CHALMERS

The impact of press releases on stock prices

Master's thesis in Complex Adaptive Systems

VICTOR EKDAHL

Department of Applied Mechanics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2015

MASTER'S THESIS IN COMPLEX ADAPTIVE SYSTEMS

The impact of press releases on stock prices

VICTOR EKDAHL

Department of Applied Mechanics
Division of Vehicle Engineering and Autonomous Systems
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2015

The impact of press releases on stock prices
VICTOR EKDAHL

Master's thesis 2015:24
ISSN 1652-8557
Department of Applied Mechanics
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
A word cloud in the shape of a dollar sign.

Chalmers Reproservice
Göteborg, Sweden 2015

The impact of press releases on stock prices
Master's thesis in Complex Adaptive Systems
VICTOR EKDAHL
Department of Applied Mechanics
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology

Abstract

Press releases are found to be important events that represent a potential explanation for up to 24% of the major stock price movements. The classifiers are able to predict price movements larger than 3% with up to 60% precision. The event study confirms that press releases have a statistically significant effect on stock prices on the first day. Trading strategies were defined and shown to be viable for the period tested, but the post-event patterns that motivate them are statistically insignificant.

To better understand stock price movements, and the ability to forecast them, press releases from companies on the New York Stock Exchange are used as data. Classifiers are applied to predict whether the stock price will react strongly to a press release or not. An event study is done to investigate patterns prior and subsequent to a press release being published. Algorithmic trading strategies are defined and tested based on the results.

Keywords: prediction, event study, news analytics, stock returns

Contents

Abstract **i**

Contents **iii**

1 Introduction **1**
1.1 Purpose . . . 1
1.2 Limitations . . . 1
1.3 Acknowledgments . . . 1

2 Literature Review **2**
2.1 An efficient market . . . 2
2.2 Types of news . . . 2
2.3 Finding the sentiment . . . 3
2.4 Lexica . . . 3
2.5 Applications . . . 3

3 Data collection **5**
3.1 Stocks and press releases . . . 5
3.1.1 Sectors and market capitalization . . . 5
3.2 The data time span . . . 6
3.3 Major price movements . . . 6
3.4 Matching between press releases and abnormal returns . . . 7
3.4.1 Initial classification of press releases . . . 8
3.5 Text processing and lexica . . . 8
3.5.1 Lexicon . . . 8
3.5.2 Annotated dictionaries . . . 9
3.5.3 Further work: Contractions and negations . . . 9
3.6 Summary . . . 9

4 Classification and event study procedures **10**
4.1 Classification . . . 10
4.1.1 Pre-processing . . . 10
4.1.2 Naive classifier . . . 10
4.1.3 Vector distance classifier . . . 11
4.1.4 Bayesian classifier . . . 11
4.1.5 Evaluation of a classifier . . . 12
4.2 Event study methodology . . . 12
4.2.1 Event definition . . . 12
4.2.2 Cumulative abnormal returns . . . 12
4.2.3 The test statistic . . . 13
4.3 Backtesting . . . 14
4.3.1 Evaluation of a backtest . . . 14
4.3.2 Position sizing technique . . . 14
4.3.3 Time frame . . . 14
4.3.4 Trading on predicted large movements . . . 15
4.3.5 Trading on post-event patterns . . . 15

5 Results **16**
5.1 Prediction of major price movements . . . 16
5.1.1 Naive classifier . . . 16
5.1.2 Vector distance classifier . . . 17
5.1.3 Bayesian classifier . . . 17
5.1.4 All results . . . 18
5.1.5 Impact of market capitalization on classification . . . 18

5.2 Event analysis of large stock price changes . . . 18
5.2.1 Categorization by market capitalization . . . 19

6 Trading strategies **20**
6.1 Trading on movements predicted to be large . . . 20
6.2 Trading on post-event patterns . . . 20

7 Discussion and conclusion **22**
7.1 Discussion . . . 22
7.2 Conclusion . . . 22

References **23**

1 Introduction

The ability to forecast a property of an asset is useful in several contexts, whether the property is future returns, volatility, or something else. The most obvious usage is that of investors and traders who attempt to understand and predict future price movements to make a profit. Several factors impact a stock price, simple pricing models assume the overall market return to have the most explanatory value. When firm-specific events occur the return diverges from the expectation given by the models. Finding an explanation for an excess, or abnormal, return is far more difficult as there are multiple potential sources to which it can be attributed. New information about a company can trigger an update of the market value in the models of the market participants. Phenomena in behavioral finance explain other abnormalities. The study of events that impact returns increases the understanding of stock price behavior.

Companies make announcements to the public through press releases. Sometimes the information delivered in a press release is important enough to lead to a significant price change. A press release is therefore interesting to analyze as an event. All quoted companies issue press releases, and there are thousands of them. The potential benefit of applying news analytics on press releases is sizable, the possible applications are many. No individual can take in all new information from the stock market, but an automated process could suggest material that is important for the person to read. That is if it is practicable to separate the important press releases from the rest. News that cause large price changes might affect the stock price behavior several days thereafter. If post-event patterns can be identified in the case of press releases that knowledge could be useful for an investor either at its own or as part of a bigger puzzle. One can also envision fully automated systems in the form of algorithmic trading where strategies expressed as algorithms are automatically executed, processing the new information as it arrives.

News analysis can be used on a text to extract its sentiment and quantify its attributes. Different techniques are appropriate depending on the sort of text. A variety of classifiers can be used to divide texts into classes. A classifier can measure some property of the text directly such as the number of positive or negative words. It can also be based on machine learning where other texts are used as a training set. Trends subsequent to a press release can be analyzed in an event study. Through summation of abnormal returns over several days after an event drifts in the stock price following a press release become apparent and the statistical significance can be assessed. To evaluate the usefulness of the results in an algorithmic trading application backtesting can be performed. In a backtest historical data are used to calculate the returns that the portfolio would have generated if it was trading during the specified period.

1.1 Purpose

The purpose of this thesis is to further the understanding of the impact of press releases on stock prices, in particular to investigate the ability to predict major price movements and subsequent returns. Those are two separate questions approached with different methods. The first is: Given a press release, to what extent can a major price movement be predicted? The second is: Given a major price movement caused by a press release, what can be said about stock price changes in the near future? It is also investigated how well the results translate into algorithmic trading strategies.

1.2 Limitations

The companies considered are those listed on the New York Stock Exchange. The last thirteen press releases are available for each stock. Thus there are restrictions both in scope and time.

1.3 Acknowledgments

Thanks to the supervisor professor Mattias Wahde for the inspiration and guidance. Thanks to Calle Ekdahl for writing the jSoupLink package which allows jSoup to be used from within Mathematica, simplifying the data collection.

2 Literature Review

The feasibility of predicting future stock returns from past stock price data has been thoroughly researched. Starting with, among others, Fama (1965) who concluded that it is pointless for an investor to look for patterns in historical prices. However, news do have an impact. The stock price reflects the value of the company and therefore reacts to new information and knowledge that affects future earnings. Niederhoffer (1971) was one of the first to study the connection between news and stock returns and found that large movements in the market was more likely on days following world events, defined as five- to eight-column headlines in the *New York Times*. An initial overreaction to extremely bad events was another observed tendency, characterized by a rise after the initial fall. Ryan and Taffler (2002) found that 65% of large stock price changes for companies could be matched to firm-related news events. Interim results, financing issues, bids, and analysts' reports are some news categories that were shown to cause share price movements. Such results indicate that there may be a predictive power in news analytics.

2.1 An efficient market

It is established that stock prices react to news, but there are differing ideas on how fast news are taken into account in the stock price. The efficient-market hypothesis, defined in its different forms by Fama (1970), claims that the price of an asset fully reflects all available information and that asset prices swiftly, or instantly, adapt to new information. If it is true it would mean that even though an investor had the tools to make predictions there would be no time frame in which she or he could use the information to trade. Prediction of future stock prices would be impossible.

Chan (2003) separates stocks that have news stories associated with them during the last month from stocks that do not, he then measures the abnormal return for both groups during several months and compare the results. He finds that stocks afflicted with bad news experience a drift in the following months, which he interprets as the stock prices being slow to reflect the new information. He also finds that stocks without news show a tendency of reversals in the next month. Engelberg, Reed and Ringenberg (2010) attempt to explain why short sellers are successful by combining trading histories with a news database, concluding that short sellers get an edge by analyzing available information. Such patterns suggest that the market might not be entirely efficient.

Antweiler and Frank (2006) used a large set of news stories from Wall Street Journal to define events which they study. They note that the event window for such studies usually covers a few days before and after an event, and uses larger event windows themselves of up to forty days to show that a few days do not capture the entire reaction. They find a tendency for drift opposite to the initial change for an extended period of time after an event. They observe that their data rejects the efficient market hypothesis.

