Title
Predicting the Abnormal Market Movement from Annual Earnings Announcements

Author
Heuschele, Sarah

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Predicting the Abnormal Market Movement from Annual Earnings Announcements

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics

by

Sarah Grace Heuschele

2022

ABSTRACT OF THE THESIS

Predicting the Abnormal Market Movement from

Annual Earnings Announcements

by

Sarah Grace Heuschele

Master of Applied Statistics

University of California, Los Angeles, 2022

Professor Nicolas Christou, Committee Co-Chair

Professor Yingnian Wu, Committee Co-Chair

Earning statements are vital in understanding the financial condition of a company. These statements provide investors the ability to make informed decisions about their involvement with a company. Event Studies were created to measure the impact of an event on the price of a security. This thesis uses the event study methodology in conjunction with many machines learning methods to predict if there will be abnormal movement to a securities price in the stock market, following their yearly earnings announcement. Machine learning methods include logistic regression with regularization, support vector machines, random forests, and neural networks. Final analysis supports use of the Lasso logistic regression for feature selection to be converted into a random forest. The feature selection allowed the model to predict abnormal price movement with 84% accuracy.

The thesis of Sarah Grace Heuschele is approved.

Maria Cha

Hongquan Xu

Nicolas Christou, Committee Co-Chair

Yingnian Wu, Committee Co-Chair

University of California, Los Angeles

2022

DEDICATION

I dedicate this work to my family, friends and partner who have shown me tremendous support and patience in the opportunity for further education.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE

Preface

1.1 Background

The stock market was created in an effort to decentralize economic power between corporations and individuals. The stock market allows for corporations and firms to raise capital needed to sustain business by giving out “little pieces” of their company to individuals. These individuals then take their piece of a company in the hopes that the value of the company will increase, therefore increasing the value of their piece, and they can turn around and sell for a profit.

Since its inception, the stock market has become an indicator of a country’s economic wealth. Countries are responsible for their own individual stock markets, but are commonly influenced by their peers, particularly the United States. The U.S. stock market has grown exponentially since its inception in 1792. It now consists of 13 different exchanges with the New York Stock Exchange holding just over $27.2 trillion in equity market [50]. All stock markets spread securities across different economic industries and sectors, providing ways to measure areas in the economy that are flourishing while others are suffer hardships. While the market facilitates trading, indices have been created to capture the overall health of a market. In the United States, the S&P 500 was formed, containing 500 companies that are believed to encapsulate the health of 11 different sectors in the market, ranging from Technology to Health Care. The S&P 500 benchmark is a market-cap-weighted-index, which means that companies with a larger market value with have a higher weight in the index. This comes from the belief that larger companies make a bigger impact in the economy of a market [22].

1.2 Introduction

The stock market heirs itself to investors at an individual level - buying one security at a time, to larger, investment institutions. While trading seems straight forward, the issue still stands that there is not a “best practices” way to predict market volatility. A common method used across firms to predict market movement is a moving average, which attempts to smooth the price of a security and remove any ad-hoc spikes in price.

In order to be listed on a U.S. market, publicly traded companies are required to disclose their earnings on a quarterly basis. These statements hold the current financial situation of a company and can indicate the profitability and earnings potential in the months to come. These earnings statements are punitive to evaluating a company and the decision to invest, sell, or remain neutral.

Earnings statements also hold power in looking at trends over time. According to an article by US News, comparison of the Price-Earnings Ratio over time indicates the growth of the company. When comparing two equal PE ratio of two high priced securities, the one with an increase in earnings, driving the cost of the security is much more valuable than a security with detreating earnings [13].

An event study is a methodology created to measure the impact of a large company event on the stock market. According to a paper in The Journal of Economic Literature, the event studies framework was developed by James Dolly to analyze the effect of stock splits on securities market price [24]. Since then, the development and use of event studies have expanded over time and to many different disciplines. Discussed in more detail, in later sections, event studies use historical trends of price movement to measure the impact of an event on a security’s price. The study works by using the historical price average as a base, accounting for volatility, and denoting if the price movement from an event differed from its historical path.

In this analysis, we will use event studies to predict the price impact of earnings statements on securities in the market. We will use historical price data, S&P 500 data and detailed accountment data to predict if there was a significant price movement the day preceding an announcement date.

CHAPTER TWO

Literature Review

2.1 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) is a theory developed by Eugene Fama that states a security's price in the market fully reflects all available information of a company [8]. In practice, this theory states that as new information is released into the market, the price of the security will immediately change to reflect this news. The EMH theory suggests that it is impossible to outperform the market, as the current security price is the *fair* price and already holds all information needed to know about the company. Figure 2.1 displays how the theory breaks up the EMH into 3 tiers. As you expand across tiers, you collect new information about

Figure 2.1 Levels of the Efficient Market Hypothesis [22]

the company as well as the information from the tier before. Here we can see that the more information known about the company, the more it is reflected in the price.

In the Weak-Form EMH, the assumption is that the only available information about current security prices, is from historical prices. This is similar to the Random Walk theory developed by Burton Malkiel in 1973 [24]. The random walk theory states that the changes in price movement all come from the same distribution and are independent of each other. This means that you cannot use historical data to predict future prices. The Efficient Market Hypothesis shares this belief, only at a weak level. From the EMH lens, if you are only given the historical prices, the price of the security is not “fair priced” as it does not show the current financial stability of a company or future company information.

The Strong-Form indicated that all information possible about a company is reflected in its current price. From historical data, current and future information as well as inside knowledge from employees and higher management are incorporated within the price.

For most large cap stocks, price reflection falls somewhere between these two extremes. The Semi-Storing EMH states that all publicly available information is held by the security. The Corporate Finance Institute shared a great example of this effect. Each month, the Non-Farm Payroll Report is released in the U.S. and each month, you can see prices adjust to reflect this new information in their price [9].

2.2 Financial & Earnings Statements

Earnings statements are required by the SEC for all publicly traded U.S. companies. This allows all traded companies to contribute to the market fairly and give investors equal insight before investing. These statements give insight into the financial stability of a company at a given point in time and help shareholders to make informed decision about their participation with the company. These statements report many key metrics that show the balance between a company’s earnings and overall performance. Metrics can denote how fast a company’s earnings are growing, how much debt they currently have, ability to make a profit, etc. New statements come out on a quarterly and yearly basis, making historical comparison fairly easy.

Within the statements, you are able to gather information about profitability, liquidity, debt, efficiency, and valuation through various ratios provided [15]. According to US News, the snapshot that these statements provide instigate investors to react, “ directly affecting stock prices in the short term” [13]. Good or bad, financial statements have the ability to change the perception of a company and therefore alter its price in the market. Days around earnings announcements, we tend to see drastic movement in prices. Figure 2.2 comes from DailyFx.com depicts the market reaction to Apple’s earnings statements in 2019.

Figure 2.2 Apple's Stock Price Following Earnings Announcement [2]

As seen in the chart, the announcement significantly dropped the price per share in the market while increasing the volume. This indicates that it was a negative earnings announcement that signaled investors to sell their shares. Highlighted by the gold box, we see a concept known as price reversion. Prices that drastically swing one way or the other, tend to return to their historical state, in this case, their long-term average price. More often than not, prices to return to their equilibrium after earnings announcements but how long that'll take is unknown. There are some cases however where we see that prices continue to move in that direction. For example, in August 2020, Tesla's stock had risen so high that they announced a 4:1 stock split in order to bring the price back down.

Analysis of financial statements and their change over time can give indication of these "runaway stocks" that show significant growth (or decline) from an early state.

2.3 Event Studies

Event studies were first introduced as a way to evaluate the impact of stock splits on the price of a company. Since then, there have been many wide adaptions that span past the finance and investing sectors. Event studies work by looking at the impact to a security on a short time horizon, meaning it only considers a few days following an event. According to a paper by S.P. Kothari and Jerold B. Warner, the importance of event studies is to provide a measure of the" unanticipated impact" of the earning announcement for shareholders [10]. Event studies measure the abnormal return associated with an event. This abnormal return is calculated using previous security and market information to predict an expected return and comparing it to the realized return following an event.

From EMH theory, we know that a stock's price is representative of all publicly available information. As shareholders, we know that financial statements are required by companies on a quarterly and yearly basis. With that, we can expect some impact when a statement comes out on the price of a security. The key word, "unanticipated" that Kothari and Warner [22] use is important. It fills in the gap between the public knowledge and the current price of the firm by adjusting the price relative to the earnings statements. Information provided on these statements is not publicly available in the three months between announcements making their impact temperamental. Event studies of previous earnings unanticipated returns can help shareholders predict and prepare for the impact of the next statement release.

In the following section, I will discuss the Event Study methodology in detail. This set the foundation for the following analysis.

CHAPTER THREE

Event Study Methodology

Event studies are developed to monitor an event in time and the effect of that event on the overall wealth of a company. The measurement of a firm's wealth is determined by the price movement of its respective stock in the market. To quantify this, event studies focus on looking at abnormal returns. These values highlight the difference between what we would expect the stock's return to be, without the event, and the realized return from that day.

