.15	0.0079	-0.0734	0.0893	0.5272	Accept
0.20	0.0061	-0.0069	0.0190	0.7231	Accept

Table 17: Hypothesis 6a, Control 2~Alternative Results

Results: Hypothesis 6b

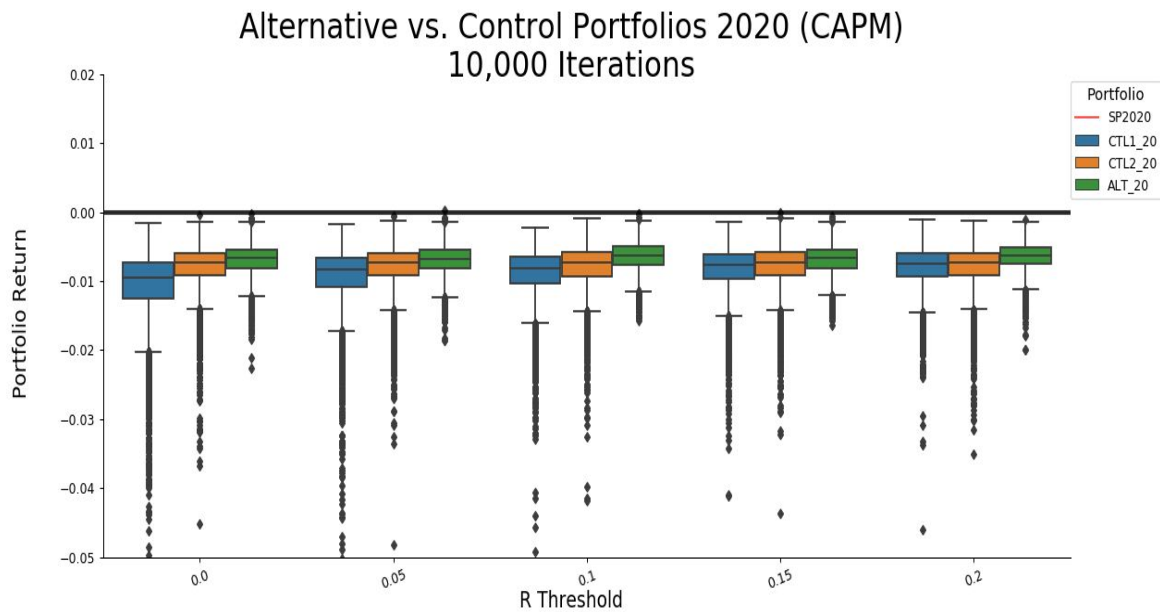

Figure 26: Hypothesis 6b, Stock Returns Threshold and Portfolio Returns Boxplots

R - Threshold	X Bar Alt	Lower 1	Upper 1	P-Value 1	Hypothesis Result (against Ctl1)
0.0	-0.0069	-0.0194	0.0055	0.3087	Accept
0.05	-0.0069	-0.0149	0.0010	0.3178	Accept
0.10	-0.0064	-0.0130	0.0002	0.2766	Accept
0.15	-0.0068	-0.0131	-0.0006	0.3538	Accept
0.20	-0.0064	-0.0129	0.0001	0.3460	Accept

Table 18: Hypothesis 6b, Control 1~Alternative Results

R - Threshold	**X Bar Alt**	**Lower 2**	**Upper 2**	**P-Value 2**	**Hypothesis Result (against Ctl2)**
0.0	-0.0069	-0.0134	-0.0005	0.4094	Accept
0.05	-0.0069	-0.0140	0.0001	0.4147	Accept
0.10	-0.0064	-0.0183	0.0055	0.4257	Accept
0.15	-0.0068	-0.0127	-0.0009	0.3916	Accept
0.20	-0.0064	-0.0123	-0.0005	0.3436	Accept

Table 19: Hypothesis 6b, Control 2~Alternative Results

Conclusions

Experiment 1 showed the potential validity in using quarterly metrics to identify recession resistant stocks. Both ROE and PEG portfolios outperformed their controls on recession data, and ROE even outperformed the controls on very bullish 2019 returns. This merits further investigation into these metrics to try to understand the relationship between company performance and stock returns.

Experiments 2 and 3 presented some very meager findings, we were hoping to get a little bit more separability between the alternative and control portfolios (especially for those stocks that beat the market by >20% at the end of 2008. We believe that using the return threshold (percentage that the stock beat the overall market) as a heuristic is a good start, but other components need to be built into the alternative portfolio criteria. Some of these include moving average comparisons between an individual stock and the overall market and the addition of changes in debt to equity ratios across time. We believe that stocks that are not highly leveraged (low D/E, or low variance of D/E) will be another useful characteristic to build into our model. Needless to say, the S&P 500 is a nice sized sample of stocks but the broader American market comprises thousands of public securities from different exchanges. This will open the scope of possible analysis.

The constant limiting factor throughout our analysis was the data. Of the daily data we obtained, only 70-75% of the stocks had corresponding quarterly data; the different quarterly treatments were inconsistent in the number of stocks tested due to this. This made it impossible to test all the treatments at once, preventing the use of completely randomized design and ANOVA techniques for Experiment 1. More data also means we can expand our hypotheses to test other treatments (metrics).

Another limiting factor was only testing on one recession period. Our data sources were limited going back further than 2007, so we only had access to the Great Recession. More periods will give more robust conclusions, and allow for a relevant multifactor conclusion across recession periods. This is where the moving average comparisons between individual stocks and the market will become useful.

Any project that looks to build portfolios or predict performance might be improved by examining its portfolio building methodology, and ours is no exception. The amount of effort put into examining portfolio weighting techniques for future data is enormous, both in theoretical discussions and practical applications. This project took a somewhat naive approach, picking a very common but flawed method in the Sharpe ratio. As there is no “perfect” method, it would be interesting to see how our hypotheses and results compare across different portfolio methodologies.

Appendix

Quarterly plots and visualizations can be found in the Appendix.zip file.

Daily visualization portfolios are contained in the following AWS S3 bucket:

- capstonefiu2020-data//visualizations/viz_portfolios
    - 386 visualization portfolios

References

1. www.investopedia.com
2. thebalance.com
3. corporatefinanceinstitute.com
4. Calculate EPS Growth Rate
5. Sharpe, W. (1966). Mutual Fund Performance. The Journal of Business, 39(1), 119-138. Retrieved March 29, 2020, from www.jstor.org/stable/2351741
6. Pav, Steven E. (2020). A Short Sharpe Course. Shortened Link
7. Lo, A. (2002). The Statistics of Sharpe Ratios. Financial Analysts Journal, 58(4), 36-52. Retrieved March 29, 2020, from www.jstor.org/stable/4480405
8. Schmid, Friedrich, & Schmidt, Rafael. (2009). Statistical Inference for Sharpe Ratio. Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds. https://www.researchgate.net/publication/265396571_Statistical_Inference_for_Sharpe_Ratio
9. https://wrds-www.wharton.upenn.edu/
10. https://github.com/RomelTorres/alpha_vantage
11. http://www.stockpup.com/data/
12. Treasury Daily Yield Curve
13. Tamhane, C. Ajit, Dunlop, Dorothy D. (2000). Statistics and Data Analysis from Elementary to Intermediate.
14. ^GSPC: Historical Prices for S&P 500
15. ^IXIC: Historical Prices for NASDAQ Composite
16. ^DJI: Historical Prices for Dow Jones Industrial Average
17. Friendly Finance with Chandra S. Bhatnagar
18. Field, Andy (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles [i.e. Thousand Oaks, Calif.]: SAGE Publications. p. 143