Initial over- and underreactions that cause asset prices to display drifts or reversals could have many explanations. Some can be found in behavioral finance which deals with market inefficiencies as a result of systematic errors from market participants. Barber and Odean (2004) show how individual investors have a tendency to buy stocks that have received attention, either through news or through extreme returns or volumes during a day. They claim that this behavior puts a pressure on the price, which explains why individual investors underperform over time when trading in this way.

2.2 Types of news

News analytics appear to be a useful tool, but news come in different forms. Leinweber and Sisk (2011) define three categories to which a news source can be sorted: News, pre-news and social media. News refers to the traditional sources such as television, radio, papers and so on. Pre-news means the material that the media outlets use as a basis for what they publish, for example it can be required filings or other information released from firms. Social media are blogs, message boards, and similar pages.

Mittermayer and Knolmayer (2006) present a system called NewsCATS which takes press releases as input and attempts to predict stock price changes, intraday. They note that press releases "have far more impact on stock prices than other news". An advantage with press releases is that they rarely are so called me-too stories which do nothing but repeat already published news.

2.3 Finding the sentiment

In addition to a wide variety of choices in the data collection process there are several methods of attaching a sentiment score to a news story. Chen and Das (2007) use five different classifiers to determine whether to label posts on the Yahoo message board as a buy, hold or sell. A vector distance classifier counts the number of appearances of each word from a lexicon to convert a news story to a word vector. Using pre-trained word vectors the news story is then classified as whatever it is closest to. An adjective-adverb phrase classifier looks for phrases within a news story that contain adjectives or adverbs, the story is classified depending on if those phrases are deemed positive or negative. A Bayesian classifier uses Bayes' theorem to compute the conditional probability that a message belongs to a certain category, by looking at how frequent different words are in the categories. The message is then assigned to the category that boasts the highest probability.

Despite their apparent simplicity, these classifiers work. Tetlock, Saar-Tsechansky and Macskassy (2008) gathered 350 000 news stories from Wall Street Journal and Dow Jones News Service and counted the number of negative and positive words in each story. They found that the fraction of negative words could be used for prediction, because of an underreaction to bad news. It should be noted that the profits from the trading strategy they designed were so small that the costs of using it would lead to a loss. This is not unexpected and, as another example, NewsCATS has the same problem.

Luss and d'Aspremont (2009) applies another technique for classification called a support vector machine (SVM). They note that it is well established that SVM perform better than, for example, Bayesian classifiers. As data they use press releases from PRNewswire. An event starts when there is a press release, and they have two goals: One is to predict whether the stock price will increase or decrease. The other is to predict abnormal returns, large movements. An event lasts between 0 and 250 minutes. The SVM is also fed with historical prices from the time of the event. The performance of the forecast when using text and when using returns is then compared. They find that the direction of the movement is not possible to predict. However, it is possible to predict whether there will be an abnormal return or not. It is shown that texts produce much better results than past returns. This is true for large abnormal returns, the accuracy of text prediction decreases with the size of the return. Using multiple kernels forecast further improved, by combining both text and returns as input. Also presented is an example of a trading strategy that bets on the occurrence of an abnormal return and makes a profit. As usual without the transaction costs taken into account.

2.4 Lexica

Not all words are equally important, some contain more information about the sentiment of a text than others. The classifiers specifically look for words of high value, stored in lexica. The decision of which lexicon to use affects the outcome of the classification. Some lexica are annotated, meaning that the words are tagged with properties such as positive or negative. Tetlock, Saar-Tsechansky and Macskassy (2008) used the Harvard-IV-4 psychosocial dictionary to identify positive and negative words. Depending on the sort of news story that is interpreted the meaning of a word may be different. While some words maintain their meaning in all contexts, others do not. In finance a word might convey a sentiment that it usually would not. Peramunetilleke and Wong (2002) received lists of words from a domain expert and used these to analyze headlines in order to predict changes in currency exchange rates. Chen and Das (2007) and Luss and d'Aspremont (2009) also used lexica that were manually put together for the purpose. Another option is to create the lexicon with the help of an algorithm that identifies suitable words. Lee, Pang and Vaithyanathan (2002) attempt to classify movie reviews and use three different lexica, two of them created by graduate students and the third through examination of the data. The third list gave the best results.

2.5 Applications

There have been many successful attempts at forecasting by analyzing news. There is more to be done however, due to the complexity of the problem and the many alternatives. What questions should be asked? What source of news, which time frame, what classifiers, and which lexicon should be chosen? Many of the decisions are taken in the data collection phase, as the data suit certain techniques more than others.

One common and very direct way of evaluating results is to use them as basis for a trading strategy. It has been exemplified how such trading strategies can be profitable, even when the methods for finding the opportunities are simple. However, transaction costs appear hard to beat.

News analytics are not limited to tasks such as forecasting of stock prices or risk management. Mitra and Mitra (2011) review applications and list several creative usages of news analytics, such as: Market surveillance for automated detection of insider trading, circuit breakers to halt trading algorithms when news arrive, and stock screening to find interesting assets. They note that news analytics are gaining acceptance, and that its use and importance in finance will likely continue to grow.

3 Data collection

As material for the analysis press releases were collected from companies on the New York Stock Exchange (NYSE). The data were transformed into a suitable format and used to create the lexicon required by the classifiers. The following sections describe that process and explore basic properties of the retrieved content.

3.1 Stocks and press releases

All financial data were obtained with Mathematica. A list of all tickers on NYSE was created. Tickers for which the historical prices were not available or partly missing during the period 4 February, 2014 to 4 March, 2015 were removed. If information about market capitalization or sector was unavailable the ticker was excluded as well. The resulting list contained the tickers of 1716 stocks.

The tickers and their corresponding fundamentals were stored in a database. Two examples can be seen in table 3.1. The fourth field contains a list of dates for which there was a major price movement. The criteria of such an event will be defined later.

Table 3.1: Five examples from the stock database.

Ticker	Market capitalization	Sector	Major movements
VALE	3.9063×10^{10}	IndustrialMetalsAndMinerals	Dates
LEG	6.036×10^{9}	HomeFurnishingsAndFixtures	Dates
BLL	9.057×10^{9}	PackagingAndContainers	Dates
HL	1.238×10^{9}	Silver	Dates
WAB	8.22×10^{9}	Railroads	Dates

Web scraping was used on Yahoo! Finance to download the press releases page for all stocks in the database. From each such page a list of links to recent press releases was extracted. Web scraping was then used to retrieve the content, which was processed and stored as in table 3.2 where two examples are shown. At least one press release could be retrieved from 1672 stocks. For 44 tickers press releases were unavailable, and they are therefore not represented in the press releases database. In total 19952 press releases were downloaded, or an average of 11.9 press releases per stock.

Table 3.2: Five examples from the press releases database.

Ticker	Class	Timestamp	Title	Content
VALE	{0,0}	{2014, 2, 21, 14, 30, 0.}	Text	Text
LEG	{0,0}	{2015, 1, 16, 9, 3, 0.}	Text	Text
BLL	{0,1}	{2015, 2, 4, 16, 7, 0.}	Text	Text
HL	{1,1}	{2014, 10, 21, 18, 20, 0.}	Text	Text
WAB	{0,0}	{2014, 7, 10, 8, 16, 0.}	Text	Text

The second field shows the initial classification, described later. The timestamps are in eastern time (ET) and include precision down to minutes. The dates of all press releases published after the market close at 4 p.m. ET are assigned the date of the next business day when appropriate.

3.1.1 Sectors and market capitalization

All stocks belong to a sector. There are 203 different sectors, which means that on average there are 8.5 companies per sector. The five largest sectors are shown in Table 3.3.

The market capitalizations were retrieved and stored on February 4, 2015. The five largest companies are shown in table 3.4. Note that both Berkshire Hathaway and Wells Fargo & Co are missing. This highlights the fact that the database is not a complete collection of all companies on NYSE. As mentioned this is due to missing data.