To set up your event study, the first step is to identify the observation and estimation time periods. These windows will surround your event date and are critical in estimating the expected return of the security and the abnormal return derived from the event. Figure 3.1 is a candlestick chart displaying the daily high, low, open, and close of a security's price over time with respect to the observation and estimation windows.

Figure 3.1 Event Study Windows

Realized returns are the actual returns that we saw in the market on that day. These values are what we use during the observation to compare to our estimate. During the estimation period, we use the actual returns of the security and the market in the period before the event to calculate alpha and beta, these values are then used in the single factor regression to find our estimated return. Each alpha and beta is specific to the security and the time period around each event. There are many methods used to calculate the estimated return from our estimation window, however we will use the popular Market Model.

The Market Model regresses a security's daily returns to the daily return of the market. According to EventStudyTools, this is the most popular model to estimate the security returns in the industry [17]. It allows us to find the relationship between how the security moves relative to the market. In investment terms, this is our beta. Values that are closer to 1/-1 move more in line with the market. The equation for the market model (3.1) is built of the regressed coefficients from our estimation period.

$$E[R]_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \tag{3.1}$$

The expected return is calculated for each security (i) at each date in the observation window (t). Since we believe that beta (β_i) captures the relationship between the security and the market, we use that to scale the return of the market (R_{mt}) each day to find the estimated return of the security $E[R]_{it}$. According to EventStudyTools, we can assume that the error (ε_{it}), which is a random variable, has an expectation of 0 and that there is no correlation between the error, the market, and the security [17]. The error is homoscedastic and is assumed not to be auto correlated [5]. The difference between the expected return from the Market Model and the realized return on that day is considered the abnormal return (3.2).

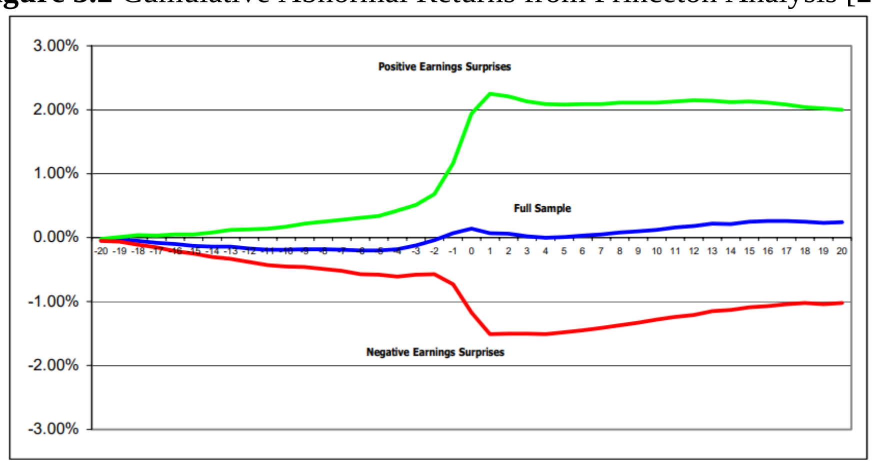
$$AR_{it} = R_{it} - E[R]_{it} \tag{3.2}$$

Taking the abnormal returns from the observation window, we look to find the cumulative abnormal return (3.3). Essentially this just requires the summation of abnormal returns over a specified time period (t_1, t_2). As mentioned by EventStudyTools, this allows us to measure the total impact of an event over time and not specific to one day [17]. This is useful as it may take investors time to react to a statement and limiting analysis to just one day will limit the ability to understand what the overall impact of the earnings statements is.

$$CAR_{(t_1 t_2)} = \sum_{t=t_1}^{t_2} AR_{it} \tag{3.3}$$

The foundation of our analysis is to see the cumulative abnormal returns for securities following a 10K yearly earnings announcement. Figure 3.2 comes from Princeton research paper titled "*Value Relevance of Analysts' Earnings Forecasts*" [3] and provides some context on the movement of cumulative abnormal returns post earnings announcements. In their analysis, they address the differences in impact by positive and negative news released into the market. Here we can see the cumulative abnormal return across their observation window of +/- 20 days. The 20 days before shows little indication of price movement. It isn't until the day of their event (t=0) where we see large movement in the cumulative abnormal returns.

Figure 3.2 Cumulative Abnormal Returns from Princeton Analysis [25]



Based strictly off the chart, the price movement seems steep and impactful. However, the final part of an event study is to determine whether or not the Cumulative Abnormal Returns are significant or can be explained by noise within the market. We are performing the hypothesis test (3.4) to determine whether the cumulative abnormal returns are zero or not, within the observation window.

$$H_0: E[CAR_i] = 0 \quad \text{vs} \quad H_A: E[CAR_i] \neq 0 \tag{3.4}$$

Since we are looking at each security and their respective earnings announcements independently, we will use the t test to calculate our test statistic (3.5).

$$T_{CAR} = \frac{CAR_{it}}{d^2 {S^2}_{AR_i}} \tag{3.5}$$

Where (CAR_{it}) is the Cumulative Abnormal Return of the security at that time, over the variance of the abnormal return (${S^2}_{AR_i}$) in the observation window. There is a scaling factor applied (d^2) that accounts for the time passed and the added volatility from the market over the interval. Using a significance level of 0.05 we will determine securities to have significantly large impacts when their test statistics are larger than 1.95.

CHAPTER FOUR

Exploratory Data Analysis

4.1 Data Collection

The Efficient Market Hypothesis suggests at a semi-strong level: a securities prices are reactive to new information. Using the event study methodology, we can measure the impact of earnings statements on the 505 securities listed on the S&P 500 from 2012 to 2016.

The following dataset was collected from Kaggle and the SEC. The data from Kaggle, titled "New York Stock Exchange" contains the historical prices and earnings statement information for the 500 companies listed on the S&P 500. Historical price data ranged from 2012 to 2016 and each company has an earnings statement each year. The data is split between three files; one of historical price data, one of security specific information (eg: GIC sector) and the last with details from their annual SEC 10K filings.

The historical price data was sourced from Yahoo Finance, the annual filings was a mix from the NASDAQ financials and index prices were found in the EDGAR SEC databases. Analysis includes 505 companies, all publicly traded on the NYSE, Nasdaq, and Cboe. The index used to benchmark was the Standard & Poor's 500 (S&P 500).

In the analysis, we adjusted for when earnings statement dates did not match up with trading days. This happens when specific companies release earnings on the same day each year, sometimes that date falls on a weekend or when markets are closed for holiday. In this case, we used the preceding trading date as the event day, in order to not miss any of the abnormal return inherited by the earnings release.

In order to get an accurate alpha and beta for an event study, you need at least 45 days of stock prices before your observation window begins. For this dataset, that removed about 30 earnings release statements that occurred in either 2012 or 2016 when the dataset is missing subsequent price information. Finally, there were 2 more observations that had missing values for their financial statements. Given the nature of financial statements, I was able to back into their values given other ratios that contain the information in a different way. This left our model with 1645 observation and 92 variables.

4.2 Data Processing

In order to perform certain machine learning techniques, it is imperative that we first rescale all variables. Rescaling offers the ability to look at different variables, across different scales and compare them equally. This means that a feature such as the total number of liabilities, can be compared to daily returns. It also helps to standardize the different scales so that features with larger values do not outweigh smaller features in many machine learning applications. For our analysis, we have utilized the z-score (4.1) for standardization. This converts each column to have a mean (μ) of 0 and a standard deviation (σ) of 1, ultimately creating equal data distributions across columns.

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

It is important to note that not all machine learning techniques applied in this analysis require standardization, however, the application of standardization does not diminish or take away from the accuracy of these models, and therefore, will be left in their standardized state.

4.3 Attribute Definitions

Attribute	Descriptions
Order Start Date	The date that the earnings statement was released.
Symbol	The NSYE’s ticker – specific to each security.