Table 3.3: *The largest sectors.*

Sector	Number
Independent oil and gas	55
Electric utilities	44
Property and casualty insurance	43
Oil and gas drilling and exploration	36
Business services	33

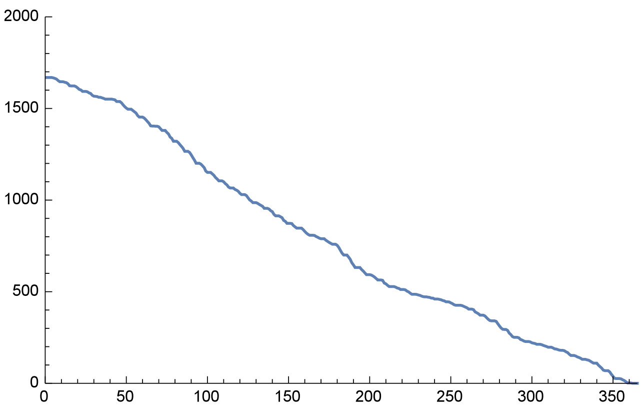
Table 3.4: *The largest companies.*

Stock	Market capitalization ($)
Exxon Mobil Corporation	3.873×10^{11}
Johnson & Johnson	2.837×10^{11}
Wal-Mart Stores Inc	2.793×10^{11}
HDFC Bank Ltd	2.789×10^{11}
China Mobile Ltd.	2.771×10^{11}

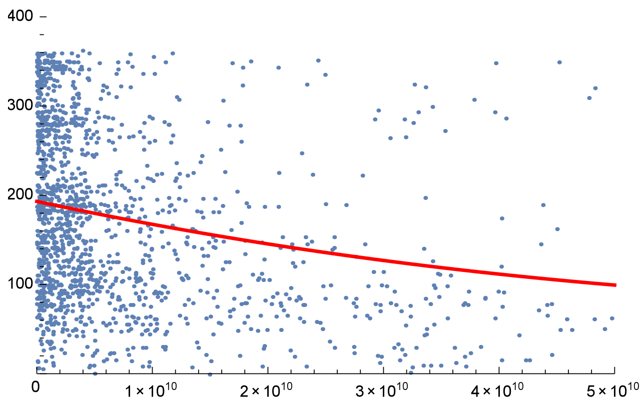
3.2 The data time span

Yahoo! Finance limits the amount of press releases provided in two ways. There are at most 13 press releases per stock and no press release is older than a year. Therefore the time period for which there is data varies between stocks from a few days to a full year.

Figure 3.1a shows the number of stocks for which there is data as a function of days. The newsflow varies widely between companies; some publish press releases rarely and others frequently. Therefore the available press releases do not cover the same amount of days for each stock. Large companies have a larger newsflow on average as can be seen in figure 3.1b where the time span covered for a stock is plotted as a function of the market capitalization. Less days covered means press releases being published more frequently, as the first of the thirteen available press releases then does not go as far back in time. Each dot represents a company and the variation is large even when the market capitalizations are similar.



(a) *Stocks covered versus days.*



(b) *Time span versus market capitalization.*

Figure 3.1: *The data time span.*

The web scraping was done February 4, 2015. Therefore no press releases older than February 4, 2014 exist in the database used here. For this period plus another month the closing prices for all stocks in the database were saved. Stock price histories obtained from Mathematica account for dividends and splits.

3.3 Major price movements

In order to predict if a press release will cause a large price change one has to define what a *large* price change is. Ryan and Taffler (2002) calculate the market-adjusted returns and define a threshold for when they can be considered major price movements. The idea is that a large abnormal return implies an unexplained change in the stock price that is possibly due to firm-specific news.

Abnormal returns can be defined and computed in different ways. Ryan and Taffler (2002) define the abnormal return for firm i at day t as

$$\mathrm{AR}_{i,t} = R_{i,t} - \mathrm{E}[R_{i,t}],$$

where

$$E[R_{i,t}] = \beta_i R_{m,t}.$$

$R_{m,t}$ is the market return on day t. The β for firm i, β_i, was computed as

$$\beta_i = \frac{\mathrm{Cov}(R_i,R_m)}{\mathrm{Var}(R_m)}$$

for the one year period mentioned before. E is the expectation value, Cov is the covariance and Var is the variance. The market index used was ˆNYA which is the ticker for NYSE Composite, an index that includes most stocks on NYSE. Returns were calculated as

$$R_{i,t} = \log(P_{i,t} + D_{i,t}) - \log(P_{i,t-1})$$

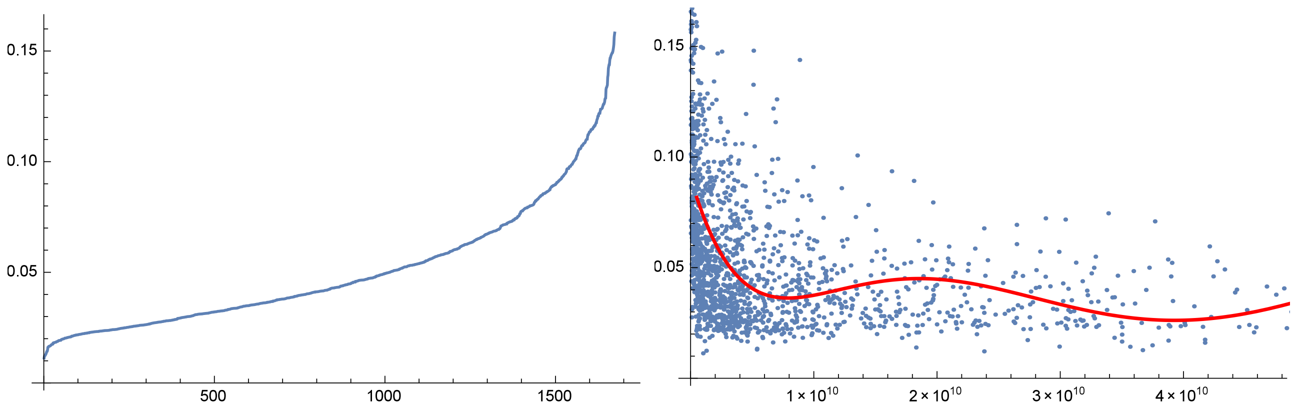
where $P_{i,t}$ is the price of stock i at time t. $D_{i,t}$ is the dividend which as mentioned is accounted for in the price histories.

An abnormal return is defined as a major price movement if it is above or below the average abnormal return by more than two standard deviations. If the log returns are normally distributed with an expected value of 0 then the expected number of major price movements per stock and year is

$$2(1 - \Phi(2))254 = 11.6,$$

where 254 is the number of days that the market was open and Φ is the cumulative distribution function for the normal distribution. The dates of all major price movements were stored in the stock database. The actual average number is indeed 11.6 per stock.

The distribution of the average return per stock on days with major price movements is shown in figure 3.2a. It varies from one percent to more than fifteen. It can be seen in figure 3.2b that the size of the average return tends to be smaller for larger companies.



(a) *Distribution of abnormal returns.* (b) *Abnormal returns versus market capitalization.*

Figure 3.2: *Major price movements.*

3.4 Matching between press releases and abnormal returns

With two lists of dates, one for press releases and one for large abnormal returns, it is possible to compare and determine to what extent they overlap. The first row of table 3.5 shows the fraction of press releases that are published on a day with an abnormal return. It also shows the fraction of abnormal returns that can be

Table 3.5: *Matching between abnormal events and press releases.*

	Press releases	Abnormal returns
Data	13.42%	24.24%
Random	5.95%	11.52%

connected to a press release. The second row shows the same measurements but with the dates of the press releases replaced by random dates in the same interval. Note that the interval varies depending on how far back the available press releases go for a specific stock.

If there is a press release published on the same day as there is an abnormal return it is either because the press release is the cause of the price change, or it is a coincidence. If p is the fraction of abnormal returns that are explained by a press release and it is assumed that a coincidence happens randomly then

$$(1-p)0.1152 + p = 0.2424 \Rightarrow p = 0.1437.$$

Thus, around 14.4% of the abnormal returns were caused by a press release.

3.4.1 Initial classification of press releases

Each press release is assigned -1 or 1 if it is published on a day where the stock has a large negative or positive abnormal return. The press release is assigned 0 if there is no major price movement. There are 1255 negative press releases, 1419 positive press releases and 17274 press releases that are not connected to an abnormal return. A press release assigned -1 or 1 will be referred to as an abnormal event. The initial classification is used to evaluate the results from the classifiers, and in some cases to train them. Note that the classification is not completely accurate since some of the press releases are matched to an abnormal return by chance. An interesting difference between the classes is that the average length of a press release classified as -1 or 1 is 1433 words, compared to 706 words for 0, an indication that press releases that cause major price movements are, in general, longer.

Another possible definition for an initial classification is a threshold. An alternative classification was done where an initial classification of 1 or -1 was given to press releases that caused returns larger than 3% or smaller than -3%.

3.5 Text processing and lexica

Most classifiers are not affected by the order of the words in a text. Instead, the text is favorably represented as a list of unique words and their multiplicity. This sort of representation is called the bag-of-words model. All words in each press release are put through the Porter stemmer in Mathematica. A stemmer reduces all forms of a word into its stem, so that they can be identified as the same. As an example, *improvement* is reduced to *improv*.

3.5.1 Lexicon

A lexicon is a list of words that indicate a sentiment and thus are helpful in the classification. Most classifiers need a lexicon as it limits what words they look for.

A random sample of 1000 press releases was extracted. All press releases classified as causing an abnormal return were selected, as was an equal number of the remaining press releases. A list of all unique words in any of these press releases was assembled. The task is to determine what words to include in the lexicon.

The fundamental property of a word helpful in the classification is that it occurs more or less frequently in one class compared to the others. To measure this property each word w is scored as

$$F(w) = 1 - \frac{\min(\mathrm{n_{w,a}}, \mathrm{n_{w,n}})}{\max(\mathrm{n_{w,a}}, \mathrm{n_{w,n}})}$$

where $n_{w,c}$ is the number of times that word w appear in class c. There are two classes, a for abnormal return and n for a normal return. n_w which is the total number of occurrences in all classes was also counted. A score of 0 corresponds to the word being equally frequent in both classes. A score of 1 means that the word only appears in one class. Approximately 20000 words were scored, a few samples can be seen in table 3.6.

Table 3.6: Scored lexicon words.

Word	$F(w)$	n_w
after-tax	0.863	83
solicitat	0.571	30
limitat	0.460	37
discontinu	0.980	50
asset	0.660	317

Note that the words have been stemmed. To reduce the large list of words into a useful lexicon certain criteria were established. Only words for which $0.4 < F(w) < 1$ and $n_w > 3$ were kept. The lexicon contains 1519 words.

3.5.2 Annotated dictionaries

There are dictionaries and lexica in which some words have been marked as positive or negative. The General Inquirer categories Positiv and Negativ contain 1915 and 2291 words, respectively. They were collected from several sources, including the Harvard IV-4 dictionary and the Lasswell value dictionary. Hu and Liu (2004) have compiled a sentiment lexicon of positive and negative words, partly for social media content. Both these alternatives will be used with the naive classifier.

3.5.3 Further work: Contractions and negations

Replacing contractions with their full form requires a list of common contractions. As of now, contractions are split into two words at the apostrophe. Negations can reverse the meaning of a sentence. Words following a negation could be tagged as negated with a list of common negations. It was not deemed necessary for this application, but could improve the results.

3.6 Summary

There are six databases in total. The stock database contains the ticker, market capitalization and sector for 1716 companies. It also provides a list of dates where there were large abnormal returns. The press release database contains information about 19952 press releases including day of publication, title and content plus an initial classification. A price history database provides the price histories for the period 4 February, 2014 to 4 March, 2015 for all companies in the stock database. The lexicon is a list of words that are useful as indicators when classifying. The General Inquirer categories Positiv and Negativ plus the sentiment lexicon by Hu and Liu (2004) are lists of words annotated either positive or negative.

4 Classification and event study procedures

There are two main questions; the first is: Given a press release, to what extent can a major price movement be predicted? The second is: Given a major price movement caused by a press release, can anything be said about how the stock price will change in the near future?

As all press releases have been assigned -1, 0, or 1 defining them as either negative, neutral, or positive the first question is a classification problem. The challenge lies in classifying press releases with no other information than the content of the press release. The results can be assessed by comparing the classification given by the algorithms to the initial classification.

The second question requires an analysis of the total impact of a press release on the stock price. The impact can be measured through an event study. The objective is to find potential drifts or reversals in the stock price after a press release has been published.

In this chapter three classifiers and the methods of evaluation are described. The methodology of an event study is also explained.

4.1 Classification

Mitra and Mitra (2011) describe five classifiers that are applied by Das and Chen (2007). The naive classifier, the vector distance classifier, and the Bayesian classifier were selected for use on the collected data. Luss and d'Aspremont (2009) found that press releases cannot be used to predict the direction of a price change. It is therefore suitable to use two classes, either there is a large movement or there is not. The goal of the classifiers is to assign 1 or 0 to each press release depending on what class it belongs to. There is also the option of combining the results from several classifiers, demanding that all of them or a majority of them come to the same conclusion. As the approach differs between classifiers, this might improve the results.

4.1.1 Pre-processing

Two of the classifiers take word vectors as input. A word vector m is a list of all words in the lexicon and their multiplicities. If there are D words in the lexicon then $m \in \mathbb{R}^D$. The press releases are stored in a bag-of-words format, including many words that do not appear in the lexicon. For all press releases the number of occurrences of each lexical word is counted, and stored in a vector.

The words in the lexicon were selected for appearing more often in one class than the other. Some words may have that quality without being useful for the classification. Conjunctions, as an example, appear in all press releases and are not expected to indicate a certain class. Press releases are of different lengths which can confuse classifiers measuring the distance between two word vectors, for even if they express the same sentiment the word vector for the longest press release is likely to have larger multiplicities. To adjust for this the term frequency-inverse document frequency (TF-IDF) is used for term weighting in all vectors. It is defined for term i in press release j as

$$\text{TF-IDF}(i,j) = \text{TF}(i,j) \times \text{IDF}(i).$$

Where $\text{TF}(i,j)$ is the multiplicity of word i in message j normalized by the total number of words in m_j,

$$\text{TF}(i,j) = \frac{n(m_j, w_i)}{n(m_j)}.$$

$\text{IDF}(i)$ is the logarithm of the number of press releases divided by the document frequency (DF) which is the number in which word i appear,

$$\text{IDF}(i) = \log \frac{N}{\text{DF}(i)},$$

where log is the natural logarithm. The IDF of words that appear in all of the press releases will be zero which makes them irrelevant.

4.1.2 Naive classifier

A naive classifier counts the number of positive and negative words in a message to determine the sentiment. A positive word is worth 1, and a negative word is worth -1. The sentiment score is the total sum of all positive and negative words. If the score is larger than the parameter T, which has to be defined, then the message is classified as 1, else 0.

The words that are to be considered positive or negative are defined in an annotated lexicon. Both the sentiment lexicon of Hu and Liu (2006) and the General inquirer dictionary will be used for this purpose. The naive classifier is often used with three classes, and in those cases a message is classified as -1 when the sentiment score is below a specified level. The classes could correspond to sell, hold or buy for example. As mentioned it is difficult to separate negative and positive press releases and therefore there are only two classes in this case.

The usefulness of a naive classifier is intuitive when analyzing social media content where the text often is opinionated and clear in its sentiment. However, there are reasons to apply it to press releases as well. For one, the average length of the press releases varies between the classes. It is not hard to imagine that longer texts include more positive and negative words. This aspect is not captured by the other classifiers, which use the word vectors, since the term frequency weighting normalizes the vector by the number of words in the press release.

4.1.3 Vector distance classifier

A vector distance classifier takes an unknown press release as input and compares its word vector to already classified press releases. A training set is chosen for this purpose. Each element in the training set comprises the word vector of a press release and its class, $C_j = (m_j, c_j)$. If v is the word vector to be classified then the angle to other vectors is defined by the cosine similarity

$$\cos(\theta_j) = \frac{v \cdot m_j}{|v| \cdot |m_j|},$$

which is computed for all C_j. The vector m_j with the largest similarity, $\max_j \cos(\theta_j)$, is selected and v is classified as c_j. That is, the press release is assumed to belong to the same class as the press release in the training set with the most similar word vector. The results depend on the lexicon and the training set.

4.1.4 Bayesian classifier

The naive Bayesian classifier computes the probability that a message belongs to a certain class. The class for which the probability is highest is then assigned to the message. In the final step Bayes' theorem is applied, hence the name.

A set of word vectors is used for training the classifier, estimating necessary parameters. Let $n(m_j, w_k)$ be the number of times word k in the lexicon appears in message j. Let $n(m_j)$ be the total number of words in message j. Now each press release is assigned to one of two classes, c_1 or c_2. Let $n(c_i, w_k)$ be the number of times that word k appear in class i, obtained through summation over all messages in the class. The probability of selecting w_k from a message in class c_i is then on average

$$p(c_i,w_k) = \frac{\sum_{m_j \in c_i} n(m_j, w_k)}{\sum_{m_j \in c_i} \sum_k n(m_j, w_k)} = \frac{n(c_i,w_k)}{n(c_i)}.$$

The probability can not be allowed to be zero and is therefore advantageously expressed as

$$p(c_i,w_k) = \frac{n(c_i,w_k) + 1}{n(c_i) + D}$$

instead, where D is the number of words in the lexicon. The conditional probability $Pr[m_j|c_i]$ is a measure of how likely message j is given class i. It can be computed as

$$Pr[m_j|c_i] = \binom{n(m_j)}{n(m,1), n(m,2), ..., n(m,D)} \times \prod_{k=1}^{D} p(c_i,w_k)^{n(m_j,w_k)}.$$

The prior probabilities $Pr[c_1]$ and $Pr[c_2]$ can be set to the fraction of press releases in the training set that belong to each class respectively. That is the last required piece before applying Bayes' theorem,

$$Pr[c_i|m_j] = \frac{Pr[m_j|c_i]Pr[c_i]}{Pr[m_j|c_1]Pr[c_1] + Pr[m_j|c_2]Pr[c_2]}.$$

Now normally one would compute $Pr[c_1|m_j]$ and $Pr[c_2|m_j]$ and classify m_j according to the highest probability. The collected data is however imbalanced, with roughly 85% of the press releases not causing an abnormal event. Such skewed priors causes the classifier to always assign class 0. One simple way of alleviating this problem is to manually set the priors to 0.5. This reduces the classifier into looking at which class is more likely to generate the message, $Pr[m_j|c_1]$ or $Pr[m_j|c_2]$.