Attribute	Descriptions
Price Return 1 Day Before	The security’s return 1 day before the earnings statement was released.
Price Return 2 Day Before	The security’s return 2 days before the earnings statement was released.
Price Return 3 Day Before	The security’s return 3 days before the earnings statement was released.
Price Return 4 Day Before	The security’s return 4 days before the earnings statement was released.
Price Return 5 Day Before	The security’s return 5 days before the earnings statement was released.
Previous Day’s Volume	The traded volume for the security the day before the earnings statement was released.
Index Return	The return of the index the day that the earnings statement was released.
Cumulative Abnormal Return	The cumulative abnormal return the day after the earnings statement was released.
Significant Indicator	(0,1) to indicate if the cumulative abnormal return is statistically significant.
Accounts Payable	The amount of money a company owes its creditors. [27]
Accounts Receivable	The amount of money a company receives from its creditors. [27]
Additional Income Expense	Additional items that are an income expense.
After Tax ROE	Efficiently of a company generating income from the equity investments of its shareholders [28].
Capital Expenditures	money spent by a business or organization on acquiring or maintaining fixed assets, such as land, buildings, and equipment.[27]
Capital Surplus	the surplus resulting after common stock is sold for more than its par value [28]
Cash Ratio	a measurement of a company's liquidity, specifically the ratio of a company's total cash and cash equivalents to its current liabilities. [28]
Cash & Cash Equivalents	reports the value of a company's assets that are cash or can be converted into cash [28]
Changes in Inventories	The change in the value of inventory held by the company.
Common Stocks	shares entitling their holder to dividends that vary in amount and may even be missed, depending on the fortunes of the company.[27]

Attribute	Descriptions
Cost of Revenue	the total cost of manufacturing and delivering a product or service [28]
Current Ratio	a liquidity ratio of a company's ability to pay short-term obligations or those due within one year. [28]
Deferred Asset Charges	an item on a company's balance sheet that reduces its taxable income in the future.[28]
Deferred Liability Charges	a liability that isn't due in the current accounting period [28]
Depreciation	a reduction in the value of an asset with the passage of time [27]
Earnings Before Interest & Tax	is a company's net income before income tax expense and interest expenses are deducted [28]
Earnings Before Tax	A company’s net income before taking out the tax expense. [28]
Equity Earnings Loss Unconsolidated Subsidiary	Equity from a subsidiary that is not included in the parent company’s consolidated financial statements. [28]
Fixed Assets	Long term assets that are not used to convert to cash (ex: equipment). [28]
Goodwill	The value of an asset above the price of purchase. [28]
Gross Margin	Net sales after the total revenue minus the company expenses. [28]
Gross Profit	Net sales – cost of goods sold [28]
Income Tax	The tax paid from the income of the company.
Intangible Assets	Assets owned by the company that are not physical objects.
Interest Expense	The cost incurred by borrowing funds.
Inventory	All things owned by a firm. [27]
Investments	As asset that a company purchases with the goal of generating income [28].
Liabilities	Debt the company owes the shareholders.
Long Term Debt	Debt owed to shareholders to be repaid in over a year.
Long Term Investments	Assets owned by the company that will churn out money in the long run.
Minority Interest	When a company owned less than 50% of another company [28].
Net Borrowing	The amount of money borrowed to finance the company.
Net Cash Flows	The amount of cash held after subtracting the cash outflow from the cash inflows.

Attribute	Descriptions
Net Cash Flows Operating	The amount of cash held from operations.
Net Cash Flows Financing	The amount of cash held from financing.
Net Cash Flows Investing	The amount of cash held from investing.
Net Income	The amount of income after subtracting out expenses and debt.
Net Income Adjustments	Adjustments to net income.
Net Income Applicable to Common Shareholders	Capital remaining that a company can choose to distribute back to its shareholders.
Net Income Cont. Operations	Capital retained from a company's routine business operations.
Net Receivables	The amount of money a company earns from its shareholders less defaulters.
Non-Recurring Items	Items that are not likely to happen again [28].
Operating Income	Income from operations.
Operating Margin	Margin earned from operations after comparing the income to the expenses of operations.
Other Assets	Additional assets owned by the company.
Other Current Assets	Additional assets that can be converted to cash in short notice.
Other Current Liabilities	Additional short-term debt owed within 12 months.
Other Equity	Stocks that represent ownership interest in the company [28].
Other Financing Activities	Additional ways the firm raises capital.
Other Investing Activities	Additional ways the firm invests in long term assets for the company.
Other Liabilities	Additional debt owed.
Other Operating Activities	Additional income from operations that are not from routine business (ex: receiving dividends).
Other Operating Items	Additional debt from operations that are not from routine business (ex: paying dividends).
Pre-Tax Margin	The amount of profit from sales.
Quick Ratio	The ability for a company to pay for its current debt [28]. It is a measure of liquidity.
Research and Development	Money allocated to research and development of new products/services by the company.
Retained Earnings	Income remaining after payoff of dividends.
Sale and Purchase of Stock	Debt related to the purchase of stock.

Attribute	Descriptions
Sales, General and Admin	Operating expenses for the sale and administration behind running the company.
Short Term Debt	Debt obligations that must be paid in a year
Short Term Investments	Assets owned by the company that are used to produce a profit in under a year.
Total Assets	All assets owned by the company.
Total Current Assets	All assets that the company can quickly convert to cash.
Total Current Liabilities	Amount of debt owed to shareholders within a year.
Total Equity	The amount of capital earned from investors.
Total Liabilities	All debt owed to shareholders.
Total Liabilities to Equity	The ratio of debt owed to income generated by shareholders.
Total Revenue	All profits earned from your service or goods.
Treasury Stock	The number of shares bought back from shareholders, to be held by the company.
Year	Year of the earnings statement.
Earnings Per Share	A company's profit divided by the number of shares outstanding as a measure of profitability [28].
Estimated shares outstanding	The number of shares left in the market, available for purchase.
Security	Long name of the security.
GICS Sector	A global classification that breaks securities into 11 categories.
GICS Sub Industry	A global classification that breaks securities into 158 categories.

4.4 Foundational Data Analysis

As we look at predicting whether or not a company will have a significant abnormal return following their earnings announcement, understanding the industry behind the security is critical. Not all types of securities react the same and consumers are not all effected by every sector. For example, if there is a negative earnings announcement and the cost of electricity is thought to rise, many consumers are going to lower the value of that company by selling their shares. On the other hand, if Amazon shows that it is making another $5 million in sales from

Amazon Prime, the Consumer Discretionary market is not likely to react as much given that isn't a ton of money compared to their bottom line as their products are considered nonessential. Table 4.1 shows the count of our securities across the 11 GIC sectors.

Table 4.1 GIC Sector Descriptions

Sector	Count	Sector	Count
Consumer Discretionary	78	Health Care	49
Consumer Staples	32	Industrials	62
Energy	31	Information Technology	61
Financials	52	Materials	24
Real Estate	27	Telecommunication Services	5
Utilities	24		

Figure 4.2 shows the proportion of significant to non-significant reactions to earnings statements by sector. Consumer discretionary has the largest number of significant reactions, however it seems that Materials has a larger proportion of their earnings statements resulting in significant price movement. Overall, we do see a trend that with an increased number of securities and earnings statements, there is a larger number of significant reactions. This is important that we don't see any potential outliers.

Figure 4.2 Proportion of Significant Earnings Releasements by GICS Sector

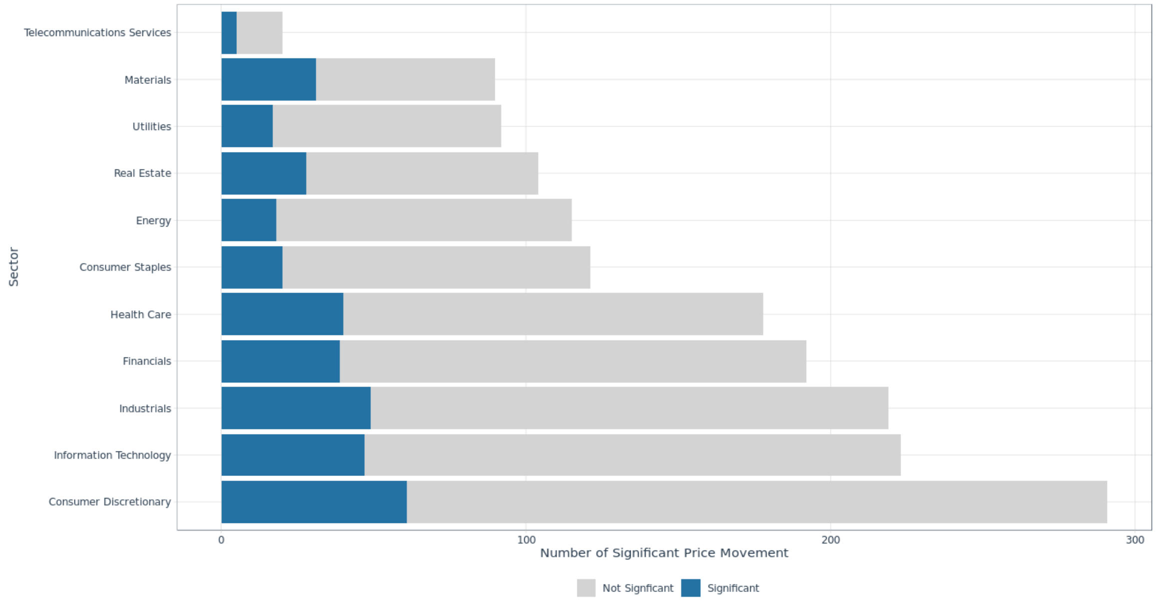
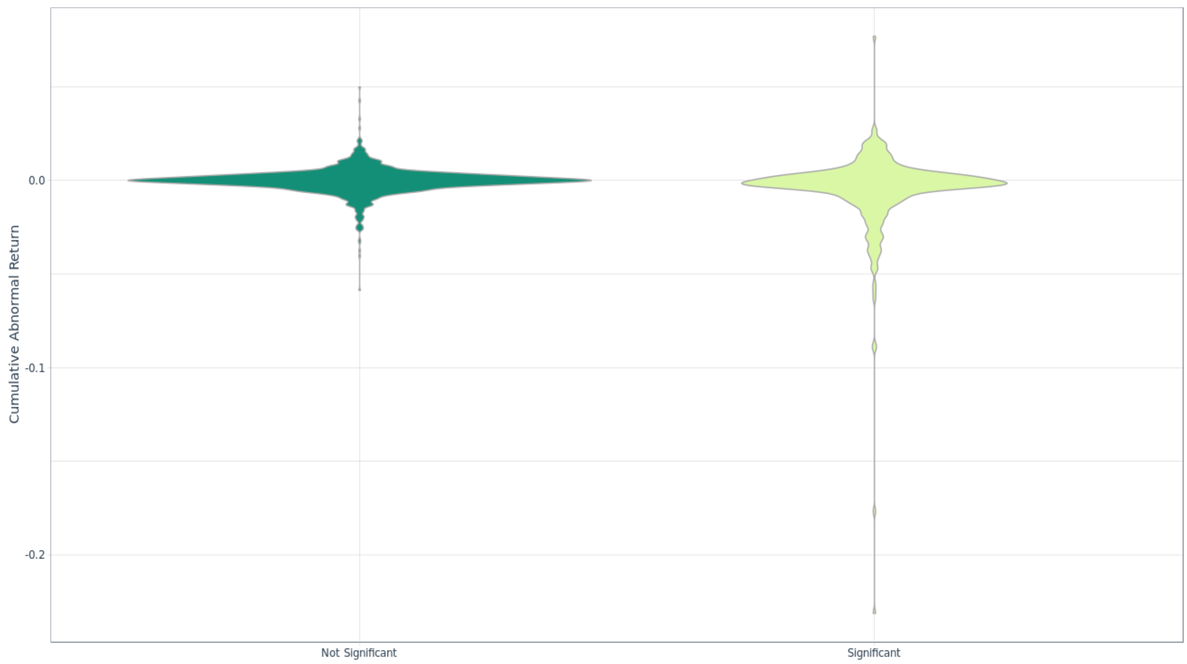