4.1.5 Evaluation of a classifier

Binary classification is commonly evaluated through calculation of precision and recall. A positive classification in this context corresponds to an abnormal event. The press releases are divided into four categories: True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Using these categories the accuracy, recall, and precision can be defined.

$$\begin{aligned}\text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN},\\ \text{Recall} &= \frac{TP}{TP+FN},\\ \text{Precision} &= \frac{TP}{TP+FP}.\end{aligned}$$

Accuracy is the overall fraction of correct predictions. Recall is the fraction of abnormal events that are classified as such. Precision is arguably the most important measure for this application, it is the fraction of positive classifications that are abnormal events. A low precision can lead to losses if a trader acts on information that turns out to be false.

4.2 Event study methodology

A press release constitutes an event and its impact on the stock price can be quantified in an event analysis. The methodology of such a study is described in Chapter 4 of Campbell, Lo and MacKinlay (1997). The framework there presented has been used among others by Antweiler and Frank (2005) in their study on the impact of news articles from the Wall street journal.

4.2.1 Event definition

The basic idea is to compute the cumulative abnormal return for a time period after an event. It is a way to identify stock price movements that can not be explained by market movements. The period for which abnormal returns are computed is called the event window (L_w). The length of the event window can vary but in this study it ranges from one to ten days. Only days where the market is open count. The window starts on the day following the press release, thus not taking the initial reaction into consideration. Prior to the event window there is an estimation window (L_e). The estimation window is set to 120 days and is used to estimate parameters needed for the calculations.

4.2.2 Cumulative abnormal returns

The cumulative abnormal return is simply the sum of abnormal returns. As the name indicates the abnormal return is the deviation from the normal, expected, return.

$$AR_{i,t} = R_{i,t} - \mathrm{E}[R_{i,t}|X_t],$$

where X_t is the information required by the chosen model for expected returns. There are several ways of defining what the normal return is, one of them is the market model. The market model expresses the stock price for stock i as a function of α_i, β_i, and the market return. In this case X_t is the market return $R_{m,t}$.

$$E[R_{i,t}|R_{m,t}] = \alpha_i + \beta_i R_{m,t} + \epsilon_{i,t}.$$

$\epsilon_{i,t}$, an error term, has mean zero and variance $\sigma^2_{\epsilon_i}$. The parameters of the market model are estimated from data in the estimation window. Note that this window comprises the 120 days preceding the event, and it differs between press releases. The ordinary least squares method for a simple regression model on the data in the estimation window, $t \in L_e$, gives

$$\begin{aligned}\hat{\beta}_i &= \frac{\text{Cov}(R_i,R_m)}{\text{Var}(R_m)}, \\ \hat{\alpha}_i &= \mu_{R_i} - \hat{\beta}_i \mu_{R_m}.\end{aligned}$$

μ_{R_i} and μ_{R_m} are the mean values of the respective returns

$$\begin{aligned}\mu_{R_i} &= \frac{1}{|L_e|} \sum_{t \in L_e} R_{i,t}, \\ \mu_{R_m} &= \frac{1}{|L_e|} \sum_{t \in L_e} R_{m,t}.\end{aligned}$$

There is a third parameter in the market model, namely the variance of the error $\sigma^2_{\epsilon_i}$. It can be computed as

$$\hat{\sigma}^2_{\epsilon_i} = \sigma^2_{R_i} - \hat{\beta}_i \text{Cov}(R_i,R_m).$$

The abnormal return for a day in the event window, $t \in L_w$, is

$$AR_{i,t} = R_{i,t} - (\hat{\alpha}_i + \hat{\beta}_i R_{m,t})$$

and thus the cumulative abnormal return is

$$\text{CAR}_{i,t} = \sum_{t \in L_w} AR_{i,t}.$$

It should be noted that all returns are log returns, defined as

$$R_{i,t} = \log(P_{i,t} + D_{i,t}) - \log(P_{i,t-1})$$

where $P_{i,t}$ is the price of asset i at day t, and $D_{i,t}$ is a possible dividend.

4.2.3 The test statistic

The abnormal return has an error, which is propagated when summation is carried out over time. It is again affected when averaged over a class of assets. To find out if the cumulative abnormal return is statistically significant a test statistic is formulated. The null hypothesis $\mathcal{H}_0$ is that the cumulative abnormal return is zero, and the aim is to reject it with a 0.01 significance level.

The variance $v^2_{i,t}$ of an abnormal return is

$$v^2_{i,t} = \sigma^2_{\epsilon_i}(1 + \frac{1}{|L_e|}(1 + \frac{(R_{m,t} - \mu_{R_m})^2}{\sigma^2_{m,t}})).$$

The first term is the disturbance term which represents the error in the market model. It is always present. The second term is the sampling error that occurs when estimating $\hat{\beta}_i$ and $\hat{\alpha}_i$, it decreases as the estimation window increases. For a large enough estimation window it therefore holds that $v^2_{i,t} \approx \sigma^2_{\epsilon_i}$.

The standardized cumulative abnormal return (SCAR) for the event window L_w with $|L_w|$ days starting at and including t can be calculated as

$$\text{SCAR}_{i,t} = \frac{\text{CAR}_{i,t}}{\hat{\sigma}_{\epsilon_i}\sqrt{|L_w|}}$$

since the abnormal return has the same variance each day. The average SCAR, ASCAR, over N stocks is simply

$$\text{ASCAR}_t = \frac{1}{N}\sum_{i=1}^{N} \text{SCAR}_{i,t}.$$

The variance of the ASCAR is $\frac{L_e-2}{N(L_e-4)}$, and a test statistic can be defined as

$$J = \sqrt{\frac{N(L_e-4)}{L_e-2}}\mathrm{ASCAR}_{M,t} \sim \mathcal{N}(0,1).$$

The squared fraction in front of the ASCAR is always positive and larger than one and removing it will not affect the validity of a rejection. When the CAR is computed, so is the ASCAR, and it is determined whether the event did or did not have an impact on the stock price.

4.3 Backtesting

Finding a pattern can be interesting in its own right, but the aim in looking for patterns is often to identify potential for profit. That is something else entirely, as slippage and commission costs enter into the equation. To investigate how well findings translate into trading strategies backtesting is used. Backtesting is the retroactive execution of a strategy on historical data. It measures the performance for a passed period of time. The platform used for these experiments is Quantopian, which enables algorithmic trading on minute time resolution for all American stocks.

4.3.1 Evaluation of a backtest

To evaluate how well a trading strategy performs the portfolio's α, β, Sharpe ratio and volatility are computed. The Sharpe ratio is defined as

$$S_i = \sqrt{T}\frac{\mu_{R_i}}{\sigma_{R_i}}$$

where T is the number of days over which it is calculated. It is the average return adjusted by the standard deviation. How to find α and β has already been shown. Note that in this case the backtesting engine returns those values automatically. Commission is set so as to be as realistic as possible. It is \$0.014 per share but at least \$1.4 in total per transaction.

4.3.2 Position sizing technique

The returns, and therefore the metrics, are affected not only by the times at which shares are bought and sold. The size of a trade has an impact as well. There are many different position sizing techniques. How the portfolio's assets are distributed over time is of great importance. If the strategy is executed over a period of time for which only a few press releases exist in the database the portfolio will mostly hold cash. Holding cash decouples the portfolio returns from the market and does not reflect how the strategy would do in a live trading situation. Therefore the results become less reliable the further back in time the strategy is tested. The initial bankroll is \$100000 for all portfolios.

For one strategy a fixed dollar amount is used as position sizing technique. When an opportunity for investing has been found the stock in question is bought or sold for \$2000, assuming that such a position can be afforded. The other strategy takes a fixed risk per day. It holds assets for n days, and each day $\frac{1}{n}$ of the portfolio value is invested in stocks that are considered promising.

4.3.3 Time frame

It is important to observe that the data only allow the mimicking of a live trading strategy for a short period. Quite soon data are missing, as thirteen press releases do not stretch over the entire year for most stocks. Ideally a strategy should be tested over many years. The difference between how a strategy performs for a limited amount of months and in general is potentially huge. Strategies that perform in bull markets may turn out to be costly in bear markets and the opposite. A backtest with the time frame limitation that exists here can merely be seen as an indication.

4.3.4 Trading on predicted large movements

The classifiers predict large price movements, but do not reveal the direction. A non-directional strategy that profits on any large movement is therefore ideal. An example of a position with this property is a long straddle: Buy a call option and a put option on the same stock with the same strike price and the same expiration date. If the change in the stock price is large enough the profit from one of the options will more than cover the loss of the other. To simulate a strategy that takes such positions would require intraday option prices, which are not available. Therefore an alternative strategy is devised.

A portfolio is constructed and its assets allocated as described here. At the start of each day a list of press releases that were published that day or after the stock market closed the preceding day is imported. Only press releases that were classified as causing large movements are imported. A delay is set to 10 minutes. If a press release was published when the market was closed the time of that press release is set to the market open plus the delay, in this case 9:40 a.m. ET. For all other press releases the time of publishment plus the delay is the time at which the press release is processed by the algorithm. This delay can be considered the time period during which, in a real time application, the press release is fetched and classified. When the backtesting reaches the time at which a press release is to be processed the return since the previous day of the stock to which the press release belongs determines the direction of the trade. If the stock has a positive return, $2000 is invested into the stock. If the stock has a negative return, stocks are sold for $2000, a short position. Five minutes before the market closes all positions are closed. The holding period of each stock therefore depends on what time during the day the press release is published. Summarized, if a press release is published and if it is classified as causing a large price movement the stock is bought or sold depending on the direction of the return so far and held until the end of the day. This is repeated every day.

4.3.5 Trading on post-event patterns

The press releases are divided into categories for the post-event analysis. If a pattern is found, for example that positive press releases have a drift upwards in the following days, the press releases in that category are supplied to the algorithm on the day that they are published. Ten minutes before the market closes the stocks are bought, and kept for a specified number of days. If $i(t)$ stocks are bought on day t and kept for n days, then $\frac{V(t)}{i(t)n}$ are invested into each stock. $V(t)$ is the total portfolio value at day t. After n days the positions are closed. Summarized, a fraction of the portfolio is invested in press releases that are believed to give abnormal returns within the forthcoming days. The positions are kept for the specified number of days and then closed.

5 Results

The results from the classification of press releases and the results from the event study analysis are presented in this chapter. The results from the trading strategies are shown in the next chapter. First the question of predicting large price movements is dealt with, and all classifiers are applied and evaluated individually and in combination.

5.1 Prediction of major price movements

The naive classifier was applied to all press releases. The basis for this classifier is that the sentiment score, the sum of positive and negative words, will differ between classes. To see if there is such a phenomenon to exploit the average sentiment score for each class was computed using the lexica described in Section 3.5.2. The results can be seen in Table 5.1.

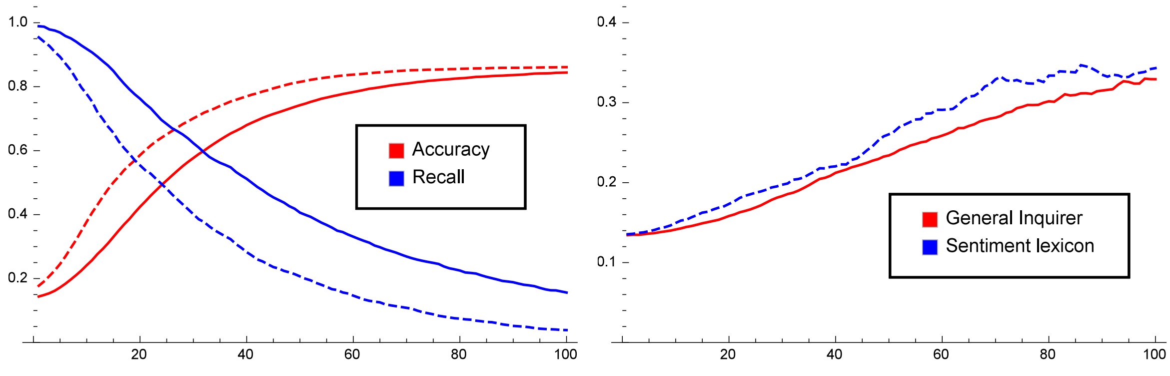
Table 5.1: Average sentiment scores, computed for three categories of press releases using two different lexica.

	General Inquirer	Sentiment lexicon
Negative	51.2	33.0
Neutral	37.2	23.0
Positive	61.3	34.5

The average sentiment score is significantly lower for neutral press releases. The difference between positive and negative press releases is however small. In other contexts such as when analyzing social media content it is expected that the sentiment score is negative for negative news. That is not the case for press releases. This is in line with the belief that it is impossible to determine the direction of a large movement from looking at the press release that caused it. Therefore the classification is executed with two classes, the neutral and the union of the positive and negative.

5.1.1 Naive classifier

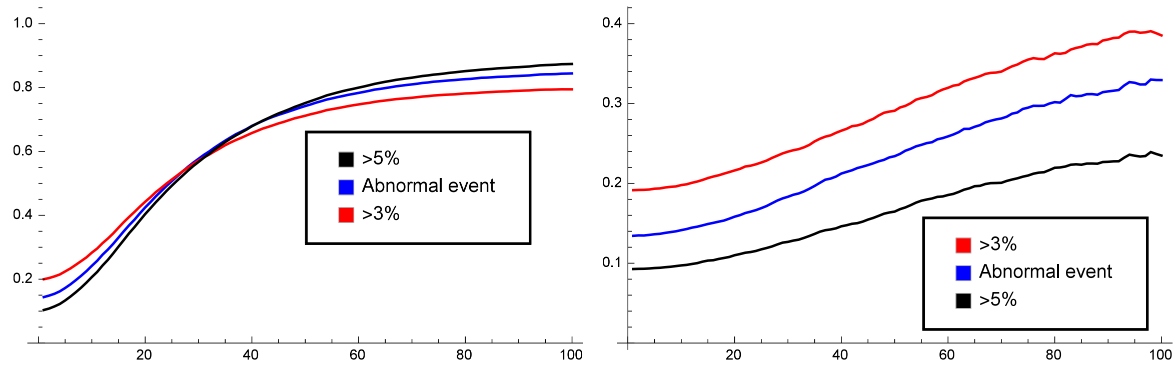
The choice of annotated dictionary and the value of parameter T, explained in Section 4.1.2, influence the outcome of the naive classifier. In figure 5.1 the accuracy, recall, and precision is shown as a function of T for both the sentiment lexicon and the General Inquirer lexicon.



(a) *Accuracy and recall as a function of T.* (b) *Precision as a function of T.*

Figure 5.1: *Accuracy, recall, and precision of the naive classifier. Dashed lines represent the sentiment lexicon, and solid lines the General Inquirer dictionary.*

The General Inquirer lexicon has a much better recall but worse precision and accuracy. Which one to prefer depends on the application. For all further computations the General Inquirer lexicon will be used, as it is the most comprehensive. Another choice that has impact on the results is the initial classification. To investigate its importance the accuracy and precision was calculated for three different alternatives, described in the data collection section. The results can be seen in Figure 5.2



(a) *Accuracy as a function of T.* (b) *Precision as a function of T.*

Figure 5.2: *Accuracy and precision of the naive classifier. The lines show attempts at different classifications.*

It is important to note that there is a difference in the number of press releases that are classified as abnormal in the different initial classifications. Almost all press releases have a sentiment score larger than 0 and thus the values at $T = 0$ correspond to the fraction of press releases that are considered abnormal. It is the probability to be right if one always classify a press release as abnormal. If the precision was a straight line at its initial value it would mean that the classifier's ability to forecast would be equal to guessing. It is not only the value itself that is of interest but also its increase. Accuracy can be misleading in a similar way for if only, as an example, 5% of the press releases have a positive classification a classifier could achieve 95% accuracy with a 0% recall by never classifying a press release as positive.

5.1.2 Vector distance classifier

To run the vector distance classifier each press release is assigned a training set of size N, selected randomly from among the other press releases. All press releases are classified and the accuracy, recall, and precision are measured. The computations are done for two different initial classifications. The results are shown in Table 5.2.

Table 5.2: Vector distance classifier.

	Abnormal events			>3%		
N	Accuracy	Recall	Precision	Accuracy	Recall	Precision
1000	0.810	0.244	0.272	0.721	0.360	0.303
10000	0.823	0.335	0.333	0.773	0.407	0.404

The precision of the vector distance classifier is similar to that of the naive classifier when $T = 80$. However, the recall is much better. That could potentially be a big deal since the total return is the profit per trade multiplied by the number of trades. Recall is a measure of how many opportunities that are identified.

There is a clear improvement in all three measures when the number of training vectors is increased. The drawback is that the computation requires more time as the time complexity for classifying x press releases is $\mathcal{O}(Nx)$ or $\mathcal{O}(x^2)$. Each word vector that is to be classified has to be compared to N other vectors, at most $x - 1$.