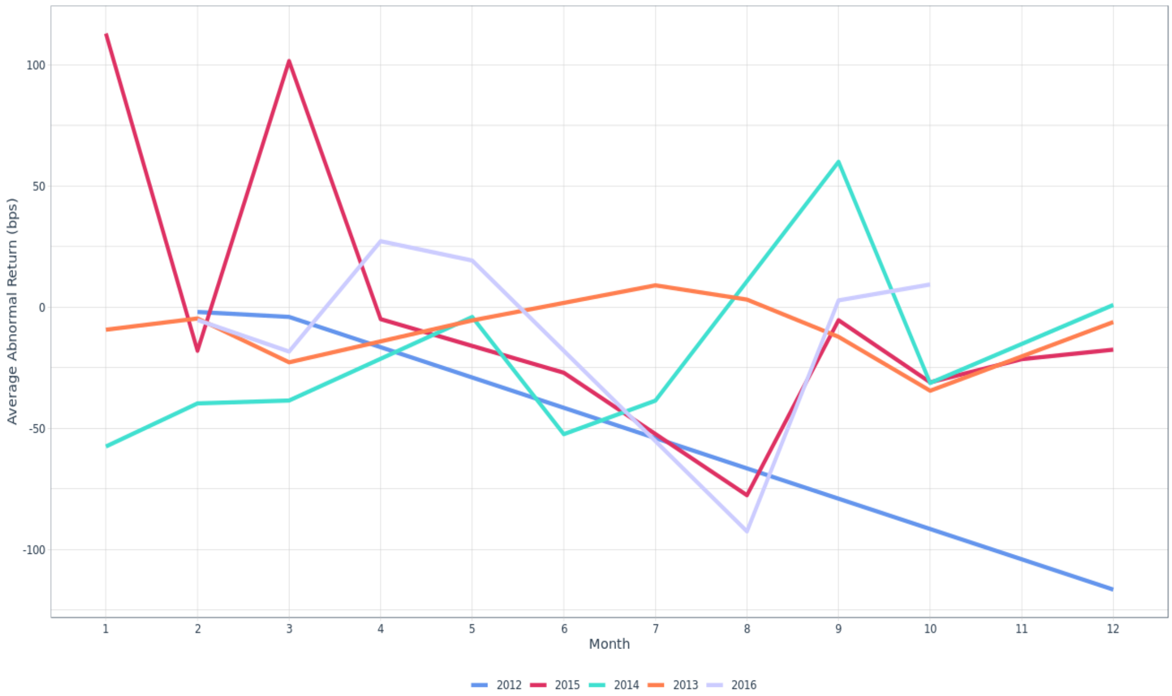
Figure 4.3 shows violin plots of the Cumulative Abnormal Return for significant and non-significant events. They both have a peak in their distribution around 0, however significant events are skewed towards a negative return. This means more often than not, if a security in the S&P 500 has a significant abnormal return, it is negative a move in the opposite direction of your investment.

Figure 4.3 Distribution of Cumulative Abnormal Return by Significant/Non Significant Events



Finally, in Figure 4.4, we can see that the Average Abnormal Returns for each month between 2012 to 2016. As the year progressed in 2012, the average abnormal return declined rapidly. We do not see that seasonality across any other year. In fact, 2013 shows little movement as the cumulative average abnormal return hugs the 0-bps return axis while 2015 shows severe volatility as we see the average abnormal return making large leaps from 100 bps movement, to 0 and back to 100 bps return across three months. This is good news because seasonality in the abnormal returns could indicate a factor outside of earnings statements that influences abnormal returns in securities.

Figure 4.4 Average Cumulative Abnormal Return by Year



CHAPTER FIVE

Methodology

5.1 Logistic Regression

Logistic regression is similar to linear regression except the output variable is binary. For our analysis, we are looking whether an earnings release for a company creates an abnormal return (1) or not (0). Logistic regression works by predicting the probability of an observation belonging to one of those two classes. To calculate these probabilities, we map real values using a sigmoid logit function (5.1) which will give us a probability between (0,1).

$$f(x) = \frac{1}{1 + e^{-(x)}} \tag{5.1}$$

Expanding on the sigmoid function, for logistic regression, we adjust to account for the linear relationship between the features, resulting in the following equation (5.2).

$$h(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \tag{5.2}$$

These functions denote probabilities between [0,1] but are not limited to those two classes. To solve the disconnect, a breakpoint of 0.5 must be established to determine which probabilities signal for class 0 and which signal class 1.

In order to measure the performance of logistic regression, we need a cost function other than the Mean Square Error. Stemming from linear regression, the MSE will try to measure a logistic model linearly, avoiding the convex and movement in the model. Instead, we use a cross-entropy log loss equation.

5.1.1 Regularization

Ridge Regression is an extension of regression that works to minimize the complexity of the model. Ridge adds a penalty that is equal to the square of the coefficient's magnitude [32]. Coefficients are then shrunk, but not eliminated. To implement ridge regression, we want to minimize its cost function (5.3), to increase our predictive power.

$$\beta_{Ridge} = \min \left(\|Y - (\theta)\|_2^2 + \lambda \|\theta\|_2^2 \right) \tag{5.3}$$

Lasso Regression is very similar to Ridge as it works to minimize complexity by applying a tuning parameter to the loss function (5.4). The main difference is that Lasso will shrink unnecessary parameters to 0, eliminating them from the model. Lasso not only addressing multicollinearity, but also performs feature selection [32]. One problem with Lasso's nature to perform feature selection is that it will eliminate any correlated variables, only retaining one. This could lead to a loss in information and understanding of the relationships within our model.

$$\beta_{Lasso} = \min \left(\|Y - (\theta)\|_2^2 + \lambda \|\theta\|_1 \right) \tag{5.4}$$

Elastic Net is a combination of Ridge and Lasso Regressions. In practice, you can approach Elastic Nets by apply two different penalty terms (5.5), one for Ridge and one for Lasso, or by making the penalty a weight [32] between the two methods (5.6).

$$\beta_{Elastic\ Net} = \min \left(\|Y - (\theta)\|_2^2 + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta\|_2^2 \right) \tag{5.5}$$

$$\beta_{Elastic\ Net} = \min \left(\|Y - (\theta)\|_2^2 + \alpha(1 - ratio) \sum_{i=1}^{m} |\theta_i| + \alpha(ratio) \sum_{i=1}^{m} |\theta_i^2| \right) \tag{5.6}$$

5.2 Support Vector Machines

Support Vector Machines (SVM) is a method of supervised learning that attempts to classify data into two groups. One of the many praises that SVM receives is its ability to withstand influence from outliers and does not rely on data assumptions [6]. SVM works to create a hyperplane that is used to divide the two groups (Figure 5.1) while creating the largest margin between the two groups (Figure 5.2).