5.1.3 Bayesian classifier

The Bayesian classifier was applied to the press releases in the same fashion as the vector distance classifier. The results are shown in Table 5.3. The most striking difference is that the Bayesian classifier requires a larger training set. The classifier is not useful for $N = 1000$ as the precision and recall are close to zero. However, the Bayesian classifier runs in constant time once its parameters have been approximated. Therefore this does not pose a problem. The Bayesian classifier is the best of the three as it boasts a higher precision than the others while retaining a high recall.

Table 5.3: Bayesian classifier.

	Abnormal events			>3%		
N	Accuracy	Recall	Precision	Accuracy	Recall	Precision
1000	0.864	0.002	0.095	0.807	0.005	0.264
10000	0.841	0.356	0.396	0.792	0.380	0.450

5.1.4 All results

The classifiers were not only evaluated individually, combinations were also assessed. That includes *any*, which classifies a message as an abnormal event if any of the classifiers do it. *Majority*, which requires two out of three classifiers to agree and *all*, which only classifies a message as an abnormal event if all classifiers were positive. The results are shown in Table 5.4.

Table 5.4: Classifier results, individually and combinations.

	Abnormal events			>3%		
$N = 10000$	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Naive	0.828	0.223	0.301	0.783	0.189	0.374
Vector distance	0.824	0.345	0.341	0.768	0.393	0.397
Bayesian	0.838	0.358	0.383	0.796	0.386	0.464
Any	0.767	0.558	0.299	0.724	0.588	0.365
Majority	0.856	0.276	0.432	0.810	0.289	0.512
All	0.868	0.091	0.519	0.814	0.091	0.604

There is a clear trend, as the recall goes down the precision goes up. The more picky one is, the more correct the predictions are. Combining the results from several classifiers greatly improves the precision. The classifiers are somewhat better at predicting large movements defined in absolute terms as $> 3\%$, but they show a clear ability of being useful in the prediction of both.

5.1.5 Impact of market capitalization on classification

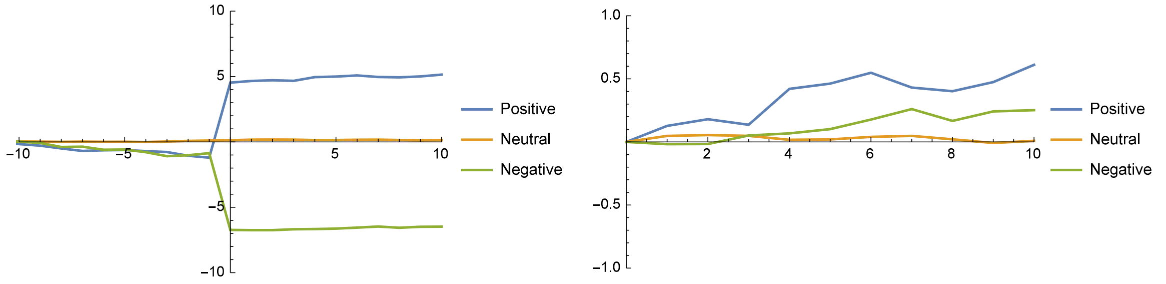
The stocks were sorted by their market capitalizations and partitioned into three equally sized groups. The groups are called *small*, *middle*, and *large*. The best precision is achieved when all classifiers agree, and this combination of classifiers is evaluated for all three groups as a way of determining the impact of market capitalization on the ability to classify. The results are shown in Table 5.5.

Table 5.5: Classifier results depending on market capitalization.

	Abnormal events			>3%		
$N = 10000$	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Small	0.850	0.082	0.510	0.720	0.091	0.672
Middle	0.872	0.100	0.491	0.839	0.101	0.549
Large	0.875	0.070	0.460	0.882	0.095	0.514

5.2 Event analysis of large stock price changes

An event study analysis can be run on any category of press releases. Dividing them into the three intuitive categories negative, neutral and positive is a good starting point. Positive and negative press releases being defined by if they caused a large movement in that direction, as determined by the initial classification of abnormal events. In Figure 5.3 the cumulative abnormal returns for these three categories are shown over time. In Figure 5.3a the summation starts ten days prior to the event and ends ten days after the event. The neutral category shows no abnormal returns at all, which is to be expected. Both the positive and negative categories show negative abnormal returns leading up to the event. When the press releases are published there is a strong reaction. The cumulative abnormal returns for the days after the event are shown in Figure 5.3b where the summation starts at the first day after the event.



(a) *CAR, summation starting ten days prior to the event.* (b) *CAR, summation starting the day after the event.*

Figure 5.3: *Cumulative abnormal returns before and after the event for three classes of press releases.*

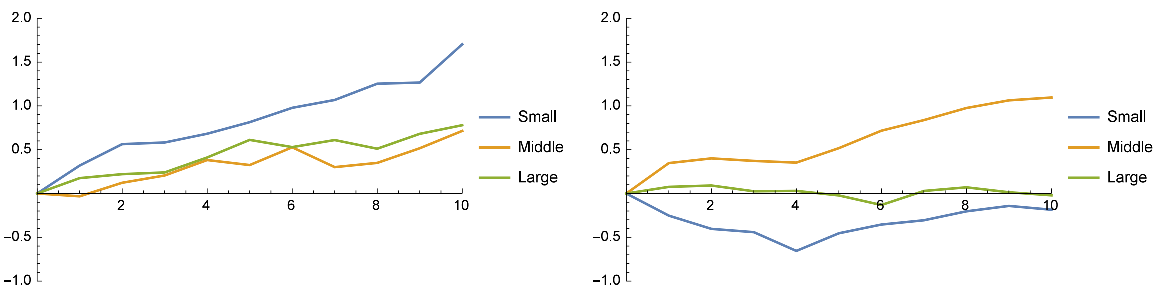
Table 5.6: Abnormal returns for positive and negative press releases.

	Positive			Negative		
Day	AR (%)	CAR (%)	P-value	AR (%)	CAR (%)	P-value
0	5.739	5.739	0.001	-5.871	-5.871	0.001
1	0.127	5.866	0.012	-0.018	-5.889	0.010
2	0.053	5.919	0.031	0.001	-5.888	0.029
3	-0.044	5.875	0.052	0.068	-5.820	0.050
4	0.284	6.160	0.066	0.017	-5.804	0.070
5	0.042	6.202	0.085	0.034	-5.770	0.090
6	0.086	6.288	0.101	0.077	-5.694	0.109
7	-0.117	6.170	0.122	0.083	-5.611	0.129
8	-0.029	6.141	0.136	-0.094	-5.705	0.140
9	0.072	6.214	0.148	0.075	-5.630	0.154
10	0.137	6.350	0.157	0.010	-5.620	0.165

5.2.1 Categorization by market capitalization

The companies are divided into three groups called small, middle, and large by their market capitalizations. The positive and negative press releases from each group are then selected, and the average abnormal returns computed. The results are shown in Figure 5.4, where results from the positive press releases are in Figure 5.4a and results from the negative press releases are in Figure 5.4b.

Positive press releases have the largest post-event cumulative abnormal returns for small companies. Negative press releases have the largest post-event cumulative abnormal returns for middle sized companies and negative cumulative abnormal returns for small companies. The results for neutral press releases are not shown, but the cumulative abnormal returns are approximately zero for all market capitalizations.



(a) *CARs for positive press releases.* (b) *CARs for negative press releases.*

Figure 5.4: *Cumulative abnormal return, post-event patterns for companies of different sizes.*

6 Trading strategies

Both strategies were executed for the period 4 October, 2014 to 4 February, 2015. The initial bankroll was \$100000 per portfolio. The strategy for large movements bets \$2000 per trade. The index, NYSE Composite, went up 1.2% during the testing period. SPY, en exchange-traded fund that reflects the movements of S&P 500, went up 4%. The platform used is Quantopian.

6.1 Trading on movements predicted to be large

Press releases were selected according to their classification and used as input to the algorithm. Two groups of press releases were added, random and cheating. The random press releases were added as a benchmark for how well the algorithm performs if given any press releases, as opposed to those selected by the classifiers. The cheating group is selected by the initial classification which of course is not in reality available when the trade is made. It is a benchmark for how well the algorithm would have performed if the classifications were perfect. The initial classification used in the classification for this strategy is the $> 3\%$. The results are shown in Table 6.1.

Table 6.1: Trading on movements predicted to reach $> 3\%$.

Classifier	# trades	Returns (%)	α	β	Sharpe	Drawdown (%)
Random	1244	-3.8	-0.13	0.02	-4.30	4.9
Cheating	1400	31.6	0.93	0.04	13.87	2.2
Naive	823	-3.4	-0.12	0.01	-3.67	3.7
Vector distance	1460	-1.0	-0.05	-0.01	-1.01	3.4
Bayesian	1170	-1.9	-0.07	0.0	-1.56	3
Majority	847	-1.5	-0.06	0.0	-1.43	2.3
All	245	0.7	0.0	0.0	0.08	1.4

The trading strategy trades in both directions, therefore the total result is the sum of the trades where the algorithm went short and where it went long. To see if one type of trade is more profitable than the other the trades were divided into long and short for two classifiers and the results are shown in Table 6.2.