Figure 5.1 Hyperplane Boundary

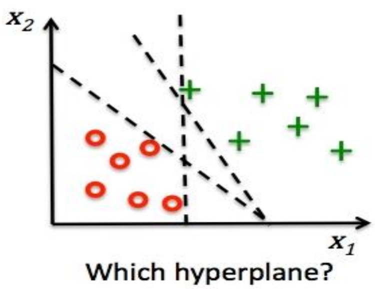
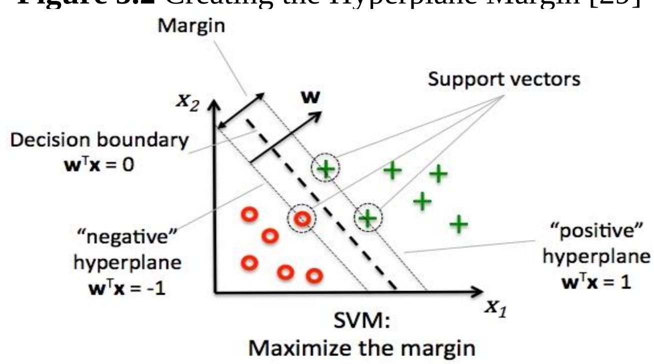


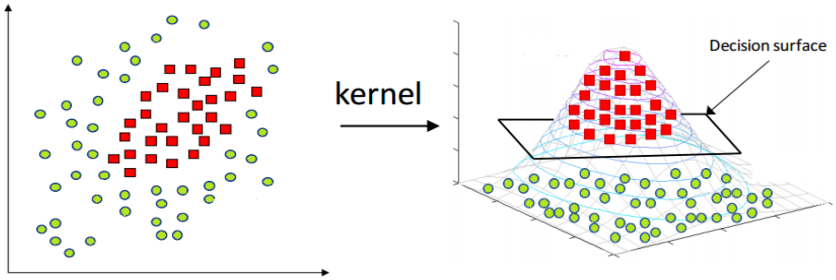
Figure 5.2 Creating the Hyperplane Margin [29]



The closest "members" from each group are denoted as the *Support Vector*, as they help to set the distance of the margins from the boundary line. To create the hyperplane boundary, the algorithm measures the distance between the support vectors and the boundary, finding the combination that maximizes this difference. To find the best boundary split, we want to minimize $\| \vec{\omega} \|$, which will maximize the margin.

In the previous examples, we see linear data, making it is easy to separate the data into two groups by a hyperplane boundary line. However, many use cases have nonlinear data (Figure 5.3). In this case, a kernel transformation is applied to the data which projects the data onto another dimension, making it easier to create a hyperplane boundary line. The kernel transformation can project a 2-dimensional space, onto a 3-dimensional space which helps facilitate the separation between the two groups.

Figure 5.3 Nonlinear SVM Kernel Transformation [30]



A kernel function is necessary to help understand the transformation of the nonlinear data onto a linear plane. The radial kernel function measures the Euclidian distance between the input and a fixed point, the hydroplane [28]. The radial kernel function (5.7) was used in our analysis and is as follows,

$$K(x, x') = \exp\left(-\frac{\| x - x' \|^2}{2\sigma^2}\right) \tag{5.7}$$

To address the overfitting, a cost parameter was applied to our SVM model. With a kernel application and a nonlinear boundary, the opportunity to overfit our model grows. To combat this, a cost parameter is applied that seeks to collect the highest accuracy without overfitting the model. This helps to optimize the model for interpretation of new data.

5.3 Random Forest

Random Forest algorithms are another type of supervised learning that handles large datasets with high dimensionality extremely well. Another strong suite of Random Forest is its ability to work with both regression and classification inputs and outputs. The random forest model helps to combat some of the downfalls of Decision Trees. The algorithm builds out multiple random trees into a “forest” based off subsets of the original dataset. These subsets are created using bootstrapping, meaning the subsets are collections of rows from our original dataset, selected at random, with replacement. This means that some observations may repeat within a tree.

Each subset of data will be used to construct its own decision tree, independent of the other subsets and their respective tree. Each tree will choose a selection of variables, at random, from our dataset to train the model. This causes each tree to be unique and provide its own estimations. An example of 3 separate tress, from a single data set, is shown in Figure 5.4. The example shows how the random selection of variables and data subsets yield unique trees.

Figure 5.4 Unique Trees in a Random Forest

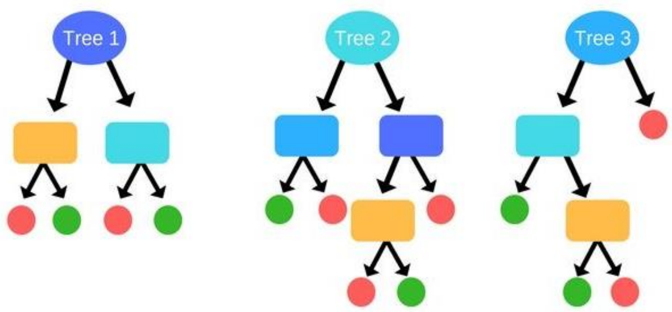
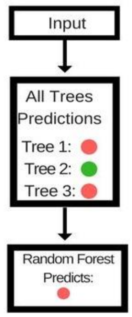


Figure 5.5 RF Prediction Process



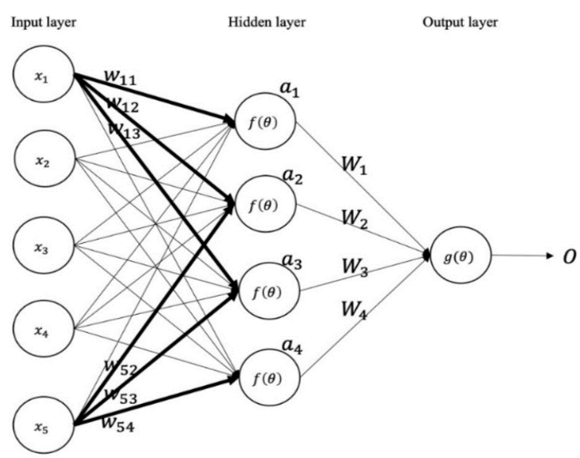
For a new observation, the data will be fed through each tree in the "forest" and its output will be collected. The final prediction will be a *majority wins* scenario in which the most commonly predicted group will be the final prediction of the data. Figure 5.5 shows the process of new data going through the random forest model. Each of the three trees produces its own production but the majority vote red, so that will be model's final output.

The algorithm is less sensitive to outliers and creates a more holistic and opportunistic model while reducing the overall variance in the prediction. This is done by first bootstrapping our data to ensure that we are not overfitting our model to the original dataset. It is reinforced with the random selection of features for each tree. Each tree is then independent of the other and the individual trees are not correlated.

5.4 Neural Networks

Neural Networks are unique in the sense that they can be used with supervised, and unsupervised learning and handle both regression and classification problems. Figure 5.6 lays out the basic structure of a Neural Network. Each model includes input, hidden, and output layers, all connected by "wires". The network is constructed by taking data from an input node and applying a weight (w_{ij}) as it travels across the wire to the hidden layers. Within the hidden layers, all weighted data is summed together along with the node's biases factor (a_i). The aggregated data is then passed through an activation function ($f(\theta)$) which allows the model to "add non-linearity" to our data [39]. Finally, the data leaves the hidden layer through another

Figure 5.6 Neural Network Properties [18]



wire and weight partnership before reaching an output node. The construction, or training, of the network is done one observation at a time, reassigning the wire weights after each iteration until an appropriate balance is found. To measure the improvement of each iteration, we try to minimize the loss function ($g(\theta)$).

When choosing an activation function, you must be aware of the vanishing gradient problem. As mentioned above, the neural network will update model weights based on the model's previous predictive performance. Backpropagation works by applying updates proportional to the partial derivative of the activation function [27]. This can be problematic for some activation functions, such as the Sigmoid function, as they have relatively small gradients. This does not allow for much movement among weight adjustments [1]. To overcome the restriction, many use the Hyperbolic Tangent, or Tanh, activation function (5.8). It is known for its speed and computational ease, but also has a larger gradient and providing a larger opportunity for the model to learn. The function suppresses our data and outputs values between [-1,1] allowing for standardization across nodes and layers.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5.8}$$

Binary Cross-Entropy, one of many loss functions, is commonly used in binary classifications [20]. The function (5.9) compares the expected class (0,1) to the predicted probability ($\hat{y}_i$) from the neural network. The function then calculates a score, penalizing the predicted probabilities based on how far from the expected class they are [23]. With each iteration of training the neural network, we adjust the weights to minimize this function.

$$Log\ Loss = \frac{1}{N}\sum_{i=1}^{N} -[y_i * \log(\hat{y}_i) + (1 - y_i) * \log\,(1 - \hat{y}_i)] \tag{5.9}$$

CHAPTER SIX

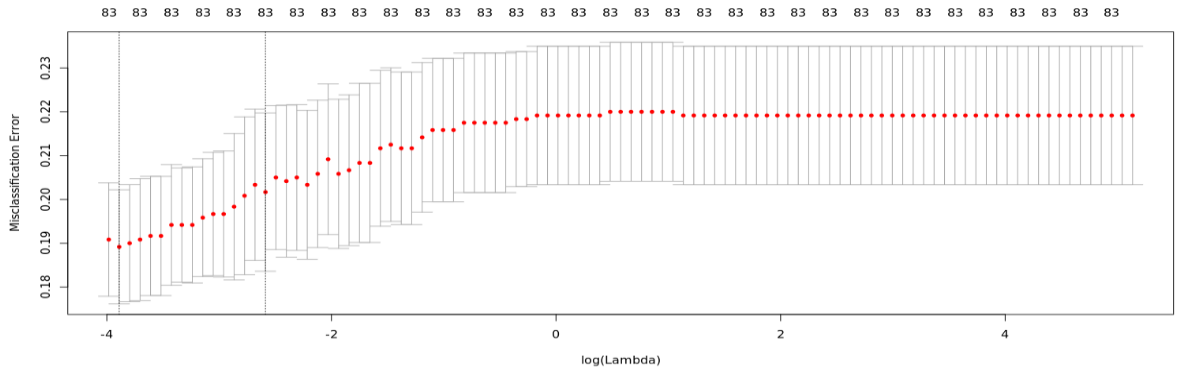
Model Analysis and Results

6.1 Logistic Regression

We started our analysis with the most basic model, logistic regression. This model will help set a baseline and reference point for the following models. Originally, all variables were included in the model, however after the initial run, it was apparent that there was some aliasing amongst the variables. We then removed 2 factors, Total Revenue and Total Liabilities & Equity, as these were being represented amongst other predictors. Then, in the second rendition of logistic regression, we looked for multicollinearity by the Variance Inflation Factor. The nature of earnings statements is that they revolve around “the bottom line” or your profitability as a company. This means many of the variables gathered from these statements show a different measure of profitability and can lead to correlation among variables. To combat this, I ran three different regularization models: Ridge, Lasso, and Elastic Net Regressions. Each of these three regularization models has a penalty parameter in the cost function in order to tune their respective model.

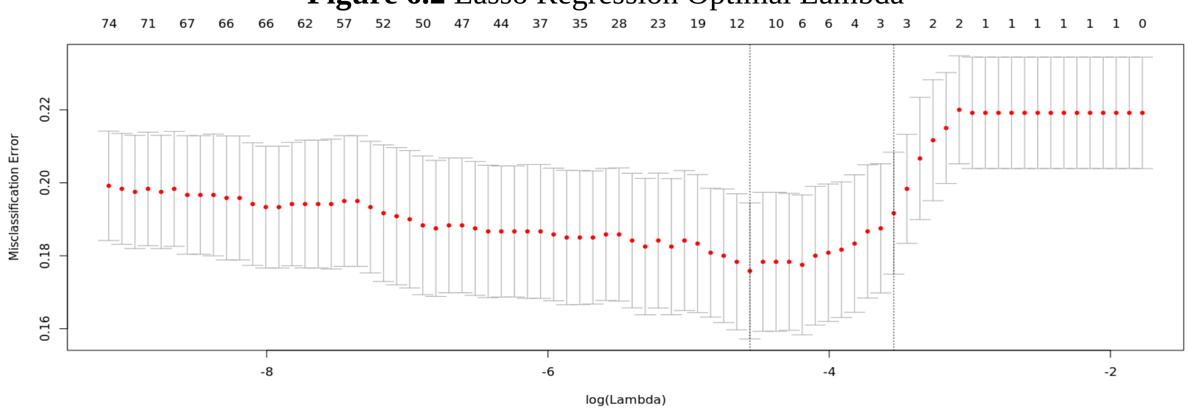
Starting with Ridge Regression, I performed a 10-fold cross-validation to find the optimal lambda for our tuning parameter. In Figure 6.1, we can see the dispersion of lambda values, in the end I chose $\lambda_{Ridge} = 0.02044961$ which minimizes percent deviance, or the error in cross-validation [11] at 0.1998. Our final ridge regression model, with our optimal lambda has an RMSE of 0.4113, an R square of 0.0115, and an error of 0.1692.

Figure 6.1 Ridge Regression Optimal Lambda



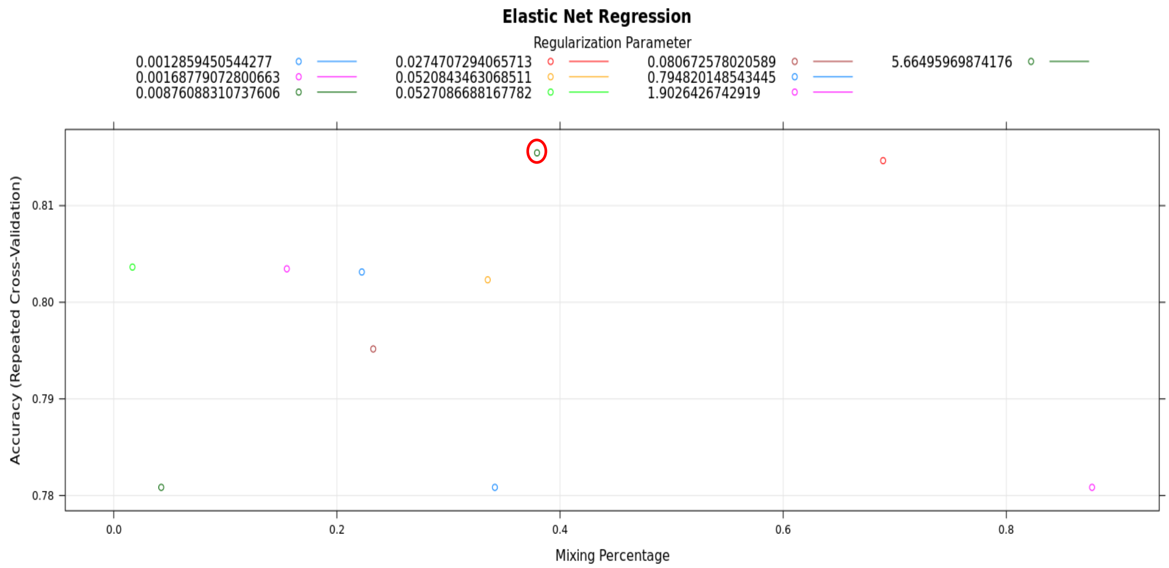
Next, we have Lasso Regression. Similar to Ridge Regression, I performed a 10-fold cross-validation to find the optimal lambda for our tuning parameter. In Figure 6.2, we can see the dispersion of lambda values, in the end I chose $\lambda_{Lasso} = 0.01041731$ which minimizes percent deviance, or the error in cross-validation [11] at 0.1998. Our final Lasso Regression model, with our optimal lambda has an RMSE of 0.4093, an R square of 0.0213, and an error of 0.1674.

Figure 6.2 Lasso Regression Optimal Lambda



Finally, we have Elastic Net Regression. After another 10-fold cross-validation, the model produced an optimal $\alpha = 0.3793799$ and $\lambda = 0.008760883$, resulting in an accuracy of 0.8154669. Figure 6.3 shows the other combinations of alpha and lambda, the chosen pair are circled. Our final elastic net regression model has an RMSE of 0.4113, and R square of 0.0115 and an error of 0.1692.

Figure 6.3 Elastic Net Regression Cross Validation



To compare across models, Table 6.4 shows the RMSE, R-squared and Error for the three types of regularization. As mentioned before, Elastic Net is a product of Ridge and Lasso "working together." Here is a great example of when one, in this case Ridge, dominates the overall model. However, Lasso has a higher R-squared and lower error then the other two making it the most optimal choice.

Figure 6.4 Comparison of Regularization Models

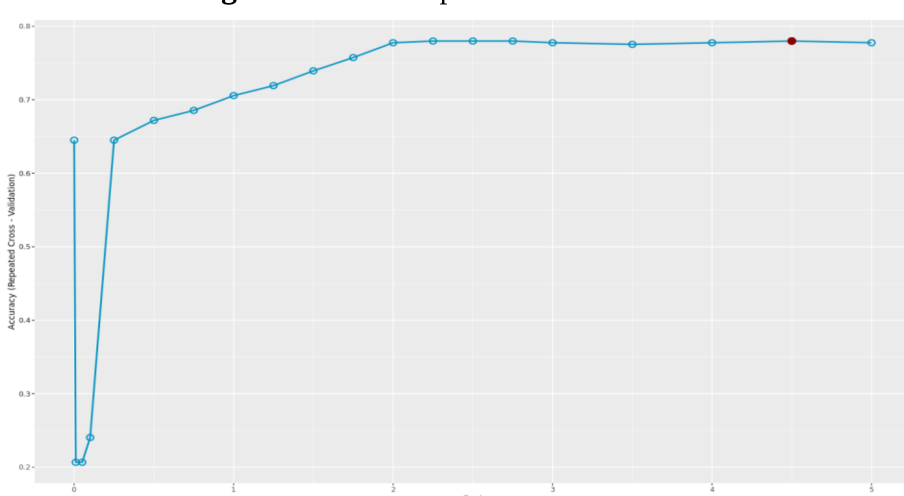
	RMSE	**R-squared**	**Error**
Ridge Regression	0.4112988	0.01148800	0.1691667
Lasso Regression	0.4092676	0.02122704	0.1675000
Elastic Net Regression	0.4112988	0.01148800	0.1691667

6.2 Support Vector Machine

When initially building the model, I chose to add a parameter that weighted the two classes. Based off the limiting number of significant events denoted in our dataset, it was important to weigh the outcomes to be forgiving that there are drastically less significant events to non-significant. I applied a weight of 0.8 to significant events and 0.2 to non-significant events.