Table 6.2: Results when trades are divided into long and short.

Classifier	# trades	Returns (%)	α	β	Sharpe	Drawdown (%)
Vector distance, long	777	0.0	-0.03	0.09	-0.40	2.2
Vector distance, short	684	-1.0	-0.04	-0.09	-1.15	3.1
Majority, long	438	0.2	-0.02	0.07	-0.33	1.3
Majority, short	409	-1.8	-0.06	-0.06	-1.89	2.6

Another categorization of press releases is that of dividing the companies into three groups based on the market capitalizations. Table 6.3 shows how the strategy performs for different company sizes using the vector distance classifier.

Table 6.3: Trading with vector distance classification on stocks divided into three groups by market capitalization.

Classifier	# trades	Returns (%)	α	β	Sharpe	Drawdown (%)
Small	414	0.0	-0.01	-0.05	-0.53	2.1
Middle	522	-1.2	-0.06	0.01	-2.55	2.3
Large	520	-0.1	-0.03	0.03	-1.32	0.9

6.2 Trading on post-event patterns

Press releases are sorted into positive, neutral, and negative categories based on their initial classification. The initial classification here used is that of abnormal events. Each day 10% of the portfolio is invested in the stocks for which press releases have been published the day before. After ten days the positions are closed. In Table 6.4 the results from executing the trading strategy for different categories are shown. Since there are so many neutral press releases a sample was randomly selected.

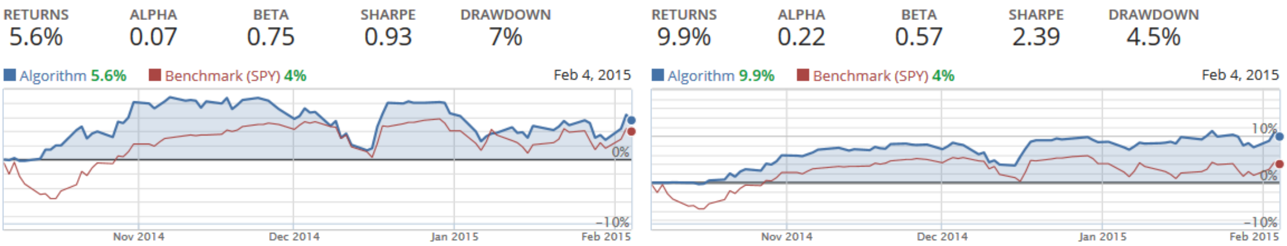
Table 6.4: Buying the stock depending on the price reaction on the press release.

Classifier	# trades	Returns (%)	α	β	Sharpe	Drawdown (%)
Negative	686	0.3	-0.10	0.85	-0.04	9.8
Neutral	967	2.5	-0.03	0.85	0.35	8.69
Positive	771	5.6	0.07	0.75	0.93	7.0

The division into three groups based on market capitalization was used for both negative and positive press releases and the results are shown in Table 6.5. In Figure 6.1 two simulations are shown as they appear in Quantopian.

Table 6.5: Dividing the input of positive (pos) and negative (neg) press releases by market capitalization.

Classifier	# trades	Returns (%)	α	β	Sharpe	Drawdown (%)
Small (pos)	221	-1.3	-0.13	0.70	-0.30	10.1
Middle (pos)	267	7.9	0.15	0.7	1.56	7.4
Large (pos)	282	9.9	0.22	0.57	2.39	4.5
Small (neg)	229	-12.7	-0.48	0.76	-1.86	14.29
Middle (neg)	225	11.1	0.24	0.77	2.03	7.4
Large (neg)	231	2.5	-0.01	0.68	0.44	6.3



(a) *All positive press releases.* (b) *Positive press releases from companies considered large.*

Figure 6.1: *Results from backtests as shown in Quantopian. Reproduced with permission.*

7 Discussion and conclusion

With the results retrieved one can discuss to what extent and in which way the understanding of press releases as a stock price driver has been improved. In the data collection phase it was discovered that press releases that cause large price movements are on average about twice as long as press releases that do not. An interesting observation as it indicates that classes can be distinguished, there are textual attributes which are more pronounced in the press releases of interest. It was also found that a press release frequently is a possible explanation for a large price movement. 13% of the press releases are published on such days, and for 24% of those days there is a press release published. Press releases appear to be important events on which automated analysis can be applied.

7.1 Discussion

The average sentiment score is approximately the same for both positive and negative press releases. This would be unexpected for other sorts of messages, where negative sentiments are reflected by a negative sentiment score. This made it impossible for the naive classifier to tell positive and negative press releases apart. One possible reason is that announcements with similar format can be either. Consider for example an earnings announcement, the text might not differ much whether the analysts' estimates are beaten or not. An interesting option could be to divide press releases into topics, and then look at how press releases from each topic affect the stock price.

The Bayesian classifier was the most successful. However, combining classifiers improved the precision at the cost of recall. Having classifiers that capture different aspects of similarity is therefore worthwhile in applications where precision is important. The results in general were uplifting in the sense that all classifiers demonstrated an ability to predict large movements. Whether the capacity of prediction is enough to be useful depends on the application. For the trading strategy the classifiers were unable to capture the large potential profit received when cheating.

The event study analysis showed statistical significance at the day of the event, rejecting the null hypothesis that press releases do not affect the stock price. An expected confirmation. There was no statistical significance on days preceding or succeeding the event. The former case is interesting, as it contradicts the idea of information leaking to the market during the days before a press release. Not only is the movement prior to positive press releases statistically insignificant, it is also negative.

The post-event patterns were positive in most cases. The strategy attempting to exploit the abnormal returns therefore took long positions. Since the market rose during the testing period a positive return was expected, but the size of the returns beat the market with a large margin in some cases. Some results are hard to explain, such as why negative press releases have positive subsequent abnormal returns for middle sized companies. There is no obvious answer, but in combination with the results not being statistically significant one has to be careful in the interpretation. A much longer period of testing would have to precede an actual implementation of the strategies.

7.2 Conclusion

Press releases were found to frequently have a significant impact on stock prices. The significant movement in the price occurs on the day the press release is published and not thereafter. It is possible to predict major price movements to a certain extent. The results from the trading strategy designed to profit from predictions indicate that the classifiers used are not good enough for this specific application. The sentiment score of a press release does not reveal whether the content is positive or negative. There are no signs of information leaking to the market or insider trading.

References

[1] W. Antweiler and M. Frank. Do U.S. stock markets typically overreact to corporate news stories? *Working paper, University of British Columbia* (2006).

[2] B. M. Barber and T. Odean. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* **21**.2 (2008), 785–818.

[3] J. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[4] W. Chan. Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics* **70**.2 (2003), 233–260.

[5] S. Das and M. Chen. Yahoo for Amazon! Sentiment extraction from small talk on the web. *Management Science* **53** (2007), 1375–1388.

[6] J. Engelberg, A. V. Reed, and M. Ringgenberg. How Are Shorts Informed? Short Sellers, News, and Information Processing. *Kenan-Flagler Business School, University of North Carolina* (2010).

[7] E. F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance* **25**.2 (1970).

[8] E. F. Fama. The Behavior of Stock-Market Prices. *The Journal of Business* **38**.1 (1965), 34–105.

[9] M. Hu and B. Liu. Mining and Summarizing Customer Reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2004).

[10] D. Leinweber and J. Sisk. "Relating news analytics to stock returns". *The Handbook of News Analytics in Finance*. Ed. by G. Mitra and L. Mitra. Wiley, 2011, pp. 149–172.

[11] R. Luss and A. d'Aspremont. Predicting Abnormal Returns From News Using Text Classification. *Working paper, ORFE, Princeton, NJ.* ().

[12] G. Mitra and L. Mitra, eds. *The Handbook of News Analytics in Finance*. John Wiley & Sons, 2011.

[13] M. Mittermayer and G. Knolmayer. Text mining system for market response to news: A survey. *Working paper* ().

[14] V. Niederhoffer. The analysis of World Events and Stock Prices. *The Journal of Business* **44**.2 (1971), 193–219.

[15] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* (2002).

[16] D. Peramunetilleke and R. K. Wong. Currency Exchange Rate Forecasting from News Headlines. *Australian Computer Science Communications* **24**.2 (2002), 131–139.

[17] P. Ryan and R. Taffler. What firm-specific news releases drive economically significant stock returns and trading volumes? *Working paper, Cranfield University* ().

[18] P. Tetlock, M. Saar-Tsechansky, and S. Macskassy. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* **63** (2008), 1437–1467.