Next, I added the radial kernel function to account for our data being nonlinear between groups. In an effort to avoid overfitting, I supplemented the model with a cost factor. This inherently transformed the boundary and margins created during SVM to move from “hard” to “soft” margins. This means that there will be some data from the opposing group within the margins in an attempt to create a stronger model for new data. I reran the SVM model created with a tuning parameter that looped through the cost parameter with values [0,5]. Figure 6.5 shows the tradeoff between the cost parameter and the accuracy of the model. To the right, we can see the peak in accuracy, denoted by the red dot.

Figure 6.5 SVM Optimal Cost Parameter



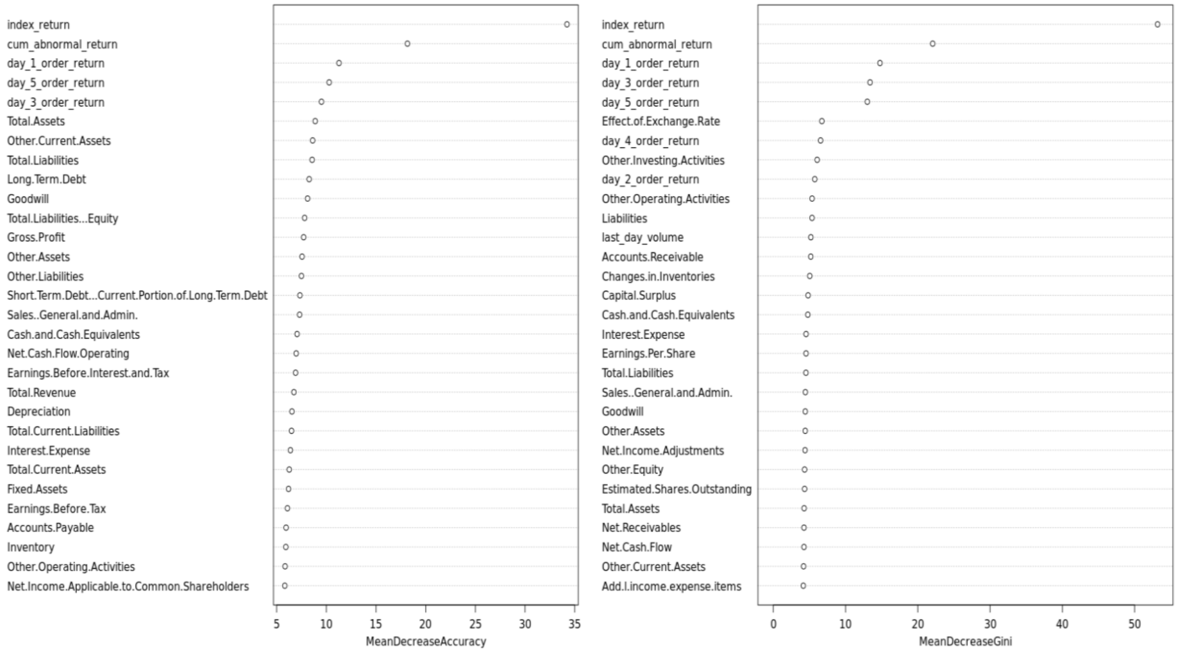
For this analysis, the optimal model was chosen based on which cost parameter yields the highest accuracy. Our final model has a cost parameter of 4.5 which gives us an accuracy measure of 0.7798. Looking at the confusion matrix of our testing data in our model, it seems that our SVM tends to lean more towards type 1 error, predicting that there will be an abnormal return when there is not one.

6.3 Random Forest

When building out your random forest model, you have the ability to tune the model with different parameters to increase its predictive power [19]. I chose to focus on the number of parameters needed to split at each node. To select the number of features to subset, the standard is to use the square root of the number of model parameters. I started with this value and pruned the number of parameters from there. For the best fit for our model, we use the number of estimators that gives the lowest "Out of Bag Error" (OOBE). In layman terms, this is the error of the data not used to create one of the trees. This allows for the model to be tested while it is being trained [25]. For our model, the lowest OOBE is 0.1591667, accompanied with 13 features.

Refitting the model with 13 features needed to split, we get an accuracy rate of 84.4%. This is an improvement from our original model with 9 features and an accuracy of 83.6%. Looking at the training data, we construct two charts that highlight the important factors in our model (Figure 6.6). The plot of the left shows the magnitude in which each variable contributes to the model's accuracy. The plot on the right shows each variable and their respective decrease in the Gini Index.

Figure 6.6 Random Forest Feature Selection



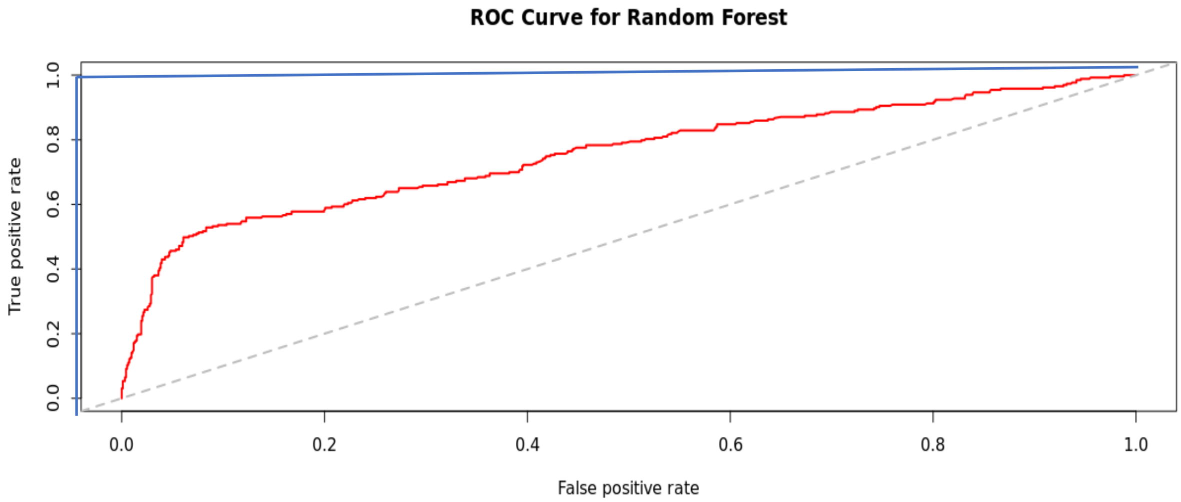
The Gini Index is a measure of the prediction probability of an element into a class [26]. Calculated by subtracting one from the sum of class probabilities. In building a tree, you want to decrease the Gini Index at each node. The chart on the right is depicting the variables that have the greatest influence on the Gini Index at each node.

$$Gini\ Index = 1 - \sum_{i=1}^{n} (P_i)^2$$

While in different orders, the top 5 important features, by both measures, include S&P 500 Index return, the Cumulative Abnormal Return, 1 day before announcement security return, 3 days before announcement security return and 5 days before announcement security return.

The Receiver Operating Characteristic (ROC) curve in Figure 6.7 shows the relationship between sensitivity and specificity captured by the model’s predictive power [4]. In a perfect model, the ROC curve would hug the top left of the plot, denoted by the blue line. However, our model’s ROC curve is much closer to the baseline, showing that its’ predictive power is not as strong. In this case, we see that the true positive rate increases at a much higher rate than the rate of false positives.

Figure 6.7 Random Forest ROC



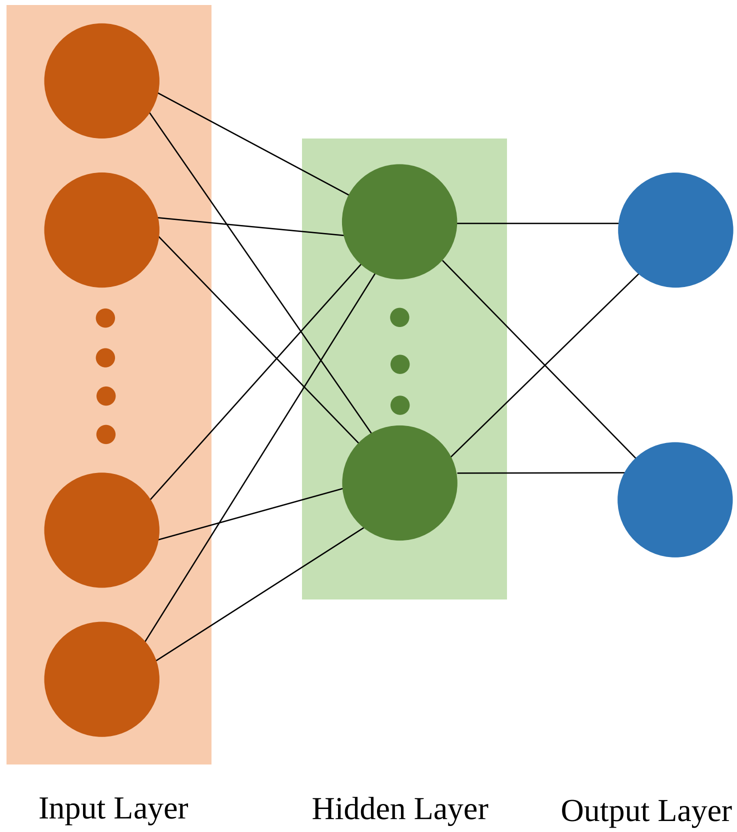
6.4 Neural Network

Our final model is a binary Artificial Neural Net. While there are many types of neural networks, this model applies best with our data as it best handles issues with heteroskedasticity which is often seen when evaluating stock price movement [24]. In current literature, there is no standard number of nodes or hidden layers for analysis. In order to find the best fit for our data, I looped through our network estimating with 1 through 60 nodes, from 1 hidden layer. The optimal number of nodes was determined by the highest prediction accuracy score. The construction of the model, displayed in Figure 6.8 contains 82 input nodes, 1 hidden layer with 20 nodes, and an output layer with two nodes for our binary output.

For the activation function, I chose to use tanh for its increased computational ease and increased power from having a higher gradient. Binary Cross-Entropy was the loss function that best fit our data and binary classification problem.

The output of our network gave us an RMSE of our network was 0.4205945 and a sensitivity rate of 81.6%

Figure 6.8 Artificial Neural Network



CHAPTER SEVEN

Conclusion

In this thesis, we evaluate numerous machine learning models in their ability to predict significant abnormal stock returns. Starting with logistic regression, we were able to apply three regularization techniques: Ridge, Lasso and Elastic Net that helped identify which factors, out of 84 are most influential in our prediction. Random Forest analysis showed similar factors to those in the regularization techniques. Support Vector Machines and Artificial Neural Networks were imperative in applying deep learning methods that allow us to apply nonlinearity to our data through kernel and activation functions, letting our models find stronger relationships among our factors. In the end, the best performing model was a Lasso logistic regression, which gave us the smallest RMSE. This model's high performance came from its feature selection, keeping only influential data for regression.

Stock prices have the reputation of being impossible to predict. This theory started with idea of a Random Walk by Burton Malkiel [24]. The idea is that every day, as new information arises, stock prices move in an unpredictable way. Given SEC oversight, we are able to anticipate large company events that have the potential to impact their price in the market. These yearly and quarterly earnings statements provide a lens into a company and their potential for economic growth. By evaluating price movement following an earnings statement, we can make predictions on how a security may behave following the next earnings release. The ability to understand their relationship will lead investors to make a profit from their trading strategy.

REFERENCES

[1] Antoniadis, P. (2022, February 28). Activation functions: Sigmoid vs Tanh. Retrieved June 2, 2022, from https://www.baeldung.com/cs/sigmoid-vs-tanh-functions#:~:text=This%20means%20that%20using%20the,use%20the%20tanh%20activation%20function.

[2] Ben Lobel, P. (2021, June 1). Trading earnings season: 3 steps for using earnings reports. Retrieved June 2, 2022, from https://www.dailyfx.com/education/understanding-the-stock-market/how-to-trade-earnings-season.html

[3] Brunnermeier, P. K. (2022). Value Relevance of Analysts' Earnings Forecasts. Lecture. Retrieved June 2, 2022, from https://www.princeton.edu/~markus/teaching/Eco467/04Lecture/04Event%20Study%20Description.pdf

[4] Chan, C. (2020, December 09). What is a ROC curve and how to interpret it. Retrieved June 2,2022, from https://www.displayr.com/what-is-a-roc-curve-how-to-interpret-it/#:~:text=The%20ROC%20curve%20shows%20the,diagonal%20(FPR%20%3D%20TP).

[5] Chen, J. (2021, November 05). Understanding the theory of mean reversion. Retrieved June 2, 2022, from https://www.investopedia.com/terms/m/meanreversion.asp

[6] Dwivedi, R. (2021, January 28). How does support Vector Machine (SVM) algorithm works in machine learning? Retrieved June 2, 2022, from

https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning

[7] Gold, I. (2019). [The Three Forms of the Efficient Market Hypothesis]. Retrieved 2022, from https://www.ig.com/en/trading-strategies/what-is-the-efficient-market-hypothesis--emh---191217#weak

[8] Inc, W. (2022, April 29). Efficient market hypothesis (EMH). Retrieved June 2, 2022, from https://www.wallstreetprep.com/knowledge/efficient-market-hypothesis-emh/

[9] Institute, C. F. (2022, January 22). Efficient Markets hypothesis. Retrieved June 2, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/efficient-markets-hypothesis/

[10] Khotari, S., & Warner, J. B. (2016). "Econometrics of Event Studies". Retrieved May 20, 2022, from https://www.bu.edu/econ/files/2011/01/KothariWarner2.pdf

[11] Jackson, S. (2017, April 10). How and when: Ridge regression with glmnet • blogr. Retrieved June 2, 2022, from https://drsimonj.svbtle.com/ridge-regression-with-glmnet

[12] Jain, V. (2019, December 30). Everything you need to know about "activation functions" in Deep learning models. Retrieved June 2, 2022, from https://towardsdatascience.com/everything-you-need-to-know-about-activation-functions-in-deep-learning-models-84ba9f82c253

[13] Likos, P. (2021). How earnings affect stock prices | investing 101 | US news. Retrieved June 2, 2022, from https://money.usnews.com/investing/investing-101/articles/how-earnings-affect-stock-prices

[14] Mahanta, J. (2017, July 12). Introduction to neural networks, advantages and applications. Retrieved June 2, 2022, from https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207

[15] Maverick, J. (2022, May 26). Why do shareholders need financial statements? Retrieved June 2, 2022, from https://www.investopedia.com/ask/answers/032615/why-do-shareholders-need-financial-statements.asp#:~:text=Financial%20statements%20are%20important%20to,and%20long%2Dterm%20financial%20obligations

[16] Mackinlay, A. C. (1997). Event Studies in Economics and Finance. Journal of Economic Literature, 35(1), 13-39. Retrieved May 20, 2022, from http://www.jstor.org/stable/2729691

[17] Neuhierl, Andreas and Scherbina, Anna D. and Schlusche, Bernd, Market Reaction to Corporate Press Releases (March 10, 2010). Available at SSRN: https://ssrn.com/abstract=1556532 or http://dx.doi.org/10.2139/ssrn.1556532

[18] Neural Network Diagram. (n.d.). Retrieved 2022, from https://www.mdpi.com/applsci/applsci-11-00217/article_deploy/html/images/applsci-11-00217-g003.png

[19] R, S. E. (2021, June 24). Random Forest: Introduction to random forest algorithm. Retrieved June 2, 2022, from https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

[20] S. (2021, September 28). An introduction to neural network loss functions. Retrieved June 2, 2022, from https://programmathically.com/an-introduction-to-neural-network-loss-functions/

[21] Simpson, J. (ed.). (2017). Oxford English Dictionary (3rd ed.). Oxford: Oxford University Press.

[22] SoFi. (2022, May 10). A brief history of the stock market. Retrieved from https://www.sofi.com/learn/content/history-of-the-stock-market/

[23] Saxena, S. (2021, March 03). Binary cross entropy/log loss for binary classification. Retrieved June 2, 2022, from https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification/#:~:text=Binary%20cross%20entropy%20compares%20each,far%20from%20the%20actual%20value.

[24] Smith, T. (2022, February 08). Random walk theory. Retrieved June 2, 2022, from https://www.investopedia.com/terms/r/randomwalktheory.asp

[25] T. Hastie, R. Tibshirani and J. Friedman, “Elements of Statistical Learning Ed. 2”, p592-593, Springer, 2009.

[26] Tyagi, N. (2020, September 30). Understanding the gini index and information gain in decision trees. Retrieved June 2, 2022, from https://medium.com/analytics-steps/understanding-the-gini-index-and-information-gain-in-decision-trees-ab4720518ba8#:~:text=You%20can%20learn%20another%20tree,classified%20for%20a%20distinct%20class.

[27] Vanishing gradient problem. (2019, May 6). In Wikipedia. https://en.wikipedia.org/wiki/Vanishing_gradient_problem

[28] Walia, A. S. (2017, August 7). Radial kernel support vector classifier. Retrieved June 2, 2022, from https://datascienceplus.com/radial-kernel-support-vector-classifier/

[29] Towardsdatascience.com. (2019). *SVM Projection*. Retrieved 2022, from https://laptrinhx.com/demystifying-support-vector-machines-401343508/

[30] Sharma, S. (2019). *SVM Projections*. Retrieved 2022, from https://medium.com/analytics-vidhya/how-to-classify-non-linear-data-to-linear-data-bb2df1a6b781